



同济大学中德工程学院

CDHAW OF TONGJI UNIVERSITY

# DSGE 模型笔记

朱彦元

Dr. Yanyuan ZHU  
[yyz@tongji.edu.cn](mailto:yyz@tongji.edu.cn)

2018 年 5 月 2 日



# 目录

第一章 前言	15
第一部分 模型构建	17
第二章 基准 New Keynesian 模型	19
2.1 家庭部门	19
2.2 企业部门	20
2.2.1 最终产品生产部门	20
2.2.2 中间产品生产部门	20
2.2.3 最终产品定价: Calvo assumption	22
2.3 外生技术冲击	22
2.4 总量均衡	23
2.4.1 家庭部门消费的决定	23
2.4.2 price dispersion	23
2.5 (非随机的) 稳定状态	24
2.6 flexible price equilibrium 和 output gap	25
2.6.1 flexible price equilibrium	25
2.6.2 output gap	26
2.7 Taylor 法则	26
2.8 完整的均衡条件	27
2.9 对数线性化	28
2.9.1 对数线性化计算	28
2.9.2 线性模型	31
第三章 Medium-Sized DSGE 模型	33
3.1 产品的生产部门	33
3.1.1 最终产品生产部门	33
3.1.2 中间产品生产部门	33
3.1.3 中间产品生产者的边际成本	34
3.1.4 中间产品生产者的定价	36
3.1.5 price dispersion index	38
3.1.6 投资的调节成本	39
3.1.7 总产出的分配	39
3.2 家庭部门	40

3.2.1 劳动力投入：同质化假定还是异质化假定？ . . . . .	40
3.2.2 劳动力承包商 . . . . .	41
3.2.3 家庭行为 . . . . .	42
3.2.4 劳工联盟的工资策略 . . . . .	42
3.2.5 wage dispersion index . . . . .	45
3.2.6 家庭预算约束条件 . . . . .	46
3.2.7 资本积累 . . . . .	46
3.2.8 家庭部门最大化问题 . . . . .	47
3.3 Price Philips Curve . . . . .	50
3.3.1 price dispersion index 的线性近似 . . . . .	50
3.3.2 两个辅助变量的线性近似 . . . . .	50
3.3.3 reset price inflation 的线性近似 . . . . .	51
3.3.4 Price Philips Curve . . . . .	51
3.4 Wage Philips Curve . . . . .	51
3.4.1 稳定状态的描述 . . . . .	52
3.4.2 wage inflation . . . . .	52
3.4.3 辅助变量 . . . . .	52
3.4.4 劳动力供应 . . . . .	53
3.4.5 边际劳动成本 . . . . .	53
3.4.6 wage price inflation . . . . .	53
3.4.7 其他辅助变量 . . . . .	54
3.5 比较两条 philips curve . . . . .	56
<b>Appendices</b>	<b>59</b>
附录 3.A scaled variables . . . . .	59
3.A.1 几个没写完的说明 . . . . .	60
附录 3.B Frisch elasticity of labor supply . . . . .	60
3.B.1 定义 . . . . .	60
3.B.2 举例 . . . . .	61
附录 3.C 调节成本的常见设定形式及比较 . . . . .	61
3.C.1 家庭部门优化条件 . . . . .	62
3.C.2 比较线性和非线性调节成本 . . . . .	63
3.C.3 比较投资调节成本和资本调节成本 . . . . .	63
<b>第四章 NK-DSGE 模型</b>	<b>65</b>
4.1 Intrdution . . . . .	65
4.1.1 模型中的六个生产部门 . . . . .	65
4.1.2 几点说明 . . . . .	65
4.2 解析模型 . . . . .	66
4.2.1 劳动承包部门 . . . . .	66
4.2.2 家庭部门 . . . . .	67
4.2.3 最终产品生产部门 . . . . .	73
4.2.4 中间产品生产部门 . . . . .	74
4.2.5 政府部门 . . . . .	78

4.2.6 中央银行	78
4.2.7 总量层面的均衡	79
4.2.8 外生过程	82
4.2.9 均衡条件的完整解集	82
4.2.10 非随机的稳定状态	84
<b>第五章 DSGE with banking</b>	<b>91</b>
5.1 Introduction	91
5.2 模型设定	91
5.2.1 家庭及生产部门	91
5.2.2 银行部门	92
5.2.3 政府(中央银行)部门	93
5.2.4 技术冲击	93
5.2.5 实际消费	93
5.3 一阶条件	94
5.4 均衡条件	95
5.5 利率	96
5.5.1 资产市场基准利率	96
5.5.2 抵押品收益率	96
5.5.3 银行间拆借利率(无抵押)	97
5.5.4 贷款利率(有抵押)	97
5.5.5 存款利率	97
5.5.6 外部融资溢价	98
5.5.7 抵押品收益率与外部融资溢价的关联	98
5.6 稳定状态	99
5.6.1 核心稳态系统	99
5.6.2 稳定状态利率	100
<b>第六章 Cash in advance</b>	<b>101</b>
6.1 家庭部门	101
6.2 企业部门	102
6.2.1 生产规模	102
6.2.2 利润的使用	102
6.3 银行部门	103
6.4 一阶条件	103
6.4.1 家庭部门	103
6.4.2 企业部门	103
6.4.3 银行部门	104
6.4.4 一阶条件的整理	104
6.5 市场均衡及部门最优行为	104
6.5.1 4个市场均衡条件	104
6.5.2 部门最优行为	105
6.6 外生冲击	106
6.7 均衡方程组	106

6.8 去趋势的均衡状态 . . . . .	107
<b>第二部分 模型求解</b>	<b>111</b>
<b>第七章 DSGE 模型求解方法简论</b>	<b>113</b>
7.1 简论的简论 . . . . .	113
7.2 pros and cons . . . . .	115
7.2.1 cons . . . . .	115
7.2.2 pros . . . . .	115
7.3 工作流程 . . . . .	116
7.4 模型设定 . . . . .	117
7.5 求解方法 . . . . .	118
7.5.1 近似线性展开 . . . . .	118
7.6 求解线性随机差分方程系统 . . . . .	120
<b>第八章 动态规划</b>	<b>121</b>
8.1 包络定理 . . . . .	121
8.2 例: 吃蛋糕 (直接求解法) . . . . .	122
8.3 吃蛋糕 (动态规划) . . . . .	123
8.3.1 有限时间段的动态规划问题 . . . . .	123
8.3.2 无限时间段的动态规划问题 . . . . .	125
8.4 随机问题 . . . . .	127
8.5 包络定理和欧拉等式的关系 . . . . .	127
<b>第九章 线性二次最优控制</b>	<b>131</b>
9.1 目标方程和状态转移方程 . . . . .	131
9.2 价值方程 . . . . .	131
9.2.1 一阶条件 . . . . .	132
9.3 政策方程 . . . . .	133
9.3.1 政策方程满足确定性等价条件 . . . . .	133
9.4 线性政策方程和二次价值方程 . . . . .	133
9.4.1 价值方程满足确定性等价条件 . . . . .	134
9.4.2 代数矩阵 Riccati 方程 . . . . .	135
9.5 数值方法 . . . . .	135
9.6 范例 . . . . .	135
9.6.1 最优线性二次控制问题 . . . . .	136
9.6.2 Matlab 程序 . . . . .	136
9.6.3 Matlab 程序输出 . . . . .	138
<b>第十章 理性期望模型</b>	<b>141</b>
10.1 简介 . . . . .	141
10.2 数值算法 . . . . .	142
10.3 一个简单的随机增长模型 . . . . .	143
10.3.1 稳定状态 . . . . .	143

10.3.2 对数线性化 . . . . .	143
10.3.3 状态——空间表现形式 . . . . .	145
10.4 特征值——特征方程分解法 . . . . .	145
10.4.1 特征值——特征方程分解法 . . . . .	145
10.4.2 解的存在性以及唯一性 . . . . .	147
10.4.3 应用 Blanchard-Kahn 算法实例 . . . . .	148
10.5 未定系数法 . . . . .	149
10.5.1 模型范例 . . . . .	150
10.5.2 未定系数法求解 . . . . .	150
10.5.3 求解系数矩阵 A, B . . . . .	151
10.5.4 存在性及唯一性的探讨 . . . . .	152
10.6 期望误差法 . . . . .	152
10.6.1 模型范例 . . . . .	153
10.6.2 期望误差法求解 . . . . .	153
10.6.3 存在性及唯一性的探讨 . . . . .	155
10.7 参数化期望法 . . . . .	155
10.7.1 参数化期望法示例 . . . . .	156
10.8 Schur 分解法 . . . . .	157
10.8.1 Schur 分解的定义 . . . . .	158
10.8.2 爆炸分块矩阵的解 . . . . .	158
10.8.3 稳定分块矩阵的解 . . . . .	159
10.8.4 原经济系统的解 . . . . .	159
<b>第十一章 扰动法</b>	<b>161</b>
11.1 扰动法简介 . . . . .	161
11.2 分析框架 . . . . .	161
11.3 求解方法 . . . . .	162
11.3.1 状态——空间形式 . . . . .	162
11.3.2 稳定状态 . . . . .	163
11.3.3 外生随机过程 . . . . .	165
11.3.4 设想解 . . . . .	166
11.3.5 一阶扰动 . . . . .	167
11.3.6 二阶扰动 . . . . .	172
11.3.7 更高阶扰动 . . . . .	174
11.3.8 例 . . . . .	175
11.4 剪枝算法 . . . . .	180
11.5 变量的变换 . . . . .	180
11.5.1 例 1 . . . . .	181
11.5.2 例 2 . . . . .	182
11.5.3 最优变量变换：参数的选取 . . . . .	183
11.5.4 对数线性和对数正态线性化 . . . . .	183
11.6 价值方程迭代中的扰动法 . . . . .	185

<b>Appendices</b>	<b>189</b>
附录 11.A Schur 补和对称正(半)定矩阵 . . . . .	189
11.A.1 Schur 补 . . . . .	189
11.A.2 利用 Schur 分解法判断可逆正定矩阵 . . . . .	191
11.A.3 利用 Schur 分解法判断可逆正半定矩阵 . . . . .	192
附录 11.B 伪逆矩阵 . . . . .	195
11.B.1 用 SVD 法做伪逆矩阵分解 . . . . .	195
11.B.2 用 SVD 法做 M 的矩阵分解 . . . . .	197
<b>第十二章 映射法</b>	<b>199</b>
12.1 简介 . . . . .	199
12.1.1 例 . . . . .	200
12.2 全局基的选取——单维基光谱法 . . . . .	202
12.2.1 单项式 . . . . .	202
12.2.2 三角序列 . . . . .	203
12.2.3 雅各比多项式 . . . . .	203
12.2.4 切比雪夫多项式 . . . . .	203
12.3 局部基的选取 (1) . . . . .	207
12.3.1 离散状态变量 . . . . .	208
12.3.2 张量与完全多项式 . . . . .	209
12.4 局部基的选取 (2): 有限元法 . . . . .	211
12.5 目标方程的选取 . . . . .	214
12.5.1 最小方差 . . . . .	214
12.5.2 子域 . . . . .	215
12.5.3 配点 . . . . .	215
12.5.4 伽辽金法 . . . . .	215
12.5.5 两个近似技巧 . . . . .	216
12.6 投影法求解 DSGE 模型举例 . . . . .	217
12.6.1 模型设定 . . . . .	217
12.6.2 社会规划者问题 . . . . .	218
12.6.3 参数校准 . . . . .	218
12.6.4 分步求解 . . . . .	219
12.7 稀疏网格 . . . . .	221
12.7.1 Smolyak 法的步骤 . . . . .	223
12.7.2 稀疏网格的一些扩展 . . . . .	228
<b>Appendices</b>	<b>231</b>
附录 12.A 超几何方程 . . . . .	231
附录 12.B 多项式 . . . . .	236
12.B.1 正交集 . . . . .	236
12.B.2 正交多项式 . . . . .	237
12.B.3 递推关系 . . . . .	239
12.B.4 雅各比矩阵和克里斯托费尔——达布关系 . . . . .	240
12.B.5 正交多项式的根 . . . . .	242

12.B.6 高斯求积 . . . . .	243
附录 12.C 常见的正交多项式类型 . . . . .	244
12.C.1 埃米特多项式 . . . . .	245
12.C.2 拉盖尔多项式 . . . . .	249
12.C.3 雅各比多项式 . . . . .	253
12.C.4 勒让德多项式 . . . . .	257
12.C.5 切比雪夫多项式 . . . . .	260
附录 12.D AR(1) 过程的离散方法 . . . . .	267
12.D.1 Tauchen(1986) 法 . . . . .	267
12.D.2 Tauchen and Hussey (1991) 法 . . . . .	270
12.D.3 Rouwenhorst(1995) 法 . . . . .	271
附录 12.E 伽辽金近似 . . . . .	272
12.E.1 举例：1 维空间中的伽辽金近似 . . . . .	272
12.E.2 伽辽金近似 . . . . .	274
12.E.3 强制型方程形式的伽辽金近似收敛 . . . . .	277
12.E.4 仅满足上极限——下极限关系方程形式的伽辽金近似收敛 . . . . .	278
附录 12.F 数值求积 . . . . .	278
12.F.1 牛顿——寇特斯法则 . . . . .	279
12.F.2 几个小技巧 . . . . .	284
12.F.3 求积中的可积奇异点问题 . . . . .	285
12.F.4 梯形法则的近似误差 . . . . .	286
12.F.5 克伦肖——柯蒂斯求积法则 . . . . .	291
12.F.6 程序实现 . . . . .	296
附录 12.G 傅里叶分析 . . . . .	298
12.G.1 傅里叶变换 . . . . .	299
12.G.2 脉冲方程的傅里叶变换 . . . . .	302
12.G.3 非脉冲方程的傅里叶变换 . . . . .	304
12.G.4 原方程的平滑和傅里叶变换方程的衰减 . . . . .	305
12.G.5 傅里叶级数 . . . . .	305
12.G.6 泊松求和式 . . . . .	311
12.G.7 傅里叶分析与卷积 . . . . .	313
12.G.8 高维傅里叶变换 . . . . .	315
12.G.9 一些指数求和的背景知识 . . . . .	315
<b>第十三章 扰动法和投影法的比较</b>	<b>317</b>
13.1 优缺点 . . . . .	317
13.2 实际研究中用哪个？ . . . . .	318
13.3 混合求解法 . . . . .	318
<b>第十四章 有界元法和有限元法</b>	<b>319</b>
14.1 边界值问题：位势方程 . . . . .	319
14.1.1 偏微分算子及椭圆边界值问题 . . . . .	319
14.1.2 边界条件 . . . . .	319
14.1.3 诺依曼边界值问题 . . . . .	322

14.2 方程空间 . . . . .	322
14.2.1 $C^k(\Omega), C^{k,\kappa}(\Omega)$ 空间 . . . . .	322
14.2.2 勒贝格 $L^p(\Omega)$ 空间 . . . . .	324
14.2.3 希尔伯特 (H) 空间 . . . . .	331
14.2.4 索伯列夫 $W^{k,p}(\Omega)$ 空间 . . . . .	335
14.3 变分法 . . . . .	357
14.3.1 算子方程 . . . . .	357
14.3.2 椭圆算子 . . . . .	361
14.3.3 算子与稳定性条件 . . . . .	363
14.3.4 含有限制条件的算子方程 . . . . .	365
14.3.5 混合算子方程 (鞍点变分问题) . . . . .	367
14.3.6 强制算子方程 . . . . .	373
14.4 变分法求解边界值问题 . . . . .	374
14.4.1 位势方程基本介绍 . . . . .	374
14.4.2 狄利克雷边界值问题 1 . . . . .	375
14.4.3 狄利克雷边界值问题 2 . . . . .	379
14.4.4 诺依曼边界值问题 . . . . .	381
14.4.5 混合边界值问题 . . . . .	384
14.4.6 罗宾边界值问题 . . . . .	385
14.5 拉普拉斯算子的基本解 . . . . .	386
14.5.1 基本解 . . . . .	386
14.5.2 拉普拉斯算子 . . . . .	387
14.5.3 三维空间的基本解 . . . . .	388
14.5.4 二维空间的基本解 . . . . .	389
14.5.5 基本解总结 . . . . .	391
14.6 边界积分算子 . . . . .	391
14.6.1 牛顿位势 . . . . .	391
14.6.2 单层位势 . . . . .	397
14.6.3 伴随双层位势 . . . . .	399
14.6.4 双层位势 . . . . .	402
14.6.5 超奇异边界积分算子 . . . . .	405
14.6.6 边界积分算子之间的关系 . . . . .	412
14.6.7 单层位势算子的椭圆性和可逆性 . . . . .	413
14.6.8 超奇异边界积分算子的椭圆性 . . . . .	419
14.6.9 Steklov-Poincaré 算子 . . . . .	422
14.6.10 双层位势的收缩属性 . . . . .	423
14.6.11 边界积分算子的映射特性 . . . . .	427
14.7 边界积分方程 . . . . .	429
14.7.1 狄利克雷边界值问题 . . . . .	430
14.7.2 诺依曼边界值问题 . . . . .	433
14.7.3 混合边界值问题 . . . . .	436
14.7.4 罗宾边界值问题 . . . . .	438
14.8 近似方法 . . . . .	438
14.8.1 伽辽金——布博诺夫法 . . . . .	438

14.8.2 线性形近似 . . . . .	441
14.8.3 算子的近似 . . . . .	442
14.8.4 伽辽金——佩特洛夫法 . . . . .	444
14.8.5 混合边界值问题 . . . . .	446
14.8.6 强制算子 . . . . .	450
14.9 有限元 . . . . .	453
14.9.1 参考元 . . . . .	453
14.9.2 形式方程 . . . . .	461
14.9.3 检测空间 . . . . .	465
<b>第十五章 误差分析</b>	<b>469</b>
15.1 理论分析 . . . . .	469
15.2 初步评估 . . . . .	469
15.3 chi 方误差检测 . . . . .	470
15.4 欧拉等式误差检测 . . . . .	470
15.5 误差值的改进 . . . . .	474
<b>第三部分 模型估计</b>	<b>475</b>
<b>第十六章 将 DSGE 模型和实际数据联系起来</b>	<b>477</b>
16.1 典型 DSGE 模型 . . . . .	477
16.1.1 对数线性化均衡条件 . . . . .	478
16.1.2 模型求解 . . . . .	480
16.1.3 状态——空间表现式 . . . . .	482
16.2 模型的性质 . . . . .	484
16.2.1 自协方差, 预测误差方差 . . . . .	484
16.2.2 光谱分析 . . . . .	487
16.2.3 冲击响应方程 . . . . .	489
16.2.4 条件矩的限制 . . . . .	490
16.3 经验验证 . . . . .	492
16.3.1 自协方差 . . . . .	495
16.3.2 光谱分析 . . . . .	496
16.3.3 冲击响应方程 . . . . .	498
16.3.4 条件矩的限制 . . . . .	499
16.4 处理趋势 . . . . .	500
<b>Appendices</b>	<b>503</b>
附录 16.A 欧拉公式 (复分析) . . . . .	503
16.A.1 欧拉公式和欧拉恒等式 . . . . .	503
16.A.2 欧拉公式的作用 . . . . .	506
附录 16.B 高斯核方程 . . . . .	509
16.B.1 高斯核 . . . . .	509
16.B.2 标准化 . . . . .	509
16.B.3 瀑布特征 . . . . .	510

16.B.4 带宽参数 . . . . .	510
16.B.5 狄拉克方程 . . . . .	511
16.B.6 高斯核和二项式系数的关系 . . . . .	512
16.B.7 中心极限定理 . . . . .	513
16.B.8 各向同性和各向异性 . . . . .	514
附录 16.C 数值线性代数基础 . . . . .	517
16.C.1 矩阵结构与算法复杂度 . . . . .	517
16.C.2 矩阵结构与线性系统求解 . . . . .	518
16.C.3 矩阵分解 . . . . .	522
16.C.4 LU 因子分解 . . . . .	522
16.C.5 Cholesky 因子分解 . . . . .	523
16.C.6 LDL 因子分解 . . . . .	525
16.C.7 块消除与 Schur 补 . . . . .	525
16.C.8 矩阵求逆引理 . . . . .	527
16.C.9 求解欠定线性系统 . . . . .	530
附录 16.D 大数法则 . . . . .	532
16.D.1 弱大数法则 . . . . .	532
16.D.2 强大数法则 . . . . .	534
16.D.3 强均匀大数法则 . . . . .	538

## 第四部分 待分类 541

第十七章 最大似然估计 543	
17.1 线性模型 . . . . .	543
17.2 参数估计 . . . . .	544
17.2.1 回归系数的估计 . . . . .	545
17.2.2 方差的估计 . . . . .	546
17.3 假设检验 . . . . .	547
17.3.1 Wald 检验 . . . . .	547
17.3.2 似然率检验 . . . . .	548
17.3.3 Anova 表 . . . . .	549
17.3.4 简单线性回归 . . . . .	550
17.4 回归诊断 . . . . .	551
17.4.1 拟合值与残差 . . . . .	551
17.4.2 标准化残差 . . . . .	552
17.4.3 Jack-knifed 残差 . . . . .	552
17.4.4 离群值的检测 . . . . .	553
17.4.5 影响力和杠杆 . . . . .	554
17.4.6 真实影响因子和库克距离 . . . . .	554
17.5 数据变换 . . . . .	555
17.5.1 变换观测数据 . . . . .	555
17.5.2 Box-Cox 变换 . . . . .	555
17.5.3 阿特金森法 . . . . .	557

17.6 简介 (待归并) . . . . .	557
17.7 近似的测度 (待归并) . . . . .	558
17.7.1 估计的偏误: 标量参数 . . . . .	558
17.7.2 估计的偏误: 向量参数 . . . . .	559
17.8 费雪信息矩阵和克拉美罗下界 (待归并) . . . . .	559
17.8.1 标量参数 . . . . .	559
17.8.2 向量参数 . . . . .	560
17.9 最大似然估计 (待归并) . . . . .	560
17.9.1 最大似然估计: to be or not to be? . . . . .	560
17.10 伯努利实验: 最大似然估计法 (待归并) . . . . .	561
17.10.1 伯努利实验: 最大似然估计的 Python 程序实现 . . . . .	562
<b>第十八章 准最大似然估计</b>	<b>573</b>
18.1 Kullback-Leibler 信息准则 . . . . .	573
18.1.1 信息 . . . . .	573
18.1.2 Kullback-Leibler 信息准则 . . . . .	574
18.2 QMLE 的渐进性质 . . . . .	577
18.3 信息矩阵等式 . . . . .	579
18.3.1 一种设定形式下的 IME . . . . .	579
18.3.2 另一种 score 方程定义下的 IME . . . . .	581
18.3.3 动态误设定 . . . . .	582
18.3.4 迭代期望法则和条件方差分解 . . . . .	583
18.4 一个例子 . . . . .	583
18.5 异方差问题 . . . . .	586
<b>第五部分 附录</b>	<b>587</b>
<b>Appendices</b>	
<b>附录 第一章 后凯恩斯主义经济学的一个小综述</b>	<b>591</b>
A.1 Introduction . . . . .	591
A.2 PK 的主要方法论 . . . . .	591
A.3 PK 下的市场结构与定价 . . . . .	591
A.3.1 PK 下的市场结构 . . . . .	591
A.3.2 PK 下的定价策略 . . . . .	591
A.4 PK 下的宏观经济 . . . . .	591
A.5 PK 视野下的经济政策 . . . . .	592
<b>附录 第二章 Keynesian, New Keynesian and New Classical Economics</b>	<b>593</b>
B.1 Keynesian 的四个核心 . . . . .	593
B.1.1 失业与有效工资理论 . . . . .	593
B.1.2 价格变化与经济波动 . . . . .	594
B.1.3 储蓄与投资, 信贷配给 . . . . .	595
B.1.4 供应与技术进步 . . . . .	595

---

B.2 New Keynesian 经济学的四个核心 . . . . .	595
B.2.1 有效工资 . . . . .	595
B.2.2 资产配给 . . . . .	595
B.2.3 信贷配给 . . . . .	595
B.2.4 货币政策 . . . . .	596
B.2.5 小结 . . . . .	596
B.3 凯恩斯的不足 . . . . .	596
B.3.1 债券和股票的区别 . . . . .	597
B.3.2 需求和供应 . . . . .	597
B.3.3 投资的决定因素 . . . . .	597
B.3.4 货币政策 . . . . .	598
B.4 小结 . . . . .	598

# 第一章 前言

还没想好要写什么...



# **第一部分**

## **模型构建**



## 第二章 基准 New Keynesian 模型

### 2.1 家庭部门

假定 cashless economy, 家庭的效用函数  $U(t, N_t)$  表示为<sup>1</sup>

$$U(C_t, N_t) = \frac{C_t^{1-\sigma}}{1-\sigma} - \psi \cdot \frac{N_t^{1+\eta}}{1+\eta}, \quad (2.1)$$

HH 问题。目标：追求效用最大化

$$\begin{aligned} & \max_{\{C_t, N_t, B_{t+1}\}} E_0 \sum_{t=0}^{\infty} \beta^t \cdot U(C_t, N_t), \\ & \text{st. } P_t \cdot C_t + B_{t+1} \leq W_t \cdot N_t + Div_t - P_t \cdot T_t + (1 + i_{t-1}) \cdot B_t, \end{aligned} \quad (2.2)$$

其中  $Div_t$  表示 dividends, 中间产品企业的（垄断）利润。 $T_t$  表示税收或转移支付。 $i_t$  为名义利率。

建 Lagrange

$$\mathcal{L} = E_0 \sum_{t=0}^{\infty} \beta^t \cdot \{U(C_t, N_t) + \lambda_t \cdot [W_t \cdot N_t + Div_t - P_t \cdot T_t + (1 + i_{t-1}) \cdot B_t - P_t \cdot C_t - B_{t+1}]\}. \quad (2.3)$$

FOCs

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial C_t} &= 0 \Rightarrow C_t^{-\sigma} = \lambda_t \cdot P_t \\ \frac{\partial \mathcal{L}}{\partial N_t} &= 0 \Rightarrow \psi \cdot N_t^\eta = \lambda_t \cdot W_t \\ \frac{\partial \mathcal{L}}{\partial B_{t+1}} &= 0 \Rightarrow \frac{\partial \{\lambda_t \cdot (1 + i_{t-1}) \cdot B_t\}}{\partial B_{t+1}} - \lambda_t = \frac{\partial \beta \cdot E_t \{\lambda_{t+1} \cdot (1 + i_t) \cdot B_{t+1}\}}{\partial B_{t+1}} - \lambda_t = 0 \end{aligned}$$

整理得

$$\psi \cdot N_t^\eta = C_t^{-\sigma} \cdot w_t \quad (2.4)$$

$$C_t^{-\sigma} = \beta \cdot E_t \{C_{t+1}^{-\sigma} \cdot (1 + i_t) \cdot \frac{P_t}{P_{t+1}}\} = \beta \cdot E_t \{C_{t+1}^{-\sigma} \cdot (1 + i_t) \cdot (1 + \pi_{t+1})^{-1}\}, \quad (2.5)$$

其中  $w_t \equiv \frac{W_t}{P_t}$  表示实际工资,  $\pi_t \equiv P_t/P_{t-1}$  表示通货膨胀率。由 (2.4) 可得 Frisch elasticity of labor supply 为  $1/\eta$ <sup>2</sup>。

<sup>1</sup> 背景：RBC(DSGE) 在经验研究中常用工作小时数而非就业人员数作为  $N_t$  的代理变量，其相关讨论早期文献可见 Hansen (1985); Rogerson (1988); 近期的综述见 Rogerson and Wallenius (2009a)。

<sup>2</sup> 假定短时期内家庭的总财富(总消费)不变, 市场上工资的变化只影响家庭的劳动力供应, Frisch labor supply elasticity(Frisch, 1932, 1959) 可表示为

$$\frac{\partial n_t}{\partial w_t} \cdot \frac{w_t}{n_t},$$

可参考 (Heer and Maussner, 2009, pp.279), 以及 Christiano et al. (2010)。

## 2.2 企业部门

### 2.2.1 最终产品生产部门

最终产品部门以中间产品  $Y_t(j), j \in [0, 1]$  的组合为投入要素，产出  $Y_t$ ，符合完全竞争假定。Dixit and Stiglitz (1977) 形式的生产规模报酬不变生产函数为

$$Y_t = \left[ \int_0^1 Y_t(j)^{\frac{\epsilon-1}{\epsilon}} dj \right]^{\frac{\epsilon}{\epsilon-1}}, \quad (2.6)$$

其中  $\epsilon > 1$  表示  $j^{th}$  中间产品的替代弹性。

最终产品厂商问题：在给定  $P_t(j)$  的情况下，通过选择  $Y_t(j)$  的投入追求利润最大化

$$\max_{Y_t(j)} P_t \cdot Y_t - \int_0^1 P_t(j) \cdot Y_t(j) dj, \quad (2.7)$$

引入式(2.6)，FOC 整理可得对  $j$  中间产品的需求函数

$$Y_t(j) = \left( \frac{P_t(j)}{P_t} \right)^{-\epsilon} \cdot Y_t, \quad (2.8)$$

进而根据完全竞争市场假定

$$\begin{aligned} P_t \cdot Y_t &\equiv \int_0^1 P_t(j) \cdot Y_t(j) dj \\ &= \left[ \int_0^1 \left( \frac{P_t(j)}{P_t} \right)^{-\epsilon} \cdot Y_t \cdot P_t(j) dj \right] \\ &= \left[ \int_0^1 P_t(j)^{1-\epsilon} dj \right] \cdot P_t^\epsilon \cdot Y_t, \end{aligned}$$

整理得最终产品价格的决定 (aggregate price index)：

$$P_t = \left[ \int_0^1 P_t(j)^{1-\epsilon} dj \right]^{\frac{1}{1-\epsilon}} \quad (2.9)$$

### 2.2.2 中间产品生产部门

中间产品生产部门假定处于垄断竞争状态。代表企业  $j$  雇佣劳动力  $N_t(j)$  生产  $Y_t(j)$ ，生产函数形式

$$Y_t(j) = A_t \cdot N_t(j), \quad (2.10)$$

其中  $A_t$  表示外生的生产率冲击，它对所有中间产品生产者都是相同的。对劳动力的总需求为  $j$  个厂商的加总。

$$N_t = \int_0^1 N_t(j) dj. \quad (2.11)$$

#### 成本最小化：边际成本与工资

中间产品生产者  $j$  的问题可以表示为两阶段优化。第一阶段为成本最小化：在给定工资  $W_t$  的基础上，选择雇佣劳动力投入  $N_t(j)$ ，生产中间品  $Y_t(j)$ ，以满足最终产品生产部门对  $Y_t(j)$  的需求，

$$\begin{aligned} &\min_{N_t(j)} W_t \cdot N_t(j), \\ &st. \quad A_t \cdot N_t(j) \geq \left( \frac{P_t(j)}{P_t} \right)^{-\epsilon} Y_t, \end{aligned} \quad (2.12)$$

第二行 LHS 和 RHS 分别表示中间产品  $Y_t(j)$  的供应和需求，见式(2.10)和式 (2.8)。

建 Lagrange

$$\mathcal{L} = W_t \cdot N_t(j) + \lambda_t \cdot \left[ \left( \frac{P_t(j)}{P_t} \right)^{-\epsilon} \cdot Y_t - A_t \cdot N_t(j) \right],$$

其中拉格朗日乘子  $\lambda_t$  表示  $j$  生产额外 1 单位  $Y_t(j)$  的影子价格（边际成本），设为  $MC_t \equiv \lambda_t$ 。

FOC:

$$MC_t = \frac{W_t}{A_t}$$

或者用实际价格形式表示

$$mc_t = \frac{MC_t}{P_t} = \frac{W_t}{A_t \cdot P_t} = \frac{w_t}{A_t}. \quad (2.13)$$

式(2.13) 反映了实际（边际成本）和实际工资的对应关系。

### 利润最大化：定价策略

在此基础上，第二阶段，中间产品生产者  $j$  对自己的产品  $Y_t(j)$  定价  $P_t(j)$ ，以追求实际利润  $\Pi_t$  最大化。

$$\begin{aligned} \Pi_t(j) &= \frac{P_t(j) \cdot Y_t(j)}{P_t} - \frac{W_t \cdot N_t(j)}{P_t} \\ &= \frac{P_t(j)}{P_t} \cdot Y_t(j) - mc_t \cdot Y_t(j) \\ &= \left( \frac{P_t(j)}{P_t} \right)^{1-\epsilon} \cdot Y_t - mc_t \cdot \left( \frac{P_t(j)}{P_t} \right)^{-\epsilon} \cdot Y_t \end{aligned} \quad (2.14)$$

粘性价格。在  $t$  时期，中间产品生产者  $j$  有  $\phi < 1$  的概率不能调整价格，维持上一期的定价  $P_{t-1}(j)$ ；有  $1 - \phi$  的概率可以调整价格，将产品售价更新为  $P_t^{\#}(j)$ 。

$$P_t(j) = \begin{cases} P_{t-1}(j) & \text{with prob. } \phi \\ P_t^{\#}(j) & \text{else } 1 - \phi \end{cases} \quad (2.15)$$

根据模型假设，中间产品生产部门由于垄断产生的利润  $\Pi_t$ ，流回到家庭部门，供消费以提升效用，满足跨期消费的 Euler equation 式(2.5)，设  $\tilde{M}_{t+s} \equiv \beta^s \cdot \frac{U_{C,t+s}}{U_{C,t}}$  作为 discount factor。从  $t$  期向前直到  $t+s$  期， $j$  不能自由调整价格的概率是  $\phi^s$ 。此外，从  $t$  期向前直到  $t+s$  期， $j$  不能自由调整价格的概率是  $\phi^s$ 。由此， $j$  生产者 forward-looking 的随机折旧因子为  $\phi^s \cdot \tilde{M}_{t+s}$ 。

先来看式(2.15)中  $p_j^{\#}(j)$  的决定。 $t$  期中间品生产者  $j$  的利润最大化问题表示为

$$\begin{aligned} &\max_{P_t(j)} E_t \sum_{s=0}^{\infty} \phi^s \cdot \tilde{M}_{t+s} \cdot \Pi_{t+s} \\ &= \max_{P_t(j)} E_t \sum_{s=0}^{\infty} (\beta \cdot \phi)^s \cdot \frac{U_{C,t+s}}{U_{C,t}} \cdot \left[ \frac{P_{t+s}(j)}{P_{t+s}} \cdot Y_{t+s}(j) - mc_{t+s} \cdot Y_{t+s}(j) \right] \\ &= \max_{P_t(j)} E_t \sum_{s=0}^{\infty} (\beta \cdot \phi)^s \cdot \frac{U_{C,t+s}}{U_{C,t}} \cdot Y_{t+s} \cdot \left[ \left( \frac{P_t(j)}{P_{t+s}} \right)^{1-\epsilon} - mc_{t+s} \cdot \left( \frac{P_t(j)}{P_{t+s}} \right)^{-\epsilon} \right], \end{aligned}$$

FOC wrt  $P_t(j)$ ，整理得

$$P_t(j) = \frac{\epsilon}{\epsilon - 1} \cdot \frac{E_t \sum_{s=0}^{\infty} (\beta \cdot \phi)^s \cdot U_{C,t+s} \cdot Y_{t+s} \cdot mc_{t+s} \cdot P_{t+s}^{\epsilon}}{E_t \sum_{s=0}^{\infty} (\beta \cdot \phi)^s \cdot U_{C,t+s} \cdot Y_{t+s} \cdot P_{t+s}^{\epsilon-1}} \quad (2.16)$$

式(2.16) RHS 分子和分母均与  $j$  无关, 即在价格粘性的情况下, forward-looking 的所有中间产品生产者  $j \in [0, 1]$ , 如果有机会调整价格 (概率  $1 - \phi$ ), 都会遵循相同的价格调整策略, 设为  $P_t^{\#} \equiv P_t^{\#}(j), \forall j$ 。上式改写为

$$P_t^{\#} = \frac{\epsilon}{\epsilon - 1} \cdot \frac{X_{1,t}}{X_{2,t}}, \quad (2.17)$$

其中辅助变量

$$X_{1,t} \equiv U_{C,t} \cdot Y_t \cdot P_t^{\epsilon} \cdot mc_t + \beta \cdot \phi \cdot E_t X_{1,t+1}, \quad (2.18)$$

$$X_{2,t} \equiv U_{C,t} \cdot Y_t \cdot P_t^{\epsilon-1} + \beta \cdot \phi \cdot E_t X_{2,t+1}. \quad (2.19)$$

为了让变量平稳, 定义  $x_{1,t} \equiv X_{1,t}/P_t^{\epsilon}$ ,  $x_{2,t} \equiv X_{2,t}/P_t^{\epsilon-1}$ , 式 (2.17) 变为

$$P_t^{\#} = \frac{\epsilon}{\epsilon - 1} \cdot \frac{x_{1,t}}{x_{2,t}} \cdot P_t. \quad (2.20)$$

或者定义 reset price 的通胀项  $(1 + \pi_t^{\#}) \equiv P_t^{\#}/P_{t-1}$ , 将式(2.20)由价格形式改写为通货膨胀率的形式

$$(1 + \pi_t^{\#}) = \frac{\epsilon}{\epsilon - 1} \cdot (1 + \pi_t) \cdot \frac{x_{1,t}}{x_{2,t}}. \quad (2.21)$$

根据(2.20), 在 flexible price 即  $\phi = 0$  的情况下,  $\forall j$  中间产品的定价为  $P_t^{\#} = \frac{\epsilon}{\epsilon - 1} \cdot (mc_t \cdot P_t)$ , 即名义的边际成本乘以 markup。

### 2.2.3 最终产品定价: Calvo assumption

Aggregate price index 式(2.9)中含有  $j$ , 为了消除中间产品生产者的价格异质性对最终产品价格的影响, 引入式(2.15), 根据 Calvo assumption (Calvo, 1983) 得

$$\begin{aligned} P_t^{1-\epsilon} &= \int_0^1 P_t(j)^{1-\epsilon} dj \\ &= \int_0^{1-\phi} (P_t^{\#})^{1-\epsilon} dj + \int_{1-\phi}^1 (P_{t-1}(j))^{1-\epsilon} dj \\ &= \int_0^{1-\phi} (P_t^{\#})^{1-\epsilon} dj + \int_0^{\phi} (P_{t-1}(j))^{1-\epsilon} dj \\ &= (1 - \phi) \cdot \int_0^1 (P_t^{\#})^{1-\epsilon} dj + \phi \cdot \int_0^1 (P_{t-1}(j))^{1-\epsilon} dj \\ &= (1 - \phi) \cdot (P_t^{\#})^{1-\epsilon} + \phi \cdot (P_{t-1})^{1-\epsilon} \end{aligned} \quad (2.22)$$

或者以 reset price inflation 形式表示为

$$(1 + \pi_t)^{1-\epsilon} = (1 - \phi) \cdot (1 + \pi_t^{\#})^{1-\epsilon} + \phi \quad (2.23)$$

## 2.3 外生技术冲击

设外生技术冲击满足 log 形式的 AR(1) 过程

$$\ln A_t = \rho_a \cdot \ln A_{t-1} + \varepsilon_{a,t} \quad (2.24)$$

其中  $0 < \rho_a < 1$ ,  $E\{\varepsilon_{a,t}\} = 0$ 。

## 2.4 总量均衡

### 2.4.1 家庭部门消费的决定

均衡状态下, 家庭持有的债券和净转移支付为零,  $B_t = T_t = 0 \forall t$ 。 $Div_t$  来自全部中间产品生产者的垄断利润之和,

$$\begin{aligned} \frac{Div_t}{P_t} &= \frac{\Pi_t}{P_t} \\ &\equiv \int_0^1 \left[ \frac{P_t(j)}{P_t} \cdot Y_t(j) - \frac{W_t}{P_t} N_t(j) \right] dj \\ &= \int_0^1 \left( \frac{P_t(j)}{P_t} \right) Y_t(j) dj - w_t \cdot \int_0^1 N_t(j) dj \\ &= \int_0^1 \left( \frac{P_t(j)}{P_t} \right) Y_t(j) dj - w_t \cdot N_t \end{aligned} \quad (2.25)$$

家庭部门预算约束条件式(2.2)因此改写为

$$\begin{aligned} C_t &= \int_0^1 \left( \frac{P_t(j)}{P_t} \right) \cdot Y_t(j) dj \\ &= \int_0^1 \left( \frac{P_t(j)}{P_t} \right) \cdot \left[ \left( \frac{P_t(j)}{P_t} \right)^{-\epsilon} \cdot Y_t \right] \\ &= [Y_t \cdot P_t^{\epsilon-1}] \cdot \int_0^1 P_t(j)^{1-\epsilon} dj \\ &= Y_t \end{aligned} \quad (2.26)$$

### 2.4.2 price dispersion

在市场出清情况下, 对  $j$  类中间产品  $Y_t(j)$  的需求式(2.8)与供应式(2.10)联立, 并加总

$$\int_0^1 A_t \cdot N_t(j) dj = \int_0^1 \left( \frac{P_t(j)}{P_t} \right)^{-\epsilon} \cdot Y_t dj,$$

整理可得 (最终产品) 总量生产函数

$$Y_t = \frac{A_t \cdot N_t}{\nu_t^p}, \quad (2.27)$$

其中  $\nu_t^p \geq 1$

$$\nu_t^p = \int_0^1 \left( \frac{P_t(j)}{P_t} \right)^{-\epsilon} dj \quad (2.28)$$

是 price dispersion index, 反映市场上各种商品价格的差异程度: 价格不一致的程度越高,  $\nu_t^p$  越大, 总产出越小; lower bound  $\nu_t^p = 1$  表示在没有 price friction 时, 所有企业对自己的产品都会设定同样的价格。这在一定程度上说明了旨在稳定物价政策的重要性。

利用 Calvo assumption 对  $\nu_t^p$  作调整, 以消除  $j$  个体企业的异质性:

$$\begin{aligned} \nu_t^p &= \int_0^{1-\phi} \left( \frac{P_t^\#}{P_t} \right)^{-\epsilon} dj + \int_{1-\phi}^1 \left( \frac{P_{t-1}(j)}{P_t} \right)^{-\epsilon} dj \\ &= (1-\phi) \cdot \int_0^1 \left( \frac{P_t^\#}{P_{t-1}} \right)^{-\epsilon} \cdot \left( \frac{P_t}{P_{t-1}} \right)^\epsilon dj + \phi \cdot \int_0^1 \left( \frac{P_{t-1}(j)}{P_{t-1}} \right)^{-\epsilon} \cdot \left( \frac{P_t}{P_{t-1}} \right)^\epsilon dj \\ &= (1-\phi) \cdot (1 + \pi_t^\#)^{-\epsilon} \cdot (1 + \pi_t)^\epsilon + \phi \cdot (1 + \pi_t)^\epsilon \cdot \int_0^1 \left( \frac{P_{t-1}(j)}{P_{t-1}} \right)^{-\epsilon} dj \\ &= (1-\phi) \cdot (1 + \pi_t^\#)^{-\epsilon} \cdot (1 + \pi_t)^\epsilon + \phi \cdot (1 + \pi_t)^\epsilon \cdot \nu_{t-1}^p. \end{aligned} \quad (2.29)$$

## 2.5 (非随机的) 稳定状态

根据第2.4节的总量均衡，可以进一步探讨非随机稳定。我们将未标注时间下角标的变量表示为其稳定状态。

根据定义式(2.24)可得

$$A = 1. \quad (2.30)$$

式(2.26)  $\Rightarrow$

$$Y = C. \quad (2.31)$$

跨期消费的 Euler 等式(2.5)  $\Rightarrow$

$$\begin{aligned} i &= \frac{1-\beta}{\beta} + \frac{1}{\beta} \cdot \pi \\ &\approx \rho + \pi, \end{aligned} \quad (2.32)$$

上式中第二行等式，假定时间贴现  $\beta \approx 1$ ，则  $\rho \equiv \frac{1-\beta}{\beta}$  表示时间贴现率。

Reset price inflation 式(2.23)  $\Rightarrow$

$$(1 + \pi^\#) = \left[ \frac{(1 + \pi)^{1-\epsilon} - \phi}{1 - \phi} \right]^{\frac{1}{1-\epsilon}}. \quad (2.33)$$

price dispersion 式(2.29)  $\Rightarrow$

$$\nu^p = \frac{(1 - \phi) \cdot \left[ \frac{1+\pi}{1+\pi^\#} \right]^\epsilon}{1 - \phi \cdot (1 + \pi)^\epsilon}. \quad (2.34)$$

边际成本式(2.21)  $\Rightarrow$

$$mc = \frac{1 - \phi \cdot \beta \cdot (1 + \pi)^\epsilon}{1 - \phi \cdot \beta \cdot (1 + \pi)^{\epsilon-1}} \cdot \frac{1 + \pi^\#}{1 + \pi} \cdot \frac{\epsilon - 1}{\epsilon} \quad (2.35)$$

基于式(2.33)、(2.34)和(2.35)，利用数值模拟的方法，考察稳态下总物价通胀  $\pi$  和 reset price 通胀  $\pi^\#$ 、price dispersion  $\nu^p$ 、边际成本  $mc$  之间的关系。

1.  $\pi = 0$  时， $\pi^\# = 0$ 。 $\pi \geq 0$  时， $\pi^\# \geq \pi$ 。见图2.1左图。
2.  $\pi = 0$  时， $\nu^p = 1$ ，price dispersion 处于最小值。 $\pi \neq 0$  时， $\nu^p > 1$ ，且相比较负通胀 ( $\pi < 0$ )， $\nu^p$  对正通胀的响应更剧烈，导致对产出的干扰更大。见图2.1中图。
3.  $\pi = 0$  时， $mc = \frac{\epsilon-1}{\epsilon}$ ，即等于 fixed price markup 的倒数。 $\pi \neq 0$  时，有  $mc < \frac{\epsilon-1}{\epsilon}$ ，说明  $\pi \neq 0$  时的 steady state markup，大于  $\pi = 0$  时的 steady state price markup。见图2.1右图。

给定  $A = 1$ ，工资决定式(2.13)  $\Rightarrow$

$$w = mc \quad (2.36)$$

即工资等于劳动投入的边际成本。

式(2.27)  $\Rightarrow$  边际产出  $mpn$

$$mpn = \frac{1}{\nu^p} \quad (2.37)$$

可见  $\pi = 0$  时， $mpn$  和  $mc$  的差距最小，体现为一个 fixed price markup  $\frac{\epsilon}{\epsilon-1}$ 。 $\pi \neq 0$  且越远离 0 时，经济体 distorted 的程度越大；此外， $\pi > 0$  时经济体的 distorted 程度，高于  $\pi < 0$  时。

劳动力的供应式(2.4)  $\Rightarrow$

$$\begin{aligned}\psi \cdot N^\eta &= C^{-\sigma} \cdot w \\ &= Y^{-\sigma} \cdot mc \\ &= \left(\frac{N}{\nu^p}\right)^{-\sigma} \cdot mc,\end{aligned}$$

整理得

$$N = \left[ \frac{1}{\psi} \cdot (\nu^p)^\sigma \cdot mc \right]^{\frac{1}{\eta+\sigma}} \quad (2.38)$$

总产出由式(2.27) (2.38) (2.35)得  $\Rightarrow$

$$\begin{aligned}Y &= \frac{N}{\nu^p}, \\ &= \frac{\left[ \frac{1}{\psi} \cdot (\nu^p)^\sigma \cdot mc \right]^{\frac{1}{\eta+\sigma}}}{\nu^p} \\ &= \left( \frac{1}{\psi} \right)^{\frac{1}{\sigma+\eta}} \cdot (\nu^p)^{-\frac{\eta}{\eta+\sigma}} \cdot \left[ \frac{1 - \phi \cdot \beta \cdot (1 + \pi)^\epsilon}{1 - \phi \cdot \beta \cdot (1 + \pi)^{\epsilon-1}} \cdot \frac{1 + \pi^\#}{1 + \pi} \cdot \frac{\epsilon - 1}{\epsilon} \right]^{\frac{1}{\eta+\sigma}}\end{aligned} \quad (2.39)$$

## 2.6 flexible price equilibrium 和 output gap

### 2.6.1 flexible price equilibrium

假定中间产品生产者可以自由调整价格, 对应  $\phi = 0$ 。将变量加上角标  $f$  以标注。此时  $\pi = \pi^\#$ , 且名义价格不会对实际变量产生任何影响。

Flexible price dispersion index 式(2.29)  $\Rightarrow$

$$v_t^{p,f} = 1, \quad (2.40)$$

即包括最终产品和中间产品在内的全部生产者都会采取同样的产品定价策略, 使 price dispersion 位于最低值 1。

Reset price inflation 式(2.21)  $\Rightarrow$

$$mc_t^f = \frac{\epsilon - 1}{\epsilon}, \quad (2.41)$$

可见边际成本是 price markup 的倒数。企业的定价策略在于: 给  $mc$  加上一个固定的 fixed price markup 权数, 作为产品售价。

Flexible price wage  $\Rightarrow$

$$\begin{aligned}mpn_t^f &= \frac{\partial Y_t^f}{\partial N_t^f} = A_t, \\ P_t^f &= mc_t^f \cdot \frac{\epsilon}{\epsilon - 1}, \\ mc_t^f &= w_t, \\ mpn_t &= p_t^f,\end{aligned}$$

整理得

$$w_t^f = A_t \cdot \frac{\epsilon - 1}{\epsilon}. \quad (2.42)$$

Flexible price labor supply  $\Rightarrow$

$$\begin{aligned}\psi(N_t^f)^\eta &= (C_t^f)^{-\sigma} \cdot w_t^f, \\ &= (Y_t^f)^{-\sigma} \cdot mc_t^f, \\ &= \left( A_t \cdot N_t^f \right)^{-\sigma} \cdot \left( \frac{\epsilon - 1}{\epsilon} \cdot A_t \right),\end{aligned}$$

整理得

$$N_t^f = \left( \frac{1}{\psi} \cdot \frac{\epsilon - 1}{\epsilon} \cdot A_t^{1-\sigma} \right)^{\frac{1}{\sigma+\eta}}. \quad (2.43)$$

Flexible price aggregate output  $\Rightarrow$

$$Y_t^f = A_t \cdot N_t^f = \left( \frac{1}{\psi} \cdot \frac{\epsilon - 1}{\epsilon} \right)^{\frac{1}{\sigma+\eta}} \cdot A_t^{\frac{1+\eta}{\sigma+\eta}} \quad (2.44)$$

可见 flexible price ( $\phi = 0$ ) 情况下，名义波动不会对真实变量产生影响。

### 2.6.2 output gap

定义 output gap  $\ln X_t \equiv \ln Y_t - \ln Y_t^f$ ，表示 flexible price ( $\phi = 0$ ) 情况下产出  $Y_t^f$  与 sticky price ( $\phi > 0$ ) 情况下产出  $Y_t$  的差。

稳定状态下， $Y^f$  的值由式(2.44)给出；

$$Y^f = \left( \frac{1}{\psi} \cdot \frac{\epsilon - 1}{\epsilon} \right)^{\frac{1}{\sigma+\eta}}$$

由此可得稳态 output gap 的值：

$$\frac{Y}{Y^f} = \left[ \frac{1 - \phi \cdot \beta \cdot (1 + \pi)^\epsilon}{1 - \phi \cdot \beta \cdot (1 + \pi)^{\epsilon-1}} \cdot \frac{1 + \pi^\#}{1 + \pi} \right]^{\frac{1}{\eta+\sigma}} \cdot (\nu^p)^{-\frac{\eta}{\eta+\sigma}} \quad (2.45)$$

output gap 的性质，分两种情况来讨论：

$$1. \pi = 0 \Rightarrow \pi^\# = \pi = 0, \nu^p = 1 \Rightarrow Y = Y^f \Rightarrow \ln X = 0,$$

$$2. \pi > 0 \Rightarrow \pi^\# > \pi, \nu^p > 1 \Rightarrow Y < Y^f \Rightarrow \ln X < 0.$$

即在 sticky price ( $\phi > 0$ ) 条件下，稳定的 output gap (对数) 为负。

## 2.7 Taylor 法则

假定中央银行的货币政策着眼于利率而非货币供应量：通过盯紧通货膨胀和产出这两个变量，灵活制定内生的货币政策以达到既定的利率目标<sup>3</sup>。常见的内生货币政策如 Taylor 法则：

$$(i_t - i) = \rho_i \cdot (i_{t-1} - i) + (1 - \rho_i) \cdot i + (1 - \rho_i) \cdot [\theta_\pi \cdot (\pi_t - \pi) + \theta_x \cdot (\ln X_t - \ln X)] + \varepsilon_{i,t}, \quad (2.46)$$

其中  $i_t$  表示名义利率。 $X_t$  表示 output gap，见第2.6.2节。去掉下角标的变量表示其稳态值。 $0 \leq \rho_i \leq 1$  表示 smoothing parameter。 $\theta_\pi \geq 0, \theta_x \geq 0$ 。外生利率冲击  $E\{\varepsilon_{i,t}\} = 0$ 。

---

<sup>3</sup> 脊按：外生利率政策会导致 indeterminacy 问题，补充一个 Appendix。

不难看出式(2.46)的 Taylor 法则是个局部调整的货币政策：如果将稳态值  $i$  和  $\pi$  视作长期目标，则中央银行根据上期名义利率（距其稳态值）的偏离程度，以及当前期目标值（距其稳态值）的偏离程度，来灵活调整当期的名义利率；当前期目标值包括通货膨胀和 output gap。一旦当期名义利率的调整目标确定，中央银行即通过向市场印发货币等手段，调节市场上的货币量，来实现其名义利率的目标。

## 2.8 完整的均衡条件

经济系统的均衡解包括下述 12 个变量  $\{C_t, N_t, w_t, mc_t, Y_t, v_t^p, i_t, \pi_t, \pi_t^\#, x_{1,t}, x_{2,t}, A_t\}$ ，以及 12 个等式：

跨期消费的 Euler equation 式(2.5)  $\Rightarrow$

$$C_t^{-\sigma} = \beta \cdot E_t \{ C_{t+1}^{-\sigma} \cdot (1 + i_t) \cdot \frac{P_t}{P_{t+1}} \} = \beta \cdot E_t \{ C_{t+1}^{-\sigma} \cdot (1 + i_t) \cdot (1 + \pi_{t+1})^{-1} \}.$$

劳动力供应式(2.4)  $\Rightarrow$

$$\psi \cdot N_t^\eta = C_t^{-\sigma} \cdot w_t.$$

工资/边际成本的决定式(2.13)  $\Rightarrow$

$$mc_t = \frac{MC_t}{P_t} = \frac{W_t}{A_t \cdot P_t} = \frac{w_t}{A_t}.$$

总消费与总产出的关系式(2.26)  $\Rightarrow$

$$C_t = Y_t.$$

总量生产函数式(2.27)  $\Rightarrow$

$$Y_t = \frac{A_t \cdot N_t}{\nu_t^p}.$$

Price dispersion index 式(2.29)  $\Rightarrow$

$$\nu_t^p = (1 - \phi) \cdot (1 + \pi_t^\#)^{-\epsilon} \cdot (1 + \pi_t)^\epsilon + \phi \cdot (1 + \pi_t)^\epsilon \cdot \nu_{t-1}^p.$$

Evolution of inflation 式(2.23)  $\Rightarrow$

$$(1 + \pi_t)^{1-\epsilon} = (1 - \phi) \cdot (1 + \pi_t^\#)^{1-\epsilon} + \phi.$$

Reset price inflation 式 (2.21)  $\Rightarrow$

$$(1 + \pi_t^\#) = \frac{\epsilon}{\epsilon - 1} \cdot (1 + \pi_t) \cdot \frac{x_{1,t}}{x_{2,t}}.$$

两个 reset price inflation 的辅助变量，式(2.18)-(2.19)  $\Rightarrow$

$$\begin{aligned} x_{1,t} &\equiv C_t^{-\sigma} \cdot Y_t \cdot mc_t + \beta \cdot \phi \cdot E_t (1 + \pi_{t+1})^\epsilon \cdot x_{1,t+1}, \\ x_{2,t} &\equiv C_t^{-\sigma} \cdot Y_t + \beta \cdot \phi \cdot E_t (1 + \pi_{t+1})^{\epsilon-1} \cdot x_{2,t+1}. \end{aligned}$$

Taylor rule 式(2.46)  $\Rightarrow$

$$(i_t - i) = \rho_i \cdot (i_{t-1} - i) + (1 - \rho_i) \cdot i + (1 - \rho_i) \cdot [\theta_\pi \cdot (\pi_t - \pi) + \theta_x \cdot (\ln X_t - \ln X)] + \varepsilon_{i,t}.$$

Exogenous productivity shock 式(2.24)  $\Rightarrow$

$$\ln A_t = \rho_a \cdot \ln A_{t-1} + \varepsilon_{a,t}.$$

## 2.9 对数线性化

### 2.9.1 对数线性化计算

求解上述均衡方程组，方法之一是利用 Dynare 等计算机软件进行计算。此外，在模型较简单的情况下，也可以围绕 zero-inflation steady state  $\pi = 0$  的点，手算对数线性的近似。

Euler equation 式(2.5) + 式(2.26)  $\Rightarrow$

$$-\sigma \cdot \ln Y_t = \ln \beta - \sigma E_t \ln Y_{t+1} + i_t - E_t \pi_{t+1},$$

设  $\ln(1+i_t) \approx i_t$ ,  $\ln(1+\pi_t) \approx \pi_t$ 。将上式中每个变量分别围绕自己的稳态值作一阶泰勒级数展开,

$$-\sigma \cdot \frac{Y_t - Y}{Y} = -\sigma \cdot E_t \frac{Y_{t+1} - Y}{Y} + (i_t - i) - E_t(\pi_{t+1} - \pi),$$

定义  $\tilde{Z}_t \equiv (Z_t - Z)/Z$  作为变量  $Z_t$  距离其稳态值  $Z$  的偏离程度,  $Z_t = (Y_t, N_t, A_t, \dots)$ ; 为了手算方便, 对于已经是 rate form 的变量, 包括利率  $i_t$ 、通货膨胀率  $\pi_t$  和  $\pi_t^\#$ 、price dispersion index  $v_t^p$ , 直接用差分代替增速形式, 如  $\tilde{i}_t \equiv i_t - i$ 。上式进一步改写为

$$\tilde{Y}_t = E_t \tilde{Y}_{t+1} - \frac{1}{\sigma} [\tilde{i}_t - E_t \tilde{\pi}_{t+1}]. \quad (2.47)$$

式(2.47)又被称为 New Keynesian IS Curve (NKIS)。Keynesian IS curve 反映了投资 (Investment) 和投资 (Saving) 之间的对应关系。这里的“基础” New Keynesian 模型中并未考虑投资, 它反映了当期消费需求和实际之间的负相关关系。与传统 IS curve 相比, NKIS 的“新”体现在其 forward-looking 的特征上: 当期需求 ( $\tilde{C}_t = \tilde{Y}_t$ ) 不只取决于实际利率 ( $\tilde{i}_t - E_t \tilde{\pi}_{t+1}$ ) 且负相关, 还取决于对未来收入 (消费) 的期望  $E_t \tilde{Y}_{t+1}$  且负相关。

边际成本与工资的关系式(2.13)  $\Rightarrow$

$$\tilde{m}c_t = \tilde{w}_t - \tilde{A}_t. \quad (2.48)$$

Price dispersion index 式(2.29)  $\Rightarrow$

$$\begin{aligned} \tilde{v}_t^p &= [(-\epsilon) \cdot (1 - \phi) \cdot (1 + \pi^\#)^{-\epsilon-1} \cdot (1 + \pi)^\epsilon] \cdot (\pi_t^\# - \pi^\#) \\ &\quad + [\epsilon \cdot (1 - \phi) \cdot (1 + \pi^\#)^{-\epsilon} \cdot (1 + \pi)^{\epsilon-1}] \cdot \pi_t \\ &\quad + [\epsilon \cdot \phi \cdot (1 + \pi)^{\epsilon-1} \cdot v^p] \cdot (\pi_t - \pi) \\ &\quad + [\phi \cdot (1 + \pi)^\epsilon] \cdot (v_{t-1}^p - v^p) \\ &= -\epsilon \cdot (1 - \phi) \cdot \tilde{\pi}_t^\# + \epsilon \cdot (1 - \phi) \cdot \tilde{\pi}_t + \epsilon \cdot \phi \cdot \tilde{\pi}_t + \phi \cdot \tilde{v}_{t-1}^p \\ &= -\epsilon \cdot (1 - \phi) \cdot \tilde{\pi}_t^\# + \epsilon \cdot \tilde{\pi}_t + \phi \cdot \tilde{v}_{t-1}^p, \end{aligned} \quad (2.49)$$

其中第三个等号用到  $\pi = \pi^\# = 0$  的假设条件。

Evolution of inflation 式(2.23)  $\Rightarrow$

$$\begin{aligned} (1 - \epsilon) \cdot \ln(1 + \pi_t) &= \ln \left[ (1 - \phi) \cdot \left( 1 + \pi_t^\# \right)^{1-\epsilon} + \phi \right], \\ (1 - \epsilon) \cdot (\pi_t - \pi) &= \left[ (1 - \phi) \cdot (1 - \epsilon) \cdot (1 + \phi^\#)^{-\epsilon} \cdot (1 + \pi)^{\epsilon-1} \right], \\ \tilde{\pi}_t &= (1 - \phi) \cdot \tilde{\pi}_t^\#. \end{aligned} \quad (2.50)$$

由式(2.50)可见, actual inflation 和 reset price inflation 呈一定比例变化, 比例  $1 - \phi$  为全部企业中, 调整价格者所占的比重。

总量生产函数式(2.27)  $\Rightarrow$

$$\begin{aligned}\tilde{Y}_t &= \tilde{A}_t + \tilde{N}_t - \tilde{\nu}_t^p \\ &= \tilde{A}_t + \tilde{N}_t - \left[ \epsilon \cdot \tilde{\pi}_t - \epsilon \cdot (1 - \phi) \cdot \tilde{\pi}_t^\# + \phi \cdot \tilde{\nu}_{t-1}^p \right] \\ &\approx \tilde{A}_t + \tilde{N}_t,\end{aligned}\quad (2.51)$$

第三行的约等号是由于, 从式(2.49)可得, 在 zero inflation steady state 下,  $\pi = \pi^\# = 0$  且  $\nu^p = 1$ , 则  $(\nu_t^p - 1) = \phi \cdot (\nu_{t-1}^p - 1)$ 。这说明  $\tilde{\nu}_t^p = 0 \forall t$ 。换句话说, 在 price dispersion index 是一个二阶量; 在我们围绕 zero inflation steady state 作一阶线性近似时, 可以省略。

flexible price 下的产出水平式(2.44)  $\Rightarrow$

$$\tilde{Y}_t^f = \frac{1 + \eta}{\sigma + \eta} \cdot \tilde{A}_t. \quad (2.52)$$

劳动力供应式(2.4)  $\Rightarrow$

$$\begin{aligned}\eta \cdot \tilde{N}_t &= -\sigma \cdot \tilde{Y}_t + \tilde{w}_t \\ &= -\sigma \cdot \tilde{Y}_t + m\tilde{c}_t + \tilde{A}_t,\end{aligned}$$

引入式(2.51)替代 LHS 的  $\tilde{N}_t$ , 引入式(2.44)替代 RHS 的  $\tilde{A}_t$ , 可得

$$\begin{aligned}\tilde{m}\tilde{c}_t &= (\sigma + \eta) \cdot \tilde{Y}_t - (1 + \eta) \cdot \tilde{A}_t \\ &= (\sigma + \eta) \cdot \tilde{Y}_t - \left( \frac{\sigma + \eta}{1 + \eta} \right) \cdot \tilde{Y}_t^f \cdot (1 + \eta) \\ &= (\sigma + \eta) \cdot \left( \tilde{Y}_t - \tilde{Y}_t^f \right) \\ &= (\sigma + \eta) \cdot \tilde{X}_t,\end{aligned}\quad (2.53)$$

式(2.53) LHS 是 fixed steady state price markup 的倒数  $(\epsilon - 1)/\epsilon$ 。

1.  $\tilde{X}_t = 0 \Rightarrow \tilde{m}\tilde{c}_t = 0 \Rightarrow m\tilde{c}_t \equiv mc = \frac{\epsilon - 1}{\epsilon} \Rightarrow$  price markups 等于 desired steady state fixed price markup  $\frac{\epsilon}{\epsilon - 1}$

2.  $\tilde{X}_t < 0 \Rightarrow Y_t < Y_t^f \Rightarrow \tilde{m}\tilde{c}_t < mc \Rightarrow$  实际边际成本小于 steady state fixed price markup,  $m\tilde{c}_t < \frac{\epsilon}{\epsilon - 1} \Rightarrow$  该经济体 is more distorted。

辅助变量  $x_{1,t}$  式(2.18)  $\Rightarrow$

$$\begin{aligned}\ln x_{1,t} &= \ln [Y_t^{1-\sigma} \cdot mc_t + \phi \cdot \beta \cdot E_t \cdot (1 + \pi_{t+1})^\epsilon \cdot x_{1,t+1}], \\ \frac{x_{1,t} - x_1}{x_1} &= \frac{1}{x_1} \cdot \{ [(1 - \sigma) \cdot Y^{-\sigma} \cdot mc] \cdot (Y_t - Y) \\ &\quad + Y^{1-\sigma} \cdot (mc_t - mc) \\ &\quad + [\phi \cdot \beta \cdot \epsilon \cdot (1 + \pi)^{\epsilon-1} \cdot x_1] \cdot E_t (\pi_{t+1} - \pi) \\ &\quad + [\phi \cdot \beta \cdot (1 + \pi)^\epsilon] \cdot E_t (x_{1,t+1} - x_1)\},\end{aligned}\quad (2.54)$$

由式(2.18)可得, 在  $\pi = 0$  时

$$x_1 = \frac{Y^{-\sigma} \cdot mc}{1 - \phi \cdot \beta}, \quad (2.55)$$

联立(2.54)-(2.55)得

$$\tilde{x}_{1,t} = (1 - \sigma) \cdot (1 - \phi \cdot \beta) \cdot \tilde{Y}_t + (1 - \phi \cdot \beta) \cdot \tilde{m}c_t + \epsilon \cdot \phi \cdot \beta \cdot E_t \tilde{\pi}_{t+1} + \phi \cdot \beta \cdot E_t \tilde{x}_{1,t+1}. \quad (2.56)$$

辅助变量  $x_{2,t}$  式(2.19)  $\Rightarrow$

$$\begin{aligned} \ln x_{2,t} &= \ln \left[ Y_t^{1-\sigma} + \phi \cdot \beta \cdot E_t (1 + \pi_{t+1})^{\epsilon-1} \cdot x_{2,t+1} \right], \\ \frac{x_{2,t} - x_2}{x_2} &= \frac{1}{x_2} \cdot \{ [(1 - \sigma) \cdot Y^{-\sigma}] \cdot (Y_t - Y) \\ &\quad + [\phi \cdot \beta \cdot E_t (\epsilon - 1) \cdot (1 + \pi)^{\epsilon-2} \cdot x_2] \cdot (\pi_{t+1} - \pi) \\ &\quad + [\phi \cdot \beta \cdot E_t (1 + \pi)^{\epsilon-1}] \cdot (x_{2,t+1} - x_2), \end{aligned} \quad (2.57)$$

由式(2.19)可得，在  $\pi = 0$  时

$$x_2 = \frac{Y^{-\sigma}}{1 - \phi \cdot \beta} \quad (2.58)$$

联立式(2.57)-(2.58)得

$$\tilde{x}_{2,t} = (1 - \sigma) \cdot (1 - \phi \cdot \beta) \cdot \tilde{Y}_t + (\epsilon - 1) \cdot \phi \cdot \beta \cdot E_t \tilde{\pi}_{t+1} + \phi \cdot \beta \cdot E_t \tilde{x}_{2,t+1}. \quad (2.59)$$

将式(2.59)-(2.59)代入式(2.21)得

$$\begin{aligned} \tilde{\pi}_t^\# - \tilde{\pi}_t &= \tilde{x}_{1,t} - \tilde{x}_{2,t} \\ &= (1 - \phi \cdot \beta) \cdot \tilde{m}c_t + \phi \cdot \beta \cdot E_t \tilde{\pi}_{t+1} + \phi \cdot \beta \cdot E_t (\tilde{x}_{1,t+1} - \tilde{x}_{2,t+1}). \end{aligned} \quad (2.60)$$

式(2.50)与(2.60)联立，可得 Reset price inflation 式

$$\tilde{\pi}_t = \left( \frac{1 - \phi}{\phi} \right) \cdot (1 - \phi \cdot \beta) \cdot \tilde{m}c_t + \beta \cdot E_t \tilde{\pi}_{t+1}. \quad (2.61)$$

式(2.61)又被称为 New Keynesian Philips Curve (NKPC)，它反映了中央银行对产出和通货膨胀的 trade off。也可以用式(2.53)中的 output gap  $\tilde{X}_t$  代替边际成本  $\tilde{m}c_t$ ，NKPC 改写为

$$\begin{aligned} \tilde{\pi}_t &= \left( \frac{1 - \phi}{\phi} \right) \cdot (1 - \phi \cdot \beta) \cdot [(\sigma + \eta) \cdot (\tilde{Y}_t - \tilde{Y}_t^f)] + \beta \cdot E_t \tilde{\pi}_{t+1} \\ &= \left( \frac{1 - \phi}{\phi} \right) \cdot (1 - \phi \cdot \beta) \cdot [(\sigma + \eta) \cdot \tilde{X}_t] + \beta \cdot E_t \tilde{\pi}_{t+1}. \end{aligned} \quad (2.62)$$

或者，around a zero inflation steady state，用 forward-looking 形式表现 NKPC

$$\tilde{\pi}_t = \left( \frac{1 - \phi}{\phi} \right) \cdot (1 - \phi \cdot \beta) \cdot \sum_{s=0}^{\infty} \beta^s \cdot \tilde{m}c_{t+s} \quad (2.63)$$

$$= \left( \frac{1 - \phi}{\phi} \right) \cdot (1 - \phi \cdot \beta) \cdot (\sigma + \eta) \cdot \sum_{s=0}^{\infty} \beta^s \cdot \tilde{X}_{t+s}. \quad (2.64)$$

Keynesian PC curve 反映了通货膨胀率和边际成本之间的对应关系。NKPC 的“新”体现在其 forward-looking 的特征上。式(2.63)表明当前通货膨胀率和对未来实际边际成本的贴现值之间呈等比关系，边际成本表现为 price markup 的倒数。在价格是完全浮动的情况下  $\phi = 0$ ，企业会根据 desired constant markups 给产品定价；如果企业预期未来的边际成本会增加，那么在定价时就会相应调低 price markups。在存在价格粘性的情况下  $\phi > 0$ ，有机会在当前  $t$  期调整价格的企

业，便会提前提高产品价格（以避免在未来  $t + s$  时间期内无法再次调整产品售价），以达到他们的 desired price markup，这造成了通货膨胀。反之亦然。

NKPC curve 的“斜率”与  $\phi$  负相关。随着  $\phi$  逐渐降低，NKPC 线越来越陡峭。当  $\phi \rightarrow 0$  时，perfectly flexible price，NKPC 线完全垂直，此时  $\tilde{Y}_t = \tilde{Y}_t^f$ ,  $\tilde{m}c_t = 0$ 。

Exogenous productivity shock 式(2.24)  $\Rightarrow$

$$\tilde{A}_t = \rho_a \cdot \tilde{A}_{t-1} + \varepsilon_{a,t}. \quad (2.65)$$

Taylor rule 式(2.46)  $\Rightarrow$

$$\tilde{i}_t = \rho_i \cdot \tilde{i}_{t-1} + (1 - \rho_i) \cdot [\phi_\pi \cdot \tilde{\pi}_t + \phi_x \cdot \tilde{X}_t]. \quad (2.66)$$

## 2.9.2 线性模型

作线性近似处理后的经济系统，表述为如下 5 个变量  $(\tilde{Y}_t, \tilde{i}_t, \tilde{\pi}_t, \tilde{Y}_t^f, \tilde{A}_t)$ ，5 个等式线性 Euler equation，反映总量需求的 NKIS 式(2.47)  $\Rightarrow$

$$\tilde{Y}_t = E_t \tilde{Y}_{t+1} - \frac{1}{\sigma} (\tilde{i}_t - E_t \tilde{\pi}_{t+1}).$$

反映总量供应的 NKPC 式(2.61)或 (2.62)  $\Rightarrow$

$$\begin{aligned} \tilde{\pi}_t &= \left( \frac{1 - \phi}{\phi} \right) \cdot (1 - \phi \cdot \beta) \cdot \tilde{m}c_t + \beta \cdot E_t \tilde{\pi}_{t+1} \\ &= \left( \frac{1 - \phi}{\phi} \right) \cdot (1 - \phi \cdot \beta) \cdot [(\sigma + \eta) \cdot \tilde{X}_t] + \beta \cdot E_t \tilde{\pi}_{t+1}. \end{aligned}$$

辅助变量，反映 flexible price 下的产出水平，式(2.52)  $\Rightarrow$

$$\tilde{Y}_t^f = \frac{1 + \eta}{\sigma + \eta} \cdot \tilde{A}_t.$$

Exogenous productivity shock 式(2.24)  $\Rightarrow$

$$\tilde{A}_t = \rho_a \cdot \tilde{A}_{t-1} + \varepsilon_{a,t}.$$

反映货币政策的 Taylor rule 式(2.46)  $\Rightarrow$

$$\tilde{i}_t = \rho_i \cdot \tilde{i}_{t-1} + (1 - \rho_i) \cdot [\phi_\pi \cdot \tilde{\pi}_t + \phi_x \cdot \tilde{X}_t].$$

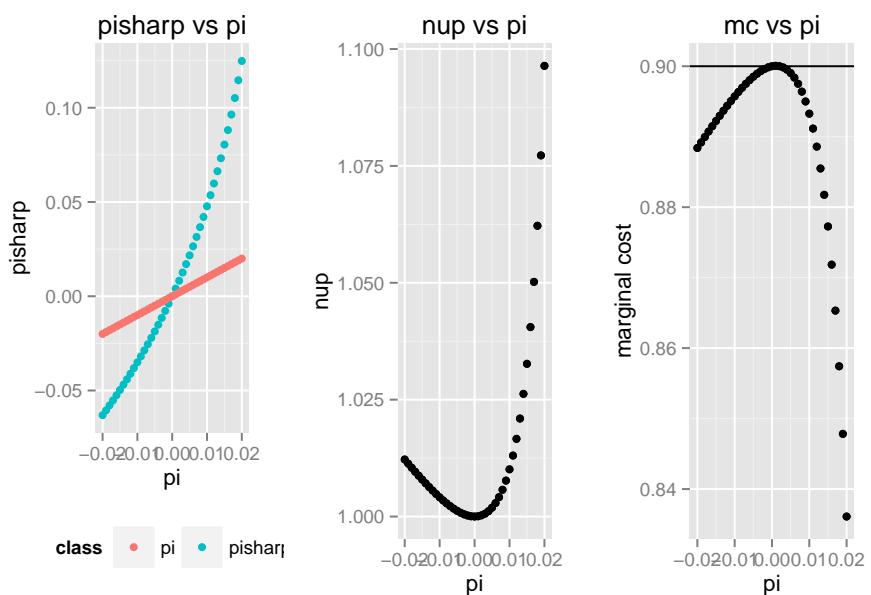


图 2.1: 数值模拟:  $\{\pi^\#, \nu^p\}$  vs  $\pi$ 。模拟过程中设参数值  $\phi = 0.25$ ,  $\epsilon = 10$ ,  $\beta = 0.99$ 。

# 第三章 Medium-Sized DSGE 模型

## 3.1 产品的生产部门

### 3.1.1 最终产品生产部门

最终产品生产部门以中间产品  $Y_t(i), i \in (0, 1)$  的组合为投入要素，产出  $Y_t$ ，符合完全竞争假定。Dixit and Stiglitz (1977) 形式的生产规模报酬不变生产函数为

$$Y_t = \left[ \int_0^1 Y_t(i)^{\frac{\varepsilon_f - 1}{\varepsilon_f}} di \right]^{\frac{\varepsilon_f}{\varepsilon_f - 1}}, \quad (3.1)$$

其中  $\varepsilon_f$  表示  $i^{th}$  中间产品的替代弹性。

最终产品厂商问题：在给定中间产品价格  $P_t(i)$  的情况下，通过选择  $Y_t(i)$  的投入追求利润最大化

$$\max_{Y_t(i)} P_t \cdot Y_t - \int_0^1 P_t(i) \cdot Y_t(i) di.$$

引入式(2.6)，FOC 整理可得对  $i$  中间产品的需求函数

$$Y_t(i) = \left( \frac{P_t(i)}{P_t} \right)^{-\varepsilon_f} \cdot Y_t, \quad (3.2)$$

进而根据完全竞争市场假定

$$\begin{aligned} P_t \cdot Y_t &\equiv \int_0^1 P_t(i) \cdot Y_t(i) di \\ &= \left[ \int_0^1 \left( \frac{P_t(i)}{P_t} \right)^{-\varepsilon_f} \cdot Y_t \cdot P_t(i) di \right] \\ &= \left[ \int_0^1 P_t(i)^{1-\varepsilon_f} di \right] \cdot P_t^{\varepsilon_f} \cdot Y_t, \end{aligned}$$

整理得最终产品价格的决定 (aggregate price index)：

$$P_t = \left[ \int_0^1 P_t(i)^{1-\varepsilon_f} di \right]^{\frac{1}{1-\varepsilon_f}} \quad (3.3)$$

### 3.1.2 中间产品生产部门

每个  $i^{th}$  中间产品有且只有一个生产者，处于垄断竞争状态，视式(3.2)-(2.9)为需求函数和价格；最终产品的总产出  $Y_t$  和总价格  $P_t$  是外生的。 $i^{th}$  中间品生产者的生产函数为

$$Y_t(i) = K_t(i)^\alpha \cdot [z_t \cdot H_t(i)]^{1-\alpha} - z_t^+ \cdot \varphi, \quad (3.4)$$

其中投入要素  $H_t(i)$  和  $K_t(i)$  分别表示同质的劳动力和实物资本服务，产出系数  $0 < \alpha < 1$ 。 $z_t$  和  $\Psi_t$  分别表示 neutral 和 investment-specific 类型的随机技术冲击，见第3.1.7节。 $\varphi$  为生产的固定成本； $z_t^+$  定义如下

$$z_t^+ = \Psi_t^{\frac{\alpha}{1-\alpha}} \cdot z_t, \quad (3.5)$$

可见沿着非随机的稳态增长路径， $Y_t(i)/z_t^+$  收敛至常数<sup>1</sup>。

$i^{th}$  中间产品生产者雇佣同质劳动服务  $H_t(i)$ ，假定生产者必须 100% 借款来支付工资，根据Christiano et al. (2005) 的 working capital channel 设定加以简化 ( $v_t = 0, \psi = 1$ )，每一单位劳动服务的成本等于

$$\text{cost} = (1 - v_t) \cdot (1 - \psi + \psi \cdot R_t) \cdot W_t = R_t \cdot W_t,$$

其中  $W_t$  和  $R_t$  分别表示总量水平上的名义工资和 working capital 借贷的名义利率<sup>2</sup>。

经济体受到的外部技术冲击有二，分别为  $z_t$  和  $\Psi_t$ ，均假定为对数形式的单位根过程<sup>3</sup>：

$$\Delta \log z_t = \mu_z + \varepsilon_{z,t}, E(\varepsilon_{z,t}) = \sigma_z^2, \quad (3.6)$$

$$\Delta \log \Psi_t = (1 - \rho_\Psi) \cdot \mu_\Psi + \rho_\Psi \cdot \Delta \log \Psi_{t-1} + \varepsilon_{\Psi,t}, E(\varepsilon_{\Psi,t}) = \sigma_\Psi^2. \quad (3.7)$$

### 3.1.3 中间产品生产者的边际成本

$i^{th}$  中间产品生产者的最优行为表现为两阶段优化。第一阶段是成本最小化，选择合适的投入要素  $\{K_t(i), H_t(i)\}$  数量，生产出满足最终产品生产部门需求的  $Y_t(i)$ 。投入要素价格分别为市场给定的  $\bar{R}_t$  及  $\bar{W}_t$ ，

$$\begin{aligned} & \min_{K_t(i), H_t(i)} \bar{W}_t \cdot H_t(i) + \bar{R}_t \cdot K_t(i), \\ & \text{st. } Y_t(i) = K_t(i)^\alpha \cdot [z_t \cdot H_t(i)]^{1-\alpha} - z_t^+ \cdot \varphi. \end{aligned}$$

建 Lagrangian

$$\mathcal{L} = \bar{W}_t \cdot H_t(i) + \bar{R}_t \cdot K_t(i) + \lambda_t \cdot \left\{ Y_t(i) - K_t(i)^\alpha \cdot [z_t \cdot H_t(i)]^{1-\alpha} + z_t^+ \cdot \varphi \right\}.$$

<sup>1</sup>膏按：模型引入固定成本的设定，以及采取一个变量乘以一个常量形式设定的考虑。模型假定中间产品生产者处于垄断地位，且假定没有进出 (no entry and exit)，以确保垄断利润得以长期持续，因此在中间产品生产函数中引入常数固定生产成本  $\varphi$ ，使得稳态利润是零，被固定成本抵消，维持 no entry 的假设成立。但这需要 fixed cost 的增速与产出增速相等，由此对常数  $\varphi$  额外乘以一个系数  $z_t^+$ 。此外经验研究的发现也肯定了这种设定的意义：在引入规模报酬递增的情况下，正的货币政策冲击产生后，劳动的生产率随着提高，这与现实的观测基本接近。

<sup>2</sup>Working Capital Channel 可追溯至Basu (1995)。Christiano et al. (2005) 对资本和劳动服务的 working capital 做了建模。近期研究可见Phaneuf et al. (2015)。根据不含 working capital channel 的 DSGE 模型模拟出来的结果，对经济体注入正的货币政策冲击会导致通货膨胀率有大幅度的提升，而现实世界中通货膨胀并没有上升的如此剧烈。Christiano et al. (2005) 因此引入了 working capital channel 设定以缓解该问题，其思路如下：正的货币政策冲击降低了名义利率，进而降低企业进行外部融资（支付员工工资）的边际成本，而边际成本对企业价格决策的制定至关重要。一系列对经济现实的 VAR-based 观察发现，扩张性货币政策冲击出现后的初期，通货膨胀率出现小幅的下降。这从侧面印证了 working capital channel 假设的重要性，除了可能存在计量意义上的设定失误之外，恐怕只有 working capital channel 的假设可以解释它了。计量意义上的设定失误，Sims (1992) 最早提出这种可能；Christiano, Lawrence J et al. (1999) 进一步探讨了该可能性；另外可见Bernanke et al. (2005)。

<sup>3</sup>膏按：式(3.6)直接将中性技术进步定义为有 drift 的随机游走过程，依据如下：

1. estimation 部分，Smets and Wouters (2003) 估计  $\log z_t$  发现它高度自相关，
2. Prescott (1986) 的经验研究，
3. Fernald (2009) 对美国经济部门的经验研究发现，1947 年第 2 季度到 2009 年第三季度 TFP 增速的一阶自回归系数为 0.0034。

FOCs:

$$\frac{\partial \mathcal{L}}{\partial H_t(i)} = 0 \Rightarrow \bar{W}_t = \lambda_t \cdot (1 - \alpha) \cdot z_t^{1-\alpha} \cdot H_t(i)^{-\alpha} \cdot K_t(i)^\alpha, \quad (3.8)$$

$$\frac{\partial \mathcal{L}}{\partial K_t(i)} = 0 \Rightarrow \bar{R}_t = \lambda_t \cdot \alpha \cdot z_t^{1-\alpha} \cdot H_t(i)^{1-\alpha} \cdot K_t(i)^{\alpha-1}. \quad (3.9)$$

式(3.8)- (3.9)可得

$$\frac{\bar{W}_t}{\bar{R}_t} = \left( \frac{\alpha}{1 - \alpha} \right)^{-1} \cdot \left( \frac{K_t(i)}{H_t(i)} \right). \quad (3.10)$$

由式(3.10)可得

$$\frac{K_t(i)}{H_t(i)} \equiv \frac{K_t}{H_t}, \quad \forall i, \quad (3.11)$$

即所有中间产品生产者以及总量层面上的投入要素之比相等。

将式(3.10)带回生产函数式(2.10), 得

$$Y_t(i) + z_t^+ \cdot \varphi = z_t^{1-\alpha} \cdot \left( \frac{\bar{W}_t}{\bar{R}_t} \right)^{-(1-\alpha)} \cdot \left( \frac{\alpha}{1 - \alpha} \right)^{-(1-\alpha)} \cdot K_t(i) = z_t^{1-\alpha} \cdot \left( \frac{\bar{W}_t}{\bar{R}_t} \right)^\alpha \cdot \left( \frac{\alpha}{1 - \alpha} \right)^\alpha \cdot H_t(i).$$

进而最优投入要素的数量

$$K_t(i) = [Y_t(i) + z_t^+ \cdot \varphi] \cdot z_t^{-(1-\alpha)} \cdot \left( \frac{\bar{W}_t}{\bar{R}_t} \right)^{1-\alpha} \cdot \left( \frac{\alpha}{1 - \alpha} \right)^{1-\alpha},$$

$$H_t(i) = [Y_t(i) + z_t^+ \cdot \varphi] \cdot z_t^{-(1-\alpha)} \cdot \left( \frac{\bar{W}_t}{\bar{R}_t} \right)^{-\alpha} \cdot \left( \frac{\alpha}{1 - \alpha} \right)^{-\alpha}.$$

$i^{th}$  中间产品生产者的成本函数

$$C(Y_t(i), K_t(i), H_t(i)) = \bar{W}_t \cdot H_t(i) + \bar{R}_t \cdot K_t(i) = (Y_t(i) + z_t^+ \cdot \varphi) \cdot z_t^{-(1-\alpha)} \cdot \left( \frac{\bar{W}_t}{1 - \alpha} \right)^{1-\alpha} \cdot \left( \frac{\bar{R}_t}{\alpha} \right)^\alpha. \quad (3.12)$$

因此名义边际成本  $S_t(i)$  等于

$$S_t(i) = \frac{\partial C(Y_t(i), K_t(i), H_t(i))}{\partial Y_t(i)} = z_t^{-(1-\alpha)} \cdot \left( \frac{\bar{W}_t}{1 - \alpha} \right)^{1-\alpha} \cdot \left( \frac{\bar{R}_t}{\alpha} \right)^\alpha, \quad (3.13)$$

类似地,  $S_t(i) \equiv S_t, \forall i$ 。

根据定义, 投入要素的市场价格

$$\bar{W}_t = (1 - v_t) \cdot (1 - \psi + \psi \cdot R_t) \cdot W_t, \quad (3.14)$$

$$\bar{R}_t = (1 - v_t) \cdot (1 - \psi + \psi \cdot r_t^k) \cdot P_t \quad (3.15)$$

引入  $v_t = 0, \psi = 1$  的设定后, 实际边际成本  $s_t$  等于

$$s_t \equiv \frac{S_t}{P_t} = \frac{\left( \frac{R_t \cdot W_t}{1 - \alpha} \right)^{1-\alpha} \cdot \left( \frac{r_t^k \cdot P_t}{\alpha} \right)^\alpha}{z_t^{1-\alpha} \cdot P_t}. \quad (3.16)$$

引入式 (3.140)、(3.139)、(3.5)后, 式(3.16)改写为 scaled 形式

$$s_t = \frac{\left( \frac{R_t \cdot (\bar{w}_t \cdot z_t^+ \cdot P_t)}{1 - \alpha} \right)^{1-\alpha} \cdot \left( \frac{\left( \frac{r_t^k}{\Psi_t} \right) \cdot P_t}{\alpha} \right)^\alpha}{z_t^{1-\alpha} \cdot P_t} = \left( \frac{\bar{w}_t \cdot R_t}{1 - \alpha} \right)^{1-\alpha} \cdot \left( \frac{\bar{r}_t^k}{\alpha} \right)^\alpha. \quad (3.17)$$

此外，根据 productive efficiency，生产额外 1 单位  $Y_t(i)$  的边际成本  $s_t$ ，等于雇佣额外 1 单位劳动的成本  $W_t \cdot R_t$ ，除以劳动的名义边际产出  $P_t \cdot \frac{\partial Y_t(i)}{\partial H_t(i)}$ ，即

$$\begin{aligned}
s_t &= \frac{W_t \cdot R_t}{P_t \cdot \frac{\partial Y_t(i)}{\partial H_t(i)}} \\
&= \frac{(\bar{w}_t \cdot z_t^+ \cdot P_t) \cdot R_t}{(1 - \alpha) \cdot \left[ \frac{K_t(i)}{H_t(i)} \right]^\alpha \cdot z_t^{1-\alpha} \cdot P_t} \\
&= \frac{\bar{w}_t \cdot R_t \cdot z_t^+ \cdot P_t}{(1 - \alpha) \cdot \left[ \frac{\frac{k_t(i)}{z_{t-1}^+ \cdot \Psi_{t-1}}}{H_t(i)} \right]^\alpha \cdot (z_t^+ \cdot \Psi^{-\frac{\alpha}{1-\alpha}})^{1-\alpha} \cdot P_t} \\
&= \frac{\bar{w}_t \cdot R_t}{(1 - \alpha) \cdot \left[ \frac{k_t(i)}{H_t(i) \cdot \mu_{z+,t} \cdot \mu_{\Psi,t}} \right]^\alpha} \\
&= \frac{\bar{w}_t \cdot R_t \cdot z_t^+ \cdot P_t}{(1 - \alpha) \cdot \left[ \frac{\frac{k_t(i)}{z_{t-1}^+ \cdot \Psi_{t-1}}}{H_t(i)} \right]^\alpha \cdot (z_t^+ \cdot \Psi^{-\frac{\alpha}{1-\alpha}})^{1-\alpha} \cdot P_t} \\
&= \frac{\bar{w}_t \cdot R_t}{(1 - \alpha) \cdot \left[ \frac{k_t}{H_t \cdot \mu_{z+,t} \cdot \mu_{\Psi,t}} \right]^\alpha}.
\end{aligned} \tag{3.18}$$

其中最后一个等号利用式(3.34)消除异质性。

### 3.1.4 中间产品生产者的定价

#### updating 生产者的定价策略

价格刚性。假定在  $t$  时期，有  $0 \leq \phi_f \leq 1$  比例的中间产品生产者  $i \in [0, 1]$  无法调整价格；有  $1 - \phi_f$  比例的生产者不可以调整价格。可以调整价格的中间产品垄断生产者，其 forward-looking 问题可以描述为，制定合理的垄断价格  $P_t^\#(i)$  以满足利润最大化

$$\max_{P_t^\#(i)} E_t \sum_{j=0}^{\infty} (\phi_f \cdot \beta)^j \cdot v_{t+j} \cdot \left[ P_t^\#(i) \cdot Y_{t+j}(i) - P_{t+j} \cdot s_{t+j} \cdot Y_{t+j}(i) \right], \tag{3.19}$$

st. 对  $i^{th}$  中间产品的需求式(3.2)。 $s_t$  的定义见式 (3.17)。中括号中的内容为  $t+j$  时刻  $i^{th}$  生产者的利润； $\beta^j \cdot v_{t+j}$  为家庭部门跨期预算约束条件的乘数<sup>4</sup>，满足  $v_t = \frac{\partial U_t(\cdot)/\partial C_t}{P_t}$ 。<sup>5</sup>

引入式(3.2)替代  $Y_{t+j}(i)$ ，整理得

$$\begin{aligned}
&\max_{P_t^\#(i)} E_t \sum_{j=0}^{\infty} (\phi_f \cdot \beta)^j \cdot \left( \frac{U_{C,t+j}}{P_{t+j}} \right) \cdot \left( \frac{P_{t+j}^\#(i)}{P_{t+j}} \right)^{-\varepsilon_f} \cdot Y_{t+j} \cdot \left[ P_t^\#(i) - P_{t+j} \cdot s_{t+j} \right] \\
&= \max_{P_t^\#(i)} E_t \sum_{j=0}^{\infty} (\phi_f \cdot \beta)^j \cdot (U_{C,t+j} \cdot Y_{t+j}) \cdot \left[ P_t^\#(i)^{-(\varepsilon_f-1)} \cdot P_{t+j}^{\varepsilon_f-1} - P_t^\#(i)^{-\varepsilon_f} \cdot P_{t+j}^{\varepsilon_f} \cdot s_{t+j} \right].
\end{aligned}$$

<sup>4</sup>中间产品生产部门的垄断利润，假设全部返回家庭部门；家庭部门基于自己的消费偏好作跨期消费决策；因此对于  $i^{th}$  来说， $\beta^j \cdot v_{t+j}$  是外生给定的。

<sup>5</sup>膏接：补一个 eqref，嵌到 hh 部门的 max problem 上去。

FOC wrt  $P_t^\#(i)$ ,

$$\begin{aligned} & [-(\varepsilon_f - 1)] \cdot \max_{P_t^\#(i)} E_t \sum_{j=0}^{\infty} (\phi_f \cdot \beta)^j \cdot (U_{C,t+j} \cdot Y_{t+j}) \cdot [P_t^\#(i)^{-\varepsilon_f} \cdot P_{t+j}^{\varepsilon_f - 1}] \\ & = -\varepsilon_f \cdot \max_{P_t^\#(i)} E_t \sum_{j=0}^{\infty} (\phi_f \cdot \beta)^j \cdot (U_{C,t+j} \cdot Y_{t+j}) \cdot [P_t^\#(i)^{-\varepsilon_f - 1} \cdot P_{t+j}^{\varepsilon_f} \cdot s_{t+j}] \end{aligned}$$

整理得

$$P_t(i)^\# = \frac{\varepsilon_f}{\varepsilon_f - 1} \cdot \frac{E_t \sum_{j=0}^{\infty} (\phi_f \cdot \beta)^j \cdot (U_{C,t+j} \cdot Y_{t+j}) \cdot (P_{t+j}^{\varepsilon_f} \cdot s_{t+j})}{E_t \sum_{j=0}^{\infty} (\phi_f \cdot \beta)^j \cdot (U_{C,t+j} \cdot Y_{t+j}) \cdot (P_{t+j}^{\varepsilon_f - 1})}. \quad (3.20)$$

式(3.20)的 RHS 与个体企业  $i^{th}$  无关；因此  $P_t(i)^\# \equiv P_t^\#$ ,  $\forall i$ 。

定义两个辅助变量  $X_{1,t}^f$ ,  $X_{2,t}^f$

$$X_{1,t}^f \equiv E_t \sum_{j=0}^{\infty} (\phi_f \cdot \beta)^j \cdot (U_{C,t+j} \cdot Y_{t+j}) \cdot (P_{t+j}^{\varepsilon_f} \cdot s_{t+j}) = U_{C,t} \cdot Y_t \cdot P_t^{\varepsilon_f} \cdot s_t + \phi_f \cdot \beta \cdot E_t X_{1,t+1}^f, \quad (3.21)$$

$$X_{2,t}^f \equiv E_t \sum_{j=0}^{\infty} (\phi_f \cdot \beta)^j \cdot (U_{C,t+j} \cdot Y_{t+j}) \cdot (P_{t+j}^{\varepsilon_f - 1}) = U_{C,t} \cdot Y_t \cdot P_t^{\varepsilon_f - 1} + \phi_f \cdot \beta \cdot E_t X_{2,t+1}^f. \quad (3.22)$$

进一步对辅助变量作 scaling

$$\begin{aligned} x_{1,t}^f & \equiv \frac{X_{1,t}^f}{P_t^{\varepsilon_f}} \\ & = U_{C,t} \cdot Y_t \cdot s_t + \phi_f \cdot \beta \cdot E_t (1 + \pi_{t+1})^{\varepsilon_f} \cdot x_{1,t+1}^f, \\ & = \psi_{z+,t} \cdot y_t \cdot s_t + \phi_f \cdot \beta \cdot E_t (1 + \pi_{t+1})^{\varepsilon_f} \cdot x_{1,t+1}^f. \end{aligned} \quad (3.23)$$

$$\begin{aligned} x_{2,t}^f & \equiv \frac{X_{2,t}^f}{P_t^{\varepsilon_f - 1}} \\ & = U_{C,t} \cdot Y_t + \phi_f \cdot \beta \cdot E_t (1 + \pi_{t+1})^{\varepsilon_f - 1} \cdot x_{2,t+1}^f \\ & = \psi_{z+,t} \cdot y_t + \phi_f \cdot \beta \cdot E_t (1 + \pi_{t+1})^{\varepsilon_f - 1} \cdot x_{2,t+1}^f. \end{aligned} \quad (3.24)$$

其中上两式的最后一个等号根据式(3.76)、(3.138)和(3.142)求得

$$U_{C,t} \cdot Y_t = (v_t \cdot P_t) \cdot (z_t^+ \cdot y_t) = \psi_{z+,t} \cdot y_t, \quad (3.25)$$

式(3.20)因此变为

$$P_t^\# = \frac{\varepsilon_f}{\varepsilon_f - 1} \cdot \frac{X_{1,t}^f}{X_{2,t}^f} = \frac{\varepsilon_f}{\varepsilon_f - 1} \cdot \frac{x_{1,t}^f}{x_{2,t}^f} \cdot P_t. \quad (3.26)$$

定义 reset price inflation  $1 + \pi_t^\# \equiv \frac{P_t^\#}{P_{t-1}}$ , 则式(3.26)两侧同时除以  $P_{t-1}$  得

$$(1 + \pi_t^\#) = \frac{\varepsilon_f}{\varepsilon_f - 1} \cdot (1 + \pi_t) \cdot \frac{x_{1,t}^f}{x_{2,t}^f}. \quad (3.27)$$

### Aggregate Price Index

利用 Calvo pricing 方法 (Calvo, 1983), aggregate price index 式(2.9)可以改写为

$$P_t^{1-\varepsilon_f} = \int_0^1 P_t(i)^{1-\varepsilon_f} di = \int_0^{1-\phi_f} P_t^{\#}{}^{1-\varepsilon_f} di + \int_{1-\phi_f}^1 p_{t-1}(i)^{1-\varepsilon_f} di = (1-\phi_f) \cdot P_t^{\#}{}^{1-\varepsilon_f} + \phi_f \cdot P_{t-1}^{1-\varepsilon_f}.$$

等式两侧同时除以  $P_{t-1}^{1-\varepsilon_f}$ , 整理得

$$(1 + \pi_t^{\#}) = \left[ \frac{(1 + \pi_t)^{1-\varepsilon_f} - \phi_f}{1 - \phi_f} \right]^{\frac{1}{1-\varepsilon_f}}. \quad (3.28)$$

式(3.27)与式(3.28)联立, 整理可得

$$\left[ \frac{1 - \phi_f \cdot (1 + \pi_t)^{\varepsilon_f-1}}{1 - \phi_f} \right]^{\frac{1}{1-\varepsilon_f}} = \frac{\varepsilon_f}{\varepsilon_f - 1} \cdot \frac{x_{1,t}^f}{x_{2,t}^f}. \quad (3.29)$$

### 3.1.5 price dispersion index

Medium-sized DSGE 模型中, 同质总产出  $Y_t^{sum}$  是  $i^{th}$  种中间产品产出  $Y_t(i)$  之和, 结合式(3.4)可得

$$\begin{aligned} Y_t^{sum} &= \int_0^1 Y_t(i) di \\ &= \int_0^1 \left[ (z_t \cdot H_t(i))^{1-\alpha} \cdot K_t(i)^\alpha - z_t^+ \cdot \varphi \right] di \\ &= \int_0^1 \left[ z_t^{1-\alpha} \cdot \left( \frac{K_t(i)}{H_t(i)} \right)^\alpha \cdot H_t(i) - z_t^+ \cdot \varphi \right] di \\ &= z_t^{1-\alpha} \cdot \left( \frac{K_t}{H_t} \right)^\alpha \cdot \int_0^1 H_t(i) di - z_t^+ \cdot \varphi \\ &= z_t^{1-\alpha} \cdot K_t^\alpha \cdot H_t^{1-\alpha} - z_t^+ \cdot \varphi, \end{aligned} \quad (3.30)$$

其中  $K_t$  和  $H_t$  分别表示经济体中同质资本服务品和同质劳动力的数量。成本最小化的中间产品企业面临相同的要素价格, 因此致力于雇佣同等比例的资本-劳动投入品进行生产, 最后一个等号因此消除了企业  $i$  的异质性。市场出清情况下对中间产品的总需求式(3.2)等于总供应式(3.30), 整理得

$$Y_t = \frac{z_t^{1-\alpha} \cdot K_t^\alpha \cdot H_t^{1-\alpha} - z_t^+ \cdot \varphi}{\int_0^1 \left( \frac{p_t(i)}{P_t} \right)^{-\varepsilon_f} di} = \frac{z_t^{1-\alpha} \cdot K_t^\alpha \cdot H_t^{1-\alpha} - z_t^+ \cdot \varphi}{\nu_t^f}. \quad (3.31)$$

其中定义了产品价格的分布指标  $\nu_t^f \geq 1$ 。当  $\nu_t^f = 1$  时, 不存在 price dispersion, 实际总产出最大。利用 calvo pricing 可得

$$\begin{aligned} \nu_t^f &\equiv \int_0^1 \left( \frac{P_t(i)}{P_t} \right)^{-\varepsilon_f} di \\ &= \int_0^{1-\phi_f} \left( \frac{P_t^{\#}}{P_t} \right)^{-\varepsilon_f} di + \int_{1-\phi_f}^1 \left( \frac{P_{t-1}(i)}{P_t} \right)^{-\varepsilon_f} di \\ &= (1 - \phi_f) \cdot \left[ \frac{1 + \pi_t}{1 + \pi_t^{\#}} \right]^{\varepsilon_f} + \phi_f \cdot (1 + \pi_t)^{\varepsilon_f}. \end{aligned} \quad (3.32)$$

式(3.28)与式(3.32)联立得

$$\nu_t^f = (1 - \phi_f) \cdot \left[ \frac{1 - \phi_f \cdot (1 + \pi_t)^{\varepsilon_f-1}}{1 - \phi_f} \right]^{\frac{\varepsilon_f}{\varepsilon_f-1}} + \phi_f \cdot (1 + \pi_t)^{\varepsilon_f} \cdot \nu_{t-1}^f. \quad (3.33)$$

此外，联立式(3.138)、(3.31)和(3.40)可得

$$y_t \equiv \frac{Y_t}{z_t^+} = \frac{1}{\nu_t^f} \cdot \left[ \frac{z_t^{1-\alpha} \cdot K_t^\alpha \cdot H_t^{1-\alpha}}{z_t^+} - \varphi \right] \\ =$$
(3.34)

结合式(3.40)、式(3.68)，可将式(3.34)变为如下 scale 形式

$$y_t = \frac{1}{\nu_t^f} \cdot \left[ \left( \frac{u_t \cdot \bar{k}_t}{\mu_{z^+, t} \cdot \mu_{\Psi, t}} \right)^\alpha \cdot (h_t \cdot \nu_t^w)^{1-\alpha} - \varphi \right]$$
(3.35)

### 3.1.6 投资的调节成本

$F(I_t, I_{t-1})$  表示投资的调节成本，定义为

$$F(I_t, I_{t-1}) = \left[ 1 - S \left( \frac{I_t}{I_{t-1}} \right) \right] \cdot I_t,$$
(3.36)

$S(I_t, I_{t-1})$  表示为投资的调节成本，也有用实物资本的调节形式表现的。adjustment cost 的详细讨论见第3.C节。隐函数  $S(\cdot)$  满足  $S(1) = S'(1) = 0$ ,  $S'' = \kappa$  是个常数，即 DSGE 模型的稳定状态与  $\kappa$  系数值无关。

### 3.1.7 总产出的分配

市场出清条件下总产出的分配

$$Y_t = C_t + \tilde{I}_t + G_t,$$
(3.37)

其中  $G_t$  表示外生的政府支出。 $C_t$  表示家庭消费支出。同质的投资品  $\tilde{I}_t$  定义式如下

$$\tilde{I}_t \equiv \frac{I_t + a(u_t) \cdot \bar{K}_t}{\Psi_t},$$
(3.38)

同质投资品  $\tilde{I}_t$  用于形成投资品  $I_t$ 。 $I_t$  被家庭部门用于增加下一个时间期的实物资本存量  $\bar{K}_{t+1}$ 。 $a(u_t) \cdot \bar{K}_t$  是实物资本的维护成本， $0 \leq u_t \leq 1$  是可变的资本利用率，反映家庭的 variable capital utilization 决策，递增且 convex 的成本函数  $a(u_t)$  表示对应于  $u_t$ ，利用实物资本存量从事生产的成本；在稳定状态下，家庭会设  $u = 1$  即实物资本  $\bar{K}_t$  全部用于生产活动，且成本函数  $a(u) = 0$ 。

unit root 的 investment specific technology shock  $\Psi_t$  程度越大，同等单位  $\tilde{I}_t$  所能形成的投资品  $I_t$  越多。 $\Psi_t$  的定义式见 (3.7)。

实物资本服务品  $K_t$  与实物资本存量  $\bar{K}_t$  和利用率  $u_t$  有关：

$$K_t \equiv u_t \cdot \bar{K}_t.$$
(3.39)

对式(3.39)作 scale 调整。根据式(3.39)和(3.133)得

$$K_t = u_t \cdot \bar{K}_t \\ = u_t \cdot (\bar{k}_t \cdot z_{t-1}^+ \cdot \Psi_{t-1}) \\ = \frac{u_t \cdot \bar{k}_t \cdot z_t^+ \cdot \Psi_t}{\mu_{z^+, t} \cdot \mu_{\Psi, t}}.$$
(3.40)

进而总产出分配式(3.37)的 scale 形式如下

$$y_t = g_t + c_t + i_t + \frac{a(u_t) \cdot \bar{k}_t}{\mu_{z^+, t} \cdot \mu_{\Psi, t}}.$$
(3.41)

### 可变的资本利用率

如果  $u_t = u$  即资本利用率是个常数，模型模拟出来的结果显示，通货膨胀率会随着外生货币政策冲击而发生很大波动；然而在实际经济运行过程中观测，往往通货膨胀的波动并没有那么大。在模型中引入可变的资本利用率  $u_t$ ，有助于解释通货膨胀对货币政策冲击的响应速度为何缓慢：价格很大程度上由生产成本所决定；生产成本又受到生产要素弹性的影响；如果弹性很大，则一个较小的成本上升即可导致较大的投入要素量变化，从而使得通货膨胀率对货币冲击的响应缓慢且温和。建模时使投入要素弹性变大的方法有很多，其一便是让实物资本利用率是可变的，可以使实物资本服务品更有弹性——如果  $a()$  函数的曲率很低 (very little curvature in the  $a$  function)，那么家庭部门可以在确保成本不大幅度提高的情况下，增加资本服务品的供应。

## 3.2 家庭部门

家庭部门拥有全部生产资料（劳动力和实物资本），向生产部门供应生产资料以获得收入（工资和垄断利润）。假定劳动力市场上存在具有不同特征的异质劳动力  $h_t(j)$ ,  $j \in (0, 1)$ ，且对于  $j' \neq j$ ,  $h_t(j)$  和  $h_t(j')$ （至少部分地）可以相互替代。引入工资粘性的设定 (Erceg et al., 2000)，存在一个垄断者为每一种  $h_t(j)$  分别定价，并且由于可替代性的存在，其垄断定价的能力受到市场竞争的限制 (Christiano et al., 2010)。

### 3.2.1 劳动力投入：同质化假定还是异质化假定？

家庭的效用函数中引入劳动力投入  $h_t^{1+\eta}$ ，反映休闲带来的效用（或者劳动带来的负效用）。幂指数的倒数  $1/\eta$  反映在保持消费水平不变的情况下，劳动力供应相对于真实工资水平的弹性。宏观经济学研究中  $h_t$  进而  $\eta$  反映何种含义，引起广泛争论。

一种观点是假定经济体中的家庭部门都是同质的， $h_t$  反映了用就业市场上一个典型劳动者工作时间（小时数），体现了家庭的 labor effort。此时  $1/\eta$  系数又称 Frisch elasticity of labor supply<sup>6</sup>，描述典型劳动者随着工资的变化，愿意增加或减少的工作时间。

另一种做法是仍然假定家庭部门的同质性， $h_t$  反映了劳动力市场上的就业人数。 $1/\eta$  描述了额外一个边际劳动者，随着工资的变化，愿意进入或是离开劳动力市场；它不反映某一个特定个人的劳动力供应的弹性。

已有大量基于家庭调查数据的微观层面经验研究发现，Frisch 弹性尽管跨国差异较大，但总的来说值比较小，这往往意味着  $\eta \geq 1$ 。早期宏观经济的经验研究往往采用这一设定，但存在不足：引入外生冲击后，RBC 模型的结果往往显示，就业的波动要远远大于工资的波动，这与实际观测到的数据不符。并且从（宏观经济学的）道理上讲，理性经济行为者应该对工资的波动做出较有弹性的应对，使得就业率波动小于工资，劳动力供应的弹性较大， $\eta < 1$ 。

宏观和微观研究中的分歧似乎可以从这个角度来理解：微观层面上的 Frisch elasticity，和宏观层面上的 labor supply elasticity，所描述的对象并不一致。Rogerson (1988); Hansen (1985) 等人论证了，在 Frisch elasticity of labor supply 等于 0 的情况下，总就业仍然可能随着实际工资的小幅度变化而出现大幅度震荡 (Rogerson and Wallenius, 2009b)。

有鉴于此，Galí (2008) 提出了新的模型设定思路，随后被 Christiano et al. (2010); Mulligan (2001); Krusell et al. (2008, 2011) 等所采用：

- 经济体中每个典型家庭都有大量成员  $j \in (0, 1)$ ,
- 任何工资水平下的 Frisch labor supply elasticity 都为 0,

<sup>6</sup>见附录3.B。

- 每个家庭成员只有两个状态，对应两种效用函数：
  - 被雇佣， $\log C_t^{employed} - l^n$ ，
  - 失业， $\log C_t^{unemployed}$ ，
 其中  $l$  代表对工作厌恶程度： $l$  越高的家庭成员（老幼病人）越厌恶工作。
- 家庭效用最大化的目标是追求内部所有成员整体效用的最大化，即所有成员无论工作与否，消费水平相等， $C_t = C_t^{employed} = C_t^{unemployed}$ 。

如果家庭需要提供  $H_t$  单位劳动力，那么家庭全部成员中， $0 \leq l \leq H_t$  的去工作， $l > H_t$  的不工作。那么对所有  $l \in [0, 1]$  求积分后，消费水平为  $C_t$ ，就业水平为  $H_t$  的典型家庭的效用函数为  $\log C_t - \frac{H_t^{1+\eta}}{1+\eta}$ 。

在这样的设定下， $H_t$  重新表示劳动者的数量， $\eta$  不再作为衡量 Frisch 弹性（设为 0）的指标，它表示在受到外部经济环境冲击的情况下，进入或离开劳动力市场的家庭成员变化的弹性。如果  $\eta$  比较大，反映出家庭内部各个成员之间厌恶工作的程度差异较大，分布较平均。此时工资的变化只会导致就业量发生较小的变化。如果  $\eta$  比较小，反映出家庭内部各个成员中，厌恶工作的差异程度较小，且集中在对于工作不工作无所谓的水平附近，此时工资的变化会导致就业量发生较大的变化。

### 3.2.2 劳动力承包商

设产品生产部门所需要的同质劳动服务  $H_t$  由劳动力承包商提供。承包商负责向家庭部门征集一系列具有不同特性  $j \in (0, 1)$  的劳动力投入  $h_t(j)$ ，满足生产部门的需求，投入产出关系符合 Dixit-Stiglitz 形式 (Dixit and Stiglitz, 1977)

$$H_t = \left[ \int_0^1 h_t(j)^{\frac{\varepsilon_w - 1}{\varepsilon_w}} dj \right]^{\frac{\varepsilon_w}{\varepsilon_w - 1}}. \quad (3.42)$$

承包商是完全竞争的，视  $W_t$  和  $H_t$  为生产部门所给定，视  $W_t(j)$  为家庭部门所给定。劳动承包商的最大化问题：

$$\max_{h_t(j)} W_t \cdot H_t - \int_0^1 W_t(j) \cdot h_t(j) dj.$$

引入式(3.42)，FOC 整理得对  $h_t(j)$  的需求函数

$$h_t(j) = \left( \frac{W_t(j)}{W_t} \right)^{-\varepsilon_w} \cdot H_t. \quad (3.43)$$

$$\begin{aligned} W_t \cdot H_t &\equiv \int_0^1 W_t(j) \cdot h_t(j) dj \\ &= \left[ \int_0^1 \left( \frac{W_t(j)}{W_t} \right)^{-\varepsilon_w} \cdot H_t \cdot W_t(j) dj \right] \\ &= \left[ \int_0^1 W_t(j)^{1-\varepsilon_w} dj \right] \cdot W_t^{\varepsilon_w} \cdot H_t, \end{aligned}$$

整理得最终工资的决定 (aggregate wage index)：

$$W_t = \left[ \int_0^1 W_t(j)^{1-\varepsilon_w} dj \right]^{\frac{1}{1-\varepsilon_w}} \quad (3.44)$$

### 3.2.3 家庭行为

经济体中存在一系列同质化家庭，相关假定及评述第3.2.1节。一个典型家庭中存在许多成员，对应异质化的劳动力特征  $j \in [0, 1]$ 。诸多家庭劳动承包商供应  $h_t(j)$  异质劳动力，满足式 (3.43)。诸多家庭的  $j^{th}$  类型劳动力汇总到垄断竞争的  $j^{th}$  劳工联盟，由联盟制定工资  $W_t(j)$ ，见下节<sup>7</sup>。

假定总量层面上，家庭的效用函数表现为

$$U(C_t, h_t) = \log(C_t - b \cdot C_{t-1}) - A \cdot \int_0^1 \frac{h_t(j)^{1+\eta}}{1+\eta} dj. \quad (3.45)$$

效用来自消费和休闲<sup>8</sup>两方面，二者对效用的正效果相加得到总效用函数，是内部可分的 (intratemporal separability)<sup>9</sup>。

#### (消费) 习惯的形成

消费的效应是跨期可分 (intertemporal separability) 的<sup>10</sup>。 $b > 0$  表示 habit persistence parameter<sup>11</sup>， $b = 0$  时当期效用仅与当期消费有关，habit formation 不存在，模型回到经典的 PIH 模式 (permanent income hypothesis)。 $0 < b < 1$  时，当期效用不仅取决于当期消费水平，还与当期消费相对于过去消费水平的变化程度有关，因此又称消费偏好的习惯形成。

宏观经济模型中引入消费习惯的形成设定，主要是为了减小传统 PIH-based 模型的模拟结果与实际情况间的背离，如：

1. Excess smoothness puzzle。消费习惯越是强 ( $b$  越接近于 1)，在 permanent income 出现波动时，消费水平随之变化的幅度就越小 (Carroll et al., 1997, 2000)。
2. Asset pricing 方面的 Equity premium puzzle。消费习惯越是强 ( $b$  越接近于 1)，消费者的行为决策看起来就更加具有风险规避的特征，从而使我们在构建模型时，不必为相对风险规避系数定义一个大的离谱的值，如  $C_t^{1-\sigma}/(1-\sigma)$  中的  $\sigma$ : 一方面可以设  $\sigma = 1$  ( $\sigma$  越大，消费者越厌恶风险)；另一方面使得消费者行为仍然表现出厌恶风险的特性 (Constantinides, 1990; Boldrin et al., 2001)。

引入消费 habit formation 设定后的 DSGE 模型，其模拟的冲击-响应结果更符合现实 (基于 VAR 对现实数据的观察): 货币政策等外生冲击可以造成消费响应的 hump-shape。而传统 PIH based DSGE 模型中消费的响应往往做不到这一点 (Fuhrer, 2000)。除了消费之外，HF-based DSGE 模型还可以生成实际利率持续下降的结果。

### 3.2.4 劳工联盟的工资策略

$j^{th}$  劳工联盟居于垄断竞争地位，满足劳动力承包商对  $h_t(j)$  的需求式(3.43)。假定价格刚性存在，在  $t$  时期有  $0 \leq \phi_w \leq 1$  比例的劳工联盟  $j \in [0, 1]$  无法调整价格。

对于剩下的可以调整价格的  $1 - \phi_w$  联盟而言，名义工资的决定式

$$W_{t+1}(j) \equiv (1 + \tilde{\pi}_{w,t+1}) \cdot W_t(j), \quad (3.46)$$

其中

$$1 + \tilde{\pi}_{w,t+1} \equiv (1 + \pi_t)^{\kappa_w} \cdot (1 + \pi)^{1-\kappa_w} \cdot \mu_{z+}, \quad 0 < \kappa_w < 1. \quad (3.47)$$

<sup>7</sup>膏按：补一个 reference。

<sup>8</sup>休闲的效应以劳动的负效果 (disutility) 形式体现，因此用负号。

<sup>9</sup>内部不可分的效用函数的例子，可见King et al. (1988a,b); Greenwood et al. (1988); Guerrón-Quintana (2008)。

<sup>10</sup>效用函数中，关于 intertemporal- 和 intratemporal (non)separability 的介绍，可见 Eric Sims(2015) 讲义。

<sup>11</sup>经验研究中常常限定  $b < 1$ ，这是出于计算方便的需要：如果  $b = 1$ ，则稳定状态下消费的边际效用就变成无穷大了。

可以调整工资的垄断劳工联盟，其 forward-looking 问题可以描述为，制定合理的垄断工资  $W_t^{\#}(j)$  以追求效用最大化

$$\max E_t \sum_{m=0}^{\infty} (\phi_w \cdot \beta)^m \cdot \left[ v_{t+m} \cdot W_{t+m}^{\#}(j) \cdot h_{t+m}(j) - A_L \cdot \frac{h_{t+m}(j)^{1+\eta}}{1+\eta} \right], \quad (3.48)$$

中括号中的内容为  $t+j$  时刻  $i^{th}$  类型劳动力供应为家庭部门带来的效用；与第3.1.4节类似， $\beta^m \cdot v_{t+m}$  为家庭部门跨期预算约束条件的乘数， $v_t = \frac{\partial U_t(\cdot)/\partial C_t}{P_t}$ 。

$t$  时刻制定的垄断工资直到  $t+m$  时段都无法再调整，因此由式(3.46)得

$$\begin{aligned} W_{t+m}^{\#}(j) &= W_{t+m-1}^{\#}(j) \cdot (1 + \tilde{\pi}_{w,t+m}) \\ &= W_{t+m-2}^{\#}(j) \cdot (1 + \tilde{\pi}_{w,t+m}) \cdot (1 + \tilde{\pi}_{w,t+m-1}) \\ &= \dots \\ &= W_t^{\#}(j) \cdot [(1 + \tilde{\pi}_{w,t+m}) \cdot \dots \cdot (1 + \tilde{\pi}_{w,t+1})]. \end{aligned} \quad (3.49)$$

并且

$$\frac{z_{t+m}^+}{z_t^+} = \frac{z_{t+m}^+}{z_{t+m-1}^+} \cdot \frac{z_{t+m-1}^+}{z_{t+m-2}^+} \cdot \dots \cdot \frac{z_{t+1}^+}{z_t^+} = \mu_{z^+, t+m} \cdot \dots \cdot \mu_{z^+, t+1}, \quad (3.50)$$

$$\frac{P_{t+m}}{P_t} = \frac{P_{t+m}}{P_{t+m-1}} \cdot \frac{P_{t+m-1}}{P_{t+m-2}} \cdot \dots \cdot \frac{P_{t+1}}{P_t} = (1 + \pi_{t+m}) \cdot \dots \cdot (1 + \pi_{t+1}), \quad (3.51)$$

式(3.49)两侧同时除以  $W_{t+m}$ ，引入 scaling 式 (3.139)、(3.50) 和 (3.51) 得

$$\frac{W_{t+m}^{\#}(j)}{W_{t+m}} = \frac{W_t^{\#}(j) \cdot (\tilde{\pi}_{w,t+m} \cdot \dots \cdot \tilde{\pi}_{w,t+1})}{W_{t+m}} \quad (3.52)$$

$$= \frac{\left(\frac{W_t^{\#}(j)}{W_t}\right) \cdot (\bar{w}_t \cdot z_t^+ \cdot P_t) \cdot (\tilde{\pi}_{w,t+m} \cdot \dots \cdot \tilde{\pi}_{w,t+1})}{\bar{w}_{t+m} \cdot z_{t+m}^+ \cdot P_{t+m}} \quad (3.53)$$

$$= \left(\frac{W_t^{\#}(j)}{W_t}\right) \cdot \left(\frac{\bar{w}_t}{\bar{w}_{t+m}}\right) \cdot \frac{[(1 + \tilde{\pi}_{w,t+m}) \cdot \dots \cdot (1 + \tilde{\pi}_{w,t+1})]}{[(1 + \pi_{t+m}) \cdot \dots \cdot (1 + \pi_{t+1})] \cdot [\mu_{z^+, t+m} \cdot \dots \cdot \mu_{z^+, t+1}]} \quad (3.54)$$

$$= \left(\frac{W_t^{\#}(j)}{W_t}\right) \cdot \left(\frac{\bar{w}_t}{\bar{w}_{t+m}}\right) \cdot \mathcal{X}_{t,m}, \quad (3.55)$$

其中定义辅助变量

$$\mathcal{X}_{t,m} = \begin{cases} \frac{[(1 + \tilde{\pi}_{w,t+m}) \cdot \dots \cdot (1 + \tilde{\pi}_{w,t+1})]}{[(1 + \pi_{t+m}) \cdot \dots \cdot (1 + \pi_{t+1})] \cdot [\mu_{z^+, t+m} \cdot \dots \cdot \mu_{z^+, t+1}]} & \text{if } m \geq 0, \\ 1 & \text{if } m = 0. \end{cases} \quad (3.56)$$

利用式(3.52)可以将式(3.43)改写为

$$h_{t+m}(j) = \left[ \left(\frac{W_t^{\#}(j)}{W_t}\right) \cdot \left(\frac{\bar{w}_t}{\bar{w}_{t+m}}\right) \cdot \mathcal{X}_{t,m} \right]^{-\varepsilon_w} \cdot H_{t+m}. \quad (3.57)$$

效用最大化问题式(3.48)改写为

$$\max E_t \sum_{m=0}^{\infty} (\phi_w \cdot \beta)^m \cdot \left[ v_{t+m} \cdot W_{t+m} \cdot \left(\frac{W_{t+m}^{\#}(j)}{W_{t+m}}\right) \cdot h_{t+m}(j) - A_L \cdot \frac{h_{t+m}(j)^{1+\eta}}{1+\eta} \right],$$

中括号中的内容进一步调整为

$$\begin{aligned} & \left( \frac{v_{t+m} \cdot W_{t+m}}{\bar{w}_{t+m}} \right) \cdot \left[ \bar{w}_t \cdot \left( \frac{W_t^\#(j)}{W_t} \right) \cdot \mathcal{X}_{t,m} \right] \cdot \left[ \left( \frac{W_t^\#(j)}{W_t} \right) \cdot \left( \frac{\bar{w}_t}{\bar{w}_{t+m}} \right) \cdot \mathcal{X}_{t,m} \right]^{-\varepsilon_w} \cdot H_{t+m} \\ & - A_L \cdot \frac{1}{1+\eta} \cdot \left[ \left( \frac{W_t^\#(j)}{W_t} \right) \cdot \left( \frac{\bar{w}_t}{\bar{w}_{t+m}} \right) \cdot \mathcal{X}_{t,m} \right]^{-\varepsilon_w \cdot (1+\eta)} \cdot H_{t+m}^{1+\eta}. \end{aligned}$$

并且由式(3.139)、(3.142)得

$$\frac{v_{t+m} \cdot W_{t+m}}{\bar{w}_{t+m}} = v_{t+m} \cdot p_{t+m} \cdot z_{t+m}^+ = \psi_{z^+, t+m}.$$

可得调整后的最大化问题式

$$\begin{aligned} \max E_t \sum_{m=0}^{\infty} (\phi_w \cdot \beta)^m \cdot \{ & \psi_{z^+, t+m} \cdot \bar{w}_t^{1-\varepsilon_w} \cdot \bar{w}_{t+m}^{\varepsilon_m} \cdot \mathcal{X}_{t,m}^{1-\varepsilon_w} \cdot \left( \frac{W_t^\#(j)}{W_t} \right)^{1-\varepsilon_w} \cdot H_{i,t} \\ & - \frac{A_L}{1+\eta} \cdot \bar{w}_t^{-\varepsilon_w \cdot (1+\eta)} \cdot \bar{w}_{t+m} \cdot \mathcal{X}_{t,m}^{-\varepsilon_w \cdot (1+\eta)} \cdot \left( \frac{W_t^\#(j)}{W_t} \right)^{-\varepsilon_w \cdot (1+\eta)} \cdot H_{t+m}^{1+\eta} \}. \end{aligned} \quad (3.58)$$

设  $w_t^\#(j) \equiv W_t^\#(j)/W_t$ , 劳工联盟 FOCs wrt  $w_t^\#$ , 整理得

$$w_t^\#(j)^{1+\varepsilon_w \cdot \eta} = \frac{A_L}{\bar{w}_t} \cdot \frac{\varepsilon_w}{\varepsilon_w - 1} \cdot \frac{E_t \sum_{m=0}^{\infty} (\beta \cdot \phi_w)^m \left[ \left( \frac{\bar{w}_t}{\bar{w}_{t+m}} \cdot \mathcal{X}_{t,m} \right)^{-\varepsilon_w} \cdot H_{t+m} \right]^{1+\eta}}{E_t \sum_{m=0}^{\infty} (\beta \cdot \phi_w)^m \cdot \psi_{z^+, t+m} \cdot \left( \frac{\bar{w}_t}{\bar{w}_{t+m}} \cdot \mathcal{X}_{t,m} \right)^{-\varepsilon_w} \cdot H_{t+m} \cdot \mathcal{X}_{t,m}} \quad (3.59)$$

可见  $W_t^\#(j) = W_t^\# = W_t, \forall j$ , 劳工联盟工资策略的异质性特征得以消除。为了进一步简化, 定义两个辅助变量<sup>12</sup>

$$x_{1,t}^w \equiv E_t \sum_{m=0}^{\infty} (\beta \cdot \phi_w)^m \left[ \left( \frac{\bar{w}_t}{\bar{w}_{t+m}} \cdot \mathcal{X}_{t,m} \right)^{-\varepsilon_w} \cdot H_{t+m} \right]^{1+\eta}, \quad (3.60)$$

$$x_{2,t}^w \equiv E_t \sum_{m=0}^{\infty} (\beta \cdot \phi_w)^m \cdot \psi_{z^+, t+m} \cdot \left( \frac{\bar{w}_t}{\bar{w}_{t+m}} \cdot \mathcal{X}_{t,m} \right)^{-\varepsilon_w} \cdot H_{t+m} \cdot \mathcal{X}_{t,m}. \quad (3.61)$$

对  $x_{1,t}^w$  的迭代简化

$$\begin{aligned} x_{1,t}^w &= H_t^{1+\eta} + (\beta \cdot \phi_w) \cdot \left[ \left( \frac{\bar{w}_t}{\bar{w}_{t+1}} \cdot \mathcal{X}_{t,1} \right)^{-\varepsilon_w} \cdot H_{t+1} \right]^{1+\eta} + (\beta \cdot \phi_w)^2 \cdot \left[ \left( \frac{\bar{w}_t}{\bar{w}_{t+2}} \cdot \mathcal{X}_{t,2} \right)^{-\varepsilon_w} \cdot H_{t+2} \right]^{1+\eta} + \dots \\ &= H_t^{1+\eta} + E_t (\beta \cdot \phi_w) \cdot \left[ \frac{\bar{w}_t}{\bar{w}_{t+1}} \cdot \frac{(1+\pi)^{1-\kappa_w} \cdot (1+\pi_t)^{\kappa_w} \cdot \mu_{z^+}}{(1+\pi_{t+1}) \cdot \mu_{z^+, t+1}} \right]^{-\varepsilon_w \cdot (1+\eta)}. \\ &\quad \left\{ H_{t+1}^{1+\eta} + (\beta \cdot \phi_w) \cdot \left[ \left( \frac{\bar{w}_{t+1}}{\bar{w}_{t+2}} \cdot \frac{(1+\pi)^{1-\kappa_w} \cdot (1+\pi_{t+1})^{\kappa_w} \cdot \mu_{z^+}}{(1+\pi_{t+2}) \cdot \mu_{z^+, t+2}} \right)^{-\varepsilon_w \cdot (1+\eta)} \cdot H_{t+2}^{1+\eta} \right] + \dots \right\} \\ &= H_t^{1+\eta} + (\beta \cdot \phi_w) \cdot E_t \left( \frac{\bar{w}_t}{\bar{w}_{t+1}} \cdot \frac{(1+\pi)^{1-\kappa_w} \cdot (1+\pi_t)^{\kappa_w} \cdot \mu_{z^+}}{(1+\pi_{t+1}) \cdot \mu_{z^+, t+1}} \right)^{-\varepsilon_w \cdot (1+\eta)} \cdot x_{1,t+1}^w \\ &= H_t^{1+\eta} + (\beta \cdot \phi_w) \cdot E_t \left( \frac{1 + \tilde{\pi}_{w,t+1}}{1 + \pi_{w,t+1}} \right)^{-\varepsilon_w \cdot (1+\eta)} \cdot x_{1,t+1}^w, \end{aligned} \quad (3.62)$$

<sup>12</sup>也可以在此基础上, 作工资的 philips 曲线, 见第3.4节。

其中最后一个等号用到了由式(3.146)而得的衍生：

$$1 + \pi_{w,t+1} = \frac{W_{t+1}}{W_t} = \frac{\bar{w}_{t+1} \cdot P_{t+1} \cdot z_{t+1}^+}{\bar{w}_t \cdot P_t \cdot z_t^+} = \frac{\bar{w}_{t+1}}{\bar{w}_t} \cdot (1 + \pi_{t+1}) \cdot \mu_{z^+, t+1}. \quad (3.63)$$

对  $x_{2,t}^w$  的迭代简化

$$\begin{aligned} x_{2,t}^w &= \psi_{z^+, t} \cdot H_t + (\beta \cdot \phi_w) \psi_{z^+, t+1} \cdot \left( \frac{\bar{w}_t}{\bar{w}_{t+1}} \right)^{-\varepsilon_w} \cdot \mathcal{X}_{t,1}^{1-\varepsilon_w} \cdot H_{t+1} \\ &\quad + (\beta \cdot \phi_w)^2 \psi_{z^+, t+2} \cdot \left( \frac{\bar{w}_t}{\bar{w}_{t+2}} \right)^{-\varepsilon_w} \cdot \mathcal{X}_{t,2}^{1-\varepsilon_w} \cdot H_{t+2} + \dots \\ &= \psi_{z^+, t} \cdot H_t + (\beta \cdot \phi_w) \cdot \left( \frac{\bar{w}_t}{\bar{w}_{t+1}} \right)^{-\varepsilon_w} \cdot \left[ \frac{(1+\pi)^{1-\kappa_w} \cdot (1+\pi_t)^{\kappa_w} \cdot \mu_{z^+}}{(1+\pi_{t+1}) \cdot \mu_{z^+, t+1}} \right]^{1-\varepsilon_w} \cdot \\ &\quad \left\{ \psi_{z^+, t+1} \cdot H_{t+1} + (\beta \cdot \phi_w) \cdot \psi_{z^+, t+2} \cdot \left( \frac{\bar{w}_{t+1}}{\bar{w}_{t+2}} \right)^{-\varepsilon_w} \cdot \left( \frac{(1+\pi)^{1-\kappa_w} \cdot (1+\pi_{t+1})^{\kappa_w} \cdot \mu_{z^+}}{(1+\pi_{t+2}) \cdot \mu_{z^+, t+2}} \right)^{1-\varepsilon_w} \cdot H_{t+2} + \dots \right\} \\ &= \psi_{z^+, t} \cdot H_t + (\beta \cdot \phi_w) \cdot E_t \left( \frac{\bar{w}_{t+1}}{\bar{w}_t} \right) \cdot \left( \frac{1+\tilde{\pi}_{w,t+1}}{1+\pi_{w,t+1}} \right)^{1-\varepsilon_w} \cdot x_{2,t+1}^w \end{aligned} \quad (3.64)$$

其中最后一个等号用到了式(3.63)。

结合辅助变量式(3.62)、(3.64)，可得 repotimizing 劳工联盟的工资

$$w_t^\# = \left[ \frac{A_L}{\bar{w}_t} \cdot \frac{\varepsilon_w}{\varepsilon_w - 1} \cdot \frac{x_{1,t}^w}{x_{2,t}^w} \right]^{\frac{1}{1+\varepsilon_w \cdot \eta}} \quad (3.65)$$

另一方面，对 Aggregate wage index 式(3.44)作 calvo pricing

$$\begin{aligned} W_t^{1-\varepsilon_w} &= \int_0^1 W_t(j)^{1-\varepsilon_w} dj \\ &= \int_0^{1-\phi_w} [W_t(j)^\#]^{1-\varepsilon_w} dj + \int_{1-\phi_w}^1 [(1+\tilde{\pi}_{w,t}) \cdot W_{t-1}(j)]^{1-\varepsilon_w} dj \\ &= (1-\phi_w) \cdot \left( W_t^\# \right)^{1-\varepsilon_w} + \phi_w \cdot [(1+\tilde{\pi}_{w,t}) \cdot W_{t-1}]^{1-\varepsilon_w}, \end{aligned}$$

其中最后一个等号消除  $j^{th}$  劳工联盟工资定价的异质性特征，见式(3.59)。式两侧同时除以  $W_t^{1-\varepsilon_w}$ ，整理得

$$w_t^\# = \left[ \frac{1 - \phi_w \cdot \left( \frac{1+\tilde{\pi}_{w,t}}{1+\pi_{w,t}} \right)^{1-\varepsilon_w}}{1 - \phi_w} \right]^{\frac{1}{1-\varepsilon_w}}. \quad (3.66)$$

联立式(3.65)-(3.66)可得工资的决定

$$\frac{A_L}{\bar{w}_t} \cdot \frac{\varepsilon_w}{\varepsilon_w - 1} \frac{x_{1,t}^w}{x_{2,t}^w} = \left[ \frac{1 - \phi_w \cdot \left( \frac{1+\tilde{\pi}_{w,t}}{1+\pi_{w,t}} \right)^{1-\varepsilon_w}}{1 - \phi_w} \right]^{\frac{1+\varepsilon_w \cdot \phi_w}{1-\varepsilon_w}}. \quad (3.67)$$

### 3.2.5 wage dispersion index

在市场出清情况下，劳动力承包商所提供的全部同质劳动  $H_t$  和家庭部门提供的全部异质劳动时间  $h_t$  的关系为：

$$h_t \equiv \int_0^1 h_t(j) dj = H_t \cdot \int_0^1 \left( \frac{W_t(j)}{W_t} \right)^{-\varepsilon_w} \cdot dj,$$

其中用到了式(3.43)。整理可得

$$H_t = \frac{h_t}{\nu_t^w}, \quad (3.68)$$

其中 wage dispersion index  $\nu_t^w \geq 1$

$$\nu_t^w = \int_0^1 \left( \frac{W_t(j)}{W_t} \right)^{-\varepsilon_w} \cdot dj \quad (3.69)$$

反映劳动力市场上各类型  $j$  劳动力工资的差异程度：差异越大， $\nu_t^w$  值越高，同质的总劳动力产出越小；lower bound  $\nu_t^p = 1$  表明在没有 wage setting friction 时，所有  $j$  劳工联盟都会设定同样的工资。利用 Calvo pricing 可得

$$\begin{aligned} v_t^w &= \int_0^1 \left[ \frac{W_t(j)}{W_t} \right]^{-\varepsilon_w} dj \\ &= \int_0^{1-\phi_w} \cdot \left[ \frac{W_t^\#(j)}{W_t} \right]^{-\varepsilon_w} dj + \int_{1-\phi_w}^1 \cdot \left[ (1 + \tilde{\pi}_{w,t}) \cdot \frac{W_{t-1}(j)}{W_t} \right]^{-\varepsilon_w} dj \\ &= (1 - \phi_w) \cdot \left[ \frac{W_t^\#}{W_t} \right]^{-\varepsilon_w} + \phi_w \cdot (1 + \tilde{\pi}_{w,t})^{-\varepsilon_w} \cdot \int_0^1 \cdot \left[ \frac{W_{t-1}}{W_{t-1}} \cdot \frac{W_{t-1}}{W_t} \right]^{-\varepsilon_w} dj \\ &= (1 - \phi_w) \cdot (w_t^\#)^{-\varepsilon_w} + \phi_w \cdot \left( \frac{1 + \tilde{\pi}_{w,t}}{1 + \pi_{w,t}} \right)^{-\varepsilon_w} \cdot v_{t-1}^w \\ &= (1 - \phi_w) \cdot \left[ \frac{1 - \phi_w \cdot \left( \frac{1 + \tilde{\pi}_{w,t}}{1 + \pi_{w,t}} \right)^{1-\varepsilon_w}}{1 - \phi_w} \right]^{\frac{\varepsilon_w}{\varepsilon_w - 1}} + \phi_w \cdot \left( \frac{1 + \tilde{\pi}_{w,t}}{1 + \pi_{w,t}} \right)^{-\varepsilon_w} \cdot v_{t-1}^w \end{aligned} \quad (3.70)$$

### 3.2.6 家庭预算约束条件

家庭面临的预算约束条件为<sup>13</sup>

$$P_t \cdot \left( C_t + \frac{1}{\Psi_t} \cdot I_t \right) + B_{t+1} + P_t \cdot P_{k',t} \cdot \Delta_t \leq \int_0^1 W_t(j) \cdot h_t(j) dj + X_t^k \cdot \bar{K}_t + R_{t-1} \cdot B_t, \quad (3.71)$$

其中  $B_t$  为家庭部门购买的无风险债券。 $R_t$  表示  $t_1$  时刻购买的债券，在  $t$  时刻兑现，所对应的名义利率。

### 3.2.7 资本积累

假定家庭部门

- 是实物资本的所有者，
- 设定实物资本的利用率 (utilization rate)，
- 并进一步在竞争市场上将其租给 (中间产品) 生产者使用，收益为资本租金，成本为维护成本 (fast depreciation)。

当期资本存量尽管是由前期投资形成的，但家庭部门的当期经济决策仍可更集中/更少地使用已经形成的资本，投入到生产活动中去。决策依据取决于对当前经济形势的观察，和/或对未来的预期。实物资本存量积累式

$$\bar{K}_{t+1} = (1 - \delta) \cdot \bar{K}_t + F(I_t, I_{t-1}) + \Delta_t, \quad (3.72)$$

<sup>13</sup>膏按：上文所需要的等式。这一章节还需要补充相关文字。

其中折旧率为常数  $\delta^{14}$ 。

将式(3.36)带回式(3.72)，资本积累式可改写为

$$\bar{K}_{t+1} = (1 - \delta) \cdot \bar{K}_t + \left[ 1 - S \cdot \left( \frac{I_t}{I_{t-1}} \right) \right] \cdot I_t + \Delta_t. \quad (3.73)$$

$\Delta_t$  表示该家庭从其他家庭购买的“净”实物资本，均衡条件下  $\Delta_t = 0$ ，模型设定中保留  $\Delta_t$  以便于测算实物资本存量的价值。 $\Delta_t$  的市场价格为  $P_t \cdot P_{k',t}$ 。

利用均衡条件下  $\Delta_t = 0$ ，以及式，将(3.133)、(3.135)，对上式作 scale

$$\bar{k}_{t+1} = \frac{(1 - \delta) \cdot \bar{k}_t}{\mu_{z^+,t} \cdot \mu_{\Psi,t}} + \left[ 1 - s \cdot \left( \frac{i_t}{i_{t-1}} \cdot \mu_{z^+,t} \cdot \mu_{\Psi,t} \right) \right] \cdot i_t \quad (3.74)$$

### 资本净收益

$t$  时期的投资带来  $t+1$  时期的实物资本积累增加； $\bar{K}_{t+1}$  每增加 1 单位，给家庭带来的净收益 (net cash payment)  $X_{t+1}^k$  为

$$X_{t+1}^k = u_{t+1} \cdot P_{t+1} \cdot r_{t+1}^k - \frac{P_{t+1}}{\psi_{t+1}} \cdot a(u_{t+1}), \quad (3.75)$$

RHS 前半部分为考虑到资本利用率之后的名义净 rental 收入（扣除掉折旧）；后半部分为使用成本 (capital utilization cost)。 $P_{t+1}/\Psi_{t+1}$  为同质投资品  $I_t$  的均衡市场价格 (名义量)，由式(3.37)-(3.38)给出。

### 3.2.8 家庭部门最大化问题

家庭部门问题可以表示为：选择投入组合  $\{C_t, I_t, \Delta_t, B_{t+1}, \bar{K}_{t+1}, u_t\}$ ，基于给定的预算约束式(3.71)、资本积累式(3.73)和资本净收益式(3.75)，来追求式(3.45)效用函数最大化。

$$\begin{aligned} \max_{\{C_t, \Delta_t, I_t, \bar{K}_{t+1}, B_{t+1}, u_t\}} \mathcal{L} &= E_0 \sum_{t=0}^{\infty} \beta \cdot \{ \ln(C_t - b \cdot C_{t-1}) - A_L \cdot \frac{\int_0^1 h_t(j)^{1+\eta} dj}{1+\eta} \\ &+ v_t \cdot \left[ \int_0^1 W_t(j) \cdot h_t(j) dj + X_{t+1}^k \cdot \bar{K}_t + R_{t-1} \cdot B_t - P_t \cdot C_t - P_t \cdot \frac{I_t}{\Psi_t} - B_{t+1} - P_t \cdot P_{k',t} \cdot \Delta_t \right] \\ &+ \omega_t \cdot \left[ \Delta_t + (1 - \delta) \cdot \bar{K}_t + \left[ 1 - S \left( \frac{I_t}{I_{t-1}} \right) \right] \cdot I_t - \bar{K}_{t+1} \right] \\ &+ \tau_t \cdot \left[ u_{t+1} \cdot P_{t+1} \cdot r_{t+1}^k - \frac{P_{t+1}}{\psi_{t+1}} \cdot a(u_{t+1}) - X_{t+1}^k \right] \}. \end{aligned}$$

取消异质性  $j \in (0, 1)$  特征。依次求解 FOCs，为表述简便，在不产生歧义的情况下省略式中的期望符号。

#### FOC wrt C

$$\frac{\partial \mathcal{L}}{\partial C_t} = 0 \Rightarrow \frac{1}{C_t - b \cdot C_{t-1}} - \frac{\beta \cdot b}{C_{t+1} - b \cdot C_t} = v_t \cdot P_t, \quad (3.76)$$

代入式(3.5)、(3.136)、(3.132)、(3.142)，进一步整理得

$$\frac{1}{c_t - \frac{b}{\mu_{z^+,t}} \cdot c_{t-1}} - \frac{\beta \cdot b}{\mu_{z^+,t+1} \cdot c_{t+1} - b \cdot c_t} = \psi_{z^+,t}. \quad (3.77)$$

<sup>14</sup> 脊按：也有将折旧率视为与  $u_t$  有关的变量， $\delta(u_t)$ ，补充 reference。

FOC wrt  $\Delta_t$

$$\frac{\partial \mathcal{L}}{\partial \Delta_t} = 0 \Rightarrow P_t \cdot P_{k',t} = \frac{\omega_t}{v_t}, \quad (3.78)$$

其中 RHS 两个影子价格之比反映了 Tobin's q, 见第3.C节。

FOC wrt I

$$\begin{aligned} \frac{\partial \omega_t \cdot [1 - S\left(\frac{I_t}{I_{t-1}}\right)] \cdot I_t}{\partial I_t} &= \frac{\partial \omega_t \cdot I_t}{\partial I_t} - \frac{[\partial \omega_t \cdot I_t \cdot S\left(\frac{I_t}{I_{t-1}}\right)]}{I_t} \\ &= \omega_t - \omega_t \cdot \frac{I_t}{I_{t-1}} \cdot S'\left(\frac{I_t}{I_{t-1}}\right) - \omega_t \cdot S\left(\frac{I_t}{I_{t-1}}\right) - \frac{\partial \beta \cdot E_t \left\{ \omega_{t+1} \cdot I_{t+1} \cdot S\left(\frac{I_{t+1}}{I_t}\right) \right\}}{\partial I_t} \\ &= \omega_t - \omega_t \cdot \frac{I_t}{I_{t-1}} \cdot S'\left(\frac{I_t}{I_{t-1}}\right) - \omega_t \cdot S\left(\frac{I_t}{I_{t-1}}\right) - \beta \cdot E_t \left\{ \omega_{t+1} \cdot I_{t+1} \cdot S'\left(\frac{I_{t+1}}{I_t}\right) \cdot (-1) \cdot \frac{I_{t+1}}{I_t^2} \right\} \\ &= \omega_t \cdot \left[ 1 - S\left(\frac{I_t}{I_{t-1}}\right) - \frac{I_t}{I_{t-1}} \cdot S'\left(\frac{I_t}{I_{t-1}}\right) \right] + \beta \cdot E_t \left\{ \omega_{t+1} \cdot \left(\frac{I_{t+1}}{I_t}\right)^2 \cdot S'\left(\frac{I_{t+1}}{I_t}\right) \right\}. \end{aligned} \quad (3.79)$$

式(3.142)、(3.141)代入式(3.79)得

$$\omega_t = (v_t \cdot P_t) \cdot (P_{k',t}) = \left( \frac{\psi_{z^+,t}}{z_t^+} \right) \cdot \left( \frac{p_{k',t}}{\Psi_t} \right) \quad (3.80)$$

此外由式(3.135)得

$$\frac{i_t}{i_{t-1}} = \frac{i_t \cdot z_t^+ \cdot \psi_t}{i_{t-1} \cdot z_{t-1}^+ \cdot \psi_{t-1}} = \frac{i_t}{i_{t-1}} \cdot \mu_{z^+,t} \cdot \mu_{\Psi,t}. \quad (3.81)$$

FOC wrt  $I_t$

$$\frac{\partial \mathcal{L}}{\partial I_t} = 0 \Rightarrow \frac{v_t \cdot P_t}{\psi_t} = \frac{\partial \omega_t \cdot [1 - S\left(\frac{I_t}{I_{t-1}}\right)] \cdot I_t}{\partial I_t}, \quad (3.82)$$

LHS 代入式(3.142), RHS 代入式(3.79)、(3.80)得

$$\begin{aligned} \frac{\psi_{z^+,t}}{z_t^+ \cdot \psi_t} &= \frac{\psi_{z^+,t} \cdot p_{k',t}}{z_t^+ \cdot \psi_t} \cdot \mathcal{A} + \beta \cdot E_t \frac{\psi_{z^+,t+1} \cdot p_{k',t+1}}{z_{t+1}^+ \cdot \psi_{t+1}} \cdot \mathcal{B}, \\ \mathcal{A} &\equiv 1 - S\left(\mu_{z^+,t} \cdot \mu_{\psi,t} \cdot \frac{i_t}{i_{t-1}}\right) - S'\left(\mu_{z^+,t} \cdot \mu_{\psi,t} \cdot \frac{i_t}{i_{t-1}}\right) \cdot \left(\mu_{z^+,t} \cdot \mu_{\psi,t} \cdot \frac{i_t}{i_{t-1}}\right), \\ \mathcal{B} &\equiv S'\left(\mu_{z^+,t+1} \cdot \mu_{\psi,t+1} \cdot \frac{i_{t+1}}{i_t}\right) \cdot \left(\mu_{z^+,t+1} \cdot \mu_{\psi,t+1}\right) \cdot \left(\frac{i_{t+1}}{i_t}\right)^2, \end{aligned} \quad (3.83)$$

等式两侧同时乘以  $z_t^+ \cdot \psi_t$ , 整理得

$$\begin{aligned} \psi_{z^+,t} &= \psi_{z^+,t} \cdot p_{k',t} \cdot \left[ 1 - S\left(\mu_{z^+,t} \cdot \mu_{\psi,t} \cdot \frac{i_t}{i_{t-1}}\right) - S'\left(\mu_{z^+,t} \cdot \mu_{\psi,t} \cdot \frac{i_t}{i_{t-1}}\right) \cdot \left(\mu_{z^+,t} \cdot \mu_{\psi,t} \cdot \frac{i_t}{i_{t-1}}\right) \right] \\ &\quad + \beta \cdot E_t \left[ (\psi_{z^+,t+1} \cdot p_{k',t+1}) \cdot (\mu_{z^+,t+1} \cdot \mu_{\psi,t+1}) \cdot S'\left(\mu_{z^+,t+1} \cdot \mu_{\psi,t+1} \cdot \frac{i_{t+1}}{i_t}\right) \cdot \left(\frac{i_{t+1}}{i_t}\right)^2 \right]. \end{aligned} \quad (3.84)$$

## FOC wrt K

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \bar{K}_{t+1}} = 0 \Rightarrow \omega_t &= \frac{\partial v_t \cdot X_t^k \cdot \bar{K}_t}{\partial \bar{K}_{t+1}} + \frac{\partial \omega_t \cdot (1-\delta) \cdot \bar{K}_t}{\partial \bar{K}_{t+1}} \\ &= \beta \cdot E_t X_{t+1}^k \cdot v_{t+1} + \beta \cdot (1-\delta) \cdot E_t \omega_{t+1},\end{aligned}\quad (3.85)$$

代入式(3.80)以消去  $\omega_t$ , 整理得

$$\begin{aligned}v_t &= \beta \cdot E_t v_{t+1} \cdot \frac{X_{t+1}^k + (1-\delta) \cdot P_{t+1} \cdot P_{k',t+1}}{P_t \cdot P_{k',t}} \\ &= \beta \cdot E_t \cdot v_{t+1} \cdot R_{t+1}^k \\ \text{其中定义 } R_{t+1}^k &\equiv \frac{X_{t+1}^k + (1-\delta) \cdot P_{t+1} \cdot P_{k',t+1}}{P_t \cdot P_{k',t}}\end{aligned}\quad (3.86)$$

$R_t^k$  表示资本的回报率 (rate of return on capital)。

对式(3.86)两侧同时乘以  $P_t \cdot z_t^+$

$$\begin{aligned}\psi_{z^+,t} &= \beta \cdot E_t \frac{v_{t+1} \cdot P_{t+1} \cdot z_{t+1}^+}{\frac{P_{t+1}}{P_t} \cdot \frac{z_{t+1}^+}{z_t^+}} \cdot R_{t+1}^k \\ &= \beta \cdot E_t \cdot \frac{\psi_{z^+,t+1}}{(1+\pi_{t+1}) \cdot \mu_{z^+,t+1}} \cdot R_{t+1}^k\end{aligned}\quad (3.87)$$

将 net cash payment  $X_{t+1}^k$  式(3.75)代入资本回报率  $R_{t+1}^k$  定义式(3.86), 并利用(3.140)、(3.141)改写为 scaled variables 形式

$$\begin{aligned}R_{t+1}^k &= \frac{u_{t+1} \cdot \frac{P_{t+1}}{P_t} \cdot r_{t+1}^k}{P_{k',t}} - \frac{\frac{P_{t+1}}{P_t} \cdot a(u_{t+1})}{P_{k',t} \cdot \psi_{t+1}} + \frac{P_{t+1}}{P_t} \cdot \frac{P_{k',t+1}}{P_{k',t}} \cdot (1-\delta) \\ &= \frac{1+\pi_{t+1}}{\mu_{\psi,t+1}} \cdot \frac{u_{t+1} \cdot \bar{r}_{t+1}^k - (1+\pi_{t+1}) \cdot a(u_{t+1}) + p_{k',t+1} \cdot (1-\delta)}{p_{k',t}}\end{aligned}\quad (3.88)$$

## FOC wrt u

家庭选择 capital utilization rate  $u_t$ , 使得同时段的 net cash payments of physical capital  $X_t^k$  最大化, 这表现为家庭问题中的静态比较。

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial u_t} = 0 \Rightarrow P_t \cdot r_t^k &= \frac{P_t}{\Psi_t} \cdot a'(u_t), \\ a'(u_t) &= r_t^k \cdot \Psi_t = \bar{r}_t^k.\end{aligned}\quad (3.89)$$

## FOC wrt B

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial B_{t+1}} = 0 \Rightarrow \frac{\partial [\beta \cdot E_t v_{t+1} \cdot R_t \cdot B_{t+1}]}{\partial B_{t+1}} - \frac{\partial v_t \cdot B_{t+1}}{\partial B_{t+1}} &= 0, \\ v_t &= \beta \cdot E_t v_{t+1} \cdot R_t,\end{aligned}\quad (3.90)$$

代入 scale 式(3.142)

$$\frac{\psi_{z^+,t}}{P_t \cdot z_t^+} = \beta \cdot \frac{\psi_{z^+,t+1}}{P_{t+1} \cdot z_{t+1}^+} \cdot R_t,$$

整理得

$$\psi_{z^+,t} = \beta \cdot E_t \frac{\psi_{z^+,t+1}}{\mu_{z^+,t+1} \cdot (1+\pi_{t+1})} \cdot R_t. \quad (3.91)$$

### 3.3 Price Philips Curve

#### 3.3.1 price dispersion index 的线性近似

式(3.33)  $\Rightarrow$

$$\begin{aligned} \ln \nu_t^f &= \ln \left[ (1 - \phi_f) \cdot (1 + \pi_t)^{\varepsilon_f} \cdot \left(1 + \pi_t^\# \right)^{-\varepsilon_f} + \phi_f \cdot (1 + \pi_t)^{\varepsilon_f} \cdot \nu_{t-1}^f \right] \\ \frac{\nu_t^f - \nu^f}{\nu^f} &= \varepsilon_f \cdot (1 - \phi_f) \cdot (1 + \pi)^{\varepsilon_f - 1} \cdot \left(1 + \pi^\# \right)^{-\varepsilon_f} \cdot [\pi_t - \pi] \\ &\quad + (-\varepsilon_f) \cdot (1 - \phi_f) \cdot (1 + \pi)^{\varepsilon_f} \cdot \left(1 + \pi^\# \right)^{-\varepsilon_f - 1} \cdot [\pi_t^\# - \pi^\#] \\ &\quad + \varepsilon_f \cdot \phi_f \cdot (1 + \pi)^{\varepsilon_f - 1} \cdot \nu^f \cdot [\pi_t - \pi], \end{aligned}$$

沿着  $\pi = 0, \pi^\# = 0, \nu^f = 1$  作线性近似

$$\hat{\nu}_t^f = \varepsilon_f \cdot \hat{\pi}_t - \varepsilon_f \cdot (1 - \phi_f) \cdot \hat{\pi}_t^\# + \phi_f \cdot \hat{\nu}_{t-1}. \quad (3.92)$$

#### 3.3.2 两个辅助变量的线性近似

对辅助变量  $x_{1,t}^f$  作线性近似。已知  $U_{c,t} = 1/C_t$ <sup>15</sup>, 式(3.23)  $\Rightarrow$

$$\begin{aligned} \frac{x_{1,t}^f - x_1^f}{x_1} &= \frac{-\frac{Y}{C^2} \cdot s \cdot (C_t - C)}{x_1^f} + \frac{\frac{s}{C} \cdot (Y_t - Y)}{x_1^f} + \frac{\frac{Y}{C} \cdot (s_t - s)}{x_1^f} \\ &\quad + \beta \cdot \phi_f \cdot (1 + \pi)^{\varepsilon_f - 1} \cdot \varepsilon_f \cdot (\pi_{t+1} - \pi) \\ &\quad + \beta \cdot \phi_f \cdot (1 + \pi)^{\varepsilon_f} \cdot \frac{x_1^f - x_1^f}{x_1^f} \end{aligned} \quad (3.93)$$

稳定状态下我们有  $\pi = \pi^\# = 0$ , 因而稳态  $x_1^f$  的值由式(3.23)给出

$$x_1^f = \frac{1}{1 - \beta \cdot \phi_f} \cdot \frac{Y \cdot s}{C}. \quad (3.94)$$

式(3.94)带回式(3.93), 进一步整理得

$$\hat{x}_{1,t}^f = (1 - \beta \cdot \phi_f) \cdot \left( \hat{Y}_t + \hat{s}_t - \hat{C}_t \right) + \beta \cdot \phi_f \cdot E_t \left( \varepsilon_f \cdot \hat{\pi}_{t+1} + \hat{x}_{1,t+1}^f \right). \quad (3.95)$$

对辅助变量  $x_{2,t}^f$  作线性近似。式(3.24)  $\Rightarrow$

$$\begin{aligned} \frac{x_{2,t}^f - x_2^f}{x_2} &= \frac{-\frac{Y}{C^2} \cdot (C_t - C)}{x_2^f} + \frac{\frac{1}{C} \cdot (Y_t - Y)}{x_2^f} + \beta \cdot \phi_f \cdot (\varepsilon_f - 1) \cdot (1 + \pi)^{\varepsilon_f - 1} \cdot (\pi_{t+1} - \pi) \\ &\quad + \beta \cdot \phi_f \cdot (1 + \pi)^{\varepsilon_f - 1} \cdot \frac{x_{2,t+1}^f - x_2^f}{x_2^f}, \end{aligned} \quad (3.96)$$

进一步整理得

$$\hat{x}_{2,t}^f = (1 - \beta \cdot \phi_f) \cdot \left( \hat{Y}_t - \hat{C}_t \right) + \beta \cdot \phi_f \cdot E_t \left[ (\varepsilon_f - 1) \cdot \hat{\pi}_{t+1} + \hat{x}_{2,t+1}^f \right]. \quad (3.97)$$

<sup>15</sup> 脊按：在写完家庭部门优化条件 FOC 之后，把  $U_{c,t}$  对应的 FOC 的连接写进去。

### 3.3.3 reset price inflation 的线性近似

式(3.28)  $\Rightarrow$

$$(1 - \varepsilon_f) \cdot \ln(1 + \pi_t) = \ln \left[ (1 - \phi_f) \cdot \left(1 - \pi_t^\# \right)^{1-\varepsilon_f} + \phi_f \right],$$

$$(1 - \varepsilon_f) \cdot (\pi_t - \pi) = \frac{(1 - \phi_f) \cdot (1 - \varepsilon_w) \cdot (1 + \pi^\#)^{-\varepsilon_w} \cdot (\pi_t^\# - \pi^\#)}{(1 - \phi_f) \cdot \left(1 - \pi_t^\# \right)^{1-\varepsilon_f} + \phi_f},$$

由此可得 reset price inflation 的线性近似式

$$\hat{\pi}_t = (1 - \phi_f) \cdot \frac{(1 + \pi^\#)^{-\varepsilon}}{(1 + \pi)^{1-\varepsilon}} \cdot \hat{\pi}_t^\# \approx (1 - \phi_f) \cdot \hat{\pi}_t^\# \quad (3.98)$$

### 3.3.4 Price Philips Curve

式(3.27)  $\Rightarrow$

$$\ln \left(1 + \pi_t^\# \right) - \ln(1 + \pi_t) = \ln \left( \frac{\varepsilon_f}{\varepsilon_f - 1} \right) + \ln x_{1,t}^f - \ln x_{2,t}^f,$$

$$\begin{aligned} \hat{\pi}_t^\# - \hat{\pi}_t &= \hat{x}_{1,t}^f - \hat{x}_{2,t}^f \\ &= (1 - \beta \cdot \phi_f) \cdot \hat{s}_t + \beta \cdot \phi_f \cdot E_t \hat{\pi}_{t+1} + \beta \cdot \phi_f \cdot E_t \left( \hat{x}_{1,t+1}^f - \hat{x}_{2,t+1}^f \right), \end{aligned} \quad (3.99)$$

将式(3.98)的代入上式替代  $\hat{\pi}_t^\#$ , 得到 price philips curve

$$\begin{aligned} \hat{\pi}_t &= \left( \frac{1 - \phi_f}{\phi_f} \right) \cdot (1 - \beta \cdot \phi_f) \cdot \hat{s}_t + \beta \cdot E_t \hat{\pi}_{t+1} \\ &= \left( \frac{1 - \phi_f}{\phi_f} \right) \cdot (1 - \beta \cdot \phi_f) \cdot \left[ \alpha \cdot (1 + \phi_f) \cdot \hat{X}_t + \hat{R}_t \right] + \beta \cdot E_t \hat{\pi}_{t+1} \end{aligned}$$

其中  $\hat{s}_t$  由式(3.18)作近似线性测得;  $X_t$  表示 output gap,  $\hat{s}_t$  和  $\hat{X}_t$  的关系参见基准 NK 模型相关部分<sup>16</sup>。

据此得到 price philips curve

$$\begin{aligned} \hat{\pi}_t &= \kappa_f \cdot \left[ \alpha \cdot (1 + \eta) \cdot \hat{X}_t + \hat{R}_t \right] + \beta \cdot E_t \hat{\pi}_{t+1}, \\ \text{其中定义系数 } \kappa_f &\equiv \frac{1 - \phi_f}{\phi_f} \cdot (1 - \beta \cdot \phi_f). \end{aligned} \quad (3.100)$$

## 3.4 Wage Philips Curve

式(3.59)的最优化问题可以调整为

$$\left[ \frac{1}{1 - \varepsilon_w} \cdot \frac{1}{w_t^\#} \right] \cdot E_t \sum_{m=0}^{\infty} \cdot (\beta \cdot \phi_w)^m \cdot \psi_{z^+, t+m} \cdot h_{t+m}^t \cdot \left[ w_t^\# \cdot \bar{w}_t \cdot \mathcal{X}_{t,m} - \frac{\varepsilon_w}{\varepsilon_w - 1} \cdot \left( \frac{A_L \cdot (h_{t+m}^t)^\eta}{\psi_{z^+, t+m}} \right) \right], \quad (3.101)$$

或者进一步变为

$$E_t \sum_{m=0}^{\infty} \cdot (\beta \cdot \phi_w)^m \cdot \psi_{z^+, t+m} \cdot h_{t+m}^t \cdot \left[ w_t^\# \cdot \bar{w}_t \cdot \mathcal{X}_{t,m} - \frac{\varepsilon_w}{\varepsilon_w - 1} \cdot MRS_{t+m}^t \right], \quad (3.102)$$

<sup>16</sup>膏按：这部分还没看明白，需要进一步搞清楚。

其中

$$MRS_{t+m}^t \equiv \frac{A_L \cdot (h_{t+m}^t)^\eta}{\psi_{z^+, t+m}} \quad (3.103)$$

表示边际劳动成本，上角标和下角标表示在  $t$  时间调整工资，在随后的  $t+1, \dots, t+m$  时间均未调整。稳定状态下 forward-looking 的劳工联盟最优策略为，将工资设为等于提供额外 1 单位劳动力（劳动时间）的边际成本乘以 price markup，即使式(3.102)中中括号内的部分等于零；线性展开见下文式(3.129)。

### 3.4.1 稳定状态的描述

稳定状态下， $w^\# = \mathcal{X} = 1$ ,  $\bar{w} = \frac{\varepsilon_w}{\varepsilon_w - 1} \cdot MRS$ ,  $1 + \tilde{\pi} = (1 + \pi) \cdot \mu_{z^+}$ 。

### 3.4.2 wage inflation

对  $(1 + \tilde{\pi}_{w,t+1})$  作线性近似。式(3.47)  $\Rightarrow$

$$\ln(1 + \tilde{\pi}_{w,t+1}) = \kappa_w \cdot \ln(1 + \pi_t) + (1 - \kappa_w) \cdot \ln(1 + \pi_t) + \ln \mu_{z^+},$$

$$\hat{\tilde{\pi}}_{w,t+1} \approx \tilde{\pi}_{w,t+1} - \tilde{\pi}_w \approx \kappa_w \cdot (\pi_t - \pi) = \kappa_w \cdot \hat{\pi}_t \quad (3.104)$$

$$\tilde{\pi}_w = \pi \cdot \mu_{z^+} \quad (3.105)$$

### 3.4.3 辅助变量

对  $\hat{\mathcal{X}}_{t,m}$  作线性近似。式(3.56)  $\Rightarrow$

$$\begin{aligned} \hat{\mathcal{X}}_{t,m \neq 0} &\approx \left[ \left( \hat{\tilde{\pi}}_{w,t+1} + \hat{\tilde{\pi}}_{w,t+2} + \dots + \hat{\tilde{\pi}}_{w,t+m} \right) - \left( \hat{\pi}_{t+1} + \hat{\pi}_{t+2} + \dots + \hat{\pi}_{t+m} \right) \right] \\ &\quad - (\hat{\mu}_{z^+,t+1} + \hat{\mu}_{z^+,t+2} + \dots + \hat{\mu}_{z^+,t+m}) \\ &= - [(\hat{\pi}_{t+1} - \kappa_w \cdot \hat{\pi}_t) + (\hat{\pi}_{t+2} - \kappa_w \cdot \hat{\pi}_{t+1}) + \dots + (\hat{\pi}_{t+m} - \kappa_w \cdot \hat{\pi}_{t+m-1})] \\ &\quad - (\hat{\mu}_{z^+,t+1} + \hat{\mu}_{z^+,t+2} + \dots + \hat{\mu}_{z^+,t+m}) \\ &\equiv - (\Delta \kappa_w \pi_{t+1} + \Delta \kappa_w \pi_{t+2} + \dots + \Delta \kappa_w \pi_{t+m}) - (\hat{\mu}_{z^+,t+1} + \hat{\mu}_{z^+,t+2} + \dots + \hat{\mu}_{z^+,t+m}) \end{aligned} \quad (3.106)$$

以及

$$\hat{\mathcal{X}}_{t,0} = 0, \quad (3.107)$$

其中定义

$$\Delta_{\kappa_w} \pi_{w,t+1} \equiv \hat{\pi}_{w,t+1}^\# - \kappa_w \cdot \hat{\pi}_t, \quad (3.108)$$

$$\Delta_{\kappa_w} \pi_{t+1} \equiv \hat{\pi}_{t+1} - \kappa_w \cdot \hat{\pi}_t. \quad (3.109)$$

### 3.4.4 劳动力供应

对  $h_{t+m}$  式(3.57)作线性近似。 $\Rightarrow$

$$\begin{aligned}\frac{\bar{w}_t}{\bar{w}_{t+m}} &= \frac{\frac{W_t}{P_t \cdot z_t^+}}{\frac{W_{t+m}}{P_{t+m} \cdot z_{t+m}^+}} \\ &= \frac{W_t}{W_{t+m}} \cdot \frac{P_{t+m}}{P_t} \cdot \frac{z_{t+m}^+}{z_t^+} \\ &= \frac{(1 + \pi_{t+1}) \cdot (1 + \pi_{t+2}) \cdots (1 + \pi_{t+m}) \cdot (\mu_{z^+, t+1} \cdot \mu_{z^+, t+2} \cdots \mu_{z^+, t+m})}{(1 + \tilde{\pi}_{w, t+1}) \cdot (1 + \tilde{\pi}_{w, t+2}) \cdots (1 + \tilde{\pi}_{w, t+m})},\end{aligned}\quad (3.110)$$

其中最后一个等号用到了式(3.47)。结合式(3.105)，对上式线性展开，整理得

$$\left( \frac{\hat{w}_t}{\bar{w}_{t+m}} \right) \approx \hat{\mu}_{z^+, t+1} \cdot \hat{\mu}_{z^+, t+2} \cdots \cdot \hat{\mu}_{z^+, t+m}. \quad (3.111)$$

将式(3.111)代入式(3.57)  $\Rightarrow$

$$\hat{h}_{t+m}^t - \hat{H}_{t+m} = \begin{cases} -\varepsilon_w \cdot [\hat{w}_t^\# - (\Delta \kappa_w \pi_{t+1} + \Delta \kappa_w \pi_{t+2} + \dots + \Delta \kappa_w \pi_{t+m})], & \text{当 } m > 0, \\ -\varepsilon_w \cdot \hat{w}_t^\#, & \text{当 } m = 0. \end{cases} \quad (3.112)$$

### 3.4.5 边际劳动成本

对  $RMS_{t+m}^t$  作线性近似。式(3.103)  $\Rightarrow$

$$\begin{aligned}\ln MRS_{t+m}^t &= \ln A_L + \eta \cdot \ln_{t+m}^t - \ln \Psi_{z^+, t+m}, \\ MRS_{t+m}^t &= -\hat{\Psi}_{z^+, t+m} + \eta \cdot (\hat{h}_{t+m}^t - \hat{H}_{t+m}) + \eta \cdot \hat{H}_{t+m}.\end{aligned}\quad (3.113)$$

### 3.4.6 wage price inflation

式(3.66)反映了，aggregate wage index  $W_t$  对劳工联盟制定工资指数  $W_t^\#$  的限制。调整得

$$1 = (1 - \phi_w) \cdot w_t^{\#, 1-\varepsilon_w} + \phi_w \cdot \left( \frac{1 + \pi_{w,t}^\#}{1 + \pi_{w,t}} \right)^{1-\varepsilon_w},$$

两侧取  $\ln$  后作近似线性展开

$$\begin{aligned}0 &= \frac{(1 - \phi_w) \cdot (1 - \varepsilon_w) \cdot w^{\#, -\varepsilon_w} \cdot (w_t^\# - w^\#)}{(1 - \phi_w) \cdot w^{\#, 1-\varepsilon_w} + \phi_w \cdot \left( \frac{\pi_w^\#}{\pi_w} \right)^{1-\varepsilon_w}} \\ &\quad + \frac{\phi_w \cdot (1 - \varepsilon_w) \cdot \left( \frac{\pi_w^\#}{\pi_w} \right)^{1-\varepsilon_w} \cdot \pi_w^{\#, -1} \cdot (\pi_{w,t}^\# - \pi_w^\#)}{(1 - \phi_w) \cdot w^{\#, 1-\varepsilon_w} + \phi_w \cdot \left( \frac{\pi_w^\#}{\pi_w} \right)^{1-\varepsilon_w}} \\ &\quad + \frac{\phi_w \cdot (\varepsilon_w - 1) \cdot \left( \frac{\pi_w^\#}{\pi_w} \right)^{1-\varepsilon_w} \cdot \pi_w^{-1} \cdot (\pi_{w,t} - \pi_w)}{(1 - \phi_w) \cdot w^{\#, 1-\varepsilon_w} + \phi_w \cdot \left( \frac{\pi_w^\#}{\pi_w} \right)^{1-\varepsilon_w}} \\ &= \frac{(1 - \varepsilon_w) \cdot \left[ (1 - \phi_w) \cdot w^{\#, 1-\varepsilon_w} \cdot \hat{w}_t^\# + \phi_w \cdot \left( \frac{\pi_w^\#}{\pi_w} \right)^{1-\varepsilon_w} \cdot (\hat{\pi}_{w,t}^\# - \hat{\pi}_{w,t}) \right]}{(1 - \phi_w) \cdot w^{\#, 1-\varepsilon_w} + \phi_w \cdot \left( \frac{\pi_w^\#}{\pi_w} \right)^{1-\varepsilon_w}}.\end{aligned}\quad (3.114)$$

稳定状态下,  $w^\# = 1$ ,  $\pi_w^\#/ \pi_w = 1$ , 且根据式(3.108)得

$$\hat{\pi}_{w,t}^\# - \hat{\pi}_{w,t} = -(\hat{\pi}_{w,t} - \kappa_w \cdot \hat{\pi}_{t-1}) = -\Delta_{\kappa_w} \hat{\pi}_{w,t}.$$

因此式(3.114)可以进一步改写为如下 wage price inflation index

$$\hat{w}_t^\# = \frac{\varepsilon_w}{1 - \varepsilon_w} \cdot \Delta_{\kappa_w} \hat{\pi}_{w,t}. \quad (3.115)$$

### 3.4.7 其他辅助变量

由(3.112)可得

$$\begin{aligned} & \sum_{m=0}^{\infty} \eta \cdot (\hat{h}_{t+m}^t - \hat{H}_{t+m}) \\ &= -\varepsilon_w \cdot \eta \cdot \hat{w}_t^\# \\ &\quad - (\beta \cdot \phi_w) \cdot \varepsilon_w \cdot \eta \cdot [\hat{w}_t^\# - \Delta_{\kappa_w} \pi_{t+1}] \\ &\quad - (\beta \cdot \phi_w)^2 \cdot \varepsilon_w \cdot \eta \cdot [\hat{w}_t^\# - \Delta_{\kappa_w} \pi_{t+1} - \Delta_{\kappa_w} \pi_{t+2}] \\ &\quad - (\beta \cdot \phi_w)^3 \cdot \varepsilon_w \cdot \eta \cdot [\hat{w}_t^\# - \Delta_{\kappa_w} \pi_{t+1} - \Delta_{\kappa_w} \pi_{t+2} - \Delta_{\kappa_w} \pi_{t+3}] \dots \\ &= -\varepsilon_w \cdot \eta \cdot [1 + (\beta \cdot \phi_w) + (\beta \cdot \phi_w)^2 + \dots + (\beta \cdot \phi_w)^m] \cdot \hat{w}_t^\# \\ &\quad + \varepsilon_w \cdot \eta \cdot [(\beta \cdot \phi_w) + (\beta \cdot \phi_w)^2 + \dots + (\beta \cdot \phi_w)^m] \cdot \Delta_{\kappa_w} \pi_{t+1} \\ &\quad + \varepsilon_w \cdot \eta \cdot [(\beta \cdot \phi_w)^2 + (\beta \cdot \phi_w)^3 + \dots + (\beta \cdot \phi_w)^m] \cdot \Delta_{\kappa_w} \pi_{t+2} \\ &\quad + \varepsilon_w \cdot \eta \cdot [(\beta \cdot \phi_w)^3 + (\beta \cdot \phi_w)^4 + \dots + (\beta \cdot \phi_w)^m] \cdot \Delta_{\kappa_w} \pi_{t+3} \\ &\quad + \dots \\ &\quad + \varepsilon_w \cdot \eta \cdot [(\beta \cdot \phi_w)^m] \cdot \Delta_{\kappa_w} \pi_{t+m} \\ &\approx -\varepsilon_w \cdot \eta \cdot (\beta \cdot \phi_w)^0 \cdot [1 + (\beta \cdot \phi_w) + (\beta \cdot \phi_w)^2 + \dots + (\beta \cdot \phi_w)^m] \cdot \hat{w}_t^\# \\ &\quad + \varepsilon_w \cdot \eta \cdot (\beta \cdot \phi_w)^1 \cdot [1 + (\beta \cdot \phi_w) + (\beta \cdot \phi_w)^2 + \dots + (\beta \cdot \phi_w)^m] \cdot \Delta_{\kappa_w} \pi_{t+1} \\ &\quad + \varepsilon_w \cdot \eta \cdot (\beta \cdot \phi_w)^2 \cdot [1 + (\beta \cdot \phi_w) + (\beta \cdot \phi_w)^2 + \dots + (\beta \cdot \phi_w)^m] \cdot \Delta_{\kappa_w} \pi_{t+2} \\ &\quad + \varepsilon_w \cdot \eta \cdot (\beta \cdot \phi_w)^3 \cdot [1 + (\beta \cdot \phi_w) + (\beta \cdot \phi_w)^2 + \dots + (\beta \cdot \phi_w)^m] \cdot \Delta_{\kappa_w} \pi_{t+3} + \dots \end{aligned}$$

由于  $[1 + (\beta \cdot \phi_w) + (\beta \cdot \phi_w)^2 + \dots + (\beta \cdot \phi_w)^m] \approx 1/(1 - \beta \cdot \phi_w)$ , 随着  $m \rightarrow \infty$  上式化简为

$$\begin{aligned} & \sum_{m=0}^{\infty} \eta \cdot (\hat{h}_{t+m}^t - \hat{H}_{t+m}) \\ &\approx -\frac{\varepsilon_w \cdot \eta}{1 - \beta \cdot \phi_w} \cdot \left\{ \hat{w}_t^\# - \left[ (\beta \cdot \phi_w) \cdot \Delta_{\kappa_w} \pi_{t+1} + (\beta \cdot \phi_w)^2 \cdot \Delta_{\kappa_w} \pi_{t+2} + \dots + (\beta \cdot \phi_w)^m \cdot \Delta_{\kappa_w} \pi_{t+m} \right] \right\}, \end{aligned} \quad (3.116)$$

将上式中定义中括号中的内容定义为  $S_{w,t}$ , 进而改写为递归形式

$$\begin{aligned} S_{w,t} &\equiv (\beta \cdot \phi_w) \cdot \Delta_{\kappa_w} \pi_{t+1} + (\beta \cdot \phi_w)^2 \cdot \Delta_{\kappa_w} \pi_{t+2} + \dots + (\beta \cdot \phi_w)^m \cdot \Delta_{\kappa_w} \pi_{t+m} \\ &\approx (\beta \cdot \phi_w) \cdot (\Delta_{\kappa_w} \pi_{t+1} + S_{w,t+1}), \end{aligned} \quad (3.117)$$

式(3.117)代回式(3.116)可得

$$\sum_{m=0}^{\infty} \eta \cdot (\hat{h}_{t+m}^t - \hat{H}_{t+m}) \approx -\frac{\varepsilon_w \cdot \eta}{1 - \beta \cdot \phi_w} \cdot [\hat{w}_t^\# - S_{w,t}], \quad (3.118)$$

结合式(3.118)和式(3.113), 作时间  $m$  的加总得

$$\begin{aligned} S_{MRS,t} &\equiv \sum_{m=0}^{\infty} (\beta \cdot \phi_w)^m \cdot MRS_{t+m}^t \\ &= \sum_{m=0}^{\infty} (\beta \cdot \phi_w)^m \cdot \left[ -\hat{\Psi}_{z^+, t+m} + \eta \cdot \hat{H}_{t+m} \right] + \sum_{m=0}^{\infty} (\beta \cdot \phi_w)^m \cdot \left[ \eta \cdot \left( \hat{h}_{t+m}^t - \hat{H}_{t+m} \right) \right], \end{aligned} \quad (3.119)$$

第二个等式中后半部分由式(3.118)给出。将前半部分定义为  $S_{o,t}$ , 进而改写为递归形式

$$\begin{aligned} S_{o,t} &\equiv \sum_{m=0}^{\infty} (\beta \cdot \phi_w)^m \cdot \left[ -\hat{\Psi}_{z^+, t+m} + \eta \cdot \hat{H}_{t+m} \right] \\ &\approx -\hat{\Psi}_{z^+, t} + \eta \cdot \hat{H}_t + (\beta \cdot \phi_w) \cdot S_{o,t+1}, \end{aligned} \quad (3.120)$$

由此, 式(3.118)、式(3.120)代回式(3.119), 进一步调整得

$$S_{MRS,t} = S_{o,t} - \frac{\varepsilon_w \cdot \eta}{1 - \beta \cdot \phi_w} \cdot \left[ \hat{w}_t^\# - S_{w,t} \right]. \quad (3.121)$$

结合式(3.106)对  $\hat{X}_{t,m}$  作沿时间  $m$  的加总, 定义为  $S_{x,t}$ , 进而改写为递归形式, 得

$$\begin{aligned} S_{x,t} &\equiv \sum_{m=0}^{\infty} (\beta \cdot \phi_w)^m \cdot \hat{X}_{t,m} \\ &= \frac{1}{1 - \beta \cdot \phi_w} \cdot \sum_{m=0}^{\infty} (\beta \cdot \phi_w)^m \cdot \left[ -\Delta \kappa_w \hat{\pi}_{t+1} - \hat{\mu}_{z^+, t+1} \right] \\ &\approx \frac{\beta \cdot \phi_w}{1 - \beta \cdot \phi_w} \cdot \left[ -\Delta \kappa_w \hat{\pi}_{t+1} - \hat{\mu}_{z^+, t+1} \right] + (\beta \cdot \phi_w) \cdot S_{x,t+1}. \end{aligned} \quad (3.122)$$

稳定状态下的最优决策, 式(3.102)的中括号中内容应当为零, 线性近似, 整理得

$$\begin{aligned} \ln \bar{w}_t^\# + \ln \bar{w}_t &= \ln \left( \frac{\bar{w}_t}{\bar{w}_{t+m}} \right) + \ln MRS_{t+m}^t - \ln \mathcal{X}_{t,m}, \\ \frac{\hat{w}_t^\# + \hat{w}_t}{1 - \beta \cdot \phi_w} + S_{x,t} - S_{MRS,t} &= 0, \end{aligned} \quad (3.123)$$

将式(3.115), (3.122), (3.121)的  $\hat{w}_t^\#$ ,  $S_{x,t}$ ,  $S_{MRS,t}$  代入上式, 得

$$\begin{aligned} &\frac{1}{1 - \beta \cdot \phi_w} \cdot \hat{w}_t + \frac{1}{1 - \beta \cdot \phi_w} \cdot \hat{w}_t^\# + S_{x,t} \\ &= S_{o,t} - \frac{1}{1 - \beta \cdot \phi_w} \cdot \varepsilon_w \cdot \eta \cdot \left[ \frac{\phi_w}{1 - \phi_w} \cdot \Delta \kappa_w \hat{\pi}_{w,t} - S_{w,t} \right]. \end{aligned} \quad (3.124)$$

将式(3.124)改为  $t$  时刻对  $t+1$  期的期望

$$\begin{aligned} &\frac{1}{1 - \beta \cdot \phi_w} \cdot (\beta \cdot \phi_w \cdot \hat{w}_{t+1}) + \frac{1}{1 - \beta \cdot \phi_w} \cdot (\beta \cdot \phi_w \cdot \hat{w}_{t+1}^\#) + (\beta \cdot \phi_w \cdot S_{x,t+1}) \\ &= (\beta \cdot \phi_w \cdot S_{o,t+1}) - \frac{1}{1 - \beta \cdot \phi_w} \cdot \varepsilon_w \cdot \eta \cdot \left[ \frac{\phi_w}{1 - \phi_w} \cdot (\beta \cdot \phi_w \cdot \Delta \kappa_w \hat{\pi}_{w,t+1}) - (\beta \cdot \phi_w \cdot S_{w,t+1}) \right], \end{aligned} \quad (3.125)$$

式(3.124)减去式(3.125), 整理得

$$\begin{aligned} &\frac{1}{1 - \beta \cdot \phi_w} \cdot (\hat{w}_t - \beta \cdot \phi_w \cdot \hat{w}_{t+1}) + \frac{1}{1 - \beta \cdot \phi_w} \cdot (\hat{w}_t^\# - \beta \cdot \phi_w \cdot \hat{w}_{t+1}^\#) + \frac{\beta \cdot \varepsilon_w}{1 - \beta \cdot \phi_w} \cdot (\Delta \kappa_w \hat{\pi}_{t+1} + \hat{\mu}_{z^+, t+1}) \\ &= (-\hat{\Psi}_{z^+, t} + \eta \cdot \hat{H}_t) - \frac{\varepsilon_w \cdot \eta}{1 - \beta \cdot \phi_w} \cdot \frac{\phi_w}{1 - \phi_w} \cdot (\Delta \kappa_w \hat{\pi}_{w,t} - \beta \cdot \phi_w \cdot \Delta \kappa_w \hat{\pi}_{w,t+1}) \\ &+ \frac{\varepsilon_w \cdot \eta \cdot \beta \cdot \phi_w}{1 - \beta \cdot \phi_w} \cdot \Delta \kappa_w \hat{\pi}_{w,t+1}, \end{aligned} \quad (3.126)$$

根据式(3.139)得

$$\frac{\bar{w}_t}{\bar{w}_{t-1}} = \frac{W_t}{W_{t-1}} \cdot \frac{P_{t-1}}{P_t} \cdot \frac{z_{t-1}^+}{z_t^+},$$

线性近似展开

$$\begin{aligned} \ln \bar{w}_t &= \ln \bar{w}_{t-1} + \ln \pi_{w,t} - \ln \pi_t - \ln \mu_{z^+,t}, \\ \hat{w}_t &= \hat{w}_{t-1} + \hat{\pi}_{w,t} - \hat{\pi}_t - \hat{\mu}_{z^+,t}, \end{aligned} \quad (3.127)$$

对式(3.109)作近似线性展开，并利用式(3.127)替代  $\hat{\pi}_t$  得

$$\begin{aligned} \Delta_{\kappa_w} \hat{\pi}_t &= \hat{\pi}_t - \kappa_w \cdot \hat{\pi}_{t-1} \\ &= -(\hat{w}_t - \hat{w}_{t-1}) + (\hat{\pi}_{w,t} - \kappa_w \cdot \hat{\pi}_{t-1}) - \hat{\mu}_{z^+,t} \\ &= (\hat{w}_{t-1} - \hat{w}_t) + \Delta_{\kappa_w} \hat{\pi}_{w,t} - \hat{\mu}_{z^+,t} \end{aligned} \quad (3.128)$$

式(3.127)-(3.128)代回式(3.126)，重写为

$$\begin{aligned} &\frac{1}{1 - \beta \cdot \phi_w} \cdot (\hat{w}_t - \beta \cdot \phi_w \cdot \hat{w}_{t+1}) + \frac{1}{1 - \beta \cdot \phi_w} \cdot \frac{\phi_w}{1 - \phi_w} \cdot (\Delta_{\kappa_w} \hat{\pi}_{w,t} - \beta \cdot \phi_w \cdot \Delta_{\kappa_w} \hat{\pi}_{w,t+1}) \\ &- \frac{\beta \cdot \phi_w}{1 - \beta \cdot \phi_w} \cdot [\Delta_{\kappa_w} \cdot \hat{\pi}_{w,t+1} + (\hat{w}_t - \hat{w}_{t+1})] \\ &= (-\hat{\Psi}_{z^+,t} + \eta \cdot \hat{H}_t) - \frac{\varepsilon_w \cdot \eta}{1 - \beta \cdot \phi_w} \cdot \frac{\phi_w}{1 - \phi_w} \cdot (\Delta_{\kappa_w} \hat{\pi}_{w,t} - \beta \cdot \phi_w \cdot \Delta_{\kappa_w} \hat{\pi}_{w,t+1}) \\ &+ \frac{\varepsilon_w \cdot \eta \cdot \beta \cdot \phi_w}{1 - \beta \cdot \phi_w} \cdot \Delta_{\kappa_w} \hat{\pi}_{w,t+1}, \end{aligned}$$

进一步调整为

$$\begin{aligned} &\frac{1}{1 - \beta \cdot \phi_w} \cdot (\hat{w}_t - \beta \cdot \phi_w \cdot \hat{w}_t) - \frac{\beta \cdot \phi_w}{1 - \beta \cdot \phi_w} \cdot (\hat{w}_{t+1} - \hat{w}_{t+1}) + \frac{1 + \varepsilon_w \cdot \eta}{1 - \beta \cdot \phi_w} \cdot \frac{\phi_w}{1 - \phi_w} \cdot (\Delta_{\kappa_w} \hat{\pi}_{w,t}) \\ &- \frac{\beta \cdot \phi_w}{1 - \beta \cdot \phi_w} \cdot \frac{1 + \varepsilon_w \cdot \eta}{1 - \phi_w} \cdot \Delta_{\kappa_w} \hat{\pi}_{w,t+1} = -\hat{\Psi}_{z^+,t} + \phi \cdot \hat{H}_t, \end{aligned}$$

再整理得 wage philips curve

$$\begin{aligned} &(1 + \varepsilon_w \cdot \eta) \cdot \Delta_{\kappa_w} \hat{\pi}_{w,t} \\ &= (1 - \beta \cdot \phi_w) \cdot \left( \frac{1 - \phi_w}{\phi_w} \right) \cdot (-\hat{\Psi}_{z^+,t} + \eta \cdot \hat{H}_t - \bar{w}_t) + \beta \cdot (1 + \eta \cdot \varepsilon_w) \cdot \Delta_{\kappa_w} \cdot \hat{\pi}_{w,t+1} \\ &= \kappa_w \cdot (-\hat{\Psi}_{z^+,t} + \eta \cdot \hat{H}_t - \bar{w}_t) + \beta \cdot (1 + \eta \cdot \varepsilon_w) \cdot \Delta_{\kappa_w} \hat{\pi}_{w,t+1}, \end{aligned} \quad (3.129)$$

其中定义系数  $\kappa_w \equiv (1 - \beta \cdot \phi_w) \cdot \left( \frac{1 - \phi_w}{\phi_w} \right)$ 。

根据工资 philips curve 式(3.129)，当期名义工资率的增速  $\Delta_{\kappa_w} \hat{\pi}_{w,t}$  与对未来名义工资率增速的预期正相关，与额外 1 单位劳动投入的边际成本  $(-\hat{\Psi}_{z^+,t} + \phi \cdot \hat{H}_t)$  正相关，与实际工资  $\hat{w}_t$  正相关。

### 3.5 比较两条 philips curve

比较 price philips curve 式(3.100)和 wage philips curve 式(3.129)。

$$\begin{aligned} \hat{\pi}_t &= \kappa_f \cdot [\alpha \cdot (1 + \eta) \cdot \hat{X}_t + \hat{R}_t] + \beta \cdot E_t \hat{\pi}_{t+1}, \\ \text{其中 } \kappa_f &\equiv \frac{1 - \phi_f}{\phi_f} \cdot (1 - \beta \cdot \phi_f), \end{aligned}$$

$$\Delta_{\kappa_w} \hat{\pi}_{w,t} = \frac{\kappa_w}{(1 + \varepsilon_w \cdot \eta)} \cdot \left( -\Psi_{z+,t} + \eta \cdot \hat{H}_t - \bar{w}_t \right) + \beta \cdot \Delta_{\kappa_w} \hat{\pi}_{w,t+1},$$

其中  $\kappa_w \equiv (1 - \beta \cdot \phi_w) \cdot \left( \frac{1 - \phi_w}{\phi_w} \right)$ .

假定 price stickiness 和 wage stickiness 的程度相等，即  $\kappa_f = \kappa_w$ 。那么，如果  $1 + \varepsilon_w \cdot \eta > 1$  的话，则 PPC 中边际成本  $(\alpha \cdot (1 + \eta) \cdot \hat{X}_t)$  的斜率  $\kappa_f$  要高于 WPC 中边际成本  $(-\Psi_{z+,t} + \eta \cdot \hat{H}_t)$  的斜率  $(\kappa_w / (1 + \varepsilon_w \cdot \eta))^{17}$ 。

若要满足  $1 + \varepsilon_w \cdot \eta > 1$ ，需要

1. 劳动的需求弹性  $\varepsilon_w$  较大，和/或
2. 劳动者提供额外 1 单位劳动的边际成本  $\eta$  较高，即随着劳动供应的增加，边际劳动成本 MRS 式(3.103)迅速上升。

假定  $j^{th}$  劳工联盟出于某种目的，考虑提高名义工资。分三种情况来讨论。

1. 若在市场上对劳动力需求函数的斜率  $\varepsilon_w$  保持不变，则  $j^{th}$  种类劳动力工资提高，对  $j^{th}$  劳动力的需求降低，边际成本 MRS 随之降低；并且 MRS 的斜率越陡峭即  $\eta$  越大，边际成本下降的幅度越大。
2. 若边际成本 MRS 的斜率  $\eta$  保持不变，则对  $j^{th}$  劳动力需求曲线的斜率越平缓， $j^{th}$  工资上升带来导致的对  $j^{th}$  劳动力需求的下降就会越剧烈。
3. 在 MRS 线斜率向上倾斜即  $\eta$  变大的情况下， $j^{th}$  工资上调也会使得边际成本大幅度下降。

因此，若  $j^{th}$  劳工联盟想要提高  $j^{th}$  工资，在弹性需求函数以及陡峭边际成本曲线的条件下，需要做好面临边际成本大幅度下降的准备。但另一方面，ceteris paribus，若当前边际成本较低，也会抑制住  $j^{th}$  劳工联盟提高工资的冲动。

---

<sup>17</sup>在弹性劳动力需求曲线和/或陡峭边际成本曲线的假设下，wpc 的斜率比 ppc 斜率平缓，其原理类似于 firm-specific capital 导致 ppc 曲线斜率降低，参见 Sveen and Weinke (2005); Altig et al. (2011); De Walque et al. (2006)。



# 附录

## 3.A scaled variables

中性技术冲击  $z_t$  的增速

$$\mu_{z,t} \equiv \frac{z_t}{z_{t-1}}. \quad (3.130)$$

investment-specific 技术冲击  $\Psi_t$  的增速

$$\mu_{\Psi,t} \equiv \frac{\Psi_t}{\Psi_{t-1}}. \quad (3.131)$$

结合式(3.5)和式 (3.131)可得, 生产成本系数  $z_t^+$  的增速

$$\mu_{z^+,t} \equiv \frac{z_{t+1}^+}{z_t^+} = \mu_{\Psi,t}^{\frac{\alpha}{1-\alpha}} \cdot \mu_{z,t}. \quad (3.132)$$

physical capital stock

$$\bar{k}_{t+1} \equiv \frac{\bar{K}_{t+1}}{z_t^+ \cdot \Psi_t}. \quad (3.133)$$

physical capital service

$$k_{t+1} \equiv \frac{K_{t+1}}{z_t^+ \cdot \Psi_t}. \quad (3.134)$$

投资品

$$i_t \equiv \frac{I_t}{z_t^+ \cdot \Psi_t}. \quad (3.135)$$

消费品

$$c_t \equiv \frac{C_t}{z_t^+}. \quad (3.136)$$

政府支出

$$g_t \equiv \frac{G_t}{z_t^+}. \quad (3.137)$$

产出

$$y_t \equiv \frac{Y_t}{z_t^+}. \quad (3.138)$$

实际工资水平

$$\bar{w}_t \equiv \frac{W_t}{z_t^+ \cdot P_t}. \quad (3.139)$$

实际资本租金

$$\bar{r}_t^k \equiv \Psi_t \cdot r_t^k. \quad (3.140)$$

$t+1$  时刻形成的资本存量, 在  $t$  时期的价格

$$p_{k',t} \equiv \Psi_t \cdot P_{k',t}. \quad (3.141)$$

### 3.A.1 几个没写完的说明

生产成本系数的调整

$$\psi_{z_t^+, t} \equiv v_t \cdot P_t \cdot z_t^+, \quad (3.142)$$

其中  $v_t$  表示家庭优化问题中，名义预算约束条件的 Lagrangian 乘子，它等于额外 1 单位货币收入带来的边际效用； $v_t \cdot P_t$  因此等于额外 1 单位消费带来的边际效用。

中间产品生产厂商调整后的产品价格

$$\tilde{p}_t \equiv \frac{\tilde{P}_t}{P_t}. \quad (3.143)$$

劳动者联盟调整后的工资

$$\tilde{w}_t \equiv \frac{\tilde{W}_t}{W_t}. \quad (3.144)$$

通货膨胀率

$$1 + \pi_t \equiv \frac{P_t}{P_{t-1}}. \quad (3.145)$$

工资增速

$$1 + \pi_{w,t} \equiv \frac{W_t}{W_{t-1}} \quad (3.146)$$

## 3.B Frisch elasticity of labor supply

### 3.B.1 定义

Frisch elasticity of labor supply 是指在保持财富的边际效用不变的情况下，工资和劳动力供应之间的弹性 (Frisch, 1932)<sup>18</sup>。Frisch 弹性  $\eta^\lambda$  的定义式可以表示如下

$$\eta^\lambda = \frac{\partial n}{\partial w} \cdot \frac{w}{n} ||_\lambda,$$

其中  $||_\lambda$  表示保持财富的边际效用不变。

定义式。考虑如下 consumer problem

$$\begin{aligned} & \max_{\{c_t, a_{t+1}, n_t\}} \sum_{t=0}^{\infty} \beta^t \cdot E_t U(c_t, n_t), \\ & \text{st. } c_t + a_{t+1} = (1+r) \cdot a_t + w_t \cdot n_t. \end{aligned}$$

建 Lagrangian

$$\mathcal{L} = \sum_{t=0}^{\infty} \beta^t E_t \{U(c_t, n_t) + \lambda_t \cdot [c_t + a_{t+1} - (1+r) \cdot a_t - w_t \cdot n_t]\}.$$

FOCs

$$\frac{\partial \mathcal{L}}{\partial c_t} = 0 \Rightarrow U_{c,t} = \lambda_t, \quad (3.147)$$

$$\frac{\partial \mathcal{L}}{\partial n_t} = 0 \Rightarrow U_{n,t} = -\lambda_t \cdot w_t, \quad (3.148)$$

$$\frac{\partial \mathcal{L}}{\partial a_{t+1}} = 0 \Rightarrow \beta \cdot E_t \lambda_{t+1} \cdot (1+r) = \lambda_t. \quad (3.149)$$

<sup>18</sup>此外还有 Marshallian elasticity of labor supply (保持收入不变的情况下)，和 Hicksian elasticity of labor supply (保持效用水平不变的情况下) 等。

式(3.147)-(3.147) $\Rightarrow$

$$\frac{\partial U(c_t(\lambda_t, w_t), n(\lambda_t, w_t))}{\partial c_t} = \lambda_t, \quad (3.150)$$

$$\frac{\partial U(c_t(\lambda_t, w_t), n(\lambda_t, w_t))}{\partial n_t} = -\lambda_t \cdot w_t. \quad (3.151)$$

式(3.147)-(3.148)对  $w_t$  求导得

$$\frac{\partial U_{cc,t}}{\partial w_t} = 0 \Rightarrow U_{cc,t} \cdot \frac{\partial c_t}{\partial w_t} + U_{cn,t} \cdot \frac{\partial n_t}{\partial w_t} = 0, \quad (3.152)$$

$$\frac{\partial U_{nn,t}}{\partial w_t} = -\lambda_t \Rightarrow U_{cn,t} \cdot \frac{\partial c_t}{\partial w_t} + U_{nn,t} \cdot \frac{\partial n_t}{\partial w_t} = -\lambda_t. \quad (3.153)$$

可见式(3.152)-(3.153)是有两个未知变量  $\left(\frac{\partial c_t}{\partial w_t}, \frac{\partial n_t}{\partial w_t}\right)$  的两个方程组。利用式(3.148)替代  $\lambda_t$ , 解得

$$\frac{\partial n_t}{\partial w_t} = \frac{\lambda_t \cdot U_{cc,t}}{U_{cn}^2 - U_{cc,t} \cdot U_{nn,t}} = \frac{-\frac{U_{n,t}}{w_t} \cdot U_{cc,t}}{U_{cn}^2 - U_{cc,t} \cdot U_{nn,t}} \quad (3.154)$$

因此 Frisch elasticity 等于

$$\eta^\lambda = \frac{\partial n_t}{\partial w_t} \cdot \frac{w_t}{n_t} = \frac{U_{n,t}}{n_t \cdot \left[U_{nn,t} - \left(\frac{U_{cn,t}^2}{U_{cc,t}}\right)\right]}. \quad (3.155)$$

### 3.B.2 举例

举例, 消费者问题:  $\max \beta^t \cdot E_t U_{ct,n_t}$ ,  $U_{ct,n_t} = \ln c_t - \alpha \cdot \frac{n_t^{1+\frac{1}{\nu}}}{1+\frac{1}{\nu}}$ , subject to  $c_t + a_{t+1} = (1+r) \cdot a_t + w_t \cdot n_t$ 。

FOCs:

$$\frac{1}{c_t} = \lambda_t, \quad (3.156)$$

$$n_t = \left(\frac{\lambda_t \cdot w_t}{\alpha}\right)^\nu. \quad (3.157)$$

$$\frac{\partial n_t}{\partial w_t} = \nu \cdot \left(\frac{\lambda_t}{\alpha}\right)^\nu \cdot w_t^{\nu-1}. \quad (3.158)$$

Frisch elasticity of labor supply 因此等于

$$\eta^\lambda = \frac{\partial n_t}{\partial w_t} \cdot \frac{w_t}{n_t} = \nu. \quad (3.159)$$

## 3.C 调节成本的常见设定形式及比较

经验研究需要将调节成本函数  $F(\cdot)$  改写为显函数形式, 常见的有三种形式

1. 线性调节成本。设  $S$  近似为常数, 式(3.36)改写为

$$F = constant \cdot I_t, \quad (3.160)$$

即调节成本和投资之间呈线性关系,  $constant$  是常数。

2. 投资调节成本。最早可见Lucas and Prescott (1971), 后经Hayashi (1982) 作进一步模型化。

$$F = I_t - \frac{S''}{2} \cdot \left(\frac{I_t}{K_t} - \delta\right)^2 \cdot \bar{K}_t, \quad (3.161)$$

3. 资本调节成本。最早可见Christiano et al. (2005), 如式(3.36)所示。

### 3.C.1 家庭部门优化条件

在 capital adjustment cost 设定下, 家庭部门优化条件表示为通过选择投入组合  $\{C_t, I_t, \Delta_t, B_{t+1}, \bar{K}_{t+1}\}$ , 基于给定的预算约束式(3.71)、资本积累式(3.73)和资本净收益式(3.75), 来追求式(3.45)效用函数最大化。

建 Lagrangian

$$\begin{aligned} \mathcal{L}_t = & E_0 \sum_{t=0}^{\infty} \beta^t \left\{ \left[ \log(C_t - b \cdot C_{t-1}) - A \cdot \int_0^1 \frac{h_t(j)^{1+\eta}}{1+\eta} dj \right] \right. \\ & + \lambda_{1,t} \cdot \left[ \int_0^1 W_t(j) \cdot h_t(j) dj + X_t^k \cdot \bar{K}_t + R_{t-1} \cdot B_t - P_t \cdot \left( C_t + \frac{1}{\Psi_t} \cdot I_t \right) - B_{t+1} - P_t \cdot P_{k',t} \cdot \Delta_t \right] \\ & + \lambda_{2,t} \cdot \left[ (1-\delta) \cdot \bar{K}_t + \left[ I_t - \frac{S''}{2} \cdot \left( \frac{I_t}{\bar{K}_t} - \delta \right)^2 \cdot \bar{K}_t \right] + \Delta_t - \bar{K}_{t+1} \right] \\ & \left. + \lambda_{3,t} \cdot \left[ u_{t+1} \cdot P_{t+1} \cdot r_{t+1}^k - \frac{P_{t+1}}{\psi_{t+1}} \cdot a(u_{t+1}) - X_{t+1}^k \right] \right\} \end{aligned}$$

FOCs:

$$\frac{\partial \mathcal{L}_t}{\partial C_t} = 0 \Rightarrow \lambda_{1,t} = \frac{U_{C,t}}{P_t}, \quad (3.162)$$

$$\frac{\partial \mathcal{L}_t}{\partial I_t} = 0 \Rightarrow \lambda_{1,t} \cdot P_t = \lambda_{2,t} \cdot \left[ 1 - S'' \cdot \left( \frac{I_t}{\bar{K}_t} - \delta \right) \right], \quad (3.163)$$

$$\frac{\partial \mathcal{L}_t}{\partial \Delta_t} = 0 \Rightarrow \lambda_{1,t} \cdot P_t \cdot P_{k',t} = \lambda_{2,t}, \quad (3.164)$$

$$\frac{\partial \mathcal{L}_t}{\partial B_{t+1}} = 0 \Rightarrow \beta \cdot E_t \lambda_{1,t+1} \cdot R_t^k = \lambda_{1,t}, \quad (3.165)$$

$$\begin{aligned} \frac{\partial \mathcal{L}_t}{\partial \bar{K}_{t+1}} = 0 \Rightarrow & \beta \cdot E_t \lambda_{1,t+1} \cdot X_{t+1}^k = \lambda_{2,t} \\ & + \beta \cdot E_t \lambda_{2,t+1} \cdot \left[ \frac{S''}{2} \cdot \left( \frac{I_{t+1}}{\bar{K}_{t+1}} - \delta \right)^2 - S'' \cdot \left( \frac{I_{t+1}}{\bar{K}_{t+1}} - \delta \right) \cdot \frac{I_{t+1}}{\bar{K}_{t+1}} - (1-\delta) \right]. \end{aligned} \quad (3.166)$$

联立式(3.163)、式(3.164)得

$$P_{k',t} = \frac{1}{1 - S'' \cdot \left( \frac{I_t}{\bar{K}_t} - \delta \right)}. \quad (3.167)$$

式(3.166)改写为

$$\begin{aligned} \lambda_{2,t} &= \beta \cdot E_t \lambda_{1,t+1} \cdot X_{t+1}^k + \beta \cdot E_t \cdot \lambda_{2,t+1} \cdot \mathcal{H}_{t+1}, \\ \mathcal{H}_{t+1} &\equiv \left[ (1-\delta) + S'' \cdot \left( \frac{I_{t+1}}{\bar{K}_{t+1}} - \delta \right) \cdot \frac{I_{t+1}}{\bar{K}_{t+1}} - \frac{S''}{2} \cdot \left( \frac{I_{t+1}}{\bar{K}_{t+1}} - \delta \right)^2 \right], \end{aligned}$$

$\mathcal{H}_{t+1}$  反映每 1 单位  $\bar{K}_{t+1}$  的增加, 可能导致  $\bar{K}_{t+2}$  的增加。等式两侧同时除以  $\lambda_{1,t}$  得

$$\frac{\lambda_{2,t}}{\lambda_{1,t}} = \frac{\beta \cdot E_t \cdot \lambda_{1,t+1} \cdot X_{t+1}^k}{\lambda_{1,t}} + \frac{\beta \cdot E_t \lambda_{2,t+1}}{\beta \cdot E_t \lambda_{2,t+1}} \cdot \frac{\beta \cdot E_t \lambda_{2,t+1}}{\lambda_{1,t}} \cdot \mathcal{H}_{t+1},$$

进一步整理得

$$\frac{U_{C,t}}{\beta \cdot E_t U_{C,t+1}} = \frac{\frac{X_{t+1}^k}{P_{t+1}} + P_{k',t+1} \cdot \mathcal{H}_{t+1}}{P_{k',t}} \quad (3.168)$$

### 3.C.2 比较线性和非线性调节成本

式(3.168)以 Euler equation 的形式反映家庭的跨期消费行为决策。定义投资的回报率 (rental rate of return on investment)  $R_{t+1}^k$

$$R_{t+1}^k = \frac{x_{t+1}^k + \left[ (1 - \delta) + S'' \cdot \left( \frac{I_{t+1}}{\bar{K}_{t+1}} - \delta \right) \cdot \frac{I_{t+1}}{\bar{K}_{t+1}} - \frac{S''}{2} \cdot \left( \frac{I_{t+1}}{\bar{K}_{t+1}} - \delta \right)^2 \right] \cdot P_{k',t+1}}{P_{k',t}}, \quad (3.169)$$

其中  $x_{t+1}^k \equiv \frac{X_{t+1}^k}{P_{t+1}}$  表示实际的 cash payment,  $X_{t+1}^k$  由式(3.75)给出。RHS 分母  $P_{k',t}$  表示每一单位新增  $\bar{K}_t$  的市场价格, 由式(3.167)给出。RHS 分子中  $P_{k',t+1}$  是以消费品形式表示的每一单位  $\bar{K}_{t+2}$  资本存量的市场价格, 乘以  $[ \cdot ]$  之后转化为消费品形式。

在充分竞争的要素市场上, 额外一单位资本的价格取决于其边际产出, 即

$$\begin{aligned} P_{k',t} &= -\frac{dC_t}{d\bar{K}_{t+1}} \\ &= -\frac{dC_t}{dI_t} \cdot \frac{dI_t}{d\bar{K}_{t+1}} \\ &= \frac{1}{\frac{d\bar{K}_{t+1}}{dI_t}} \\ &= 1(\text{线性调节成本}), \text{ 或者} \\ &\frac{1}{1 - S'' \cdot \left( \frac{I_t}{\bar{K}_t - \delta} \right)} (\text{资本调节成本}), \end{aligned}$$

其中  $\frac{dC_t}{dI_t}$  为消费和实物资本积累 (投资) 之间的边际技术替代率 (MRTS), 由式(3.37)-(3.38)给出;  $\frac{I_t}{\bar{K}_{t+1}}$  为投资和实物资本积累之间的边际技术替代率 (MRTS), 由式(3.73)给出。这样, 模型中 Tobin's q 与  $P_{k',t}$  相关, 为资本的市场价格除以投资品的价格, 其中资本的市场价格等于完全竞争市场假定下的资本的边际成本。或者定义

$$q_t \equiv \frac{\lambda_{2,t}}{\lambda_{1,t}} = P_t \cdot P_{k',t}, \quad (3.170)$$

后一等号由式(3.164)得。拉格朗日乘子  $\lambda_{1,t}$ 、 $\lambda_{2,t}$  分别表示额外 1 单位实物资本和消费的边际效用 (影子价格)。 $\lambda_{2,t}/\lambda_{1,t}$  因此表示为了在明天多拥有 1 单位的实物资本, 个人愿意放弃多少单位的当前消费, 即资本相对于消费的价格。

### 3.C.3 比较投资调节成本和资本调节成本

VAR-based evidence: 通过对 actual data 的 observation 可以看出, 在一个正的货币冲击发生后, 一方面实际利率持久走低, 另一方面投资却呈现出 hump-shaped pattern。为了让模型能够更好解释这一现象, 就需要在建构模型的时候, 使得(3.169)中的真实回报率  $R_t^k$  随着扩张性货币政策冲击而降低。情境分析如下。

1. 如果  $S'' = 0$ , 即不存在 adjustment cost。此时  $q_t = p_{k',t} = 0$ , 式(3.169)中唯一能让  $R_{t+1}^k$  下降的便是  $x_{t+1}^k$ , 此时
  - (a) 若 rate of return on capital  $RRoc \approx (1 - \alpha) \cdot K_{t+1}^{\alpha-1} \cdot H_{t+1}^{1-\alpha} + (1 - \delta) \approx 1/\beta$ , 稳定状态下, 不考虑经济增长情况下的年均  $RRoc \approx 1.03$ , 剩余的内生部分的 RRoc 只占全部 RRoc 的很小一部分。这意味着, 为了让 RRoc 下降一小部分, 需要让内生  $RRoc(K_{t+1}^{\alpha-1} \cdot H_{t+1}^{1-\alpha})$  下降非常大的一部分, 这意味着需要让投资大幅度增加——而在

在现实中很难做到：一方面投资的大幅度上升会压迫消费发生大幅度下降，可是现实世界中消费并没有发生如此大的降低<sup>19</sup>；另一方面作为流量的投资的大幅度变化，也只能导致作为存量的实物资本的小幅度提升。

- (b) VAR based evidence 表明，正的货币政策冲击发生后，就业上升，这意味着 hours worked  $H_t$  上升，这会迫使内生 RROC 上升，压迫实物资本存量，调低投资的回报率。

总之，不引入调节成本的模型，会产生反事实结论：正的货币政策冲击发生后，投资反而会大幅度扩张，而现实数据观测中并未支持这一结论。

2. 如果  $S'' > 0$ ，则内生 RROC 要比  $S'' = 0$  情况下的内生 RROC 大得多，这是由于当调节成本存在时，资本这就速度加快，实物资本存量降低，资本的边际产出、进而边际回报变大。分两种情况来说明。

- (a) 资本调节成本  $+S'' > 0$  的设定如式(3.161)，一次正货币政策冲击所产生的响应有
  - i. contemporary response of investment  $I_t$ ，即投资响应中最大的一次响应发生在冲击产生的当期而非随后。
  - ii. a hump-shaped response of capital price  $P_{k',t}$ 。这是由于受到即时变化的投资流响应的影响：一个正的货币政策冲击会产生持续数个时间段的  $P_{k',t+1}/pP_{k',t}$  上升。将这一现象带回式(3.169)可见，对未来  $P_{k',t+1}$  的语气上升，及对投资收益率的预期上升，会导家庭产生追加投资的冲动，且短期内这种冲动更强，出现 hump shaped，这使得在均衡状态下，冲击出现当期的投资响应最强，随后减弱。
- (b) 投资调节成本  $+S'' > 0$  的设定如式(3.36)，投资响应是 hump shaped 且最大响应发生在 shock 出现的时间之后而非当期。这是由于相比资本调节成本设定而言，投资调节成本设定下，当期投资相对于上期投资若想要发生大规模变动，其调节成本会变得更高。hump shaped impulse responses of investment 导致当期 growth rate 相对于上期 growth rate 呈现更强的正相关，且至少会持续一小段时期。

---

<sup>19</sup>膏按：habit formation？见第3.2.3节。

# 第四章 NK-DSGE 模型

## 4.1 Introduction

### 4.1.1 模型中的六个生产部门

1. 劳动承包部门。labor contractor，在充分竞争环境中，将家庭部门提供的异质劳动力转化为同质劳动力，供中间产品生产部门使用。
2. 家庭部门。存在一系列家庭，家庭行为包括
  - (a) 供应(异质)劳动力给劳动力承包商
  - (b) 对应向下倾斜的需求曲线，决定(异质)劳动力的工资
  - (c) 投资于实物资本
  - (d) 做实物资本的使用率决策
  - (e) 将资本服务品租借给生产部门，赚取资本租金
  - (f) 消费，购买最终产品
  - (g) 购买政府部门的债券。
3. 中间产品生产部门。存在一系列中间产品生产部门，在垄断竞争环境中，使用来自家庭部门的资本服务品，和来自劳动承包商的劳动力投入，生产异质的中间产品，以供最终产品生产部门使用。
4. 最终产品生产部门。存在一个最终产品生产者，在充分竞争环境中，将中间产品生产部门生产的异质中间产品转化为最终产品，创造产出。
5. 政府部门。
  - (a) 决定(外生)政府支出的规模
  - (b) 通过收取一揽子税和/或发放债券来平衡收支。
6. 中央银行。通过 Taylor rule 来推行货币政策。

### 4.1.2 几点说明

1. cashless economy，效用函数中无 cash。
2. 中性的财政政策，政府收一揽子税和发债的目标是确保收支平衡。
3. 打破利率的 zero lower bound，即不再假定利率必须  $\geq 0$ 。
4. 假定是封闭经济体，无进出口<sup>1</sup>。

<sup>1</sup>DSGE 框架中加入进出口的框架，如Svensson (2010)。

## 4.2 解析模型

### 4.2.1 劳动承包部门

经济体中存在一系列异质化家庭，将它们标记为标准化的  $l \in [0, 1]$ ，每个家庭供应劳动力  $N_t(l)$ 。劳动供应商将异质的  $N_t(l)$  打包为同质劳动  $N_{d,t}$ ，以供生产部门使用。打包技术如下

$$N_{d,t} = \left( \int_0^1 N_t(l)^{\frac{\epsilon_w - 1}{\epsilon_w}} dl \right)^{\frac{\epsilon_w}{\epsilon_w - 1}}, \quad (4.1)$$

其中  $\epsilon_w$  表示不同家庭  $l$  所提供劳动力之间的替代弹性，设  $\epsilon_w > 1$  即它们是替代品。

#### 对 $l^{th}$ 劳动力的需求

劳动承包商的最大化问题。根据给定  $\{W_t, W_t(l)\}$  和既定产出条件式(4.1)，选择异质劳动力  $\{N_t(l)\}$  以实现利润最大化。其中  $W_t$  表示总工资水平， $W_t(l)$  表示  $l^{th}$  类型劳动力工资，

$$\max_{N_t(l)} W_t \cdot N_{d,t} - \int_0^1 W_t(l) \cdot N_t(l) dl.$$

FOC wrt  $N_t(l)$

$$\frac{\partial \left\{ W_t \cdot \left[ \int_0^1 N_t(l)^{\frac{\epsilon_w - 1}{\epsilon_w}} dl \right]^{\frac{\epsilon_w}{\epsilon_w - 1}} - \int_0^1 W_t(l) \cdot N_t(l) dl \right\}}{\partial N_t(l)} = 0,$$

$$W_t \cdot \frac{\epsilon_w}{\epsilon_w - 1} \cdot [.]^{\frac{1}{\epsilon_w - 1}} \cdot \frac{\epsilon_w - 1}{\epsilon_w} \cdot N_t(l)^{-\frac{1}{\epsilon_w}} = W_t(l),$$

$$W_t \cdot \frac{\epsilon_w}{\epsilon_w - 1} \cdot N_{d,t}^{\frac{1}{\epsilon_w}} \cdot \frac{\epsilon_w - 1}{\epsilon_w} \cdot N_t(l)^{-\frac{1}{\epsilon_w}} = W_t(l),$$

$$N_t(l) = N_{d,t} \cdot \left( \frac{W_t(l)}{W_t} \right)^{-\epsilon_w}. \quad (4.2)$$

式(4.2)为 (劳动承包商) 对  $l^{th}$  家庭劳动力的需求函数，可见  $N_t(l)$  一方面取决于 (生产部门) 对  $N_{d,t}$  的需求，一方面取决于  $l^{th}$  劳动力的相对工资。

#### 总工资水平

劳动承包商处于充分竞争环境中，利润为 0。

$$W_t \cdot N_{d,t} = \int_0^1 W_t(l) N_t(l) dl.$$

引入式(4.2)替换掉  $N_t(l)$

$$W_t \cdot N_{d,t} = \int_0^1 W_t(l)^{1-\epsilon_w} dl \cdot (W_t \cdot N_{d,t})^{\epsilon_w}.$$

整理得总工资水平  $W_t$  的决定式，它是一个关于  $W_t(l)$  的函数

$$W_t^{1-\epsilon_w} = \int_0^1 W_t(l)^{1-\epsilon_w} dl. \quad (4.3)$$

### 工资分布指标

根据定义，家庭部门异质劳动力供应的加总为  $N_t = \int_0^1 N_t(l) dl$ ，代入式(4.2)用  $N_{d,t}$  代替  $N_t(l)$

$$N_t = N_{d,t} \cdot \int_0^1 \left( \frac{W_t(l)}{W_t} \right)^{-\epsilon_w} dl = N_{d,t} \cdot v_t^w, \quad (4.4)$$

上式反映了家庭部门的总劳动供应和劳动承包商生产同质劳动之间的关系，其中

$$v_t^w \equiv \int_0^1 \left( \frac{W_t(l)}{W_t} \right)^{-\epsilon_w} \quad (4.5)$$

表示工资分布指标，衡量不同家庭劳动工资相对于总工资的差异程度，设  $v_t^w \geq 1$ 。 $v_t^w \rightarrow 1$ ，工资差异越小，劳动承包商的劳动供应  $N_{d,t}$  接近于家庭总劳动供应；反之亦然。

### 4.2.2 家庭部门

1. 家庭部门存在 Calvo 工资摩擦，即  $l^{th}$  家庭的劳动力  $N_t(l)$  是异质的，这导致工资收入  $W_t(l)$  也是异质的 (Calvo, 1983)。
2. 家庭  $l$  的效用  $U_t(l)$  来自于消费  $C_t$  和劳动 (休闲)  $N_t(l)$ 。家庭与家庭效用的不同仅体现在工资  $W_t(l)$  和劳动  $N_t(l)$  的差异上，其他条件相同。
3. 家庭是实物资本的持有者，一方面决定利用实物资本的强度 (capital utilization intensity, Greenwood et al. (1988))，利用实物资本的强度越高，资本折旧率越大；另一方面决定将多少 (利用率调整后的) 资本服务租借给生产企业，以换取租金。
4. 家庭决定购买政府债券的数量。

### 资本形成

实物资本积累式

$$K_{t+1} = Z_t \cdot \left[ 1 - \frac{\kappa}{2} \cdot \left( \frac{I_t}{I_{t-1}} - 1 \right)^2 \right] \cdot I_t + [1 - \delta(u_t)] \cdot K_t, \quad (4.6)$$

其中

- $Z_t$  表示外部投资冲击，影响从流量投资向存量实物资本的转化效率。
- 常数  $\kappa$  反映投资的调整成本 (adjustment cost, Lucas and Prescott (1971); Hayashi (1982))，我们取 Christiano et al. (2005) 的设定形式。
- $u_t$  表示实物资本的利用率。
- 折旧率变量  $\delta(u_t)$  的函数形式如下

$$\delta(u_t) = \delta_0 + \delta_1 \cdot (u_t - 1) + \frac{\delta_2}{2} \cdot (u_t - 1)^2. \quad (4.7)$$

### 资本服务品

家庭选择  $u_t$  和  $K_t$ ，向生产部门供应资本服务 (capital services)，函数形式如下

$$\hat{K}_t = u_t \cdot K_t. \quad (4.8)$$

## 效用函数

$t$  期效用函数

$$U_t(l) = \ln(C_t - b \cdot C_{t-1}) - \psi_t \cdot \frac{N_t(l)^{1+\chi}}{1+\chi} \quad (4.9)$$

其中

- 消费对效用的贡献表现为习惯形成 (habit formation of consumption), 如 Fuhrer (2000); Bouakez et al. (2005)), 用参数  $b \geq 0$  来表示。 $b \rightarrow 0$ , 上期消费习惯对当期效用的影响就越小。
- 消费和劳动 (休闲) 对效用的贡献是内部可分的 (intratemporal separable), 如 Christiano et al. (2005))。参数  $\chi$  表示劳动力供应的 Frisch 弹性的倒数, 见节 3.B。
- 外生的偏好冲击  $\psi_t$  影响家庭对消费带来的效用和劳动带来的负效用之间的权衡。

## 预算约束条件

家庭预算约束条件, 当期支出不得超过当期收入。

$$P_t \cdot C_t + P_t \cdot I_t + B_{t+1} \leq W_t(l) \cdot N_t(l) + R_t^n \cdot \hat{K}_t + \Pi_t^n - P_t \cdot T_t + (1 + i_{t-1}) \cdot B_t, \quad (4.10)$$

其中

- $P_t$  表示价格水平。
- $t$  期初, 家庭购买  $B_t$  名义国债,  $t$  期末时, 根据上一期名义利率  $i_{t-1}$  获得利息。
- 家庭以名义租金率  $R_t^n$  将资本服务品租借给生产部门。
- $\Pi_t^n$  为 (垄断竞争的中间产品) 生产者的名义盈利, 转回到作为资本所有者的家庭部门。
- $T_t$  表示政府的一揽子税/转移支付。

## 家庭最大化问题

$t^{th}$  家庭决策为, 选择  $\{C_t, N_t(l), W_t(l), u_t, K_{t+1}, B_{t+1}, I_t\}$ , 以实现 forward-looking 形式的效用最大化

$$\max_{\{C_t, N_t(l), W_t(l), u_t, K_{t+1}, B_{t+1}, I_t\}} E_t \sum_{s=0}^{\infty} \beta^s \cdot \nu_{t+s} \cdot U_{t+s} \quad (4.11)$$

$$= \max_{\{C_t, N_t(l), W_t(l), u_t, K_{t+1}, B_{t+1}, I_t\}} E_t \sum_{s=0}^{\infty} \beta^s \cdot \nu_{t+s} \cdot \left\{ \ln(C_{t+s} - b \cdot C_{t+s-1}) - \psi_{t+s} \cdot \frac{N_{t+s}(l)^{1+\chi}}{1+\chi} \right\} \quad (4.12)$$

其中  $\beta$  为时间的折旧系数,  $\nu_{t+s}$  为外生 inter-temporal 偏好冲击。最大化的约束条件包括预算约束条件(4.10), 资本形成式(4.6)。效用表达式(4.9)中引入式(4.8)替换  $\hat{K}_{t+s}$ , 引入式(4.2)替换  $N_t(l)$ 。

### 与工资设定无关的一阶条件

将家庭最大化问题中，首先提取与工资设定无关的控制变量  $\{C_t, u_t, K_{t+1}, B_{t+1}, I_t\}$ ，建 Lagrangian

$$\begin{aligned}\mathcal{L} = E_t \sum_{t=0}^{\infty} \beta^s \cdot & \{ \nu_{t+s} \cdot \ln(C_{t+s} - b \cdot C_{t+s-1}) + \dots \\ & + \lambda_{t+s}^n \cdot [W_{t+s}(l) \cdot N_{t+s}(l) + R_{t+s}^n \cdot u_{t+s} \cdot K_{t+s} + \Pi_{t+s}^n - P_{t+s} \cdot T_{t+s} + (1 + i_{t+s-1}) \cdot B_{t+s}] \\ & + \mu_{t+s} \cdot \left[ Z_{t+s} \cdot \left( 1 - \frac{\kappa}{2} \cdot \left( \frac{I_{t+s}}{I_{t+s-1}} - 1 \right)^2 \right) \cdot I_{t+s} + (1 - \delta(u_{t+s})) \cdot K_{t+s} \right] \}\end{aligned}$$

FOC wrt  $C_t$

$$\frac{\partial \mathcal{L}}{\partial C_t} = 0 \Rightarrow \frac{\partial \nu_t \ln(C_t - b \cdot C_{t-1})}{\partial C_t} = \lambda_t^n \cdot P_t,$$

$$\lambda_t^n \cdot P_t = \frac{\nu_t - b \cdot \beta \cdot E_t \nu_{t+1}}{C_t - b \cdot C_{t-1}}. \quad (4.13)$$

FOC wrt  $u_t$

$$\frac{\partial \mathcal{L}}{\partial u_t} = 0 \Rightarrow \lambda_t^n \cdot R_t^n = \mu_t \cdot \delta'(u_t). \quad (4.14)$$

FOC wrt  $B_{t+1}$

$$\frac{\partial \mathcal{L}}{\partial B_{t+1}} = 0 \Rightarrow \lambda_t^n = \beta \cdot (1 + i_t) \cdot E_t \lambda_{t+1}^n. \quad (4.15)$$

FOC wrt  $I_t$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial I_t} = 0 \Rightarrow \lambda_t^n \cdot P_t = & \frac{\partial \mu_t \cdot Z_t \cdot (\cdot) \cdot I_t}{\partial I_t} \\ = & \mu_t \cdot Z_t \cdot \left[ 1 - \frac{\kappa}{2} \cdot \left( \frac{I_t}{I_{t-1}} - 1 \right)^2 \right] + \mu_t \cdot Z_t \cdot I_t \cdot \left( -\frac{\kappa}{2} \right) \cdot 2 \cdot \left( \frac{I_t}{I_{t-1}} - 1 \right) \cdot \frac{1}{I_{t-1}} \\ & + \frac{\partial \beta \cdot E_t \left\{ \mu_{t+1} \cdot Z_{t+1} \cdot I_{t+1} \cdot \left[ 1 - \frac{\kappa}{2} \cdot \left( \frac{I_{t+1}}{I_t} - 1 \right)^2 \right] \right\}}{\partial I_t} \\ = & \mu_t \cdot Z_t \cdot \left[ 1 - \frac{\kappa}{2} \cdot \left( \frac{I_t}{I_{t-1}} - 1 \right)^2 \right] + Z_t \cdot u_t \cdot I_t \cdot \left( -\frac{\kappa}{2} \right) \cdot 2 \cdot \left( \frac{I_t}{I_{t-1}} - 1 \right) \cdot \frac{1}{I_{t-1}} \\ & + \beta \cdot E_t \cdot Z_{t+1} \cdot \mu_{t+1} \cdot I_{t+1} \cdot \left( -\frac{\kappa}{2} \right) \cdot \left( \frac{I_{t+1}}{I_t} - 1 \right) \cdot \left( \frac{-I_{t+1}}{I_t^2} \right),\end{aligned}$$

$$\begin{aligned}\lambda_t^n \cdot P_t = & Z_t \cdot \mu_t \cdot \left[ 1 - \frac{\kappa}{2} \cdot \left( \frac{I_t}{I_{t-1}} - 1 \right)^2 - \kappa \cdot \left( \frac{I_t}{I_{t-1}} - 1 \right) \cdot \frac{I_t}{I_{t-1}} \right] \\ & + \beta \cdot E_t Z_{t+1} \cdot \mu_{t+1} \cdot \kappa \cdot \left( \frac{I_{t+1}}{I_t} - 1 \right) \cdot \left( \frac{I_{t+1}}{I_t} \right)^2.\end{aligned} \quad (4.16)$$

FOC wrt  $K_{t+1}$

$$\frac{\partial \mathcal{L}}{\partial K_{t+1}} = 0 \Rightarrow \frac{\partial \lambda_t^n \cdot R_t^n \cdot u_t \cdot K_t}{\partial K_{t+1}} + \frac{\mu_t \cdot (1 - \delta(u_t)) \cdot K_t}{\partial K_{t+1}} = \frac{\partial \mu_t \cdot K_{t+1}}{\partial K_{t+1}},$$

$$\mu_t = \beta \cdot E_t \left\{ \lambda_{t+1}^n \cdot R_{t+1}^n \cdot u_{t+1} + \mu_{t+1} \cdot [1 - \delta(u_{t+1})] \right\}. \quad (4.17)$$

将名义量式(4.13)-(4.16)折算为实际量。定义消费的边际效用  $\lambda_t$  和资本服务品的实际租金  $R_t$  如下

$$\lambda_t \equiv \lambda_t^n \cdot P_t, \quad (4.18)$$

$$R_t \equiv \frac{R_t^n}{P_t}. \quad (4.19)$$

此外，定义通货膨胀率  $\pi_t$  如下

$$1 + \pi_t \equiv \frac{P_t}{P_{t-1}}. \quad (4.20)$$

我们有 FOC wrt 消费 (Euler equation)

$$\lambda_t = \frac{\nu_t}{C_t - b \cdot C_{t-1}} - b \cdot \beta \cdot E_t \frac{\nu_{t+1}}{C_{t+1} - b \cdot C_t}. \quad (4.21)$$

FOC wrt 资本利用率

$$\lambda_t \cdot R_t = \mu_t \cdot \delta'(u_t). \quad (4.22)$$

FOC wrt 债券

$$\lambda_t = \beta \cdot E_t \lambda_{t+1} \cdot \left( \frac{1 + i_t}{1 + \pi_{t+1}} \right). \quad (4.23)$$

FOC wrt 投资

$$\begin{aligned} \lambda_t = & \mu_t \cdot Z_t \cdot \left\{ 1 - \frac{\kappa}{2} \cdot \left( \frac{I_t}{I_{t-1}} - 1 \right)^2 - \kappa \cdot \left( \frac{I_t}{I_{t-1}} - 1 \right) \cdot \left( \frac{I_t}{I_{t-1}} \right) \right\} \\ & + \beta \cdot E_t \mu_{t+1} \cdot Z_{t+1} \cdot \kappa \cdot \left( \frac{I_{t+1}}{I_t} - 1 \right) \cdot \left( \frac{I_{t+1}}{I_t} \right)^2 \end{aligned} \quad (4.24)$$

FOC wrt 资本存量

$$\mu_t = \beta \cdot E_t \{ \lambda_{t+1} \cdot R_{t+1} \cdot u_{t+1} + \mu_{t+1} \cdot (1 - \delta(u_{t+1})) \} \quad (4.25)$$

现在来看有关工资设定部分的一阶条件。在任一时间，假定全部  $l \in [0, 1]$  家庭中有  $(1 - \phi_w)$  比例可以调整工资，设为  $W_t^{\#}(l)$ ；有  $\phi_w$  比例不能调整工资，只能根据上期工资和上期通胀率制定当期工资，可表示如下

$$W_t(l) = \begin{cases} W_t^{\#}(l) & \text{能调整工资，概率 } 1 - \phi_w, \\ (1 + \pi_{t-1})^{\zeta_w} \cdot W_{t-1}(l) & \text{不能调整工资，概率 } \phi_w. \end{cases} \quad (4.26)$$

其中通货膨胀率的传导系数  $0 \leq \zeta_w \leq 1$ 。

### 工资设定的 backward indexation 分析

假定某家庭  $l$  能够在  $t$  期调整价格至  $W_t^{\#}(l)$ ，随后直到  $t+s, s \geq 0$  期均不能调整价格。根据式(4.26)，随着  $s = (0, 1, 2 \dots \infty)$ ， $l$  家庭的工资  $W_{t+s}(l)$  依次为

$$W_t(l) = W_t^{\#}(l),$$

$$W_{t+1}(l) = (1 + \pi_t)^{\zeta_w} \cdot W_t(l) = (1 + \pi_t)^{\zeta_w} \cdot W_t^{\#}(l),$$

$$W_{t+2}(l) = (1 + \pi_{t+1})^{\zeta_w} \cdot W_{t+1}(l) = [(1 + \pi_{t+1}) \cdot (1 + \pi_t)]^{\zeta_w} \cdot W_t^{\#}(l),$$

$$W_{t+3}(l) = (1 + \pi_{t+2})^{\zeta_w} \cdot W_{t+2}(l) = [(1 + \pi_{t+2}) \cdot (1 + \pi_{t+1}) \cdot (1 + \pi_t)]^{\zeta_w} \cdot W_t^{\#}(l),$$

...

或者改写为

$$\begin{aligned} W_{t+s}(l) &= \left[ \prod_{j=0}^{s-1} (1 + \pi_{t+j}) \right]^{\zeta_w} \cdot W_t^{\#}(l) \\ &= \left( \frac{P_{t+s-1}}{P_{t-1}} \right)^{\zeta_w} \cdot W_t^{\#}(l). \end{aligned} \quad (4.27)$$

其中第二行根据式(4.20)求得。该式描述  $t+s$  期,  $l^{th}$  家庭的工资与  $t$  期工资和往期通货膨胀 (或往期价格水平) 的关系。

### 涉及工资设定的一阶条件

$t$  期有机会调整工资的家庭  $l$ , 从 forward-looking 的角度, 通过设定  $W_t^{\#}(l)$  追求  $t$  到  $t+s$  期的效用之和最大化,

$$\max_{\{W_t^{\#}(l)\}} E_t \sum_{s=0}^{\infty} (\beta \cdot \phi_w)^s \cdot \left\{ \nu_{t+s} \cdot \left[ \dots - \psi_{t+s} \cdot \frac{N_{t+s}(l)^{1+\chi}}{1+\chi} \right] + \lambda_{t+s}^n \cdot W_{t+s}(l) \cdot N_{t+s}(l) \right\}, \quad (4.28)$$

$W_t^{\#}(l)$  信息在  $t+s$  期仍然有效的概率是  $\phi_w^s$ 。 $(\beta \cdot \phi_w)^s$  为主观折旧系数。

根据式(4.2), 用劳动承包商对  $N_{d,t}$  的需求代替异质劳动力供应  $N_t(l)$ ; 根据式(4.27), 用  $t$  期的调整工资  $W_t^{\#}(l)$  代替  $t+s$  期工资  $W_{t+s}(l)$ , 得

$$N_{t+s}(l) = N_{d,t+s} \cdot \left( \frac{W_{t+s}(l)}{W_{t+s}} \right)^{-\epsilon_w} = N_{d,t+s} \cdot \left( \frac{W_t^{\#}(l)}{W_{t+s}} \right)^{-\epsilon_w} \cdot \left( \frac{P_{t+s-1}}{P_{t-1}} \right)^{-\epsilon_w \cdot \zeta_w}. \quad (4.29)$$

式(4.28)变为

$$\begin{aligned} E_t \sum_{s=0}^{\infty} (\beta \cdot \phi_w)^s \cdot \left\{ -\nu_{t+s} \cdot \psi_{t+s} \cdot \frac{N_{d,t+s}^{1+\chi}}{1+\chi} \cdot \left( \frac{W_t^{\#}(l)}{W_{t+s}} \right)^{-\epsilon_w \cdot (1+\chi)} \cdot \left( \frac{P_{t+s-1}}{P_{t-1}} \right)^{-\epsilon_w \cdot \zeta_w \cdot (1+\chi)} \right. \\ \left. + \lambda_{t+s}^n \cdot \left[ \left( \frac{W_t^{\#}(l)}{W_{t+s}} \right)^{-\epsilon_w} \cdot \left( \frac{P_{t+s-1}}{P_{t-1}} \right)^{-\epsilon_w \cdot \zeta_w} \cdot N_{d,t+s} \right] \right\} \end{aligned} \quad (4.30)$$

求解一阶条件, 整理得

$$W_t^{\#}(l)^{1+\epsilon_w \cdot \chi} = \frac{\epsilon_w}{\epsilon - 1} \cdot \frac{E_t \sum_{s=0}^{\infty} (\beta \cdot \phi_w)^s \cdot \nu_{t+s} \cdot \psi_{t+s} \cdot W_{t+s}^{\epsilon_w \cdot (1+\chi)} \cdot N_{d,t+s}^{1+\chi} \cdot \left( \frac{P_{t+s-1}}{P_{t-1}} \right)^{-\epsilon_w \cdot \zeta_w \cdot (1+\chi)}}{E_t \sum_{s=0}^{\infty} (\beta \cdot \phi_w)^s \cdot \lambda_{t+s}^n \cdot W_{t+s}^{\epsilon_w} \cdot N_{d,t+s} \cdot \left( \frac{P_{t+s-1}}{P_{t-1}} \right)^{\zeta_w \cdot (1-\epsilon_w)}}$$

式右侧与个体家庭  $l$  无关, 可见  $W_t^{\#}(l) = W_t^{\#}$ ,  $\forall l$ 。定义两个辅助变量  $H_{1,t}, H_{2,t}$ , 上式变为

$$W_t^{\#,1+\epsilon_w \cdot \chi} = \frac{\epsilon_w}{\epsilon_w - 1} \cdot \frac{H_{1,t}}{H_{2,t}}, \quad (4.31)$$

其中

$$H_{1,t} \equiv E_t \sum_{s=0}^{\infty} (\beta \cdot \phi_w)^s \cdot \nu_{t+s} \cdot \psi_{t+s} \cdot W_{t+s}^{\epsilon_w \cdot (1+\chi)} \cdot N_{d,t+s}^{1+\chi} \cdot \left( \frac{P_{t+s-1}}{P_{t-1}} \right)^{-\epsilon_w \cdot \zeta_w \cdot (1+\chi)}, \quad (4.32)$$

$$H_{2,t} \equiv E_t \sum_{s=0}^{\infty} (\beta \cdot \phi_w)^s \cdot \lambda_{t+s}^n \cdot W_{t+s}^{\epsilon_w} \cdot N_{d,t+s} \cdot \left( \frac{P_{t+s-1}}{P_{t-1}} \right)^{\zeta_w \cdot (1-\epsilon_w)}. \quad (4.33)$$

将名义量的式(4.33)-(4.31)改写为实际量。定义实际工资

$$w_t \equiv \frac{W_t}{P_t}, \quad (4.34)$$

两个辅助变量分别变为

$$\begin{aligned} H_{1,t} &= E_t \sum_{s=0}^{\infty} (\beta \cdot \phi_w)^s \cdot \nu_{t+s} \cdot \psi_{t+s} \cdot w_{t+s}^{\epsilon_w \cdot (1+\chi)} \cdot N_{d,t+s}^{1+\chi} \cdot \left( \frac{P_{t+s-1}}{P_{t-1}} \right)^{-\epsilon_w \cdot \zeta_w \cdot (1+\chi)} \cdot P_{t+s}^{\epsilon_w \cdot (1+\chi)} \\ &= \nu_t \cdot \psi_t \cdot w_t^{\epsilon_w \cdot (1+\chi)} \cdot N_{d,t}^{1+\chi} \cdot P_t^{\epsilon_w \cdot (1+\chi)} + (\beta \cdot \phi_w) \cdot E_t \left( \frac{P_t}{P_{t-1}} \right)^{-\epsilon_w \cdot \zeta_w \cdot (1+\chi)} \cdot H_{1,t+1} \\ &= \nu_t \cdot \psi_t \cdot w_t^{\epsilon_w \cdot (1+\chi)} \cdot N_{d,t}^{1+\chi} \cdot P_t^{\epsilon_w \cdot (1+\chi)} + (\beta \cdot \phi_w) \cdot E_t (1 + \pi_t)^{-\epsilon_w \cdot \zeta_w \cdot (1+\chi)} \cdot H_{1,t+1}, \end{aligned} \quad (4.35)$$

$$\begin{aligned} H_{2,t} &= E_t \sum_{s=0}^{\infty} (\beta \cdot \phi_w)^s \cdot \lambda_{t+s}^n \cdot W_{t+s}^{\epsilon_w} \cdot N_{d,t+s} \cdot \left( \frac{P_{t+s-1}}{P_{t-1}} \right)^{\zeta_w \cdot (1-\epsilon_w)} \\ &= \lambda_t \cdot w_t^{\epsilon_w} \cdot N_{d,t} \cdot P_t^{\epsilon_w - 1} + (\beta \cdot \phi_w) \cdot E_t \left( \frac{P_t}{P_{t-1}} \right)^{\zeta_w \cdot (1-\epsilon_w)} \cdot H_{2,t+1} \\ &= \lambda_t \cdot w_t^{\epsilon_w} \cdot N_{d,t} \cdot P_t^{\epsilon_w - 1} + (\beta \cdot \phi_w) \cdot E_t (1 + \pi_t)^{\zeta_w \cdot (1-\epsilon_w)} \cdot H_{2,t+1}. \end{aligned} \quad (4.36)$$

继续调整辅助变量，定义

$$\begin{aligned} h_{1,t} &\equiv \frac{H_{1,t}}{P_t^{\epsilon_w \cdot (1+\chi)}} = \nu_t \cdot \psi_t \cdot w_t^{\epsilon_w \cdot (1+\chi)} \cdot N_{d,t}^{1+\chi} \\ &\quad + (\beta \cdot \phi_w) \cdot E_t (1 + \pi_t)^{-\epsilon_w \cdot \zeta_w \cdot (1+\chi)} \cdot (1 + \pi_{t+1})^{\epsilon_w \cdot (1+\chi)} \cdot h_{1,t+1}, \end{aligned} \quad (4.37)$$

$$\begin{aligned} h_{2,t} &\equiv \frac{H_{2,t}}{P_t^{\epsilon_w - 1}} = \lambda_t \cdot w_t^{\epsilon_w} \cdot N_{d,t} \\ &\quad + (\beta \cdot \phi_w) \cdot E_t (1 + \pi_t)^{\zeta_w \cdot (1-\epsilon_w)} \cdot (1 + \pi_{t+1})^{\epsilon_w - 1} \cdot h_{2,t+1}. \end{aligned} \quad (4.38)$$

式(4.31)进一步调整为

$$w_t^{\#,1+\epsilon_w \cdot \chi} = \frac{\epsilon_w}{\epsilon_w - 1} \cdot \frac{h_{1,t}}{h_{2,t}}, \quad (4.39)$$

其中定义实际调整工资

$$w_t^\# \equiv \frac{W_t^\#}{P_t}. \quad (4.40)$$

式(4.37)-(4.39)一道构成了家庭调整工资的最优决策。

### 相对调整工资

出于研究的需要，我们常常更感兴趣于调整工资  $w^\#$  与当期总工资水平  $w_t$  的比值，而非  $w^\#$  值本身。因此定义一组新的辅助变量

$$\begin{aligned} \hat{h}_{1,t} &\equiv \frac{h_{1,t}}{w^{\#, \epsilon_w \cdot (1+\chi)}} \\ &= \nu_t \cdot \psi_t \cdot \left( \frac{w_t}{w_t^\#} \right)^{\epsilon_w \cdot (1+\chi)} \cdot N_{d,t}^{1+\chi} \\ &\quad + (\beta \cdot \phi_w) \cdot E_t (1 + \pi_t)^{-\epsilon_w \cdot \zeta_w \cdot (1+\chi)} \cdot (1 + \pi_{t+1})^{\epsilon_w \cdot (1+\chi)} \cdot \left( \frac{w_{t+1}^\#}{w_t^\#} \right)^{\epsilon_w \cdot (1+\chi)} \cdot \hat{h}_{1,t+1}, \end{aligned} \quad (4.41)$$

$$\begin{aligned}
\hat{h}_{2,t} &\equiv \frac{h_{2,t}}{w^{\#, \epsilon_w}} \\
&= \lambda_t \cdot \left( \frac{w_t}{w_t^\#} \right)^{\epsilon_w} \cdot N_{d,t} \\
&+ (\beta \cdot \phi_w) \cdot E_t (1 + \pi_t)^{\zeta_w \cdot (1 - \epsilon_w)} \cdot (1 + \pi_{t+1})^{\epsilon_w - 1} \cdot \left( \frac{w_{t+1}^\#}{w_t^\#} \right)^{\epsilon_w} \cdot \hat{h}_{2,t+1}.
\end{aligned} \tag{4.42}$$

调整工资决定式(4.39)变为

$$w_t^{\#, 1 + \epsilon_w \cdot \chi} = \frac{\epsilon_w}{\epsilon_w - 1} \cdot \frac{h_{1,t}}{h_{2,t}} = \frac{\epsilon_w}{\epsilon_w - 1} \cdot \frac{\frac{h_{1,t}}{w_t^{\#, \epsilon_w \cdot (1 + \chi)}}}{\frac{h_{2,t}}{w_t^{\#, \epsilon_w}}} \cdot \frac{w_t^{\#, \epsilon_w \cdot (1 + \chi)}}{w_t^{\#, \epsilon_w}}$$

整理得

$$w_t^\# = \frac{\epsilon_w}{\epsilon_w - 1} \cdot \frac{\hat{h}_{1,t}}{\hat{h}_{2,t}}. \tag{4.43}$$

这样，式(4.21)-(4.25)，以及(4.41)-(4.43)一道，描述了家庭的最优行为决策。

### 4.2.3 最终产品生产部门

经济体中存在一个最终产品生产者和一系列异质化的中间产品生产者  $j \in [0, 1]$ ，每个中间产品上生产一种中间产品  $Y_{t,j}$ ，以  $P_t(j)$  的价格提供给最终产品生产者。最终产品生产者将  $Y_{t,j}$  打包为最终产出，满足投入产出关系

$$Y_t = \left( \int_0^1 Y_t(j)^{\frac{\epsilon_p - 1}{\epsilon_p}} dj \right)^{\frac{\epsilon_p}{\epsilon_p - 1}}, \tag{4.44}$$

其中  $\epsilon_p$  表示不同中间产品之间的替代弹性，设  $\epsilon_p > 1$  即它们是替代品。

#### 对 $j^{th}$ 中间产品的需求

在完全竞争市场假定下，最终产品生产者的利润最大化问题。根据给定价格  $\{P_t(j)\}$ ，决定对中间产品  $\{Y_t(j)\}$  的需求，

$$\max_{Y_t(j)} P_t \cdot Y_t - \int_0^1 P_t(j) \cdot Y_t(j) dj,$$

引入式(4.44)替代  $Y_t$ ，FOC wrt  $Y_t(j)$  可得

$$Y_t(j) = \left( \frac{P_t(j)}{P_t} \right)^{-\epsilon_p} \cdot Y_t. \tag{4.45}$$

#### 总价格水平

根据完全竞争假定，利润为 0，

$$P_t \cdot Y_t = \int_0^1 P_t(j) Y_t(j) dj,$$

引入(4.45)，可得经济体的总价格水平

$$P_t^{1-\epsilon_p} = \int_0^1 P_t(j)^{1-\epsilon_p} dj. \tag{4.46}$$

#### 4.2.4 中间产品生产部门

$j^{th}$  中间产品生产者的产出函数

$$Y_t(j) = A_t \cdot \hat{K}_t(j)^\alpha \cdot N_{d,t}(j)^{1-\alpha} - F, \quad (4.47)$$

$F$  代表固定成本,  $F > 0$  意味着在垄断竞争的市场条件下, 中间产品生产者的稳态利润为 0, no negative production, 使得 no entry no exit。详见第3.1.2节。

#### 生产的边际成本

$j^{th}$  中间产品生产者的最大化问题可以描述为, 以  $R_t^n$  的价格从家庭部门租用资本服务  $\hat{K}_t(j)$ , 以  $W_t$  的价格从劳动承包商那里获取同质劳动  $N_{d,t}(j)$ ; 产出  $Y_t(j)$  以  $P_t(j)$  的价格出售给最终产品生产者以获取利润。市场摩擦的存在导致  $j$  无法随意调整产品价格, 但可以通过调整投入品的数量实现成本最小化

$$\min_{\{\hat{K}_t(j), N_{d,t}(j)\}} W_t \cdot N_{d,t}(j) + R_t^n \cdot \hat{K}_t(j),$$

满足约束条件

$$A_t \cdot \hat{K}_t(j)^\alpha \cdot N_{d,t}(j)^{1-\alpha} - F \geq \left( \frac{P_t(j)}{P_t} \right)^{-\epsilon_p} \cdot Y_t,$$

LHS 表示中间产品的供应, RHS 表示需求。

建 Lagrangian

$$\mathcal{L} = W_t \cdot N_{d,t}(j) + R_t^n \cdot \hat{K}_t(j) - \varphi_t(j) \cdot \left[ A_t \cdot \hat{K}_t(j)^\alpha \cdot N_{d,t}(j)^{1-\alpha} - F - \left( \frac{P_t(j)}{P_t} \right)^{-\epsilon_p} \cdot Y_t \right].$$

FOCs

$$\frac{\partial \mathcal{L}}{\partial N_{d,t}(j)} = 0 \Rightarrow W_t = \varphi_t \cdot (1 - \alpha) \cdot A_t \cdot \hat{K}_t(j)^\alpha \cdot N_{d,t}(j)^{1-\alpha},$$

$$\frac{\partial \mathcal{L}}{\partial \hat{K}_t(j)} = 0 \Rightarrow R_t^n = \varphi_t \cdot \alpha \cdot A_t \cdot \hat{K}_t(j)^{\alpha-1} \cdot N_{d,t}(j)^{1-\alpha}.$$

两式相除可得

$$\frac{W_t}{R_t^n} = \frac{1 - \alpha}{\alpha} \cdot \frac{\hat{K}_t(j)}{N_{d,t}(j)},$$

即所有中间产品生产者, 在外部给定的投入要素价格  $\{W_t, R_t^n\}$  下, 会有相同的资本-劳动投入比,

$$\frac{\hat{K}_t(j)}{N_{d,t}(j)} = \frac{\hat{K}_t}{N_{d,t}}, \quad \forall j. \quad (4.48)$$

带回上式, 消除异质化特征可得

$$\frac{W_t}{R_t^n} = \frac{1 - \alpha}{\alpha} \cdot \frac{\hat{K}_t}{N_{d,t}},$$

根据式(4.34)、(4.34)可以将上式改写为实际量

$$\frac{w_t}{R_t} = \frac{1 - \alpha}{\alpha} \cdot \frac{\hat{K}_t}{N_{d,t}}. \quad (4.49)$$

### 对投入要素的需求

式(4.49)带回  $j^{th}$  的生产函数, 可得

$$\begin{aligned} Y_t(j) &= A_t \cdot \hat{K}_t(j)^\alpha \cdot \left[ \frac{R_t^n}{W_t} \cdot \frac{1-\alpha}{\alpha} \cdot \hat{K}_t(j) \right]^{1-\alpha} \\ &= A_t \cdot \left( \frac{W_t}{R_t^n} \right)^{-(1-\alpha)} \cdot \left( \frac{1-\alpha}{\alpha} \right)^{-(1-\alpha)} \cdot \hat{K}_t(j) - F, \end{aligned}$$

$$\begin{aligned} Y_t(j) &= A_t \cdot \left[ \frac{W_t}{R_t^n} \cdot \frac{\alpha}{1-\alpha} \cdot N_{d,t}(j) \right]^\alpha \cdot N_{d,t}^{1-\alpha} - F \\ &= A_t \cdot \left( \frac{W_t}{R_t^n} \right)^\alpha \cdot \left( \frac{1-\alpha}{\alpha} \right)^{-\alpha} \cdot N_{d,t}(j) - F. \end{aligned}$$

整理可得中间产品生产者  $j$  对投入要素  $\hat{K}_t(j), N_{d,t}(j)$  的需求

$$\hat{K}_t(j) = \frac{1}{A_t} \cdot \left( \frac{\hat{K}_t}{N_{d,t}} \right)^{1-\alpha} \cdot [Y_t(j) + F], \quad (4.50)$$

$$N_{d,t}(j) = \frac{1}{A_t} \cdot \left( \frac{\hat{K}_t}{N_{d,t}} \right)^\alpha \cdot [Y_t(j) + F]. \quad (4.51)$$

### 边际成本

根据式(4.50)-(4.51)可求得  $j$  的成本函数

$$Cost_t(j) = W_t \cdot N_{d,t}(j) + R_t^n \cdot \hat{K}_t(j). \quad (4.52)$$

$j$  生产额外 1 单位  $Y_t(j)$  的边际成本

$$\begin{aligned} MC_t(j) &= \frac{\partial Cost_t(j)}{\partial Y_t(j)} = W_t \cdot \frac{\partial N_{d,t}(j)}{\partial Y_t(j)} + R_t^n \cdot \frac{\partial \hat{K}_t(j)}{\partial Y_t(j)} \\ &= \frac{1}{A_t} \cdot \left( \frac{\hat{K}_t}{N_{d,t}} \right)^{-\alpha} \cdot \left[ W_t + R_t^n \cdot \left( \frac{\hat{K}_t}{N_{d,t}} \right) \right] = \frac{1}{A_t} \cdot \left( \frac{\hat{K}_t}{N_{d,t}} \right)^{-\alpha} \cdot \left[ W_t + W_t \cdot \frac{\alpha}{1-\alpha} \right] \quad (4.53) \\ &= \frac{W_t}{(1-\alpha) \cdot A_t \cdot \left( \frac{\hat{K}_t}{N_{d,t}} \right)^{-\alpha}}. \end{aligned}$$

类似地, RHS 也不涉及  $j^{th}$  生产者的异质性特征, 可见  $MC_t(j) = MC_t, \forall j$ 。定义实际边际成本  $mc_t(j)$  如下

$$mc_t \equiv \frac{MC_t}{P_t} = \frac{w_t}{(1-\alpha) \cdot A_t \cdot \left( \frac{\hat{K}_t}{N_{d,t}} \right)^{-\alpha}} \quad (4.54)$$

### 利润函数

根据 product efficiency, 每投入额外 1 单位某种投入要素的成本, 应等于该要素的边际产出与边际成本的乘积

$$\begin{aligned} W_t &= MC_t \cdot \frac{\partial Y_t(j)}{\partial N_{d,t}(j)} = MC_t \cdot (1-\alpha) \cdot A_t \cdot \hat{K}_t(j)^\alpha \cdot N_{d,t}(j)^{-\alpha}, \\ R_t^n &= MC_t \cdot \frac{\partial Y_t(j)}{\partial \hat{K}_t(j)} = MC_t \cdot \alpha \cdot A_t \cdot \hat{K}_t(j)^{\alpha-1} \cdot N_{d,t}(j)^{1-\alpha}. \end{aligned}$$

对于  $N_{d,t}(j)$  和  $\hat{K}_t(j)$  单位的要素投入来说,

$$\begin{aligned} W_t \cdot N_{d,t}(j) &= (1 - \alpha) \cdot A_t \cdot \hat{K}_t(j)^\alpha \cdot N_{d,t}(j)^{1-\alpha}, \\ R_t^n \cdot \hat{K}_t(j) &= \alpha \cdot A_t \cdot \hat{K}_t(j)^{\alpha-1} \cdot N_{d,t}(j)^{1-\alpha}. \end{aligned}$$

成本函数式(4.52)变为

$$\begin{aligned} Cost_t(j) &= W_t \cdot N_{d,t}(j) + R_t^n \cdot \hat{K}_t(j) \\ &= MC_t \cdot A_t \cdot \hat{K}_t(j)^\alpha \cdot N_{d,t}(j)^{1-\alpha} \\ &= MC_t \cdot (Y_t(j) + F). \end{aligned}$$

中间产品生产者  $j$  的利润式

$$\Pi_t^n(j) = P_t(j) \cdot Y_t(j) - MC_t \cdot Y_t(j) - MC_t \cdot F, \quad (4.55)$$

在任一时间  $t$ , 假定全部  $j \in [0, 1]$  生产者中有  $(1 - \phi_p)$  比例可以调整价格, 设为  $P_t^{\#}(j)$ ; 有  $\phi_p$  比例不能调整, 只能根据上期价格和上期通胀率制定当期产品价格, 可表示如下

$$P_t(l) = \begin{cases} P_t^{\#}(l) & \text{能调整价格, 概率 } 1 - \phi_p, \\ (1 + \pi_{t-1})^{\zeta_p} \cdot P_{t-1}(l) & \text{不能调整价格, 概率 } \phi_p. \end{cases} \quad (4.56)$$

其中通货膨胀率的传导系数  $0 \leq \zeta_p \leq 1$ 。

### 价格设定的 backward indexation 分析

假定某生产者  $j$  能够在  $t$  期调整价格至  $P_t^{\#}(j)$ , 随后直到  $t+s, s \geq 0$  期均不能调整价格。根据式(4.56), 随着  $s = (0, 1, 2 \dots \infty)$ ,  $j$  产品价格  $P_{t+s}(j)$  依次为

$$\begin{aligned} P_t(j) &= P_t^{\#}(j), \\ P_{t+1}(j) &= (1 + \pi_t)^{\zeta_p} \cdot P_t(l) = (1 + \pi_t)^{\zeta_p} \cdot P_t^{\#}(j), \\ P_{t+2}(j) &= (1 + \pi_{t+1})^{\zeta_p} \cdot P_{t+1}(j) = [(1 + \pi_{t+1}) \cdot (1 + \pi_t)]^{\zeta_p} \cdot P_t^{\#}(j), \\ P_{t+3}(j) &= (1 + \pi_{t+2})^{\zeta_p} \cdot P_{t+2}(j) = [(1 + \pi_{t+2}) \cdot (1 + \pi_{t+1}) \cdot (1 + \pi_t)]^{\zeta_p} \cdot P_t^{\#}(j), \\ &\dots \end{aligned}$$

或者改写为

$$\begin{aligned} P_{t+s}(j) &= \left[ \prod_{\rho=0}^{s-1} (1 + \pi_{t+\rho}) \right]^{\zeta_p} \cdot P_t^{\#}(j) \\ &= \left( \frac{P_{t+s-1}}{P_{t-1}} \right)^{\zeta_p} \cdot P_t^{\#}(j). \end{aligned} \quad (4.57)$$

该式描述  $t+s$  期,  $j^{th}$  产品价格与  $t$  期价格和往期通货膨胀 (或往期价格水平) 的关系。

### 利润最大化问题

$t$  期有机会调整工资的中间产品生产者  $j$ , 从 forward-looking 的角度, 通过设定  $P_t^{\#}(j)$  追求从  $t$  到  $t+s$  期的利润最大化

$$\max_{\{P_t^{\#}\}} E_t \sum_{s=0}^{\infty} (\beta \cdot \phi_p)^s \cdot \left( \frac{\lambda_{t+s}^n}{\lambda_t^n} \right) \cdot \Pi_{t+s}^n \quad (4.58)$$

$\lambda_t^n$  反映消费的边际效用。 $t$  期价格到  $t+s$  期仍然有影响的概率是  $\phi_p^s$ 。 $(\beta^s \cdot \lambda_{t+s}^n / \lambda_t^n)$  表示名义的随机折旧系数。将式(4.57)、(4.45)代入式(4.55)可得,

$$\begin{aligned}
 \Pi_{t+s}^n &= P_{t+s}(j) \cdot Y_{t+s}(j) - MC_{t+s} \cdot Y_{t+s}(j) - MC_{t+s} \cdot F \\
 &= P_{t+s}(j) \cdot \left( \frac{P_{t+s}(j)}{P_{t+s}} \right)^{-\epsilon_p} \cdot Y_{t+s} - MC_{t+s} \cdot \left( \frac{P_{t+s}(j)}{P_{t+s}} \right)^{-\epsilon_p} \cdot Y_{t+s} - MC_{t+s} \cdot F \\
 &= P_{t+s}(j)^{1-\epsilon_p} \cdot P_{t+s}^{\epsilon_p} \cdot Y_{t+s} - MC_{t+s} \cdot P_{t+s}(j)^{-\epsilon_p} \cdot P_{t+s}^{\epsilon_p} \cdot Y_{t+s} - MC_{t+s} \cdot F \\
 &= \left[ \left( \frac{P_{t+s-1}}{P_{t-1}} \right)^{\zeta_p} \cdot P_t^\#(j) \right]^{1-\epsilon_p} \cdot P_{t+s}^{\epsilon_p} \cdot Y_{t+s} \\
 &\quad - MC_{t+s} \cdot \left[ \left( \frac{P_{t+s-1}}{P_{t-1}} \right)^{\zeta_p} \cdot P_t^\#(j) \right]^{-\epsilon_p} \cdot P_{t+s}^{\epsilon_p} \cdot Y_{t+s} - MC_{t+s} \cdot F.
 \end{aligned} \tag{4.59}$$

式(4.59)带回式(4.58), 最大化问题式变为

$$\begin{aligned}
 \max_{P_t^\#(j)} E_t \sum_{s=0}^{\infty} (\beta \cdot \phi_p)^s \cdot \frac{\lambda_{t+s}^n}{\lambda_t^n} \cdot \{ & \left[ \left( \frac{P_{t+s-1}}{P_{t-1}} \right)^{\zeta_p} \cdot P_t^\#(j) \right]^{1-\epsilon_p} \cdot P_{t+s}^{\epsilon_p} \cdot Y_{t+s} \\
 & - MC_{t+s} \cdot \left[ \left( \frac{P_{t+s-1}}{P_{t-1}} \right)^{\zeta_p} \cdot P_t^\#(j) \right]^{-\epsilon_p} \cdot P_{t+s}^{\epsilon_p} \cdot Y_{t+s} - MC_{t+s} \cdot F \}.
 \end{aligned} \tag{4.60}$$

FOC wrt  $P_t^\#(j)$

$$\begin{aligned}
 (1 - \epsilon_p) \cdot E_t \sum_{s=0}^{\infty} (\beta \cdot \phi_p)^s \cdot \frac{\lambda_{t+s}^n}{\lambda_t^n} \cdot \left( \frac{P_{t+s-1}}{P_{t-s}} \right)^{\zeta_p \cdot (1-\epsilon_p)} \cdot P_t^{\#, -\epsilon_p} \cdot P_{t+s}^{\epsilon_p} \cdot Y_{t+s} \\
 = -\epsilon_p \cdot E_t \sum_{s=0}^{\infty} (\beta \cdot \phi_p)^s \cdot \frac{\lambda_{t+s}^n}{\lambda_t^n} \cdot MC_{t+s} \left( \frac{P_{t+s-1}}{P_{t-s}} \right)^{-\zeta_p \cdot \epsilon_p} \cdot P_t^{\#, -\epsilon_p - 1} \cdot P_{t+s}^{\epsilon_p} \cdot Y_{t+s},
 \end{aligned}$$

整理得

$$P_t^\# = \frac{\epsilon_p}{\epsilon_p - 1} \cdot \frac{E_t \sum_{s=0}^{\infty} (\beta \cdot \phi_p)^s \cdot \lambda_{t+s}^n \cdot MC_{t+s} \left( \frac{P_{t+s-1}}{P_{t-s}} \right)^{-\zeta_p \cdot \epsilon_p} \cdot P_{t+s}^{\epsilon_p} \cdot Y_{t+s}}{E_t \sum_{s=0}^{\infty} (\beta \cdot \phi_p)^s \cdot \lambda_{t+s}^n \cdot \left( \frac{P_{t+s-1}}{P_{t-s}} \right)^{\zeta_p \cdot (1-\epsilon_p)} \cdot P_{t+s}^{\epsilon_p} \cdot Y_{t+s}},$$

引入式(4.18)和(4.54), 上式改写为名义量形式。

$$P_t^\#(j) = \frac{\epsilon_p}{\epsilon_p - 1} \cdot \frac{E_t \sum_{s=0}^{\infty} (\beta \cdot \phi_p)^s \cdot \lambda_{t+s} \cdot mc_{t+s} \cdot \left( \frac{P_{t+s-1}}{P_{t-s}} \right)^{-\zeta_p \cdot \epsilon_p} \cdot P_{t+s}^{\epsilon_p} \cdot Y_{t+s}}{E_t \sum_{s=0}^{\infty} (\beta \cdot \phi_p)^s \cdot \lambda_{t+s} \cdot \left( \frac{P_{t+s-1}}{P_{t-s}} \right)^{\zeta_p \cdot (1-\epsilon_p)} \cdot P_{t+s}^{\epsilon_p} \cdot Y_{t+s}}, \tag{4.61}$$

类似地, RHS 与个体  $j$  无关,  $P^\#(j) = P^\#$ ,  $\forall j$ 。定义两个辅助变量  $X_{1,t}, X_{2,t}$ , 上式变为

$$P_t^\# = \frac{\epsilon_p}{\epsilon_p - 1} \cdot \frac{X_{1,t}}{X_{2,t}}, \tag{4.62}$$

其中

$$\begin{aligned}
 X_{1,t} &\equiv E_t \sum_{s=0}^{\infty} (\beta \cdot \phi_p)^s \cdot \lambda_{t+s} \cdot mc_{t+s} \left( \frac{P_{t+s-1}}{P_{t-s}} \right)^{-\zeta_p \cdot \epsilon_p} \cdot P_{t+s}^{\epsilon_p} \cdot Y_{t+s} \\
 &= \lambda_t \cdot mc_t \cdot P_t^{\epsilon_p} \cdot Y_t + (\beta \cdot \phi_p) \cdot (1 + \pi_t)^{-\zeta_p \cdot \epsilon_p} \cdot E_t X_{1,t+1},
 \end{aligned} \tag{4.63}$$

$$\begin{aligned} X_{2,t} &\equiv E_t \sum_{s=0}^{\infty} (\beta \cdot \phi_p)^s \cdot \lambda_{t+s} \cdot \left( \frac{P_{t+s-1}}{P_{t-s}} \right)^{\zeta_p \cdot (1-\epsilon_p)} \cdot P_{t+s}^{\epsilon_p} \cdot Y_{t+s} \\ &= \lambda_t \cdot P_t^{\epsilon_p-1} \cdot Y_t + (\beta \cdot \phi_p) \cdot (1 + \pi_t)^{\zeta_p \cdot (1-\epsilon_p)} \cdot E_t X_{2,t+1}. \end{aligned} \quad (4.64)$$

继续调整辅助变量，定义

$$x_{1,t} = \frac{X_{1,t}}{P_t^{\epsilon_p}} = \lambda_t \cdot m c_t \cdot Y_t + (\beta \cdot \phi_p) \cdot (1 + \pi_t)^{-\zeta_p \cdot \epsilon_p} \cdot E_t (1 + \pi_{t+1})^{\epsilon_p} \cdot x_{1,t+1} \quad (4.65)$$

$$x_{2,t} = \frac{X_{2,t}}{P_t^{\epsilon_p-1}} = \lambda_t \cdot Y_t + (\beta \cdot \phi_p) \cdot (1 + \pi_t)^{\zeta_p \cdot (1-\epsilon_p)} \cdot E_t (1 + \pi_{t+1})^{\epsilon_p-1} x_{2,t+1}. \quad (4.66)$$

调整价格的决定式(4.62)变为

$$P_t^\# = \frac{\epsilon_p}{\epsilon_p - 1} \cdot P_t \cdot \frac{x_{1,t}}{x_{2,t}}. \quad (4.67)$$

### 调整价格的通胀率

定义调整价格的通胀率为

$$(1 + \pi_t^\#) \equiv \frac{P_t^\#}{P_{t-1}}. \quad (4.68)$$

由式(4.67)得调整价格通胀率表示的决定式

$$(1 + \pi_t^\#) = \frac{\epsilon_p}{\epsilon_p - 1} \cdot (1 + \pi_t) \cdot \frac{x_{1,t}}{x_{2,t}}. \quad (4.69)$$

## 4.2.5 政府部门

### 政府支出

用  $G_t$  定义政府支出；设  $\ln G_t$  满足 AR(1) 过程

$$\ln G_t = (1 - \rho_G) \cdot \ln G + \rho_G \cdot \ln G_{t-1} + s_g \cdot \varepsilon_{G,t}, \quad (4.70)$$

其中  $G$  表示稳定状态下的政府支出。 $0 \leq \rho_G < 1$ 。 $\varepsilon_{G,t}$  表示政府支出的外生冲击， $s_g$  表示这个外生冲击的标准差。

### 政府预算约束

政府部门的总支出不得超过总收入。总支出由当期政府支出  $G_t$  和上一期利率  $i_{t-1}$  价格支付的当期初发放债券  $D_t$  的利息组成。总收入由两部分构成，分别为（一揽子）税收  $T_t$  和当期新发放债券数量  $\Delta D_{t+1} = D_{t+1} - D_t$ 。

$$P_t \cdot G_t + i_{t-1} \cdot D_t \leq P_t \cdot T_t + D_{t+1} - D_t, \quad (4.71)$$

## 4.2.6 中央银行

中央银行的货币政策，假定遵循 partial adjustment 形式的 Taylor rule

$$i_t = (1 - \rho_i) \cdot i + \rho_i \cdot i_{t-1} + (1 - \rho_i) \cdot [\phi_\pi \cdot (\pi_t - \pi) + \phi_y \cdot (\ln Y_t - \ln Y_{t-1})] + s_i \cdot \varepsilon_{i,t}, \quad (4.72)$$

类似地，设  $0 \leq \rho_i < 1$ ； $i$  和  $\pi$  分别代表稳定状态下的利率和通胀率；假定经济体处于 determinacy region，且  $\phi_\pi > 0$ ， $\phi_y > 0$ 。

### 4.2.7 总量层面的均衡

#### 总利润

经济总体层面，总利润  $\Pi_t^n$  来自  $j$  中间产品生产者的利润之和<sup>2</sup>，

$$\begin{aligned}
 \Pi_t^n &= \int_0^1 \Pi_t^n(j) dj \\
 &= \int_0^1 \left[ P_t(j) \cdot Y_t(j) - W_t \cdot N_{d,t}(j) - R_t^n \cdot \hat{K}_t(j) \right] dj \\
 &= \int_0^1 P_t(j) \cdot Y_t(j) dj - W_t \cdot \int_0^1 N_{d,t}(j) dj - R_t^n \cdot \int_0^1 \hat{K}_t(j) dj \\
 &= \int_0^1 P_t(j) \cdot Y_t(j) dj - W_t \cdot N_{d,t} - R_t^n \cdot u_t \cdot \hat{K}_t, \\
 &= \left[ \int_0^1 P_t(j)^{1-\epsilon_p} dj \right] \cdot P_t^{\epsilon_p} \cdot Y_t - W_t \cdot N_{d,t} - R_t^n \cdot u_t \cdot K_t \\
 &= P_t \cdot Y_t - W_t \cdot N_{d,t} - R_t^n \cdot u_t \cdot K_t
 \end{aligned} \tag{4.73}$$

第二行用到了式(4.55)。第四行用到了在市场出清的均衡条件下的假定，投入要素的总供给等于总需求

$$N_{d,t} = \int_0^1 N_{d,t}(j) dj, \tag{4.74}$$

$$\hat{K}_t = u_t \cdot K_t = \int_0^1 \hat{K}_t(j) dj. \tag{4.75}$$

第五行根据式(4.45)，利用  $Y_t$  和相对价格代替对异质中间产品  $Y_t(j)$  的需求，以消除异质性。第六行根据式(4.46)，用总物价水平  $P_t$  替代中间产品价格  $P_t(j)$  以消除异质性。

式(4.73)两侧同时除以  $P_t$ ，可得实际利润表述式

$$\Pi_t = Y_t - w_t \cdot N_{d,t} - R_t \cdot u_t \cdot K_t. \tag{4.76}$$

#### 家庭总预算约束

家庭部门的总预算约束来自  $l$  个家庭预算约束的加总，总支出不得超过总收入。由式(4.10)得

$$\begin{aligned}
 &P_t \cdot C_t + P_t \cdot I_t + B_{t+1} \\
 &= \int_0^1 W_t(l) \cdot N_t(l) dl + R_t^n \cdot u_t \cdot K_t + \Pi_t^n - P_t \cdot T_t + (1 + i_{t-1}) \cdot B_t \\
 &= \int_0^1 W_t(l)^{1-\epsilon_w} \cdot W_t^{\epsilon_w} N_{d,t} dl + R_t^n \cdot u_t \cdot K_t + \Pi_t^n - P_t \cdot T_t + (1 + i_{t-1}) \cdot B_t \\
 &= W_t^{\epsilon_w} \cdot N_{d,t} \cdot \int_0^1 W_t(l)^{1-\epsilon_w} dl + R_t^n \cdot u_t \cdot K_t + \Pi_t^n - P_t \cdot T_t + (1 + i_{t-1}) \cdot B_t \\
 &= W_t \cdot N_{d,t} + R_t^n \cdot u_t \cdot K_t + \Pi_t^n - P_t \cdot T_t + (1 + i_{t-1}) \cdot B_t \\
 &= P_t \cdot Y_t - P_t \cdot T_t + (1 - i_{t-1}) \cdot B_t
 \end{aligned} \tag{4.77}$$

第三行用到了式(4.2)，市场出清均衡下，用劳动承包者的同质劳动产出  $N_{d,t}$  替代家庭的劳动力供应  $N_t(l)$ ，以消除异质性。第五行用根据式(4.3)，利用总工资水平  $W_t$  替代异质家庭劳动力工资  $W_t(l)$ 。第六行根据式(4.73)替代总利润  $\Pi_t^n$ 。

<sup>2</sup>最终产品部门假定为完全竞争的，利润为零。

### 总产出的使用

国债市场出清要求家庭部门持有全部政府债券，即  $D_t \equiv B_t, \forall t$ 。

政府预算约束条件式(4.71)调整为

$$P_t \cdot T_t = P_t \cdot G_t + (1 + i_{t-1}) \cdot D_t - D_{t+1}.$$

家庭部门总预算约束条件式(4.77)中用  $D_t$  替代  $B_t$ ，调整得

$$P_t \cdot T_t = P_t \cdot Y_t + (1 - i_{t-1}) \cdot D_t - D_{t-1} - P_t \cdot C_t - P_t \cdot I_t.$$

两式联立可得总产出的使用

$$Y_t = C_t + I_t + G_t. \quad (4.78)$$

### 总量生产函数

经济总体层面， $j$  中间产品的总供应等于总需求。供应端由式(4.47)可得

$$\begin{aligned} \int_0^1 Y_t(j) dj &= \int_0^1 \left[ A_t \cdot \hat{K}_t(j)^\alpha \cdot N_{d,t}(j)^{1-\alpha} - F \right] dj \\ &= A_t \cdot \left( \frac{\hat{K}_t}{N_{d,t}} \right)^\alpha \cdot \int_0^1 N_{d,t}(j) dj - F \\ &= A_t \cdot \hat{K}_t^\alpha N_{d,t}^{1-\alpha} - F, \end{aligned} \quad (4.79)$$

其中第二行利用了式(4.48)，第三行利用了式(4.74)。

需求端由式(4.45)可得

$$\int_0^1 Y_t(j) dj = Y_t \cdot \int_0^1 \left( \frac{P_t(j)}{P_t} \right)^{-\epsilon_p} dj \quad (4.80)$$

定义一个价格分布指标  $\nu_t^p$

$$\nu_t^p \equiv \int_0^1 \left( \frac{P_t(j)}{P_t} \right)^{-\epsilon_p} dj. \quad (4.81)$$

联立式(4.79)-(4.80)可得总量生产函数

$$Y_t = \frac{A_t \cdot \hat{K}_t^\alpha \cdot N_{d,t}^{1-\alpha} - F}{\nu_t^p}. \quad (4.82)$$

### 价格分布指标的演化

根据 Calvo assumption(Calvo, 1983), 将式(4.56)引入式(4.81), 可得价格分布指标随时间的演化,

$$\begin{aligned}
\nu_t^p &\equiv \int_0^1 \left( \frac{P_t(j)}{P_t} \right)^{-\epsilon_p} dj \\
&= \int_0^{1-\phi_p} \left( \frac{P_t^\#}{P_t} \right)^{-\epsilon_p} dj + \int_{1-\phi_p}^1 \left[ \frac{(1+\pi_{t-1})^{\zeta_p} \cdot P_{t-1}(j)}{P_t} \right]^{-\epsilon_p} dj \\
&= (1-\phi_p) \cdot \int_0^1 \left( \frac{P_t^\#}{P_t} \right)^{-\epsilon_p} dj + (1+\pi_{t-1})^{-\zeta_p \cdot \phi_p} \int_0^{\phi_p} \left( \frac{P_{t-1}(j)}{P_{t-1}} \right)^{-\epsilon_p} \cdot \left( \frac{P_{t-1}}{P_t} \right)^{-\epsilon_p} dj \\
&= (1-\phi_p) \cdot \left( \frac{P_t^\#}{P_{t-1}} \right)^{-\epsilon_p} \cdot \left( \frac{P_{t-1}}{P_t} \right)^{-\epsilon_p} + (1+\pi_{t-1})^{-\zeta_p \cdot \epsilon_p} \cdot (1+\pi_t)^{\epsilon_p} \cdot \phi_p \cdot \int_0^1 \left( \frac{P_{t-1}(j)}{P_{t-1}} \right)^{-\epsilon_p} dj \\
&= (1-\phi_p) \cdot \left( \frac{1+\pi_t^\#}{1+\pi_t} \right)^{-\epsilon_p} + \phi_p \cdot (1+\pi_{t-1})^{-\zeta_p \cdot \epsilon_p} \cdot (1+\pi_t)^{\epsilon_p} \cdot \phi_p \cdot \nu_{t-1}^p
\end{aligned} \tag{4.83}$$

### 总物价水平的演化

根据 Calvo assumption (Calvo, 1983), 将式(4.56)引入式(4.46), 可得总物价水平随时间的演化

$$\begin{aligned}
P_t^{1-\epsilon_p} &= \int_0^1 P_t(j)^{1-\epsilon_p} dj \\
&= \int_0^{1-\phi_p} P_t^{\#,1-\epsilon_p} dj + \int_{1-\phi_p}^1 (1+\pi_{t-1})^{\zeta_p \cdot (1-\epsilon_p)} \cdot P_{t-1}(j)^{1-\epsilon_p} dj \\
&= (1-\phi_p) \cdot P_t^{\#,1-\epsilon_p} + \phi_p \cdot (1+\pi_{t-1})^{\zeta_p \cdot (1-\epsilon_p)} \cdot \int_0^1 P_{t-1}(j)^{1-\epsilon_p} dj \\
&= (1-\phi_p) \cdot P_t^{\#,1-\epsilon_p} + \phi_p \cdot (1+\pi_{t-1})^{\zeta_p \cdot (1-\epsilon_p)} \cdot P_{t-1}^{1-\epsilon_p},
\end{aligned}$$

式两侧同时除以  $P_{t-1}^{1-\epsilon_p}$ , 整理得

$$(1+\pi_t)^{1-\epsilon_p} = (1-\phi_p) \cdot \left( 1 + \pi_t^\# \right)^{1-\epsilon_p} + \phi_p \cdot (1+\pi_{t-1})^{\zeta_p \cdot (1-\epsilon_p)} \tag{4.84}$$

### 总工资水平的演化

根据 Calvo assumption(Calvo, 1983), 将式(4.26)引入式(4.26), 可得总工资水平随时间的演化

$$\begin{aligned}
W_t^{1-\epsilon_w} &= \int_0^1 W_t(l)^{1-\epsilon_w} dl \\
&= \int_0^{1-\phi_w} W_t(l)^{\#,1-\epsilon_w} dl + \int_{1-\phi_w}^1 (1+\pi_{t-1})^{\zeta_w \cdot (1-\epsilon_w)} \cdot W_{t-1}(l)^{1-\epsilon_w} dl \\
&= (1-\phi_w) \cdot W_t^{\#,1-\epsilon_w} + (1+\pi_{t-1})^{\zeta_w \cdot (1-\epsilon_w)} \cdot \phi_w \cdot \int_0^1 W_{t-1}(l)^{1-\epsilon_w} dl \\
&= (1-\phi_w) \cdot W_t^{\#,1-\epsilon_w} + \phi_w \cdot (1+\pi_{t-1})^{\zeta_w \cdot (1-\epsilon_w)} \cdot W_{t-1}^{1-\epsilon_w},
\end{aligned}$$

式两侧同时除以  $P_t^{1-\epsilon_w}$  化为实际量, 整理得

$$w_t^{1-\epsilon_w} = (1-\phi_w) \cdot w_t^{\#,1-\epsilon_w} + \phi_w \cdot (1+\pi_{t-1})^{\zeta_w \cdot (1-\epsilon_w)} \cdot w_{t-1}^{1-\epsilon_w} \cdot (1+\pi_t)^{-(1-\epsilon_w)} \tag{4.85}$$

### 4.2.8 外生过程

模型中设定六个外生冲击，假定分为两组。第一组为政府支出冲击式(4.70)、货币政策冲击式(4.72)和影响劳动力供应的 Intratemporal 偏好冲击  $\psi_t$ ，均假定为对数形式的 AR(1) 过程，伴随着非随机均值。 $\psi_t$  设定形式如下

$$\ln \psi_t = (1 - \rho_\psi) \cdot \ln \psi + \rho_\psi \cdot \ln \psi_{t-1} + s_\psi \cdot \varepsilon_{\psi,t}. \quad (4.86)$$

第二组为产出效率冲击  $A_t$ 、投资的边际效率冲击  $Z_t$  和影响消费和休闲所带来效用的 Intertemporal 偏好冲击  $\nu_t$ ，均假定为对数形式的 AR(1)，其非随机均值设定为 1，对应的  $\log(1) = 0$ ，

$$\ln A_t = \rho_A \cdot \ln A_{t-1} + s_A \cdot \varepsilon_{A,t}, \quad (4.87)$$

$$\ln Z_t = \rho_Z \cdot \ln Z_{t-1} + s_Z \cdot \varepsilon_{Z,t}, \quad (4.88)$$

$$\ln \nu_t = \rho_\nu \cdot \ln \nu_{t-1} + s_\nu \cdot \varepsilon_{\nu,t}. \quad (4.89)$$

### 4.2.9 均衡条件的完整解集

根据模型设定，均衡状态由 26 个变量构成，对应 26 个方程，其中包括 33 个待决参数。

26 个变量有

$$\{\lambda_t, \mu_t, C_t, i_t, \Pi_t, R_t, u_t, Z_t, I_t, v_t, \psi_t, w_t, w_t^\#, \hat{h}_{1,t}, \hat{h}_{2,t}, N_{d,t}, \hat{K}_t, K_t, mc_t, \pi_t^\#, x_{1,t}, x_{2,t}, Y_t, G_t, A_t, \nu_t^p\}$$

33 个待决参数有

$$\{\beta, b, \alpha, \delta_0, \delta_1, \delta_2, \pi, \kappa, \epsilon_w, \chi, \phi_w, \zeta_w, \alpha, \zeta_p, \phi_p, \epsilon_p, F, \rho_i, \phi_\pi, \phi_y, s_i, \rho_A, \rho_Z, \rho_G, \rho_v, \rho_\psi, s_A, s_Z, s_v, s_G, s_\psi, G, \psi\}$$

26 个方程如下。来自于家庭部门经济行为决策的方程 1-9；实物资本、资本服务品的形成方程 10-11；生产部门经济行为方程 12-20；政府部门外生支出方程 21；中央银行部门货币政策方程 22；余下 4 个外生冲击方程 23-26。

1. 家庭部门 FOC wrt 消费 (Euler equation) , 式(4.21)  $\Rightarrow$

$$\lambda_t = \frac{\nu_t}{C_t - b \cdot C_{t-1}} - b \cdot \beta \cdot E_t \frac{\nu_{t+1}}{C_{t+1} - b \cdot C_t}.$$

2. 家庭部门 FOC wrt 资本利用率, 式(4.22)  $\Rightarrow$

$$\lambda_t \cdot R_t = \mu_t \cdot \delta'(u_t).$$

3. 家庭部门 FOC wrt 债券, 式(4.23)  $\Rightarrow$

$$\lambda_t = \beta \cdot E_t \lambda_{t+1} \cdot \left( \frac{1 + i_t}{1 + \pi_{t+1}} \right).$$

4. 家庭部门 FOC wrt 投资, 式(4.24)  $\Rightarrow$

$$\begin{aligned} \lambda_t = & \mu_t \cdot Z_t \cdot \left\{ 1 - \frac{\kappa}{2} \cdot \left( \frac{I_t}{I_{t-1}} - 1 \right)^2 - \kappa \cdot \left( \frac{I_t}{I_{t-1}} - 1 \right) \cdot \left( \frac{I_t}{I_{t-1}} \right) \right\} \\ & + \beta \cdot E_t \mu_{t+1} \cdot Z_{t+1} \cdot \kappa \cdot \left( \frac{I_{t+1}}{I_t} - 1 \right) \cdot \left( \frac{I_{t+1}}{I_t} \right)^2. \end{aligned}$$

5. 家庭部门 FOC wrt 资本存量, 式(4.25)  $\Rightarrow$

$$\mu_t = \beta \cdot E_t \{ \lambda_{t+1} \cdot R_{t+1} \cdot u_{t+1} + \mu_{t+1} \cdot (1 - \delta(u_{t+1})) \}.$$

6. 家庭调整工资的设定, 式(4.43)  $\Rightarrow$

$$w_t^\# = \frac{\epsilon_w}{\epsilon_w - 1} \cdot \frac{\hat{h}_{1,t}}{\hat{h}_{2,t}}.$$

7. 家庭调整工资的辅助变量之一, 式(4.41)  $\Rightarrow$

$$\begin{aligned} \hat{h}_{1,t} &= \nu_t \cdot \psi_t \cdot \left( \frac{w_t}{w_t^\#} \right)^{\epsilon_w \cdot (1+\chi)} \cdot N_{d,t}^{1+\chi} \\ &+ (\beta \cdot \phi_w) \cdot E_t (1 + \pi_t)^{-\epsilon_w \cdot \zeta_w \cdot (1+\chi)} \cdot (1 + \pi_{t+1})^{\epsilon_w \cdot (1+\chi)} \cdot \left( \frac{w_{t+1}^\#}{w_t^\#} \right)^{\epsilon_w \cdot (1+\chi)} \cdot \hat{h}_{1,t+1}, \end{aligned}$$

8. 家庭调整工资的辅助变量之二, 式(4.42)  $\Rightarrow$

$$\begin{aligned} \hat{h}_{2,t} &= \lambda_t \cdot \left( \frac{w_t}{w_t^\#} \right)^{\epsilon_w} \cdot N_{d,t} \\ &+ (\beta \cdot \phi_w) \cdot E_t (1 + \pi_t)^{\zeta_w \cdot (1-\epsilon_w)} \cdot (1 + \pi_{t+1})^{\epsilon_w - 1} \cdot \left( \frac{w_{t+1}^\#}{w_t^\#} \right)^{\epsilon_w} \cdot \hat{h}_{2,t+1}. \end{aligned}$$

9. 总工资水平的演化, 反映工资水平和调整工资水平的关系, 式(4.85)  $\Rightarrow$

$$w_t^{1-\epsilon_w} = (1 - \phi_w) \cdot w_t^{\#,1-\epsilon_w} + \phi_w \cdot (1 + \pi_{t-1})^{\zeta_w \cdot (1-\epsilon_w)} \cdot w_{t-1}^{1-\epsilon_w} \cdot (1 + \pi_t)^{-(1-\epsilon_w)}$$

10. 实物资本的积累, 式(4.6)  $\Rightarrow$

$$K_{t+1} = Z_t \cdot \left[ 1 - \frac{\kappa}{2} \cdot \left( \frac{I_t}{I_{t-1}} - 1 \right)^2 \right] \cdot I_t + [1 - \delta(u_t)] \cdot K_t.$$

11. 资本服务品的决定, 式(4.8)  $\Rightarrow$

$$\hat{K}_t = u_t \cdot K_t.$$

12. 要素相对价格和资本-劳动比的关系, 式(4.49)  $\Rightarrow$

$$\frac{w_t}{R_t} = \frac{1 - \alpha}{\alpha} \cdot \frac{\hat{K}_t}{N_{d,t}}.$$

13. 生产的边际成本, 式(4.54)  $\Rightarrow$

$$mc_t \equiv \frac{MC_t}{P_t} = \frac{w_t}{(1 - \alpha) \cdot A_t \cdot \left( \frac{\hat{K}_t}{N_{d,t}} \right)^{-\alpha}}.$$

14. 生产者调整价格的行为决策, 以调整通胀率形式表示, 式(4.69)  $\Rightarrow$

$$(1 + \pi_t^\#) = \frac{\epsilon_p}{\epsilon_p - 1} \cdot (1 + \pi_t) \cdot \frac{x_{1,t}}{x_{2,t}}.$$

15. 生产者调整价格的辅助变量之一, 式(4.65)  $\Rightarrow$

$$x_{1,t} = \lambda_t \cdot mc_t \cdot Y_t + (\beta \cdot \phi_p) \cdot (1 + \pi_t)^{-\zeta_p \cdot \epsilon_p} \cdot E_t (1 + \pi_{t+1})^{\epsilon_p} \cdot x_{1,t+1}.$$

16. 生产者调整价格的辅助变量之二, 式(4.66)  $\Rightarrow$

$$x_{2,t} = \lambda_t \cdot Y_t + (\beta \cdot \phi_p) \cdot (1 + \pi_t)^{\zeta_p \cdot (1-\epsilon_p)} \cdot E_t (1 + \pi_{t+1})^{\epsilon_p - 1} \cdot x_{2,t+1}.$$

17. 价格分布指标的演化, 式(4.83)  $\Rightarrow$

$$\nu_t^p = (1 - \phi_p) \cdot \left( \frac{1 + \pi_t^\#}{1 + \pi_t} \right)^{-\epsilon_p} + \phi_p \cdot (1 + \pi_{t-1})^{-\zeta_p \cdot \epsilon_p} \cdot (1 + \pi_t)^{\epsilon_p} \cdot \phi_p \cdot \nu_{t-1}^p.$$

18. 总物价水平的演化, 反映通胀水平和调整价格通胀水平的关系, 式(4.84)  $\Rightarrow$

$$(1 + \pi_t)^{1-\epsilon_p} = (1 - \phi_p) \cdot \left( 1 + \pi_t^\# \right)^{1-\epsilon_p} + \phi_p \cdot (1 + \pi_{t-1})^{\zeta_p \cdot (1-\epsilon_p)}$$

19. 总产出的使用, 式(4.78)  $\Rightarrow$

$$Y_t = C_t + I_t + G_t.$$

20. 外生政府支出, 式(4.70)  $\Rightarrow$

$$\ln G_t = (1 - \rho_G) \cdot \ln G + \rho_G \cdot \ln G_{t-1} + s_g \cdot \varepsilon_{G,t}.$$

21. 部分调整 Taylor 法则形式的货币政策, 式(4.72)  $\Rightarrow$

$$i_t = (1 - \rho_i) \cdot i + \rho_i \cdot i_{t-1} + (1 - \rho_i) \cdot [\phi_\pi \cdot (\pi_t - \pi) + \phi_y \cdot (\ln Y_t - \ln Y_{t-1})] + s_i \cdot \varepsilon_{i,t}.$$

22. 期内偏好冲击, 影响劳动力供应, 式(4.86)  $\Rightarrow$

$$\ln \psi_t = (1 - \rho_\psi) \cdot \ln \psi + \rho_\psi \cdot \ln \psi_{t-1} + s_\psi \cdot \varepsilon_{\psi,t}.$$

23. 产出效率冲击, 影响总产出, 式(4.87)  $\Rightarrow$

$$\ln A_t = \rho_A \cdot \ln A_{t-1} + s_A \cdot \varepsilon_{A,t}.$$

24. 投资冲击, 影响投资的边际效率, 式(4.88)  $\Rightarrow$

$$\ln Z_t = \rho_Z \cdot \ln Z_{t-1} + s_Z \cdot \varepsilon_{Z,t}.$$

25. 跨期偏好冲击, 影响消费和休闲带来的效用, 式(4.89)  $\Rightarrow$

$$\ln \nu_t = \rho_\nu \cdot \ln \nu_{t-1} + s_\nu \cdot \varepsilon_{\nu,t}.$$

#### 4.2.10 非随机的稳定状态

来看如何求得相关解释变量的稳态值。

1. 式(4.23)  $\Rightarrow$

$$\lambda = \beta \cdot \lambda \cdot \frac{1+i}{i+\pi},$$

调整得

$$i = \left( \frac{1+\pi}{\beta} \right) - 1. \quad (4.90)$$

2. 式(4.24)  $\Rightarrow$

$$\lambda = \mu \cdot Z,$$

其中设稳定状态下, 投资的边际效率冲击为 1,

$$Z \equiv 1, \quad (4.91)$$

整理得

$$\lambda = \mu, \quad (4.92)$$

即 Tobin's q = 1, 详见第3.C节。

3. 设稳定状态下，资本利用率为 1，

$$u \equiv 1, \quad (4.93)$$

式(4.7)  $\Rightarrow$

$$\delta(u) = \delta_0, \quad (4.94)$$

进而式(4.25)  $\Rightarrow$

$$\mu = \beta \cdot [\lambda \cdot R \cdot u + \mu \cdot (1 - \delta(u))] = \beta \cdot (\lambda \cdot R + \mu \cdot (1 - \delta_0)),$$

整理得

$$R = \frac{1}{\beta} - (1 - \delta_0). \quad (4.95)$$

4. 式(4.7)  $\Rightarrow$

$$\delta'(u_t) = \delta_1 + \delta_2 \cdot (u_t - 1), \quad (4.96)$$

以及，若引入稳定状态下的假定式(4.93)  $u \equiv 1$ ，上式变为

$$\delta'(1) = \delta_1. \quad (4.97)$$

式(4.22)  $\Rightarrow$

$$\lambda \cdot R = \mu \cdot \delta_1,$$

引入式(4.92)、式(4.95)，整理得

$$\delta_1 = \frac{1}{\beta} - (1 - \delta_0). \quad (4.98)$$

5. 式(4.84)  $\Rightarrow$

$$(1 + \pi)^{1-\epsilon_p} = (1 - \phi_p) \cdot (1 + \pi^\#)^{1-\epsilon_p} + \phi_p \cdot (1 + \pi)^{\zeta_p \cdot (1-\epsilon_p)},$$

整理得

$$(1 + \pi^\#) = \left[ \frac{(1 + \pi)^{1-\epsilon_p} - \phi_p \cdot (1 + \pi)^{\zeta_p \cdot (1-\epsilon_p)}}{1 - \phi_p} \right]^{\frac{1}{1-\epsilon_p}}, \quad (4.99)$$

由式(4.99)不难看出，只有当  $\pi = 0$ (稳定状态下的通货膨胀为 0)，或  $\zeta_p = 1$ (所有生产者都可以调整价格)，才能得到  $\pi^\# = \pi$ 。

6. 式(4.83)  $\Rightarrow$

$$\nu^p = (1 - \phi_p) \cdot \left( \frac{1 + \pi^\#}{1 + \pi} \right)^{-\epsilon_p} + \phi_p \cdot (1 + \pi)^{\epsilon_p \cdot (1 - \zeta_p)} \cdot \nu^p,$$

整理得

$$\nu^p = \frac{(1 - \phi_p) \cdot \left( \frac{1 + \pi^\#}{1 + \pi} \right)^{-\epsilon_p}}{1 - \phi_p \cdot (1 + \pi)^{\epsilon_p \cdot (1 - \zeta_p)}}. \quad (4.100)$$

同上分析，由式(4.100)可见，只有当  $\pi = 0$  或  $\zeta_p = 1$  时，可得  $\nu^p = 1$ ，即价格完全一致。

7. 式(4.65)- (4.66)  $\Rightarrow$

$$\begin{aligned} x_1 &= \lambda \cdot mc \cdot Y + \beta \cdot \phi_p \cdot (1 + \pi)^{-\zeta_p \cdot \epsilon_p} \cdot (1 + \pi)^{\epsilon_p} x_1, \\ x_2 &= \lambda \cdot Y + \beta \cdot \phi_p \cdot (1 + \pi)^{\zeta_p \cdot (1 - \epsilon_p)} \cdot (1 + \pi)^{\epsilon_p - 1} \cdot x_2, \end{aligned}$$

整理得

$$x_1 = \frac{\lambda \cdot mc \cdot Y}{1 - \beta \cdot \phi_p \cdot (1 + \pi)^{\epsilon_p \cdot (1 - \epsilon_p)}}, \quad (4.101)$$

$$x_2 = \frac{\lambda \cdot Y}{1 - \beta \cdot \phi_p \cdot (1 + \pi)^{(\epsilon_p - 1) \cdot (-1 + \zeta_p)}}, \quad (4.102)$$

进而我们有

$$\frac{x_1}{x_2} = mc \cdot \frac{1 - \beta \cdot \phi_p \cdot (1 + \pi)^{(\epsilon_p - 1) \cdot (-1 + \zeta_p)}}{1 - \beta \cdot \phi_p \cdot (1 + \pi)^{\epsilon_p \cdot (1 - \zeta_p)}}. \quad (4.103)$$

同上分析, 由式(4.103)可见, 当  $\pi = 0$  或  $\zeta_p = 1$  时,  $mc = \frac{x_1}{x_2}$  是个常量, 厂商基于相同的边际成本定价。否则  $\frac{x_1}{x_2}$  是个变量, 厂商在边际成本上做一个额外的 price markup, 作为定价依据。

8. 对式(4.69)求非随机稳态, 并引入式(4.103)可得

$$\begin{aligned} (1 + \pi^\#) &= \frac{\epsilon_p}{\epsilon_p - 1} \cdot (1 + \pi) \cdot \frac{x_1}{x_2} \\ &= \frac{\epsilon_p}{\epsilon_p - 1} \cdot (1 + \pi) \cdot mc \cdot \frac{1 - \beta \cdot \phi_p \cdot (1 + \pi)^{(\epsilon_p - 1) \cdot (-1 + \zeta_p)}}{1 - \beta \cdot \phi_p \cdot (1 + \pi)^{\epsilon_p \cdot (1 - \zeta_p)}}, \end{aligned}$$

整理可得稳态边际成本

$$mc = \frac{\epsilon_p - 1}{\epsilon_p} \cdot \frac{1 + \pi^\#}{1 + \pi} \cdot \frac{1 - \beta \cdot \phi_p \cdot (1 + \pi)^{\epsilon_p \cdot (1 - \zeta_p)}}{1 - \beta \cdot \phi_p \cdot (1 + \pi)^{(1 - \epsilon_p) \cdot (1 - \zeta_p)}}. \quad (4.104)$$

9. 假定条件  $u = 1$  式(4.91)引入式(4.8)  $\Rightarrow$

$$\hat{K} = K. \quad (4.105)$$

10. 式(4.105)引入式(4.49)  $\Rightarrow$

$$\frac{\hat{K}}{N_d} = \frac{K}{N_d} = \frac{\alpha}{1 - \alpha} \cdot \frac{w}{R} \quad (4.106)$$

11. 关于固定成本  $F$  值的选取。模型设定中常见两种方案。第一种方案是设  $F = 0$ , 即不存在固定成本。对于  $F \neq 0$  的第二种方案, 将  $F$  值设定为使得厂商的稳态利润为 0。具体说来, 式(4.76)对应的稳定状态

$$\Pi = Y - w \cdot N_d - R \cdot K = 0,$$

其中利用到了  $u = 1$  的条件式(4.93)。进而有

$$\frac{Y}{N_d} = w + R \cdot \frac{K}{N_d}. \quad (4.107)$$

12. 式(4.49)  $\Rightarrow$

$$\begin{aligned} R_t &= w_t \cdot \frac{\alpha}{1 - \alpha} \cdot \left( \frac{\hat{K}_t}{N_{d,t}} \right)^{-1} \\ &= mc_t \cdot (1 - \alpha) \cdot A_t \cdot \left( \frac{\hat{K}_t}{N_{d,t}} \right)^\alpha \cdot \frac{\alpha}{1 - \alpha} \cdot \left( \frac{\hat{K}_t}{N_{d,t}} \right)^{-1} \\ &= \alpha \cdot A_t \cdot mc_t \cdot \left( \frac{\hat{K}_t}{N_{d,t}} \right)^{\alpha-1}, \end{aligned} \quad (4.108)$$

其中第二行根据式(4.54), 用  $mct$  替代  $w_t$ 。

设稳定状态下, 产出效率冲击为 1(对数值为 0)

$$A = 1, \quad (4.109)$$

代入式(4.109)的 A, 式(4.105)以替代  $\hat{K}$ , 根据式(4.104)求得的  $mc$ 。式(4.108)进一步调整为

$$R = \alpha \cdot mc \cdot \left( \frac{K}{N_d} \right)^{\alpha-1},$$

即稳定状态下的资本-劳动投入比表示为

$$\frac{K}{N_d} = \left( \frac{\alpha \cdot mc}{R} \right)^{\frac{1}{1-\alpha}}. \quad (4.110)$$

13. 式(4.54)  $\Rightarrow$

$$w = (1 - \alpha) \cdot mc \cdot \left( \frac{K}{N_d} \right)^\alpha. \quad (4.111)$$

14. 式(4.85)  $\Rightarrow$

$$w^{1-\epsilon_w} = (1 - \phi_w) \cdot w^{\#} \cdot 1^{-\epsilon_w} + \phi_w \cdot (1 + \pi)^{\zeta_w \cdot (1 - \epsilon_w)} \cdot (1 + \pi)^{\epsilon_w - 1} \cdot w^{1-\epsilon_w},$$

整理得

$$w^{\#} = \left\{ \frac{w^{1-\epsilon_w} \cdot [1 - \phi_w \cdot (1 + \pi)^{(\epsilon_w - 1) \cdot (1 - \zeta_w)}]}{1 - \phi_w} \right\}^{\frac{1}{1-\epsilon_w}}. \quad (4.112)$$

同上分析, 由式(4.112)可见, 当  $\pi = 0$  或  $\zeta_w = 1$ (所有家庭都可以调整工资) 时, 可得  $w^{\#} = w$ 。

15. 设稳定状态下, 影响跨期偏好冲击的变量对数为 0,

$$\nu \equiv 1, \quad (4.113)$$

式(4.113)引入式(4.21)  $\Rightarrow$

$$\begin{aligned} \lambda &= \frac{\nu}{c \cdot (1 - b)} - b \cdot \beta \cdot \frac{\nu}{c \cdot (1 - b)} \\ &= \frac{1}{c} \cdot \frac{1 - b \cdot \beta}{1 - b}. \end{aligned} \quad (4.114)$$

16. 式(4.6)  $\Rightarrow$

$$K = Z \cdot I + (1 - \delta(u)) \cdot K,$$

引入式(4.91)、式(4.94)得

$$I = \delta_0 \cdot K. \quad (4.115)$$

17. 式(4.78)  $\Rightarrow$

$$Y = C + I + G,$$

定义稳定的政府支出占全部 GDP 的比重  $\omega \equiv \frac{G}{Y}$ , 引入式(4.115), 上式变为

$$(1 - \omega) \cdot Y = C + IC + \delta_0 \cdot K.$$

式两侧同时除以  $N_d$  得

$$(1 - \omega) \cdot \frac{Y}{N_d} = \frac{C}{N_d} + \delta_0 \cdot \frac{K}{N_d},$$

引入式(4.107)以替代  $Y/N_d$ , 整理得

$$\begin{aligned} \frac{C}{N_d} &= (1 - \omega) \cdot \left( \frac{Y}{N_d} \right) - \delta_0 \cdot \frac{K}{N_d} \\ &= (1 - \omega) \cdot \left( w + R \cdot \frac{K}{N_d} \right) - \delta_0 \cdot \frac{K}{N_d} \end{aligned} \quad (4.116)$$

18. 式(4.41)-(4.42)  $\Rightarrow$

$$\begin{aligned} \hat{h}_1 &= \nu \cdot \psi \cdot \left( \frac{w}{w^\#} \right)^{\epsilon_w \cdot (1+\chi)} \cdot N_d^{1+\chi} + \beta \cdot \phi_w \cdot (1 + \pi)^{-\zeta_w \cdot \epsilon_w \cdot (1+\chi)} \cdot (1 + \pi)^{\epsilon_w \cdot (1+\chi)} \cdot \hat{h}_1, \\ \hat{h}_2 &= \lambda \cdot \left( \frac{w}{w^\#} \right)^{\epsilon_w} \cdot N_d + \beta \cdot \phi_w \cdot (1 + \pi)^{\zeta_w \cdot (1-\epsilon_w)} \cdot (1 + \pi)^{\epsilon_w - 1} \cdot \hat{h}_2, \end{aligned}$$

整理得

$$\hat{h}_1 = \frac{\nu \cdot \psi \cdot \left( \frac{w}{w^\#} \right)^{\epsilon_w \cdot (1+\chi)} \cdot N_d^{1+\chi}}{1 - \beta \cdot \phi_w \cdot (1 + \pi)^{\epsilon_w \cdot (1-\zeta_w) \cdot (1+\chi)}}, \quad (4.117)$$

$$\hat{h}_2 = \frac{\lambda \cdot \left( \frac{w}{w^\#} \right)^{\epsilon_w} \cdot N_d}{1 - \beta \cdot \phi_w \cdot (1 + \pi)^{(1-\zeta_w) \cdot (\epsilon_w - 1)}}, \quad (4.118)$$

进而, 结合  $\nu = 1$  的稳态假定式(4.113)我们有

$$\frac{\hat{h}_1}{\hat{h}_2} = \frac{\psi}{\lambda} \cdot \left( \frac{w}{w^\#} \right)^{\epsilon_w \cdot \chi} \cdot N_d^\chi \cdot \frac{1 - \beta \cdot \phi_w \cdot (1 + \pi)^{(1-\zeta_w) \cdot (\epsilon_w - 1)}}{1 - \beta \cdot \phi_w \cdot (1 + \pi)^{\epsilon_w \cdot (1-\zeta_w) \cdot (1+\chi)}}, \quad (4.119)$$

19. 利用式(4.119)可得, 式(4.43)  $\Rightarrow$

$$\begin{aligned} w^\# &= \frac{\epsilon_w}{\epsilon_w - 1} \cdot \frac{\hat{h}_1}{\hat{h}_2} \\ &= \frac{\epsilon_w}{\epsilon_w - 1} \cdot \frac{\psi}{\lambda} \cdot \left( \frac{w}{w^\#} \right)^{\epsilon_w \cdot \chi} \cdot N_d^\chi \cdot \frac{1 - \beta \cdot \phi_w \cdot (1 + \pi)^{(1-\zeta_w) \cdot (\epsilon_w - 1)}}{1 - \beta \cdot \phi_w \cdot (1 + \pi)^{\epsilon_w \cdot (1-\zeta_w) \cdot (1+\chi)}} \end{aligned} \quad (4.120)$$

式(4.120)分两种情境分析。情境一, 当  $\pi = 0$  时,  $w = w^\#$ , 式子变为

$$w^\# = \frac{\epsilon_w}{\epsilon_w - 1} \cdot N_d^\chi = w,$$

整理得

$$w = \left[ \frac{\epsilon_w}{\epsilon_w - 1} \right] \cdot \left[ \frac{\psi}{\lambda} \cdot N_d^\chi \right], \quad (4.121)$$

即垄断竞争条件下的稳态工资为一个 markup(RHS 前半部分), 乘以劳动和消费的边际替代率 (RHS 后半部分)。

情景二, 对于更通用的情况, 将式改写为关于  $N_d$  的函数

$$N_d^\chi = \frac{\epsilon_w - 1}{\epsilon_w} \cdot \frac{\lambda}{\psi} \cdot w^\# \cdot \left( \frac{w^\#}{w} \right)^{-\epsilon_w \cdot \chi} \cdot \frac{1 - \beta \cdot \phi_w \cdot (1 + \pi)^{\epsilon_w \cdot (1-\zeta_w) \cdot (1+\chi)}}{1 - \beta \cdot \phi_w \cdot (1 + \pi)^{(1-\zeta_w) \cdot (\epsilon_w - 1)}},$$

引入式(4.114)以替代  $\lambda$ , 两侧再同乘  $N_d$  得

$$N_d^{1+\chi} = \frac{\epsilon_w - 1}{\epsilon_w} \cdot \frac{N_d}{C} \cdot \frac{1}{\psi} \cdot \frac{1 - b \cdot \beta}{1 - b} \cdot w^\# \cdot \left( \frac{w^\#}{w} \right)^{-\epsilon_w \cdot \chi} \cdot \frac{1 - \beta \cdot \phi_w \cdot (1 + \pi)^{\epsilon_w \cdot (1-\zeta_w) \cdot (1+\chi)}}{1 - \beta \cdot \phi_w \cdot (1 + \pi)^{(1-\zeta_w) \cdot (\epsilon_w - 1)}}, \quad (4.122)$$

其中  $N_d/C$  的值由式(4.116)给出。

20. 式(4.82)  $\Rightarrow$

$$Y \cdot \nu^p = A \cdot K^\alpha \cdot N_d^{1-\alpha} - F,$$

其中利用了式(4.93)、式(4.109)和(4.105)。进一步调整得

$$F = N_d \cdot \left[ \left( \frac{K}{N_d} \right)^\alpha - \frac{Y}{N_d} \cdot \nu^p \right] \quad (4.123)$$

其中  $K/N_d$ ,  $Y/N_d$  和  $N_d$  的值分别由式(4.106)、(4.107)和(4.122)测算而得。



# 第五章 DSGE with banking

## 5.1 Introduction

将货币和资产价格纳入到一般均衡模型的分析框架中去，在 DSGE 模型中加入银行部门，主要基于Goodfriend and McCallum (2007); Gilchrist (2007) 的工作。Chadha et al. (2008) 对模型做了两点扩展，一是把政府（包括中央银行）部门的预算约束条件也考虑进来，二是对于受到流动性约束的家庭部门，对货币需求与银行部门的贷款供应的比值是变化的。进而Chadha and Corrado (2012) 进一步探讨了当央行对商业银行准备金所支付的利率，也可以作为宏观审慎政策的工具。

## 5.2 模型设定

### 5.2.1 家庭及生产部门

家庭受到流动性约束，决定消费的数量以及劳动力的供应量。此外家庭消费不只取决于存款，也取决于他们可能从银行获得的贷款（流动性）。

出于简化模型的考虑，家庭部门也是最终产品的生产者，追求利润最大化，生产活动表现为 Cobb-Douglas 形式，资本、劳动力投入和生产效率的冲击共同影响产出。产品遵循标准的Yun (1996) 设定，即为 Calvo 定价形式的垄断竞争 (Calvo, 1983)。

家庭的效用函数

$$U = E_0 \sum_{t=0}^{\infty} \beta^t \cdot [\phi \cdot \log(c_t) + (1 - \phi) \cdot \log(1 - n_t^s - m_t^s)], \quad (5.1)$$

其中

- $c_t$  表示实际消费，
- $n_t^s$  和  $m_t^s$  分别表示投入到产品生产和银行部门的劳动力供应， $1 - n_t^s - m_t^s$  进而表示休闲，
- $0 < \phi < 1$  区分消费和休闲带来的相对效用满足。

家庭效用取决于三个约束条件。第一为预算约束条件，对应拉格朗日乘子  $\lambda_t$

$$\begin{aligned} q_t \cdot (1 - \delta) \cdot K_t + \frac{B_t}{P_t^A} + \frac{H_{t-1}}{P_t^A} + w_t \cdot (n_t^s + m_t^s) + c_t^A \cdot \left( \frac{P_t}{P_t^A} \right)^{1-\theta} \\ - w_t \cdot (n_t + m_t) - \frac{H_t}{P_t^A} - tax_t - q_t \cdot K_{t+1} - \frac{B_{t+1}}{P_t^A \cdot (1 + R_t^B)} - c_t = 0, \end{aligned} \quad (5.2)$$

其中

- $K_t$  为  $t$  期初的实物资本存量， $q_t$  为实物资本价格， $\delta$  为实物资本折旧，

- $B_{t+1}$  为  $t$  期末家庭持有的债券量,  $P_t^A$  表示消费品价格指数, 变量加上上角标  $A$  代表总量层面的因素, 个体家庭视为给定条件。
- $H_t$  为  $t$  期末家庭持有的名义货币。
- $w_t$  表示实际工资,
- $P_t$  表示总物价水平, 家庭作为垄断产品的生产者, 所生产产品的需求替代弹性为  $\theta$ , 生产活动需要雇佣劳动力  $n_t$  和  $m_t$ ,
- $tax_t$  表示政府的实际税收, 以一揽子税形式表示
- $R_t^B$  为政府债券的名义利率, 由银行部门的贷款生产函数决定, 见式<sup>1</sup>。

第二为供应-需求的等价约束条件, 家庭部门(最终产品生产部门)市场出清要求总产出等于总供应, 对应拉格朗日乘子  $\xi_t$

$$K_t^\eta \cdot (A_{1,t} \cdot n_t)^{1-\eta} = c_t^A \cdot \left(\frac{P_t}{P_t^A}\right)^{-\theta} \quad (5.3)$$

等式左侧为生产函数。 $\eta$  代表资本的产出弹性。 $A_{1,t}$  为产品生产中的技术冲击, 均值增速为  $\gamma$ 。

第三为存款-货币的等价约束条件。

货币出现在效用函数和模型中, 充当交易媒介, 要求家庭用所持有的货币支付当期消费支出。家庭持有的货币, 假定全部以存款的形式保存在商业银行中。则假定家庭消费与存款呈现如下对应关系<sup>2</sup>

$$c_t = v_t \cdot \frac{D_t}{P_t^A}, \quad (5.4)$$

其中  $D_t$  表示名义存款, 一定数量的存款最终被用于消费, 以比例  $v_t$  表示。

### 5.2.2 银行部门

银行部门吸收家庭部门存款  $D_t$ , 用途一是为家庭提供高能货币  $H_t$ (high-powered money), 满足家庭的消费需求; 二是为家庭提供贷款  $L_t$ , 满足家庭生产活动的资金需求, 银行的资产负债平衡式

$$D_t = H_t + L_t. \quad (5.5)$$

根据式(5.4)和(5.5)可以间接推得经济体对  $H_t$  的需求<sup>3</sup>。假定

$$H_t = rr \cdot D_t, \quad (5.6)$$

其中  $0 < rr < 1$  表示商业银行选择将多少比例的存款留用以满足家庭消费需求, 设为常数。余下的  $1 - rr$  部分可用于贷款  $L_t$ 。联立式(5.5)-(5.5)可得储蓄和贷款之间的关系

$$D_t = \frac{L_t}{1 - rr}. \quad (5.7)$$

银行部门的贷款生产函数由此可以设定如下: 以家庭部门提供的抵押品和银行部门的贷款监控为投入, 以贷款为产出, 满足关系

$$\frac{L_t}{P_t^A} = F \cdot \underbrace{\left( b_{t+1} + A_{3,t} \cdot k \cdot q_t \cdot K_{t+1} \right)}_{\text{抵押品}}^\alpha \cdot \underbrace{\left( A_{2,t} \cdot m_t \right)}_{\text{贷款监控}}^{1-\alpha}, \quad (5.8)$$

<sup>1</sup>膏按: 补一个 reference。

<sup>2</sup>已有大量经验研究, 从微观层面上探讨银行的存款和贷款供应间的关系, 如Kashyap et al. (2002)

<sup>3</sup>进而, 中央银行通过调整 policy rate 来引导商业银行的放贷行为, 以满足这一需求; 而不是通过诸如印钞等直接手段。

其中  $0 \leq \alpha \leq 1$  表示抵押品的产出弹性,  $F$  表示贷款生产函数的中性技术水平,  $b_{t+1}$  为家庭的贴现实际债券

$$b_{t+1} = \frac{B_{t+1}}{P_t^A \cdot (1 + R_t^B)}. \quad (5.9)$$

式(5.8)中第一个括号内表示抵押品, 银行在发放贷款时, 要求借款人(家庭)提供抵押担保, 担保来自两部分, 一为债券抵押  $b_{t+1}$ ; 二为实物资本抵押  $q_t \cdot K_{t+1}$ 。相较于债券抵押品而言, 实物资本抵押品需要大量额外的成本支出, 包括物理状况检查、市场价格评估等, 因此需要做折扣, 以参数  $0 < k < 1$  来表示。此外实物资本品还受到均值为 0 的技术冲击的干扰, 用  $A_{3,t}$  来表示<sup>4</sup>。

第二个括号表示贷款监控, 银行雇佣额外的劳动力  $m_t$ , 监督借款人对贷款的利用情况, 以降低风险。类似地, 假定贷款监控也受技术冲击  $A_{2,t}$  干扰, 均值的增速为  $\gamma$ 。

### 5.2.3 政府(中央银行)部门

政府(包括中央银行)部门的当期预算约束条件为

$$Transfer_t = g_t - tax_t = \frac{H_t - H_{t-1}}{P_t^A} + \frac{\frac{B_{t+1}}{(1+R_t^B)} - B_t}{P_t^A}, \quad (5.10)$$

即政府的转移支付  $Transfer_t$  为全部支出  $g_t$  与收入  $tax_t$  之差。

### 5.2.4 技术冲击

$$\begin{aligned} A_{1,t} &= A_{1,0} \cdot (1 + \gamma)^t, \\ A_{2,t} &= A_{2,0} \cdot (1 + \gamma)^t, \end{aligned}$$

为简化模型, 设  $A_{1,0} = A_{2,0} = 1$ , 则我们有

$$A_{1,t} = (1 + \gamma)^t, \quad (5.11)$$

$$A_{2,t} = (1 + \gamma)^t. \quad (5.12)$$

此外

$$E(A_{3,t}) = 0. \quad (5.13)$$

### 5.2.5 实际消费

(5.7) (5.8) (5.9) 带回到(5.4)可得实际消费

$$\begin{aligned} c_t &= v_t \cdot \frac{D_t}{P_t^A} = \frac{v_t}{1 - rr} \cdot \frac{L_t}{P_t^A} \\ &= \frac{v_t}{1 - rr} \cdot F \cdot (b_{t+1} + A_{3,t} \cdot k \cdot q_t \cdot K_{t+1})^\alpha \cdot (A_{2,t} \cdot m_t)^{1-\alpha} \\ &= v_t \cdot \frac{D_t}{P_t^A} = \frac{v_t}{1 - rr} \cdot \frac{L_t}{P_t^A} \\ &= \frac{v_t}{1 - rr} \cdot F \cdot \left( \frac{B_{t+1}}{P_t^A \cdot (1 + R_t^B)} + A_{3,t} \cdot k \cdot q_t \cdot K_{t+1} \right)^\alpha \cdot (A_{2,t} \cdot m_t)^{1-\alpha}. \end{aligned} \quad (5.14)$$

---

<sup>4</sup>出于简化模型的考虑, 我们假定均衡状态下不存在违约的情况。在不完全市场条件下, 考虑违约的有限承诺情况可见 (Kocherlakota, 1996)。

结合式(5.14), 由银行部门贷款生产函数式(5.8)可得抵押品  $(b_{t+1} + A_{3,t} \cdot k \cdot q_t \cdot K_{t+1})$  的边际产出

$$\frac{\partial(L_t/P_t^A)}{\partial(b_{t+1} + A_{3,t} \cdot k \cdot q_t \cdot K_{t+1})} = \alpha \cdot \frac{\frac{L_t}{P_t^A}}{(b_{t+1} + A_{3,t} \cdot k \cdot q_t \cdot K_{t+1})} = \frac{\Omega_t}{\frac{v_t}{1-r}}, \quad (5.15)$$

其中  $\Omega_t$  定义为

$$\Omega_t = \frac{\alpha \cdot c_t}{(b_{t+1} + A_{3,t} \cdot k \cdot q_t \cdot K_{t+1})}. \quad (5.16)$$

进而可得额外 1 单位 (向金融部门) 劳动力供应  $m_t$ , 债券  $B_{t+1}$ , 实物资本  $K_{t+1}$  导致消费  $c_t$  的边际变化

$$\frac{\partial c_t}{\partial m_t} = (1 - \alpha) \cdot A_{2,t} \cdot \frac{c_t}{m_t}. \quad (5.17)$$

$$\frac{\partial c_t}{\partial B_{t+1}} = \frac{\Omega_t}{P_t^A \cdot (1 + R_t^B)}. \quad (5.18)$$

$$\frac{\partial c_t}{\partial K_{t+1}} = \Omega_t \cdot A_{3,t} \cdot k \cdot q_t. \quad (5.19)$$

### 5.3 一阶条件

建拉格朗日

$$\begin{aligned} \mathcal{L} = E_0 \sum_{t=0}^{\infty} \beta^t \cdot & \{ [\phi \cdot \log(c_t) + (1 - \phi) \cdot \log(1 - n_t^s - m_t^s)] \\ & + \lambda_t \cdot \left[ q_t \cdot (1 - \delta) \cdot K_t + \frac{B_t}{P_t^A} + \frac{H_{t-1}}{P_t^A} + w_t \cdot (n_t^s + m_t^s) + c_t^A \cdot \left( \frac{P_t}{P_t^A} \right)^{1-\theta} \right. \\ & \left. - w_t \cdot (n_t + m_t) - \frac{H_t}{P_t^A} - tax_t - q_t \cdot K_{t+1} - \frac{B_{t+1}}{P_t^A \cdot (1 + R_t^B)} - c_t \right] \\ & + \xi_t \left[ \cdot K_t^\eta \cdot (A_{1,t} \cdot n_t)^{1-\eta} - c_t^A \cdot \left( \frac{P_t}{P_t^A} \right)^{-\theta} \right] \} \end{aligned} \quad (5.20)$$

家庭部门优化可以表示为

$$\max_{\{m_t^s, n_t^s, m_t, n_t, P_t, K_{t+1}, B_{t+1}, c_t, H_t\}} \mathcal{L}.$$

FOC wrt  $m_t^s$  或  $n_t^s$ , 得劳动力的供应

$$\frac{\partial \mathcal{L}}{\partial m_t^s} = 0 \Rightarrow \lambda_t = \frac{1 - \phi}{(1 - n_t^s - m_t^s) \cdot w_t}. \quad (5.21)$$

FOC wrt  $m_t$ , 得银行部门对劳动力的需求

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial m_t} = 0 \Rightarrow & \frac{\phi}{c_t} \cdot \frac{\partial c_t}{\partial m_t} - \lambda_t \cdot w_t - \lambda_t \cdot \frac{\partial c_t}{\partial m_t} = 0, \\ & w_t = \left( \frac{\phi}{c_t \cdot \lambda_t} - 1 \right) \cdot (1 - \alpha) \cdot A_{2,t} \cdot \frac{c_t}{m_t}. \end{aligned} \quad (5.22)$$

FOC wrt  $n_t$ , 得最终产品生产对劳动力的需求

$$\frac{\partial \mathcal{L}}{\partial n_t} = 0 \Rightarrow w_t = \left( \frac{\xi_t}{\lambda_t} \right) \cdot A_{1,t} \cdot (1 - \eta) \cdot \left( \frac{K_t}{A_{1,t} \cdot n_t} \right)^\eta. \quad (5.23)$$

FOC wrt  $P_t$ , 得影子价格的比值

$$\frac{\partial \mathcal{L}}{\partial P_t} = 0 \Rightarrow \frac{\xi_t}{\lambda_t} = - \left( \frac{1-\theta}{\theta} \right) \cdot \left( \frac{P_t}{P_t^A} \right). \quad (5.24)$$

FOC wrt  $K_{t+1}$ , 得

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial K_{t+1}} &= 0 \Rightarrow \\ \phi \cdot \frac{\partial c_t}{\partial K_{t+1}} + \beta \cdot E_t \lambda_{t+1} \cdot q_{t+1} \cdot (1-\delta) - \lambda_t \cdot q_t - \lambda_t \cdot \frac{\partial c_t}{\partial K_{t+1}} + \beta \cdot E_t \xi_{t+1} \cdot \eta \cdot \left( \frac{K_{t+1}}{A_{1,t+1} \cdot n_{t+1}} \right)^{\eta-1} &= 0, \\ \left( \frac{\phi}{\lambda_t \cdot c_t} - 1 \right) \cdot \Omega_t \cdot A_{3,t} \cdot k \cdot q_t - q_t + \beta \cdot (1-\delta) \cdot E_t \frac{\lambda_{t+1}}{\lambda_t} \cdot q_{t+1} + \beta \cdot \eta \cdot E_t \frac{\xi_{t+1}}{\lambda_{t+1}} \cdot \frac{\lambda_{t+1}}{\lambda_t} \cdot \left( \frac{K_{t+1}}{A_{1,t+1} \cdot n_{t+1}} \right)^{\eta-1} &= 0. \end{aligned} \quad (5.25)$$

FOC wrt  $B_{t+1}$ , 得

$$\begin{aligned} \beta \cdot E_t \left( \frac{\lambda_{t+1}}{P_{t+1}^A} \right) - \frac{\lambda_t}{P_t^A \cdot (1+R_t^B)} + \phi \cdot \frac{1}{c_t} \cdot \frac{\partial c_t}{\partial B_{t+1}} - \lambda_t \cdot \frac{\partial c_t}{\partial B_{t+1}} &= 0, \\ \left( \frac{\phi}{c_t \cdot \lambda_t} - 1 \right) \cdot \left[ \frac{\Omega_t}{P_t^A \cdot (1+R_t^B)} \right] + \beta \cdot E_t \frac{\lambda_{t+1}}{\lambda_t} \cdot \frac{1}{P_{t+1}^A} - \frac{1}{P_t^A \cdot (1+R_t^B)} &= 0, \\ \left( \frac{\phi}{c_t \cdot \lambda_t} - 1 \right) \cdot \Omega_t + \beta \cdot E_t \frac{\lambda_{t+1}}{\lambda_t} \cdot \frac{P_t^A}{P_{t+1}^A} \cdot (1+R_t^B) - 1 &= 0. \end{aligned} \quad (5.26)$$

FOC wrt  $c_t$ , 得

$$\frac{U_{c,t}}{\lambda_t} - 1 = \frac{\phi}{c_t \cdot \lambda_t} - 1 = 0, \quad (5.27)$$

其中  $U_{c,t} = \phi/c_t$  表示消费的边际效用, 由式(5.1)推得。

## 5.4 均衡条件

综上, 模型的均衡条件由 14 个等式构成:

- 家庭部门的预算约束条件式(5.2),
- 家庭部门生产活动的市场出清条件式(5.3),
- 银行部门的资产负债平衡式(5.5),
- 消费存款平衡式(5.4),
- 商业银行的贷款管理 (贷款生产函数) 式(5.8),
- 反映贴现实际债券的变量  $b_{t+1}$  定义式(5.9),
- 反映抵押品边际产出 (放贷) 的变量  $\Omega_t$  定义式(5.16),
- 6 个一阶条件式(5.21), (5.22), (5.23), (5.24), (5.25), (5.26),
- 政府 (包括中央银行) 部门的预算约束条件式(5.10)。

随后的任务就变成了, 在给定

- 3 个外生技术冲击  $\{A_{i,t}\}, i = 1, 2, 3$ , 式(5.11), (5.12), (5.13),

- 由政策制定者设定的政策变量  $\{H_t, g_t, b_t, \text{Transfer}_t\}$ ,

- 外生资本存量  $K_t^5$ ,

的情况下，探讨一组 14 个内生变量的变化，包括

- 家庭部门的行为决策，包括消费、劳动力供应、购入债券  $\{c, n, m, B\}$ ,
- 银行部门的存贷款，以及抵押品对放贷的边际产出  $\{D, L, \Omega\}$ ,
- 政府（含中央银行）的税收  $tax$ ,
- 市场价格和利率，包括劳动力和资本价格，总物价水平，影子价格和国债利率  $\{w, q, P, \lambda, \xi, R^B\}$ 。

## 5.5 利率

第5.4节的均衡条件中，除了政府债券利率  $R_t^B$  之外，并不包括其他利率。这是由于根据模型设定，产品生产部门和银行最终由家庭运营，利润也归家庭所有。

### 5.5.1 资产市场基准利率

首先根据Goodfriend (2005)，假想存在一种完全无风险债券。由于无风险，因此借贷者不需要任何抵押品。根据家庭部门的最优消费——储蓄（投资）决策，这种债券的投资回报率反映该项投资的影子价格，我们设为基准利率  $R_t^T$ 。基于家庭部门优化决策，无风险债券的利率  $R_t^T$  由类似于式(5.26)的方式决定：

$$1 + R_t^T = \frac{1 - \left(\frac{U_{c,t}}{\lambda_t} - 1\right) \cdot \Omega_t}{\beta \cdot E_t \frac{\lambda_{t+1}}{\lambda_t} \cdot \frac{P_t^A}{P_{t+1}^A}},$$

代入家庭的消费最优决策式(5.26)，调整得

$$1 + R_t^T = \frac{1}{\beta} \cdot E_t \frac{\lambda_t}{\lambda_{t+1}} \cdot \frac{P_{t+1}^A}{P_t^A}. \quad (5.28)$$

### 5.5.2 抵押品收益率

#### 流动性债券收益率

无风险利率  $R_t^T$  定义式(5.28)带回式(5.26)

$$\frac{1 + R_t^B}{1 + R_t^T} = 1 - \left(\frac{U_{c,t}}{\lambda_t} - 1\right) \cdot \Omega_t.$$

$\frac{U_{c,t}}{\lambda_t}$  用于比较消费的边际效用与投资的影子价格，可得  $\left(\frac{\phi}{c_t \cdot \lambda_t} - 1\right) > 0$ ； $\Omega_t$  用于近似描述抵押品的边际价值，根据式(5.14)中  $0 < \alpha < 1$  可得  $\Omega > 0$ 。因此我们有  $R_t^T > R_t^B$ ，二者之差反映了流动性债券的收益率，定义为  $LSY_t^B$

$$LSY_t^B \equiv R_t^T - R_t^B \approx \left(\frac{\phi}{c_t \cdot \lambda_t} - 1\right) \cdot \Omega_t. \quad (5.29)$$

<sup>5</sup>即是说在模型的均衡条件下，每一期资本存量均等于其稳态值  $K_t = K$ 。需要指出的是， $K$  本身是由模型所内生决定的。

### 流动性资本收益率

如贷款生产函数式(5.8)所定义, 债券和实物资本都可以用于贷款抵押, 而实物资本比起债券来具有成本劣势, 用系数  $0 < k < 1$  表示。因此可定义流动性资本收益率  $LSY_t^B$ , 基于式(5.29)我们有<sup>6</sup>

$$LSY_t^K = k \cdot LSY_t^B = k \cdot \left( \frac{\phi}{c_t \cdot \lambda_t} - 1 \right) \cdot \Omega_t. \quad (5.30)$$

### 5.5.3 银行间拆借利率 (无抵押)

根据模型设定, 某一商业银行可以在银行间市场以利率  $R_t^{IB}$  筹集资金, 进而贷款给家庭部门, 获得回报  $R_t^T$ 。假定不存在套利行为, 如果该银行的操作有利可图, 即净收益可以弥补全部成本支出的话, 上述情况即可能发生。

如式(5.8)所示, 贷款成本包括抵押品成本和监控成本两部分。假定该银行在银行间市场向其他银行的借款无需抵押品, 则成本主要指监控成本, 即以工资  $w_t$  雇佣劳动力  $m_t$  对贷出给家庭部门的款项做监控的劳动成本支出。

银行的最优行为表现为两阶段优化。第一阶段为成本最小化, 无抵押贷款生产函数的实际边际成本等于劳动的市场价格 (工资), 除以劳动投入的边际产出, 后者由式(5.8)求得:

$$\text{real marginal cost} = \frac{w_t}{\partial \left( \frac{L_t}{P_t^A} \right) / \partial m_t} = \frac{w_t \cdot m_t \cdot v_t}{(1 - \alpha) \cdot (1 - rr) \cdot c_t}. \quad (5.31)$$

第二阶段, 追求利润最大化的银行行为最终会使得下式成立

$$(1 + R_t^{IB}) \cdot (1 + \text{real marginal cost}) = 1 + R_t^T, \quad (5.32)$$

即银行间拆借利率  $R_t^{IB}$  乘以贷款的边际生产成本, 等于基准利率  $R_t^T$ 。类似地, 整理可得

$$R_t^T - R_t^{IB} = \frac{w_t \cdot m_t \cdot v_t}{(1 - \alpha) \cdot (1 - rr) \cdot c_t} \quad (5.33)$$

### 5.5.4 贷款利率 (有抵押)

实际由银行部门向非银行部门贷出的款项, 往往是有抵押贷款。有抵押贷款的成本等于实际边际成本乘以监控的要素份额  $(1 - \alpha)$ 。利润最大化银行会令抵押贷款利率  $R_t^L$  的值等于

$$1 + R_t^L = (1 + R_t^{IB}) \cdot [1 + (1 - \alpha) \cdot \text{real marginal cost}],$$

即抵押贷款的利率应该等于无抵押的银行间拆借利率, 乘以贷款监控的成本支出。进一步整理得

$$\begin{aligned} 1 + R_t^L &= (1 + R_t^{IB}) \cdot \left( 1 + \frac{w_t \cdot m_t \cdot v_t}{(1 - rr) \cdot c_t} \right), \\ R_t^L - R_t^{IB} &= \frac{w_t \cdot m_t \cdot v_t}{(1 - rr) \cdot c_t}. \end{aligned} \quad (5.34)$$

### 5.5.5 存款利率

银行的贷款来自于家庭部门储蓄 (存款)  $D_t$ 。全部存款中有  $rr$  比例作为高能货币  $H_t$  用于日常消费支出, 不产生利息; 余下  $(1 - rr)$  部分可被银行作为贷款  $L_t$ 。因此, 存款利率  $R_t^D$  的值应为

$$\frac{R_t^D}{1 - rr} = R_t^{IB}, \quad (5.35)$$

即贷存比乘以无抵押银行间拆借利率。

<sup>6</sup>更为详尽的说明, 见Goodfriend (2005)。

### 5.5.6 外部融资溢价

#### 无抵押贷款的外部融资溢价

家庭部门向银行做无抵押贷款，需要支付的利息包括两部分，一为无抵押银行间拆借利率  $R_t^{IB}$ ，一为银行贷款生产函数的实际边际成本式(5.31)，二者之和应等于资产市场的无抵押贷款基准利率  $R_t^T$ 。从这个意义上来说，贷款的实际边际成本即反映了无抵押贷款外部融资的溢价 (uncollateralized external financial premium, UEFP)<sup>7</sup>，我们用  $R_t^T$  和  $R_t^{IB}$  的利差来表示，由式(5.33)得

$$UEFP_t = \text{real marginal cost} = R_t^T - R_t^{IB} = \frac{w_t \cdot m_t \cdot v_t}{(1 - \alpha) \cdot (1 - rr) \cdot c_t}. \quad (5.36)$$

#### 有抵押贷款的外部融资溢价

类似地，家庭向银行做抵押贷款，需要支付的利息包括两部分，一为银行间拆借利率  $R_t^{IB}$ ，一为有抵押贷款生产函数的边际成本，二者之和应等于有抵押贷款利率  $R_t^L$ 。此时有抵押贷款的边际成本反映了抵押贷款的外部融资溢价 (collateralized external financial premium, CEFP)，我们用  $R_t^L$  和  $R_t^{IB}$  的利差来表示，由式(5.34)得

$$CEFP_t = (1 - \alpha) \cdot \text{real marginal cost} = R_t^L - R_t^{IB} = \frac{w_t \cdot m_t \cdot v_t}{(1 - rr) \cdot c_t} \quad (5.37)$$

#### 比较

联立 (5.36)- (5.37) 可得

$$CEFP_t = (1 - \alpha) \cdot UEFP_t \quad (5.38)$$

外部融资溢价  $CEFP_t < UEFP$ ，或利差  $(R_t^L - R_t^{IB}) < (R_t^T - R_t^{IB})$ ，反映了家庭部门当提供抵押品时所做的借款，只需要支付贷款生产函数边际成本中的一部分  $(1 - \alpha)$ ，这一部分被银行用来对贷出款项做监控；余下的  $\alpha$  部分得以减免。

### 5.5.7 抵押品收益率与外部融资溢价的关联

由于我们假定银行部门是充分竞争的，并且银行的贷款生产函数是规模报酬不变的，那么均衡状态下银行的利润为 0，即

$$\underbrace{(R_t^T - R_t^{IB}) \cdot \frac{L_t}{P_t}}_{\text{利息收入}} = \underbrace{b_{t+1} \cdot LSY_t^B + q_t \cdot K_{t+1} \cdot LSY_t^K}_{\substack{\text{债券抵押品利息} \\ \text{实物资本抵押品利息}}} + \underbrace{w_t \cdot w_t}_{\text{监控成本支出}}.$$

从这个意义上讲， $LSY_t^{B,K}$  可以看做是银行支付给家庭部门借款人的“租金”：借款人向银行提供抵押品，抵押品作为投入要素进入银行部门的贷款生产函数中。抵押品的回报率因此低于假想中的无风险债券的基准利率  $R_t^T$ ；两种利率之差体现了抵押品作为贷款生产函数的投入要素的回报，这种回报以利率优惠的形式，反映在银行对家庭借款人收取的贷款利息减免中。

流动性抵押品的收益率，与外部融资溢价之间的关联，可以由式(5.30)、(5.22)、(5.37)联立求得

$$LSY_t^K = k \cdot LSY_t^B = k \cdot \frac{w_t \cdot m_t}{(1 - \alpha) \cdot c_t} = k \cdot \left[ \frac{1 - rr}{(1 - \alpha) \cdot v_t} \right] \cdot CEFP_t. \quad (5.39)$$

<sup>7</sup> 关于外部融资溢价，见 Bernanke et al. (1999)。

## 5.6 稳定状态

### 5.6.1 核心稳态系统

如前述，模型的均衡条件包括 22 个变量。假定稳定状态下是确定的 (deterministic) 并且

1. 通胀率为零  $P^A = P = 1$ ,
2. 资产价格  $q = 1$ ,
3. Transfer =  $g = \text{tax} = 0$ 。

由式(5.11), (5.12), (5.13)可得

$$g_{A_1} = g_{A_2} = 1 + \gamma, g_{A_3} = 0, \quad (5.40)$$

由式(5.40)可得， $\{c_t, K_t, w_t, \lambda_t\}_{t=0}^{\infty}$  的增速均为  $1 + \gamma$ 。

根据定义式(5.16)可知

$$\Omega = \frac{\alpha}{\frac{b}{c} + k \cdot \frac{K}{c}} \quad (5.41)$$

根据一阶条件可知：

式(5.21)  $\Rightarrow$

$$\lambda = \frac{1 - \phi}{(1 - n - m) \cdot w} \quad (5.42)$$

式(5.22)  $\Rightarrow$

$$w = (1 - \alpha) \cdot \left( \frac{\phi}{\lambda \cdot c} - 1 \right) \cdot \frac{c}{m} \quad (5.43)$$

式(5.24)  $\Rightarrow$

$$\frac{\xi}{\lambda} = -\frac{1 - \theta}{\theta} \quad (5.44)$$

式(5.23)结合式(5.44)  $\Rightarrow$

$$w = -(1 - \eta) \cdot \frac{1 - \theta}{\theta} \cdot \left( \frac{K}{n} \right)^{\eta} \quad (5.45)$$

式(5.25)  $\Rightarrow$

$$k \cdot \Omega \cdot \left( \frac{\phi}{\lambda \cdot c} - 1 \right) - 1 + \frac{\beta}{1 + \gamma} \cdot \left[ (1 - \delta) - \eta \cdot \frac{1 - \theta}{\theta} \cdot \left( \frac{K}{n} \right)^{\eta-1} \right] = 0 \quad (5.46)$$

家庭部门总预算约束式(5.2)改写为

$$\begin{aligned} q_t \cdot [(1 - \delta) \cdot K_t - K_{t+1}] &- \left[ \frac{H_t}{P_t^A} - \frac{H_{t-1}}{P_t^A} + \frac{B_{t+1}}{P_t^A \cdot (1 + R_t^B)} - \frac{B_t}{P_t^A} + \text{tax}_t \right] \\ &+ w_t \cdot [(n_t^s + m_t^s) - (n_t + m_t)] + c_t^A \cdot \left( \frac{P_t}{P_t^A} \right)^{1-\theta} = c_t, \end{aligned}$$

其中等式左侧第二个中括号部分，根据政府部门的预算约束条件式(5.10)等于  $g_t$ 。第三个中括号，根据市场出清条件假定为 0。第四部分，根据另一个市场出清条件式(5.3)予以替代，得

$$q_t \cdot [(1 - \delta) \cdot K_t - K_{t+1}] - g_t + K_t^{\eta} \cdot (A_{1,t} \cdot n_t)^{1-\eta} \cdot \frac{P_t}{P_t^A} = c_t,$$

则稳定状态下我们有

$$\left( \frac{K}{n} \right)^{\eta} - \delta \cdot \left( \frac{K}{n} \right) = \frac{c}{n} \quad (5.47)$$

式(5.14)  $\Rightarrow$

$$1 = \frac{v \cdot F}{1 - rr} \cdot \left( \frac{b}{c} + k \cdot \frac{K}{c} \right)^\alpha \cdot \left( \frac{m}{c} \right)^{1-\alpha} \quad (5.48)$$

则 22 个变量的稳定状态系统，可以改写为含有 7 个核心变量  $\{c, K, w, \lambda, \Omega, m, n\}$  的 7 个方程组：(5.41), (5.42), (5.43), (5.45), (5.46), (5.47), (5.48)。

### 5.6.2 稳定状态利率

# 第六章 Cash in advance

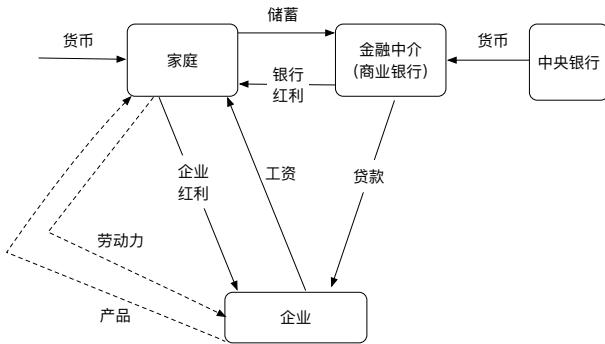


图 6.1: CIA DSGE 模型结构图

注: 实线表示名义资金流动。虚线表示实际物质流动。

## 6.1 家庭部门

假定一个永生家庭, 追求期望效用的最大化

$$\max E_0 \sum_{t=0}^{\infty} \beta^t [(1 - \phi) \cdot \ln C_t + \phi \cdot \ln(1 - H_t)]. \quad (6.1)$$

其中变量  $C_t$  表示消费,  $0 < H_t < 1$  表示劳动时间占全部可支配时间的比重。系数  $0 < \beta < 1$  表示时间折旧,  $0 < \phi < 1$  表示消费  $C_t$  相对于休闲  $(1 - H_t)$ , 对效用的重要程度。

家庭效用最大化受到 2+1 个条件限制。

第 1 个是家庭的现金平衡条件, 假定  $t$  时期家庭需要存留足够数量的现金, 以满足同期全部消费需求:

$$W_t \cdot H_t + M_t - D(t) \geq P_t \cdot C_t, \quad (6.2)$$

其中等式左侧  $W_t$  表示由市场给定的劳动力平均工资,  $M_t$  表示  $t-1$  期末经济体中的基础货币存量,  $D_t$  表示家庭部门的储蓄, 假定储蓄全部以存款形式  $D_t$  存入银行部门,  $M_t$ 。等式右侧  $P_t$  表示市场给定的最终产品价格。cash balance 是 CIA-DSGE 模型的核心设定之一。

第 2 个是家庭的资源约束条件

$$M_{t+1} \geq (M_t - D_t + W_t \cdot H_t - P_t \cdot C_t) + R_{H,t} \cdot D_t + (F_t + B_t), \quad (6.3)$$

等式右侧分为三个部分。第一部分为现金平衡条件如式(6.2), 第二部分为家庭存款  $D_t$  的收益,  $R_{H,t}$  为银行提供给家庭的无风险名义存款利率, 第三部分为家庭部门收到的分红<sup>1</sup>, 分别来自银行部门

<sup>1</sup>模型假定银行和企业的资产最终由家庭部门所有, 如Christiano and Eichenbaum (1992)。

$B_t$  和企业部门  $F_t$ 。根据资源约束条件,  $t$  期家庭需要决定全部收入中提取多少现金存留在手上, 以满足  $t$  期的消费需求; 剩余部分以  $D_t$  形式存入银行, 获取利润, 以供  $t+1$  期消费所用。

第 3 个是家庭净资产非负条件, 假定所有时期家庭的储蓄均大于 0

$$D_t \geq 0. \quad (6.4)$$

## 6.2 企业部门

假定完全竞争市场环境下, 一家典型企业的生产函数

$$Y_t = K_t^\alpha \cdot (A_t \cdot N_t)^{1-\alpha}, \quad (6.5)$$

符合 Cobb-Douglas 形式。投入要素有 2。第 1 个投入要素为资本存量  $K_t$ , 通过企业往期投资  $I_{t-s}$ ,  $s > 0$ , 经折旧后累积形成,

$$K_{t+1} = I_t + (1 - \delta) \cdot K_t, \quad (6.6)$$

其中常系数  $0 < \delta < 1$  表示自然折旧率。

第 2 个投入要素为劳动力  $N_t$ 。基于市场给定的名义工资率  $W_t$ , 企业向银行借贷  $L_t$ , 用于支付全部劳动工资

$$L_t \geq W_t \cdot N_t, \quad (6.7)$$

银行借贷需要支付市场利率  $R_{F,t}$ 。

此外式(6.5)中,  $\alpha$  表示资本的产出弹性。外生变量  $A_t$  用于描述劳动扩张型技术进步, 见第6.6节。

### 6.2.1 生产规模

如前文所述, 完全竞争市场上, 利润最大化企业以  $R_{F,t}$  利率从银行获取贷款  $L_t$  用于成本支出, 即支付员工工资  $W_t \cdot N_t$ , 展开生产。企业将逐渐增加生产规模, 直到雇佣额外 1 单位劳动力的边际收益, 等于向银行借贷所需支付的利息。

$$\frac{\partial [P_t \cdot Y_t - W_t \cdot N_t]}{\partial N_t} + \frac{\partial [(1 - R_{F,t}) \cdot L_t]}{\partial N_t} = 0. \quad (6.8)$$

### 6.2.2 利润的使用

对于已经生成的利润,  $t$  期企业核心决策为, 决定利润中多少比例留作投资  $I_t$ , 以增加  $t+1$  期的资本存量  $K_{t+1}$ , 用于扩大再生产; 剩余利润以红利  $F_t$  形式返回家庭部门, 用于  $t+1$  期的消费  $C_{t+1}$ 。企业行为可以表示如下

$$\max E_0 \sum_{t=0}^{\infty} \beta^{t+1} \cdot \frac{F_t}{P_{t+1} \cdot C_{t+1}}. \quad (6.9)$$

企业利润最大化生产活动受到 2 个条件限制。第 1 个是预算约束条件,

$$P_t \cdot (Y_t - I_t) + L_t \geq F_t + W_t \cdot N_t + R_{F,t} \cdot L_t, \quad (6.10)$$

等式右侧表示全部生产收入和来自银行的贷款; 左侧支出部分分别为给家庭部门的红利, 劳动力工资, 和向银行支付的贷款利息。

第 2 个是非负借贷条件式(6.7), 确保当期向银行借款数足以支付工资。

## 6.3 银行部门

类似地,  $t$  期的银行利润以分红  $B_t$  形式返还给家庭部门, 用于  $t+1$  期的消费。利润最大化的银行追求

$$\max E_t \sum_{t=0}^{\infty} \beta^{t+1} \cdot \frac{B_t}{P_{t+1} \cdot C_{t+1}}. \quad (6.11)$$

银行利润最大化行为受到 2 个条件限制。第 1 个是预算约束条件

$$B_t + R_{H,t} \cdot D_t \leq R_{F,t} \cdot L_t + D_t + X_t - L_t, \quad (6.12)$$

其中  $X_t$  表示  $t$  期中央银行对市场释放的基础货币, 定义为  $t+1$  期末基础货币存量和  $t$  期末基础货币存量的差,

$$X_t = M_{t+1} - M_t. \quad (6.13)$$

第 2 个是资产负债平衡约束, 要求同期的银行负债不得超过资产

$$X_t + D_t \leq L_t. \quad (6.14)$$

## 6.4 一阶条件

### 6.4.1 家庭部门

根据最大化式(6.1)及约束条件式(6.2)-(6.4), 建家庭部门 Lagrangian

$$\begin{aligned} \mathcal{L} = & E_0 \sum_{t=0}^{\infty} \beta^t \cdot \{ [(1-\phi) \cdot \ln C_t + \phi \cdot \ln(1-H_t)] \\ & + \lambda_t \cdot [M_t - D_t + W_t \cdot H_t - P_t \cdot C_t + R_{H,t} \cdot D_t + (F_t + B_t - M_{t+1})] \}. \end{aligned}$$

FOCs:

$$\frac{\partial \mathcal{L}}{\partial C_t} = 0 \Rightarrow \lambda_t = \frac{1-\phi}{P_t \cdot C_t}, \quad (6.15)$$

$$\frac{\partial \mathcal{L}}{\partial H_t} = 0 \Rightarrow \lambda_t \cdot W_t = \frac{\phi}{1-H_t}, \quad (6.16)$$

$$\frac{\partial \mathcal{L}}{\partial M_{t+1}} = 0 \Rightarrow \frac{\partial \beta \cdot E_t \cdot (\lambda_{t+1} \cdot R_{H,t+1} \cdot D_{t+1})}{\partial M_{t+1}} = \lambda_t. \quad (6.17)$$

### 6.4.2 企业部门

利用将式(6.6)、(6.5) 代入(6.10)替换  $I_t$ 、 $Y_t$ , 与式 (6.7)共同构成约束条件

$$P_t \cdot \left[ K_t^\alpha \cdot (A_t \cdot N_t)^{1-\alpha} - K_{t+1} + (1-\delta) \cdot K_t \right] \geq F_t + W_t \cdot N_t + R_{F,t} \cdot L_t - L_t.$$

建企业部门 Lagrangian

$$\begin{aligned} \mathcal{L} = & E_t \sum_{t=0}^{\infty} \beta^{t+1} \cdot \left\{ \frac{F_t}{P_{t+1} \cdot C_{t+1}} \right. \\ & \left. + \lambda_t \cdot \left\{ P_t \cdot \left[ K_t^\alpha \cdot (A_t \cdot N_t)^{1-\alpha} - K_{t+1} + (1-\delta) \cdot K_t \right] - F_t - W_t \cdot N_t - R_{F,t} \cdot L_t + L_t \right\} \right\}. \end{aligned}$$

FOCs:

$$\frac{\partial \mathcal{L}}{\partial F_t} = 0 \Rightarrow \lambda_t = E_t \frac{1}{P_{t+1} \cdot C_{t+1}}, \quad (6.18)$$

$$\frac{\partial \mathcal{L}}{\partial N_t} = 0 \Rightarrow P_t \cdot K_t^\alpha \cdot (1-\alpha) \cdot A_t^{1-\alpha} \cdot N_t^{-\alpha} = W_t, \quad (6.19)$$

$$\frac{\partial \mathcal{L}}{\partial K_{t+1}} = 0 \Rightarrow \lambda_t \cdot P_t = \beta \cdot E_t \lambda_{t+1} \cdot P_{t+1} \cdot \left[ \alpha \cdot K_{t+1}^{\alpha-1} \cdot (A_{t+1} \cdot N_{t+1})^{1-\alpha} + (1-\delta) \right]. \quad (6.20)$$

### 6.4.3 银行部门

建立银行部门 Lagrangian

$$\begin{aligned} \mathcal{L} = E_t \sum_{t=0}^{\infty} \beta^{t+1} \cdot & \left\{ \frac{B_t}{P_{t+1} \cdot C_{t+1}} \right. \\ & \left. + \lambda_t \cdot (D_t + R_{F,t} \cdot L_t - R_{H,t} \cdot D_t - L_t + X_t - B_t) \right\}. \end{aligned}$$

### 6.4.4 一阶条件的整理

式(6.15)代入式(6.16)以替换  $\lambda_t$ , 可得劳动力供应的决定式

$$W_t = \frac{\phi}{1-\phi} \cdot \frac{P_t \cdot C_t}{1-H_t}. \quad (6.21)$$

式(6.15)代入式(6.17)以替换  $\lambda_t$ ; 结合式(6.12)、(6.13)可得跨期消费的 Euler equation

$$\frac{1}{P_t \cdot C_t} = R_{H,t} \cdot \beta \cdot E_t \frac{1}{P_{t+1} \cdot C_{t+1}}. \quad (6.22)$$

(6.18)代入式(6.19)以替换  $\lambda_t$ , 可得劳动力需求的决定式

$$W_t = (1-\alpha) \cdot P_t \cdot K_t^\alpha \cdot A_t^{1-\alpha} \cdot N_t^{-\alpha}. \quad (6.23)$$

(6.18)代入式(6.20)以替换  $\lambda_t$ , 可得产品市场上跨期消费决策的 Euler equation

$$E_t \frac{P_t}{P_{t+1} \cdot C_{t+1}} = \beta \cdot E_t \cdot \left\{ \left( \frac{P_{t+1}}{P_{t+2} \cdot C_{t+2}} \right) \cdot \left[ \alpha \cdot K_{t+1}^{\alpha-1} \cdot (A_{t+1} \cdot N_{t+1})^{1-\alpha} + (1-\delta) \right] \right\}. \quad (6.24)$$

## 6.5 市场均衡及部门最优行为

### 6.5.1 4 个市场均衡条件

假定下述 4 个完全竞争市场均处于均衡条件, 完全出清。

#### 劳动力市场均衡

家庭部门的劳动力供应等于企业部门的劳动力需求

$$H_t = N_t. \quad (6.25)$$

#### 产品市场均衡

企业部门的产出完全用于企业再投资和家庭部门的消费

$$Y_t = C_t + I_t. \quad (6.26)$$

### 货币市场均衡

家庭部门对现金（消费）的需求，等于市场上的货币供应

$$P_t \cdot C_t = M_{t+1}. \quad (6.27)$$

### 信贷市场均衡

商业银行的负债等于资产，即式(6.14)改写为

$$X_t + D_t = L_t. \quad (6.28)$$

## 6.5.2 部门最优行为

将几个市场的均衡行为纳入部门最优行为的决策分析。

### 劳动力市场

劳动力市场均衡条件下，式(6.7)取等号，可求得平均工资的决定

$$W_t = \frac{L_t}{N_t}. \quad (6.29)$$

将劳动力市场出清式(6.25)和均衡工资决定式(6.29)代入一阶条件式(6.21)，分别替代  $W_t, H_t$  可得家庭部门的同期最优消费——劳动供应决策。

$$\frac{L_t}{N_t} = \frac{\phi}{1 - \phi} \cdot \frac{P_t \cdot C_t}{1 - N_t}. \quad (6.30)$$

### 产品市场

产品市场最优条件，见跨期消费平滑 Euler equation 式(6.24)。

### 货币市场

式(6.13)代入货币市场均衡条件式(6.27)可得

$$P_t \cdot C_t = M_t + X_t, \quad (6.31)$$

即全部现金需求等于市场上的货币存量和当期新增货币投放量之和。

### 信贷市场

首先由信贷市场均衡式(6.28)可见，根据银行利润为零的假定

$$R_{H,t} \cdot D_t = R_{F,t} \cdot (L_t - X_t),$$

由此我们有

$$R_{F,t} = R_{H,t} \equiv R_t. \quad (6.32)$$

式(6.5)、(6.29)和(6.32)代入式(6.8)分别替换  $Y_t, L_t, R_{F,t}$ ，整理后，可得信贷市场最优条件

$$R_t = \frac{P_t \cdot K_t^\alpha \cdot (1 - \alpha) \cdot A_t^{1-\alpha} \cdot N_t^{-\alpha}}{W_t}. \quad (6.33)$$

## 6.6 外生冲击

经济模型中存在两种随机外生冲击。第 1 种是作用于实体经济部门的劳动增强型技术冲击，假定符合下述形式

$$\ln A_t = \gamma + \ln A_{t-1} + \varepsilon_{A,t}, \quad \varepsilon_{A,t} \sim \mathcal{N}(0, \sigma_A^2), \quad \text{corr}(\varepsilon_{A,t}, \varepsilon_{A,s}) = 0, \forall s \neq t, \quad (6.34)$$

$\gamma > 0$  表示确定性趋势。不可提前预支的当期技术波动用 innovation 项  $\varepsilon_{A,t}$  来反映。

由定义式(6.34)可得外生技术冲击的增速

$$\frac{A_t}{A_{t-1}} = \exp(\gamma + \varepsilon_{A,t}). \quad (6.35)$$

第 2 种是作用于非实体非实体经济部门的货币投放增速的冲击。假定货币投放增速表述为

$$m_t \equiv \frac{M_{t+1}}{M_t}, \quad (6.36)$$

随机冲击符合下述形式

$$\ln m_t = (1 - \rho) \cdot \ln m^* + \rho \cdot \ln m_{t-1} + \varepsilon_{m,t}, \quad \varepsilon_{m,t} \sim \mathcal{N}(0, \sigma_m^2), \quad \text{corr}(\varepsilon_{m,t}, \varepsilon_{m,s}) = 0, \forall s \neq t. \quad (6.37)$$

$m^*$  表示  $m_t$  的无条件均值。 $m^*$  和  $0 < \rho < 1$  一道用于描述常规货币政策，进而二者的变化反映了货币政策的调整。模型引入随机 innovation  $\varepsilon_{m,t}$  表示当期影响货币供应增速变化的波动因素 (Sims et al., 1982)。

## 6.7 均衡方程组

模型均衡方程组由 9 个内生变量和 2 个外生变量  $\{K_{t+1}, N_t, D_t, C_t, L_t, P_t, W_t, Y_t, R_t, A_t/A_{t-1}, m_t\}$  共同构成<sup>2</sup>。对应一组结构方程如下，每个方程的经济学含义在第6.8节做简要描述：

式(6.29)、(6.33)代入(6.33)以替代  $W_t, R_t$  得

$$\frac{1}{P_t \cdot C_t} = \frac{\beta \cdot P_t \cdot K_t^\alpha \cdot (1 - \alpha) \cdot A_t^{1-\alpha} \cdot N_t^{1-\alpha}}{L_t \cdot E_t P_{t+1} \cdot C_{t+1}}. \quad (6.38)$$

式(6.24)  $\Rightarrow$

$$E_t \frac{P_t}{P_{t+1} \cdot C_{t+1}} = \beta \cdot E_t \cdot \left\{ \left( \frac{P_{t+1}}{P_{t+2} \cdot C_{t+2}} \right) \cdot \left[ \alpha \cdot K_{t+1}^{\alpha-1} \cdot (A_{t+1} \cdot N_{t+1})^{1-\alpha} + (1 - \delta) \right] \right\}.$$

式(6.30)  $\Rightarrow$

$$\frac{L_t}{N_t} = \frac{\phi}{1 - \phi} \cdot \frac{P_t \cdot C_t}{1 - N_t}.$$

式(6.6)、(6.5)代入式(6.26)以替换  $I_t, Y_t \Rightarrow$

$$K_t^\alpha \cdot (A_t \cdot N_t)^{1-\alpha} = C_t + K_{t+1} - (1 - \delta) \cdot K_t. \quad (6.39)$$

式(6.27)  $\Rightarrow$

$$P_t \cdot C_t = M_{t+1}.$$

式(6.13)代入(6.28)以替代  $X_t \Rightarrow$

$$L_t = D_t + M_{t+1} - M_t. \quad (6.40)$$

<sup>2</sup>有的时候还需要考虑一个派生变量，通货膨胀  $\pi_t = \frac{P_t}{P_{t-1}}$ 。

式(6.5)  $\Rightarrow$

$$Y_t = K_t^\alpha \cdot (A_t \cdot N_t)^{1-\alpha},$$

式(6.29)  $\Rightarrow$

$$W_t = \frac{L_t}{N_t}.$$

式(6.33)  $\Rightarrow$

$$R_t = \frac{P_t \cdot K_t^\alpha \cdot (1-\alpha) \cdot A_t^{1-\alpha} \cdot N_t^{-\alpha}}{W_t}.$$

外生冲击式(6.34)  $\Rightarrow$

$$\ln A_t = \gamma + \ln A_{t-1} + \varepsilon_{A,t}, \quad \varepsilon_{A,t} \sim \mathcal{N}(0, \sigma_A^2), \quad \text{corr}(\varepsilon_{A,t}, \varepsilon_{A,s}) = 0, \forall s \neq t,$$

外生冲击式(6.37)  $\Rightarrow$

$$\ln m_t = (1 - \rho) \cdot \ln m^* + \rho \cdot \ln m_{t-1} + \varepsilon_{m,t}, \quad \varepsilon_{m,t} \sim \mathcal{N}(0, \sigma_m^2), \quad \text{corr}(\varepsilon_{m,t}, \varepsilon_{m,s}) = 0, \forall s \neq t.$$

## 6.8 去趋势的均衡状态

第6.7节的方程组是非平稳的。非平稳来自两方面因素，第一是随机技术冲击和货币冲击的趋势影响。即便假定冲击均为零，第二方面因素在于各变量之间没有统一的稳定状态：实际变量的增速等于  $A_t$  的增速（除了劳动力  $N_t$ ）；名义变量的增速等于  $M_t$  的增速；总物价水平的增速等于  $M_t/A_t$  的增速。对上节的变量做去趋势调整：

$$\begin{cases} (\text{名义变量}) \quad \hat{\mathcal{B}}_t = \frac{\mathcal{B}_t}{A_t}, \text{ 其中 } \mathcal{B}_t = [Y_t, C_t, I_t, K_{t+1}]', \\ (\text{实际变量}) \quad \hat{\mathcal{C}}_t = \frac{\mathcal{C}_t}{M_t}, \text{ 其中 } \mathcal{C}_t = [W_t, D_t, L_t]', \\ (\text{名义变量}) \quad \hat{P}_t = \frac{P_t \cdot A_t}{M_t}, \\ (\text{其他}) \quad \{N_t, R_t\} \text{ 保持不变, 根据式(6.36)用 } m_t - 1 \text{ 替代 } \frac{X_t}{M_t}. \end{cases}$$

由去趋势变量构成的新方程组如下。

式(6.38)  $\Rightarrow$

$$\begin{aligned} \frac{1}{\left(\frac{\hat{P}_t \cdot M_t}{A_t}\right) \cdot \left(\hat{C}_t \cdot A_t\right)} &= \frac{\beta \cdot \left(\frac{\hat{P}_t \cdot M_t}{A_t}\right) \cdot \left(\hat{K}_t^\alpha \cdot A_{t-1}^\alpha\right) \cdot (1-\alpha) \cdot A_t^{1-\alpha} \cdot N_t^{1-\alpha}}{\left(\hat{L}_t \cdot M_t\right) \cdot E_t \cdot \left(\frac{\hat{P}_{t+1} \cdot M_{t+1}}{A_{t+1}}\right) \cdot \left(\hat{C}_{t+1} \cdot A_{t+1}\right)}, \\ \frac{1}{\hat{P}_t \cdot \hat{C}_t \cdot M_t} &= \frac{(1-\alpha) \cdot \beta \cdot \hat{P}_t \cdot M_t \cdot \hat{K}_t^\alpha \cdot N_t^{1-\alpha} \cdot \left(\frac{A_t}{A_{t-1}}\right)^{-\alpha}}{\left(\hat{L}_t \cdot M_t\right) \cdot E_t M_{t+1} \cdot \hat{P}_{t+1} \cdot \hat{C}_{t+1}}, \\ \frac{1}{\hat{P}_t \cdot \hat{C}_t} &= \frac{(1-\alpha) \cdot \beta \cdot \frac{\hat{P}_t \cdot \hat{K}_t^\alpha \cdot N_t^{1-\alpha}}{\hat{L}_t \cdot m_t} \cdot \exp[-\alpha \cdot (\gamma + \varepsilon_{A,t})]}{E_t \hat{P}_{t+1} \cdot \hat{C}_{t+1}}. \end{aligned} \tag{6.41}$$

反映信贷市场的最优跨期决策，使得今天放弃 1 单位消费用于储蓄，导致明天消费增加的现值，等于今天 1 单位消费带来的满足。

式(6.24)  $\Rightarrow$

$$\frac{P_t}{E_t P_{t+1} \cdot C_{t+1}} = \frac{\frac{\hat{P}_t \cdot M_t}{A_t}}{E_t \left(\frac{\hat{P}_{t+1} \cdot \hat{M}_{t+1}}{A_{t+1}}\right) \cdot \left(\hat{C}_{t+1} \cdot A_{t+1}\right)} = \frac{\hat{P}_t}{E_t \hat{P}_{t+1} \cdot \hat{C}_{t+1} \cdot m_t \cdot A_t}$$

进而我们有

$$\begin{aligned} E_t \frac{\hat{P}_t}{\hat{P}_{t+1} \cdot \hat{C}_{t+1} \cdot m_t} &= \beta \cdot E_t \frac{\hat{P}_{t+1}}{\hat{P}_{t+2} \cdot \hat{C}_{t+2} \cdot m_{t+1} \cdot \left(\frac{A_{t+1}}{A_t}\right)} \cdot \left[ \alpha \cdot \hat{K}_{t+1}^{\alpha-1} \cdot N_{t+1}^{1-\alpha} \cdot \left(\frac{A_{t+1}}{A_t}\right)^{1-\alpha} + (1-\delta) \right], \\ E_t \frac{\hat{P}_t}{\hat{P}_{t+1} \cdot \hat{C}_{t+1} \cdot m_t} &= \beta \cdot E_t \frac{\hat{P}_{t+1}}{\hat{P}_{t+2} \cdot \hat{C}_{t+2} \cdot m_{t+1}} \cdot \left\{ \alpha \cdot \hat{K}_{t+1}^{\alpha-1} \cdot N_{t+1}^{1-\alpha} \cdot \exp[-\alpha \cdot (\gamma + \varepsilon_{A,t+1})] \right. \\ &\quad \left. + (1-\delta) \cdot \exp[-1 \cdot (\gamma + \varepsilon_{A,t+1})] \right\}. \end{aligned} \quad (6.42)$$

反映产品市场的跨期最优决策，确保经济体整体的跨期消费平滑。

式(6.30)  $\Rightarrow$

$$\frac{\hat{L}_t}{N_t} = \frac{\phi}{1-\phi} \cdot \frac{\hat{P}_t \cdot \hat{C}_t}{1-N_t}. \quad (6.43)$$

反映劳动力市场的当期最优决策：家庭部门的劳动力供应（企业部门的劳动力需求），和消费——休闲边际替代率之间的关系。

式(6.39)  $\Rightarrow$

$$\hat{K}_t^\alpha \cdot N_t^{1-\alpha} \cdot \exp[-\alpha \cdot (\gamma + \varepsilon_{A,t})] = \hat{C}_t + \hat{K}_{t+1} - (1-\delta) \cdot \hat{K}_t \cdot \exp[-1 \cdot (\gamma + \varepsilon_{A,t})]. \quad (6.44)$$

反映总量上的资源约束条件：当期产出全部用于当期消费和当期投资；当期投资贡献于下一期的资本存量积累。

式(6.27)  $\Rightarrow$

$$\hat{P}_t \cdot \hat{C}_t = m_t. \quad (6.45)$$

反映货币市场的最优决策：货币的名义消费需求等于名义货币供应量。名义货币供应量等于名义现有货币存量与当期货币注入量之和。

式(6.40)  $\Rightarrow$

$$\hat{L}_t = (m_t - 1) + \hat{D}_t. \quad (6.46)$$

反映信贷市场的最优决策：商业银行追求资产——负债平衡。

式(6.5)  $\Rightarrow$

$$\hat{Y}_t = \hat{K}_t^\alpha \cdot N_t^{1-\alpha} \cdot \exp[-\alpha \cdot (\gamma + \varepsilon_{A,t})]. \quad (6.47)$$

反映经济整体的投入产出关系。

式(6.29)  $\Rightarrow$

$$\hat{W}_t = \frac{\hat{L}_t}{N_t}. \quad (6.48)$$

反映企业对商业银行的贷款需求，与企业生产活动中的劳动力需求（家庭部门的劳动力供应）之间的关系。

式(6.33)  $\Rightarrow$

$$R_t = (1-\alpha) \cdot \frac{\hat{P}_t \cdot \hat{K}_t^\alpha \cdot N_t^{1-\alpha}}{\hat{L}_t} \cdot \exp[-\alpha \cdot (\gamma + \varepsilon_{A,t})]. \quad (6.49)$$

反映信贷市场的最优决策：均衡利率水平表示额外 1 单位劳动力投入的边际产出效果，等于借贷用于雇佣此额外 1 单位劳动力所需支付的成本。

外生技术冲击和货币冲击，如式(6.35)、(6.37)所示。

此外，实际观测到的经济增速和稳定状态下的经济增速之间的关系如下

$$\frac{Y_t}{Y_{t-1}} = \frac{\hat{Y}_t}{\hat{Y}_{t-1}} \cdot \exp(\gamma + \varepsilon_{A,t}). \quad (6.50)$$

类似地，实际观测到的通货膨胀和稳定状态下的通货膨胀之间关系如下

$$\frac{P_t}{P_{t-1}} = \frac{\hat{P}_t}{\hat{P}_{t-1}} \cdot m_{t-1} \cdot \exp[-1 \cdot (\gamma + \varepsilon_{A,t})] \quad (6.51)$$



## **第二部分**

### **模型求解**



# 第七章 DSGE 模型求解方法简论

## 7.1 简论的简论

过去的三十多年来，宏观经济学研究经历了一场飞速变革。这场变革始于 Kydland and Prescott (1982) 利用 RBC 模型研究美国经济。这种研究方法逐渐成为宏观经济学的标准范式之一 (An and Schorfheide, 2007; Fernández-Villaverde, 2010)。

随后 RBC 模型逐渐扩展到新凯恩斯主义模型。经典教材如 Galí (2008); Woodford (2011)。然而新凯恩斯主义模型也远非完美无缺，随着新的问题逐渐被发现，学术界对模型做了进一步的修正和扩展，使模型对现实的拟合程度越来越高，如

- 根据基准黏性价格假定所生成的一些重要经济变量的时间序列，与实际观察到的经济现实相比，出入较大——真实经济世界中的通货膨胀，产出，实际工资等时间序列数据都具有很高的持续性——为了使得模型与现实相贴近，模型设定中就需要引入看起来异常高的名义粘性。
- 因此产生了一系列对基准新凯恩斯模型的扩展，它们的基本思路是，加入 Calvo (1983) 定价机制，从而可以在控制名义粘性不至于过高的情况下，有效提高模型生成的通货膨胀时间序列数据的持续性 (Rabanal and Rubio-Ramírez, 2005)。

这样新凯恩斯主义分析框架得到了中央银行等政策制定者的青睐。不同中央银行开发了一系列自己的 DSGE 模型，如

- 欧洲中央银行 (ECB) 的 NAWM (New Area-Wide Model)。  
在 Fagan et al. (2005) 的 Area-Wide Model (ARM) 和 Smets and Wouters (2003) 的基础上，Christoffel et al. (2008) 开发了欧元区的 NAWM，一个基于微观经济基础的开放经济模型，广泛应用于欧洲中央银行研究人员对经济系统的预测。
- 加拿大银行 (Bank of Canada) 的 ToTEM (Term-of-Trade Economic Model)，见 Murchison and Rennison (2006)。
- 英格兰银行 (Bank of England) 的 BEQM (Bank of England Quarterly Model)，见 Harrison et al. (2005)。
- 挪威银行 (Norges Bank) 的 NEMO (Norwegian Economy Model)，见 Brubakk et al. (2006)。
- 智利中央银行 (Central Bank of Chile) 的 MAS (Model of Analysis and Simulation)，见 Medina and Soto (2007)。
- 瑞典中央银行 (Sveriges Riksbank) 的 RAMSES (Riksbanks Aggregated Model for Studies of the Economy in Sweden)，见 Adolfson et al. (2007, 2008a)。

- 美联储的 SIGMA。

在Obstfeld and Rogoff (1995) 的开放经济体模型基础上, Erceg et al. (2006) 简历的多国家开放经济模型。

- 国际货币基金组织 (IMF) 的 GEM (Global Economic Model), 见Bayoumi et al. (2004)。

上述中央银行开发的经济预测模型, 多得以下于 DSGE 模型的长处, 具有如下特征:

1. 动态性。关注变量随时间的变化路径, 而不是某单独时间点上的情况。
2. 整体性。致力于解释、预测整体经济运行, 而非仅仅是局部市场。
3. 重视部门均衡。根据经济学理论, 在不同市场中, 重视市场调节机制作用下的供需平衡。
4. 模型中引入随机干扰。

从广义的宏观经济学模型角度来看, DSGE 模型可以分为两大类: RBC 模型和新凯恩斯主义模型。RBC 模型致力于研究在灵活价格环境下的经济周期波动, 常表现为两部分的组合: 一部分是随机内生经济增长模型作为内核, 另一部分是外部真实技术冲击。Kydland and Prescott (1982) 对 RBC 模型 (进而 DSGE 模型) 研究做出了开创性的贡献, 进一步评述可见Cooley (1995)。

RBC 关注实际经济变量, 对货币问题未作深入讨论。而在现实经济世界中, 货币的重要性越发得到重视。于是就有了新凯恩斯主义模型, 大致说来, 它对 RBC 模型做了两方面的扩展, 一是些内部机制的调整, 二是外部冲击来源的选择。就前者而言, 包括比如

1. 将货币和货币发行机构 (中央银行) 引入到分析框架中来,
2. 不再持有完全竞争假定, 在一些部门的行为分析中假定不完全竞争, 如:
  - (a) 产品/服务市场, 和/或劳动力市场是不完全竞争的,
  - (b) 允许私有部门的消费和投资决策在存在刚性的情况下展开: 通过引入黏性价格、黏性工资的设定, 将名义刚性引入模型设定中来。这意味着货币政策会对实际经济变量产生影响。

如何构建一个 DSGE 模型, 进而如何将其应用到宏观经济分析中去, 仍需要我们在理论模型和计算求解两个维度上展开深入研究。就后者而言, 很显然理论模型设定的细节不同, 需要不同的求解方法。但总的说来, 求解流程及方法大同小异: 识别假设条件 → 推出 (一阶) 均衡条件 → 构建结构方程组 → 形成随机差分方程系统 (通常是非线性的) → 对非线性系统做近似线性化 → 求得近似解 → (计算冲击响应方程或二阶矩以) 检验近似解的有效度。

本章主要对后半部分做简要介绍, 大致包括 DSGE 模型的求解方法、参数估计、模型检验等。已有一系列文献对此作了综述, 随着 DSGE 模型类型的不同, 这些文献各有侧重, 分别关注不同的求解方法, 如

- Canova (2009, 2011); Balke et al. (2012) 主要介绍常见的宏观计量经济学方法。
- An and Schorfheide (2007); Fernández-Villaverde (2010); Del Negro and Schorfheide (2011); Herbst and Schorfheide (2015) 主要介绍贝叶斯估计法在 DSGE 模型中的应用。
- Galí (2008); Woodford (2011) 主要关注 DSGE 模型的一种: 新凯恩斯主义模型。
- Tovar Mora (2009) 探讨 DSGE 模型在中央银行中的应用。

## 7.2 pros and cons

### 7.2.1 cons

目前学界基本达成共识，DSGE 模型可作为宏观经济学研究的标准框架之一。但仍有局限，试从以下几个角度做简要介绍，分别为前提假定，求解方法，解的复杂性，以及利用解做决策的过程。

**前提假定。**一系列假定条件，如完美市场范式、市场效率、个体如家庭的理性期望等，引发争议。尤其是，过分简化、因而脱离实际的前提设定使得 DSGE 模型未能有效预测 2008 年的金融危机，使其广受质疑 (Buiter, 2009)。作为回应，近年来的 DSGE 模型逐渐引入一些与现实更加贴近的假定，如个体行为的经验法则 (rule-of-thumb)，金融市场的摩擦，行为个体的异质性特征等。

**求解方法。**对本质上是高度非线性的经济系统而言，不恰当的近似线性化处理会导致损失许多重要信息，这些信息原本是宏观经济学研究的重要对象。Buiter (2009) 甚至称之为“对宏观经济学模型的阉割”。

**解的复杂性，以及向公众解释政策含义的难度。**线性近似的最优条件和约束条件构成复杂而冗长的方程系统，在这系统中，第一，前向期望项变得难于辨认和解读，更难于向公众说明这一方程系统背后的经济学含义。第二，难于识别外生冲击如财政、货币政策的决策，是通过怎样的传导机制进入经济系统，并产生何种影响。第三，大多数情况下，难以直接求得系统的解析解，就需要借助计算机来测算近似的数值解，而采用何种数值算法更适合这一系统，数值近似解的精确程度如何，就成了又一个难题。这些都决定了模型系统及系统的解难以为公众们所理解，也因此难以在不同政策制定者之间形成共识。

**门槛高。**综上，DSGE 模型的构建、求解、预测等一系列工作的展开，均要求研究者及政策制定者受过良好的宏观经济学训练，并有相当程度的建模能力、统计学知识和编程水平。

### 7.2.2 pros

尽管如此，采用 DSGE 模型作为宏观经济学研究分析框架的优势也很突出。

**微观基础。**传统宏观模型往往不对个体行为做深入设定。与之相比，DSGE 模型在微观形为基础方面做了深入探索：行为个体基于理性预期的假定采取最优行动，决定要素价格和资源配置，进而影响公共部门的目标和约束条件。这样所生成的一系列局部最优条件，如家庭的劳动力供应决策、消费决策，企业的劳动力需求决策、产品定价决策等，为政策制定者提供重要参考依据。“优化”行为也意味着，个人和企业基于他们对未来期可能出现状况的预测，展开当前期的行为；从这种意义上来说，这种“理性预期”的行为模式便不同于“经验法则”。

**稳健性。**基于微观基础的 DSGE 模型使得缩减形式 (reduced form) 的参数与更深层次的结构形式 (structural form) 的参数之间，形成更紧密关联，使得模型参数较少可能随着政策的变动而发生变化——从此意义上说，DSGE 模型有助于更稳健地回应卢卡斯批判 (Lucas, 1976)<sup>1</sup>。

**模型研究与政策制定的契合度。**通过近似分散化经济体 (decentralized economy) 中典型个体的效用函数，DSGE 模型产生与中心化经济 (centralized economy) 模型中的福利定律相一致的结果。从而，DSGE 模型可以提供一套连贯的用于政策讨论与分析的工具，评估不同决策的效果，从而选取更好的决策付诸实施。在这一过程中，政策分析的展开与模型的设定条件是紧密相关的。

**工具包。**DSGE 模型的吸引人之处不仅在于理论分析框架，更在于为宏观经济学的经验研究提供了一套可量化的政策分析和预测工具。伴随着理论建模的进展，也涌现出了很多新的经验工具和算法，致力于让模型分析的结果更贴近实际观察到的数据。随着二者相符程度的不断提升，DSGE 模型的可靠性越来越得到政策制定机构如中央银行的认可 (An and Schorfheide, 2007)，其预测效

<sup>1</sup>递归宏观经济学模型中的稳健控制论综述，可见 Hansen and Sargent (2004)。简约式和结构式计量经济学方法论的争论，可见 Jarrow and Protter (2004)；一个更全面的综述可见 Angrist and Pischke (2008)。

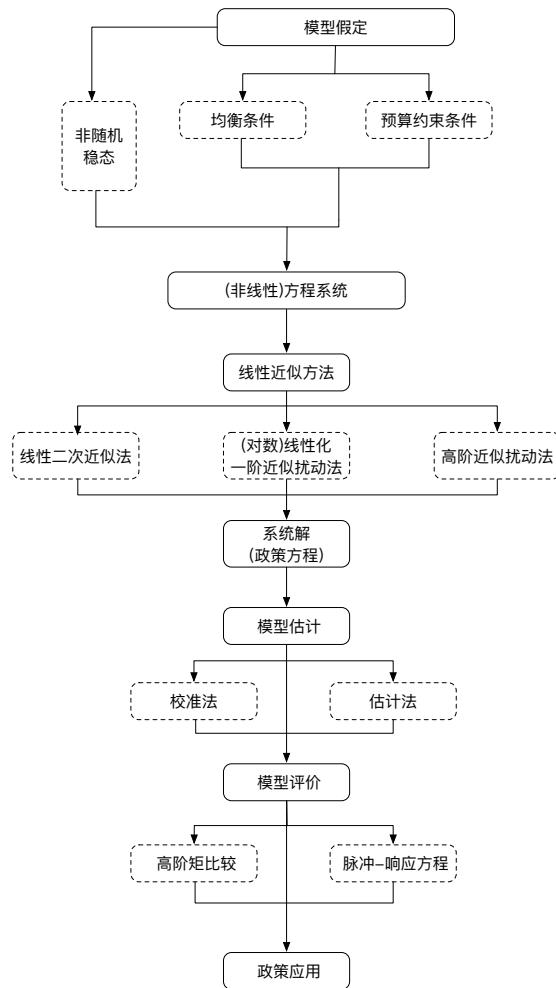
力不弱于 VAR 模型Edge et al. (2010)。此外，模型估计方面的研究也有快速进展 (Schorfheide, 2011; Fernández-Villaverde et al., 2016)。

总之，DSGE 模型提供了一整套集成的政策分析框架。但仍需要指出两点。一，简化。经济模型始终是对现实世界的描述，这种描述是简化的，并不追求事无巨细的完整重现。那么研究人员基于 DSGE 模型给出的政策性建议都必须立足于常识，同时牢记简化了的模型并不覆盖纷繁复杂现实世界的每一个细节。二，主观性。一方面由于 DSGE 模型的复杂性，另一方面中央银行建立新的或改进已有的 DSGE 模型，目的往往是为自己的货币政策的有效性背书。在此过程中，主观性总是不可避免的。那么需要强调的是，基于 DSGE 模型的经济研究“不能代替专家意见”，政策制定者还需要考虑“坊间消息和模型之外的信息”(Bernanke, 2007)。

### 7.3 工作流程

设定前提假定条件 → 求得系统解 → 将模型生成的方针数据与实际观测到的数据向比较 → 政策性建议，在分析无限时间周期的，由多个典型行为个体组成的 DSGE 模型时，常见的工作流程见下文，或如图7.1所示。

图 7.1: DSGE 模型的工作流程



1. 构建经济模型 (假设条件)。

2. 求解一阶均衡条件。一阶均衡条件与预算约束条件等一道构成非线性随机差分方程系统。
3. 由于该系统常常难以求得解析解，需要围绕这一个给定的点做数值近似。这一给定点，常常设定为模型的非随机稳定状态。
4. 非随机稳定状态可由如下方法之一求得：
  - (a) (对数) 线性化近似，围绕稳态，得到一个状态——空间形式表述的线性差分系统，进而借助一些常见算法求得系统的近似数值解；
  - (b) 围绕非随机稳定状态，做二阶或更高阶的近似。
5. 参数校准和/或估计。
6. 针对外生随机冲击，计算方差、做方差分解，求得有关变量的脉冲——响应方程。
7. 比较模型生成数据和史记观测到的数据以评价模型的解释力。
8. 政策参考。

下一节首先讨论模型设定。

## 7.4 模型设定

模型的前提假定应设定为与所研究的问题密切相关，并且必须指明模型中所讨论的问题是从定性还是定量的角度展开的。围绕所研究的问题，模型应该提供有所助益的回答，如：哪些冲击对经济体产生何种影响，它们的传导机制是怎样的，如何制定相关政策以应对这些冲击？

在一般均衡模型中，经济体中的每一个部门都是单独设定的，典型的部门如私人部门和公共部门：

1. 私人部门包括如消费（家庭）部门和企业部门等。通常来说，消费者被简化为一个典型家庭，基于自身偏好，做关于消费和休闲的一生最优决策 (Christiano et al., 2010)。企业在一定的生产条件限制下追求利润最大化。在为私人部门建模时，必须假定预期的形式，如理性预期或者经验法则等。
2. 公共部门包括如财政政策和货币政策等。通常来说，政府是财政政策的制定者，大多数情况下财政政策的效果受到李嘉图等价 (Ricardian equivalence) 的影响。中央银行设定货币政策。在为公共部门建模时，必须假定政策的执行方式，比如公共部门的政策制定是否依据某准则，是否受某种预算约束条件的限制，行为是否遵循某种利润最大化原则等。
3. 在经济模型中，个体行动也要受到市场机制和制度性要素的考量，因此也须作出设定，比如哪些市场是充分/非充分竞争的，决策者的政策制定是否需要遵循一些经验法则等。
4. 外生随机冲击的识别和建模。

上述模型设定都必须转换成适当的数学表达式，一方面使其能被纳入到模型分析框架中去，另一方面更重要的是，数学表达式的设定需要使模型有解。如，关于偏好和冲击的凹或凸的方程式，非庞兹骗局的约束条件等。

在模型设定后，下一步工作是要针对经济体中的家庭、企业、公共部门等，分别求得一阶最优条件和预算约束条件方程。这些方程常常包括当期和/或前向（后向）元素，又称非线性理性期望系统。通常来说，我们需要借助一定的数值近似方法，将非线性系统做近似线性化，进而求解，这一求解方法的讨论见下节。

## 7.5 求解方法

对于动态模型如新古典主义经济增长模型，最常见的解法是动态规划 (Dynamic Programming)，又称价值方程迭代，简要介绍见第九章。这种方法的优势在于算法可靠度高，解的收敛特性好。但不足也较明显，如计算速度慢，存在维数灾难 (curse of dimensionality, Bellman (1957)) 等。

动态规划法的不足导致新算法的出现，典型如扰动法 (perturbation methods)、投影法 (projection methods) 等。这些新方法在保持较好收敛性的基础上，提供了更快的求解速度。扰动法在宏观经济学研究中的广泛应用，综述可见如 Stokey et al. (1989); Ljungqvist and Sargent (2004)。投影法的简要介绍见 Aruoba et al. (2006)。

本章内容重在对求解方法做简要介绍。大致说来，求解方法分为两步。第一步，近似线性化，将由均衡条件和约束条件构成的非线性方程系统转化为线性方程系统。不同方法背后核心思路基本一致，即将变量围绕其非随机稳定状态附近做线性近似。第二步常常被视为求解 DSGE 模型的核心，即根据近似线性系统的实际特征，采用相应的数值算法求解，作为原非线性系统的近似解<sup>2</sup>。

### 7.5.1 近似线性展开

在 King et al. (1988a,b, 2002) 的开创性工作之后，(对数) 线性近似法逐渐成为主要标准。以将变量沿着其非随机稳态做一阶泰勒级数展开为例，假定  $y = f(x)$ ，其中  $x, y$  分别是动态模型中的前定和非前定变量 (predetermined and non-predetermined variables)， $f$  是个平滑的非线性函数<sup>3</sup>。

定义  $x_t$  的确定性 (非随机) 稳定状态为  $\bar{x}$ ，我们有  $\bar{y} = \bar{x}$ 。

对  $f$  围绕  $\bar{x}$  做泰勒级数展开

$$y \approx \sum_{n=0}^{\infty} \frac{f^{(n)}(\bar{x})}{n!} (x - \bar{x})^n, \quad (7.1)$$

其中  $f^{(n)}$  表示  $f$  对  $x$  的第  $n$  次求导。

#### (1 阶) 线性化近似

根据(7.1)，1 阶泰勒级数展开

$$\begin{aligned} y &\approx f(\bar{x}) + f_x(\bar{x})(x - \bar{x}), \\ \Rightarrow (y - \bar{y}) &\approx f_x(\bar{x})(x - \bar{x}), \\ \Rightarrow \frac{y - \bar{y}}{\bar{y}} &\approx \underbrace{\frac{f_x(\bar{x})\bar{x}}{\bar{y}}}_{\text{常系数}} \frac{x - \bar{x}}{\bar{x}}, \\ \Rightarrow \tilde{y} &\approx \underbrace{\frac{f_x(\bar{x})\bar{x}}{\bar{y}}}_{\text{常系数}} \tilde{x}, \end{aligned} \quad (7.2)$$

根据定义， $\tilde{y}_t$  和  $\tilde{x}_t$  分别表示  $y_t$  和  $x_t$  相对于其确定性稳态  $\bar{y}$  和  $\bar{x}$  的偏离百分比。

<sup>2</sup> 扰动法和投影法，从本质上来说都是局域算法。全局算法如车比雪夫多项式 (Chebyshev-polynomial)、有限元 (finite-elements) 等的综述可见 Judd (1998)。

<sup>3</sup> 这里我们采用 Klein (2000) 的定义： $x_t$  是前定变量，当且仅当它有外生的 1 期前向预测偏误，即  $x_{t+1} - E_t X_{t+1} = C\varepsilon_{t+1}$  是外生的。

根据这一定义，一个前定变量由它的滞后项和当期外生冲击所共同决定，如外生冲击变量。另一方面，不属于此情况的变量成为非前定变量，常常是前向变量。

### (1 阶) 对数线性化近似

对模型做变型,  $y = \exp(\ln y) = f(\exp(\ln x))$ 。

两侧取对数,  $\ln y = \ln f(\exp(\ln x))$ 。

围绕  $\ln \bar{x}$  作 1 阶泰勒级数展开

$$\begin{aligned} \ln y &\approx \ln f(\exp(\ln \bar{x})) + f_x(\bar{x}) \frac{d \exp(\ln x)}{d \bar{x}} (\ln x - \ln \bar{x}), \\ &\Rightarrow \ln y \approx \ln \bar{y} + f(\exp(\ln \bar{x})) (\ln x - \ln \bar{x}), \\ &\Rightarrow \ln \left( \frac{y}{\bar{y}} \right) \approx \underbrace{\frac{f_x(\bar{x}) \bar{x}}{\bar{y}}}_{\text{常系数}} \ln \left( \frac{x}{\bar{x}} \right), \end{aligned} \quad (7.3)$$

其中  $\ln \left( \frac{y}{\bar{y}} \right)$  和  $\ln \left( \frac{x}{\bar{x}} \right)$  分别表示对数形式的偏离百分比。

一阶泰勒级数展开后的线性化近似和对数线性化近似, 本质上相似。为了说明这一点, 见下式

$$\tilde{x}_t = \frac{x_t - \bar{x}}{\bar{x}} = \frac{x_t}{\bar{x}} - 1 \approx \ln \left( \frac{x_t}{\bar{x}} \right) = \ln x_t - \ln \bar{x}.$$

此外 Uhlig (1999) 提出另一种类似对数线性化的方法, 从而在特定情况下无需求导: 对于常数  $a$  和接近 0 的  $\tilde{x}_t$  和  $\tilde{y}_t$ ,

$$\begin{aligned} \exp(x + ay) &\approx 1 + x + ay, \\ \Rightarrow xy &\approx 0, \\ \Rightarrow E_t(a \exp(x_{t+1})) &\approx E_t(ax_{t+1}). \end{aligned}$$

对数线性化方法的本质在于将变量  $x_t$  用  $x \exp(\tilde{x}_t)$  来代替, 此时  $\tilde{x}_t = \ln(x_t/\bar{x})$  表示它相对于稳态的对数偏差。利用此方法, 可以将非线性方程近似为围绕稳态的线性方程。

### 2 阶近似

根据(7.1), 以 2 阶展开为例

$$y \approx \underbrace{f(\bar{x})}_{0 \text{ 阶展开}} + \underbrace{f_x(\bar{x})(x - \bar{x})}_{1 \text{ 阶展开}} + \underbrace{\frac{1}{2} f_{xx}(\bar{x})(x - \bar{x})^2}_{2 \text{ 阶展开}} + \underbrace{O^3}_{\text{误差项}}, \quad (7.4)$$

其中线性近似的误差项  $O^3$  是由忽略了 3 阶及以上阶数近似所造成的。

(7.4) 为代表的线性近似法具有如下特征:

1. 在一阶泰勒级数展开中,  $y$  的条件期望值等于稳态下的  $\bar{y}$ , 这意味着不确定性 (以方差的二阶或更高阶矩等形式表现) 在线性近似过程中不起作用。
2. 在高阶泰勒级数展开中,  $y$  的条件期望值与方程  $f$  的曲率和变量  $x$  的方差有关。
3. 如果研究的目的是求得冲击响应方程以及变量的二阶距, 那么我们只分析模型的一阶属性即可, 只需要对非线性系统做一阶泰勒级数展开。如,  $\text{var}(y)$  可以通过(7.4)算出:  $\text{var}(y) = E(y - Ey)^2 = f_x^2 \text{var}(x)$ 。

在什么情况下做一阶展开, 什么情况下需要做更高阶展开?

1. 在很多情况下, 对非线性系统做一阶线性近似即可。尽管有研究发现更高阶的近似会提供更精确的近似解, 但精度提升的幅度有限。

2. 然而的确存在一些情况，仅仅用一阶近似处理 DSGE 模型是不够的，尤其是当研究目标涉及到分析一些政策的福利效果时——这些福利政策往往不会对模型产生一阶影响。

入一个 refer-  
ce。

我们将在随后章节中进一步展开探讨这个问题。

对非线性方程系统做线性近似，生成以状态——空间形式表现的变型系统。这样，一方面可以更方便的展开后续经验研究，如模型估计，卡曼滤波，经济系统预测等。更为重要的是另一方面，搭配一些数值求解方法，可以求得系统的近似解。随后我们讨论如何在理性预期的情境下，运用数值算法求解变型系统。

## 7.6 求解线性随机差分方程系统

如前文所述，由各部门均衡条件和预算约束条件所组成的非线性系统，经由一定的（对数）线性化处理后变型为新的系统，又称线性随机方程差分系统，

新系统有一组含有理性期望条件的线性随机差分方程构成，可表述为如下状态——空间形式，

$$A_0 E_t Y_{t+1} = A_1 Y_t + B_0 \varepsilon_t, \quad (7.5)$$

其中  $A_0, A_1, B_0$  是线性化系统的系数矩阵， $Y_t, \varepsilon_t$  分别是内生变量和外生变量构成的向量。

在构建 DSGE 模型时，外生变量  $\varepsilon_t$  常取如下两个假定之一。假定  $\varepsilon_t$  是 iid 冲击向量， $E(\varepsilon_t) = 0, E(\varepsilon_t \varepsilon_t^\top) = \Sigma, E(\varepsilon_t \varepsilon_s^\top) = \Sigma \forall s \neq t$ ，或者假定  $\varepsilon_t$  是与 iid 外生冲击有关的 AR(1) 过程。

这样一个线性化方程构成的动态系统，描述了模型中变量的运动路径。通常来说系统中含有前向及后向要素。随着模型参数设定的不同，系统解存在三种情况：

- 无解。系统没有稳定的理性预期解。
- 不确定 (indeterminacy)。系统存在不止一个解的多重均衡。
- 确定的。有且只有一个解，又称稳定解。

如果前两种情况出现，意味着模型参数设定不当，需要重新调整，以使得系统存在稳定解。

已经有大量文献探讨理性期望系统(7.5)的求解方法。系统的解构成一个回应机制 (feedback rule)，将当期内生变量与模型内的状态变量联系起来。这方面的重要工作框架由 Blanchard and Kahn (1980) 所奠定，随后经由许多人的努力，得到进一步深化和扩展，见第十一章。

如果 DSGE 模型致力于考察政策的福利效果，如期望效用等，则仅仅用一阶近似线性转换可能不够精确。此时需要对效用函数等其他函数做二阶近似，以确保二次项所含有的重要经济学信息不被忽略掉 (Kim and Kim, 2003; Schmitt-Grohé and Uribe, 2004)，见第十一章。

Carl Whelan 对  
forward-looking  
backward-  
looking 的描述

# 第八章 动态规划

简要介绍动态规划 (dynamic programming)。更详细的介绍，见Adda and Cooper (2003)。Python 中的程序实现见Stachurski (2008)。

## 8.1 包络定理

包络定理 (Envelope Theorem) 研究当模型中的参数发生变化时，模型中某一变量的最大值 (或最小值) 如何随着发生变化。

**Theorem 8.1** (包络定理). 假定我们有

$$v(a) = \max_{\{x\}} f(x, a), \quad (8.1)$$

那么下式成立

$$\frac{dv(a)}{da} = \frac{\partial f(x, a)}{\partial a} \Big|_{x=x^*(a)}, \quad (8.2)$$

其中

$$x^*(a) = \arg \max_{\{x\}} f(x, a). \quad (8.3)$$

证明. 设

$$v(a) \equiv f[x^*(a), a],$$

两侧同时对  $a$  求 (偏) 导

$$\frac{dv(a)}{da} = \frac{\partial f[x^*(a), a]}{\partial x} \frac{\partial x^*(a)}{\partial a} + \frac{\partial f[x^*(a), a]}{\partial a}. \quad (8.4)$$

上式中  $x^a$  是  $f(x, a)$  取极大值时的系数，满足

$$\frac{\partial f[x^a, a]}{\partial x} = 0,$$

代回(8.4)有

$$\frac{dv(a)}{da} = \frac{\partial f[x^*(a), a]}{\partial a}, \quad (8.5)$$

等价于(8.2).

□

或者也可以从有约束条件的优化问题中推得包络定理

**Theorem 8.2** (包络定理 (有约束条件的优化问题)). 设

$$\begin{aligned} m(a) &= \max_{\{x\}} f(x, a), \\ s.t. &\begin{cases} g(x, a) = 0, \\ x \geq 0. \end{cases} \end{aligned}$$

使  $\mathcal{L}(x, a, \lambda)$  为相应的拉格朗日方程, 其中  $x^*(a), \lambda^*(a)$  为库恩塔克条件 (*Kuhn-Tucker condition*) 的最优解。则我们有

$$\frac{dm(a)}{da} = \frac{\partial \mathcal{L}(a)}{\partial a} \Big|_{x^*(a), \lambda^*(a)}. \quad (8.6)$$

## 8.2 例: 吃蛋糕 (直接求解法)

小明有一块蛋糕, 大小是  $W_1$ , 可以在  $t = 1, 2, \dots, T$  期内吃完。问小明怎么吃可以实现效用最大化?

小明的最优决策可以表现为贴现效用的求和加总

$$\sum_{t=1}^T \beta^{t-1} u(c_t),$$

其中  $c_t$  表示  $t$  期消费即吃掉蛋糕的大小。 $u(c_t)$  表示  $t$  期效用, 设方程满足可导、严格单调、严格凹 (concave), 即稻田条件 (Inada condition) (Inada, 1963)

$$\lim_{c \rightarrow 0} u'(c) \rightarrow \infty, \quad \lim_{c \rightarrow \infty} u'(c) \rightarrow 0.$$

引入稻田条件的作用是确保小明每个时段  $t$  都至少会吃一小口蛋糕, 即从数学角度上来讲, 不存在角点解 (corner solution)。蛋糕的大小随时间发生变化, 满足

$$W_{t+1} = W_t - c_t, \quad t = 1, 2, \dots, T. \quad (8.7)$$

若是按照直接求解法的思路, 小明吃蛋糕问题可以表示为

$$\max_{\{c_t\}_{t=1}^T, \{W_t\}_{t=1}^{T+1}} \sum_{t=1}^T \beta^{t-1} u(c_t),$$

给定约束条件

$$\begin{aligned} \sum_{t=1}^T c_t + \sum_{t=1}^{T+1} W_t &= \sum_{t=1}^T W_t, \\ \Rightarrow \sum_{t=1}^T c_t + W_{T+1} &= W_t, \end{aligned}$$

即小明吃掉的所有蛋糕, 和最终剩余的蛋糕数量  $W_{T+1}$  之和, 等于蛋糕的初始大小  $W_1$ 。

鉴于此, 可以将问题改写为

$$\max_{\{c_t\}_{t=1}^T, \{W_t\}_{t=2}^{T+1}} \sum_{t=1}^T \beta^{t-1} u(c_t), \quad (8.8)$$

约束条件为

$$\begin{cases} \sum_{t=1}^T c_t + W_{T+1} = W_1, \\ c_t \geq 0 \quad \forall t = 1, \dots, T, \\ W_{T+1} \geq 0. \end{cases} \quad (8.9)$$

在随后的分析中我们不再讨论  $c_t \geq 0$  的约束条件, 因为假定稻田条件  $\lim_{c \rightarrow 0} u'(c) \rightarrow \infty$  成立, 就已经满足该项。

构建拉格朗日方程方程

$$\mathcal{L} = \sum_{t=1}^T \beta^{t-1} u(c_t) + \lambda_t \left[ W_1 - \sum_{t=1}^T c_t - W_{T+1} \right] + \phi_t W_{T+1}.$$

两个一阶条件 FOCs 给出最大化问题的解

$$\frac{\partial \mathcal{L}}{\partial c_t} = 0 \Rightarrow \beta^{t-1} u'(c_t) = \lambda_t, \quad t = 1, \dots, T, \quad (8.10)$$

其中拉格朗日乘子  $\lambda_t$  对应预算约束条件, (8.9)第一行, 即

$$\lambda_t = \beta^t u'(c_{t+1}).$$

将两期最优消费决策联系起来, 可得

$$u'(c_t) = \beta u'(c_{t+1}), \quad (8.11)$$

这称为跨 (1) 期消费的欧拉等式 (Euler equation)。

$$\frac{\partial \mathcal{L}}{\partial W_{T+1}} = 0 \Rightarrow \lambda_t = \phi_t, \quad (8.12)$$

其中拉格朗日乘子  $\phi_t$  对应横截条件, (8.9)第三行。已知根据稻田条件的,  $\lambda_t > 0$  严格成立, 则  $\phi_t > 0 \forall t$ 。这意味着若要满足库恩-塔克 (Kuhn-Tucker condition)

$$\phi_t W_{T+1} = 0, \quad (8.13)$$

只能使得  $W_{T+1} \equiv 0$ 。

进一步解读欧拉等式(8.11)。在均衡状态下, 当期  $t$  减少 1 单位消费的边际效用下降, 应当等于  $t$  期额外 1 单位储蓄的增加, 在  $t+1$  期带来边际效用提升的贴现值。将跨 1 期扩展到跨多期的情况, 任意两期  $t, t'$  之间  $t, t' = 1, \dots, T, t' > t$  的跨期消费转移都不会导致总福利水平的变化。这样一来, (已经处于最优行为决策下的) 小明, 全部福利水平最大便只取决于初始蛋糕的大小  $W_1$ , 对应福利  $V_T(W_1)$ , 又称价值方程 (value function):

$$V_T(W_1) = \max_{\{c_t\}_{t=1}^T} \sum_{t=1}^T \beta^{t-1} u(c_t) + \lambda_t \left[ W_1 - \sum_{t=1}^T c_t - W_{T+1} \right] + \phi_t W_{T+1} \quad (8.14)$$

那么为了回答初始蛋糕大小  $W_1$  对小明总福利水平  $V_T(W_1)$  的影响, 可以求导数

$$\frac{dV_T(W_1)}{dW_1} = V'_T(W_1) = \lambda_t.$$

## 8.3 吃蛋糕 (动态规划)

### 8.3.1 有限时间段的动态规划问题

我们先从容易理解的  $T < \infty$  入手, 熟悉动态规划问题的基本原理。然后将问题扩展到  $T \rightarrow \infty$  的情况, 见第8.3.1节。

从  $t = 1$  期来看, 给定  $W_1$ , 吃蛋糕问题可以改写为如下形式

$$V_T(W_1) = \max_{\{c_t\}_{t=1}^T} u(c_1) + \beta V_{T-1}(W_2), \quad W_2 = W_1 - C_1, \quad (8.15)$$

$t = 1$  期小明的决策是, 消费多少蛋糕  $c_1$ , 这会传导到  $t = 2$  期影响到  $W_2$ ; 以此类推。那么,  $c_t$  最优消费决策应当使  $t$  期效用的变化, 与  $t+1$  期效用变化的贴现相等, 这称为贝尔曼等式 (Bellman equation)。

另一方面, 对贝尔曼等式(8.15)求 FOC, 可得欧拉等式

$$u'(c_1) = \beta V'_{T-1}(W_2). \quad (8.16)$$

现在设效用函数的显形式为  $u(c) = \ln c$ , 那么通过逆推法, 可以进一步计算价值方程。

$T = 1$  时, 小明的最优决策是当期吃掉全部蛋糕,  $c_1 = W_1$ , 因此我们有

$$V_1(W_1) = \ln W_1. \quad (8.17)$$

$T = 2$  时, 小明的最优决策根据(8.15)表示为

$$V_2(W_1) = \max_{W_2} \underbrace{\ln(W_1 - W_2)}_{u(c_1)} + \beta \underbrace{\ln W_2}_{u(c_2)}, \quad (8.18)$$

约束条件

$$c_1 + W_2 = c_1 + c_2 = W_1, \quad (8.19)$$

欧拉等式

$$u'(c_1) = \beta V'_2(W_1),$$

FOC

$$\frac{1}{c_1} = \beta \frac{1}{c_2}, \quad (8.20)$$

联立(8.19)和(8.20)可得最优消费决策  $\{c_t\}_{t=1}^2$

$$c_1 = \frac{1}{1+\beta} W_1, \quad c_2 = \frac{\beta}{1+\beta} W_1, \quad (8.21)$$

价值方程因此为

$$V_2(W_1) = \ln \left( \frac{1}{1+\beta} W_1 \right) + \beta \ln \left( \frac{\beta}{1+\beta} W_1 \right), \quad (8.22)$$

有时我们将其进一步简化为

$$\begin{aligned} V_2(W_1) &= A_2 + B_2 \ln W_2, \\ A_2 &= \ln \left( \frac{1}{1+\beta} \right) + \beta \ln \left( \frac{\beta}{1+\beta} \right), \\ B_2 &= 1 + \beta. \end{aligned}$$

$T = 3$  时, 小明的最优决策分三部分, 依次为

1.

$$V_3(W_1) = \max_{W_2} \underbrace{\ln(W_1 - W_2)}_{u(c_1)} + \beta V_2(W_2), \quad (8.23)$$

FOC

$$V'_3(W_1) = 0 \Rightarrow \frac{1}{c_1} = \beta V'_2(W_2). \quad (8.24)$$

2.

$$V_2(W_2) = \max_{W_3} \underbrace{\ln(W_2 - W_3)}_{u(c_2)} + \beta V_1(W_3), \quad (8.25)$$

FOC

$$V'_2(W_2) = 0 \Rightarrow \frac{1}{c_2} = \beta V'_1(W_3). \quad (8.26)$$

包络条件

$$\frac{1}{c_2} = V'_2(W_2). \quad (8.27)$$

3.

$$V_1(W_3) = \ln W_3, \quad (8.28)$$

因为  $W_4 = 0$ 。

FOC

$$V'_1(W_3) = 0 \Rightarrow \frac{1}{c_3} = V'_1(W_3). \quad (8.29)$$

在此基础上我们有，我们可以用逆推的方法求得最优消费决策  $\{c_t\}_{t=1}^3$

1. 联立(8.26)和(8.29)可得

$$\frac{1}{c_2} = \beta \frac{1}{c_3},$$

2. 联立(8.24)和(8.27)可得

$$\frac{1}{c_1} = \beta \frac{1}{c_2},$$

3. 资源约束条件

$$c_1 + c_2 + c_3 = W_1.$$

因此我们有

$$\begin{aligned} c_1 &= \frac{1}{1 + \beta + \beta^2} W_1, \\ c_2 &= \frac{\beta}{1 + \beta + \beta^2} W_1, \\ c_3 &= \frac{\beta^2}{1 + \beta + \beta^2} W_1, \end{aligned}$$

对应的价值方程为

$$V_3(W_1) = \ln c_1 + \beta (\ln c_2 + \beta \ln c_3). \quad (8.30)$$

### 8.3.2 无限时间段的动态规划问题

随着  $T \rightarrow \infty$ ，无法再使用上一节介绍过的逆推法，情况变得较为复杂。无限时间段的吃蛋糕问题，小明追求效用最大化

$$\begin{aligned} &\max_{\{c_t\}_{t=1}^{\infty}, \{W_t\}_{t=2}^{\infty}} \sum_{t=1}^{\infty} \beta^{t-1} u(c_t), \\ &\text{s.t. } W_{t+1} = W_t - c_t, \quad t = 1, 2, \dots \end{aligned}$$

我们将这个问题改写为动态规划的形式

$$V(W) = \max_{c \in [0, W]} u(c) + \beta V(W - c),$$

其中  $V(W)$  为当期小明采取最优行为决策所可能得到的最大效用； $V(W - c)$  为下期小明采取最优行为决策所可能得到的最大效用：显然， $c$  是控制变量， $W$  是状态变量。设  $W' := W - c$ ，那么动态规划问题改写为如下贝尔曼等式形式的泛函方程

$$V(W) = \max_{W' \in [0, W]} u(W - W') + \beta V(W'). \quad (8.31)$$

求解无限时段的动态规划问题(8.31)可分三部，核心思路是找到一个  $V(W)$  使得对于所有  $W$  的值，泛函式均成立。出于简化模型的考虑，我们取消时间下角标——因为在无限时段下，这成为一个静态(stationary)或者无关时间(time-invariant)的问题。三部如下

## 1. (8.31)的 FOC

$$V'(W) = 0 \Rightarrow u'(c) = \beta V'(W'). \quad (8.32)$$

难点：在不清楚原方程形式  $V(W)$  的情况下，如何计算  $V'(W')$ 。

## 2. 包络条件

$$V'(W) = u'(c) \Leftrightarrow V'(W') = u'(c'). \quad (8.33)$$

包络条件的前提是假定行为人处于最优决策条件下。

## 3. 联立 FOC(8.32)和包络条件(8.33)

$$u'(c) = \beta u'(c'), \quad (8.34)$$

即欧拉等式。由此我们可以定义一组策略方程

$$\begin{aligned} c &= \phi(W), \\ W' &= e(W) = W - \phi(W), \end{aligned} \quad (8.35)$$

使满足欧拉等式

$$u'(\phi(W)) = \beta u'(W - \phi(W)). \quad (8.36)$$

现在以  $u(c) = \ln c$  为例做进一步求解<sup>1</sup>。首先，我们猜测  $V(W)$  可能是以下线性形式

$$V(W) = A + B \ln W, \quad (8.37)$$

线性形的猜测是为了让计算过程简化，更多非线性形的讨论可见 Hansen and Sargent (2004)。代回(8.31)可得

$$A + B \ln W = \max_{W' \in [0, W]} u(W - W') + \beta V(A + B \ln W'). \quad (8.38)$$

类似地，取 FOC 作为最优条件

$$\begin{aligned} \frac{1}{W - W'} &= \beta B \frac{1}{W'}, \\ \Rightarrow W' &= \frac{\beta B}{1 + \beta B} W, \end{aligned}$$

代回(8.38)可得

$$\begin{aligned} A + B \ln W &= \ln\left(\frac{W}{1 + \beta B}\right) + \beta \left[ A + B \ln\left(\frac{\beta B}{1 + \beta B} W\right) \right] \\ &= \ln W - \ln(1 + \beta B) + \beta A + \beta B \ln\left(\frac{\beta B}{1 + \beta B}\right) + \beta B \ln W \\ &= \underbrace{-[\ln(\beta B) + 1] \ln(1 + \beta B)}_A + \underbrace{\beta [A + B \ln(\beta B)]}_B + (1 + \beta B) \ln W, \end{aligned} \quad (8.39)$$

LHS 和 RHS 相对应，可得系数 A B 的值

$$\begin{aligned} A &= \frac{1}{(1 - \beta)^2} \left\{ \beta \ln\left(\frac{\beta B}{1 - \beta}\right) - (1 - \beta) \ln\left(\frac{1}{1 - \beta}\right) \left[ 1 + \ln\left(\frac{\beta}{1 - \beta}\right) \right] \right\}, \\ B &= \frac{1}{1 - \beta}, \end{aligned}$$

这样我们最终得到了策略方程(8.35)的计算式

$$\begin{aligned} c &= \phi(W) = (1 - \beta) W, \\ W' &= e(W) = \beta W, \end{aligned} \quad (8.40)$$

<sup>1</sup>通常来说，我们无法求得价值方程的显性形式，但  $u(c) = \ln c$  可能是少数的例外之一。

## 8.4 随机问题

现在将不确定性引入动态规划问题的分析中去，如外生偏好冲击，效用函数变为  $\varepsilon_t u(c_t)$ ，其中  $\varepsilon$  是一个随机变量。在  $t$  期做最优消费决策的行为主体知道当期的  $\varepsilon_t$ ，但不掌握对未来期  $\varepsilon$  的全部信息。一个最简单的设定是  $\varepsilon \in \{\varepsilon_l, \varepsilon_h\}$ ，两种冲击状态满足  $\varepsilon_h > \varepsilon_l > 0$ 。

偏好冲击表现为一阶马尔科夫过程，即当期冲击状态遇上一期的冲击水平有关。我们常用 Tauchen (1986) 等方法将 AR(1) 过程做离散化处理，见第 12.D 节。用  $p_{ij}$  表示当期冲击状态  $\varepsilon_i$  在下一期变为  $\varepsilon_j$  的概率，据此可以构建完整的转移矩阵  $P$ ，其中的元素例如  $p_{lh} \equiv \Pr(\varepsilon' = \varepsilon_h | \varepsilon = \varepsilon_l)$ ：

表 8.1: 偏好冲击的转移矩阵

		$t+1$	
		$\varepsilon_h$	$\varepsilon_l$
$t$	$\varepsilon_h$	$p_{hh}$	$p_{hl}$
	$\varepsilon_l$	$p_{lh}$	$p_{ll}$

这样，随机吃蛋糕问题就变为

$$V(W, \varepsilon) = \max_{W'} \varepsilon u(W - W') + \beta E_{\varepsilon'|\varepsilon} V(W', \varepsilon'), \quad \varepsilon \in \{\varepsilon_l, \varepsilon_h\},$$

对应转移矩阵  $P$ 。

FOC 描述最优状态

$$\varepsilon u'(W - W') = \beta E_{\varepsilon'|\varepsilon} V(W', \varepsilon') \quad (8.41)$$

由包络定理可得

$$V_1(W, \varepsilon) = \varepsilon u'(W - W'),$$

进而我们有

$$V_1(W', \varepsilon') = E_{\varepsilon'|\varepsilon} \varepsilon' u'(W' - W''). \quad (8.42)$$

联立包络定理(8.42)和 FOC 条件(8.41)可得跨(多)期欧拉等式

$$\varepsilon u'(W - W') = \beta E_{\varepsilon'|\varepsilon} \varepsilon' u'(W' - W''). \quad (8.43)$$

如果我们设政策方程满足一定形式(常常设为线性形)

$$W' = \varphi(W, \varepsilon),$$

代回(8.43)，欧拉等式变为

$$\varepsilon u'(W - \varphi(W, \varepsilon)) = \beta E_{\varepsilon'|\varepsilon} \varepsilon' u' \{ \varphi(W, \varepsilon) - \varphi(\varphi(W, \varepsilon), \varepsilon') \}. \quad (8.44)$$

## 8.5 包络定理和欧拉等式的关系

上文以小明吃蛋糕问题为例，对动态规划做了一般性介绍。下面以常见的 NCGT(Neo-Classical Growth Model，如 Ramsey Model) 为例，介绍家庭部门的多期优化问题，它是动态经济学模型的基础组成部分之一。

$t$  期行为人的目标，追求  $t, t+1, \dots, T$  的效用最大化

$$\max \sum_{n=0}^{T-t} \beta^n u(c_{t+n}), \quad (8.45)$$

$$m_{t+1} = (m_t - c_t) R + y_{t+1}.$$

根据动态规划的思路，将问题改写为贝尔曼等式形式

$$V_t(m_t) = \max_{\{c_t\}} u(c_t) + \beta V_{t+1}[(m_t - c_t) R + y_{t+1}]. \quad (8.46)$$

FOC 得到最优条件

$$0 = u'(c_t) + \left( \frac{dm_{t+1}}{dc_t} \right) \beta V'_{t+1}(m_{t+1}),$$

其中  $\left( \frac{dm_{t+1}}{dc_t} \right) = -R$ ,  $R$  反映在(8.45)中。因此上式变为

$$u'(c_t) = \beta R V'_{t+1}(m_{t+1}). \quad (8.47)$$

根据(8.45)，行为人基于已有的信息  $m_t$  做出当期消费决策  $c_t$ ，因此可以定义一个方程

$$c_t = c_t(m_t), \quad (8.48)$$

用于描述在给定  $m_t$  时，行为人可以达到效用最大化的最有消费决策。(8.48)代回(8.47)，写为

$$u'(c_t(m_t)) - \beta R V'_{t+1}([m_t - c_t(m_t)] R + y_{t+1}) = 0. \quad (8.49)$$

上式中的两个方程都是一阶形式；在原方程形式未知的情况下很难直接求解。我们因此采取近似算法。

定义一个新的方程

$$\underline{V}_t(m_t, c_t) = u(c_t) + \beta \underline{V}_{t+1}((m_t - c_t) R + y_{t+1}), \quad (8.50)$$

同样地，对上式求偏导我们有

$$\underline{V}_t^c \equiv \frac{\partial \underline{V}_t}{\partial c_t} = u'(c_t) - \beta R \underline{V}_{t+1}^c((m_t - c_t) R + y_{t+1}), \quad (8.51)$$

$$\underline{V}_t^m \equiv \frac{\partial \underline{V}_t}{\partial m_t} = \beta R \underline{V}_{t+1}^m(m_{t+1}). \quad (8.52)$$

根据定义式(8.50)，结合(8.46)可见

$$V_t(m_t) = \underline{V}_t(m_t, c_t), \quad (8.53)$$

引入链式法则 (Chain rule)(14.60)，上式变为

$$\begin{aligned} V'_t(m_t) &\equiv V_t^m(m_t) \equiv \frac{dV_t(m_t)}{dm_t} \\ &= \underline{V}_t^m(m_t, c_t(m_t)) + \frac{\partial c_t(m_t)}{\partial m_t} \underline{V}_t^c(m_t, c_t(m_t)). \end{aligned} \quad (8.54)$$

理解(8.54)的一个关键是：行为人是处于最优决策状态的，这意味着他总是在衡量价值方程，以及价值方程相对于消费的偏导，使其满足最优条件(8.49)<sup>2</sup>，这使得(8.51)中的  $\underline{V}_t^c = 0$ ，即

$$u'(c_t) = \beta R \underline{V}_{t+1}^c((m_t - c_t) R + y_{t+1}), \quad (8.55)$$

<sup>2</sup>有时这种最优决策也会受到资源约束条件的限制，如流动性约束等，我们在这里暂不讨论这种情况。

代回(8.54)变为

$$V_t'(m_t) = V_t^m(m_t, c_t(m_t)) = \beta R V_{t+1}^m(m_{t+1}). \quad (8.56)$$

此外(8.56)的 RHS 和(8.47)的 RHS 相等，因此有

$$V_t'(m_t) = u'(c_t). \quad (8.57)$$

结合(8.56)和(8.57)，又有

$$u'(c_t) = \beta R u'(c_{t+1}). \quad (8.58)$$

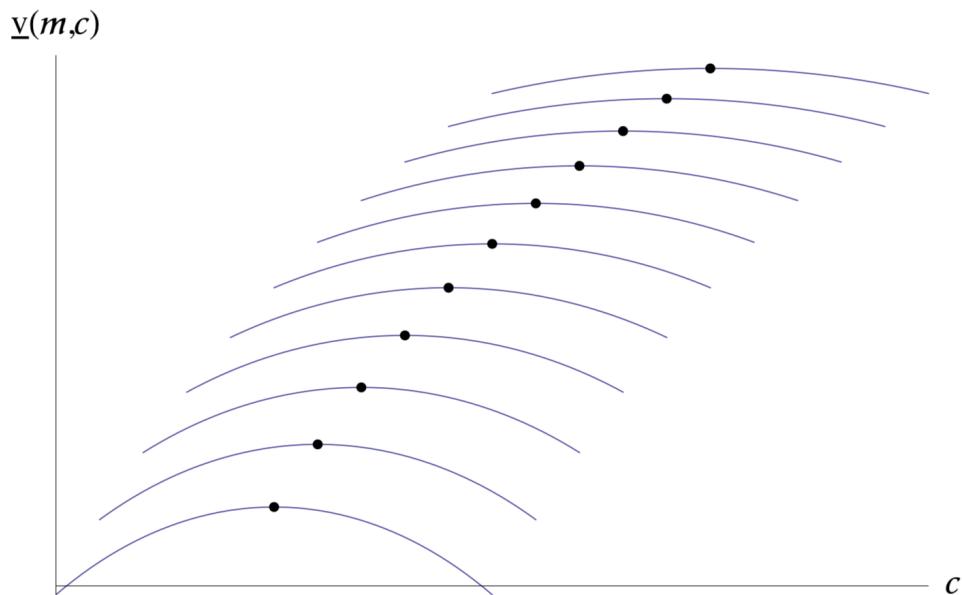
至此可见，包络定理指的是，价值方程对任意变量的全求导，等于最优消费决策下，行为人消费该变量的边际效用为 0 的点。因此，可将  $c_t(m_t)$  的最优决策看成是一个常数，对应  $c_t'(m_t) = 0$ ，并直接将贝尔曼等式对  $m_t$  求导：

$$V_t(m_t) = u(c_t(m_t)) + \beta V_{t+1}((m_t - c_t(m_t)))R + y_{t+1}, \quad (8.59)$$

$$V_t'(m_t) = \beta R V_{t+1}'(m_{t+1}). \quad (8.60)$$

根据包络定理，不同的  $m_t$  值成为影响行为人效用最大化的关键，如图8.1所示。

图 8.1: 不同  $m$  取值下的包络定理





# 第九章 线性二次最优控制

多数情况下，均衡条件较为复杂，无法直接得出解析解。我们需要引入线性二次最优控制理论 (optimal linear-quadratic control, LQ)，将均衡条件用状态-空间形式表现出来 (state-space representation)。这里对线性二次最优控制论问题做简要介绍。

假定一个经济系统，由一系列状态变量的向量  $x_t$  和一系列控制变量的向量  $u_t$  构成； $x_t$  受到外生随机冲击  $\varepsilon_t$  的干扰，外生冲击的方差-协方差矩阵为  $\Sigma$ 。这样一个线性二次最优控制问题可以表述如下：经济系统的决策者通过选择  $u_t$  来影响  $x_t$  的走向，由此我们来构建目标方程和状态转移方程。

## 9.1 目标方程和状态转移方程

目标方程 (objective function)：

$$\max_{\{u_t\}} E \sum_{t=0}^{\infty} \beta^t [x_t^T R x_t + u_t^T Q u_t], \quad (9.1)$$

其中  $\beta$  是时间贴现系数，对称矩阵  $R$  和  $Q$  分别对应目标方程中状态变量和控制变量的权重。

状态转移方程 (transition function)：

$$x_{t+1} = Ax_t + Bu_t + \varepsilon_{t+1}, \quad (9.2)$$

当期状态由上期状态变量、上期控制变量所共同决定，二者的权重分别由矩阵  $A$  和  $B$  所反映。此外，当期状态还受到当期随机冲击的影响，假定是一个 Gaussian 过程  $\{\varepsilon_t\}_{t=0}^{\infty} \sim \mathcal{N}(0, \Sigma)$ 。

## 9.2 价值方程

在给定时间贴现系数  $\beta$  的情况下，如果已知系数矩阵  $A, B, R, Q$ ，我们可以求得系统的解。而对于系数矩阵未知的情况，问题则较为复杂。我们引入动态规划 (dynamic programming, DP) 的思路，定义价值方程  $V(x_t)$  表示当期状态变量的价值。结合状态转移方程(9.2)和目标方程(9.1)，迭代形式表现的价值方程 (value function) 优化问题如下

$$\begin{aligned} V(x_t) &= \max_{\{u_t\}} \{x_t^T R x_t + u_t^T Q u_t + \beta E[V(x_{t+1})]\} \\ &= \max_{\{u_t\}} \{x_t^T R x_t + u_t^T Q u_t + \beta E[V(Ax_t + Bu_t + \varepsilon_{t+1})]\}, \end{aligned} \quad (9.3)$$

为了求解线性二次最优控制问题 (9.3)，我们需要两方面信息。一是政策方程 (policy function) 的具体形式  $u_t = g(x_t)$ ，即基于当期状态，决策者如何选择最优控制变量，见第9.3节。二是基于当期状态，价值方程的具体形式。

对于后者，方案设计如下：首先我们假设

$$V(x_t) = x_t^T P x_t + d, \quad (9.4)$$

即价值方程是关于状态变量的二项式形式（第9.4节验证价值方程的二项式形式假设是否成立），系数表示为幂等矩阵  $P$ ，满足  $P = P^T = P^2$ ， $d$  是个常数矩阵。

随后，将(9.4)代回(9.3)可得 Bellman equation 形式的价值方程

$$\begin{aligned} V(x_t) &= \max_{\{u_t\}} \left\{ x_t^T R x_t + u_t^T Q u_t + \beta E \left[ (Ax_t + Bu_t + \varepsilon_{t+1})^T P (Ax_t + Bu_t + \varepsilon_{t+1}) \right] + \beta d \right\} \\ &\equiv \max_{\{u_t\}} \left\{ x_t^T R x_t + u_t^T Q u_t + \beta E [\mathcal{X}] + \beta d \right\}, \end{aligned} \quad (9.5)$$

其中

$$\begin{aligned} E[\mathcal{X}] &= x_t^T A^T P A x_t + x_t^T A^T P B u_t + u_t^T B^T P A x_t + u_t^T B^T P B u_t \\ &\quad + (x_t^T A^T P \varepsilon_{t+1} + u_t^T B^T P \varepsilon_{t+1} + \varepsilon_{t+1}^T P A x_t + \varepsilon_{t+1}^T P B u_t) + \varepsilon_{t+1}^T P \varepsilon_{t+1}, \end{aligned}$$

根据定义式  $E[\varepsilon] = 0$ ，上式进一步简化为

$$E[\mathcal{X}] = x_t^T A^T P A x_t + x_t^T A^T P B u_t + u_t^T B^T P A x_t + u_t^T B^T P B u_t + \varepsilon_{t+1}^T P \varepsilon_{t+1}. \quad (9.6)$$

### 9.2.1 一阶条件

决策者最优行为可由(9.5)的一阶条件求得

$$\begin{aligned} \frac{\partial V(x_t)}{\partial u_t} &= \frac{\partial u_t^T Q u_t}{\partial u_t} + \beta \frac{\partial E[\mathcal{X}]}{\partial u_t} \\ &= \frac{\partial u_t^T Q u_t}{\partial u_t} + \beta \left( \frac{\partial x_t^T A^T P B u_t}{\partial u_t} + \frac{\partial u_t^T B^T P A x_t}{\partial u_t} + \frac{\partial u_t^T B^T P B u_t}{\partial u_t} \right) \\ &= 2Q u_t + 2\beta B^T P A x_t + 2\beta B^T P B u_t = 0, \end{aligned} \quad (9.7)$$

其中为了求得最后一行等式，我们做一些矩阵运算见第9.2.1节。

#### 计算一阶条件所需的部分矩阵运算

对于对称的常系数矩阵  $\Gamma$ ，首先我们有<sup>1</sup>

$$\frac{d}{du}(u^T \Gamma u) = u^T (\Gamma + \Gamma^T) = 2\Gamma u,$$

于是

$$\begin{aligned} \frac{\partial u_t^T Q u_t}{\partial u_t} &= 2Q u_t, \\ \frac{\partial u_t^T B^T P B u_t}{\partial u_t} &= 2B^T P B u_t. \end{aligned}$$

---

<sup>1</sup>这是由于

$$\begin{aligned} \text{tr}(d(u^T \Gamma u)) &= \text{tr}(d((u^T \Gamma)(u))) = \text{tr}(u^T \Gamma du + d(u^T \Gamma)u) = \text{tr}(u^T \Gamma du + d(u^T \Gamma)^T u) \\ &= \text{tr}(u^T \Gamma du) + \text{tr}(u^T + d(u^T \Gamma)) = \text{tr}(u^T \Gamma du) + \text{tr}(u^T \Gamma^T du) = \text{tr}(u^T (\Gamma + \Gamma^T) du). \end{aligned}$$

其次我们有<sup>2</sup>

$$\begin{aligned}\frac{d}{du} \Gamma u &= \Gamma^T, \\ \frac{d}{du} u^T \Gamma &= \Gamma,\end{aligned}$$

于是

$$\frac{\partial x_t^T A^T P B u_t}{\partial u_t} + \frac{\partial u_t^T B^T P A x_t}{\partial u_t} = (x_t^T A^T P B)^T + (B^T P A x_t) = 2 B^T P A x_t.$$

### 9.3 政策方程

重新整理(9.7)，可以得到政策方程  $u_t = g(x_t)$  的近似线性表达形式

$$u_t = - (Q + \beta B^T P B)^{-1} \beta B^T P A x_t. \quad (9.8)$$

换句话说，决策者的最优行为可以描述为如下政策方程：

$$u_t = -F x_t, \quad \text{其中} \quad (9.9)$$

$$F = \beta (Q + \beta B^T P B)^{-1} B^T P A. \quad (9.10)$$

基于政策方程(9.9)，控制变量  $u_t$  随着观测到的状态  $x_t$  做线性调整，调整依据是系数矩阵  $F$ 。 $F$  的确是一个非线性的函数形式，其值取决于两组矩阵：基础矩阵  $A, B, Q$  和（假设为二次形式）价值方程中的  $P$ 。一旦我们算出  $P$ ，便可以进一步算出  $F$  的值，从而求得政策方程的完整形式<sup>3</sup>。

#### 9.3.1 政策方程满足确定性等价条件

政策方程的政策意义价值还在于，它不受外生随机冲击  $\varepsilon$  的干扰<sup>4</sup>，这是由于在这样一个线性二次系统中，确定性等价 (certainty equivalence) 成立，见第9.3.1、9.4.1节。

### 9.4 线性政策方程和二次价值方程

如前文所述，线性政策方程 (9.9)-(9.10) 是基于（假设的）二次价值方程 (9.5)-(9.6) 得出的。现在我们反证，基于这样的线性政策方程的确可以得到二次形式的价值方程。一旦反证成功，我们便可以进一步求得  $P$  和  $d$ 。

---

<sup>2</sup>这是由于

$$\text{tr}(d u^T \Gamma) = \text{tr}(d(\Gamma^T u)^T) = \text{tr}(d(\Gamma^T u)).$$

<sup>3</sup>另一种方案是，先求得  $F$ ，根据  $F$  测算出  $P$ ，如 (Hansen and Sargent, 2004, Ch.2)

<sup>4</sup>除非以下情况出现：如第一，外生冲击彼此相关（所以我们要在模型设定中假设不相关）。第二，目标方程(9.1)并不是二次形式。

将(9.9)代入(9.5)<sup>5</sup>

$$\begin{aligned}
 V(x_t) &= x_t^T R x_t + (-F x_t)^T Q (-F x_t) + \beta d + \beta E[\varepsilon_{t+1}^T P \varepsilon_{t+1}] \\
 &\quad + \beta E \left[ x_t^T A^T P A x_t + x_t^T A^T P B (-F x_t) + (-F x_t)^T B^T P A x_t + (-F x_t)^T B^T P B (-F x_t) \right] \\
 &= x_t^T R x_t + x_t^T F^T Q F x_t + \beta d + \beta E[\varepsilon_{t+1}^T P \varepsilon_{t+1}] \\
 &\quad + \beta E \left[ x_t^T A^T P A x_t - x_t^T A^T P B F x_t - x_t^T F^T B^T P A x_t + x_t^T F^T B^T P B F x_t \right] \\
 &= x_t^T R x_t + x_t^T F^T Q F x_t + \beta E \left[ x_t^T A^T P A x_t - 2x_t^T A^T P B F x_t + x_t^T F^T B^T P B F x_t \right] \\
 &\quad + \beta E[\varepsilon_{t+1}^T P \varepsilon_{t+1}] + \beta d.
 \end{aligned} \tag{9.11}$$

联立(9.4)和(9.11)，我们有

$$\begin{cases} d = \beta E[\varepsilon_{t+1}^T P \varepsilon_{t+1}] + \beta d, \\ x_t^T P x_t = x_t^T R x_t + x_t^T F^T Q F x_t + \beta E \left[ x_t^T A^T P A x_t - 2x_t^T A^T P B F x_t + x_t^T F^T B^T P B F x_t \right]. \end{cases}$$

#### 9.4.1 价值方程满足确定性等价条件

来看第一个等式，整理可得<sup>6</sup>

$$d = \frac{\beta}{1 - \beta} \text{tr}(P\Sigma), \tag{9.12}$$

可见随机冲击  $\varepsilon$  尽管对价值方程产生影响，但不是通过  $F$  至政策方程进而影响价值方程，而是通过常数项  $d$  影响价值方程的。因此，对于最优价值方程而言，确定性等价条件依然成立。

---

<sup>5</sup>第三个等号所需条件：根据定义  $P = P^T$ ，因此  $x_t^T F^T B^T P A x_t$  和  $x_t^T A^T P B F x_t$  都是标量，且满足

$$x_t^T F^T B^T P A x_t = x_t^T A^T P B F x_t.$$

<sup>6</sup>这是由于

$$E[\varepsilon_{t+1}^T P \varepsilon_{t+1}] = \text{tr}(E[\varepsilon_{t+1}^T P \varepsilon_{t+1}]) = \text{tr}(P E[\varepsilon_{t+1}^T \varepsilon_{t+1}]) = \text{tr}(P E[\Sigma_{t+1}]).$$

### 9.4.2 代数矩阵 Riccati 方程

第二个等式整理，并引入最优政策方程(9.10)替换  $F$  得<sup>7</sup>

$$\begin{aligned}
 P &= R + F^T Q F + \beta \cdot (A^T P A - 2A^T P B F + F^T B^T P B F) \\
 &= R + \beta A^T P A - 2\beta A^T P B F + F^T (Q + \beta B^T P B) F \\
 &= R + \beta A^T P A - 2\beta A^T P B (Q + \beta B^T P B)^{-1} \beta B^T P A \\
 &\quad + \beta A^T P B ((Q + \beta B^T P B)^{-1})^T (Q + \beta B^T P B) (Q + \beta B^T P B)^{-1} \beta B^T P A \\
 &= R + \beta A^T P A - \beta^2 A^T P B (Q + \beta B^T P B)^{-1} \beta B^T P A. \tag{9.13}
 \end{aligned}$$

(9.13)又称线性矩阵 Riccati 方程 (linear matrix Riccati equation)，它表明价值方程中的  $P$  是一个与基础矩阵  $A, B, R, Q$  有关的函数，呈非线性关系。

此外，由(9.13)可以看出，一个线性的最优政策方程(9.9)-(9.10)，的确预示着二次形式的价值方程 (9.4),(9.13)。

## 9.5 数值方法

在线性二次控制中，在利用解析法推得线性代数 Riccati 方程后，需要依赖数值计算的方法，对  $P$  的矩阵 Riccati 差分方程做迭代近似。假定期初猜测值为  $P_j$ ，经过 1 次迭代， $P_{j+1}$  的值更新至

$$P_{j+1} = R + \beta A^T P_j A - \beta^2 A^T P_j B (Q + \beta B^T P B)^{-1} \beta B^T P_j A. \tag{9.14}$$

重复迭代，直至  $P$  收敛到某一值。根据数值模拟的  $P$  值计算  $F$  值，进而得到最优政策方程和价值方程。

## 9.6 范例

假定这样一个线性二次控制问题。中央银行试图通过控制利率  $r_t$  来影响通货膨胀率  $\pi_t$  和产出  $y_t$ ，收益方程 (payoff function)

$$\mathcal{L}_t = \pi_t^2 + y_t^2 + 0.1r_t^2. \tag{9.15}$$

经济系统的结构，由两个状态变量的运动定律构成

$$\pi_{t+1} = 0.75\pi_t - 0.5r_t + \varepsilon_{\pi,t+1}, \tag{9.16a}$$

$$y_{t+1} = 0.25y_t - 0.5r_t + \varepsilon_{y,t+1}. \tag{9.16b}$$

---

<sup>7</sup>最后一个等号的计算依据如下：

$$\begin{aligned}
 &\beta A^T P B \left( (Q + \beta B^T P B)^{-1} \right)^T (Q + \beta B^T P B) (Q + \beta B^T P B)^{-1} \beta B^T P A \\
 &= \beta A^T P B \left( (Q + \beta B^T P B)^T \right)^{-1} (Q + \beta B^T P B) (Q + \beta B^T P B)^{-1} \beta B^T P A \\
 &= \beta A^T P B \left( (Q + \beta B^T P B)^T \right)^{-1} (Q + \beta B^T P B)^T (Q + \beta B^T P B)^{-1} \beta B^T P A \\
 &= \beta^2 A^T P B (Q + \beta B^T P B)^{-1} B^T P A.
 \end{aligned}$$

作为决策者，中央银行的最大化问题为

$$\max_{\{r_t\}} E \sum_{t=0}^{\infty} \beta^t \mathcal{L}_t, \quad \text{s.t. (9.16a) - (9.16b)}$$

### 9.6.1 最优线性二次控制问题

将这一问题改写为最优线性二次控制的一般形式：

$$\begin{aligned} & \max_{\{u_t\}} E \sum_{t=0}^{\infty} \beta^t (x_t^T R x_t + u_t^T Q u_t), \\ & \text{s.t. } x_{t+1} = Ax_t + Bu_t + \varepsilon_t, \end{aligned}$$

其中状态变量，控制变量，外生扰动变量分别用  $x_t, u_t, \varepsilon_t$  表示

$$x_t \equiv \begin{pmatrix} \pi_t \\ y_t \end{pmatrix}, \quad u_t \equiv r_t, \quad \varepsilon_t = \begin{pmatrix} \varepsilon_{\pi,t} \\ \varepsilon_{y,t} \end{pmatrix},$$

基础系数矩阵  $A, B, R, Q$  分别为

$$A = \begin{pmatrix} 0.75 & 0 \\ 0 & 0.25 \end{pmatrix}, \quad B = \begin{pmatrix} -0.5 \\ -0.5 \end{pmatrix}, \quad R = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad Q = 0.1.$$

如前文所述，我们需要使用数值方法，递归处理线性矩阵 Riccati 方程以近似价值方程系数矩阵  $P$ ，进而测算政策方程的系数矩阵  $F$ ，对应 (9.14) (9.10)。

### 9.6.2 Matlab 程序

```

1 clear;
2 %时间贴现系数的确定
3 beta = 0.99;
4
5 %定义矩阵
6 R=zeros(2,2);
7 Q=zeros(1,1);
8 A=zeros(2,2);
9 B=zeros(2,1);
10
11 %为矩阵赋值
12 R(1,1)=1;
13 R(2,2)=1;
14 Q(1,1)=0.1;
15 A(1,1)=0.75;
16 A(2,2)=0.25;
17 B(1,1)=-0.5;
18 B(2,1)=-0.5;
19
20 %matrix Riccati equation迭代

```

```

21 %d描述$P_j$和邻近迭代$P_{j+1}$之间元素的最大绝对值偏差
22 d = 1;
23 %i记录迭代次数
24 i = 0;
25 % 最大绝对值偏差d保存在D; 迭代次数i保存在I
26 D=0;
27 I=0;
28 %迭代P的初始值定义为P0
29 %说明: 定义P0和Q的非零值, 为了使得linear matrix Riccati equation迭代操作中
30 %%$(Q+\beta B^T P_j B) \neq 0$. 只有这个值不等于0, 才能使得该矩阵是可逆的。
31 P0=zeros(2,2);
32 P0(1,1)=-0.000001;
33 P0(2,2)=-0.000001;
34
35
36 format long
37
38 %matrix Riccati Equation 迭代
39 %迭代新生成的$P_{j+1}$写入矩阵P1。
40 %两个P值的差值, 写入矩阵Pd。
41 %当d<0.0000000001时, 迭代终止。否则继续进行。
42 %每次迭代的d和i写入D和I, 以备最终程序输出。
43 while d > 0.0000000001
44 P1 = R + beta * A' * P0 * A - beta^2 * A' * P0 * B * (inv(Q+beta * B' *
45     P0 * B)) * (B' * P0 * A);
46 Pd = P1 - P0;
47 d = max(abs(Pd));
48 d = max(d');
49 D=[D d];
50 P0 = P1;
51 i = i+1;
52 I = [I i];
53 end
54
55 %根据迭代生成的P值, 计算policy function的矩阵F
56 P=P0;
57 F = -inv(Q + beta * B' * P * B) * (beta * B' * P * A);
58 ID = [I(2:length(I))', D(2:length(I))'];
59 disp('      i          d');
60 disp(ID);
61 disp('  SOLUTIONS');
```

```

61 disp('F');
62 disp(F);
63 disp('P');
64 disp(P);

```

### 9.6.3 Matlab 程序输出

运行 Matlab 程序，输出如下

```

1 i d
2 1.000000000000000 1.000000938124847
3 2.000000000000000 0.325234163624664
4 3.000000000000000 0.080558191789241
5 4.000000000000000 0.018875706161742
6 5.000000000000000 0.004338686564976
7 6.000000000000000 0.000992730600232
8 7.000000000000000 0.000226907809828
9 8.000000000000000 0.000051851748175
10 9.000000000000000 0.000011848232754
11 10.000000000000000 0.000002707312049
12 11.000000000000000 0.000000618616947
13 12.000000000000000 0.000000141352999
14 13.000000000000000 0.000000032298933
15 14.000000000000000 0.000000007380254
16 15.000000000000000 0.000000001686376
17 16.000000000000000 0.000000000385334
18 17.000000000000000 0.000000000088048
19
20 SOLUTIONS
21 F
22 0.744954171236066 0.175909878487998
23
24 P
25 1.430293038776173 -0.106183304364736
26 -0.106183304364736 1.044189928712962

```

由此可得最优政策方程

$$u_t = -Fx_t, \quad r_t = -F \begin{pmatrix} \pi_t \\ y_t \end{pmatrix}, \quad F = (0.745, 0.176) \quad (9.17)$$

$$r_t = 0.745\pi_t + 0.176y_t. \quad (9.18)$$

价值方程

$$V(\pi_t, y_t) = 1.430\pi_t^2 + 1.044y_t^2 - 0.212y_t. \quad (9.19)$$

根据最优政策方程，当通胀水平和/或产出水平高于稳定状态时，中央银行应当提高利率<sup>8</sup>。利率对通胀波动的响应幅度高于对产出波动的响应 (0.745 相对于 0.176)，这与模型的假设条件有关：(9.16a)-(9.16b)假定通胀波动比产出波动更为持久 (0.75 相对于 0.25)，从而过去时段的物价波动，比起产出波动来，更有可能对当期经济系统产生影响。

根据价值方程，首先同样地，通胀波动二次项的系数大于产出波动二次项的系数。交互项的系数为负，反映了当通胀和产出的波动同方向变化时，比如通货膨胀伴随产出增加（或通货紧缩伴随产出减少），中央银行更容易通过调节利率这一政策工具来稳定经济运行。反之，如果两种波动反方向变化，稳定经济运行将更为困难。

---

<sup>8</sup>注意式中的变量  $z_t = (\tilde{z}_t - z)/z$ ，表示实际观测到的变量，相对于稳态水平的 deviation。



# 第十章 理性期望模型

## 10.1 简介

利用动态规划 (dynamic programming) 的方法求解 Ramsey 随机增长模型，一个核心假定是只存在一个典型的经济行为个体，他追求自身利益最大化的行为带来社会福利最大化。我们称这样的经济体为 centralized economy，均衡状态处于 Pareto optimality，这个经济行为人称为 social planner。

然而现实中更常见 decentralized economy 的情况，即存在异质的多个经济个体，他们的最大目标各异，如厂商追求利润最大化，家庭追求效用最大化，劳动者做劳动——休闲的最优决策，等等。的确，根据福利经济学第二定律 (the second fundamental theorem of welfare economics, (Mas-Colell et al., 1995, p.151)), 在某些极端情况下，完全竞争的 decentralized 均衡状态可以导致 social planner 的 Pareto optimality；但对于更一般的情况，当 decentralized economies 中存在局部摩擦如价格/工资粘性时，decentralized 均衡并不必然导致 centralized 的 Pareto optimality。换句话说，我们无法通过求解 social planner 问题来推得 decentralized 均衡。我们只能直接从不同异质经济个体的一阶条件 (FOC) 入手，构建这样一组线性随机一阶差分方程，即理性期望模型。

求解理性期望模型的核心在线性化：在将多阶自回归改写成一阶自回归形式之后，如何将原本是高度非线性的一阶自回归系统做线性近似。对于稍微复杂一些的系统而言，往往无法直接求得解析解，替代方案为：首先将模型变量围绕其稳定状态做对数线性化处理求得解析式；随后根据一定的数值算法，求解线性化方程组，将相关内生变量改写为自 VAR 形式。

此种求解思路主要来自 King et al. (1988a,b)<sup>1</sup>，大致说来分为四个步骤：

1. 计算稳定状态，
2. 将解释变量围绕稳定状态做近似，求解析式，
3. 模型参数校准，
4. 根据数值算法求得政策方程，将内生变量与外生变量和前定变量联系起来。

该思路在宏观经济研究中得到了较为广泛的应用，主要得益于其优点：存在多种可供选择的数值算法，可以在一定程度上近似非线性一阶条件的线性表达式，而只需要付出一定的计算机处理时间。也正因如此，该思路也存在着适用性的局限。

1. 围绕稳定状态做对数线性近似，其前提假定是模型接近对数线性形式。而模型非线性的程度越高，模型的规模越大，考虑的因素越是多（比如消费者的风厌恶程度越高，外生冲击的种类越多），对数线性近似导致的失真情况就越严重。
2. 稳定状态无法在模型内部求得，并且对于存在多重稳态的经济系统来说，情况会变的更复杂。

---

<sup>1</sup> 此外可见 King and Rebelo (1999)。

## 10.2 数值算法

针对一阶线性近似的变型系统，常见的数值算法，大致说来有<sup>2</sup>

1. 特征值——特征向量分解法 (eigenvalue-eigenvector decomposition)

最早由Blanchard and Kahn (1980) 提出，因此也称为 Blanchard-Kahn Algorithm。模型要求将所有内生变量分为两类，一类为状态变量，主要指前定变量。其他变量归入第二类跳跃变量 (jump variables)。通过特征根——特征变量分解，该算法求得跳跃变量爆炸根 (explosive root) 的数量，进而判断系统解是否存在，以及是否唯一。见第10.4节。

对于满足一些假定条件的（比如非奇异方块矩阵）常规系数矩阵，特征根——特征向量分解法提供了一种较好的求解思路。由此优点出发，在算法方面，一些后续研究作出不断改进 (King and Watson, 1998, 2002; Anderson and Moore, 1985; Anderson, 1998; Sims, 2002; Klein, 2000; Söderlind, 1999)。

2. 未定系数法 (undetermined coefficients)

最早由McCallum (1983) 提出，随后的一系列重要扩展包括Uhlig (1999); Binder and Pesaran (1995); Christiano (2002) 等。这种方法不再将变量区分为前定与非前定变量，根据未定系数近似系统的数值解。见第10.5节。

其基本思路是：第一，假定系统存在 1 个解，根据这个解，内生变量是关于状态变量的线性方程，并且这个假定不能使完全随机的猜测。第二，将猜测解代回结构方程系统中，构建关于待定系数的方程组。第三，利用二次方程矩阵的解法，求得该方程组的解，进而整个线性系统的解。

未定系数法具有可操作性和计算速度等优势。但存在不足：第一，需要预先假定模型存在唯一解。第二，仅当系统中没有冗余变量时（系数矩阵的列线性不相关），状态——空间表现形式处于最简规模，该方法才适用。否则值为 0 的特征值会导致泡沫解的出现 (Canova, 2011, p.57)。

3. 期望误差法 (expectational errors)

由Sims (2002) 提出，也不再做前定、非前定变量的区分，更可以进一步探讨理性期望下期望误差的性质。见第10.6节。

4. 参数化期望算法 (PEA, Parameterised Expectations Algorithm)，见第10.7节。

5. Schur 分解法 (Schur Decomposition)，又称 QZ 分解法，作为更为通用的形式，主要用于奇异（不可逆）系数矩阵的情况 (King and Watson, 1998, 2002; Söderlind, 1999; Klein, 2000)，见第10.8节。

大致说来，各种方法之间的主要区别在于

- 构建稳定解模块的方式，
- 求解过程中对理性期望的处理，
- 保留大部分的非线性成分，交给数值算法去近似处理，
- 对前定和非前定变量的区分等。

---

<sup>2</sup>一个更为详尽的综述，见 (Milani, 2012)。

根据King et al. (1988a,b) 的四步骤求解思路, 第10.3节首先建立一个简单的随机增长模型。在此基础上, 第10.3.1节计算稳定状态, 第10.3.2节做对数线性化近似, 第10.4-10.8节分别介绍几种主要的算法。

## 10.3 一个简单的随机增长模型

在这样一个简单随机经济增长模型中, 典型个体追求最大化问题

$$\begin{aligned} \max_{\{C_t\}} E \sum_{t=0}^{\infty} \beta^t \cdot \left( \frac{C_t^{1-\sigma}}{1-\sigma} \right), \quad \text{s.t.} \\ C_t + K_{t+1} = A_t \cdot K_t^\alpha + (1-\delta) \cdot K_t, \\ \ln A_t = \rho \cdot \ln A_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim i.i.d.(0, \sigma^2), 0 < \rho < 1. \end{aligned}$$

模型的均衡解以  $\{C_t, K_t, Y_t\}_{t=0}^{\infty}$  的形式展现。

求解一阶条件我们有

$$\begin{cases} C_t^{-\sigma} &= \beta E [C_{t+1}^{-\sigma} \cdot (\alpha \cdot A_{t+1} \cdot K_{t+1}^{\alpha-1} + 1 - \delta)] \\ C_t + K_{t+1} &= A_t \cdot K_t^\alpha + (1 - \delta) \cdot K_t \\ \ln A_t &= \rho \cdot \ln A_{t-1} + \varepsilon_t \end{cases}$$

### 10.3.1 稳定状态

在稳定状态下我们有

$$\begin{bmatrix} C_t \\ K_t \\ Y_t \end{bmatrix} \equiv \begin{bmatrix} \bar{C} \\ \bar{K} \\ \bar{Y} \end{bmatrix}, \quad A_t \equiv \bar{A} = 1, \quad \forall t.$$

进而我们有

$$\begin{cases} 1 &= \beta \cdot [\alpha \cdot \bar{K}^{\alpha-1} + (1 - \delta)], \\ \bar{C} + \bar{K} &= \bar{K}^\alpha + (1 - \delta) \cdot \bar{K}. \end{cases}$$

整理后得稳定状态

$$\begin{cases} \bar{C} &= \bar{K}^\alpha - \delta \cdot \bar{K}, \\ \bar{K} &= \left( \frac{1 - (1 - \delta) \cdot \beta}{\alpha \cdot \bar{K}} \right)^{\frac{1}{\alpha-1}}, \\ \bar{Y} &= \bar{K}^\alpha. \end{cases}$$

### 10.3.2 对数线性化

对数线性的定义式可表示为, 对于一个变量  $X_t$ :

$$\tilde{X}_t \equiv \frac{X_t - \bar{X}}{\bar{X}} \approx \ln X_t - \ln \bar{X},$$

对于含有不止一个变量的复杂方程  $f(X_t, Y_t)$ , 对其稳定状态  $(\bar{X}, \bar{Y})$  做一阶泰勒级数展开的方式为

$$\ln f(X_t, Y_t) \approx \ln f(\bar{X}, \bar{Y}) + \left[ \frac{\frac{\partial f(X_t, Y_t)}{\partial X_t}|_{\{\bar{X}, \bar{Y}\}}}{f(\bar{X}, \bar{Y})} \right] \cdot (X_t - \bar{X}) + \left[ \frac{\frac{\partial f(X_t, Y_t)}{\partial Y_t}|_{\{\bar{X}, \bar{Y}\}}}{f(\bar{X}, \bar{Y})} \right] \cdot (Y_t - \bar{Y}).$$

### 外生技术冲击的对数线性化

$$\tilde{A}_t \approx \rho \cdot \tilde{A}_{t-1} + \varepsilon_t. \quad (10.1)$$

#### Euler equation 的对数线性化

等式两侧取对数

$$-\sigma \cdot \ln C_t = \ln \beta - \sigma \cdot E \ln C_{t+1} + E \ln [\alpha \cdot A_{t+1} \cdot K_{t+1}^{\alpha-1} + (1 - \delta)]$$

RHS 第三部分，围绕  $\{\bar{K}, \bar{A}\}$  做一阶泰勒展开

$$\begin{aligned} & \ln [\alpha \cdot A_{t+1} \cdot K_{t+1}^{\alpha-1} + (1 - \delta)] \\ & \approx \ln [\alpha \bar{A} \bar{K}^{\alpha-1} + (1 - \delta)] \\ & + \frac{\frac{\partial}{\partial K} [\alpha \cdot \bar{A} \cdot \bar{K}^{\alpha-1} + (1 - \delta)] \cdot \frac{K_{t+1} - \bar{K}}{\bar{K}} \cdot \bar{K}}{\alpha \cdot \bar{A} \cdot \bar{K}^{\alpha-1} + (1 - \delta)} + \frac{\frac{\partial}{\partial A} [\alpha \cdot \bar{A} \cdot \bar{K}^{\alpha-1} + (1 - \delta)] \cdot \frac{A_{t+1} - \bar{A}}{\bar{A}} \cdot \bar{A}}{\alpha \cdot \bar{A} \cdot \bar{K}^{\alpha-1} + (1 - \delta)} \\ & = \ln [\alpha \cdot \bar{A} \cdot \bar{K}^{\alpha-1} + (1 - \delta)] + \frac{\alpha \cdot \bar{A} \cdot (\alpha - 1) \cdot \bar{K}^{\alpha-1} \cdot \tilde{K}_{t+1} + \alpha \cdot \bar{K}^{\alpha-1} \cdot \bar{A} \cdot \tilde{A}_{t+1}}{\alpha \cdot \bar{A} \cdot \bar{K}^{\alpha-1} + (1 - \delta)} \\ & = \ln [\alpha \cdot \bar{A} \cdot \bar{K}^{\alpha-1} + (1 - \delta)] + \frac{\alpha \cdot \bar{K}^{\alpha-1} \cdot [(\alpha - 1) \cdot \tilde{K}_{t+1} + \tilde{A}_{t+1}]}{\alpha \cdot \bar{A} \cdot \bar{K}^{\alpha-1} + (1 - \delta)} \\ & = -\ln \beta + [1 - \beta \cdot (1 - \delta)] \cdot [(\alpha - 1) \cdot \tilde{K}_{t+1} + \tilde{A}_{t+1}], \end{aligned}$$

进而我们有

$$-\sigma \cdot \tilde{C}_t = -\sigma \cdot \tilde{C}_{t+1} + [1 - \beta \cdot (1 - \delta)] \cdot [(\alpha - 1) \cdot \tilde{K}_{t+1} + \tilde{A}_{t+1}] \quad (10.2)$$

#### 预算约束条件的对数线性化

等式两侧取对数

$$\ln(C_t + K_{t+1}) = \ln[A_t \cdot K_t^\alpha + (1 - \delta) \cdot K_t]$$

LHS  $\Rightarrow$

$$\ln(C_t + K_{t+1}) \approx \ln(\bar{C} + \bar{K}) + \frac{\frac{C_t - \bar{C}}{\bar{C}} \cdot \bar{C} + \frac{K_{t+1} - \bar{K}}{\bar{K}} \cdot \bar{K}}{\bar{C} + \bar{K}} = \ln(\bar{C} + \bar{K}) + \frac{\bar{C} \cdot \tilde{C}_t + \bar{K} \cdot \tilde{K}_{t+1}}{\bar{C} + \bar{K}}.$$

RHS  $\Rightarrow$

$$\begin{aligned} & \ln[A_t \cdot K_t^\alpha + (1 - \delta) \cdot K_t] \\ & \approx \ln[\bar{A} \cdot \bar{K}^\alpha + (1 - \delta) \cdot \bar{K}] + \frac{\frac{\partial}{\partial K} [\bar{A} \bar{K}^\alpha + (1 - \delta) \cdot \bar{K}] \cdot \frac{K_t - \bar{K}}{\bar{K}} \cdot \bar{K}}{\bar{A} \cdot \bar{K}^\alpha + (1 - \delta) \cdot \bar{K}} + \frac{\frac{\partial}{\partial A} [\bar{A} \bar{K}^\alpha + (1 - \delta) \cdot \bar{K}] \cdot \frac{A_t - \bar{A}}{\bar{A}} \cdot \bar{A}}{\bar{A} \cdot \bar{K}^\alpha + (1 - \delta) \cdot \bar{K}} \\ & = \ln[\bar{A} \cdot \bar{K}^\alpha + (1 - \delta) \cdot \bar{K}] + \frac{[\alpha \cdot \bar{A} \cdot \bar{K}^\alpha + (1 - \delta) \cdot \bar{K}] \cdot \tilde{K}_t + \bar{A} \cdot \bar{K}^\alpha \cdot \tilde{A}_t}{\bar{A} \cdot \bar{K}^\alpha + (1 - \delta) \cdot \bar{K}} \\ & = \ln[\bar{A} \cdot \bar{K}^\alpha + (1 - \delta) \cdot \bar{K}] + \frac{\frac{1}{\beta} \cdot \bar{K} \cdot \tilde{K}_t + \bar{A} \cdot \bar{K}^\alpha \cdot \tilde{A}_t}{\bar{A} \cdot \bar{K}^\alpha + (1 - \delta) \cdot \bar{K}} \\ & = \ln(\bar{C} + \bar{K}) + \frac{\frac{1}{\beta} \cdot \bar{K} \cdot \tilde{K}_t + \bar{K}^\alpha \cdot \tilde{A}_t}{\bar{C} + \bar{K}}. \end{aligned}$$

$LHS = RHS \Rightarrow$

$$\bar{C} \cdot \tilde{C}_t + \bar{K} \cdot \tilde{K}_{t+1} = \frac{1}{\beta} \cdot \bar{K} \cdot \tilde{K}_t + \bar{K}^\alpha \cdot \tilde{A}_t,$$

因此我们有

$$\begin{aligned}\tilde{K}_{t+1} &= -\frac{\bar{C}}{\bar{K}} \cdot \tilde{C}_t + \frac{1}{\beta} \cdot \tilde{K}_t + \bar{K}^{\alpha-1} \cdot \tilde{A}_t \\ &= -\frac{1-\beta \cdot [1-\delta \cdot (1-\alpha)]}{\alpha \cdot \beta} \cdot \tilde{C}_t + \frac{1}{\beta} \cdot \tilde{K}_t + \frac{1-\beta \cdot (1-\delta)}{\alpha \cdot \beta} \cdot \tilde{A}_t.\end{aligned}\quad (10.3)$$

### 10.3.3 状态——空间表现形式

一阶差分式(10.1)、(10.2)、(10.3)共同描述这样一个动态的经济系统  $\{C_t, K_t, A_t\}_{t=0}^\infty$ 。可以将其改写为如下状态——空间表现形式 (state-space representation)

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \sigma \end{bmatrix} E \begin{bmatrix} \tilde{A}_{t+1} \\ \tilde{K}_{t+1} \\ \tilde{C}_{t+1} \end{bmatrix} = \begin{bmatrix} \rho & 0 & 0 \\ \frac{1-\beta \cdot (1-\delta)}{\alpha \cdot \beta} & \frac{1}{\beta} & -\frac{1-\beta \cdot [1-\delta \cdot (1-\alpha)]}{\alpha \cdot \beta} \\ \rho \cdot [1-\beta \cdot (1-\delta)] & -[1-\alpha] \cdot [1-\beta \cdot (1-\delta)] & \sigma \end{bmatrix} \begin{bmatrix} \tilde{A}_t \\ \tilde{K}_t \\ \tilde{C}_t \end{bmatrix}. \quad (10.4)$$

## 10.4 特征值——特征方程分解法

### 10.4.1 特征值——特征方程分解法

根据 Blanchard-Kahn 算法，要将经济系统中的  $p+m+k$  个变量分类。一类是  $k$  个外生变量  $z_t$ 。一类是内生变量  $x_t$ 。内生变量再分为  $m$  个控制变量 (跳跃变量)  $x_t^j$  和  $p$  状态变量 (前定变量)  $x_t^s$ 。(10.4)可以改写为

$$\mathcal{T}_0 E x_{t+1} = \mathcal{T}_1 x_t + \Psi z_t, \quad x_t \equiv \begin{bmatrix} x_t^s & x_t^j \end{bmatrix}' . \quad (10.5)$$

假定系数矩阵  $\mathcal{T}_0$  是可逆的，上式进一步调整为

$$\begin{aligned}E x_{t+1} &= Ax_t + Bz_t, \quad \text{其中 } A \equiv \mathcal{T}_0^{-1} \mathcal{T}_1, \quad B \equiv \mathcal{T}_0^{-1} \Psi, \quad \text{或者} \\ E \begin{bmatrix} x_{t+1}^s \\ x_{t+1}^j \end{bmatrix} &= \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_t^s \\ x_t^j \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} z_t\end{aligned}\quad (10.6)$$

Blanchard-Kahn 算法的核心就在于，对系数矩阵  $A$  做 Jordan decomposition。对于可对角化 (diagonalizable) 的系数矩阵  $A$ ，假定其特征向量是序列不相关的<sup>3</sup>，我们有  $AP = P\Lambda$ ，改写为 Jordan canonical form

$$A = P\Lambda P^{-1}, \quad (10.7)$$

其中  $\Lambda$  是对角矩阵，对角元素对应特征根  $\Lambda_{ii} = \lambda_i$ ，其他元素均为 0。 $P$  的每一列为对应  $\lambda_i$  的特征向量。我们用  $\bar{p}$  和  $\bar{m}$  来表示用稳定和不稳定特征根来为矩阵分块的情况，后面会简要探讨当  $\bar{p} \neq p, \bar{m} \neq m$  时可能会出现的问题，见第10.4.2节。

根据 Blanchard-Kahn 条件，不稳定特征根 (即  $|\lambda| > 1$ ) 的数量应该恰好等于经济系统中控制变量的数量，以确保相图中鞍点稳定性的存在。如果不稳定特征根的数量少于控制变量的数量，经济系统超稳定，出现未定问题 (indeterminacy)<sup>4</sup>。如果多于，经济系统爆炸性发展，会违反横截条件。一个较为详细的讨论见第10.4.2节。

<sup>3</sup>对于  $n$  个长度为  $m$  的向量之间线性不相关，是指对于由这  $n$  个向量构成的  $m \times n$  矩阵  $X$  来说， $\det(X) \neq 0$ 。

<sup>4</sup>对未定经济系统问题的探讨，可见如 Benhabib and Farmer (1999)。

在本章的动态 Ramsey 经济模型中，内生变量是二维的，由一个控制变量和一个前定状态变量构成。因此需要恰好有一个不稳定特征根和一个稳定特征根。

将特征矩阵做进一步的分解，根据特征根绝对值从低到高做重新排列（相应地，需要调整特征矩阵的列）

$$\Lambda = \begin{bmatrix} \Lambda_s & 0 \\ \bar{p} \times \bar{p} & \bar{p} \times \bar{m} \\ 0 & \Lambda_e \\ \bar{m} \times \bar{p} & \bar{m} \times \bar{m} \end{bmatrix}, \quad P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}.$$

其中对角矩阵  $\Lambda_s(\Lambda_e)$  的对角元素是绝对值小于（大于）1 的特征根。系统(10.6)由此改写为

$$Ew_{t+1} = \Lambda w_t + \bar{B} z_t, \quad \text{其中 } w_t \equiv P^{-1} x_t, \bar{B} \equiv P^{-1} B. \quad (10.8)$$

对应地我们有

$$w_t = \begin{bmatrix} w_{1,t} \\ \bar{p} \times 1 \\ w_{2,t} \\ \bar{m} \times 1 \end{bmatrix}, \quad \bar{B} = \begin{bmatrix} \bar{B}_1 \\ \bar{p} \times \bar{k} \\ \bar{B}_2 \\ \bar{m} \times k \end{bmatrix}.$$

根据上述分析，由(10.8)我们有

$$Ew_{1,t+1} = \Lambda_s w_{1,t} + \bar{B}_1 z_t, \quad (10.9a)$$

$$Ew_{2,t+1} = \Lambda_e w_{2,t} + \bar{B}_2 z_t. \quad (10.9b)$$

根据模型设定，(10.9a)总是稳定的，因为  $\Lambda_s$  的对角元素由绝对值小于 1 的特征根构成。(10.9b)是爆炸的，因为  $\Lambda_e$  的对角元素由绝对值大于 1 的特征根构成。因此我们先来考察(10.9b)成立需要满足的条件。

我们对  $w_{2,t}$  做 forward-looking 迭代，求解等式。(10.9b)意味着

$$\begin{aligned} w_{2,t} &= (\Lambda_e)^{-1} Ew_{2,t+1} - \bar{B}_2 z_t, \\ w_{2,t+1} &= (\Lambda_e)^{-1} Ew_{2,t+2} - \bar{B}_2 z_{t+1}, \\ &\vdots \\ w_{2,t+T} &= (\Lambda_e)^{-1} Ew_{2,t+T} - \bar{B}_2 z_{t+T}. \end{aligned}$$

根据 forward-looking 迭代我们有

$$w_{2,t} = \lim_{T \rightarrow \infty} \Lambda_e^{-T} E(w_{2,t+T}) - \sum_{s=0}^T \Lambda_e^{-s-1} \bar{B}_2 E(z_{t+s}). \quad (10.10)$$

由于  $\Lambda_e$  中所有对角元素绝对值都大于 1，对于  $T \rightarrow \infty$  可得  $\Lambda_e^T \rightarrow 0$ ，因此改写上式， $w_{2,t}$  的值可以计算如下

$$w_{2,t} = - \sum_{s=0}^T \Lambda_e^{-s-1} \bar{B}_2 E(z_{t+s}). \quad (10.11)$$

回到反映  $w_t$  和  $x_t$  关系的(10.8)中，根据定义

$$x_t \equiv P w_t \Rightarrow \begin{bmatrix} x_{1,t} \\ \bar{p} \times 1 \\ x_{2,t} \\ \bar{m} \times 1 \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} w_{1,t} \\ \bar{p} \times 1 \\ w_{2,t} \\ \bar{m} \times 1 \end{bmatrix},$$

等价于

$$x_{1,t} = P_{11} w_{1,t} + P_{12} w_{2,t}, \quad (10.12a)$$

$$x_{2,t} = P_{21} w_{1,t} + P_{22} w_{2,t}. \quad (10.12b)$$

如果假定  $\bar{p} = p, \bar{m} = m$ , 即系统中控制变量的数量等于不稳定特征根的数量, 并且假定  $P_{11}$  是可逆分块矩阵, 则  $x_{1,t} = x_t^s, x_{2,t} = x_t^j$ , 对于二者不相等的情况, 见第10.4.2节。对于最终形成 Blanchard-Kahn 算法的求解思路:

1. 从初始状态  $t = 0$  出发, 对应给定的  $x_{t=0}^s$ 。

2. 结合生成的  $z_t$ , 根据(10.11)计算  $w_{2,t}$ :

$$w_{2,t} = - \sum_{s=0}^T \Lambda_e^{-s-1} \bar{B}_2 E(z_{t+s}).$$

3. 根据(10.12a)得到  $w_{1,t}$  的值:

$$w_{1,t} = P_{11}^{-1} x_t^s - P_{11}^{-1} P_{12} w_{2,t},$$

其中前定变量  $x_t^s$  由上一期的状态求得。

4. 根据(10.12b)得到控制变量  $x_t^j$  的值, 又称政策方程:

$$x_t^j = P_{21} w_{1,t} + P_{22} w_{2,t}.$$

5. 根据 (10.6)得到  $t+1$  期状态变量的期望值  $x_{t+1}^s$ :

$$E x_{t+1}^s = A_{11} x_t^s + A_{12} x_t^j + B_1 z_t,$$

6. 重复以上步骤。

### 10.4.2 解的存在性以及唯一性

如前文所述, 当  $p = \bar{p}, m = \bar{m}$  时, 控制变量 (状态变量) 的数量等于爆炸 (稳定) 特征根的数量, 此时系统均衡解存在且唯一。

但当  $m < \bar{m}$  时,  $p > \bar{p}$ , 情况有所不同。我们仍然可以根据(10.12a)得到  $w_{1,t}$  的值, 尽管  $w_{1,t}$  的值不唯一。但在根据(10.12b)测算  $x_t^j$  时, 会受到较大限制, 比如在  $t = 0$  时:

$$x_{2,t_0} = P_{21} w_{1,t_0} + P_{22} w_{2,t_0},$$

此时,  $x_{2,t_0}$  的  $\bar{m}$  行中, 最上面的  $\bar{m} - m \equiv p - \bar{p}$  行是状态变量, 因此是提前给定的。这意味着我们没有足够数量的前定 (状态) 变量用于求解方程。换句话说, 对整个经济系统而言, 当跳跃变量的数量少于爆炸特征根时, 系统的均衡解不存在。

当  $m > \bar{m}$  时,  $p < \bar{p}$ , 在  $t = 0$  时(10.12a)变为

$$w_{1,t_0} = P_{11}^{-1} x_{1,t_0} - P_{11}^{-1} P_{12} w_{2,t_0},$$

$x_{1,t_0}$  的  $\bar{p}$  行中, 最下面的  $\bar{p} - p \equiv m - \bar{m}$  行是跳跃变量, 这会导致解出的  $w_{1,t}$  并不唯一。这意味着我们可以随意选择这  $\bar{p} - p$  行跳跃变量的初始值。换句话说, 对整个经济系统而言, 当跳跃变量的数量多于爆炸特征根时, 系统可能存在无数个均衡解。

### 10.4.3 应用 Blanchard-Kahn 算法实例

以随机 Ramsey 增长模型(10.4)为例,  $z_t \equiv \tilde{A}_t, x_t^j \equiv \tilde{C}_t, x_t^s \equiv \tilde{K}_t$ ,

$$\begin{bmatrix} 1 & 0 \\ (1-\alpha) \cdot [1 - \beta \cdot (1 - \delta)] & \sigma \end{bmatrix} E \begin{bmatrix} \tilde{K}_{t+1} \\ \tilde{C}_{t+1} \end{bmatrix} = \begin{bmatrix} \frac{1}{\beta} & -\frac{1-\beta \cdot [1-\delta \cdot (1-\alpha)]}{\alpha \cdot \beta} \\ 0 & \sigma \end{bmatrix} \begin{bmatrix} \tilde{K}_t \\ \tilde{C}_t \end{bmatrix} + \begin{bmatrix} \frac{1-\beta \cdot (1-\delta)}{\alpha \cdot \beta} \\ \rho \cdot [1 - \beta \cdot (1 - \delta)] \end{bmatrix} \tilde{A}_t$$

在 Matlab 中, 首先定义系数  $\beta, \alpha, \sigma, \delta, \rho$  的值。

```

1 clear;
2
3 %参数设定
4 beta = 0.9;
5 alpha = 0.75;
6 sigma = 1;
7 delta = 0.3;
8 rho = 0.95;
```

进而计算稳定状态。

```

1 %稳态值
2 kbar = ((1-(1-delta)*beta)/(alpha * beta))^(1/(alpha - 1));
3 cbar = kbar^alpha - delta * kbar;
4 ybar = kbar^alpha;
```

根据测算结果,  $\bar{K} = 11.0766, \bar{C} = 2.7486, \bar{Y} = 6.0716$ 。

输入矩阵  $\mathcal{T}_0, \mathcal{T}_1, \mathcal{T}_2$ :

```

1 %定义矩阵 T0, T1, PSI
2 Tau0=zeros(2,2);
3 Tau1=zeros(2,2);
4 Psi=zeros(2,1);
5 Tau0(1,1)=1;
6 Tau0(1,2)=0;
7 Tau0(2,1)=(1-alpha) * (1-beta * (1-delta));
8 Tau0(2,2)=sigma;
9 Tau1(1,1)=1/beta;
10 Tau1(1,2)=-(1-beta * (1-delta * (1-alpha))) / (alpha * beta);
11 Tau1(2,1)=0;
12 Tau1(2,2)=sigma;
13 Psi(1,1)=(1-beta * (1-delta)) / (alpha * beta);
14 Psi(2,1)=rho * (1-beta * (1-delta));
```

Matlab 测算出的矩阵值如下:

$$\mathcal{T}_0 = \begin{bmatrix} 1.0000 & 0.0000 \\ 0.0925 & 1.0000 \end{bmatrix}, \quad \mathcal{T}_1 = \begin{bmatrix} 1.1111 & -0.2481 \\ 0.0000 & 1.0000 \end{bmatrix}, \quad \Psi = \begin{bmatrix} 0.5481 \\ 0.3515 \end{bmatrix}.$$

计算  $A$  和  $B$ :

```

1 %计算矩阵A,B
2 A=zeros(2,2);
3 B=zeros(2,2);
4 A = inv(Tau0) * Tau1;
5 B = inv(Tau0) * Psi;

```

测算结果：

$$A = \begin{bmatrix} 1.1111 & -0.2481 \\ -0.1028 & 1.0230 \end{bmatrix}, \quad B = \begin{bmatrix} 0.5481 \\ 0.3008 \end{bmatrix}.$$

测算  $A$  的特征值和特征向量：

```

1 %对A做 Jordan decomposition,
2 %分解为特征值和特征向量
3 %对A做 Jordan decomposition,
4 %分解为特征值和特征向量
5 [VE, Lambda] = eig(A); %MU储存特征值
6 P=inv(VE); %P储存normalized特征向量

```

特征值和特征向量如下：

$$P = \begin{bmatrix} 0.7049 & -0.8340 \\ 0.4805 & 0.9806 \end{bmatrix}, \quad \Lambda = \begin{bmatrix} 1.2327 & 0 \\ 0 & 0.9014 \end{bmatrix},$$

不难看出， $\lambda_i, i = (1, 2)$  分别是不稳定和稳定的特征根，满足 Blanchard-Kahn 条件，系统是鞍点稳定的。

将特征根矩阵沿着对角线元素（特征根）从低到高的顺序排列。

to be finished...

## 10.5 未定系数法

根据 Christiano (2002)，假定经济模型以这样的状态——空间形式表现

$$\alpha_0 E x_{t+1} + \alpha_1 \cdot x_t + \alpha_2 \cdot x_{t-1} + \beta \cdot z_t = 0, \quad t \geq 0, \quad (10.13)$$

$$z_t = R \cdot z_{t-1} + \varepsilon_t,$$

其中

- $x_t \in \mathbb{R}^n$   $\Rightarrow$  在  $t$  时间决定的内生变量向量， $x_{-1}$  是提前给定的。
- $z_t \in \mathbb{R}^k$   $\Rightarrow$  外生技术冲击变量的向量，满足  $\varepsilon_t \sim i.i.d.(0, \Sigma)$ 。
- 系数矩阵  $\alpha_0 \in \mathbb{R}^{n \times n}$ ,  $\alpha_1 \in \mathbb{R}^{n \times n}$ ,  $\alpha_2 \in \mathbb{R}^{n \times n}$ ,  $\beta \in \mathbb{R}^{n \times k}$ ,  $R \in \mathbb{R}^{k \times k}$ 。

与第10.4节的 Blanchard-Kahn 法相比，未定系数法不再将内生变量做状态变量和跳跃变量的区分。好处是让算法（看起来）简化，但这种简化有一定的额外成本：需要预设全部状态变量的初始值  $x_{t=0}$ （与之相比，Blanchard-Kahn 算法则只需要  $x_{t=0}^s$ ）：如果初始时间的经济系统恰好完全处于稳定状态，这是没问题的；否则便只能针对每一个均衡方程，分别设定其对应变量的初始条件。

(10.13)描述了这样一个经济系统，系统解表现为一个反馈机制：当前内生向量  $x_t$  与上期  $x_{t-1}$  和当期外生冲击  $z_t$  线性相关，因此假定下式

$$x_t = \underset{n \times n}{A} \cdot x_{t-1} + \underset{n \times k}{B} \cdot z_t, \quad (10.14)$$

在静态均衡条件下我们有  $z_t \equiv 0 \quad \forall t$ ，这要求系数矩阵  $A$  的所有特征值绝对值都小于 1<sup>5</sup>。现在目标变成了，对于给定的初始值  $x_{-1}$ ，找到系数矩阵  $A, B$ ，使 (10.14) 与 (10.13) 一致。

### 10.5.1 模型范例

第10.3节经济系统可改写为(10.13)形式，其中

$$\begin{aligned} x_t &= \begin{bmatrix} \tilde{K}_{t+1} \\ \tilde{C}_{t+1} \end{bmatrix}, \quad z_t = \tilde{A}_t, \quad R = \rho, \\ \alpha_0 &= \begin{bmatrix} 0 & 0 \\ 0 & \sigma \end{bmatrix}, \quad \alpha_1 = \begin{bmatrix} -1 & -\frac{1-\beta \cdot [1-\delta \cdot (1-\alpha)]}{\alpha \cdot \beta} \\ -(1-\alpha) \cdot [1-\beta \cdot (1-\delta)] & \delta \end{bmatrix}, \quad \alpha_2 = \begin{bmatrix} \frac{1}{\beta} & 0 \\ 0 & 0 \end{bmatrix}, \quad \beta = \begin{bmatrix} \frac{1-\beta \cdot (1-\delta)}{\alpha \cdot \beta} \\ \rho \cdot [1-\beta \cdot (1-\delta)] \end{bmatrix}. \end{aligned} \quad (10.15)$$

### 10.5.2 未定系数法求解

由(10.14)得

$$x_{t+1} = Ax_t + RBz_t = A(Ax_{t-1} + Bz_t) + RBz_t = A^2x_{t-1} + B(R + A)z_t.$$

带回(10.14)，用  $x_{t-1}$  替代  $x_{t+1}$  和  $x_t$

$$\underbrace{(\alpha_0 A^2 + \alpha_1 A + \alpha_2)}_{\equiv \mathcal{A}} x_{t-1} + \underbrace{[\alpha_0 B(R + A) + \alpha_1 B + \beta]}_{\equiv \mathcal{B}} z_t = 0. \quad (10.16)$$

Deterministic 状态下， $Ez_t = 0, \forall t$ 。(10.16)成立需要满足  $x_{t-1} = 0$  或  $\mathcal{A} = 0$ 。在现实经济世界中， $x_{t-1} = 0$  并无研究必要，因此需要满足  $\mathcal{A} = 0$ 。换句话说，可以通过下式求得未定系数  $A$  的值

$$\alpha_0 A^2 + \alpha_1 \cdot A + \alpha_2 = 0. \quad (10.17)$$

在 stochastic 状态下，(10.16)成立还需要  $\mathcal{B} = 0$ 。因此基于得到的  $A$  值，可通过下式求得  $B$

$$\alpha_0 B(R + A) + \alpha_1 B + \beta = 0. \quad (10.18)$$

因此，问题的关键就成了如何通过二项式(10.17)求解  $A$ 。或者更进一步：

1. 存在性：是否存在  $A$  的解，以及如果存在的话，有几个，见第10.5.3节。
2. 唯一性：如果存在多个解，哪一个满足静态均衡约束条件，即全部特征值的绝对值均  $< 1$ ，见第10.5.4节。

<sup>5</sup>回忆一下 Blanchard-Kahn 算法中将内生变量分为状态和跳跃两部分；对应相同数量的系数矩阵稳定根和不稳定根。未定系数法中所有内生变量  $x_t$  都是在  $t$  期决定的，这使得我们不再有多余的自由度用于处理不稳定根。

### 10.5.3 求解系数矩阵 A, B

#### 求解系数矩阵 A

将 deterministic 状态下的(10.13)改写为如下 AR(1) 过程。已知

$$\begin{aligned}\alpha_0 x_{t+1} + \alpha_1 x_t + \alpha_2 x_{t-1} &= 0, \\ x_t - x_t &= 0,\end{aligned}$$

改写为矩阵形式

$$\begin{aligned}\mathcal{T}_0 Y_{t+1} + \mathcal{T}_1 Y_t &= 0, \quad \forall t \geq 0, \\ Y_t &\equiv \begin{bmatrix} x_t \\ x_{t-1} \end{bmatrix}, \quad \mathcal{T}_0 = \begin{bmatrix} \alpha_0 & 0_{n \times n} \\ 0_{n \times n} & I_n \end{bmatrix}, \quad \mathcal{T}_1 = \begin{bmatrix} \alpha_1 & \alpha_2 \\ -I_n & 0_{n \times n} \end{bmatrix},\end{aligned}\tag{10.19}$$

其中值得注意的是,  $Y_0$  由  $n$  个初始状态  $x_{-1}$  所决定。

做两个假定。首先假定  $\mathcal{T}_0$  可逆, 因此  $\alpha_0$  也是可逆矩阵, 由此(10.19)改写为

$$Y_{t+1} = -\mathcal{T}_0 \mathcal{T}_1 Y_t, \tag{10.20}$$

其次, 假定  $(-\mathcal{T}_0 \mathcal{T}_1)$  有  $2n$  个线性不相关的特征向量。

基于这两个假定, 我们可以采取特征值——特征向量分解方法:

$$-\mathcal{T}_0 \mathcal{T}_1 = \frac{P}{(2n \times 2n)} \frac{\Lambda}{(2n \times 2n)} \frac{P^{-1}}{(2n \times 2n)}, \tag{10.21}$$

其中对角矩阵  $\Lambda$  的对角元素为  $-\mathcal{T}_0 \mathcal{T}_1$  的特征值,  $P$  是对应的特征向量。

假设  $-\mathcal{T}_0 \mathcal{T}_1$  有  $\bar{n}$  个稳定特征值, 构分成块对角矩阵  $\Lambda_s$ , 余下的  $2n - \bar{n}$  个不稳定特征值构成分块对角矩阵  $\Lambda_e$ 。重新排列  $\Lambda$ :

$$\Lambda_{2n \times 2n} = \begin{bmatrix} \Lambda_s & 0_{\bar{n} \times (2n - \bar{n})} \\ 0_{(2n - \bar{n}) \times \bar{n}} & \Lambda_e \end{bmatrix}.$$

对于(10.20)- (10.21), 定义  $W_t \equiv P^{-1} Y_t$ , 我们有

$$P^{-1} Y_t = \Lambda P^{-1} Y_{t-1} \Leftrightarrow W_t = \Lambda W_{t-1}. \tag{10.22}$$

对(10.22)做 backward-looking 迭代

$$\begin{aligned}W_t &= \Lambda W_{t-1} = \Lambda(\Lambda W_{t-2}) = \dots = \Lambda^t W_{t_0} \\ &= \begin{bmatrix} \Lambda_s^t & 0 \\ 0 & \Lambda_e^t \end{bmatrix} W_{t_0} = \begin{bmatrix} \Lambda_s^t & 0 \\ 0 & \Lambda_e^t \end{bmatrix} \begin{bmatrix} W_{1,t_0} \\ W_{2,t_0} \end{bmatrix}_{(2n - \bar{n}) \times n}.\end{aligned}\tag{10.23}$$

根据定义,  $\{x_t\}_{t=0}^\infty$  是平稳过程  $\Rightarrow Y_t$  是  $x_t$  的线性方程,  $\{Y_t\}_{t=0}^\infty$  是平稳过程  $\Rightarrow W_t$  是  $Y_t$  的线性方程,  $\{W_t\}_{t=0}^\infty$  是平稳过程。由于  $\lim_{t \rightarrow \infty} \Lambda_s^t \approx 0$ ,  $W_t$  的前  $\bar{n}$  行一定是平稳的, 初始  $W_{1,t_0}$  可以取任意值; 由于  $\lim_{t \rightarrow \infty} \Lambda_e^t \approx \infty$ ,  $W_t$  的后  $2n - \bar{n}$  行是不平稳的, 式(10.23)成立便要求初始  $W_{2,t_0} = 0$ 。

如果假定  $\bar{n} \equiv n$ , 即(10.19)中系数矩阵的稳定特征值的数量等于经济系统(10.13)中内生变量的数量 (对于  $\bar{n} \neq n$  情况的讨论见第10.5.4节。), 我们有

$$\begin{bmatrix} W_{1,t_0} \\ W_{2,t_0} \end{bmatrix}_{n \times n} = \begin{bmatrix} (P^{-1})_{11} & (P^{-1})_{12} \\ (P^{-1})_{21} & (P^{-1})_{22} \end{bmatrix} \begin{bmatrix} x_0 \\ x_{-1} \end{bmatrix}_{n \times n},$$

其中要求  $W_{2,t_0} = 0$ , 即

$$(P^{-1})_{21} x_0 + (P^{-1})_{22} x_{-1} = 0.$$

对于可逆矩阵  $(P^{-1})_{21}$ , 上式改写为

$$x_0 = -((P^{-1})_{21})^{-1} (P^{-1})_{22} x_{-1}, \quad (10.24)$$

结合 deterministic 状态的(10.14)与(10.24), 可得系数矩阵  $A$

$$A = -((P^{-1})_{21})^{-1} (P^{-1})_{22}. \quad (10.25)$$

### 求解系数矩阵 $B$

将求得的系数矩阵  $A$  代入(10.18)。对于  $n \times k$  的矩阵  $B$  和  $\beta$ , 等式两侧向量化, 我们有

$$\begin{aligned} 0 &= \text{vec}((\alpha_0 A + \alpha_1) B + \alpha_0 B R) + \text{vec}(\beta) \\ &= \text{vec}((\alpha_0 A + \alpha_1) B) + \text{vec}(\alpha_0 B R) + \text{vec}(\beta) \\ &= [I_k \otimes (\alpha_0 A + \alpha_1)] \text{vec}(B) + (R^T \alpha_0) \text{vec}(B) + \text{vec}(\beta) \\ &= [I_k \otimes (\alpha_0 A + \alpha_1) + R^T \alpha_0] \text{vec}(B) + \text{vec}(\beta), \end{aligned}$$

其中  $\text{vec}$  表示对矩阵向量化;  $\otimes$  表示 Kronecker 乘<sup>6</sup>。进而我们有

$$\text{vec}(B) = -[I_k \otimes (\alpha_0 A + \alpha_1) + R' \otimes \alpha_0]^{-1} \text{vec}(\beta), \quad (10.26)$$

#### 10.5.4 存在性及唯一性的探讨

如前文所述,  $\bar{n} = n$  时, 经济系统存在唯一均衡解。我们需要做的是通过对初始状态  $x_{t_0}$  的选择, 使得在  $W_{t_0} = \begin{bmatrix} W_{1,t_0}^T, & W_{2,t_0}^T \\ \bar{n} \times n & (2n-\bar{n}) \times n \end{bmatrix}^T$  当中 (对应  $W_{2,t_0}$  的) 后  $2n - \bar{n}$  行等于 0。

当  $\bar{n} < n$  时,  $2n - \bar{n} > n$ , 我们需要使后  $2n - \bar{n}$  行等于 0, 但我们只有  $n$  个自由变量。此时均衡解不存在。

当  $\bar{n} > n$  时,  $2n - \bar{n} < n$ , 我们需要使后  $2n - \bar{n}$  行等于 0, 对应  $n$  个自由变量。这导致存在很多组  $x_0$  的值可以带来均衡, 均衡解不唯一。

## 10.6 期望误差法

根据Sims (2002), 假定经济模型以这样的状态——空间形式展现

$$\underset{(n \times n)(n \times 1)}{\mathcal{T}_0} x_t = \underset{(n \times n)}{\mathcal{T}_1} x_{t-1} + \underset{(n \times k)(k \times 1)}{\Psi} u_t + \underset{(n \times r)(r \times 1)}{\Pi} \eta_t, \quad t \geq 0, \quad (10.27)$$

其中

---

<sup>6</sup>对于矩阵  $A \in \mathbb{R}^{k \times l}$ ,  $B \in \mathbb{R}^{l \times m}$ ,  $C \in \mathbb{R}^{m \times n}$ , 我们有矩阵向量化

$$\text{vec}(A) = [A_{1,1}, \dots, A_{k,1}, A_{1,2}, \dots, A_{k,2}, \dots, A_{1,l}, \dots, A_{k,l}]^T,$$

以及以下向量化运算

$$\text{vec}(A + B) = \text{vec}(A) + \text{vec}(B),$$

$$\text{vec}(AB) = (I_m \otimes A) \text{vec}(B) = (B^T \otimes I_k) \text{vec}(A),$$

$$\text{vec}(ABC) = (I_n \otimes AB) \text{vec}(C) = (C^T \otimes A) \text{vec}(B) = (C^T B^T \otimes I_k) \text{vec}(A).$$

- 向量  $x_t$  表示内生变量，对应系数矩阵  $\mathcal{T}_0, \mathcal{T}_1$ ； $x_{-1}$  是给定的。在期望误差法中， $x_t$  包含一部分在  $t$  期对  $t+1$  期的期望值。
- $u_t$  表示外生随机冲击过程，假定  $u_t \sim i.i.d.(0, \Sigma)$ ，对应系数矩阵  $\Psi$ ，
- $\eta_t = x_t - E_{t-1}x_t$  是期望误差向量，反映  $t$  期实际状态  $x_t$  与  $t-1$  期对  $t$  期状态的期望的偏差，满足  $E_t\eta_{t+1} = 0$ ，对应系数矩阵  $\Pi$ 。

### 10.6.1 模型范例

第10.3节经济系统可改写为(10.27)形式，其中

$$\begin{aligned} x_t &= \begin{bmatrix} E_t \tilde{K}_{t+1} \\ \tilde{C}_t \\ E_t \tilde{C}_{t+1} \\ \tilde{A}_t \end{bmatrix}, \quad u_t = \varepsilon_t, \quad \eta_t = \tilde{C}_t - E_{t-1} \tilde{C}_t, \\ \mathcal{T}_0 &= \begin{bmatrix} 1 & \frac{1-\beta \cdot [1-\delta \cdot (1-\alpha)]}{\alpha \cdot \beta} & 0 & -\frac{1-\beta \cdot (1-\delta)}{\alpha \cdot \beta} \\ (\alpha-1) \cdot [1-\beta \cdot (1-\delta)] & \delta & -\delta & \rho \cdot [1-\beta \cdot (1-\delta)] \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad (10.28) \\ \mathcal{T}_1 &= \begin{bmatrix} \frac{1}{\beta} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \Psi = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \Pi = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}. \end{aligned}$$

### 10.6.2 期望误差法求解

假定  $\mathcal{T}_0$  是可逆矩阵，(10.27)可以改写为

$$\begin{aligned} x_t &= \mathcal{T}_0^{-1} \mathcal{T}_1 x_{t-1} + \mathcal{T}_0^{-1} (\Psi u_t + \Pi \eta_t) \\ &= Ax_{t-1} + \mathcal{T}_0^{-1} (\Psi u_t + \Pi \eta_t), \quad (10.29) \\ \text{其中 } A &\equiv \mathcal{T}_0^{-1} \mathcal{T}_1, x_0 \text{ 是给定的, 并且 } t \geq 1. \end{aligned}$$

进一步假定  $A$  的所有特征向量都是线性不相关的，由此我们可以对  $A$  做特征值——特征向量分解

$$A = \underset{n \times n}{P} \Lambda P^{-1},$$

代回(10.29)，调整得

$$\begin{aligned} P^{-1} x_t &= P^{-1} (P \Lambda P^{-1}) x_{t-1} + P^{-1} \mathcal{T}_0^{-1} (\Psi u_t + \Pi \eta_t), \text{进而} \\ w_t &= \Lambda w_{t-1} = Q (\Psi u_t + \Pi \eta_t), \text{其中} \quad (10.30) \\ w_t &\equiv P^{-1} x_t, \quad Q \equiv P^{-1} \mathcal{T}_0^{-1}. \end{aligned}$$

类似地，将对角矩阵  $\Lambda$  按特征值从小到大顺序重新排列

$$\Lambda \underset{n \times n}{=} \begin{bmatrix} \Lambda_s & 0 \\ \bar{n} \times \bar{n} & \bar{n} \times (n-\bar{n}) \\ 0 & \Lambda_e \\ (n-\bar{n}) \times \bar{n} & (n-\bar{n}) \times (n-\bar{n}) \end{bmatrix}, \quad (10.31)$$

其中  $\Lambda_s$  为绝对值小于 1 的特征值，设为  $\bar{n} < n$  个。

根据(10.31)，经济系统(10.30)可以改写为

$$\begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix} = \begin{bmatrix} \Lambda_s & 0 \\ \bar{n} \times \bar{n} & \bar{n} \times (n - \bar{n}) \\ 0 & \Lambda_e \\ (n - \bar{n}) \times \bar{n} & (n - \bar{n}) \times (n - \bar{n}) \end{bmatrix} \begin{bmatrix} w_{1,t-1} \\ w_{2,t-1} \end{bmatrix} + \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} (\Psi u_t + \Pi \eta_t), \quad t \geq 1. \quad (10.32)$$

上半部分为稳定分块，下半部分为不稳定分块。我们先从不稳定分块开始求解。

### 不稳定分块求解

提取不稳定分块

$$w_{2,t} = \Lambda_e w_{2,t-1} + Q_2 (\Psi u_t + \Pi \eta_t),$$

调整为前向形式

$$w_{2,t} = \Lambda_e^{-1} w_{2,t+1} - \Lambda_e^{-1} Q_2 (\Psi u_{t+1} + \Pi \eta_{t+1}), \quad (10.33)$$

进一步调整为前向迭代形式

$$w_{2,t} = \underbrace{\lim_{T \rightarrow \infty} \Lambda_e^{-T} w_{2,t+1}}_{\mathcal{A}} - \underbrace{\sum_{s=1}^T \Lambda_e^{-s} Q_2 (\Psi u_{t+s} + \Pi \eta_{t+s})}_{\mathcal{B}}, \quad (10.34)$$

其中等式右侧

- $\mathcal{A} = 0$ 。这是由于首先分块矩阵  $\Lambda_e$  对应的所有元素，即特征值的绝对值都大于 1， $\Lambda_e^\infty \rightarrow 0$ ，其次静态均衡  $E_t(w_{2,t+\infty})$  是有界的。
- $\mathcal{B} = 0$ 。这是由于首先  $u_t$  是一个均值为 0 的随机过程，满足  $E(u_t) = 0$ ，其次在理性期望条件下， $\eta_t$  的条件均值为 0。

基于上述分析，我们有不稳定分块的值

$$w_{2,t} = 0. \quad (10.35)$$

### 稳定分块求解

提取稳定分块

$$w_{1,t} = \underbrace{\Lambda_s w_{1,t-1}}_{\bar{n} \times \bar{n}} + \underbrace{Q_1 \left( \begin{smallmatrix} \Psi & u_t \\ (\bar{n} \times k)(k \times 1) & (n \times r)(r \times 1) \end{smallmatrix} \right)}_{\bar{n} \times n} \eta_t.$$

为了得到  $w_{1,t}$  的值，我们首先需要替代期望误差  $\eta_t$ ，随后求解稳定分块。

### 期望误差求解

假定  $k = r$ (对于  $k \neq r$  的讨论，见第10.6.3节。)。基于(10.34)-(10.35)我们有

$$Q_2 (\Psi u_t + \Pi \eta_t) = 0, \quad (10.36)$$

这意味着期望误差  $\eta_t$  随着同期外部冲击  $u_t$  的变化而反向变化，所谓“理性预期”。如果假定  $Q_2 \Pi$  是可逆矩阵，那么可以将  $\eta_t$  写为关于 (并且只是关于)  $u_t$  的函数

$$\eta_t = -(Q_2 \Pi)^{-1} Q_2 \Psi u_t. \quad (10.37)$$

### 稳定分块求解

$\eta_t$  的决定式(10.37)代回稳定分块决定式可得

$$\begin{aligned} w_{1,t} &= \Lambda_s w_{1,t-1} + Q_1 (\Psi u_t + \Pi \eta_t) \\ &= \Lambda_s w_{1,t-1} + Q_1 (\Psi - \Pi (Q_2 \Pi)^{-1} Q_2 \Psi) u_t. \end{aligned} \quad (10.38)$$

### 内生变量向量 $x_t$ 求解

联立两个分块  $w_{1,t}, w_{2,t}$  的解(10.38)与(10.35)，经济系统(10.32)可以进一步改写为

$$\begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix} = \begin{bmatrix} \Lambda_s & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} w_{1,t-1} \\ w_{2,t-1} \end{bmatrix} + \begin{bmatrix} Q_1 (\Psi - \Pi (Q_2 \Pi)^{-1} Q_2 \Psi) \\ 0 \end{bmatrix} u_t.$$

根据定义式(10.30)用  $x_t$  替代上式中的  $w_t$ ，可得  $x_t$  的解

$$x_t = P \begin{bmatrix} \Lambda_s & 0 \\ 0 & 0 \end{bmatrix} P^{-1} x_{t-1} + P \begin{bmatrix} Q_1 (\Psi - \Pi (Q_2 \Pi)^{-1} Q_2 \Psi) \\ 0 \end{bmatrix} u_t. \quad (10.39)$$

### 10.6.3 存在性及唯一性的探讨

如前文所述，经济系统存在均衡解的条件由(10.36)给出：根据该方程，期望误差  $\eta_t$  根据出现的外生冲击  $u_t$  而灵活调整。如果经济系统中对  $\eta_t$  的限定条件太多——导致  $\eta_t$  无法灵活调整，均衡解可能不存在。

如果  $r < k$ ，即对于  $\eta_t$  向  $u_t$  的调整存在过多的限定条件，这可能会使得系统解不存在。换句话说，经济系统有解的充分必要条件是： $r \geq k$ ，即  $Q_2 \Psi$  的列空间包含在  $Q_2 \Pi$  的列空间内。

如果  $r > k$ ，即对于  $\eta_t$  向  $u_t$  的调整存在过多的限定条件，这可能会使得存在多个  $\eta_t$  同时满足式(10.37)，对应的只是唯一的  $Q_2 \Pi \eta_t$ ，而非唯一的  $\eta_t$ 。多重  $\eta_t$  值使得我们有多个  $Q_1 \Pi \eta_t$ ，对应(10.38)中多个稳定分块矩阵  $w_t$ (进而  $x_t$ )的解。为了使系统存在唯一解，需要使  $Q_1 \Pi$  的行空间包含在  $Q_2 \Pi$  内，即存在这样一个矩阵  $\Phi$  满足  $Q_1 \Pi = \Phi_2 \Pi$ 。

## 10.7 参数化期望法

参数化期望法 (PEA) 最早由Den Haan and Marcet (1990) 提出。PEA 所依赖的渐进式趋同结果，由Marcet and Sargent (1989); Marcet and Marshall (1994) 等人所讨论。算法方面，Den Haan and Marcet (1994) 提出精确度测试(见第15.3节)；Christiano and Fisher (2000) 等人讨论了网格算法。基于Den Haan and Marcet (1990)，本节介绍 PEA 算法的基本思路和一个简单范例。

大多数动态经济学模型建立在一个或一组 Euler 方程的基础上。这个(些)Euler 方程将一组当期变量和一组未来期变量的条件期望联系在一起，如

$$f(x_t) = E_t h(x_{t+1}, u_{t+1}, z_t), \quad (10.40)$$

其中  $\{x_t, u_t, z_t\}$  分别代表当前  $t$  期的内生变量、状态变量和随机扰动。 $E_t$  表示在当前  $t$  期，对未来时间期比如  $t+1$  状态的条件期望。对于经验模型来说，研究的重点在于选取合适的显函数形式来描述  $f()$  以及  $h()$ ，对于非线性系统来说，难点在于如何对之作线性近似，以求得系统的解。

一个常见的求解思路是后向求解法。简单来说，根据这个思路，如果(10.40)可以改写为

$$x_t = f^{-1}(E_t h(x_{t+1}, u_{t+1}, z_t)),$$

那么在我们知道方程  $f^{-1}$  近似形式的情况下，一旦我们得到未来期的条件期望  $E_t h(x_{t+1}, u_{t+1}, z_t)$ ，便可以后向求解当前期的  $x_t$ 。换句话说，一旦我们知道了  $x_{t+1}$  的值，便可以后向求得  $x_t$ 。

与之相反，PEA 法的求解思路是前向的，致力于用已知的  $x_t$  求解  $x_{t+1}$ ：即便  $E_t h(x_{t+1}, u_{t+1}, z_t)$  是一个关于  $x_{t+1}$  的函数，可是从定义来看，它仍然是由经济个体在  $t$  时刻的决策  $x_t$  所决定的。因此，PEA 用前向递归的方式，基于当前期已知信息，求解未来期的经济系统，将(10.40)改写为

$$E_t h(x_{t+1}, u_{t+1}, z_t) = m(\beta, Z_t^1) \quad (10.41)$$

其中  $Z_t^1$  是  $t$  期状态变量的集合。换句话说，PEA 假定未来期的期望值是关于当前期已知信息集的函数，当前期已知信息集包括前定变量和当期冲击等。根据这一假定，对未来的期望得以“参数化”。

PEA 法成功与否的关键有两点：

- 能提供多大的精确程度，基于当期信息集，用  $m()$  来近似参数化期望  $E_t h()$ 。 $\Rightarrow$  Den Haan and Marcer (1990) 假定  $m()$  是多项式形式，多项式  $m()$  近似  $E_t h()$  的精确度？
- 参数向量  $\beta$  的精确度。提高多项式的非线性程度有助于提升近似精度，然而进一步的精确度提升更离不开参数向量  $\beta$  的估计。

### 10.7.1 参数化期望法示例

假定一个典型的随机增长模型，经济个体需要做效用最大化决策

$$\max_{c_t} E_t \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\sigma}}{1-\sigma},$$

约束条件

$$c_t + k_t - \mu k_{t-1} = A_t k_{t-1}^{\alpha}, \quad (10.42)$$

其中

$$\ln A_{t+1} = \rho \ln A_t + \varepsilon_t$$

表示随机技术冲击， $\beta$  表示时间贴现， $1 - \mu$  表示折旧率， $\sigma$  表示风险规避系数。FOC  $\Rightarrow$

$$MU_{c_t} = \beta E_t R_{t+1} MU_{c_{t+1}},$$

其中  $MU_{c_t}$  表示  $t$  期消费带来的边际效用。上式也可表示为如下 Euler 方程形式

$$c_t^{-\sigma} = \beta E_t [c_{t+1}^{-\sigma} (\alpha A_{t+1} k_t^{\alpha-1} + \mu)]. \quad (10.43)$$

根据 PEA 法，将右侧的期望值改写为一个与  $t$  期状态变量有关的多项式方程  $m()$ ，状态变量包括  $t-1$  期前定变量  $k_{t-1}$  和当期技术冲击  $A_t$ ，上式因而改写为

$$c_t^{-\sigma} = \beta m(k_{t-1}, A_t, \delta), \quad (10.44)$$

对应系数变量  $\delta$ 。

假定多项式  $m(\cdot)$  以一阶显函数形式表现如下<sup>7</sup>

$$m(k_{t-1}, A_t; \delta) = \delta_1 k_{t-1}^{\delta_2} A_t^{\delta_3}, \quad (10.45)$$

<sup>7</sup>也可以用更高阶显函数的形式设定  $m()$ ，如

$$m(k_{t-1}, A_t; \delta) = \delta_1 k_{t-1}^{\delta_2} A_t^{\delta_3} (k_{t-1} A_t)^{\delta_4} (k_{t-1}^2)^{\delta_5} (A_t^2)^{\delta_6}.$$

随后，为向量  $\delta$  赋一组初始值，基于初始值，PEA 进行迭代，直到逼近“真实的”系数值位置。

在给定以下三组信息

- 初始资本存量  $k_{t=0}$ ,
- 在利用计算机生成的一组随机游走时间序列中挑选  $A_{t=1}$ ,
- 初始赋值  $\delta_1, \delta_2, \delta_3$

之后，我们可以根据(10.44)求得  $c_{t=1}$ 。下一步，利用预算约束式(10.42)可生成  $k_{t=1}$ 。自此我们有了经济系统在第 1 期  $t = 1$  的全部信息  $\{c_{t=1}, k_{t=1}, A_{t=1}\}$ 。

对于  $t = 2$ ，重复上述过程

- 在利用计算机生成的同一组随机游走时间序列中挑选  $A_{t=2}$ ,
- 根据(10.44)求得  $c_{t=2}$ ,
- 根据(10.42)求得  $k_{t=2}$ ,

以此类推，直到第  $t = n$  期。

对于给定的  $\{\delta\}$  和技术冲击时间序列  $A_t$ ，利用 PEA 方法生成的  $\{k_t, c_t, y_t\}_{t=1}^n$ ，并不总是与 Euler 描述的经济个体期望相一致。个体基于当期信息对未来期作出期望，并采取行动如(10.44)以致力于实现这种期望，但实际导致的结果往往与当初的 Euler 方程期望(10.43)相悖。

随后的工作就变为，选取“正确”的系数  $\{\delta\}$ ，使得

$$E_t [c_{t+1}^{-\sigma} (\alpha A_{t+1} k_t^\alpha + u)] - \delta_1 k_{t-1}^{\delta_2} A_t^{\delta_3} = \nu_t \approx 0, \quad (10.46)$$

参数选取的标准：选取合适的  $\{\delta\}$  组合，使得生成的经济时间序列数据中，对应  $\min \sum \nu_t^2$ 。这就涉及到非线性最小二乘法 (NLS)：首先猜测一组  $\{\delta_1, \delta_2, \delta_3\}$ ，利用其生成一组  $\{c_t, k_t, y_t\}$ ，运行 NLS 回归测算  $\min \sum \nu_t^2$  对应的  $\{\delta'_1, \delta'_2, \delta'_3\}$ 。 $\delta$  和  $\delta'$  的差异反映了基于对未来期期望的行为，在未来期未必会产生如当初所设想的结果。这就需要修正我们的预测模型，主要通过调整系数  $\delta$  来实现：将  $\{\delta'_1, \delta'_2, \delta'_3\}$  替代原有的  $\{\delta_1, \delta_2, \delta_3\}$ ，生成新的一组时间序列数据，NLS 测算  $\min \sum \nu_t^2$  对应的  $\{\delta''_1, \delta''_2, \delta''_3\}$ ，重复上述过程，直到  $\sum \nu_t^2 \approx 0$ ，此时  $\delta$  达到不动点，从经济意义上来说，消费者通过预测方程产生当期消费需求，当期消费需求又引出未来期消费需求，这一组消费需求都证明了最初的预测方程是正确的。

可以从学习算法 (learning algorithm) 的角度来理解 PEA。在初始阶段，经济系统中的行为人持有有限信息。他知道一阶条件 (Euler 方程) 的形式，但不知道如何合理展开当期行动以符合这种期望。行为人只有通过试错法，形成、并逐次调整自己的行动，这构成一个学习过程：起初他只掌握少量当前状态变量的信息，只能凭直觉猜测作出行动，并对未来做初步的预测。随着信息掌握的越来越丰富，他得意不断修正作出的直觉猜测，使得预测越来越精确，可以更多地凭经验行事，更少地依赖直觉（直觉和经验作为互补品），直到对未来的预测  $E_t [c_{t+1}^{-\sigma} (\alpha A_{t+1} k_t^\alpha + u)]$  与实际情况  $c_{t+1}^{-\sigma} (\alpha A_{t+1} k_t^\alpha + u)$  相吻合。

## 10.8 Schur 分解法

上文提到了一系列求解非线性系统的算法，如特征值分解法、未定系数法、期望误差法等。它们的展开都依赖于这样的假设条件： $\mathcal{T}_0$  是非奇异矩阵 (可逆的)，进而  $\mathcal{T}_1$  是可对角化的矩阵。该假设有助于将经济系统转化为更易于处理的标量方程组的形式。

然而对于更为一般的形式， $\mathcal{T}_0$  是奇异矩阵的情况，我们无法将经济系统对角化。此时就需要介绍 Schur 分解法（又称 QZ 分解法，见第 11.A 节），将经济系统三角化。本节简要介绍 Schur 分解法在求解非线性经济系统中的应用。

### 10.8.1 Schur 分解的定义

对于这样一个非线性经济系统

$$\mathcal{T}_0 x_{t+1} = \mathcal{T}_1 x_t, \quad t \geq 0, \quad x_t \in \mathbb{R}^{n \times n}, \quad (10.47)$$

其中  $x_{t=0}$  为给定的初始状态。假定存在

- 两个酉矩阵  $Q, Z$  (unitary matrix, 即满足  $QQ^T = ZZ^T = I$ )，和
- 两个上三角矩阵  $\Omega, \Lambda$ ，分别对应对角元素  $\omega_{ii}, \lambda_{ii}$ ，

使得满足

$$\mathcal{T}_0 = Q^T \Omega Z, \quad \mathcal{T}_1 = Q^T \Lambda Z, \quad (10.48)$$

我们称这种方法为 Schur 分解法，其中  $\Omega, \Lambda$  分别又称为系数矩阵  $\mathcal{T}_0, \mathcal{T}_1$  的 Schur 式 (Schur form)<sup>8</sup>。

这样一个非线性经济系统的广义特征根解集 (generalized eigenvalue set) 为  $\left(\frac{\lambda_{ii}}{\omega_{ii}}\right)$ ，分子为系数矩阵  $\mathcal{T}_1$  的特征根三角矩阵  $\Lambda$  的对角元素，分母为系数矩阵  $\mathcal{T}_0$  的特征根三角矩阵  $\Omega$  的对角元素；两组对角元素的绝对值均经过从左上到右下升序重新排列。

根据 Schur 分解法，我们可以利用系数矩阵  $\mathcal{T}_0, \mathcal{T}_1$  来测算  $Q, Z, \Omega, \Lambda$ ，并进一步求得经济系统的解。(10.47) 左乘  $Q$ ，整理得

$$\Omega Z^T x_{t+1} = \Lambda Z^T x_t \Rightarrow \Omega w_{t+1} = \Lambda w_t, \quad t \geq 0, \quad (10.49)$$

其中定义了变量

$$w_t \equiv Z^T x_t, \quad (10.50)$$

可见(10.49)是一个关于  $\{w\}_t$  的 (三角矩阵) 系统。新系统的特征根对角矩阵中，对角元素的特征根的绝对值  $\left|\frac{\lambda_{ii}}{\omega_{ii}}\right|$  从左上到右下升序排列。据此可将新系统分为两部分：1 个稳定分块矩阵  $w_{1,t}$ ，对应  $\left|\frac{\lambda_{ii}}{\omega_{ii}}\right| < 1$ ，和 1 个爆炸分块矩阵  $w_{2,t}$ ，对应  $\left|\frac{\lambda_{ii}}{\omega_{ii}}\right| > 1$ ，(10.49)因此改写为

$$\begin{bmatrix} \Omega_{11} & \Omega_{12} \\ 0 & \Omega_{22} \end{bmatrix} \begin{bmatrix} w_{1,t+1} \\ w_{2,t+1} \end{bmatrix} = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ 0 & \Lambda_{22} \end{bmatrix} \begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix}. \quad (10.51)$$

### 10.8.2 爆炸分块矩阵的解

(10.51) 中的爆炸分块可以表示为

$$\Omega_{22} w_{2,t+1} = \Lambda_{22} w_{2,t},$$

根据模型设定，上三角矩阵  $\Lambda$  的对角分块  $\Lambda_{22}$  是非奇异矩阵，因此我们有

$$w_{2,t} = \Lambda_{22}^{-1} \Omega_{22} w_{2,t+1},$$

进一步做前向迭代有

$$w_{2,t} = \lim_{T \rightarrow \infty} (\Lambda_{22}^{-1} \Omega_{22})^T w_{2,t+T}. \quad (10.52)$$

<sup>8</sup>考虑到现实经济世界的特性，我们暂不考虑复根的情况。

(10.52)中, 一方面系数矩阵  $(\Lambda_{22}^{-1}\Omega_{22})$  的特征根矩阵, 对角元素  $\left(\frac{\omega_{ii}}{\lambda_{ii}}\right)$  对应原系统广义特征根  $\left(\frac{\lambda_{ii}}{\omega_{ii}}\right)$  的倒数, 在爆炸分块中,  $\left|\frac{\omega_{ii}}{\lambda_{ii}}\right| < 1$ , 因此  $\left(\frac{\omega_{ii}}{\lambda_{ii}}\right)^T \rightarrow 0$ 。另一方面, 经济系统的稳定性要求  $\lim_{T \rightarrow \infty} (w_{w,t+T}) < \infty$ 。由此我们有, 在稳定均衡条件下, (10.52)成立需要满足条件

$$w_{2,t} = 0. \quad (10.53)$$

### 10.8.3 稳定分块矩阵的解

(10.51)中的稳定分块

$$\Omega_{11}w_{1,t+1} + \Omega_{12}w_{2,t+1} = \Lambda_{11}w_{1,t} + \Lambda_{12}w_{2,t},$$

引入爆炸分块的解(10.53)得

$$\Omega_{11}w_{1,t+1} = \Lambda_{11}w_{1,t},$$

类似地, 根据模型设定, 上三角矩阵  $\Omega$  的对角分块  $\Omega_{11}$  是非奇异矩阵, 因此我们有

$$w_{1,t} = \Omega_{11}^{-1}\Lambda_{11}w_{1,t-1},$$

进一步做后向迭代得

$$w_{1,t} = (\Omega_{11}^{-1}\Lambda_{11})^t w_{1,t=0}, \quad (10.54)$$

其中  $w_{1,t=0}$  由已知的初始状态  $x_{t=0}$  得到。

### 10.8.4 原经济系统的解

合并稳定分块和爆炸分块的解(10.54)(10.53), 可得新系统  $w_t$  的解。在此基础上, 考虑到酉矩阵  $Z$  的可逆性, 根据(10.50)可进一步测得原经济系统  $x_t$  的解:

$$x_t = Zw_t = Z \begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix} = Z \begin{bmatrix} (\Omega_{11}^{-1}\Lambda_{11})^t w_{1,t=0} \\ 0 \end{bmatrix}. \quad (10.55)$$



# 第十一章 扰动法

## 11.1 扰动法简介

对于 DSGE 模型的非线性系统，另一种求解方法是扰动法 (perturbation)，即根据隐函数定理 (implicit function theorem)，围绕经济系统的非随机稳定状态做 TSE(泰勒级数展开 Taylor Series Expansion)<sup>1</sup>。扰动法在经济学研究中的应用<sup>2</sup>，可以追溯至 Judd and Guu (1993)<sup>3</sup>，从那时起至今二十多年间，扰动法在宏观经济学研究中受到越来越多的重视，主要原因有四。

1. 解的精确度。一方面在近似点附近，扰动法的局部解的精确度常常是比较高的，另一方面扰动法生成的状态变量值也具有较好的全局解特性 (Judd, 1998; Judd and Guu, 2001; Caldara et al., 2012)，并且扰动法的解还可以较方便地交由其他求解方法做进一步处理，如价值方程的迭代，见第11.6节。
2. 解的经济意义。扰动法的近似解相对直观易懂，例如 DSGE 模型的二阶近似解中含有一个外生冲击的冲击标准差项，会对经济系统的动态路径产生影响，这呼吁经济政策制定者采取有效的预防措施。而相反地，持有确定性等价 (certainty equivalence) 假定的一阶线性近似法 (第十章)，由于无法将外生技术冲击的波动纳入到分析框架中，从而在分析福利、风险等问题上存在不足。
3. 如前所述，传统线性求解技术基本上类似于一阶线性扰动法，这这些前期研究的宝贵经验能够为我们研究高阶扰动求解法提供帮助。
4. 一批新的计算程序有助于显著提高高阶扰动法数值计算的效率，如Matlab平台下的Dynare (以及一个独立软件平台Dynare++)，Mathematica平台下的Perturbation AIM等。

## 11.2 分析框架

扰动法致力于近似求解状态——空间系统

$$\mathcal{H}(d) = \mathbf{0}, \quad (11.1)$$

其中  $d : \Omega \rightarrow \mathcal{R}^m$  表示由一系列均衡条件和预算约束条件组成的方程系统，对应  $n$  个状态变量组成的向量  $\mathbf{x}$  以及相关系数  $\theta$ 。求解的基本思路是对围绕状态变量的非随机稳定点做泰勒级数展开。

<sup>1</sup>需要指出的是，扰动法的算法有多种，这里以宏观经济学研究中得到广泛采用的方法为例。

<sup>2</sup>扰动法在物理学和自然科学中得到广泛应用，至少可以上溯到 19 世纪。随着 20 世纪上半叶量子力学的发展，扰动法更成为自然科学研究的核心方法之一，相关介绍可见 Simmonds and Mann (2013); Bender and Orszag (2013)。

<sup>3</sup>关于采用扰动法进行经济学研究的严格数学证明，如可解性等，以及更高级的求解技术如 Padé 近似等，可见 Judd (1998); Judd and Guu (2001); Jin and Judd (2002)。

以二阶泰勒级数展开为例：

$$d_i^2(\mathbf{x}, \theta) = \theta_{i,0} + \theta_{i,1}(\mathbf{x} - \mathbf{x}_0)^\top + (\mathbf{x} - \mathbf{x}_0)\theta_{i,2}(\mathbf{x} - \mathbf{x}_0)^\top, \quad i = 1, 2, \dots, m, \quad (11.2)$$

其中

- $\mathbf{x}^\top$  表示向量  $\mathbf{x}$  的转置，
- $\mathbf{x}_0$  表示非随机稳态，
- $\theta_{i,0}$  是一个矢量，
- $\theta_{i,1}$  是一个  $n$  维向量，
- $\theta_{i,2}$  是一个  $n \times n$  矩阵，
- 系数  $\theta_{i,0}, \theta_{i,1}, \theta_{i,2}$  可根据隐函数定理，由  $d$  的导数求得。

与传统线性近似法相比 ((King et al., 1988a,b, 2002), 本文第十章)，线性近似与一阶扰动法基本等同。而更高阶扰动法是将一阶扰动(近似线性)化的经济系统结构一般化到更一般的形式，引入额外项，从而使得近似式具有更高的解释效力。

**Remark 11.1** (扰动法的线性和对数线性化). 通常来说，线性化 (*linearization*)是指对状态变量(或对状态变量针对模型中的某些量做预处理后)进行线性调整。类似地，对数线性化 (*log-linearization*)描述状态变量距离其稳定状态的对数偏离程度，以某一变量  $x \in \mathbf{x}$  举例，定义  $\hat{x} = \log(x/\bar{x})$ ，其中  $\bar{x}$  是稳态值。则二阶近似

$$d_i^2(\hat{\mathbf{x}}, \theta) = \theta_{i,0} + \theta_{i,1}(\hat{\mathbf{x}} - \hat{\mathbf{x}}_0)^\top + (\hat{\mathbf{x}} - \hat{\mathbf{x}}_0)\theta_{i,2}(\hat{\mathbf{x}} - \hat{\mathbf{x}}_0)^\top, \quad i = 1, 2, \dots, m,$$

如果  $x_0$  是确定性稳态，我们有  $\hat{x}_0 = 0$ ，则对于  $\forall x \in \mathbf{x}$  上式可改写为

$$d_i^2(\hat{\mathbf{x}}, \theta) = \theta_{i,0} + \theta_{i,1}\hat{\mathbf{x}}^\top + \hat{\mathbf{x}}\theta_{i,2}\hat{\mathbf{x}}_0^\top, \quad i = 1, 2, \dots, m,$$

比起线性化来，对数线性化形式的系统解更容易解读，并且有时候更精确。

**Remark 11.2** (常规扰动和奇异扰动). 有时候我们需要区分常规扰动和奇异扰动。

常规扰动 (*regular perturbations*) 往往指，外部环境的一个小变动引发经济系统解的一个小变动，例如新凯恩斯主义模型中，一个货币政策冲击(标准差的变化) 印发经济系统均衡路径的变化，进而带来产出、通胀等内生变量的小变化 (Gali, 2008; Woodford, 2011)。DSGE 模型大多数研究常规扰动的经济现象，本论文也以常规扰动的介绍为主。

奇异扰动法 (*singular perturbations*) 往往指，外部环境的小变动引发经济系统的大波动，如市场出清价格。如不完全竞争市场模型中，一种新资产的出现可能带来经济系统解的大调整，对于研究金融市场摩擦和/或不完全竞争市场的经济学家来说，就需要予以额外关注 (Judd and Guu, 1993)。Judd (1998) 介绍了如何采用分叉法 (*bifurcation*) 求解奇异扰动的经济系统问题。

## 11.3 求解方法

### 11.3.1 状态——空间形式

在第11.2节分析框架的基础上，我们来看如何将扰动法应用到求解典型经济系统中，主要参考自Schmitt-Grohé and Uribe (2004)。

对于由一系列均衡条件和预算约束条件构成的非线性经济系统，可以表述为如下含有期望的状态——空间形式

$$E_t \mathcal{H}(y, y', x, x') = 0, \quad (11.3)$$

其中

- $x, y$  分别是  $n_x \times 1$  和  $n_y \times 1$  的状态向量和控制向量， $n = n_x + n_y$ ,  $x', y'$  表示  $t+1$  期的变量。
- 运算符  $\mathcal{H} : \mathcal{R}^{n_y} \times \mathcal{R}^{n_y} \times \mathcal{R}^{n_x} \times \mathcal{R}^{n_x} \rightarrow \mathcal{R}^n$  表示方程系统，其中部分方程含有期望项。
- 状态向量  $x$  可以分解成两块，分别表示内生状态变量和外生变量：

$$x = \begin{bmatrix} x_1 \\ (n_x - n_\varepsilon \times 1) \\ x_2 \\ (n_\varepsilon \times 1) \end{bmatrix}$$

### 11.3.2 稳定状态

为了用扰动法分析经济系统(11.3)，首先要求得模型的非随机稳定状态，可表示为向量集合  $(\bar{x}, \bar{y})$ ，满足

$$\mathcal{H}(\bar{y}, \bar{y}, \bar{x}, \bar{x}) = 0. \quad (11.4)$$

通常来说，可以直接求得  $(\bar{x}, \bar{y})$  的解析解，或使用一些常规的非线性求解方法（见下文），我们假定稳态解存在且唯一<sup>4</sup>。此外，也可能系统不存在唯一的稳态解，而是有一个均衡路径解 (BGP, Balanced Growth Path)，即模型中的内生变量，除了劳动力等少量例外之外，以大于 0 的某一相同速度增长，这就需要对变量做 scaling (或称为去趋势 detrending)，以便在随后进一步应用局部扰动法。具体说来，设

$$\hat{x}_t = \frac{x_t}{\mu_t},$$

其中  $\mu_t$  是含有趋势的变量，如某一有偏技术进步过程，用额外一个方程来表示。经过 scale 或称 detrend 处理后的  $\hat{x}_t$  可代回系统中做扰动法处理，并且如果有必要，在下一阶段的近似求解及方针过程中，将趋势重新加回来，变成  $x_t$ 。DSGE 模型中对变量做 scaling 处理的例子可见如Fernández-Villaverde and Rubio-Ramírez (2007); Christiano et al. (2010)。

在求解稳态  $(\bar{x}, \bar{y})$  时，有些小窍门可以使得计算过程更简单。举三个例子。

#### 系统缩减

第一个例子是方程系统的缩减 (system reduction)。将由  $n$  个方程,  $n$  个未知变量的原系统，通过替换掉其中一部分变量，缩减为含有  $n'$  个方程,  $n'$  个未知变量的新系统,  $n' < n$ 。例如Christiano et al. (2005) 利用资本存量的运动式

$$k_{t+1} = i_t + (1 - \delta) k_t$$

在经济系统模型中替换掉投资  $i_t$ 。通过方程系统缩减，在下一步扰动法求解非线性系统过程中，工作复杂度进一步降低。如Sikorski (1985) 便指出，方程系统维度的增加常伴随着求解过程复杂程度的指数上升。

---

<sup>4</sup>的确存在着无解或存在多个解的可能，相关讨论见Galor (2007)。

### 变量标准化与参数赋值

第二个例子是通过变量的标准化 (normalization) 确定参数值。可将一个或数个变量设为固定值，进而其他变量是关于这几个固定变量有关的方程，对非线性系统进行简化。以简单的随机内生经济增长模型来说：先来看一个变量标准化的例子：典型家庭的效用函数表现为 log CRRA 形式

$$E_0 \sum_{t=0}^{\infty} \beta^t \left( \ln c_t - \psi \frac{l_t^{1+\eta}}{1+\eta} \right),$$

生产函数

$$y_t = A_t k_t^\alpha l_t^{1-\alpha},$$

资本存量的运动法则

$$k_{t+1} = i_t + (1 - \delta)k_t = (y_t - c_t) + (1 - \delta)k_t.$$

通过计算 FOC，可得模型一组静态最优条件，其中劳动力供应式

$$w_t = \psi c_t l_t^\eta.$$

可以将非随机稳态的劳动力供应予以标准化，设为  $\bar{l} = 1$ 。对应地，由其他全部均衡条件构成的新系统中，内生变量均表示为与此标准化劳动量  $\bar{l}$  有关的函数。进而，通过静态优化条件，我们可以得到参数  $\psi$  的值

$$\psi = \frac{\bar{w}}{\bar{c}\bar{l}^\eta} = \frac{\bar{w}}{\bar{c}}.$$

### 多变量标准化与参数赋值

第三个例子是多变量标准化。在上个模型已经引入  $\bar{l} = 1$  的基础上，在引入第二个标准化，设  $\bar{y} = 1$ ，从而

$$\bar{y} = 1 = \bar{A} \bar{k}^\alpha \bar{l}^{1-\alpha} = \bar{A} \bar{k}^\alpha,$$

我们因此有

$$\begin{aligned} \bar{A} &= \frac{1}{\bar{k}^\alpha}, \\ \bar{w} &= (1 - \alpha) \frac{\bar{y}}{\bar{l}} = 1 - \alpha. \end{aligned}$$

跨期消费的 Euler 方程变为

$$\frac{1}{\bar{c}} = \frac{1}{\bar{c}} \beta (1 + \bar{r} - \delta),$$

因此

$$\bar{r} = \frac{1}{\beta} - 1 + \delta.$$

此外由于

$$\bar{r} = MPK = \alpha \frac{\bar{y}}{\bar{k}} = \alpha \frac{\bar{y}}{\bar{k}},$$

可得稳态资本存量和稳态消费

$$\begin{aligned} \bar{k} &= \frac{\alpha}{\bar{r}} = \frac{\alpha}{\frac{1}{\beta} - 1 + \delta}, \\ \bar{c} &= \bar{y} - \delta \bar{k} = 1 - \delta \frac{\alpha}{\frac{1}{\beta} - 1 + \delta}. \end{aligned}$$

最后我们有

$$\psi = \frac{\bar{w}}{\bar{c}} = \frac{1 - \alpha}{1 - \delta \frac{\alpha}{\frac{1}{\beta} - 1 + \delta}}.$$

### 11.3.3 外生随机过程

设外生冲击变量  $x_2$  可以表示为如下过程

$$\mathbf{x}'_2 = \mathbf{C}(\mathbf{x}_2) + \sigma\eta_\varepsilon\varepsilon', \quad (11.5)$$

其中

- 运算符  $\mathbf{C}$  可能反映非线性关系。在这里假定围绕稳定状态  $(\bar{x}, \bar{y})$  所做的  $\mathbf{C}$  的 Hessian matrix 中，所有特征值绝对值都小于 1。
- $\varepsilon'$  包含  $n_\varepsilon$  个外生波动，常假定为  $\mathcal{N}(0, I)$ ，即均值为 0，二阶矩小于  $\infty$  的分布。在研究中常常需要引入额外的矩限定条件。
- $\eta_\varepsilon$  是一个  $n_\varepsilon \times n_\varepsilon$  的关于波动  $\varepsilon$  的方差协方差矩阵， $\sigma \geq 0$  是一个关于  $\eta_\varepsilon$  的扰动参数 (Schmitt-Grohé and Uribe, 2004)，我们将在随后深入讨论。

因此经验研究中，常将(11.5)写成

$$\mathbf{x}'_2 = \mathbf{C}\mathbf{x}_2 + \sigma\eta_\varepsilon\varepsilon', \quad (11.6)$$

其中  $\mathbf{C}$  是一个  $n_\varepsilon \times n_\varepsilon$  矩阵，其全部特征值绝对值小于 1。

#### 外生冲击的波动的线性特征

(11.5)-(11.6)均假定波动是以线性方式作用于随机过程  $\mathbf{x}_2$  的。这一假定看似随意，但不失一般特性。如我们不再假定这一线性关系，设

$$\mathbf{x}_{2,t} = \mathbf{D}(\mathbf{x}_{2,t-1}, \sigma\eta_\varepsilon\varepsilon_t).$$

将扰动  $\varepsilon_t$  写入经济系统  $\mathbf{x}_{2,t}$  中

$$\tilde{\mathbf{x}}_{2,t} = \begin{bmatrix} \mathbf{x}_{2,t-1} \\ \varepsilon_t \end{bmatrix}, \quad \tilde{\varepsilon}_t = \begin{bmatrix} \mathbf{0}_{n_\varepsilon \times 1} \\ \varepsilon_t \end{bmatrix},$$

则我们有

$$\mathbf{x}_{2,t} = \tilde{\mathbf{D}}(\tilde{\mathbf{x}}_{2,t}, \sigma\eta_\varepsilon).$$

外生随机冲击的过程，可以重新表示为

$$\begin{bmatrix} \mathbf{x}_{2,t} \\ \varepsilon_{t+1} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{D}}(\tilde{\mathbf{x}}_{2,t}, \sigma\eta_\varepsilon) \\ 0 \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{n_\varepsilon \times 1} \\ \varepsilon_{t+1} \end{bmatrix},$$

或者用递归形式来表示

$$\tilde{\mathbf{x}}'_2 = \mathbf{C}(\tilde{\mathbf{x}}_2) + \tilde{\varepsilon}', \quad \tilde{\varepsilon} \sim iid(0, I).$$

举例来说明。时变波动性 (time-varying volatility) 对于理解总量层面上的变量变化具有重要影响 (Bloom, 2009; Fernández-Villaverde et al., 2011)。假定生产率是一个随机波动的过程，满足

$$\log a_t = \rho_a \log a_{t-1} + \lambda_t \nu_t,$$

其中波动  $\nu_t \sim \mathcal{N}(0, 1)$ ，波动的标准差  $\lambda_t$  表示为另一个自回归过程

$$\log \lambda_t = \bar{\lambda} + \rho_\lambda \log \lambda_{t-1} + \psi \eta_t, \quad \eta_t \sim \mathcal{N}(0, 1).$$

需要指出的是，在这个模型中，扰动参数  $\psi$  不止影响生产率的波动  $\nu_t$ ，还影响波动的标准差  $\lambda_t$ 。

时变波动模型中常常含有大量状态变量。对于每一个随机过程，都需要追踪其水平值及方差随时间的变化情况。这使得适合采用扰动法来做模型近似求解，求他方法如映射法等，则难以处理如此多的变量。

此外需要指出的是，如(11.6)所示，尽管一个模型可能有许多个冲击，但在模型设定上只需要一个扰动参数  $\psi$  即可，不同冲击的相对水平及联动 (comovements) 是由系数矩阵  $\eta_\varepsilon$  来调节的。如果  $\sigma = 0$ ，则模型回到确定性经济增长模型中去。

### 扰动参数的讨论

(11.6)的经济模型中，引入扰动参数  $\sigma$  反映随机过程的标准差。然而事实上这并不是唯一的建模方案。首先，在模型中其他位置用其他参数来反映波动，可能会使得模型的解更加精确，和/或模拟结果与现实情况更贴近，如Hansen et al. (2008) 构建了一个含有 Epstein-Zin 偏好的经济模型，并围绕跨期弹性等于 1 时的点做扰动。其次，比起离散时间模型来，连续时间建模中对扰动参数的处理有所不同，并且后者在控制方差方面有优势。

### 扰动法近似解与仿真结果的稳定性

存在这样的可能：扰动法求得的近似解呈现收敛特征，但根据近似解所生成的仿真数据序列却是爆炸的。一种常见的处理方法如Samuelson (1970) 和Jin and Judd (2002) 等，在在外生过程的波动中引入一个有界的设定，从而避免扰动法近似解中可能存在的仿真结果稳定性问题。另一种新出现的方法称剪枝法，见下文。

#### 11.3.4 设想解

非线性经济系统(11.3)设想中的解，应该由如下一组方程构成，分别是控制变量的决策式和状态变量的决策式：

$$\mathbf{y} = \mathbf{g}(\mathbf{x}; \delta), \quad (11.7)$$

$$\mathbf{x}' = \mathbf{h}(\mathbf{x}; \delta) + \sigma \eta \varepsilon', \quad (11.8)$$

其中两个运算符  $\mathbf{g} : \mathcal{R}^{n_x} \times \mathcal{R}^+ \mapsto \mathcal{R}^{n_y}$ ,  $\mathbf{h} : \mathcal{R}^{n_x} \times \mathcal{R}^+ \mapsto \mathcal{R}^{n_x}$ 。系数矩阵  $\eta$  可做如下分解

$$\eta_{(n_x \times n_\varepsilon)} = \begin{bmatrix} \mathbf{0} \\ (n_x - n_\varepsilon \times n_\varepsilon) \eta_\varepsilon \\ (n_\varepsilon \times n_\varepsilon) \end{bmatrix},$$

其中前半部分 ( $n_x - n_\varepsilon$  行) 由当期状态决定，这些当期状态影响下一期的内生状态。后半部分 ( $n_\varepsilon$  行) 由下一期的外生变量决定，下一期的外生变量由当期外生变量和下一期的波动共同决定。

扰动法求解经济系统的目地是：围绕某一适当的不动点，对  $\mathbf{g}$  和  $\mathbf{h}$  做泰勒级数展开。分三个步骤逐次展开。

步骤一，常用非随机稳态作为不动点，对应  $\mathbf{x}_t \equiv \bar{\mathbf{x}}$ ,  $\sigma = 0$ 。根据第11.3.2的方法，我们可以计算稳态值  $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ 。

步骤二，将确定性稳态  $(\bar{\mathbf{x}}; 0)$  代入决策方程组(11.7)-(11.8)有

$$\mathbf{g}(\bar{\mathbf{x}}; 0) = \bar{\mathbf{y}}, \quad (11.9)$$

$$\mathbf{h}(\bar{\mathbf{x}}; 0) = \bar{\mathbf{x}}. \quad (11.10)$$

步骤三，引入关于  $\mathcal{H}$  的未知解，定义为一个新的运算符  $F : \mathcal{R}^{n_x+1} \mapsto \mathcal{R}^{n_x}$ ，式(11.3)改写为

$$\begin{aligned} F(\mathbf{x}, \sigma) &\equiv E_t \mathcal{H}(\mathbf{y}', \mathbf{y}, \mathbf{x}', \mathbf{x}) \\ &= E_t \mathcal{H}(\mathbf{g}(\mathbf{h}(x; \sigma) + \delta\eta\varepsilon'; \sigma), \mathbf{g}(\mathbf{x}; \sigma), \mathbf{h}(x; \sigma) + \delta\eta\varepsilon', \mathbf{x}) \\ &= \mathbf{0}. \end{aligned}$$

由于  $F(\mathbf{x}; \sigma) = \mathbf{0} \quad \forall \mathbf{x}, \sigma$ ，因此我们有

$$F_{x_i^k \sigma^j}(\mathbf{x}; \sigma) = 0, \quad \forall x, \sigma, i, k, j, \quad (11.11)$$

其中  $F_{x_i^k \sigma^j}(\mathbf{x}; \sigma)$  表示用  $F(\mathbf{x}; \sigma)$  对  $\mathbf{x}$  中第  $i$  个元素  $x$  求导  $k$  次，对  $\sigma$  求导  $j$  次，求导都是围绕  $(\mathbf{x}; \sigma)$  进行的。根据模型定义，对于所有可能的状态  $x \in \mathbf{x}$  和扰动参数值  $\sigma$ ，均应满足均衡条件  $F(\cdot) = \mathbf{0}$ 。

这带来两个问题需要做进一步的讨论。

### 可求导性的探讨

(11.11)成立的前提是假定在一个固定点，比如稳态  $(\bar{x}; 0)$  附近， $F(\cdot)$  的所有导数都存在。在 DSGE 模型中，进入  $F(\cdot)$  的各个模块常常是平滑的，如效用函数、生产函数等，这使得该假定可以被接受 (Judd, 1998, pp463)<sup>5</sup>。

可求导性假定指出了扰动法适用范围的限制：如果在一个经济系统中，部分变量是离散的，或相关均衡条件不可导，便无法使用扰动法予以近似。此外有两点需要补充。第一，模型中出现的期望项，有时候会将看起来是时间离散的问题转化为时间连续的问题，举例来说，要不要上大学的决策可以有一个大学学费的随机冲击，或一个努力程度的变量（描述该学生学习或申请奖学金的努力程度）来平滑掉。第二，即便最坏的情况发生了，可求导性假定不成立，但这也并不是说扰动法完全没有价值——它仍然可以成为一个寻找替代求解方法的猜测值。

补一个 ref

### 导数的计算

在扰动法近似求解过程中，导数的计算是核心环节之一，然而这往往是一项繁琐的工作，完全依手算较不明智，时间较长、错误率较高 (Judd, 1998, ch.7)。这时需要借助计算机和支持符号运算的计算软件，如 Mathematica、Python 等。因此现在常用结合二者优点的自动微分 (AD, automatic differentiation) 技术，将链式法则 (chain rule) 应用到一系列基本运算中去。自动微分法在 DSGE 模型中的运用，相关讨论可见 Bastani and Guerrieri (2008)。

#### 11.3.5 一阶扰动

围绕  $(\bar{x}; 0)$  对  $\mathbf{g}$  和  $\mathbf{h}$  做一阶扰动有

$$\begin{aligned} \mathbf{y}' &= \mathbf{g}(\mathbf{x}; \sigma) = \mathbf{g}(\bar{x}; 0) + \mathbf{g}_x(\bar{x}; 0)(\mathbf{x} - \bar{\mathbf{x}})' + \mathbf{g}_\sigma(\bar{x}; 0)\sigma \\ \mathbf{x}' &= \mathbf{h}(\mathbf{x}; \sigma) = \mathbf{h}(\bar{x}; 0) + \mathbf{h}_x(\bar{x}; 0)(\mathbf{x} - \bar{\mathbf{x}})' + \mathbf{h}_\sigma(\bar{x}; 0)\sigma, \end{aligned}$$

结合(11.9)-(11.10)的稳态值，可得

$$\begin{aligned} \mathbf{g}(\mathbf{x}; \sigma) - \bar{\mathbf{y}} &= \mathbf{g}_x(\bar{x}; 0)(\mathbf{x} - \bar{\mathbf{x}})' + \mathbf{g}_\sigma(\bar{x}; 0)\sigma \\ \mathbf{h}(\mathbf{x}; \sigma) - \bar{\mathbf{x}} &= \mathbf{h}_x(\bar{x}; 0)(\mathbf{x} - \bar{\mathbf{x}})' + \mathbf{h}_\sigma(\bar{x}; 0)\sigma, \end{aligned} \quad (11.12)$$

<sup>5</sup>对这一假定的经验验证可能会较为复杂，需要做更深入的研究，可参考经典讨论如 Santos (1992)。

即，在已知  $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$  的基础上，只需要再得到四组未知系数的值，总共  $n \times (n_x + 1)$  个，我们便可测算出任意一点  $(\mathbf{x}; \sigma)$  对应的  $(\mathbf{x}', \mathbf{y}')$ ：

$$\begin{bmatrix} \mathbf{g}_{\mathbf{x}}(\bar{\mathbf{x}}; 0) \\ \mathbf{h}_{\mathbf{x}}(\bar{\mathbf{x}}; 0) \\ \mathbf{g}_{\sigma}(\bar{\mathbf{x}}; 0) \\ \mathbf{h}_{\sigma}(\bar{\mathbf{x}}; 0) \end{bmatrix}$$

这一系列未知系数，可以通过(11.11)求得，具体来说，分两组

- 利用  $F_x(\bar{\mathbf{x}}; 0) = 0$  ( $n \times n_x$  个方程) 求得  $(\mathbf{g}_x(\bar{\mathbf{x}}; 0), \mathbf{h}_x(\bar{\mathbf{x}}; 0))$ ，
- 利用  $F_{\sigma}(\bar{\mathbf{x}}; 0) = 0$  ( $n$  个方程) 求得  $(\mathbf{g}_{\sigma}(\bar{\mathbf{x}}; 0), \mathbf{h}_{\sigma}(\bar{\mathbf{x}}; 0))$ 。

在做进一步求导之前，我们先介绍一下张量 (tensor) 的表示形式。

**Remark 11.3** (张量形式). 张量 (tensor)，或称爱因斯坦求和约定 (Einstein summation notation)，常用于物理学研究中，致力于将执行扰动近似所需的代数计算工作维持在一个可控的范围内。这主要包括两部分：一是省略  $\Sigma, \partial$  等运算符号，二是省略求导所围绕的固定点的表述。在 1 个  $m$  维空间中，第  $n$  个张量是 1 个有  $n$  个 index 和  $m^n$  个元素的运算符，符合相关的转换规则。如在此例中， $\mathcal{H}$  对  $\mathbf{y}$  求导形成一个  $n \times n_y$  的导数矩阵， $[\mathcal{H}]^i_{\alpha}$  表示其中第  $i$  行第  $\alpha$  个元素。当前一个矩阵的下角标也是后一个矩阵的上角标时，我们常常省略求和符号，如

$$[\mathcal{H}_y]^i_{\alpha} [\mathbf{g}_x]_{\beta} [\mathbf{h}_x]^{\beta}_j = \sum_{\alpha=1}^{n_y} \sum_{\beta=1}^{n_x} \frac{\partial \mathcal{H}^i}{\partial y^{\alpha}} \frac{\partial \mathbf{g}^{\alpha}}{\partial x^{\beta}} \frac{\partial \mathbf{h}^{\beta}}{\partial x^j}.$$

类似地，高阶导数常表示为  $[\mathcal{H}_{y'y'}]_{\alpha\gamma}^i$  的形式。其中  $\mathcal{H}_{y'y'}$  是个  $n$  行， $n_y$  列， $n_y$  页的三围数组， $[\mathcal{H}_{y'y'}]_{\alpha\gamma}^i$  是二阶导数矩阵的第  $i$  行，第  $\alpha$  列，第  $\gamma$  页元素。

### $F_x(\cdot)$ 的求解

先来看(11.11)中的前半部分。 $(\mathbf{g}_x(\bar{\mathbf{x}}; 0), \mathbf{h}_x(\bar{\mathbf{x}}; 0))$  构成  $[F_x(\bar{\mathbf{x}}; 0)]_j^i$  的解，表示为

$$\begin{aligned} [F_x(\bar{\mathbf{x}}; 0)]_j^i &= [\mathcal{H}_{y'}]^i_{\alpha} [\mathbf{g}_x]_{\beta} [\mathbf{h}_x]^{\beta}_j \\ &\quad + [\mathcal{H}_y]^i_{\alpha} [\mathbf{g}_x]_j^{\alpha} \\ &\quad + [\mathcal{H}_{x'}]^i_{\beta} [\mathbf{h}_x]^{\beta}_j \\ &\quad + [\mathcal{H}_x]^i_j \\ &= 0, \end{aligned} \tag{11.13}$$

$$i = 1, \dots, n, \quad j, \beta = 1, \dots, n_x, \quad \alpha = 1, \dots, n_y.$$

(11.13)构成一个  $n \times n_x$  个方程组成的系统，其中含有一组系数和一组待解未知数。系数包含  $([\mathcal{H}_{y'}]^i_{\alpha}, [\mathcal{H}_y]^i_{\alpha}, [\mathcal{H}_{x'}]^i_{\beta}, [\mathcal{H}_x]^i_j)$ 。稳定状态  $(\bar{\mathbf{x}}; 0)$  对应  $(\bar{\mathbf{y}}', \bar{\mathbf{y}}, \bar{\mathbf{x}}', \bar{\mathbf{x}})$ ，可以对  $\mathcal{H}$  分别求导数测算这组系数。未知数有  $n \times n_x$  个，由元素  $(\mathbf{g}_x(\bar{\mathbf{x}}; 0), \mathbf{h}_x(\bar{\mathbf{x}}; 0))$  构成。

根据上述分析，可将方程系统(11.13)写为一个二次形式的矩阵系统

$$\frac{A}{\tilde{n} \times \tilde{n}} P^2 - \frac{B}{\tilde{n} \times \tilde{n}} P - \frac{C}{\tilde{n} \times \tilde{n}} = \frac{0}{\tilde{n} \times \tilde{n}}, \tag{11.14}$$

其中系数矩阵  $A, B, C$  包含  $([\mathcal{H}_{y'}]_{\alpha}^i, [\mathcal{H}_y]_{\alpha}^i, [\mathcal{H}_{x'}]_{\beta}^i, [\mathcal{H}_x]_j^i)$ 。 $P$  包含待解未知数  $[\mathbf{h}_x]_j^{\beta}$ , 是与状态向量  $\mathbf{x}$  中内生状态向量  $\mathbf{x}_1$  的运动法则式有关的元素。

至此我们可以采用标准的二次矩阵方程求解法来求  $P$ , 如 QZ 分解法。

**Remark 11.4** (QZ 分解法求解二次矩阵方程系统). 援引 ([Uhlig, 1999, pp.43-45](#)) 的方法。对于二次矩阵系统([11.14](#)), 定义两个矩阵

$$\underset{2\tilde{n} \times 2\tilde{n}}{D} = \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix}, \quad \underset{2\tilde{n} \times 2\tilde{n}}{D} = \begin{bmatrix} B & C \\ I & 0 \end{bmatrix}.$$

定义两个酉矩阵 (*unitary matrix*)  $Q, Z$ <sup>6</sup>。

设两个上三角矩阵  $\Phi, \Sigma$ , 对角元素分别为  $\phi_{ii}, \sigma_{ii}$ 。

对  $D$  和  $F$  分别作扩展 Schur 分解 (QZ 分解):

$$\begin{aligned} Q^T \Sigma Z &= D, \\ Q^T \Phi Z &= F, \end{aligned}$$

其中确保  $\Sigma$  和  $\Phi$  是对角矩阵, 并且对角元素的比值  $\left| \frac{\phi_{ii}}{\sigma_{ii}} \right|$  按照从左上到右下的升序排列<sup>7</sup>, 从而可以将系统分为两部分, 前半部分是稳定分块, 对应  $\left| \frac{\phi_{ii}}{\sigma_{ii}} \right| < 1$ ; 后半部分是不稳定分块, 对应  $\left| \frac{\phi_{ii}}{\sigma_{ii}} \right| > 1$ 。在排序之后, 可以将矩阵输入 *Mathematica*, 做 QZ 分解计算。测得的矩阵

$$\underset{2\tilde{n} \times 2\tilde{n}}{Z} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}.$$

进而我们有

$$P = -Z_{21}^{-1} Z_{22}.$$

对于  $2\tilde{n}$  个扩展特征值比 (*generalized eigenvalue ratio*)  $\left| \frac{\phi_{ii}}{\sigma_{ii}} \right|$  来说:

1. 如果稳定的扩展特征值比数量  $< \tilde{n}$ , 那么  $P$  值存在且唯一, 对于任何  $\tilde{n}$  维度向量中, 有  $\lim_{m \rightarrow \infty} P^m \mathbf{x} \rightarrow \mathbf{0}$ 。
2. 如果稳定的扩展特征值比数量  $> \tilde{n}$ , 那么  $P$  值可能有多个, 我们选取任何一个满足如下条件的  $P$  值:  $\tilde{n}$  维度向量中,  $\lim_{m \rightarrow \infty} P^m \mathbf{x} \rightarrow \mathbf{0}$ 。

之所以会出现二次形式的矩阵系统, 其原因在于, 通常说来满足模型均衡条件的内生变量变化路径并不唯一, 而是有很多 ([Uhlig, 1999; Galor, 2007](#)), 其中有一些是稳定路径, 有一些是不稳定路径, 而只有稳定路径才能满足适当的横截条件。显然, 对于存在多重解的经济系统而言, 我们需要找出其中的稳定路径解。许多 DSGE 模型中存在  $\tilde{n}$  个扩展特征根, 从而确保系统有且只有 1 个稳定解。如  $< \tilde{n}$ , 扩展特征根的数量不足, 会导致均衡动态路径出现内在不稳定的状况。如  $> \tilde{n}$ , 会存在太阳黑子现象 ([Lubik and Schorfheide, 2003](#))。这段对解的存在性和唯一性的分析是基于一阶扰动近似而言的, 但对高阶扰动近似过程也同样有效, 换句话说: 如果一阶扰动近似的解是唯一的 (不唯一的), 则高阶扰动近似的解也是唯一的 (不唯一的)。

<sup>6</sup>满足条件  $Q^H Q = Z^H Z = I_{2\tilde{n}}$ , 其中  $H$  表示埃米特转置。

<sup>7</sup>这是因为,  $\left| \frac{\phi_{ii}}{\sigma_{ii}} \right|$  的每一种排列顺序, 都存在一个 QZ 分解的解。

**Remark 11.5** (二次系统的递归分解). 对于(11.13)的经济系统, 还可以分解为两部分, 以递归形式分两步求解。第一步, 定义一个分块系统

$$\begin{aligned} [F_x(\bar{x}; 0)]_j^i &= [\mathcal{H}_{y'}]_\alpha^i [\mathbf{g}_x]_\beta^\alpha [\mathbf{h}_x]_j^\beta \\ &\quad + [\mathcal{H}_y]_\alpha^i [\mathbf{g}_x]_j^\alpha \\ &\quad + [\mathcal{H}_{x'}]_\beta^i [\mathbf{h}_x]_j^\beta \\ &\quad + [\mathcal{H}_x]_j^i \\ &= 0, \end{aligned} \tag{11.15}$$

$$i = 1, \dots, n, \quad j, \beta = 1, \dots, \tilde{n}, \quad \alpha = 1, \dots, n_y.$$

与(11.13)相比, (11.15)的系统中未知数的数量减少了, 从  $n \times n_x$  个变为  $n \times \tilde{n}$  个, 由元素  $\left( \begin{array}{c} \mathbf{g}_x(\bar{x}; 0), \mathbf{h}_x(\bar{x}; 0) \\ \tilde{n} \times n_y \quad \tilde{n} \times n_x \end{array} \right)$  构成, 二者都与  $\tilde{n}$  个内生变量  $x_1$  有关。

采用上文提到的办法, 求得对应(11.15)系统的解  $\left( \begin{array}{c} \mathbf{g}_x(\bar{x}; 0), \mathbf{h}_x(\bar{x}; 0) \\ \tilde{n} \times n_y \quad \tilde{n} \times n_x \end{array} \right)$ 。

第二步, 将前一步求得的解代入新系统

$$\begin{aligned} [F_x(\bar{x}; 0)]_j^i &= [\mathcal{H}_{y'}]_\alpha^i [\mathbf{g}_x]_\beta^\alpha [\mathbf{h}_x]_j^\beta \\ &\quad + [\mathcal{H}_y]_\alpha^i [\mathbf{g}_x]_j^\alpha \\ &\quad + [\mathcal{H}_{x'}]_\beta^i [\mathbf{h}_x]_j^\beta \\ &\quad + [\mathcal{H}_x]_j^i \\ &= 0, \end{aligned} \tag{11.16}$$

$$i = 1, \dots, n, \quad j, \beta = \tilde{n} + 1, \dots, n_x, \quad \alpha = 1, \dots, n_y,$$

并进一步求解与  $n_\varepsilon$  个外生随机变量  $x_2$  有关的未知数, 由元素  $\left( \begin{array}{c} \mathbf{g}_x(\bar{x}; 0), \mathbf{h}_x(\bar{x}; 0) \\ n_\varepsilon \times n_y \quad n_\varepsilon \times n_x \end{array} \right)$  构成。

采用递归分解法求解二次系统, 有三个优点。第一, 可以显著改善计算时间: (11.15)比(11.13)减少了  $n_\varepsilon \times n$  个未知数的计算, 待求解系统规模小了很多 (Rubio-Ramírez and Fernández-Villaverde, 2008)。第二, (11.16)的线性系统, 比起(11.13)的非线性系统来, 求解更为便利。第三, 对于含有  $n_x$  个状态变量的向量  $\mathbf{x}$  而言, 有时我们只需关注其中的  $\tilde{n}$  个内生变量而非全部。这取决于特定的研究目标, 一个例子是, 在一定的初始条件下, 计算经济系统向某一稳定状态运动的确定性转移路径。另一个例子是, 基于一阶扰动近似解, 描绘经济系统的冲击——响应机制。

### $F_\sigma(\cdot)$ 的求解

再来看(11.11)中的后半部分。 $(\mathbf{g}_\sigma(\bar{x}; 0), \mathbf{h}_\sigma(\bar{x}; 0))$  构成  $[F_\sigma(\bar{x}; 0)]^i$  的解, 表示为

$$\begin{aligned} [F_\sigma(\bar{x}; 0)]^i &= E_t \{ [\mathcal{H}_{y'}]_\alpha^i [\mathbf{g}_x]_\beta^\alpha [\mathbf{h}_\sigma]^\beta \\ &\quad + [\mathcal{H}_{y'}]_\alpha^i [\mathbf{g}_x]_\beta^\alpha [\eta]_\phi^\beta [\varepsilon']^\phi \\ &\quad + [\mathcal{H}_{y'}]_\alpha^i [\mathbf{g}_\sigma]^\alpha \\ &\quad + [\mathcal{H}_y]_\alpha^i [\mathbf{g}_\sigma]^\alpha \\ &\quad + [\mathcal{H}_{x'}]_\beta^i [\mathbf{h}_\sigma]^\beta \\ &\quad + [\mathcal{H}_{x'}]_\beta^i [\eta]_\phi^\beta [\varepsilon']^\phi \} \\ &= 0, \end{aligned} \tag{11.17}$$

$$i = 1, \dots, n, \quad \alpha = 1, \dots, n_y, \quad \beta = 1, \dots, n_x, \quad \phi = 1, \dots, n_\varepsilon.$$

一个 ref, 见下  
的范例。

进而我们有

$$[F_\sigma(\bar{x}; 0)]^i = [\mathcal{H}_{y'}]_\alpha^i [\mathbf{g}_x]_\beta^\alpha [\mathbf{h}_\sigma]^\beta + [\mathcal{H}_{y'}]_\alpha^i [\mathbf{g}_\sigma]^\alpha + [\mathcal{H}_y]_\alpha^i [\mathbf{g}_\sigma]^\alpha + [\mathbf{f}_{x'}]_\beta^i [\mathbf{h}_\sigma]^\beta = 0, \quad (11.18)$$

(11.17) 构成一个  $n$  个方程组成的系统，是关于未知数  $(\mathbf{g}_\sigma(\bar{x}; \sigma), \mathbf{h}_\sigma(\bar{x}; \sigma))$  的线性齐次方程组，唯一满足条件的解只能是：

$$\begin{aligned} \bar{\mathbf{g}}_\sigma(\bar{x}; 0) &= 0, \\ \bar{\mathbf{h}}_\sigma(\bar{x}; 0) &= 0. \end{aligned} \quad (11.19)$$

### 确定性等价条件

(11.19) 代入 (11.12) 得

$$\begin{aligned} \mathbf{g}(\mathbf{x}; \sigma) - \bar{\mathbf{y}} &= \mathbf{g}_x(\bar{x}; 0)(\mathbf{x} - \bar{\mathbf{x}})', \\ \mathbf{h}(\mathbf{x}; \sigma) - \bar{\mathbf{x}} &= \mathbf{h}_x(\bar{x}; 0)(\mathbf{x} - \bar{\mathbf{x}})'. \end{aligned} \quad (11.20)$$

这被称为确定性等价条件 (certainty equivalence)。Simon (1956); Theil (1957)。根据确定性等价条件，模型的一阶近似解等于在完全期望条件下，或者  $\sigma = 0$  条件下的解。确定性等价条件承认外生冲击对系统的决策式 (decision rule) 的影响，但它排除外生冲击的标准差 (不确定性) 对系统决策式的影响。

确定性等价假定提出的理由很简单：典型个体的风险规避程度，通常可以通过 concave 效用函数的二阶求导给出，从而一阶扰动近似即可满足需要。但甚至早至 Leland (1968); Sandmo (1970) 的研究便已指出，为了应对不确定性带来的风险，储蓄等个人的预防性行为 (precautionary behaviors) 的分析只有对效用函数做三阶求导才能推得，换句话说，二阶以上的扰动近似是必要的。与之相比，一阶线性扰动近似只涉及到模型的均衡条件 (尤其是效用函数的一阶导数，比如跨期消费决策的 Euler 方程)，以及对这些均衡条件的一阶求导 (包括对效用函数的二次求导)，而不包含更高阶导数了。

确定性等价存在以下几点不足：第一，难于分析不确定性给福利效果带来的影响。一方面模型的动态路径的确受到来自波动的方差 (标准差) 的影响，波动的方差本身即是波动的来源之一。但另一方面，模型中的行为人，基于确定性等价假定，无法采取预防性措施应对方差变化带来的不确定性风险，这使得相关福利分析难以进行。第二，对应于前一点，任何基于确定性等价的近似解，也因此无法用于评估资产的风险溢价。这并不符合最初构建模型时的设想：在一般均衡框架下，福利的测算和资产的定价二者之间存在着密切的关联，任何对前者的研究都自然涉及到后者 (Alvarez and Jermann, 2004)。第三，确定性等价假定使得研究者无法探讨波动 (volatility) 变化的经济效果。

**Remark 11.6** (线性二次 (LQR) 法，近似 LQR 法，与扰动法)。 Kydland and Prescott (1982) 及其后相当数量的研究，探讨另一种 DSGE 模型的求解方法，可称之为线性二次调节法 (LQR, linear quadratic regulator)。对于一个最优控制问题，有  $n_x$  个状态变量  $\mathbf{x}_t$ ，和  $n_u$  个控制变量  $\mathbf{u}_t$ ，定义为  $\mathbf{w}_t = [\mathbf{x}_t, \mathbf{u}_t]^\top$ ，是个  $n_w = n_x + n_u$  个维度的行向量。最优控制问题可以表示为

$$\begin{aligned} \max E_0 \sum_{t=0}^{\infty} \beta^t r(\mathbf{w}_t) \\ s.t. \quad \mathbf{x}_{t+1} = A(\mathbf{w}_t, \varepsilon_t), \end{aligned}$$

其中

- 运算符  $r$  表示回报方程，

- 运算符  $A$  汇总模型的均衡条件和预算约束条件,
- $\varepsilon_t$  为  $n_\varepsilon$  个冲击组成的向量, 均值为 0, 方差  $< \infty$ <sup>8</sup>。

做两个假设。第一个假设是回报方程是二次形式的

$$r(\mathbf{w}_t) = B_0 + \underset{1 \times n_w}{B_1} \mathbf{w}_t + \mathbf{w}_t^\top \underset{n_w \times n_w}{B_w} \mathbf{w}_t.$$

第二个假设是  $A$  是线性形式的

$$\mathbf{x}_{t+1} = \underset{n_x \times n_w}{B_3} \mathbf{w}_{t+1} + \underset{n_x \times n_\varepsilon}{B_4} \varepsilon_t.$$

这便构成了一个随机贴现线性——二次控制问题 (*stochastic discounted linear-quadratic regulator problem*), 有大量研究围绕这个问题做了深入探讨, 可参考综述 [Anderson et al. \(1996\)](#); [Hansen and Sargent \(2004, 2013\)](#)。已知该问题的最优决策是个关于状态和冲击的线性方程

$$\mathbf{u}_t = F_w \mathbf{w}_t + F_\varepsilon \varepsilon_t,$$

其中  $F_w$  可用 *Riccati* 方程来求解,  $F_\varepsilon$  可用 *Sylvester* 方程来求解 ([Anderson et al., 1996, pp. 182-183, 202-205](#))。本讲义的第九章也对求解方法做了介绍。 $F_w$  不受冲击的方差的影响, 换句话说, 如果  $\text{var}(\varepsilon_t) = 0$ , 那么最优决策问题简化为

$$\mathbf{u}_t = F_w \mathbf{w}_t.$$

最优控制问题中的 *LQR* 法, 经济系统分为两个部分, 分别计算  $F_w$  和  $F_\varepsilon$ , 这使得研究者可以深入探讨一系列相关问题。然而需要指出的是, *LQR* 法仍然属于一阶扰动法的一种, 因此也持有确定性等价的假定, 受到该假定不足的限制。

[Kydland and Prescott \(1982\)](#) 构建经济模型来分析社会规划者问题, 研究框架符合上述最优调节问题 (*optimal regulator problem*) 的范式, 也因此可以求出关于  $\mathbf{w}_t$  的线性方程  $A(\cdot)$ 。但在 [Kydland and Prescott \(1982\)](#) 中, 报酬方程  $r(\cdot)$  并未设定成二次形式; 他们只是对社会规划者的目标方程做了二次近似。大多数后续研究都使用类似的方法, 对目标方程围绕稳态做 TSE, 可以归类称之为近似 *LQR* 问题。进而, 可以用价值方程来求解近似 *LQR* 问题, 见 [Díaz-Giménez \(1999\)](#)。

当  $A(\cdot)$  是线性的时, 近似 *LQR* 问题的解, 与对模型均衡条件做一阶扰动法近似的解, 是相同的。原因很简单: 导数总是唯一的, 并且两种方法都致力于求得模型的近似线性解。然而, 近似 *LQR* 法在学术界渐渐不受青睐, 主要原因有: 第一, *DSGE* 的模型常常很难将  $A(\cdot)$  写为线性形式。第二, 当经济体不是处于帕累托最优状态时, 我们往往很难构建社会规划者问题。尽管的确有研究者常识构建调整社会规划者问题, 引入一系列非最优的限制条件, 如 [Benigno and Woodford \(2003\)](#), 但比起对模型均衡条件做扰动近似的方法而言, 前者的工作量更多也更为复杂。第三个也许是更重要的原因, 扰动法可以更便利地做高阶近似, 引入非线性元素, 打破确定性等价的假定, 从而使得模型可以分析更多经济现象。我们将在下一节探讨二阶扰动法。

### 11.3.6 二阶扰动

在求得一阶扰动法的解之后, 我们可以迭代求得二阶扰动的系统解。具体来说

---

<sup>8</sup> 在这一状态——空间表示形式中, 模型可以做适当扩展, 以便  $\varepsilon_t$  也会对当期的回报方程, 和当期变量 (包括状态变量  $\mathbf{x}_t$  和当期控制变量  $\mathbf{u}_t$  在内) 产生影响。

1. 对  $\mathbf{g}(\mathbf{x}; \sigma)$  围绕  $(\bar{\mathbf{x}}; 0)$  做二阶近似:

$$\begin{aligned} [\mathbf{g}(\mathbf{x}; \sigma)]^i &= [\mathbf{g}(\bar{\mathbf{x}}; 0)]^i + [\mathbf{g}_x(\bar{\mathbf{x}}; 0)]_a^i [(\mathbf{x} - \bar{\mathbf{x}})]_a + [\mathbf{g}_\sigma(\bar{\mathbf{x}}; 0)]_\sigma^i [\sigma] \\ &\quad + \frac{1}{2} [\mathbf{g}_{xx}(\bar{\mathbf{x}}; 0)]_{ab}^i [(\mathbf{x} - \bar{\mathbf{x}})]_a [(\mathbf{x} - \bar{\mathbf{x}})]_b \\ &\quad + \frac{1}{2} [\mathbf{g}_{x\sigma}(\bar{\mathbf{x}}; 0)]_a^i [(\mathbf{x} - \bar{\mathbf{x}})]_a [\sigma] \\ &\quad + \frac{1}{2} [\mathbf{g}_{\sigma x}(\bar{\mathbf{x}}; 0)]_a^i [(\mathbf{x} - \bar{\mathbf{x}})]_a [\sigma] \\ &\quad + \frac{1}{2} [\mathbf{g}_{\sigma\sigma}(\bar{\mathbf{x}}; 0)]^i [\sigma] [\sigma], \end{aligned} \tag{11.21}$$

$i = 1 \dots n_y, \quad a, b = 1, \dots n_x.$

2. 对  $\mathbf{h}(\mathbf{x}; \sigma)$  围绕  $(\bar{\mathbf{x}}; 0)$  做二阶近似:

$$\begin{aligned} [\mathbf{h}(\mathbf{x}; \sigma)]^j &= [\mathbf{h}(\bar{\mathbf{x}}; 0)]^j + [\mathbf{h}_x(\bar{\mathbf{x}}; 0)]_a^j [(\mathbf{x} - \bar{\mathbf{x}})]_a + [\mathbf{g}_\sigma(\bar{\mathbf{x}}; 0)]_\sigma^j \\ &\quad + \frac{1}{2} [\mathbf{h}_{xx}(\bar{\mathbf{x}}; 0)]_{ab}^j [(\mathbf{x} - \bar{\mathbf{x}})]_a [(\mathbf{x} - \bar{\mathbf{x}})]_b \\ &\quad + \frac{1}{2} [\mathbf{h}_{x\sigma}(\bar{\mathbf{x}}; 0)]_a^j [(\mathbf{x} - \bar{\mathbf{x}})]_a [\sigma] \\ &\quad + \frac{1}{2} [\mathbf{h}_{\sigma x}(\bar{\mathbf{x}}; 0)]_a^j [(\mathbf{x} - \bar{\mathbf{x}})]_a [\sigma] \\ &\quad + \frac{1}{2} [\mathbf{h}_{\sigma\sigma}(\bar{\mathbf{x}}; 0)]^j [\sigma] [\sigma], \end{aligned} \tag{11.22}$$

$j = 1 \dots n_x, \quad a, b = 1, \dots n_x.$

两个方程系统中共有 6 个待求解未知数  $(\mathbf{g}_{xx}, \mathbf{g}_{x\sigma}, \mathbf{g}_{\sigma\sigma}, \mathbf{h}_{xx}, \mathbf{h}_{x\sigma}, \mathbf{h}_{\sigma\sigma})_{(\bar{\mathbf{x}}; 0)}$ , 可同样利用式(11.11)求得。具体说来, 分为三组:

- 利用  $F_{xx}(\bar{\mathbf{x}}; 0) = 0$  求解  $(\mathbf{g}_{xx}(\bar{\mathbf{x}}; 0), \mathbf{h}_{xx}(\bar{\mathbf{x}}; 0))$ ,
- 利用  $F_{\sigma\sigma}(\bar{\mathbf{x}}; 0) = 0$  求解  $(\mathbf{g}_{\sigma\sigma}(\bar{\mathbf{x}}; 0), \mathbf{h}_{\sigma\sigma}(\bar{\mathbf{x}}; 0))$ ,
- 利用  $F_{\sigma x}(\bar{\mathbf{x}}; 0) = 0$  求解  $(\mathbf{g}_{\sigma x}(\bar{\mathbf{x}}; 0), \mathbf{h}_{\sigma x}(\bar{\mathbf{x}}; 0))$ .

### $F_{xx}(\cdot)$ 的求解

一阶导数  $F_x(\cdot) = 0$ , 式(11.13)的 LHS 表示为 4 个部分相加。因此二阶导数

$$[F_{xx}(\bar{\mathbf{x}}; 0)]_{jk}^i = \mathcal{A} + \mathcal{B} + \mathcal{C} + \mathcal{D} = 0, \tag{11.23}$$

其中

$$\begin{aligned} \mathcal{A} &= \left( [\mathcal{H}_{y'y'}]_{\alpha\gamma}^i [\mathbf{g}_x]_\delta^\gamma [\mathbf{h}_x]_k^\delta + [\mathcal{H}_{y'y'}]_{\alpha\gamma}^i [\mathbf{g}_x]_k^\gamma + [\mathcal{H}_{y'x'}]_{\alpha\delta}^i [\mathbf{h}_x]_k^\delta + [\mathcal{H}_{y'x}]_{\alpha k}^i \right) [\mathbf{g}_x]_\beta^\alpha [\mathbf{h}_x]_j^\beta \\ &\quad + [\mathcal{H}_{y'}]_\alpha^i [\mathbf{g}_{xx}]_{\beta\delta}^\alpha [\mathbf{h}_x]_k^\delta [\mathbf{h}_x]_j^\beta + [\mathcal{H}_{y'}]_\alpha^i [\mathbf{g}_x]_\beta^\alpha [\mathbf{h}_{xx}]_{jk}^\beta, \\ \mathcal{B} &= [\mathcal{H}_{yy'}]_{\alpha\gamma}^i [\mathbf{g}_x]_\delta^\gamma [\mathbf{h}_x]_k^\delta + [\mathcal{H}_{yy}]_{\alpha\gamma}^i [\mathbf{g}_x]_k^\gamma + [\mathcal{H}_{yx'}]_{\alpha\delta}^i [\mathbf{h}_x]_k^\delta + [\mathcal{H}_{yx}]_{\alpha k}^i [\mathbf{g}_x]_j^\alpha + [\mathcal{H}_y]_\alpha^i [\mathbf{g}_{xx}]_{jk}^\alpha, \\ \mathcal{C} &= \left\{ [\mathcal{H}_{x'y'}]_{\beta\gamma}^i [\mathbf{g}_x]_\delta^\gamma [\mathbf{h}_x]_k^\delta + [\mathcal{H}_{x'y}]_{\beta\gamma}^i [\mathbf{g}_x]_k^\gamma + [\mathcal{H}_{x'x'}]_{\beta\delta}^i [\mathbf{h}_x]_k^\delta + [\mathcal{H}_{x'x}]_{\beta k}^i \right\} [\mathbf{h}_x]_j^\beta + [\mathcal{H}_{x'}]_\beta^i [\mathbf{h}_{xx}]_{jk}^\beta, \\ \mathcal{D} &= [\mathcal{H}_{xy'}]_{j\gamma}^i [\mathbf{g}_x]_\delta^\gamma [\mathbf{h}_x]_k^\delta + [\mathcal{H}_{xy}]_{j\gamma}^i [\mathbf{g}_x]_k^\gamma + [\mathcal{H}_{xx'}]_{j\delta}^i [\mathbf{h}_x]_k^\delta + [\mathcal{H}_{xx}]_{jk}^i, \end{aligned}$$

$i = 1, \dots, n, \quad j, k, \beta, \delta = 1, \dots, n_x, \quad \alpha, \gamma = 1, \dots, n_y.$

上式中, 所有  $\mathcal{H}$  的导数, 以及关于  $\mathbf{g}(\bar{\mathbf{x}}; 0)$  和  $\mathbf{h}(\bar{\mathbf{x}}; 0)$  的一阶导数, 都是已知数。因此上式是一个  $n \times n_x \times n_x$  个方程组成的系统, 含有以  $\mathbf{g}_{xx}$  和  $\mathbf{h}_{xx}$  为元素的  $(n_x + n_y) \times n_x \times n_x$  个未知数。

$F_{\sigma\sigma}(\cdot)$  的求解

一阶导数  $F_\sigma(\cdot) = 0$ , 式(11.17)的 LHS 表示为 6 个部分相加。因此二阶导数

$$[F_{\sigma\sigma}(\bar{\mathbf{x}}; 0)]^i = \mathcal{A} + \mathcal{B} + \mathcal{C} + \mathcal{D} + \mathcal{E} + \mathcal{F} = 0, \quad (11.24)$$

其中

$$\begin{aligned} \mathcal{A} &= [\mathcal{H}_{y'}]_\alpha^i [\mathbf{g}_x]_\beta^\alpha [\mathbf{h}_{\sigma\sigma}]^\beta, \\ \mathcal{B} &= [\mathcal{H}_{y'y'}]_{\alpha\gamma}^i [\mathbf{g}_x]_\delta^\gamma [\eta]_\zeta^\delta [\mathbf{g}_x]_\beta^\alpha [\eta]_\phi^\beta [I]_\zeta^\phi + [\mathcal{H}_{y'x'}]_{\alpha\delta}^i [\eta]_\zeta^\delta [\mathbf{g}_x]_\beta^\alpha [\eta]_\phi^\beta [\varepsilon']^\phi, \\ \mathcal{C} &= [\mathcal{H}_{y'}]_\alpha^i [\mathbf{g}_{\sigma\sigma}]^\alpha, \\ \mathcal{D} &= [\mathcal{H}_y]_\alpha^i [\mathbf{g}_{\sigma\sigma}]^\alpha, \\ \mathcal{E} &= [\mathcal{H}_{x'}]_\beta^i [\mathbf{h}_{\sigma\sigma}]^\beta, \\ \mathcal{F} &= [\mathcal{H}_{x'y'}]_{\beta\gamma}^i [\mathbf{g}_x]_\delta^\gamma [\eta]_\zeta^\delta [\eta]_\phi^\beta [I]_\zeta^\phi + [\mathcal{H}_{x'x'}]_{\beta\delta}^i [\eta]_\zeta^\delta [\eta]_\phi^\beta [I]_\zeta^\phi, \\ i &= 1 \dots n, \quad \alpha, \gamma = 1, \dots n_y, \quad \beta, \delta = 1, \dots n_x, \dots \phi, \zeta = 1, \dots n_\varepsilon. \end{aligned}$$

上式是一个  $n$  个方程构成的线性系统, 含有由  $\mathbf{g}_{\sigma\sigma}$  和  $\mathbf{h}_{\sigma\sigma}$  构成的  $n$  个未知数。

不为零的  $\mathbf{g}_{\sigma\sigma}$  和  $\mathbf{h}_{\sigma\sigma}$  意味着对风险的纠正 (correction for risk): 风险来自于经济系统对确定性等价的悖离。

 $F_{\sigma x}(\cdot)$  的求解

一阶导数  $F_\sigma(\bar{\mathbf{x}}; 0) = 0$ , 式(11.17)进一补对  $\mathbf{x}$  求导, 可得二阶导数

$$\begin{aligned} [F_{\sigma x}(\bar{\mathbf{x}}; 0)]_j^i &= [\mathcal{H}_{y'}]_\alpha^i [\mathbf{g}_x]_\beta^\alpha [\mathbf{h}_{\sigma x}]^\beta j + [\mathcal{H}_{y'}]_\alpha^i [\mathbf{g}_x]_\gamma^\alpha [\mathbf{h}_x]_j^\gamma + [\mathcal{H}_y]_\alpha^i [\mathbf{g}_{\sigma x}]_j^\alpha + [\mathcal{H}_{x'}]_\beta^i [\mathbf{h}_{\sigma x}]_j^\beta = 0, \\ i &= 1, \dots n, \quad \alpha = 1, \dots n_y, \quad \beta, \gamma, j = 1, \dots n_x, \end{aligned} \quad (11.25)$$

其中所有包括  $\mathbf{g}_\sigma(\bar{\mathbf{x}}; 0), \mathbf{h}_\sigma(\bar{\mathbf{x}}; 0)$  的项都等于 0(确定性等价条件)。

上式是一个  $n \times n_x$  个方程构成的线性系统, 含有由  $\mathbf{g}_{\sigma x}$  和  $\mathbf{h}_{\sigma x}$  构成的  $n \times n_x$  个未知数。系统成立的条件需要满足

$$\begin{aligned} \mathbf{g}_{\sigma x} \bar{\mathbf{x}}; 0 &= 0, \\ \mathbf{h}_{\sigma x} \bar{\mathbf{x}}; 0 &= 0. \end{aligned} \quad (11.26)$$

## 11.3.7 更高阶扰动

如前文所介绍的那样, 在 1 阶扰动近似解的基础上, 我们可以继续迭代, 依次求得  $2, 3, \dots, n$  阶的近似解。如果  $\mathbf{g}$  和  $\mathbf{h}$  在  $\bar{\mathbf{x}}$  附近有解, 那么以  $n$  阶近似生成的一组变量时间序列将会有无限项, 以某个半径  $r$  逐渐收敛到  $\bar{\mathbf{x}}$ 。

$r$  可以无限大, 此时的变量序列有可能收敛到任意一处。 $r$  也可以是个有限的值, 此时变量序列的边界存在不可删除的奇异点。不幸的是, 对大多数 DSGE 模型来说, 收敛半径  $r$  的值是未知的, 这方面尚需进一步深入的研究 (Swanson et al., 2006; Aldrich and Kung, 2011)。

即便该变量序列是收敛的, 以下 2 个问题也值得重视。

1. 在一个  $j$  阶近似中, 我们有可能会找不到  $\mathbf{g}$  和  $\mathbf{h}$  这两个决策式的“正确”形式。例如Aruoba et al. (2006) 的研究表明, 在对一个随机内生经济增长模型做了 5 阶扰动近似后, 所生成的两个近似决策式——消费的决定式和资本积累的决定式——不再是全局 concave 的, 这与经济学理论的假设不符。并且它们都呈现出震荡特征 (oscillating patterns)。

2. 随着扰动近似阶数的不断增加，近似解向某一个特定点（如稳定状态）的收敛，可能不是单调的。事实上我们可以较容易地构建这样一个模型，来说明当经济运行状况远离稳定状态时， $j + 1$  阶扰动近似的结果可能比  $j$  阶近似的结果更差。

这 2 个问题都不是致命性的，但在实际研究开展过程中，它们的确值得研究者予以关注，并使用必要的检验方法，确保其负面影响最小化。例如，查验不同阶近似线性所生成的变量数据序列，与实际观测到的数据序列之间的近似程度。

在后文中我们将讨论如何评估不同阶近似解的精确程度，以及是否有必要采取更高阶扰动来求近似解。例如，为了更好处理含有时变波动性的模型，我们需要至少 3 阶，甚至更高阶的扰动近似。再如 Levintal (2017a) 便指出，对暗含灾难性风险的经济模型来说，至少 5 阶的扰动近似是必须的。

### 11.3.8 例

#### 模型设定

以经典的随机 NCGT 模型为例，介绍二阶扰动法的应用。假定  $u(c_t) = \log c_t$ ,  $\delta = 1$ 。经济模型构成一个非线性系统，由均衡条件（跨期消费决策的 Euler 方程），预算约束条件和外生技术过程组成：

$$\begin{aligned}\frac{1}{c_t} &= \beta E_t \frac{\alpha \exp(z_{t+1}) k_{t+1}^{\alpha-1}}{c_{t+1}}, \\ c_t + k_{t+1} &= \exp(z_t) k_t^\alpha, \\ z_t &= \rho z_{t-1} + \eta \varepsilon_t.\end{aligned}$$

$\delta = 1$  使得技术冲击的收入效果和替代效果互相抵消，并且消费和投资相对于产出的比例恒定。在此基础上，我们可以得到关于消费和资本积累的最优决策方程解析形式：

$$\begin{aligned}c_t &= (1 - \alpha\beta) \exp(z_t) k_t^\alpha, \\ k_{t+1} &= \alpha\beta \exp(z_t) k_t^\alpha\end{aligned}\tag{11.27}$$

现在，假定我们不知道解析形式的最优决策方程。构建猜测的最优决策方程为

$$\begin{aligned}c_t &= c(k_t, z_t), \\ k_{t+1} &= k(k_t, z_t).\end{aligned}$$

把猜测的决策方程代回均衡系统并做适当简化得

$$\begin{aligned}\frac{1}{c(k_t, z_t)} &= \beta E_t \frac{\alpha \exp(\rho z_t + \eta \varepsilon_{t+1}) k(k_t, z_t)^{\alpha-1}}{c(k(k_t, z_t), \rho z_t + \eta \varepsilon_{t+1})}, \\ c(k_t, z_t) + k(k_t, z_t) &= \exp(z_t) \cdot k_t^\alpha.\end{aligned}\tag{11.28}$$

根据扰动法，将猜测的决策方程  $c(\cdot), k(\cdot)$  做关于扰动参数  $\sigma$  的近似：

$$\begin{aligned}c_t &= c(k_t, z_t; \sigma), \\ k_{t+1} &= k(k_t, z_t; \sigma),\end{aligned}$$

以及将扰动参数引入外生的技术过程

$$z_t = \rho z_{t-1} + \sigma \eta \varepsilon_t.$$

### 稳定状态

很显然，当  $\sigma = 0$  或  $z_t = 0 \forall t$  时，随机增长模型回复到传统的确定性增长模型中去<sup>9</sup>。此时系统有且只有一个稳定状态，对应  $(\bar{k}, \bar{c})$ ，满足

$$\begin{aligned}\bar{k} &= k(\bar{k}, 0; 0) = (\alpha\beta)^{\frac{1}{1-\alpha}}, \\ \bar{c} &= c(\bar{c}, 0; 0) = (\alpha\beta)^{\frac{\alpha}{1-\alpha}} - (\alpha\beta)^{\frac{1}{1-\alpha}}.\end{aligned}\quad (11.29)$$

### 对模型做扰动近似

对猜测的消费决策方程和猜测的资本存量决策方程分别做二阶扰动展开，我们有

$$\begin{aligned}c_t &= c(k_t, z_t; \sigma) \\ &= \underbrace{\bar{c}}_{0 \text{ 阶}} + \underbrace{(c_k(k_t - \bar{k}) + c_z z_t + c_\sigma \sigma)}_{1 \text{ 阶}} \\ &\quad + \underbrace{\frac{1}{2} c_{kk} (k_t - \bar{k})^2 + c_{kz}(k_t - \bar{k})z_t + c_{k\sigma}(k_t - \bar{k})\sigma + \frac{1}{2} c_{zz} (z_t - \bar{z})^2 + c_{z\sigma} z_t \sigma + \frac{1}{2} c_{\sigma\sigma} \sigma^2}_{2 \text{ 阶}}.\end{aligned}\quad (11.30)$$

$$\begin{aligned}k_{t+1} &= k(k_t, z_t; \sigma) \\ &= \underbrace{\bar{k}}_{0 \text{ 阶}} + \underbrace{(k_k(k_t - \bar{k}) + k_z z_t + k_\sigma \sigma)}_{1 \text{ 阶}} \\ &\quad + \underbrace{\frac{1}{2} k_{kk} (k_t - \bar{k})^2 + k_{kz}(k_t - \bar{k})z_t + k_{k\sigma}(k_t - \bar{k})\sigma + \frac{1}{2} k_{zz} (z_t - \bar{z})^2 + k_{z\sigma} z_t \sigma + \frac{1}{2} k_{\sigma\sigma} \sigma^2}_{2 \text{ 阶}}.\end{aligned}\quad (11.31)$$

与一阶近似线性展开相比，二阶展开项中包含关于经济系统的更多信息，是一阶近似所无法分析的，举例说明。

1. 风险纠正。 $c_{\sigma\sigma}\sigma^2, k_{\sigma\sigma}\sigma^2$  对应风险纠正 (correction for risk)，描述在面对不确定性时，经济系统的稳定能力。
2. 从对称到不对称。在一阶近似中， $c_z z_t$  和  $k_z z_t$  项暗示着，对系统施加同等程度的一个正冲击和一个负冲击，产生的经济效果互为镜像。正因如此，基于一阶线性近似的 DSGE 研究常常只报告正负冲击之一的冲击响应 (IRFs)：另一个冲击只有正负符号的不同。

而二阶近似则不同。交互项  $c_{kz}(k_t - \bar{k})z_t$  和  $k_{kz}(k_t - \bar{k})z_t$  意味着，外部冲击的经济效果不只取决于外生随机冲击程度的大小  $z_t$ ，更与当期资本存量  $k_t$  有关，而后者是一个内生变量。经验研究表明，当衡量如一个金融部门外生冲击的效果时，很可能也需要考虑到同期经济体中的财产存量。

为了求解(11.30)-(11.31)，我们需要得到两组值。一组是变量的稳定状态，已于(11.29)得到。另一组是决策方程的一阶和二阶导数，需要我们回到非线性均衡条件系统(11.28)，改写为矩阵形式

$$\mathcal{H}(k_t, z_t; \sigma) = \begin{bmatrix} \frac{1}{c(k_t, z_t; \sigma)} - \beta E_t \frac{\alpha \exp(\rho z_t + \eta \varepsilon_{t+1}) k(k_t, z_t; \sigma)^{\alpha-1}}{c(k(k_t, z_t; \sigma), \rho z_t + \eta \varepsilon_{t+1})} \\ c(k_t, z_t; \sigma) + k(k_t, z_t; \sigma) - \exp(z_t) \cdot k_t^\alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad (11.32)$$

<sup>9</sup>对于  $z_t = 0 \forall t$  的情况，可能由以下两个因素之一导致：初始状态  $z_0 = 0$ ，或者随着  $t \rightarrow \infty$ ,  $z_t \rightarrow 0$ 。

或者代入近似系统的运算符  $F(\cdot)$ , 定义为关于  $\mathcal{H}(\cdot)$  的紧凑形式

$$F(k_t, z_t; \sigma) = \mathcal{H} \left( \underbrace{c(k_t, z_t; \sigma)}_1, \underbrace{c(k(k_t, z_t; \sigma), z_{t+1}; \sigma)}_2, \underbrace{k_t}_3, \underbrace{k(k_t, z_t; \sigma)}_4, \underbrace{z_t}_5; \underbrace{\sigma}_6 \right), \quad (11.33)$$

其中下角标编号 1 至 6 依次表示运算符  $\mathcal{H}$  中的元素。

### 一阶展开

围绕稳态  $(\bar{k}, 0; 0)$ , 对非线性系统  $F(k_t, z_t; \sigma)$  求一阶偏导

$$\begin{aligned} F_k &= \mathcal{H}_1 c_k + \mathcal{H}_2 c_k k_k + \mathcal{H}_3 + \mathcal{H}_4 k_k = 0, \\ F_z &= \mathcal{H}_1 c_z + \mathcal{H}_2 (c_k k_z + c_z \rho) + \mathcal{H}_4 k_z + \mathcal{H}_5 = 0, \\ F_\sigma &= \mathcal{H}_1 c_\sigma + \mathcal{H}_2 (c_k k_\sigma + c_\sigma) + \mathcal{H}_4 k_\sigma + \mathcal{H}_6 = 0, \end{aligned}$$

其中  $\mathcal{H}_i, i = 1, 2 \dots 6$  表示  $\mathcal{H}$  对第  $i$  个元素求偏导。

这构成了 6 个方程组成的系统<sup>10</sup>, 对应 6 个系数  $\mathcal{H}_i, i = 1, 2 \dots 6$ , 6 个待求解未知数  $c_k, c_z, k_k, k_z, c_\sigma, k_\sigma$ 。

求解过程分三步。第一步, 求  $c_k, k_k$ 。提取  $F_k = 0$  和  $F_z = 0$  对应的 4 个等式构成一个子系统, 写为二次矩阵形式

$$\begin{bmatrix} \mathcal{H}_1^1 \\ \mathcal{H}_1^2 \end{bmatrix} c_k + \begin{bmatrix} \mathcal{H}_2^1 \\ \mathcal{H}_2^2 \end{bmatrix} c_k k_k + \begin{bmatrix} \mathcal{H}_3^1 \\ \mathcal{H}_3^2 \end{bmatrix} + \begin{bmatrix} \mathcal{H}_4^1 \\ \mathcal{H}_4^2 \end{bmatrix} k_k = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

由(11.32)的上半部分和下半部分可得,  $\mathcal{H}_3^1 = 0, \mathcal{H}_2^2 = 0$ , 上式重写为

$$\begin{bmatrix} \mathcal{H}_1^1 \\ \mathcal{H}_1^2 \end{bmatrix} c_k + \begin{bmatrix} \mathcal{H}_2^1 \\ 0 \end{bmatrix} c_k k_k + \begin{bmatrix} 0 \\ \mathcal{H}_3^2 \end{bmatrix} + \begin{bmatrix} \mathcal{H}_4^1 \\ \mathcal{H}_4^2 \end{bmatrix} k_k = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad (11.34)$$

(11.34)中, 提取下半部分

$$\begin{aligned} \mathcal{H}_1^2 c_k + \mathcal{H}_3^2 + \mathcal{H}_4^2 k_k &= 0, X \\ \Rightarrow c_k &= \frac{1}{\mathcal{H}_1^2} (\mathcal{H}_4^2 k_k) - \frac{\mathcal{H}_3^2}{\mathcal{H}_1^2}, \end{aligned}$$

代回上半部分, 替换其中的  $c_k$ , 我们有

$$\begin{aligned} \mathcal{H}_1^1 c_k + \mathcal{H}_2^1 c_k k_k + \mathcal{H}_4^1 k_k &= 0, \\ \Rightarrow -\frac{\mathcal{H}_2^1 \mathcal{H}_4^1}{\mathcal{H}_1^2} k_k + \left( \mathcal{H}_4^2 - \frac{\mathcal{H}_1^1 \mathcal{H}_4^2 + \mathcal{H}_2^1 \mathcal{H}_3^2}{\mathcal{H}_1^2} \right) k_k - \frac{\mathcal{H}_1^1 \mathcal{H}_3^2}{\mathcal{H}_1^2} &= 0. \end{aligned}$$

上式成为一个二次矩阵形式  $AP^2 - BP - C = 0$ , 对应  $k_k = P$ 。采用 Uhlig (1999) 的方法 (见 Remark 11.4), 我们可以求得  $k_k$  的值, 进而  $c_k$ 。

此外需要指出的是, 二次矩阵系统的  $P$  值有两个, 对应两种情况。

1.  $k_k > 1$ , 此时的系统解是不稳定解。回顾一下资本的决策方程:

$$k_{t+1} = \bar{k} + k_k (k_t - \bar{k}) \dots,$$

很显然如果  $k_k > 1$ , 那么  $t$  期变量与稳态的偏离程度  $k_t - \bar{k}$ , 会导致  $t+1$  期变量距稳态更大程度的偏离。

<sup>10</sup>F 有 2 个维度, 3 个一阶导数对应 6 个方程。

2.  $0 < k_k < 1$ , 此时的系统解是稳定解,  $t$  期变量距稳态的偏离, 在不考虑其他冲击影响的情况下, 会在随后时间里逐渐收敛至稳态。

第二步, 将求得的  $c_k$  和  $k_k$  代入  $F_z = 0$  的 2 个方程中, 重复上一步的方法, 求得  $c_z$  和  $k_z$  的值。需要指出的是, 由  $F_z = 0$  这 2 个方程所构成的子系统是线性的。

第三步, 将求得的  $c_k, k_k, c_z, k_z$  代入 2 个方程  $F_\sigma = 0$  中, 构成一个线性齐次的关于  $c_\sigma$  和  $k_\sigma$  组成的系统。因此, 系统解为  $c_\sigma = 0$  和  $k_\sigma = 0$ 。进而, 我们得到一阶扰动法线性近似下的确定性等价条件。

## 二阶展开

下面, 对  $F(k_t, z_t; \sigma)$  围绕稳态  $(\bar{k}, 0; 0)$  做二阶线性近似, 得到一组  $6 \times 2 = 12$  个方程组成的系统

$$\begin{cases} F_{kk} = 0, \\ F_{kz} = F_{zk} = 0, F_{k\sigma} = F\sigma k = 0, \\ F_{zz} = 0, \\ F_{z\sigma} = F_{\sigma z} = 0, \\ F_{\sigma\sigma} = 0, \end{cases} \quad (11.35)$$

其中的系数部分包括:

- 稳态值:  $(\bar{c}, \bar{k})$ , 已经求得。
- 一阶导数的系数, 包括两部分:
  - $(c_k, k_k, c_z, k_z)$ , 已在前文中做一阶展开部分求得, 是已知数。
  - $c_\sigma = k_\sigma = 0$ , 确定性等价条件。
- 二阶导数的系数, 包括两部分:
  - $(c_{k\sigma}, k_{k\sigma}, c_{z\sigma}, k_{k\sigma}, )$ , 所有二次求导中含有  $\sigma$  的交互项, 它们的值都为 0;
  - $(c_{kk}, k_{kk}, c_{kz}, k_{kz}, c_{zz}, k_{zz}, c_{\sigma\sigma}, k_{\sigma\sigma})$ , 是待求解系数。

基于上述分析, 将一系列等于 0 的条件代回消费和资本的最优决策式(11.30)-(11.31), 我们有近似的最优决策式:

$$\begin{aligned} c_t = & \bar{c} + c_k (k_t - \bar{k}) + c_z z_t \\ & + \frac{1}{2} c_{kk} (k_t - \bar{k})^2 + c_{kz} (k_t - \bar{k}) z_t + \frac{1}{2} c_{zz} z_t^2 + \frac{1}{2} c_{\sigma\sigma} \sigma^2, \end{aligned} \quad (11.36)$$

$$\begin{aligned} k_{t+1} = & \bar{k} + k_k (k_t - \bar{k}) + k_z z_t \\ & + \frac{1}{2} k_{kk} (k_t - \bar{k})^2 + k_{kz} (k_t - \bar{k}) z_t + \frac{1}{2} k_{zz} z_t^2 + \frac{1}{2} k_{\sigma\sigma} \sigma^2. \end{aligned} \quad (11.37)$$

## 数值解

对于哪怕一个基准的随机经济增长模型, 利用扰动法得到的近似系统(11.36)-(11.37)也已经较为复杂, 难以表示为解析形式了。因此我们求助于数值方法。下面分几个步骤分别介绍。第一步, 参数校准。采用文献中常见的校准值。

- $\delta = 1$ 。折旧系数。如前文所述，这使得模型得以简化，但也因此使得模型经济与任何实际经济有较大不同。
- $\beta = 0.99$ 。时间贴现。
- $\alpha = 0.33$ 。资本的产出弹性。
- $\rho = 0.95$ 。外生生产率自回归过程中的持续系数 (persistence parameter)。
- $\eta = 0.01$ 。生产率自回归过程中，波动的标准差。

第二步，计算稳定状态。将校准参数值代入(11.29)中得  $\bar{c} = 0.388$ ,  $\bar{k} = 0.188$ 。

第三步，计算一阶、二阶扰动解。

- 一阶扰动解。 $c_k = 0.680$ ,  $c_z = 0.388$ ,  $k_k = 0.330$ ,  $k_z = 0.188$ 。
- 二阶扰动解。 $c_{kk} = -2.420$ ,  $c_{kz} = 0.680$ ,  $c_{zz} = 0.388$ ,  $c_{\sigma\sigma} = 0$ ,  $k_{kk} = -1.174$ ,  $k_{kz} = 0.330$ ,  $k_{zz} = 0.188$ ,  $k_{\sigma\sigma} = 0$ 。其中  $c_{\sigma\sigma} = k_{\sigma\sigma} = 0$  的情况，我们将在下文中做进一步说明，见 [reference](#)
- 此外我们有  $c_\sigma = k_\sigma = c_{k\sigma} = k_{k\sigma} = c_{z\sigma} = k_{z\sigma} = 0$  (理论模型中的确定性等价条件)。

第四步，关于消费和资本存量的近似决策式可由(11.36)-(11.37)求得。

**Remark 11.7** (风险纠正系数的讨论). 数值模拟过程中，两个风险纠正的系数值都是 0，对应  $(1/2)c_{\sigma\sigma}\sigma^2 = (1/2)k_{\sigma\sigma}\sigma^2 = 0$ 。这个看似奇怪的结果是和标准的新古典主义经济增长模型的特性有关的：模型中的风险主要来自外生技术冲击造成的生产过程风险。面对这种风险，一方面从局部部门来说，典型家庭积累的资本越多，他所面对的风险就越大，另一方面从整体来说，经济体中唯一可以用作净储蓄的财产就是资本。因此，任何风险的增大 (即技术冲击的标准差的增大) 产生了两种相互对冲的效果：一方面尽量减缓资本积累以免受到明日生产过程风险的损害，另一方面尽量增加储蓄 (资本积累) 以应付未来可能出现的负面冲击。

两种效果对冲后对经济运行的影响，分两种情况来分析。第一种情况，如果构建经济模型时设定较低的风险厌恶度，对应 CRRA 效用函数如对数形式，则二者完全抵消，系统解只收到技术冲击的波动  $z_t$  影响，而不受到波动的标准差的影响。第二种情况，当风险厌恶度较高，或者存在着多种财产储蓄形式时——如模型中本国典型家庭还有另一种储蓄选择：购买外国债券，并且外国债券的收益与本国技术冲击之间并非完全替代关系——则风险纠正项便很可能是个显著不为 0 的值了。

第五步，比较解析决策式与近似决策式。一个自然出现的问题就是：采用扰动法近似的决策式 (近似系统解)，在多大程度上贴近原系统的解？我们用解析决策式来代表原系统，可得

$$\begin{aligned} c_t &= 0.6734 \exp(z_t) k_t^{0.33}, \\ k_{t+1} &= 0.327 \exp(z_t) k_t^{0.33}. \end{aligned}$$

画出解析决策式、一阶扰动近似决策式、二阶扰动近似决策式，不难看出：

1. 在稳态 ( $\bar{k} = 0.188$ ) 附近，一阶近似的解几乎与解析解完全一致，但当  $k_t$  越是远离  $\bar{k}$ ，二者的差异越明显：当  $k_t = 0.1412$  时，偏差约 1%。
2. 比较起来，二阶近似解在全局内，与解析解相比更加贴近，如  $k_t = 0.1412$  时，偏差约 0.13%。

这表明，扰动法，尤其是高阶扰动法提供的近似解，具有更好的全局特性。

**Remark 11.8** (扰动近似解的精确度检验). 0.13% 的偏差程度已经足够精确了，还是仍然太高？问题的答案取决于实际研究工作中的具体情况。例如，在商业周期研究中的精确度要求，通常低于关于福利水平的研究。其原因在于，在近似计算一系列矩的方差时（如消费的均值和方差等），随着矩的不同，矩的误差也有所不同，各项误差之间有时会相互抵消一部分。而福利水平常常是一个关于财富分配的非线性方程，在近似计算分配状况时，一点微小的误差可能会引起福利水平近似值的较大变化，从而偏离实际值。

## 11.4 剪枝算法

如前文所述，通常来讲，越高阶的扰动近似解，与原非线性系统的贴近程度越高。但高阶扰动法带来的一个附属问题是，在实际研究中，尽管高阶扰动近似求得的线性决策方程的确是稳定的，但根据这个决策方程所生成的序列数据却有可能是不稳定的（爆炸的）。其原因在于，高阶扰动项在系统中加入了额外的不动点，从而使得围绕这些不动点所做的近似解变得不稳定了（Kim et al., 2008; Den Haan and De Wind, 2012）。

举例说明。假设某一资本的近似决策方程可表示为

$$k_{t+1} = a_0 + a_1 k_t + a_2 k_t^2 + \dots + b_1 \varepsilon_t + \dots,$$

为简化模型考虑，不列出技术外生冲击  $z_t$  的情形。以递归形式表述，上式可以改写为

$$k_{t+1} = a_0 + a_1 (a_0 + a_1 k_{t-1} + a_2 k_{t-1}^2) + a_2 (a_0 + a_1 k_{t-1} + a_2 k_{t-1}^2)^2 + \dots + b_1 \varepsilon_t + \dots$$

上式中涉及到  $k_{t-1}$  的 3 次甚至 4 次方。因此在仿真过程中，随着  $t$  越来越大，所生成的时间序列数据就越容易出现  $(k_{t+1} - \bar{k}) \rightarrow \infty$  的情况，即产生不稳定的爆炸路径。在这种情况下，GMM、SMM 等无条件的矩匹配估计方法也不再适用了<sup>11</sup>，原因在于它们需要满足一系列条件，如矩是平稳且有极限的，并且是非周期的遍历（ergodic）概率分布。

鉴于这种情况，Kim et al. (2008) 提出要对近似解做剪枝（pruning），即是说，在递归形式中去掉所有比近似解更高阶的项——以本例中的二阶扰动近似法为例，就是要删除一切高于二阶的状态项和/或扰动项。Kim et al. (2008) 证明由剪枝后的近似解所生成的时间序列数据，就不再出现爆炸路径的情况。

Andreasen et al. (2016) 对 Kim et al. (2008) 的剪枝法做了进一步扩展，应用到了任意高的阶数的情况。他们首先证明一个剪枝状态——空间系统中的确存在一阶以及二阶无条件矩，进而提供了关于这些一、二阶矩与冲击响应方程（IRFs）的闭合表达式（closed-form expressions）。其研究价值在于，可以省去研究者计算数值并进行仿真的工作：这些数据仿真已经被不少 DSGE 的扩展 IRFs 研究证明是不可靠的了<sup>12</sup>。此外，Andreasen et al. (2016) 还讨论了更高阶无条件矩阵（如偏度 skewness，峰度 kurtosis）等的存在条件。

## 11.5 变量的变换

上文中介绍了，在扰动法求解 DSGE 模型过程中，可以用对数形式的变量而非变量本身作为研究对象。Jin and Judd (2002) 介绍了对数线性化方法，作为一种变量的特殊变化形式，为什么要比变量本身更适合作为 DSGE 模型的研究对象，以及更通用的变量变换形式是怎么样的，如何

<sup>11</sup> 关于 GMM、SMM 等方法在 DSGE 研究中的应用，可见 Ruge-Murcia (2007)。

<sup>12</sup> 关于扩展 IRFs 的定义，可见如 Koop et al. (1996)。

更好的运用他们，更多讨论见Fernández-Villaverde and Rubio-Ramírez (2006)。这里简要介绍变量变换的通用方法及其用途。

已知对于  $d(x)$ ，围绕  $a$  点做 TSE

$$d(x) \approx d(a) + \frac{\partial d(a)}{\partial a}(x - a) + H.O.T.$$

其中 H.O.T. 表示高阶展开项。据此，也可以将变量从  $x$  转换为反函数  $Y(x)$ ，围绕  $b = Y(a)$  点做 TSE

$$g(y) = h(d(x(y))) = g(b) + \frac{\partial g(b)}{\partial b}(Y(x) - b) + H.O.T.$$

根据第一个等式，我们可以通过 TSE 求得关于状态  $x$  的近似方程  $d(\cdot)$ ，进而求解运算符  $\mathcal{H}(\cdot)$ 。那么同样地，在对  $x$  做变量变换后，根据第二个等式我们也能找到关于状态  $Y(x)$  的未知方程  $g(\cdot)$ ，进而求解运算符  $\mathcal{H}(\cdot)$ 。在进行近似线性化过程之前，变量变换工作首先将重心放在状态形式的选择上来，将原本是高度线性化的问题转化为接近线性的，从而提高扰动法求解的精度。

### 11.5.1 例 1

以上文提到的随机内生经济增长模型为例，求资本积累的近似决定式，假定满足如下一阶扰动形式

$$k_{t+1} = \bar{k} + a_1(k_t - \bar{k}) + b_1 z_t,$$

其中  $a, b$  系数通过对  $F(k_t, z_t; \sigma)$  求导得出，稳态  $\bar{k}$  也是已知。上式调整，化为总量形式

$$(k_{t+1} - \bar{k}) = a_1(k_t - \bar{k}) + b_1 z_t. \quad (11.38)$$

类似地，对(11.38)做对数线性化

$$\begin{aligned} \log k_{t+1} - \log \bar{k} &= a_2 (\log k_t - \log \bar{k}) + b_2 \cdot z_t, \\ \Rightarrow \hat{k}_{t+1} &= a_2 \hat{k}_t + b_2 z_t, \quad \hat{x}_t \equiv \log x_t - \log x_0. \end{aligned} \quad (11.39)$$

**Proposition 11.1.** (11.38)和(11.39)等价。

证明. 分三步予以证明。第一，已知变量的线性近似形式

$$k_{t+1} = d(k_t, z_t; \sigma) = d(\bar{k}, 0; 0) + d_1(\bar{k}, 0; 0)(k_t - \bar{k}) + d_2(\bar{k}, 0; 0)z_t,$$

则我们有

$$d_1(\bar{k}, 0; 0) = a_1, \quad d_2(\bar{k}, 0; 0) = b_1.$$

第二步，引入变量的变化，设  $h = \log d$ ，其中  $Y(x) = \log x$ 。则根据Judd and Guu (2003) 我们有

$$\log k_{t+1} - \log \bar{k} = d_1(\bar{k}, 0; 0)(\log k_t - \log \bar{k}) + \frac{1}{\bar{k}} d_2(\bar{k}, 0; 0)z_t.$$

第三步，比较上面两个等式，系数相等，我们有

$$a_2 = a_1, \quad b_2 = \frac{1}{\bar{k}} b_1.$$

□

这个例子中有 3 点值得注意。

1. 从  $(a_1, b_1)$  到  $(a_2, b_2)$  的计算涉及到  $\bar{k}$ , 而  $\bar{k}$  已经在对水平变量一阶扰动法求解中求得。因此, 当我们求得了一节线性系统的近似解后, 可以很方便得到变量变换(对数线性化)后的系统近似解。
2. 对于含有常规设定(如效用函数、生产函数等)的随机内生经济增长模型而言, 通常来说无需引入其他额外假设, 即可进行变量变换如对数线性化。
3. 变量变换可用于任何阶数的扰动近似。

### 11.5.2 例 2

我们来看一个更通用的例子。已知某一系统围绕稳态( $x = a$ )的一阶扰动近似解为

$$d(x) \approx d(a) + \frac{\partial d(a)}{\partial a}(x - a),$$

那么对于  $Y(x)$  的反函数  $X(y)$ , 围绕  $b = Y(a)$ , 对变量变换后的新系统  $g(y) = h(d(X(y)))$  做一阶扰动, 我们有

$$g(y) = h(d(X(y))) = g(b) + g_\alpha(b)(Y(x)^\alpha - b^\alpha),$$

其中  $g_\alpha = [h]_A [d_i]^A [x]_\alpha^i$  为张量形式, 可由微积分运算中的链式法则求得。

根据 Judd and Guu (2003) 的方法, 我们可以把任一幂的近似式写为

$$k_{t+1}(k_t, z_t; \gamma, \zeta, \varphi)^\gamma - \bar{k}^\gamma = a_3 \left( k_t^\zeta - \bar{k}^\zeta \right) + b_3 z_t^\varphi,$$

其中设  $\varphi \geq 1$  以确保  $z_t^\varphi$  是实数。

幂形式方程的优点在于, 我们只需要 3 个参数值  $\gamma, \zeta, \varphi$  就可以描述许多种非线性结构的系统。并且对于  $\gamma \rightarrow 0, \zeta \rightarrow 0, \varphi \rightarrow 1$  的情况, 该方程求极限就变成了对数转换。此时作变量转换  $h = d^\gamma, y = x^\gamma, x = y^{\frac{1}{\zeta}}$ 。此外, 由于

$$k_{t+1}(k_t, z_t; \gamma, \zeta, \varphi)^\gamma - \bar{k}^\gamma = \frac{\gamma}{\zeta} \bar{k}^{\gamma-\zeta} a_1 \left( k_t^\zeta - \bar{k}^\zeta \right) + \frac{\gamma}{\varphi} \bar{k}^{\gamma-1} z_t^\varphi,$$

于是我们有

$$a_3 = \frac{\gamma}{\zeta} \bar{k}^{\gamma-\zeta} a_1, \quad b_3 = \frac{\gamma}{\varphi} \bar{k}^{\gamma-1} b_1.$$

对于  $\gamma = \zeta, \varphi = 1$  的情况, 系统中只有一个自由系数, 则

$$k_{t+1}(k_t, z_t; \gamma)^\gamma - \bar{k}^\gamma = a_4 \left( k_t^\zeta - \bar{k}^\zeta \right) + b_4 z_t,$$

定义  $\tilde{k}_{t+1} \equiv k_t^\gamma - \bar{k}^\gamma$ , 我们于是有

$$\tilde{k}_{t+1} = a_4 \tilde{k}_t + b_4 z_t, \tag{11.40}$$

因此

$$a_4 = a_1, \quad b_4 = \bar{k}^{\gamma-1} b_1.$$

通过变量变换, 将原本是非线性的方程系统转换为线性形式(11.40), 这有助于我们展开后续分析, 以及采用卡曼滤波对模型做进一步的估计。

### 11.5.3 最优变量变换：参数的选取

我们介绍了如何通过几个幂参数值的设定，对变量做变换，将 DSGE 模型用更通用的一阶线性形式表现出来。随后的问题就变成了，如何选取最合适的幂参数。参数取值可以遵循以下两个策略之一。第一是最优策略，致力于追求最高的精确度，基于似然方程等方法的经验研究，目标的确为追求最大化解，对应最高的精确度，但这往往导致计算成本过于高昂。第二是次优策略，将参数取值向最优方向做适度改进，以追求模型精度的适度提高。实际经验研究中往往采取后者，在精度提升和计算成本控制之间寻求平衡点。

有鉴于此，Fernández-Villaverde and Rubio-Ramírez (2006) 提出一个合理判定原则，致力于选取合适的参数值来改变变量值，通过尽可能减少 Euler 方程的误差来提高模型解的精确度<sup>13</sup>。

### 11.5.4 对数线性和对数正态线性化

如前文所述，标准的对数线性近似法，常常是对差分变量取对数后做一阶扰动近似。金融学研究中有时也会遇到另一种对数正态线性化方法 (lognormal-linearization)，二者有所不同，容易引起混淆，有必要建模做简要说明。

假定经济模型中一个代表性家庭，效用函数和预算约束条件分别为

$$\max E_0 \sum_{t=0}^{\infty} \beta^t \log C_t,$$

$$W_{t+1} = R_{t+1}(W_t - C_t),$$

其中  $W_t$  为总财富，其初始水平  $W_0$  是给定的。

非线性均衡系统由一阶条件和预算约束条件组成：

$$\frac{1}{C_t} = \beta E_t \frac{1}{C_{t+1}} R_{t+1},$$

$$W_{t+1} = R_{t+1} (W_t - C_t).$$

因此稳定状态下我们有

$$R = \frac{1}{\beta},$$

$$W = \frac{1}{\beta} (W - C).$$

#### (对数偏移形式的) 对数线性化

定义  $\hat{x}_t = x_t - x = \log X_t - \log X$  表示对数变量  $x_t$  距离其稳态  $x$  的偏离。则对于 Euler 方程来说

$$\begin{aligned} \log C_{t+1} &= \log C_t + \log R_{t+1} + \log \beta, \\ \Rightarrow \log C + \left( \frac{C_{t+1} - C}{C} \right) &= \log C + \left( \frac{C_t - C}{C} \right) + \log R + \left( \frac{R_{t+1} - R}{R} \right), \\ \Rightarrow \hat{c}_{t+1} &= \hat{c}_t + \hat{r}_{t+1}, \\ \Rightarrow E_t \Delta \hat{c}_{t+1} &= E_t \hat{r}_{t+1}. \end{aligned}$$

---

<sup>13</sup> 此外他们还发现，最优参数值的选取还与经济模型外生冲击的标准差有关。这是一个重要发现：变量变换随着模型中不确定程度的变化而调整。这又一次印证了确定性等价条件的不适用。

对于预算约束条件来说

$$\begin{aligned}\log W_{t+1} &= \log R_{t+1} + \log(W_t - C_t), \\ \Rightarrow \log W + \left( \frac{W_{t+1} - W}{W} \right) &= \log R + \left( \frac{R_{t+1} - R}{R} \right) + \log(W - C) + \left( \frac{(W_t - W)(C_t - C)}{W - C} \right), \\ \Rightarrow E_t \hat{w}_{t+1} &= E_t \hat{r}_{t+1} + \frac{\hat{w}_t w - \hat{c}_t c}{w - c}, \\ \Rightarrow E_t \Delta \hat{w}_{t+1} &= E_t \hat{r}_{t+1} + \left( 1 - \frac{1}{\rho} \right) (\hat{c}_t - \hat{w}_t), \quad \rho \equiv \frac{w - c}{w}.\end{aligned}$$

### (对数形式的) 对数线性化

由上两式可得, 对数形式的线性化

$$E_t \Delta c_{t+1} = E_t r_{t+1} + \log \beta, \tag{11.41}$$

$$E_t \Delta w_{t+1} = E_t r_{t+1} + \left( 1 - \frac{1}{\rho} \right) (c_t - w_t) + \kappa, \quad \kappa \equiv -r - \left( 1 - \frac{1}{\rho} \right) (c - w). \tag{11.42}$$

### 对数正态线性化

对数正态线性化下的预算约束条件式与对数线性化下的(11.42)一致。但在 Euler 方程上的设定有所不同: 它假定变量  $\frac{C_t}{C_{t+1}} R_{t+1}$  是对数正态分布的。对于任何一个对数正态分布的变量我们有

$$\log E_t X_t = E_t \log X_t + \frac{1}{2} Var_t \log X_t.$$

那么回到 Euler 方程我们有

$$\begin{aligned}1 &= \beta E_t \left( \frac{C_t}{C_{t+1}} \right) R_{t+1}, \\ \Rightarrow 0 &= \log \beta + \log E_t \left( \frac{C_t}{C_{t+1}} \right) R_{t+1}, \\ \Rightarrow 0 &= \log \beta + E_t \log \left( \frac{C_t}{C_{t+1}} \right) R_{t+1} + \frac{1}{2} Var_t \log \left( \frac{C_t}{C_{t+1}} \right) R_{t+1}, \\ \Rightarrow E_t \Delta c_{t+1} &= \log \beta + E_t r_{t+1} + \underbrace{\frac{1}{2} [Var_t \Delta c_{t+1} + Var_t r_{t+1} - 2Cov(\Delta c_{t+1}, r_{t+1})]}_{\chi}.\end{aligned} \tag{11.43}$$

由上可见, 在对数正态线性化中, 对于不含有期望的项, 我们用标准对数线性化方法来处理; 对于含有期望的项, 我们假定它呈对数正态分布, 其结果是一个确定的值, 不涉及到泰勒级数展开的近似处理。

### 比较

比较(11.41)和(11.43)可见, 对数正态线性化中多出了一项  $\chi$ , 使得确定性等价条件不成立。这一新特性使对数正态线性化具有易于解释的优点。如对于一种资产  $i$  和定价核 (pricing kernel)  $M_t$  来说<sup>14</sup>, 则定价方程

$$\begin{aligned}1 &= E_t M_{t+1} R_{i,t+1}, \\ \Rightarrow 0 &= E_t \log(M_{t+1} R_{i,t+1}) + \frac{1}{2} Var_t \log(M_{t+1} R_{i,t+1}), \\ \Rightarrow E_t r_{i,t+1} &= -E_t m_{t+1} - \frac{1}{2} Var_t m_{t+1} - Cov(m_{t+1}, r_{i,t+1}) - \frac{1}{2} Var_t r_{i,t+1}.\end{aligned}$$

<sup>14</sup>关于定价核的介绍, 可见如 (Cochrane, 2000, Ch.1)。

此外对于无风险债券  $f$  来说，有

$$1 = E_t M_{t+1} R_{f,t+1}, \\ r_{f,t+1} = -E_t m_{t+1} - \frac{1}{2} Var_t m_{t+1}.$$

则两式相减我们可得回报

$$E_t r_{i,t+1} - r_{f,t+1} = -\frac{1}{2} Var_t r_{i,t+1} - Cov(m_{t+1}, r_{i,t+1}).$$

但对数正态线性化也存在一定不足：

1. 我们不清楚在一个一般均衡模型中，变量  $\frac{C_t}{C_{t+1}} R_{t+1}$  在多大程度上是接近对数正态分布的。
2. 对数正态线性化过程中同时使用了两种方法，分别是对数正态假设和对数线性化。从扰动法的角度来看，这种处理方法并非逻辑自洽，缺乏坚实的理论支撑，在经验研究中也无法确保呈收敛态势。
3. 对数线性化中我们可以通过求解二次矩阵系统来计算全部相关系数。而在对数正态线性化中，只能通过计算二阶矩的状态来估计系数；然而对许多研究来说，如何计算这些二阶距，这是个问题。
4. 对数线性化很容易做更高阶的扰动近似，但对数正态线性化则很难做高阶近似。

## 11.6 价值方程迭代中的扰动法

某些情况下我们需要在 DSGE 模型中对价值方程做扰动，如当偏好呈现出递归特征时，或者当我们需要对福利效果做评估时。这个扰动的价值方程，可以用作很好的初始值猜测，供我们展开随后的价值方程迭代——如前文所述，在一些高维问题中，如果缺乏合适的初始值设定，跌倒过程的结果可能导致收敛过慢，甚至不收敛。这里举例说明对将扰动法应用到价值方程迭代过程中去。

考虑这样一个价值方程问题<sup>15</sup>：

$$V(k_t, z_t) = \max_{c_t} [(1 - \beta) \log c_t + \beta E_t V(k_{t+1}, z_{t+1})],$$

约束条件及外部冲击过程

$$c_t + k_{t+1} = \exp(z_t) k_t^\alpha + (1 - \delta) k_t, \\ z_t = \rho z_{t-1} + \eta \varepsilon_t, \quad \varepsilon_t \sim i.i.d \mathcal{N}(0, 1).$$

引入扰动参数  $\sigma$ ，将上式改写为

$$V(k_t, z_t; \sigma) = \max_{c_t} [\log c_t + \beta E_t V(\exp(z_t) k_t^\alpha + (1 - \delta) k_t - c_t, \rho z_t + \sigma \eta \varepsilon_{t+1}; \sigma)],$$

围绕确定性稳态  $(\bar{k}, 0; 0)$  做二阶扰动近似，该问题的解由两部分构成：1 个价值方程  $V(k_t, z_t; \sigma)$  和 1 个消费的策略方程  $c(k_t, z_t; \sigma)$ 。

先来看价值方程，

$$V(k_t, z_t; \sigma) = V_{ss} + V_{1,ss}(k_t - \bar{k}) + V_{2,ss}z_t + V_{3,ss}\sigma \\ + \frac{1}{2} V_{11,ss}(k_t - \bar{k})^2 + V_{12,ss}(k_t - \bar{k})z_t + V_{13,ss}(k_t - \bar{k})\sigma \\ + \frac{1}{2} V_{22,ss}z_t^2 + V_{23,ss}z_t\sigma + \frac{1}{2} V_{33,ss}\sigma^2,$$

<sup>15</sup>出于简化模型的考虑，上式中将  $\log c_t$  标准化（乘以  $(1 - \beta)$ ），这样价值方程的稳态  $V_{ss} = \log c$ ，其中  $c$  对应稳态消费水平。通过标准化，价值方程和模型的效用函数的单位相同。

其中

$$\begin{aligned} V_{ss} &= V(\bar{k}, 0; 0), \\ V_{i,ss} &= V_i(\bar{k}, 0; 0), \quad i = 1, 2, 3, \\ V_{ij,ss} &= V_{ij}(\bar{k}, 0; 0), \quad i, j = 1, 2, 3, \\ V_{3,ss} &= V_{13,ss} = V_{23,ss} = 0, \quad \text{确定性等价条件.} \end{aligned}$$

由此价值方程的二阶扰动近似可以表示为

$$\begin{aligned} V(k_t, z_t; \sigma) &= V_{ss} + V_{1,ss}(k_t - \bar{k}) + V_{2,ss}z_t \\ &\quad + \frac{1}{2}V_{11,ss}(k_t - \bar{k})^2 + V_{12,ss}(k_t - \bar{k})z_t + \frac{1}{2}V_{22,ss}z_t^2 + \frac{1}{2}V_{33,ss}\sigma^2, \end{aligned} \tag{11.44}$$

其中  $V_{33,ss} \neq 0$ , 这与前面介绍随机新古典主义经济增长模型中 Remark 11.7 中, 对效用函数做扩展 Schur 分解 (LQ 近似) 的情况不同。

类似地, 消费的策略方程

$$c_t = c(k_t, z_t; \sigma) = c_{ss} + c_{1,ss}(k_t - \bar{k}) + c_{2,ss}z_t + c_{3,ss}\sigma, \tag{11.45}$$

其中

$$\begin{aligned} c_{ss} &= c(\bar{k}, 0; 0), \\ c_{i,ss} &= c_i(\bar{k}, 0; 0), \quad i = 1, 2, 3, \\ c_{3,ss} &= 0. \end{aligned}$$

$c_{3,ss} = 0$  是由于, 效用函数的一阶导数只与  $V_{1,2}$  有关<sup>16</sup>。

对(11.44)和(11.45)构成的系统求解, 遵循常规算法:

- 对线性项的系数, 用  $V_{ss}$  依次对控制  $c_t$ , 状态  $k_t, z_t$  和扰动参数  $\sigma$  求导, 随后在  $\sigma = 0$  的情况下求解。
- 对二次项的系数, 用  $V_{ss}$  做二次求导, 引入上一环节求得的一阶系数, 然后在  $\sigma = 0$  的情况下求解。

对价值方程做扰动近似, 有如下若干有点

1. 可以测度经济周期波动产生的福利成本。在稳态  $(k_t, z_t) = (\bar{k}, 0)$  下我们有

$$V(\bar{k}, 0; 0) = V_{ss} + \underbrace{\frac{1}{2}V_{33,ss}\sigma^2}_{\text{福利成本}},$$

划线部分反映了在二阶层面上出现的福利成本, 它是二者之差: 一个是  $V_{ss} = V(\bar{k}, 0)$ , 即在  $(\bar{k}, 0)$  的稳态下所测量的价值方程, 此时我们知道正处于稳定状态。一个是  $V(\bar{k}, 0; 0)$  的均衡状态下所测量的价值方程, 此刻我们掌握更多信息, 不只知道我们当前正处在稳定状态, 更知道未来也会继续处在这一点上。

此外福利成本  $V_{33,ss}$  也并非在全部情况下都是负数。在一些情况下它甚至可能为正, 如一个包括休闲决策的随机内生经济增长模型。进一步的模型描述及经验验证见 Cho et al. (2015)。

---

<sup>16</sup>但也并不绝对, 比如当模型中存在预防性消费 (precautionary consumption) 时, 效用函数的一阶导数也需要考虑  $V_3$ , 相关讨论见 Kimball (1990)。

2. 方便将福利成本  $V_{33,ss}$  转换为消费单位, 从经济学意义上来说更有意义。具体说来, 定义系数  $\tau$  反映消费减少的量, 使得家庭部门对于以下 2 个选项是无差异的: 1 是确定性条件下消费  $(1 - \tau)c_t$  个单位, 1 是在不确定性条件下消费  $c_t$ , 即  $\tau$  满足

$$\log(1 - \tau)c = \log c + \frac{1}{2}V_{33,ss}\sigma^2,$$

其中使用到了  $V_{ss} = \log c$  的条件。整理得  $\tau$  的决定

$$\tau = 1 - \exp\left(\frac{1}{2}V_{33,ss}\sigma^2\right) \quad (11.46)$$

数值解。采用与第11.3.8节一样的参数校准值, 我们有

$$\begin{aligned} V &= -0.540 + 0.026(k_t - 0.188) + 0.250z_t - 0.069(k_t - 0.188)^2, \\ c_t &= 0.388 + 0.680(k_t - 0.188) + 0.388z_t. \end{aligned} \quad (11.47)$$

其中第一, 价值方程迭代的消费政策方程, 与模型均衡条件下的扰动近似消费决策方程相同。第二,  $V_{kz} = V_{zz} = V_{\sigma\sigma} = 0$ , 针对这一组校准参数, 经济周期波动的福利成本是 0<sup>17</sup>。

初始值猜测。利用(11.47)生成的初始值  $V_0$  做初始猜测值做后续价值方程迭代, 可以较快出现收敛。

或者采取混合策略。将两种方法共同作用于运算符  $\mathcal{H}$ , 一个是模型均衡条件, 一个是根据最优决策方程而估算的价值方程  $V(k_t, z_t) = (1 - \beta)\log c_t + \beta E_t V(k_{t+1}, z_{t+1})$ 。混合策略有助于以较低的计算成本对价值方程和决策方程作出近似<sup>18</sup>。

---

<sup>17</sup>均衡条件下消费决策方程  $c_t = 0.673 \exp(z_t)k_t^{0.33}$ , 对应当期效用  $u_t = \log c_t = z_t + \log 0.673 + 0.33 \log k_t$ 。其中  $z_t$  的无条件均值是 0, 以及  $\log k_t$  的决策方程满足确定性等价条件, 因此改变  $z_t$  的方差不会产生(无条件的)福利成本。

<sup>18</sup>也可以将价值方程的若干导数对方在一起, 如  $(1 - \beta)c_t^{-1} - \beta E_t V_{1,t+1} = 0$  等, 进而寻找这个(由价值方程导数构成的)方程系统的扰动近似解。这种方法可能会对寻找价值方程的更高阶近似有帮助。



# 附录

## 11.A Schur 补和对称正(半)定矩阵

### 11.A.1 Schur 补

本节作为 (Boyd and Vandenberghe, 2004, A.5.5.5) 的详细说明。假定矩阵  $M \in \mathbb{R}^{n \times n}$  可做如下分解

$$M_{n \times n} = \begin{bmatrix} A_{p \times p} & B_{p \times q} \\ C_{q \times p} & D_{q \times q} \end{bmatrix}, \quad n = p + q, \quad p, q, n > 0.$$

Schur 补的思路在于，通过去掉一个变量分块求解线性方程。假定一个线性系统  $(x, y)$

$$Ax + By = c, \tag{11.48a}$$

$$Cx + Dy = d, \tag{11.48b}$$

改写为矩阵形式

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} c \\ d \end{bmatrix}. \tag{11.49}$$

可以利用高斯消元法 (Gaussian elimination) 求得系统解。

### A 的 Schur 补以及系统解

假定  $D$  是可逆的，则(11.48b)改写为

$$y = D^{-1}(d - Cx),$$

代回(11.48a)替代  $y$  得

$$(A - BD^{-1}C)x = c - BD^{-1}d.$$

进一步假定  $(A - BD^{-1}C)$  是可逆的，我们有系统解

$$x = \underbrace{(A - BD^{-1}C)}_{A \text{ 的 Schur 补}}^{-1} (c - BD^{-1}d), \tag{11.50a}$$

$$y = D^{-1}d - D^{-1}C \left( \underbrace{(A - BD^{-1}C)}_{A \text{ 的 Schur 补}}^{-1} (c - BD^{-1}d) \right). \tag{11.50b}$$

我们将上式中标记出的  $(A - BD^{-1}C)$  定义为  $M$  中分块矩阵  $A$  的 Schur 补。

采用  $A$  的 Schur 补, 我们有

$$\begin{aligned} x &= (A - BD^{-1}C)^{-1}c - (A - BD^{-1}C)^{-1}BD^{-1}d, \\ y &= -D^{-1}C(A - BD^{-1}C)^{-1}c + \left(D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1}\right)d, \end{aligned}$$

改写为矩阵形式

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & - (A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & \left(D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1}\right) \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix}, \quad (11.51)$$

与(11.49)联立不难看出,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & - (A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & \left(D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1}\right) \end{bmatrix} \quad (11.52)$$

对式右侧做 matrix reflection 可得

$$\begin{aligned} &\begin{bmatrix} (A - BD^{-1}C)^{-1} & - (A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & \left(D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1}\right) \end{bmatrix} \\ &= \begin{bmatrix} (A - BD^{-1}C)^{-1} & 0 \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} \end{bmatrix} \begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix} \\ &= \begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix} \begin{bmatrix} (A - BD^{-1}C)^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix} \begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix}, \end{aligned}$$

并且第三行三个矩阵中, 上下三角矩阵都是可逆的; 中间对角矩阵对角元素也是可逆的。

代回(11.52)右侧, 且等式两侧同时求逆运算, 可以得到  $M$  的解

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix}. \quad (11.53)$$

即, 利用  $A$  的 Schur 补求解系统,  $(x, y)$  的值由(11.50a)-(11.50b)给出,  $M$  的值由(11.53)给出。这种算法的好处在于, 在假定  $D$  是可逆矩阵的情况下, 只需要求得  $A$  的 Schur 补, 即可求解系统。

#### D 的 Schur 补以及系统解

类似地, 若假定  $A$  是可逆的, (11.48a)改写为

$$x = A^{-1}(c - By),$$

代回(11.48b)替代  $x$  得

$$(D - CA^{-1}B)y = D - CA^{-1}c.$$

进一步假定  $(D - CA^{-1}B)$  是可逆的, 我们有系统解

$$y = \underbrace{(D - CA^{-1}B)}_{D \text{ 的 Schur 补}}(c - BD^{-1}d) \quad (11.54a)$$

$$y = A^{-1}c - A^{-1}B \left( \underbrace{\left( D - CA^{-1}B \right)^{-1}}_{D \text{ 的 Schur 补}} (d - CA^{-1}c) \right). \quad (11.54b)$$

我们将上式中标记出的  $(D - CA^{-1}B)$  定义为  $M$  中分块矩阵  $D$  的 Schur 补。

采用  $D$  的 Schur 补, 我们有

$$\begin{aligned} x &= \left( A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} \right) c - \left( A^{-1}B(D - CA^{-1}B)^{-1} \right) d, \\ y &= -(D - CA^{-1}B)^{-1}CA^{-1}c + (D - CA^{-1}B)^{-1}d, \end{aligned}$$

改写为矩阵形式

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix}, \quad (11.55)$$

与(11.49)联立可见

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix} \quad (11.56)$$

对式右侧做 matrix reflection 可得

$$\begin{aligned} &\begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix} \\ &= \begin{bmatrix} A^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ 0 & (D - CA^{-1}B)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \\ &= \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ 0 & (D - CA^{-1}B)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}, \end{aligned}$$

同样地, 第三行三个矩阵中, 上下三角矩阵都是可逆的; 中间对角矩阵对角元素也是可逆的。

代回(11.56)右侧, 且等式两侧同时求逆运算, 可以得到  $M$  的解

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & CA^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix} \begin{bmatrix} I & 0 \\ A^{-1}B & I \end{bmatrix}. \quad (11.57)$$

### 11.A.2 利用 Schur 分解法判断可逆正定矩阵

对于符合第10.8描述的可逆矩阵  $M$ , 假定其分块对角  $A, D$  都是可逆的, 并且  $C = B^T$ , 那么可以将其改写为如下类似于分块——对角矩阵的形式

$$\begin{aligned} M &= \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \\ &= \begin{bmatrix} I & BC^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A - BC^{-1}B^T & 0 \\ 0 & C \end{bmatrix} \begin{bmatrix} I & BC^{-1} \\ 0 & I \end{bmatrix}^T. \end{aligned} \quad (11.58)$$

其中第二行利用了  $A$  的 Schur 补。我们可以利用这一分块对角形式, 查看对称矩阵  $M$  是否是正定的, 即  $M \succ 0$ 。

**Proposition 11.2.** 对于任一对称矩阵  $M = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$ , 如果  $C$  是可逆的, 则

1. 当且仅当

$$\begin{cases} C \succ 0, \text{ 并且} \\ A - BC^{-1}B^T \succ 0 \end{cases}$$

时, 我们有  $M \succ 0$ 。

2. 如果  $C \succ 0$ , 那么当且仅当  $A - BC^{-1}B^\top \succeq 0$  时, 我们有  $M \succeq 0$ 。

证明. 略。  $\square$

**Proposition 11.3.** 对于任一对称矩阵  $M = \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}$ , 如果  $A$  是可逆的, 则

1. 当且仅当

$$\begin{cases} A \succ 0, \text{ 并且} \\ C - B^\top AB \succeq 0 \end{cases}$$

时, 我们有  $M \succ 0$ 。

2. 如果  $A \succ 0$ , 那么当且仅当  $C - B^\top AB \succeq 0$  时, 我们有  $M \succeq 0$ 。

证明. 略。  $\square$

### 11.A.3 利用 Schur 分解法判断可逆正半定矩阵

当  $C$  和  $A$  都是奇异分块矩阵, 我们无法直接求得  $C^{-1}$  和  $A^{-1}$ 。我们首先介绍判断方法, 随后用该方法判断  $M$  是否是正半定的。

#### 判定方法

1. (通常是 nonconvex) 的 quadratic 优化问题: 对于满足一定形式的  $x$  和 (非奇异) 的  $P$ , 式

$$\min_{\{x\}} f(x) = \frac{1}{2} x^\top Px + x^\top b$$

是否存在最小值, 以及最小值是多少, 见 Proposition 11.4。

2. 对于奇异的  $P$ , 我们要计算其对角分块的伪逆矩阵和相应的 Schur 补, 进而判断最小值, 见 Proposition 11.5。

**Proposition 11.4** (非奇异系数矩阵 (quadratic 优化问题)). 如果  $P$  是一个对称的非奇异矩阵, 那么只有当  $P \succeq 0$  时, 方程  $f(x) = (\frac{1}{2})x^\top Px + x^\top b$  才具有唯一的最优 (极小) 值  $x^* = P^{-1}b$ , 使得  $f(x^*) = f(P^{-1}b) = -\frac{1}{2}b^\top P^{-1}b$ 。

证明. 已知

$$\frac{1}{2}(x + P^{-1}b)^\top P(x + P^{-1}b) = \frac{1}{2}x^\top Px + x^\top b + \frac{1}{2}b^\top P^{-1}b,$$

因此我们有

$$f(x) = \frac{1}{2}x^\top Px + x^\top b = \frac{1}{2}(x + P^{-1}b)^\top P(x + P^{-1}b) - \frac{1}{2}b^\top P^{-1}b. \quad (11.59)$$

如果假定  $P$  中有负的特征根  $-\lambda$ , ( $\lambda > 0$ ), 对应特征向量  $u$ , 根据定义我们有  $Pu = -\lambda u$ 。另一方面, 设  $x + P^{-1}b := \alpha u$ , 其中  $\alpha$  是任一不为 0 的实数。则我们有

$$(x + P^{-1}b)^\top P(x + P^{-1}b) = (\alpha u)^\top P(\alpha u) = \alpha^2 u^\top Pu = \alpha^2 u^\top (-\lambda u) = -\alpha^2 \lambda \|u\|^2,$$

其中  $\|u\|$  表示向量的范数 (norm)。则(11.59)进一步改写为

$$f(x) = \underbrace{-\frac{1}{2}\alpha^2 \lambda \|u\|^2}_{\geq 0} - \frac{1}{2}b^\top P^{-1}b, \quad (11.60)$$

由于  $\alpha$  的取值可以至任意大, 在  $P$  存在负的特征根的情况下  $f(x)$  不可能有最小值。

因此,  $f(x)$  最小值存在的先决条件是满秩矩阵  $P \succeq 0$ 。在  $P \succeq 0$  并且(11.59)右侧第一部分恒大于等于 0(由(11.60)右侧第一部分恒大于等于 0 推得)的情况下,  $f(\cdot)$  的最小值只能出现在  $x^* = -P^{-1}b$ , 对应  $f(x^*) = -\frac{1}{2}b^\top P^{-1}b$ 。 $\square$

**Proposition 11.5** (奇异矩阵). 如果  $P$  是一个对称的奇异矩阵, 那么只有当  $P \succeq 0$  并且  $(I - PP^\dagger)b = 0$  时, 方程  $f(x) = (\frac{1}{2})x^\top Px + x^\top b$  才具有唯一的最优(极小)值  $x^* = -P^\dagger b$ , 使得  $f(x^*) = -\frac{1}{2}b^\top P^\dagger b$ 。

进而, 如果  $P = U^\top \Sigma U$  是一个关于  $P$  的 SVD, 那么系统最优值  $f(x^*)$  对应  $x \in \mathbb{R}^{n \times n}$ :

$$x^* = -P^\dagger b + U^\top \begin{bmatrix} 0 \\ z \end{bmatrix}, \quad \text{对于任一 } z \in \mathbb{R}^{n-r \times n-r}, \quad r := \text{rank}(P).$$

证明. 对不满秩( $r < n$ )的奇异矩阵  $P$  做分块对角化

$$P = U^\top \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} U,$$

其中  $U$  是正交矩阵,  $\Sigma_r \in \mathbb{R}^{r \times r}$  是可逆对角矩阵。由此可得

$$\begin{aligned} f(x) &= \frac{1}{2}x^\top Px + x^\top b \\ &= \frac{1}{2}x^\top U^\top \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} Ux + x^\top U^\top Ub \\ &= \frac{1}{2}(Ux)^\top \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} (Ux) + (Ux)^\top Ub. \end{aligned}$$

定义

$$Ux := \begin{bmatrix} y \\ z \end{bmatrix}, \quad Ub := \begin{bmatrix} c \\ d \end{bmatrix} \quad y, c \in \mathbb{R}^r, \quad z, d \in \mathbb{R}^{(n-r)},$$

上式可进一步调整为

$$\begin{aligned} f(x) &= \frac{1}{2} \begin{bmatrix} y^\top & z^\top \end{bmatrix} \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} + \begin{bmatrix} y^\top & z^\top \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} \\ &= \frac{1}{2}y^\top \Sigma_r y + y^\top c + z^\top d. \end{aligned} \tag{11.61}$$

根据(11.61), 我们逐次分析  $f(x)$  最小值存在的两个条件:

1. 对于  $y = 0$  的情况, 我们有  $f(x) = z^\top d$ , 此时如果  $d \neq 0$  取任意值,  $f(x)$  将没有最小值。因此为了让  $f(x)$  有最小值, 需要使  $d = 0$ 。根据第11.B节可得,  $Ub = \begin{bmatrix} c \\ 0 \end{bmatrix}$ , 进而  $b = \text{range}(P)$ <sup>19</sup>; 因此  $f(x)$  存在最小值的条件之一是  $b = 0$ , 等价于  $(I - PP^\dagger)b = 0$ 。
2. 将  $b = 0$  代回原式, 我们有  $f(x) = \frac{1}{2}y^\top \Sigma_r y + y^\top c$ , 由于分块矩阵  $\Sigma_r$  是可逆的, 根据 Proposition 11.4,  $f(x)$  存在最小值的第二个条件是  $\Sigma_r \succeq 0$ , 等价于  $P \succeq 0$ 。

假定  $f(x)$  最小值存在的两个条件都得到了满足。来看  $x^*$  的取值。

<sup>19</sup> 第11.B节中的  $U$  对应本节中的  $U^\top$ 。

1. 对于  $d = 0$ ,

$$\frac{\partial f}{\partial y} = \Sigma_r y + c = 0,$$

可得  $y^* = -\Sigma_r^{-1}c$ .

2. 假定  $z = 0$ ,

$$Ux^* = \begin{bmatrix} y^* \\ z \end{bmatrix} = \begin{bmatrix} -\Sigma_r^{-1}c \\ 0 \end{bmatrix},$$

由正交矩阵性质我们有  $U^\top = U^{-1}$ , 进而

$$\begin{aligned} x^* &= -U^{-1}\Sigma_r^{-1}c \\ &= -U^\top \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} c \\ 0 \end{bmatrix} \\ &= -U^\top \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix} Ub \\ &= -P^\dagger b, \end{aligned} \tag{11.62}$$

并且不难看出, 任一  $z \in R^{n-r}$  对  $x^* = -P^\dagger b + U^\top \begin{bmatrix} 0 \\ z \end{bmatrix}$  的取值都不产生影响。

进一步看最优极小值  $f(x^*)$

$$\begin{aligned} f(x^*) &= \frac{1}{2}(x^*)^\top Px^* + (x^*)^\top b \\ &= \frac{1}{2}(-P^\dagger b)^\top P(-P^\dagger b) + (-P^\dagger b)^\top b \\ &= \frac{1}{2}b^\top (P^\dagger)^\top PP^\dagger b - b^\top (P^\dagger)^\top b \\ &= b^\top \left(-\frac{1}{2}(P^\dagger)^\top\right) b \\ &= -\frac{1}{2}b^\top P^\dagger b, \end{aligned} \tag{11.63}$$

并且同样地, 任一  $z \in R^{n-r}$  对  $f(x^*)$  的取值也不产生影响。  $\square$

### 判断 $M$ 是否正半定

在此基础上, 回到最初的问题上来: 如何判断对称矩阵  $M = \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}$  是正半定矩的。

这个问题等同于: 判断方程  $f(x, y)$  是否有最小值:

$$f(x, y) = \begin{bmatrix} x^\top & y^\top \end{bmatrix} \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

根据 Proposition (11.5) 可知: 对于给定的常数向量  $y$ , 只有当  $A \succeq 0, (I - AA^\dagger)By = 0, C - B^\top A^\dagger B \succeq 0$  时,  $f$  存在最小值, 最小值为

$$f(x^*, y) = -y^\top B^\top A^\dagger B y + y^\top C y = y^\top (C - B^\top A^\dagger B) y.$$

**Theorem 11.1.** 对于任一对称矩阵  $M = \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}$ , 下述条件之间等价

1.  $M \succeq 0$ ,
2.  $A \succeq 0, (I - AA^\dagger B) = 0, C - B^\top A^\dagger B \succeq 0$ ,
3.  $C \succeq 0, (I - CC^\dagger B = 0), A - B^\top C^\dagger B \succeq 0$ ,

并且可以对  $M$  做如下矩阵分解

$$M = \begin{cases} = \begin{bmatrix} I & BC^\dagger \\ 0 & I \end{bmatrix} \begin{bmatrix} A - BC^\dagger B^\top & 0 \\ 0 & C \end{bmatrix} \begin{bmatrix} I & 0 \\ C^\dagger B^\top & I \end{bmatrix}, & \text{或} \\ = \begin{bmatrix} I & 0 \\ B^\top A^\dagger & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & C - B^\top A^\dagger B \end{bmatrix} \begin{bmatrix} I & A^\dagger B \\ 0 & I \end{bmatrix}. \end{cases}$$

## 11.B 伪逆矩阵

又称 Moore-Penrose Pseudo inverse。可见如 (Bapat, 2012, Ch.5)。

### 11.B.1 用 SVD 法做伪逆矩阵分解

一个方块矩阵  $M \in \mathbb{R}^{n \times n}$  可以做如下 SVD 分解 (singular value decomposition)，以求得伪逆矩阵  $M^\dagger$

$$M = U\Sigma V^\top,$$

其中

- $U, V$  是正交矩阵， $U$  和  $V$  的 column 分别是  $MM^\top$  和  $M^\top M$  的特征向量。需要指出的是，对  $M$  做 SVD，得到的  $U$  和  $V$  并不唯一。
- 对角矩阵  $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r, 0, 0, \dots, 0)$ ：
  - $\sigma_1, \sigma_2, \dots, \sigma_r$  表示  $M$  的秩，并且  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$ 。
  - $\sigma_i$  表示  $M$  的奇异值。 $\sigma_i, i = 0, 1, \dots, r$  表示  $MM^\top$  和  $M^\top M$  的非零特征值的平方根。

如果  $M = U\Sigma V^\top$  是某个  $M$  的 SVD，则我们将  $M$  的伪逆矩阵  $M^\dagger$  定义如下

$$M^\dagger = V\Sigma^\dagger U^\top, \quad \Sigma^\dagger = \text{diag}(\sigma_1^{-1}, \sigma_2^{-1}, \dots, \sigma_r^{-1}, 0, 0, \dots, 0). \quad (11.64)$$

不难看出，当  $M$  是满秩 ( $r = n$ ) 的（等价于  $M$  是可逆的）时， $M^\dagger = M^{-1}$ 。因此  $M^\dagger$  是个  $M$  的“广义逆矩阵”。对  $M$  做 SVD， $U$  和  $V$  有所不同，但同一  $M$  只有唯一的  $M^\dagger$  与之相对应。此外，根据伪逆矩阵的性质，我们有  $MM^\dagger M = M$ ,  $M^\dagger MM^\dagger = M^\dagger$ ，并且  $M$  和  $M^\dagger$  都是对称矩阵；并且

$$\begin{cases} MM^\dagger = U\Sigma V^\top V\Sigma^\dagger U^\top = U\Sigma\Sigma^\dagger U^\top = U \begin{bmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{bmatrix} U^\top, \\ M^\dagger M = V\Sigma^\dagger U^\top U\Sigma V^\top = V\Sigma^\dagger\Sigma V^\top = V \begin{bmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{bmatrix} V^\top, \end{cases}$$

因此我们有

$$\begin{cases} (MM^\dagger)^2 = MM^\dagger, \\ (M^\dagger M)^2 = M^\dagger M, \end{cases}$$

即  $(MM^\dagger)$  和  $(M^\dagger M)$  是对称的正交投影：

1.  $MM^\dagger$  是  $\text{range}(M)$  的正交投影,
2.  $M^\dagger M$  是  $\ker(M)^\perp$  的正交投影,  $\ker(M)^\perp$  表示  $\ker(M)$  的正交补 (orthogonal complement)。

证明. 由于  $\text{range}(M^\dagger M) \subseteq \text{range}(M)$ , 即  $M^\dagger M$  的值域是  $M$  值域的子集, 则在  $\text{range}(M)$  值域中, 对于任一  $y = Mx$ , 由于  $MM^\dagger M = M$ , 我们有

$$MM^\dagger y = MM^\dagger Mx = Mx = y,$$

因此  $\text{range}(MM^\dagger) \subseteq \text{range}(M)$ 。此外, 根据  $MM^\dagger M = M \Rightarrow \ker(M^\dagger M) \subseteq \ker(M)$ 。因而有  $\ker(M^\top) = \ker(M)$ 。

由于  $M^\dagger M$  是埃米特矩阵 (Hermitian),  $\text{range}(M^\dagger M) = \ker(M^\dagger M)^\perp = \ker(M)^\perp$ 。  $\square$

值域  $\text{range}(M) = \text{range}(MM^\dagger)$  中包括所有向量  $y \in \mathbb{R}^{n \times n}$ , 满足  $U^T y = \begin{bmatrix} z \\ 0 \end{bmatrix}, z \in \mathbb{R}^r$ 。

证明. 对于  $y = Mx$ , 我们有

$$U^T y = U^T Mx = U^T U \Sigma V^T x = \Sigma V^T x = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0_{n-r} \end{bmatrix} V^T x = \begin{bmatrix} z \\ 0 \end{bmatrix},$$

反过来, 如果  $U^T y = \begin{bmatrix} z \\ 0 \end{bmatrix}$ , 则我们首先有  $y = U \begin{bmatrix} z \\ 0 \end{bmatrix}$ , 进而

$$\begin{aligned} MM^\dagger y &= U \begin{bmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{bmatrix} U^T y \\ &= U \begin{bmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{bmatrix} U^T U \begin{bmatrix} z \\ 0 \end{bmatrix} \\ &= U \begin{bmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{bmatrix} \begin{bmatrix} z \\ 0 \end{bmatrix} \\ &= U \begin{bmatrix} z \\ 0 \end{bmatrix} \end{aligned}$$

这说明  $y \in \text{range}(M)$ , 即  $y$  在  $M$  的值域之中。  $\square$

类似地, 值域  $\text{range}(M^\dagger M) = \ker(M)^\perp$  中包括所有向量  $y \in \mathbb{R}^{n \times n}$ , 满足  $V^T y = \begin{bmatrix} z \\ 0 \end{bmatrix}, z \in \mathbb{R}^{r \times r}$ 。

证明. 对于  $y = M^\dagger MU$ , 我们有

$$y = M^\dagger MU = V \begin{bmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{bmatrix} V^T U = V \begin{bmatrix} z \\ 0 \end{bmatrix}.$$

反过来, 如果  $V^T y = \begin{bmatrix} z \\ 0 \end{bmatrix}$ , 则  $y = \begin{bmatrix} z \\ 0 \end{bmatrix}$ , 进而

$$M^T M V \begin{bmatrix} z \\ 0 \end{bmatrix} = V \begin{bmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{bmatrix} V^T V \begin{bmatrix} z \\ 0 \end{bmatrix} = V \begin{bmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{bmatrix} \begin{bmatrix} z \\ 0 \end{bmatrix} = y.$$

这说明  $y \in \text{range}(M^\dagger M)$ 。  $\square$

### 11.B.2 用 SVD 法做 M 的矩阵分解

对于  $M$  是一个对称实矩阵的情况。每一个  $M = U\Sigma V^\top$  的 SVD 分解常常只有不超过 1 个  $\Sigma$ , 但有多个  $U$  和  $V$ , 并且通常来说不存在  $U = V$  的情况。然而, 如果对称矩阵  $M \succeq 0$ , 那么

1.  $M$  的特征值非负,
2.  $M$  的非零特征值数量, 与奇异值数量相等,
3. 可以用 SVD 法分解:  $M = U\Sigma V^\top$ 。

如果  $M$  是一个对称复矩阵, 结果与之相类似, 除了此时

1.  $U$  和  $V$  是埃尔米特矩阵, 并且
2.  $M^\top M$  和  $MM^\top$  是埃尔米特正交映射。

对于  $M$  是一个对称正规矩阵 (normal matrix) 的情况, 即方块矩阵满足  $MM^\top = M^\top M$ , 此时  $M$  的 SVD 分解与分块对角分解之间关系密切, 并且  $M$  的伪逆矩阵可以直接由分块对角分解求得。

证明. 对实正规矩阵  $M$  做分块对角分解:

$$M = U\Lambda U^\top,$$

其中  $U$  是正交矩阵,  $\Lambda = \text{diag}(B_1, B_2, \dots, B_n)$  是实系数分块对角矩阵。

分块  $B_j, j = 0, 1, \dots, n$  可以是下属两种形式之一。第一是一个  $2 \times 2$  的矩阵  $B_j = \begin{bmatrix} \lambda_j & -\mu_j \\ \mu_j & \lambda_j \end{bmatrix}$ ; 第二是一个一维分块  $B_j = (\lambda_j)$ 。

假定  $B_1, B_2, \dots, B_p$  是前一种形式, 且  $\lambda_{2p+1}, \lambda_n$  是标量。我们已知  $\lambda_j \pm i\mu_j$ , 并且  $\lambda_{2p+k}$  表示  $A$  的特征值。

令  $\rho_{2j-1} = \rho_{2j} = (\lambda_j^2 + \mu_j^2)^{1/2}$  for  $j = 1, \dots, p$ , 以及  $\rho_{2p+j} = \lambda_j$ , 并且假定这些分块已经按照降序排列  $\rho_1 \geq \rho_2 \geq \dots \geq \rho_n$ 。那么可见

$$UU^\top = U^\top U = U\Lambda U^\top U\Lambda^\top U^\top = U\Lambda\Lambda^\top U^\top,$$

并且

$$\Lambda\Lambda^\top = \text{diag}(\rho_1^2, \rho_2^2, \dots, \rho_n^2),$$

因此可见:  $M$  的奇异值  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$  就是矩阵  $MM^\top$  特征值的非负平方根  $\rho_1 \geq \rho_2 \geq \dots \geq \rho_n$ , 满足  $\sigma_j = \rho_j, 1 \leq j \leq n$ 。

接着定义对角矩阵

$$\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r, 0, 0, \dots, 0),$$

其中  $r = \text{rank}(M)$ ,  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ , 并且定义

$$\Theta = \text{diag}(\sigma_1^{-1}B_1, \sigma_2^{-1}B_2, \dots, \sigma_p^{-1}B_p, 1, 1, \dots, 1).$$

由此可见,  $\Theta$  是个正交矩阵, 并且

$$\Lambda = \Theta\Sigma = (B_1, B_2, \dots, B_p, \lambda_{2p+1}, \lambda_{2p+2}, \dots, \lambda_r, 0, 0, \dots, 0).$$

进而

$$M = U\lambda U^\top = U\Theta\Sigma U^\top.$$

由于  $U$  和  $\Theta$  都是正交矩阵，因此  $V \equiv U\Theta$  也是正交矩阵，并且  $M = V\Sigma U^\top$  是  $M$  的 SVD。根据 SVD 的性质我们有

$$M = V\Sigma U^\top = U\Sigma^\dagger V^\top = U\Sigma^\dagger\Theta^\top U^\top,$$

其中最后一个等式根据  $V = U\Theta \Rightarrow V^\top = \Theta^\top U^\top$ 。

$\Theta$  是正交矩阵  $\Rightarrow \Theta^\top = \Theta^{-1} \Rightarrow \Sigma^\dagger\Theta^\top = \Sigma^\dagger\Theta^{-1} = \Lambda^\dagger$

因此我们有  $M^\dagger = U\Lambda^\dagger U^\top$ 。

来看  $\Lambda^\dagger$  的性质。如果我们将  $\Lambda_r$  写为  $\Lambda_r = (B_1, B_2, \dots, B_p, \lambda_{2p+1}, \dots, \lambda_r)$ ，则  $\Lambda$  是可逆矩阵，并且  $\Lambda^\dagger = \begin{bmatrix} \Lambda_r^{-1} & 0 \\ 0 & 0 \end{bmatrix}$ 。

由上可见，一个(实系数)正规矩阵  $M$  的伪逆矩阵，可以通过对  $M$  做分块对角分解求得。□

# 第十二章 映射法

## 12.1 简介

映射法 (projection methods), 又称加权残差法 (weighted residual methods)。利用映射法求解 DSGE 模型是指, 找到一组合适的系数向量  $\theta = \{\theta_0, \theta_1, \dots, \theta_j\}$  和一组基方程 (basis function)  $\Psi_i(\mathbf{x})$  后, 假定  $\theta$  和  $\Psi_i(\mathbf{x})$  呈现线性组合方程  $d^j(\mathbf{x}|\theta)$

$$d^j(\mathbf{x}|\theta) = \sum_{i=0}^j \theta_i \Psi_i(\mathbf{x}), \quad (12.1)$$

进而用  $d^j(\mathbf{x}|\theta)$  来近似  $d(\mathbf{x})$ , 进而求解  $\mathcal{H}(d) = 0$ , 即是说, 定义一个残差方程 (residual function)  $R(\mathbf{x}|\theta) = \mathcal{H}(d^j(\mathbf{x}|\theta))$ , 通过  $\theta$  系数的选取实现残差方程的最小化

$$\min_{\{\theta\}} |R(\mathbf{x}|\theta)|, \quad (12.2)$$

这称为对  $\mathcal{H}$  做  $\Psi$  的映射, 以求最佳系数  $\theta$ 。

在操作层面上, 这意味着首先如(12.1), 选取某种恰当的基  $\{\Psi_i(\mathbf{x})\}_{i=0}^\infty$ , 用基与系数内积并求和的方法构建  $d(\mathbf{x}|\theta)$ ; 进而如(12.2), 对  $\mathcal{H}$  做  $\Psi$  的映射, 求得  $\theta$ 。就这两个环节来说, 显然, 选取不同形式的基和映射算法的组合, 会出现不同的映射法, 在不同的文献中, 有时也会给它们以一些专有的称呼, 以与其他映射法相区分。

经济学家很早就将映射理论应用到经验研究中去了, 进入 20 世纪 90 年代后, 映射法逐渐成为一种成熟的宏观经济学研究方法, 主要归功于 Judd (1992); Gaspar and Judd (1997); Judd (1998) 的贡献<sup>1</sup>。

大致说来, 映射法比扰动法更易于描述, 但算法实现上则较为困难。下面是一个映射法的基本算法描述:

**Algorithm 12.1** (映射法的基本算法). 可简述如下:

- 对于正整数  $j < \infty$ , 定义  $j+1$  个线性不相关方程  $\Psi_i : \Omega \mapsto \mathbb{R}, i = 0, 1, \dots, j$ 。则  $\{\Psi_i\}_0^j = \psi_0, \psi_1, \dots, \psi_j$  称为基方程, 是与状态变量向量  $\mathbf{x}$  有关的方程。
- 对于正整数  $m < \infty$  ( $m$  是方程系统  $d(\mathbf{x})$  所映射的维度), 系数向量  $\theta^l = [\theta_0^l, \theta_1^l, \dots, \theta_j^l]$ ,  $l = 1, 2, \dots, m$ 。将  $m$  组系数向量合并组成系数矩阵

$$\begin{matrix} \theta \\ \{m \times (j+1)\} \end{matrix} = \begin{pmatrix} \theta^1 \\ \theta^2 \\ \vdots \\ \theta^m \end{pmatrix} = \begin{pmatrix} \theta_0^1 & \theta_1^1 & \dots & \theta_j^1 \\ \theta_0^2 & \theta_1^2 & \dots & \theta_j^2 \\ \vdots & \vdots & \ddots & \vdots \\ \theta_0^m & \theta_1^m & \dots & \theta_j^m \end{pmatrix}.$$

<sup>1</sup>与扰动法相比, 映射法的思路更为现代, 但在自然科学、工程学研究中也有许多年的悠久历史了。映射法的前身之一光谱法 (spectral methods) 至少可追溯至 Lanczos (1938), 另一个前身有限元法 (finite elements methods, FEM) 来自于波音工程师 Clough (1960) 的开创性工作; 关于有限元法的早期发展可见 Clough and Wilson (1999)。

## 3. 定义内积求和方程

$$d^{l,j}(\cdot|\theta^l) = \sum_{i=0}^j \phi_i^l \psi_i(\cdot),$$

则对应地我们有

$$d^j(\cdot|\theta) = \begin{pmatrix} d^{1,j}(\cdot|\theta^1) \\ d^{2,j}(\cdot|\theta^2) \\ \vdots \\ d^{m,j}(\cdot|\theta^m) \end{pmatrix}.$$

4. 将  $d^j(\cdot|\theta)$  作为  $d(\cdot)$  的近似, 代入  $\mathcal{H}(\cdot) = 0$ , 构建残差方程

$$R(\cdot|\theta) = \mathcal{H}(d^j(\cdot|\theta)).$$

文有注释, 对  
距离方程其实  
三角方程的一  
。到时把讲义  
容补充到这里  
。

5. 定义度量方程 (metric function)<sup>2</sup>  $\rho(R(\cdot|\theta), \mathbf{0})$  作为目标方程, 在  $\theta$  中找到对应的  $\hat{\theta}$ , 满足条件

$$\hat{\theta} = \arg \min_{\theta \in \mathbb{R}^{m \times (j+1)}} \rho(R(\cdot|\theta), \mathbf{0}).$$

## 12.1.1 例

举例说明。以经典的随机内生经济增长模型为例, 第11.3.8节中我们介绍了二阶扰动方法的应用。使用近似的模型, 这里来介绍如何用映射法近似求解系统<sup>3</sup>。

第一步, 改写系统。模型中的状态变量  $c_t, k_{t+1}$  分别表示为

$$\begin{aligned} c_t &= d^1(k_t, z_t), \\ k_{t+1} &= d^2(k_t, z_t), \end{aligned}$$

由此将经济系统的均衡条件 (由跨期消费的欧拉等式和资源约束条件构成, 即  $m = 2$ ) 改写为

$$\begin{aligned} \mathcal{H}(d(k_t, z_t)) &= \mathbf{0}, \quad \forall k_t, z_t \Rightarrow \\ \begin{cases} u'(d^1(k_t, z_t)) - \beta E_t \left\{ u'(d^1(d^2(k_t, z_t), z_{t+1})) \left( \alpha \exp(\rho z_t + \sigma \varepsilon_{t+1}) (d^2(k_t, z_t))^{\alpha-1} + 1 - \delta \right) \right\} = 0, \\ d^1(k_t, z_t) + d^2(k_t, z_t) - \exp(z_t) k_t^\alpha - (1 - \delta) k_t = 0. \end{cases} \end{aligned}$$

第二步, 改写  $d(x)$  为  $d(x|\theta)$  的形式, 根据  $m = 2$  有  $\theta = [\theta^1, \theta^2]^\top$ , 进而定义

$$\begin{aligned} c_t &= d^{1,j}(k_t, z_t | \theta^1) = \sum_{i=0}^j \theta_i^1 \psi_i(k_t, z_t), \\ k_{t+1} &= d^{2,j}(k_t, z_t | \theta^2) = \sum_{i=0}^j \theta_i^2 \psi_i(k_t, z_t). \end{aligned}$$

其中  $j + 1$  个基方程  $\psi_i(k_t, z_t), i = 0, 1, \dots, j$  的形式, 将在下文讨论。

<sup>2</sup> 度量方程又称距离方程 (distance function), 描述集合中一对元素之间的距离。度量方程的一种, 离散度量方程 (discrete metric function) 可以表示如下

$$d(x, y) = \begin{cases} 0 & x = y, \\ 1 & x \neq y. \end{cases}$$

<sup>3</sup> 第11.3.8节中的  $\eta$ , 在这里用  $\sigma$  来表示。

第三步，构建残差方程  $R(k_t, z_t | \theta) = [R(k_t, z_t | \theta^1), R(k_t, z_t | \theta^2)]^\top \Rightarrow$

$$\begin{aligned} R(k_t, z_t | \theta^1) &= u' \left( \sum_{i=0}^j \theta_i^1 \psi_i(k_t, z_t) \right) - \beta E_t \left\{ u' \left( \sum_{i=0}^j \theta_i^1 \psi_i \left( \sum_{i=0}^j \theta_i^2 \psi_i(k_t, z_t), \rho z_t + \sigma \varepsilon_{t+1} \right) \right) \times \right. \\ &\quad \left. \left( \alpha \exp(\rho z_t + \sigma \varepsilon_{t+1}) \left( \sum_{i=0}^j \theta_i^2 \psi_i(k_t, z_t) \right)^{\alpha-1} + 1 - \delta \right) \right\}, \\ R(k_t, z_t | \theta^2) &= \sum_{i=0}^j \theta_i^1 \psi_i(k_t, z_t) + \sum_{i=0}^j \theta_i^2 \psi_i(k_t, z_t) - \exp(z_t) k_t^\alpha - (1 - \delta) k_t. \end{aligned}$$

第四步，构建度量方程  $\rho(R(\cdot|\theta), \mathbf{0})$ ，计算

$$\hat{\theta} = \arg \min_{\theta \in \mathbb{R}^{m \times (j+1)}} \rho(R(\cdot|\theta), \mathbf{0}),$$

使得相对于  $\hat{\theta}$  的任意  $m \times (j+1)$  个点  $(k_l, z_l)$ ，都有度量  $\rho = 0$ 。为了实现这点，需要令对应的  $m \times (j+1)$  个残差方程最小化，换句话说，求解下述含有  $m \times (j+1)$  个未知数的方程系统<sup>4</sup>

$$R(k_l, z_l | \theta) = \mathbf{0}, \quad l \in \mathbb{R}^{m \times (j+1)}.$$

如上所述，采用映射法从事宏观经济学研究，关键的两点在于选择合适的基方程  $\Psi(\cdot)$  以及度量方程  $\rho(\cdot)$ ，在下文分别讨论两个方程的选取策略之前，我们先来回顾一下，“映射法”和计量经济学如 OLS 回归中的“映射”之间有何异同，以及映射法和参数化期望法之间的联系与区别。

reference 和 reference

### 与计量经济学中的“映射”的关系

计量经济学中也有“映射”的概念，与映射法的思路接近。例如，对于给定的变量  $(X, Y)$ ，线性回归考虑构建一个未知的条件期望方程  $E(Y|X)$ 。为了求解该条件期望方程，我们可以将  $E(\cdot)$  近似为两个关于解释变量  $X$  的单项式 (monomial)，一个是常系数  $\theta_0$ ，一个是关于  $X$  的线性方程对应系数  $\theta_1$

$$E(Y|X) \approx \underbrace{\theta_0}_{\text{单项式 1}} + \underbrace{\theta_1 X}_{\text{单项式 2}},$$

这两个单项式可以合并构成另一个单项式的基的前两个元素，也可以合并构成另一个多项式的基的前两个元素，如切比雪夫多项式<sup>5</sup>。从而，计量经济学中“映射”法所构建的残差方程可以表示为

$$R(Y, X | \theta_0, \theta_1) = Y - \theta_0 - \theta_1 X,$$

随后的工作就变为，通过插入实际观测到的数据  $\{Y, X\}_{t=1}^T$ ，寻找合适的系数  $(\theta_0, \theta_1)$  使得残差方程的平方值最小  $\min R(Y, X | \theta_0, \theta_1)^2$ 。

可见计量经济学研究中的“映射”，和这里介绍的“映射”法，研究思路相近，区别在于以 OLS 为例的前者使用实际观测到的数据做系数测算，后者则使用来自经济学理论模型的条件  $\mathcal{H}(d)$ 。

### 映射法与参数化期望法

在第10.7节我们介绍了参数化期望法 (PEA)。不难看出，参数化期望法与映射法也有相通之处。但二者还是存在着不同：

<sup>4</sup>本讲义略过方程解的存在性与唯一性的讨论。

<sup>5</sup>超几何方程的一般介绍，见附录12.A；超几何方程的另一种表现形式是（正交）多项式，见附录12.B。附录12.C.5对切比雪夫多项式也做了介绍。

1. 在模型构建方面，映射法将原方程系统表述为一组新的近似方程系统，新系统由一系列含有条件期望的基方程灵活组合而成，如(12.1)。在参数化期望法中，近似方程系统却需要假定单项式，或者由一系列单项式构成的方程系统来实现，如(10.45)。后者往往不是一个最佳策略<sup>6</sup>。
2. 在特定条件下，利用参数化期望法迭代求得的系数值，即便能够求得，常常与实际情况下的最优选择相差较远；或者即便最终收敛到最优解，但收敛速度往往较慢，甚至是不稳健的。

## 12.2 全局基的选取——单维基光谱法

本节讨论如何选取一组合适的基方程  $\Psi = \psi_0, \psi_1, \dots, \psi_j$ 。通常说来，选取策略分为两大类，第一类是选取在状态变量域  $\Omega$  的绝大部分范围内都是非零且平滑的基，称为全局基。另一类是选取在  $\Omega$  中的一小部分范围内非零且平滑的基，称为局部基 (local basis)。基于全局基的映射法常称为光谱法 (spectral method)，基于局部基的映射法常称为有限元法 (finite elements method, FEM)。我们先介绍光谱法。

Judd (1992) 首次将光谱法应用到经济学研究中来。光谱基作为全局基的代表，主要优点在于，便于构建近似算法并求解。然而缺点在于，在处理状态变量的某些局部子域时，效果未尽人意——如傅里叶序列 (Fourier series) 中的吉布斯现象 (Gibbs phenomenon)<sup>7</sup>。关于光谱法的介绍，可見Shen et al. (2011)。

单维基 (unidimensional basis) 是光谱法中最常见的一种基方程。大致说来，有以下几种

- 单项式
- 三角序列
- 雅各比多项式
- 切比雪夫多项式

当方程系统中只有一个状态变量时，使用单维基有助于使系统求解尽可能简单。但模型常常较为复杂，不止一个状态变量，这就需要我们采用其他一些特殊类型的基方程。

### 12.2.1 单项式

单项式可以作为光谱法全局基的备选项之一：

$$1, x, x^2, x^3, \dots$$

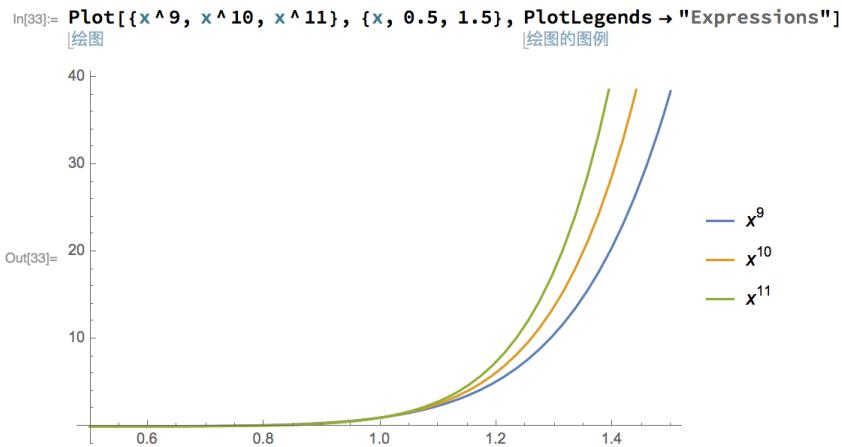
单项式直观，形象，易于理解。尽管单项式的基并不是由正交方程所构成的，但根据斯通-魏尔斯拉定理 (Stone-Weierstrass Theorem)<sup>8</sup>，由一系列可测度且有界的方程所构成的紧致集合 (compact set) 空间中，我们总是能够将任一闭区间内的连续方程近似为若干单项式的线性组合。

然而，采用单项式作为基，存在两个棘手的问题。第一个是多重共线性。单项式都是 (近似) 多重共线的，如图(12.1)所示，在  $x \in [0.5, 1.5]$  区间内，单项式方程  $x^9$  和  $x^{10}$  的曲线非常近似，这使得我们哪怕增添更高一次的单项式方程  $x^{11}$ ，计算出的近似解也许不会很快逼近实际方程。第二个问题是，随着  $x$  值的变化，多项式的值变化较大：如  $\frac{1.5^{11}}{0.5^{11}} = \frac{86.4976}{0.000488281} = 177147$ ，容易造成计算误差。

<sup>6</sup> Christiano and Fisher (2000) 因此建议对参数化期望法进行改进，用如车比雪夫多项式替代(10.45)的方程形式的猜测。

<sup>7</sup> 第12.G.5节。

<sup>8</sup> cf. 维基百科词条[https://en.wikipedia.org/wiki/Stone-Weierstrass\\_theorem](https://en.wikipedia.org/wiki/Stone-Weierstrass_theorem)。

图 12.1: 单项式方程  $x^9, x^{10}, x^{11}$  在  $[0.5, 1.5]$  区间内的图形

单项式基的上述缺点，使得我们寻求利用正交多项式构建基方程。内积形式的正交多项式的值相对于  $x$  的变化往往较为平缓，并且在加入更高次多项式元素后，系统会出现足够大的变化，有助于更精确近似原系统中的方程  $d(\cdot)$ 。

### 12.2.2 三角序列

光谱法全局基的另一选项是三角序列 (trigonometric series)，如

$$\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{2\pi}}, \frac{\sin x}{\sqrt{2\pi}}, \dots, \frac{\cos kx}{\sqrt{2\pi}}, \frac{\sin kx}{\sqrt{2\pi}}, \dots$$

三角序列适合用于分析有周期性特征的方程，因而在自然科学和工程学中得到广泛应用。遗憾的是除了时间序列分析等领域之外，经济学研究的问题较少涉及到周期性。此外，目前将周期性方程近似为非周期性方程的方法尚未成熟。因此我们不做深入探讨。

### 12.2.3 雅各比多项式

雅各比多项式(第12.C.3节)也适合作为光谱法全局基的选项之一。

雅各比多项式一族中，还包括盖根鲍尔多项式 (Gegenbauer)，勒让德 (Legendre)，切比雪夫多项式 (Chebyshev)，他们可以相互转换。[\(Boyd and Petschek, 2013, Table 1\)](#) 比较了几种多项式在处理不同优化指标时的性能排序，从而提供了重要参考：在绝大多数情况下，对于求解 DSGE 模型的工作来说，最适于采用切比雪夫多项式构建基方程。

### 12.2.4 切比雪夫多项式

切比雪夫多项式(第12.C.5节)作为光谱法全局基的选项之一，具有许多优点：

1. 便于在各种表述形式之间互相转换，如罗德里格斯公式、三项递推关系、母方程、二阶导数线性方程等。
2. 可以通过余弦换算，迅速计算系数变化后的值。
3. 切比雪夫插值 (interpolation) 换算的结果，比其他几种多项式插值的结果更为稳健。
4. 切比雪夫多项式是平滑的，并且有界 (闭区间  $[-1, 1]$ )。

5. 切比雪夫差值的误差，可以由一系列定理得到较好的限制。

第一类切比雪夫多项式  $T_n(x)$  满足关系  $T_n(0) = 1, T_n(1) = x, T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$ ，因此我们可得到一组多项式序列  $1, x, 2x^2 - 1, 4x^3 - 3x, 8x^4 - 8x^2 + 1, \dots$ ，随着  $n = 0, 1, 2, \dots, 7$ ，我们绘制出了  $T_n(x)$  的曲线，见12.2。不难看出， $n = 0$  时是条平行线， $n = 1$  时是 45 度斜线， $n = 2$  时是一条抛物线，随着  $n$  逐渐增加，切比雪夫多项式的曲线形状呈波浪状。

图12.2在 Mathematica 中输入如下命令生成：

```
t0 = Plot[{ChebyshevT[0, x]}, {x, -1, 1}, PlotLabels -> "Expressions"];
t1 = Plot[{ChebyshevT[1, x]}, {x, -1, 1}, PlotLabels -> "Expressions"];
t2 = Plot[{ChebyshevT[2, x]}, {x, -1, 1}, PlotLabels -> "Expressions"];
t3 = Plot[{ChebyshevT[3, x]}, {x, -1, 1}, PlotLabels -> "Expressions"];
t4 = Plot[{ChebyshevT[4, x]}, {x, -1, 1}, PlotLabels -> "Expressions"];
t5 = Plot[{ChebyshevT[5, x]}, {x, -1, 1}, PlotLabels -> "Expressions"];
t6 = Plot[{ChebyshevT[6, x]}, {x, -1, 1}, PlotLabels -> "Expressions"];
t7 = Plot[{ChebyshevT[7, x]}, {x, -1, 1}, PlotLabels -> "Expressions"];
GraphicsGrid[{{t0, t1}, {t2, t3}, {t4, t5}, {t6, t7}}]
```

第  $n$  次切比雪夫多项式  $T_n(x)$  有  $n$  个根，对应

$$x_k = \cos\left(\frac{2k-1}{2n}\pi\right), \quad k = 1, 2, \dots, n, -1 \leq x_k \leq 1, \quad (12.3)$$

如图12.3所示，所有根都在闭区间  $[-1, 1]$  之间。

大多数 DSGE 模型的状态变量域  $x \in [a, b]$ ,  $a, b \neq \pm 1$ ，在实际研究中我们常采用如下线性关系转换，将它转换为  $[-1, 1]$  区间内的值，以符合切比雪夫多项式的要求：

$$2\frac{x-a}{b-a} - 1, \quad (12.4)$$

详见第12.2.4节。例如  $(a = -7, b = 5)$  的某状态变量线性转换，如图(12.4)所示。

使用切比雪夫多项式作为映射基，有两个定理值得关注：

1. 切比雪夫插值定理 (Chebyshev interpolation theorem)，见附录12.56：如果近似方程恰好是第  $n_1$  次切比雪夫多项式的根，那么随着  $n_1 \rightarrow \infty$ ，近似误差逐渐减少到足够小的程度。根据切比雪夫插值定理，可以使用切比雪夫多项式的根作为正交配点 (orthogonal collocation)，见12.4节 FEM 方法介绍。
2. 切比雪夫截断定理 (Chebyshev truncation theorem)，见附录12.57。进而在特定情况下，切比雪夫向未知原方程做几何收敛 (geometric convergence)，可以表示为

$$d(x) - d^j(x|\theta) \sim O(\theta_j),$$

即是说，当切比雪夫近似停止于第  $j$  次多项式时，对应的截断误差与  $j$  次切比雪夫多项式系数  $\theta_j$  是同次的。

根据该定理，我们可以设计一种数值检测机制：如截断误差大于某一阈值，则意味着  $j$  次切比雪夫近似  $T_j(x)$  的精度不够，需要再增加一次近似至  $T_{j+1}(x)$ 。我们将在下文进一步讨论该问题。

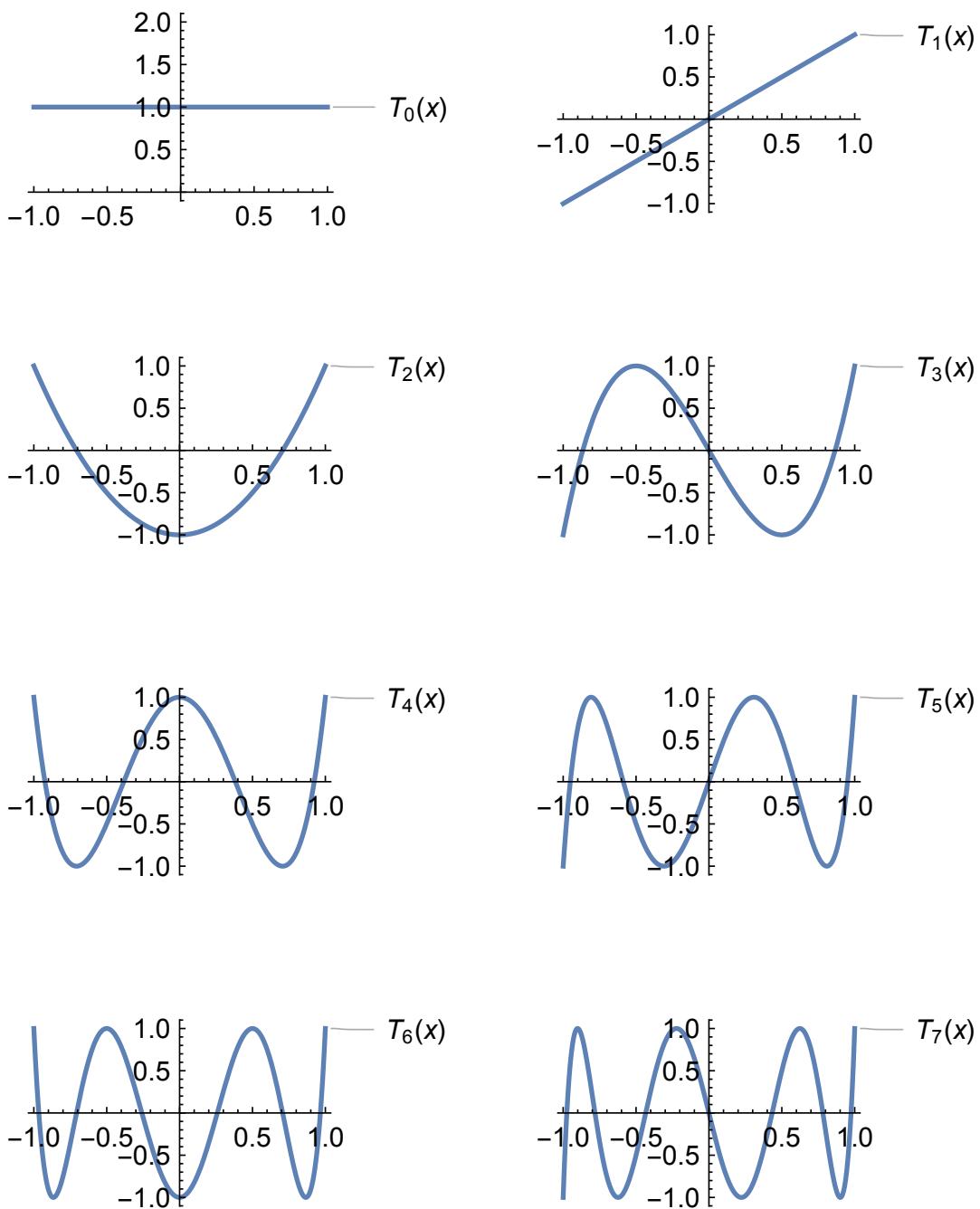
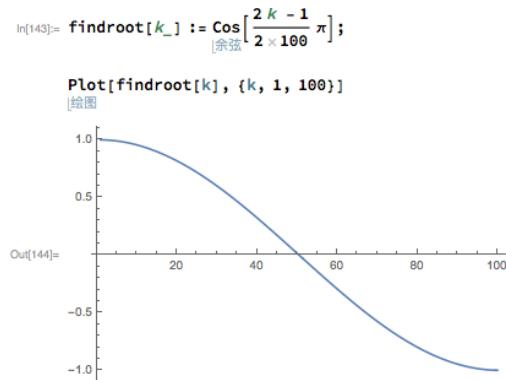
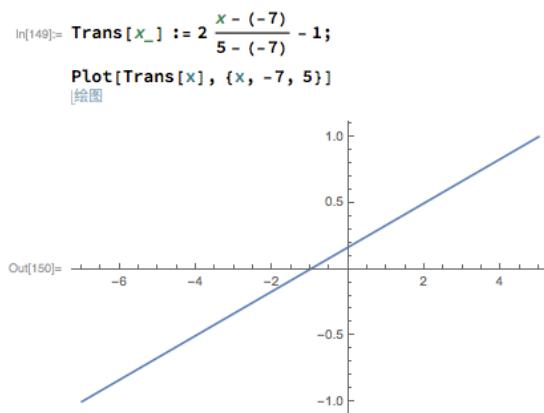
图 12.2: 切比雪夫多项式  $T_n(x)$ 

图 12.3: 切比雪夫多项式  $T_n(x)$  的根  $x_k, n = 100, k = 1, 2, \dots, n$ 

说明: 计算等式依据(12.3)。

图 12.4: 切比雪夫多项式  $T_n(x)$  的根  $x_k, n = 100, k = 1, 2, \dots, n$ 

说明: 计算等式依据(12.4)。

### 变量的变换

如前所述, 12.4 提供了一种可将  $x \in [a, b], a < -1, b > 1$  转化为  $[-1, 1]$  区间中变量的方法, 以便进行切比雪夫多项式近似。在扰动法中我们也讨论了变量的变换, 见 11.5 节。二者尽管技术细节有所不同, 但核心思路是相同的: 都是为了尽可能提高近似的精确度。本节以随机 NCGT 模型为例, 介绍为何通过变量变化有助于提高映射法的近似精度。

原目标为寻找经济系统方程组

$$\begin{aligned} c_t &= d^1(k_t, z_t), \\ k_{t+1} &= d^2(k_t, z_t) \end{aligned}$$

的近似解。我们可以通过变量变换, 改为寻求方程组

$$\begin{aligned} \log c_t &= d^1(\log k_t, z_t), \\ \log k_{t+1} &= d^2(\log k_t, z_t) \end{aligned}$$

的近似解, 用映射法表示为

$$\begin{aligned} \log c_t &= d^{1,j}(\log k_t, z_t | \theta^1) = \sum_{i=0}^j \theta_i^1 \psi_i(\log k_t, z_t), \\ \log k_{t+1} &= d^{2,j}(\log k_t, z_t | \theta^2) = \sum_{i=0}^j \theta_i^2 \psi_i(\log k_t, z_t). \end{aligned}$$

### 伯依德原则

前文简要介绍了利用切比雪夫多项式从事经济学研究所具有的理论优势。但近些年来的经验研究, 尤其是基于 DSGE 模型的经济学研究, 越来越多地使用基于切比雪夫多项式的映射法, 切比雪夫多项式作为基方程的备选方案, 其巨大优势得到了越来越广泛的认可 (Aruoba et al., 2006; Caldara et al., 2012)。

(Boyd, 2001, p.10) 用一种近似于开玩笑的方式, 总结了这几十年来的研究经验, 将之命名为“道德准则一号”:

1. 当不确定用什么基时, 就用切比雪夫多项式。除非模型呈现出较强的周期性, 这时可以考虑用傅里叶序列。
2. 除非你确定其他某种基方程更好, 不然就用切比雪夫多项式。
3. 除非你非常非常确定其他某种基方程更好, 不然就用切比雪夫多项式。

## 12.3 局部基的选取 (1)

第 12.2 节所介绍的几种方案, 全部都是单维度基方程, 相对简单, 易于我们理解用映射法求解 DSGE 模型的基本思路。然而现实世界中大多数经济问题都是多维度的, 几乎全部 DSGE 模型都研究一个以上的状态变量。这就需要我们探讨多维基的选择方案。

如何选取多维基成为映射法的关键问题。然而映射法受到维数灾难 (curse of dimensionality) 的强烈冲击 (Bellman, 1957)。随着解释变量数量的增加——如中等规模 DSGE 模型可能要处理超过 20 组解释变量——应用映射法求解模型变得异常困难, 这要求我们用更高的技巧去选取多维基方程。

### 12.3.1 离散状态变量

前文一系列分析均暗含假定，状态变量是连续的。然而在很多 DSGE 模型中，至少一部分变量是离散的：

- 状态变量本身是离散的，如
  - 财政政策，政府可能处于主权债务违约或未违约状态 (Bocola, 2016)，
  - 货币政策，可能是积极的或者消极的 (Leeper, 1991)，
- 出于计算求解的考虑，益于将连续状态变量做离散化处理，如
  - 外生随机过程 (技术冲击、偏好冲击等) 的离散化。

研究发现，有限马尔科夫链 (finite Markov chain)能够产生与连续过程相同的样本矩；经验分析表明，在大多数情况下，含有 5-7 个状态的马尔科夫链足够模拟一般情况下的随机过程信息，供经济学定量分析使用 (Tauchen, 1986; Kopecky and Suen, 2010)。因此状态变量的离散化问题可以理解为，针对某一个连续变量，我们寻找另一个 (离散的) 决策方程来描述它<sup>9</sup>。

仍然以含有随机 NCGT 模型为例，假定外生技术冲击  $z_t$  是一个一阶自回归过程 AR(1)

$$z_t = \rho z_{t-1} + \varepsilon_t,$$

其中  $\varepsilon_t \sim N(0, \sigma_z^2)$  是一个平稳分布 (stationary distribution)。

可以表示为含有  $n$  个点的马尔科夫过程  $z_t \in \{z_1, z_2, \dots, z_n\}$ ，对应转移矩阵  $P_{z,z'}$  (transition matrix)

$$P_{z,z'} = \begin{pmatrix} p_{11} & \dots & p_{1n} \\ p_{21} & \dots & p_{2n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \dots & p_{nn} \end{pmatrix}, \quad (12.5)$$

其中转移概率 (transition probability)  $p_{i,j}$  表示当前期处于位置  $i$  马尔科夫链下一时期转移到位置  $j$  的概率。

将 AR(1) 过程  $z_t$  离散化的常见算法，见附录第12.D节。

(注意，建议阅读这个附录，以便弄明白是怎么回事，还有附带的 matlab 代码。)

将技术冲击离散化后，我们的研究目标就变为，寻找  $2 \times n$  个决策方程：

$$\begin{aligned} c(k, z_m) &= d^{c,m,j}(k | \theta^{m,c}) = \sum_{i=0}^j \theta_i^{m,c} \psi_i(k), \\ k(k, z_m) &= d^{k,m,j}(k | \theta^{m,k}) = \sum_{i=0}^j \theta_i^{m,k} \psi_i(k), \end{aligned}$$

其中  $m = 1, 2, \dots, n$ 。举例来说，我们首先寻找当今天的技术水平是  $z_1$  时资本和消费的决策方程，进而寻找当今天的技术水平是  $z_2$  时资本和消费的决策方程，随后  $z_3, \dots, z_m, \dots, z_n$ 。为了让数值运算不至于太过复杂，我们常将  $n$  控制在一个比较小的值上，比如 5 到 7 之间。

在求得两个决策方程后，代回到原系统的欧拉方程

$$u'(c_t) = \beta E_t [u'(c_{t+1}) (\alpha \exp(z_{t+1}) k_{t+1}^{\alpha-1} + 1 - \delta)] \quad (12.6)$$

<sup>9</sup> 我们不对马尔科夫链做过多介绍 (来不及把笔记敲进硬盘了)。讲义可参考帝国理工大学 Emma J McCoy M3S4/M4S4 - Applied Probability 的讲义第六章:Markov Chains, <http://101.96.10.63/wwwf.imperial.ac.uk/~ejm/M3S4/NOTES3.pdf>。教材可参考Privault (2013)。

中，我们有

$$\begin{aligned} u' \left( d^{c,m,j} (k|\theta^{m,c}) \right) &= \beta \sum_{l=0}^n p_{ml} \left[ u' \left( d^{c,l,j} (d^{k,m,j} (k|\theta^{m,k}) |\theta^{l,c}) \right) \right. \\ &\quad \left. \left( \alpha \exp(z_{t+1}) (d^{k,m,j} (k|\theta^{m,k}))^{\alpha-1} + 1 - \delta \right) \right] \end{aligned} \quad (12.7)$$

(12.7)中有两点值得注意

1. 近似算法中  $2 \times n$  个决策方程都是当期的，不过上式中  $t+1$  的决策方程，依然考虑到  $t+1$  期可能出现的技术水平变化。
2. 由于我们将随机过程进行离散化近似，对应地，积分形式的 RHS 被简化为离散求和形式，乘以转移矩阵(12.5)中的相应元素<sup>10</sup>。

因此，在存在多维度问题的 DSGE 模型中，常常可以对其中至少一部分状态变量如技术冲击做离散化处理，其优点在于简单直观，操作过程透明，并且不是特别耗费计算资源。以及在一定意义上，求解 DSGE 模型往往依赖于混合策略：对一部分连续状态变量做离散化处理，对剩余的连续变量采用其他近似策略，如张量、完全多项式等，见下文。

### 12.3.2 张量与完全多项式

张量 (tensors) 将一组单维的基方程，用克罗内克乘积<sup>11</sup>的组合在一起，构成多维基方程。例如一个经济系统中有两个状态变量，实物资本  $k_t$  和人力资本  $h_t$ ，每个状态变量分别对应 3 个切比雪夫多项式：

$$\begin{aligned} \psi_0^k(k_t), \psi_1^k(k_t), \psi_2^k(k_t), \\ \psi_0^h(h_t), \psi_1^h(h_t), \psi_2^h(h_t). \end{aligned}$$

则我们可以构建一个张量作为基方程

$$\begin{aligned} \psi_0^k(k_t)\psi_0^h(h_t), \psi_0^k(k_t)\psi_1^h(h_t), \psi_0^k(k_t)\psi_2^h(h_t), \\ \psi_1^k(k_t)\psi_0^h(h_t), \psi_1^k(k_t)\psi_1^h(h_t), \psi_1^k(k_t)\psi_2^h(h_t), \\ \psi_2^k(k_t)\psi_0^h(h_t), \psi_2^k(k_t)\psi_1^h(h_t), \psi_2^k(k_t)\psi_2^h(h_t). \end{aligned}$$

进而，对于一个有  $n$  个状态变量的方程系统  $d : [-1, 1]^n \rightarrow \mathbb{R}$ ，我们试图用  $j$  次切比雪夫多项式予以近似，则可用如下张量形式表现

$$d^j(\cdot|\theta) = \sum_{i_1=0}^j \sum_{i_2=0}^j \dots \sum_{i_n=0}^j \theta_{i_1, i_2, \dots, i_n} \psi_{i_1}^1(\cdot) * \dots \psi_{i_n}^n(\cdot),$$

其中  $\psi_{i_\kappa}^\kappa$  是第  $\kappa$  个状态变量的  $i_\kappa$  次切比雪夫多项式， $\kappa = 1, 2, \dots, n$ ；系数向量  $\theta$  为  $\{\theta_{i_1}, \theta_{i_2}, \dots, \theta_{i_n}\}$ 。

采用张量基方程有如下优点，

1. 易于构建
2. 传递性：如果组成张量的单维基方程是正交的，那么由单维基乘积组成的张量基方程也是正交的。

<sup>10</sup>也有一些研究直接从积分形式入手，探讨一系列求积法，可参考 Judd (1998); Judd et al. (2011a)。

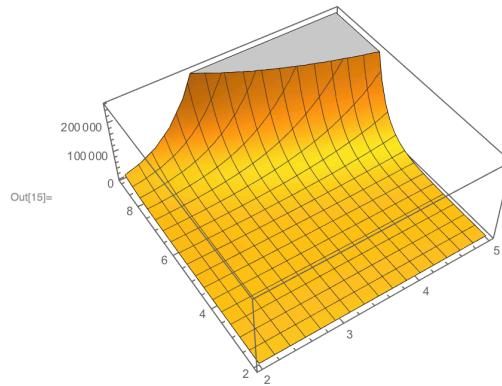
<sup>11</sup>见式(12.51)。

最显著的不足在于维数灾难：待求解系数  $\theta_{i_1, \dots, i_n}$  的个数随着  $j$  和  $n$  的增加而指数增加，呈  $(j+1)^n$  形。上例中  $j = 2, n = 2$  对应需要近似计算 9 个系数；图12.1。

$$(j+1)^n = \begin{cases} 256 & (j, n) = (3, 4) \\ 3125 & (j, n) = (4, 5) \\ 46656 & (j, n) = (5, 6) \\ \dots \end{cases}$$

图 12.1: 张量代求解系数的数量 ( $j = 1)^n$

```
In[14]:= f[j_, x_] := (j + 1)^x;
Plot3D[f[j, x], {j, 2, 5}, {x, 2, 9}]
绘制三维图形
```



在实际研究中，当模型里有  $n > 3$  个连续状态变量和 1 个哪怕是中等规模的  $j$  时，张量基的方法便不易展开了。一个方法是去掉张量集合中的部分元素，以减轻计算负担。Gaspar and Judd (1997) 建议使用完全多项式 (complete polynomials)：

$$\begin{aligned} \mathcal{P}_\kappa^n &\equiv \{\psi_{i_1}^1 * \dots * \psi_{i_n}^n\}, |\mathbf{i}| \leq \kappa, \\ |\mathbf{i}| &= \sum_{l=1}^n i_l, \quad 0 \leq i_1, \dots, i_n, \end{aligned} \tag{12.8}$$

即是说，首先预设一个正整数值  $\kappa$ ，进而在张量中，只选取各个基方程的级数之和小于  $\kappa$  所对应的那部分元素。这是基于下述认识：对于描述原方程系统  $d$  的目标而言，绝大多数信息都已经在完全多项式  $\mathcal{P}_\kappa^n$  中体现出来了；余下的部分  $\psi_{i_1}^1 * \dots * \psi_{i_n}^n, |\mathbf{i}| > \kappa$  只能给基方程产生有限的信息，却是以大量额外计算时间为代价的。

举例来说，对于  $j = 4, n = 3$  的张量系统，有  $(4+1)^3 = 125$  个元素。我们设  $\kappa = 6$ ，提取  $|\mathbf{i}| \leq 6$  的部分构建完全多项式  $\mathcal{P}_6^3$ ，则只需要近似计算其中的 87 个系数了。

不幸的是，这仍然太多。在下文中我们进一步介绍斯莫尔亚克稀疏网格算法。

## 12.4 局部基的选取 (2): 有限元法

作为局部基方法的一种，有限元法 (finite elements method, FEM) 由McGrattan (1996) 最先在经济学研究中倡导并予以实践<sup>12</sup>。

有限元法常应用在一些关键任务的工业领域，如航空航天，核电厂施工等，这是由它的突出优势所决定的：可以很方便描述某些局部域内的行为，以及达到相当高的精确度。相应地，有限元法的主要不足在于难以编程，计算求解速度慢。

利用有限元法展开分析，常见的步骤如下。第一步是确定状态变量的值域  $\Omega$ 。值域的选取方面，一些规则是天然的（如  $k_t > 0$ ），另一些则不是 ( $k_t < \bar{k}$ )，后者我们需要额外注意：如，可以将  $\bar{k}$  设的足够高，以至于在模型仿真过程中所有  $k_t$  都在上限  $\bar{k}$  的下方。随着仿真过程的实际展开， $\bar{k}$  的值也可能做微调。

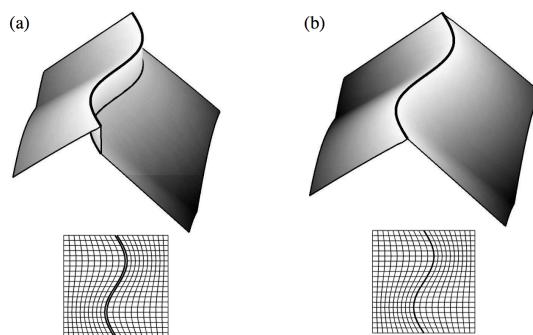
第二步是将  $\Omega$  分成若干彼此不重叠的有限个元。各个元相接的部分称为结点 (nodes)。分元的原则总体上来讲灵活性很高，可以

- 均等划分，直观简便
- 根据变量处于某种状态的频率，划分为一系列小块（频率高）和大块（频率低）。对“频率”的设定又分为两种情况
  - 来自理论模型系统的特征
  - 利用迭代法，对所划分元的事后验证
- 在  $\Omega$  中，将  $d(\cdot)$  变化幅度较大的部分划分为小元，相对变化幅度不大的部分划分为大元，从而在每一个元内部，原本是非线性的  $d(\cdot)$  接近线性。

见下文，  
reference。

在工程实践中，得益于灵活的分元，有限元法可以更好处理 kinks 或者一些特殊限制条件<sup>13</sup>。面对这些 kinks、或限制条件，一方面光谱法的操作难度较大，另一方面由于不满足可微、连续等条件，扰动法很难展开。

图 12.1: 弱 kinks 和强 kinks 的示例图



图片来源: (Fries and Belytschko, 2010, Figure 2)。

将  $\Omega$  域分为有限个元素的最优方法 (网格生成 grid generation)，已有大量工程学、数学方面的讨论，如Thompson et al. (1985)。在内生 NCGT 模型中应用网格优化法生成有限个不均等的元以优化计算时间，可见Fernández-Villaverde and Rubio-Ramírez (2004)。

<sup>12</sup>有限元法的数理知识，可见教材Hughes (2000); Brenner and Scott (2008)。可参考 Joseph E. Flaherty 的讲义 CSCI, MATH 6860: Finite Element Analysis <http://www.cs.rpi.edu/~flaherje/FEM/index4.html>。

<sup>13</sup>kinks 通常是指，分元后，两个元的衔接处 (interface) 出现偏移 (弱 kinks) 甚至断裂 (强 kinks) 的情况，见图12.1。

第三步，在每一个元中选择对应的基，用于构建政策方程。如果前述步骤中元素的划分已经有效，则这一步中的基只设定线性形式即可。例如， $\Omega$  中诸元的节点表示为  $\{k_0, k_1, \dots, k_j\}$ ，则可以定义基方程  $\psi_i(k)$  为三角函数形式， $i = 1, 2, \dots, j - 1$ ：

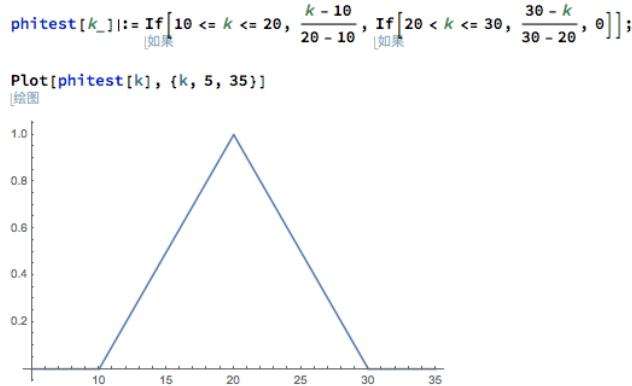
$$\psi_0(k) = \begin{cases} \frac{k_0 - k}{k_1 - k}, & \text{如果 } k \in [k_0, k_1], \\ 0, & \text{否则} \end{cases} \quad (12.9a)$$

$$\psi_i(k) = \begin{cases} \frac{k - k_{i-1}}{k_i - k_{i-1}}, & \text{如果 } k \in [k_{i-1}, k_i], \\ \frac{k_{i+1} - k}{k_{i+1} - k_i}, & \text{如果 } k \in [k_i, k_{i+1}], \\ 0, & \text{否则} \end{cases} \quad (12.9b)$$

$$\psi_j(k) = \begin{cases} \frac{k - k_{j-1}}{k_j - k_{j-1}}, & \text{如果 } k \in [k_{j-1}, k_j], \\ 0, & \text{否则} \end{cases} \quad (12.9c)$$

之所以说(12.9)的基方程是三角形式，举例说明，见图12.2。

图 12.2: 三角方程形式



第四步，与其他映射法类似，生成决策方程系统作为原方程系统的近似，在有限元分析中我们常称之为分段线性近似 (piecewise linear approximation)。

$$d^{n,j}(\cdot | \theta^n) = \sum_{i=0}^j \theta_i^n \psi_i(\cdot),$$

将近似系统代回  $\mathcal{H}$ 。其中， $\theta$  和  $\psi$  由切比雪夫多项式的相关算法求得：以本例中的有限元节点  $k_i$  和  $k_{i+1}$  为例，分别对应基方程

$$\begin{aligned} \psi_i(k) &= \frac{k_{i+1} - k}{k_{i+1} - k_i}, \\ \psi_{i+1}(k) &= \frac{k - k_i}{k_{i+1} - k_i}, \end{aligned}$$

那么由节点  $k_i$  和  $k_{i+1}$  划出的元中，值  $d^{n,j}(\cdot | \theta^n)$  由一个近似方程  $\hat{d}$  表示

$$\begin{aligned} \hat{d}(k | k_i, k_{i+1}, \theta_i^n, \theta_{i+1}^n) &= \theta_i^n \psi_i(k) + \theta_{i+1}^n \psi_{i+1}(k) = \theta_i^n \frac{k_{i+1} - k}{k_{i+1} - k_i} + \theta_{i+1}^n \frac{k - k_i}{k_{i+1} - k_i} \\ &= \frac{(\theta_{i+1}^n - \theta_i^n)k + \theta_i^n k_{i+1} - \theta_{i+1}^n k_i}{k_{i+1} - k_i}, \end{aligned}$$

不难看出, 近似方程  $\hat{d}$  是线性的, 斜率的符号由  $\theta_{i+1}^n - \theta_i^n$  来决定。

另一方面采用类似的方法, 以有限元结点  $k_{i-1}$  和  $k_i$  划出的元中, 我们可以得到另一个线性近似方程  $\hat{d}$

$$\hat{d}(k|k_{i-1}, k_i, \theta_{i-1}^n, \theta_i^n) = \frac{(\theta_i^n - \theta_{i-1}^n)k + \theta_{i-1}^nk_i - \theta_i^nk_{i-1}}{k_i - k_{i-1}}$$

两个元在  $k_i$  处相交, 此时我们有

$$\begin{cases} \hat{d}(k_i|k_i, k_{i+1}, \theta_i^n, \theta_{i+1}^n) = \theta_i^n, \\ \hat{d}(k_i|k_{i-1}, k_i, \theta_{i-1}^n, \theta_i^n) = \theta_i^n, \end{cases}$$

即两个方程等于一个相同的值  $\theta_i^n$ , 这确保了作为各个  $\hat{d}$  的加总的  $d$  方程是连续的。

上述介绍也说明为何有限元法提供的分段线性近似是个较为理想的近似策略。假定我们设对应的度量方程  $\rho(\cdot)$  的形式为<sup>14</sup>, 使得在划分元所依据的每个节点  $k_i$  上, 都有残差方程  $R(\cdot) = 0$ 。对于三角函数形式的分段基方程  $\psi_i(\cdot)$  而言, 这意味着当状态变量处在  $k_i$  节点上时, 对系数  $\theta_i^n$  的取值应当使得近似方程系统与原方程系统相等:

$$\hat{d}^{n,j}(\cdot|\theta^n) = d^n(\cdot).$$

需要指出的是, 有限元分析中的分段线性近似法也表明, 系数  $\theta_i^n$  的选取, 与状态变量处于节点  $k_i$  之外其他位置时  $d^n(\cdot)$  的值无关。从这个意义上讲, 利用有限元法求解大型非线性方程系统所得到的近似系统, 都是稀疏的 (sparse)。这成为一个可为现代非线性求解方法所利用的特征。

第五步, (如有需要) 对结果的改进。如果第一轮有限元分析近似解的精度不达标, 我们可以根据需要对近似解做反复改进, 进行第二轮、第三轮甚至更多有限元分析, 在计算时间和内存允许的范围内, 尽可能提高近似解的精度。事实上这是有限元法研究的又一突出优势。在现有研究文献中, 改进常常分为三大类型。

1. h——改进 (h-refinement), 在全部域  $\Omega$  内, 将第一轮所分元 (如  $A, B, C, \dots$ ) 再均等分为更小的元 (如  $A1, A2, A3, \dots$ ), 对这些细分元再反复迭代使用有限元法, 改善近似解, 直至精度达标。
2. r——改进 (r-refinement), 第二、第三甚至更多轮有限元分解的重点针对存在着明显非线性特征的局部。
3. p——改进 (p-refinement), 不对现有元再作细分, 而是在元内部通过加入更多个基方程 (如更多的切比雪夫多项式) 来提高近似的阶数 (order); 如果元内现有阶数已足够高, 则应用有限元法和光谱法组合的混合策略, 常称为光谱元法 (spectral elements)。在自然科学和工程学领域, 光谱元法已得到了广泛应用 ([Šolín et al., 2003](#))。

有时, h——和 p——改进被混合在一起使用, 称 hp——有限元法 (hp-adaptive finite elements), 可以使近似解以指数速度向真实解收敛 ([Ciarlet, 2002](#))。尽管它的编程难度更高, 计算时间更长, 但 hp——有限元法可能是目前已知最强大的 DSGE 求解工具了, 有助于求解甚至是最复杂的 DSGE 模型。关于这种方法的详细介绍, 可参考 [Babuška and Suri \(1994\)](#); [Demkowicz \(2006\)](#); [Demkowicz et al. \(2007\)](#)。

<sup>14</sup> 度量方程的选取方案有数种, 这里只举其中之一; 更多讨论见下节。

## 12.5 目标方程的选取

如前所述，映射法研究中也需要选取度量方程  $\rho$  作为目标方程。在未对  $\mathcal{H}$  作过多限定的前提下，可视  $R(\cdot|\theta)$  为最简单的单维情况<sup>15</sup>。此时对  $\rho(R(\cdot|\theta), \mathbf{0})$  的选取目标可设定为，使用加权残差法，选取合适的  $\theta$  向量，使得加权残差之的积分最接近于零，对应某个权重方程  $\phi_i : \Omega \mapsto \mathbb{R}$ ：

$$\rho(R(\cdot|\theta), \mathbf{0}) = \begin{cases} 0 & \text{如果 } \int_{\Omega} \phi_i(\mathbf{x}) R(\cdot|\theta) d\mathbf{x}, \quad i = 1, 2, \dots, j+1, \\ 1, & \end{cases}$$

这样问题变为在给定  $j+1$  个权重方程  $\phi_i$  的前提下，选择  $\theta$  的值来求解积分系统

$$\int_{\Omega} \phi_i(\mathbf{x}) R(\cdot|\theta) d\mathbf{x}, \quad i = 0, 1, \dots, j+1, \quad (12.10)$$

对此我们有一系列常用解法可供选择，如规模较小的系统可用牛顿算法 (Newton algorithm)，规模较大的系统可用莱文贝格——马夸特方法 (Levenberg-Marquardt algorithm)，等。然而需要指出的是，系统(12.10)可能无实解或者有多个解。关于如何将映射法应用到经济学经验研究中的理论依据，到目前为止我们所知不多——应用数学研究中大量关于映射法的文献，涉及解的存在性、收敛特性等问题的研究，主要针对自然科学和工程学领域，它们并不完全适用于经济学。事实上，对诸如(12.10)的经济系统而言，需要确保解满足 DSGE 模型的横截条件 (transversality condition)，从而使得状态变量处于稳定域内——在实际求解过程中，这边需要我们选择合适的初始猜测系数  $\theta_0$ ，或是在求解过程中加入边界条件。

与基方程  $\psi_i$  类似，关于权重方程  $\phi_i$  也存在一系列选择方案。下面我们介绍一下经济学研究中常见的权重方程的设定方法。

### 12.5.1 最小方差

将研究目标理解为如下变分法问题 (variational problem)

$$\min_{\theta} \int_{\Omega} R^2(\cdot|\theta) d\mathbf{x},$$

一阶条件为

$$\int_{\Omega} \frac{\partial R(\mathbf{x}|\theta)}{\partial \theta_{i-1}} R(\cdot|\theta) d\mathbf{x}, \quad i = 1, 2, \dots, j+1.$$

则权重方程的选取方案之一是将其定义为

$$\phi_i(\mathbf{x}) = \frac{\partial R(\mathbf{x}|\theta)}{\partial \theta_{i-1}}. \quad (12.11)$$

映射法研究中，利用变分法问题设定权重方程，其思路与计量经济学中的回归问题相近似（见第12.1.1节）。

优缺点：

1. 优点：直观，易于理解。

2. 缺点：

(a) 最小方差及几种变体都要求计算  $\frac{\partial R(\mathbf{x}|\theta)}{\partial \theta_{i-1}}$ ，导致计算成本高。

(b) 最小方差问题过于复杂，条件苛刻，难于数值求解。

<sup>15</sup>随着  $\mathcal{H}$  的限定条件变多， $R(\cdot|\theta) = \mathcal{H}(d^j(\cdot|\theta))$  可能变成多维度的，但在这里我们暂不做讨论。

### 12.5.2 子域

将状态变量的域  $\Omega$  通过一系列灵活的划分原则，分为  $j + 1$  个子域 (subdomain)  $\Omega_i, i = 1, 2, \dots, j + 1$ ，则研究目标可以理解为如下  $j + 1$  个求积问题

$$\int_{\Omega_i} R(\cdot|\theta) d\mathbf{x} = 0, \quad i = 1, 2, \dots, j + 1.$$

则子域法将权重设为如下  $j + 1$  个分段方程

$$\phi_i(\mathbf{x}) = \begin{cases} 1 & \text{如果 } \mathbf{x} \in \Omega_i, \\ 0. & \end{cases}$$

优缺点基本同于第12.5.1节的最小方差法。

### 12.5.3 配点

#### 配点法

配点法 (collocation method)，又称伪光谱法 (pseudospectral method) 或选点法 (selected points method)，是指在状态变量域  $\Omega$  中选取  $j + 1$  个配点  $\mathbf{x}_i$ ，将权重方程定义为

$$\psi_i(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}_i), i = 1, 2, \dots, j + 1,$$

其中  $\delta$  是狄拉克方程 (Delta Dirac function, 第16.B.5节)，满足

$$\delta(\mathbf{y}_i) = \begin{cases} +\infty, & \text{如果 } \mathbf{y}_i \equiv \mathbf{x} - \mathbf{x}_i = 0, \\ 0, & \text{如果 } \mathbf{y}_i \neq 0, \end{cases}$$

$$\int_{-\infty}^{+\infty} \delta(\mathbf{y}_i) d\mathbf{y}_i = 1.$$

配点法假定当状态变量恰好处于选点时  $\mathbf{x} = \mathbf{x}_i$ ，权重的值为 0，残差方程因此也为 0。此时不必做复杂的求积计算，只需求解  $j + 1$  个方程系统

$$R(\mathbf{x}_i|\theta) = 0, \quad j = 1, 2, \dots, j + 1.$$

当原方程系统  $\mathcal{H}$  表现出较明显的非线性特征时，配点近似方法具有较大优势。

#### 配点的选取：正交配点法

$j + 1$  个配点的选取方法有很多，一个常用方法称正交配点法 (orthogonal collocation)，是指在值域  $\Omega$  中对于向量  $\mathbf{x}$  对应的每一组状态变量，都用  $j + 1$  次切比雪夫多项式的根 (有  $j + 1$  个) 予以表示。根据切比雪夫插值定理 (定理12.56)，利用正交配点法生成的近似方程系统，会收敛至、甚至有时是均匀收敛 (uniform convergence) 至未知方程  $d$ 。

### 12.5.4 伽辽金法

第四种权重方程  $\phi_i(x)$  的选取方法为，设它等于近似计算过程中求得的基方程  $\psi_{i-1}(x)$ ，满足

$$\phi_i = \psi_{i-1}(x),$$

又称伽辽金近似法 (Galerkin method, 第12.E节)。

对应地，残差  $R(\cdot|\theta)$  的加权积分为 0

$$\int_{\Omega} \psi_i(x) R(\cdot|\theta) dx = 0, \quad i = 1, \dots, j+1,$$

即是说残差项正交于每一个对应的基方程。

伽辽金法的编程代码难写，但优点在于近似结果高度精确且稳健。如果基方程  $\psi_i(x), i = 1, \dots, j+1$  在  $J$  上完备<sup>16</sup>，那么随着  $j \rightarrow \infty$ ，伽辽金近似会逐渐点收敛 (pointwise) 至真实解

$$\lim_{j \rightarrow \infty} d^j(\cdot|\theta) = d(\cdot),$$

相关证明可参考 Theorem 12.60, Theorem 12.61。

实际研究结果表明， $j$  阶伽辽金近似的精确度，可以达到  $j+1$  或者  $j+2$  阶配点近似 (伪光谱法) 的程度。

### 12.5.5 两个近似技巧

考虑这样一个经典的非线性方程系统

$$\int_{\Omega} \phi_i(x) R(\cdot|\theta) dx, \quad i = 1, \dots, j+1, \tag{12.12}$$

当系数向量  $\theta$  中的元素较多时，和/或我们很难得到适当的初始猜测值  $\theta_0$  时，近似求解的计算会变得很困难。实际近似过程中，我们常用以下两个技巧来提高计算速度与近似解的稳健性。

#### 系统变形

求解系统(12.12)最大的瓶颈在于系统的非线性，并且求解难度随着非线性的程度而指数上升。因此在进行数值近似之前，常可以对系统做变形来降低非线性的程度。例如跨期消费的欧拉方程

$$\frac{1}{C_t} = \beta E_t \frac{1}{C_{t+1}} R_{t+1},$$

其中  $R_{t+1}$  表示资本的总回报率。Judd (1992) 建议系统变形如下

$$\beta C_t = \left( E_t \frac{1}{C_{t+1}} R_{t+1} \right)^{-1},$$

这样 LHS 是线性形式，RHS 比原式更接近线性形式了。对于某个状态变量  $x_t$  而言，在原来的系统中，残差  $R(\cdot|\theta)$  的计算方式如下

$$R(\cdot|\theta) = \frac{1}{C_t(x_t|\theta)} - \beta E_t \left[ \frac{1}{C_{t+1}(x_t|\theta)} R_{t+1}(x_t|\theta) \right],$$

若是采用系统变形的思路，变形后的残差  $\tilde{R}(\cdot|\theta)$  计算式为

$$\tilde{R}(\cdot|\theta) = \beta C_t(x_t|\theta) - \left[ E_t \frac{1}{C_{t+1}(x_t|\theta)} R_{t+1}(x_t|\theta) \right]^{-1}.$$

---

<sup>16</sup>完备 (completeness) 的定义，见第332页脚注。

### 系统规模缩减

如前所述，我们说(12.12)的规模较大时，往往是指系数  $\theta$  的数量较多。在近似求解过程中，常常可以分步骤进行：首先在大系统中提取一组核心变量构成小系统，近似求解小系统，将近似解输入大系统中求得最终的完整解，这个思路可称为分步求解法。当系统中含有正交基方程（如切比雪夫多项式）时，分步求解法显得非常可靠：不再直接用  $j+1$  个基方程来直接求解大系统，而是首先提取出  $j'+1 << j+1$  个基方程做近似求解小系统，将其作为猜测接代入大系统中，展开下一个阶段的近似。

举例来说，我们想要求一个  $m$  维系统中，含有 10 个切比雪夫多项式的， $10 \times m$  个方程的系统。首先我们做 3 个切比雪夫多项式， $3 \times m$  个方程的小系统求解，将近似得到的系数作为大方程系统的初始猜测值的一部分， $\theta_0 = [\theta^3, 0_{1 \times m}, \dots, 0_{1 \times m}]$ ，即第一组系数  $\theta^3$  来自小系统的计算，其余更多系数的初始猜测值设为 0。由于前 3 个和后 7 个切比雪夫多项式正交，那么引入额外的 7 个 0 不会导致系统解发生太大改变：将  $\theta^3$  作初始猜测值因而是非常理想的。此外，由于切比雪夫多项式具有快速收敛的特点，越是高阶的切比雪夫多项式，其对应的系数就越接近于 0，那么将  $\theta^3$  作为初始猜测值的一部分，也是有信息含量的<sup>17</sup>。

## 12.6 投影法求解 DSGE 模型举例

举一个实际案例，介绍如何应用投影法，即切比雪夫多项式和正交配点等技术，来求解 DSGE 模型。

### 12.6.1 模型设定

考虑一个经典的含有内生劳动力供应决定机制的随机内生经济增长模型。假定这是个中心化模型 (centralized economy)<sup>18</sup>。典型家庭的消费和劳动力供应决策，由前向 CCRA 效用函数来表现

$$E_0 \sum_{t=1}^{\infty} \beta^{t-1} \frac{[c_t^{\tau} (1 - \ell_t)^{1-\tau}]^{1-\eta}}{1 - \eta},$$

其中

- $0 < \beta < 1$  表示时间贴现，
- $\eta$  是跨期替代弹性，反映家庭的风险厌恶程度，
- $1/\tau$  是劳动力供应的 Frisch 替代弹性，见第3.B节。
- $E_0$ ,  $t = 0$  时刻的条件期望符。

经济体中有且只有一种最终商品，总量生产函数

$$y_t = \exp(Z_t) k_t^{\alpha} \ell_t^{1-\alpha}, \quad (12.13)$$

其中

---

<sup>17</sup>上例将整个系统截为两段分步展开研究。随着实际需要，可以分为更多段，在写投影法的代码时，可将这些段写为与特定数量的多项式有关，进而在一个 loop 循环中迭代运行程序，使得由  $j'$  向  $j$  的近似以较高速度或者较低速度进行。

<sup>18</sup>用中心化经济模型来介绍投影法的应用，主要原因是... 懒... 投影法一样可以用于求解去中心化的经济模型，又称竞争均衡模型。并且事实上，投影法的最突出优势恰恰在于，它更适合用于分析常常处于非帕累托最优状态的分散经济体，如见第219页的讨论。

- $k_t$  是总的实物资本存量，积累形式满足运动法则

$$k_{t+1} = (1 - \delta) k_t + i_t,$$

$i_t$  表示当期的实物资本投资，

- $\ell_t$  是总劳动力，
- $z_t$  表示生产率，是一个随机  $AR(1)$  过程，满足

$$z_t = \rho z_{t-1} + \epsilon_t, \quad |\rho| < 1, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2),$$

经济体中的总资源约束条件

$$y_t = c_t + i_t.$$

### 12.6.2 社会规划者问题

社会规划者问题可表示为，在给定初始条件  $\{k_0, z_0\}$  的情况下，追求价值方程  $V(k_t, z_t)$  的最大化，即贝尔曼等式 (Bellman equation)

$$\begin{aligned} V(k_t, z_t) &= \max_{\{c_t, \ell_t\}} \frac{\left[ c_t^\tau (1 - \ell_t)^{1-\tau} \right]^{1-\eta}}{1-\eta} + \beta E_t V_{k_{t+1}, z_{t+1}}, \\ \text{s.t. } &\begin{cases} k_{t+1} = \exp(z_t) k_t^\alpha \ell_t^{1-\alpha} + (1 - \delta) k_t - c_t, \\ z_t = \rho z_{t-1} + \epsilon_t. \end{cases} \end{aligned} \tag{12.14}$$

实际求解过程中我们只需要考虑状态变量  $k_t$  和决策变量  $\ell_t$ ； $c_t$  则时一个关于  $k_t$  和  $\ell_t$  的变量。为了说明这一点，首先我们有

$$\begin{aligned} MPL &:= \frac{\partial y_t}{\partial \ell_t} = (1 - \alpha) \exp(z_t) k_t^\alpha \ell_t^{-\alpha}, \\ MUC &:= \tau c_t^{\tau-1} (1 - \ell_t)^{1-\tau} [.]^{-\eta}, \\ MUL &:= -\tau (1 - \tau) c_t^\tau (1 - \ell_t)^{-\tau} [.]^{-\eta}, \end{aligned}$$

其中  $MPL$  表示劳动的边际产出， $MUC, MUL$  分别表示消费和劳动的边际效用 (负效用)。

进而可得

$$\begin{aligned} MUC + MUL \frac{1}{MPL} &= 0, \\ \hookrightarrow [.]^{-\eta} \tau c_t^{\tau-1} (1 - \ell_t)^{1-\tau} &= \frac{[.]^{-\eta} (1 - \tau) c_t^\tau (1 - \ell_t)^{-\tau}}{(1 - \alpha) \exp(z_t) k_t^\alpha \ell_t^{-\alpha}}, \\ \hookrightarrow \frac{\tau}{1 - \tau} (1 - \ell_t) &= \frac{c_t}{(1 - \alpha) \exp(z_t) k_t^\alpha \ell_t^{-\alpha}}, \\ \hookrightarrow c_t &= \frac{\tau}{1 - \tau} (1 - \alpha) \exp(z_t) k_t^\alpha \ell_t^{-\alpha} (1 - \ell_t). \end{aligned} \tag{12.15}$$

### 12.6.3 参数校准

求解模型需要为参数赋值。本节先用参数校准，在下文介绍 DSGE 的估计方法时将进一步讨论参数赋值的方法。

$$(\beta, \eta, \tau, \alpha, \delta, \rho, \sigma) = (0.991, 5.000, 0.357, 0.300, 0.0196, 0.950, 0.007).$$

大多数参数值的选取都与经验研究中的发现相近。唯一的例外是  $\eta = 5$ , 接近一系列经验测算值中的最高值。 $\eta$  的值越高, 意味着消费者的风险厌恶程度越高, 消费行为越是缺乏耐心。在 DSGE 模型中之所以取如此高的  $\eta$ , 是为了使模型中的行为人对市场中的波动更加敏感, 经济行为具有更高的预防倾向 (precautionary behaviors), 进而决策法则的曲率 (curvature) 更高。而曲率越高, 就表示方程系统的非线性程度越高, 从而投影法近似的难度就越大。

除此而外, 还需要对  $AR(1)$  的外生技术冲击过程  $z_t$  做离散化处理 (discretization), 见第12.D节。设  $n = 5$ , 即  $\{z_1, z_2, z_3, z_4, z_5\}$ 。无条件标准差设为  $\pm 3$ 。进而我们有转移矩阵  $P_{z,z'}$ , 矩阵中的元素  $p_{m,n}$  表示当前技术水平  $z_m$ , 在下一时间段变为  $z_n$  的概率。

#### 12.6.4 分步求解

在无法直接求得系统(12.6.2)解析解的情况下, 近似求解的目标可以分步表示如下, 所需的动态规划知识, 见第八章。

第一阶段, 近似价值方程  $V(k_t)$  和劳动力供应的决策法则  $\ell(k_t)$ , 设为  $\{V^j(k_t|\theta^{V,j}), \ell^j(k_t|\theta^{\ell,j})\}$ , 其中系数为  $\{\theta^{\ell,j}, \theta^{V,j}\}$ 。

设 5 个马尔科夫链用  $j = 1, 2, 3, 4, 5$  表示, 11 个切比雪夫结点  $T_i(\cdot)$  用  $i = 0, 1, \dots, 10$  表示, 定义式如(12.96)。那么价值方程和劳动力供应的决策法则可以近似表示为

$$V^j(k_t|\theta^{V,j}) = \sum_{i=0}^{10} \theta_i^{V,j} T_i(k_t), \quad (12.16)$$

$$\ell^j(k_t|\theta^{\ell,j}) = \sum_{i=0}^{10} \theta_i^{\ell,j} T_i(k_t), \quad (12.17)$$

第二阶段, 已知  $\ell^j(k_t|\theta^{\ell,j})$  的近似值(12.17)计算而得, 在此基础上可计算近似值  $\{y^j(k_t), c^j(k_t), k^j(k_t)\}$ :

- 由总量生产函数(12.13)可得  $y^j(k_t)$

$$y^j(k_t) = \exp(z_t) k_t^\alpha [\ell^j(k_t|\theta^{\ell,j})]^{1-\alpha}. \quad (12.18)$$

- 由(12.15)可得  $c^j(k_t)$

$$c^j(k_t) = \frac{\tau}{1-\tau} (1-\alpha) \exp(z_t) k_t^\alpha [\ell^j(k_t|\theta^{\ell,j})]^{-\alpha} [1 - \ell^j(k_t|\theta^{\ell,j})]. \quad (12.19)$$

- 下一时段的资本存量由(12.14)给出

$$k^j(k_t) = \exp(z_t) K_t^\alpha [\ell^j(k_t|\theta^{\ell,j})]^{1-\alpha} + (1-\delta) k_t - c^j(k_t). \quad (12.20)$$

值得指出的是, 上面的求解过程中, 我们设一个状态变量  $k_t$ , 一个决策变量  $\ell_t$  为第一阶段投影近似的对象。而类似地, 也可以将分步求解的方案设计为, 第一阶段投影近似消费和资本存量的决策方程, 第二阶段投影近似劳动力决策和价值方程。这两个方案是等价的。这从另一个侧面印证了投影法在求解分散化的去中心经济系统时 (各个部门的子系统由竞争均衡条件彼此联系起来) 时, 具有高度的灵活性。在实际研究中具体采取哪种方案, 可以视建模难度、计算 (编程) 难度和研究目标之间的权衡取舍而灵活决定。

第三阶段, 求解近似系数  $\{\theta^{\ell,j}, \theta^{V,j}\}$ 。

为了得到系统的近似解，需要确定  $\{\theta^{\ell,j}, \theta^{V,j}\}$  的值：这可以通过价值方程来计算。将贝尔曼等式(12.14)改写为

$$V^j(k_t|\theta^{V,j}) = \frac{\left\{ [c^j(k_t)]^\tau \left[ 1 - \ell_t^j(k_t|\theta^{\ell,j}) \right]^{1-\tau} \right\}^{1-\eta}}{1-\eta} + \underbrace{\beta E_t V^j(K_{t+1}|\theta^{V,j})}_{=: \mathcal{A}}, \quad (12.21)$$

上式已经在最优决策下，因此省去  $\max$  符。根据离散化的设定，当期价值方程  $V^j(\cdot|\cdot)$  在下期变为  $V^m(\cdot|\cdot)$  的概率，由转移矩阵  $P_{z,z'}$  给出，对应其中的概率元素  $p_{jm}$ 。以  $p_{jm}$  为权重，(12.21)RHS 的后半部分  $\mathcal{A}$  可以改写为

$$\mathcal{A} := \beta E_t V^j(K_{t+1}|\theta^{V,j}) = \beta \sum_{m=1}^5 p_{jm} V^j(k^j(k_t)|\theta^{V,j}),$$

这样我们将所有对于下一时段的期望值  $E_t [\cdot]$  写为当期项的加权和，对应权重概率  $p_{jm}$ 。依次引入(12.16), (12.17), (12.19), (12.20) 替换  $V^j(k_t|\theta^{V,j}), \ell^j(k_t|\theta^{\ell,j}), c^j(k_t), k^j(k_t)$ , (12.21)最终改写为

$$\sum_{i=0}^{10} \theta_i^{V,j} T_i(k_t) = \frac{1}{1-\eta} \left\{ [c^j(k_t)]^\tau \left[ 1 - \sum_{i=0}^{10} \theta_i^{\ell,j} T_i(k_t) \right]^{1-\tau} \right\}^{1-\eta} + \beta \sum_{m=1}^t p_{j,m} \sum_{i=0}^{10} \theta_i^{V,j} T_i(k_t). \quad (12.22)$$

此外，模型中跨期消费的欧拉等式满足

$$u'(c_t) = \beta E_{\epsilon'}|_\epsilon V'(k'_t),$$

那么用类似的方式可得欧拉等式的近似方程

$$\frac{\left[ c_t^\tau \left( 1 - \sum_{i=0}^{10} \theta_i^{\ell,j} T_i(k_t) \right)^{1-\tau} \right]^{1-\eta}}{c_t} = \beta E_t \sum_{m=1}^5 p_{jm} \sum_{i=0}^{10} \theta_i^{V,j} T'_i(k^j(k_t)), \quad (12.23)$$

其中  $T'_i(k^j(k_t)) := \frac{\partial}{\partial k^j(k_t)} T_i(k^j(k_t))$ 。

将贝尔曼等式的近似式(12.22)，和欧拉等式的近似式(12.23)连起来，构建一个残差方程系统

$$R(k_t, z_j | \theta^{V,j}, \theta^{\ell,j}) = \begin{cases} \mathcal{A}_1, \\ \mathcal{A}_2, \end{cases} \quad (12.24)$$

其中

$$\begin{aligned} \mathcal{A}_1 &:= \sum_{i=0}^{10} \theta_i^{V,j} T_i(k_t) - \frac{\left\{ [c^j(k_t)]^\tau \left[ 1 - \sum_{i=0}^{10} \theta_i^{\ell,j} T_i(k_t) \right]^{1-\tau} \right\}^{1-\eta}}{1-\eta} - \beta \sum_{m=1}^t p_{j,m} \sum_{i=0}^{10} \theta_i^{V,j} T_i(k_t), \\ \mathcal{A}_2 &:= \frac{\left[ c_t^\tau \left( 1 - \sum_{i=0}^{10} \theta_i^{\ell,j} T_i(k_t) \right)^{1-\tau} \right]^{1-\eta}}{c_t} - \beta E_t \sum_{m=1}^5 p_{jm} \sum_{i=0}^{10} \theta_i^{V,j} T'_i(k^j(k_t)), \end{aligned}$$

$\theta$  是  $\theta_i^{V,j}$  和  $\theta_i^{\ell,j}$  的集合， $z_j$  是离散化处理后的外生冲击变量 (因而取消时间下角标)。

根据这样的设定不能看出，如果我们

- 用 11 个切比雪夫多项式近似价值方程式(12.16)，

- 用 11 个切比雪夫多项式近似劳动力供应的决策方程(12.17),
- 用 5 个状态水平描述外生技术冲击  $z_j$ ,

那么  $\theta$  共有  $11 \times 2 \times 5 = 110$  个元素。这 100 个系数值的求解, 可用牛顿——拉夫森 (Newton-Raphson method) 做近似求解, 程序实现如见第17.10.1节。为了提高计算效率, 可采取前述的分步求解法: 先取 3 个切比雪夫多项式计算近似解, 代入 11 个多项式的大系统中作为初始猜测值的前 3 项, 余下部分设为 0, 再进行下一轮的近似计算。

图 12.1: 随机内生经济增长模型的投影法近似求解

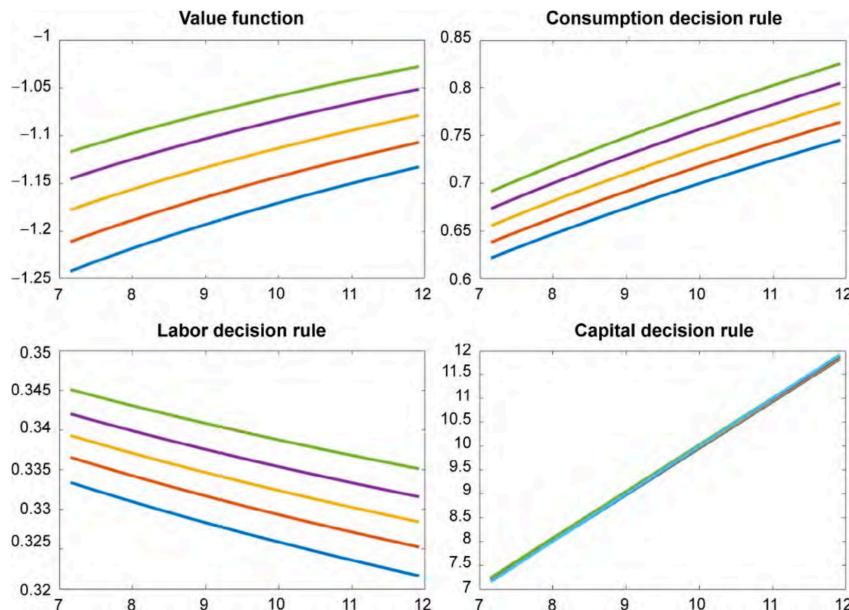


图12.1描述了用投影法求解随机新古典内生经济增长模型的模拟结果, 从左上到右下分别表示价值方程的近似, 消费、劳动力供应、资本积累决策方程的近似。五条颜色的线段分别表示 5 种技术冲击水平。横轴是状态变量  $k_t$ 。

- 价值方程的近似, 符合模型设定, 即价值方程是关于  $k_t$  和  $z_t$  的凹函数 (concave)。
- 值得注意的是右下图, 资本积累的决策方程, 基本呈近似线性特征, 这暗示着模型设定中资本积累式的线性型设定是比较合理的。

在求得 1 个价值方程和 3 个决策方程的近似式的基础上, 我们便可以进一步对模型做仿真, 计算冲击响应方程, 并展开福利效果分析了。

投影法做近似求解可以达到非常高的精度, 对误差项精度的进一步解读, 可见下一节。总的说来, 图12.1的近似解足以代替(离散生产率水平设定下的)新随机新古典主义增长模型的解析解。做一个 ref

## 12.7 稀疏网格

高维 DSGE 模型的求解普遍受到维数灾难的限制, 是公认难题。在利用完全多项式来近似求解高维系统的过程中, 一种可以更好处理维数灾难的投影法是稀疏网格法 (sparse grid), 又称 Smolyak 算法。

数学上的 Smolyak 算法介绍，可见 Smolyak (1963); Delvos (1982); Barthelmann et al. (2000)，一个全面的文献综述可见 Bungartz and Griebel (2004)。

Krueger and Kubler (2004); Malin et al. (2011) 等首次尝试将 Smolyak 算法引入经济学研究中去，尤其是 DSGE 模型的求解。随后，一系列相关工作逐渐展开，如

1. Fernández-Villaverde et al. (2015) 利用 Smolyak 算法求解一个有着利率零下限设定的 5 个状态变量的新凯恩斯模型。
2. Fernández-Villaverde and Levintal (2016) 利用 Smolyak 算法求解一个存在大规模灾难性风险设定的 12 个状态变量的新凯恩斯模型。
3. Gordon (2011) 用 Smolyak 算法求解一个动态的异质行为人模型，异质性表现为总量层面上的分布方程，以一个状态变量直接进入模型当中。
4. Malin et al. (2011) 对一个大规模、高度非线性 DSGE 模型做出精确估计，模型中含有 20 个连续状态变量。高度非线性表现一系列生产函数和效用函数的曲率上。
5. 除了 Smolyak 算法之外，稀疏网格矩阵的另一种方法，是基于遍历集合 (ergodic set) 来求解高维非线性系统，解释效果和精度也较为理想，可参考 Maliar et al. (2011); Judd et al. (2011b,a) 等。一个全面的介绍可见 Maliar and Maliar (2015)。

这里以 Krueger and Kubler (2004); Malin et al. (2011) 为例，介绍用 Smolyak 算法求解 DSGE 模型的基本思路。研究的目标依旧是近似求解一个由  $n$  个状态变量构成的方程，如决策方程、价值方程、期望方程等，表示为<sup>19</sup>

$$d : [-1, 1]^n \mapsto \mathbf{R}^m.$$

Smolyak 算法的求解思路为，找到一个由若干点组成的网格  $G(q, n)$ ,  $q > n$ ，以及一个近似方程  $d(x|\theta, q, n) : [-1, 1]^n \mapsto \mathbf{R}$ ，这个近似方程由系数向量  $\theta$  予以描述，使得

- 在网格  $G(q, n)$  中的任意一点  $x_i \in G(q, n)$ ，待求解的未知方程  $d(\cdot)$  和近似方程  $d(\cdot|\theta, q, n)$  相等，即  $\mathcal{H}(\cdot)$  算子得到充分满足

$$d(x_i) = d(\cdot|\theta, q, n), \forall x_i \in G(q, n).$$

- 在不属于网格的点  $x_i \notin G(q, n)$ ，近似方程  $d(\cdot|\theta, q, n)$  接近原方程  $d(\cdot)$ ，即残差接近于 0。

不难看出， $q$  的作用在于描述网格的大小，以及衡量近似解的精确程度。

Smolyak 算法的核心在于，如何选取恰当的网格  $G(q, n)$ ，以确保系数向量  $\theta$  不会随着  $n$  而爆炸。

Smolyak 编程较难，但有突出优点：

- 在多种基于多项式近似的算法中，它（几乎）是最优的算法 (Barthelmann et al., 2000)，
- 它提供全局解，即对许多方程空间来说都（几乎）是普遍最优的。

<sup>19</sup> 两个简单的扩展：一是

$$d : [-k, k]^n \mapsto [-1, 1]^n, k \neq 1,$$

可参考如第 12.7.1 节。二是

$$d : [-k, k]^n \mapsto \mathbf{R}^m, m \neq n, m > 0, n > 0.$$

### 12.7.1 Smolyak 法的步骤

如上文所述, Smolyak 算法的核心是选定适宜的稀疏网格  $G(q, n)$  和近似方程  $d(\cdot | \theta, q, n)$ 。以下是常见步骤。

#### 状态变量的域转换

任意状态变量  $\{\tilde{x}_\ell\}_{\ell=1}^n : [a, b]^n$ , 可用如下方法线性平移至单位域  $[-1, 1]^n$  中, 变为  $\{x_\ell\}_{\ell=1}^n : [-1, 1]^n$ :

$$x_\ell = 2 \frac{\tilde{x}_\ell - a}{b - a} - 1, \quad \ell = 1, \dots, n. \quad (12.25)$$

#### 选定多项式的阶

定义多项式的阶,  $m_i, i = 1, 2, \dots$  从而在随后的步骤中, 用  $m_{i-1}$  阶多项式来近似原方程  $d(\cdot)$ :

$$m_i = \begin{cases} 1 & i = 1, \\ 2^{i-1} + 1, & i = 2, 3, \dots \end{cases} \quad (12.26)$$

#### 建立克伦肖——柯蒂斯结点

构建复数集合  $g^i = \{\xi_1^i, \dots, \xi_{m_i}^i\} \subset [-1, 1]$ , 每个上角标为  $i$  的集合中都包含有  $m_i$  个克伦肖——柯蒂斯结点 (Clenshaw-Curtis nodes, 见第12.F.5节)。 $\xi_j^i, j = 1, 2, \dots, m_i$  的值是切比雪夫多项式(12.96), 取值范围在上下极值 (extrema)  $[-1, 1]$  之间, 满足

$$\xi_j^i = -\cos\left(\frac{j-1}{m_i-1}\right)\pi, \quad j = 1, \dots, m_i. \quad (12.27)$$

举例来说,

- 当  $i = 1$  时,  $m_1 = 1$ , 初始集合设为  $g^1 = \{\xi^1\} = \{0\}$ , 只有一个值。
- 当  $i = 2$  时,  $m_2 = 3$ , 3 阶切比雪夫多项式构成  $g^2 = \{-1, 0, 1\}$ 。
- 当  $i = 3$  时,  $m_3 = 5$ ,  $g^3 = \{-1, -\cos(\frac{\pi}{4}), 0, -\cos(\frac{3\pi}{4}), 1\}$ 。
- 当  $i = 4$  时,  $m_4 = 9$ ,  $g^4 = \{-1, -\cos\frac{\pi}{8}, -\cos\frac{\pi}{4}, -\sin\frac{\pi}{8}, 0, \sin\frac{\pi}{8}, -\cos\frac{3}{4}\pi, \cos\frac{\pi}{8}, 1\}$ ,

以此类推。可见随着  $i$  的逐渐增加,  $m_i$  指数增加, 集合  $g^{i-1}$  中的值都包含在了  $g^i$  中。(向下兼容性)。这个性质对 Smolyak 算法来说至关重要。

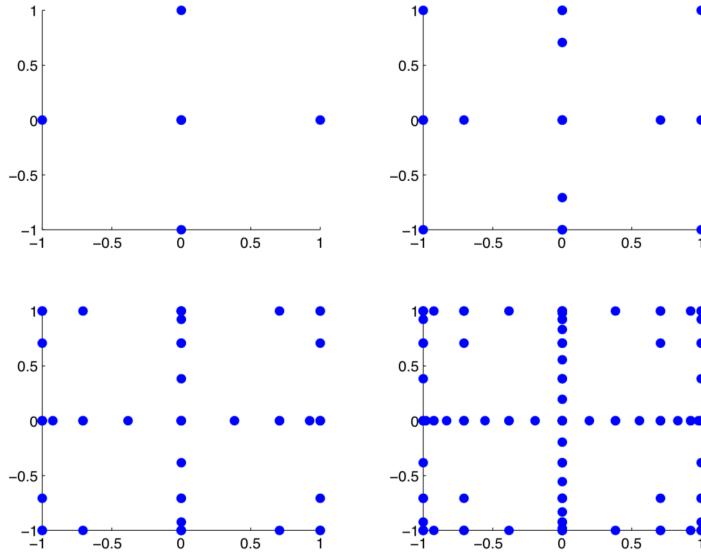
#### 构建稀疏网格

给定系统中状态变量的数量  $n$ , 对于任意  $q > n, q \in \mathcal{Z}$ , 可将稀疏网格  $G(q, n)$  定义为一个由一系列笛卡尔乘积 (cartesian product) 所组成的并集

$$G(q, n) = \bigcup_{q-n+1 \leq |\Upsilon| \leq q} (g^{i_1} \times g^{i_2} \times \dots \times g^{i_n}), \quad |\Upsilon| = \sum_{\ell=1}^n i_\ell.$$

以一个  $n = 2$  个连续状态变量的 DSGE 模型为例, 我们选取  $q = 3$ , 那么稀疏网格

$$\begin{aligned} G(3, 2) &= \bigcup_{2 \leq |\Upsilon| \leq 3} (g^{i_1} \times g^{i_2}) \\ &= (g^1 \times g^1) \cup (g^1 \times g^2) \cup (g^2 \times g^1) \\ &= \{(-1, 0), (0, 1), (0, 0), (0, -1), (1, 0)\}, \end{aligned} \quad (12.28)$$

图 12.1: 稀疏网格  $G(3, 2)$ ,  $G(4, 2)$ ,  $G(5, 2)$ ,  $G(6, 2)$ 

来源: Krueger and Kubler (2004) Figure 1。

将这些坐标绘制出来, 见图12.1左上角。

如果选  $q = 4$ , 那么

$$\begin{aligned} G(4, 2) &= \bigcup_{3 \leq |\Gamma| < 4} (g^{i_1} \times g^{i_2}) \\ &= (g^1 \times g^2) \cup (g^1 \times g^3) \cup (g^2 \times g^2) \cup (g^3 \times g^1) \\ &= \left\{ (-1, -1), (-1, 0), (-1, 1), \left(-\cos \frac{3}{4}\pi, 0\right), \left(-\cos \frac{1}{4}\pi, 0\right), \right. \\ &\quad (0, -1), \left(0, -\cos \frac{3}{4}\pi\right), \left(0, -\cos \frac{1}{4}\pi\right), (0, 0), (0, 1), (1, -1), (1, 0), (1, 1) \left. \right\}, \end{aligned}$$

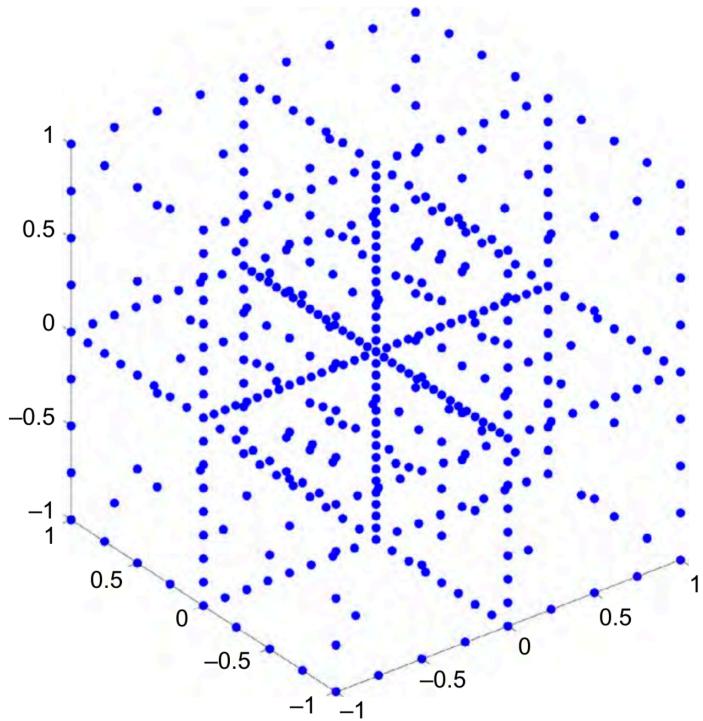
坐标对应图12.1右上角。以此类推, 图12.1的左下、右下分别绘出  $G(5, 2)$  和  $G(6, 2)$  的坐标。 $G(5, 3)$  的网格点, 见图12.2。

从图12.1和12.2不难看出稀疏网格的两个重要特征:

1. 网格点在全部值域空间中的分布是不均匀的, 越是靠近边缘的地区, 和越是靠近域中心的位置, 网格点越是密集。
2. 一个  $q = n + 2$  的稀疏网格中, 有网格点数  $2n(n - 1) + 4n + 1$  个。总结一般规律: 网格  $G(q, n)$  的势 (cardinality) 随着  $n^2$  而呈多项式级数型增长。若  $q = n + 3$ , 则势随  $n^3$  呈多项式技术型增长。事实上在给定  $n$  不变的情况下,  $q - n$  的值越大, 对应的网格点越多, 数值求解越耗费资源。

幸运的是, 对大多数 DSGE 系统而言,  $q = n + 2$  或者  $q = n + 3$  就足以达到较为理想的近似精度了。

在构建稀疏网格的过程中, 之所以选用切比雪夫结点构建网格点的集合, 是因为在克伦肖——柯蒂斯求积过程中, 切比雪夫点具有很理想的嵌套性特征 (nestedness), 该特征对于将

图 12.2: 稀疏网格  $G(5, 3)$ 

$\text{card}(G(q, n))$  控制在可操作范围内来说至关重要。例如，矩形网格的节点数是  $5^n$  个，总点数随着系统维度  $n$  的增加而指数上升；若  $n = 2$ ，这意味着需要在  $[-1, 1]^2$  的二维空间中绘制 25 个结点。而得益于切比雪夫结点的嵌套性，在部署 Smolyak 算法的过程中，我们只需要考虑全部 25 个结点中的 13 个就够了——网格变得“稀疏”了。随着  $n$  的变化，普通矩形网格点数和稀疏网格点数，如下表所示，相对于普通网格（如矩形网格）的指数增速，稀疏网格在控制网格点数量的增加（势）上有很大优势。

表 12.1: 网格结点的数量比较

$n$	$\text{card}(G)^*$	$5^n^{**}$
2	13	25
3	25	125
4	41	625
5	61	3125
12	313	244,140,625

\* 稀疏网格结点的势  $\text{card}(G(q, n))$ ,  $q = n + 2$ 。

\*\* 普通网格结点的势，等于  $5^n$ 。

### 构建张量积方程

我们将多元多项式的乘积简化表示为张量积 (tensor product) 的形式 (第12.3.2节), 记为  $p^{|\Upsilon|}(x|\theta)$

$$p^{|\Upsilon|}(x|\theta) \equiv \sum_{\ell_1=1}^{m_{i_1}} \dots \sum_{\ell_n=1}^{m_{i_n}} (\theta_{\ell_1 \dots \ell_n}) [\psi_{\ell_1}(x_1) \dots \psi_{\ell_n}(x_n)],$$

其中

- $|\Upsilon| \equiv \sum_{\ell=1}^n i_\ell$ ,
- $x \in [-1, 1], \quad x = \{x_1, \dots, x_n\}$ ,
- $\theta$  是含有全部系数的堆栈,

$$\theta = \{\theta_{\ell_1, \dots, \ell_n}\},$$

- 设基方程  $\psi_i(x_i)$  等于上一阶的切比雪夫多项式<sup>20</sup>,

$$\psi_i(x_i) = T_{i-1}(x_i), \quad i = 1, \dots, n$$

其中根据切比雪夫多项式的性质, 设初始点  $T_0(x_i) = 1$ 。

那么, 对于一个  $n = 2, q = 3$  的 DSGE 系统而言,  $|\Upsilon| = \{2, 3\}$ , 分两种情况。

$$p^{|\Upsilon|, i=1, 2} = \begin{cases} p^{[2]}(x|\theta) = p^{1,1}(x|\theta), \\ p^{[3]}(x|\theta) = p^{1,2}(x|\theta) + p^{2,1}(x|\theta), \end{cases} \quad (12.29)$$

其中

$$\begin{aligned} & \begin{cases} i = 1 \Rightarrow m_1 = 1, \\ i = 2 \Rightarrow m_2 = 3, \end{cases} \\ & p^{1,1}(x|\theta) = \sum_{\ell_1=1}^{m_1} \dots \sum_{\ell_n=1}^{m_1} \theta_{\ell_1 \ell_2} \psi_{\ell_1}(x_1) \psi_{\ell_2}(x_2) = \theta_{11}, \\ & p^{1,2}(x|\theta) = \sum_{\ell_1=1}^{m_1} \dots \sum_{\ell_n=1}^{m_2} \theta_{\ell_1 \ell_2} \psi_{\ell_1}(x_1) \psi_{\ell_2}(x_2) = \theta_{11} + \theta_{12} \psi_2(x_2) + \theta_{13} \psi_3(x_2), \\ & p^{2,1}(x|\theta) = \sum_{\ell_1=1}^{m_2} \sum_{\ell_n=1}^{m_1} \theta_{\ell_1 \ell_2} \psi_{\ell_1}(x_1) \psi_{\ell_2}(x_2) = \theta_{11} + \theta_{21} \psi_2(x_1) + \theta_{31} \psi_3(x_1) = \theta_{11} + \theta_{21} T_1(x_1) + \theta_{31} T_2(x_1). \end{aligned}$$

或者可以通过简化系数堆栈  $\theta$ , 来进一步简化上式。对于任意一个网格中的点  $k_1, \dots, k_n$ , 有

$$\theta = \theta_{\ell_1 \dots \ell_n} = \frac{2^n}{(k_1 - 1) \dots (k_n - 1)} \frac{1}{c_{\ell_1} \cdot c_{\ell_n}} \sum_{j_1=1}^{k_1} \dots \sum_{j_n=1}^{k_n} \frac{1}{c_{j_1} \cdot c_{j_n}} [\psi_{\ell_1}(\xi_1) \psi_{\ell_d}(\xi_n)] d(\xi_1, \dots, \xi_n), \quad (12.30)$$

其中

---

<sup>20</sup>这里选取切比雪夫多项式作为基方程, 是因为切比雪夫多项式已经成为 Smolyak 算法求解 DSGE 系统所普遍采用的准则了。但也有其他选取方式的有益尝试, 例如Nobile et al. (2008) 将有限元和 Smolyak 方法相结合, 把值域  $\Omega$  分为若干元, 在每个元中构建局部基方程再做加权求和。

- $c_j = \begin{cases} 2 & j = 1, \\ 1 & \forall j = 2, 3, \dots \end{cases}$

•  $\xi_k \in g^i$  表示克伦肖——柯蒂斯结点，如式(12.27)。

不难看出，这种张量积的近似测算方法，本质上来说就是在值域中，以克伦肖——柯蒂斯结点作插值。

### 构建插值方程

在稀疏网格  $G(q, n)$  中构建 Smolyak 插值方程，方程的值是张量  $p^{|\Upsilon|}(x|\theta)$  的加权求和

$$d(x|\theta, q, n) = \sum_{\max(n, q-n+1) \leq |\Upsilon| \leq q} (-1)^{q-|\Upsilon|} \binom{n-1}{q-|\Upsilon|} p^{|\Upsilon|}(x|\theta).$$

以上文的 DSGE 模型为例， $n = 2, q = 3$ ，对应的稀疏网格为  $G(3, 2)$  见式(12.28)，图12.1左上角

$$G(3, 2) = \{(-1, 0), (0, 1), (0, 0), (0, 1), (1, 0)\},$$

张量积  $p^{|\Upsilon|}(x|\theta)$  见式 (12.29)。

那么我们有近似方程

$$\begin{aligned} d(x|\theta, q, n) &= d(x|\theta, 3, 2) = \sum_{2 \leq |\Upsilon| \leq 3} (-1)^{3-|\Upsilon|} \binom{1}{3-|\Upsilon|} p^{|\Upsilon|}(x|\theta) \\ &= (-1) \binom{1}{1} p^{[2]}(x|\theta) + (-1)^0 \binom{1}{0} p^{[3]}(x|\theta) \\ &= p^{1,2}(x|\theta) + p^{2,1}(x|\theta) - p^{1,1}(x|\theta) \\ &= \theta_{11} + \theta_{21}T_1(x_1) + \theta_{31}T_2(x_1) + \theta_{12}T_1(x_2) + \theta_{13}T_2(x_2), \end{aligned}$$

上式中的系数  $\theta$  由(12.30)给出，具体来说

$$\begin{aligned} \theta_{21} &= \frac{1}{2} [d(1, 0) - d(-1, 0)], \\ \theta_{12} &= \frac{1}{2} [d(0, 1) - d(0, -1)], \\ \theta_{31} &= \frac{1}{4} [d(1, 0) + d(-1, 0)] - \frac{1}{2} d(0, 0), \\ \theta_{13} &= \frac{1}{4} [d(0, 1) + d(0, -1)] - \frac{1}{2} d(0, 0), \end{aligned}$$

只有常数项  $\theta_{11}$  的计算方式稍有些不同，这是为了确保 Smolyak 差值方程满足  $d(0, 0) = d(x|\theta, q, n)$ ：

$$\theta_{11} = \frac{1}{4} [d(0, 1) + d(0, -1) + d(1, 0) + d(-1, 0)].$$

利用这种思路，我们可以检验 Smolyak 算法的代码是否正确，具体说来就是看以下条件是否成立：在稀疏网格中的任意一点，插值近似方程的值是否等于原方程。以网格点  $(-1, 0)$  为例

$$\begin{aligned} d((-1, 0)|\theta, q, n) &= \theta_{11} + \theta_{21}T_1(-1) + \theta_{31}T_2(-1) + \theta_{12}T_1(0) + \theta_{13}T_2(0) \\ &= \theta_{11} - \theta_{21} + \theta_{31} - \theta_{13} \\ &= \frac{1}{4} [d(0, 1) + d(0, -1) + d(1, 0) + d(-1, 0)] - \frac{1}{2} [d(1, 0) - d(-1, 0)] \\ &\quad + \frac{1}{4} [d(1, 0) + d(-1, 0)] - \frac{1}{2} d(0, 0) - \frac{1}{4} [d(0, 1) + d(0, -1)] + \frac{1}{2} d(0, 0) \\ &= d(-1, 0) \end{aligned}$$

通过上述分析，我们可以看到 Smolyak 近似方程的两个重要性质

1.  $\theta$  系数的数量恰好等于稀疏网格的势

$$n = 2, q = 3 \Rightarrow \text{card}(\mathbf{G}(3, 2)) = 5 \Rightarrow \theta = \{\theta_{11}, \theta_{21}, \theta_{31}, \theta_{12}, \theta_{13}\}.$$

2. 对原未知方程的 Smolyak 近似方程  $d(x|\theta, q, n)$  表现为一个多项式方程的形式，有一组小于等于  $q - n$  阶的单项式构成。

### 求解多项式的系数

上一步骤中我们以  $(-1, 0)$  为例，求解  $\mathbf{G}(3, 2)$  中的近似方程  $d((-1, 0)|\theta, 3, 2)$ 。同理，对于所有网格点  $x_i \in \mathbf{G}(q, n)$ ，都有且仅有一条近似方程  $d(x_i|\theta, q, n)$  与之相对应，例如

$$\begin{aligned} \mathbf{G}(3, 2) &= \{(-1, 0), (0, 1), (0, 0), (0, 1), (1, 0)\} \\ d((-1, 0)|\theta, 3, 2) &= \theta_{11} - \theta_{21} + \theta_{31} - \theta_{13}, \\ d((0, 1)|\theta, 3, 2) &= \dots \\ &\vdots \end{aligned}$$

这一组近似方程组合起来构成一个标准的非线性方程系统，定义为算子  $\mathcal{H}(\cdot)$ ，满足

$$\mathcal{H}(d(x_i|\theta, q, n)) = 0, \quad \forall x_i \in \mathbf{G}(q, n), \forall i \in n, \quad (12.31)$$

其中向量  $\theta = \{\theta_{11}, \theta_{12}, \theta_{13}, \theta_{21}, \theta_{31}\}$  是代求解系数，可以通过一系列标准的非线性系统近似方法来求解，例如 Krueger and Kubler (2004); Malin et al. (2011) 建议用时间迭代法 (time iteration method) 进行求解，<sup>21</sup>。

### 12.7.2 稀疏网格的一些扩展

Brumm and Scheidegger (2017) 对时间迭代做出扩展：

1. 将时间迭代嵌入到适应性稀疏网格 (adaptive sparse grid) 中去。

适应性稀疏网格在局部中自动精炼，从而可以描述一系列更复杂的情况，如陡峭梯度 (steep gradients)，不可求导 (non-differentiabilities) 等情况。

2. 采用混合型并行计算策略 (fully hybrid parallel computing)，利用大规模并行计算 (massively parallel processing) 的最新进展，提高数值求解的计算效率。

除了 Krueger and Kubler (2004); Malin et al. (2011) 所建议的 Smolyak 算法之外，在应用稀疏网格求解高维度、非线性 DSGE 方程系统方面，还有一系列研究做了有益的尝试。其中值得注意的是 Judd et al. (2014) 所做的重要改进。

1. 使用不连续集 (disjoint-set generoator) 替代  $\mathbf{g}^i$ ，
2. 使用拉格朗日插值法，
3. 构建各向异性网格 (anisotropic grid)，根据各向异性特征，针对不同的状态变量  $x_i$ ，可以有不同数量的网格点、基方程与之相对应。

这点之所以显得非常重要，是因为 DSGE 研究中常常会遇到这种情况：沿着某些变量的维度探讨经济个体的行为决策时，会比其他维度的分析更难。

<sup>21</sup>他们建议初始猜测值取为模型本身的一阶扰动项。不过对于 Smolyak 算法而言，初始值的选取并不是至关重要的。

记里有各向异  
的介绍，敲到  
盘里之后，在  
处做一个 ref。

4. Judd et al. (2014) 用可自由求导的不动点迭代方法 (derivative-free fixed point iteration method), 替代Krueger and Kubler (2004); Malin et al. (2011) 时间迭代法, 来做  $\mathcal{H}(\cdot) = 0$  的近似线性求解。



# 附录

## 12.A 超几何方程

**Definition 12.1** (超几何序列). 一组序列  $\{c_n\}$ , 如果  $\frac{c_{n+1}}{c_n}$  是一个关于  $n$  的有理方程, 则我们称  $\{c_n\}$  是一个超几何序列 (*Hypergeometric Series*)。

以阶乘形式表示为

$$\frac{c_{n+1}}{c_n} = \frac{(n+a_1)(n+a_2)\dots(n+a_p)}{(n+b_1)(n+b_2)\dots(n+b_q)} \frac{z}{(n+1)}, \quad n=0,1,2\dots, \quad (12.32)$$

如果  $z=1$ ,  $\{c_n\}$  又称首一超几何序列 (monic hypergeometric series); 如果  $z\neq 1$ , 则称非首一超几何序列。分母中的  $n+1$  项可以是来自阶乘本身, 也可以不是, 对于后者的情况, 我们也可以在分子中加入一个额外的  $(n+a_i)$  项予以调节, 即选取一个  $i \in 0, 1 \dots n+1$  使得  $a_i = 1$ 。

对(12.32)做迭代我们有

$$c_n = \frac{(a_1)_n (a_2)_n \dots (a_p)_n}{(b_1)_n (b_2)_n \dots (b_q)_n} \frac{z^n}{n!} c_0, \quad (12.33)$$

其中  $n!$  表示  $n$  的阶乘。 $(a_1)_n$  表示伯赫哈默尔符号 (Pochhammer symbol) 的定义如下

$$(a_1)_n := \begin{cases} 1 & \text{如果 } n=0, \\ a_1(a_1+1)(a_1+2)\dots(a_1+n-1) & \text{如果 } n=1,2,3\dots. \end{cases}$$

对超几何序列  $\{c_n\}$  的求和为

$$\sum_{n=0}^{\infty} c_n = c_0 \sum_{n=0}^{\infty} \frac{(a_1)_n (a_2)_n \dots (a_p)_n}{(b_1)_n (b_2)_n \dots (b_q)_n} \frac{z^n}{n!},$$

这引出超几何方程的定义。

**Definition 12.2** (超几何方程). 超几何方程 (*hypergeometric function*)  ${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; x)$  定义为超几何序列  $\{c_n\}$  的和:

$${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; x) = \sum_{k=0}^n c_n = \sum_{n=0}^{\infty} \frac{(a_1)_n (a_2)_n \dots (a_p)_n}{(b_1)_n (b_2)_n \dots (b_q)_n} \frac{z^n}{n!} \quad (12.34)$$

显然, 上式成立要求分母的系数  $\neq 0$ 。对于分子的系数, 如果某一个  $i$  次系数  $a_i = -N$ , 其中  $N$  是个非负整数, 那么超几何方程  ${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; x)$  是一个关于  $z$  的多项式, 详见定义12.3; 反之, 则超几何序列的收敛速度  $\rho$  满足

$$\rho = \begin{cases} \infty & \text{如果 } p < q + 1, \\ 1 & \text{如果 } p = q + 1, \\ 0 & \text{如果 } p > q + 1, \end{cases}$$

这是由超几何序列的性质所决定的：由(12.32)可得

$$\lim_{n \rightarrow \infty} \left| \frac{c_{n+1}}{c_n} \right| = \begin{cases} \infty & \text{如果 } p < q + 1, \\ |z| & \text{如果 } p = q + 1, \\ \infty & \text{如果 } p > q + 1, \end{cases}$$

我们最为关注  $p = q + 1$  的情况：随着  $|z|$  的取值不同，超几何函数  ${}_qF_q(a_1, \dots, a_{q+1}; b_1, \dots, b_q; z)$  呈现出不同的收敛特点：

$${}_qF_q(a_1, \dots, a_{q+1}; b_1, \dots, b_q; z) \begin{cases} \text{收敛,} & \text{如果 } |z| < 1, \\ \text{收敛,} & \text{如果 } |z| = 1, \operatorname{Re} \left( \sum^q b_i - \sum^{q+1} a_j \right) > 0, \\ \text{有条件收敛,} & \text{如果 } |z| = 1, z \neq -1, -1 < \operatorname{Re} \left( \sum^q b_i - \sum^{q+1} a_j \right) \leq 0, \\ \text{不收敛,} & \text{如果 } \operatorname{Re} \left( \sum^q b_i - \sum^{q+1} a_j \right) \leq -1. \end{cases}$$

**Definition 12.3** (广义超几何方程). 广义超几何方程 (*generalized hypergeometric function*) 是超几何方程的一种特殊形式，定义式为

$${}_2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!}, \quad (12.35)$$

其中如果  $a = -N$ ,  $N = 0, 1, 2 \dots$ , 那么我们有

$$(a)_n = (-N)_n = (-N)(-N+1)(-N+2)\dots(-N+n-1) = 0, n = N+1, N+2, N+3 \dots$$

那么(12.35)变为

$${}_2F_1(-N, b; c; z) = \sum_{n=0}^N \frac{(-N)_n (b)_n}{(b)_n} \frac{z^n}{n!}, \quad N = 0, 1, 2 \dots \quad (12.36)$$

如果  $a_n$  不是个常数，那么超几何序列想要获得收敛特性，需要满足如下情况，第一种情况是  $|z| < 1$ ，第二种情况是  $|z| = 1$  以及  $\operatorname{Re}(c - a - b) > 0$ 。

许多函数都可以表示为超几何方程的形式。如

$$\begin{aligned} \ln(1+x) &= \sum_{n=0}^{\infty} \frac{(-1)^n}{n+1} x^{n+1} = \sum_{n=0}^{\infty} \frac{(1)_n (1)_n}{(2)_n} \frac{(-1)^n x^{n+1}}{n!} = \sum_{n=0}^{\infty} \frac{n! n!}{(n+1)!} \frac{(-1)^n x^{n+1}}{n!} = x {}_2F_1\left(\frac{1}{2}, \frac{1}{2}; -x\right), \\ \arctan x &= \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} x^{2n+1} = \sum_{n=0}^{\infty} \frac{\left(\frac{1}{2}\right)_n (1)_n}{\left(\frac{3}{2}\right)_n} \frac{(-1)^n x^{2n+1}}{n!} = x {}_2F_1\left(\frac{1}{2}, \frac{1}{2}; -x^2\right), \end{aligned}$$

最后一个等式是由于  $\frac{\left(\frac{1}{2}\right)_n}{\left(\frac{3}{2}\right)_n} = \frac{\frac{1}{2} \frac{3}{2} \dots \frac{2n-1}{2}}{\frac{3}{2} \frac{5}{2} \dots \frac{2n+1}{2}} = \frac{\frac{1}{2}}{\frac{2n+1}{2}} = \frac{1}{2n+1}$

**Theorem 12.1** (二项式定理). 作为超几何方程的特殊形式之一，可以写成如下形式，称为二项式定理 (*binomial theorem*)

$${}_2F_1\left(\begin{matrix} a & b \\ b & \end{matrix}; z\right) = {}_1F_0\left(\begin{matrix} a \\ \underline{b} \end{matrix}; z\right) = \sum_{n=0}^{\infty} \frac{(a)_n}{n!} z^n = \sum_{n=0}^{\infty} \binom{-a}{n} \frac{(a)_n}{n!} (-z)^n = (1-z)^{-a}, \quad |z| < 1. \quad (12.37)$$

以及还有一类特殊的超几何方程

$${}_0F_0\left(\begin{matrix} \underline{\phantom{a}} \\ \underline{\phantom{a}} \end{matrix}; z\right) = \sum_{t=0}^{\infty} \frac{z^n}{n!} = e^z, \quad z \in \mathbb{C}. \quad (12.38)$$

以及当系数趋近于无穷时：

$$\lim_{b \rightarrow \infty} {}_2F_1\left(\begin{matrix} a & b \\ c & \end{matrix}; \frac{z}{b}\right) = \lim_{b \rightarrow \infty} \sum_{n=0}^{\infty} \frac{(a)_n z^n}{(c)_n n!} \frac{(b)_n}{b^n} = \sum_{n=0}^{\infty} \frac{(a)_n}{(c)_n} \frac{z^n}{n!} = {}_1F_1\left(\begin{matrix} a \\ c \end{matrix}; z\right),$$

$$\lim_{c \rightarrow \infty} {}_2F_1\left(\begin{matrix} a & b \\ c & \end{matrix}; cz\right) = \lim_{c \rightarrow \infty} \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{n!} z^n \frac{c^n}{(c)_n} = \sum_{n=0}^{\infty} (a)_n (b)_n \frac{z^n}{n!} = {}_2F_0\left(\begin{matrix} a & b \\ - & \end{matrix}; z\right).$$

**Theorem 12.2** (超几何方程与欧拉积分表达式的转换). 如果  $\operatorname{Re}(c) > \operatorname{Re}(b) > 0$ , 那么对于沿着从 1 到  $\infty$  实数轴展开的所有  $z \in \mathbb{C}$ , 满足  $\arg t = \arg(1-t) = 0$  并且  $(1-zt)^{-a}$  有主值 (principal value) 的条件<sup>22</sup>, 我们可以将超几何方程转换为欧拉积分表达式 (Euler's integral representation)

$${}_2F_1\left(\begin{matrix} a & b \\ c & \end{matrix}; z\right) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1} (1-t)^{c-b-1} (1-zt)^{-a} dt. \quad (12.39)$$

证明. 假设  $|z| < 1$ , 那么由二项式定理(12.37)可得

$$(1-zt)^{-a} = \sum_{n=0}^{\infty} \frac{(a)_n}{n!} z^n t^n,$$

这表明

$$\int_0^t t^{b-1} (1-t)^{c-b-1} (1-zt)^{-a} dt = \underbrace{\sum_{n=0}^{\infty} \frac{(a)_n}{n!} z^n t^n}_{\text{RHS 甲}} \underbrace{\int_0^t t^{b-1} (1-t)^{c-b-1} dt}_{\text{RHS 乙}},$$

RHS 分为两部分。乙是一个贝塔积分 (Beta integral), 满足条件<sup>23</sup>

$$\int_0^t t^{b-1} (1-t)^{c-b-1} dt = B(n+b, c-b) = \frac{\Gamma(n+b)\Gamma(c-b)}{\Gamma(n+c)}.$$

此外 Gamm 方程可以转换为 Pochhammer 符号的表达形式：

$$\frac{\Gamma(n+b)}{\Gamma(b)} = b(b+1)(b+2) \dots (b+n-1) = (b)_n,$$

那么

$$\frac{\Gamma(n+b)\Gamma(c-b)}{\Gamma(n+c)} = \Gamma(c-b) \frac{\Gamma(b)}{\Gamma(c)} \frac{(b)_n}{(c)_n},$$

作为 RHS 乙代回原式有

$$\frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^t t^{b-1} (1-t)^{c-b-1} (1-zt)^{-a} dt = \sum_0^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!} = {}_2F_1\left(\begin{matrix} a & b \\ c & \end{matrix}; z\right).$$

□

**Theorem 12.3** (欧拉积分表达式与高斯求和公式的转换). 对于  $\operatorname{Re}(c-a-b) > 0$ , 我们有欧拉积分表达式与高斯求和公式 (Gauss summation formula)的转换

$${}_2F_1\left(\begin{matrix} a & b \\ c & \end{matrix}; 1\right) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}. \quad (12.40)$$

<sup>22</sup>膏按：反正这一串前提限定条件我没看懂... 我的理解是，经济学研究中，能够熟悉下面的转换公式就可以了——这真是一本不负责任的“讲义”啊...

<sup>23</sup> 其中  $\Gamma(\cdot)$  是 Gamma 方程, 是阶乘函数在实数与复数上的扩展, 定义为

$$\Gamma(a) = \begin{cases} (n-1)! & n \text{ 是正整数}, \\ \int_0^{\infty} t^{a-1} \exp(-t) dt & z \in \mathbb{C}, \operatorname{Re}(c) > 0. \end{cases}$$

证明. 假设  $\lim_{z \rightarrow 1}$ , 通过(12.39)我们有

$$\begin{aligned} \lim_{z \rightarrow 1} {}_2F_1\left(\begin{matrix} a & b \\ c & \end{matrix}; z\right) &\approx {}_2F_1\left(\begin{matrix} a & b \\ c & \end{matrix}; 1\right) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1}(1-t)^{c-a-b-1} dt \\ &= \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} B(b, c-a-b) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \frac{\Gamma(b)\Gamma(c-a-b)}{\Gamma(c-a)} = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}. \end{aligned}$$

□

**Theorem 12.4** (高斯求和公式与朱世杰——范德蒙德求和公式的转换). 当  $a = -n, n \in \{0, 1, 2, \dots\}$  (并且  $\lim_{z \rightarrow 1}$ ) 时, 高斯求和公式(12.40)缩减成为朱世杰——范德蒙德求和公式 (*Chu-Vandermonde summation*)

$${}_2F_1\left(\begin{matrix} -n & b \\ c & \end{matrix}; 1\right) = \frac{(c-b)_n}{(c)_n}, n = 0, 1, 2, \dots \quad (12.41)$$

证明.

$$\frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} = \frac{\Gamma(c)\Gamma(c-b+n)}{\Gamma(c+n)\Gamma(c-b)} = \frac{(c-b)_n}{c(n)}.$$

□

**Theorem 12.5** (超几何方程与法夫转换公式). 我们可以用欧拉积分表达式证明, 超几何方程可以写成法夫转换公式 (*Pfaff's transformation formula*)的形式

$${}_2F_1\left(\begin{matrix} a & b \\ c & \end{matrix}; z\right) = (1-z)^{-a} {}_2F_1\left(\begin{matrix} a & c-b \\ c & \end{matrix}; \frac{z}{z-1}\right). \quad (12.42)$$

证明. 定义  $t = 1-s$ , 代入欧拉积分表达式(12.39)我们有

$$\begin{aligned} {}_2F_1\left(\begin{matrix} a & b \\ c & \end{matrix}; z\right) &= \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-zt)^{-a} dt \\ &= - \int_0^1 (1-s)^{b-1}s^{c-b-1}(1-z+zs)^{-a} ds \\ &= (1-z)^{-a} \int_0^1 s^{c-b-1}(1-s)^{b-1} \left(1 - \frac{zs}{z-1}\right)^{-a} ds \\ &= (1-z)^{-a} {}_2F_1\left(\begin{matrix} a & c-b \\ c & \end{matrix}; \frac{z}{z-1}\right). \end{aligned}$$

□

**Theorem 12.6** (超几何方程与欧拉转换公式). 我们可以用法夫转换公式证明, 超几何方程可以写成欧拉转换公式 (*Euler's transformation formula*)的形式

$${}_2F_1\left(\begin{matrix} a & b \\ c & \end{matrix}; z\right) = (1-z)^{c-a-b} {}_2F_1\left(\begin{matrix} c-a & c-b \\ c & \end{matrix}; z\right). \quad (12.43)$$

证明. 两次使用法夫转换公式

$$\begin{aligned} {}_2F_1\left(\begin{matrix} a & b \\ c & \end{matrix}; z\right) &= (1-z)^a {}_2F_1\left(\begin{matrix} a & c-b \\ c & \end{matrix}; \frac{z}{z-1}\right) \\ &= (1-z)^a (1-z)^{b-c} {}_2F_1\left(\begin{matrix} c-a & c-b \\ c & \end{matrix}; \frac{\frac{z}{z-1}}{\frac{z}{z-1}-1}\right) \\ &= (1-z)^{c-a-b} {}_2F_1\left(\begin{matrix} c-a & c-b \\ c & \end{matrix}; z\right) \end{aligned}$$

□

**Theorem 12.7** (法夫——萨尔舒茨求和公式). 一个  ${}_3F_2$  超几何方程又可以写成法夫——萨尔舒茨求和公式 (*Pfaff-Saalschütz summation formula*) 的形式

$${}_3F_2\left(\begin{matrix} -n & a & b \\ c & 1+a+b-c-n \end{matrix}; 1\right) = \frac{(c-a)_n(c-b)_n}{(c)_n(c-a-b)_n}, \quad n=0,1,2,\dots \quad (12.44)$$

证明. 欧拉转换公式(12.43)又可以写成

$$(1-z)^{a+b-c} {}_2F_1\left(\begin{matrix} a & b \\ c \end{matrix}; z\right) = {}_2F_1\left(\begin{matrix} c-a & c-b \\ c \end{matrix}; z\right),$$

其中 LHS →

$$\begin{aligned} (1-z)^{a+b-c} {}_2F_1\left(\begin{matrix} a & b \\ c \end{matrix}; z\right) &= \sum_{j=0}^{\infty} \frac{(c-a-b)_j}{j!} z^j \sum_{k=0}^{\infty} \frac{(a)_k(b)_k}{(c)_k} \frac{z^k}{k!} \\ &= \sum_{n=0}^{\infty} \sum_{k=0}^n \frac{(a)_k(b)_k(c-a-b)_{n-k}}{(c)_k k!(n-k)!} z^n. \end{aligned}$$

RHS →

$${}_2F_1\left(\begin{matrix} c-a & c-b \\ c \end{matrix}; z\right) = \sum_{n=0}^{\infty} \frac{(c-a)_n(c-b)_n}{(c)_n} \frac{z^n}{n!}.$$

比较两个式子中  $z^n$  对应的系数, 我们有

$$\sum_{k=0}^n \frac{(a)_k(b)_k(c-a-b)_{n-k}}{(c)_k k!(n-k)!} z^n = \frac{(c-a)_n(c-b)_n}{(c)_n n!}, \quad n=0,1,2,\dots, \quad k \in \{0,1,2,\dots,n\}.$$

□

其中由于

$$\frac{n!}{(n-k)!} = n(n-1)\dots(n-k+1) = (-1)^k(-n)(-n+1)\dots(-n+k-1) = (-1)^k(-n)^k,$$

以及

$$\begin{aligned} (c-a-b)_{n-k} &= \frac{(c-a-b)_n}{(c-a-b+n-k)(c-a-b+n-k-1)\dots(c-a-b+n-1)} \\ &= \frac{(c-a-b)_n}{(-1)^k(1+a+b-c-n)_k}. \end{aligned}$$

因此

$$\sum_{k=0}^n \frac{(a)_k(b)_k(-n)_k}{(c)_k k!(1+a+b-c-n)_k} = \frac{(c-a)_n(c-b)_n}{(c)_n(c-a-b)_n}, \quad n=0,1,2,\dots, \quad k \in \{0,1,2,\dots,n\},$$

进而(12.44)成立。

**Theorem 12.8** (法夫——萨尔舒茨求和公式与高斯求和公式的转换). 随着  $n \rightarrow \infty$ ,  ${}_3F_2$  的法夫——萨尔舒茨求和公式(12.44)逐渐缩减到  ${}_2F_1$  的高斯求和公式(12.40)的形式

$$\lim_{n \rightarrow \infty} {}_3F_2\left(\begin{matrix} -n & a & b \\ c & 1+a+b-c-n \end{matrix}; 1\right) = {}_2F_1\left(\begin{matrix} a & b \\ c \end{matrix}; 1\right). \quad (12.45)$$

证明. 根据

$$(a)_n = a(a+1)\dots(a+n-1) = \frac{\Gamma(a+n)}{\Gamma(a)},$$

我们有

$$\begin{aligned}\frac{(c-a)_n(c-b)_n}{(c)_n(c-a-b)_n} &= \frac{\Gamma(c-a+n)\Gamma(c-b+n)\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)\Gamma(c+n)\Gamma(c-a-b+n)} \\ &= \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} \underbrace{\frac{\Gamma(c-a+n)\Gamma(c-b+n)}{\Gamma(c+n)\Gamma(c-a-b+n)}}\end{aligned}$$

上式中下括号部分可由斯特灵公式 (Stirling formula) 近似：

$$\lim_{n \rightarrow 0} \frac{\Gamma(c-a+n)\Gamma(c-b+n)}{\Gamma(c+n)\Gamma(c-a-b+n)} \sim n^{c-a+c-b-c-c+a+b} = n^0 = 1,$$

即

$$\lim_{n \rightarrow \infty} \frac{(c-a)_n(c-b)_n}{(c)_n(c-a-b)_n} = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)},$$

由此得到

$$\begin{aligned}_2F_1\left(\begin{matrix} a & b \\ c & \end{matrix}; 1\right) &= \lim_{n \rightarrow \infty} {}_3F_2\left(\begin{matrix} -n & a & b \\ c & 1+a+b-c-n & \end{matrix}; 1\right) \\ &= \lim_{n \rightarrow \infty} \frac{(c-a)_n(c-b)_n}{(c)_n(c-a-b)_n} \\ &= \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}.\end{aligned}$$

□

## 12.B 多项式

### 12.B.1 正交集

数学意义上来说，正交 (orthogonal) 意味着垂直 (perpendicular)。如，一个三维向量空间中的向量集  $\{x, y, z\}$  是正交的，意思就是说任意两个正交向量的点乘 (dot product) 都为零， $x.y = 0, x.z = 0, y.z = 0$ 。因此我们又将  $\{x, y, z\}$  称为正交集合。在  $\{x, y, z\}$  构成的三维空间中，任何一点都可以表示为  $x, y$  或  $z$  向量中的一项——换句话说， $\{x, y, z\}$  共同构成了一个三维空间的基 (basis)。

如果  $\{x, y, z\}$  都是单位向量，则进一步称之为标准正交向量集。

**Definition 12.4** (黎曼——斯蒂尔杰斯积分). 在区间  $[a, b]$  内，假定对于一个实变量  $\lambda$  的实函数  $\alpha(\lambda)$ ，存在一个对应的实函数  $f(\lambda)$ ，则黎曼——斯蒂尔杰斯积分 (*Riemann-Stieltjes integral*)  $\int_a^b f(\lambda) d\alpha(\lambda)$  描述了当将区间  $[a, b]$  分割为无限多个块 (数量用  $\pi$  表示) 时，下述和的极限值

$$\int_a^b f(\lambda) d\alpha(\lambda) = \sum_{\lambda_i \in \pi} f(c_i) (\alpha(\lambda_{i+1}) - \alpha(\lambda_i)), \quad c_i \in [\lambda_{i+1}, \lambda_i]. \quad (12.46)$$

黎曼——斯蒂尔杰斯积分值若要存在，需要  $f(\cdot)$  是连续方程，以及非递减方程  $\alpha(\cdot)$  在  $[a, b]$  中有界。

黎曼——斯蒂尔杰斯积分常常直接写作简化形式

$$\int_a^b f(\lambda) w(\lambda) d\lambda, \quad (12.47)$$

其中  $w(\cdot)$  称权重方程 (weight function)。

## 12.B.2 正交多项式

任意一个无限多项式序列  $\{p_n(\lambda)\}_{n=0}^{\infty}$  向量 (其中  $p_n(\lambda)$  表示其中的第  $n$  次多项式), 都构成无限维度向量空间中的一个基。

**Definition 12.5** (正交多项式). 如果一个多项式序列  $\{p_n(\lambda)\}_{n=0}^{\infty}$  在区间  $(a, b)$  中, 关于权重方程  $w(\lambda)$ , 满足如下关系, 那么我们说它是正交多项式 (*orthogonal polynomial*):

$$\int_a^b w(\lambda) p_m(\lambda) p_n(\lambda) d\lambda = h_n \delta_{mn} \quad (12.48)$$

- $h_n$  称范数 (norm), 正交多项式  $p(\lambda)$  的范数也表示为  $|p|$ , 定义为 (具体细节见第14.2.2。)

$$h_n = |p_n(\lambda)| = \int_a^b w(\lambda) p_n(\lambda)^2 dx > 0, n = 0, 1, 2 \dots \quad (12.49)$$

- 权重方程  $w(\lambda)$  应当在  $(a, b)$  区间内连续且为正, 以便矩 (moment)  $\mu_n$  存在。

- 矩 (moment)  $\mu_n$  定义为

$$\mu_n = \int_a^b w(x) \lambda^n dx, \quad i = 0, 1, 2 \dots \quad (12.50)$$

- $\delta_{mn}$  是克罗内克 (Kronecker product), 表达如下关系

$$\delta_{m,n} = \begin{cases} 0 & m \neq n, \\ 1 & m = n. \end{cases} \quad (12.51)$$

- 积分  $\langle f, g \rangle$  表示多项式  $f$  和  $g$  的内积 (inner product)

$$\langle p, q \rangle = \int_a^b w(\lambda) f(\lambda) g(\lambda) d\lambda. \quad (12.52)$$

- 区间  $(a, b)$  称正交区间 (可以是一个无限区间, 即  $a \rightarrow -\infty, b \rightarrow +\infty$ )。

如果正交多项式  $\{p_n(\lambda)\}_{n=0}^{\infty}$  对于  $k = 0, 1, 2 \dots n$  都有  $h_k = 1$ , 则称标准正交多项式 (orthonormal polynomial)。

如果第  $n$  次多项式  $p_n(\lambda) = k_n \lambda^n + \chi$ , 其中  $\chi$  表示低于  $n$  次的其他项, 并且对于每一个  $n = 0, 1 \dots$  都有  $k_n = 1$ , 则称首一正交多项式 (monic orthogonal polynomial)。

标准正交多项式  $p_n(\lambda)$  和首一正交多项式  $\hat{p}_n(\lambda)$  之间可以相互转换 (如见定理12.12)。

$$\hat{p}_n(\lambda) = \frac{p_n(\lambda)}{\sqrt{\langle p_n(\lambda), p_n(\lambda) \rangle}}. \quad (12.53)$$

我们可以根据格拉姆-施密特正交化过程 (Gram-Schmidt orthogonality process), 把正交多项式序列  $\{p_n(\lambda)\}_{n=0}^{\infty}$  转化为正交集合。举例说明, 取  $w(\lambda) = 1, (a, b) = (0, 1), p_0(\lambda) = 1$ , 求首一多项式序列  $\{p_n(\lambda)\}$ 。

从序列  $\{1, \lambda, \lambda^2 \dots\}$  开始。第一步求  $p_1(\lambda)$

$$\begin{aligned} p_1(\lambda) &= \lambda - \frac{\langle \lambda, p_0(\lambda) \rangle}{\langle p_0(\lambda), p_0(\lambda) \rangle} p_0(\lambda) \\ &= \lambda - \frac{\langle \lambda, 1 \rangle}{\langle 1, 1 \rangle} \\ &= \lambda - \frac{\int_0^1 w(\lambda) (\lambda \cdot 1) d\lambda}{\int_0^1 w(\lambda) (1 \cdot 1) d\lambda} \\ &= \lambda - \frac{1}{2}. \end{aligned}$$

第一步求  $p_2(\lambda)$

$$\begin{aligned} p_2(\lambda) &= \lambda^2 - \frac{\langle \lambda^2, p_0(\lambda) \rangle}{\langle p_0(\lambda), p_0(\lambda) \rangle} p_0(\lambda) - \frac{\langle \lambda^2, p_1(\lambda) \rangle}{\langle p_1(\lambda), p_1(\lambda) \rangle} p_1(\lambda) \\ &= \lambda^2 - \frac{\langle \lambda^2, 1 \rangle}{\langle 1, 1 \rangle} - \frac{\langle \lambda^2, \lambda - \frac{1}{2} \rangle}{\langle \lambda - \frac{1}{2}, \lambda - \frac{1}{2} \rangle} \left( \lambda - \frac{1}{2} \right) \\ &= \lambda^2 - \frac{\frac{1}{3}}{1} - \frac{\frac{1}{12}}{\frac{1}{12}} \left( \lambda - \frac{1}{2} \right) \\ &= \lambda^2 - \lambda + \frac{1}{6}. \end{aligned}$$

则  $p_0(\lambda) = 1$ ,  $p_1(\lambda) = \lambda - \frac{1}{2}$ ,  $p_2(\lambda) = \lambda^2 - \lambda + \frac{1}{6}$  就是首一正交多项式在  $(0,1)$  区间内, 对应权重方程  $w(\lambda) = 1$  的前三项。以此类推, 可以继续求得  $p_3(\lambda), p_4(\lambda) \dots$

$$\begin{aligned} p_3(\lambda) &= \lambda^3 - \frac{3}{2}\lambda^2 + \frac{3}{5}\lambda - \frac{1}{20}, \\ p_4(\lambda) &= \lambda^4 - 2\lambda^3 + \frac{9}{7}\lambda^2 - \frac{2}{7}\lambda + \frac{1}{70}, \\ p_5(\lambda) &= \lambda^5 - \frac{5}{2}\lambda^4 + \frac{20}{9}\lambda^3 - \frac{5}{6}\lambda^2 + \frac{5}{42}\lambda - \frac{1}{252} \dots \end{aligned}$$

根据(12.49)计算  $h_n$ ; 与首一正交多项式  $p_n(\lambda)$  对应的正交多项式  $q_n(\lambda)$  由(12.53)给出

$$\begin{aligned} q_1(\lambda) &= \frac{p_1(\lambda)}{\sqrt{h_1}} = 2\sqrt{3} \left( x - \frac{1}{2} \right), \\ q_2(\lambda) &= \frac{p_2(\lambda)}{\sqrt{h_2}} = 6\sqrt{5} \left( x^2 - x + \frac{1}{6} \right), \\ q_3(\lambda) &= \frac{p_3(\lambda)}{\sqrt{h_3}} = 20\sqrt{7} \left( \lambda^3 - \frac{3}{2}\lambda^2 + \frac{3}{5}\lambda - \frac{1}{20} \right) \dots \end{aligned}$$

### 正定

如果  $|p| > 0$ ,  $|q| > 0$ ,  $\forall p, q \in \mathbb{P}$ , 则  $\langle \cdot, \cdot \rangle$  是正定的。换句话说, 内积是正定的充分必要条件为, 它的汉克尔矩阵 (Hankel moment matrix) 的行列式为正

$$\det \begin{pmatrix} \mu_0 & \mu_1 & \mu_2 & \dots & \mu_{k-1} \\ \mu_1 & \mu_2 & \mu_3 & \dots & \mu_k \\ \vdots & \ddots & & & \vdots \\ \mu_{k-1} & \mu_k & \mu_{k+1} & \dots & \mu_{2k-2} \end{pmatrix} > 0, k = 1, 2, \dots \quad (12.54)$$

其中  $\mu_i$  的定义如(12.50)所示。

**Theorem 12.9** (正交多项式的存在性). 如果某内积  $\langle \cdot, \cdot \rangle$  在  $\mathbb{P}$  中正定, 那么有且只有一个首一正交的无限多项式与每一个  $\alpha(\lambda)$  相对应。

证明. 略。 □

**Theorem 12.10** (正交多项式的最小化特性). 如果  $q_k(\lambda)$  是一个  $k$  次的首一多项式, 则只有当  $q_k$  等于一个常数与另一个同样关于  $\alpha(\lambda)$  的多项式  $p_k$  的乘积时, 以下最小值才存在

$$\min_{q_k} \int_a^b q_k(\lambda)^2 d\alpha(\lambda).$$

证明. 略。 □

### 12.B.3 递推关系

**Theorem 12.11** (首一正交多项式的三项递推关系). 对于首一正交多项式序列  $\{p_n(\lambda)\}$  而言, 存在着系数  $\alpha_k$  和  $\gamma_k$ ,  $k = 1, 2, \dots$ , 使得其满足三项递推关系 (*three-term recurrence relation*)

$$\begin{aligned} p_{k+1}(\lambda) &= (\lambda - \alpha_{k+1}) p_k(\lambda) - \gamma_k p_{k-1}(\lambda), \quad k = 1, 2, \dots, \\ p_{-1}(\lambda) &:= 0, \quad p_0(\lambda) = 1, \\ \text{其中 } \alpha_{k+1} &= \frac{\langle \lambda p_k, p_k \rangle}{\langle p_k, p_k \rangle}, \quad \gamma_k = \frac{\langle p_k, p_k \rangle}{\langle p_{k-1}, p_{k-1} \rangle}. \end{aligned} \tag{12.55}$$

证明. 对于  $k = 1, 2, \dots$  设

$$F(\lambda) := p_{k+1}(\lambda) - (\lambda - \alpha_{k+1}) p_k(\lambda) + \gamma_k p_{k-1}(\lambda),$$

由  $p_{k-1}, p_k, p_{k+1}$  都是首一多项式可得  $F(\lambda) \in \mathbb{P}_n[\lambda]$ 。此外, 由  $p_{k-1}, p_k, p_{k+1}$  的正交特性, 我们分三种情况展开分析。

第一种情况, 当  $0 < l < k - 1$  时, 有

$$\langle F(\lambda), p_l(\lambda) \rangle = \underbrace{\langle p_{k+1}(\lambda), p_l(\lambda) \rangle}_{=0} - \underbrace{\langle p_k(\lambda), (\lambda - \alpha_{k+1}) p_l(\lambda) \rangle}_{=0} + \underbrace{\langle \gamma_k p_{k-1}(\lambda), p_l(\lambda) \rangle}_{=0} = 0,$$

其中用到了  $\langle \lambda p, q \rangle = \langle p, \lambda q \rangle$ , 以及  $\langle p_k, p_l \rangle = 0, \forall 0 < l < k - 1$ 。

第二种情况, 当  $l = k - 1$  时, 有

$$\begin{aligned} \langle F(\lambda), p_l(\lambda) \rangle &= \langle F(\lambda), p_{k-1}(\lambda) \rangle \\ &= \underbrace{\langle p_{k+1}(\lambda), p_{k-1}(\lambda) \rangle}_{=0} - \langle p_k(\lambda), (\lambda - \alpha_{k+1}) p_{k-1}(\lambda) \rangle + \langle \gamma_k p_{k-1}(\lambda), p_{k-1}(\lambda) \rangle \\ &= -\langle p_k(\lambda), \lambda p_{k-1}(\lambda) \rangle + \underbrace{\langle p_k(\lambda), \alpha_{k+1} p_{k-1}(\lambda) \rangle}_{=0} + \gamma_k \langle p_{k-1}(\lambda), p_{k-1}(\lambda) \rangle = 0, \end{aligned}$$

因此我们有

$$\gamma_k = \frac{\langle p_k(\lambda), p_k(\lambda) \rangle}{\langle p_{k-1}(\lambda), p_{k-1}(\lambda) \rangle}.$$

第三种情况, 当  $l = k$  时, 有

$$\begin{aligned} \langle F(\lambda), p_l(\lambda) \rangle &= \langle F(\lambda), p_k(\lambda) \rangle \\ &= \underbrace{\langle p_{k+1}(\lambda), p_k(\lambda) \rangle}_{=0} + \langle (-\lambda + \alpha_{k+1}) p_k(\lambda), p_k(\lambda) \rangle + \underbrace{\langle \gamma_k p_{k-1}(\lambda), p_k(\lambda) \rangle}_{=0} \\ &= -\langle \lambda p_k(\lambda), p_k(\lambda) \rangle + \alpha_{k+1} \langle p_k(\lambda), p_k(\lambda) \rangle = 0, \end{aligned}$$

因此我们有

$$\alpha_{k+1} = \frac{\langle \lambda p_k(\lambda), p_k(\lambda) \rangle}{\langle p_k(\lambda), p_k(\lambda) \rangle}.$$

□

**Theorem 12.12** (标准正交多项式的三项递推关系). 对于标准正交多项式  $\{p_n(\lambda)\}$  而言, 存在着系数  $\alpha_k$  和  $\beta_k$ ,  $k = 0, 1, 2, \dots$ , 使得其满足三项递推关系 (*three-term recurrence relation*)

$$\sqrt{\beta_{k+1}} p_{k+1}(\lambda) = (\lambda - \alpha_{k+1}) - \sqrt{\beta_k} p_{k-1}(\lambda), \quad k = 1, 2, \dots$$

$$p_{-1}(\lambda) := 0, \quad p_0(\lambda) = \beta_0^{-\frac{1}{2}}, \quad \beta_0 = \int_a^b d\alpha(\lambda),$$

其中  $\alpha_{k+1} = \langle \lambda p_k, p_k \rangle$ ,  $\beta_k$  通过计算  $|p_k| = 1$  求得.

证明. 假定有一个首一正交多项式  $\{p_k(\lambda)\}_{n=0}^{\infty}$  满足三项递推关系(12.55), 那么根据(12.53), 对应的标准正交多项式  $\{\hat{p}_n(\lambda)\}_{n=0}^{\infty}$  满足

$$\hat{p}_k(\lambda) = \frac{p_k(\lambda)}{\sqrt{\langle p_k, p_k \rangle}} = \frac{p_k(\lambda)}{|p_k|},$$

最后一个等号由(12.49)-(12.52)联立求得。上式代回(12.55)可得

$$\hat{p}_{k+1}(\lambda) |p_k| = \left( \lambda |p_{k+1}| - \frac{\langle \lambda p_k, p_k \rangle}{|p_k|} \right) \hat{p}_k - \left( \frac{|p_k|^2}{|p_{k-1}|} \right) \hat{p}_{k-1},$$

进一步整理

$$\frac{|p_{k+1}|}{|p_k|} \hat{p}_{k+1} = (\lambda - \langle \lambda \hat{p}_k, \hat{p}_k \rangle) \hat{p}_k - \frac{|p_k|}{|p_{k-1}|} \hat{p}_{k-1},$$

由此我们有

$$\langle \lambda \hat{p}_k, \hat{p}_k \rangle = \frac{\langle \lambda p_k, p_k \rangle}{\langle p_k, p_k \rangle} = \alpha_{k+1}, \quad \sqrt{\beta_{k+1}} = \frac{\langle p_{k+1}, p_{k+1} \rangle}{\langle p_k, p_k \rangle} = \gamma_{k+1}.$$

在求得首一多项式  $\{p_n(\lambda)\}_{n=0}^{\infty}$  的三项递归关系系数  $(\alpha_k, \gamma_k)$  后, 对应的标准多项式  $\{\hat{p}_n(\lambda)\}_{n=0}^{\infty}$  的三项递归关系系数变为  $(\alpha_k, \gamma_k^2)$ 。  $\square$

#### 12.B.4 雅各比矩阵和克里斯托费尔——达布关系

如果存在一个标准正交多项式  $\{\hat{p}_n(\lambda)\}_{n=0}^{\infty}$ , 则也存在一个无限的对称三角对角矩阵 (tridiagonal matrix), 用  $J_{\infty}$  表示。由于它的所有  $k$  次对角元素均为正, 因此  $J_{\infty}$  又称无限雅各比矩阵 (infinite Jacobi matrix)

$$J_{\infty} = \begin{pmatrix} \alpha & \sqrt{\beta_1} & & \dots \\ \sqrt{\beta_1} & \alpha_2 & \sqrt{\beta_2} & \dots \\ & \sqrt{\beta_2} & \alpha_3 & \sqrt{\beta_3} & \dots \\ \vdots & \ddots & \ddots & \ddots & \dots \end{pmatrix} \quad (12.56)$$

它的  $k$  次领先主子矩阵 (leading principle submatrix) 称为  $J_k$ 。以  $k = 3$  为例

$$J_3 = \begin{pmatrix} \alpha & \sqrt{\beta_1} & & \\ \sqrt{\beta_1} & \alpha_2 & \sqrt{\beta_2} & \\ & \sqrt{\beta_2} & \alpha_3 & \sqrt{\beta_3} \\ & & \sqrt{\beta_3} & \alpha_4 \end{pmatrix}.$$

进而, 所有的正交多项式都可用雅各比矩阵表现出来:

**Theorem 12.13** (克里斯托费尔——达布关系). 对于正交多项式  $p_k(\lambda)$ ,  $k = 0, 1, 2, \dots$ , 我们有如下克里斯托费尔——达布关系 (Christoffel-Darboux relation)

$$\sum_{i=0}^k p_i(\lambda) p_i(\mu) = \begin{cases} \sqrt{\beta_{k+1}} \frac{p_{k+1}(\lambda)p_k(\mu) - p_k(\lambda)p_{k+1}(\mu)}{\lambda - \mu} & \lambda \neq \mu, \\ \sqrt{\beta_{k+1}} \left[ \left( \frac{d}{d\lambda} p_{k+1}(\lambda) \right) p_k(\lambda) - \left( \frac{d}{d\lambda} p_k(\lambda) \right) p_{k+1}(\lambda) \right] & \lambda = \mu. \end{cases} \quad (12.57)$$

类似地, 对于首一正交多项式, 我们有

$$\sum_{i=0}^k (\gamma_k \gamma_{k-1} \dots \gamma_{i+1}) p_i(\lambda) p_i(\mu) = \frac{p_{k+1}(\lambda)p_k(\mu) - p_k(\lambda)p_{k+1}(\mu)}{\lambda - \mu}, \quad \lambda \neq \mu.$$

证明. 由三项递推关系定理(12.11)可得, 正交多项式满足如下关系

$$\begin{aligned} p_{n+1}(\lambda) - A_n \lambda p_n(\lambda) &= B_n p_n(\lambda) + c_{n-1}(\lambda) p_{n-1}(\lambda), \quad n = 1, 2, 3 \dots \\ A_n &= \frac{k_{n+1}}{k_n}, \quad C_n = -\frac{A_n}{A_{n-1}} \frac{h_n}{h_{n-1}}, \end{aligned}$$

其中  $k_n$  为  $n$  次项对应的系数: 将  $p_{n+1}(\lambda) - A_n \lambda p_n(\lambda)$  看做一个多项式,  $p_{n+1}(\lambda) - A_n \lambda p_n(\lambda) = \sum_{k=0}^n c_k p_k(\lambda)$ 。由此我们有

$$\begin{aligned} p_{n+1}(\lambda) p_n(\mu) &= (A_n \lambda + B_n) p_n(\lambda) p_n(\mu) + c_n p_{n-1}(\lambda) p_n(\mu), \\ p_{n+1}(\mu) p_n(\lambda) &= (A_n \mu + B_n) p_n(\mu) p_n(\lambda) + c_n p_{n-1}(\mu) p_n(\lambda). \end{aligned}$$

上两式相减得

$$p_{n+1}(\lambda) p_n(\mu) - p_{n+1}(\mu) p_n(\lambda) = A_n (\lambda - \mu) p_n(\lambda) p_n(\mu) + C_n [p_{n-1}(\lambda) p_n(\mu) - p_{n-1}(\mu) p_n(\lambda)]$$

↪ 裂项和 (telescoping sum)<sup>24</sup>

$$(\lambda - \mu) p_n(\lambda) p_n(\mu) = \frac{p_{n+1}(\lambda) p_n(\mu) - p_{n+1}(\mu) p_n(\lambda)}{A_n} - \frac{C_n}{A_n} [p_{n-1}(\lambda) p_n(\mu) - p_{n-1}(\mu) p_n(\lambda)],$$

↪

$$(\lambda - \mu) \frac{p_n(\lambda) p_n(\mu)}{h_n} = \frac{p_{n+1}(\lambda) p_n(\mu) - p_{n+1}(\mu) p_n(\lambda)}{A_n h_n} - \frac{p_{n-1}(\lambda) p_n(\mu) - p_{n-1}(\mu) p_n(\lambda)}{A_{n-1} h_{n-1}},$$

↪

$$\begin{aligned} (\lambda - \mu) \sum_{k=1}^n \frac{p_k(\lambda) p_k(\mu)}{h_k} &= \sum_{k=1}^n \frac{p_{k+1}(\lambda) p_k(\mu) - p_{k+1}(\mu) p_k(\lambda)}{A_k h_k} - \sum_{k=1}^n \frac{p_k(\lambda) p_{k-1}(\mu) - p_{k-1}(\lambda) p_k(\mu)}{A_{k-1} h_{k-1}} \\ &= \frac{p_{n+1}(\lambda) p_n(\mu) - p_{n+1}(\mu) p_n(\lambda)}{A_n h_n} - \frac{k_0^2 (\lambda - \mu)}{h_0}, \end{aligned}$$

↪

$$\begin{aligned} (\lambda - \mu) \sum_{k=0}^n \frac{p_k(\lambda) p_k(\mu)}{h_k} &= \frac{p_{n+1}(\lambda) p_n(\mu) - p_{n+1}(\mu) p_n(\lambda)}{A_n h_n} \\ &= \frac{1}{h_n} \frac{k_n}{k_{n+1}} [p_{n+1}(\lambda) p_n(\mu) - p_n(\lambda) p_{n+1}(\mu)], \quad \lambda \neq \mu. \end{aligned}$$

另, 对于  $\mu \rightarrow \lambda$  时的极限, 上式变为

$$\begin{aligned} \lim_{\mu \rightarrow \lambda} \frac{p_{n+1}(\lambda) p_n(\mu) - p_{n+1}(\mu) p_n(\lambda)}{\lambda - \mu} &= \lim_{\mu \rightarrow \lambda} \frac{p_n(\lambda) [p_{n+1}(\lambda) - p_{n+1}(\mu)] - p_{n+1}(\lambda) [p_n(\lambda) - p_n(\mu)]}{\lambda - \mu} \\ &\approx \left[ \frac{d}{d\lambda} p_{n+1}(\lambda) \right] p_n(\lambda) - \left[ \frac{d}{d\lambda} p_n(\lambda) \right] p_{n+1}(\lambda). \end{aligned}$$

□

因此, 对于矩阵形式的正交多项式  $p_k(\lambda) = (p_0(\lambda), p_1(\lambda), \dots, p_k(\lambda))^\top$  来说, 它可以写成如下的三项递推关系

$$\lambda p_k = J_k p_k + \eta_k p_k(\lambda) e^{(k)},$$

其中  $e^{(k)}$  为单位矩阵 (identity matrix)  $\eta_k = \sqrt{\beta_k}$  的最后一列。 $p_k(\lambda)$  的根  $\theta_j^k$  是  $J_k$  的特征值。

<sup>24</sup>裂项和是指一个求和方程中, 后续各项相互抵消, 只剩下初始项和最高项的情况, 如

$$S = \sum_{i=1}^{n-1} (a_i - a_{i+1}) = (a_1 - a_2) + (a_2 - a_3) + \dots + (a_{n-1} - a_n) = a_1 - a_n.$$

### 12.B.5 正交多项式的根

**Theorem 12.14** (正交多项式根的特征). 正交多项式序列  $\{p_n(\lambda)\}$  中, 第  $n$  次多项式  $p_n(\lambda)$  在区间  $(a, b)$  中, 正好有  $n$  项唯一且简单的实根。

证明. 根据定义, 第  $n$  次多项式  $p_n(\lambda)$  至多有  $n$  个实根。

假设  $p_n(\lambda)$  有  $m \leq n$  个奇数实根  $\lambda_1, \lambda_2, \dots \in \lambda_m(a, b)$ , 则  $p_n(\lambda)(\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_m)$  的符号不变。这意味着

$$\int_a^b w(\lambda)p_n(\lambda)(\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_m) d\lambda \neq 0.$$

另一方面由正交条件得如果  $m < n$ , 则

$$\int_a^b w(\lambda)p_n(\lambda)(\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_m) d\lambda = 0.$$

因此, 我们得出  $m = n$ , 即  $p_n(x)$  在空间  $(a, b)$  中有  $n$  个唯一的奇数次实根。  $\square$

**Theorem 12.15.** 如果  $\{p_n(\lambda)\}_{n=0}^{\infty}$  是一个在  $(a, b)$  区间中, 关于权重方程  $w(\lambda)$  的正交多项式序列, 那么  $p_n(\lambda)$  的根和  $p_{n+1}(\lambda)$  的根彼此错落分布。

证明. 由范的定义可得,

$$h_n = \int_a^b w(\lambda)p_n(\lambda) d\lambda > 0, \quad n = 0, 1, 2 \dots$$

由克里斯托费尔——达布公式 (Theorem 12.13) 可得

$$h_n \frac{k_n}{k_{n+1}} \left\{ p_n(\lambda) \left[ \frac{d}{d\lambda} p_{n+1}(\lambda) \right] - p_{n+1}(\lambda) \left[ \frac{d}{d\lambda} p_n(\lambda) \right] \right\} = \sum_{k=0}^n \frac{p_k(\lambda)^2}{h_k} > 0,$$

因此我们有

$$\frac{k_n}{k_{n+1}} \left\{ p_n(\lambda) \left[ \frac{d}{d\lambda} p_{n+1}(\lambda) \right] - p_{n+1}(\lambda) \left[ \frac{d}{d\lambda} p_n(\lambda) \right] \right\} = \sum_{k=0}^n \frac{p_k(\lambda)^2}{h_k} > 0.$$

现在, 假定  $p_n(\lambda)$  中有两个连续的根  $\lambda_{n,k}$  和  $\lambda_{n,k+1}$  满足  $\lambda_{n,k} < \lambda_{n,k+1}$ 。由于  $p_n(\lambda)$  中全部  $n$  个根都是简单的实数, 因此  $p'_n(\lambda_{n,k})$  和  $p'_{n+1}(\lambda_{n,k+1})$  的符号应当相反。于是我们有

$$\begin{aligned} p_n(\lambda_{n,k}) &= p_n(\lambda_{n,k+1}) = 0, \\ p'_n(\lambda_{n,k})p'_n(\lambda_{n,k+1}) &< 0, \\ p'_{n+1}(\lambda_{n,k})p'_{n+1}(\lambda_{n,k+1}) &< 0, \end{aligned}$$

根据多项式序列的连续性特征, 我们有: 在  $\lambda_{n,k}$  和  $\lambda_{n,k+1}$  之间, 至少应当有一个根落在  $(\lambda_{n,k}, \lambda_{n,k+1})$  区间中。证毕。  $\square$

进而, 如果  $p_n(\lambda)$  和  $p_{n+1}(\lambda)$  中的根分别用  $\{\lambda_{n,k}\}_{k=1}^n$  和  $\{\lambda_{n+1,k}\}_{k=1}^{n+1}$  来表示, 则我们有

$$a < \lambda_{n+1,1} < \lambda_{n,1} < \lambda_{n+1,2} < \lambda_{n,2} < \dots < \lambda_{n+1,n} < \lambda_{n,n} < \lambda_{n+1,n+1} < b.$$

## 12.B.6 高斯求积

在前面的章节中，我们用  $\lambda$  符号来表示变量。随后我们改用  $x$ ，主要原因是打字起来方便。。。

**Theorem 12.16** (高斯求积公式). 如果在区间  $(a, b)$  中有一个连续方程  $f(x)$ ，对应  $a < x_1 < x_2 < \dots < x_n$  的  $n$  个点。那么存在且只存在一个 ( $\leq n - 1$  次的) 多项式  $p$ ，使得  $p(x_i) = f(x_i)$ ，那么可用高斯积公式 (Gauss quadrature formula) 来近似  $f(x)$  的值，满足

$$\int_a^b w(x)f(x) dx = \sum_{i=1}^n \lambda_{n,i} f(x_i), \quad \lambda_{n,i} := \int_a^b \frac{w(x)p_n(x)}{(x-x_i)p'_n(x_i)} dx, i = 1, 2, \dots, n. \quad (12.58)$$

证明. 可以使用拉格朗日插值法 (Lagrange interpolation) 求解多项式  $p$ 。定义

$$p(x) := (x - x_1)(x - x_2) \dots (x - x_n)$$

以及

$$\begin{aligned} P(x) &:= \sum_{i=1}^n f(x_i) \frac{p(x)}{(x - x_i)p'(x_i)} \\ &= \sum_{i=1}^n f(x_i) \frac{(x - x_1)(x - x_2) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x_i - x_n)}{(x_i - x_1)(x_i - x_2) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)}. \end{aligned}$$

定义一个区间  $(a, b)$  内关于权重方程  $w(x)$  的正交多项式序列  $\{p_n(x)\}_{n=0}^\infty$ 。那么对于  $x_1 < x_2 < \dots < x_n$ ，我们有  $n$  个多项式  $p_n(x)$  的实根。如果  $f(x)$  是一个  $\leq 2n - 1$  次的多项式，那么  $f(x) - P(x)$  就是一个至少包括这些  $x_1 < x_2 < \dots < x_n$  实根在内的  $\leq 2n - 1$  次的多项式。因此，我们将  $f(x)$  定义为

$$f(x) = P(x) + r(x)p_n(x),$$

其中  $r(x)$  是一个  $\leq n - 1$  次的多项式。引入  $P(x)$  的定义式，将上式改写为

$$f(x) = \sum_{i=1}^n f(x_i) \frac{p(x)}{(x - x_i)p'_n(x_i)} + r(x)p_n(x).$$

这意味着

$$\int_a^b w(x)f(x) dx = \sum_{i=1}^n f(x_i) \int_a^b \frac{w(x)p_n(x)}{(x - x_i)p'_n(x_i)} dx + \int_a^b w(x)r(x)p_n(x) dx.$$

由于  $r(x)$  是一个  $\leq n - 1$  次的多项式，因此我们有  $\int_a^b w(x)r(x)p_n(x) dx = 0$ 。调整后的高斯求积如(12.58)所示。在已知  $x_1 < x_2 < \dots < x_n$  所对应的  $p_n(x)$  的  $n$  个实根，如果对应的  $f(x_i)$  也是已知的，则高斯求积式给出了  $\leq 2n - 1$  次多项式  $f(x)$  的积分值。

如果  $f(x)$  不是一个  $\leq 2n - 1$  的多项式，高斯求积式转而求积分的近似：

$$\int_a^b w(x)f(x) dx \approx \sum_{i=1}^n \lambda_{n,i} f(x_i), \quad \lambda_{n,i} := \int_a^b \frac{w(x)p_n(x)}{(x - x_i)p'_n(x_i)} dx, \quad i = 1, 2, \dots, n.$$

□

我们将  $\{\lambda_{n,i}\}_{n=0}^\infty$  称作克里斯托费尔数 (Christoffel numbers)，它不受到  $f(x)$  方程的影响。

**Theorem 12.17** (克里斯托费尔数的值全部为正)。

证明. 将  $\lambda_{n,i}$  改写为

$$\lambda_{n,i} = \int_a^b w(x)l_{n,i}(x)dx, \quad l_{n,i} := \frac{p_n(x)}{(x-x_i)p'_n(x_i)}, \quad i = 1, 2, \dots, n.$$

那么对应  $p_n(x)$  的实根  $\{x_{n,k}\}_{k=1}^n$ ,  $\leq 2n-2$  次的多项式  $l_{n,i}^2 - l_{n,i}(x)$  等于零。换句话说对于某个  $\leq n-2$  次的多项式  $q$  来说, 我们有

$$l_{n,i}^2 - l_{n,i}(x) = p_n(x)q(x).$$

根据正交条件, 上式变为

$$\int_a^b w(x)(l_{n,i}^2 - l_{n,i}(x))dx = \int_a^b w(x)p_n(x)q(x)dx = 0.$$

由此我们有

$$\lambda_{n,i} = \int_a^b w(x)l_{n,i}(x)dx = \int_a^b w(x)l_{n,i}(x)^2 dx > 0.$$

□

## 12.C 常见的正交多项式类型

本节介绍几种常见的正交多项式, 包括

- 埃米特多项式 (Hermite polynomials)
- 拉盖尔多项式 (Laguerre polynomials)
- 雅各比多项式 (Jacobi polynomials)
- 勒让德多项式 (Legendre polynomials)
- 车比雪夫多项式 (Chebyshev polynomials, 1st kind, 2nd kind)。

表 12.C.1: 经典正交多项式

名称	$p_n(x)$	$w(x)$	(a,b)
Hermite	$H_n(x)$	$\exp(-x^2)$ (Normal dist.)	$(-\infty, +\infty)$
Laguerre	$L_n^\alpha(x)$	$\exp(-\alpha)x^\alpha$ (Gamma dist.)	$[0, +\infty)$
Jacobi	$P_n^{\alpha, \beta}(x)$	$(1-x)^\alpha(1+x)^\beta$ (Beta dist.)	$(-1, 1)$
Legendre <sup>*</sup>	$P_n(x)$	1 (Uniform dist.)	$[-1, 1]$
Chebyshev <sup>**</sup> , 1st kind	$T_n(x)$	$\frac{1}{\sqrt{1-x^2}}$	$[-1, 1]$
Chebyshev, 2nd kind	$U_n(x)$	$\sqrt{1-x^2}$	$[-1, 1]$

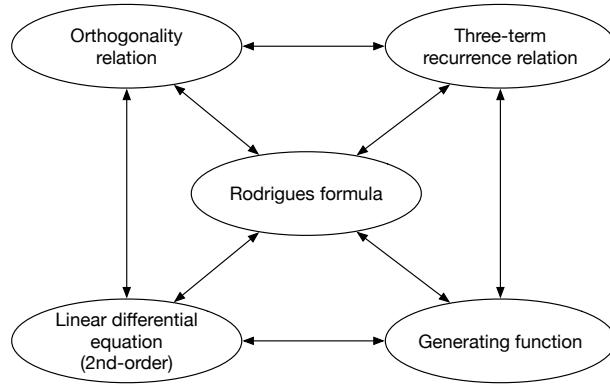
<sup>\*</sup> Legendre 是 Jacobi 的一种特殊形式,  $\alpha = \beta = 0$ 。

<sup>\*\*</sup> Chebyshev 是 Legendre 的一种特殊形式。

这些多项式往往满足如下条件 (图12.C.1)

- 罗德里格斯公式 (Rodrigues formula)
- 正交条件 (orthogonality)
- 三项递推关系 (three-term recurrence relation)
- 二阶线性微分方程 (2nd order linear differential equation)
- 母函数 (generating function)。

图 12.C.1: 正交多项式满足的关系



### 12.C.1 埃米特多项式

**Theorem 12.18** (埃米特多项式的罗德里格斯公式). 在区间  $(-\infty, +\infty)$  内, 关于正态分布 (normal distribution)  $w(x) = \exp(-x^2)$  的埃米特正交多项式 (Hermite polynomial)  $H_n(x)$ , 可由罗德里格斯公式予以定义 (Rodrigues formula)

$$H_n(x) = \frac{(-1)^n}{w(x)} D^n w(x) = (-1)^n \exp(x^2) D^n \exp(-x^2), \quad n = 0, 1, 2, \dots \quad (12.59)$$

其中  $(-1)^n$  项是为了保证  $\{D^n w(x)\}$  的每一个首项系数都为正。 $D = \frac{d}{dx}$  是微分符,  $D^n$  是第  $n$  次求导。 $D^n$  遵循莱布尼兹法则 (Leibniz rule)

$$D^n [f(x)g(x)] = \sum_{k=0}^n = \sum_{k=0}^n \binom{n}{k} D^k f(x) D^{n-k} g(x), \quad n = 0, 1, 2, \dots, \quad (12.60)$$

其中  $\binom{n}{k} = \frac{n!}{(n-k)!}$  是帕斯卡三角 (Pascal triangle identity) 中的二项式系数。帕斯卡三角满足关系

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}.$$

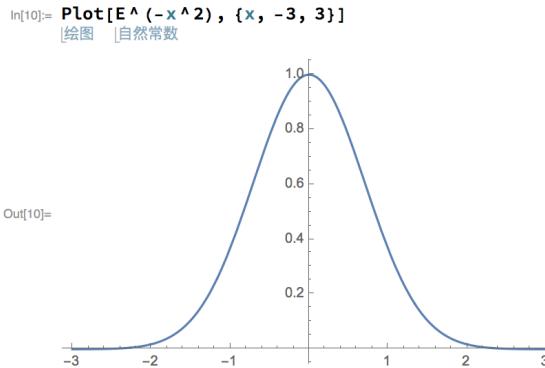
帕斯卡三角的简单证明:

$$\begin{aligned}
 & \binom{n}{k} + \binom{n}{k-1} \\
 &= \frac{n!}{k!(n-k)!} + \frac{n!}{(k-1)!(n-k+1)!} \\
 &= n! \left\{ \frac{n-k+1}{k!(n-k+1)!} + \frac{k}{k!(n-k+1)!} \right\} \\
 &= \frac{(n+1)!}{k!(n+1-k)!} \\
 &= \binom{n+1}{k}.
 \end{aligned}$$

正态分布  $w(x) = \exp(-x^2)$  在区间  $(-3, 3)$  内, 如图 12.C.2 所示。

**Theorem 12.19.** 由 (12.59) 定义的埃米特多项式  $H_n(x)$  是一个关于  $x$  的  $n$  次多项式, 并且  $H_0(1) = x$ ,  $H_n(x)$  的首项系数  $k_n = 2^n$ ,  $H_{2n}(x)$  是偶方程,  $H_{2n+1}(x)$  是奇方程。

图 12.C.2: 正态分布



$w(x) = \exp(-x^2)$  在  $(-3, 3)$  区间内的正态分布。

证明. (12.59) $\Rightarrow$

$$\begin{aligned}
 D^{n+1}w(x) &= D[D^n w(x)] \\
 &= D[(-1)^n H_n(x)w(x)] \\
 &= (-1)^n [w'(x)H_n(x) + w(x)H'_n(x)] \\
 &= (-1)^n [-2x \exp(-x^2)H_n(x) + w(x)H_n(x)] \\
 &= (-1)^{n+1}w(x)[2xH_n(x) - H'_n(x)], \quad n = 0, 1, 2 \dots
 \end{aligned} \tag{12.61}$$

由此, (12.59) $\Rightarrow$

$$H_{n+1}(x) = \begin{cases} 1 & n = 0 \\ \frac{(-1)^{n+1}}{w(x)}D^{n+1}w(x) = 2xH_n(x) - H'_n(x) & n = 1, 2 \dots \end{cases} \tag{12.62}$$

(12.62)决定了  $H_n(x)$  是一个  $n$  次多项式, 并且  $H_n(x)$  的首项系数  $k_n = 2^n$ , 可以写出如下序列

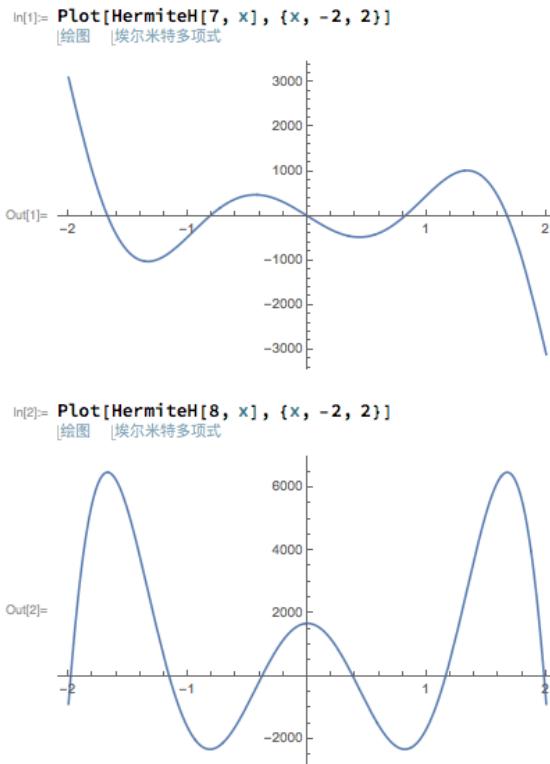
$$\begin{aligned}
 p_0(x) &= 1, \\
 p_1(x) &= 2x, \\
 p_2(x) &= 4x^2 - 2, \\
 p_3(x) &= 8x^3 - 12x, \\
 p_4(x) &= 16x^4 - 48x^2 + 12, \\
 p_5(x) &= 32x^5 - 160x^3 + 120x, \\
 p_6(x) &= 64x^6 - 480x^4 + 720x^2 - 120, \\
 p_7(x) &= 128x^7 - 1344x^5 + 3360x^3 - 1680x, \\
 &\vdots
 \end{aligned}$$

由上式可得, 对于偶数次  $2n = 2, 4, 6 \dots$  的情况我们有  $p_{2n}(x) = p_{2n}(-x)$  是偶方程; 对于基数次  $2n + 1 = 1, 3, 5 \dots$  的情况我们有  $p_{2n+1}(x) = -p_{2n}(-x)$  是奇方程, 如图12.C.3。  $\square$

**Theorem 12.20** (埃米特多项式的正交条件). 埃米特多项式  $H_n(x)$  满足正交条件 (*orthogonality condition*)

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-x^2) H_m(x) H_n(x) dx = 2^n n! \delta_{mn}, \quad m, n = 0, 1, 2 \dots \tag{12.63}$$

图 12.C.3: 埃米特多项式



上图和下图分别表示  $n = 7, n = 8$  时,  $H_n(x)$  在  $(-2, 2)$  区间内的值。

证明. 由埃米特多项式的罗德里格斯定义式(12.59)我们有, 当  $m < n$  时

$$\begin{aligned} \int_{-\infty}^{\infty} \exp(-x^2) H_m(x) H_n(x) dx &= \int_{-\infty}^{\infty} \exp(-x^2) H_m(x) [(-1)^n \exp(x^2) D^n \exp(-x^2)] dx \\ &= \int_{-\infty}^{\infty} (-1)^n \exp(-x^2) H_m(x) dx, \end{aligned}$$

对上式做  $n$  次求导, 积分内的值变为零。

当  $m = n$  时  $\Rightarrow$

$$\begin{aligned} \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-x^2) H_m(x) H_n(x) dx &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-x^2) H_n(x) H_n(x) dx \\ &= \frac{1}{\sqrt{\pi}} (-1)^n \int_{-\infty}^{\infty} H_n(x) D^n \exp(-x^2) dx \\ &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} D^n H_n(x) \exp(-x^2) dx \\ &= \frac{k_n n!}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-x^2) dx \\ &= k_n n!. \end{aligned}$$

证毕。 □

**Theorem 12.21** (埃米特多项式的三项递推关系). 埃米特多项式的三项递推关系为

$$H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x), n = 1, 2, 3 \dots \quad (12.64)$$

证明. 由正态分布  $w(x) = \exp(-x^2)$  我们有

$$w'(x) = -2xw(x).$$

代入莱布尼兹法则(12.60) $\Rightarrow$

$$D^{n+1}w(x) = D^n w'(x) = D^n [-2xw(x)] = (-2x)D^n w(x) + (-2n)D^{n-1}w(x).$$

将上式代回(12.62) $\Rightarrow$

$$\begin{aligned} H_{n+1}(x) &= \frac{(-1)^{n+1}}{w(x)} D^{n+1}w(x) \\ &= (-1) \frac{(-1)^n}{w(x)} (-2x)D^n w(x) + (-1)^2 \frac{(-1)^{n-1}}{w(x)} (-2n)D^{n-1}w(x) \\ &= \frac{(-1)^n}{w(x)} w(x)D^n w(x) - 2n \frac{(-1)^{n-1}}{w(x)} D^{n-1}w(x) \\ &= 2xH_n(x) - 2nH_{n-1}(x), n = 1, 2, 3 \dots \end{aligned}$$

□

**Theorem 12.22** (埃米特多项式的二阶线性微分方程). 埃米特多项式的二阶线性微分方程 (*second order linear differential equation*) 为

$$H_n''(x) + 2nH_n(x) - 2xH'_n(x) = 0. \quad (12.65)$$

证明. 联立埃米特多项式的罗德里格斯公式(12.59)和三项递归关系式(12.64)我们有

$$\begin{aligned} H'_n(x) &= 2nH_{n-1}(x), n = 1, 2, 3 \dots \\ H'_{n+1}(x) &= 2(n+1)H_n(x), n = 0, 1, 2 \dots \end{aligned}$$

对(12.62)再做一次求导, 并引入上式替换  $H'_{n+1}(x)$  我们有

$$\begin{aligned} H'_{n+1}(x) &= 2xH'_n(x) + 2H_n(x) - H''_n(x), \\ \hookrightarrow 2(n+1)H_n(x) &= 2xH'_n(x) + 2H_n(x) - H''_n(x), \end{aligned}$$

这意味着  $H_n(x)$  构成一个二阶线性微分方程系统

$$y''(x) - 2xy'(x) + 2ny(x) = 0. \quad (12.66)$$

□

**Theorem 12.23** (埃米特多项式的母方程). 埃米特多项式的母方程 (*generating function*) 为

$$\exp(2xt - t^2) = \sum_{n=0}^{\infty} \frac{H_n(x)}{n!} t^n. \quad (12.67)$$

证明. 设  $F(t) = \exp(-(x-t)^2) = \exp(-x^2) \exp(2xt - t^2)$ . 围绕  $\tilde{t} = 0$  对  $F(t)$  做泰勒级数展开近似

$$F(t) \approx \sum_{n=0}^{\infty} \frac{F^n(0)}{n!} t^n.$$

设  $u = x - t$ ,  $\lim_{t \rightarrow 0} u \approx x$ . 则

$$\begin{aligned} F^n(0) &= \frac{d^n}{dt^n} \exp(-(x-t)^2) = \left[ (-1)^n \frac{d^n}{du^n} \exp(-u^2) \right]_{u=x} \\ &= (-1)^n D^n \exp(-x^2) = \exp(-x^2) H_n(x), n = 0, 1, 2 \dots \end{aligned}$$

↪

$$\begin{aligned} \exp(-x^2) \exp(-2xt - t^2) &= \exp(-(x-t)^2) \\ &= F(t) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} t^n = \exp(-x^2) \sum_{t=0}^{\infty} \frac{H_n(x)}{n!} t^n. \end{aligned}$$

□

## 12.C.2 拉盖尔多项式

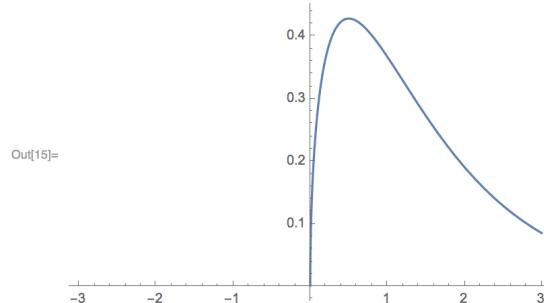
**Theorem 12.24** (拉盖尔多项式的罗德里格斯公式). 在  $(0, \infty)$  区间内, 关于伽玛分布 (*Gamma distribution*)  $w(x) = \exp(-x)x^\alpha$  的拉盖尔多项式 (*Laguerre polynomial*)  $L_n^{(\alpha)}(x)$ , 可由罗德里格斯公式予以定义

$$L_n^{(\alpha)}(x) = \frac{1}{n!} \frac{1}{w(x)} D^n [w(x)x^n] = \frac{1}{n!} \exp(x)x^{-\alpha} D^n [\exp(-x)x^{n+\alpha}], \quad n = 0, 1, 2 \dots \quad (12.68)$$

当  $\alpha = 1/2$  时伽玛分布  $w(x) = \exp(-x)x^\alpha$  在区间  $(-3, 3)$  内, 如图12.C.4所示。

图 12.C.4: 伽玛分布

In[15]:= Plot[E^(-x) x^(1/2), {x, -3, 3}]  
|绘图 |自然常数



$w(x) = \exp(-x)x^\alpha$  在  $(-3, 3)$  区间内的正态分布 ( $\alpha = 1/2$ )。

**Theorem 12.25.** 拉盖尔多项式  $L^{(n)}(x)$  是一个关于  $x$  的  $n$  次多项式。 $L^{(\alpha)}(0) = \frac{(\alpha+1)_n}{n!}$ ,  $L^{(n)}(x)$  的首项系数  $k_n = \frac{(-1)^n}{n!}$ 。

证明. 根据莱布尼兹法则(12.60)我们有

$$\begin{aligned} D^n [\exp(-x)x^{n+\alpha}] &= \sum_{k=0}^n \binom{n}{k} D^k \exp(-x) D^{n-k} x^{n+\alpha} \\ &= \sum_{k=0}^n \binom{n}{k} (-1)^k \exp(-x) x^{\alpha+k} (n+\alpha)(n+\alpha-1)\dots(n+k+1) \\ &= \exp(-x)x^\alpha \sum_{k=0}^n (-1)^k \binom{n}{k} \frac{\Gamma(n+\alpha+1)}{\Gamma(k+\alpha+1)} x^k. \end{aligned}$$

最后一个等式的简单数学计算:

$$\frac{\Gamma(n+\alpha+1)}{\Gamma(k+\alpha+1)} = \frac{\int_0^\infty \exp(-t)t^{n+\alpha} dt}{\int_0^\infty \exp(-t)t^{k+\alpha} dt} \approx \frac{(n+\alpha)!}{(k+\alpha)!}$$

因此我们有

$$L_n^{(\alpha)}(x) = \sum_{k=0}^n (-1)^k \binom{n+\alpha}{n-k} \frac{x^k}{k!}, \quad \text{其中 } \binom{n+\alpha}{n-k} = \frac{\Gamma(n+\alpha+1)}{\Gamma(k+\alpha+1)} (n-k)! \quad (12.69)$$

由上式可见,  $L_n^{(\alpha)}(x)$  是一个  $n$  次多项式。此外, 由于

$$(-1)^k \binom{n+\alpha}{n-k} = \frac{(-1)^k}{(n-k)!} \frac{(n+\alpha)!}{(\alpha+k)!} = \frac{(-1)^k}{(n-k)!} \frac{(\alpha+1)_n}{(\alpha+1)_k} = \frac{(\alpha+1)_n}{n!} \frac{(-n)_k}{(\alpha+1)_k},$$

我们可得拉盖尔多项式的超几何方程

$$L_n^{(\alpha)}(x) = \frac{(\alpha+1)_n}{n!} \sum_{k=0}^n \frac{(-n)_k}{(\alpha+1)_k} \frac{x^k}{k!} = \binom{n+\alpha}{n} {}_1F_1\left(\begin{matrix} -n \\ \alpha+1 \end{matrix}; x\right). \quad (12.70)$$

由(12.70)得  $x=0$  时

$$L_n^{(\alpha)}(0) = \binom{n+\alpha}{n} = \frac{(\alpha+1)_n}{n!}, \quad n=0, 1, 2 \dots$$

$\hookrightarrow$

$$k_n = \frac{(-1)^n}{n!}, \quad n=0, 1, 2 \dots$$

□

**Theorem 12.26** (拉盖尔多项式的正交条件). 拉盖尔多项式  $L_n^{(\alpha)}(x)$  的正交条件满足

$$\int_0^\infty \exp(-x) x^\alpha L_m^{(\alpha)}(x) L_n^{(\alpha)}(x) dx = \frac{\Gamma(n+\alpha+1)}{n!} \delta_{mn}, \alpha > -1. \quad (12.71)$$

证明. 定义矩

$$\mu_n := \int_0^n \exp(-x) x^{n+\alpha} dx,$$

如果对于所有  $n=0, 1, 2 \dots$ ,  $\mu_n$  都存在, 那么(12.71)的 LHS 积分是收敛的, 因此需要  $\alpha > -1$ 。

此外根据定义我们有

$$\mu_n = \Gamma(n+\alpha+1).$$

进而, 由拉盖尔多项式的罗德里格斯公式(12.68)我们有

$$\begin{aligned} & \int_0^\infty \exp(-x) x^\alpha L_m^{(\alpha)}(x) L_n^{(\alpha)}(x) dx \\ &= \int_0^\infty \exp(-x) x^\alpha L_m^{(\alpha)}(x) \left\{ \frac{1}{n!} \exp(-x) x^{n+\alpha} D^n [\exp(-x) x^{n+\alpha}] \right\} dx \\ &= \frac{-1}{n!} \int_0^\infty L_m^{(\alpha)}(x) D^n [\exp(-x) x^{n+\alpha}] dx \\ &= \frac{(-1)^n}{n!} \int_0^\infty L_m^{(\alpha)}(x) D^n [\exp(-x) x^{n+\alpha}] dx \end{aligned}$$

当  $m < n$  时, 对上式做  $n$  次求导, 积分内的值变为零。当  $m = n$  时  $\Rightarrow$

$$\begin{aligned} & \frac{(-1)^n}{n!} \int_0^\infty L_m^{(\alpha)}(x) D^n [\exp(-x) x^{n+\alpha}] dx \\ &= \frac{(-1)^n}{n!} k_n n! \int_0^\infty \exp(-x) x^{n+\alpha} dx \\ &= \frac{\Gamma(n+\alpha+1)}{n!}. \end{aligned}$$

□

**Theorem 12.27** (拉盖尔多项式的母方程). 拉盖尔多项式  $L_n^{(\alpha)}(x)$  的母方程为

$$(1-t)^{-\alpha-1} \exp\left(-x \frac{t}{1-t}\right) = \sum_{n=0}^{\infty} L_n^{(\alpha)}(x) t^n. \quad (12.72)$$

证明. 将(12.70)代入(12.72)RHS  $\Rightarrow$

$$\begin{aligned} & \sum_{n=0}^{\infty} L_n^{(\alpha)}(x) t^n \\ &= \sum_{n=0}^{\infty} \frac{(\alpha+1)_n}{n!} t^n \sum_{k=0}^n \frac{(-n)_k}{(\alpha+1)_k} \frac{x^k}{k!} \\ &= \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} \frac{(\alpha+1)_n}{(\alpha+1)_k} \frac{(-1)^k x^k t^n}{k!(n-k)!} \\ &= \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} \frac{(\alpha+1)_{n+k}}{(\alpha+1)_k} \frac{(-1)^k x^k t^{n+k}}{k! n!} \\ &= \sum_{k=0}^{\infty} \frac{(-xt)^k}{k!} \sum_{n=0}^{\infty} \frac{(\alpha+k+1)_n}{n!} t^n \\ &= \sum_{k=0}^{\infty} \frac{(-xt)^k}{k!} (1-t)^{(-\alpha-k-1)} \\ &= (1-t)^{-\alpha-1} \sum_{k=0}^{\infty} \frac{1}{k!} \left(-x \frac{t}{1-t}\right)^k \\ &= (1-t)^{-\alpha-1} \exp\left(-x \frac{t}{1-t}\right). \end{aligned}$$

□

**Theorem 12.28.** 由拉盖尔多项式的母方程(12.72)我们有

$$L_n^{(\alpha+\beta+1)}(x+y) = \sum_{k=0}^n L_k^{(\alpha)}(x) L_{n-k}^{(\beta)}(y). \quad (12.73)$$

证明. (12.72) $\Rightarrow$

$$\begin{aligned} \sum_{n=0}^{\infty} L_n^{(\alpha+\beta+1)}(x+y) t^n &= (1-t)^{-\alpha-\beta-2} \exp\left[-(x+y) \frac{t}{1-t}\right] \\ &= \left[(1-t)^{-\alpha-1} \exp\left(-x \frac{t}{1-t}\right)\right] \left[(1-t)^{(-\beta-1)} \exp\left(-y \frac{1}{1-t}\right)\right] \\ &= \left[\sum_{k=0}^{\infty} L_k^{(\alpha)}(x) t^k\right] \left[\sum_{m=0}^{\infty} L_m^{(\beta)}(y) t^m\right] \\ &= \sum_{n=0}^{\infty} \left[ \sum_{k=0}^n L_k^{(\alpha)}(x) L_{n-k}^{(\beta)}(y) \right] t^n. \end{aligned}$$

证毕。 □

**Theorem 12.29** (拉盖尔多项式的三项递推关系). 拉盖尔多项式  $L_n^{(\alpha)}(x)$  的三项递推关系为

$$(n+1)L_{n+1}^{(\alpha)}(x) + (x-2n-\alpha-1)L_n^{(\alpha)}(x) + (n+\alpha)L_{n-1}^{(\alpha)}(x) = 0, \quad n = 1, 2, 3 \dots \quad (12.74)$$

证明. 首先由(12.69)得

$$\begin{aligned}
 \frac{d}{dx} L_n^{(\alpha)}(x) &= \frac{d}{dx} \sum_{k=0}^n (-1)^k \binom{n+\alpha}{n-k} \frac{x^k}{k!} \\
 &= \sum_{k=1}^n (-1)^k \binom{n+\alpha}{n-k} \frac{x^{k-1}}{(k-1)!} \\
 &= \sum_{k=0}^{n-1} (-1)^{k+1} \binom{n+\alpha}{n-k-1} \frac{x^k}{(k+1)!} \\
 &= -L_{n-1}^{(\alpha+1)}(x), n = 1, 2, 3 \dots
 \end{aligned} \tag{12.75}$$

为简化表述, 根据拉盖尔多项式的母方程(12.72), 定义

$$F(x, t) := (1-t)^{-\alpha-1} \exp\left(-x \frac{t}{1-t}\right) = \sum_{n=0}^{\infty} L_n^{(\alpha)}(x) t^n.$$

则我们有第一个偏导数

$$\begin{aligned}
 \frac{\partial F(x, t)}{\partial x} &= (-t)(1-t)^{-\alpha-2} \exp\left(-x \frac{t}{1-t}\right), \\
 \Rightarrow (1-t) \frac{\partial F(x, t)}{\partial x} + tF(x, t) &= 0, \\
 \Rightarrow (1-t) \sum_{n=0}^{\infty} \frac{d}{dx} L_n^{(\alpha)}(x) t^n + t \sum_{n=0}^{\infty} L_n^{(\alpha)}(x) t^n &= 0, \\
 \Rightarrow \underbrace{\sum_{n=0}^{\infty} \frac{d}{dx} L_n^{(\alpha)}(x) t^n}_{n \rightarrow n+1} - \sum_{n=0}^{\infty} \frac{d}{dx} L_n^{(\alpha)}(x) t^{n+1} + \sum_{n=0}^{\infty} L_n^{(\alpha)}(x) t^n &= 0, \\
 \Rightarrow \frac{d}{dx} L_{n+1}^{(\alpha)}(x) - \frac{d}{dx} L_n^{(\alpha)}(x) + L_n^{(\alpha)}(x) &= 0,
 \end{aligned} \tag{12.76}$$

上式代入(12.75)有

$$\begin{aligned}
 &-L_n^{(\alpha+1)}(x) - \frac{d}{dx} L_n^{(\alpha)}(x) + L_n^{(\alpha)}(x) = 0, \\
 \Rightarrow \frac{d}{dx} L_n^{(\alpha)}(x) &= L_n^{(\alpha)}(x) - L_n^{(\alpha+1)}(x), \quad n = 0, 1, 2 \dots
 \end{aligned}$$

第二个偏导数

$$\begin{aligned}
 \frac{\partial F(x, t)}{\partial t} &= \left[ (\alpha+1)(1-t)^{-\alpha-2} + (1-t)^{-\alpha-1} \frac{-x(1-t)-xt}{(1-t)^2} \right] \exp\left(-x \frac{t}{1-t}\right) \\
 &= \left( \alpha+1 - \frac{x}{1-t} \right) (1-t)^{-\alpha-2} \exp\left(-x \frac{t}{1-t}\right), \\
 \Rightarrow (1-t)^2 \frac{\partial F(x, t)}{\partial t} + [x - (\alpha+1)(1-t)] F(x, t) &= 0, \\
 \Rightarrow (1-t)^2 \sum_{n=1}^{\infty} n L_n^{(\alpha)}(x) t^{(n-1)} + [x - (\alpha+1)(1-t)] \sum_{n=0}^{\infty} L_n^{(\alpha)}(x) t^n &= 0,
 \end{aligned}$$

将各项拆出

$$\begin{aligned}
 &\sum_{n=1}^{\infty} n L_n^{(\alpha)}(x) t^{n-1} - 2 \sum_{n=1}^{\infty} n L_n^{(\alpha)}(x) t^n + \sum_{n=1}^{\infty} n L_n^{(\alpha)}(x) t^{n+1} \\
 &+ x \sum_{n=0}^{\infty} L_n^{(\alpha)}(x) t^n + (\alpha+1) \sum_{n=0}^{\infty} L_n^{(\alpha)}(x) t^{n+1} - (\alpha+1) \sum_{n=0}^{\infty} L_n^{(\alpha)}(x) t^n = 0,
 \end{aligned}$$

按照  $t$  的幂次重新排列组合, 得三项递推关系式(12.74)。  $\square$

**Theorem 12.30** (拉盖尔多项式的二阶线性微分方程). 拉盖尔多项式  $L_n^{(\alpha)}(x)$  的二阶线性微分方程为

$$x \frac{d^2}{dx^2} L_n^{(\alpha)}(x) + (\alpha + 1 - x) \frac{d}{dx} L_n^{(\alpha)}(x) + n L_n^{(\alpha)}(x) = 0.$$

证明. 拉盖尔多项式的三项递推关系(12.74) $\Rightarrow$

$$x L_n^{(\alpha)}(x) + (n+1) [L_{n+1}^{(\alpha)}(x) - L_n^{(\alpha)}(x)] - (n+\alpha) \left[ \frac{d}{dx} L_n^{(\alpha)}(x) - \frac{d}{dx} L_{n-1}^{(\alpha)}(x) \right] = 0.$$

对  $x$  求导  $\Rightarrow$

$$L_n^{(\alpha)}(x) + x \frac{d}{dx} L_n^{(\alpha)}(x) + (n+1) \frac{d}{dx} L_{n+1}^{(\alpha)}(x) - (n+1) \frac{d}{dx} L_n^{(\alpha)}(x) - (n+\alpha) \frac{d}{dx} L_n^{(\alpha)}(x) + (n+\alpha) \frac{d}{dx} L_{n-1}^{(\alpha)}(x) = 0,$$

引入(12.76)替换上式中的  $\frac{d}{dx} L_{n+1}^{(\alpha)}(x)$  和  $\frac{d}{dx} L_n^{(\alpha)}(x)$

$$\begin{aligned} L_n^{(\alpha)}(x) + x \frac{d}{dx} L_n^{(\alpha)}(x) - (n+1) L_n^{(\alpha)}(x) + (n+\alpha) L_{n-1}^{(\alpha)}(x) &= 0, \\ x \frac{d}{dx} L_n^{(\alpha)}(x) &= n L_n^{(\alpha)}(x) - (n+\alpha) L_{n-1}^{(\alpha)}(x). \end{aligned} \quad (12.77)$$

上式继续对  $x$  求导  $\Rightarrow$

$$\begin{aligned} \frac{d}{dx} L_n^{(\alpha)}(x) + x \frac{d^2}{dx^2} L_n^{(\alpha)}(x) &= n \frac{d}{dx} L_n^{(\alpha)}(x) - (n+\alpha) \frac{d}{dx} L_{n-1}^{(\alpha)}(x) \\ &= (n+\alpha) \left( \frac{d}{dx} L_n^{(\alpha)}(x) - \frac{d}{dx} L_{n-1}^{(\alpha)}(x) \right) - \alpha \frac{d}{dx} L_n^{(\alpha)}(x) \\ &= -(n+\alpha) L_{n-1}^{(\alpha)}(x) - \alpha \frac{d}{dx} L_n^{(\alpha)}(x) \\ &= x \frac{d}{dx} L_n^{(\alpha)}(x) - n L_n^{(\alpha)}(x) - \alpha \frac{d}{dx} L_n^{(\alpha)}(x), \end{aligned} \quad (12.78)$$

$\hookrightarrow$

$$\begin{aligned} \frac{d^2}{dx^2} L_n^{(\alpha)}(x) + (1+\alpha-x) \frac{d}{dx} L_n^{(\alpha)}(x) + n L_n^{(\alpha)}(x) &= 0, \\ \Rightarrow x y''(x) + (1+\alpha-x)y'(x) + ny(x) &= 0, \quad n = 0, 1, 2 \dots \end{aligned}$$

□

### 12.C.3 雅各比多项式

**Theorem 12.31** (雅各比多项式的罗德里格斯公式). 在  $(-1, 1)$  区间内, 关于贝塔分布 (Beta distribution)  $w(x) = (1-x)^\alpha (1+x)^\beta$  的雅各比多项式 (Jacobi polynomial)  $P_n^{(\alpha, \beta)}(x)$ , 可由罗德里格斯公式予以定义

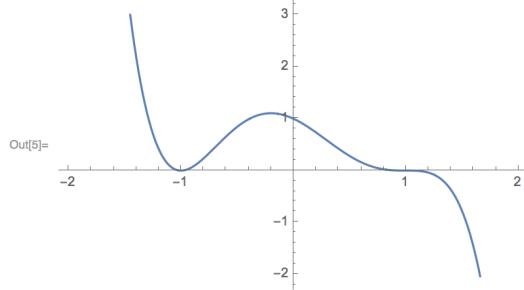
$$\begin{aligned} P_n^{(\alpha, \beta)}(x) &= \frac{(-1)^n}{2^n n!} \frac{1}{w(x)} D^n [w(x)(1-x^2)^n] \\ &= \frac{(-1)^n}{2^n n!} (1-x)^{-\alpha} (1+x)^{-\beta} D^n [(1-x)^{n+\alpha} (1+x)^{n+\beta}], \quad \alpha, \beta > -1, n = 0, 1, 2 \dots \end{aligned} \quad (12.79)$$

当  $(\alpha, \beta) = (3, 2)$  时贝塔分布  $w(x) = (1-x)^\alpha (1+x)^\beta$  在区间  $(-2, 2)$  内, 如图12.C.5所示。

**Theorem 12.32.** 雅各比多项式  $P_n^{(\alpha, \beta)}(x)$  是一个关于  $x$  的  $n$  次多项式。 $P_n^{(\alpha, \beta)}(x)$  是奇方程, 并且  $P_n^{(\alpha, \beta)}(1) = \binom{n+\alpha}{n}$ ,  $P_n^{(\alpha, \beta)}(-1) = (-1)^n \binom{n+\beta}{n}$ ,  $n = 0, 1, 2 \dots$

图 12.C.5: 贝塔分布

```
In[5]:= Plot[(1 - x)^3 (1 + x)^2, {x, -2, 2}]
```



$w(x) = (1 - x)^\alpha (1 + x)^\beta$  在  $(-2, 2)$  区间内的正态分布 ( $\alpha = 3, \beta = 2$ )。

证明. 根据莱布尼兹法则(12.60)我们有

$$\begin{aligned} & D^n [(1 - x)^{n+\alpha} (1 + x)^{n+\beta}] \\ &= \sum_{k=0}^n \binom{n}{k} D^k (1 - x)^{n+\alpha} D^{n-k} (1 + x)^{n+\beta} \\ &= \sum_{k=0}^n \binom{n}{k} (-1)^k (n + \alpha)(n + \alpha - 1)(n + \alpha - 2) \dots (n + \alpha - k + 1) (1 - x)^{n+\alpha-k} \\ &\quad \times (n + \beta)(n + \beta - 1)(n + \beta - 2) \dots (\beta + k + 1) (1 + x)^{\beta+k} \\ &= n! \sum_{k=0}^n (-1)^k \binom{n + \alpha}{k} \binom{n + \beta}{n - k} (1 - x)^{n+\alpha-k} (1 + x)^{\beta+k}, \quad n = 0, 1, 2 \dots \end{aligned}$$

代回(12.79) $\Rightarrow$

$$\begin{aligned} P_n^{(\alpha, \beta)}(x) &= \frac{(-1)^n}{2^n n!} (1 - x)^{-\alpha} (1 + x)^{-\beta} D^n [(1 - x)^{n+\alpha} (1 + x)^{n+\beta}] \\ &= \frac{(-1)^n}{2^n} \sum_{k=0}^n (-1)^k \binom{n + \alpha}{k} \binom{n + \beta}{n - k} (1 - x)^{n-k} (1 + x)^k, \quad n = 0, 1, 2 \dots \end{aligned} \tag{12.80}$$

这表明雅各比多项式  $P_n^{(\alpha, \beta)}(x)$  是一个关于  $x$  的  $n$  次多项式。

(12.79)  $\Rightarrow P_n^{(\alpha, \beta)}(x)$  的对称性 (略):

$$P_n^{(\alpha, \beta)}(-x) = (-1)^n P_n^{(\alpha, \beta)}(x), \quad n = 0, 1, 2 \dots \tag{12.81}$$

(12.79)  $\Rightarrow$

$$\begin{aligned} P_n^{(\alpha, \beta)}(1) &= \frac{(-1)^n}{2^n} \sum_{k=0}^n (-1)^k \frac{(n + \alpha)!}{(n + \alpha - k)! k!} \frac{(n + \beta)!}{(n + \beta + k)! (n - k)!} (1 + x)^k = \binom{n + \alpha}{n}, \\ P_n^{(\alpha, \beta)}(-1) &= (-1)^n \binom{n + \beta}{n}. \end{aligned}$$

□

**Theorem 12.33** (雅各比多项式的超几何方程). 雅各比多项式的  $P_n^{(\alpha, \beta)}(x)$  的超几何方程可表示为

$$P_n^{(\alpha, \beta)}(x) = \binom{n + \alpha}{n} {}_2F_1 \left( \begin{matrix} -n & n + \alpha + \beta + 1 \\ \alpha + 1 & \end{matrix}; \frac{1 - x}{2} \right), \quad n = 0, 1, 2 \dots \tag{12.82}$$

证明. 对于  $x \neq 1$  的情况, (12.80)⇒

$$P_n^{(\alpha, \beta)}(x) = \left(\frac{x-1}{2}\right)^n \sum_{k=0}^{\infty} \binom{n+\alpha}{n} \binom{n+\beta}{n-k} \underbrace{\left(\frac{x+1}{x-1}\right)^k}_{}, \quad n = 0, 1, 2 \dots$$

其中

$$\begin{aligned} \left(\frac{x+1}{x-1}\right)^k &= \left(1 + \frac{2}{x-1}\right)^k, \quad k = 0, 1, 2 \dots \\ &= \sum_{i=0}^k \binom{k}{i} \left(\frac{2}{x-1}\right)^i \\ &= \left(\frac{2}{x-1}\right)^n \sum_{i=0}^n \sum_{k=i}^n \binom{n+\alpha}{k} \binom{n+\beta}{n-k} \binom{k}{i} \left(\frac{2}{x-1}\right)^i \\ &= \left(\frac{2}{x-1}\right)^n \sum_{i=0}^n \sum_{k=0}^{n-i} \binom{n+\alpha}{i+k} \binom{n+\beta}{n-i-k} \binom{i+k}{k} \left(\frac{2}{x-1}\right)^i \\ &= \left(\frac{2}{x-1}\right)^n \sum_{i=0}^n \sum_{k=0}^n \binom{n+\alpha}{n-i+k} \binom{n+\beta}{i-k} \binom{n-i+k}{n-i} \left(\frac{2}{x-1}\right)^i \\ &= \sum_{i=0}^n \sum_{k=0}^n \binom{n+\alpha}{n-i+k} \binom{n+\beta}{i-k} \binom{n-i+k}{n-i} \left(\frac{2}{x-1}\right)^i \\ &= \sum_{i=0}^n \sum_{k=0}^n \frac{\Gamma(n+\alpha+1)}{(n-i+k)! \Gamma(i-k+\alpha+1)} \frac{\Gamma(n+\beta+1)}{(i-k)! \Gamma(n-i+k+\beta+1)} \frac{(n-i+k)!}{(n-i)! k!} \left(\frac{x-1}{2}\right)^i \\ &= \frac{\Gamma(n+\alpha+1) \Gamma(n+\beta+1)}{n!} \sum_{i=0}^n \underbrace{\frac{(-n)_i}{\Gamma(i+\alpha+1) \Gamma(n-i+\beta+1)}}_{\sum_{k=0}^n \frac{(-i)_k (-i-\alpha-1)_k}{(n-i+\beta+1)_k k!}} \left(\frac{2}{x-1}\right)^i \end{aligned} \tag{12.83}$$

其中, 首先根据朱世杰——范德蒙德求和公式我们有

$$\sum_{k=0}^n \frac{(-i)_k (-i-\alpha-1)_k}{(n-i+\beta+1)_k k!} = {}_2F_1\left(\begin{matrix} -i & -i-\alpha-1 \\ n-i+\beta+1 & \end{matrix}; 1\right) = \frac{(n+\alpha+\beta+1)_i}{(n-i+\beta+1)_i},$$

其次

$$\Gamma(n-i+\beta+1)(n-i+\beta+1)_i = \Gamma(n+\beta+1),$$

因此(12.83)进一步改写为

$$\begin{aligned} P_n^{(\alpha, \beta)}(x) &= \frac{\Gamma(n+\alpha+1)}{n!} \sum_{i=0}^n \frac{(-n)_i (n+\alpha+\beta+1)_i}{\Gamma(i+\alpha+1) i!} \left(\frac{1-x}{2}\right)^i \\ &= \frac{\Gamma(n+\alpha+1)}{\Gamma(\alpha+1) n!} \sum_{i=0}^n \frac{(-n)_i (n+\alpha+\beta+1)_i}{(\alpha+1)_i i!} \left(\frac{1-x}{2}\right)^i \\ &= \binom{n+\alpha}{n} {}_2F_1\left(\begin{matrix} -n & n+\alpha+\beta+1 \\ \alpha+1 & \end{matrix}; \frac{1-x}{2}\right), \quad n = 0, 1, 2 \dots \end{aligned} \tag{12.84}$$

此外由奇函数的对称性质(12.81)我们有

$$P_n^{(\alpha, \beta)}(-x) = (-1)^n \binom{n+\beta}{n} {}_2F_1\left(\begin{matrix} -n & n+\alpha+\beta+1 \\ \beta+1 & \end{matrix}; \frac{1+x}{2}\right), \quad n = 0, 1, 2 \dots$$

□

**Theorem 12.34** (雅各比多项式的首项系数). 雅各比多项式  $P_n^{(\alpha, \beta)}(x)$  的首项系数  $k_n$  为

$$k_n = \frac{n + \alpha + \beta + 1}{2^n n!}, \quad n = 0, 1, 2 \dots \quad (12.85)$$

证明. 由雅各比多项式的超几何方程(12.84)得

$$k_n = \binom{n + \alpha}{n} \frac{(-n)_n (n + \alpha + \beta + 1)_n}{(\alpha + 1)_n n!} \frac{(-1)^n}{2^n} = \frac{n + \alpha + \beta + 1}{2^n n!}, \quad n = 0, 1, 2 \dots$$

□

**Theorem 12.35** (雅各比多项式的正交条件). 雅各比多项式  $P_n^{(\alpha, \beta)}(x)$  满足如下正交关系

$$\int_{-1}^1 (1-x)^\alpha (1+x)^\beta P_m^{(\alpha, \beta)}(x) P_n^{(\alpha, \beta)}(x) dx = \frac{2^{\alpha+\beta+1} \Gamma(n + \alpha + 1) \Gamma(n + \beta + 1)}{(2n + \alpha + \beta + 1) \Gamma(n + \alpha + \beta + 1) n!} \delta_{mn}, \quad (12.86)$$

for  $\alpha > -1, \beta > -1, m, n \in \{0, 1, 2 \dots\}$

证明. 当  $m = n$  时, 由雅各比多项式的罗德里格斯公式(12.79)得

$$\begin{aligned} & \int_{-1}^1 (1-x)^\alpha (1+x)^\beta P_m^{(\alpha, \beta)}(x) P_n^{(\alpha, \beta)}(x) dx \\ &= \int_{-1}^1 (1-x)^\alpha (1+x)^\beta \left( P_n^{(\alpha, \beta)}(x) \right)^2 dx \\ &= \frac{(-1)^n}{2^n n!} \int_{-1}^1 P_n^{(\alpha, \beta)}(x) D^n [(1-x)^{n+\alpha} (1+x)^{1+\beta}] dx \\ &= \frac{(-1)^n}{2^n n!} \int_{-1}^1 D^n P_n^{(\alpha, \beta)}(x) (1-x)^{n+\alpha} (1+x)^{1+\beta} dx \\ &= \frac{(n + \alpha + \beta_1)_n}{2^n n!} \int_{-1}^1 (1-x)^{n+\alpha} (1+x)^{n+\beta} dx \\ &= \frac{\Gamma(2n + \alpha + \beta + 1)}{\Gamma(\alpha + \beta + n + 1) 2^{2n} n!} \underbrace{\int_{-1}^1 (1-x)^{n+\alpha} (1+x)^{n+\beta} dx}_{n = 0, 1, 2 \dots} \end{aligned}$$

设  $2t := 1 - x$  我们有

$$\begin{aligned} & \int_{-1}^1 (1-x)^{n+\alpha} (1+x)^{n+\beta} dx \\ &= \int_0^1 (2t)^{n+\alpha} (2t)^{n+\beta} dt \\ &= 2^{(2n+\alpha+\beta+1)} \int_0^1 t^{n+\alpha} (1-t)^{n+\beta} dt \\ &= 2^{(2n+\alpha+\beta+1)} \underbrace{B(n + \alpha + 1, n + \beta + 1)}_{\Gamma(n + \alpha + 1) \Gamma(n + \beta + 1)} \\ &= 2^{(2n+\alpha+\beta+1)} \frac{\Gamma(n + \alpha + 1) \Gamma(n + \beta + 1)}{\Gamma(2n + \alpha + \beta + 2)} \\ &= 2^{(2n+\alpha+\beta+1)} \frac{\Gamma(n + \alpha + 1) \Gamma(n + \beta + 1)}{(2n + \alpha + \beta + 1) \Gamma(2n + \alpha + \beta + 1)}, \end{aligned}$$

代回上式我们有

$$\begin{aligned}
 & \int_{-1}^1 (1-x)^\alpha (1+x)^\beta P_m^{(\alpha, \beta)}(x) P_n^{(\alpha, \beta)}(x) dx \\
 &= \frac{\Gamma(2n+\alpha+\beta+1)}{\Gamma(\alpha+\beta+n+1) 2^{2n} n!} \int_{-1}^1 (1-x)^{n+\alpha} (1+x)^{n+\beta} dx, \quad n = 0, 1, 2 \dots \\
 &= \frac{\Gamma(2n+\alpha+\beta+1)}{\Gamma(\alpha+\beta+n+1) 2^{2n} n!} \left[ 2^{(2n+\alpha+\beta+1)} \frac{\Gamma(n+\alpha+1) \Gamma(n+\beta+1)}{(2n+\alpha+\beta+1) \Gamma(2n+\alpha+\beta+1)} \right] \\
 &= \frac{2^{\alpha+\beta+1} \Gamma(n+\alpha+1) \Gamma(n+\beta+1)}{(2n+\alpha+\beta+1) \Gamma(n+\alpha+\beta+1) n!}.
 \end{aligned}$$

对于  $m < n$  的情况 (略)。  $\square$

**Theorem 12.36** (雅各比多项式的二阶线性微分方程). 雅各比多项式  $P_n^{(\alpha, \beta)}(x)$  的二阶线性微分方程形式为

$$\begin{aligned}
 & (1-x)^2 \frac{d^2}{dx^2} P_n^{(\alpha, \beta)}(x) + [\beta - \alpha - (\alpha + \beta + 2)x] \frac{d}{dx} P_n^{(\alpha, \beta)}(x) + n(n + \alpha + \beta + 1) P_n^{(\alpha, \beta)}(x) = 0, \\
 & \hookrightarrow (1-x)^2 y''(x) + [\beta - \alpha - (\alpha + \beta + 2)x] y'(x) + n(n + \alpha + \beta + 1)y(x) = 0.
 \end{aligned} \tag{12.87}$$

证明. 略。提示: 由雅各比多项式  $P_n^{(\alpha, \beta)}(x)$  的超几何方程(12.82)我们有

$$\begin{aligned}
 \frac{d}{dx} P_n^{(\alpha, \beta)}(x) &= \binom{n+\alpha}{n} \frac{(-n)(n+\alpha+\beta+1)}{(\alpha+1)} \left( -\frac{1}{2} \right) {}_2F_1 \left( \begin{matrix} -n+1 & n+\alpha+\beta+1 \\ \beta+1 & \end{matrix}; \frac{1-x}{2} \right) \\
 &= \frac{n+\alpha+\beta+1}{2} \binom{n+\alpha}{n-1} {}_2F_1 \left( \begin{matrix} -n+1 & n+\alpha+\beta+1 \\ \beta+1 & \end{matrix}; \frac{1-x}{2} \right) \\
 &= \frac{n+\alpha+\beta+1}{2} P_{n-1}^{(\alpha+1, \beta+1)}(x), \quad n = 1, 2, 3 \dots
 \end{aligned}$$

$\square$

**Theorem 12.37** (雅各比多项式的母方程). 雅各比多项式  $P_n^{(\alpha, \beta)}(x)$  的母方程为

$$\frac{2^{\alpha+\beta}}{R(1+R-t)^\alpha (1+R+t)^\beta} = \sum_{n=0}^{\infty} P_n^{(\alpha, \beta)}(x) t^n, \quad R := \sqrt{1-2x+t^2}. \tag{12.88}$$

**Theorem 12.38** (雅各比多项式的三项递推关系). 雅各比多项式  $P_n^{(\alpha, \beta)}(x)$  的三项递推关系可表示为

$$\begin{aligned}
 \mathcal{A} P_{n+1}^{\alpha, \beta}(x) &= \mathcal{B} P_n^{\alpha, \beta}(x) + \mathcal{C} P_{n-1}^{\alpha, \beta}(x), \\
 \mathcal{A} &= 2(n+1)(n+\alpha+\beta+1)(2n+\alpha+\beta), \\
 \mathcal{B} &= (2n+\alpha+\beta+1)(\alpha^2-\beta^2) + (2n+\alpha+\beta)(2n+\alpha+\beta+1)(2n+\alpha+\beta+2)x, \\
 \mathcal{C} &= -2(n+\alpha)(n+\beta)(2n+\alpha+\beta+2).
 \end{aligned} \tag{12.89}$$

证明. 略。可参考 (Shen et al., 2011, p.74)。  $\square$

#### 12.C.4 勒让德多项式

**Theorem 12.39** (勒让德多项式的罗德里格斯公式). 在  $(-1, 1)$  区间内, 关于  $w(x) = 1$  的均匀分布 (uniform distribution)  $w(x) = 1$  的勒让德多项式  $P_n(x)$ , 可由罗德里格斯公式予以定义

$$P_n(x) = \frac{(-1)^n}{2^n n!} \frac{1}{w(x)} D^n [w(x)(1-x^2)^n] = \frac{(-1)^n}{2^n n!} D^n [(1-x^2)^n], \quad n = 0, 1, 2 \dots \tag{12.90}$$

**Theorem 12.40** (勒让德多项式是雅各比多项式的特例; 勒让德多项式的超几何方程). 勒让德多项式(12.90)是雅各比多项式(12.79)的特例  $\alpha = \beta = 0$ :

$$P_n(x) = P_n^{(\alpha=0, \beta=0)}(x) = {}_2F_1\left(\begin{matrix} -n & n+1 \\ \alpha+1 & \end{matrix}; \frac{1-x}{2}\right), \quad n = 0, 1, 2 \dots \quad (12.91)$$

证明. 将  $\alpha = 0, \beta = 0$  代入(12.82)可得。  $\square$

**Theorem 12.41.** 勒让德多项式  $P_n(x)$  是一个关于  $x$  的  $n$  次多项式。 $P_n(x)$  是奇方程, 并且  $P_n(1) = 1, P_n(-1) = (-1)^n$ 。

证明. 由 Theorem 12.40 可得  $P_n(x) = P_n^{(\alpha=0, \beta=0)}(x)$ 。因此根据雅各比多项式的相关性质 (Theorem 12.19), 可证。  $\square$

**Theorem 12.42** (勒让德多项式的首项系数). 勒让德多项式  $P_n(x)$  的首项系数  $k_n$  为

$$k_n = \frac{(2n)!}{2^n (n!)^2}. \quad (12.92)$$

证明. 由勒让德多项式的超几何方程(12.91)得

$$k_n = \frac{(-n)_n (n+1)_n}{n! (-1)^n} \frac{(-1)^n}{2^n} = \frac{(2n)!}{2^n (n!)^2}.$$

$\square$

**Theorem 12.43** (勒让德多项式的正交条件). 勒让德多项式  $P_n(x)$  满足如下正交关系

$$\int_{-1}^1 P_m(x) P_n(x) dx = \frac{2}{2n+1} \delta_{mn}, \quad m, n \in \{0, 1, 2 \dots\} \quad (12.93)$$

证明. 由勒让德多项式的罗德里格斯公式(12.90)可得

$$\begin{aligned} \int_{-1}^1 P_m(x) P_n(x) dx &= \frac{(-1)^n}{2^n n!} \int_{-1}^1 P_m(x) D^n [(1-x^2)^n] dx \\ &= \frac{(-1)^n}{2^n n!} \int_{-1}^1 D^n [P_m(x)(1-x^2)^n] dx \end{aligned}$$

当  $m < 0$  时,  $\int_{-1}^1 P_m(x) P_n(x) dx = 0$ 。当  $m = n$  时,

$$\begin{aligned} \int_{-1}^1 P_m(x) P_n(x) dx &= \int_{-1}^1 D^n [P_n(x)(1-x^2)^n] dx \\ &= k_n n! \int_{-1}^1 (1-x^2)^n dx \\ &= \frac{(2n)!}{2^n (n!)^2} \int_{-1}^1 (1-x^2)^n dx. \end{aligned}$$

定义  $1 - x := 2t, n = 0, 1, 2 \dots \Rightarrow$

$$\begin{aligned} \int_{-1}^1 (1 - x^2)^n dx &= \int_{-1}^1 (1 - x)^n (1 + x)^n dx \\ &= \int_{-1}^1 (2t)^n (2 - 2t)^n 2 dx \\ &= 2^{2n+1} B(n+1, n+1) \\ &= 2^{2n+1} \frac{\Gamma(n+1)\Gamma(n+1)}{\Gamma(2n+2)} \\ &= \frac{2^{2n+1} (n!)^2}{(2n+1)!}, \end{aligned}$$

$\hookrightarrow$

$$\int_{-1}^1 [P_n(x)]^2 dx = \frac{(2n)!}{2^{2n} (n!)^2} \frac{2^{2n+1} (n!)^2}{(2n+1)!} = \frac{2}{2n+1}, \quad n = 0, 1, 2 \dots$$

□

**Theorem 12.44** (勒让德多项式的母方程). 勒让德多项式  $P_n(x)$  的母方程为

$$\sum_{n=0}^{\infty} P_n(x) t^n = (1 - 2xt + t^2)^{-\frac{1}{2}}. \quad (12.94)$$

证明. 由勒让德多项式的超几何方程(12.91)得

$$\begin{aligned} \sum_{n=0}^{\infty} P_n(x) t^n &= \sum_{n=0}^{\infty} {}_2F_1\left(\begin{matrix} -n & n+1 \\ 1 & \end{matrix}; \frac{1-x}{2}\right) t^n \\ &= \sum_{n=0}^{\infty} \sum_{k=0}^n \frac{(-n)_k (n+1)_k}{(1)_n k!} \left(\frac{1-x}{2}\right)^k t^n \\ &= \sum_{n=0}^{\infty} \sum_{k=n}^{\infty} \frac{(-n)_k (n+1)_k}{k! k!} \left(\frac{1-x}{2}\right)^k t^n \\ &= \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} \frac{(-n-k)_k (n+k+1)_k}{k! k!} \left(\frac{1-x}{2}\right)^k t^{n+k} \end{aligned}$$

□

**Theorem 12.45** (勒让德多项式的三项递推关系). 勒让德多项式  $P_n(x)$  的三项递推关系为

$$(n+1)P_{n+1}(x) - x(2n+1)P_n(x) + nP_{n-1}(x) = 0, \quad n = 1, 2, 3 \dots \quad (12.95)$$

证明. 定义  $F(x, t) := (1 - 2xt + t^2)^{-\frac{1}{2}}$ 。由勒让德多项式  $P_n(x)$  的母方程(12.94)我们有

$$\frac{\partial}{\partial t} F(x, t) = -\frac{1}{2} (1 - 2xt + t^2)^{-\frac{3}{2}} (-2x + 2t) = \frac{x-t}{(1 - 2xt + t^2)^{\frac{3}{2}}} = \sum_{n=1}^{\infty} n P_n(x) t^{n-1},$$

进而

$$\begin{aligned} (1 - 2xt + t^2) \frac{\partial}{\partial t} F(x, t) &= (x-t) F(x, t), \\ \hookrightarrow (1 - 2xt + t^2) \sum_{n=1}^{\infty} n P_n(x) t^{n-1} &= (x-t) (1 - 2xt + t^2)^{-\frac{1}{2}}. \end{aligned}$$

拆分上式

$$\underbrace{\sum_{n=1}^{\infty} nP_n(x)t^{n-1}}_{:=A} - 2x \underbrace{\sum_{n=1}^{\infty} nP_n(x)t^n}_{:=B} + \underbrace{\sum_{n=1}^{\infty} nP_n(x)t^{n+1}}_{:=C} = x \underbrace{\sum_{n=0}^{\infty} P_n(x)t^n}_{:=D} - \underbrace{\sum_{n=0}^{\infty} nP_n(x)t^{n+1}}_{:=E},$$

按照  $t$  的幂次重新整理

$$\begin{aligned} A &= \sum_{n=1}^{\infty} nP_n(x)t^{n-1}, \\ B + D &= -2x \sum_{n=1}^{\infty} nP_n(x)t^n - x \sum_{n=0}^{\infty} P_n(x)t^n = -x \sum_{n=0}^{\infty} (2n+1)P_n(x)t^n, \\ C + E &= \sum_{n=0}^{\infty} (n+1)P_n(x)t^{n+1}, \\ \hookrightarrow & \\ \sum_{n=1}^{\infty} nP_n(x)t^{n-1} - x \sum_{n=0}^{\infty} (2n+1)P_n(x)t^n + \sum_{n=0}^{\infty} (n+1)P_n(x)t^{n+1} &= 0, \end{aligned}$$

再次整理可得(12.95)。  $\square$

### 12.C.5 切比雪夫多项式

更多数学上的证明, 可参考Boyd (2001); Fornberg (1996); Mason and Handscomb (2003); Shen et al. (2011)。

在  $[-1, 1]$  区间内, 关于  $w(x) = (1 - x^2)^{-\frac{1}{2}}$  的第一类切比雪夫多项式 (the first kind Chebyshev polynomial)  $T_n(x)$  定义为

$$T_n(x) = \cos(n\theta), \quad x = \cos \theta, \quad n = 0, 1, 2 \dots \quad (12.96)$$

在  $[-1, 1]$  区间内, 关于  $w(x) = (1 - x^2)^{\frac{1}{2}}$  的第二类切比雪夫多项式 (the second kind Chebyshev polynomial)  $U_n(x)$  定义为

$$U_n(x) = \frac{\sin(n+1)\theta}{\sin \theta}, \quad x = \cos \theta, \quad n = 0, 1, 2 \dots \quad (12.97)$$

**Theorem 12.46.** 第一类切比雪夫多项式  $T_n(x)$  的首项系数为

$$k_n = 2^{n-1}. \quad (12.98)$$

证明.  $T_0(x) = 1, T_1(x) = x$  代入三项递推关系(12.103)有

$$\begin{aligned} T_0(x) &= 1, \\ T_1(x) &= x, \\ T_2(x) &= 2x^2 - 1, \\ T_3(x) &= 4x^3 - 3x, \\ T_4(x) &= 8x^4 - 8x^2 + 1, \\ T_5(x) &= 16x^5 - 20x^3 + 5x, \\ &\vdots \end{aligned}$$

可见  $k_n = 2^{n-1}$ .  $\square$

**Theorem 12.47** (切比雪夫多项式的罗德里格斯公式). 第一类切比雪夫多项式  $T_n(x)$  的罗德里格斯公式 定义为

$$T_n(x) = \frac{(-1)^n 2^n n!}{(2n)!} (1-x^2)^{-\frac{1}{2}} D^n (1-x^2)^{\frac{n-1}{2}} \quad (12.99)$$

第二类切比雪夫多项式  $U_n(x)$  的罗德里格斯公式

$$U_n(x) = \frac{(-1)^n (n_1)! 2^n}{(2n+1)!} (1-x^2)^{\frac{1}{2}} D^n (1-x^2)^{\frac{n+1}{2}} \quad (12.100)$$

**Theorem 12.48** (切比雪夫多项式的正交条件). 第一类、第二类切比雪夫多项式  $T_n(x), U_n(x)$  的正交条件分别为 (*Gradshteyn and Ryzhik, 2014, Entry 7.343.1, pp.807-808*)<sup>25</sup>

$$\int_{-1}^1 (1-x^2)^{-\frac{1}{2}} T_m(x) T_n(x) dx = \int_0^\pi \cos(m\theta) \cos(n\theta) d\theta = \begin{cases} 0 & m \neq n, \\ \frac{\pi}{2} & m = n = 0, \\ \pi & m = n \neq 0. \end{cases} \quad (12.101)$$

$$\int_{-1}^1 (1-x^2)^{-\frac{1}{2}} U_m(x) U_n(x) dx = \int_0^\pi \sin(m+1)\theta \sin(n+1)\theta d\theta = \begin{cases} 0 & m \neq n, \\ \frac{\pi}{2} & m = n = 0, \\ \pi & m = n \neq 0. \end{cases} \quad (12.102)$$

证明. 略。 □

**Theorem 12.49** (切比雪夫多项式的三项递推关系). 第一类、第二类切比雪夫多项式  $T_n(x), U_n(x)$  的三项递推关系分别为

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), \quad n = 1, 2, 3 \dots \quad (12.103)$$

$$U_{n+1}(x) = 2xU_n(x) - U_{n-1}(x), \quad n = 1, 2, 3 \dots \quad (12.104)$$

证明. (12.96)  $\Rightarrow$

$$T_{n+1}(x) + T_{n-1}(x) = \cos(n+1)\theta + \cos(n-1)\theta = 2\cos\theta \cos(n\theta) = 2xT_n(x).$$

(12.97)  $\Rightarrow$

$$U_{n+1}(x) + U_{n-1}(x) = \frac{\sin(n+2)\theta}{\sin\theta} \frac{\sin n\theta}{\sin\theta} = 2 \frac{\cos\theta \sin(n+1)\theta}{\sin\theta} = 2xU_n(x).$$

□

**Theorem 12.50** (第一类、第二类切比雪夫多项式的关系). 第一类、第二类切比雪夫多项式的关系为

$$\begin{cases} T_0(x) = U_0(x) = 1, \\ T_1(x) = x, \quad U_1(x) = 2x, \\ T_n(x) = U_n(x) - xU_{n-1}(x), \quad n = 1, 2, 3 \dots \end{cases} \quad (12.105)$$

证明.  $n = 0$  时, 可由定义式求得。 $n \leq 1$  时

$$U_n(x) - xU_{n-1}(x) = \frac{\sin(n+1)\theta}{\sin\theta} - \frac{\cos\theta \sin n\theta}{\sin\theta} = \frac{\sin\theta \cos n\theta}{\sin\theta} = \cos n\theta = T_n(x).$$

□

<sup>25</sup>Gradshteyn and Ryzhik (2014) 的补充材料可参考如Moll (2015, 2016)。

**Theorem 12.51** (第一类切比雪夫多项式的母方程). 第一类切比雪夫多项式  $T_n(x)$  的母方程为

$$\sum_{n=0}^{\infty} T_n(x)t^n = \frac{1-x}{1-2xt+t^2}, \quad |t| < 1. \quad (12.106)$$

证明. 将第一类切比雪夫多项式的三项递推关系(12.103)两侧分别乘以  $t^{n+1}$ , 并沿着  $n = 1, 2, 3 \dots$  求和

$$2x \sum_{n=1}^{\infty} T_n(x)t^{n+1} - \sum_{n=1}^{\infty} T_{n-1}(x)t^{n+1} = \sum_{n=1}^{\infty} T_{n+1}(x)t^{n+1}.$$

定义  $F(x, t) := \sum_{n=0}^{\infty} T_n(x)t^n$ ,  $|t| < 1$ , 则上式变为

$$\begin{aligned} [F(x, t) - T_1(x) - T_0(x)] &= 2xt[F(x, t) - T_0(x)] - t^2 F(x, t), \\ \hookrightarrow (1-2xt+t^2)F(x, t) &= T_0(x) + T_1(x)t - 2xtT_0(x) = 1 - xt, \\ \hookrightarrow F(x, t) &= \sum_{n=0}^{\infty} T_n(x)t^n = \frac{1-x}{1-2xt+t^2}, \quad |t| < 1. \end{aligned}$$

□

**Theorem 12.52** (第二类切比雪夫多项式的母方程). 第一类切比雪夫多项式  $U_n(x)$  的母方程为

$$\sum_{n=0}^{\infty} U_n(x)t^n = \frac{1}{1-2xt+t^2}, \quad |t| < 1. \quad (12.107)$$

将第二类切比雪夫多项式的三项递推关系(12.104)两侧分别乘以  $t^{n+1}$ , 并沿着  $n = 1, 2, 3 \dots$  求和

$$2x \sum_{n=1}^{\infty} U_n(x)t^{n+1} - \sum_{n=1}^{\infty} U_{n-1}(x)t^{n+1} = \sum_{n=1}^{\infty} U_{n+1}(x)t^{n+1}.$$

定义  $G(x, t) := \sum_{n=0}^{\infty} U_n(x)t^n$ ,  $|t| < 1$ , 则上式变为

$$\begin{aligned} [G(x, t) - U_1(x) - U_0(x)] &= 2xt[G(x, t) - U_0(x)] - t^2 F(x, t), \\ \hookrightarrow (1-2xt+t^2)G(x, t) &= U_0(x) + U_1(x)t - 2xtU_0(x) = 1, \\ \hookrightarrow G(x, t) &= \sum_{n=0}^{\infty} U_n(x)t^n = \frac{1}{1-2xt+t^2}, \quad |t| < 1. \end{aligned}$$

**Theorem 12.53** (第一类和第二类切比雪夫多项式的转换). 第一类切比雪夫多项式  $T_n(x)$  和第二类切比雪夫多项式  $U_n(x)$  的转换, 满足

$$U_n(x) = \sum_{k=0}^n T_k(x)x^{n-k}, \quad |t| < 1, \quad n = 0, 1, 2 \dots \quad (12.108)$$

证明. 对于  $|t| < 1$  我们有

$$\begin{aligned} & \sum_{n=0}^{\infty} \left[ \sum_{k=0}^{\infty} T_k(x) x^{n-k} \right] t^n \\ &= \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} T_k(x) x^{n-k} t^n \\ &= \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} T_k(x) x^n t^{n+k} \\ &= \sum_{k=0}^{\infty} T_k(x) t^k \sum_{n=0}^{\infty} (xt)^n \\ &= \frac{1-xt}{1-2xt+t^2} \frac{1}{1-xt} \\ &= \frac{1}{1-2xt+t^2} \\ &= \sum_{n=0}^{\infty} U_n(x) t^n. \end{aligned}$$

去掉等式两侧的求和符号, 证毕。  $\square$

**Theorem 12.54** (勒让德多项式和第二类切比雪夫多项式的转换). 勒让德多项式  $P_n(x)$  和第二类切比雪夫多项式  $U_n(x)$  的转换, 满足

$$U_n(x) = \sum_{k=0}^{\infty} P_k(x) P_{n-k}(x), \quad |t| < 1, \quad n = 0, 1, 2 \dots \quad (12.109)$$

证明. 对于  $|t| < 1$  我们有

$$\begin{aligned} & \sum_{n=0}^{\infty} \left[ \sum_{k=0}^{\infty} P_k(x) P_{n-k}(x) \right] t^n \\ &= \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} P_k(x) P_{n-k}(x) t^n \\ &= \sum_{k=0}^{\infty} P_k(x) t^k \sum_{n=0}^{\infty} P_n(x) t^n \\ &= (1-2xt-t^2)^{-\frac{1}{2}} (1-2xt-t^2)^{-\frac{1}{2}} \\ &= \sum_{n=0}^{\infty} U_n(x) t^n. \end{aligned}$$

去掉等式两侧的求和符号, 证毕。  $\square$

**Theorem 12.55** (第一类切比雪夫多项式与次的关系). 1 个  $n$  次第一类切比雪夫多项式  $T_n(x)$  满足关系

$$\begin{aligned} \int_{-1}^1 T_n(x) dx &= \frac{(-1)^{n-1} - 1}{(n-1)(n+1)}, \quad n \geq 2, \\ \int_{-\infty}^{\infty} T_n(x) dx &= - \int_{-\infty}^{\infty} \cos(n\theta) \sin \theta d\theta. \end{aligned} \quad (12.110)$$

证明. 略。  $\square$

**Theorem 12.56** (切比雪夫插值定理). 切比雪夫插值 (*Chebyshev interpolation*)。

证明。假定在区间  $[a, b]$  内有 1 个方程  $f(x)$ 。则关于结点 (nodes)  $\{x_0, x_1, \dots, x_n\} \in [a, b], x_i \neq x_j, i \neq j$  的拉格朗日插值 (Lagrange interpolation) 可以定义为有且只有一个  $\leq n$  次的多项式  $P_n(x)$ , 满足  $P(x_i) = f(x_i), i = 0, 1, \dots, n$ 。

如果在区间  $[a, b]$  内,  $f^{(n)}$  连续并且  $f^{(n+1)}$  存在, 那么拉格朗日插值的误差 (interpolation error) 可以表示为

$$f(x) - p(x) = \frac{f^{(n+1)}(\zeta)}{(n+1)!} q(x), \quad \zeta \in [a, b], \quad (12.111)$$

其中  $q(x)$  定义为

$$q(x) = \prod_{i=0}^n (x - x_i) = (x - x_0)(x - x_1) \dots (x - x_n).$$

我们的研究目标是, 选取合适的多项式  $P(x)$  作为原方程  $f(x)$  的近似, 近似的判定标准是, 误差(12.111)越小, 近似越精确, 或者换句话说,  $P_n(x)$  收敛到  $f(x)$ 。这可以分为两个问题来描述。

第一个问题, 通常来说, 随着  $n \rightarrow \infty$ ,  $f^{(n+1)}$  可能较大, 我们很难保证(12.111)中的插值误差足够小。

第二个问题, 关于  $n + 1$  个结点  $\{x_0, x_1, \dots, x_n\}$  的选取。最直观的方案一是等距法 (equidistance), 即在  $[a, b]$  区间内按照均等距离选取这  $n + 1$  个节点。然而均等距离选取的结点效果并不理想:  $x$  值越接近区间中值  $\frac{b-a}{2}$ ,  $|q(x)|$  越小; 反之  $x$  越接近区间两端,  $|q(x)|$  越大。例如, 设  $x_0 = a, x_n = b, x_k - x_{k-1} = h, k = 1, 2, \dots, n$ , 那么我们有

$$|q(x)| \leq \frac{h}{2} \frac{h}{2} 2h \dots nh = \frac{h^{n+1} n!}{4}, \quad x \in [a, b].$$

等距法对于降低插值误差(12.111)的目标来说, 并不理想。

于是我们提出方案二, 利用 (切比雪夫) 正交多项式, 致力于追求

$$\min \left\{ \max_{x \in [a, b]} |q(x)| \right\}.$$

将  $q(x)$  视作一个在  $[a, b]$  区间内的首一正交多项式, 满足关系

$$\max q(x) = \prod_{i=0}^n (x - x_i) = (x - x_0)(x - x_1) \dots (x - x_n) = \frac{T_{n+1}(x)}{2^n},$$

其中 RHS 分母的  $2^n$  是  $n + 1$  次切比雪夫多项式  $T_{n+1}(x)$  的首项系数,  $\frac{T_{n+1}(x)}{2^n}$  由此变为首一多项式。当  $a = -1, b = 1$  时, 根据上式我们有

$$\max_{x \in [-1, 1]} q(x) = \frac{T_{n+1}(x)}{2^n} \leq \frac{1}{2^n},$$

最后一个不等式是根据切比雪夫多项式的定义, 在  $[-1, 1]$  之内  $T_n(x)$  的极值为  $T_n(-1) = -1, T_n(1) = 1$ 。

首一正交切比雪夫多项式  $\frac{T_{n+1}(x)}{2^n} = \prod_{i=0}^n (x - x_i)$  的根为

$$x_k = \cos \left( \frac{2k+1}{2(n+1)} \pi \right), k = 0, 1, \dots, n. \quad (12.112)$$

因此(12.111)改写为

$$\begin{aligned} |f(x) - p(x)| &= \left| q(x) \frac{f^{(n+1)}(\zeta)}{(n+1)!} \right| \\ &\leq \min \left\{ \max_{x \in [-1, 1]} |q(x)| \right\} \left| \frac{f^{(n+1)}(\zeta)}{(n+1)!} \right| \\ &= \left| \frac{T_{n+1}(x)}{2^n} \frac{f^{(n+1)}(\zeta)}{(n+1)!} \right| \\ &= \frac{1}{2^n (n+1)!} \sup_{\{\zeta \in [-1, 1]\}} |f^{(n+1)}(\zeta)|. \end{aligned} \quad (12.113)$$

当  $a \neq -1, b \neq 1$  时, (12.112)  $\Rightarrow$

$$x_k = \frac{b+a}{2} + \frac{b-a}{2} \cos\left(\frac{2k-1}{2(n+1)}\pi\right), k = 1, 2, \dots, n+1. \quad (12.114)$$

则(12.113)  $\Rightarrow$

$$|f(x) - p(x)| \leq \left(\frac{b-a}{2}\right)^{n+1} \frac{1}{2^n(n+1)!} \sup_{\{\zeta \in [-1,1]\}} |f^{(n+1)}(\zeta)|. \quad (12.115)$$

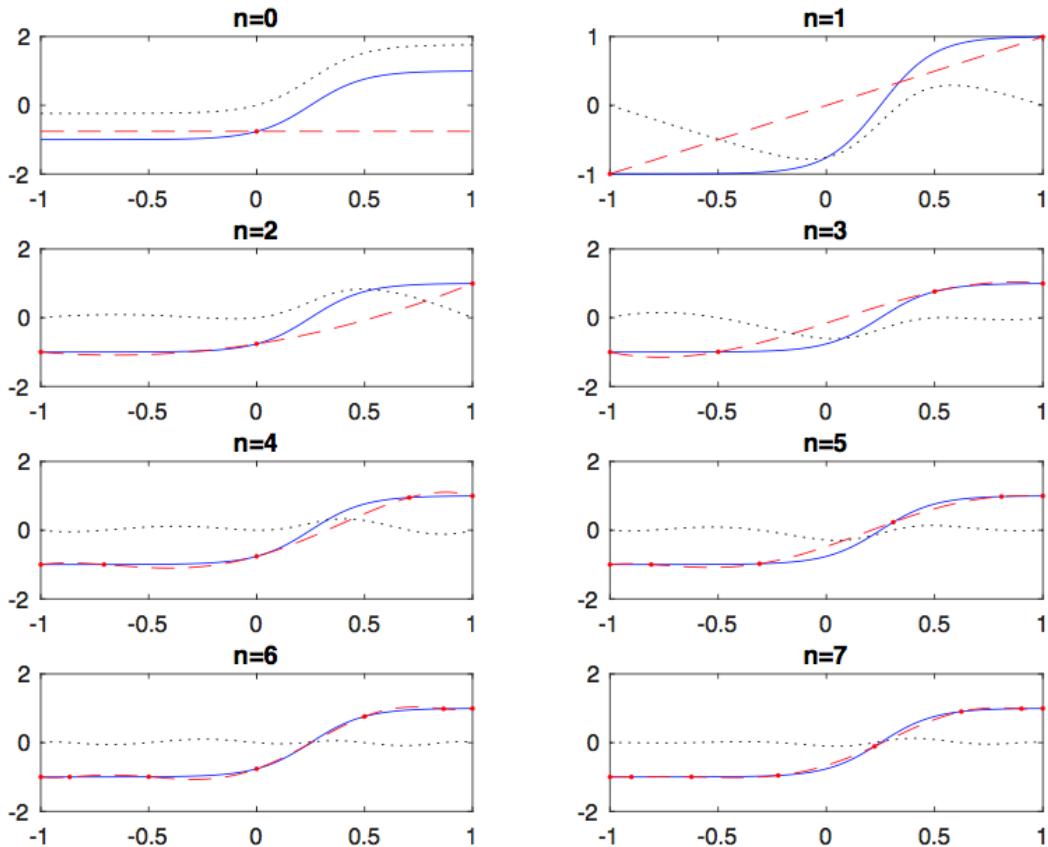
进而我们有切比雪夫插值定理 (Chebyshev interpolation theorem)

$$\lim_{n \rightarrow \infty} ()^2 \quad (12.116)$$

□

例如, 假定  $f(x) = \tanh(4x - 1)$ , 我们采用切比雪夫多项式作基方程, 用  $p_n(x)$  作以近似, 见图12.C.6, Matlab 代码如下

图 12.C.6: 切比雪夫插值近似



```
x = chebfun('x');
f = tanh(4 * x - 1);
n0=0;n1=1;n2=2;n3=3;n4=4;n5=5;n6=6;n7=7;
pn0 = chebfun(f,n0+1);
```

```

pn1 = chebfun(f, n1+1);
pn2 = chebfun(f, n2+1);
pn3 = chebfun(f, n3+1);
pn4 = chebfun(f, n4+1);
pn5 = chebfun(f, n5+1);
pn6 = chebfun(f, n6+1);
pn7 = chebfun(f, n7+1);

figure(1);
subplot(4,2,1),plot(f, '-b'), hold on, plot(pn0, '.--r'), hold on, plot(f-
pn0, ':k'), title('n=0'),
subplot(4,2,2),plot(f, '-b'), hold on, plot(pn1, '.--r'), hold on, plot(f-
pn1, ':k'), title('n=1'),
subplot(4,2,3),plot(f, '-b'), hold on, plot(pn2, '.--r'), hold on, plot(f-
pn2, ':k'), title('n=2'),
subplot(4,2,4),plot(f, '-b'), hold on, plot(pn3, '.--r'), hold on, plot(f-
pn3, ':k'), title('n=3'),
subplot(4,2,5),plot(f, '-b'), hold on, plot(pn4, '.--r'), hold on, plot(f-
pn4, ':k'), title('n=4'),
subplot(4,2,6),plot(f, '-b'), hold on, plot(pn5, '.--r'), hold on, plot(f-
pn5, ':k'), title('n=5'),
subplot(4,2,7),plot(f, '-b'), hold on, plot(pn6, '.--r'), hold on, plot(f-
pn6, ':k'), title('n=6'),
subplot(4,2,8),plot(f, '-b'), hold on, plot(pn7, '.--r'), hold on, plot(f-
pn7, ':k'), title('n=7'),

```

**Theorem 12.57** (切比雪夫截断定理). 切比雪夫截断定理 (*Chebyshev truncation theorem*)是指, 我们用  $j$  次切比雪夫多项式来近似未知的原方程系统  $f(x)$ , 近似的误差小于等于一系列未被纳入考虑的更高次切比雪夫多项式系数之和, 换句话说, 定义  $d^j(\cdot|\theta) = \sum_{i=0}^j \theta_i \psi_i(x)$ , 则截断误差 (*truncation error*)

$$d(x) - d^j(x|\theta) \leq \sum_{i=j+1}^{\infty} |\theta_i|, \quad \forall x \in [-1, 1], \quad \forall j. \quad (12.117)$$

证明. 略。 □

**Theorem 12.58.** 第  $n$  次一类切比雪夫多项式  $T_n(x)$  有  $n$  个根

$$x_i = \cos\left(\frac{2i-1}{2n}\pi\right), \quad i = 1, 2, \dots, n, \quad (12.118)$$

并且进而可将  $T_n(x)$  写为

$$T_n(x) = k_n \prod_{i=0}^n (x - x_i) = 2^{n-1} \prod_{i=0}^n (x - x_i), \quad i = 1, 2, \dots, n. \quad (12.119)$$

证明. cf. (Boyd, 2001, p.49). □

## 12.D AR(1) 过程的离散方法

当未来存在不确定性时，求解经济个体行为最大化的问题便涉及条件期望的计算。以消费——储蓄决策问题为例，可表示为如下贝尔曼方程 (Bellman equation)

$$\begin{aligned} V(a, \lambda) &= \max_{\{c, a'\}} \{u(c) + \beta E[V(a', \lambda')|\lambda]\}, \quad \text{s.t.} \\ a(1+r) + \omega \exp(\lambda) &= a' + c, \\ \lambda' &= (1-\rho)\mu_\lambda + \rho\lambda + \varepsilon, \quad \varepsilon \sim N(0, \sigma_\varepsilon^2), \\ c \geq 0, a' > 0. \end{aligned} \tag{12.120}$$

其中状态矩  $\lambda$  的无条件均值和方差为  $\lambda \sim N(\mu_\lambda, \sigma_\lambda^2)$ 。在给定  $\lambda$  的情况下， $\lambda'$  是一个 AR(1) 过程，满足条件均值和方差  $\lambda' \sim N((1-\rho)\mu_\lambda + \rho\lambda, \sigma_\varepsilon^2)$ 。 $\sigma_\varepsilon$  和  $\sigma_\lambda$  的关系满足

$$\sigma_\lambda = \frac{\sigma_\varepsilon}{\sqrt{1-\rho^2}}.$$

在此例中，涉及条件期望的计算目标是由  $f(\lambda'|\lambda)$  加权后的价值方程的积分值

$$E[V(a', \lambda')|\lambda] = \int_{-\infty}^{\infty} V(a', \lambda') f(\lambda'|\lambda) d\lambda', \tag{12.121}$$

其中  $f(\lambda'|\lambda)$  指给定  $\lambda$  的情况下， $\lambda'$  的条件密度 (conditional density)。

上式是一个与当前期状态有关的无限维度问题，需要借助数值近似算法。一个可行的近似方案，思路为：将状态空间中原本是连续的状态，离散化变为有限个点，对这些点对应的条件期望做近似。换句话说，就是将关于连续的  $\lambda$  的马尔科夫链，转变为离散的有限马尔科夫链，我们将新的马尔科夫链用  $\tilde{\lambda}$  来予以区分。有限个数的  $\tilde{\lambda}$  取值可以来自  $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$  集合，对应转移矩阵  $P$ ，含有转移概率  $p_{i,j}$ ，满足

$$p_{i,j} = \text{Prob}(\tilde{\lambda}' = \lambda_j | \tilde{\lambda} = \lambda_i), \quad i, j = 1, 2, \dots, n.$$

通过这种方式，原本是含有条件期望的积分计算问题(12.121)，被转换为离散化的求和问题，可以借助计算机，用某些特定的数值算法求得。下面分别介绍三种常见算法。

### 12.D.1 Tauchen (1986) 法

Tauchen (1986) 的方法可表示如下：

首先是  $\lambda$  的选取。在  $\Lambda$  中选取均匀分布的  $n$  个点  $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ ，其中  $\lambda_1$  和  $\lambda_n$  分别对应上下边界，值等于无条件均值  $\mu_\lambda$  加上或减去无条件标准差  $\sigma_\lambda$  的  $m$  倍数

$$\begin{aligned} \lambda_1 &= \mu_\lambda - m\sigma_\lambda, \\ \lambda_2 &= \mu_\lambda + m\sigma_\lambda, \end{aligned}$$

经验研究中， $m$  的取值通常在 2 到 3 之间。

随后是转移矩阵  $P$  的选取。设

$$\omega = \lambda_j - \lambda_{j-1},$$

进而

$$p_{i,j} = \begin{cases} \text{Prob}\left[(1-\rho)\mu_\lambda + \rho\lambda_i + \varepsilon \leq \lambda_{\{j=1\}} + \frac{\omega}{2}\right] & \text{如果 } j = 1, \\ \text{Prob}\left[\lambda_j - \frac{\omega}{2} \leq (1-\rho)\mu_\lambda + \rho\lambda_i + \varepsilon \leq \lambda_j + \frac{\omega}{2}\right] & \text{如果 } j = 2, 3, \dots, n-1, \\ 1 - \text{Prob}\left[\lambda_{\{j=n\}} + \frac{\omega}{2} \leq (1-\rho)\mu_\lambda + \rho\lambda_i + \varepsilon\right] & \text{如果 } j = n. \end{cases}$$

上式进一步整理为

$$p_{i,j} = \begin{cases} \Phi\left(\frac{\lambda_1 + \frac{\omega}{2} - (1-\rho)\mu_\lambda - \rho\lambda_i}{\sigma_\varepsilon}\right) & \text{如果 } j = 1, \\ \Phi\left(\frac{\lambda_j + \frac{\omega}{2} - (1-\rho)\mu_\lambda - \rho\lambda}{\sigma_\varepsilon}\right) - \Phi\left(\frac{\lambda_{j-1} + \frac{\omega}{2} - (1-\rho)\mu_\lambda - \rho\lambda}{\sigma_\varepsilon}\right) & \text{如果 } j = 2, 3, \dots, n-1, \\ 1 - \Phi\left(\frac{\lambda_n - \frac{\omega}{2} - (1-\rho)\mu_\lambda - \rho\lambda}{\sigma_\varepsilon}\right) & \text{如果 } j = n, \end{cases}$$

其中  $\Phi(\cdot)$  表示标准正态累积分布函数 (standard normal cumulative distribution function)。

Tauchen (1986) 法, 将  $\tilde{\lambda}'|\tilde{\lambda} = \lambda_i$  作为对  $\lambda'|\lambda = \lambda_i$  的条件概率的离散近似。 $n$  越大, 对节点的选取越密, 近似的效果越精确。Tauchen (1986) 指出当  $n = 9$  时所获得的近似精度基本已能满足经验研究的一般需要<sup>26</sup>。

举例说明。对于一个随机 AR(1) 连续过程

$$z_t = 0.95z_{t-1} + \varepsilon_t$$

以及正态分布  $N(0, 0.007^2)$  而言, 假定我们想要对它做离散化近似, 对应  $n = 9$  个点,  $m = 3$ 。

我们可以根据 Tauchen (1986) 的思路编写代码, 或者使用现有程序。以 Matlab 为例, Iskander Karibzhanov 在他的个人主页<sup>27</sup>提供了名为 tauchen.m 的程序。Matlab 代码如下

```
>> [y,P,d]=tauchen(5,0,0.95,0.007,3)

y =
-0.0673    -0.0336         0    0.0336    0.0673
```

```
P =
0.9727    0.0041    0.0000    0.0000    0.0000
0.0273    0.9806    0.0082    0.0000    0.0000
0.0000    0.0153    0.9837    0.0153    0.0000
0    0.0000    0.0082    0.9806    0.0273
0        0    0.0000    0.0041    0.9727
```

```
d =
0.0361    0.2392    0.4494    0.2392    0.0361
```

即, 我们得到了

$$z_t \in \{-0.0673, -0.0336, 0, 0.0336, 0.0673\}$$

<sup>26</sup>然而随着外生过程的持续性越来越强 (high persistence) 即  $\rho \rightarrow 1$ , 近似的精度就越来越低, 这就需要我们提高  $n$  的数量以弥补精度的下降, 或是采取更适合离散化高持续性过程的算法, 如第12.D.3节的Rouwenhorst (1995) 法。

<sup>27</sup><http://karibzhanov.com>。

对应转移矩阵  $P$

$$P_{z,z'} = \begin{pmatrix} 0.9727 & 0.0041 & 0 & 0 & 0 \\ 0.0273 & 0.9806 & 0.0082 & 0 & 0 \\ 0 & 0.0153 & 0.9837 & 0.0153 & 0 \\ 0 & 0 & 0.0082 & 0.9806 & 0.0273 \\ 0 & 0 & 0 & 0.0041 & 0.9727 \end{pmatrix}$$

需要注意的是  $P_{z,z'}$  对角元素接近于 1，这说明随机过程的当期状态与上期状态高度相关（对应  $\rho = 0.95$ ）。Tauchen (1986) 增加  $n$  的值到 9 后，相关度仍然比较高（见下），这需要我们探讨其他数值算法，见第12.D.3节。

```
>> [y,P,d]=tauchen(9,0,0.95,0.007,3)

y =
-0.0673 -0.0504 -0.0336 -0.0168 0 0.0168 0.0336 0.0504 0.0673

P =
0.7644 0.0592 0.0001 0 0 0 0 0
0.2347 0.7405 0.0747 0.0001 0 0 0 0
0.0009 0.1997 0.7569 0.0931 0.0002 0 0 0
0 0.0006 0.1679 0.7669 0.1147 0.0002 0 0
0 0 0.0004 0.1396 0.7702 0.1396 0.0004 0 0
0 0 0 0.0002 0.1147 0.7669 0.1679 0.0006 0
0 0 0 0 0.0002 0.0931 0.7569 0.1997 0.0009
0 0 0 0 0 0.0001 0.0747 0.7405 0.2347
0 0 0 0 0 0 0.0001 0.0592 0.7644

d =
0.0108 0.0428 0.1144 0.2064 0.2513 0.2064 0.1144 0.0428 0.0108
```

即，我们得到了

$$z_t \in \{-0.0673, -0.0504, -0.0336, -0.0168, 0, 0.0168, 0.0336, 0.0504, 0.0673\},$$

对应转移矩阵  $P$

$$P_{z,z'} = \begin{pmatrix} 0.7644 & 0.0592 & 0.0001 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.2347 & 0.7405 & 0.0747 & 0.0001 & 0 & 0 & 0 & 0 & 0 \\ 0.0009 & 0.1997 & 0.7569 & 0.0931 & 0.0002 & 0 & 0 & 0 & 0 \\ 0 & 0.0006 & 0.1679 & 0.7669 & 0.1147 & 0.0002 & 0 & 0 & 0 \\ 0 & 0 & 0.0004 & 0.1396 & 0.7702 & 0.1396 & 0.0004 & 0 & 0 \\ 0 & 0 & 0 & 0.0002 & 0.1147 & 0.7669 & 0.1679 & 0.0006 & 0 \\ 0 & 0 & 0 & 0 & 0.0002 & 0.0931 & 0.7569 & 0.1997 & 0.0009 \\ 0 & 0 & 0 & 0 & 0 & 0.0001 & 0.0747 & 0.7405 & 0.2347 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.0001 & 0.0592 & 0.7644 \end{pmatrix}$$

## 12.D.2 Tauchen and Hussey (1991) 法

同样对矩  $\lambda$  而言, Tauchen and Hussey (1991) 使用高斯——埃米特求积 (Gauss-Hermite quadrature) 去近似价值方程的条件期望积分, (12.121) 改写为

$$E[V(a', \lambda')|\lambda] \approx \frac{1}{\sqrt{\pi}} \sum_{j=1}^n \omega_j V(a', \sqrt{2}\sigma_\lambda \hat{\lambda}'_j + \mu_\lambda), \quad j = 1, 2, \dots, n \quad (12.122)$$

其中  $\hat{\lambda}'_j$  是第  $n$  次埃米特多项式  $H_n(\lambda)$  的根 (第12.C.1节), 满足

$$\lambda_j = \sqrt{2}\sigma_\varepsilon \hat{\lambda}_j + \mu_\lambda,$$

$\omega_j$  是对应  $\hat{\lambda}'_j$  的权重。

研究目标与(12.121)相同: 给定  $\lambda$  的情况下有 AR(1) 过程  $\lambda'$ , 计算加权价值方程积分

$$E[V(a', \lambda')|\lambda] = \int_{-\infty}^{\infty} V(a', \lambda') f(\lambda'|\lambda) d\lambda'.$$

原本可以采用上节12.D.1介绍的Tauchen (1986) 算法做近似求解, 但需注意  $\lambda'$  的条件均值是一个关于  $\lambda$  的方程, 这意味着对于每个不同的  $\lambda$ , 都会有一组不同的  $\lambda'$ , 进而迭代算出的集合  $\Lambda$  可能有无数个。为了避免这个问题, Tauchen and Hussey (1991) 建议使用如下转换

$$\begin{aligned} E[V(a', \lambda')|\lambda] &= \int_{-\infty}^{\infty} V(a', \lambda') f(\lambda'|\lambda) d\lambda' \\ &= \int_{-\infty}^{\infty} V(a', \lambda') \frac{f(\lambda'|\lambda)}{f(\lambda'|\mu_\lambda)} f(\lambda'|\mu_\lambda) d\lambda'. \end{aligned}$$

其中  $f(\lambda'|\mu_\lambda)$  表示在  $\lambda$  恰好等于其无条件均值  $\mu_\lambda$  的情况下,  $\lambda'$  的条件密度。将高斯——埃米特求积(12.122)代入上式, 可得

$$\begin{aligned} E[V(a', \lambda')|\lambda] &\approx \frac{1}{\sqrt{\pi}} \sum_{j=1}^n \omega_j V(a', \lambda_j) \frac{f(\lambda_j|\lambda)}{f(\lambda_j|\mu_\lambda)}, \quad \text{其中} \\ \lambda_j &= \sqrt{2}\sigma_\varepsilon \hat{\lambda}_j + \mu_\lambda, \\ \hat{\lambda}_j, j = 1, 2, \dots, n &\text{表示第 } n \text{ 次埃米特多项式的根,} \\ \omega_j &\text{表示第 } n \text{ 次埃米特多项式的根对应的权重.} \end{aligned} \quad (12.123)$$

那么, 对于  $\lambda = \lambda_i$ , 我们有

$$E[V(a', \lambda'|\lambda)] \approx \sum_{j=1}^n \tilde{\omega}_{i,j} V(a', \lambda_i), \quad \tilde{\omega}_{i,j} = \frac{1}{\sqrt{\pi}} \omega_j \frac{f(\lambda_j|\lambda_i)}{f(\lambda_j|\mu_\lambda)}, \quad (12.124)$$

上式意味着，我们选取  $\tilde{\lambda}$  的可能集合为

$$\Lambda = \left\{ \tilde{\lambda} | \lambda = \sqrt{2}\sigma_{\varepsilon}\hat{\lambda}_j, j = 1, 2, \dots, n \right\},$$

选取  $\tilde{w}$  的可能集合为转移矩阵  $P$ : 需要注意的是

$$\sum_{j=1}^n \tilde{\omega}_{i,j} \neq 1,$$

因此  $P$  中的元素

$$p_{i,j} = \frac{\tilde{\omega}_{i,j}}{\sum_{j=1}^n \tilde{\omega}_{i,j}}.$$

(12.124)因此最终改写为

$$E \left[ V(a', \tilde{\lambda}' | \lambda_i) \right] \approx \sum_{j=1}^n p_{i,j} V(a', \lambda_j). \quad (12.125)$$

由(12.125)可见，利用Tauchen and Hussey (1991) 近似条件期望用到高斯——埃米特求积的知识，但在实际计算过程中并不需要做高斯——埃米特求积<sup>28</sup>。

与Tauchen (1986) 法相比，Tauchen and Hussey (1991) 的不同在于权重方程改用条件密度  $\omega(\lambda') = f(\lambda' | \mu_{\lambda})$ ，如(12.123)。与无条件密度相比，有条件密度在区间的中央赋予更多权重，在区间的两端赋予更小权重。这样一来，有条件密度作为权重方程，可以更好平衡两组互相冲突的原则：第一，一个好的近似总是会在无条件均值附近留下较多权重，第二，一个好的近似总是会使得在区间的两端， $\frac{f(\lambda' | \lambda)}{\omega(\lambda')}$  比值不至于相对于  $w(\lambda')$  而增长过快。

Iskander Karibzhanov 也提供了名为 tauchenHussey.m 的 Matlab 程序。

### 12.D.3 Rouwenhorst(1995) 法

如前文所述，对于高持续性  $\rho \rightarrow 1$ (如  $\rho = 0.9$ ) 的过程来说，Tauchen (1986) 近似法的精度较低。Kopecky and Suen (2010) 研究发现Rouwenhorst (1995) 法在此时更为理想。

在区间  $[\mu_{\lambda} - \nu, \mu_{\lambda} + \nu]$  内选取对称且均匀分布的  $n$  个点构成  $\Lambda$ 。用如下方法计算  $P$ : 通过选择  $p$  和  $p$ ，使得当时

$$P_{n=2} = \begin{pmatrix} p & 1-p \\ 1-p & q \end{pmatrix}, \quad P_{n=3} = \begin{pmatrix} p^2 & 2p(1-q) & (1-p)^2 \\ p(1-q) & pg + (1-p)(1-q) & q(1-p) \\ (1-q)^2 & 2q(1-p) & q^2 \end{pmatrix},$$

进而用递归方法，根据  $P_{n-1}$  计算  $P_n$ ，分两步。首先将  $n \times n$  矩阵相加

$$p \begin{pmatrix} P_{n-1} & \mathbf{0} \\ \mathbf{0}^\top & 0 \end{pmatrix} + (1-p) \begin{pmatrix} \mathbf{0} & P_{n-1} \\ 0 & \mathbf{0}^\top \end{pmatrix},$$

$$(1-q) \begin{pmatrix} \mathbf{0}^\top & 0 \\ P_{n-1} & \mathbf{0} \end{pmatrix} + q \begin{pmatrix} 0 & \mathbf{0}^\top \\ \mathbf{0} & P_{n-1} \end{pmatrix},$$

随后将除了第一和最后一行之外的其他行都除以 2。这样矩阵  $P_n$  的条件概率之和等于 1。

<sup>28</sup>然而随着  $n \rightarrow \infty$ ，我们有

$$\lim_{n \rightarrow \infty} \sum_{j=1}^n \tilde{\omega}_{i,j} \rightarrow 1, \quad \forall i = 1, 2, \dots, n,$$

因此(12.125)的Tauchen and Hussey (1991) 算出的条件期望近似依然是向高斯——埃米特求积法(12.122)收敛的。

$p^{n-1}$  表示当期已处于最低状态时，下期仍处在最低状态的概率； $q^{n-1}$  表示当期已处于最高状态时，下期仍处在最高状态的概率。 $(1-p)^{n-1}$  表示当期已处于最低状态时，下期由最低转为最高状态的概率； $(1-q)^{n-1}$  表示当期已处于最高状态时，下期由最高转为最低状态的概率。对于任何  $p \neq q$  的情况，都会导致生成的冲击序列中出现异方差 (heteroscedasticity)。

不论  $n$  和  $\Lambda$  的取值是怎样的， $\tilde{\lambda}$  过程的一届序列相关性总是  $p + q - 1$ 。因此，如果我们设定  $p + q = \pi$ ，就可以使得所生成的离散过程，具有和连续过程相同的一阶持续性  $\rho$ 。

此外， $\tilde{\lambda}$  的方差也总是  $\frac{\nu^2}{n-1}$ 。因此对于某个给定的  $n$  值，1 个合适的  $\nu$  值取值是使得

$$\text{var}(\tilde{\lambda}) = \text{var}(\lambda) \Rightarrow \nu = \left( \frac{n-1}{\rho^2-1} \right)^{\frac{1}{2}} \sigma_{\varepsilon}.$$

Iskander Karibzhanov 的 matlab 代码：

```
>> rouwenhorst(0.95,0.007,5)

ans =

-0.0448    -0.0224         0    0.0224    0.0448

>> rouwenhorst(0.95,0.007,7)

ans =

-0.0549    -0.0366    -0.0183         0    0.0183    0.0366    0.0549

>> rouwenhorst(0.95,0.007,9)

ans =

-0.0634    -0.0476    -0.0317    -0.0159         0    0.0159    0.0317    0.0476    0.0634
```

## 12.E 伽辽金近似

### 12.E.1 举例：1 维空间中的伽辽金近似

以一个 1 维空间的有界极值问题为例，介绍伽辽金近似法 (Galerkin approximation method)。假定  $u(x)$  是下述方程系统的解

$$\begin{cases} -u'' + u = f, & \text{in } (0, 1) \\ u(0) = u(1) = 0, \end{cases} \quad (12.126)$$

我们的目标是找到一个对  $u$  的近似解。对上式两侧乘以一个检验方程  $\nu$ ，然后分段积分，可得(12.126)的弱条件形式

$$\int_0^1 (u' \nu' + u \nu) dx = \int_0^1 f \nu dx, \quad \forall \nu \in H_0^1(0, 1), \quad (12.127)$$

其中  $H_0^1(0, 1)$  是索伯列夫空间 (Sobolev space)。

条件补一个  
ference

一个 reference

如果方程  $u$  是常规方程, 即连续且二阶可导, 那么(12.126)与(12.127)等价。这意味着我们可以利用(12.127), 构建一个多项式近似  $u$ 。将域  $[0, 1]$  划分为  $N+1$  个均匀分布 (uniform distribution) 的子域, 每个域对应子空间  $(x_j, x_{j+1})$ , 满足

$$x_j = \frac{j}{N+1} \quad j = 0, 1,$$

此外, 定义一个包含检验方程  $\nu$  的空间  $V_N$ , 满足

$$V_N = \{\nu \in C^0 : \nu|_{[x_j, x_{j+1}]} \text{ 是线性方程, 并且 } \nu(0) = \nu(1) = 0\},$$

其中  $C^0$  表示连续方程的空间。

$V_N$  满足如下特征: 首先,  $V_N$  是个有限维度空间,  $N < \infty$ 。其次,  $V_N \subset H_0^1(0, 1) \forall N$ 。事实上,  $\nu(x) \in V_N$  由且仅由他所对应的有限个数的点  $(x_1, x_2, \dots, x_N)$  所决定。

定义一个伽辽金近似方程  $u_N \in V_N$  作为对  $u$  的近似, 代入(12.127), 我们有

$$\int_0^1 (u'_N \nu' + u_N \nu) dx = \int_0^1 f \nu dx, \quad \forall \nu \in V_N, \quad (12.128)$$

由下节可知, 存在且只存在一个  $u_N$  方程满足(12.128)条件。并且由于  $V_N$  是个有限维度空间, 因此  $u_N$  可以通过一些线性方程系统求解方法算得。基方程  $\psi_j \in V_N$  可以采取不同形式, 例如可以选拉格朗日多项式 (Lagrange polynomial)

$$\psi_j(x_i) = \delta_{ij}, \quad i, j = 1, 2, \dots, N, \quad (12.129)$$

其中  $\delta_{ij}$  表示克罗内克乘积。因此我们有近似方程

$$u_N = \sum_{j=1}^N U_j \psi_j, \quad U_j \in \mathbb{R}, \quad (12.130)$$

其中  $U_j := u_N(x_j)$ 。

既然任何一个检测方程  $\nu \in V_N$  都与  $\psi_j$  呈线性组合关系, (12.128)等价于

$$\int_0^1 (u'_N \psi'_k + u_N \psi_k) dx = \int_0^1 f \psi_k dx, \quad k = 1, 2, \dots, N. \quad (12.131)$$

再引入(12.130), 替换上式 LHS 中的  $u_N$ , 我们有

$$\sum_{j=1}^N U_j \int_0^1 (\psi'_j \psi'_k + \psi_j \psi_k) dx = \int_0^1 f \psi_k dx, \quad k = 1, 2, \dots, N. \quad (12.132)$$

(12.132)构成一个  $N$  个方程的线性系统

$$AU = F,$$

$$\begin{cases} \text{系数矩阵 } A = (a_{kj}) \in \mathbb{R}^{N \times N}, \quad a_{kj} = \int_0^1 (\psi'_j \psi'_k + \psi_j \psi_k) dx, \\ \text{系数矩阵 } F = (F_k) \in \mathbb{R}^N, \quad F_k = \int_{0,1} f \psi_k dx, \\ \text{待求解矩阵 } U = (U_j) \in \mathbb{R}^N \end{cases} \quad (12.133)$$

通过求解线性方程系统(12.133), 可得矩阵  $U$  的解, 进而根据(12.130)可得  $u_N$  的值。

通过观察  $\psi_j(x_i)$  的性质(12.129)可见, 矩阵  $A$  是个对称的三角对角矩阵, 对角元素及其附近不等于 0 的元素分别为

$$\begin{cases} a_{jj} = \frac{2}{h} + \frac{2}{3}h, \\ a_{jj-1} = a_{jj+1} = -\frac{1}{h} + \frac{h}{6}, \end{cases}$$

由此，待求解方程系统为  $N + 1$  个等式

$$\begin{cases} U_0 := 0, \\ \frac{-U_{j-1} + 2U_j - U_{j+1}}{h} + \frac{h}{6}U_{j-1} + \frac{2}{3}hU_j + \frac{h}{6}U_{j+1} = F_j, & j = 1, 2, \dots, N, \\ U_{j+1} := 0. \end{cases}$$

矩阵  $A$  是对称正定的（即  $C^\top AC > 0 \forall C \neq 0$ ；由  $a_{ij}$  积分项中的双线性形式推得），由此可得  $A$  是可逆矩阵。因此， $U$  进而  $u_N$  的解是唯一的。

上式中两侧同时除以  $h$ ，我们得到(12.126)的有限差分近似形式

$$\underbrace{\frac{-U_{j-1} + 2U_j - U_{j+1}}{h^2}}_{\approx u''(x_j)} + \underbrace{\frac{1}{6}U_{j-1} + \frac{2}{3}U_j + \frac{1}{6}U_{j+1}}_{\approx u(x_j)} = \underbrace{\frac{1}{h}F_j}_{\approx f(x_j)}, \quad j = 1, 2, \dots, N,$$

也就是说，在这个 1 维空间的例子中，伽辽金近似法就是将原求解问题转换为一个已知数量  $N$  的有限差分近似问题。

收敛问题。对于给定数量  $N$ ，我们定义一个伽辽金近似  $u_N \in V_N$ ，作为对原方程系统  $u$  的近似。随着  $N \rightarrow \infty$ ，正交多项式的级数逐渐增加，可用来近似任何连续方程。换句话说，当空间集合  $V_N$  以下方式近似  $u$  时，伽辽金近似  $u_N$  逐渐收敛至  $u$ ：

$$\lim_{N \rightarrow \infty} \|u - v\| = \lim_{N \rightarrow \infty} \inf_{v \in V_N} \|u - v\| \rightarrow 0.$$

## 12.E.2 伽辽金近似

第12.E.1节以 1 维空间为例介绍了伽辽金近似的基本思路和方法。这节我们从 1 维扩展到通用情况，即在一个希尔伯特空间 (Hilbert space) 中，双线性形式下的伽辽金近似法。

将一个希尔伯特空间定义为  $V$ ，在其中定义一个双线性形式的连续方程  $a(\cdot, \cdot)$  和一个线性方程  $L$ ，我们的研究目标是找到一个近似解  $u \in V$ ，求解原方程系统

$$a(u, \nu) = \langle L, \nu \rangle, \quad \forall \nu \in V, \tag{12.134}$$

其中  $\langle \cdot, \cdot \rangle$  表示在  $V'$  和  $V$  中的双内积 (duality inner product)<sup>29</sup>。下面我们来分析，为了确保  $u$  存在唯一解，方程  $a(\cdot, \cdot)$  所应呈现的形式，即关于强制形式 (coercive forms) 的讨论。

**Definition 12.6** (方程的强制形式). 如果存在一个常数  $\alpha > 0$ ，使得满足

$$a(u, u) \geq \alpha \|u\|_V^2, \quad \forall u \in V, \tag{12.135}$$

那么我们说方程  $a(\cdot, \cdot)$  是一个  $V$  空间中的强制形式。

举例说明，以(12.134)为例，对于二阶线形椭圆方程

$$\begin{cases} -\sum_{i,j=1}^n \frac{\partial}{\partial x_i} \left( a_{ij} \frac{\partial u}{\partial x_j} \right) = f, & \text{in } \Omega \subset \mathbb{R}^n, \\ u = 0, & \text{on } \partial\Omega \end{cases}$$

其中系数  $a_{ij} = a_{ij}(x)$  是有界方程。那么存在任意  $\gamma > 0$  使得

$$\gamma |\xi|^2 \leq \sum_{i,j=1}^n a_{ij} \xi_i \xi_j, \quad \forall x \in \Omega, \forall \xi \in \mathbb{R}^n, \tag{12.136}$$

或者说，可以以(12.134)的形式来表述，其中

<sup>29</sup>如  $\langle A, B \rangle := A'B' + AB$ ,  $\forall (A, B) \in V$ ,  $(A', B') \in V'$ 。注意和(12.52)的定义有所区别。

- 希尔伯特空间  $V$

$$V = H_0^1(\Omega) = \left\{ \nu \in L^2(\Omega) : \frac{\partial \nu}{\partial x_j} \in L^2(\Omega), \quad j = 1, 2, \dots, n, \text{ 以及在 } \partial\Omega \text{ 上有 } \nu = 0 \right\},$$

- $\nu \in V$  的范数 (norm) 为

$$\|\nu\|_{H^1} = \|\nu\|_{L^2} + \|\nabla \nu\|_{L^2},$$

- $a(u, \nu)$  为

$$a(u, \nu) = \sum_{i,j=1}^n \int_{\Omega} a_{i,j} \frac{\partial u}{\partial x_i} \frac{\partial \nu}{\partial x_j} dx,$$

- $L$  满足

$$\langle L, \nu \rangle = \int_{\Omega} f \nu dx.$$

结合椭圆方程条件(12.136)、系数  $a_{ij}$  的有界特性以及庞加莱不等式，可得  $a$  函数形式是强制且连续的。线性方程形式  $L$  如果满足  $f \in L^2$ ，则我们说它是连续的。

再举一个线性弹性方程的例子。对于应变——应力关系方程

$$\begin{cases} -\mu \Delta \mathbf{u} - (\lambda + \mu) \nabla \operatorname{div} \mathbf{u} = \mathbf{f}, & \text{in } \Omega \subset \mathbb{R}^3, \\ \mathbf{u} = 0, & \text{on } \partial\Omega \end{cases}$$

其中出于简化考虑，设满足齐次狄利克雷条件 (homogeneous Dirichlet condition)。大于 0 的系数  $\lambda$  和  $\mu$  分别为拉梅第一参数和拉梅第二参数 (Lamé elasticity parameters)。向量方程  $\mathbf{u}$  和  $\mathbf{f}$  未知。这个问题的弱形式可以表述为带有  $V = H_0^1(\Omega)^3$  条件的式(12.134)，其中  $a(\mathbf{u}, \mathbf{v})$  满足条件

$$a(\mathbf{u}, \mathbf{v}) = \int_{\Omega} \{2\mu \varepsilon_{i,j}(\mathbf{u}) \varepsilon_{i,j}(\mathbf{v}) + \lambda \operatorname{div} \mathbf{u} \operatorname{div} \mathbf{v}\} dx, \quad \varepsilon_{i,j} := \frac{1}{2} \left( \frac{\partial u_i}{\partial x_j} \frac{\partial v_j}{\partial x_i} \right).$$

进而，利用科恩不等式 (Korn's inequality)，我们有双线性形式的方程  $a(., .)$  (12.134) 是强制形式。

方程形式  $a(., .)$  (12.134) 的解得存在性和唯一性，可由拉克斯—密格拉蒙定理 (Lax-Milgram theorem, Theorem 14.14) 予以证明。

在此基础上，上述条件也证实了伽辽金近似方程向原方程收敛。事实上，的确存在着一些重要范例如斯托克斯公式 (Stokes equations) 等，其中的双线性形式方程并不是强制的，但满足一个更弱的条件，称为上极限——下极限条件 (inf-sup condition)<sup>30</sup>。下面先介绍上线——下线极限条件定义。在第12.E.3节介绍双线性形式伽辽金近似的收敛问题。

**Definition 12.7** (上极限——下极限关系问题). 如果双线性形式的方程  $a(u, \nu)$  在  $V$  空间上满足条件：对于某一  $\alpha > 0$ ，有

$$\sup_{\nu \in V} \frac{a(u, \nu)}{\|\nu\|_V} \geq \alpha \|u\|_V, \quad \forall u \in V, \tag{12.137a}$$

$$\sup_{u \in V} \frac{a(u, \nu)}{\|u\|_V} \geq \alpha \|\nu\|_V, \quad \forall \nu \in V, \tag{12.137b}$$

那么我们说，它满足上极限——下极限关系。

根据这一定义，我们有以下推论：

<sup>30</sup>关于稳态斯托克斯公式弱形式的分析，以及上极限——下极限条件，可参考 Chen Long 的讲义 <https://www.math.uci.edu/~chenlong/226/inf-sup.pdf>。

1. 如果  $a$  是对称方程, (12.137a)和(12.137b)等价。

2. 两个方程也可以写作

$$\inf_{u \in V} \sup_{\nu \in V} \frac{a(u, \nu)}{\|u\|_V \|v\|_V}. \quad (12.138)$$

3. 如果  $a$  是强制方程, 那么它满足上极限——下极限关系, 这是由于

$$\sup_{\nu \in V} \frac{a(u, \nu)}{\|\nu\|_V} \geq \frac{a(u, \nu)}{\|u\|_V} \geq \alpha \|u\|_V. \quad (12.139)$$

4. 双线性方程  $a$  的上极限——下极限关系可以表示为关于  $a$  的线性算子  $A$  及其伴随算子  $A^*$ ,

$$A : V \rightarrow V', \quad A^* : V' \rightarrow V,$$

两个算子定义为

$$\langle Au, \nu \rangle_{V' \times V} = a(u, \nu),$$

$$\langle u, A^* \nu \rangle_{V' \times V} = a(u, \nu),$$

从而(12.137a)-(12.137b)等价于

$$\|Au\|_{V'} \geq \alpha \|u\|_V, \quad \forall u \in V, \quad (12.140a)$$

$$\|A^* \nu\|_{V'} \geq \alpha \|\nu\|_V, \quad \forall \nu \in V. \quad (12.140b)$$

5. 当  $V = \mathbb{R}^n$  时,  $a$  的强制性意味着, 其对应的矩阵  $A$  正定;  $a$  若满足上极限——下极限关系, 则  $A$  进一步是可逆的。

**Theorem 12.59.** 对于连续的双线性形式方程  $a$ , 当且仅当线性算子  $A$  是双射 (bijective) 的时,  $a$  满足上极限——下极限关系。 $A$  是双射的意思是说, 原方程系统(12.134)对于任意  $L \in V$  都有唯一解, 进而  $A$  有一个连续逆 (continuous inverse), 即  $\|u\|_V \leq C \|L\|_{V'}^{31}$ 。

证明. 根据泛函分析的基本原理可得, 对于  $W \subset V$ , 我们定义  $W^0 \subset V'$ , 满足

$$W^0 \subset \{L \in V' : \langle L, \nu \rangle = 0, \quad \forall \nu \in W\},$$

那么我们有

$$(Ker A)^0 = \overline{\Im A^*}, \quad (12.141a)$$

$$(Ker A^*)^0 = \overline{\Re A}, \quad (12.141b)$$

其中  $\Im(\cdot)$  表示复数的虚部 (imaginary part); 对应地, 复数的实部 (real part) 表示为  $\Re(\cdot)$ 。

如果假定  $a$  满足上极限——下极限关系, 那么由(12.140a)可得  $A$  是单射 (injective) 的, 由(12.140b)可得  $A^*$  也是单射的。进而, 如果我们能证明  $\Im A$  是封闭集, 则根据(12.141b)可以证明定理。

假定  $Au_n \rightarrow w$ , 则由(12.140a)可得

$$\|A(u_n - u_m)\|_{V'} \geq \alpha \|u_n - u_m\|_V,$$

由此可见  $\{u_n\}$  是一个柯西序列 (Cauchy sequence),  $u_n$  收敛向某一个  $u \in V$ , 并且由  $A$  的连续性特征可得,  $w = Au \in \Im A$ 。

同理可得, 如果  $A$  是双射的, 那么  $A^*$  也是双射的,  $A$  和  $A^*$  都有一个连续逆。因此(12.137a)-(12.137b)成立。  $\square$

<sup>31</sup>关于连续逆, 可参考讲义 Li Shiu-Tang 的讲义 <http://www.math.utah.edu/li/Right%20continuous%20inverse.pdf>。

在此基础上，我们为原方程系统(12.134)作伽辽金近似解。假定在  $V$  空间内，我们有一个有限维度的子空间  $V_N$ 。则对原方程系统的伽辽金近似  $u_N \in V_N$  可定义为

$$a(u_N, \nu) = \langle L, \nu \rangle, \quad \forall \nu \in V_N. \quad (12.142)$$

为了让  $u_N$  定义良好 (well defined)，我们需要对  $a$  作出限定。根据 Theorem 12.59 可得当且仅当  $a$  在  $V_N$  子空间中满足上极限——下极限关系时，伽辽金近似解  $u_N$  才是存在且唯一的。尤其是，伽辽金近似解  $u_N$  对于强制形式的方程  $a$  是定义良好的。

从这个角度出发，有必要指出一下两种  $V$  中方程形式的本质区别：一种是强制形式，另一种是满足上极限——下极限关系，但并不是强制形式的：

**Remark 12.1.** 如果  $a$  在  $V$  上强制，那么它在  $V$  的任何一个子空间中，尤其是  $V_N$  中都是强制的，此时伽辽金近似解  $u_N$  定义完备。

但  $V$  上的上极限——下极限关系，无法传导到  $V$  的子空间中，那么对于不满足强制性的方程形式，还需要在子空间  $V_N$  中重新检验，以确保  $u_N$  定义完备。

### 12.E.3 强制型方程形式的伽辽金近似收敛

假定我们已知双线性形式方程  $a(.,.)$  是连续和强制的。因此我们有

$$a(u, \nu) \leq M \|u\|_V \|\nu\|_V, \quad \forall u, \nu \in V, \quad (12.143)$$

其中  $M$  定义为一个连续性常数 (continuity constant)。在采用伽辽金近似法求得近似解  $u_N$  后，一个自然出现的问题是：在  $V_N$  被恰当选取的前提下，是否会出现  $\lim_{N \rightarrow \infty} u_N \rightarrow u$  的收敛？

显然，如果答案是肯定的，那么

$$d(u, V_n) = \inf_{\nu \in V_n} \|u - \nu\|_V \rightarrow 0, \quad \text{随着 } N \rightarrow \infty. \quad (12.144)$$

换句话说，收敛与否的问题，等价于探讨在子空间  $V_N$  中(12.144)是否成立<sup>32</sup>，以及它是不是收敛的充要条件。根据齐亚引理 (Céa's lemma)，答案是肯定的。

**Lemma 12.1** (齐亚引理). 如果  $a$  是一个连续且强制的方程，那么我们有

$$\|u - u_N\|_V \leq \frac{M}{\alpha} \inf_{\nu \in V_N} \|u - \nu\|_V \quad (12.145)$$

证明. 将伽辽金近似解 (12.142) 和(12.143)联立，构成误差方程

$$a(u - u_N, \nu) = 0, \quad \forall \nu \in V_N. \quad (12.146)$$

由(12.143), , (12.144)和 (12.146)我们有

$$\alpha \|u - u_N\|^2 \leq a(u - u_N, u - u_N) = a(u - u_N, u - \nu) \leq M \|u - u_N\|_V \|u - \nu\|_V,$$

因此

$$\|u - u_N\| \leq M \|u - u_N\|_V \|u - \nu\|_V, \quad \forall \nu \in V_N,$$

证毕。 □

根据齐亚引理，在子空间  $V_N$  中，伽辽金近似解  $u_N$  是对原方程系统解  $u$  最好的近似，它除了受一个常数影响外，但只受双线性方程  $a$  影响，而与子空间  $V_N$  的情况无关。根据这一特征，我们可以引出伽辽金近似解的收敛特性。

<sup>32</sup>子空间  $V_N$  中的所有方程，都是对  $u$  的近似。

**Theorem 12.60** (伽辽金近似解的收敛 (强制形式)). 如果  $a$  是个连续且强制的方程, 并且空间  $V_N$  满足 (12.144), 那么我们有

$$\lim_{N \rightarrow \infty} u_N = u.$$

来讨论一个特殊情况,  $a$  还额外具有对称性。那么在  $V$  上,  $a$  定义了一个标量积 (scalar product), 这个标量积与原方程系统相同。此时, 误差方程(12.146)表明伽辽金近似解  $u_N$  恰好就是  $u$  在子空间  $V_N$  上的正交映射, 对应标量积方程  $a$ 。此时标量积  $a$  所对应的范数提供了最佳近似解。并且对应地, 齐亚引理中的常数项  $M/\alpha$  变为  $\sqrt{M/\alpha}$ 。

#### 12.E.4 仅满足上极限——下极限关系方程形式的伽辽金近似收敛

现在假定  $a$  不是强制的, 但满足上极限——下极限关系(12.137a)-(12.137b)。已知原方程系统(12.134)有唯一解, 我们来讨论一下此时伽辽金近似的收敛问题。如 Remark 12.1 所强调的, 上极限——下极限关系无法从  $V$  向下传递到子空间  $V_N$ , 因此为了让伽辽金近似是定义良好的, 我们需要假定  $a$  也在子空间  $V_N$  中满足上极限——下极限关系, 即存在一个系数  $\beta > 0$ , 使得

$$\sup_{\nu \in V_N} \frac{a(u, \nu)}{\|\nu\|_V} \geq \beta \|u\|_V, \quad \forall u \in V_N. \quad (12.147a)$$

$$\sup_{u \in V_N} \frac{a(u, \nu)}{\|u\|_V} \geq \beta \|\nu\|_V, \quad \forall \nu \in V_N \quad (12.147b)$$

我们证明(12.147a)即可。这是因为既然子空间  $V_N$  是有限维度的, 那么一旦(12.147a)成立, (12.147b)也同样成立。

收敛若要存在, 需要假定  $\beta$  不依赖于  $N$ 。进而, Babuška (1971) 对齐亚引理 12.1 作以扩展, 据此可以对 Theorem 12.60 作扩展。

**Lemma 12.2** (扩展齐亚引理). 如果  $a$  是连续方程, 满足(12.147a)的上极限——下极限关系, 那么

$$\|u - u_N\|_V \leq \left( \frac{\beta + M}{\beta} \right) \inf_{\nu \in V_N} \|u - \nu\|_V,$$

尤其是当  $\beta$  不依赖于  $N$  时, 误差项中的常数也不依赖于  $N$ 。

证明. 取  $\nu \in V_N$ 。由(12.147a)及误差方程(12.146)可得

$$\beta \|\nu - u_N\|_V \leq \sup_{w \in V_N} \frac{a(\nu - u_N, w)}{\|w\|_V} = \sup_{w \in V_N} \frac{a(\nu - u, w)}{\|w\|_V} \leq M \|\nu - u\|_V.$$

□

由此我们有收敛定理:

**Theorem 12.61** (伽辽金近似解的收敛 (上极限——下极限形式)). 如果  $a$  是个连续方程, 满足上极限——下极限关系(12.147a),  $\beta$  不依赖于  $N$ , 并且子空间  $V_N$  满足(12.144), 那么我们有

$$\lim_{N \rightarrow \infty} u_N = u.$$

## 12.F 数值求积

我们常常需要计算一个定积分的值, 如

$$\int_a^b f(x) dx,$$

若直接用解析法求解较为困难，则常常采取数值积分 (numerical integration) 的思路，将积分式转化为有限个方程求和的方式做近似求解，如

$$\int_a^b f(x) dx \approx \sum_{n=1}^N w_n f(x_n), \quad (12.148)$$

也即，我们在原方程的取值区间  $[a, b]$  中抽取  $N$  个点， $n = 0, 1, \dots, N$  的  $x_n$  值代回方程中对应值  $f(x_n)$ 。 $w_n$  表示相应的权重系数。

因此，数值积分也常常称为数值求积 (numerical quadrature)。集合  $\{x_n\}, \{w_n\}$  常常称为求积点 (quadrature points) 集合和求积权重 (quadrature weights) 集合；合适的选取  $\{x_n\}, \{w_n\}$  的算法称为求积法则 (quadrature rule)。对于同一个积分求解问题(12.148)往往存在一系列法则；评价不同法则之间好坏的标准在于，看哪个法则能用最少的样本点  $N$  来对(12.148)作出最精确的近似，同时确保计算成本可控，如编程难度、计算时间等。

先从最基本的牛顿——寇特斯法则开始介绍。

### 12.F.1 牛顿——寇特斯法则

近似求解积分问题(12.148)的基本思路是，在区间  $[a, b]$  中找到一个多项式方程  $P(x)$  来近似  $f(x)$ 。由于多项式的求和计算往往较为简单 (多项式的介绍见第12.B节)，这会简化(12.148)RHS 的计算时间。我们将这种思路称为牛顿——寇特斯 (求积) 法则 (Newton—Cotes Rule)。牛顿——寇特斯法则随着多项式的次 (degree, order) 而呈层级特征：使用  $p$  次多项式的牛顿——寇特斯法则可称为  $p$  阶牛顿——寇特斯法则。

如果  $f$  不是多项式，那么在  $[a, b]$  区间内寻找近似多项式  $P(x)$  会较为困难。一个近似方案是将空间  $[a, b]$  划分为  $N$  个子区间 (对应  $N + 1$  个点)，在每个子空间中分别寻找可以近似  $f$  的多项式，进而将多个多项式加权求和，作为整个空间中  $f$  的近似方程。 $N$  越大，子空间的个数越多，划分就越精细，每个子空间中的近似就越精确，进而整个空间中的近似求积就越精确。

#### 矩形法则

我们先从  $p = 0$  的情况开始理解牛顿——寇特斯法则，根据黎曼求和 (Riemann sums)，我们可以将方程  $f(x)$  理解为一组矩形的集合，对  $f(x)$  在区间  $[a, b]$  中求积就是求曲线下方的面积，近似等于各个矩形面积的和。随着  $N$  趋近于无穷大，每个矩形的宽度无限接近于 0，近似就越精确。

在每个子空间中分别用 1 个  $p = 0$  次多项式来近似  $f$ ，0 次多项式是个常数，换句话说我们在每个子空间中用一个常数来近似  $f$ 。这又称为矩形求积法则 (rectangular quadrature rule)。

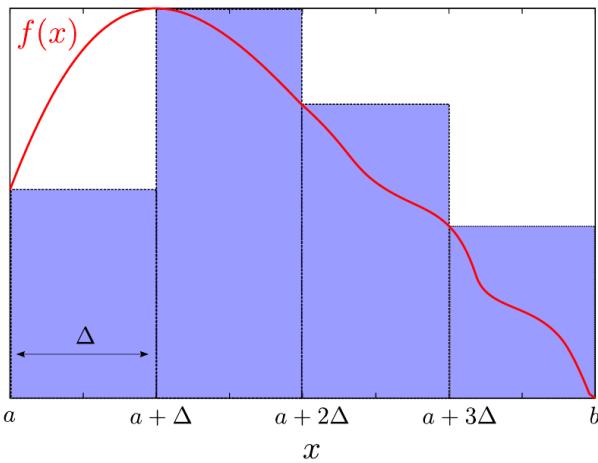
图12.F.1绘出了  $N = 4$  情况下，利用矩形法则对  $\int_a^b f(x) dx$  的近似。不难看出，将区间  $[a, b]$  划分为  $N$  个等宽子区间，每个子区间的宽度都是  $\Delta = (b - a) / N$ 。最左侧第一个矩形，左边长 (高) 为  $f(a)$ ，宽为  $\Delta$ ，对应面积为  $f(a) \cdot \Delta$ 。左数第二个矩形，左边长 (高) 为  $f(a + \Delta)$ ，宽也是  $\Delta$ ，面积为  $(a + \Delta) \cdot \Delta$ ，以此类推直到第  $N = 4$  个矩形为止。将这些子区间中矩形的面积加总，可得矩形法则的近似表达式

$$\mathcal{I}_{N=4}^{rect} \approx f(a) \cdot \Delta + f(a + \Delta) \cdot \Delta + f(a + 2\Delta) \cdot \Delta + f(a + 3\Delta) \cdot \Delta,$$

扩展到更一般的  $N \in \mathcal{N}$  的情况，利用  $N$  段矩形法则对积分  $\int_a^b f(x) dx$  的近似为

$$\mathcal{I}_N^{rect} = \sum_{n=0}^{N-1} f(a + n \cdot \Delta) \cdot \Delta. \quad (12.149)$$

矩形法则(12.149)成为对(12.148)的近似求积法则之一：

图 12.F.1: 矩形法则 ( $N = 4$ )

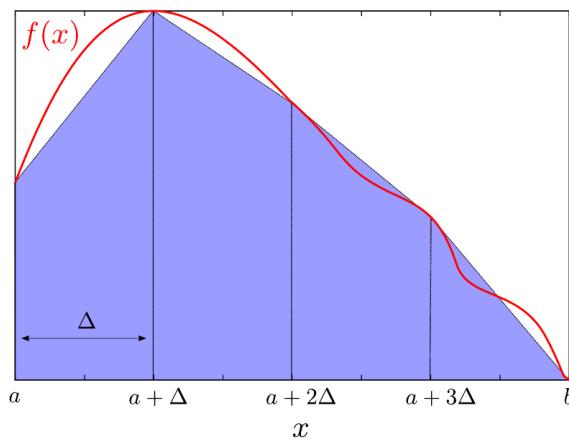
- 求积权重集合  $\{w_n\}_{n=0}^{N-1}$  对应  $w_n \equiv \Delta \forall n$ ,
- 求积点集合  $\{x_n\}_{n=0}^{N-1}$  对应  $x_n = a + n \cdot \Delta, n = 0, 1, \dots, N - 1$ 。

值得注意的是，在矩形法则下，我们只需要左边长，从而无需计算  $f(b)$ 。

### 梯形法则

从图12.F.1中不难看出，利用矩形法则近似曲线下方阴影的面积，近似效果并不理想。一个可能的改进方案是利用梯形代替矩形做近似，又称梯形近似法则。

在每个子空间中分别用 1 个  $p = 1$  次多项式来近似  $f$ ，1 次多项式是条斜线，连接子区间的两个端点（始点和终点）。这又称为梯形求积法则 (trapezoidal quadrature rule)。

图 12.F.2: 梯形法则 ( $N = 4$ )

见图12.F.2所示，左数第 1 个梯形的面积等于

$$\frac{1}{2} [f(a) + f(a + \Delta)] \cdot \Delta,$$

左数第 2 个梯形的面积等于

$$\frac{1}{2} [f(a + \Delta) + f(a + 2\Delta)] \cdot \Delta,$$

以此类推直到第  $N = 4$  个梯形。将子区间中梯形的面积加总，可得

$$\begin{aligned}\mathcal{I}_{N=4}^{trap} &= \frac{1}{2} [f(a) + f(a + \Delta)] \cdot \Delta + \frac{1}{2} [f(a + \Delta) + f(a + 2\Delta)] \cdot \Delta \\ &\quad + \frac{1}{2} [f(a + 2\Delta) + f(a + 3\Delta)] \cdot \Delta + \frac{1}{2} [f(a + 3\Delta) + f(a + 4\Delta)] \cdot \Delta \\ &= \left[ \frac{1}{2} f(a) + f(a + \Delta) + f(a + 2\Delta) + f(a + 3\Delta) + \frac{1}{2} f(a + b) \right] \cdot \Delta,\end{aligned}$$

扩展到更一般的  $N \in \mathbb{N}$  的情况，我们有  $N$  段梯形法则对  $\int_a^b f(x) dx$  的近似为

$$\mathcal{I}_N^{trap} = \frac{1}{2} f(a) \cdot \Delta + \Delta \cdot \sum_{n=1}^{N-1} f(a + n \cdot \Delta) + \frac{1}{2} f(b) \cdot \Delta. \quad (12.150)$$

梯形法则(12.150)成为对(12.148)的又一种近似求积法则：

- 求积点集合  $\{x_n\}_{n=1}^{N-1}$  对应  $x_n = a + n \cdot \Delta, n = 0, 1, \dots, N$ ,
- 求积权重集合  $\{w_n\}_{n=1}^N$  对应

$$w_n = \begin{cases} \frac{1}{2}\Delta, & n = 0, N, \\ \Delta, & n = 1, 2, \dots, N-1, \end{cases}$$

值得注意的是，比起矩形法则来，当利用梯形法则做近似求积时，需计算  $N+1$  个求积抽样点，比矩形法则多出的 1 个点为总区间中的末端  $b$ ，对应  $f(b)$ 。

### 更高阶牛顿——寇特斯法则

已知用  $p = 0$  阶多项式（常数）近似子区间中的  $f$ ，称矩形法则。用  $p = 1$  阶多项式（斜线）近似，称梯形法则。那么我们可以进一步迭代计算更高阶  $p > 1$  的牛顿——寇特斯求积近似，方法如下

1. 在子区间  $[x_n, x_{n+1}]$  中，配  $p+1$  个等距离分布的点（包括子区间的起始点  $x_n$  和终结点  $x_{n+1}$ ），将子区间分为  $p$  个部分。例如，
  - 对  $p = 2$  而言，配 3 个点  $x_n, \frac{1}{2} \cdot [x_n + x_{n+1}], x_{n+1}$ ，
  - 对  $p = 3$  而言，配 4 个点  $x_n, \frac{1}{3} \cdot [x_n + x_{n+1}], \frac{2}{3} \cdot [x_n + x_{n+1}], x_{n+1}$ ，以此类推。
2. 在子空间中针对这  $p+1$  个配点，找到唯一的多项式  $P(x)$  以近似  $f(x)$ 。例如，
  - $p = 0$ ， $P(x)$  是个常数。矩形法则。
  - $p = 1$ ， $P(x)$  是条斜线，连接子区间的两个端点  $f(x_n), f(x_{n+1})$ 。梯形法则。
  - $p = 2$ ， $P(x)$  是一条经过如下 3 个点的抛物线，3 个点的坐标依次为

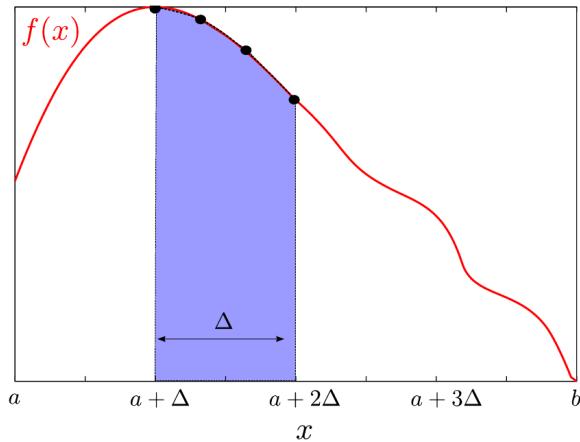
$$\begin{cases} (x_n, f(x_n)), \\ (x_m, f(x_m)), \quad x_m \equiv \frac{x_n+x_{n+1}}{2}, \\ (x_{n+1}, f(x_{n+1})). \end{cases}$$

- $p = 3$ ,  $P(x)$  是一条经过如下 4 个点的曲线 (图12.F.3), 4 个点的坐标依次为

$$\begin{cases} \left( x_n, f(x_n) \right), \\ \left( x_m, f(x_m) \right), & x_m \equiv \frac{1}{3}(x_n + x_{n+1}), \\ \left( x_{m+1}, f(x_{m+1}) \right), & x_{m+1} \equiv \frac{2}{3}(x_n + x_{n+1}), \\ \left( x_{n+1}, f(x_{n+1}) \right), \end{cases}$$

以此类推。无论哪一个例子中,  $P(x)$  都是这样一个多项式, 其系数是抽样方程  $f(x)$  值

图 12.F.3: 3 阶牛顿——寇特斯法则, 对应子区间中的 4 个配点



的线性组合。

3. 从  $x_n$  到  $x_{n+1}$  对  $P(x)$  求积分, 作为这个子区间内方程  $f(x)$  的近似。

4. 将所有子区间内的近似方程组合起来, 作为总区间中的最终近似求积法则。

$p = 2$  时的牛顿——寇特斯求积, 又称辛普森法则 (Simpson's rule)

$$\mathcal{I}_N^{simp} = \sum_{n=0}^{N-1} \frac{\Delta}{6} \left[ f(a + n \cdot \Delta) + 4f\left(a + \left(n + \frac{1}{2}\right) \cdot \Delta\right) + f(a + (n + 1)) \cdot \Delta \right]. \quad (12.151)$$

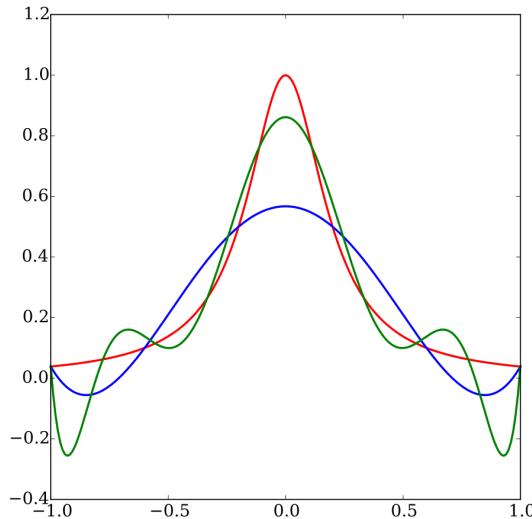
### 龙格现象

一个自然出现的问题: 更高阶的牛顿——寇特斯法则会不会带来更精确的近似解? 很遗憾, 答案是否定的。这是由于龙格现象 (Runge phenomenon): 当我们试图利用子空间中等距分布的点做更高阶多项式近似时, 在插值点附近的多项式方程会出现较大幅度的震荡, 从而影响最终求积估计的精确度, 可参考[维基百科词条](#)。图12.F.4中, 蓝线和绿线分别表示  $p = 5$  和  $p = 9$  时的差值多项式近似, 对应  $N = 6$ 。红线表示龙格方程 (Runge function)。

### 误差的收敛

比较不同求积法则的优劣, 可用启发式误差 “分析”<sup>33</sup>, 观察近似误差随着配点数  $N$  增加的衰减速度。具体说来

<sup>33</sup>加上引号是指, 这种方法只是一种较为常用的检验措施, 而不宜理解为某种严谨的学术性系统性表达。

图 12.F.4: 龙格现象 ( $N = 6$ )

- 考虑区间中某一宽度为  $\Delta$  的特定子区间，在其中用一个  $p$  阶多项式  $P(x)$  来近似  $f$

$$P(x) = c_0 + c_1x + c_2x^2 + \dots + c_p x^p.$$

- 那么子区间中存在一个点  $x_0$ ，使得  $f$  围绕  $x_0$  做泰勒级数展开的前  $p+1$  项，与  $P(x)$  一致<sup>34</sup>。可将  $f$  写成如下形式

$$f(x) = \underbrace{c_0 + c_1x + \dots + c_p x^p}_{\equiv P(x)} + c_{p+1}x^{p+1} + \dots$$

- 由此可见，原方程  $f(x)$  及其近似多项式  $P(x)$  之间的误差是一个从  $p+1$  阶开始的多项式

$$f(x) - P(x) = c_{p+1}x^{p+1} + c_{p+2}x^{p+2} + \dots,$$

- 那么在这个子区间  $[x_n, x_{n+1}]$  内的全部误差可求积得出

$$\begin{aligned} \int [f(x) - P(x)] dx &= \int [c_{p+1}x^{p+1} + c_{p+2}x^{p+2} + \dots] dx \\ &\propto \Delta^{p+2} + \text{更高阶项}, \end{aligned}$$

最后一行是说，对  $x^{p+1}$  沿着某个宽度为  $\Delta$  的子区间求积，等于某个和  $\Delta^{p+2}$  成比例的值。

- 将一个子区间中的情况扩展到其他子区间，每个子区间中的误差项都与  $\Delta^{p+2}$  成定比例。更进一步地，由于  $\Delta$  与  $N$  反比例相关 ( $N$  是整个区间中划分的子区间的数量)

$$\Delta \sim \frac{1}{N},$$

<sup>34</sup>例如

- $p = 0$ ，矩形法则， $x_0$  可以是子区间的起始点。
- $p = 1$ ，梯形法则， $x_0$  的存在性可由中值定理予以证明： $x_n$  和  $x_{n+1}$  之间必然存在一点，该点上的方程值  $f(x_0)$  的导数，等于连接  $f(x_n)$  和  $f(x_{n+1})$  两点的斜线的斜率。
- $p > 1$ ，更高阶求和法则中  $x_0$  点存在性的证明，也可用类似思路证得。

那么每个子区间中的误差项可近似表示为

$$\Delta \propto \frac{1}{N^{p+2}}.$$

#### 6. 将 $N$ 个子区间的误差项汇总

$$\varepsilon \propto \frac{N}{N^{p+2}} = \frac{1}{N^{p+2}}, \quad (12.152)$$

例如矩形法则 ( $p = 0$ ) 的近似误差  $\propto \frac{1}{N}$ , 梯形法则 ( $p = 1$ ) 的近似误差  $\propto \frac{1}{N^2}$ 。可见误差与  $N$  成反比:  $N$  越大, 划分的子区间数量越多, 近似误差越小。

举例, 假定我们要用求积法则近似计算

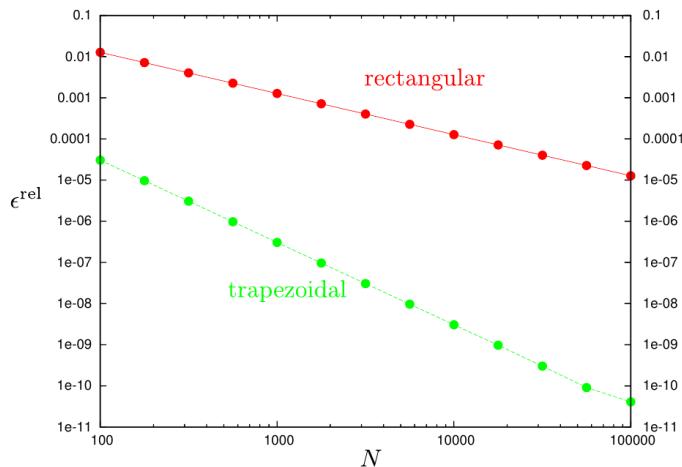
$$\mathcal{I} = \int_1^2 \log^2 x \, dx,$$

对应相对误差定义为  $\varepsilon^{rel}$

$$\varepsilon^{rel} \equiv \frac{|\mathcal{I}_N^{approx} - \mathcal{I}|}{\mathcal{I}}$$

应用矩形法则和梯形法则, 查看近似误差随着  $N$  增大的收敛情况如图12.F.5, 可以看出

图 12.F.5: 比较矩形法则和梯形法则下, 近似误差随  $N$  值增大的收敛情况



$$\text{原求积方程 } \int_1^2 \log^2 x \, dx.$$

这两种方法代表的牛顿——寇特斯法则下误差项随着  $N$  而作线性收敛: 随着  $N$  增大 1,000 倍, 矩形法则误差降低约 1,000 倍,  $\Delta\varepsilon^{rect} \propto \frac{1}{\Delta N}$ ; 梯形法则降低约 1,000,000 倍,  $\Delta\varepsilon^{trap} \propto \frac{1}{(\Delta N)^2}$ 。梯度法则优于矩形法则。

但图12.F.5也揭示出牛顿——寇特斯法则的不足: 即便是对于  $\log^2 x$  这种很平滑的方程, 梯形法则也需要大约 1,000 个样本, 才能将总体误差控制在  $10^{-6}$  的水平上(为了达到类似的精度, 矩形法则甚至需要  $10^6$  个样本)。从计算成本来看, 牛顿——寇特斯法则恐怕不是最理想的方案, 需要加以改进。如下文介绍的克伦肖——柯蒂斯法则, 可以用少得多的样本量达到同样  $10^{-6}$  级别的误差水平, 哪怕所处理的原方程更加复杂。我们做了一个小的数值计算实验来说明这个问题, 见表12.G.1。

## 12.F.2 几个小技巧

以牛顿——寇特斯法则为例, 介绍几个小技巧。它们对其他如克伦肖——柯蒂斯法则也适用。

### 定区间单位的换算

对于方程  $\int_a^b f(t) dt$  的求积问题，常常需要将区间从  $[a, b]$  转换为  $[-1, 1]$ 。可以设一个新的方程  $g(u), u \in [-1, 1]$ ，满足

$$g(u) = f\left[a + \frac{b-a}{2}(u+1)\right], \quad u \in [-1, 1], \quad (12.153)$$

那么

$$\int_a^b f(t) dt \Leftrightarrow \int_{-1}^1 g(u) du.$$

### 无限区间求积

有时会遇到求含有无限区间的积分问题如

$$\int_0^\infty f(x) dx,$$

这需要我们先将区间  $[0, \infty)$  映射到一个有界区间内，如

$$x : [0, \infty) \mapsto u : [0, 1],$$

然后再应用求解定积分的求积法则。映射的方法有很多种，其中之一是设

$$x \equiv \frac{u}{1-u}, \Rightarrow dx = \frac{du}{(1-u)^2},$$

求积问题因此转换为

$$\int_0^\infty f(x) dx = \int_0^1 \frac{1}{(1-u)^2} f\left(\frac{u}{1-u}\right) du.$$

需要指出的是，上式在转换之后，含有一个奇异点 (singularity): $\lim_{u \rightarrow 1} f(x) \rightarrow \infty$ 。然而若是在前提假设中假定  $f(x)$  在  $x \rightarrow \infty$  处消失 (这个假定是为了确保积分  $\int_0^\infty f(x) dx$  收敛)，则这一奇异点也就不存在了。

### 12.F.3 求积中的可积奇异点问题

除了上节提到的情况之外，还应该注意到，有些奇异点是可积的 (integrable singular points)，例如这个求积问题

$$\mathcal{I} = \int_0^1 \frac{\exp(x)}{\sqrt{x}} dx, \quad (12.154)$$

方程在区间  $(0, 1]$  内定义良好，除了起始点 0 的情况，称之为可积奇异点<sup>35</sup>。存在可积奇异点的积分方程并不是全都不能求解，针对具体问题的不同，有一些求解技巧可供选择。试举例如下。

#### 提取奇异点

有时可以在积分中把奇异点单独提取出来做求积运算，如 (12.154) 可以改写为

$$\mathcal{I} = \underbrace{\int_0^1 \frac{1}{\sqrt{x}} dx}_{=: \mathcal{I}_1} + \underbrace{\int_0^1 \frac{\exp(x) - 1}{\sqrt{x}} dx}_{=: \mathcal{I}_2}.$$

<sup>35</sup>注意区分可积和不可积奇异点。后者如积分方程  $\int_0^1 \frac{\exp(x)}{x} dx$  的起始点 0 的情况，此时积分不存在，因此无法估计。

其中  $\mathcal{I}_1$  可以用解析法求得

$$\mathcal{I}_1 = \int_0^1 \frac{1}{\sqrt{x}} dx = |2\sqrt{x}|_0^1 = 2.$$

含有可积奇异点的  $\mathcal{I}_2$  无法用解析法求解，但在区间起始点 ( $= 0$ ) 处是非奇异的：这是由于，将被积方程沿着  $x \rightarrow 0$  做泰勒级数展开，可得

$$\begin{aligned} \frac{\exp(x) - 1}{\sqrt{x}} &\approx \frac{x - \frac{1}{2}x^2 + \frac{1}{6}x^3 + \dots}{\sqrt{x}} = x^{\frac{1}{2}} - \frac{1}{2}x^{\frac{3}{2}} + \frac{1}{6}x^{\frac{5}{2}} + \dots, \\ &\hookrightarrow \lim_{x \rightarrow 0} \frac{\exp(x) - 1}{\sqrt{x}} \rightarrow 0, \end{aligned}$$

因此可用常规求积法则进一步近似求解  $\mathcal{I}_2$ 。

### 奇异点消除

若被积方程的分母中含有可积奇异点，那么可以考虑用雅各比变换 (Jacobian transformation)。例如对(12.154)，设

$$u \equiv \sqrt{x} \Rightarrow du = \frac{dx}{2\sqrt{x}},$$

求积运算因此变为

$$\int_0^1 \frac{\exp(x)}{\sqrt{x}} dx = 2 \int_0^1 \exp(u^2) du,$$

新的积分方程中没有奇异点，因此可进一步用常规数值求积法则来近似。

### Epsilon 扩展

如果可积奇异点无法利用前述两种方法消除，我们可以尝试加入一个平滑参数  $\epsilon$ 。对应地，取极限  $\lim \epsilon \rightarrow 0$  以消除奇异点。如将(12.154)改写为

$$\mathcal{I}_\epsilon = \int_0^1 \frac{\exp(x)}{(x^2 + \epsilon^2)^{\frac{1}{4}}} dx.$$

对于有限值的  $\epsilon < \infty$ ，上式非奇异，并且  $\lim_{\epsilon \rightarrow 0} \mathcal{I}_\epsilon \rightarrow \mathcal{I}$ 。

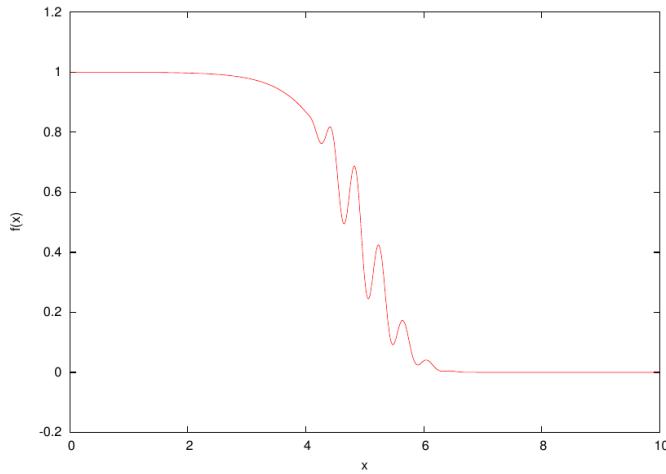
### 适应性求积

通常来说，被积方程在不同值域段的表现不同，如图 12.F.6 所示。如果我们应用梯形法则求解整个区间  $[0, 10]$  内的积分，那么在子区间  $[4, 6]$  中可能需要将宽度  $\Delta$  设的小一些，在  $[0, 4]$  和  $[6, 10]$  中将  $\Delta$  设的大一些。这就产生了适应性求积 (adaptive quadrature) 的概念：随着子区间中方程变化的程度不同，应用不同的求积法则来做近似，以求近似精度和计算成本的平衡。举例来说，此时的梯形法则变为

#### 12.F.4 梯形法则的近似误差

如前文所述，牛顿——寇特斯求积法则 (第12.F.1节) 的近似误差随着分段数  $N$  的衰减速度并不够令人满意；若是尝试利用更高阶多项式被积方程做近似，还会受到龙格现象的干扰。因此在实际应用中，常常并不会直接使用牛顿——柯蒂斯法则。一个更好的方案是克伦肖——柯蒂斯求积法则 (Clenshaw-Curtis Rule)。在了解傅里叶分析 (第12.G节) 的基本要点之后，我们可以对这个法则有大致介绍。首先来看梯形法则的近似误差分析。

图 12.F.6: 适应性求积



现在假定我们的目标是求解积分  $\mathcal{I} = \int_0^T f(t) dt$ 。根据梯形法则(12.150)可以做  $N$  个点的近似

$$I_N^{trap} = \frac{T}{N} \left\{ \underbrace{\left[ \frac{1}{2}f(0) + \frac{1}{2}f(T) \right]}_{=:A} + \underbrace{\sum_{n=1}^{N-1} f\left(\frac{nT}{N}\right)}_{=:B} \right\}, \quad (12.155)$$

近似的精度与  $N$  有关:  $N$  越大, 划分的点越多, 近似越精确。我们将精度定义为误差项  $\varepsilon_N^{trap}$

$$\varepsilon_N^{trap} = |\mathcal{I} - I_N^{trap}|.$$

现在引入傅里叶分析的概念。在实区间  $[0, T]$  中的方程  $f(t)$ , 可写为傅里叶空间中的综合表达式

$$f(t) = \sum_{\nu=-\infty}^{\infty} \tilde{f}_{\nu} \exp(i\nu\omega_0 t), \quad \omega_0 \equiv \frac{2\pi}{T}, \quad (12.156)$$

其中傅里叶级数的系数  $\tilde{f}_{\nu}$  值为

$$\tilde{f}_{\nu} = \frac{1}{T} \int_0^T f(t) \exp(-i\nu\omega_0 t) dt. \quad (12.157)$$

对  $\mathcal{I}$  做傅里叶变换可得

$$\begin{aligned} \mathcal{I} &= \int_0^T f(t) dt = \int_0^T \sum_{\nu=-\infty}^{\infty} \tilde{f}_{\nu} \exp(i\nu\omega_0 t) dt \\ &= \int_0^T \sum_{\nu=-\infty}^{\infty} \frac{1}{T} \int_0^T f(t) \exp(-i\nu\omega_0 t) dt \exp(i\nu\omega_0 t) dt \\ &= T \tilde{f}_0, \end{aligned}$$

即积分的值等于  $T$  个  $\nu = 0$  时的系数  $\tilde{f}_{\nu=0}$  值之和。

需要强调的是, 在对  $f(t)$  利用梯形法则做牛顿——寇特斯求积时, 我们并不知道系数  $\{\tilde{f}_{\nu}\}_{\nu}$  的值——若是知道, 也就不必再做傅里叶转换了。我们知道的是, 存在一个关于被积方程  $f(t)$  的傅里叶综合表达形式(12.156), 可供我们展开进一步的分析。

现在的任务是要将傅里叶综合式(12.156)代回梯形法则(12.155)。(12.155)RHS 分为  $A$  和  $B$  两部分,

- $\mathcal{A}$  等于  $t = 0$  时对(12.156)的计算：

$$\mathcal{A} = \sum_{m=-\infty}^{\infty} \tilde{f}_m,$$

分两种情况予以说明

- 当  $f(t)$  正好是一个  $T$ -周期方程时， $f(0) = f(T)$ ，显然这是正确的。
- 当  $f(t)$  的周期不等于  $T$  时， $\mathcal{A}$  就等于设  $t = 0$  时，对(12.156)的近似估计，见第12.G.5节。

- $\mathcal{B}$  的求和值，可以用一系列不同的  $t = \frac{nT}{N}$  来替代。

$$\mathcal{B} = \sum_{n=1}^{N-1} \sum_{m=-\infty}^{\infty} \tilde{f}_m \exp\left(im\omega_0 \frac{nT}{N}\right).$$

因此我们有

$$\begin{aligned} I_N^{trap} &= \frac{T}{N} \{ \mathcal{A} + \mathcal{B} \} = \frac{T}{N} \left\{ \sum_{m=-\infty}^{\infty} \tilde{f}_m + \sum_{n=1}^{N-1} \sum_{m=-\infty}^{\infty} \tilde{f}_m \exp\left(im\omega_0 \frac{nT}{N}\right) \right\} \\ &= \frac{T}{N} \sum_{n=0}^{N-1} \left\{ \sum_{\nu=-\infty}^{\infty} \tilde{f}_{\nu} \exp\left(i\nu\omega_0 \frac{nT}{N}\right) \right\} = \frac{T}{N} \sum_{n=0}^{N-1} \left\{ \sum_{\nu=-\infty}^{\infty} \tilde{f}_{\nu} \exp\left(2\pi i \frac{\nu n}{N}\right) \right\}, \end{aligned} \quad (12.158)$$

上式中引入一个新定义方程  $K_N(\nu)$  做进一步简化

$$I_N^{trap} = T \sum_{\nu=-\infty}^{\infty} \tilde{f}_{\nu} K_N(\nu), \quad (12.159)$$

$$\begin{aligned} K_N(\nu) &\equiv \frac{1}{N} \sum_{n=0}^{N-1} \exp\left(2\pi i \frac{\nu n}{N}\right) \\ &= \frac{1}{N} [1 + \xi + \xi^2 + \dots + \xi^{N-1}], \quad \xi \equiv \exp\left(2\pi i \frac{\nu}{N}\right), \end{aligned} \quad (12.160)$$

$K_N(\nu)$  的取值分为两种情况

- $\nu = 0$ ，或者  $\nu$  是一个  $N$  的整数倍是，有  $\xi = 1 \Rightarrow K_N(\nu) = 1$ ，
- $\nu \neq 0$  且  $\nu$  不是  $N$  的整数倍时，有  $\xi \neq 1$ ，以及

$$K_N(\nu) \approx \frac{1}{N} \left[ \frac{1 - \xi^N}{1 - \xi} \right] = \frac{1}{N} \frac{1 - \exp(2\pi i \nu)}{1 - \exp\left(\frac{2\pi i \nu}{N}\right)} = 0.$$

由此(12.159)进一步改写为

$$I_N^{trap} = T \sum_{\nu=-\infty}^{\infty} \tilde{f}_{\nu} K_N(\nu) = T \sum_{p=-\infty}^{\infty} \tilde{f}_{pN}, \quad p \in \mathcal{Z}.$$

可见对  $[0, T]$  区间中对周期方程求积问题  $\int f(t) dt$  做  $N$  个点的梯形法则近似，其值与傅里叶级数的系数  $\{\tilde{f}_{\nu}\}_{\nu}$  有关，但并不是全部系数，而仅仅是其中的一部分系数之和，具体说来是那些  $N$  的整数倍所对应的系数  $\sum_p \tilde{f}_{pN}$ ，它们在傅里叶域中对应的傅里叶系数，是指频率为  $N\omega_0$  整数倍的

若干条弦波曲线的傅里叶系数。尤其是，其中只有  $\nu = 0$  那条对应的项才是我们真正需要的；其余的部分都是误差项

$$I_N^{trap} = \underbrace{T \tilde{f}_{\nu=0}}_{\mathcal{I}} + T \sum_{p \neq 0} \tilde{f}_{pN}, \Leftrightarrow \varepsilon_N^{trap} \equiv |\mathcal{I} - I_N^{trap}|,$$

$$\varepsilon_N^{trap}$$

$\varepsilon_N^{trap}$  从而可以理解为第  $\pm N$  个，第  $\pm 2N$  个，…个傅里叶级数的系数之和

$$\varepsilon_N^{trap} \equiv |\mathcal{I} - I_N^{trap}| = \left| \sum_{p=-\infty, p \neq 0}^{\infty} \tilde{f}_{pN} \right|. \quad (12.161)$$

### 非周期方程的傅里叶级数收敛

同样地，分析即使进展到(12.161)这一步，我们还是不知道傅里叶级数系数  $\{\tilde{f}_{pN}\}$  的值，因而无法直接数值计算做近似求积。不过我们可以用傅里叶分析的平滑——衰减特性（第12.G.4节），来观察误差随着  $N$  的增大而衰减的速度有多快（收敛速度）。

随着被积方程  $f(t)$  是周期方程还是非周期方程，我们分两种情况来分析。首先考虑  $f(t)$  是平滑方程，但不满足  $f(0) = f(T)$  的情况。此时我们利用(12.157)计算的傅里叶级数系数  $\tilde{f}_\nu$  是一个不连续方程。

为了在这种情况下做收敛分析，一个较原始的方案是<sup>36</sup>：将  $[0, T]$  中的这一段  $f(t)$  截取出来，以  $T$  周期为单位向前向后做重复扩展，所形成的一个新周期方程定义为  $f^{per}(t)$ ，作为  $f(t)$  的  $T$ -周期扩展，见图12.F.7。

由佩利——维纳诸定理 (Theorem 12.62) 可见，不连续方程  $f(t)$  的傅里叶系数的强度 (magnitude)  $\tilde{f}_\nu$ ，以等比于  $\frac{1}{\nu}$  的速度衰减。那么在理想状况下，根据(12.161)，梯形法则的近似误差  $\varepsilon_N^{trap}$  也应该以等比于  $\frac{1}{N}$  的速度衰减。然而实际上(12.161)中  $\nu \in \mathbb{Z}$  可以是正值也可以是负值，可能导致求和后的  $\tilde{f}_\nu$  被抵消掉。为了解决这个问题，我们在傅里叶系数求和式(12.157)中区分  $\nu > 0$  和  $\nu < 0$  的情况，作分部积分 (Boyd, 2001, Sec 2.9)：

$$\begin{aligned} \tilde{f}_\nu &= \frac{1}{T} \int_0^T f(t) \exp(-i\nu\omega_0 t) dt \\ &= \frac{1}{T} \left\{ \underbrace{\frac{1}{i\nu\omega_0} \left[ f(t) \exp(-i\nu\omega_0 t) \right]_0^T}_{=: \mathcal{A}} + \underbrace{\frac{1}{i\nu\omega_0} \int_0^T f'(t) \exp(-i\nu\omega_0 t) dt}_{=: \mathcal{B}} \right\}, \end{aligned}$$

其中

$$\mathcal{A} = \left| \cdot \right|_0^T = f(T) \underbrace{\exp(-i\nu\omega_0 T)}_{=: 1} - f(0) \underbrace{\exp(-i\nu\omega_0 0)}_{=: 1} = f(T) - f(0) \equiv \Delta f.$$

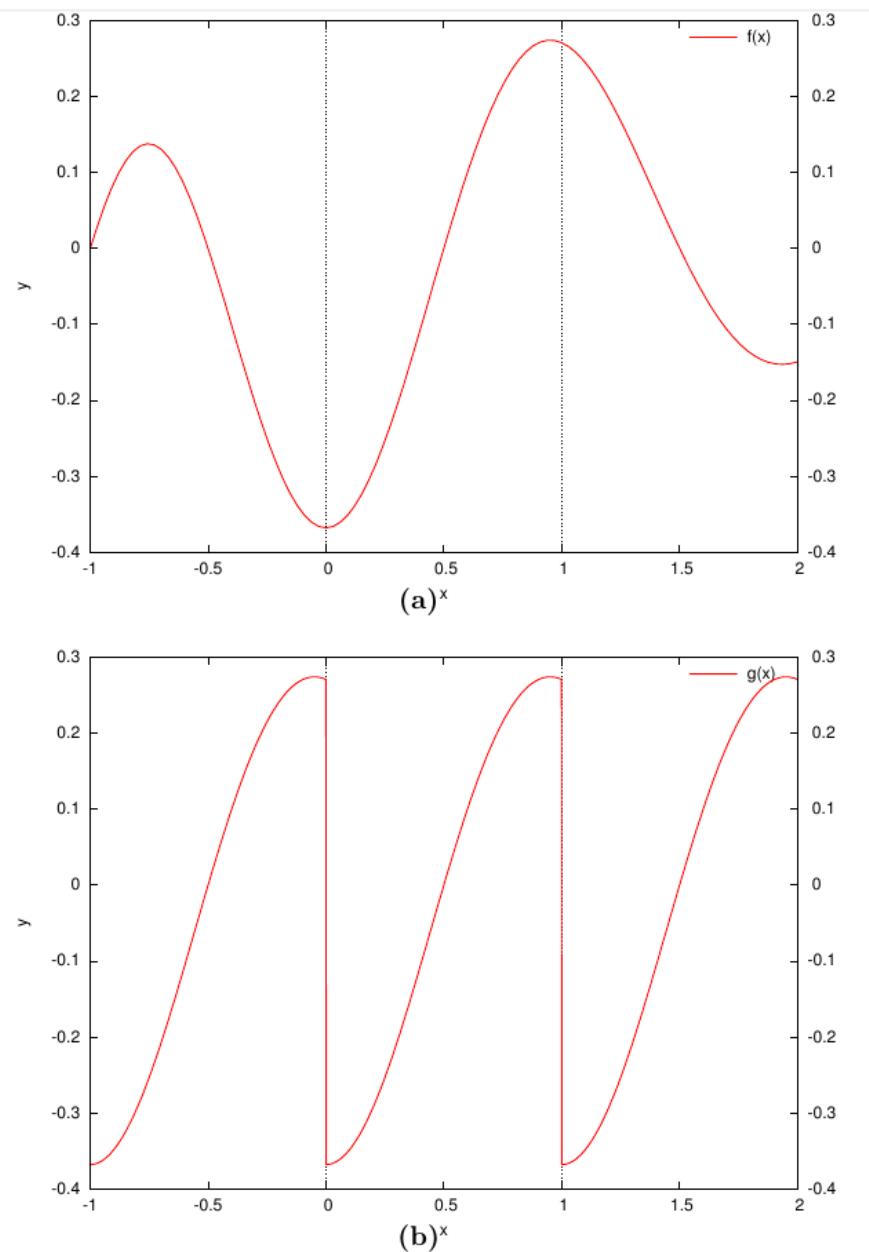
那么我们有

$$\tilde{f}_\nu = \underbrace{\frac{\Delta f}{i\nu\omega_0 T}}_{=: \frac{C_1}{\nu}} + \underbrace{\frac{1}{i\nu\omega_0} \int_0^T f'(t) \exp(-i\nu\omega_0 t) dt}_{=: \frac{C_2}{\nu^2}}.$$

对上式再次分部积分，采用类似的思路，有

$$\tilde{f}_\nu = \underbrace{\frac{\Delta f}{i\nu\omega_0 T}}_{=: \frac{C_1}{\nu}} + \underbrace{\frac{\Delta f'}{(i\nu\omega_0)^2}}_{=: \frac{C_2}{\nu^2}} + \underbrace{\frac{1}{(i\nu\omega_0)^2} \int_0^T f''(t) \exp(-i\nu\omega_0 t) dt}_{=: \frac{C_3}{\nu^3}},$$

<sup>36</sup>注意这里讨论的  $T$ -周期扩展方案较为直观易于理解，但精度上有所欠缺。在下文中我们将对  $f(t)$  做更精致的周期扩展见第12.F.5节。两种方案的比较见第294页脚注。

图 12.F.7: 非周期方程的  $T$ -周期扩展  $t \in [0, 1]$ 

其中  $\Delta f' \equiv f'(T) - f'(t)$ 。

继续进行下去可以看到, 傅里叶系数  $\tilde{f}_\nu$  是一与  $\nu$  的幂次的倒数之和<sup>37</sup>

$$\tilde{f}_\nu = \frac{C_1}{\nu} + \frac{C_2}{\nu^2} + \frac{C_3}{\nu^3} + \frac{C_4}{\nu^4} \dots \quad (12.162)$$

这样一来, 若是取与正值  $\nu$  相对应的负值  $-\nu$  两两配对, 将  $\tilde{f}_\nu$  和  $\tilde{f}_{-\nu}$  相加, 则 RHS 中奇数项相互抵消, 偶数项部分保留下来

$$\tilde{f}_\nu + \tilde{f}_{-\nu} = \frac{C'_2}{\nu^2} + \frac{C'_4}{\nu^4} + \dots \quad (12.163)$$

其中  $C'_2 \equiv 2C_2$ , 以此类推 (常数项不重要, 重要的是分母随着  $\nu$  值变化而表现出的渐进特性)。

将(12.163)代回(12.161)可得

$$\begin{aligned} \varepsilon_N^{trap} &= \left| \sum_{p=-\infty, p \neq 0}^{\infty} \tilde{f}_{pN} \right| = \left| \sum_{p=1}^{\infty} (\tilde{f}_{pN} + \tilde{f}_{-pN}) \right| \\ &= \frac{C'_2}{N^2} \sum_{p=1}^{\infty} \frac{1}{p^2} = \frac{C'_2 \pi^2}{N^2 \cdot 6}, \end{aligned}$$

忽略掉常数项可得, 梯形法则下对非周期方程  $f(t)$  求积的近似误差  $\varepsilon_N^{trap} \sim \frac{1}{N^2}$ , 即等比收敛至  $\frac{1}{N^2}$ 。

### 周期方程的傅里叶级数收敛

来考虑  $f(t) \in C^\infty$  是平滑的  $T$ -周期方程的情况。此时

$$f(0) = f(T), f'(0) = f'(T), f''(0) = f''(T), \dots$$

根据佩利——维纳诸定理 (Theorem 12.62), 这类方程在傅里叶域中, 傅里叶系数的强度 (magnitude)  $|\tilde{f}_\nu|$  的衰减速度比任何多项式在  $\nu$  的衰减速度都快, 例如  $|\tilde{f}_N| \sim \exp(-\alpha|N|)$ 。此时误差项可以写为

$$\varepsilon_N^{trap} = \left| \sum_{p \neq 0} \tilde{f}_{pN} \right| = \sum_{p \neq 0} \exp(-\alpha N|p|) \sim \exp(-\alpha N),$$

即周期方程  $f(t)$  的梯形法则近似误差项等比例收敛至  $\exp(-\alpha N)$ 。

### 12.F.5 克伦肖——柯蒂斯求积法则

上节分析了为什么梯形法则下的傅里叶级数近似误差可以很快速的指数收敛 (如果原方程是周期方程), 或是以较慢的速度收敛 (如果原方程是非周期方程)。对于非周期方程的情况, 能不能让近似误差收敛的速度加快? 克伦肖——柯蒂斯求积法则 (Clenshaw-Curtis Rule)的作用就在于, 将非周期方程转换为周期方程的形式, 从而加快梯形法则误差项的收敛速度。大致说来分为三个步骤

- 将非周期方程转换为周期方程,
- 将求积式改写为傅里叶系数和,
- 应用求积法则做近似,

以下分别作以说明。

<sup>37</sup>巧合的是, 这恰好成为佩利——维纳诸定理 (Theorem 12.62) 的一种证明: 如果  $\Delta f = 0$  但  $\Delta f' \neq 0$ , 那么对  $f(t)$  的  $T$ -周期扩展中就同事含有连续的方程和不连续的倒数, 那么其对应的不连续点就会出现在  $p = 1$  处; 进而由上面的扩展可得, 随着  $|\nu| \rightarrow \infty$ , 有  $|\tilde{f}_\nu| \sim \frac{1}{\nu^2}$ 。

### 将非周期方程 $f(t)$ 转换为周期方程 $g(\theta)$

假定我们要在  $[-1, 1]$  区间中数值求积

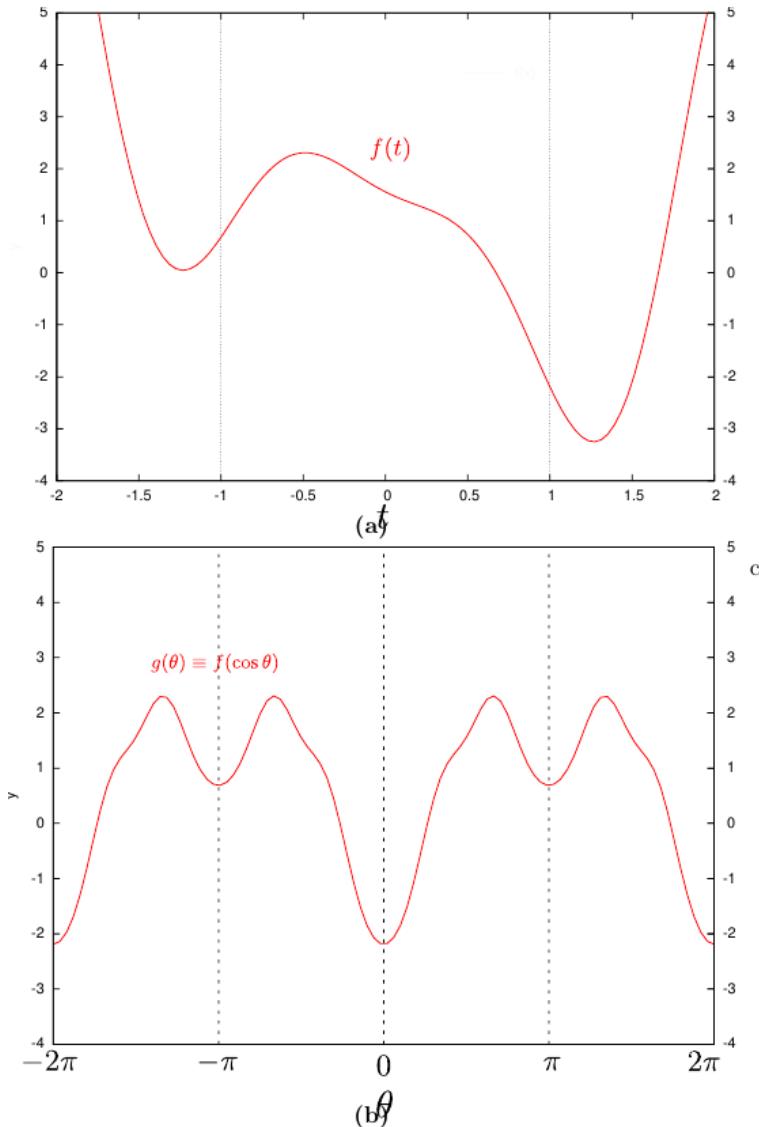
$$\mathcal{I} = \int_{-1}^1 f(t) dt, \quad (12.164)$$

如果原空间是  $[a, b]$  而不是  $[-1, 1]$ , 可以参考第12.F.2的转换方法。

$[-1, 1]$  区间恰好是  $\cos \theta, \theta \in [0, \pi]$  的取值范围, 因此可做参数化处理, 设  $t = \cos \theta$ , 定义新方程如图12.F.8所示。

$$g(\theta) \equiv f(\cos \theta). \quad (12.165)$$

图 12.F.8:  $f(t)$  和  $g(\theta)$



$g(\theta)$  有如下特征:

- 是个周期方程, 周期  $T = 2\pi$ ,
- 是个偶方程, 满足  $g(-\theta) = g(\theta)$ ,

- $g(\theta)$  随着  $\theta$  从 0 到  $\pi$  的路径, 勾勒出  $f(t)$  随着  $t$  从 1 到 -1 的变化路径,
- $f(t)$  在实域  $(-\infty, 1)$  和  $(1, \infty)$  中的变化路径, 无法由  $g(\theta)$  表现出来。

$g(\theta)$  的傅里叶级数表现形式包含一组弦波之和, 每一个弦波的频率都是基准频率  $\omega_0 = \frac{2\pi}{T} = 1$  的整数倍

$$\begin{aligned} g(\theta) &= \sum_{\nu=-\infty}^{\infty} \tilde{g}_{\nu} \exp(i\nu\theta), \\ \tilde{g}_{\nu} &= \frac{1}{2\pi} \int_0^{2\pi} g(\theta) \exp(-i\nu\theta) d\theta. \end{aligned} \quad (12.166)$$

或者  $g(\theta)$  也可以用傅里叶余弦级数的形式表示为

$$g(\theta) = \frac{\tilde{a}_0}{2} + \sum_{\nu=1}^{\infty} \tilde{a}_{\nu} \cos(\nu\theta). \quad (12.167)$$

两组傅里叶系数  $\{\tilde{a}_{\nu}\}$ ,  $\{\tilde{g}_{\nu}\}$  之间的关系为

$$\tilde{a}_0 = 2\tilde{g}_0, \quad \tilde{a}_{\nu} = 2(\tilde{g}_{\nu} + \tilde{g}_{-\nu}) = 2\tilde{g}_{\nu}, \quad (12.168)$$

这是由于  $g(\theta) = g(-\theta) \Rightarrow \tilde{g}_{\nu} = \tilde{g}_{-\nu}$ 。

此外  $\{\tilde{a}_{\nu}\}$  可以表示为如下形式

$$\tilde{a}_{\nu} = \frac{1}{\pi} \int_0^{2\pi} g(\theta) \cos(\nu\theta) d\theta. \quad (12.169)$$

由(12.169)可见, 每个系数  $\tilde{a}_{\nu}$  都是一个被积方程  $g(\theta) \cos(\nu\theta)$  关于其域  $\theta \in [0, 2\pi]$  的积分。由于在这个域中,  $g(\theta)$  和  $\cos(\nu\theta)$  都是偶函数, 那么  $g(\theta) \cos(\nu\theta)$  也是一个偶函数。这意味着积分(12.169)可由一个简单的  $N$  个点的梯形法则来做近似求解, 并且近似误差随着  $N$  值的增大而呈指数速度衰减。

### 原域求积问题改写为傅里叶域求和问题

目标: 将关于  $f(t)$  的求积问题(12.164)改写为关于  $g(\theta)$  的求积问题, 用  $t = \cos \theta$  替换  $t$  可得

$$\mathcal{I} = \int_{-1}^1 f(t) dt = \int_0^{\pi} \underbrace{f(\cos \theta)}_{=g(\theta)} \underbrace{d\cos \theta}_{=\sin \theta d\theta} = \int_0^{\pi} g(\theta) \sin \theta d\theta. \quad (12.170)$$

尽管被积方程  $g(\theta) \sin \theta$  是周期方程, 但积分区间  $[0, \pi]$  只是完整周期  $[0, 2\pi]$  的一半, 因此我们不能直接对(12.170)应用梯形法则来做 (指数衰减的) 求积。为了解决这个问题, 我们将  $g(\theta)$  的傅里叶余弦级数表达式(12.167)代入(12.170)有

$$\begin{aligned} \mathcal{I} &= \int_0^{\pi} g(\theta) \sin \theta d\theta = \int_0^{\pi} \left[ \frac{\tilde{a}_0}{2} + \sum_{\nu=1}^{\infty} \tilde{a}_{\nu} \cos(\nu\theta) \right] \sin \theta d\theta \\ &= \frac{\tilde{a}_0}{2} \underbrace{\int_0^{\pi} \sin \theta d\theta}_{=2} + \sum_{\nu=1}^{\infty} \tilde{a}_{\nu} \underbrace{\int_0^{\pi} \sin \theta \cos(\nu\theta) d\theta}_{=\frac{1+\cos(\nu\pi)}{1-\nu^2}} \\ &= \tilde{a}_0 + \sum_{\nu=1}^{\infty} \tilde{a}_{\nu} \frac{1 + \cos(\nu\pi)}{1 - \nu^2}, \end{aligned} \quad (12.171)$$

其中

$$\cos(\nu\pi) = (-1)^\nu = \begin{cases} -1 & \nu \text{ 是奇数} \\ 1 & \nu \text{ 是偶数} \end{cases}$$

代回上式得

$$\mathcal{I} = \tilde{a}_0 + \sum_{\nu=1, \nu \text{ 是偶数}}^{\infty} \frac{2\tilde{a}_\nu}{1-\nu^2}, \quad (12.172)$$

也即用  $g(\theta)$  的傅里叶余弦系数  $\{\tilde{a}_\nu\}$  来描述  $f(t)$ , 进而计算求积问题。如果假定  $f(t)$  平滑, 那么  $g(\theta)$  继承这一特性, 是个平滑的周期方程, 其傅里叶余弦系数  $\tilde{a}_\nu$  随着  $\nu$  值而衰减的速度比任何多项式都快<sup>38</sup>。

在实际数值计算过程中, 很难考虑到  $\nu \rightarrow \infty$  的情况。对此, 常常要对(12.172)中的求和式做一个  $\nu$  等于某个有限值比如  $N$  的截断 (truncation), 从而最终得到对求积问题(12.164)的克伦肖——柯蒂斯法则

$$\mathcal{I} \approx I_N^{CC} \equiv \tilde{a}_0 + \sum_{\nu=1, \text{偶数}}^N \frac{2\tilde{a}_\nu}{1-\nu^2}, \quad (12.173)$$

或者用向量形式, 表示为权重向量  $W$  和傅里叶级数系数  $\tilde{a}$  的点乘

$$\mathcal{I} = \mathbf{W} \cdot \tilde{\mathbf{a}},$$

$$\tilde{\mathbf{a}} = \begin{pmatrix} \tilde{a}_0 \\ \tilde{a}_1 \\ \tilde{a}_2 \\ \vdots \\ \tilde{a}_N \end{pmatrix}, \quad \mathbf{W} = \begin{pmatrix} W_0 \\ W_1 \\ W_2 \\ \vdots \\ W_N \end{pmatrix}, \quad W_\nu \equiv \begin{cases} 1 & \nu = 0, \\ \frac{2}{1-\nu^2} & \nu \neq 0, \nu \text{ 是偶数}, \\ 0 & \nu \text{ 是奇数}. \end{cases} \quad (12.174)$$

### 应用求积法则做近似

在求得(12.173)的基础上, 需要计算傅里叶级数的系数  $\{\tilde{a}_\nu\}_\nu$ 。常见的思路是利用  $N$  个点的梯形法则, 对傅里叶级数积分式(12.169)作近似<sup>39</sup>。

由于(12.169)也是一个沿着周期区间对周期方程求积的计算, 那么其近似误差同样也随着  $N$  的增大而等比指数衰减。

式(12.169)在对  $\tilde{a}_\nu$  做梯形法则近似的过程中, 取样  $N$  个  $\nu$  值的点, 在原空间中的  $f(\cos \theta)$  和傅里叶空间中的  $g(\theta)$  一一对应。进而在(12.173)中, 便是将这  $N$  个取样点对应的值加权求和, 即也是一个  $N$  个点的求积法则。

### 求积法则再探讨

回顾一下上节介绍的克伦肖——柯蒂斯法则, 利用该法则求解积分问题(12.173), 其中

- $\mathcal{I}$  是一个关于  $g(\theta)$  在傅里叶域  $[0, 2\pi]$  中取样后求积的方程,
- $\{\tilde{a}_\nu\}_\nu$  是一组关于方程  $g(\theta) \equiv f(\cos \theta)$  的傅里叶系数, 可以根据(12.169)利用  $N$  个点的梯形法则近似计算得出, 也即  $\tilde{a}_\nu$  是一个关于  $g(\theta)$  的方程。

<sup>38</sup>与之相比, 图12.F.7的  $T$ -周期扩展方案则不能保证近似式一定收敛, 这是由于它在每个周期的最末一点不连续, 因此其傅里叶系数只能以等比于  $\nu^{-1}$  的代数速度收敛。

<sup>39</sup>事实上在这里我们可以取  $M$ -点梯形法则来近似(12.169),  $M \neq N$  并且和  $N$  没有明显关联。此时近似误差的衰减速度是  $\exp(-\#M)$ 。另一方面, (12.173)的衰减速度是  $\exp(-\#N)$ 。最终我们取  $\min\{\exp(-\#M), \exp(-\#N)\}$  作为全部近似求积的总误差收敛速度。

本节我们进一步讨论如何

- 将  $\tilde{a}_\nu$  转换为与样本方程  $f(t)$  有关的方程，
- 将  $\mathcal{I}$  转换为与样本方程  $f(t)$  有关的方程。

先来看第一个问题。回顾(12.165) (12.166) (12.167) (12.168) (12.169)可见， $\tilde{a}_\nu$  是个在一个周期内  $[0, 2\pi]$ ，对周期方程  $f(\cos \theta) \cos(\nu\theta)$  求积的运算

$$\tilde{a}_\nu = \frac{1}{2\pi} \int_0^{2\pi} f(\cos \theta) \cos(\nu\theta) d\theta,$$

那么可以用  $N$  个点的梯形法则对上式做近似估计，估计的误差随着取样点数  $N$  的增加而呈指数速度等比衰减<sup>40</sup>。由于被积方程  $f(\cos \theta) \cos(\nu\theta)$  是个偶方程，我们可以取原周期的一半  $[0, \pi]$  求积，并将常数乘以 2

$$\tilde{a}_\nu = \frac{2}{\pi} \int_0^\pi f(\cos \theta) \cos(\nu\theta) d\theta.$$

对上式应用  $N$  点梯形法则

$$\tilde{a}_\nu = \frac{2}{N} \sum_{n=0}^N f(\cos \theta_n) \cos(\nu\theta_n), \quad \theta_n \equiv \frac{n\pi}{N}, \quad (12.175)$$

其中  $\sum_{n=0}^N$  是一个简化表述，表示首项  $n = 0$  和末项  $n = N$  要在前面乘以  $1/2$ 。

(12.175)也可以进一步简化表述为

$$\begin{aligned} \tilde{a}_\nu &= \sum_n \Lambda_{\nu n} f(t_n), \\ t_n &\equiv \cos\left(\frac{\pi n}{N}\right), \quad \Lambda_{\nu n} \equiv \begin{cases} \frac{1}{N} \cos\left(\frac{n\nu\pi}{N}\right), & n = 0 \\ \frac{2}{N} \cos\left(\frac{n\nu\pi}{N}\right), & n = 1, 2, \dots, N-1 \\ \frac{1}{N} \cos\left(\frac{n\nu\pi}{N}\right), & n = N, \end{cases} \end{aligned} \quad (12.176)$$

其中  $t_n$  可称为切比雪夫点 (Chebyshev points)。在(12.176)中，用到同一组二次项值  $\cos\left(\frac{n\nu\pi}{N}\right)$  来对  $f(t)$  做取样，计算不同的系数值  $\tilde{a}_\nu$ 。这意味着，可以将(12.176)改写为以下矩阵形式

$$\tilde{\mathbf{a}} = \mathbf{\Lambda} \cdot \mathbf{f}, \quad (12.177)$$

其中

- $\tilde{\mathbf{a}}$  是  $(N+1) \times 1$  的傅里叶系数向量

$$\tilde{\mathbf{a}} = [\tilde{a}_0, \tilde{a}_1, \dots, \tilde{a}_N]^\top.$$

- $\mathbf{f}$  是个  $(N+1) \times 1$  的抽样值向量

$$\mathbf{f} = [f(t_0), f(t_1), \dots, f(t_N)]^\top, \quad t_n \equiv \frac{n\pi}{N}.$$

<sup>40</sup> 通常来讲，误差项呈指数速度快速收敛的前提是原方程  $f(t)$  平滑：克伦肖——柯蒂斯法则的周期化算法(12.165)消除了时间周期最末端一点的不连续性（从而使得这种算法优于前文提出的原始周期化方案），但无法消除  $f(t)$  在周期内其他位置可能存在的不连续性。

从技术上来讲， $f(t)$  的平滑只能确保误差项的收敛速度超过  $N$  值对应的任意多项式。在必要的前提下，确保误差项指数收敛的更严格前提是假定  $f(t)$  时解析的 (analytic)，即  $f$  在所关注域内的任意一点上，任意级泰勒级数均收敛。

在数值近似的实际操作过程中，平滑方程假定和解析方程假定的区别并不大，从而我们可以在一般意义上仅假定  $f(t)$  是平滑的，就可以确保傅里叶级数的近似误差呈指数收敛。

- $\Lambda$  是一个  $(N+1) \times (N+1)$  矩阵，含有元素  $\Lambda_{\nu n}$

$$\Lambda = \frac{2}{N} \begin{pmatrix} \frac{1}{2} & 1 & 1 & 1 & \dots & \frac{1}{2} \\ \frac{1}{2} & \cos \frac{\pi}{N} & \cos \frac{2\pi}{N} & \cos \frac{3\pi}{N} & \dots & \frac{1}{2} \cos \pi \\ \frac{1}{2} & \cos \frac{2\pi}{N} & \cos \frac{4\pi}{N} & \cos \frac{6\pi}{N} & \dots & \frac{1}{2} \cos 2\pi \\ \frac{1}{2} & \cos \frac{3\pi}{N} & \cos \frac{6\pi}{N} & \cos \frac{9\pi}{N} & \dots & \frac{1}{2} \cos 3\pi \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2} & \cos \pi & \cos 2\pi & \cos 3\pi & \dots & \frac{1}{2} \cos N\pi \end{pmatrix} \quad (12.178)$$

现在来看第二个问题，用抽样样本值  $f$  来表示  $\mathcal{I}$ 。已知(12.174)将求积问题  $\mathcal{I}$  表示为域傅里叶系数  $\tilde{a}_\nu$  有关的方程，(12.177)将  $\tilde{a}_\nu$  表示为与抽样值  $f$  有关的方程，抽样点为切比雪夫点  $t_N$ 。那么两式联立可得用  $f$  表示的  $\mathcal{I}$ ，是一个求积法则

$$\mathcal{I} = \mathbf{W}^T \mathbf{\Lambda} f,$$

或者进一步简化

$$\mathcal{I} = \mathbf{w} \cdot \mathbf{f}, \quad \mathbf{w} \equiv \lambda^T \mathbf{W}, \quad (12.179)$$

$\mathbf{w}$  是一个  $(N+1) \times 1$  的权重向量。

(12.179)等价于

$$\mathcal{I} = \int_{-1}^1 f(t) dt = \sum_{n=0}^N w_n f(t_n), \quad (12.180)$$

这成为一个  $(N+1)$  个点的求积法则，利用  $N+1$  个抽样  $f$  的加权求和来近似  $\mathcal{I}$ 。其中权重向量  $\mathbf{w}$  的计算步骤为

- 利用(12.174)计算  $\mathbf{W}$ ，
- 利用(12.178)计算  $\mathbf{\Lambda}$ ，
- 利用  $\mathbf{w} = \mathbf{\Lambda}^T \mathbf{W}$  算得  $\mathbf{w}$ 。

当  $N$  是偶数时，克伦肖——柯蒂斯求积法则中权重  $\mathbf{w}$  的显性形式

$$w_m = \begin{cases} \frac{1}{N^2-1} & m = 0 \\ \frac{2}{N} \left\{ 1 + \left[ \sum_{n=1}^{\frac{N}{2}-1} \left( \frac{2}{1-4n^2} \right) \cos \left( \frac{2mn\pi}{N} \right) \right] + \frac{\cos m\pi}{1-N^2} \right\} & m = 1, \dots, N-1 \\ \frac{1}{N^2-1} & m = N \end{cases} \quad (12.181)$$

## 12.F.6 程序实现

举例，求积问题

$$\mathcal{I} = \int_{-1}^1 f(x) dx, \quad f(x) = x \cdot \sin(2 \cdot \exp(2 \cdot \sin(2 \cdot \exp(2x))))$$

Mathematica 可用内建函数 NIntegrate[ ] 来计算求解，如

```
In[1]:= func[x_] := x*Sin[2 Exp[2 Sin[2 Exp[2 x]]]]
In[2]:= NIntegrate[func[x], x, -1, 1, Method -> "ClenshawCurtisRule"]
Out[2]= 0.336733
```

Matlab 中并没有直接内建的用于克伦肖——柯蒂斯数值积分程序。我们用 `chebfun` 的程序来计算，代码如下。

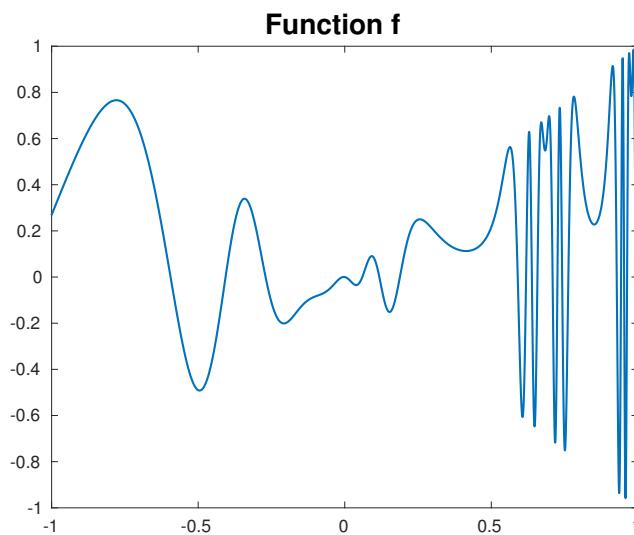
```

1 % 假定要对方程  $f(x)$  在  $[-1, 1]$  求积
2 x = chebfun('x');
3 f = @(x) x.*sin(2*exp(2*sin(2*exp(2*x)))); % 定义函数
4 % 输入 help @查询命令
5 % .* 表示按元素相乘 * 表示矩阵乘
6 fc = chebfun(f);
7 LW = 'linewidth';
8 FS = 'fontsize';
9 MS = 'markersize';
10 figure, plot(fc, LW, 1.2)
11 title('Function f', FS, 16)

```

输出  $f(x), x \in [-1, 1]$  见图12.F.9。

图 12.F.9:  $f(x), x \in [-1, 1]$



```

1 % 输出积分的数值近似值
2 Ichebfun = sum(fc)

```

返回值 0.336732834781728。

作克伦肖——柯蒂斯法则近似，和 Gauss-Legendre 近似 (Glaser et al., 2007)

```

1 % 求积近似过程中，使用了多少插值点
2 Npts = length(fc)
3
4 [s,w] = chebpts(Npts); % 使用 chebshев points
5 % s 是在  $[-1, 1]$  之间插值的若干点 (N)
6 % w 是对应的权重矩阵

```

```

7 % 两个矩阵相乘，得到求积的值，命名为 Iclenshawcurtis，等同于 sum(fc) 的命
    令
8 Iclenshawcurtis = w * f(s)
9
10 % 也可以尝试用 Legendre points (gauss-legendre 求积算法)
11 [s,w] = legpts(Npts);
12 Igauss = w * f(s)

```

绘制图形，描述将两种近似方法测得的近似误差，随着  $N$  值增大的指数衰减情况，输出结果见图12.F.10。

```

1 % 绘出采用 Gauss-legendre 算法，误差项的收敛情况。semilogy 与 plot 命令类
    似，除了 semilogy 输出 log 形式的值
2 semilogy(NN, err, '.-' , LW, 1, MS, 16), grid on
3 ylim([1e-18 1])
4 xlabel('Npts', FS, 12), ylabel('Error', FS, 12)
5 title('Gauss quadrature convergence', FS, 16), toc % toc 输出计算时间
6
7 % 现在将 Gauss-Legendre 算法和 Clenshaw-Curtis 算法的误差收敛情况都绘制出来
8 hold on, tic, err = [];
9 for Npts = NN
10     [s,w] = chebpts(Npts);
11     Iclenshawcurtis = w * f(s);
12     err = [err abs(Iclenshawcurtis-Ichebfun)];
13 end
14 semilogy(NN, err, '.-r' , LW, 1, MS, 16)
15 title('Gauss-Legendre and Clenshaw-Curtis', FS, 16)
16 legend('Gauss-Legendre', 'Clenshaw-Curtis', 'location', 'southwest'),
    toc

```

## 12.G 傅里叶分析

我们可以这样理解分析 (analysis): 将研究对象分解成小块来分别研究。那么我们可以这样理解傅里叶分析 (Fourier analysis): 将研究对象 (方程) 分解为小块，即不同的弦波 (sinosoids)，每个小块都以有限的速度变化。例如这个方程

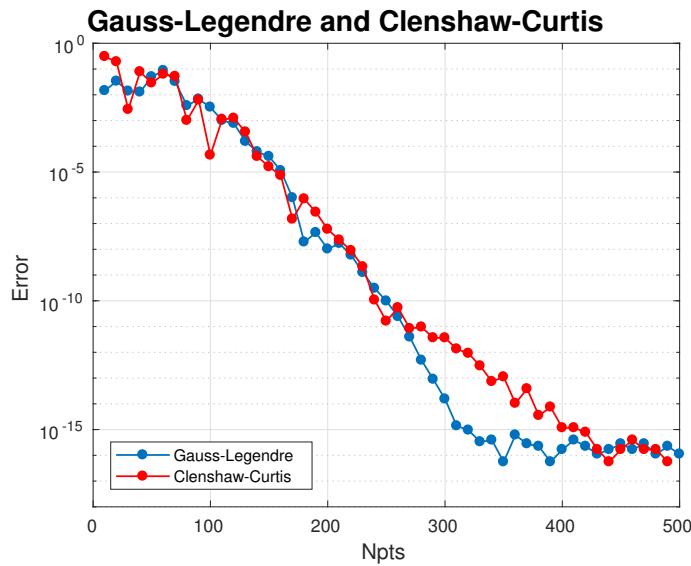
$$f(t) = 3 \cos 2\pi t + 19 \sin 4\pi t - 0.14 \cos 7\pi t,$$

可以分解为

$$f(t) = 3A + 19B - 0.14C,$$

其中  $A, B, C$  分别表示角频率 (angular frequency) 为  $2\pi, 4\pi, 7\pi$  的弦波。

图 12.F.10: 求积法则误差的衰减



但对于这样的方程

$$f(t) = \exp(-\alpha|t|),$$

$$f(t) = \begin{cases} 1, & |t| < 1 \\ 2, & |t| > 1 \end{cases},$$

甚至更复杂一些的方程，该如何做傅里叶分析？

在开始正式介绍之前，有必要做一些简单的概念界定，见表12.G.1。

表 12.G.1: 傅里叶分析的常见概念界定

	连续方程 $f(t)$	离散方程 $f_n \equiv f(n\Delta t), n \in \mathcal{Z}$
无限域 $-\infty < t < \infty$	傅里叶变换 (Fourier transform)	半离散傅里叶变换 (Semidiscrete Fourier transform)
有限域 $-\frac{T}{2} < t < \frac{T}{2}$	傅里叶级数 (Fourier series)	离散傅里叶变换 (Discrete Fourier transform)

## 12.G.1 傅里叶变换

我们先从表12.G.1左上角的傅里叶变换开始介绍，即连续方程  $f(t)$  在无限时间区间  $(-\infty, \infty)$  的求积问题。

$$\tilde{f}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-i\omega t) f(t) dt, \quad (12.182)$$

其中  $\tilde{f}(\omega)$  是一个关于频率  $\omega$  的方程，方程形式反映了在频率域  $\omega$  中表现出的非线性（指数）强烈程度。

现在基于傅里叶变换式(12.182)来回顾前面的问题：如何将方程  $E_\alpha(t) = \exp(-\alpha|t|)$  分解为一组弦波组合的形式？其基本思路是，通过将所有关于频率  $\omega$  的弦波曲线加总来重写原方程

$\exp(-\alpha|t|)$ , 某一弦波的频率越大, 振幅 (amplitude)。

$$\begin{aligned}\tilde{E}_\alpha(\omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-i\omega t) \exp(-\alpha|t|) dt \\ &= \frac{1}{2\pi} \left\{ \int_{-\infty}^0 \exp[(-i\omega + \alpha)t] dt + \int_0^{\infty} \exp[(-i\omega - \alpha)t] dt \right\} \\ &= \frac{1}{2\pi} \left[ \frac{1}{\alpha - i\omega} - \frac{1}{\alpha + i\omega} \right] \\ &= \frac{\alpha}{\pi(\alpha^2 + \omega^2)},\end{aligned}$$

注意, 这里所说的“频率” $\omega$  越大或越小, 是相对于参数  $\alpha$  而言的:  $\alpha$  越大, 指数方程  $E_\alpha(t)$  越快衰减至 0, 我们就越需要增加  $\tilde{E}_\alpha(\omega)$  的振幅, 利用更高振幅的弦波来近似原方程。下面将作进一步说明。

### 一些概念

$$y = A \sin(Bx + c) + D,$$

其中

- $A$  表示振幅,
- $\frac{B}{2\pi}$  表示周期, 频率 =  $\frac{1}{\text{周期}}$ ,
- $-\frac{C}{B}$  表示相位的 (水平) 移动,
- $D$  表示相位的 (垂直) 移动。

图 12.G.1: 傅里叶变换

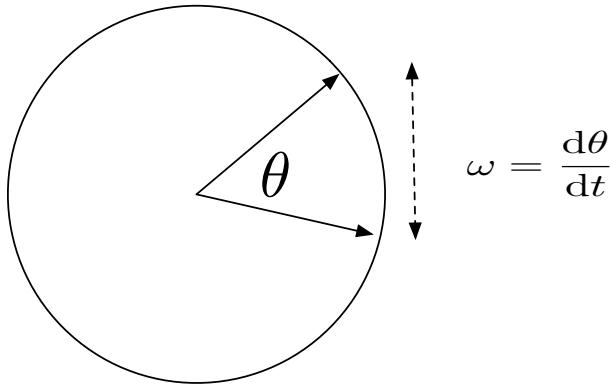


图12.G.1中

- $\omega = \frac{2\pi}{T} = 2\pi\nu = \frac{ds}{dt} \cdot \frac{1}{r} = \frac{d\theta}{dt}$
- $\omega$ , 角频率 (angular frequency), 是对物体旋转速度快慢的度量, 单位如弧度/秒 (radian/second)。
- $T$ , 周期, 单位如秒。
- $\nu$ , 线频率 (linear frequency), 单位如赫兹 (Hertz),  $1\text{Hertz} = 1\text{次}/\text{秒}$ 。
- 旋转的半径, 单位如米。

### 线频率和角频率

对于一个给定的弦波方程如  $\sin \omega t$  或  $\exp[-i\omega t]$  等，严格说来我们需要将其中的频率  $\omega$  理解为角频率 (angular frequency)，表示相角旋转速度的快慢，单位如角每秒 (radian per second)。线频率 (linear frequency) 可写为  $\nu = \frac{\omega}{2\pi}$ ，表示这一过程自我重复的频率，单位如赫兹。例如，一个钟摆 4 秒钟转一圈，那么

- 周期  $T = 4$ ，
- 角频率  $\omega = \frac{2\pi}{T} = \frac{\pi}{2} \approx 1.57 \text{ Rad/s}$ ，
- 线频率  $\nu = \frac{1}{T} = 0.25 \text{ Hz}$ 。

### 域

(12.182)对原方程  $f(t)$  作傅里叶变换，生成近似方程  $\tilde{f}(\omega)$ ，二者等价但处于不同的域中。 $f(t)$  的解释变量是 (时间) $t$ ， $\tilde{f}(\omega)$  的解释变量是频率  $\omega$ ，从这个角度可以说， $f(t)$  是在时间域中， $\tilde{f}(\omega)$  是在频率域中。

福利也分析也可用于原方程  $f$  的解释变量不是时间的情况，例如可以是方位  $x$ ，对应的傅里叶变换处于空间频率  $k$  中，有时也称波数 (wavenumber)。更多例子见下表。

表 12.G.2: 原方程域和傅里叶变换方程域

	原变量	原域	傅里叶变量	傅里叶域
信号处理	时间 $t$	时间域	频率 $\omega$	频率域
光学	方位 $x$	实空间	波数 $k$	$k$ 空间
量子力学	方位 $x$	方位空间	动量 (momentum) $p$	动量空间
固体物理学	点阵向量 $L$	实空间	布洛赫向量 (Bloch vector) $k$	晶体动量空间 (crystal momentum space)

### 单位

实际研究过程中需要牢记  $f(t)$  和  $\tilde{f}(\omega)$  的单位不同。以(12.182)为例，RHS 中有  $dt$ ，LHS 中没有。因此

$$\tilde{f} \text{ 的单位} = f \text{ 的单位} \times \text{时间} = \frac{f \text{ 的单位}}{\text{频率}},$$

如果  $f$  的单位是伏特 (Volt)，那么  $\tilde{f}$  的单位是伏特秒，或者伏特每赫兹。

### 傅里叶变换的性质

由定义(12.182)可得，傅里叶变换具有如下性质，我们将分别作说明。

- 导数的傅里叶变换
- 傅里叶变换的导数
- 实值方程的傅里叶变换

导数的傅里叶变换。对于某一给定方程  $f(x)$ ，如果我们可以求得其傅里叶变换  $\tilde{f}(k)$ ，那么  $f(x)$  的导数也很容易写出。首先写出  $f(t)$  的傅里叶综合表达式 (Fourier synthesis equation, 第12.206节)

$$f(x) = \int_{-\infty}^{\infty} \tilde{f}(k) \exp(ikx) dk,$$

两侧同时对  $x$  求导

$$\frac{d}{dx} f(x) = \int_{-\infty}^{\infty} ik\tilde{f}(k) \exp(ikx) dk,$$

上式的 RHS 是对  $(ik\tilde{f}(k))$  的逆傅里叶变换 (Definition 14.32); 因此可以将 RHS 看作是  $\frac{d}{dx} f(x)$  的傅里叶变换, 即

$$\mathcal{F}[f(x)] = \tilde{f}(k) \Rightarrow \mathcal{F}\left[\frac{d}{dx} f(x)\right] = ik\tilde{f}(k). \quad (12.183)$$

也可以重复上述过程求解高阶导数, 如

$$\mathcal{F}[f(x)] = \tilde{f}(k) \Rightarrow \mathcal{F}\left[\frac{d}{dx} f(x)\right] = ik\tilde{f}(k) \Rightarrow \mathcal{F}\left[\frac{d^2 f}{dx^2}\right] = -k^2 \tilde{f}(k). \quad (12.184)$$

傅里叶变换的导数。类似地, 对傅里叶变换(12.182)

$$\tilde{f}(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-ikx) f(x) dx$$

两侧求  $k$  的导数有

$$\frac{d}{dk} \tilde{f}(k) = \int_{-\infty}^{\infty} -ixf(x) \exp(-ikx) dx,$$

可见  $\frac{d}{dk} \tilde{f}(k)$  可以看成是方程  $x \cdot f(x)$  的傅里叶变换。因此有

$$\mathcal{F}[f(x)] = \tilde{f}(k) \Rightarrow \mathcal{F}[xf(x)] = i \frac{d}{dk} \tilde{f}(k). \quad (12.185)$$

不难看出, (12.183) 和(12.185)等价。

实值方程的傅里叶变换。如果  $f(x)$  是一个实值方程, 那么它对应的傅里叶变换  $\tilde{f}(k)$  中所含有的信息中, 相对于  $k < 0$  的部分就是冗余信息:  $\tilde{f}(k)|_{k<0}$  的信息可由  $\tilde{f}(k)|_{k>0}$  的部分而获得。以定义式为例, 对于  $k < 0$  的情况下有

$$\tilde{f}(-k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) \exp(ikx) dx,$$

由于假定  $f(x)$  是一个实值方程, 进一步可得

$$\tilde{f}(-k) = \left[ \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) \exp(-ikx) dx \right]^* = \tilde{f}^*(k).$$

## 12.G.2 脉冲方程的傅里叶变换

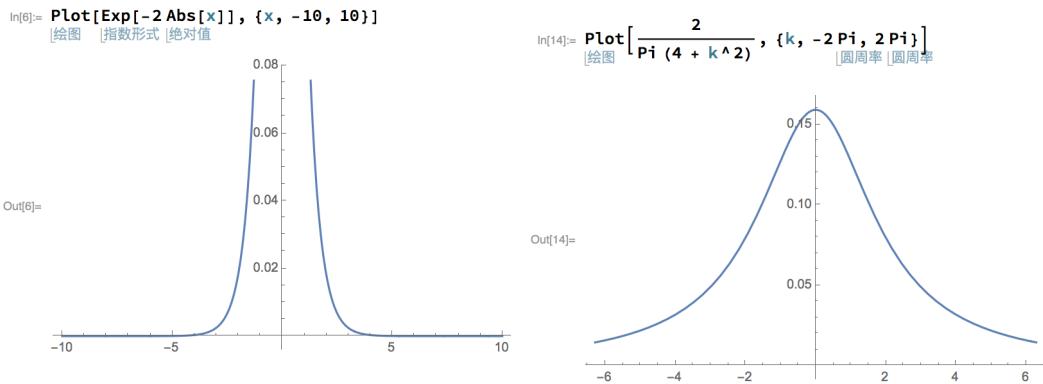
介绍几种常见的傅里叶变换。

### 洛伦兹变换

以前文提到的方程为例

$$E_{\alpha}(x) = \exp(-\alpha|x|), \quad \tilde{E}_{\alpha}(k) = \frac{\alpha}{\pi(\alpha^2 + k^2)}. \quad (12.186)$$

$E_{\alpha}(x)$  和  $\tilde{E}_{\alpha}(k)$  都可以被看作是某种脉冲方程 (pulse functions): 它们都在初始点  $x_0, k_0$  处附近有最大值, 又随着  $x \rightarrow \infty$  或  $k \rightarrow \infty$  而逐渐降低直至 0, 称为洛伦兹变换 (Lorentzian function)。两个方程的宽度均与参数  $\alpha$  有关, 但影响方向相反:  $\alpha$  的值越大, 实空间中的脉冲越弱、宽度越窄, 傅里叶空间中的脉冲越强、宽度越宽, 如图12.G.2所示。

图 12.G.2: 洛伦兹方程 ( $\alpha = 2$ )

我们将脉冲的一个半峰全宽 (full width at half maximum, FWHM) 定义为，在方程的一个周期中，前后两个方程值等于一个峰值一半的点之间的距离。根据这个定义， $E_\alpha(x)$  的峰值为  $1(x=0)$ ，并且在  $x \pm (\ln 2)/\alpha$  时，值为峰值的一半，因此有

$$FWHM [E_\alpha(x)] = \frac{2 \ln 2}{\alpha}, \quad (12.187)$$

$\tilde{E}_\alpha(k)$  的峰值为  $\frac{1}{\pi\alpha}(k=0)$ ，在  $k = \pm\alpha$  时，值为峰值的一半，

$$FWHM [\tilde{E}_\alpha(k)] = 2\alpha. \quad (12.188)$$

(12.187) 和 (12.188) 联立有

$$FWHM [E_\alpha(x)] \cdot FWHM [\tilde{E}_\alpha(k)] = 4 \ln 2, \quad (12.189)$$

注意：单个原方程及洛伦兹变换方程的 FWHM 与  $\alpha$  有关；两个 FWHM 的乘积则与  $\alpha$  无关：所有洛伦兹脉冲方程族  $\{E_\alpha(t)\}$  都具有这样的特征。

### 高斯变换

设一个宽度为  $\sigma$  的高斯方程 (Gaussian function)<sup>41</sup>

$$G_\sigma(x) = \exp\left(-\frac{x^2}{\sigma^2}\right),$$

那么其傅里叶变换

$$\begin{aligned} \tilde{G}_\sigma(k) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-ikx) G_\sigma(x) dx \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \underbrace{\exp\left(-ikx - \frac{x^2}{\sigma^2}\right)}_{=: \mathcal{A}} dx, \end{aligned}$$

<sup>41</sup>高斯方程族中有不同的类型，简要介绍见第12.G.9节。

RHS 的被积方程

$$\begin{aligned}
 \mathcal{A} &= \frac{1}{\sigma^2} (x^2 + ikx\sigma^2) \\
 &= \frac{1}{\sigma^2} \left( x + \frac{1}{2}ik\sigma^2 \right)^2 + \frac{1}{4}k^2\sigma^2 \\
 &= \frac{1}{2\pi} \exp\left(-\frac{1}{4}k^2\sigma^2\right) \underbrace{\int_{-\infty}^{\infty} \exp\left[-\frac{1}{\sigma^2}\left(x + \frac{1}{2}ik\sigma^2\right)^2\right] dx}_{=\sigma\cdot\sqrt{\pi}} \\
 &= \frac{\sigma}{2\sqrt{\pi}} \exp\left(-\frac{1}{4}k^2\sigma^2\right),
 \end{aligned}$$

可见  $\tilde{G}_\sigma(k)$  也是一个  $k$  空间中的高斯方程，其宽度与原方程  $G_\sigma(x)$  的宽度成等比例

$$\tilde{G}_\sigma \propto \exp\left(-\frac{k^2}{\tilde{\sigma}^2}\right) = G_{\tilde{\sigma}}(k), \quad \tilde{\sigma} \equiv \frac{2}{\sigma}.$$

此外

$$\begin{aligned}
 FWHM[G_\sigma(x)] &= 2\sqrt{\ln 2}\sigma, \\
 FWHM[G_\sigma(x)] \cdot FWHM[\tilde{G}_\sigma(x)] &= \left(2\sqrt{\ln 2}\sigma\right) \cdot \left(\frac{4\sqrt{\ln 2}}{\sigma}\right) = 8(\ln 2)^2,
 \end{aligned} \tag{12.190}$$

可见单个原高斯方程及高斯变换方程的 FWHM 与  $\alpha$  有关；两个 FWHM 的乘积则与  $\alpha$  无关：所有高斯脉冲方程族  $\{G_\sigma(x)\}$  都具有这样的特征。

### 12.G.3 非脉冲方程的傅里叶变换

洛伦兹变换和高斯变换都属于脉冲方程，特点是初始点附近有峰值，随着  $x(k)$  逐渐增大而慢慢回落至 0。如何求得非脉冲方程的傅里叶变换，如  $f(x) = 1$ ,  $f(x) = x$ , 或  $f(x) = x^2$  呢？

一种求解思路是，考虑(12.186)在  $\lim_{\alpha \rightarrow 0}$  时的情况，实值方程趋近于 1

$$\lim_{\alpha \rightarrow 0} E_\alpha(x) \rightarrow 1.$$

而它的洛伦兹变换  $\tilde{E}_\alpha(k)$  的求解则较为复杂，第一它的宽度变窄了  $FWHM \tilde{E}_\alpha = 2\alpha$ ；第二它的高度变高了  $\tilde{E}_\alpha(0) = \frac{1}{\pi\alpha}$ 。随着  $\alpha \rightarrow 0$  的极限情况出现， $\lim_{\alpha \rightarrow 0} \tilde{E}_\alpha(k)$  变得极窄、极高，就成为狄拉克方程 (Delta Dirac function, 第16.B.5节)

$$\lim_{\alpha \rightarrow 0} \tilde{E}_\alpha(k) = \lim_{\alpha \rightarrow 0} \frac{\alpha}{\pi(\alpha^2 + k^2)} \equiv \delta(k) = \begin{cases} +\infty & k = 0, \\ 0 & k \neq 0, \end{cases} \Rightarrow \int_{-\infty}^{\infty} \delta(k) dk = 1. \tag{12.191}$$

对应地，非脉冲方程  $f(x) = 1$  的傅里叶变换可以写作

$$f(x) \equiv 1 \Rightarrow \tilde{f}(k) = \delta(k), \tag{12.192}$$

可以这么理解： $f(x) \equiv 1$  是一个频率为 0 的弦波。对  $f(x)$  做傅里叶转换，分解为若干弦波之和，在这一系列弦波中，我们将除了第一条（对应频率为 0）之外的其他全部弦波曲线的系数值均设为 0。

在此基础上，利用(12.192)和(12.185)，进一步可得到另外两个非脉冲方程的傅里叶变换

$$f(x) = x \Rightarrow \tilde{f}(k) = i\delta'(k), \tag{12.193}$$

$$f(x) = x^2 \Rightarrow \tilde{f}(k) = -\delta''(k), \tag{12.194}$$

其中  $\delta'(k)$  是狄拉克方程的导数，基于(12.191)用分部积分法有

$$\int f(u)\delta'(u)du = -\int f'(u)\delta(u)du = -f'(0),$$

可见  $\delta'$  可以被看做是和  $\delta$  相近的方程，二者只有一点不同：利用分部求积计算  $\delta'$  时产生了  $-f'(0)$  的变化，而  $\delta$  只产生  $-f$  的变化。

需要指出的是，对非脉冲方程的傅里叶变换如(12.192)、(12.193)、(12.194)，都不算是“好”的方程——它们甚至不是方程，而只是分布（比如狄拉克方程单独出现往往没有实际意义，它更多是与其他定义良好的方程配对出现在积分式中）：这么说是因为  $f(x) = \{1, x, x^2\} \notin L^1$ ，即它们不满足  $\int_{-\infty}^{\infty} |f(x)| dx < \infty$ <sup>42</sup>。因此，(12.194)等傅里叶变换，往往更适合以算子的形式出现。

#### 12.G.4 原方程的平滑和傅里叶变换方程的衰减

通过前面的分析不难总结出一个一般规律：原方程  $f(t)$  越是（口语意义上的）不“平滑”，即随着解释变量  $t$  的变动而剧烈变动，其傅里叶变换  $\tilde{f}(\omega)$  随着  $\omega$  的衰减（decay）速度就越慢。反之亦然： $f(t)$  越平滑， $\tilde{f}(\omega)$  的衰减就越快。可以将  $f(t)$  的平滑特性表现为  $f$  的连续性和导数：

**Theorem 12.62** (佩利——维纳诸定理). 如果  $f(t)$  和它的前  $p-1$  个导数都是连续的，但第  $p$  个导数是不连续的，并且变差有界 (bounded variation)，那么  $\tilde{f}(\omega)$  的衰减速度不低于  $|\omega|^{-(p+1)}|_{\omega \rightarrow \infty}$ 。这称为佩利——维纳诸定理 (Paley-Wiener theorems)。

证明. 一种证明方式见第291页脚注。更详实的证明可参考如 (Yosida, 1978, Sec. VI.4), (Agranovich, 2015, Theorem 4.1.2)。□

尤其是，若  $f(t) \in C^\infty$ <sup>43</sup>，那么在  $\omega$  值较大时， $\tilde{f}(\omega)$  的衰减速度比任何多项式都快。这样的方程包括诸如  $\exp(-\omega)$ ,  $\exp(-\sqrt{\omega})$ ,  $\exp(-\omega^2)$  等。

举例如洛伦兹方程  $\exp(-\alpha|t|)$  连续，但其一阶导数不连续，表现为在  $t = 0$  处有跳跃，如图12.G.2左侧。因此  $p = 1$ 。对应地，我们会看到它的傅里叶变换以等比于  $\omega^{-2}$  的速度衰减（当  $\omega$  的值较大时）。

再比如，方程  $\exp\left(-\frac{t^2}{\sigma^2}\right) \in C^\infty$ ，那么其傅里叶变换  $\tilde{F} = \exp\left(-\frac{1}{4}\sigma^2\omega^2\right)$  的衰减速度快于任意多项式。

#### 12.G.5 傅里叶级数

现在来看表12.G.1左下一栏，对应有限空间中连续方程  $f(t)$  的傅里叶分析，称为傅里叶级数。

假设  $f(t)$  是个关于  $T$  的周期方程 (periodic function)，即每隔  $T$  单位时间 (秒, 分, 小时, 天等) 就自我重复一遍：

$$f(t+T) \equiv f(t), \quad \forall t.$$

假定  $f(t)$  符合傅里叶分析的条件，对应傅里叶变换

$$\tilde{f}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-i\omega t) f(t) dt, \quad (12.195)$$

随着角频率  $\omega$  和基准频率 (base frequency)  $\omega_0 = \frac{2\pi}{T}$  的关系不同，需要分两种情况来分析：

<sup>42</sup>  $L^1$  表示勒贝格空间，见第(14.2.2)节。

<sup>43</sup> 空间  $C^\infty$  中的方程连续，所有导数也都连续，可参考第14.2.1节。

1.  $\omega$  是  $\omega_0$  的整数倍。

那么(12.195)中的被积方程周期为  $T$ , 每个宽度为  $T$  的子区间对总积分的作用相同, 并且由于这样的子区间有无数个, 我们有  $\tilde{f}(\omega) = \infty$ 。

2.  $\omega$  是  $\omega_0$  的整数倍。

那么(12.195)中的被积方程不是一个周期方程。我们可以将  $f(t)$  理解为两部分: 一部分是  $f(t)$ , 这是一个关于  $T$  的周期方程, 另一部分是  $\exp(-i\omega t)$ , 它只有在某些特定的时期才是周期方程, 这些特定的时期是指不等于  $T$  的整数倍的时期。两部分的乘积导致被积方程不是一个周期方程。

每个宽度为  $T$  的子区间都给总积分值产生同等程度的影响, 但影响的大小 (phase, 相) 是随机的。各个随机相的影响加总后相互抵消,  $\tilde{f}(\omega) = 0$ 。

将上述两种情况汇总可见, 在  $f(t)$  是一个周期方程的前提下

$$\tilde{f}(\omega) = \begin{cases} \infty, & \omega \text{ 是 } \omega_0 \text{ 的整数倍}, \\ 0, & \text{否则}. \end{cases}$$

我们可以用三种数学形式将上式表达出来:

1. 利用狄拉克方程, 可将  $T$ -周期方程的傅里叶变换表示为

$$\tilde{f}(\omega) = \sum_{\nu=-\infty}^{\infty} \tilde{f}_{\nu} \cdot \delta(\omega - \nu\omega_0), \quad \omega_0 \equiv \frac{2\pi}{T}, \nu \in \mathcal{Z}. \quad (12.196)$$

2. 对  $T$ -周期方程  $f(t)$  的傅里叶分解, 只包括那些频率为  $\omega_{\nu}$  的弦波曲线, 满足  $\omega_{\nu} = \nu \cdot \omega_0 = \frac{2\nu\pi}{T}$ ,  $\nu \in \mathcal{Z}$ , 记作

$$f(t) = \sum_{\nu \in \mathcal{Z}} \tilde{f}_{\nu} \cdot \exp(i\nu\omega_0 t), \quad \omega_0 = \frac{2\nu\pi}{T}, \nu \in \mathcal{Z}, \quad (12.197)$$

我们又将(12.197)称为傅里叶级数 (Fourier series) 的表达形式之一: 复杂指数表达形式。

3. 除此之外, 也可以将傅里叶级数表达为另一种形式, 即一组正弦函数和一组余弦函数之和

$$f(t) = \sum_{\nu=0}^{\infty} \tilde{a}_{\nu} \cdot \cos(\nu\omega_0 t) + \sum_{\nu=1}^{\infty} \tilde{b}_{\nu} \cdot \sin(\nu\omega_0 t), \quad \nu \in \mathcal{Z}^{+}. \quad (12.198)$$

### 两种表达形式的比较

(12.198)中的  $\nu$  需要取正整数; (12.197)则正负均可。除此而外, 指数形式(12.197)和弦波形式(12.198)等价, 可以被看作是同一个  $T$ -周期方程  $f(t)$  的傅里叶级数分解。

来看两组系数  $\{\tilde{a}_{\nu}, \tilde{b}_{\nu}\}_{\nu \in \mathcal{Z}^{+}}$  和  $\{\tilde{f}_{\nu}\}_{\nu \in \mathcal{Z}}$  之间的关系。已知复杂指数  $\exp(i\nu\omega_0 t)$  可表示为关于正弦和余弦函数的方程

$$\exp(i\nu\omega_0 t) = \cos(\nu\omega_0 t) + i \sin(\nu\omega_0 t), \quad (12.199)$$

以及三角函数关系: 余弦函数是偶函数, 正弦函数是奇函数

$$\cos(-A) = \cos(A), \quad \sin(-A) = -\sin(A), \quad (12.200)$$

结合(12.197)-(12.200)可得

$$\begin{aligned}\tilde{a}_0 &= \tilde{f}_0, \\ \tilde{a}_\nu &= \left( \tilde{f}_\nu + \tilde{f}_{-\nu} \right), \quad \nu > 0, \\ \tilde{b}_\nu &= i \left( \tilde{f}_\nu - \tilde{f}_{-\nu} \right), \quad \nu > 0,\end{aligned}\tag{12.201}$$

或者反过来

$$\tilde{f}_\nu = \begin{cases} \frac{1}{2} \left( \tilde{a}_\nu + i\tilde{b}_\nu \right), & \nu < 0, \\ \tilde{a}_0, & \nu = 0, \\ \frac{1}{2} \left( \tilde{a}_\nu - i\tilde{b}_\nu \right), & \nu > 0. \end{cases}\tag{12.202}$$

比较两种表达式形式：复杂指数形式(12.197)和正弦余弦形式(12.198)：

1. (12.198)的好处在于

- (a) 若  $T$ -周期  $f(t)$  是实值方程，那么其傅里叶变换系数  $\{\tilde{a}_\nu, \tilde{b}_\nu\}_{\nu \in \mathbb{Z}^+}$  也都是实值系数，从而无需考虑复数的情况。
- (b) 结合具体  $f(t)$  的奇偶性，可以作灵活调整，从而简化后续的计算工作。例如若  $f(t)$  是偶函数，那么它的傅里叶变换只包含余弦项；全部正弦项都消失了，因此无需计算  $\tilde{b}_\nu$ 。

2. (12.197)的好处在于

- (a) 只需计算一组系数  $\{\tilde{f}_\nu\}_{\nu \in \mathbb{Z}}$ ，而非两组系数  $\{\tilde{a}_\nu, \tilde{b}_\nu\}_{\nu \in \mathbb{Z}^+}$ 。
- (b) 指数形式傅里叶级数的推算公式(12.197)更直观和易于理解。
- (c) 求导更容易。事实上对(12.197)求导后生成的傅里叶级数仍然是(12.197)，只是原来的系数  $\{\tilde{f}_\nu\}_{\nu \in \mathbb{Z}}$  现在变成了  $\{i\omega_0 \tilde{f}_\nu\}_{\nu \in \mathbb{Z}}$ 。反观(12.198)，求导操作更为复杂。

基于上述考虑，在大多数情况下我们采用复杂指数形式(12.197)展开傅里叶级数分析<sup>44</sup>，使得更为直观和容易理解；然而这两种形式之间的互相转换仍然很简单，如式(12.201)-(12.202)所示。

### 复杂指数形式的系数计算

假定一个实值方程  $f(t)$ ，在  $t \in [0, T]$  时域内的傅里叶级数，其系数  $\{\tilde{f}_\nu\}_{\nu \in \mathbb{Z}}$  的计算式为

$$\tilde{f}_\nu = \frac{1}{T} \int_0^T f(t) \exp(-i\nu\omega_0 t) dt.$$

例如，设  $f(t) = \cos^2(3t)$ ,  $T = \frac{\pi}{3}$ ，见图(12.G.3)。

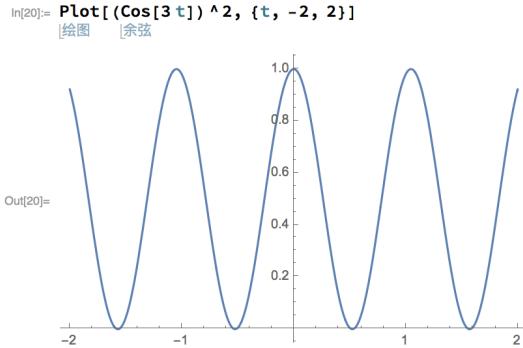
基准频率  $\omega_0 = \frac{2\pi}{T} = 6$ 。已知

$$\begin{aligned}\cos(3t) &= \frac{1}{2} [\exp(3it) + \exp(-3it)], \\ \Rightarrow f(t) &= \cos^2(3t) = \frac{1}{4} [\exp(6it) + 2 + \exp(-6it)], \\ \Rightarrow \tilde{f}_\nu &= \frac{1}{4T} \int_0^T \exp(-i\nu\omega_0 t) [\exp(i\omega_0 t) + 2 + \exp(-i\omega_0 t)] dt.\end{aligned}\tag{12.203}$$

根据正交性质(第12.G.9节)可得，上式变为

$$\tilde{f}_\nu = \frac{1}{4} [\delta_{\nu,1} + 2\delta_{\nu,0} + \delta_{\nu,-1}],$$

<sup>44</sup>除了个别情况，比如利用切比雪夫光谱法来分析实值偶函数  $f(t)$  时。

图 12.G.3:  $f(t) = \cos^2(3t)$ ,  $t \in (-2, 2)$ 

换句话说，傅里叶级数的系数  $\tilde{f}_\nu$  只有在  $\nu = \{-1, 0, 1\}$  时不等于 0。那么，将  $f(t)$  用傅里叶综合形式（第12.206节）写为

$$\begin{aligned} f(t) &= \sum_{\nu} \tilde{f}_{\nu} \exp(-i\nu\omega_0 t) = \frac{1}{4} \exp(i\omega_0 t) + \frac{1}{2} + \frac{1}{4} \exp(-i\omega_0 t) \\ &= \frac{1}{2} [1 + \cos(\omega_0 t)] = \frac{1}{2} [1 + \cos(6t)], \end{aligned} \quad (12.204)$$

结合三角函数关系，上式还可以做进一步简化

$$\cos^2(3t) = \frac{\cos(6t) + 1}{2}.$$

### 傅里叶余弦级数

(12.204) 将  $f(t) = \cos^2(3t)$  分解为一个余弦项和一个常数项之和，没有正弦项。事实上若方程  $f(t)$  是偶函数，傅里叶变换后常常是没有正弦项的。反之亦然。我们将两种情况分别称为傅里叶余弦级数 (Fourier cosine series) 和傅里叶正弦级数 (Fourier sine series)。

有时候原始方程本身既不是奇函数也不是偶函数，但有可能通过坐标点的移动来变成奇或偶函数。如  $T$ -周期方程

$$f(t) = \begin{cases} 0, & 0 < t < \frac{T}{2} \\ 1, & \frac{T}{2} < t < T \end{cases}$$

非奇非偶。可以设一个新函数  $g(t) \equiv f(t + \frac{\pi}{4})$ ，使得  $g(t)$  变成偶函数。

更一般地说，所有方程都可以分解为奇函数、偶函数组合的形式，如

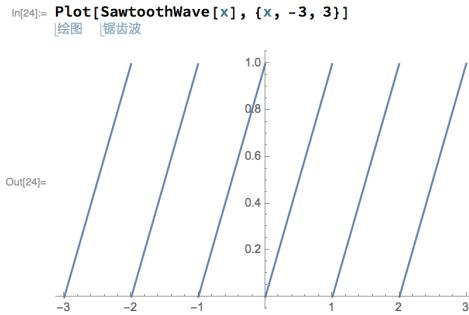
$$\begin{aligned} f(t) &= f_{\text{偶}}(t) + f_{\text{奇}}(t), \\ f_{\text{偶}}(t) &\equiv \frac{1}{2} [f(t) + f(-t)], \\ f_{\text{奇}}(t) &\equiv \frac{1}{2} [f(t) - f(-t)]. \end{aligned}$$

举例来说如锯齿波 (sawtooth wave)， $T$ -周期方程  $f(t)$  在每一个周期内都满足  $f(t) = t$ ,  $0 < t < T$ ，也即， $f(t)$  的单位与  $t$  的单位相同，都是时间，见图12.G.4。

该方程的傅里叶级数

$$\begin{aligned} f(t) &= \sum_{\nu=-\infty}^{\infty} \tilde{f}_{\nu} \cdot \exp(i\nu\omega_0 t), \quad \omega_0 = \frac{2\pi}{T}, \\ \hookrightarrow \tilde{f}_{\nu} &= \frac{1}{T} \int_0^T f(t) \exp(-i\nu\omega_0 t) dt = \frac{1}{T} \int_0^T t \exp(-i\nu\omega_0 t) dt. \end{aligned}$$

图 12.G.4: 锯齿波



上式中  $\tilde{f}_\nu$  的值随着  $\nu$  的取值而不同，具体说来：

$$\tilde{f}_0 = \frac{T}{2}, \quad \nu = 0,$$

$$\tilde{f}_\nu = \frac{1}{T} \left[ -\frac{1}{i\nu\omega_0} |t \exp(-i\nu\omega_0 t)|_0^T + \frac{1}{i\nu\omega_0} \int_0^T \exp(-i\nu\omega_0 t) dt \right] = -\frac{1}{i\nu\omega_0}, \quad \nu \neq 0.$$

代回可得

$$\underbrace{\tilde{f}_\nu}_{\text{(时间)}} = \underbrace{\frac{T}{2}}_{\text{时间}} - \underbrace{\frac{1}{i\omega_0} \sum_{\nu=-\infty}^{\infty} \frac{1}{\nu} \exp(i\omega_0 t)}_{\text{单位: (1/角频率)= 时间}}$$

不难看出，在角频率  $\omega$  的值较大时， $\tilde{f}_\nu$  以等比于  $\frac{1}{|\omega|}$  的速度衰减。这是由  $f(t)$  方程的非连续性所决定的（佩利——维纳诸定理，Theorem 12.62）。

上式也可改为正弦余弦形式

$$f(t) = \frac{T}{2} - \underbrace{\frac{T}{\pi} \sum_{\nu=1}^{\infty} \frac{1}{\nu} \sin\left(\frac{2\nu\pi t}{T}\right)}_{=: \mathcal{A}}, \quad (12.205)$$

其中  $\mathcal{A}$  符合傅里叶级数的特征。若设  $g(t) = f(t) - \frac{T}{2}$ ，则  $g(t)$  就是个傅里叶正弦级数了——只要  $g(t)$  是一个奇函数的话——将图 12.G.4 垂直下移  $\frac{\pi}{2}$  个单位不难看出， $g(t)$  也是个奇函数。

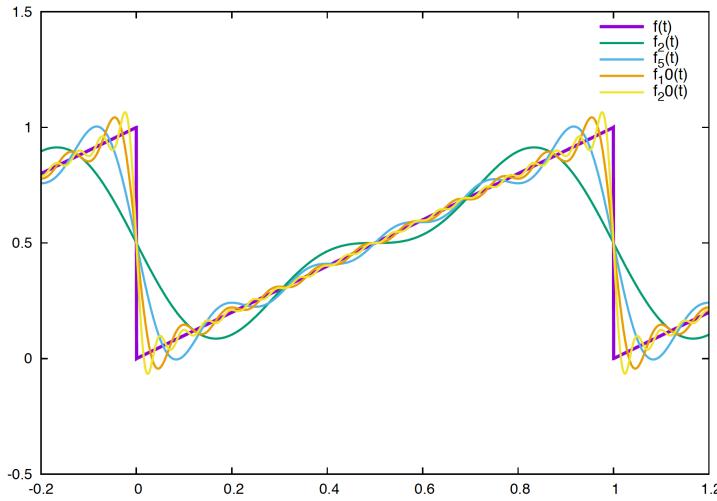
### 吉布斯现象

前文可见，通过加总一系列弦波方程（假定其中每一个弦波都是平滑的），可以生成图 12.G.4 一般参差状不连续的锯齿方程，但这是有限定条件的：如果我们只截取全部傅里叶级数中的一段，即只加总有限个项（总共有无数个项加总）的话，这意味着只对  $f(t)$  作有限个傅里叶综合的话，就会产生吉布斯现象（Gibbs phenomenon）：这是指，在对非连续方程作不完全傅里叶综合时，在不连续点附近所出现的震荡。如图 12.G.5 所示，原方程  $f(t)$  为锯齿方程， $f_N(t)$ ,  $n = 2, 5, 10, 20$  分别为对  $f(t)$  所做的不完全傅里叶综合

$$f_N(t) = \sum_{\nu=-N}^N \tilde{f}_\nu \cdot \exp(i\nu\omega_0 t).$$

不难看出，在远离不连续点（如  $t = 0, 1$ ）的地方， $N$  越是大，傅里叶级数越能对原方程作精确近似。而反之则不然：在接近不连续点的地方，傅里叶级数的近似精度并为随着  $N$  的升高而显著提升。

图 12.G.5: 吉布斯现象



### 傅里叶级数的收敛

图12.G.5也揭示了傅里叶级数的又一个重要特征：如果非连续方程  $f(t)$  在  $t^*$  点不连续，那么它的傅里叶级数  $\sum \tilde{f}_\nu \exp(i\nu\omega_0 t^*)$  收敛至这一不连续的中点，即

$$\lim_{N \rightarrow \infty} \sum_{\nu=-N}^N \tilde{f}_\nu \cdot \exp(i\nu\omega_0 t^*) \approx \frac{1}{2} [f(t_-^*) + f(t_+^*)],$$

$$f(t_-^*) \equiv \lim_{\varepsilon \rightarrow 0} f(t^* - \varepsilon), \quad f(t_+^*) \equiv \lim_{\varepsilon \rightarrow 0} f(t^* + \varepsilon), \quad \varepsilon > 0.$$

根据这一性质，如果我们在  $[0, T]$  区间内，对一个并非以  $[0, T]$  为周期的不连续方程作傅里叶级数的近似，那么用  $t = 0$  点作傅里叶近似，可得  $\frac{1}{2} [f(0) + f(T)]$ ，这在图12.G.5中有所反映。

### 傅里叶分析是无损转换过程

傅里叶分析的又一个重要特征表现为，对原方程的傅里叶转换是无损 (lossless) 的，即  $f(t)$  中的信息在  $f(t) \rightarrow \tilde{f}(\omega)$  的过程中没有损耗。为了说明这一点，首先来介绍傅里叶综合 (Fourier synthesis) 的概念：存在关于  $f(t)$  的傅里叶级数  $\tilde{f}(\omega)$ ，并且可以根据傅里叶级数向回逆推得到原方程  $f(t)$ ，表示为

$$f(t) = \int_{-\infty}^{\infty} \exp(i\omega t) \cdot \tilde{f}(\omega) d\omega, \quad (12.206)$$

即将一系列指数方程 (或弦波曲线) 加权加总，权重由与频率  $\omega$  有关的傅里叶变换  $\tilde{f}(\omega)$  给出；加总后的方程即原方程  $f(t)$ 。

**Theorem 12.63** (帕塞瓦尔定理 (傅里叶变换)). 我们说(12.206)是精确无损的，是指原域中的  $f(t)$  和傅里叶域中的  $\tilde{f}(\omega)$  之间相互转换不会产生信息损耗，这称为帕塞瓦尔定理 (Parseval's theorem)。

证明. 设两方程的内积 (inner product) 满足

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f^*(t) g(t) d(t),$$

引入傅里叶综合形式(12.206)分别替代 RHS 被积方程中的  $f^*$  和  $g(t)$ , 有

$$\begin{aligned}
 \langle f, g \rangle &= \int_{-\infty}^{\infty} f^*(t) g(t) dt \\
 &= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} \exp(-i\omega t) \tilde{f}^*(\omega) d\omega \right\} \cdot \left\{ \int_{-\infty}^{\infty} \exp(i\omega' t) \tilde{g}(\omega') d\omega' \right\} dt \\
 &= \int_{-\infty}^{\infty} \tilde{f}^*(\omega) \underbrace{\int_{-\infty}^{\infty} \tilde{g}(\omega') \left\{ \int_{-\infty}^{\infty} \exp[i(\omega' - \omega)t] dt \right\} d\omega'}_{=2\pi\delta(\omega' - \omega)} d\omega \\
 &= \int_{-\infty}^{\infty} \tilde{f}^*(\omega) \int_{-\infty}^{\infty} \tilde{g}(\omega') 2\pi\delta(\omega' - \omega) d\omega' d\omega \\
 &= 2\pi \int_{-\infty}^{\infty} \tilde{f}^*(\omega) \tilde{g}(\omega) d\omega,
 \end{aligned} \tag{12.207}$$

也就是说, 原方程的内积  $\langle f, g \rangle$  等于它们两个的傅里叶变换相乘  $\tilde{f}^*(\omega) \tilde{g}(\omega)$  后的积。  $\square$

**Theorem 12.64** (普朗歇尔定理 (傅里叶变换)). 在帕塞瓦尔定理 (*Theorem 12.63*) 的基础上, 设  $f(t) \equiv g(t)$ , 那么由式(12.207)可得

$$\langle f, g \rangle = \int_{-\infty}^{\infty} |f(t)|^2 dt = 2\pi \int_{-\infty}^{\infty} |\tilde{f}(\omega)|^2 d\omega, \tag{12.208}$$

又称普朗歇尔定理 (*Plancherel's theorem*)。

现在将区间  $(-\infty, \infty)$  缩小到某个定区间  $[0, T]$ , 来看傅里叶级数中方程转换的信息损耗问题。之所以关注定区间  $[0, T]$ , 要么是因为原方程  $f(t)$  是  $T$ -周期方程, 要么是我们的研究中只关注特定区间内方程值的变化。

**Theorem 12.65** (帕塞瓦尔定理 (傅里叶级数)). 在  $[0, T]$  区间内, 由 *Theorem 12.63*可得

$$\int_0^T f^*(t) g(t) dt = T \sum_{\nu=-\infty}^{\infty} \tilde{f}_\nu^* \tilde{g}_\nu. \tag{12.209}$$

**Theorem 12.66** (普朗歇尔定理 (傅里叶级数)). 在  $[0, T]$  区间内, 由 *Theorem 12.64*可得

$$\int_0^T |f(t)|^2 dt = T \sum_{\nu=-\infty}^{\infty} |\tilde{f}_\nu|^2. \tag{12.210}$$

根据帕塞瓦尔定理和普朗歇尔定理, 在数值计算过程中, 我们既可以求解实区间中的原始方程, 也可以求解傅里叶空间中的变换方程, 二者等价。实际求解哪一种取决于哪个更容易计算。

## 12.G.6 泊松求和式

采用类似于帕塞瓦尔定理的思路, 也有另一种在实域或傅里叶域近似求积的方法, 称泊松求和式 (Poisson summation formula)。

**Theorem 12.67.** 具体说来, 假定关于方程  $f(t)$ , 我们想要求得一组等宽分布的时间点 (对应宽度  $\Delta t$ ) 下  $f(t)$  值的和。根据泊松求和式, 我们也可以计算傅里叶域中一组等宽分布 (对应宽度  $\frac{2\pi}{\Delta t}$ ) 的  $\tilde{f}(\omega)$  的值, 两种方法等价。

$$\sum_{n=-\infty}^{\infty} f(n\Delta t) = \frac{2\pi}{\Delta t} \sum_{\nu=-\infty}^{\infty} \tilde{f}\left(\frac{2\pi\nu}{\Delta t}\right). \tag{12.211}$$

证明. 用傅里叶综合表达式改写(12.211)LHS

$$\begin{aligned} \sum_{n=-\infty}^{\infty} f(n\Delta t) &= \sum_{n=-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} \tilde{f}(\omega) \exp(in\omega\Delta t) d\omega \right] \\ &= \int_{-\infty}^{\infty} \tilde{f}(\omega) \underbrace{\sum_{n=-\infty}^{\infty} \exp(in\omega\Delta t)}_{=2\pi \sum_{\nu} \delta(\omega\Delta t - 2\nu\pi)} d\omega \\ &= \int_{-\infty}^{\infty} \tilde{f}(\omega) 2\pi \underbrace{\sum_{\nu} \delta(\omega\Delta t - 2\nu\pi)}_{=:A} d\omega, \end{aligned}$$

最后一行等式中, 将  $f(n\Delta t)$  理解为一个关于  $\omega\Delta t - 2\nu\pi$  的狄拉克方程, 可分两种情况进一步分析

- 如果  $\omega\Delta t$  是  $2\pi$  的整数倍 ( $\nu \in \mathbb{Z}$  可以是任意整数), 那么  $\sum_{n=-\infty}^{\infty} f(n\Delta t) = \infty$ 。
- 否则,  $A$  就是一组狄拉克方程  $\delta(\omega\Delta t - 2\nu\pi)$  之和, 每个狄拉克方程对应不同的  $\nu$  值; 进而  $\sum_{n=-\infty}^{\infty} f(n\Delta t) = 0$ , 各项之间相互抵消。

由狄拉克方程的性质

$$\delta(ax - b) = \frac{1}{a} \delta\left(x - \frac{b}{a}\right)$$

可进一步得到

$$\begin{aligned} \sum_{n=-\infty}^{\infty} f(n\Delta t) &= 2\pi \sum_{\nu} \int_{-\infty}^{\infty} \tilde{f}(\omega) \delta(\omega\Delta t - 2\nu\pi) d\omega \\ &= \frac{2\pi}{\Delta t} \sum_{\nu} \int_{-\infty}^{\infty} \tilde{f}(\omega) \delta\left(\omega - \frac{2\nu\pi}{\Delta t}\right) d\omega \\ &= \frac{2\pi}{\Delta t} \sum_{\nu=-\infty}^{\infty} \tilde{f}\left(\nu \frac{2\pi}{\Delta t}\right), \end{aligned}$$

证毕。  $\square$

两个值域中各有一组点, 两组点的单位相同<sup>45</sup>。为了简化表达, 设  $\Delta\omega \equiv \frac{2\pi}{\Delta t}$ , 则(12.211)改写为

$$\sum_{n=-\infty}^{\infty} f(n\Delta t) = \Delta\omega \sum_{\nu=-\infty}^{\infty} \tilde{f}(\nu \cdot \Delta\omega), \quad (12.212)$$

值得留意的是  $\Delta t$  和  $\Delta\omega$  呈反比例, 即, 在原时间域中, 若是对原方程  $f(t)$  取密集的配点求  $\sum f(\cdot)$ , 对应较小的  $\Delta t$  值, 那么在对应的傅里叶域中, 就涉及到较高的频率值  $\Delta\omega$ ——换句话说, 原域中的配点越是密, 在稀疏频率域中对应的点就越是稀疏。

随着  $f(\cdot)$  的形式不同, 泊松求和式的一些具体应用如下。

### 高斯方程的泊松求和

在前面介绍过, 一个高斯方程的傅里叶变换也是一个高斯方程(第12.G.2节)。结合本节的知识我们可以作进一步判断: 对高斯方程形式原方程配点越宽(如, 实域宽度  $\sigma$ ), 其变换后的高斯形式方程越窄(如, 傅里叶域对应的宽度  $\frac{2}{\sigma}$ )。

<sup>45</sup>  $\tilde{f}$  的单位 =  $f$  的单位 / 频率 =  $f$  的单位  $\times$  (时间), 其中时间用  $\Delta t$  来表示。因此  $\tilde{f}$  的单位 /  $\Delta t$  =  $f$  的单位。

因此, 设一个以  $x$  为参数的关于  $t$  的高斯方程  $T_x(t) = \exp(-t^2\pi x)$ , 要求  $\sum T_x(t)$ 。在  $\sum T_x(t)$  的收敛速度较慢的情况下, 我们可以利用泊松求和式, 在傅里叶空间中进行求和计算<sup>46</sup>。具体说来,  $T_x(t)$  的傅里叶变换

$$\begin{aligned}\tilde{T}_x(\omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-i\omega t) T_x(t) dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-i\omega t - t^2\pi x) dt \\ &= \frac{1}{2\pi\sqrt{x}} \exp\left(-\frac{\omega^2}{4\pi x}\right),\end{aligned}$$

### 雅各比方程

雅各比方程族 (Jacobian  $\theta$  function) 中的一种形式可以表示为

$$\theta(x) = \sum_{n=-\infty}^{\infty} \exp(-n^2\pi x) = \sum_{n=-\infty}^{\infty} T_x(n). \quad (12.213)$$

利用泊松求和式(12.211), 以及设  $\Delta t = 1$ , 有

$$\theta(x) = 2\pi \sum_{\nu=-\infty}^{\infty} \tilde{T}(2\nu\pi) = \frac{1}{\sqrt{x}} \underbrace{\sum_{\nu=-\infty}^{\infty} \exp\left(-\frac{\pi\nu^2}{x}\right)}_{=\theta\left(\frac{1}{x}\right)}, \quad (12.214)$$

即雅各比方程(12.213)的函数方程 (functional equation) 形式

$$\begin{aligned}\theta(x) &= \frac{1}{\sqrt{x}} \theta\left(\frac{1}{x}\right), \\ \Leftrightarrow \sum_{n=-\infty}^{\infty} \exp(-n^2\pi x) &= x^{-\frac{1}{2}} \sum_{\nu=-\infty}^{\infty} \exp\left(-\frac{\nu^2\pi}{x}\right).\end{aligned} \quad (12.215)$$

为了说明为何傅里叶级数求和的形式即(12.215)RHS 更适于进行数值计算, 我们做了一个小实验, 设  $x = 0.04$ ,  $n = 0, 1, \dots, 11$ 。不难看出, LHS 需要  $n = 11$ , 前后 23 项的求和才能达到小数点后 6 位的精度; 而对于雅各比方程形式的傅里叶级数求和, RHS 只需要  $n = 1$  就能达到同样的精度, 换句话说, 由(12.215)我们有

$$\underbrace{\theta(0.04)}_{4.999999971, n=11} = \underbrace{\frac{1}{\sqrt{0.04}}}_{5.0} \underbrace{\theta\left(\frac{1}{0.04}\right)}_{1.000000000, \nu=1}$$

### 12.G.7 傅里叶分析与卷积

傅里叶分析的又一个重要特征是, 它是以卷积 (convolution) 乘的形式出现的。如上文所示, 由两个方程  $f(t)$  和  $g(t)$  构成的卷积, 近似表示为一系列  $g(\tau)$  的加权和,  $g(\tau)$  是对  $g(t)$  的近似 (copy) 之一, 对应权重  $f(\tau)$ , 我们设这个卷积为  $C(t)$

$$C(t) \equiv f \times g = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau.$$

对  $C(t)$  作傅里叶变换, 有

$$\begin{aligned}\tilde{C}(\omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} C(t) \exp(-i\omega t) d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\tau)g(t-\tau) d\omega \exp(-i\omega t) dt.\end{aligned}$$

<sup>46</sup>快速收敛的原因可能如,  $\pi x$  的值较小, 从而导致  $\sum T_x(t)$  收敛缓慢。

表 12.G.3: 雅各比方程的泊松求和

	LHS		RHS	
$n(\nu)$	$\exp(-n^2\pi x)$	$\sum_n \exp(-n^2\pi x)$	$\exp(-\nu^2\pi/x)$	$\sum_\nu \exp(-\nu^2\pi/x)$
0	1.0000000000	1.0000000000	1.0000000000	5.0000000000
1	0.881911378	2.763822757	7.77304E-35	5.0000000000
2	0.604922563	3.973667882	3.6506E-137	5.0000000000
3	0.322718983	4.619105849	1.0359E-307	5.0000000000
4	0.133905721	4.886917291	0	5.0000000000
5	0.043213918	4.973345128	0	5.0000000000
6	0.010846711	4.995038549	0	5.0000000000
7	0.002117495	4.999273539	0	5.0000000000
8	0.000321512	4.999916562	0	5.0000000000
9	3.79683E-05	4.999992498	0	5.0000000000
10	3.48734E-06	4.999999473	0	5.0000000000
11	2.49126E-07	4.999999971	0	5.0000000000
x=	0.04			

注: 基于式(12.215)计算。excel 表格见文件夹中 data/20180306-poisson.xlsx .

将上式中的  $f(\tau)$  和  $g(t - \tau)$  替换为傅里叶综合形式

$$\begin{aligned}\tilde{C}(\omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} \exp(i\omega_1 \tau) \tilde{f}(\omega_1) d\omega_1 \right\} \left\{ \int_{-\infty}^{\infty} \exp[i\omega_2(t - \tau)] \tilde{g}(\omega_2) d\omega_2 \right\} \exp(-i\omega t) d\tau dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} \exp[i(\omega_1 - \omega_2)\tau] d\tau \right\} \left\{ \int_{-\infty}^{\infty} \exp[i(\omega_2 - \omega)t] dt \right\} \tilde{f}(\omega_1) \tilde{g}(\omega_2) d\omega_1 d\omega_2 \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [2\pi\delta(\omega_1 - \omega_2)][2\pi\delta(\omega_2 - \omega)] \tilde{f}(\omega_1) \tilde{g}(\omega_2) d\omega_1 d\omega_2 \\ &= 2\pi \tilde{f}(\omega) \tilde{g}(\omega),\end{aligned}$$

上式告诉我们， $f$  和  $g$  的卷积的频率为  $\omega$  的傅里叶变换系数  $\tilde{C}(\omega)$ ，就等于  $f$  和  $g$  分别的  $\omega$  频率傅里叶系数之积。这意味着在频率域中进行卷积运算，要比在原始实值域中做卷积运算更方便一些。

## 12.G.8 高维傅里叶变换

更高维度傅里叶变换的方法，与一维傅里叶变换类似。例如对二维方程  $f(x, y)$  而言，第一步保持  $x$  不变，做关于  $y$  的傅里叶变换，可得一个混合了实空间和傅里叶空间的方程  $\tilde{f}(x, k_y)$

$$\tilde{f}(x, k_y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-ik_y y) f(x, y) dy.$$

第二步保持  $k_y$  不变，再对  $\tilde{f}(x, k_y)$  做关于  $x$  的傅里叶变换，得到  $\tilde{f}(k_x, k_y)$

$$\begin{aligned}\tilde{f}(k_x, k_y) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-ik_x x) \tilde{f}(x, k_y) dx \\ &= (2\pi)^{-2} \int_{-\infty}^{\infty} \exp(-ik_x x) \int_{-\infty}^{\infty} \exp(-ik_y y) f(x, y) dy dx \\ &= (2\pi)^{-2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp[-i(k_x x + k_y y)] f(x, y) dy dx \\ &= (2\pi)^{-2} \int \exp(-i \cdot \mathbf{k} \cdot \mathbf{x}) d\mathbf{x},\end{aligned}$$

其中积分符号表示解释变量  $\mathbf{x}$  的全部取值区间。

$D$  维方程也可用类似地方式求得，前缀系数此时变为  $(2\pi)^{-D}$ 。

## 12.G.9 一些指数求和的背景知识

### 连续、有限时间域

假定一个连续、有限的时间区间，宽度为  $T$ ，那么方程  $f(t)$  在傅里叶域中的基准频率为  $\omega_0 = \frac{2\pi}{T}$ 。进而我们有

$$\frac{1}{T} \int_0^T \exp(n_1 - n_2) \omega_0 t dt = \delta(n_1, n_2), \quad (12.216)$$

其中  $\delta(n_1, n_2)$  表示克罗内克乘积，见(12.51)。

### 正交性

(12.216)中，如果  $n_1 \neq n_2$ ，那么我们可以说，方程  $f_{n_1}(t) = \exp(in_1 \omega_0 t)$  和  $f_{n_2}(t) = \exp(in_2 \omega_0 t)$  在内积的意义上正交 (orthogonal)，可表示为

$$\langle f, g \rangle = \frac{1}{T} \int f^* g dt.$$

更多正交性条件的介绍，见第12.C节。

### 连续、无限时间域

在连续、无限时间域中, (12.216)就变成了

$$\int_{-\infty}^{\infty} [i(\omega - \omega')t] dt = 2\pi\delta(\omega - \omega'). \quad (12.217)$$

### 离散时间域

上文在第12.G.7节做泊松求和时, 用到了(12.217), 以处理连续、无限时间域中的求积问题。对应的离散傅里叶变换版本可写作

$$\sum_{n=-\infty}^{\infty} \exp(inkx) = 2\pi \sum_{\nu=-\infty}^{\infty} \delta(xk - 2\nu\pi),$$

RHS 中, 如果  $x$  等于  $\frac{2\pi}{k}$  的整数倍, 那么 LHS 是  $\infty$ :  $\exp(inkx) = 1 \forall n$ ,  $\sum_n 1 = \infty$ 。如果不是整数倍, LHS 的无数个  $\exp(inkx)$  相互抵消, LHS = 0。

来看是整数倍, 进而 LHS =  $\infty$  时的情况。若将左侧进一步扩展为  $\int f(x) \exp(inkx) dx$ (对  $x$  求积分), 那么我们将得到一个  $< \infty$  的数值, 数值的大小取决于  $x = \frac{2\nu\pi}{k}$  时对应的  $f(x)$  的值。

### 高斯求积

第12.G.2节介绍高斯变换时, 使用了一个较为复杂的高斯方程。高斯方程族中, 一个基本的高斯求积问题可以表示为

$$\int_{-\infty}^{\infty} \exp(-x^2) dx = \sqrt{\pi}. \quad (12.218)$$

稍稍做个扩展, 在指数项中放入一个系数  $\alpha$ , 则(12.218)调整为

$$\int_{-\infty}^{\infty} \exp(-\alpha x^2) dx = \sqrt{\frac{\pi}{\alpha}}. \quad (12.219)$$

再继续做扩展, 一个高斯求积

$$I(\alpha, \beta, \gamma) = \int_{-\infty}^{\infty} \exp(-\alpha x^2 + \beta x + \gamma) dx,$$

求解方法如下:

- 提取出常数项  $\gamma$

$$\begin{aligned} I(\alpha, \beta, \gamma) &= \exp(\gamma) \int_{-\infty}^{\infty} \exp\left(\underbrace{-\alpha x^2 + \beta x}_{=-\alpha(x - \frac{\beta}{2\alpha})^2 + \frac{\beta^2}{4\alpha}}\right) dx \\ &= \exp\left(\gamma + \frac{\beta^2}{4\alpha}\right) \int_{-\infty}^{\infty} \exp\left(-\alpha\left[x - \frac{\beta}{2\alpha}\right]^2\right) dx. \end{aligned}$$

- 设  $y \equiv x - \frac{\beta}{2\alpha}$ 。上式变为

$$\begin{aligned} I(\alpha, \beta, \gamma) &= \exp\left(\gamma + \frac{\beta^2}{4\alpha}\right) \underbrace{\int_{-\infty}^{\infty} \exp(-\alpha y^2) dy}_{=\sqrt{\frac{\pi}{\alpha}}} \\ &= \sqrt{\frac{\pi}{\alpha}} \exp\left(\gamma + \frac{\beta^2}{4\alpha}\right). \end{aligned}$$

# 第十三章 扰动法和投影法的比较

## 13.1 优缺点

扰动法

### 1. 显著优势:

- (a) 相对易于理解、易于编程。
- (b) 较高的计算效率。

例如 12 个左右状态变量的 DSGE 模型, 若是采用 3 阶扰动法做近似 (matlab + dynare 平台), 可以在笔记本上实现秒级的求解速度。

### 2. 明显不足:

- (a) 本质上来说还是局部近似。

泰勒级数近似方法有时只有在近似点附近表现良好, 稍微远离近似点就会产生较大误差, 当 DSGE 模型的非线性程度较高时 (常常如此), 这一现象表现的更明显。尽管扰动法近似解也具有一定程度上的全局性特征 (Swanson et al., 2006; Aruoba et al., 2006; Caldara et al., 2012), 但在每一个具体应用中, 都需要针对求得的局部近似解, 作全局的稳健性检验, 这不得不说是一个额外负担。

- (b) 精确度。

对于一些追求高精度近似的 DSGE 研究来说, (哪怕是高阶) 扰动法的近似精度可能无法达标。

- (c) 前提假定过强。

应用扰动法的前提是假定方程是连续、可导的。而连续可导性条件在引入利率的 DSGE 模型中往往不能满足, 往往存在 Kinks、限制条件 (binding constraints) 等, 如图 12.1。<sup>1</sup>。

投影法

### 1. 显著优势:

全局解, 高精度。

这是受益于 (切比雪夫) 多项式和有限元分析法的性质, 使得投影法的近似解是全局解, 并且具有很高的精确度 (Aruoba et al., 2006; Caldara et al., 2012), 这使得它能处理甚至是最复杂的 DSGE 系统, 如包括一些突发限制条件, 非常规形状, 局部特殊性等。

---

<sup>1</sup>对此, 一些研究从模型设定本身来处理不连续的情况, 如引入惩罚方程 (penalty function), 如 Preston and Roca (2007)。事实上, 由许多国家央行主持的引入负利率的政策实验也表明, 零下界 (zero lower bound, ZLB) 的限定条件可能更接近于惩罚方程的一种形式, 而非传统意义上的 kinks。

2. 明显不足

计算成本偏高。

尤其是编写代码的难度，具有很强的维数灾难的特征！

## 13.2 实际研究中用哪个？

在扰动法和投影法一众可选的近似求解方案挑选哪一个展开实际研究，应当视情况而定。

1. 如果要求解一个含有 20 个左右状态变量的中等规模新凯恩斯模型，那么扰动法可能是理想的求解方案。

因为新凯恩斯模型往往定义良好，在大多数情况下，用局部近似求解法足以达到目的。

- 一阶近似可提供对经济周期相关信息的估计，如方差、协方差等。
- 二、三甚至更高阶近似，在经过谨慎处理之后，可提供福利效果分析的估计 (Levintal, 2017a)。

2. 如果要求解含有（金融等）部门摩擦、较高风险厌恶程度的 DSGE 模型，同时状态变量的数量较少，那么更适于采用投影法。

3. 处于实际研究目的的需要，有时甚至可以用扰动法、投影法分别求解同一个系统，比较两者在编程时间、程序计算时间、近似精度等方面的表现。

## 13.3 混合求解法

通过比较扰动法和投影法的优缺点，我们也可以探讨一种混合求解策略，将二者的长处结合起来。如 (Judd, 1998, Sec 5.6) 提出如下混合算法的思路，进而作了举例说明：

1. 用扰动法，构建一组待求解 DSGE 系统的基方程，
2. 应用格拉姆-施密特正交化过程 (Gram-Schmidt ortorthogonality proprocess, 第12.B.2节)，构建另一组基方程，与第一步算得的基方程正交。
3. 对第二步算得的基方程做投影法近似。

近期的一系列研究也采用了混合求解的思路，如Maliar et al. (2012); Fernández-Villaverde and Levintal (2016); Levintal (2017b) 等。Fernández-Villaverde and Levintal (2016) 发现，在计算含有罕见风险的大约 12 个状态变量的 DSGE 模型时，混合策略比单纯的扰动法、投影法都具有更高的精确度。

# 第十四章 有界元法和有限元法

## 14.1 边界值问题：位势方程

我们从二阶偏微分方程入手，介绍边界值问题 (boundary value problem)。一个合适的例子是位势方程 (potential equation)。

### 14.1.1 偏微分算子及椭圆边界值问题

定义有界域  $\Omega \in \mathbb{R}^d, d = 2, 3$ , 边界  $\Gamma = \partial\Omega$ , 外代数单位向量空间 (exterior unit normal vector)  $\underline{n}(x)$  对于  $x \in \Gamma$  几乎处处存在。对于  $x \in \Omega$ , 我们考虑一个线性二阶偏微分的自伴随算子<sup>1</sup> (self-adjoint operator)  $L$ , 作用于实值标量方程  $u$

$$(L u)(x) := - \sum_{i,j=1}^d \frac{\partial}{\partial x_j} \left[ a_{ji}(x) \frac{\partial}{\partial x_i} u(x) \right] + a_0(x) u(x), \quad (14.1)$$

其中  $a_{ji}(x), i, j = 1, \dots, d, x \in \Omega$  表示系数方程, 假定为平滑的并满足  $a_{ij}(x) = a_{ji}(x)$ 。由此可以构建一个对称的系数矩阵  $A(x)$ , 满足

$$A(x) = (a_{ij}(x))_{i,j=1}^d, \quad x \in \Omega,$$

对应实数特征根  $\lambda_k(x)$ 。

当且仅当  $\lambda_k(x) > 0$  对于所有  $k = 1, \dots, d$  都成立时, 我们称偏微分算子  $L$  在某一个  $x \in \Omega$  上是椭圆 (elliptic) 的。

更进一步, 如果  $\forall x \in \Omega$  该条件都成立, 那么我们称  $L$  在  $\Omega$  上是椭圆的。

如果存在一个一致下界 (uniform lower bound)  $\lambda_0 > 0$ , 满足

$$\lambda_k(x) \geq \lambda_0, \quad \forall k = 1, \dots, d, \forall x \in \Omega, \quad (14.2)$$

那么我们称  $L$  在  $\Omega$  上一致椭圆 (uniformly elliptic)。

### 14.1.2 边界条件

边界条件的分析, 可以从散度定理开始。

**Theorem 14.1** (散度定理). 散度定理 (*divergence theorem*), 又称奥斯特罗格拉德斯基——高斯定理 (*Ostrogradsky-Gauss theorem*)、高斯散度定理 (*Gauss' theorem*)等, 是指

$$\int_{\Omega} \frac{\partial}{\partial x_i} f(x) dx = \int_{\Gamma} [\gamma_0^{int} f(x)] n_i(x) ds_x, \quad i = 1, \dots, d, \quad (14.3)$$

---

<sup>1</sup>有限维内积向量空间  $V$  中, 自伴随算子  $A$  是一个从  $V$  到  $V$  自身的线性映射  $\langle Au, \nu \rangle = \langle \nu, Au \rangle, \forall \nu, w \in V$ 。

其中  $\gamma_0^{int} f(x)$  是某个给定方程  $f(x), x \in \Omega$  的内界迹 (interior boundary trace, Theorem 14.10), 满足

$$\gamma_0^{int} f(x) := \lim_{\Omega \ni \tilde{x} \mapsto x \in \Gamma} f(\tilde{x}), \quad \forall x \in \Gamma = \partial\Omega. \quad (14.4)$$

假定两个足够光滑的方程  $u, \nu \in \Omega$ , 通过设定  $f(x) = u(x)\nu(x)$ , 可以将散度定理(14.3)改写为分部积分 (integration by parts)的形式

$$\int_{\Omega} u(x) \frac{\partial}{\partial x_i} \nu(x) x + \int_{\Omega} \nu(x) \frac{\partial}{\partial x_i} u(x) x = \int_{\Gamma} [\gamma_0^{int} u(x)] [\gamma_0^{int} \nu(x)] n_i(x) ds_x.$$

重新调整上式, 将  $\nu(x)$  视作检测方程 (test function), 两侧乘以(14.1)中的二阶偏微分算子  $(L u)(x)$ , 在  $\Omega$  中求积

$$\begin{aligned} (L u)(x) \nu(x) &:= - \sum_{i,j=1}^d \frac{\partial}{\partial x_j} \left[ a_{ji}(x) \frac{\partial}{\partial x_i} u(x) \right] \nu(x) + a_0(x) \underbrace{u(x) \nu(x)}_{\equiv f(x)}, \\ &\hookrightarrow \int_{\Omega} (L u)(x) \nu(x) dx = - \sum_{i,j=1}^d \int_{\Omega} \frac{\partial}{\partial x_j} \left[ a_{ji}(x) \frac{\partial}{\partial x_i} u(x) \right] \nu(x) dx, \end{aligned} \quad (14.5)$$

使用分部积分  $\hookrightarrow$

$$\begin{aligned} \int_{\Omega} (L u)(x) \nu(x) dx &= \underbrace{\sum_{i,j=1}^d \int_{\Omega} a_{ji}(x) \frac{\partial}{\partial x_i} u(x) \frac{\partial}{\partial x_j} \nu(x) dx}_{:= a(u, \nu)} \\ &\quad - \sum_{i,j=1}^d \int_{\Gamma} n_j(x) \left[ \gamma_0^{int}(x) \left( a_{ji}(x) \frac{\partial}{\partial x_i} u(x) \right) \right] [\gamma_0^{int} \nu(x)] ds_x, \end{aligned}$$

由此, 我们由散度定理 (Theorem 14.1) 推导出格林第一恒等式 (Green's first identity)

$$\begin{aligned} a(u, \nu) &:= \sum_{i,j=1}^d \int_{\Omega} a_{ji}(x) \frac{\partial}{\partial x_i} u(x) \frac{\partial}{\partial x_j} \nu(x) dx \\ &= \int_{\Omega} (L u)(x) \nu(x) dx + \sum_{i,j=1}^d \int_{\Gamma} n_j(x) \underbrace{\left[ \gamma_0^{int}(x) \left( a_{ji}(x) \frac{\partial}{\partial x_i} u(x) \right) \right]}_{\gamma_1^{int} u(x)} [\gamma_0^{int} \nu(x)] ds_x, \quad (14.6) \\ &= \int_{\Omega} (L u)(x) \nu(x) dx + \int_{\Gamma} \underbrace{[\gamma_1^{int} u(x)]}_{\gamma_1^{int} u(x)} [\gamma_0^{int} \nu(x)] ds_x, \end{aligned}$$

其中定义  $\gamma_1^{int}$  为内部共形导数 (interior co-normal derivative, Mikhailov (2006, 2009); Ancona (2009))

$$\gamma_1^{int} u(x) := \lim_{\Omega \ni \tilde{x} \mapsto x \in \Gamma} \left[ \sum_{i,j=1}^d n_j(x) a_{ji}(\tilde{x}) \frac{\partial}{\partial \tilde{x}_i} u(\tilde{x}) \right], \quad x \in \Gamma. \quad (14.7)$$

将格林第一恒等式(14.6)中的  $u, \nu$  互换位置, 我们有

$$a(\nu, u) = \int_{\Omega} (L \nu)(x) u(x) dx + \int_{\Gamma} [\gamma_1^{int} \nu(x)] [\gamma_0^{int} u(x)] ds_x,$$

由上式和(14.6)我们有格林第二恒等式 (Green's second identity):  $a(u, \nu) = a(\nu, u) \leftrightarrow$

$$\begin{aligned} &\int_{\Omega} (L u)(x) \nu(x) dx + \int_{\Gamma} [\gamma_1^{int} u(x)] [\gamma_0^{int} \nu(x)] ds_x \\ &= \int_{\Omega} (L \nu)(x) u(x) dx + \int_{\Gamma} [\gamma_1^{int} \nu(x)] [\gamma_0^{int} u(x)] ds_x, \quad \forall u, \nu \in \Omega, \text{且 } u, \nu \text{ 足够平滑}. \end{aligned} \quad (14.8)$$

下面来考虑一个特殊情况， $a_{ij}(x) = \delta_{ij}$ ， $\delta_{ij}$  是克罗内克乘积 (Kronecker product)，见(12.51)。  
(14.1) 的二阶偏微分算子  $(Lu)(x)$  变为拉普拉斯算子

$$(Lu)(x) = -\Delta u(x) := -\sum_{i=1}^d \frac{\partial^2}{\partial x_i^2} u(x), \quad x \in \mathbb{R}^d. \quad (14.9)$$

内部共形导数  $\gamma_1^{int}$  (14.7) 变为

$$\gamma_1^{int} u(x) = \frac{\partial}{\partial n_x} u(x) := \underline{n}(x) \nabla u(x), \quad x \in \Gamma. \quad (14.10)$$

对边界域  $\Gamma = \partial\Omega$  分解成三个不相交集合的并集 (disjoint union)

$$\Gamma = \bar{\Gamma}_D \cup \bar{\Gamma}_N \cup \bar{\Gamma}_R,$$

对应地，边界值问题变为两部分：第一部分，在  $\Omega$  中，基于给定的方程  $f(x)$ ，寻找偏微分算子  $(Lu)(x)$ ，使得

$$(Lu)(x) = f(x), \quad x \in \Omega. \quad (14.11)$$

第二部分，在  $\Gamma$  中，基于给定的方程  $g(x)$ ，寻找内界迹  $\gamma_0^{int} u(x)$  或者内共形导数  $\gamma_1^{int}(x)$ 。随着  $\Gamma$  的取值范围不同，分为三种情况：

$$\gamma_0^{int} u(x) = g_D(x), \quad x \in \Gamma = \Gamma_D, \quad (14.12a)$$

$$\gamma_1^{int} u(x) = g_N(x), \quad x \in \Gamma = \Gamma_N, \quad (14.12b)$$

$$\kappa(x) \gamma_0^{int} u(x) + \gamma_1^{int} u(x) = g_R(x), \quad x \in \Gamma = \Gamma_R. \quad (14.12c)$$

**Definition 14.1** (边界值条件). 于是我们有以下几种不同的边界值条件：

- $\Gamma = \Gamma_D$  : (14.11) + (14.12a) → 狄利克雷边界值条件 (Dirichlet boundary value condition),
- $\Gamma = \Gamma_N$  : (14.11) + (14.12b) → 诺依曼边界值条件 (Neumann boundary value condition),
- $\Gamma = \Gamma_R$  : (14.11) + (14.12c) → 罗宾边界值条件 (Robin boundary value condition),
- 混合边界值条件，以上几种情况的组合。

有时候我们还需要将线性罗宾边界值条件扩展为非线性的情况，(14.12c) →

$$G(\gamma_0^{int} u(x), x) + \gamma_1^{int} u(x) = g_R(x), \quad x \in \Gamma = \Gamma_R, \quad (14.13)$$

其中  $G(u, \cdot)$  是某个给定的非线性方程，如  $u(x)^3$ 。

对于边界值问题的解  $u(x)$ ，还需要注意以下几点

1.  $u(x)$  的存在性和唯一性，相关讨论可参考如Ladyzhenskaya (1968)，
2. 观测到的数据需要是充分平滑的，以确保  $u(x)$  充分可微 (sufficiently differentiable)

$$u \in C^2(\Omega) \cap C^1(\Omega \cup \Gamma_N \cup \Gamma_R) \cap C(\Omega \cup \Gamma_D).$$

### 14.1.3 諾依曼邊界值問題

对于諾依曼邊界值條件的解，其存在性和唯一性需要做进一步讨论。

假定  $\nu_1(x) = 1, x \in \Omega$  是关于  $\nu_1(x)$  的齊次諾依曼邊界值問題的一個解，(14.9)、(14.10)  $\Rightarrow$  关于  $u(x)$  的諾依曼邊界值問題

$$\begin{aligned} (L\nu_1)(x) &= 0, \quad x \in \Omega, \\ \gamma_1^{int}\nu_1(x) &= 0, \quad x \in \Gamma. \end{aligned} \tag{14.14}$$

(14.14) 代入格林第二恒等式(14.8)可得正交条件

$$\int_{\Omega} (Lu)(x) dx + \int_{\Gamma} \gamma_1^{int} u(x) ds_x = 0, \tag{14.15}$$

諾依曼邊界值條件(14.11)、(14.12b) $\Rightarrow$

$$\begin{aligned} (Lu)(x) &= f(x), \quad x \in \Omega, \\ \gamma_1^{int} u(x) &= g_N(x), \quad x \in \Gamma. \end{aligned} \tag{14.16}$$

对于给定的  $f$  和  $g_N$ ，根据正交条件(14.15)，我们可以假设諾依曼問題的可解性条件 (solvability condition)

$$\int_{\Omega} f(x) dx + \int_{\Gamma} g_N(x) ds_x = 0. \tag{14.17}$$

换句话说，如果关于  $\nu(x)$  的齊次諾依曼邊界問題解是  $\nu_1(x) = 1, x \in \Omega$ ，那么关于  $u$  的諾依曼邊界值問題(14.16)的解并不唯一：不只包括一个解  $u(x)$ ，还包括另一个解  $\tilde{u}(x)$ ，满足关系

$$\tilde{u}(x) = u(x) + \alpha, \quad x \in \Omega,$$

其中常数  $\alpha \in \mathbb{R}$  的值是唯一的，取决于为了使第一个解  $u(x)$  成为諾依曼邊界值問題(14.16)的解，而需要在系统中加入的規模調整条件，如

$$\int_{\Omega} u(x) dx = 0, \quad \text{或者} \quad \int_{\Gamma} \gamma_0^{int} u(x) ds_x = 0.$$

## 14.2 方程空间

在进一步介绍邊界值問題的弱形式之前，一些与之紧密相关的方程空间的知识是必需的。相关教材，可参考如McLean (2000); Adolfson et al. (2008b); Tartar (2007); Mazya (2009a,b)。

### 14.2.1 $C^k(\Omega), C^{k,\kappa}(\Omega)$ 空间

给定  $d \in \mathbb{N}$ 。作如下定义：

- 向量 (vector)  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_d), \alpha_i \in \mathbb{N}_0$ 。
- 多重指标 (multi-index) 的绝对值  $|\alpha| = \sum_{i=1}^d \alpha_i$ 。
- 阶乘 (factorial)  $\alpha! = \alpha_1! \alpha_2! \dots \alpha_d!$ 。

给定  $x \in \mathbb{R}^d$  我们有

$$x^{|\alpha|} = x_1^{\alpha_1} x_2^{\alpha_2} \dots x_d^{\alpha_d}.$$

给定一个充分平滑的实值方程  $u$ , 其相对于  $x$  的  $\alpha$  阶偏微分导数

$$D^\alpha u(x) := \left( \frac{\partial}{\partial x_1} \right)^{\alpha_1} \left( \frac{\partial}{\partial x_2} \right)^{\alpha_2} \cdots \left( \frac{\partial}{\partial x_d} \right)^{\alpha_d} u(x_1, x_2, \dots, x_d).$$

给定一个开放子集  $\Omega \subseteq \mathbb{R}^d$ , 对于某个标量  $k \in \mathbb{N}_0$ 。则  $C^k(\Omega)$  表示在  $\Omega$  域中有界且  $k$  次连续可导的方程空间。对于某个方程  $u \in \Omega$ ,  $u$  的范数 (norm) 值是有限的

$$\|u\|_{C^k(\Omega)} := \sum_{|\alpha| \leq k} \sup_{x \in \Omega} |D^\alpha u(x)| < \infty,$$

随着  $k \rightarrow \infty$ ,  $C^\infty(\Omega)$  是个有界且无限阶连续可积的方程空间。

对于方程  $u(x), x \in \Omega$ , 我们将  $u$  的支撑 (support) 定义为  $\text{supp } u$

$$\text{supp } u := \overline{x \in \Omega : u(x) \neq 0}.$$

进而定义  $C_0^\infty(\Omega)$  为  $C^\infty(\Omega)$  中的紧支撑 (compact support) 方程空间。

定义  $C^{k,\kappa}, k \in \mathbb{N}_0, \kappa \in (0, 1)$  为霍德耳连续方程空间 (Hölder continuous function space), 对应范数为

$$\|u\|_{C^{k,\kappa}(\Omega)} := \|u\|_{C^k(\Omega)} + \sum_{|\alpha|=k} \sup_{x,y \in \Omega, x \neq y} \frac{|D^\alpha u(x) - D^\alpha u(y)|}{|x-y|^\kappa}$$

当  $\kappa = 1$  时,  $C^{k,1}$  用来表示第  $|\alpha| = k$  次偏导数  $D^\alpha u(x)$  是利普希茨连续方程 (Lipschitz continuous) 的方程  $u \in C^k(\Omega)$  所组成的空间。

我们用  $\Gamma$  来表示开放集  $\Omega \subset \mathbb{R}^d$  的边界

$$\Gamma := \partial\Omega = \bar{\Omega} \cup (\mathbb{R}^d \setminus \Omega).$$

当  $d \geq 2$ ,  $\Gamma = \partial\Omega$  可以视作利普希茨方程的局部图, 随着所处在  $\Gamma$  中不同的位置, 对应不同的笛卡尔坐标系。一个最简单的例子是假定一个利普希茨方程  $\gamma : \mathbb{R}^{d-1} \mapsto \mathbb{R}$ , 满足

$$\Omega := \{x \in \mathbb{R}^d : x_d < \gamma(\tilde{x}), \quad \forall \tilde{x} = (x_1, \dots, x_{d-1}) \in \mathbb{R}^{d-1}\}.$$

由利普希茨方程的性质我们有,  $\gamma(\cdot)$  满足

$$|\gamma(\tilde{x}) - \gamma(\tilde{y})| \leq L |\tilde{x} - \tilde{y}|, \quad \forall \tilde{x}, \tilde{y} \in \mathbb{R}^{d-1},$$

那么  $\Omega$  被称为一个利普希茨亚图 (Lipschitz hypograph), 对应边界  $\Gamma$

$$\Gamma = \{x \in \mathbb{R}^d : x_n = \gamma(\tilde{x}), \quad \forall \tilde{x} \in \mathbb{R}^{d-1}\}.$$

**Definition 14.2** (利普希茨域). 可参考 [Heinonen \(2005\)](#)。某一个开集  $\Omega \subset \mathbb{R}^d, d \geq 2$  在满足如下条件时, 被称为利普希茨域 (Lipschitz domain), 或有利普希茨边界的域 (domain with Lipschitz boundary):

- $\Omega$  的边界  $\Gamma = \partial\Omega$  是紧凑的, 并且
- $\exists$  有限的索引族 (index family)  $\{W_j\}$  和  $\{\Omega_j\}$ , 满足 <sup>2</sup>

<sup>2</sup>对于集合  $I$  和  $S$ , 某个方程  $x$

$$\begin{aligned} x : I &\mapsto S \\ i &\mapsto x_i = x(i) \end{aligned}$$

被称作用  $I$  索引的  $S$  中元素的族 (family of elements), 也表示为  $\{x_i\}_{i \in I}, x_i \subset S$ 。

- 索引族  $\{W_j\}$  是对  $\Gamma$  的有限开覆盖 (*finite open cover*)

$$W_j \in \mathbb{R}^d, \quad \& \Gamma \subseteq \bigcup_j W_j.$$

- 每一个索引族  $\Omega_j, \forall j$  都可以通过一定的操作变换 (*transformation*) 为利普希茨图, 如旋转 (*rotation*) 和平移 (*translation*) 等。
- $W_j \cap \Omega = W_j \cap \Omega_j, \quad \forall j.$

需要注意的是, 利普希茨边界  $\Gamma = \partial\Omega$  作局部表达的方案, 即对  $W_j$  和  $\Omega_j$  的选取, 通常来讲并不是唯一的<sup>3</sup>。

如果方程  $\gamma(\cdot)$  中参数的选取使得满足  $\gamma \in C^k(\mathbb{R}^{d-1})$ , 我们称这个利普希茨边界  $\Gamma$  是  $k$  次可微的; 如果满足  $\gamma \in C^{k,\kappa}(\mathbb{R}^{d-1})$ , 我们称  $\Gamma$  为霍德尔连续 (Hölder continuous); 如果  $\gamma$  仅仅是在某个局域内满足该条件, 则我们称对应的  $\Gamma$  为分段平滑边界 (piecewise smooth boundary)。

### 14.2.2 勒贝格 $L^p(\Omega)$ 空间

数学上  $L^p(\Omega)$  空间 ( $L^p(\Omega)$  space) 又称勒贝格空间 (Lebesgue space), 指  $\Omega$  上一组测度方程 (measurable function) 的等价类的集合, 这些测度方程都是  $p$  次勒贝格可积方程 (Lebesgue integrable function, Definition 14.3)<sup>4</sup>。

**Definition 14.3** (勒贝格可积方程). 勒贝格可积方程 (*Lebesgue integrable function*) 是指该方程的绝对值的  $p$  次幂的积分是有限的, 如

$$\int_{\Omega} |u(x)|^p dx < \infty.$$

#### $\sigma$ 代数

**Definition 14.4** (sigma 代数). 设  $S$  是一个非空集合, 另有一个集合  $\Sigma$  中的所有元素都是  $S$  的子集, 那么我们将满足以下条件的  $\Sigma$  成为  $S$  上的一个  $\sigma$  代数 ( $\sigma$ -algebra) (Bogachev, 2007a,b, p. 4)

- $S$  在  $\Sigma$  中

$$S \in \Sigma,$$

- 如果一个集合  $A$  在  $\Sigma$  中, 那么它的补集 (*complement*) 也在  $\Sigma$  中

$$A \in \Sigma \Rightarrow A^c \in \Sigma,$$

- 如果  $n$  个集合  $A_1, \dots, A_n$  都在  $\Sigma$  中, 那么他们的并集 (*union*) 也在  $\Sigma$  中

$$(A_n \in \Sigma, \quad \forall n \in \mathbb{N}) \Rightarrow \bigcup_{i=1}^n A_i \in \Sigma.$$

**Definition 14.5** (幂集). 对于任一集合  $S$  的幂集 (*powerset*)是指这样的一个集合, 包括空集  $\emptyset$ 、 $S$  本身和  $S$  的所有集合, 常常表示为  $\mathcal{P}(S)$  或  $2^S$ 。

在公理集合论 (*axiomatic set theory*)例如 ZFC 集合论 (*ZFC axioms*)假定了任何集合的幂集均存在。

$\mathcal{P}(S)$  上的全部子集称为  $S$  上的集族 (*family of sets over S*)。

<sup>3</sup>如非利普希茨的例子, 可参考McLean (2000)。

<sup>4</sup>对应地,  $\ell^p$  空间是由  $p$  次可和序列组成的空间。

假定一个有限集  $S$  有  $n$  个元素, 表示为  $|S| = n$  或  $\text{card}(S) = n$ , 即  $S$  的势 (cardinality) 是  $n$ 。那么  $S$  的幂集里有  $\text{card}(\mathcal{P}(S)) = 2^n$  个元素。例如  $S = \{a, b, c\}$ ,  $\text{card}S = 3$ 。 $S$  的全部子集包括

$$\left\{ \begin{array}{l} \{\{a\}, \{b\}, \{c\}, \\ \{\{a, b\}, \{a, c\}, \{b, c\}, \end{array} \right.$$

因此  $|\mathcal{P}(S)|$  为 6 个  $S$  的子集, 加上  $\emptyset$  和  $S$  自身, 共  $2^3 = 8$  个。

对于非空集合  $S$  来说,  $S$  上的一个  $\sigma$  代数是指其幂集 (powerset, Definition 14.4节) 的一个子集  $\Sigma$ ,  $\Sigma$  中的元素在经过有限个补集、并集、交集 (intersection) 这三种运算后, 依然属于  $\Sigma$ 。即是说,  $\Sigma$  对这三种运算是封闭 (closed) 的。

### 测度, 可测空间, 测度空间

**Definition 14.6** (测度). 对于一个集合  $S$  表示函数的定义域, 有关于  $S$  的可测空间  $\Sigma$ ,  $\Sigma$  中的元素是  $S$  上的子集族 (family of subsets), 并且  $\Sigma$  是一个  $\sigma$  代数 (第 14.4 节)。沿着扩展实数线 (extended real number line)<sup>5</sup> 定义一个测度方程  $\mu \in \Sigma$ , 如果  $\mu$  满足如下条件, 我们称它为一个测度 (measure) :

- 非负性 (non negative)

$$\forall E \in \Sigma : \mu(E) \geq 0,$$

- 空集合的测度为 0

$$\mu(\emptyset) = 0, \quad (14.18)$$

- 可数可加性 (countable additivity), 又称  $\sigma$  可加性 ( $\sigma$  additivity) :  $\{E_i\}_{i=1}^{\infty} = \{E_1, E_2, \dots\}$  为  $\Sigma$  中可数个两两不相交序列的集合 (pairwise disjoint sets in  $\Sigma$ ), 则所有  $E_i$  并集的测度等于每个  $E_i$  的测度之和

$$\mu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{k=1}^{\infty} \mu(E_i).$$

**Definition 14.7** (可测空间). 进而, 我们称  $(S, \Sigma)$  为一个可测空间 (measurable space)。 $\Sigma$  中的所有元素  $\{E_n\}_{n=1}^{\infty}$  成为可测集 (measurable sets)。

**Definition 14.8** (测度空间). 一个三元组  $(\mu, S, \Sigma)$  称为测度空间 (measure space)。测度空间满足如下性质

- 测度  $\mu$  是单调方程 (monotonic)

$$(\text{可测集合 } E_1, E_2 \in \Sigma, \quad E_1 \subseteq E_2) \Rightarrow (\mu(E_1) \leq \mu(E_2)).$$

- 无限个可测集合的并集的测度

- $\mu$  是可数的次可加方程 (countably subadditive)。在  $\Sigma$  中的任何可数可测集合  $\{E_n\}_{n=1}^{\infty}$  (可以不满足两两不相交), 有

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} \mu(E_n).$$

---

<sup>5</sup> 扩展实数线是指实数集合  $\mathbb{R}$  加上  $-\infty$  和  $+\infty$ , 常写作  $\overline{\mathbb{R}}$  或  $[-\infty, +\infty]$ 。

- $\mu$  是连续 (*continuous*) 方程。在  $\Sigma$  中的任何可数可测集合  $\{E_n\}_{i=n}^{\infty}$  (可以不满足两两不相交), 满足  $E_n \subset E_{n+1} \quad \forall n \in \mathbb{N}$ , 则集合的并集  $\bigcup E_n$  也是可测的, 并且满足

$$\mu \left( \bigcup_{n=1}^{\infty} E_n \right) = \lim_{n \rightarrow \infty} \mu(E_n).$$

- 无限个可测集合的交集的测度

- $\mu$  是连续方程。在  $\Sigma$  中的任何可数可测集合  $\{E_n\}_{i=n}^{\infty}$  (可以不满足两两不相交), 满足  $E_n \supset E_{n+1} \quad \forall n \in \mathbb{N}$ , 则集合的交集  $\bigcap E_n$  也是可测的, 并且满足

$$\mu \left( \bigcap_{n=1}^{\infty} E_n \right) = \lim_{n \rightarrow \infty} \mu(E_n),$$

需要指出的是, 对于交集的情况, 若无下述假设, 该性质一般不成立: 可测集合  $\{E_n\}_{n=1}^{\infty}$  中应当至少有一个  $E_n$  有有限测度。举例来说, 如果我们设  $E_n = [n, \infty) \subset \mathbb{R} \quad \forall n \in \mathbb{N}$ , 则这些可测集合全都具有无限测度, 满足  $E_n \supset E_{n+1}$ , 然而

$$\left( \bigcap_{n=1}^{\infty} E_n = \emptyset \right) \Rightarrow \left( \mu \left( \bigcap_{n=1}^{\infty} E_n \right) = \mu(\emptyset) = 0 \right).$$

**Definition 14.9** (计数测度). 在一个测度空间  $(S, \Sigma, \mu)$  中, 任意子集  $E \in \Sigma$  的计数测度 (*counting measure*) 定义为

$$\mu(E) = \begin{cases} \text{card}(E), & \text{如果 } E \text{ 是有限子集} \\ +\infty, & \text{如果 } E \text{ 是无限子集} \end{cases}$$

利用计数测度这种直观的方法, 我们可以在一个测度空间中, 通过将  $S$  的全部可测子集作  $\Sigma$  代数, 从而将  $S$  映射进入这个测度空间中。然而, 只有当空间  $S$  是可数的时, 它在测度空间  $(S, \Sigma, \mu)$  中的计数测度才是  $\sigma$  有限 ( $\sigma$  finite) 的。

## 范数

范数 (norm) 是一个具有“长度”或“大小”概念的方程, 是指赋予一个向量空间 (vector space) 内每个向量以一个非负的长度或大小; 零向量 (zero vector) 的赋值为 0。半范数 (seminorm) 是指可以对某些非零向量赋值 0。

**Definition 14.10** (范数). 假定  $V$  是域  $F$  上的向量空间  $V \subset F$ 。 $V$  上的  $p$  范数 ( $p$ -norm) 是指这样一个方程  $p : V \mapsto \mathbb{R}$ , 使得  $\forall a \in F, \forall \mathbf{u}, \mathbf{v} \in V$ , 以下 4 个关系均得到满足

- 绝对齐次 (*absolutely homogeneous*) 或称绝对可标量化 (*absolutely scalable*), 数乘线性

$$p(a\mathbf{v}) = |a| p(\mathbf{v}),$$

- 次可加 (*subadditivity*) 或称三角不等式 (*triangle inequality*)

$$p(\mathbf{u} + \mathbf{v}) \leq p(\mathbf{u}) + p(\mathbf{v}),$$

- 严格非负

$$p(\mathbf{u}) \geq 0,$$

- 确定 (*definite*)。只有零向量的范数是 0, 反之范数是 0 的向量是零向量。

$$(p(\nu) = 0) \Rightarrow (\nu = 0),$$

即  $\nu$  是个空向量。

只满足前 3 条关系的方程  $p: V \mapsto \mathbb{R}$ , 我们称之为  $V$  上的半范数 (*seminorm*)。换句话说, 所有范都是半范。

**Definition 14.11** (商空间). 每个向量空间  $V$  及其半范  $p$  都生成一个赋范向量空间 (*normed vector space*)  $(\frac{V}{W})$ , 我们称之为商空间 (*quotient space*), 其中  $W$  是  $V$  的子空间  $W \subset V$ , 包括所有满足  $p(\nu) = 0$  的向量  $\nu \in V$ 。

对应地, 商空间中的范数定义为  $p(W + \nu) = p(\nu)$ 。

**Definition 14.12** (等价范). 一个向量空间  $V$  中的两个范 (或两个半范)  $p$  和  $q$ , 当满足下述条件时, 可称为等价范 (*equivalent norms*):

$$cq(\nu) \leq p(\nu) \leq Cq(\nu), \quad \exists \text{实常数} c, C, \quad \forall \nu \in V.$$

**Definition 14.13** (平凡半范). 平凡半范数 (*trivial seminorm*) 是指所有满足如下关系的半范

$$p(\nu) = 0, \quad \forall \nu \in V.$$

一个向量空间  $V$  中的每一个线性泛函  $f$  (linear form), 都定义了 1 个半范  $\nu \mapsto |f(\nu)|$ 。

### 有限维度的可数勒贝格空间 ( $p \geq 1$ )

$x = \{x_1, \dots, x_n\} \in \mathbb{R}^n$  的  $p$  范数 ( $p$ -norm) 或者  $L^p$ ,  $\forall p \geq 1$ , 可定义为

$$\|x\|_p = (\|x_1\|^p + \|x_2\|^p + \dots + \|x_n\|^p)^{\frac{1}{p}},$$

包括一些特殊情况如

- $p = 1$  是个网格距离 (grid distance), 又称出租车距离 (taxicab distance), 曼哈顿距离 (Manhattan distance) 等。
- $p = 2$  是个欧几里得范数 (Euclidean norm)。
- $p = \infty$  是个  $L^\infty$  范数 ( $L^\infty$  norm), 又称最大范数 (maximum norm), 均匀范数 (uniform norm), 切比雪夫距离 (Chebyshev distance) 等。

不同范数  $p$  之间的关系:

- $p = 1$  的曼哈顿范数, 从不小于任何欧几里得范数  $p = 2$ 。换句话说, 任何向量  $x$  的欧几里得范数都受限于它的 1 范数

$$\|x\|_n \leq \|x\|_1, \quad n \geq 1, n \in \mathbb{N}.$$

扩展。任何向量  $x$  的  $p$  范数并不随着  $p$  的增加而增加

$$\|x\|_{p+a} \leq \|x\|_p, \quad \forall \text{向量 } x, \quad \forall p \geq 1, a \geq 0, p, a \in \mathbb{N}.$$

扩展。柯西——施瓦茨不等式 (Cauchy-Schwarz inequality, Definition 14.21) 可得

$$\|x\|_1 \leq \sqrt{n}\|x\|_2, \quad n = \dim(x).$$

扩展。

$$\|x\|_p \leq \|x\|_r \leq n^{(\frac{1}{r} - \frac{1}{p})} \|x\|_p, \quad \forall x \in \mathbb{C}^n, \quad 0 < r < p, r, p \in \mathbb{N}.$$

### 有限维度的可数勒贝格空间 ( $0 \leq p \leq 1$ )

略。

### 无限维度的勒贝格空间 ( $p$ 不可数)

我们设  $p \leq 1$ 。 $p$  范数可以扩展到分析由无数个元素构成的向量，向量集合构成可数无限维的  $p$  范数列空间，用  $l^p$  表示。一些特殊情况如

- $l^1$  由绝对收敛 (absolute convergence) 序列构成的空间，如
  - $\sum_{n=0}^{\infty} |a_n| = L$ , 其中  $L$  是某个实数,  $\{a_n\}_{n=0}^{\infty}$  是一个实数或复数序列。
  - $\int_0^{\infty} |f(x)| dx = L$ , 其中  $\int_0^{\infty} f(x) dx$  是关于某个方程  $f(x)$  的不定积分。
- $l^2$  是一个由平方可加数列构成的空间，即一个希尔伯特空间 (Hilbert space)，见第14.2.3节。
- $l^{\infty}$  是一个由有界数列 (bounded sequence) 构成的空间。

$l^p$  数列空间反映这样的向量空间结构：通过一个坐标一个坐标的向量加总 (或标量相乘) 而组成。如可数无限维实数 (复数) 数列  $\mathbf{x} = \{x_n\}_{n=1}^{\infty}, \mathbf{y} = \{y_n\}_{n=1}^{\infty}$  中的向量和和标量乘

$$(x_1, \dots, x_n, x_{n+1}, \dots) + (y_1, \dots, y_n, y_{n+1}, \dots) = (x_1 + y_1, \dots, x_n + y_n, x_{n+1} + y_{n+1}, \dots), \\ \lambda(x_1, \dots, x_n, x_{n+1}, \dots) = (\lambda x_1, \dots, \lambda x_n, \lambda x_{n+1}, \dots),$$

对应的  $p$  范数

$$\|\mathbf{x}\|_p = (|x_1|^p + \dots + |x_n|^p + |x_{n+1}|^p + \dots)^{\frac{1}{p}}.$$

则我们有定义

**Definition 14.14** (可数无限维数列空间). 定义  $l^p$  为一个包括所有实 (复) 数无限数列的空间 (*countably infinite dimensional sequence space*)，并且这些数列的  $p$  范数必须是有限的：这是因为存在着  $p$  范数是  $\infty$  的无限数列，他们不应当包括在  $l^p$  空间中，如  $(1, 1, 1, \dots)$ ,  $1 \leq p < \infty$ 。

随着  $p$  值的增加， $l^p$  集合的大小增加的更快。例如数列  $\mathbf{x} = (1, \frac{1}{2}, \dots, \frac{1}{n}, \frac{1}{n+1}, \dots)$  为例，它不在  $l^1$  空间中 ( $p = 1$  时范数不收敛)，而可能在某个  $p > 1$  的  $l^p$  的空间中 (范数收敛)

$$\|\mathbf{x}\|_1 = \left( 1 + \frac{1}{2} + \dots + \frac{1}{n} + \frac{1}{n+1} + \dots \right) \rightarrow \infty, \\ \|\mathbf{x}\|_p = \left( 1 + \left(\frac{1}{2}\right)^p + \dots + \left(\frac{1}{n}\right)^p + \left(\frac{1}{n+1}\right)^p + \dots \right)^{\frac{1}{p}}, \quad p > 1.$$

$p = \infty$  时， $l^{\infty}$  范数

$$\|\mathbf{x}\|_{\infty} = \sup(|x_1|, \dots, |x_n|, |x_{n+1}|, \dots),$$

对应  $l^{\infty}$  数列空间，包括全部实 (复) 数无限维数列，但要求他们都是有界的，即  $\infty$  范数收敛。

### 无限维度的不可数勒贝格空间

上面我们讨论了有限维度和可数无限维度的勒贝格空间。然而当空间维度无限并且不可数 (即不存在可数的基) 时，我们无法像前文的方法来定义范数、进而描述空间。但如果该空间是勒贝格可积的 (Lebesgue integrable, 见 Definition 14.3)，仍然可以利用下述办法进行描述。

给定可测空间  $(\Omega, \Sigma, \mu)$  (见 Definition 14.8)，以及  $p \in \mathbb{R}, p \geq 1$ 。考虑所有从  $\Omega$  到域  $\mathbb{F} = (\mathbb{R}, \mathbb{C})$  的可测函数 (measurable function) 集合  $u(x)$ ，方程绝对值的  $p$  次幂在  $\Omega$  上有界，可积

$$L^p(\Omega) = \left\{ u(x); \|u\|_{L^p(\Omega)} \equiv \left( \int_{\Omega} |u(x)|^p dx \right)^{\frac{1}{p}} < \infty \right\},$$

集合中的方程  $u, \nu \in L^p(\Omega)$  具有性质

- 可加

$$(u + \nu)(x) = u(x) + \nu(x),$$

- 数乘线性

$$u(\lambda x) = \lambda u(x), \forall \text{标量 } \lambda \in \mathbb{F},$$

- 范数满足不等式

$$\|u + \nu\|_p^p \leq 2^{p-1} (\|u\|_p^p + \|\nu\|_p^p),$$

如果存在某一个集合  $K$  构成了零测度 (zero measure)  $\mu(K) = 0$ , 那么只有当在这个零测度下  $u, \nu \in L^p(\Omega)$  才有所区别时, 我们说  $u$  和  $\nu$  互相识别。

事实上, 2 个可以扩展到更多个方程的情况, 比如 3 个, 即是说三角不等式 (triangle inequality) 对于范数形式的可积方程  $\|\cdot\|_p$  依然成立, 可由闵可夫斯基不等式 (Minkowski inequality) 证得。它表明  $L^p$  空间是一个赋范向量空间 (normed vector space)。对于一个测度空间  $\Omega \in L^p$ , 设  $1 \leq p \leq \infty$ ,  $u, \nu \in L^p(\Omega)$  中的元素, 此时我们有

**Definition 14.15** (三角不等式).  $L^p(\Omega)$  中的三角不等式 (*triangle inequality*)

$$\|u + \nu\|_p \leq \|u\|_p + \|\nu\|_p, \quad 1 < p < \infty$$

式中等号存在的条件当且仅当  $u$  和  $\nu$  是严格线性相关的, 即  $\exists \lambda \geq 0 \Rightarrow u = \lambda\nu$ , 或者  $\nu = 0$ 。

**Definition 14.16** (闵可夫斯基不等式). 通过取可数测度 (见 Definition 14.9 节), 闵可夫斯基不等式 (*Minkowski inequality*) 可表示为数列和向量的形式

$$\left( \sum_{k=1}^n |x_k + y_k|^p \right)^{\frac{1}{p}} \leq \left( \sum_{k=1}^n |x_k|^p \right)^{\frac{1}{p}} + \left( \sum_{k=1}^n |y_k|^p \right)^{\frac{1}{p}}, \quad \forall \{\mathbf{x}\}_n, \{\mathbf{y}\}_n = \mathbb{R} \text{ 或 } \mathbb{C}, \quad n = \dim(S).$$

证明. 首先要证明如果  $u$  和  $\nu$  都有有限的  $p$  范数, 那么  $(u + \nu)$  的  $p$  范数也是有限的, 且满足不等式关系

$$|u + \nu|_p \leq 2^{p-1} (|u|^p + |\nu|^p)$$

证明方式为: 给定  $p > 1$ , 则  $h(x) = x^p$  是一个在  $\mathbb{R}^+$  上的凸方程 (convex)。由方程的凸性质可得

$$\left| \frac{1}{2}u + \frac{1}{2}\nu \right|^p \leq \left| \frac{1}{2}|u| + \frac{1}{2}|\nu| \right|^p \leq \frac{1}{2}|u|^p + \frac{1}{2}|\nu|^p,$$

→

$$|u + \nu|^p \leq \frac{1}{2}|2u|^p + \frac{1}{2}|2\nu|^p = 2^{p-1} (|u|^p + |\nu|^p),$$

可见  $|u + \nu|_p$  是有限范数。

在证明了  $|u + \nu|_p$  范数有限后, 我们有: 如果  $|u + \nu|_p = 0$ , 闵可夫斯基不等式直接变为等号且成立。现在假定  $|u + \nu|_p \neq 0$ , 使用三角不等式和霍德尔不等式 (Hölder's inequality, Definition

(14.26))，我们有

$$\begin{aligned}
\|u + \nu\|_p^p &= \left[ \left( \int |u + \nu|^p d\mu \right)^{\frac{1}{p}} \right]^p \\
&= \int |u + \nu| |u + \nu|^{p-1} d\mu \\
&\leq \int (|u| + |\nu|) |u + \nu|^{p-1} d\mu \\
&= \int |u| |u + \nu|^{p-1} d\mu + \int |\nu| |u + \nu|^{p-1} d\mu \\
&= \left[ \left( \int |u|^p d\mu \right)^{\frac{1}{p}} + \left( \int |\nu|^p d\mu \right)^{\frac{1}{p}} \right] \left( \int \left( |u + \nu|^{(p-1) \cdot (\frac{p}{p-1})} \right) d\mu \right)^{1-\frac{1}{p}} \\
&= (\|u\|_p + \|\nu\|_p) \frac{\|u + \nu\|_p^p}{\|u + \nu\|_p},
\end{aligned}$$

↪

$$\|u + \nu\|_p \leq \|u\|_p + \|\nu\|_p.$$

□

有时我们需要积分形式的闵可夫斯基不等式 (Minkowski integral inequality):

**Definition 14.17** (闵可夫斯基积分不等式). 假定存在两个  $\sigma$  可测度空间  $(S_1, \mu_1)$  和  $(S_2, \mu_2)$ ，并且  $F : S_1 \times S_2 \mapsto \mathbb{R}$  是可测方程，那么我们有

$$\left[ \int_{S_2} \left| \int_{S_1} F(x, y) \mu_1(dx) \right|^p \mu_2(dy) \right]^{\frac{1}{p}} \leq \left[ \int_{S_1} \left| \int_{S_2} F(x, y) \mu_2(dy) \right|^p \mu_1(dx) \right]^{\frac{1}{p}}, \quad p < \infty,$$

满足该条件的  $\|\cdot\|_p$  构成半范数 (seminorm, Definition 14.10)，对应半赋范向量空间 (semi-normed vector space)  $\mathcal{L}^p(\Omega, \mu)$ 。之所以称之为半范数，是因为该空间中存在非零向量  $u$  满足  $\|u\|_p = 0$ 。

我们可以用标准的拓扑方法，从半范向量空间  $\mathcal{L}^p(\Omega, \mu)$  中得到一个赋范向量空间。在  $\mathcal{L}^p(\Omega, \mu)$  中，考虑所有使  $\|u\|_p = 0$  的向量集合

$$\mathcal{N} = \{u; \|u\|_p = 0\}.$$

$\mathcal{N}$  可以看作是一个映射  $f \mapsto \|f\|_p$  的零向量空间。则对于可测度方程  $u$  而言：

$$\|u\|_p = 0 \iff \mu(u \neq 0) \iff u_{\mu\text{-几乎处处}} = 0,$$

$\mu$ -几乎处处 表示在测度  $\mu$  的意义上几乎处处有界 (almost everywhere)。从这个意义上来看， $\mathcal{N}$  是一个 kernel  $\|\cdot\|_p$ ，并且不依赖于  $p$

$$\mathcal{N} \equiv \ker(\|\cdot\|_p) = \{u : u_{\mu\text{-几乎处处}} = 0\}.$$

则我们可以定义一个关于  $\mathcal{L}^p(\Omega, \mu)$  和 kernel  $\mathcal{N}$  的商空间

$$L^p(\Omega, \mu) \equiv \frac{\mathcal{L}^p(\Omega, \mu)}{\mathcal{N}},$$

商空间  $L^p(\Omega, \mu)$  中的某个  $u$  可以看做是与  $\mathcal{L}^p(\Omega, \mu)$  中的  $f$  相差 1 个  $\mathcal{N}$  中对应元素的等价类。

由此可见， $L^p(\Omega, \mu)$  就是  $\Omega$  上关于测度  $\mu$  的  $L^p$  空间。对应的  $\|\cdot\|_p$  成为  $L^p(\Omega, \mu)$  的  $p$  范数。需要指出的是，严格来说  $L^p$  空间中的元素并非某个具体方程，而是由一个方程族构成的等价类。当我们取出  $L^p$  中的元素作计算的时候，参与计算的其实是从这个方程组中抽取的一个代表方程。

$p = \infty$  时对应的空间  $L^\infty(S, \mu)$  也可以用类似方法求得：

$$\|f\|_\infty \equiv \inf \{C \geq 0 : |f(x)| \leq C \text{ 对于几乎所有 } x\},$$

$$\exists q < \infty \Rightarrow f \in L^\infty(S, \mu) \cap L^q(S, \mu) \Rightarrow \|f\|_\infty = \lim_{p \rightarrow \infty} \|f\|_p.$$

勒贝格空间  $L^p(S, \mu)$  的完备性 (completeness) 通常称为里兹——费舍定理 (Riesz-Fischer theorem)<sup>6</sup>。

对于  $1 \leq p \leq \infty$  的情况，勒贝格空间  $L^p(S, \mu)$  是一个完备赋范向量空间，常称为巴拿赫空间 (Banach space)。所有  $L^p$  空间都是巴拿赫空间。

### 加权勒贝格空间

有时候会遇到加权勒贝格空间的情况。

**Definition 14.18** (加权勒贝格空间). 考虑一个测度空间  $L^p(S, \sigma, \mu)$ ，其中有一个可测方程  $w : S \rightarrow [0, \infty)$ 。有时我们也将  $L^p(S, w d\mu)$  称为  $w$ - 加权勒贝格空间 ( $w$ -weighted Lebesgue space)，其中测度  $d\nu \equiv w d\mu$ 。由此我们有测度的定义

$$\nu(A) \equiv \int_A w(x) d\mu(x), \quad \forall A \in \Sigma.$$

在此基础上，加权勒贝格空间  $L^p(S, w d\mu)$  的范数

$$w = \frac{d\nu}{d\mu} \Rightarrow \|\mu\|_{L^p(S, w d\mu)} \equiv \left( \int_S w(x) |\mu(x)|^p d\mu(x) \right)^{\frac{1}{p}}.$$

从这个角度上说， $L^p(s, d\nu) \equiv L^p(s, w d\nu)$ .

### 14.2.3 希尔伯特 (H) 空间

有且只有  $p = 2$  时的特殊形式空间  $L^2(\Omega)$ ，是希尔伯特空间 (Hilbert space)。

作为 (完备赋范的) 内积向量空间 (inner product space, Definition 14.19)，希尔伯特空间是有限维欧几里得空间的一个扩展：从  $\mathbb{R}$  扩展到  $\mathbb{R}$  和  $\mathbb{C}$ ，从有限维度到无限维度，但保留了完备性 (completeness) 特征 (一般来说，非欧几里得空间往往破坏了完备性)。希尔伯特空间与欧几里得空间相仿，有长度和角度的概念，因而可以引申出正交性和垂直性，从而为基于正交多项式的傅里叶级数等提供表达方式。

任何一个希尔伯特空间都是巴拿赫空间，反之则未必。

#### 例：欧几里得空间

假设所有的希尔伯特空间都是复数 (实际应用中大多数是实数)。二维欧几里得空间  $\mathbb{R}^2$  中，向量  $\mathbf{x}, \mathbf{y}$  构成一个希尔伯特空间

$$\mathbf{x} \cdot \mathbf{y} = \sum_{k=1}^n \overline{x_k} y_k,$$

对应范数

$$\|\cdot\| = \sqrt{\langle \mathbf{x}, \mathbf{y} \rangle}.$$

三维欧几里得空间  $\mathbb{R}^3$  中，以笛卡尔坐标系表示的  $\mathbf{x}, \mathbf{y}$  向量的点乘为

$$\mathbf{x} \cdot \mathbf{y} = (x_1, x_2, x_3) \cdot (y_1, y_2, y_3) = x_1 y_1 + x_2 y_2 + x_3 y_3,$$

点乘具有如下性质：

<sup>6</sup>证明略。完备性 (completeness) 的含义，见第332页。

- 对称性

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x},$$

- 首项线性

$$(ax + by) \cdot z = ax \cdot z + by \cdot z, \quad a, b \text{ 是任意标量}, \quad \mathbf{x}, \mathbf{y}, \mathbf{z} \text{ 是任意向量},$$

- 正定

$$\mathbf{x} \cdot \mathbf{x} \begin{cases} \geq 0 & \forall \mathbf{x} \geq 0, \\ = 0 & \text{iff. } \mathbf{x} = 0. \end{cases}$$

**Definition 14.19** (内积空间). 满足上述三个条件的 (实数) 向量乘称为 (实数) 内积 (*inner product*), 用  $\langle \cdot, \cdot \rangle$  表示。给定一个实数或复数域  $\mathbb{F}$  中的向量空间  $V$ , 则我们将内积形式的向量空间  $\langle \cdot, \cdot \rangle : V \times V \mapsto \mathbb{F}$  表示为内积空间 (*inner product space*)。内积空间满足三个性质:  $\forall$  向量  $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ , 以及  $\forall$  标量  $a \in \mathbb{F}$

- 共轭对称 (*conjugate symmetry*)

$$\langle \mathbf{x}, \mathbf{y} \rangle = \begin{cases} \langle \mathbf{y}, \mathbf{x} \rangle & \mathbb{F} = \mathbb{R}, \\ \overline{\langle \mathbf{y}, \mathbf{x} \rangle} & \mathbb{F} = \mathbb{C}. \end{cases}$$

其中标有上横线  $\overline{(\cdot)}$  的部分表示复数共轭 (*complex conjugate*)。

- 首项线性

$$\begin{aligned} \langle a\mathbf{x}, \mathbf{y} \rangle &= a\langle \mathbf{x}, \mathbf{y} \rangle, \\ \langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle &= \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle. \end{aligned}$$

- 正定

$$\begin{aligned} \langle \mathbf{x}, \mathbf{x} \rangle &> 0, \\ \langle \mathbf{x}, \mathbf{x} \rangle = 0 &\Leftrightarrow \mathbf{x} = 0. \end{aligned}$$

任何有限维内积空间都也是希尔伯特空间。在欧几里得空间内, 两个向量的内积大小与两方面因素有关, 一为向量的长度 (即范数)  $\|\mathbf{x}\|$ , 一为  $\mathbf{x}$  和  $\mathbf{y}$  之间的夹角  $\theta$ , 满足

$$\mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos \theta.$$

欧几里得空间  $\mathcal{R}^3$  中, 对  $n \in \mathcal{N}$  个向量  $\mathbf{x}_n$  求和构成一个数学级数  $\sum_{n=0}^{\infty} \mathbf{x}_n$ , 当每个向量的范之和收敛到一个小于正无穷的向量  $L$  时, 我们称这个级数仍然是绝对收敛 (absolutely convergence) 的

$$\sum_{k=0}^N \|\mathbf{x}_k\| < \infty.$$

一个绝对收敛的向量数列  $\sum_{k=0}^N \mathbf{x}_k$ , 收敛至某个极限向量  $\mathbf{L} \in R^3$

$$\left\| \mathbf{L} - \sum_{k=0}^N \mathbf{x}_k \right\| \rightarrow 0, \text{ 随着 } N \rightarrow \infty,$$

这称为欧几里得空间的完备性 (completeness of Euclidean space)。

类似地，在欧几里得空间中，复数平面 (complex plane)  $\mathbb{C}$  由量 (magnitude) 的形式予以反映，即复绝对值 (complex modulus)  $|z|$ ，定义为  $z$  与其共轭复数 (complex conjugate)  $\bar{z}$  乘积的平方根

$$|z| = \sqrt{z\bar{z}}.$$

如果  $z = x + iy$ ,  $x = \Re(z)$ ,  $y = \Im(z)$ , 复绝对值为常见的二元欧几里得空间的长度

$$\|z\| = \sqrt{\Re(z)^2 + \Im(z)^2} = \sqrt{x^2 + y^2}.$$

两个复数  $z, w$  的内积

$$\langle z, w \rangle = z\bar{w},$$

或者对于复数空间  $z, w \in \mathbb{C}^2$ , 即  $z = (z_1, z_2), w = (w_1, w_2)$ , 对应内积

$$\langle z, w \rangle = z_1\bar{w}_1 + z_2\bar{w}_2,$$

其中  $\Re(\langle z, w \rangle) \in \mathbb{R}^4$ 。这个内积埃米特对称 (Hermitian symmetric)，即是说

$$\langle w, z \rangle = \overline{\langle z, w \rangle}.$$

希尔伯特空间  $H$  是一个实数 (或复数) 内积向量空间，其中的向量可以内积形式表示为  $\langle x, y \rangle$ ，满足如下特性：

- 对称性

$$\langle y, x \rangle = \begin{cases} \langle x, y \rangle & \text{实数向量,} \\ \overline{\langle x, y \rangle} & \text{复数向量.} \end{cases}$$

- 首项线性

$$(ax + by) \cdot z = ax \cdot z + by \cdot z, \quad a, b \text{是任意标量, } x, y, z \text{是任意向量,}$$

- 正定<sup>7</sup>

$$x \cdot x = \begin{cases} \geq 0 & \forall x \geq 0, \\ = 0 & \text{iff. } x = 0. \end{cases}$$

由对称性和首项线性可得第二项系数是反线性的 (antilinear):

$$\langle x, ay + bz \rangle = \bar{a}\langle x, y \rangle + \bar{b}\langle x, z \rangle.$$

希尔伯特空间的范是一个实值方程

$$\|\cdot\| = \sqrt{\langle x, y \rangle}$$

**Definition 14.20** (对偶空间). 已知内积定义

$$\langle u, v \rangle_{\Omega} := \int_{\Omega} u(x) v(x) dx,$$

根据闵可夫斯基不等式 (Definition 14.16),  $L^p(\Omega)$  中的三角不等式可以扩展到更一般的形式：

$$\|v\|_{L^q(\Omega)} = \sup_{0 \neq u \in L^p(\Omega)} \frac{|\langle u, v \rangle_{\Omega}|}{\|u\|_{L^p(\Omega)}}, \quad 1 \leq p < \infty,$$

其中  $p, q$  是伴随参数，满足

$$\frac{1}{p} + \frac{1}{q} = 1,$$

不难看出，RHS 满足三角不等式关系。则  $L^q(\Omega)$  和  $L^p(\Omega)$  构成一组对偶空间 (dual space)。

---

<sup>7</sup>省略部分复数形式的表述，以使方程结构更紧凑。

当  $p = q = 2$  时,  $L^2(\Omega)$  就成为包括全部平方可积方程的空间。此时霍德尔不等式 (Definition 14.26) 就变成了柯西——施瓦茨不等式。

**Definition 14.21** (柯西——施瓦茨不等式). 内积空间 (见 Definition 14.19)  $\langle \cdot, \cdot \rangle : V \times V \mapsto \mathbb{F}$  中, 对于任意两个向量  $\forall \mathbf{u}, \mathbf{\nu} \in V$ , 内积绝对值的平方, 满足三角不等式

$$|\langle \mathbf{u}, \mathbf{\nu} \rangle|^2 \leq \langle \mathbf{u}, \mathbf{u} \rangle \cdot \langle \mathbf{\nu}, \mathbf{\nu} \rangle,$$

两侧同时开平方根, 将 RHS 改写为向量范的形式, 我们有柯西——施瓦茨不等式 (Cauchy-Schwarz inequality)

$$\begin{aligned} |\langle \mathbf{u}, \mathbf{\nu} \rangle| &\leq \|\mathbf{u}\| \|\mathbf{\nu}\|, \\ \hookrightarrow \int_{\Omega} |u(x) \nu(x)| dx &\leq \|u\|_{L^2(\Omega)} \|\nu\|_{L^2(\Omega)}, \\ \hookrightarrow |\langle u, u \rangle|_{L^2(\Omega)} &= \|u\|_{L^2(\Omega)}^2, \forall u = \nu, u \in L^2(\Omega). \end{aligned}$$

其中等号成立的条件, 只有一下两种之一:  $\mathbf{u}, \mathbf{\nu}$  线性无关 (linearly independent), 即平行;  $\mathbf{\nu}$  是零向量或是标量。

证明. 柯西——施瓦茨不等式的证明方法有很多种 (Wu and Wu, 2011), 我们取其中一种。

$$\mathbf{\nu} = 0 \Rightarrow |\langle \mathbf{u}, \mathbf{\nu} \rangle| = \|\mathbf{u}\| \|\mathbf{\nu}\| \forall \mathbf{u} \in V.$$

$\mathbf{u} \neq 0, \mathbf{\nu} \neq 0$ . 设一个向量  $\mathbf{z}$  满足

$$\mathbf{z} := \mathbf{u} - \mathbf{u}_{\mathbf{\nu}} = \mathbf{u} - \underbrace{\frac{\langle \mathbf{u}, \mathbf{\nu} \rangle}{\langle \mathbf{\nu}, \mathbf{\nu} \rangle} \mathbf{\nu}}_{\text{标量}}.$$

对  $\mathbf{z}$  和  $\mathbf{\nu}$  作内积, 由内积空间的性质之一——首项线性得

$$\begin{aligned} \langle \mathbf{z}, \mathbf{\nu} \rangle &= \left\langle \mathbf{u} - \frac{\langle \mathbf{u}, \mathbf{\nu} \rangle}{\langle \mathbf{\nu}, \mathbf{\nu} \rangle} \mathbf{\nu}, \mathbf{\nu} \right\rangle \\ &= \langle \mathbf{u}, \mathbf{\nu} \rangle - \left\langle \frac{\langle \mathbf{u}, \mathbf{\nu} \rangle}{\langle \mathbf{\nu}, \mathbf{\nu} \rangle} \mathbf{\nu}, \mathbf{\nu} \right\rangle \\ &= \langle \mathbf{u}, \mathbf{\nu} \rangle - \frac{\langle \mathbf{u}, \mathbf{\nu} \rangle}{\langle \mathbf{\nu}, \mathbf{\nu} \rangle} \langle \mathbf{\nu}, \mathbf{\nu} \rangle \\ &= 0. \end{aligned}$$

$\langle \mathbf{z}, \mathbf{\nu} \rangle = 0 \Rightarrow \mathbf{z} = 0$ , 作为  $\mathbf{u}$  向  $\mathbf{\nu}$  所在平面 (plane) 所做的正交映射, 反映了  $\mathbf{u}$  和  $\mathbf{\nu}$  线性无关。因此我们对  $\mathbf{z}$  的定义式继续使用勾股定理

$$\begin{aligned} \mathbf{u} &= \frac{\langle \mathbf{u}, \mathbf{\nu} \rangle}{\langle \mathbf{\nu}, \mathbf{\nu} \rangle} \mathbf{\nu} + \mathbf{z} \\ \hookrightarrow \|\mathbf{u}\|^2 &= \left| \frac{\langle \mathbf{u}, \mathbf{\nu} \rangle}{\langle \mathbf{\nu}, \mathbf{\nu} \rangle} \right|^2 \|\mathbf{\nu}\|^2 + \|\mathbf{z}\|^2 \\ &= \frac{\left| \langle \mathbf{u}, \mathbf{\nu} \rangle \right|^2}{(\|\mathbf{\nu}\|^2)^2} \|\mathbf{\nu}\|^2 + \|\mathbf{z}\|^2 \\ &= \frac{\left| \langle \mathbf{u}, \mathbf{\nu} \rangle \right|^2}{\|\mathbf{\nu}\|^2} + \|\mathbf{z}\|^2 \\ &\geq \frac{\left| \langle \mathbf{u}, \mathbf{\nu} \rangle \right|^2}{\|\mathbf{\nu}\|^2} \\ \hookrightarrow |\langle \mathbf{u}, \mathbf{\nu} \rangle| &\leq \|\mathbf{u}\| \|\mathbf{\nu}\| \end{aligned} \tag{14.19}$$

□

### 14.2.4 索伯列夫 $W^{k,p}(\Omega)$ 空间

#### 微分的类以及平滑方程

$H^s(\Omega)$  中包括  $L^p$  空间中的具有弱可导性的平滑方程<sup>8</sup>，常用于求解偏微分方程 PDEs。

**Definition 14.22** (微分的类以及平滑方程). 我们可以根据方程的微分性质，对方程作分类 (*differentiability classification*)。一个实数集合  $\mathbb{R}$  上的开区间中，实值方程  $f \in \mathbb{R}$ 。

如果微分方程  $f', f'', \dots, f^{(k)}$  都存在并且  $f', f'', \dots, f^{(k-1)}$  连续，我们称  $f$  属于  $\mathbb{C}^k$  类方程。当  $k \rightarrow \infty$  时  $f$  的所有  $k$  次微分都存在且连续，我们称之为  $\mathbb{C}^\infty$  类方程，无限可微方程 (*infinitely differentiable*)，或者称之为平滑方程 (*smooth function*)。

如果  $f$  是平滑的，并且  $f$  沿着域中任意一点作泰勒级数展开都收敛至该点，则我们称  $f$  是  $\mathbb{C}^\omega$  类方程，或称之为解析方程 (*analytic function*)。可见  $\mathbb{C}^\omega \subset \mathbb{C}^\infty$ 。

举例来说，

- $\mathbb{C}^0$  中包括所有连续方程，
- $\mathbb{C}^1$  中包括所有一次可微方程，并且这些方程的一次导数是连续的，称连续可导 (*continuously differentiable*)。进而

$$\begin{aligned} & - \forall f \in \mathbb{C}^1 \Rightarrow f' \text{ 存在且 } f' \in \mathbb{C}^0 \\ & - \forall f \in \mathbb{C}^k \Rightarrow f', f'', \dots, f^k \text{ 存在且 } f' \in \mathbb{C}^{k-1} \end{aligned}$$

#### 分部积分

**Definition 14.23** (分部积分公式). 分部积分公式 (*integration by parts formula*)是指，如果  $u = u(x), u'(x) = du/dx$ , 以及  $v = v(x), v'(x) = dv/dx$ , 那么

$$\begin{aligned} \int_a^b u(x)v'(x) dx &= [u(x)v(x)]_a^b - \int_a^b u'(x)v(x) dx \\ &= [u(b)v(b) - u(a)v(a)] - \int_a^b u'(x)v(x) dx, \end{aligned}$$

或者用更紧凑的表现形式

$$\int ud(v) = uv - \int vd(u).$$

#### 广义积分

我们定义  $L^{1,\text{loc}}(\Omega)$  为局部可积 (locally integrable) 的方程空间，即方程  $u \in L^{1,\text{loc}}(\Omega)$  在任意一个封闭有界子集  $K \subset \Omega$  中可导。

例，设  $\Omega = (0, 1)$ ,  $u(x) = \frac{1}{x}$ 。由于

$$\int_0^1 u(x) dx \approx \lim_{\epsilon \rightarrow 0} \int_\epsilon^1 \frac{1}{x} dx \approx \lim_{\epsilon \rightarrow 0} \log \frac{1}{\epsilon} = \infty,$$

可见  $u \notin L^1(\Omega)$ 。可由 Mathematica 算得

<sup>8</sup>“平滑”方程 (smoothness) 的定义包括很多种，由弱到强有

- 连续性，
- 可导性 (可导方程必连续)，
- 它的最高一阶导数也是连续的，

等等。索伯列夫空间中方程设定为“弱”可导形式，是为了使得空间完备，是一个巴拿赫空间。

	1 Limit [Integrate [1/x, {x, ee, 1}], ee -> 0]
	2 Limit [Log[1/x], x -> 0]

对于任一闭区间  $K := [a, b] \subset (0, 1) = \Omega$ ,  $0 < a < b < 1$  我们有

$$\int_K u(x) dx = \int_a^b \frac{1}{x} dx = \ln \frac{b}{a} < \infty,$$

可见  $u \in L^{1,\text{loc}}(\Omega)$ 。

此外对于  $\phi, \psi \in C_0^\infty(\Omega)$ , 根据分部积分 (Definition 14.23) 我们有

$$\int_\Omega \phi(x) \frac{\partial}{\partial x_i} \psi(x) dx = - \int_\Omega \frac{\partial}{\partial x_i} \phi(x) \psi(x) dx,$$

上式对于哪怕是平滑方程  $\phi, \psi$  也适用。由此可得广义偏导数的定义

**Definition 14.24** (广义偏导数). 设  $u \in L^{1,\text{loc}}(\Omega)$ . 如果  $\exists \nu \in L^{1,\text{loc}}(\Omega)$ , 使得满足

$$\int_\Omega \nu(x) \varphi(x) dx = - \int_\Omega u(x) \frac{\partial}{\partial x_i} \varphi(x) dx, \quad (14.20)$$

其中  $\varphi(x) \in C_0^\infty(\Omega)$ , 那么我们说  $\nu(x)$  是  $u(x)$  在  $\Omega$  中关于  $x_i$  的广义偏导数 (*generalized partial derivative*), 写作  $\nu(x) := \partial u(x) / \partial x_i$ 。

类似地,  $u$  的第  $\alpha$  阶广义偏导数  $\nu(x) = D^\alpha u(x)$  记作

$$\int_\Omega u(x) D^\alpha \varphi(x) dx = (-1)^{|\alpha|} \int_\Omega \nu(x) \varphi(x) dx,$$

其中多重指数 (*multi-index*)  $\alpha = (\alpha_1, \dots, \alpha_n)$ ,  $x = (x_1, \dots, x_n)$ ,  $|\alpha| = \alpha_1 + \dots + \alpha_n$ , 积分操作符  $D^\alpha$  是下述形式的缩写

$$D^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \dots x_n^{\alpha_n}},$$

例, 设  $u(x) = |x|$ ,  $x \in \Omega = (-1, 1)$ . 对于任一  $\varphi \in C_0^\infty(\Omega)$ , 我们有

$$\begin{aligned} & \int_{-1}^1 u(x) \frac{\partial}{\partial x} \varphi(x) dx \\ &= - \int_{-1}^0 x \frac{\partial}{\partial x} \varphi(x) dx + \int_0^1 x \frac{\partial}{\partial x} \varphi(x) dx \\ &= \left\{ -[x \varphi(x)]_{-1}^0 + \int_{-1}^0 \varphi(x) dx \right\} + \left\{ [x \varphi(x)]_0^1 - \int_0^1 \varphi(x) dx \right\} \\ &= \int_{-1}^0 \varphi(x) dx - \int_0^1 \varphi(x) dx \\ &= - \int_{-1}^1 \operatorname{sgn}(x) \varphi(x) dx, \end{aligned}$$

其中

$$\operatorname{sgn}(x) := \begin{cases} 1 & x > 0, \\ -1 & x < 0. \end{cases}$$

则方程  $u(x) = |x|$  的 1 阶广义偏导数  $\nu(x)$  为

$$\nu(x) := \frac{\partial}{\partial x} u(x) = \operatorname{sgn}(x) \in L^{1,\text{loc}}(\Omega).$$

递归方法计算 2 阶广义偏导数:

$$\int_{-1}^1 \operatorname{sgn}(x) \frac{\partial}{\partial x} \varphi(x) dx = - \int_{-1}^0 \frac{\partial}{\partial x} \varphi(x) dx + \int_0^1 \frac{\partial}{\partial x} \varphi(x) dx = -2\varphi(0), \quad (14.21)$$

然而  $\nu \in L^{1,\text{loc}}(\Omega)$  满足

$$\int_{-1}^1 \nu(x) \varphi(x) dx = 2\varphi(0), \quad \forall \varphi \in C_0^\infty(\Omega).$$

我们将在随后讨论  $\operatorname{sgn}(x)$  作为分布概念时的广义积分, 见349页第14.2.4节。

### 整数阶次的单维索伯列夫空间

索伯列夫相关教材, 可参考如Adolfson et al. (2008b); Tartar (2007); Mazya (2009a,b) 等。

在  $\mathbb{R}^n$  的一个开放子集  $\Omega$  中, 对于一个给定的非负整数  $k$ , 我们有索伯列夫空间或  $W^{k,p}(\Omega)$ 。它是一种希尔伯特空间的特例, 其中

- (内积形式表示的) 方程向量空间都是可微的,
- 范数是可微方程范数的组合, 包括方程本身的  $L^p$  范数, 以及方程直到某一给定次导数的范数的组合。

单维索伯列夫空间  $W^{k,p}$  中的方程即是  $\mathbb{R}$  中的方程, 可定义为单维勒贝格空间  $L^p(\mathbb{R})$  中方程  $f$  的子集,  $f$  满足如下特征: 对于给定的  $p \in \mathbb{N}, 1 \leq p \leq \infty$ , 其中的  $f^{(k-1)}$  需要几乎处处可导, 并且几乎处处等于其勒贝格积分的  $k-1$  次导数。

$f$  本身以及  $f$  的直至第  $k$  阶弱导数是有限的  $L^p$  范数

$$\begin{aligned} \|f\|_{W^{k,p}} &= \left( \sum_{i=0}^k \|f^{(i)}\|_p^p \right)^{\frac{1}{p}} \\ &= \sum_{i=0}^k \left( \int |f^{(i)}(t)|^p dt \right)^{\frac{1}{p}} \\ &= \|f^{(k)}\|_p + \|f\|_p, \end{aligned}$$

最后一个等式表明, 单维索伯列夫空间的范数, 等于方程序列自身的范数、以及其最高一阶导数的范数之和。

带有这一范数  $\|\cdot\|_{W^{k,p}}$  的单维索伯列夫空间  $W^{k,p}$  是一个巴拿赫空间。

### 整数阶次的单维索伯列夫空间 ( $p=2$ )

$p=2$  的单维索伯列夫空间  $W^{k,2}$  非常重要, 因为它与傅里叶级数关系密切, 并且构成了希尔伯特空间  $H^k = W^{k,2}$ 。

$H^k$  空间可以定义如下: 可由帕塞瓦尔定理 (Parseval theorem) 予以证明 (证明略。)

$$H^k(\mathbb{T}) = \left\{ f \in L^2(\mathbb{T}) : \sum_{n=-\infty}^{\infty} (1 + n^2 + n^4 + \dots + n^{2k}) |\hat{f}(n)|^2 < \infty \right\},$$

其中  $\hat{f}$  是方程  $f$  的傅里叶级数 (Fourier series), 它快速衰减 (decay)。 $\mathbb{T}$  表示环面 (torus)。

此时的  $H^k$  空间可以理解为在  $L^2$  空间中取内积的形式:

$$\langle \mathbf{u}, \mathbf{\nu} \rangle_{H^k} = \sum_{i=0}^k \langle D^i \mathbf{u}, D^i \mathbf{\nu} \rangle_{L^2}.$$

### 多维索伯列夫空间

**Definition 14.25** (索伯列夫空间). 假设  $\exists k \in \mathbb{N}_0, 1 \leq p < \infty$ , 那么多维索伯列夫空间  $W^{k,p}(\Omega)$  定义为在  $\Omega$  上的全部方程集合, 使得对于每一个多元指数 (multiple index)  $\alpha$ , 方程的混合偏导数 (分部积分)  $f^{(\alpha)}$  都存在, 并且  $f^{(\alpha)} \in L^p(\Omega), \|f\|_{L^p(\Omega)} < \infty$ 。从此意义上我们有多维索伯列夫空间  $W^{k,p}(\Omega)$  的定义式:

$$\begin{aligned} W^{k,p}(\Omega) &:= \overline{C^\infty(\Omega)}^{\|\cdot\|_{W^{k,p}(\Omega)}} \\ &= \{u \in L^p(\Omega) : D^\alpha u \in L^p(\Omega), \forall |\alpha| \leq k, k \in \mathbb{N}\}. \end{aligned} \quad (14.22)$$

多维索伯列夫空间  $W^{k,p}(\Omega)$  的范数定义方式有多重, 最常见的两种如下 (并且这两种设定是等价的)

$$\begin{aligned} \|u\|_{W^{k,p}(\Omega)} &:= \begin{cases} \left( \sum_{|\alpha| \leq k} \|D^\alpha u\|_{L^p(\Omega)}^p \right)^{\frac{1}{p}} & 1 \leq p \leq +\infty, \\ \max_{|\alpha| \leq k} \|D^\alpha u\|_{L^p(\Omega)} & p = +\infty. \end{cases} \\ \|u\|'_{W^{k,p}(\Omega)} &:= \begin{cases} \sum_{|\alpha| \leq k} \|D^\alpha u\|_{L^p(\Omega)} & 1 \leq p \leq +\infty, \\ \sum_{|\alpha| \leq k} \|D^\alpha u\|_{L^\infty(\Omega)} & p = +\infty. \end{cases} \end{aligned} \quad (14.23)$$

有着上述定义和范数的无限维索伯列夫空间  $W^{k,p}(\Omega)$  是一个巴拿赫空间。同时, 对于  $p < \infty$  的情况而言, 它也是一个可分空间 (separable space)<sup>9</sup>。

习惯上, 我们将 1 个索伯列夫空间  $W^{k,2}(\Omega)$  写作希尔伯特空间形式  $H^k(\Omega)$ , 对应范数  $\|\cdot\|_{W^{k,2}(\Omega)}$ 。

### 分数次阶索伯列夫空间: Sobolev-Slobodeckij 空间法

关于分数次阶索伯列夫空间, 可参考 Di Nezza et al. (2012)。前面介绍的几种索伯列夫空间的情况, 均假定  $k \in \mathcal{N}$ 。然而有时候我们需要处理  $k$  是分数的情况 (fractional order Sobolev space)。大致说来, 有两种方法可以处理分数次阶索伯列夫空间, 一种是 Sobolev-Slobodeckij 空间法, 一种是贝塞尔位势空间法, 后者需要一些关于分布的傅里叶变换知识, 我们将在随后介绍, 见 352 页 Definition 14.34。

Sobolev-Slobodeckij 空间是勒贝格空间中的霍德耳条件 (Hölder condition, Definition 14.26) 的广义化。

根据  $W^{k,p}(\Omega)$  或者  $\dot{W}^{k,p}(\Omega)$  空间的定义 (14.22) 以及相应的范数 (14.23), 对于  $0 < s \in \mathbb{R}$ , 的情况, 设  $s = k + \kappa, k \in \mathbb{N}_0$ , 我们有 Sobolev-Slobodeckij 范数 (Sobolev-Slobodeckij norm)

$$\|u\|_{W^{s,p}(\Omega)} := \left\{ \|u\|_{W^{k,p}(\Omega)}^p + |u|_{W^{s,p}(\Omega)}^p \right\}^{\frac{1}{p}},$$

其中  $|u|_{W^{s,p}(\Omega)}^p$  是 Sobolev-Slobodeckij 半范数 (Sobolev-Slobodeckij seminorm)

$$|u|_{W^{s,p}(\Omega)}^p = \sum_{|\alpha|=k} \int_{\Omega} \int_{\Omega} \frac{|D^\alpha u(x) - D^\alpha u(y)|^p}{|x-y|^{d+p\kappa}} dx dy.$$

对于  $p = 2$  的情况,  $W^{s,2}(\Omega)$  成为一个内积形式的希尔伯特空间:

$$\langle u, v \rangle_{W^{k,2}(\Omega)} := \sum_{|\alpha| \leq k} \int_{\Omega} D^\alpha u(x) D^\alpha v(x) dx, \quad s = k \in \mathbb{N}_0, \quad (14.24)$$

<sup>9</sup>膏按: FEM!!!!

$$\langle u, \nu \rangle_{W^{s,2}(\Omega)} := \langle u, \nu \rangle_{W^{k,2}(\Omega)} + \sum_{|\alpha|=k} \int_{\Omega} \int_{\Omega} \frac{[D^\alpha u(x) - D^\alpha u(y)][D^\alpha \nu(x) - D^\alpha \nu(y)]}{|x-y|^{d+2\kappa}} dx dy,$$

$s = k + \kappa, k \in \mathbb{N}_0, \kappa \in (0, 1).$

对于  $s < 0, 1 < p < \infty$  的情况，索伯列夫空间  $W^{s,p}(\Omega)$  通过对偶空间  $\dot{W}^{-s,q}(\Omega)$  形式得以定义，其中  $\frac{1}{p} + \frac{1}{q} = 1$ ，对应范数

$$\|u\|_{W^{s,p}(\Omega)} := \sup_{0 \neq \nu \in \dot{W}^{-s,q}(\Omega)} \frac{|\langle u, \nu \rangle_{\Omega}|}{\|\nu\|_{W^{-s,q}(\Omega)}},$$

同样的， $\dot{W}^{-s,p}(\Omega)$  是  $W^{-s,q}(\Omega)$  的对偶空间。

### 索伯列夫空间的性质：嵌入定理

介绍一些索伯列夫空间  $W^{s,p}(\Omega)$  的性质，这些性质有助于更好理解下文介绍的有界元法和有限元法。第一个性质可表示为索伯列夫空间的嵌入定理，我们先从霍德尔不等式入手。

**Definition 14.26** (霍德尔不等式). 霍德尔不等式 (*Hölder inequality*) 可表示为

$$|u(x)\nu(x)|_1 \leq \|u\|_p \|\nu\|_q, \quad 1 \leq p, q \leq \infty, \frac{1}{p} + \frac{1}{q} = 1.$$

证明. 见 Definition 14.27。 □

**Definition 14.27** (广义霍德尔不等式). 霍德尔不等式 (Definition 14.26) 中向量的维度  $n = 1$ 。对于  $n \geq 2$  的情况，如  $\{u_i\}_{i=1}^n$ ，我们有广义霍德尔不等式 (*generalized Hölder inequality*)

$$\int_{\Omega} \left| \prod_{i=1}^n u_i(x_i) \right| \leq \prod_{i=1}^n \|u_i\|_{p_i}, \quad p_i \geq 1, \sum_{i=1}^n \frac{1}{p_i} = 1. \quad (14.26)$$

证明.  $n = 1$  时。略。

$n \geq 2$  并且若  $(n-1)$  的情况。已知满足霍德尔不等式条件，我们的目标就成了，希望知道  $n$  的情况是否依然满足。分两种情况来分析。首先来看  $n < \infty$ ，对应  $p_n < \infty$ 。设

$$p := \frac{p_n}{p_n - 1}, \quad q := p_n \Rightarrow \frac{1}{p} + \frac{1}{q} = 1.$$

此时我们有

$$\begin{aligned} \left\| \prod_{i=1}^n u_i \right\|_1 &\leq \left\| \prod_{i=1}^{n-1} u_i \right\|_p \left\| \prod_{i=1}^n u_n \right\|_q \\ &\leq \left( \left\| \prod_{i=1}^{n-1} u_i^p \right\|_1 \right)^{\frac{1}{p}} \left\| u_n \right\|_{p_n}, \end{aligned}$$

我们设  $p'_i := \frac{p_i}{p}$ ，上式进一步变为

$$\left\| \prod_{i=1}^n u_i \right\|_1 \leq \left( \left\| \prod_{i=1}^{n-1} u_i^p \right\|_{p'_i} \right)^{\frac{1}{p}} \left\| u_n \right\|_{p_n},$$

进一步，已知

$$\left( \left\| \prod_{i=1}^{n-1} u_i^p \right\|_{p'_i} \right)^{\frac{1}{p}} \equiv \left[ \prod_{i=1}^{n-1} |u_i|^{(p \cdot p'_i)} \right]^{\left( \frac{1}{p_i} \cdot \frac{1}{p} \right)} = \prod_{i=1}^{n-1} \|u_i\|_{p_i},$$

则上式进一步变为

$$\left\| \prod_{i=1}^n u_i \right\|_1 \leq \prod_{i=1}^{n-1} \|u_i\|_{p_i} \|u_n\|_{p_n} = \prod_{i=1}^n \|u\|_{p_i}, \quad 2 \leq n < \infty.$$

$n = \infty$  的情况,

$$\because p_n = \infty \Rightarrow \sum_{i=1}^{n-1} \frac{1}{p_i} = 1,$$

$$\begin{aligned} \therefore \left\| \prod_{i=1}^n u_i \right\| &\leq \left\| \prod_{i=1}^{n-1} u_i \right\|_1 \|u_n\|_\infty \\ &\leq \prod_{i=1}^n \|u\|_{p_i}, \quad n = \infty. \end{aligned}$$

□

第一个性质可表示为索伯列夫空间的嵌入定理 (embedding theorem of Sobolev)。

**Definition 14.28** (嵌入). 数学上, 嵌入 (*embedding*) 是指某个物件 (*instance*)  $X$  被嵌入到另一个物件  $Y$  中去, 用保留结构的映射 (*structure-preserving map*)  $f : X \mapsto Y$  表示。这里的物件指数学结构, 如群、子群等。所保留的具体数学“结构”因物件  $X$  和  $Y$  的种类而异。如在范畴论 (*category theorem*) 中, 一个保留结构的映射往往称为一个态射 (*morphism*)。

**Theorem 14.2** (索伯列夫空间的嵌入定理  $C = C(n, p)$ ). 给定有界的开放集  $\Omega \subset \mathbb{R}^n$ , 则

$$W_0^{1,p}(\Omega) \subset L^{\frac{np}{n-p}}(\Omega), \quad n \geq 3, 1 \leq p < n,$$

并且  $W_0^{1,p}(\Omega)$  被连续嵌入到空间  $L^{\frac{np}{n-p}}(\Omega)$  中, 以下一组关系始终满足

$$\|u\|_{L^{\frac{np}{n-p}}(\Omega)} \leq C(n, p) \|Du\|_{L^p(\Omega)}, \quad \forall u \in W_0^{1,p}(\Omega), \quad (14.27)$$

其中  $C(n, p) \in (0, +\infty)$  是个和  $n, p$  有关的常数,  $Df = (D^{e_1}u, \dots, D^{e_n}u) \in L^p(\Omega) \times \dots \times L^p(\Omega)$ 。

证明. 参考 (Brenner and Scott, 2008, Theorem 1.4.6), (McLean, 2000, Theorem 3.26)。

1. 证明对于  $\forall u \in C_0^\infty$  来说, (14.27) 成立。由广义霍德尔不等式 (Definition 14.27) 我们有, 如果存在方程  $u(x) \in L^p(\Omega)$ , 满足  $u_j(x) \in L^{p_j}(\Omega), j = 1, \dots, m$ , 并且  $\sum_{j=1}^m \frac{1}{p_j} = 1$ , 那么

$$\int_{\Omega} u_1(x) \dots u_m(x) dx \leq \|u_1\|_{L^{p_1}(\Omega)} \dots \|u_m\|_{L^{p_m}(\Omega)} \quad (14.28)$$

2. 在  $p = 1$  时, 由  $u \in C_0^\infty(\Omega)$  可知, 方程  $u(x)$  可以表示为下述积分形式

$$\begin{aligned} u(x) &= \int_{-\infty}^{x_i} D^{e_i} u(x_1, \dots, x_{i-1}, t, x_{i+1}, \dots, x_n) dt, \\ \hookrightarrow |u(x)| &\leq \int_{-\infty}^{x_i} |D^{e_i} u| dt \leq \int_{-\infty}^{\infty} |D^{e_i} u| dx_i, \\ \hookrightarrow |u(x)|^n &\leq \prod_{i=1}^n \int_{-\infty}^{\infty} |D^{e_i} u| dx_i, \end{aligned}$$

从而我们有

$$|u(x)|^{\frac{n}{n-1}} \leq \left( \prod_{i=1}^n \int_{-\infty}^{\infty} |D^{e_i} u| dx_i \right)^{\frac{1}{n-1}}. \quad (14.29)$$

下面对数列  $x = \{x_i\}_{i=1}^n$  求偏导。先从  $x_1$  开始, 根据(14.29)我们有

$$\begin{aligned} \int_{-\infty}^{\infty} |u(x)|^{\frac{n}{n-1}} dx_1 &\leq \int_{-\infty}^{\infty} \left( \prod_{i=1}^n \int_{-\infty}^{\infty} |D^{e_i} u| dx_i \right)^{\frac{1}{n-1}} dx_1 \\ &= \left( \int_{-\infty}^{\infty} |D^{e_1} f| dx_i \right)^{\frac{1}{n-1}} \int_{-\infty}^{\infty} \left( \prod_{i=2}^n \int_{-\infty}^{\infty} |D^{e_i} u| dx_i \right)^{\frac{1}{n-1}} dx_1 \quad (14.30) \\ &\leq \left( \int_{-\infty}^{\infty} |D^{e_1} f| dx_i \right)^{\frac{1}{n-1}} \prod_{i=2}^n \left( \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |D^{e_i} u| dx_i dx_1 \right)^{\frac{1}{n-1}}. \end{aligned}$$

在(14.30)的基础上, 继续对  $x_2, \dots, x_n$  作偏导, 可得

$$\int_{\mathbb{R}^n} |u(x)|^{\frac{n}{n-1}} dx \leq \left( \prod_{i=1}^n \int_{\mathbb{R}^n} |D^{e_i} u| dx_i \right)^{\frac{1}{n-1}}. \quad (14.31)$$

(14.31)  $\Rightarrow$

$$\begin{aligned} \|u(x)\|_{L^{\frac{n}{n-1}}(\Omega)} &\leq \left( \prod_{i=1}^n \int_{\mathbb{R}^n} |D^{e_i} u| dx_i \right)^{\frac{1}{n}} \\ &\leq \int_{\Omega} \left( \sum_{i=1}^n |D^{e_i} u| \right) dx \\ &\leq \frac{1}{\sqrt{n}} \int_{\Omega} |Du| dx \\ &= \frac{1}{\sqrt{n}} \|Du\|_{L^1(\Omega)}, \quad \forall u \in C_0^\infty(\Omega). \end{aligned}$$

3. 下面考虑  $1 < p < n$  的情况。已知  $u \in C_0^\infty(\Omega)$ , 则用  $|u|^\gamma$ ,  $\gamma > 1$  代替(2)中的  $u$ , 我们有

$$\| |u|^\gamma \|_{L^{\frac{n}{n-1}}(\Omega)} \leq \frac{1}{\sqrt{n}} \int_{\Omega} |D|u|^\gamma| dx = \frac{\gamma}{\sqrt{n}} \int_{\Omega} |u|^{\gamma-1} |Du| dx, \quad (14.32)$$

设  $\frac{1}{p} + \frac{1}{q} = 1$ ,  $p, q > 0$ , 由霍德尔不等式 (Definition 14.27) 得, (14.32)  $\Rightarrow$

$$\begin{aligned} \| |u|^\gamma \|_{L^{\frac{n}{n-1}}(\Omega)} &\leq \frac{\gamma}{\sqrt{n}} \int_{\Omega} |u|^{\gamma-1} |Du| dx \leq \frac{\gamma}{\sqrt{n}} \| |u|^{\gamma-1} \|_q \| Du \|_p, \\ \hookrightarrow \| |u| \|_{L^{\frac{\gamma n}{n-1}}(\Omega)}^\gamma &\leq \frac{\gamma}{\sqrt{n}} \| |u| \|_{L^{q(\gamma-1)}(\Omega)}^{\gamma-1} \| Du \|_p. \end{aligned} \quad (14.33)$$

下面选取  $\gamma$  的值, 使得

$$\gamma \frac{n}{n-1} \equiv (\gamma-1)q \implies \gamma = \frac{p(n-1)}{n-p} \implies \gamma \frac{n}{n-1} \equiv (\gamma-1)q = \frac{np}{n-p},$$

代入(14.33), 调整得

$$\|u\|_{L^{\frac{np}{n-p}}(\Omega)} \leq \frac{\gamma}{\sqrt{n}} \|Du\|_p, \quad \forall u \in C_0^\infty(\Omega),$$

联系(14.27)可得  $C(n, p)$  的值

$$C(n, p) = \frac{np - n}{\sqrt{n}(n-p)}.$$

4. 前面的论证过程均假设  $u \in C_0^\infty$ 。如果现在假设  $u_\ell \in W_0^{1,p}(\Omega)$ , 则我们利用  $u_m \in C_0^\infty(\Omega)$  来近似  $u_\ell \in W^{1,p}(\Omega)$ , 将  $u_\ell - u_m$  代入(14.27)中。可见  $\{u_m\} \in L^{\frac{np}{n-p}}(\Omega)$  是一个柯西数列。因此我们也能证明  $u_\ell \in L^{\frac{np}{n-p}}(\Omega)$ , 并且满足条件(14.27)。

□

前面讨论的都是  $k = 1$  时,  $n, p$  变化,  $C = C(n, p)$ 。现在来将  $k$  的变化也考虑进来,  $C = C(k, n, p)$ 。

**Corollary 14.1** (索伯列夫空间的嵌入定理  $C = C(k, n, p)$ ). 如果  $kp < n$ , 则索伯列夫空间  $W_0^{k,p}(\Omega)$  连续嵌入到  $L^{\frac{np}{n-kp}}(\Omega)$  中, 对应常数  $C(k, n, p)$  满足

$$\|u\|_{L^{\frac{np}{n-kp}}(\Omega)} \leq C(k, n, p) \|u\|_{W_0^{k,p}(\Omega)}. \quad (14.34)$$

证明. 假设  $kp < n$ ,  $u \in W_0^{k,p}(\Omega)$ 。

- 根据  $D^\alpha u \in L^p(\Omega)$ ,  $\forall |\alpha| \leq k$ , 从索伯列夫不等式可得, 对于  $|\beta| \leq k - 1$

$$\|D^\beta u\|_{L^{p^*}(\Omega)} \leq C \|u\|_{W^{k,p}(\Omega)}, \quad p^* := \frac{np}{n-p}$$

因此有  $u \in W^{k-1,p^*}(\Omega)$ 。

- 用类似的方法, 我们可以进一步证明  $u \in W^{k-2,p^{**}}(\Omega)$ , 对于  $|\gamma| \leq k - 2$

$$\|D^\gamma u\|_{L^{p^{**}}(\Omega)} \leq C \|u\|_{W^{k,p^*}(\Omega)}, \quad p^{**} := \frac{1}{p^*} - \frac{1}{n} = \frac{1}{p} - \frac{2}{n}.$$

- 以此类推, 最终经过  $k$  次迭代后, 我们得以证明, 对于  $u \in W^{0,q}(\Omega) = L^q(\Omega)$ , (14.34) 成立

$$\|u\|_{L^{\frac{np}{n-kp}}(\Omega)} \leq C \|u\|_{W_0^{k,p}(\Omega)}, \quad \frac{1}{q} = \frac{1}{p} - \frac{k}{n}.$$

□

前面讨论的是将  $W_0^{k,p}(\Omega)$  嵌入到  $L^p(\Omega)$  中去。如果  $\Omega$  是一个  $C^k - domain$ , 则我们可以利用延拓算子  $E$ , 将索伯列夫嵌入定理从  $W_0^{k,p}(\Omega)$  延伸到  $W^{k,p}(\Omega)$  空间。

即是说, 如果  $u \in W^{k,p}(\Omega)$ , 则可以考虑一个  $Eu \in W_0^{k,p}(\Omega')$ , 其中  $\Omega'$  是  $\Omega$  的延拓 (extension),  $\Omega$  是  $\Omega'$  的限制 (restriction):  $\Omega' \supset \Omega$ 。进而, 如果  $kp < n$ , 那么  $W_0^{k,p}(\Omega') \in L^{\frac{np}{n-kp}}(\Omega')$ 。

因此  $u$  属于限制  $\Omega$  所对应的勒贝格空间:  $u \in L^{\frac{np}{n-kp}}(\Omega)$ 。

一方面在  $\Omega$  中我们有  $Eu = u$ , 另一方面根据  $\Omega$  的具体情况我们有  $\|Eu\|_{W^{k,p}(\Omega')} \leq C \|u\|_{W^{k,p}(\Omega)}$ 。所以我们有

**Corollary 14.2** (索伯列夫空间的嵌入定理  $C = C(k, n, p, \Omega)$ ). 设  $\Omega \subset \mathbb{R}^n$  是一个有界的  $C^k - domain$ 。如果  $kp < n$ , 一系列索伯列夫空间  $W^{k,p}(\Omega)$  被依次嵌入到对应的  $L^{\frac{np}{n-kp}}(\Omega)$  空间中去。这几乎是说, 存在某个常数  $C = C(k, n, p, \Omega)$ , 使得

$$\|u\|_{L^{\frac{np}{n-kp}}(\Omega)} \leq C \|u\|_{W^{k,p}(\Omega)}, \quad \forall u \in W^{1,p}(\Omega). \quad (14.35)$$

证明. □

### 索伯列夫空间的性质: 范数等价

第二个性质是索伯列夫空间的范数等价定理 (norm equivalence theorem of Sobolev)。先来介绍范数等价的定义, 以及一般意义上的范数等价定理。

**Definition 14.29** (范数等价). 一个向量空间  $V \in \mathbb{F} = \mathbb{R} \cup \mathbb{C}$  中有两个范数  $\|\cdot\|_a$  和  $\|\cdot\|_b$ 。范数等价 (*equivalence of norms*) 是指,  $\|\cdot\|_b$  总是在  $\|\cdot\|_a$  的某乘数倍范围之内, 换句话说: 存在正常数  $c, C$  使得  $\forall x \in V$ , 都满足

$$c\|x\|_a \leq \|x\|_b \leq C\|x\|_a.$$

**Theorem 14.3** (范数等价定理).  $\exists V \in \mathbb{F}^\infty$ , 其中  $\nu_1, \dots, \nu_n$  是  $V$  的基,  $n$  是  $V$  的维度。因此每个  $x \in V$  都有唯一的表示形式, 如

$$x = \sum_{i=1}^n x_i \nu_i, \quad (14.36)$$

上述内积形式中,  $\{\nu_i\}_{i=1}^n$  是标准正交基,  $\{x_i\}_{i=1}^n$  是一组关于  $\{\nu_i\}_{i=1}^n$  的坐标 (标量)。

定义  $V$  中的范数为

$$\|x\|_* := \max_{i=1, \dots, n} |x_i|,$$

我们将有范数  $\|x\|_*$  的  $V$  称为完整向量空间 (*complete*)。

则范数等价定理为:  $V \in \mathbb{F}^\infty$  中的所有范是等价的。

证明. 使  $\|x\|$  为  $V$  中任一范数。我们的目标是证明  $\|x\|$  和  $\|x\|_*$  等价, 即存在常数  $c, C$  使得不等式(14.36)成立。分两步来证明。

第一步来看对  $\|x\| \leq C\|x\|_*$  的证明。根据定义我们有

$$\begin{aligned} \|x\| &= \left\| \sum_{i=1}^n x_i \nu_i \right\| \\ &\leq \sum_{i=1}^n \|x_i \nu_i\| = \sum_{i=1}^n |x_i| \|\nu_i\| \\ &\leq \underbrace{n \max_{i=1, \dots, n} \|\nu_i\|}_{\text{常数}} \max_{i=1, \dots, n} |x_i| \end{aligned}$$

设常数  $C := n \max_{i=1, \dots, n} \|\nu_i\|$ , 则上式变为

$$\begin{aligned} \|x\| &\leq C \max_{i=1, \dots, n} |x_i|, \\ \hookrightarrow \|x\| &\leq C\|x\|_*. \end{aligned}$$

第二步, 来看  $c\|x\|_* \leq \|x\|$  的证明。首先需要证明  $S = \{x : \|x\|_* = 1\}$  是个紧密空间 (*compact space*), 方法为证明  $S$  完整且完全有界 (*totally bounded*)。

完整性的证明: 由三角不等式 (Definition 14.15) 得

$$\|x\|_* - \|y\|_* \leq \|x - y\|_*,$$

$\hookrightarrow$  方程  $x : x \mapsto \|x\|_*$  是连续的。

$\hookrightarrow$  连续方程  $x$  构成的向量空间  $S$  是  $V$  中的封闭子集。

$\hookrightarrow$  已知  $V$  是个完整空间,  $\therefore V$  的封闭子集  $S$  也是个完整空间。

完全有界的证明: 设  $\varepsilon > 0$ , 选则常数  $m > 1/\varepsilon$ 。因此,  $S$  被  $O(m^d) < \infty$  个球体  $B$  的集合所覆盖, 这些球都是半径为  $1/m$ , 范数为  $\|\cdot\|_*$ , 表示为

$$B_{j_1, \dots, j_n} := \left\{ \sum_{i=1}^n x_i \nu_i : \frac{j_i}{m} - \varepsilon < x_i < \frac{j_i}{m} + \varepsilon, \quad i = 1, \dots, m, \quad j_i = -m, \dots, m, \right\}$$

$\therefore S$  完全有界。

其次，设方程  $f(x) = \|x\|, x \in [0, \infty)$ 。对于  $x \in S$ ，我们有  $f(x)$  也是一个连续方程，这是由于

$$\begin{aligned}\|f(x) - f(y)\| &\leq \|\|x\| - \|y\|\| \\ &\leq \|x - y\| \\ &\leq C \|x - y\|_*\end{aligned}$$

$\therefore$  紧密空间  $S$  中的方程  $f$  有最小值。

最后，由  $x \in S$  得，设  $c := \min_{x \in S} f(x) > 0$ 。 $x \in S \Rightarrow \frac{x}{\|x\|_*} \in S \Rightarrow$

$$f\left(\frac{x}{\|x\|_*}\right) \equiv \left\| \frac{x}{\|x\|_*} \right\| \geq c,$$

$\therefore \|x\| \geq c\|x\|_*$ . □

当  $k = 1, p = 2$  时，索伯列夫空间  $W^{1,2}(\Omega)$  的范数(14.23)变为

$$\begin{aligned}\|u\|_{W^{1,2}(\Omega)} &= \left\{ \|Du\|_{L^2(\Omega)}^2 \right\}^{\frac{1}{2}} \\ &= \left\{ \|u\|_{L^2(\Omega)}^2 + \|\nabla u\|_{L^2(\Omega)}^2 \right\}^{\frac{1}{2}},\end{aligned}$$

并且半范数为

$$|u|_{W^{1,2}(\Omega)} = \|\nabla u\|_{L^2(\Omega)},$$

则我们有索伯列夫空间的范数等价定理。

**Theorem 14.4** (索伯列夫空间的范数等价定理). 设一个有界的线性方程  $f : W^{1,2}(\Omega) \mapsto \mathbb{R}$ ，满足

$$0 \leq |f(u)| \leq c_f \|u\|_{W^{1,2}(\Omega)}, \quad \forall u \in W^{1,2}(\Omega),$$

$c_f > 0$  是个常数。如果对于某个常数  $\iota$ ,  $f(\iota) = 0$  iff.  $\iota \equiv 0$ ，那么在  $W^{1,2}(\Omega)$  空间中，所有满足上述条件的方程  $f$  的范数是等价的：

$$\|u\|_{W^{1,2}(\Omega),f} := \left\{ |f(u)|^2 + \|\nabla u\|_{L^2(\Omega)}^2 \right\}^{\frac{1}{2}}. \quad (14.37)$$

证明. 首先证明  $f(u) \leq c_f \|u\|_{L^2(\Omega)}$ : 根据前提假定， $f$  是一个线性有界方程，则

$$\begin{aligned}\|u\|_{W^{1,2}(\Omega),f}^2 &= \underbrace{|f(u)|^2}_{\leq c_f^2 \|u\|_{W^{1,2}(\Omega)}^2} + \underbrace{\|\nabla u\|_{L^2(\Omega)}^2}_{\leq \|u\|_{W^{1,2}(\Omega)}^2} \leq (1 + c_f^2) \|u\|_{L^2(\Omega)}^2,\end{aligned}$$

则我们有

$$f(u) \leq c_f \|u\|_{L^2(\Omega)}.$$

第二步，对  $|f(u)| \geq 0$  的证明，较为间接。已知  $\#c_0$  使得满足下述条件

$$\|u\|_{W^{1,2}(\Omega)} \leq c_0 \underbrace{\|u\|_{W^{1,2}(\Omega),f}}_{\equiv \|f(u)\|},$$

则假设存在一个数列  $\{u_n\}_{n \in \mathbb{N}} \in W^{1,2}(\Omega)$ ，满足

$$n \leq \frac{\|u_n\|_{W^{1,2}(\Omega)}}{\|u_n\|_{W^{1,2}(\Omega),f}}.$$

把  $\{u_n\}_{n \in \mathbb{N}}$  标准化为  $\{\bar{u}_n\}_{n \in \mathbb{N}}$

$$\{\bar{u}_n\}_{n \in \mathbb{N}} := \frac{u_n}{\|u_n\|_{W^{1,2}(\Omega)}},$$

从而使得我们有

$$\begin{aligned}\|\bar{u}_n\|_{W^{1,2}(\Omega)} &= 1, \\ \|\bar{u}_n\|_{W^{1,2}(\Omega),f} &= \frac{\|u_n\|_{W^{1,2}(\Omega),f}}{\|u_n\|_{W^{1,2}(\Omega)}} \leq \frac{1}{n},\end{aligned}$$

那么随着  $n \rightarrow \infty$ ,

$$\lim_{n \rightarrow \infty} \|\bar{u}_n\|_{W^{1,2}(\Omega),f} = \lim_{n \rightarrow \infty} |f(\bar{u}_n)| \leq \lim_{n \rightarrow \infty} \left( \frac{1}{n} \right) = 0,$$

代回(14.37)有

$$\begin{aligned}\lim_{\bar{u}_n \rightarrow \infty} \|u\|_{W^{1,2}(\Omega),f} &= \lim_{n \rightarrow \infty} \left\{ |f(\bar{u}_n)|^2 + \|\nabla \bar{u}_n\|_{L^2(\Omega)}^2 \right\}^{\frac{1}{2}}, \\ &\hookrightarrow \lim_{n \rightarrow \infty} \|\nabla \bar{u}_n\|_{L^2(\Omega)} = 0.\end{aligned}$$

由于标准化数列  $\{\bar{u}_n\}_{n \in \mathbb{N}}$  在  $W^{1,2}(\Omega)$  中有界, 以及由于  $W^{1,2}(\Omega) \hookrightarrow L^2(\Omega)$  的嵌入是紧凑的, 则  $\exists$  子数列  $\{\bar{u}'_n\}_{n' \in \mathbb{N}} \subset \{\bar{u}_n\}_{n \in \mathbb{N}}$ , 在  $L^2(\Omega)$  中收敛。

若是定义

$$\bar{u} := \lim_{n' \rightarrow \infty} \bar{u}_{n'} \in L^2(\Omega),$$

则我们有

$$\begin{aligned}\bar{u} &\in W^{1,2}(\Omega), \\ \|\nabla \bar{u}\|_{L^2(\Omega)} &= 0,\end{aligned}$$

换句话说,  $\bar{u}$  是个常数。此外

$$0 \leq |f(\bar{u})| = |f\left(\lim_{n' \rightarrow \infty} \bar{u}_{n'}\right)| = \lim_{n' \rightarrow \infty} |f(\bar{u}_{n'})| = 0.$$

已知  $\bar{u}$  是个常数,  $f(\bar{u}) = 0$ , 则我们有  $\bar{u} = 0$ 。<sup>10</sup>

$$\therefore |f(u)| \geq |f(\bar{u})| = 0.$$

对应地, 已知  $W^{1,2}(\Omega), W^{1,2}(\Gamma)$  中的等价范分别为

$$\begin{aligned}\|u\|_{W^{1,2}(\Omega),\Omega} &:= \left\{ \left[ \int_{\Omega} u(x) dx \right]^2 + \|\nabla u\|_{L^2(\Omega)}^2 \right\}^{\frac{1}{2}}, \\ \|u\|_{W^{1,2}(\Omega),\Gamma} &:= \left\{ \left[ \int_{\Gamma} u(x) ds_x \right]^2 + \|\nabla u\|_{L^2(\Omega)}^2 \right\}^{\frac{1}{2}},\end{aligned}\tag{14.38}$$

那么  $\hat{W}^{1,2}(\Omega)$  中的等价范为  $\|\nabla u\|_{L^2(\Omega)}$ 。

□

<sup>10</sup>然而需要注意的是, 这里有

$$\|\bar{u}\|_{W^{1,2}(\Omega)} = \lim_{n' \rightarrow \infty} \|\bar{u}_{n'}\|_{W^{1,2}(\Omega)} = 1.$$

下面需要介绍庞加莱不等式。

**Theorem 14.5** (庞加莱不等式 ( $u \in W_0^{k,p}(\Omega)$ )). 设一个有界域  $\Omega \subset \mathbb{R}^n$ 。则存在  $C_p = C(p, n, M) > 0$ , 使得  $\forall u \in W_0^{k,p}(\Omega)$  都满足庞加莱不等式 (Poincaré inequality)

$$\|u\|_{L^p(\Omega)} \leq C_p \|\nabla u\|_{L^p(\Omega)}. \quad (14.39)$$

证明. 1. 证明对于  $\forall u \in C_0^\infty(\Omega)$  (14.5) 成立。

2. 若  $u \in W^{1,p}(\Omega)$ , 则从  $u$  中选取某一数列  $\{u_k\} \subset C_0^\infty(\Omega)$ , 使得该数列随着  $k \rightarrow \infty$  而在  $W^{1,p}$  范数上收敛至  $u$ , 即

$$\begin{aligned} \lim_{k \rightarrow \infty} \|u_k - u\|_{L^p(\Omega)} &= 0, \\ \lim_{k \rightarrow \infty} \|\nabla u_k - \nabla u\|_{L^p(\Omega)} &= 0. \end{aligned}$$

上式等价于

$$\lim_{k \rightarrow \infty} \|u_k\|_{L^p(\Omega)} = \|u\|_{L^p(\Omega)}, \quad (14.40a)$$

$$\lim_{k \rightarrow \infty} \|\nabla u_k\|_{L^p(\Omega)} = \|\nabla u\|_{L^p(\Omega)} \quad (14.40b)$$

3. 既然对于数列  $\{u_k\}$  中的每一个  $u_k$  都满足(14.39), 则我们有以下一组不等式成立

$$\|u_k\|_{L^p(\Omega)} \leq C_p \|\nabla u_k\|_{L^p(\Omega)}, \quad \forall u_k \in C_0^\infty,$$

从而当  $k \rightarrow \infty$  时, (14.39) 对  $\forall u \in C_0^\infty(\Omega)$  都成立。

基于上述思路, 来看  $C_p$  的值。由于在  $\Gamma = \partial\Omega$  上有  $u(\mathbf{x}) = 0$ , 以  $u = 0$  为边界, 由散度定理 (divergence theorem, Theorem 14.1) 有

$$\begin{aligned} \int_{\Omega \cap \{u > 0\}} \operatorname{div}(u^p \mathbf{x}) d\mathbf{x} &= 0, \\ \int_{\Omega \cap \{u < 0\}} \operatorname{div}(u^p \mathbf{x}) d\mathbf{x} &= 0, \end{aligned} \quad (14.41)$$

其中

$$\operatorname{div}(u^p \mathbf{x}) = p u \nabla u \cdot \mathbf{x} + n u^p.$$

(14.41)  $\Rightarrow$

$$\begin{aligned} \int_{\Omega \cap \{u > 0\}} u^p d\mathbf{x} &= -\frac{p}{n} \int_{\Omega \cap \{u > 0\}} u^{p-1} \nabla u \cdot \mathbf{x} d\mathbf{x}, \\ \int_{\Omega \cap \{u > 0\}} u^p d\mathbf{x} &= -\frac{p}{n} \int_{\Omega \cap \{u > 0\}} u^{p-1} \nabla u \cdot \mathbf{x} d\mathbf{x}. \end{aligned} \quad (14.42)$$

$\Omega$  有界  $\Rightarrow M := \max_{\mathbf{x} \in \Omega} |\mathbf{x}| < \infty$ , 则由施瓦茨不等式 (Schwarz inequality) 我们有

$$\begin{aligned} \int_{\Omega \cap \{u > 0\}} u^p d\mathbf{x} &= \left| \frac{p}{n} \int_{\Omega \cap \{u > 0\}} u^{p-1} \nabla u \cdot \mathbf{x} d\mathbf{x} \right| \\ &\leq \frac{pM}{n} \left( \int_{\Omega \cap \{u > 0\}} |u^{p-1}|^q d\mathbf{x} \right)^{\frac{1}{q}} \left( \int_{\Omega \cap \{u > 0\}} |\nabla u|^p d\mathbf{x} \right)^{\frac{1}{p}} \\ &\leq \frac{pM}{n} \left( \int_{\Omega \cap \{u > 0\}} |u^p|^q d\mathbf{x} \right)^{\frac{1}{q}} \|\nabla u\|_{L^p(\Omega \cap \{u > 0\})}, \end{aligned}$$

类似地，我们也有

$$\int_{\Omega \cap \{u < 0\}} u^p d\mathbf{x} \leq \frac{pM}{n} \left( \int_{\Omega \cap \{u < 0\}} |u^p|^q d\mathbf{x} \right)^{\frac{1}{q}} \|\nabla u\|_{L^p(\Omega \cap \{u < 0\})}.$$

上两式代回(14.39)，可得常数  $C_p$  的值

$$C_p = \frac{pM}{n}$$

□

上面的分析主要针对  $u \in W_0^{k,p}(\Omega)$  的情况。我们可以进一步将庞加莱不等式扩展到  $u \in W^{k,p}(\Omega)$  的分析中。

**Corollary 14.3** (庞加莱不等式 ( $u \in W^{k,p}(\Omega)$ )).  $k = 1$  时，由

$$\|u\|_{W^{1,p}(\Omega)} = \|\nabla u\|_{L^p(\Omega)},$$

可得

$$\|u\|_{W^{1,p}(\Omega)} = \left( \|u\|_{L^p(\Omega)}^p + \|\nabla u\|_{L^p(\Omega)}^p \right)^{\frac{1}{p}},$$

因此庞加莱不等式为

$$\|\nabla u\|_{L^p(\Omega)}^p \leq \|u\|_{W^{1,p}(\Omega)}^p \leq (C_p^p + 1)^{\frac{1}{p}} \|\nabla u\|_{L^p(\Omega)}, \quad \forall u \in W^{1,p}(\Omega). \quad (14.43)$$

### 索伯列夫空间的 Bramble-Hilbert 引理

第三个特征是索伯列夫空间的 Bramble-Hilbert 引理，它有助于我们分析（分段）多项式试探空间 (trial space) 的近似属性 (Jovanović and Süli, 2014, Sec 2.3.1)。

则根据庞加莱不等式(14.5)，等价范(14.38)可以改写为

$$\int_{\Omega} |u(x)|^2 dx \leq c_P \left\{ \left[ \int_{\Omega} u(x) dx \right]^2 + \int_{\Omega} |\nabla u(x)|^2 dx \right\}, \quad \forall u \in W^{1,2}(\Omega). \quad (14.44)$$

则我们有 Bramble-Hilbert 引理 (Bramble-Hilbert Lemma)

**Lemma 14.1** (Bramble-Hilbert 引理). 设  $k \in \mathbb{N}_0$ ，一个有界线性泛函  $f : W^{k+1,2}(\Omega) \mapsto \mathbb{R}$  满足

$$|f(\nu)| \leq c_f \|\nu\|_{W^{k+1,2}(\Omega)}, \quad \forall \nu \in W^{k+1,2}(\Omega).$$

用  $\mathcal{P}(\Omega)$  表示所有在  $\Omega$  中定义的  $k$  次多项式。如果以下条件得到满足

$$f(q) = 0, \quad \forall q \in \mathcal{P}(\Omega),$$

那么我们可得

$$|f(\nu)| \leq C(C_p) C_f |\nu|_{W^{k+1,2}(\Omega)}, \quad (14.45)$$

其中  $C(C_p)$  是一个与庞加莱不等式(14.44)中系数  $C_p$  有关的常数。

证明. 以  $k = 1$  的情况为例做出证明 ( $k > 1$  的情况，证明过程与之相似)，对应  $\mathcal{P}_1(\Omega)$  是在  $\Omega$  中定义的线性方程空间。此时我们有

$$|f(\nu)| = |f(\nu) + f(q)| = |f(\nu + q)| \leq C_f \|\nu + q\|_{W^{2,2}(\Omega)},$$

其中最后一个不等式来自于庞加莱不等式(14.44)。

由索伯列夫空间的定义有

$$\begin{aligned}\|\nu + q\|_{W^{2,2}(\Omega)}^2 &= \|\nu + q\|_{L^2(\Omega)}^2 + |\nu + q|_{W^{1,2}(\Omega)}^2 + |\nu + q|_{W^{2,2}(\Omega)}^2 \\ &= \|\nu + q\|_{L^2(\Omega)}^2 + \|\nabla(\nu + q)\|_{L^2(\Omega)}^2 + |\nu|_{W^{2,2}(\Omega)}^2,\end{aligned}$$

其中最后一个等式，根据线性方程  $q(x) \in \mathcal{P}_1(\Omega)$  的二阶导数为 0。再次使用庞加莱不等式(14.44)，上式变为

$$\begin{aligned}\|\nu + q\|_{W^{2,2}(\Omega)}^2 &= \|\nu + q\|_{L^2(\Omega)}^2 + \|\nabla(\nu + q)\|_{L^2(\Omega)}^2 + |\nu|_{W^{2,2}(\Omega)}^2 \\ &\leq C_p \left\{ \left[ \int_{\Omega} (\nu + q) \right]^2 + \|\nabla(\nu + q)\|_{L^2(\Omega)}^2 \right\} + \|\nabla(\nu + q)\|_{L^2(\Omega)}^2 + |\nu|_{W^{2,2}(\Omega)}^2 \\ &= C_p \left[ \int_{\Omega} (\nu + q) \right]^2 + (1 + C_p) \underbrace{\|\nabla(\nu + q)\|_{L^2(\Omega)}^2}_{+ |\nu|_{W^{2,2}(\Omega)}^2},\end{aligned}\tag{14.46}$$

对划线标记部分再次使用庞加莱不等式，有

$$\begin{aligned}\|\nabla(\nu + q)\|_{L^2(\Omega)}^2 &= \sum_{i=1}^d \int_{\Omega} \left| \frac{\partial}{\partial x_i} [\nu(x) + q(x)] \right|^2 dx \\ &\leq C_p \sum_{i=1}^d \left\{ \left[ \int_{\Omega} \frac{\partial}{\partial x_i} [\nu(x) + q(x)] dx \right]^2 + \sum_{j=1}^d \int_{\Omega} \left[ \frac{\partial^2}{\partial x_i \partial x_j} [\nu(x) + q(x)] \right]^2 dx \right\} \\ &= C_p \sum_{i=1}^d \left[ \int_{\Omega} \frac{\partial}{\partial x_i} [\nu(x) + q(x)] dx \right]^2 + C_p \sum_{i=1}^d \sum_{j=1}^d \int_{\Omega} \left[ \frac{\partial^2}{\partial x_i \partial x_j} [\nu(x) + q(x)] \right]^2 dx \\ &= C_p \sum_{i=1}^d \left[ \int_{\Omega} \frac{\partial}{\partial x_i} [\nu(x) + q(x)] dx \right]^2 + C_p |\nu|_{W^{2,2}(\Omega)}^2,\end{aligned}$$

代回(14.46)变为

$$\begin{aligned}\|\nu + q\|_{W^{2,2}(\Omega)}^2 &\leq C_p \left[ \underbrace{\int_{\Omega} (\nu(x) + q(x)) dx}_{+ [1 + (1 + C_p) C_p] \sum_{i=1}^d \underbrace{\int_{\Omega} \frac{\partial}{\partial x_i} [\nu(x) + q(x)] dx}_{+ [1 + (1 + C_p) C_p] |\nu|_{W^{2,2}(\Omega)}^2}} \right]^2 \\ &+ [1 + (1 + C_p) C_p] |\nu|_{W^{2,2}(\Omega)}^2.\end{aligned}\tag{14.47}$$

给定  $u(x) = \nu(x) + q(x)$ ，Bramble-Hilbert 引理(14.44)若要成立，需要在(14.47)中，通过选择  $q(x) \in \mathcal{P}_1(\Omega)$  的值，使得满足下两个条件

$$\int_{\Omega} [\nu(x) + q(x)] dx = 0,\tag{14.48a}$$

$$\int_{\Omega} \frac{\partial}{\partial x_i} [\nu(x) + q(x)] dx = 0.\tag{14.48b}$$

这涉及到系数  $a_0$  和  $a_i, i = 1, \dots, d$  的值

$$q(x) = a_0 + \sum_{i=1}^d a_i x_i.$$

(14.48b)  $\Rightarrow$

$$\begin{aligned} & \int_{\Omega} \frac{\partial}{\partial x_i} \nu(x) dx + \int_{\Omega} \frac{\partial}{\partial x_i} q(x) dx = 0, \\ & \hookrightarrow \int_{\Omega} \frac{\partial}{\partial x_i} \nu(x) dx + |\Omega| a_i = 0, \\ & \hookrightarrow a_i = -\frac{1}{|\Omega|} \int_{\Omega} \frac{\partial}{\partial x_i} \nu(x) dx, \quad i = 1, \dots, d. \end{aligned}$$

(14.48b)  $\Rightarrow$

$$\begin{aligned} & \int_{\Omega} \left[ \nu(x) + a_0 + \sum_{i=1}^d a_i x_i \right] dx = 0, \\ & \hookrightarrow + |\Omega| a_0 + \int_{\Omega} \left[ \nu(x) + \sum_{i=1}^d a_i x_i \right] dx = 0, \\ & \hookrightarrow a_0 = -\frac{1}{|\Omega|} \int_{\Omega} \left[ \nu(x) + \sum_{i=1}^d a_i x_i \right] dx. \end{aligned}$$

对于  $k \in \mathbb{N}$  的情况，证明过程同上。  $\square$

### 索伯列夫空间的分布

由第14.2.4节的介绍可见，并不是  $L^{loc}(\Omega)$  中的所有方程都有广义偏导数。然而可以从分布的角度来重新解读广义偏导数 (McLean, 2000)。

**Definition 14.30** (分布). 对于  $\Omega \subseteq \mathbb{R}^d$ , 定义一个试探空间 (*test space*)  $\mathcal{D}(\Omega) := C_0^\infty(\Omega)$ , 空间中的一个复值连续线性泛函  $T \in \mathcal{D}(\Omega)$  可以称为一个分布 (*distribution*)。 $T$  在  $\mathcal{D}(\Omega)$  中连续, 是指对于  $x \in \Omega$  中的任何  $\varphi_k(x) \rightarrow \varphi(x)$ , 总有  $\mathcal{D}'(\Omega)$  中的  $T(\varphi_k) \rightarrow T(\varphi)$ ;  $\mathcal{D}'(\Omega) \subset \mathcal{D}(\Omega)$  表示  $\mathcal{D}(\Omega)$  空间中, 所有分布的集合。

对于  $x \in \Omega, u(x) \in L^{1,loc}(\Omega)$ , 分布常常定义为

$$T_u(\varphi) := \int_{\Omega} u(x) \varphi(x) dx, \quad \varphi \in \mathcal{D}(\Omega). \quad (14.49)$$

符合(14.49)形式的分布, 又称正则分布 (*regular distribution*)。

正则分布中, 局部可积方程  $u(x) \in L^{1,loc}(\Omega)$  由子集空间  $\mathcal{D}'(\Omega)$  予以识别。因此,  $T_u(\varphi) \in \mathcal{D}'(\Omega)$  有时也简写为  $u \in \mathcal{D}'(\Omega)$ 。

不符合正则类型的分布, 称为奇异分布 (*singular distribution*)。一种奇异分布的例子是狄拉克分布 (*Dirac distribution*):

$$\delta_{x_0}(\varphi) = \varphi(x_0), \quad x_0 \in \Omega, \varphi \in \mathcal{D}(\Omega).$$

现在我们回到式(14.21), 来看如何计算  $u(x)$  的广义偏导数  $\nu(x) := \frac{\partial u(x)}{\partial x} = \operatorname{sgn}(x)$ 。根据分部积分法有

$$\int_{-1}^1 \nu(x) \frac{\partial}{\partial x} \varphi(x) dx = -2\varphi(0) = - \int_{-1}^1 \frac{\partial}{\partial x} \nu(x) \varphi(x) dx, \quad \forall \varphi(x) \in \mathcal{D}(\Omega),$$

可见分布意义上的  $\nu(x)$  的广义偏导数变为

$$\frac{\partial}{\partial x} \nu(x) = 2\varphi(0) = \delta_0(\varphi) \in \mathcal{D}'(\Omega),$$

更高阶的广义偏导数为

$$(D^\alpha T_u)(\varphi) = (-1)^{|\alpha|} T_u(D^\alpha \varphi), \quad \varphi \in \mathcal{D}(\Omega).$$

在掌握了分布的基本概念之后，下面介绍在满足一些关于  $\Omega$  的正则条件假设的情况下， $H^s(\Omega)$  空间近似等价于  $W^{s,2}(\Omega)$  空间，其中  $H^s(\Omega)$  的基是分布的傅里叶变换<sup>11</sup>。我们从急减方程 (rapidly decreasing functions) 空间  $\mathcal{S}(\mathbb{R}^d)$  开始，进而介绍傅里叶变换。

**Definition 14.31** (急减方程空间). 我们将满足以下形式的方程  $\varphi(x) \in C^\infty(\mathbb{R}^d)$  称为急减方程 (rapidly decreasing function):

$$\|\varphi\|_{k,\ell} := \sup_{x \in \mathbb{R}^d} \left( |x|^k \right) \sum_{|\alpha| \leq \ell} |D^\alpha \varphi(x)| < \infty, \quad k, \ell \in \mathbb{N}_0,$$

即是说，急减方程  $\varphi(x)$  和它的导数，比任意多项式的减速更快。

急减方程组成的空间，称为急减空间 (rapidly decreasing space)，表示为  $\mathcal{S}(\mathbb{R}^d)$ 。

类似于  $\mathcal{D}'(\Omega) \subset \mathcal{D}(\Omega)$ ，我们定义缓增分布空间 (tempered distribution space)  $\mathcal{S}'(\mathbb{R}^d) \subset \mathcal{S}(\mathbb{R}^d)$ ，作为全部复值线性泛函  $T \in \mathcal{S}(\mathbb{R}^d)$  组成的子集。

举例来说，对于方程  $\varphi(x) := \exp(-|x|^2)$ ，我们有

$$\varphi \in \mathcal{S}(\mathbb{R}^d), \quad \varphi \notin \mathcal{D}(\Omega) = C_0^\infty(\mathbb{R}^d).$$

**Definition 14.32** (傅里叶变换). 对于方程  $\varphi(x) \in \mathcal{S}(\mathbb{R}^d)$ ，我们将傅里叶变换 (Fourier transform)  $\widehat{\varphi}(x) \in \mathcal{S}(\mathbb{R}^d)$  定义如下

$$\widehat{\varphi}(\xi) := (\mathcal{F}\varphi)(\xi) = (2\pi)^{-\frac{d}{2}} \int_{\mathbb{R}^d} \exp[-i\langle x, \xi \rangle] \varphi(x) dx, \quad \xi \in \mathbb{R}^d, \quad (14.50)$$

其中  $\mathcal{F}$  是可逆映射  $\mathcal{F}: \mathcal{S}(\mathbb{R}^d) \mapsto \mathcal{S}(\mathbb{R}^d)$ 。

对应地，逆傅里叶变换 (inverse Fourier transform) 为

$$(\mathcal{F}^{-1}\widehat{\varphi})(x) = (2\pi)^{-\frac{d}{2}} \int_{\mathbb{R}^d} \exp[i\langle x, \xi \rangle] \widehat{\varphi}(\xi) dx, \quad x \in \mathbb{R}^d. \quad (14.51)$$

通常来说， $\varphi \in \mathcal{D}(\mathbb{R}^d)$  并不必然意味着  $\widehat{\varphi} \in \mathcal{D}(\mathbb{R}^d)$ 。

此外对于  $\varphi \in \mathcal{S}(\mathbb{R}^d)$ ，我们进一步有

$$D^\alpha (\mathcal{F}\varphi)(\xi) = (-i)^{|\alpha|} \mathcal{F}(x^\alpha \varphi)(\xi), \quad (14.52a)$$

$$\xi^\alpha (\mathcal{F}\varphi)(\xi) = (-i)^{|\alpha|} \mathcal{F}(D^\alpha \varphi)(\xi). \quad (14.52b)$$

**Lemma 14.2** (傅里叶变换的旋转对称). 傅里叶变换保持旋转对称结构 (rotational symmetries)，即对于  $u \in \mathbb{R}^d$ ，我们有

$$\widehat{u}(\xi) = \widehat{u}(|\xi|), \forall \xi \in \mathbb{R}^d, \quad \text{iff. } u(x) = u(|x|), \forall x \in \mathbb{R}^d.$$

证明. 先来看  $d = 2$  的证明，对应极坐标系 (polar coordinates)

$$\xi = \begin{pmatrix} |\xi| \cos \psi \\ |\xi| \sin \psi \end{pmatrix}, \quad x = \begin{pmatrix} r \cos \phi \\ r \sin \phi \end{pmatrix}.$$

<sup>11</sup>傅里叶分析的初步介绍，见第12.G节。

根据(14.50)我们有

$$\begin{aligned}\hat{u}(\xi) &= \hat{u}(|\xi|, \psi) \\ &= \frac{1}{2\pi} \int_0^\infty \int_0^{2\pi} \exp\{-ir|\xi|[\cos\phi \cos\psi + \sin\phi \sin\psi]\} u(r) r d\phi dr \\ &= \frac{1}{2\pi} \int_0^\infty \underbrace{\int_0^{2\pi} \exp\{-ir|\xi|[\cos(\phi - \psi)]\} u(r) r d\phi dr}_{:=\mathcal{A}}.\end{aligned}\quad (14.53)$$

现在将坐标旋转  $\psi_0 \in [0, 2\pi]$  度, 变为  $\psi + \psi_0$ , 定义  $\tilde{\phi} := \phi - \psi_0$ , (14.53)变为

$$\begin{aligned}\hat{u}(|\xi|, \psi + \psi_0) &= \frac{1}{2\pi} \int_0^\infty \int_0^{2\pi} \exp\{-ir|\xi|[\cos(\phi - \psi_0 - \psi)]\} u(r) r d\phi dr \\ &= \frac{1}{2\pi} \int_0^\infty \underbrace{\int_{-\psi_0}^{2\pi-\psi_0} \exp\{-ir|\xi|[\cos(\tilde{\phi} - \psi)]\} u(r) r d\tilde{\phi} dr}_{:=\mathcal{B}}.\end{aligned}\quad (14.54)$$

比较上两式

$$\begin{aligned}&\int_{-\psi_0}^0 \exp\{-ir|\xi|[\cos(\tilde{\phi} - \psi)]\} d\tilde{\phi} = \int_{2\pi-\psi_0}^{2\pi} \exp\{-ir|\xi|[\cos(\tilde{\phi} - \psi)]\} d\tilde{\phi}, \\ &\hookrightarrow \mathcal{A} = \mathcal{B}, \\ &\hookrightarrow \hat{u}(|\xi|, \psi) = \hat{u}(|\xi|, \psi + \psi_0), \quad \forall \psi_0 \in [0, 2\pi], \\ &\hookrightarrow \hat{u}(\xi) = \hat{u}(|\xi|).\end{aligned}\quad (14.55)$$

再来看  $d = 3$  的证明, 对应三维坐标系 (spherical coordinates)

$$\xi = \begin{pmatrix} |\xi| \cos\psi \sin\vartheta \\ |\xi| \sin\psi \sin\vartheta \\ |\xi| \cos\vartheta \end{pmatrix}, \quad x = \begin{pmatrix} r \cos\phi \sin\theta \\ r \sin\phi \sin\theta \\ r \cos\theta \end{pmatrix}.$$

根据(14.50)我们有

$$\begin{aligned}\hat{u}(|\xi|, \psi, \vartheta) &= \frac{1}{(2\pi)^{\frac{3}{2}}} \int_0^\infty \int_0^{2\pi} \int_0^\pi \exp\{-ir|\xi|[\cos(\phi - \psi) \sin\theta \sin\vartheta + \cos\theta \cos\vartheta]\} \\ &\quad u(r) r^2 \sin\theta d\theta d\phi dr.\end{aligned}$$

由  $d = 2$  时证得的(14.55), 我们有

$$\hat{u}(|\xi|, \psi + \psi_0, \vartheta) = \hat{u}(|\xi|, \psi, \vartheta), \quad \forall \psi_0 \in [0, 2\pi].$$

给定半径  $\rho$  和一个  $\vartheta \in [0, \pi]$ , 我们有  $\hat{u}(\xi) = \hat{u}(|\xi|) = \hat{u}(\rho)$

$$\begin{aligned}\xi_1^2 + \xi_2^2 &= \rho^2 \sin^2 \vartheta, \\ \xi_3 &= \rho \cos \vartheta.\end{aligned}$$

对原三维坐标作变形, 我们有  $\hat{u}(\xi) = \hat{u}(\rho)$

$$\begin{aligned}\xi_1^2 + \xi_3^2 &= \rho^2 \sin^2 \vartheta, \\ \xi_2 &= \rho \cos \vartheta.\end{aligned}$$

□

**Definition 14.33** (分布的傅里叶变换). 对于某一个缓增分布空间中的分布  $T \in \mathcal{S}'(\mathbb{R}^d)$ , 其傅里叶变换  $\widehat{T} \in \mathcal{S}'(\mathbb{R}^d)$  可定义如下

$$\widehat{T}(\varphi) := T(\widehat{\varphi}), \quad \varphi \in \mathcal{S}(\mathbb{R}^d).$$

从  $\mathcal{F} : \mathcal{S}'(\mathbb{R}^d) \mapsto \mathcal{S}(\mathbb{R}^d)$  是一个可逆映射, 逆傅里叶变换定义为

$$(\mathcal{F}^{-1}T)(\varphi) := T(\mathcal{F}^{-1}\varphi), \quad \varphi \in \mathcal{S}(\mathbb{R}^d).$$

分布的傅里叶变换, 也满足(14.52a)-(14.52b)。

**Definition 14.34** (贝塞尔位势空间). 对于  $s \in \mathbb{R}, u \in \mathcal{S}(\mathbb{R}^d)$ , 定义一个有界的线性贝塞尔位势算子 (*Bessel potential operator*)  $\mathcal{J} : \mathcal{S}(\mathbb{R}^d) \mapsto \mathcal{S}(\mathbb{R}^d)$

$$\mathcal{J}^s u(x) := (2\pi)^{-\frac{d}{2}} \int_{\mathbb{R}^d} \left(1 + |\xi|^2\right)^{\frac{s}{2}} \widehat{u}(\xi) \exp[i \langle x, \xi \rangle] d\xi, \quad x \in \mathbb{R}^d. \quad (14.56)$$

由傅里叶变换可得

$$(\mathcal{F}\mathcal{J}^s u)(\xi) = \underbrace{\left(1 + |\xi|^2\right)^{\frac{s}{2}}}_{(\mathcal{F}u)(\xi)},$$

可见在傅里叶空间中,  $\mathcal{J}^s$  执行的操作近似于傅里叶乘子 (Fourier multiplier), 即对原式乘以一个与  $|\xi|^s$  有关的方程  $\mathcal{O}(|\xi|^s)$ 。从这个意义上讲, 类似于(14.52b), 我们可以将贝塞尔算子  $\mathcal{J}^s$  近似看做  $s$  阶微分符号。

**Definition 14.35** (分布意义上的索伯列夫空间). 在缓增分布空间 (*tempered distribution space*)  $\mathcal{S}'(\mathbb{R}^d)$  中, 给定分布  $T \in \mathcal{S}'(\mathbb{R}^d)$ 。定义有界线性算子  $\mathcal{J}^s : \mathcal{S}'(\mathbb{R}^d) \mapsto \mathcal{S}'(\mathbb{R}^d)$

$$(\mathcal{J}^s T)(\varphi) := T(\mathcal{J}^s \varphi), \quad \varphi \in \mathcal{S}(\mathbb{R}^d).$$

进而, 分布意义上的索伯列夫空间  $H^s(\mathbb{R}^d)$  可以表示为, 由全部分布  $\nu \in \mathcal{S}'(\mathbb{R}^d)$  所组成的内积空间, 其中  $\nu$  满足  $\mathcal{J}^s \nu \in L^2(\mathbb{R}^d)$ , 内积形式表示为

$$\langle u, \nu \rangle_{H^s(\mathbb{R}^d)} := \langle \mathcal{J}^s u, \mathcal{J}^s \nu \rangle_{L^2(\mathbb{R}^d)},$$

对应范数

$$\|u\|_{H^s(\mathbb{R}^d)}^2 := \|\mathcal{J}^s u\|_{H^s(\mathbb{R}^d)}^2 = \int_{\mathbb{R}^d} (1 + |\xi|^s)^s |\widehat{u}(\xi)|^2 d\xi.$$

**Theorem 14.6** (索伯列夫空间). 对于所有  $s \in \mathbb{R}$ , 下式均成立

$$H^s(\mathbb{R}^d) = W^{s,2}(\mathbb{R}^d).$$

对于有界域  $\Omega \subset \mathbb{R}^d$ , 我们将索伯列夫空间  $H^s(\mathbb{R}^d)$  定义如下形式

$$\begin{aligned} H^s(\Omega) &:= \{\nu = \widetilde{\nu}|_\Omega : \widetilde{\nu} \in H^s(\mathbb{R}^d)\}, \\ \|\nu\|_{H^s(\Omega)} &:= \inf_{\widetilde{\nu} \in H^s(\mathbb{R}^d), \widetilde{\nu}|_\Omega = \nu} \|\widetilde{\nu}\|_{H^s(\mathbb{R}^d)}. \end{aligned}$$

此外还可以对索伯列夫空间作如下定义

$$\begin{aligned} \widetilde{H}^s(\Omega) &:= \overline{C_0^\infty(\Omega)}^{\|\cdot\|_{H^s(\mathbb{R}^d)}}, \\ H_0^s(\Omega) &:= \overline{C_0^\infty(\Omega)}^{\|\cdot\|_{H^s(\Omega)}}, \end{aligned}$$

根据 (McLean, 2000, Theorem 3.33), 上述定义对于几乎全部  $s \in \mathbb{R}_+$  都成立。

**Theorem 14.7.** 假定有一个利普希茨域 (*Definition 14.2*)  $\Omega \in \mathbb{R}^d$ 。对于  $s \geq 0$  我们有

$$\tilde{H}^s(\Omega) \in H_0^s(\Omega).$$

进一步有

$$\tilde{H}^s(\Omega) = H_0^s(\Omega), \quad s \notin \left\{ \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \dots \right\}.$$

进一步有

$$\tilde{H}^s(\Omega) = [H^{-s}(\Omega)]', H^s(\Omega) = [\tilde{H}^{-s}(\Omega)]', \quad \forall s \in \mathbb{R}.$$

**Definition 14.36** (均匀法锥条件). 两个索伯列夫空间  $H^s(\Omega)$  和  $W^{s,2}(\Omega)$  相等, 即范式等价的充分条件是存在一个有界的线性延拓算子

$$E_\Omega : W^{2,s}(\Omega) \mapsto W^{2,s}(\mathbb{R}^d),$$

这要求存在一个有边界的域  $\Omega \in \mathbb{R}^d$ , 这称为均匀法锥条件 (*uniform cone condition*)。

证明. 略。见 ([Adolfson et al., 2008b](#), Theorem 4.6, 4.7)。  $\square$

**Theorem 14.8.** 对于有界的利普希茨域  $\Omega \subset \mathbb{R}^d$ , 我们有

$$\begin{aligned} \|q\|_{L^2(\Omega)} &\leq C_1 \left\{ \|q\|_{H^{-1}(\Omega)} + \|\nabla q\|_{[H^{-1}(\Omega)]^d} \right\}, \quad \forall q \in L^2(\Omega), \\ \|q\|_{L^2(\Omega)} &\leq C_2 \|\nabla q\|_{[H^{-1}(\Omega)]^d}, \quad \forall q \in L^2(\Omega), \int_{\Omega} q(x) dx = 0. \end{aligned}$$

我们可以用有界的线性算子表现从一个索伯列夫空间到另一个索伯列夫空间的映射, 这涉及到插值空间。关于插值空间的介绍可见如[Adolfson et al. \(2008b\)](#)。这里介绍与有限元法关系较为密切的插值定理 (interpolation theorem)。

**Theorem 14.9** (插值定理). 设一个有界的线性算子  $A : H^{\alpha_1}(\Omega) \rightarrow H^{\beta}(\Omega)$ , 其范数为

$$\|A\|_{\alpha_1, \beta} := \sup_{0 \neq \nu \in H^{\alpha_1}(\Omega)} \frac{\|A\nu\|_{H^\beta(\Omega)}}{\|\nu\|_{H^{\alpha_1}(\Omega)}}.$$

现在假设  $\alpha_2 > \alpha_1$ , 对应有界线性算子  $A : H^{\alpha_2}(\Omega) \rightarrow H^{\beta}(\Omega)$ , 范数

$$\|A\|_{\alpha_2, \beta} := \sup_{0 \neq \nu \in H^{\alpha_2}(\Omega)} \frac{\|A\nu\|_{H^\beta(\Omega)}}{\|\nu\|_{H^{\alpha_2}(\Omega)}}.$$

则  $A : H^\alpha(\Omega) \mapsto H^\beta(\Omega)$ ,  $\forall \alpha \in [\alpha_1, \alpha_2]$  都有界, 对应范数

$$\|A\|_{\alpha, \beta} \leq \left( \|A\|_{\alpha_1, \beta} \right)^{\frac{\alpha - \alpha_2}{\alpha_1 - \alpha_2}} \left( \|A\|_{\alpha_2, \beta} \right)^{\frac{\alpha_2 - \alpha_1}{\alpha_2 - \alpha_1}}.$$

同样, 设  $A : H^\alpha \mapsto H^{\beta_1}(\Omega)$  有界, 对应范数  $\|A\|_{\alpha, \beta_1}$ 。设  $A : H^\alpha \mapsto H^{\beta_2}(\Omega)$  有界, 对应范数  $\|A\|_{\alpha, \beta_2}$ ,  $\beta_1 < \beta_2$ 。

则  $A : H^\alpha \mapsto H^\beta(\Omega)$  有界  $\forall \beta \in [\beta_1, \beta_2]$ , 对应范数

$$\|A\|_{\alpha, \beta} \leq \left( \|A\|_{\alpha, \beta_1} \right)^{\frac{\beta - \beta_2}{\beta_1 - \beta_2}} \left( \|A\|_{\alpha, \beta_2} \right)^{\frac{\beta - \beta_1}{\beta_2 - \beta_1}}.$$

### 流形索伯列夫空间

**Definition 14.37** (流形). 流形 (*manifold*) 是局部具有欧式空间性质的空间，包括各种纬度的曲线曲面，例如球体、弯曲的平面等。流形的局部和欧式空间是同构的 (局部线性)。流形学习假设所处理的数据点分布在嵌入于外维欧式空间的一个潜在的流形体上，或者说这些数据点可以构成这样一个潜在的流形体<sup>12</sup>。

设一个有边界的域  $\Omega \subset \mathbb{R}^d, d = 2, 3$ ，其边界用  $\Gamma = \partial\Omega$  来表示。对  $\Gamma$  作局部分段参数化 (piecewise parameterization) 处理，对应一组低维空间  $\mathbb{R}^{d-1}$  中的参数域 (parameter domain)  $\mathcal{T}_i$ :

$$\begin{aligned}\Gamma &= \bigcup_{i=1}^J \Gamma_i, \\ \Gamma_i &:= \{x \in \mathbb{R}^d : x = \chi_i(\xi), \xi \in \mathcal{T}_i \subset \mathbb{R}^{d-1}\}.\end{aligned}\tag{14.57}$$

与(14.57)同步，考虑一个单位数列  $\{\varphi_i\}_{i=1}^p$ ，其中  $\varphi_i \in C_0^\infty(\mathbb{R}^d)$  是非负的 cutoff 方程，满足

$$\begin{cases} \sum_{i=1}^J \varphi_i(x) = 1 & \text{对于 } x \in \Gamma \\ \varphi_i(x) = 0 & \text{对于 } x \in \Gamma \setminus \Gamma_i. \end{cases}\tag{14.58}$$

进而，在  $\Gamma$  中定义一个方程  $\nu(x)$ ，满足

$$\begin{aligned}\nu(x) &= \sum_{i=1}^J \varphi_i(x) \nu(x) = \sum_{i=1}^J \nu_i(x), \quad \text{对于 } x \in \Gamma, \\ \nu_i(x) &:= \varphi_i(x) \nu(x).\end{aligned}\tag{14.59}$$

将关于  $x$  的局部分段参数化(14.57)代入  $\nu_i(x)$  的定义式，我们可以定义一组新的方程  $\{\tilde{\nu}_i(\xi)\}$ ，其中

$$\begin{aligned}\nu_i(x) &= \varphi_i(x) \nu(x) \\ &= \varphi_i(\chi_i(\xi)) \nu(\chi_i(\xi)) \\ &=: \tilde{\nu}_i(\xi), \quad \xi \in \mathcal{T}_i \subset \mathbb{R}^{d-1}.\end{aligned}$$

通过这种方法， $\Omega \subset \mathbb{R}^d$  中关于变量  $x$  的问题  $\nu(x)$ ，被表示为分段参数域  $\mathcal{T}_i \subset \mathbb{R}^{d-1}$  中关于局部参数  $\chi_i(\xi)$  的问题  $\tilde{\nu}_i(\xi)$ ，后者使得我们可能建立相应的索伯列夫空间，进行分析。这要求  $\tilde{\nu}_i$  可导，即局部参数  $\chi_i(\xi)$  可导，满足链式法则 (chain rule)<sup>13</sup>。如果要求  $|s| \leq k$  次的导数，需要假定局部参数  $\chi_i \in C^{k-1,1}(\mathcal{T}_i)$ ，例如，对于利普希茨域  $\mathcal{T}_i$  中的局部参数  $\chi_i \in C^{0,1}(\mathcal{T}_i)$ ，我们需要研究对应的索伯列夫空间  $H^s(\mathcal{T}_i), |s| \leq 1$ 。

**Definition 14.38** (流形索伯列夫空间 ( $0 \leq s \leq k$ )). 进而对于  $0 \leq s \leq k$ ，我们的研究对象是整个索伯列夫空间  $H^s(\Gamma)$ ，其范数定义为

$$\|\nu\|_{H_\chi^s(\Gamma)} := \left\{ \sum_{i=1}^J \|\tilde{\nu}_i\|_{H^s(\mathcal{T}_i)}^2 \right\}^{\frac{1}{2}}.\tag{14.61}$$

<sup>12</sup>从词源上来看，最早由黎曼给出了德文名称 *mannigfaltigkeit*，英文翻译为 *manifold*，字面意义为“多层”。中国第一个拓扑学家江泽涵把这个词翻译为“流形”，取自文天祥《正气歌》，“天地有正气，杂然赋流形”，其原始出处为《易经》，“大哉乾元，万物资始，乃统天。云行雨施，品物流形。”

<sup>13</sup>简单说来，链式法则可以表示为

$$f(g(x))' = f'(g(x)) \cdot g'(x).\tag{14.60}$$

**Lemma 14.3** (流形索伯列夫空间的等价范 ( $s = 0$ )).  $s = 0$  时, 流形索伯列夫空间  $H_\chi^0(\Gamma)$  的等价范为

$$\|\nu\|_{L^2(\Gamma)} := \left\{ \int_\Gamma |\nu(x)|^2 ds_x \right\}^{\frac{1}{2}}.$$

证明. 首先来看  $H_\chi^0(\Gamma)$  空间。 $s = 0$  时(14.61) $\Rightarrow$

$$\begin{aligned} \|\nu\|_{H_\chi^0(\Gamma)}^2 &= \sum_{i=1}^J \|\tilde{\nu}_i\|_{H^0(\Gamma_i)}^2 \\ &= \sum_{i=1}^J \int_{\Gamma_i} [\varphi_i(\chi_i(\xi)) \nu(\chi_i(\xi))]^2 d\xi. \end{aligned}$$

再来看  $L^2(\Gamma)$  空间  $\Rightarrow$

$$\begin{aligned} \|\nu\|_{L^2(\Gamma)}^2 &= \int_\Gamma [\nu(x)]^2 ds_x \\ &= \sum_{i=1}^J \int_{\Gamma_i} \varphi_i(x) [\nu(x)]^2 ds_x \\ &= \sum_{i=1}^J \int_{\Gamma_i} \varphi_i(\chi_i(\xi)) [\nu(\chi_i(\xi))]^2 (\det \chi_i(\xi)) d\xi. \end{aligned}$$

□

**Lemma 14.4** (流形索伯列夫空间的等价范 ( $0 < s < 1$ )).  $0 < s < 1$  时, 流形索伯列夫空间  $H_\chi^s(\Gamma)$  的等价范, 可以用 Sobolev-Slobodeckij 范数 (Section 14.2.4) 来表示

$$\|\nu\|_{H^s(\Gamma)} := \left\{ \|\nu\|_{L^2(\Gamma)}^2 + \int_\Gamma \int_\Gamma \frac{[\nu(x) - \nu(y)]^2}{|x-y|^{d-1+2s}} ds_x ds_y \right\}^{\frac{1}{2}},$$

需要指出的是, 上式并不是唯一一种定义等价范的方法。其他方法如, 根据索伯列夫范数等价定理 (Theorem 14.4), 流形索伯列夫空间  $H^{s=1/2}(\Gamma)$  的等价范也可以写作

$$\|\nu\|_{H^{\frac{1}{2}}(\Gamma), \Gamma} := \left\{ \left[ \int_\Gamma \nu(x) ds_x \right]^2 + \int_\Gamma \int_\Gamma \frac{[\nu(x) - \nu(y)]^2}{|x-y|^d} ds_x ds_y \right\}^{\frac{1}{2}}.$$

**Lemma 14.5** (流形索伯列夫空间的等价范 ( $s < 0$ )).  $s < 0$  时的流形索伯列夫空间  $H_\chi^s(\Gamma)$ , 可由其对偶空间 (dual space)  $H_\chi^{-s}(\Gamma)$  进行分析

$$H^s(\Gamma) := [H^{-s}(\Gamma)]',$$

其范数为

$$\begin{aligned} \|w\|_{H^s(\Gamma)} &:= \sup_{0 \neq \nu \in H^{-s}(\Gamma)} \frac{\langle w, \nu \rangle_\Gamma}{\|\nu\|_{H^{-s}(\Gamma)}} \\ &= \sup_{0 \neq \nu \in H^{-s}(\Gamma)} \frac{\int_\Gamma w(x) \nu(x) ds_x}{\|\nu\|_{H^{-s}(\Gamma)}}, \end{aligned} \tag{14.62}$$

其中  $\langle w, \nu \rangle_\Gamma$  表示对偶配对 (duality pairing)。

**Lemma 14.6** (流形索伯列夫空间的开放子集). 对于充分平滑的边界  $\Gamma = \partial\Omega$ , 设一个开放子集  $\Gamma_0 \subset \Gamma$ 。则  $H^s(\Gamma_0)$  可以定义如下

1.  $s \geq 0$  时的索伯列夫空间  $H^s(\Gamma_0)$  及范数

$$H^s(\Gamma_0) := \left\{ \nu = \tilde{\nu}|_{\Gamma_0} : \tilde{\nu} \in H^s(\Gamma) \right\},$$

$$\|\nu\|_{H^s(\Gamma_0)} := \inf_{\tilde{\nu} \in H^s(\Gamma) : \tilde{\nu}|_{\Gamma_0} = \nu} \|\tilde{\nu}\|_{H^s(\Gamma)}$$

2.  $s < 0$  时, 首先定义一个索伯列夫空间  $\tilde{H}^s(\Gamma_0)$

$$\tilde{H}^s(\Gamma_0) := \left\{ \nu = \tilde{\nu}|_{\Gamma_0} : \tilde{\nu} \in H^s(\Gamma), \text{supp } \tilde{\nu} \subset \Gamma_0 \right\},$$

其中  $\text{supp}$  表示支撑集 (*support*), 是一个定义在集合  $\Gamma_0$  上的实值函数  $\tilde{\nu}$  的子集, 满足  $\tilde{\nu}$  恰好在这个子集上非零<sup>14</sup>。进而利用对偶空间的属性

$$H^s(\Gamma_0) := [\tilde{H}^{-s}(\Gamma_0)]',$$

$$\tilde{H}^s(\Gamma_0) := [H^{-s}(\Gamma_0)].$$

**Lemma 14.7** (分段平滑索伯列夫空间). 对于某一封闭边界  $\Gamma = \partial\Omega$ , 假定它是分段平滑的

$$\Gamma = \bigcup_{i=1}^J \bar{\Gamma}_i, \quad \Gamma_i \cap \Gamma_j = \emptyset \forall i \neq j.$$

对应的索伯列夫空间  $H_{pw}^s(\Gamma)$  及分段平滑方程的范数为:

1.  $s > 0 \Rightarrow$

$$H_{pw}^s(\Gamma) := \left\{ \nu \in L^2(\Gamma) : \nu|_{\Gamma_i} \in H^s(\Gamma_i), i = 1, \dots, J \right\},$$

$$\|\nu\|_{H_{pw}^s(\Gamma)} := \left\{ \sum_{i=1}^J \|\nu|_{\Gamma_i}\|_{H^s(\Gamma_i)}^2 \right\}^{\frac{1}{2}}.$$

2.  $s < 0 \Rightarrow$

$$H_{pw}^s(\Gamma) := \Pi_{j=1}^J \tilde{H}_{\Gamma_j}^s, \tag{14.63}$$

$$\|w\|_{H_{pw}^s(\Gamma)} := \sum_{j=1}^J \|w|_{\Gamma_j}\|_{\tilde{H}_{\Gamma_j}^s}. \tag{14.64}$$

**Lemma 14.8.** (14.63) 的分段平滑索伯列夫空间  $H_{pw}^s(\Gamma), s < 0$  中, 方程  $w \in H_{pw}^{s<0}(\Gamma)$  满足

$$\|w\|_{H_{pw}^s(\Gamma)} \leq \|w\|_{H_{pw}^{s<0}(\Gamma)}$$

证明. 由对偶空间的定义(14.62)、扩展三角不等式等有

$$\begin{aligned} \|w\|_{H^s(\Gamma)} &= \sup_{0 \neq \nu \in H^{-s}(\Gamma)} \frac{|\langle w, \nu \rangle_\Gamma|}{\|\nu\|_{H^{-s}(\Gamma)}} \\ &\leq \sup_{0 \neq \nu \in H^{-s}(\Gamma)} \sum_{j=1}^J \frac{|\langle w, \nu \rangle_{\Gamma_j}|}{\|\nu\|_{H^{-s}(\Gamma_j)}} \\ &\leq \sup_{0 \neq \nu \in H^{-s}(\Gamma)} \sum_{j=1}^J \frac{|\langle w|_{\Gamma_j}, \nu|_{\Gamma_j} \rangle_{\Gamma_j}|}{\|\nu|_{\Gamma_j}\|_{H^{-s}(\Gamma_j)}} \\ &\leq \sum_{j=1}^J \sup_{0 \neq \nu \in H^{-s}(\Gamma_j)} \frac{|\langle w|_{\Gamma_j}, \nu|_{\Gamma_j} \rangle_{\Gamma_j}|}{\|\nu|_{\Gamma_j}\|_{H^{-s}(\Gamma_j)}} \\ &= \|w\|_{H_{pw}^s(\Gamma)}. \end{aligned}$$

<sup>14</sup>例如一个拓扑空间 (如实值轴)  $X$ , 连续方程  $f \in X$ 。此时  $\text{supp } f$  定义为这样一个闭集  $C$ , 满足 1)  $f$  在  $X \setminus C$  中为 0; 2) 不存在  $C$  的真闭子集也满足这一条件, 即  $C$  是所有这样的子集中最小的一个。

□

对于一个利普希茨域  $\Omega \subset \mathbb{R}^d$ , 其利普希茨边界可表示为  $\Gamma = \partial\Omega$ 。对于  $u(x) \in \Omega$ , 对应的内界迹 (interior boundary trace) 为  $\gamma_0^{int} u \in \Gamma$ , 见(14.4)。则  $H^s(\Omega)$  和  $H^s(\Gamma)$  两个索伯列夫空间的关系, 常用下述迹定理、逆迹定理予以描述。

**Theorem 14.10** (索伯列夫空间的迹定理). 设  $\Omega \subset \mathbb{R}^d$  是一个  $C^{k-1,1}$  域,  $\frac{1}{2} < s \leq k$ , 则内界迹

$$\gamma_0^{int} : H^s(\Omega) \mapsto H^{s-\frac{1}{2}}(\Gamma)$$

是有界的, 满足

$$\|\gamma_0^{int} \nu\|_{H^{s-\frac{1}{2}}(\Gamma)} \leq C_T \|\nu\|_{H^s(\Omega)}, \quad \forall \nu \in H^s(\Omega).$$

**Lemma 14.9.** 若  $k = 1, s \in (\frac{1}{2}, 1]$ , 则我们可以利用 Theorem 14.10 求得迹算子  $\gamma_0^{int} : H^s(\Omega) \mapsto H^{s-\frac{1}{2}}(\Gamma)$ 。

若  $s \in (\frac{1}{2}, \frac{3}{2})$ , 该方法也适用 (McLean, 2000, Theorem 3.38)。

**Theorem 14.11** (索伯列夫空间的逆迹定理). 设  $\Omega \subset \mathbb{R}^d$  是一个  $C^{k-1,1}$  域,  $\frac{1}{2} < s \leq k$ , 则内界迹

$$\gamma_0^{int} : H^s(\Omega) \mapsto H^{s-\frac{1}{2}}(\Gamma)$$

有一个连续的右逆算子

$$\varepsilon : H^{s-\frac{1}{2}}(\Gamma) \mapsto H^s(\Omega),$$

满足如下关系

$$\begin{aligned} \gamma_0^{int} \varepsilon w &= w, \quad \forall w \in H^{s-\frac{1}{2}}(\Gamma), \\ \|\varepsilon w\|_{H^s(\Omega)} &\leq C_{IT} \|w\|_{H^{s-\frac{1}{2}}(\Gamma)}, \quad \forall w \in H^{s-\frac{1}{2}}(\Gamma). \end{aligned}$$

**Definition 14.39** (索伯列夫迹空间). 从这个意义上来说, 对于某个索伯列夫空间  $H^{s+\frac{1}{2}}(\Omega)$ ,  $s > 0$  来说, 它的迹空间可以表示为  $H^s(\Gamma)$ , 对应的范数

$$\|\nu\|_{H^s(\Gamma), \gamma_0} := \inf_{V \in H^{s+\frac{1}{2}}(\Omega), \gamma_0^{int} V = \nu} \|V\|_{H^{s+\frac{1}{2}}(\Omega)}.$$

需要指出的是, 利普希茨域  $\Omega \subset \mathbb{R}^d$  中, 只有当  $|s| < 1$  时, 以下范式等价  $\|\nu\|_{H^s(\gamma), \gamma_0} = \|\nu\|_{H^s(\Gamma)}$ 。

## 14.3 变分法

弱形式边界值问题常常表现为带有算子方程 (operator equations) 的变分问题。对于变分问题, 我们常将其表示为表面积位势和体积位势 (surface and volume potentials) 的偏微分方程, 为了求解方程, 就需要求得有界的积分算子方程的解, 以建立完备的柯西数列解。本节介绍一些泛函分析的基本知识, 进而探讨算子方程解的存在性和唯一性。

### 14.3.1 算子方程

假设一个希尔伯特空间  $X$ , 空间中的内积形式  $\langle \cdot, \cdot \rangle_X$ , 对应范数  $\|\cdot\|_X = \sqrt{\langle \cdot, \cdot \rangle_X}$ 。 $X$  的对偶空间 (dual space) 表示为  $X'$ ;  $X$  和  $X'$  以  $\langle \cdot, \cdot \rangle$  的形式呈对偶配对 (duality pairing)。则我们有

$$\|f\|_{X'} = \sup_{0 \neq \nu \in X} \frac{|\langle f, \nu \rangle|}{\|\nu\|_X}, \quad \forall f \in X'. \tag{14.65}$$

定义一个有界的自伴随 (self-adjoint) 线性算子  $A : X \mapsto X'$ , 满足

$$\|A\nu\|_{X'} \leq C_2^A \|\nu\|_X, \quad \forall \nu \in X, \quad (14.66)$$

$A$  是自伴随的, 是指

$$\langle Au, \nu \rangle = \langle u, A\nu \rangle, \quad \forall u, \nu \in X. \quad (14.67)$$

那么, 边界值问题可以表示为, 对于某个给定的  $f \in X'$ , 寻找算子方程的解  $u \in X$ , 使满足

$$Au = f. \quad (14.68)$$

算子方程(14.68)的解, 通常难于直接求得。替代方案是建立一个变分问题, 寻找变分问题的解  $u \in X$ , 使满足

$$\langle Au, \nu \rangle = \langle f, \nu \rangle, \quad \forall \nu \in X. \quad (14.69)$$

**Theorem 14.12.** 算子方程(14.68)的解  $u \in X$ , 和变分问题(14.69)的解  $u \in X$ , 二者等价。

证明. 证明过程分为两个部分。

1. 显然, 算子方程(14.68)的解  $u \in X$ , 构成变分问题(14.69)的解。
2. 反过来, 假定已求得变分问题(14.69)的解  $u \in X$ , 由(14.65)可得

$$\|Au - f\|_{X'} = \sup_{0 \neq \nu \in X} \frac{|\langle Au - f, \nu \rangle|}{\|\nu\|_X},$$

代回(14.69)可得

$$\|Au - f\|_{X'} = 0 \Rightarrow Au = f,$$

即是说  $u \in X$  同时也是算子方程(14.68)的解。

□

基于算子  $A$ , 可定义一个双线性映射 (bilinear form)  $a(u, \nu)$

$$\begin{aligned} a(u, \nu) &:= \langle Au, \nu \rangle, \quad \forall u, \nu \in X, \\ A : X \mapsto X' &\Rightarrow a(\cdot, \cdot) : X \times X \mapsto \mathbb{R}, \end{aligned} \quad (14.70)$$

反之亦然, 通过双线性映射  $a(u, \nu)$  可以定义算子  $A : X \mapsto X'$ , 见 Lemma 14.10。

**Lemma 14.10.** 设一个有界的双线性映射  $a(\cdot, \cdot) : X \times X \mapsto \mathbb{R}$ , 满足

$$|a(u, \nu)| \leq C_2^A \|u\|_X \|\nu\|_X, \quad \forall u, \nu \in X.$$

对于其中任一  $u \in X$ , 都存在一个元素  $Au \in X'$ , 使得满足

$$\langle Au, \nu \rangle = a(u, \nu), \quad \forall \nu \in X.$$

则我们有: 算子  $A : X \mapsto X'$  是一个有界的线性算子, 满足

$$\|Au\|_{X'} \leq C_2^A \|u\|_X, \quad \forall u \in X.$$

证明. 对于任一给定的  $u \in X$ , 我们定义一个  $X$  中的有界双线性映射  $\langle f_u, \nu \rangle := a(u, \nu)$ , 即我们有  $f_u \in X'$ 。通过映射  $u \in X \mapsto f_u \in X'$ , 我们定义一个线性算子  $A : X \mapsto X'$ , 使得  $Au = f_u \in X'$ , 并且满足

$$\|Au\|_{X'} = \|f_u\|_{X'} = \sup_{0 \neq \nu \in X} \frac{|\langle f_u, \nu \rangle|}{\|\nu\|_X} = \sup_{0 \neq \nu \in X} \frac{|a(u, \nu)|}{\|\nu\|_X} \leq C_2^A \|u\|_X.$$

□

因此我们可以把求解变分问题(14.69)的工作，转化为求解下述最小化问题的工作。

**Lemma 14.11.** 设线性算子  $A : X \mapsto X'$  是 1) 自伴随的，即  $\langle Au, \nu \rangle = \langle u, A\nu \rangle$ ，和 2) 正半定的，即  $\langle Au, \nu \rangle \geq 0$ ， $\forall \nu \in X$ 。我们设一个泛函  $F$

$$F(\nu) := \frac{1}{2} \langle A\nu, \nu \rangle - \langle f, \nu \rangle, \quad \forall \nu \in X. \quad (14.71)$$

那么变分问题(14.69)的解  $u \in X$ ，等价于如下最小化问题的解

$$F(u) = \min_{\nu \in X} F(\nu). \quad (14.72)$$

证明. 设  $u, \nu \in X$ ，任一  $t \in \mathbb{R}$ 。进而

$$\begin{aligned} F(u + t\nu) &= \frac{1}{2} \langle A(u + t\nu), u + t\nu \rangle - \langle f, u + t\nu \rangle \\ &= \frac{1}{2} [\langle Au, u \rangle + \langle Au, t\nu \rangle + \langle A\nu, u \rangle + \langle At\nu, t\nu \rangle] - [\langle f, u \rangle + \langle f, t\nu \rangle] \\ &= \frac{1}{2} \langle Au, u \rangle + \frac{1}{2} t \langle Au, \nu \rangle + \frac{1}{2} t \langle A\nu, u \rangle + \frac{1}{2} t^2 \langle A\nu, \nu \rangle - \langle f, u \rangle - t \langle f, \nu \rangle \\ &= \left[ \frac{1}{2} \langle Au, u \rangle - \langle f, u \rangle \right] + t [\langle Au, \nu \rangle - \langle f, \nu \rangle] + \frac{1}{2} t^2 \langle A\nu, \nu \rangle \\ &= F(u) + t [\langle Au, \nu \rangle - \langle f, \nu \rangle] + \frac{1}{2} t^2 \langle A\nu, \nu \rangle. \end{aligned}$$

1. 假设  $u \in X$  是变分问题(14.69)的解，那么  $\langle Au, \nu \rangle = \langle f, \nu \rangle$ ，上式变为

$$\begin{aligned} F(u + t\nu) &= F(u) + \underbrace{\frac{1}{2} t^2 \langle A\nu, \nu \rangle}_{\geq 0}, \\ \Rightarrow F(u) &\leq F(u + t\nu), \quad \forall \nu \in X, t \in \mathbb{R}, \end{aligned}$$

由此可见  $u \in X$  同时也是最小化问题(14.72)的解。

2. 现在假设  $u \in X$  是最小化问题(14.72)的解。那么以下条件也成立

$$\frac{d}{dt} F(u + t\nu) \Big|_{t=0} = 0, \quad \forall \nu \in X,$$

由此可得

$$\langle Au, \nu \rangle = \langle f, \nu \rangle, \quad \forall \nu \in X,$$

可见  $u \in X$  同时也是变分问题(14.69)的解。

□

现在来证明变分问题(14.69)、最小化问题(14.72)的解  $u \in X$  存在且唯一，可由里兹表现定理证明。

**Theorem 14.13** (里兹表现定理). 任一线性有界泛函  $f \in X'$  均可表现为下述形式

$$\langle f, \nu \rangle = \langle u, \nu \rangle_X,$$

其中  $u \in X$  由  $f \in X'$  所唯一确定 (*uniquely determined*)，并且满足

$$\|u\|_X = \|f\|_{X'}, \quad (14.73)$$

这称为里兹表现定理 (*Riesz representation theorem*)。

证明. 设某一给定的泛函  $f \in X'$ , 我们可以通过求解变分问题(14.69)找到解  $u \in X$

$$\langle Au, \nu \rangle = \langle f, \nu \rangle, \quad \forall \nu \in X,$$

并且根据 Lemma 14.11, 这也等同于求解最小化问题(14.72)

$$F(u) = \min_{\nu \in X} F(\nu),$$

其中泛函  $F$  由(14.71)给出

$$F(\nu) := \frac{1}{2} \langle A\nu, \nu \rangle - \langle f, \nu \rangle, \quad \forall \nu \in X.$$

1. 证明  $u \in X$  是最小化问题和变分问题的解。 (14.71)  $\Rightarrow$

$$\begin{aligned} F(\nu) &= \frac{1}{2} \langle \nu, \nu \rangle_X - \langle f, \nu \rangle \\ &\geq \frac{1}{2} \|\nu\|_X^2 - \|f\|_{X'} \|\nu\|_X \\ &= \frac{1}{2} \underbrace{[\|\nu\|_X - \|f\|_{X'}]^2}_{\geq 0} - \frac{1}{2} \|f\|_{X'}^2 \\ &\geq -\frac{1}{2} \|f\|_{X'}, \quad \forall \nu \in X, \end{aligned}$$

可见  $F(\nu), \forall \nu \in X$  有下界 (infimum), 定义为  $\alpha$

$$\alpha := \inf_{\nu \in X} F(\nu) \in \mathbb{R}.$$

设存在一个数列  $\{u_k\}_{k \in \mathbb{N}} \subset X$ , 随着  $k \rightarrow \infty$ , 满足  $F(u_k) \rightarrow \alpha$ 。由数列的性质可得

$$\|u_k - u_\ell\|_X^2 + \|u_k + u_\ell\|_X^2 = 2 \left\{ \|u_k\|_X^2 + \|u_\ell\|_X^2 \right\},$$

进而我们有  $\|u_k - u_\ell\|_X^2 \geq 0$ , 以及

$$\begin{aligned} \|u_k - u_\ell\|_X^2 &= 2\|u_k\|_X^2 + 2\|u_\ell\|_X^2 - \|u_k + u_\ell\|_X^2 \\ &= 4 \underbrace{\left\{ \frac{1}{2}\|u_k\|_X^2 - \langle f, u_k \rangle \right\}}_{=F(u_k)} + 4 \underbrace{\left\{ \frac{1}{2}\|u_\ell\|_X^2 - \langle f, u_\ell \rangle \right\}}_{=F(u_\ell)} - 8 \left[ \frac{1}{2} \left\| \frac{1}{2}(u_k + u_\ell) \right\|_X^2 - \langle f, \frac{1}{2}(u_k + u_\ell) \rangle \right] \\ &= 4F(u_k) + 4F(u_\ell) - 8F\left(\frac{1}{2}(u_k + u_\ell)\right) \\ &\leq 4\alpha + 4\alpha - 8\alpha \rightarrow 0, \quad k, \ell \rightarrow \infty. \end{aligned}$$

由此可见,  $\{u_k\}_{k \in \mathbb{N}}$  是一个柯西数列 (Cauchy sequence)。此外由于  $X$  是一个希尔伯特空间, 我们得到极限值

$$u = \lim_{k \rightarrow \infty} u_k \in X.$$

对应得泛函  $F(u)$  的值, 求解过程如下。由

$$\begin{aligned} |F(u_k) - F(u)| &\leq |F(u_k - u)| \\ &\leq \frac{1}{2} |\langle u_k, u_k \rangle_X - \langle u, u \rangle_X| + |\langle f, u_k - u \rangle| \\ &= \frac{1}{2} |\langle u_k, u_k - u \rangle_X + \langle u, u_k - u \rangle_X| + |\langle f, u_k - u \rangle| \\ &\leq \left\{ \frac{1}{2} \|u_k\|_X + \frac{1}{2} \|u\|_X + \|f\|_{X'} \right\} \|u_k - u\|_X \end{aligned}$$

可得

$$F(u) = \lim_{k \rightarrow \infty} F(u_k) = \alpha.$$

可见  $u \in X$  是最小化问题(14.71)和变分问题(14.69)的解。

2. 证明  $u \in X$  是最小化问题和变分问题的唯一解。假定还存在另一个  $\tilde{u} \in X$ , 也是(14.71)和(14.69)的解, 那么我们有

$$\langle \tilde{u}, \nu \rangle_X = \langle f, \nu \rangle, \quad \forall \nu \in X.$$

将上式代回变分问题(14.69) $\Rightarrow$

$$\langle u - \tilde{u}, \nu \rangle_X = 0, \quad \forall \nu \in X.$$

若设  $\nu = u - \tilde{u} \Rightarrow$

$$\|u - \tilde{u}\|_X^2 = 0,$$

因此我们有  $u = \tilde{u}$ , 即  $u \in X$  是最小化问题和变分问题的唯一解。

3. 证明范式等价  $\|u\|_X = \|f\|_{X'}$ 。已知

$$\begin{aligned} \|u\|_X^2 &= \langle u, u \rangle_X = \langle f, u \rangle \leq \|u\|_X \|f\|_{X'}, \\ \hookrightarrow \|u\|_X &\leq \|f\|_{X'}, \end{aligned}$$

$$\|f\|_{X'} = \sup_{0 \neq \nu \in X} \frac{|\langle f, \nu \rangle|}{\|\nu\|_X} = \sup_{0 \neq \nu \in X} \frac{|\langle u, \nu \rangle_X|}{\|\nu\|_X} \leq \|u\|_X,$$

则我们有  $\|u\|_X \leq \|f\|_{X'}$ , &  $\|f\|_{X'} \leq \|u\|_X \Rightarrow \|u\|_X = \|f\|_{X'}$ .

□

**Definition 14.40** (里兹映射). 若里兹表现定理 (Theorem 14.13) 成立, 那么我们将映射  $J : X' \mapsto X, u = Jf$  称为里兹映射 (Riesz map), 满足如下变分问题

$$\langle Jf, \nu \rangle_X = \langle f, \nu \rangle, \quad \forall \nu \in X, \tag{14.74}$$

并且其范数为

$$\|Jf\|_X = \|f\|_{X'}. \tag{14.75}$$

### 14.3.2 椭圆算子

里兹表现定理 (Theorem 14.13) 探讨了算子方程(14.68)及变分问题(14.69)的解  $u \in X$  的存在性以及唯一性。除此以外, 为了确保解得唯一存在, 我们还需要对算子  $A$  和双线性形式  $a(., .)$  做出进一步设定。

**Definition 14.41** (椭圆算子). 一个算子  $A : X \mapsto X'$  被称作  $X$ -椭圆算子, 如果它满足

$$\langle A\nu, \nu \rangle \geq C_1^A \|\nu\|_X^2, \quad \forall \nu \in X, \tag{14.76}$$

其中  $0 \leq C_1^A \in \mathbb{R}$ 。

**Theorem 14.14** (拉克斯—密格拉蒙定理). 设  $A : X \mapsto X'$  是一个有界的  $X$ -椭圆算子。对于任一  $f \in X'$ , 算子方程(14.68)都存在一个唯一解  $u \in X$ , 满足

$$\|u\|_X \leq \frac{1}{C_1^A} \|f\|_{X'}. \quad (14.77)$$

证明. 设存在一个里兹映射 (Riesz map) 算子  $J : X' \mapsto X$ , 满足(14.74)定义。那么算子方程(14.68)等价于下属定点方程

$$u = u - \varrho J(Au - f) = T_\varrho u + \varrho Jf,$$

其中算子  $T_\varrho := I - \varrho JA : X \mapsto X$ , 参数  $0 < \varrho \in \mathbb{R}$ 。对应范数

$$\begin{aligned} \|T_\varrho u\|_X^2 &= \|(I - \varrho JA)u\|_X^2 \\ &= \|u\|_X^2 - 2\varrho \underbrace{\langle JAu, u \rangle_X}_{=: \mathcal{A}} + \varrho^2 \underbrace{\|JAu\|_X^2}_{=: \mathcal{B}} \\ &\leq \left[ 1 - 2\varrho C_1^A + \varrho^2 (C_2^A)^2 \right] \|u\|_X^2, \end{aligned} \quad (14.78)$$

其中, 由里兹映射算子  $J$  的性质(14.74)和  $X$ -椭圆算子  $A$  的性质(14.76)我们有

$$\mathcal{A} := \langle JAu, u \rangle_X = \langle Au, u \rangle \geq C_1^A \|u\|_X^2,$$

由里兹映射算子  $J$  的范数(14.75)和  $X$ -椭圆算子  $A$  的范数(14.66)我们有

$$\mathcal{B} := \|JAu\|_X = \|Au\|_{X'} \leq C_2^A \|u\|_X.$$

若设  $\varrho \in \left(0, \frac{2C_1^A}{(C_2^A)^2}\right)$ , 则算子  $T_\varrho$  是一个  $X$  中的收缩映射 (contraction mapping), 并且由收缩映射定理 (Banach's contraction mapping theorem, Palais (2007)) 可得, 算子方程(14.68)的解  $x \in X$  是唯一的。进而, 对于唯一的解  $u \in X$ , 根据椭圆算子  $A$  的定义(14.76)和里兹表现定理我们有

$$\begin{aligned} C_1^A \|u\|_X^2 &\leq \langle Au, u \rangle \\ &= \langle f, u \rangle \\ &\leq \|f\|_{X'} \|u\|_X. \end{aligned}$$

□

根据拉克斯—密格拉蒙定理 (Theorem 14.14), 我们可以定义一个逆算子  $A^{-1} : X' \mapsto X$ , 有

$$\|A^{-1}f\|_X \leq \frac{1}{C_1^A} \|f\|_{X'}, \quad \forall f \in X'.$$

**Lemma 14.12** ( $X$ -椭圆算子  $A$  的逆算子  $A^{-1}$  也是一个椭圆算子). 设  $A : X \mapsto X'$  是一个有界(14.66), 自伴随(14.67)的  $X$ -椭圆算子(14.77)。那么对于  $\forall \nu \in X$  我们有

$$\langle A^{-1}f, f \rangle \geq \frac{1}{C_2^A} \|f\|_{X'}^2, \quad \forall f \in X'.$$

证明. 定义一个算子  $B := JA : X \mapsto X$ , 满足

$$\|B\nu\|_X = \|JA\nu\|_X = \|A\nu\|_{X'} \leq C_2^A \|\nu\|_X, \quad \forall \nu \in X.$$

由于  $\forall u, \nu \in X$ , 都有以下关系成立

$$\langle Bu, \nu \rangle_X = \langle JAu, \nu \rangle = \langle Au, \nu \rangle = \langle u, A\nu \rangle = \langle u, JA\nu \rangle_X = \langle u, B\nu \rangle_X,$$

可见  $B$  是一个自伴随的椭圆算子, 满足

$$\langle B\nu, \nu \rangle_X = \langle A\nu, \nu \rangle \geq C_1^A \|\nu\|_X^2, \quad \forall \nu \in X.$$

因此可以定义一个可逆的自伴随算子  $B^{\frac{1}{2}}$ , 满足  $B = B^{\frac{1}{2}} B^{\frac{1}{2}}$ , 逆算子  $B^{-\frac{1}{2}} := (B^{\frac{1}{2}})^{-1}$ 。对应的范数

$$\begin{aligned} \|B^{\frac{1}{2}}\nu\|_X^2 &= \langle B\nu, \nu \rangle_X \leq \|B\nu\|_X \|\nu\|_X \leq C_2^A \|\nu\|_X^2, \\ \hookrightarrow \|B^{\frac{1}{2}}\nu\|_X &\leq \sqrt{C_2^A} \|\nu\|_X, \quad \forall \nu \in X. \end{aligned}$$

因此, 对于任一  $f \in X'$ , 我们有

$$\begin{aligned} \|f\|_{X'} &= \sup_{0 \neq \nu \in X} \frac{|\langle f, \nu \rangle|}{\|\nu\|_X} \\ &= \sup_{0 \neq \nu \in X} \frac{|\langle Jf, \nu \rangle_X|}{\|\nu\|_X} \\ &= \sup_{0 \neq \nu \in X} \frac{|\langle B^{-\frac{1}{2}}Jf, B^{\frac{1}{2}}\nu \rangle_X|}{\|\nu\|_X} \\ &\leq \sup_{0 \neq \nu \in X} \frac{\|B^{-\frac{1}{2}}Jf\|_X \|B^{\frac{1}{2}}\nu\|_X}{\|\nu\|_X} \\ &\leq \sqrt{C_2^A} \|B^{-\frac{1}{2}}Jf\|_X, \end{aligned}$$

进而

$$\|f\|_{X'}^2 \leq C_2^A \|B^{-\frac{1}{2}}Jf\|_X^2 = C_2^A \langle B^{-1}Jf, Jf \rangle_X = C_2^A \langle A^{-1}f, f \rangle,$$

其中我们使用到了如下关系

$$\begin{aligned} \|B^{-\frac{1}{2}}Jf\|_X^2 &= \|(B^{\frac{1}{2}})^{-1}Jf\|_X^2 \\ &= \langle B^{-1}Jf, Jf \rangle_X \\ &= \langle A^{-1}f, f \rangle. \end{aligned}$$

□

### 14.3.3 算子与稳定性条件

设  $\Pi$  是一个巴拿赫空间 (Banach space), 设  $B : X \mapsto \Pi'$  是一个有界的线性算子, 满足条件

$$\|B\nu\|_{\Pi'} \leq C_2^B \|\nu\|_X, \quad \forall \nu \in X. \quad (14.79)$$

算子  $B$  意味着如下双线性形式  $b(., .) : X \times \Pi \mapsto \mathbb{R}$

$$b(\nu, q) := \langle B\nu, q \rangle, \quad (\nu, q) \in X \times \Pi.$$

$B$  的核或称零空间 (kernel, null space) 定义为

$$\ker B := \{\nu \in X : B\nu = 0\}. \quad (14.80)$$

$\ker B$  在希尔伯特空间  $X$  中的正交补 (orthogonal complement) 为

$$(\ker B)^\perp := \{w \in X : \langle w, \nu \rangle_X = 0, \forall \nu \in \ker B\} \subset X. \quad (14.81)$$

进而我们有  $\ker B$  的正交空间

$$(\ker B)^0 := \{f \in X' : \langle f, \nu \rangle = 0, \forall \nu \in \ker B\} \subset X'. \quad (14.82)$$

对于某一给定的  $g \in \Pi'$ , 我们想要求得以下算子方程的解  $u \in X$

$$Bu = g. \quad (14.83)$$

将  $B : X \mapsto \Pi'$  的值域或称像 (range, image), 定义为

$$\text{Im}_X B := \{B\nu \in \Pi', \forall \nu \in X\}.$$

则算子方程(14.83)要求是可解的 (solvability condition), 即要求  $g$  在  $B$  的值域中

$$g \in \text{Im}_X B. \quad (14.84)$$

将  $B$  的伴随算子 (adjoint operator) 定义为  $B' : X \mapsto \Pi'$ , 满足

$$\langle \nu, B'q \rangle := \langle B\nu, q \rangle, \quad \forall (\nu, q) \in X \times \Pi.$$

由  $B$  的性质 (14.80), (14.81), (14.82) 可得伴随算子  $B'$  的性质

$$\ker B' := \{q \in \Pi : \langle B\nu, q \rangle = 0, \forall \nu \in X\}, \quad (14.85)$$

$$(\ker B')^\perp := \{p \in \Pi : \langle p, q \rangle_\Pi = 0, \forall q \in \ker B'\}, \quad (14.86)$$

$$(\ker B')^0 := \{g \in \Pi' : \langle g, q \rangle = 0, \forall q \in \ker B'\}. \quad (14.87)$$

$\text{Im}_X B$  的性质, 由闭值域定理 (closed range theorem) 给出

**Theorem 14.15** (闭值域定理). 设  $X$  和  $\Pi$  是巴拿赫空间, 有界线性算子  $B : X \mapsto \Pi'$ 。则以下属性等价

- $\text{Im}_X B$  是  $\Pi'$  中的闭集,
- $\text{Im}_\Pi B'$  是  $X'$  中的闭集,
- $\text{Im}_X B = (\ker B')^0$ ,
- $\text{Im}_\pi B' = (\ker B)^0$ 。

证明. 略。可参考 (Muscat, 2014, Proposition 11.30)。  $\square$

可求解性条件(14.84) $\Rightarrow$

$$\langle g, q \rangle = 0, \forall q \in \ker B' \subset \pi. \quad (14.88)$$

若可求解性条件(14.84)或(14.88)得到满足, 则算子方程(14.83)存在至少一个解  $u \in X$ 。但解并不唯一: 我们可以加入任一  $u_0 \in \ker B$ , 使得  $u + u_0$  也是  $B(u + u_0) = g$  的解。因此我们需要引入额外的假设条件  $u \in (\ker B)^\perp$ , 以确保解的唯一性。

**Theorem 14.16** (算子方程解的唯一存在性). 设希尔伯特空间  $X$  和  $\Pi$ 。有界的线性算子  $B : X \rightarrow \Pi'$ 。进而我们有稳定性条件

$$C_S \|v\|_X \leq \sup_{0 \neq q \in \Pi} \frac{\langle Bv, q \rangle}{\|q\|_\Pi}, \quad \forall v \in (\ker B)^\perp, \quad (14.89)$$

那么对于一个给定的  $g \in \text{Im}_X B$ , 算子方程  $Bu = g$  存在一个唯一的解  $u \in (\ker B)^\perp$ , 满足

$$\|u\|_X \leq \frac{1}{C_S} \|g\|_{\Pi'}.$$

证明. 已知根据假设条件  $g \in \text{Im}_X B$ , 算子方程  $Bu = g$  存在唯一的一个解  $u \in (\ker B)^\perp$ , 满足

$$\langle Bu, q \rangle = \langle g, q \rangle, \quad \forall q \in \Pi.$$

现在设存在第二个解  $\bar{u} \in (\ker B)^\perp$ , 满足

$$\langle B\bar{u}, q \rangle = \langle g, q \rangle, \quad \forall q \in \Pi,$$

则我们有

$$\langle B(u - \bar{u}), q \rangle = 0, \quad \forall q \in \Pi.$$

显然  $u - \bar{u} \in (\ker B)^\perp$  也满足稳定性条件(14.89)

$$0 \leq C_S \|u - \bar{u}\|_X \leq \sup_{0 \neq q \in \Pi} \frac{\langle B(u - \bar{u}), q \rangle}{\|q\|_\Pi}, \quad \forall v \in (\ker B)^\perp = 0,$$

$$\hookrightarrow u = \bar{u}.$$

把唯一解  $u$  代回(14.89)我们有

$$\begin{aligned} C_S \|u\|_X &\leq \sup_{0 \neq q \in \Pi} \frac{\langle Bu, q \rangle}{\|q\|_\Pi} \\ &= \sup_{0 \neq q \in \Pi} \frac{\langle g, q \rangle}{\|q\|_\Pi} \\ &\leq \|g\|_{\Pi'}. \end{aligned}$$

□

#### 14.3.4 含有限制条件的算子方程

经验研究中我们常常需要求得带有约束条件  $Bu = g$  的算子方程  $Au = f$  的解。常见的求解思路分为四步。

1. 关于限定条件  $Bu = g$ , 假定可求解条件(14.84)成立

$$g \in \text{Im}_X B := \{B\nu \in \Pi' : \forall \nu \in X\}.$$

对于给定的  $g \in \Pi'$ , 定义流形  $V_g$

$$V_g := \{\nu \in X : B\nu = g\}.$$

此外我们定义零空间  $V_0$

$$V_0 = \ker B := \{\nu \in X : B\nu = 0\}.$$

2. 关于算子方程  $Au = f$ , 同样假定可求解条件(14.84)成立

$$f \in \text{Im}_{V_g} A := \{A\nu \in X', \forall \nu \in V_g\}.$$

对于给定的  $f \in X'$ , 构建变分问题

$$\langle Au, \nu \rangle = \langle f, \nu \rangle, \quad \forall \nu \in V_0, \quad (14.90)$$

求解该问题, 得到解  $u \in V_g$ 。

3. 所求得解  $u \in V_g$  的唯一性, 见 Theorem 14.17。

4. 对于唯一存在解  $u \in V_g$ , 可以将  $u$  的范数, 和给定的  $f \in X', g \in \Pi'$  的范数联系起来。

假定对于给定的  $g \in \text{Im}_X B$ ,  $\exists u_g \in V_g$ , 我们有

$$\|u_g\|_X \leq C_B \|g\|_{\Pi'}, \quad C_B > 0 \in \mathbb{R}. \quad (14.91)$$

, 则范数之间的关联见 Corollary 14.4。

**Theorem 14.17** (带约束算子方程解的唯一存在性). 设一个有界线性  $V_0$ -椭圆算子  $A : X \mapsto X'$

$$\langle A\nu, \nu \rangle \geq C_1^A \|\nu\|_X^2, \quad \forall \nu \in V_0 := \ker B,$$

其中  $B : X \mapsto \Pi'$ 。并且有给定的  $f, g$

$$\begin{aligned} f \in \text{Im}_{V_g} A &:= \{A\nu \in X', \forall \nu \in V_g\}, \\ g \in \text{Im}_X B &:= \{B\nu \in \Pi', \forall \nu \in X\}. \end{aligned} \quad (14.92)$$

那么作为带有约束条件  $Bu = g$  的算子方程  $Au = f$ , 有且只有一个解  $u \in X$ 。

证明. 由已知条件  $g \in \text{Im}_X B$  可得, 约束条件存在至少一个解  $u_g \in X$  满足  $Bu_g = g$ 。除此而外, 我们还需要求得  $u_0 := u - u_g \in V_0$ , 作为以下算子方程的解

$$Au_0 = f - Au_g,$$

算子方程等价于以下变分问题

$$\langle Au_0, \nu \rangle = \langle f - Au_g, \nu \rangle, \quad \forall \nu \in V_0.$$

[存在性] 由已知条件  $f \in \text{Im}_{V_g} A$  可得,  $f - Au_g \in \text{Im}_{V_0} A$ , 那么方程  $Au_0 = f - Au_g$  至少存在一个解  $u_0 \in V_0$ 。

[唯一性] 现在来证明  $u_0 \in V_0$  是唯一的。设  $\bar{u}_0 \in V_0$  是算子方程的另一个解, 满足  $A\bar{u}_0 = f - Au_g$ 。由已知条件  $A$  的  $V_0$ -椭圆特性可得

$$\begin{aligned} 0 \leq C_1^A \|u_0 - \bar{u}_0\|_X^2 &\leq \langle A(u_0 - \bar{u}_0), u_0 - \bar{u}_0 \rangle = \langle Au_0 - A\bar{u}_0, u_0 - \bar{u}_0 \rangle = 0, \\ \Rightarrow u_0 &\in X = \bar{u}_0 \in X. \end{aligned}$$

$u_g \in V_g$  可能不是唯一的解, 但  $u \in X = u_0 + u_g$  却是变分问题的唯一最终解, 并不受 (可能是多重的)  $u_g \in V_g$  的影响。这是由于, 对于满足约束条件方程  $B\hat{u}_g = g$  的某一个解  $\hat{u}_g \in X$  而言, 这意味着存在唯一一个  $\hat{u}_0 \in V_0$ , 构成算子方程  $A(\hat{u}_0 + \hat{u}_g) = f$  的解, 进而

$$\begin{aligned} B(u_g - \hat{u}_g) &= Bu_g - B\hat{u}_g = g - g = 0 \in \Pi', \\ \Leftrightarrow u_g - \hat{u}_g &\in \ker B = V_0. \end{aligned}$$

由于

$$\begin{aligned} A(u_0 + u_g) &= f, \\ A(\hat{u}_0 + \hat{u}_g) &= f, \end{aligned}$$

我们因而有

$$A(u_0 + u_g - \hat{u}_0 - \hat{u}_g) = 0.$$

显然,  $u_0 - \hat{u}_0 + (u_g - \hat{u}_g) \in V_0$ , 由  $A$  的  $V_0$ -椭圆特性我们有

$$u_0 - \hat{u}_0 + (u_g - \hat{u}_g) = 0,$$

因此我们可得解的唯一性

$$u = u_0 + u_g = \hat{u}_0 + \hat{u}_g.$$

□

**Corollary 14.4.** 已知  $u \in V_g$  是有限制条件的算子方程唯一解 (Theorem 14.17), 并且满足假定(14.91)。那么  $\|u\|_X$  和给定的  $f \in X'$ ,  $g \in \Pi'$  的范数之间关系为

$$\|u\|_X \leq \frac{1}{C_1^A} \|f\|_{X'} + \left(1 + \frac{C_2^A}{C_1^A}\right) C_B \|g\|_{\Pi'}.$$

证明. 由 Theorem 14.17 得, 算子方程  $Au = f$  的解表现为  $u = u_g + u_0$  的形式, 其中  $u_0 \in V_0$  是以下变分问题的唯一解

$$\langle Au_0, \nu \rangle = \langle f - Au_g, \nu \rangle, \quad \forall \nu \in V_0.$$

由算子  $A$  的  $V_0$ -椭圆特性可得

$$\begin{aligned} C_1^A \|u_0\|_X^2 &\leq \langle Au_0, u_0 \rangle = \langle f - Au_g, u_0 \rangle \leq \|f - Au_g\|_{X'} \|u_0\|_X, \\ \hookrightarrow \|u_0\|_X &\leq \frac{1}{C_1^A} [\|f\|_{X'} + C_2^A \|u_g\|_X], \\ \Rightarrow & \end{aligned}$$

$$\begin{aligned} \|u\|_X &= \|u_0 + u_g\|_X \\ &\leq \|u_0\|_X + \|u_g\|_X \\ &\leq \frac{1}{C_1^A} \|f\|_{X'} + \left(1 + \frac{C_2^A}{C_1^A}\right) \|u_g\|_X \\ &\leq \frac{1}{C_1^A} \|f\|_{X'} + \left(1 + \frac{C_2^A}{C_1^A}\right) C_B \|g\|_{\Pi'}. \end{aligned}$$

□

### 14.3.5 混合算子方程 (鞍点变分问题)

第14.3.4节讨论了如何构建带有限制条件的算子方程来求解变分问题。除此而外的另一种方法是引入拉格朗日乘子  $p \in \Pi$ , 构建扩展变分问题, 对于  $\forall (\nu, q) \in X \times \Pi$ , 求解  $(u, p) \in X$ , 使其满足

$$\langle Au, \nu \rangle + \langle Bu, p \rangle = \langle f, \nu \rangle, \tag{14.93a}$$

$$\langle Bu, q \rangle = \langle g, q \rangle, \tag{14.93b}$$

其中  $u \in V_g$  是  $Au = f$  的解。

与上节相同, (14.93b)同样可以用于描述限制条件  $Bu = g$ 。但不同的是(14.93a)可理解为另一个变分法问题: 将  $\nu \in V_0$  作为检验方程, 求解  $u_0 \in V_0$ 。显然, 这种混合算子求解变分问题的研究思路, 可行前提之一是确保存在拉格朗日乘子  $p \in \Pi$ , 使得  $\forall \nu \in X$  都满足等式(14.93a)。具体来说就是, 对于一组  $(\nu, q) \in X \times \Pi$ , 定义一个拉格朗日泛函

$$\mathcal{L}(\nu, q) := \frac{1}{2} \langle A\nu, \nu \rangle - \langle f, \nu \rangle + \langle B\nu, q \rangle - \langle g, q \rangle,$$

应当使得 Theorem (14.18)成立。

**Theorem 14.18.** 设  $A : X \mapsto X'$  为一个有界线性算子,  $\forall \nu \in X$  都具有自伴随  $\langle A\nu, n\nu \rangle = \langle \nu, A\nu \rangle$ 、正半定  $\langle A\nu, \nu \rangle \geq 0$  的特性。设另一个有界线性算子  $B : X \mapsto \Pi'$ 。

当且仅当

$$\mathcal{L}(u, q) \leq \mathcal{L}(u, p) \leq \mathcal{L}(\nu, p) \quad \forall (\nu, q) \in X \times \Pi \quad (14.94)$$

时,  $(u, p)$  成为变分问题(14.93a)-(14.93b)的一个解。

证明. 假设有一组解  $(u, p) \in X \times \Pi$ 。

1. 证(14.94)的后半部分。

$$\begin{aligned} & \mathcal{L}(\nu, p) - \mathcal{L}(u, p) \\ &= \frac{1}{2} \langle A\nu, \nu \rangle - \langle f, \nu \rangle + \langle B\nu, p \rangle - \langle g, p \rangle - \frac{1}{2} \langle Au, u \rangle + \langle f, u \rangle - \langle Bu, p \rangle + \langle g, p \rangle \\ &= \underbrace{\frac{1}{2} \langle A(u - \nu), (u - \nu) \rangle}_{\text{正半定, } \geq 0} + \underbrace{\langle Au, (u - \nu) \rangle + \langle B(u - \nu), p \rangle - \langle f, u - \nu \rangle}_{(14.93a), = 0} \\ &\geq 0, \end{aligned}$$

$$\therefore \mathcal{L}(u, p) < \mathcal{L}(\nu, p), \quad \forall \nu \in X.$$

2. 证(14.94)的前半部分。

$$\begin{aligned} & \mathcal{L}(u, p) - \mathcal{L}(u, q) \\ &= \frac{1}{2} \langle Au, u \rangle - \langle f, u \rangle + \langle Bu, p \rangle - \langle g, p \rangle - \frac{1}{2} \langle Au, u \rangle + \langle f, u \rangle - \langle Bu, q \rangle + \langle g, q \rangle \\ &= \underbrace{\langle Bu, p - q \rangle - \langle g, p - q \rangle}_{(14.93b), = 0} \\ &= 0, \end{aligned}$$

$$\therefore \mathcal{L}(u, q) \leq \mathcal{L}(\nu, p), \quad \forall q \in \Pi.$$

3. 假设已知某个  $p \in \Pi$  是方程的解。构建如下最小化问题

$$\mathcal{L}(u, p) \leq \mathcal{L}(\nu, p), \quad \forall \nu \in X, \quad (14.95)$$

求解  $u \in X$  满足式(14.93a)。

设某个  $u \in X$  是最小化问题的解, 我们有  $\forall w \in X$  满足以下两式

$$\frac{d}{dt} \mathcal{L}(u + tw, p) |_{t=0} = 0. \quad (14.96)$$

以及

$$\begin{aligned}\mathcal{L}(u + tw, p) &= \underbrace{\frac{1}{2} \langle Au, u \rangle - \langle f, u \rangle + \langle Bu, p \rangle - \langle g, p \rangle}_{=\mathcal{L}(u, p)} \\ &\quad + \frac{1}{2} t^2 \langle Aw, w \rangle + t [\langle Au, w \rangle + \langle Bw, p \rangle - \langle f, w \rangle],\end{aligned}\tag{14.97}$$

进而(14.96)可得, (14.97) $\Rightarrow$

$$\langle Au, w \rangle + \langle Bw, p \rangle - \langle f, w \rangle = 0, \forall w \in X,$$

满足式(14.93a)。

4. 对于任一  $q \in \Pi$ , 证明(14.93b)。

(a) 定义  $\tilde{q} := p + q$ , 进而

$$\begin{aligned}0 &\leq \mathcal{L}(u, p) - \mathcal{L}(u, p + q) \\ &= \frac{1}{2} \langle Au, u \rangle - \langle f, u \rangle + \langle Bu, p \rangle - \langle g, p \rangle \\ &\quad - \frac{1}{2} \langle Au, u \rangle + \langle f, u \rangle - \langle Bu, p + q \rangle + \langle g, p + q \rangle \\ &= -\langle Bu, q \rangle + \langle g, q \rangle.\end{aligned}$$

(b) 定义  $\tilde{q} := p - q$ , 进而

$$0 \leq \mathcal{L}(u, p) - \mathcal{L}(u, p - q) = \langle Bu, q \rangle - \langle g, q \rangle.$$

(c)

$$\therefore \langle Bu, q \rangle = \langle g, q \rangle, \quad \forall q \in \Pi. \tag{14.98}$$

□

由此可见, 扩展变分问题(14.93a)-(14.93b)中, 组合  $(u, p) \in X \times \Pi$  是一个拉格朗日泛函  $\mathcal{L}(.,.)$  的鞍点。从这意义上说, 扩展变分问题也常称为鞍点变分问题。下面来探讨  $(u, p)$  解的唯一性。

**Theorem 14.19** (混合算子方程 (鞍点变分问题) 的解). 假设巴拿赫空间  $X, \Pi$ , 有界算子  $A : X \mapsto X'$ ,  $B : X \mapsto \Pi'$ 。设  $A$  满足  $V_0$ -椭圆特性

$$\langle A\nu, \nu \rangle \geq C_1^A \|\nu\|_{X'}^2, \quad \forall \nu \in V_0 = \ker B,$$

设稳定性条件

$$C_s \|q\|_{\Pi} \leq \sup_{0 \neq \nu \in X} \frac{\langle B\nu, q \rangle}{\|\nu\|_{X'}}, \quad \forall q \in \Pi. \tag{14.99}$$

那么对于  $g \in \text{Im}_X B, f \in \text{Im}_{V_g} A$ , 扩展变分问题(14.93a)-(14.93b)都存在唯一的解  $(u, p) \in X \times \Pi$ , 满足如下关系

$$\|u\|_X \leq \frac{1}{C_1^A} \|f\|_{X'} + \left(1 + \frac{C_2^A}{C_1^A}\right) C_B \|g\|_{\Pi'}, \tag{14.100}$$

$$\|p\|_{\Pi} \leq \frac{1}{C_S} \left(1 + \frac{C_2^A}{C_1^A}\right) \{\|f\|_{X'} + C_B C_2^A \|g\|_{\Pi'}\}. \tag{14.101}$$

证明. 1. 证明对于变分问题

$$\begin{aligned}\langle Au, \nu \rangle &= \langle f, \nu \rangle, \quad \forall \nu \in V_0, \\ \langle Bu, q \rangle &= \langle g, q \rangle, \quad \forall q \in P_i\end{aligned}$$

存在唯一的解  $u \in X$ 。

唯一解的存在性由 Theorem 14.17 证得。唯一解的范数不等式由 Corollary 14.4 给出，对应式(14.100)。

2. 证明对于变分问题

$$\langle B\nu, p \rangle = \langle f - Au, \nu \rangle, \quad \forall \nu \in X$$

存在唯一的解  $p \in \Pi$ 。

(a) 解的存在性。我们有  $f - Au \in (\ker B)^0$ ，进而根据闭值域定理 (Theorem 14.15) 我们有  $f - Au \in \text{Im}_\Pi(B')$ ，进而变分问题的解是  $p \in \Pi$ 。

(b) 解的唯一性。假定变分问题有两个解  $p, \hat{p} \in \Pi$ ，满足

$$\begin{aligned}\langle B\nu, p \rangle &= \langle f - Au, \nu \rangle, \quad \forall \nu \in X, \\ \langle B\nu, \hat{p} \rangle &= \langle f - Au, \nu \rangle, \quad \forall \nu \in X.\end{aligned}$$

两式相减  $\Rightarrow$

$$\langle B\nu, p - \hat{p} \rangle = 0, \quad \forall \nu \in X.$$

代入稳定性条件(14.99)有

$$\begin{aligned}0 \leq C_s \|p - \hat{p}\|_\Pi &\leq \sup_{0 \neq \nu \in X} \frac{\langle B\nu, p - \hat{p} \rangle}{\|\nu\|_X} \leq 0, \quad \forall q \in \Pi, \\ \hookrightarrow p &= \hat{p} \in \Pi.\end{aligned}$$

(c) 计算唯一解  $p \in \Pi$  的范数。再次代入稳定性条件(14.99)有

$$C_s \|p\|_\Pi \leq \sup_{0 \neq \nu \in X} \frac{\langle B\nu, p \rangle}{\|\nu\|_X} = \sup_{0 \neq \nu \in X} \frac{\langle f - Au, \nu \rangle}{\|\nu\|_X} \leq \|f\|_{X'} + C_2^A \|u\|_X,$$

代入(14.100)替换  $\|u\|_X$ ，我们得(14.100)。

□

设  $A : X \mapsto X'$  是个  $V_0$ -椭圆算子时，Theorem 14.19 成立。此外，若假设  $A$  是个  $X$ -椭圆算子，即

$$\langle A\nu, \nu \rangle \geq C_1^A \|\nu\|_X^2, \quad \forall \nu \in X, \tag{14.102}$$

Theorem 14.19 依然成立。

Theorem 14.19 探讨了求扩展变分问题的解  $(u, p) \in X \times \Pi$ 。事实上求解过程可以进一步简化，

$u$  是一个关于  $p$  的方程。已知

$$\begin{aligned} & \langle Au, \nu \rangle + \langle B\nu, p \rangle = \langle f, \nu \rangle, \\ \hookrightarrow & \langle B\nu, p \rangle = \langle \nu, Bp \rangle = \langle B'p, \nu \rangle, \\ \hookrightarrow & \langle Au, \nu \rangle + \langle B'p, \nu \rangle = \langle f, \nu \rangle, \\ \hookrightarrow & Au + B'p = f, \\ \hookrightarrow & u = A^{-1}(f - B'p) \\ \hookrightarrow & \langle Bu, q \rangle = \langle g, q \rangle, \\ \hookrightarrow & \langle BA^{-1}(f - B'p), q \rangle = \langle g, q \rangle, \\ \hookrightarrow & \langle BA^{-1}f - g, q \rangle = \langle BA^{-1}B'p, q \rangle, \end{aligned}$$

即对于任一  $p \in \Pi$ , 都存在唯一的一个解  $u = A^{-1}(f - B'p) \in X$ 。这样一来, 原本寻找  $(u, p) \in X \times \Pi$  的变分问题, 就变成了一个新的(椭圆)变分问题: 寻找解  $p \in \Pi$  使满足

$$\langle BA^{-1}f - g, q \rangle = \langle BA^{-1}B'p, q \rangle. \quad (14.103)$$

为了探讨(椭圆)变分问题(14.103)的解  $u$  的唯一性, 首先我们检验它是否符合拉克斯—密格拉蒙定理(Theorem 14.14)的前提假; 如果是, 则应用该定理。检验过程见 Lemma 14.13.

**Lemma 14.13.** 设 Theorem 14.19 的假设条件均得到满足。

那么算子  $S := BA^{-1}B'$  有界, 并且由稳定条件(14.99)可得  $S$  是  $\Pi$ -椭圆的, 满足

$$\langle Sq, q \rangle \geq C_1^S \|q\|_{\Pi}^2, \quad \forall q \in \Pi. \quad (14.104)$$

证明. 已知即对于任一  $p \in \Pi$ , 对于如下变分问题

$$\langle Au, \nu \rangle = \langle B\nu, q \rangle \quad \forall \nu \in X,$$

都存在唯一的一个解  $u = A^{-1}(f - B'p) \in X$ .

1.  $S$  的有界性。已知  $A : X \mapsto X'$  的  $X$ -椭圆特性(14.102), 那么根据 Theorem 14.19 可证得存在唯一解  $u \in X$  满足

$$\begin{aligned} \|u\|_X &= \|A^{-1}B'q\|_X \\ &\leq \frac{1}{C_1^A} \|B'q\|_{X'} \\ &\leq \frac{C_2^B}{C_1^A} \|q\|_{\Pi}, \quad \forall q \in \Pi. \end{aligned} \quad (14.105)$$

由(14.105)可得

$$\begin{aligned} \|Sq\|_{\Pi'} &= \|BA^{-1}B'q\|_{\Pi'} \\ &= \|Bu\|_{\Pi'} \\ &\leq C_2^B \|u\|_X \\ &\leq \frac{(C_2^B)^2}{C_1^A} \|q\|_{\Pi}, \quad \forall q \in \Pi, \end{aligned}$$

即  $S : \Pi \mapsto \Pi'$  有界。

2.  $S$  的椭圆性。

$$\begin{aligned}
 \langle Sq, q \rangle &= \langle BA^{-1}B'q, q \rangle \\
 &= \langle Bu, q \rangle \\
 &= \langle Au, u \rangle \\
 &\geq C_1^A \|u\|_X^2,
 \end{aligned} \tag{14.106}$$

根据稳定条件(14.99),

$$\begin{aligned}
 C_S \|q\|_\Pi &\leq \sup_{0 \neq \nu \in X} \frac{\langle B\nu, q \rangle}{\|\nu\|_X} \\
 &= \sup_{0 \neq \nu \in X} \frac{\langle Au, \nu \rangle}{\|\nu\|_X} \\
 &\leq C_2^A \|u\|_X,
 \end{aligned}$$

代回(14.106)最后得

$$\langle Sq, q \rangle \geq C_1^A \|u\|_X^2 \geq \frac{(C_2^B)^2}{C_1^A} \|q\|_\Pi^2,$$

由此证得(14.105),  $C_1^S := \frac{(C_2^B)^2}{C_1^A}$ 。

□

进而, 我们可以使用拉克斯一密格拉蒙定理 (Theorem 14.14) 求得椭圆问题(14.103)的唯一解  $p \in \Pi$ 。

在此基础上, 回到混合算子方程 (鞍点变分问题) (14.93a) -(14.93b)上来:

**Theorem 14.20** (混合算子方程 (鞍点变分问题) 的解 (续)). 设巴拿赫空间  $X, \Pi$ 。有界算子  $A : X \mapsto X'$ ,  $B : X \mapsto \Pi'$ 。假设  $A$  是  $X$ -椭圆的, 满足稳定条件(14.99)。对于给定的  $f \in X'$ ,  $g \in \Pi'$ , 混合算子方程 (鞍点变分问题) (14.93a) -(14.93b) 存在唯一的解  $(u, p) \in X \times \Pi$ , 满足

$$\|p\|_\Pi \leq \frac{1}{C_1^S} \|BA^{-1}f - g\|_{\Pi'} \leq \frac{1}{C_1^S} \left[ \frac{C_2^B}{C_1^A} \|f\|_{X'} + \|g\|_{\Pi'} \right], \tag{14.107}$$

$$\|u\|_X \leq \frac{1}{C_1^A} \left[ 1 + \frac{(C_2^B)^2}{C_1^A C_1^S} \right] \|f\|_{X'} + \frac{C_2^B}{C_1^A C_1^S} \|g\|_{\Pi'}. \tag{14.108}$$

证明. 1. 根据 Theorem 14.19 可证得(14.107)。

2. 根据拉克斯一密格拉蒙定理 (Theorem 14.14), 可以证明混合算子方程 (鞍点变分问题)(14.93a) -(14.93b) 存在唯一解, 满足

$$\langle Au, \nu \rangle = \langle f - B'p, \nu \rangle, \quad \forall \nu \in X.$$

由算子  $A$  的  $X$ -椭圆特性可得

$$C_1^A \|u\|_X^2 \leq \langle Au, u \rangle = \langle f - B'p, u \rangle \leq \|f - B'p\|_{X'} \|u\|_X,$$

由此我们有

$$\begin{aligned}
 \|u\|_X &\leq \frac{1}{C_1^A} \|f - B'p\|_{X'} \\
 &\leq \frac{1}{C_1^A} \|f\|_{X'} + \frac{C_2^B}{C_1^A} \|p\|_\Pi.
 \end{aligned}$$

将(14.107)代入上式, 可得(14.108)。

□

### 14.3.6 强制算子方程

前面介绍有界线性算子  $A : X \mapsto X'$ , 假设它具有  $X$ -椭圆的特性, 如(14.76)。这一假设过强。在多数情况下, 我们用强制算子 (coercive operator) 予以替代。

**Definition 14.42** (强制算子). 如果存在一个紧凑算子  $C : X \mapsto X'$ , 使得和  $A : X \mapsto X'$  一道满足 *Gårding 不等式* (*Gårding inequality*)

$$\langle (A + C) \nu, \nu \rangle \geq C_1^A \|\nu\|_X^2, \quad \forall \nu \in X. \quad (14.109)$$

那么我们称  $A$  是一个强制算子 (*coercive operator*)。

*Gårding 不等式* 的详细介绍, 可参考 (Jovanović and Süli, 2014, Theorem 2.4)。证明可见如 (Agranovich, 2015, Theorem 8.1.1), (Renardy and Rogers, 2004, Therorem 9.17)。

紧凑算子的定义。对于  $C : X \mapsto Y$ , 若  $X$  中单位球体 (unit sphere) 的像 (image) 在  $Y$  中相对紧凑, 则我们称  $C$  为紧凑算子 (compact operator)。一个值得关注的特性是, 紧凑算子和有界线性算子的乘也是紧凑的。

根据 Riesz-Schauder 定理 (Riesz-Schauder theorem, (Yosida, 1978, Sec. X.5), (Muscat, 2014, Theorem 14.18)), 我们有弗雷德霍姆二择一定理 (Fredholm alternative theorem)如下。

**Theorem 14.21** (弗雷德霍姆二择一定理). 设  $K : X \mapsto X$  是一个紧凑算子。那么以下两种情况之一会出现:

- 齐次方程 (*homogeneous equation*)

$$(I - K) u = 0$$

有一个不平凡解 (*nontrivial solution*)  $u \in X$ , 或

- 非其次方程

$$(I - K) u = g$$

对于每一个给定的  $g \in X$ , 都有唯一的解  $u \in X$ , 满足关系

$$\|u\|_X \leq c \|g\|_X.$$

证明. 略。可见 (Agranovich, 2015, Sec. 18.1)。  $\square$

基于弗雷德霍姆二择一定理 (Theorem 14.21), 我们可以探讨当有界线性算子  $A$  是强制的时, 算子方程  $Au = f$  的解。

**Theorem 14.22** (强制算子方程的解). 假设一个有界线性算子  $A : X \mapsto X'$ , 具有强制性 (Definition 14.42)、内射性 (*injective*, 即  $Au = 0 \Rightarrow u = 0$ )。那么强制算子方程  $Au = f$  存在唯一的解  $u \in X$ , 并且满足条件

$$\|u\|_X \leq c \|f\|_{X'}.$$

证明. 定义一个线性算子  $D := A + C : X \mapsto X'$ , 由(14.109)可得, 线性有界算子  $D$  也是  $X$ -椭圆的。

通过拉克斯—密格拉蒙定理 (Theorem 14.14 拉克斯—密格拉蒙定理) 可得, 逆算子  $D^{-1} : X' \mapsto X$ 。这样, 我们可以将原强制算子方程  $Au = f$  转换为新的强制算子方程

$$\begin{aligned} Bu &= D^{-1} Au = D^{-1} f, \text{ 其中线性有界算子} \\ B &:= D^{-1} A = D^{-1} (D - C) = I - D^{-1} C : X \mapsto X. \end{aligned} \quad (14.110)$$

由假设条件  $C : X \mapsto X'$  是紧凑算子, 和  $D : X \mapsto X'$ 、进而  $D^{-1} : X' \mapsto X$  是线性有界算子, 可得  $D^{-1}C : X \mapsto X$  是紧凑算子。进而, 可以根据弗雷德霍姆二择一定理 (Theorem 14.21) 证得强制算子方程(14.110)解的唯一存在性。

由假设条件  $A : X \mapsto X'$  的内射性, 可得齐次方程  $D^{-1}Au = 0$  的所有解  $u \in X$  都是平凡解 (trivial solution)。因此, 非齐次方程  $Bu = D^{-1}f$  存在唯一解  $u \in X$ , 满足

$$\|u\|_X \leq c \|D^{-1}f\|_X \leq \tilde{c} \|f\|_{X'}.$$

□

## 14.4 变分法求解边界值问题

在介绍了常见的变分方法 (第14.3节) 之后, 本届关注如何利用变分法分析求解二阶椭圆边界值问题, 尤其是位势方程 (potential equations, 第14.1节) 的边界值。应用变分法构建的边界值问题的若方程, 是有限元分析的重要基础之一。

### 14.4.1 位势方程基本介绍

如第14.1节所讨论, 先介绍三个算子, 分别为实值标量的偏微分算子(14.1)

$$(Lu)(x) := - \sum_{i,j=1}^d \frac{\partial}{\partial x_j} \left[ a_{ji}(x) \frac{\partial}{\partial x_i} u(x) \right], \quad x \in \Omega \in \mathbb{R}^d, \quad (14.111)$$

内界迹算子(14.4)

$$\gamma_0^{int} u(x) := \lim_{\Omega \ni \tilde{x} \rightarrow x \in \Gamma} u(\tilde{x}), \quad x \in \Gamma = \partial\Omega,$$

与之相对应的内部共形导数(14.7)

$$\gamma_1^{int} u(x) := \lim_{\Omega \ni \tilde{x} \rightarrow x \in \Gamma} \left[ \sum_{i,j=1}^d n_j(x) a_{ji}(\tilde{x}) \frac{\partial}{\partial \tilde{x}_i} u(\tilde{x}) \right], \quad x \in \Gamma = \partial\Omega. \quad (14.112)$$

当  $u, \nu \in H^1(\Omega)$  时,  $Lu \in \tilde{H}^{-1}(\Omega)$ , 格林第一恒等式 (14.6)化简为

$$\begin{aligned} a(u, \nu) &= \int_{\Omega} (Lu)(x) \nu(x) dx + \int_{\Gamma} [\gamma_1^{int} u(x)] [\gamma_0^{int} \nu(x)] ds_x \\ &= \langle Lu, \nu \rangle_{\Omega} + \langle \gamma_1^{int} u, \gamma_0^{int} \nu \rangle_{\Gamma}, \end{aligned} \quad (14.113)$$

其中对称双线性泛函  $a(.,.)$  的定义

$$a(u, \nu) := \langle u, \nu \rangle = \sum_{i,j=1}^d \int_{\Omega} a_{ji}(x) \frac{\partial}{\partial x_i} u(x) \frac{\partial}{\partial x_j} \nu(x) dx \quad (14.114)$$

**Lemma 14.14.** 假设  $a_{ij} \in L^{\infty}(\Omega)$ ,  $i, j = 1, \dots, d$  满足

$$\|a\|_{L^{\infty}(\Omega)} := \max_{i,j=1,\dots,d} \sup_{x \in \Omega} |a_{ij}(x)|. \quad (14.115)$$

那么我们可得线性形泛函  $a(.,.) : H^1(\Omega) \times H^1(\Omega) \mapsto \mathbb{R}$  有界, 并且满足

$$\begin{aligned} |a(u, \nu)| &\leq c_2^A \|u\|_{H^1(\Omega)} \|\nu\|_{H^1(\Omega)}, \quad \forall u, \nu \in H^1(\Omega), \\ c_2^A &:= d \|a\|_{L^{\infty}(\Omega)}. \end{aligned} \quad (14.116)$$

证明. 由(14.114), (14.115)可得

$$\begin{aligned} |a(u, \nu)| &= \left| \sum_{i,j=1}^d \int_{\Omega} a_{ji}(x) \frac{\partial}{\partial x_i} u(x) \frac{\partial}{\partial x_j} \nu(x) dx \right| \\ &\leq \|a\|_{L^\infty(\Omega)} \int_{\Omega} \left\{ \sum_{i=1}^d \left| \frac{\partial}{\partial x_i} u(x) \right| \sum_{j=1}^d \left| \frac{\partial}{\partial x_j} \nu(x) \right| \right\} dx, \end{aligned}$$

连续两次使用柯西——施瓦茨不等式 (Definition 14.21), 上式变为

$$\begin{aligned} |a(u, \nu)| &\leq \|a\|_{L^\infty(\Omega)} \left\{ \int_{\Omega} \left[ \sum_{i=1}^d \left| \frac{\partial}{\partial x_i} u(x) \right| \right]^2 dx \right\}^{\frac{1}{2}} \left\{ \int_{\Omega} \left[ \sum_{j=1}^d \left| \frac{\partial}{\partial x_j} \nu(x) \right| \right]^2 dx \right\}^{\frac{1}{2}} \\ &\leq \|a\|_{L^\infty(\Omega)} \left\{ \int_{\Omega} d \sum_{i=1}^d \left| \frac{\partial}{\partial x_i} u(x) \right|^2 dx \right\}^{\frac{1}{2}} \left\{ \int_{\Omega} d \sum_{j=1}^d \left| \frac{\partial}{\partial x_j} \nu(x) \right|^2 dx \right\}^{\frac{1}{2}} \\ &= \underbrace{d \|a\|_{L^\infty(\Omega)}}_{=:c_2^A} \|\nabla u\|_{L^2(\Omega)} \|\nabla \nu\|_{L^2(\Omega)}. \end{aligned}$$

□

可见, 由 Lemma 14.14 可得, (14.116) 又进一步表示为

$$|a(u, \nu)| \leq c_2^A \|u\|_{H^1(\Omega)} \|\nu\|_{H^1(\Omega)}, \quad \forall u, \nu \in H^1(\Omega). \quad (14.117)$$

**Lemma 14.15** (双线性算子的半椭圆特性). 设  $L$  是一个如(14.111)所定义的一致椭圆偏微分算子。用双线性形式  $(a., .)$  (14.114) 表示, 我们有

$$a(\nu, \nu) \geq \lambda_0 |\nu|_{H^1(\Omega)}^2, \quad \forall \nu \in H^1(\Omega), \quad (14.118)$$

其中常数  $\lambda_0 > 0$ , 见椭圆算子(14.2)。

证明. 设一个  $w_i(x)$

$$w_i(x) := \frac{\partial}{\partial x_i} \nu(x), \quad i = 1, \dots, d,$$

$$\begin{aligned} \hookrightarrow a(\nu, \nu) &= \sum_{i,j=1}^d \int_{\Omega} a_{ji}(x) \frac{\partial}{\partial x_i} \nu(x) \frac{\partial}{\partial x_j} \nu(x) dx \\ &= \int_{\Omega} [A\underline{w}(x), \underline{w}(x)] dx \\ &\geq \lambda_0 \int_{\Omega} [\underline{w}(x), \underline{w}(x)] dx \\ &= \lambda_0 \|\nabla \nu\|_{L^2(\Omega)}^2 = \lambda_0 \|\nu\|_{W^{1,2}(\Omega)}^2 = \lambda_0 \|\nu\|_{H^1(\Omega)}^2. \end{aligned}$$

□

## 14.4.2 狄利克雷边界值问题 1

### 将狄利克雷边界值问题改写为变分问题

回顾一下第14.1节的狄利克雷边界值问题 (14.11)-(14.12a): 基于给定的  $f$  和  $g$ , 求解

$$\begin{aligned} (Lu)(x) &= f(x), \quad x \in \Omega, \\ \gamma_0^{\text{int}} u(x) &= g(x), \quad x \in \Gamma, \end{aligned} \quad (14.119)$$

对应弱形式下的流形空间

$$\begin{aligned} V_g &:= \left\{ \nu \in H^1(\Omega) : \gamma_0^{\text{int}} \nu(x) = g(x), \quad x \in \Gamma \right\}, \\ V_0 &= H_0^1(\Omega). \end{aligned} \quad (14.120)$$

根据第14.3介绍的知识，我们可以将狄利克雷边界值问题(14.119)-(14.120)改写为变分问题：

$$a(u, \nu) = \langle f, \nu \rangle_{\Omega}, \quad \forall \nu \in V_0, \quad (14.121)$$

其中双线性泛函  $a(u, \nu)$  由格林第一恒等式(14.113)定义。研究目标是寻找变分问题的解  $u \in V_g$ 。由此狄利克雷边界问题表现为流形  $V_g$  中的一个附属条件，因此又称为基本边界条件 (essential boundary conditions)。

### 解的存在性与唯一性

由上所述可以看出，狄利克雷边界值问题对应的变分问题，可以归入带限制条件的算子方程问题类型，可用相应的变分方法求解 (第(14.3.4)节)。借助 Theorem 14.17 和 Corollary 14.4，我们可以证明  $u \in V_g$  存在且唯一。

**Theorem 14.23** (狄利克雷边界值问题的弱形式解). 给定  $f \in H^{-1}(\Omega), g \in H^{\frac{1}{2}}(\Gamma)$ ，变分问题(14.119)存在唯一的解  $u \in H^1(\Omega)$ ，满足

$$\|u\|_{H^1(\Omega)} \leq \frac{1}{c_1^A} \|f\|_{H^{-1}(\Omega)} + \left(1 + \frac{c_2^A}{c_1^A}\right) c_{IT} \|g\|_{H^{\frac{1}{2}}(\Gamma)}. \quad (14.122)$$

证明. 与 Theorem 14.17 类似，假设解  $u$  由两部分相加而得， $u := u_0 + u_g$ ，其中  $u_g$  是  $\gamma_0^{\text{int}} u_g = g$  的解， $u_0$  是  $\gamma_1^{\text{int}}(u_0 + u_g) = f$  的解。显然， $u$  的唯一存在，可由  $u_g$  和  $u_0$  的唯一存在所分别证得。

1. 证明  $u_g$  的唯一存在性。根据给定条件  $g \in H^{\frac{1}{2}}(\Gamma)$ ，应用逆迹定理 Theorem 14.11，可得存在唯一一个有界延拓  $u_g \in H^{\Omega}$ ，满足

$$\begin{aligned} \gamma_0^{\text{int}} u_g &= g, \\ \|u_g\|_{H^1(\Omega)} &\leq c_{IT} \|g\|_{H^{\frac{1}{2}}(\Gamma)}. \end{aligned}$$

2. 在求得唯一解  $u_g$  的基础上，继续求唯一解  $u_0$ 。变分问题(14.121)变为，寻找解  $u_0 \in V_0 = H_0^1(\Omega)$ ，使满足新的变分问题

$$a(u_0, \nu) = \langle f, \nu \rangle_{\Omega} - a(u_g, \nu), \quad \forall \nu \in V_0. \quad (14.123)$$

- (a) 由范数等价定理 (Theorem 14.3-14.4, (14.38)) 得， $H^1(\Omega)$  中的等价范为

$$\|u_0\|_{W^{1,2}(\Omega), \Gamma} = \left\{ \left[ \int_{\Gamma} \gamma_0^{\text{int}} u_0(x) ds_x \right]^2 + \|\nabla u_0\|_{L^2(\Omega)}^2 \right\}^{\frac{1}{2}},$$

即  $a(., .)$  有界。

- (b) 由 Lemma 14.15 (14.118) 可得

$$\begin{aligned} a(u_0, u_0) &\geq \lambda_0 \|u_0\|_{H^1(\Omega)} \\ &= \lambda_0 \|\nabla u_0\|_{W^{1,2}(\Omega), \Gamma}^2 \\ &\geq c_1^A \|u_0\|_{H^1(\Omega)}^2, \end{aligned} \quad (14.124)$$

即  $a(., .)$  是个  $V_0$ -椭圆算子。

(c)  $a(.,.)$  有界且  $V_0$ -椭圆, 满足拉克斯—密格拉蒙定理 (Theorem 14.14) 所需的前提条件, 根据该定理, 变分问题(14.123)有唯一解  $u_0 \in V_0$ , 满足

$$\begin{aligned} c_1^A \|u_0\|_{H^1(\Omega)}^2 &\leq a(u_0, u_0) \\ &= \langle f, u_0 \rangle_\Omega - a(u_g, u_0), \end{aligned}$$

代入 Lemma 14.10, 上式进一步变为

$$\begin{aligned} c_1^A \|u_0\|_{H^1(\Omega)}^2 &\leq \left( \|f\|_{H^{-1}(\Omega)} + c_2^A \|u_g\|_{H^1(\Omega)} \right) \|u_0\|_{H^1(\Omega)}, \\ \hookrightarrow \|u_0\|_{H^1(\Omega)} &\leq \frac{1}{c_1^A} \|f\|_{H^{-1}(\Omega)} + \frac{c_2^A}{c_1^A} \|u_g\|_{H^1(\Omega)}. \end{aligned}$$

3. 在此基础上我们有

$$\begin{aligned} \|u_0\|_{H^1(\Omega)} + \|u_g\|_{H^1(\Omega)} &\leq \frac{1}{c_1^A} \|f\|_{H^{-1}(\Omega)} + \left( 1 + \frac{c_2^A}{c_1^A} \right) \|u_g\|_{H^1(\Omega)}, \\ \hookrightarrow \|u\|_{H^1(\Omega)} &\leq \frac{1}{c_1^A} \|f\|_{H^{-1}(\Omega)} + \left( 1 + \frac{c_2^A}{c_1^A} \right) c_{IT} \|g\|_{H^{\frac{1}{2}}(\Gamma)}. \end{aligned} \quad (14.125)$$

□

### 共形导数

变分问题(14.121)的唯一解  $u \in V_g$ , 常常又称为狄利克雷边界值问题(14.119)的弱形式解。在此基础上, 对于  $f \in \tilde{H}^{-1}(\Omega)$ , 可通过构建变分问题进一步求解共形导数  $\gamma_1^{\text{int}} u \in H^{-\frac{1}{2}}(\Gamma)$

$$\langle \gamma_1^{\text{int}} u, z \rangle_\Gamma = a(u, \varepsilon_z) - \langle f, \varepsilon_z \rangle_\Omega, \quad \forall z \in H^{\frac{1}{2}}(\Gamma), \quad (14.126)$$

其中  $\varepsilon_z$  是逆迹定理 Theorem 14.11 所定义的有界延拓算子。(14.126)的唯一可解条件由 Theorem 14.16 给出, 由此我们可以假定如下稳定性条件

$$\|w\|_{H^{-\frac{1}{2}}(\Gamma)} = \sup_{0 \neq z \in H^{\frac{1}{2}}(\Gamma)} \frac{\langle w, z \rangle_\Gamma}{\|z\|_{H^{\frac{1}{2}}(\Gamma)}}, \quad \forall w \in H^{-\frac{1}{2}}(\Gamma). \quad (14.127)$$

在此基础上我们有共形导数的解

**Lemma 14.16** (共形导数的解). 设给定  $f \in \tilde{H}^{-1}(\Omega), g \in H^{\frac{1}{2}}(\Gamma)$ , 我们有  $u \in H^1(\Gamma)$  是狄利克雷边界值问题(14.121)的唯一解。

那么相应的共形导数  $\gamma_1^{\text{int}} u \in H^{-\frac{1}{2}}(\Gamma)$  满足

$$\|\gamma_1^{\text{int}} u\|_{H^{-\frac{1}{2}}(\Gamma)} \leq c_{IT} \left\{ \|f\|_{\tilde{H}^{-1}(\Omega)} + c_2^A |u|_{H^1(\Omega)} \right\}. \quad (14.128)$$

证明. 稳定性条件(14.127)⇒

$$\|\gamma_1^{\text{int}} u\|_{H^{-\frac{1}{2}}(\Gamma)} = \sup_{0 \neq z \in H^{\frac{1}{2}}(\Gamma)} \frac{\langle \gamma_1^{\text{int}} u, z \rangle_\Gamma}{\|z\|_{H^{\frac{1}{2}}(\Gamma)}},$$

变分问题(14.126)、Lemma (14.10)⇒

$$\begin{aligned} \|\gamma_1^{\text{int}} u\|_{H^{-\frac{1}{2}}(\Gamma)} &= \sup_{0 \neq z \in H^{\frac{1}{2}}(\Gamma)} \frac{|a(u, \varepsilon_z) - \langle f, \varepsilon_z \rangle_\Omega|_\Gamma}{\|z\|_{H^{\frac{1}{2}}(\Gamma)}}, \\ &\leq \left\{ c_2^A |u|_{H^1(\Omega)} + \|f\|_{\tilde{H}^{-1}(\Omega)} \right\} \underbrace{\sup_{0 \neq z \in H^{\frac{1}{2}}(\Gamma)} \frac{\|\varepsilon_z\|_{H^1(\Omega)}}{\|z\|_{H^{\frac{1}{2}}(\Gamma)}}}_{=: \mathcal{A}} \end{aligned}$$

由逆迹定理 Theorem 14.11 可得

$$\mathcal{A} \leq c_{\text{IT}},$$

因此

$$\|\gamma_1^{\text{int}} u\|_{H^{-\frac{1}{2}}(\Gamma)} \leq c_{\text{IT}} \left\{ \|f\|_{\tilde{H}^{-1}(\Omega)} + c_2^A |u|_{H^1(\Omega)} \right\}.$$

□

### 带有齐次偏微分方程的狄利克雷边界值问题弱解

来考虑一类特殊的狄利克雷边界值问题，即含有齐次偏微分方程  $f \equiv 0$  的情况，其解  $u$  对于边界积分算子的分析具有重要意义。

**Corollary 14.5** (带有齐次偏微分方程的狄利克雷边界值问题弱解). 设  $u \in H^1(\Omega)$  是以下带有齐次偏微分方程的狄利克雷边界值问题的弱形式解

$$(Lu)(x) = 0, \quad x \in \Omega, \\ \gamma_0^{\text{int}} u(x) = g(x), \quad x \in \Gamma,$$

其中  $L$  是一个一致椭圆的二阶偏微分算子。

那么我们有

$$a(u, u) \geq c \|\gamma_1^{\text{int}} u\|_{H^{-\frac{1}{2}}(\Gamma)}^2. \quad (14.129)$$

证明. 1. 当  $f \equiv 0$  时，共形导数算子和逆迹的关系，由 Lemma 14.16 的 (14.128) 变为

$$\|\gamma_1^{\text{int}} u\|_{H^{-\frac{1}{2}}(\Gamma)}^2 \leq [c_{\text{IT}} c_2^A]^2 |u|_{H^1(\Omega)}^2.$$

2. 由双线性算子的半椭圆特性 Lemma 14.15 (14.118) 可得

$$\lambda_0 |u|_{H^1(\Omega)}^2 \leq a(u, u).$$

□

### 狄利克雷边界值问题的强解简述

当  $\Omega$  是利普希茨域时，基于给定的  $f$  和  $g$ ，我们可以引入更严格的假设条件，构建强形式的正则解如  $u$ 、 $\gamma_1^{\text{int}} u$  等。

**Theorem 14.24.** 设  $\Omega \subset \mathbb{R}^d$  是个有界的利普希茨域，边界为  $\Gamma = \partial\Omega$ 。设  $u \in H^1(\Omega)$  是狄利克雷边界值问题的弱形式解

$$(Lu)(x) = f(x), \quad x \in \Omega, \\ \gamma_0^{\text{int}} u(x) = g(x), \quad x \in \Gamma.$$

如果给定的  $f$  和  $g$  满足  $f \in L^2(\Omega)$ ,  $g \in H^1(\Gamma)$ ，那么我们有  $u \in H^{\frac{3}{2}}(\Omega)$ ,  $\gamma_1^{\text{int}} u \in L^2(\Gamma)$ ，并且

$$\|u\|_{H^{\frac{3}{2}}(\Omega)} \leq c_1 \left\{ \|f\|_{L^2(\Omega)} + H^1(\Gamma) \right\}, \\ \|\gamma_1^{\text{int}} u\|_{L^2(\Gamma)} \leq c_2 \left\{ \|f\|_{L^2(\Omega)} + H^1(\Gamma) \right\}.$$

我们甚至可以使假定条件更加严格，如  $\Gamma = \partial\Omega$  是平滑或分段平滑的边界， $\Omega$  是凹的， $f \in L^2(\Omega)$  等。若  $g = \gamma_0^{\text{int}} u$  是方程解  $u_g \in H^2(\Omega)$  的迹，则我们有  $u \in H^2(\Omega)$ 。更多强形式解的讨论，可参见 Demkowicz (2006); Demkowicz et al. (2007)。

### 14.4.3 狄利克雷边界值问题 2

如第14.3.5节所述，狄利克雷边界值问题(14.119)也可以改写为鞍点变分问题，即混合算子方程，共形导数对应拉格朗日乘子 (Babuška, 1973; Bramble, 1981)。

从格林第一恒等式(14.113)入手，设拉格朗日乘子

$$\lambda := \gamma_1^{\text{int}} u \in H^{-\frac{1}{2}}(\Gamma),$$

进而鞍点变分问题表示为，寻找解  $(u, \lambda) \in H^1(\Omega) \times H^{-\frac{1}{2}}(\Gamma)$ ，使得满足

$$\begin{aligned} a(u, \nu) - b(\nu, \lambda) &= \langle f, \nu \rangle_{\Omega} \quad \forall \nu \in H^1(\Omega), \\ b(u, \mu) &= \langle g, \mu \rangle_{\Gamma} \quad \forall \mu \in H^{-\frac{1}{2}}(\Gamma), \end{aligned} \tag{14.130}$$

其中定义了一个新的双线性泛函算子

$$b(\nu, \mu) := \langle \gamma_0^{\text{int}} \nu, \mu \rangle_{\Gamma}, \quad (\nu, \mu) \in H^1(\Gamma) \times H^{-\frac{1}{2}}(\Gamma).$$

#### 解的唯一存在性

鞍点变分形式的狄利克雷边界值问题(14.130)，解的存在性和唯一性，可由 Theorem 14.19 证得。使用该定理之前，需要确保两个前提条件得到满足。一是双线性泛函  $a(., .)$  的椭圆特性，二是解的稳定性条件。

1. 椭圆性。类似于式(14.124)，由 Lemma 14.15 (14.118) 可得

$$a(., .) \geq c_1^A \| \cdot \|_{H^1(\Omega)}^2,$$

此外由于

$$\ker B := \left\{ \nu \in H^1(\Omega) : \langle \gamma_0^{\text{int}} \nu, \mu \rangle_{\Gamma} = 0, \quad \forall \mu \in H^{-\frac{1}{2}}(\Gamma) \right\} = H_0^1(\Omega),$$

我们因此有， $a(., .)$  是一个  $\ker B$ -椭圆 (或  $H_0^1$ -椭圆) 的双线性形。(通常来说，我们需要求得一个扩展双线性形  $\tilde{a}(., .)$ ，使得满足  $H^1(\Omega)$ -椭圆性质，相关讨论见第14.4.3节。)

2. 解的稳定性条件可以表示为

$$c_S \|\mu\|_{H^{-\frac{1}{2}}(\Gamma)} \leq \sup_{0 \neq \nu \in H^1(\Omega)} \frac{\langle \gamma_0^{\text{int}} \nu, \mu \rangle_{\Gamma}}{\|\nu\|_{H^1(\Omega)}}, \quad \forall \mu \in H^{-\frac{1}{2}}(\Gamma), \tag{14.131}$$

其证明见可见 Lemma 14.17。

3. 应用定理 Theorem 14.19，求得鞍点变分问题的唯一解。

**Lemma 14.17** (鞍点变分形式狄利克雷边界值问题解的稳定条件). 稳定条件(14.131)成立。

证明. 已知给定的任一  $\mu \in H^{-\frac{1}{2}}(\Gamma)$ 。由里兹表现定理 (Theorem 14.13) 可得，存在唯一的一个  $u_{\mu} \in H^{\frac{1}{2}}(\Gamma)$ ，满足

$$\begin{aligned} \langle u_{\mu}, \nu \rangle_{H^{\frac{1}{2}}(\Gamma)} &= \langle \mu, \nu \rangle_{\Gamma} \quad \forall \nu \in H^{\frac{1}{2}}(\Gamma), \\ \|u_{\mu}\|_{H^{\frac{1}{2}}(\Gamma)} &= \|\mu\|_{H^{-\frac{1}{2}}(\Gamma)}. \end{aligned}$$

由逆迹定理 Theorem 14.11 可得，存在一个延拓算子  $\varepsilon u_{\mu} \in H^1(\Omega)$ ，满足

$$\|\varepsilon u_{\mu}\|_{H^1(\Omega)} \leq c_{\text{IT}} \|u_{\mu}\|_{H^{\frac{1}{2}}(\Gamma)}.$$

那么, 对于  $\nu = \varepsilon u_\mu \in H^1(\Omega)$  我们有

$$\begin{aligned} \frac{\langle \nu, \mu \rangle_\Gamma}{\|\nu\|_{H^1(\Omega)}} &= \frac{\langle u_\mu, \mu \rangle_\Gamma}{\|\varepsilon u_\mu\|_{H^1(\Omega)}} = \frac{\langle u_\mu, u_\mu \rangle_{H^{\frac{1}{2}}(\Gamma)}}{\|\varepsilon u_\mu\|_{H^1(\Omega)}} \\ &\geq \frac{1}{c_{IT}} \|u_\mu\|_{H^{\frac{1}{2}}(\Gamma)} = \frac{1}{c_{IT}} \|u\|_{H^{-\frac{1}{2}}(\Gamma)} \end{aligned}$$

$\therefore$  稳定性条件(14.131)成立。  $\square$

### 调整鞍点变分问题

需要注意的是, 在鞍点变分问题(14.119)中的双线性形式算子  $a(.,.)$  是  $H_0^1(\Omega)$ -椭圆的。我们常常需要将它扩展为一个  $H^1(\Omega)$ -椭圆的算子  $\tilde{a}(.,.)$ , 对应新的调整鞍点变分问题。调整思路如下:

已知, 用拉格朗日乘子  $\lambda := \gamma_1^{\text{int}} u \in H^{-\frac{1}{2}}(\Gamma)$  来描述问题解  $u$  的共形导数, 那么利用格林第二恒等式(14.8), 可得狄利克雷边界值问题的正交条件(14.15)

$$\begin{aligned} &\int_{\Omega} f(x) dx + \int_{\Gamma} \lambda(x) ds_x = 0, \\ &\hookrightarrow \int_{\Gamma} \lambda(x) ds_x \int_{\Gamma} \mu(x) ds_x = - \int_{\Omega} f(x) dx \int_{\Gamma} \mu(x) ds_x, \quad \forall \mu \in H^{-\frac{1}{2}}(\Gamma). \end{aligned} \tag{14.132}$$

另一方面, 根据狄利克雷边界条件有

$$\begin{aligned} &\gamma_0^{\text{int}} u = g, \\ &\hookrightarrow \int_{\Gamma} \gamma_0^{\text{int}} u(x) ds_x = \int_{\Gamma} g(x) ds_x, \\ &\hookrightarrow \int_{\Gamma} \gamma_0^{\text{int}} u(x) ds_x \int_{\Gamma} \gamma_0^{\text{int}} \nu(x) ds_x = \int_{\Gamma} g(x) ds_x \int_{\Gamma} \gamma_0^{\text{int}} \nu(x) ds_x, \quad \forall \nu \in H^1(\Omega). \end{aligned} \tag{14.133}$$

将(14.132)、(14.133)代入(14.130), 得到调整鞍点变分问题: 寻找解  $(u, \lambda) \in H^1(\Omega) \times H^{-\frac{1}{2}}(\Gamma)$ , 使得  $\forall (\nu, \mu) \in H^1(\Omega) \times H^{-\frac{1}{2}}(\Gamma)$  均满足

$$\begin{aligned} &\underbrace{\int_{\Gamma} \gamma_0^{\text{int}} u(x) ds_x \int_{\Gamma} \gamma_0^{\text{int}} \nu(x) ds_x + a(u, \nu) - b(\nu, \lambda)}_{=: \tilde{a}(u, \nu)} = \langle f, \nu \rangle_{\Omega} + \int_{\Gamma} g(x) ds_x \int_{\Gamma} \gamma_0^{\text{int}} \nu(x) ds_x, \\ &b(u, \mu) + \int_{\Gamma} \lambda(x) ds_x \int_{\Gamma} \mu(x) ds_x = \langle g, \mu \rangle_{\Gamma} - \int_{\Omega} f(x) dx \int_{\Gamma} \mu(x) ds_x. \end{aligned} \tag{14.134}$$

下面的问题就是, 调整鞍点变分问题(14.134)的解是否存在, 是否唯一, 以及是否与原鞍点变分问题(14.130)的解一致。换句话说, 两个鞍点变分问题是否等价。

**Theorem 14.25** (变分问题等价). 调整鞍点变分问题(14.134)有唯一的解  $(u, \lambda) \in H^1(\Omega) \times H^{-\frac{1}{2}}(\Gamma)$ , 并且与鞍点变分问题(14.130)的解一致。即, 两个变分问题等价。

证明. 1. 证明扩展双线性形  $\tilde{a}(u, \nu)$  有界。

$$a(u, \nu) \text{ 有界} \Rightarrow$$

$$\tilde{a}(u, \nu) := \int_{\gamma} \gamma_0^{\text{int}} u(x) ds_x \int_{\gamma} \gamma_0^{\text{int}} \nu(x) ds_x + a(u, \nu), \quad \forall u, \nu \in H^1(\Omega)$$

有界。

2. 由  $a(.,.)$  的半椭圆性质 (Lemma 14.15) 和(14.38)有

$$\begin{aligned}\tilde{a}(\nu, \nu) &= \left[ \int_{\Gamma} \gamma_0^{\text{int}} \nu(x) ds_x \right]^2 + a(\nu, \nu) \\ &\geq \min\{1, \lambda_0\} \|\nu\|_{W^{1,2}(\Omega), \Gamma}^2 \\ &\geq c_1^{\tilde{A}} \|\nu\|_{H^1(\Omega)}^2, \quad \forall \nu \in H^1(\Omega),\end{aligned}$$

由此可得  $\tilde{a}(u, \nu)$  是  $H^1(\Omega)$ -椭圆。

3. 前提条件得到满足, 可通过 Theorem 14.19、Theorem 14.20 证得, 调整鞍点变分问题(14.134)有唯一解。

4. 对于  $(\nu, \mu) \equiv (1, 1)$  的特殊情况, (14.134)变为

$$\begin{aligned}|\Gamma| \int_{\Gamma} \gamma_0^{\text{int}} u(x) ds_x - \int_{\Gamma} \lambda(x) ds_x &= \int_{\Omega} f(x) dx + |\Gamma| \int_{\Gamma} g(x) ds_x, \\ \int_{\Gamma} \gamma_0^{\text{int}} u(x) ds_x + |\Gamma| \int_{\Gamma} \lambda(x) ds_x &= \int_{\Gamma} g(x) ds_x - |\Gamma| \int_{\Omega} f(x) dx.\end{aligned}$$

进而

$$(1 + |\Gamma|^2) \int_{\Gamma} \gamma_0^{\text{int}} u(x) ds_x = (1 + |\Gamma|^2) \int_{\Gamma} g(x) ds_x$$

即(14.133)。从而有

$$|\Gamma| \int_{\Gamma} \lambda(x) ds_x = -|\Gamma| \int_{\Gamma} f(x) ds_x,$$

即(14.132)。

因此可见,  $(u, \lambda)$  也是鞍点变分问题(14.130)的解的解; 两个问题等价。

□

#### 14.4.4 諾依曼边界值问题

##### 将諾依曼边界值问题改写为变分问题

回顾一下第14.1节的諾依曼边界值问题(14.11)-(14.12b): 基于给定的  $f$  和  $g$ , 求解

$$\begin{aligned}(Lu)(x) &= f(x), \quad x \in \Omega, \\ \gamma_1^{\text{int}} u(x) &= g(x), \quad x \in \Gamma.\end{aligned}\tag{14.135}$$

假定  $f$  和  $g$  满足可求解性条件 (14.17)

$$\int_{\Omega} f(x) dx + \int_{\Gamma} g(x) ds_x = 0.\tag{14.136}$$

基于前文的分析可见, 諾依曼边界值问题(14.135)的解  $u \in H^1(\Omega)$  将与某一个常数有关。为了将此常数予以确定, 可以在  $H^1(\Omega)$  中定义一个测试空间  $H_*^1(\Omega)$ , 测试方程  $\nu(x)$  用于规模调节。

$$H_*^1(\Omega) := \left\{ \nu \in H^1(\Omega) : \int_{\Omega} \nu(x) dx = 0 \right\}.\tag{14.137}$$

从而构建变分问题, 求解  $u \in H_*^1(\Omega)$  使得满足

$$a(u, \nu) = \langle f, \nu \rangle_{\Omega} + \langle g, \gamma_0^{\text{int}} \nu \rangle_{\Gamma}, \quad \forall \nu \in H_*^1(\Omega).\tag{14.138}$$

### 解的存在性与唯一性

**Theorem 14.26** (诺依曼边界值问题的变分法求解). 给定  $f \in \tilde{H}^{-1}(\Omega)$ ,  $g \in H^{-\frac{1}{2}}(\Gamma)$ , 满足可求解性条件(14.136)。

则诺依曼边界值的变分问题(14.138)存在唯一的解  $u \in H_*^1(\Omega)$ , 满足

$$\|u\|_{H^1(\Omega)} \leq \frac{1}{\tilde{c}_1^A} \left\{ \|f\|_{\tilde{H}^{-1}(\Omega)} + c_T \|g\|_{H^{-\frac{1}{2}}(\Gamma)} \right\}$$

证明. 1. 证  $a(.,.)$  有界且椭圆。

对于  $\nu \in H_*^1(\Omega)$ , 它在  $H^1(\Omega)$  中的等价范, 可由(14.38)求得

$$\|\nu\|_{W^{1,2}(\Omega),\Omega} = \left\{ \left[ \int_{\Omega} \nu(x) dx \right]^2 + \|\nabla \nu\|_{L^2(\Omega)}^2 \right\}^{\frac{1}{2}},$$

进而利用 Lemma 14.15 可得,

$$a(\nu, \nu) \geq \lambda_0 \|\nabla \nu\|_{W^{1,2}(\Omega),\Omega}^2 \geq \tilde{c}_1^A \|\nu\|_{H^1(\Omega)}^2, \quad \forall \nu \in H_*^1(\Omega), \quad (14.139)$$

因此可得双线性形式  $a(.,.)$  的  $H_*^1$ -椭圆特性。

2. 根据拉克斯—密格拉蒙定理 (Theorem 14.14), 证明变分问题(14.138)存在唯一解。

3. 将求得的变分问题唯一解  $u \in H_*^1(\Omega)$  代回椭圆条件 (14.139)

$$\begin{aligned} \tilde{c}_1^A \|u\|_{H^1(\Omega)}^2 &\leq a(u, u) = \langle f, u \rangle_{\Omega} + \langle g, \gamma_0^{\text{int}} u \rangle_{\Gamma} \\ &\leq \|f\|_{\tilde{H}^{-1}(\Omega)} \|u\|_{H^1(\Omega)} + \|g\|_{H^{-\frac{1}{2}}(\Gamma)} \|\gamma_0^{\text{int}} u\|_{H^{\frac{1}{2}}(\Gamma)}. \end{aligned}$$

由迹定理 Theorem 14.10 得, 上式变为

$$\tilde{c}_1^A \|u\|_{H^1(\Omega)} \leq \|f\|_{\tilde{H}^{-1}(\Omega)} + \|g\|_{H^{-\frac{1}{2}}(\Gamma)} c_T \|u\|_{H^1(\Gamma)}.$$

□

### 鞍点变分问题

类似地, 我们也可以构建一个与(14.138)等价的鞍点变分问题。此时, 用于规模调节的测试空间  $H_*^1(\Omega)$  以副条件 (side condition) 的情况出现。使用一个拉格朗日乘子, 寻找解  $(u, \lambda) \in H^1(\Omega) \times \mathbb{R}$ , 满足

$$\begin{aligned} a(u, \nu) + \lambda \int_{\Omega} \nu(x) dx &= \langle f, \nu \rangle_{\Omega} + \langle g, \gamma_0^{\text{int}} \nu \rangle_{\Gamma}, \\ \int_{\Omega} u(x) dx &= 0, \quad \forall \nu \in H^1(\Omega). \end{aligned} \quad (14.140)$$

**Theorem 14.27** (诺依曼边界值问题的鞍点变分法求解). 诺依曼边界值的鞍点变分问题(14.140)有唯一解  $(u, \lambda) \in H^1(\Omega) \times \mathbb{R}$ 。

证明. 1. 双线性形式  $b(.,.)$  有界

$$b(\nu, \mu) := \mu \int_{\Omega} \nu(x) dx, \quad \forall \nu \in H^1(\Omega), \mu \in \mathbb{R},$$

并且有  $\ker B = H_*^1(\Omega)$ 。

2. 进而由椭圆性(14.139)得, 双线性形  $a(\cdot, \cdot)$  是  $\ker B$ -椭圆的。

3. 证明满足稳定性条件

$$c_S |\mu| \leq \sup_{0 \neq \nu \in H^1(\Omega)} \frac{b(\nu, \mu)}{\|\nu\|_{H^1(\Omega)}}, \quad \forall \mu \in \mathbb{R}. \quad (14.141)$$

对于任一给定的  $\mu \in \mathbb{R}$ , 定义  $\nu^* := \nu \in H^1(\Omega)$ , 可以证得(14.141), 其中  $c_S = |\Omega|^{-\frac{1}{2}}$ 。

4. 根据定理 Theorem 14.19 可求得, 诺依曼边界值的鞍点变分问题存在唯一解  $(u, \lambda) \in H^1(\Omega) \times \mathbb{R}$ 。

5. 对于测试方程  $\nu \equiv 1$  的特殊情况, 根据诺依曼边界值问题的可求解条件(14.136)我们有拉格朗日乘子的值

$$\lambda = 0.$$

□

### 调整鞍点变分问题

如前所述, 鞍点变分问题(14.140)也可以改写为一个新的鞍点变分问题: 基于给定的任一  $f \in \tilde{H}^{-1}(\Omega), g \in H^{-\frac{1}{2}}(\Gamma)$ , 寻找  $(u, \lambda) \in H^1(\Omega) \times \mathbb{R}$ , 使满足

$$\begin{aligned} a(u, \nu) + \lambda \int_{\Omega} \nu(x) dx &= \langle f, \nu \rangle_{\Omega} + \langle g, \gamma_0^{\text{int}} \nu \rangle_{\Gamma}, \\ \int_{\Gamma} u(x) dx - \lambda &= 0, \quad \forall \nu \in H^1(\Omega). \end{aligned} \quad (14.142)$$

利用第二行等式求得拉格朗日乘子  $\lambda \in \mathbb{R}$  代入第一行, 我们得到一个新的调整变分问题, 寻找  $u \in H^1(\Omega)$  使满足

$$a(u, \nu) + \int_{\Omega} u(x) dx \int_{\Omega} \nu(x) dx = \langle f, \nu \rangle_{\Omega} + \langle g, \gamma_0^{\text{int}} \nu \rangle_{\Gamma}, \quad \forall \nu \in H^1(\Omega). \quad (14.143)$$

**Theorem 14.28** (诺依曼边界值的调整变分问题解). 基于给定的任一  $f \in \tilde{H}^{-1}(\Omega), g \in H^{-\frac{1}{2}}(\Gamma)$ , 调整变分问题(14.143)有唯一的解  $u \in H^1(\Omega)$ 。

如果给定的  $f \in \tilde{H}^{-1}(\Omega), g \in H^{-\frac{1}{2}}(\Gamma)$  满足可求解条件(14.136), 那么调整变分问题(14.143)的解  $u \in H_*^1(\Omega)$ ; 换句话说, 调整变分问题(14.143)和变分问题(14.140)等价。

证明. 1. 由(14.143)可见, 调整双线性形  $\tilde{a}(\cdot, \cdot)$  写为

$$\tilde{a}(u, \nu) := a(u, \nu) + \int_{\Omega} u(x) dx \int_{\Omega} \nu(x) dx.$$

由  $a(\cdot, \cdot)$  的半椭圆属性 Lemma 14.15 我们有

$$\begin{aligned} \tilde{a}(\nu, \nu) &\geq \lambda_0 \|\nabla \nu\|_{L^2(\Omega)}^2 + \left[ \int_{\Omega} \nu(x) dx \right]^2 \\ &\geq \min\{\lambda_0, 1\} \|\nu\|_{W^{1,2}(\Omega), \Omega}^2 \\ &\geq \hat{c}_1^A \|\nu\|_{W^{1,2}(\Omega), \Omega}^2, \quad \forall \nu \in H^1(\Omega), \end{aligned} \quad (14.144)$$

可得  $\tilde{a}(\cdot, \cdot)$  是  $H^1(\Omega)$ -椭圆且有界的。

2. 满足前提条件后, 可由拉克斯-密格拉蒙定理 (Theorem 14.14) 证得, 调整变分问题(14.143)有唯一解  $u \in H^1(\Omega)$ 。

3. 设测试方程  $\nu(x) \equiv 1$ 。调整变分问题(14.143)变为

$$\begin{aligned} |\Omega| \int_{\Omega} u(x) dx &= \langle f, 1 \rangle_{\Omega} + \langle g, 1 \rangle_{\Gamma}, \\ &= \int_{\Omega} f(x) dx + \int_{\Gamma} g(x) ds_x = 0, \end{aligned}$$

其中最后一个等式用到可求解性条件(14.136)。由此我们有  $u \in H_*^1(\Omega)$ , 也是变分问题(14.140)的解；换句话说，两个问题等价。

□

### 诺依曼边界值问题的通解

利用变分法求解诺依曼边界值问题(14.135), 所得到的  $u \in H_*^1(\Omega)$  是弱形式解。更一般意义上的通解  $\tilde{u} \in H^1(\Omega)$  可写为

$$\tilde{u} := u + \alpha,$$

其中  $\alpha \in \mathbb{R}$  是任意常数。

### 14.4.5 混合边界值问题

回顾一下第14.1节的混合边界值问题(14.11), (14.12a)-(14.12b)

$$\begin{aligned} (Lu)(x) &= f(x), \quad x \in \Omega, \\ \gamma_0^{\text{int}} u(x) &= g_D(x), \quad x \in \Gamma_D, \\ \gamma_1^{\text{int}} u(x) &= g_N(x), \quad x \in \Gamma_N, \end{aligned} \tag{14.145}$$

假定  $\Gamma = \bar{\Gamma}_D \cup \bar{\Gamma}_N$ 。对应地, 可由格林第一恒等式(14.6)建立变分问题: 求解  $u \in H^1(\Omega)$ ,  $\gamma_0^{\text{int}} u(x) = g_D(x), x \in \Gamma_D$ , 使满足

$$a(u, \nu) = \langle f, \nu \rangle_{\Omega} + \langle g_N, \gamma_0^{\text{int}} \nu \rangle_{\Gamma_N}, \quad \forall \nu \in H_0^1(\Omega, \Gamma_D), \tag{14.146}$$

其中

$$H_0^1(\Omega, \Gamma_D) := \left\{ \nu \in H^1(\Omega) : \gamma_0^{\text{int}} \nu(x) = 0, \quad x \in \Gamma_D \right\}.$$

混合边界值变分问题(14.146)的唯一解, 可由以下定理证明

**Theorem 14.29** (混合边界值变分问题的唯一解). 给定  $f \in \tilde{H}^{-1}(\Omega)$ ,  $g_D \in H^{\frac{1}{2}}(\Gamma_D)$ ,  $g_N \in H^{-\frac{1}{2}}(\Gamma_N)$ , 那么混合边界值变分问题(14.146)有唯一解  $u \in H^1(\Omega)$ , 满足

$$\|u\|_{H^1(\Omega)} \leq c \left[ \|f\|_{\tilde{H}^{-1}(\Omega)} + \|g_D\|_{H^{\frac{1}{2}}(\Gamma_D)} + \|g_N\|_{H^{-\frac{1}{2}}(\Gamma_N)} \right]. \tag{14.147}$$

证明. 1. 定义两个延拓算子。

(a) 给定  $g_D \in H^{\frac{1}{2}}(\Gamma_D)$ , 定义一个有界的延拓  $\tilde{g}_D \in H^{\frac{1}{2}}(\Gamma)$ , 满足

$$\|\tilde{g}_D\|_{H^{\frac{1}{2}}(\Gamma)} \leq c \|g_D\|_{H^{\frac{1}{2}}(\Gamma_D)}.$$

(b) 定义第二个延拓  $u_{\tilde{g}_D} \in H^{\frac{1}{2}}(\Omega)$ , 使得  $\gamma_0^{\text{int}} u_{\tilde{g}_D} = \tilde{g}_D$ , 并利用逆迹定理 Theorem 14.11 得

$$\|u_{\tilde{g}_D}\|_{H^1(\Omega)} \leq c_{\text{IT}} \|\tilde{g}_D\|_{H^{\frac{1}{2}}(\Gamma)}.$$

2. 基于(14.146), 构建新的变分问题。

$$a(u_0, \nu) = \langle f, \nu \rangle_{\Omega} + \langle g_N, \gamma_0^{\text{int}} \nu \rangle_{\Gamma_N} - a(u_{\tilde{g}_D}, \nu), \quad \forall \nu \in H_0^1(\Omega, \Gamma_D),$$

研究目标是求得唯一解  $u_0 \in H_0^1(\Omega, \Gamma_D)$ 。

根据(14.38)可定义  $H^1(\Omega)$  中的等价范

$$\|\nu\|_{W^{1,2}(\Omega), \Gamma_D} := \left\{ \left[ \int_{\Gamma_D} \gamma_0^{\text{int}} \nu(x) ds_x \right]^2 + \|\nabla \nu\|_{L^2(\Omega)}^2 \right\}^{\frac{1}{2}}.$$

由双线性算子  $a(.,.)$  的半椭圆特性 (Lemma 14.15) 可得

$$\begin{aligned} a(\nu, \nu) &\geq \lambda_0 \|\nabla \nu\|_{L^2(\Omega)}^2 \\ &= \lambda_0 |\nu|_{W^{1,2}(\Omega), \Gamma_D}^2 \\ &\geq c_1^A \|\nu\|_{W^{1,2}(\Omega)}^2. \end{aligned}$$

有界, 椭圆, 假设条件满足。由拉克斯一密格拉蒙定理 (Theorem 14.14) 可得, 变分问题(14.146)有唯一解  $u_0 \in H_0^1(\Omega, \Gamma_D)$ 。

3. 将求得的唯一解  $u_0$  分别代回半椭圆 (Lemma 14.15) 和变分问题, 可得

$$\begin{aligned} c_1^A \|u_0\|_{H^1(\Omega)}^2 &\leq a(u_0, u_0) \\ &= \langle f, u_0 \rangle_{\Omega} + \langle g_N, \gamma_0^{\text{int}} u_0 \rangle_{\Gamma_N} - a(u_{\tilde{g}_D}, u_0) \\ &= [\|f\|_{\tilde{H}^{-1}(\Omega)} + c_2^A \|u_{\tilde{g}_D}\|_{H^1(\Omega)}] \|u_0\|_{H^1(\Omega)} + \|g_N\|_{H^{-\frac{1}{2}}(\Gamma_N)} \|\gamma_0^{\text{int}} u_0\|_{\tilde{H}^{\frac{1}{2}}(\Gamma_N)}, \end{aligned}$$

证得(14.147)。

□

#### 14.4.6 罗宾边界值问题

回顾一下第14.1节的罗宾边界值问题: 基于给定的  $f$  和  $g$ , 求解

$$\begin{aligned} (Lu)(x) &= f(x), \quad x \in \Omega, \\ \gamma_1^{\text{int}} u(x) + \kappa(x) \gamma_0^{\text{int}} u(x) &= g(x), \quad x \in \Gamma. \end{aligned} \tag{14.148}$$

由格林第一恒等式(14.6)建立变分问题: 求解  $u \in H^1(\Omega)$ , 使满足

$$a(u, \nu) + \int_{\Gamma} \kappa(x) \gamma_0^{\text{int}} u(x) \gamma_0^{\text{int}} \nu(x) ds_x = \langle f, \nu \rangle_{\Omega} + \langle g, \gamma_0^{\text{int}} \nu \rangle_{\Gamma}, \quad \forall \nu \in H^1(\Omega). \tag{14.149}$$

**Theorem 14.30** (罗宾边界值变分问题的唯一解). 给定  $f \in \tilde{H}^{-1}(\Omega), g \in H^{-\frac{1}{2}}(\Gamma)$ , 假定  $\kappa(x) \geq \kappa_0 > 0, \forall x \in \Gamma$ 。那么罗宾边界值变分问题(14.149)存在唯一解  $u \in H^1(\Omega)$ , 满足

$$\|u\|_{H^1(\Omega)} \leq c \left[ \|f\|_{\tilde{H}^{-1}(\Omega)} + \|g\|_{H^{-\frac{1}{2}}(\Gamma)} \right]. \tag{14.150}$$

证明. 根据(14.38), 定义  $H^1(\Omega)$  中的等价范

$$\|\nu\|_{W^{1,2}(\Omega), \Gamma} := \left\{ \|\gamma_0^{\text{int}} \nu\|_{L^2(\Gamma)}^2 + \|\nabla \nu\|_{L^2(\Omega)}^2 \right\}^{\frac{1}{2}}$$

结合  $\kappa(x) \geq \kappa_0 > 0, \forall x \in \Gamma$ , 以及半椭圆特性 Lemma 14.15, 有

$$\begin{aligned} a(\nu, \nu) + \int_{\Gamma} \kappa(x) [\gamma_0^{\text{int}} \nu(x)]^2 ds_x &\geq \lambda_0 \|\nabla \nu\|_{L^2(\Omega)}^2 + \kappa_0 \|\gamma_0^{\text{int}} \nu\|_{L^2(\Gamma)}^2 \\ &\geq \min\{\lambda_0, \kappa_0\} \|\nu\|_{H^1(\Omega), \Gamma}^2 \\ &\geq c_1^A \|\nu\|_{H^1(\Omega)}^2. \end{aligned}$$

进而, 根据拉克斯—密格拉蒙定理 (Theorem 14.14), 可得罗宾边界值变分问题(14.149)的唯一解。由此可得(14.150)。  $\square$

## 14.5 拉普拉斯算子的基本解

### 14.5.1 基本解

回顾一下偏微分方程(14.11)

$$(Lu)(x) = f(x), \quad x \in \Omega \subset \mathbb{R}^d,$$

其中  $L$  是椭圆线性二阶偏微分算子(14.1)

$$(Lu)(x) = - \sum_{i,j=1}^d \frac{\partial}{\partial x_j} \left[ a_{ji}(x) \frac{\partial}{\partial x_i} u(x) \right].$$

对应的内部共形导数(14.7)

$$\gamma_1^{\text{int}} u(x) = \sum_{i,j=1}^d n_j(x) a_{ji}(x) \frac{\partial}{\partial x_i} u(x), \quad x \in \Gamma.$$

由格林第二恒等式(14.8)可得,  $\forall y \in \Omega \subset \mathbb{R}^d$  测试方程  $\nu(y)$  和对应  $y$  的偏微分方程解  $u(y)$  满足

$$\begin{aligned} a(u, \nu) = a(\nu, u) \Rightarrow \\ \int_{\Omega} (L\nu)(y) u(y) dy = \int_{\Gamma} \gamma_1^{\text{int}} u(y) \gamma_0^{\text{int}} \nu(y) ds_y - \int_{\Gamma} \gamma_1^{\text{int}} \nu(y) \gamma_0^{\text{int}} u(y) ds_y + \int_{\Gamma} f(y) \nu(y) dy. \end{aligned}$$

将测试方程定义为  $\nu(y) := U^*(x, y)$ ,  $u(x)$  和  $u(y)$  的关系可以表示如下

$$\int_{\Omega} (L_y U^*)(x, y) u(y) dy = u(x), \quad \forall x \in \Omega,$$

那么偏微分方程(14.11)的解  $u(x), x \in \Omega$ , 根据里兹表现定理14.13, 可以改写为如下表现方程

$$u(x) = \int_{\Omega} U^*(x, y) f(y) dy + \int_{\Gamma} U^*(x, y) \gamma_1^{\text{int}} u(y) ds_y - \int_{\Gamma} U^*(x, y) \gamma_0^{\text{int}} u(y) ds_y. \quad (14.151)$$

由此可见, 为了基于表现方程(14.151)求得任何偏微分形式方程(14.11)的解  $u(x)$ , 我们需要以下两方面的信息

- $x \in \Gamma$  对应的柯西数列  $[\gamma_0^{\text{int}} u(x), \gamma_1^{\text{int}} u(x)]$ ,
- 连接  $u(x)$  和  $u(y)$  线性二次偏微分算子  $(L_y U^*)(x, y)$ , 对应(14.151)。

对于前者，关键在于构建合适的边界积分式 (boundary integral equations) 以生成完整柯西数列 (complete cauchy data)，我们将在第14.6节讨论。

对于后者，从分布意义上来看，偏微分方程(14.11)的解  $u(x)$  可以理解为

$$u(x) = \int_{\Gamma} \delta_0(y - x) u(y) dy, \quad x \in \Omega,$$

那么定义

$$(L_y U^*)(x, y) = \delta_0(y - x), \quad x, y \in \mathbb{R}^d, \quad (14.152)$$

我们将(14.152)中的解  $U^*(x, y)$  称为基本解 (fundamental solution)。问题的关键就在于，利用(14.152)求得唯一的基本解  $U^*(x, y)$ ，进而带回到表现方程(14.151)中。

关于不同形式的偏微分算子 (尤其是存在分段常系数的偏微分算子) 是否存在基本解的问题，相关证明可见如Hörmander (1983a,b, 1994a,b)，我们不做过多讨论。对经济学应用研究来说，我们更多关注当偏微分算子是拉普拉斯算子 (Laplace operator)时对应的基本解。

### 14.5.2 拉普拉斯算子

拉普拉斯算子

$$(Lu)(x) := -\Delta u(x), \quad x \in \mathbb{R}^d, d = 2, 3.$$

对应的基本解  $U^*(x, y)$  就是下述偏微分方程的分布解

$$-\Delta_y U^*(x, y) = \delta_0(y - x), \quad x, y \in \mathbb{R}^d.$$

由于拉普拉斯算子与旋转 (rotation) 和平移 (translation) 无关，可以定义  $z := y - x, U^*(x, y) = \nu(z)$ 。则 (14.152) 变为

$$-\Delta \nu(z) = \delta_0(z), \quad z \in \mathbb{R}^d, \quad (14.153)$$

我们的任务是根据(14.153)求得基本解。

由 Definition 14.32 的傅里叶变换(14.50)和(14.52b)我们有

$$\begin{aligned} |\xi|^2 \widehat{\nu}(\xi) &= (2\pi)^{-\frac{d}{2}}, \\ \hookrightarrow \widehat{\nu}(\xi) &= (2\pi)^{-\frac{d}{2}} |\xi|^{-\frac{1}{2}} \in \mathcal{S}'(\mathbb{R}^d), \end{aligned}$$

其中  $\mathcal{S}'(\mathbb{R}^d)$  表示缓增分布空间(Definition 14.31)。

可见，一个缓增空间中的  $\nu(\xi) \in \mathcal{S}'(\mathbb{R}^d)$ ，和它所对应的傅里叶变换  $\widehat{\nu}(\xi)$  的关系为

$$\langle \widehat{\nu}, \varphi \rangle_{L^2(\mathbb{R}^d)} = \langle \nu, \widehat{\varphi} \rangle_{L^2(\mathbb{R}^d)}, \quad \forall \varphi \in \mathcal{S}(\mathbb{R}^d).$$

已知

$$\varphi(\xi) = (2\pi)^{-\frac{d}{2}} \int_{\mathbb{R}^d} \exp[i\langle z, \xi \rangle] \widehat{\varphi}(z) dz,$$

那么

$$\langle \widehat{\nu}, \varphi \rangle_{L^2(\mathbb{R}^d)} = (2\pi)^{-d} \int_{\mathbb{R}^d} |\xi|^{-\frac{1}{2}} \int_{\mathbb{R}^d} \exp[i\langle z, \xi \rangle] \widehat{\varphi}(z) dz d\xi.$$

上式的问题在于，积分  $\int_{\mathbb{R}^d} |\xi|^{-\frac{1}{2}} d\xi$  不存在，因此无法将两个求积操作合并。一个解决方案是，利用

$$\Delta_z \exp(i\langle z, \xi \rangle) = -|\xi|^2 \exp(i\langle z, \xi \rangle),$$

首先根据  $|\xi|$  的值作分步骤积分，进而调整两个积分操作的顺序，随后重复分步骤积分操作

$$\begin{aligned}
 & \langle \nu, \widehat{\varphi} \rangle_{L^2(\mathbb{R}^d)} \\
 &= \langle \widehat{\nu}, \varphi \rangle_{L^2(\mathbb{R}^d)} = (2\pi)^{-d} \int_{\mathbb{R}^d} |\xi|^{-\frac{1}{2}} \int_{\mathbb{R}^d} \exp[i\langle z, \xi \rangle] \widehat{\varphi}(z) dz d\xi \\
 &= (2\pi)^{-d} \int_{|\xi| \leq 1} |\xi|^{-\frac{1}{2}} \int_{\mathbb{R}^d} \exp[i\langle z, \xi \rangle] \widehat{\varphi}(z) dz d\xi + (2\pi)^{-d} \int_{|\xi| > 1} |\xi|^{-\frac{1}{2}} \int_{\mathbb{R}^d} \left[ -\Delta_z \frac{\exp(i\langle z, \xi \rangle)}{|\xi|^2} \right] \widehat{\varphi}(z) dz d\xi \\
 &= (2\pi)^{-d} \int_{|\xi| \leq 1} |\xi|^{-\frac{1}{2}} \int_{\mathbb{R}^d} \exp[i\langle z, \xi \rangle] \widehat{\varphi}(z) dz d\xi + (2\pi)^{-d} \int_{|\xi| > 1} |\xi|^{-\frac{1}{4}} \int_{\mathbb{R}^d} \exp(i\langle z, \xi \rangle) [-\Delta_z \widehat{\varphi}(z)] dz d\xi \\
 &= (2\pi)^{-d} \int_{\mathbb{R}^d} \widehat{\varphi}(z) \int_{|\xi| \leq 1} \frac{\exp[i\langle z, \xi \rangle]}{|\xi|^2} d\xi dz + (2\pi)^{-d} \int_{\mathbb{R}^d} [-\Delta_z \widehat{\varphi}(z)] \int_{|\xi| > 1} \frac{\exp(i\langle z, \xi \rangle)}{|\xi|^4} d\xi dz \\
 &= \int_{\mathbb{R}^d} \widehat{\varphi}(z) (2\pi)^{-d} \left[ \int_{|\xi| \leq 1} \frac{\exp[i\langle z, \xi \rangle]}{|\xi|^2} d\xi - \Delta_z \int_{|\xi| > 1} \frac{\exp[i\langle z, \xi \rangle]}{|\xi|^4} d\xi \right] dz.
 \end{aligned} \tag{14.154}$$

### 14.5.3 三维空间的基本解

对于三维空间  $d = 3$  的情况，我们可以建立三维坐标系

$$\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} = \begin{pmatrix} r \cos \varphi \sin \theta \\ r \sin \varphi \sin \theta \\ r \cos \theta \end{pmatrix}, \quad r \in (0, \infty), \varphi \in (0, 2\pi), \theta \in (0, \pi).$$

将傅里叶变换的旋转对称 (Lemma 14.2) 代回(14.154)有

$$\begin{aligned}
 \nu(z) &= \nu(|z|) = (2\pi)^{-3} \left[ \underbrace{\int_{|\xi| \leq 1} \frac{\exp(i\langle z, \xi \rangle)}{|\xi|^2} d\xi}_{=: \mathcal{A}} - \underbrace{\Delta_z \int_{|\xi| > 1} \frac{\exp(i\langle z, \xi \rangle)}{|\xi|^4} d\xi}_{=: \mathcal{B}} \right], \\
 \mathcal{A} &= \int_0^{2\pi} \int_0^\pi \int_0^1 \exp[i|z|r \cos \theta] \sin \theta dr d\theta d\varphi, \\
 \mathcal{B} &= \Delta_z \int_0^{2\pi} \int_0^\pi \int_1^\infty \frac{\exp[i|z|r \cos \theta]}{r^2} \sin \theta dr d\theta d\varphi, \\
 \hookrightarrow \nu(z) &= \nu(|z|) = (2\pi)^{-2} \left[ \int_0^\pi \int_0^1 \exp[i|z|r \cos \theta] \sin \theta dr d\theta - \Delta_z \int_0^\pi \int_1^\infty \frac{\exp[i|z|r \cos \theta]}{r^2} \sin \theta dr d\theta \right].
 \end{aligned}$$

设  $\iota := \cos \theta \in (-1, 1)$ ，那么

$$\begin{aligned}
 & \int_0^\pi \exp[i|z|r \cos \theta] \sin \theta d\theta \\
 &= \int_{-1}^1 \exp[i|z|r\iota] d\iota \\
 &= \frac{\exp[i|z|r] - \exp[-i|z|r]}{i|z|r} \\
 &= \frac{2}{|z|r} \sin |z|r \\
 \hookrightarrow \nu(z) &= (2\pi)^{-2} \left[ \underbrace{\int_0^1 \frac{\sin |z|r}{|z|r} dr}_{=: \mathcal{C}} - \underbrace{\Delta_z \int_1^\infty \frac{\sin |z|r}{|z|r^3} dr}_{=: \mathcal{D}} \right]
 \end{aligned}$$

先来看  $\mathcal{C}$ 。设  $\zeta := |z|r$  我们有

$$\begin{aligned}\mathcal{C} &= \int_0^1 \frac{\sin |z|r}{|z|r} dr \\ &= |z|^{-1} \int_0^{|z|} \frac{\sin \zeta}{\zeta} d\zeta \\ &= \frac{\text{Si}(|z|)}{|z|},\end{aligned}$$

其中 Si 表示三角积分 (trigonometric integral) 中正弦积分 (sine integral) 的一种, 定义为

$$\text{Si}(x) := \int_0^x \frac{\sin t}{t} dt,$$

以及

$$\text{sinc}(x) := \begin{cases} \frac{\sin x}{x} \\ \frac{\sin \pi x}{\pi x} \end{cases}$$

分别称非标准化 sinc 方程 (unnormalized sinc function) 和标准化 sinc 方程 (normalized sinc function)。

再来看  $\mathcal{D}$ 。由于

$$\begin{aligned}\int \frac{\sin ax}{x^3} dx &= -\frac{1}{2} \frac{\sin ax}{x^2} + \frac{a}{2} \int \frac{\cos ax}{x^2} dx \\ &= -\frac{1}{2} \frac{\sin ax}{x^2} - \frac{a}{2} \frac{\cos ax}{x} - \frac{a^2}{2} \int \frac{\sin ax}{x} dx, \\ \hookrightarrow \mathcal{D} &= \int_1^\infty \frac{\sin |z|r}{|z|r^3} dr \\ &= \left[ -\frac{1}{2} \frac{\sin |z|r}{|z|r^2} - \frac{1}{2} \frac{\cos |z|r}{r} \right]_1^\infty - \frac{|z|}{2} \int_1^\infty \frac{\sin |z|r}{r} dr \\ &= \frac{1}{2} \frac{\sin |z|}{|z|} + \frac{1}{2} \cos |z| - \frac{|z|}{2} \left[ \frac{\pi}{2} - \text{Si}(|z|) \right].\end{aligned}$$

$$\begin{aligned}\hookrightarrow \nu(z) &= (2\pi)^{-2} [\mathcal{C} - \Delta_z \mathcal{D}] \\ &= (2\pi)^{-2} \left\{ \frac{\text{Si}(|z|)}{|z|} - \Delta_z \left[ \frac{1}{2} \frac{\sin |z|}{|z|} + \frac{1}{2} \cos |z| - \frac{\pi}{4} |z| + \frac{|z|}{2} \text{Si}(|z|) \right] \right\} \\ &= \frac{1}{8\pi} \Delta_z |z| + \underbrace{\frac{1}{2\pi^2} \left\{ \frac{\text{Si}(|z|)}{|z|} - \Delta_z \left[ \frac{1}{2} \frac{\sin |z|}{|z|} + \frac{1}{2} \cos |z| + \frac{1}{2} |z| \text{Si}(|z|) \right] \right\}}_{=0} \\ &= \frac{1}{4\pi} \frac{1}{|z|}.\end{aligned}$$

这样我们有三维空间中拉普拉斯算子的基本解为

$$U^*(x, y) = \nu(z) = \frac{1}{4\pi} \frac{1}{|x-y|}, \quad x, y \in \mathbb{R}^3. \quad (14.155)$$

#### 14.5.4 二维空间的基本解

对于二维空间  $d = 2$  的情况, 首先要对基本解的逆傅里叶变换作一定的正则化处理 (Vladimirov, 1971)。定义一个缓增空间中的分布  $\mathcal{P} \frac{1}{|x|^2} \in \mathcal{S}'(\mathbb{R}^2)$ , 满足

$$\langle \mathcal{P} \frac{1}{|x|^2}, \varphi(x) \rangle_{L^2(\mathbb{R}^2)} = \int_{x \in \mathbb{R}^2: |x| \leq 1} \frac{\varphi(x) - \varphi(0)}{|x|^2} dx + \int_{x \in \mathbb{R}^2: |x| \geq 1} \frac{\varphi(x)}{|x|^2} dx, \quad \forall \varphi \in \mathcal{S}(\mathbb{R}^2).$$

那么

$$\begin{aligned} 2\pi \langle \nu, \widehat{\varphi} \rangle_{L^2(\mathbb{R}^2)} &= \langle \mathcal{P} \frac{1}{|\xi|^2}, \varphi \rangle_{L^2(\mathbb{R}^2)} \\ &= \int_{\xi \in \mathbb{R}^2: |\xi| \leq 1} \frac{\varphi(\xi) - \varphi(0)}{|\xi|^2} d\xi + \int_{\xi \in \mathbb{R}^2: |\xi| \geq 1} \frac{\varphi(\xi)}{|\xi|^2} d\xi, \quad \forall \varphi \in \mathcal{S}(\mathbb{R}^2), \end{aligned}$$

其中

$$\begin{aligned} \varphi(\xi) &= \frac{1}{2\pi} \int_{\mathbb{R}^2} \exp(i\langle z, \xi \rangle) \widehat{\varphi}(z) dz, \\ \varphi(0) &= \frac{1}{2\pi} \int_{\mathbb{R}^2} \widehat{\varphi}(z) dz, \end{aligned}$$

$$\begin{aligned} &\hookrightarrow (2\pi)^2 \langle \nu, \widehat{\varphi} \rangle_{L^2(\mathbb{R}^2)} \\ &= \int_{\xi \in \mathbb{R}^2: |\xi| \leq 1} \frac{1}{|\xi|^2} \int_{\mathbb{R}^2} [\exp(i\langle z, \xi \rangle) - 1] \widehat{\varphi}(z) dz d\xi + \int_{\xi \in \mathbb{R}^2: |\xi| > 1} \frac{1}{|\xi|^2} \int_{\mathbb{R}^2} \exp(i\langle z, \xi \rangle) \widehat{\varphi}(z) dz d\xi. \end{aligned}$$

同样地，我们无法调整两个积分的先后顺序。采用类似于(14.154)的思路，根据  $|\xi|$  作分步骤积分，进而调整两个积分操作的顺序，然后重复分步骤积分运算，得

$$(2\pi)^2 \langle \nu, \widehat{\varphi} \rangle_{L^2(\mathbb{R}^2)} = \int_{\mathbb{R}^2} \widehat{\varphi}(z) \left[ \int_{\xi \in \mathbb{R}^2: |\xi| \leq 1} \frac{\exp[i\langle z, \xi \rangle] - 1}{|\xi|^2} d\xi + \int_{\xi \in \mathbb{R}^2: |\xi| > 1} \frac{\exp[i\langle z, \xi \rangle]}{|\xi|^2} d\xi \right] dz.$$

根据傅里叶变换的旋转对称 (Lemma 14.2)，由上式可得

$$\nu(z) = \nu(|z|) = (2\pi)^{-2} \left[ \int_{\xi \in \mathbb{R}^2: |\xi| \leq 1} \frac{\exp[i\langle z, \xi \rangle] - 1}{|\xi|^2} d\xi + \int_{\xi \in \mathbb{R}^2: |\xi| > 1} \frac{\exp[i\langle z, \xi \rangle]}{|\xi|^2} d\xi \right].$$

建立二维坐标系

$$\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \begin{pmatrix} r \cos \varphi \\ r \sin \varphi \end{pmatrix}, \quad r \in (0, \infty), \varphi = (0, 2\pi),$$

上式变为

$$\begin{aligned} \nu(z) = \nu(|z|) &= (2\pi)^{-2} \left\{ \int_0^1 \int_0^{2\pi} \frac{1}{r} [\exp(ir|z| \cos \varphi) - 1] d\varphi dr + \int_1^\infty \int_0^{2\pi} \frac{1}{r} \exp(ir|z| \cos \varphi) d\varphi dr \right\} \\ &= (2\pi)^{-1} \left\{ \int_0^1 \frac{1}{r} [J_0(r|z|) - 1] dr + \int_1^\infty \frac{1}{r} [J_0(r|z|) - 1] dr \right\}, \end{aligned}$$

其中我们用到了一阶贝塞尔方程 (Bessel function, Definition 14.34, 或参考 (Gradshteyn and Ryzhik, 2014, Sec. 8.411))

$$J_0(s) := (2\pi)^{-1} \int_0^{2\pi} \exp(is \cos \varphi) d\varphi.$$

定义  $r := \frac{s}{\varrho}$ , 上式进一步变为

$$\begin{aligned} \nu(z) &= \frac{1}{2\pi} \int_0^\varrho \frac{J_0(s) - 1}{s} ds + \frac{1}{2\pi} \int_\varrho^\infty \frac{J_0(s)}{s} ds \\ &= \frac{1}{2\pi} \int_0^1 \frac{J_0(s) - 1}{s} ds + \frac{1}{2\pi} \int_1^\infty \frac{J_0(s)}{s} ds + \frac{1}{2\pi} \int_\varrho^1 \frac{1}{s} ds \\ &= -\frac{1}{2\pi} \log |z| - \frac{c_0}{2\pi}, \end{aligned}$$

其中  $c_0$  是个常数

$$c_0 := \int_0^1 \frac{1 - J_0(s)}{s} ds - \int_1^\infty \frac{J_0(s)}{s} ds.$$

由于任何常数都满足齐次拉普拉斯方程的条件，在计算二维空间拉普拉斯方程的基本解时，我们可以忽略常数项。因此基本解变为

$$U^*(x, y) = \nu(z) = -\frac{1}{2\pi} \log |x - y|, \quad x, y \in \mathbb{R}^2. \quad (14.156)$$

### 14.5.5 基本解总结

小结：拉普拉斯算子在二维、三维空间中的基本解(14.155), (14.156)为

$$U^*(x, y) = \begin{cases} -\frac{1}{2\pi} \log |x - y| & d = 2, \\ \frac{1}{4\pi} \frac{1}{|x - y|} & d = 3. \end{cases} \quad (14.157)$$

对应地，对于给定的  $x \in \Omega$ , 下述形式偏微分方程

$$-\Delta u(x) = f(x), \quad x \in \Omega \subset \mathbb{R}^d$$

的解  $u(x)$ , 都以表现方程的形式出现

$$u(x) = \int_{\Omega} U^*(x, y) f(y) dy + \int_{\Gamma} U^*(x, y) \frac{\partial}{\partial n_y} u(y) ds_y - \int_{\Gamma} \frac{\partial}{\partial n_y} U^*(x, y) u(y) ds_y. \quad (14.158)$$

## 14.6 边界积分算子

回顾一下第14.5节提出的问题，我们的目标是求解  $d = 2, 3$  下的泊松方程 (Poisson equation)

$$-\Delta u(x) = f(x), \quad x \in \Omega \subset \mathbb{R}^d,$$

$f(x) = 0$  时的齐次形式泊松方程又称拉普拉斯方程 (Laplace function)。

泊松方程的解满足表现形式 (representation formula) (14.151)

$$u(x) = \int_{\Omega} U^*(x, y) f(y) dy + \int_{\Gamma} U^*(x, y) \gamma_1^{\text{int}} u(y) ds_y - \int_{\Gamma} U^*(x, y) \gamma_0^{\text{int}} u(y) ds_y, \quad (14.159)$$

为了求解(14.159)，首先要求得拉普拉斯算子的基本解  $U^*(x, y)$ , 见(14.157)给出

$$U^*(x, y) = \begin{cases} -\frac{1}{2\pi} \log |x - y| & d = 2, \\ \frac{1}{4\pi} \frac{1}{|x - y|} & d = 3. \end{cases}$$

在此基础上，需要构建适宜的边界积分方程，以  $x \in \Gamma$  生成完整的柯西数  $[\gamma_0^{\text{int}} u(x), \gamma_1^{\text{int}} u(x)]$ 。而这就需要我们探讨表面积位势 (surface potential) 和体积位势 (volume potential) 的映射特征。

### 14.6.1 牛顿位势

泊松方程表现式(14.159)中，对于某一给定的方程  $f(y), y \in \Omega$ , 我们可以定义  $f(y)$  的体积位势，或称牛顿位势算子 (Newton potential) 为  $\tilde{N}_0 f$ , 满足

$$(\tilde{N}_0 f)(x) := \int_{\Omega} U^*(x, y) f(y) dy, \quad x \in \mathbb{R}^d. \quad (14.160)$$

由内积形式可得

$$\langle \tilde{N}_0 \varphi, \psi \rangle_{\Omega} = \int_{\Omega} \psi(x) \int_{\Omega} U^*(x, y) \varphi(y) dy dx = \langle \varphi, \tilde{N}_0 \psi \rangle_{\Omega}, \quad \varphi, \psi \in \mathcal{S}(\mathbb{R}^d),$$

可见  $\tilde{N}_0 \varphi \in \mathcal{S}(\mathbb{R}^d)$ 。

进而, 牛顿位势算子  $\tilde{N}_0 : \mathcal{S}'(\mathbb{R}^d) \mapsto \mathcal{S}'(\mathbb{R}^d)$  可以定义为

$$\langle \tilde{N}_0 f, \psi \rangle_{\Omega} := \langle f, \tilde{N}_0 \psi \rangle_{\Omega}, \quad \forall \psi \in \mathcal{S}(\mathbb{R}^d).$$

**Theorem 14.31** (牛顿位势算子的映射). 牛顿位势算子  $\tilde{N}_0 : \tilde{H}^{-1}(\Omega) \mapsto H^1(\Omega)$  定义了一个连续映射

$$\|\tilde{N}_0 f\|_{H^1(\Omega)} \leq c \|f\|_{\tilde{H}^{-1}(\Omega)}. \quad (14.161)$$

证明. 对于  $\varphi \in C_0^\infty(\Omega)$  我们可得

$$\|\varphi\|_{H^{-1}(\mathbb{R}^d)}^2 = \int_{\mathbb{R}^d} \frac{|\hat{\varphi}(\xi)|^2}{1 + |\xi|^2} d\xi,$$

其中  $\hat{\varphi}(\xi)$  表示傅里叶变换

$$\hat{\varphi}(\xi) = (2\pi)^{-\frac{d}{2}} \int_{\mathbb{R}^d} \exp[-i\langle x, \xi \rangle] \varphi(x) dx.$$

由  $\text{supp } \varphi \subset \Omega$  可得 (支撑集 supp 的定义, 见356页脚注)

$$\begin{aligned} \|\varphi\|_{H^{-1}(\mathbb{R}^d)} &= \sup_{0 \neq \nu \in H^1(\mathbb{R}^d)} \frac{\langle \varphi, \nu \rangle_{L^2(\mathbb{R}^d)}}{\|\nu\|_{H^1(\mathbb{R}^d)}} \\ &\leq \sup_{0 \neq \nu \in H^1(\Omega)} \frac{\langle \varphi, \nu \rangle_{L^2(\Omega)}}{\|\nu\|_{H^1(\Omega)}} \\ &= \|\varphi\|_{\tilde{H}^{-1}(\Omega)}. \end{aligned} \quad (14.162)$$

由定义(14.160), 我们定义一个  $u(x)$

$$u(x) := (\tilde{N}_0 \varphi)(x) = \int_{\Omega} U^*(x, y) \varphi(y) dy, \quad x \in \mathbb{R}^d. \quad (14.163)$$

设  $\Omega \subset B_R(0)$ , 以及一个有紧支撑的非负单调递增 cutoff 方程  $\mu \in C_0^\infty([0, \infty))$ , 满足  $\mu(r) = 1, r \in [0, 2R]$ 。进而定义

$$u_\mu(x) := \int_{\Omega} \mu(|x - y|) U^*(x, y) \varphi(y) dy, \quad x \in \mathbb{R}^d. \quad (14.164)$$

$$\begin{aligned} &x, y \in \Omega, \\ &\rightarrow |x - y| \geq 0, \\ &\rightarrow \mu(|x - y|) = 1, \\ &\rightarrow u_\mu(x) = u(x), \quad x \in \Omega, \\ &\rightarrow \|u\|_{H^1(\Omega)} = \|u_\mu\|_{H^1(\Omega)} \leq \|u_\mu\|_{H^1(\mathbb{R}^d)} \end{aligned}$$

以及

$$\|u_\mu\|_{H^1(\mathbb{R}^d)}^2 = \int_{\mathbb{R}^d} (1 + |\xi|^2) |\widehat{u}_\mu(\xi)|^2 d\xi. \quad (14.165)$$

现在来计算  $u_\mu(x)$  的傅里叶变换

$$\begin{aligned}
 \widehat{u}_\mu(x) &= (2\pi)^{-\frac{d}{2}} \int_{\mathbb{R}^d} \exp[-i\langle x, \xi \rangle] u_\mu(x) dx \\
 &= (2\pi)^{-\frac{d}{2}} \int_{\mathbb{R}^d} \exp[-i\langle x, \xi \rangle] \int_{\mathbb{R}^d} \mu(|x-y|) U^*(x, y) \varphi(y) dy dx \\
 &= (2\pi)^{-\frac{d}{2}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \exp[-i\langle z+y, \xi \rangle] \mu(|z|) U^*(z+y, y) \varphi(y) dy dz \\
 &= \underbrace{(2\pi)^{-\frac{d}{2}} \int_{\mathbb{R}^d} \exp[-i\langle y, \xi \rangle] \varphi(y) dy}_{=: \tilde{\varphi}(\xi)} \underbrace{\int_{\mathbb{R}^d} \exp[-i\langle z, \xi \rangle] \mu(|z|) U^*(z, 0) dz}_{=: I(|\xi|)}.
 \end{aligned} \tag{14.166}$$

求解(14.166)需要进一步求  $I(|\xi|)$  的值。已知  $u(|z|)$  和  $U^*(z, 0)$  都是只与  $|z|$  有关的方程，我们可以利用傅里叶变换的旋转对称 Lemma 14.2，在三维坐标  $\xi = (0, 0, |\xi|)^\top$  中测算  $I(|\xi|)$ 。以  $d = 3$  为例，建立坐标系

$$z = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} r \cos \phi \sin \theta \\ r \sin \phi \sin \theta \\ r \cos \theta \end{pmatrix} \quad r \in [0, \infty), \phi \in [0, 2\pi), \theta \in [0, \pi).$$

将(14.156)代入  $I(|\xi|)$

$$\begin{aligned}
 I(|\xi|) &= \frac{1}{4\pi} \int_{\mathbb{R}^d} \exp[-i\langle z, \xi \rangle] \frac{\mu(|z|)}{|z|} dz \\
 &= \frac{1}{4\pi} \int_0^\infty \int_0^{2\pi} \int_0^\pi \exp[-i|\xi|r \cos \theta] \frac{\mu(r)}{r} r^2 \sin \theta d\theta d\phi dr \\
 &= \frac{1}{2} \int_0^\infty r \mu(r) dr \underbrace{\int_0^\pi \exp[-ir|\xi| \cos \theta] \sin \theta d\theta dr}.
 \end{aligned}$$

定义  $\iota := \cos \theta$ , 我们有

$$\begin{aligned}
 \int_0^\pi \exp[-ir|\xi| \cos \theta] \sin \theta d\theta &= \int_{-1}^1 \exp[-ir|\xi|\iota] d\iota \\
 &= \left[ -\frac{1}{ir|\xi|} \exp[-ir|\xi|\iota] \right]_{-1}^1 = \frac{2 \sin r|\xi|}{r|\xi|}.
 \end{aligned}$$

由此可得

$$I(|\xi|) = \frac{1}{|\xi|^2} \int_0^\infty \mu(r) \sin r|\xi| dr.$$

下面根据  $|\xi|$  的值，作分部求积。

1. 来看  $|\xi| > 1$  的情况。定义  $s := r|\xi|$  可得

$$I(|\xi|) = \frac{1}{|\xi|^2} \int_0^\infty \mu\left(\frac{s}{|\xi|}\right) \sin s ds.$$

根据定义可知  $0 \leq \mu(r) \leq 1$  且有紧支撑，则

$$I(|\xi|) \leq c_1(R) \frac{1}{|\xi|^2}, \quad |\xi| \geq 1. \tag{14.167}$$

此外考虑到

$$(1 + |\xi|^2)^2 \leq 4|\xi|^4, \quad |\xi| \geq 1,$$

结合(14.165), (14.166), (14.167)可得

$$\begin{aligned}
 \|u_\mu\|_{H^1(\mathbb{R}^d), |\xi|>1}^2 &= \int_{|\xi|>1} (1+|\xi|^2) |\widehat{u}_\mu(\xi)|^2 d\xi \\
 &= \int_{|\xi|>1} (1+|\xi|^2) |\widehat{\varphi}(\xi) I(|\xi|)|^2 d\xi \\
 &\leq [c_1(R)]^2 \int_{|\xi|>1} \frac{1+|\xi|^2}{|\xi|^4} |\widehat{\varphi}(\xi)|^2 d\xi \\
 &\leq 4 [c_1(R)]^2 \int_{|\xi|>1} \frac{1}{1+|\xi|^2} |\widehat{\varphi}(\xi)|^2 d\xi.
 \end{aligned} \tag{14.168}$$

2. 来看  $|\xi| \leq 1$  的情况。

$$\begin{aligned}
 I(|\xi|) &= \int_0^\infty \mu(r) \frac{\sin r|\xi|}{|\xi|} dr \\
 &\leq c_2(R), \quad |\xi| \leq 1.
 \end{aligned} \tag{14.169}$$

结合(14.165), (14.166), (14.169)可得

$$\begin{aligned}
 \|u_\mu\|_{H^1(\mathbb{R}^d), |\xi|\leq 1}^2 &= \int_{|\xi|\leq 1} (1+|\xi|^2) |\widehat{u}_\mu(\xi)|^2 d\xi \\
 &= \int_{|\xi|\leq 1} (1+|\xi|^2) |\widehat{\varphi}(\xi) I(|\xi|)|^2 d\xi \\
 &\leq 2 [c_2(R)]^2 \int_{|\xi|\leq 1} |\widehat{\varphi}(\xi)|^2 d\xi \\
 &\leq 4 [c_2(R)]^2 \int_{|\xi|\leq 1} \frac{1}{1+|\xi|^2} |\widehat{\varphi}(\xi)|^2 d\xi.
 \end{aligned} \tag{14.170}$$

将(14.168), (14.170)汇总, 代回(14.165), (14.166)可得

$$\begin{aligned}
 \|u_\mu\|_{H^1(\mathbb{R}^d)}^2 &= \int_{\xi \in \mathbb{R}^d} (1+|\xi|^2) |\widehat{u}_\mu(\xi)|^2 d\xi \\
 &\leq c \int_{\xi \in \mathbb{R}^d} \frac{1}{1+|\xi|^2} |\widehat{\varphi}(\xi)|^2 d\xi \\
 &= c \|\varphi\|_{H^{-1}(\mathbb{R}^d)}^2.
 \end{aligned} \tag{14.171}$$

将(14.164), (14.162)代入上式, 有

$$\begin{aligned}
 \|\widetilde{N}_0 \varphi\|_{H^1(\mathbb{R}^d)}^2 &= \|u_\mu\|_{H^1(\mathbb{R}^d)}^2 \leq c \|\varphi\|_{H^{-1}(\mathbb{R}^d)}^2 \\
 &\leq c \|\varphi\|_{\widetilde{H}^{-1}(\Omega)}^2.
 \end{aligned} \tag{14.172}$$

因此

$$\frac{|\langle \widetilde{N}_0 f, \varphi \rangle_\Omega|}{\|\varphi\|_{\widetilde{H}^{-1}(\Omega)}} = \frac{|\langle f, \widetilde{N}_0 \varphi \rangle_\Omega|}{\|\varphi\|_{\widetilde{H}^{-1}(\Omega)}} \leq \frac{\|f\|_{\widetilde{H}^{-1}(\Omega)} \|\widetilde{N}_0 \varphi\|_{H^1(\Omega)}}{\|\varphi\|_{\widetilde{H}^{-1}(\Omega)}} \leq c \|f\|_{\widetilde{H}^{-1}(\Omega)}, \quad \forall \varphi \in C_0^\infty(\Omega).$$

对  $\|\cdot\|_{\widetilde{H}^{-1}(\Omega)}$  取闭包, 由对偶配对可得(14.161)。  $\square$

**Theorem 14.32** ( $\mathbb{R}^d$  中的牛顿位势通解). 牛顿位势  $\widetilde{N}_0 \widetilde{f}$  是下述偏微分方程的通解

$$-\Delta_x (\widetilde{N}_0 \widetilde{f}) = \widetilde{f}(x) = \begin{cases} f(x) & x \in \Omega, \\ 0 & x \in \mathbb{R}^d \setminus \overline{\Omega}. \end{cases} \tag{14.173}$$

证明. 对于  $\varphi \in C_0^\infty(\mathbb{R}^d)$ , 依次作分段积分, 改变积分符顺序, 使用基本解的对称特性, 可得

$$\begin{aligned}
\int_{\mathbb{R}^d} \left[ -\Delta_x \left( \tilde{N}_0 \tilde{f} \right) (x) \right] \varphi(x) dx &= \int_{\mathbb{R}^d} \underbrace{\left( \tilde{N}_0 \tilde{f} \right) (x)}_{(14.160)} [-\Delta_x \varphi(x)] dx \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} U^*(x, y) \tilde{f}(y) dy [-\Delta_x \varphi(x)] dx \\
&= \int_{\mathbb{R}^d} \tilde{f}(y) \int_{\mathbb{R}^d} U^*(x, y) [-\Delta_x \varphi(x)] dx dy \\
&= \int_{\mathbb{R}^d} \tilde{f}(y) \int_{\mathbb{R}^d} \underbrace{[-\Delta_x U^*(x, y)]}_{=: \delta_0(x-y)} \varphi(x) dx dy \\
&= \int_{\mathbb{R}^d} \tilde{f}(y) \underbrace{\int_{\mathbb{R}^d} \delta_0(x-y) \varphi(x) dx}_{\text{ }} dy \\
&= \int_{\mathbb{R}^d} \tilde{f}(y) \varphi(y) dy.
\end{aligned}$$

对  $\|\cdot\|_{H^1(\mathbb{R}^d)}$  取  $C_0^\infty(\mathbb{R}^d)$  的闭包, 再使用配偶配对, 可得(14.173)成立。  $\square$

**Corollary 14.6** ( $\Omega$  中的牛顿位势通解). 将 Theorem 14.32 中的牛顿位势通解进一步限定到有界域  $\Omega \subset \mathbb{R}^d$  的情况。

此时位势方程  $\tilde{N}_0 f$  是以下偏微分方程的通解

$$-\Delta_x \tilde{N}_0 f(x) = f, \quad x \in \Omega.$$

内界迹算子  $\gamma_0^{int}(\tilde{N}_0 f)$  可以表示为

$$\gamma_0^{int}(\tilde{N}_0 f)(x) = \lim_{\Omega \ni \tilde{x} \mapsto x \in \Gamma} (\tilde{N}_0 f)(\tilde{x}). \quad (14.174)$$

内界迹算子  $\gamma_0^{int}(\tilde{N}_0 f)$  定义了一个有界线性算子  $N_0 f$

$$N_0 = \gamma_0^{int} \tilde{N}_0 : \tilde{H}^{-1}(\Omega) \mapsto H^{\frac{1}{2}}(\Gamma), \quad (14.175)$$

$N_0 f$  的范数满足

$$\|N_0 f\|_{H^{\frac{1}{2}}(\Gamma)} \leq c_2^N \|f\|_{\tilde{H}^{-1}(\Omega)} \quad \forall f \in \tilde{H}^{-1}(\Omega). \quad (14.176)$$

**Lemma 14.18** (牛顿位势算子方程的表现式). 设给定  $f \in L^\infty$ 。那么牛顿位势表示如下等式, 又称弱奇异表面积分 (*weakly singular surface integral*)

$$(N_0 f)(x) = \gamma_0^{int} (\tilde{N}_0 f)(x) = \int_{\Omega} U^*(x, y) f(y) dy, \quad x \in \Gamma.$$

证明. 关于奇异积分、弱奇异积分, 可参考 Neri (1971); Vainikko (1993)。关于表面积分, 可参考 Callahan, 2010, Ch.10)。

对于给定的  $\tilde{x} \in \Omega, x \in \Gamma$ , 设  $\exists \varepsilon > 0$ , 满足  $|\tilde{x} - x| < \varepsilon$ 。则

$$\begin{aligned}
&\left| \int_{\Omega} U^*(\tilde{x}, y) f(y) dy - \int_{y \in \Omega: |y-x| > \varepsilon} U^*(x, y) f(y) dy \right| \\
&\leq \underbrace{\left| \int_{y \in \Omega: |y-x| > \varepsilon} [U^*(\tilde{x}, y) - U^*(x, y)] f(y) dy \right|}_{=: \mathcal{A}} + \underbrace{\left| \int_{y \in \Omega: |y-x| \leq \varepsilon} U^*(\tilde{x}, y) f(y) dy \right|}_{=: \mathcal{B}},
\end{aligned} \quad (14.177)$$

其中

1.  $\mathcal{A} \Rightarrow$

$$\lim_{\Omega \ni \tilde{x} \mapsto x \in \Gamma} \left| \int_{y \in \Omega: |y-x| > \varepsilon} [U^*(\tilde{x}, y) - U^*(x, y)] f(y) dy \right| = 0.$$

2.  $\mathcal{B} \Rightarrow$

$$\begin{aligned} \left| \int_{y \in \Omega: |y-x| \leq \varepsilon} U^*(\tilde{x}, y) f(y) dy \right| &\leq \|f\|_{L^\infty(\Omega \cap B_\varepsilon(x))} \int_{\Omega \cap B_\varepsilon(x)} |U^*(\tilde{x}, y)| dy \\ &\leq \|f\|_{L^\infty(\Omega)} \underbrace{\int_{B_\varepsilon(x)} |U^*(\tilde{x}, y)| dy}_{=: \mathcal{C}} \end{aligned}$$

3. 分  $d = 2, 3$  两种情况来讨论  $\mathcal{C}$  的值

(a)  $d = 2 \Rightarrow$

$$\begin{aligned} \mathcal{C} &:= \int_{B_\varepsilon(x)} |U^*(\tilde{x}, y)| dy = \frac{1}{2\pi} \int_{|y-\tilde{x}| < 2\varepsilon} |\log |y-\tilde{x}|| dy \\ &= \frac{1}{2} \int_0^{2\pi} \int_0^{2\varepsilon} |\log r| r dr d\varphi \\ &= \varepsilon^2 [1 - 2 \log(2\varepsilon)]. \end{aligned} \tag{14.178}$$

(b)  $d = 3 \Rightarrow$

$$\begin{aligned} \mathcal{C} &:= \int_{B_\varepsilon(x)} |U^*(\tilde{x}, y)| dy = \frac{1}{4\pi} \int_{|y-\tilde{x}| < 2\varepsilon} \frac{1}{|y-\tilde{x}|} dy \\ &= \frac{1}{4\pi} \int_0^{2\pi} \int_0^\pi \int_0^{2\varepsilon} \frac{1}{r} r^2 \sin \psi dr d\psi d\varphi \\ &= 2\varepsilon^2. \end{aligned} \tag{14.179}$$

取极限  $\tilde{x} \rightarrow x, \varepsilon \rightarrow 0$ , 可证得。  $\square$

**Lemma 14.19.** 算子

$$N_1 := \gamma_1^{int} \tilde{N}_0 : \tilde{H}^{-1}(\Omega) \mapsto H^{-\frac{1}{2}}(\Gamma)$$

是有界算子, 即是说, 满足不等式条件

$$\|N_1 f\|_{H^{-\frac{1}{2}}(\Gamma)} = \left\| \gamma_1^{int} \tilde{N}_0 f \right\|_{H^{-\frac{1}{2}}(\Gamma)} \leq c \|f\|_{\tilde{H}^{-1}(\Omega)}.$$

证明. 偏微分方程

$$-\Delta u(x) = f(x), \quad x \in \Omega$$

的通解是  $u = \tilde{N}_0 f \in H^1(\Omega)$ 。

对于任意给定的  $w \in H^{\frac{1}{2}}(\Gamma)$ , 构建一个有界延拓  $\varepsilon w \in H^1(\Omega)$ , 满足

$$\|\varepsilon w\|_{H^1(\Omega)} \leq c_{IT} \|w\|_{H^{\frac{1}{2}}(\Gamma)}.$$

$\square$

由格林第一恒等式(14.6)可得

$$\begin{aligned} \langle \gamma_1^{int} u, w \rangle_\Gamma &= \langle u, \varepsilon_w w \rangle \\ &= \int_{\Omega} \nabla u(x) \nabla \varepsilon_w(x) dx - \langle f, \varepsilon_w \rangle_\Omega. \end{aligned}$$

代入牛顿位势算子的映射(14.161)(Theorem 14.31) 中, 我们有

$$\begin{aligned} |\langle \gamma_1^{\text{int}} u, w \rangle_{\Gamma}| &\leq \left\{ \|u\|_{H^1(\Omega)} + \|f\|_{\tilde{H}^{-1}(\Omega)} \right\} \|\varepsilon_w\|_{H^1(\Omega)} \\ &\leq (c+1)c_{\text{IT}} \|f\|_{\tilde{H}^{-1}(\Omega)} \|w\|_{H^{-\frac{1}{2}}(\Gamma)}. \end{aligned}$$

### 14.6.2 单层位势

设  $w \in H^{-\frac{1}{2}}(\Gamma)$  为一个给定的密度方程。考虑如下单层位势算子 (single layer potential)

$$u(x) := (\tilde{V}w)(x) := \int_{\Gamma} U^*(x, y) w(y) ds_y, \quad x \in \Omega \cup \Omega^c, \forall y \in \Gamma. \quad (14.180)$$

则基于给定的密度  $w \in H^{-\frac{1}{2}}(\Gamma)$ , 单层位势算子  $(\tilde{V}w)(x) \in H^1(\Omega)$  就是齐次偏微分方程的解, 如下

**Lemma 14.20.** (14.180) 定义的单层位势算子  $u(x) = (\tilde{V}w)(x)$ , 构成如下齐次偏微分方程的解

$$-\Delta u(x) = 0, \quad x \in \Omega \cup \Omega^c.$$

对于给定的密度方程  $w \in H^{-\frac{1}{2}}(\Gamma)$ , 我们有对应的  $u \in H^1(\Omega)$ , 满足

$$\|u\|_{H^1(\Omega)} = \|\tilde{V}w\|_{H^1(\Omega)} \leq c \|w\|_{H^{-\frac{1}{2}}(\Gamma)}. \quad (14.181)$$

证明. 对于  $x \in \Omega \cup \Omega^c$  和  $y \in \Gamma$ , 可得齐次偏微分方程的基本解  $U^*(x, y) \in C^\infty$ 。因此, 我们可以调整差分和积分的顺序

$$\begin{aligned} -\Delta_x u(x) &= -\Delta_x \int_{\Omega} U^*(x, y) f(y) dy \\ &= \int_{\Omega} [-\Delta_x U^*(x, y)] f(y) dy = 0. \end{aligned}$$

在此基础上, 对于  $\varphi \in C^\infty(\Omega)$ , 可得

$$\begin{aligned} \int_{\Omega} u(x) \varphi(x) dx &= \int_{\Omega} \underbrace{\int_{\Gamma} U^*(x, y) w(y) ds_y}_{:=u(x)} \varphi(x) dx \\ &= \int_{\Gamma} w(y) \underbrace{\int_{\Omega} U^*(x, y) \varphi(x) dx}_{:=(N_0\varphi)(y)} ds_y, \\ &= \int_{\Gamma} w(y) (N_0\varphi)(y) ds_y \end{aligned}$$

其中定义  $(N_0\varphi)(y) ds_y$  为

$$(N_0\varphi)(y) ds_y := \gamma_0^{\text{int}} \int_{\Omega} U^*(x, y) \varphi(x) dx, \quad x \in \Gamma.$$

代入牛顿位势在  $\Omega$  中的通解(14.176)(Corollary 14.6), 上式进一步变为

$$\begin{aligned} \int_{\Omega} u(x) \varphi(x) dx &= \int_{\Gamma} w(y) (N_0\varphi)(y) ds_y \\ &\leq \|w\|_{H^{-\frac{1}{2}}(\Gamma)} \|N_0\varphi\|_{H^{\frac{1}{2}}(\Gamma)} \\ &\leq c_2^N \|w\|_{H^{-\frac{1}{2}}(\Gamma)} \|\varphi\|_{\tilde{H}^{-1}(\Omega)} \end{aligned}$$

对  $\|\cdot\|_{\tilde{H}^{-1}(\Omega)}$  取  $c^\infty(\Omega)$  的闭包, 再使用配偶配对, 可证(14.181)成立。  $\square$

单层位势算子(14.180)定义了一个有界的线性映射  $\tilde{V} : H^{-\frac{1}{2}}(\Gamma) \mapsto H^1(\Omega)$ 。因此，内界迹算子  $\tilde{V}w \in H^{-1}(\Omega)$  定义良好。对应地，我们也可以写出线性算子  $V = \gamma_0^{int}\tilde{V} : H^{-\frac{1}{2}}(\Gamma) \mapsto H^{\frac{1}{2}}(\Gamma)$ ，其中  $V$  满足

$$\|Vw\|_{H^{\frac{1}{2}}(\Gamma)} \leq c_2^V \|w\|_{W^{-\frac{1}{2}}(\Gamma)}, \quad \forall w \in H^{-\frac{1}{2}}(\Gamma). \quad (14.182)$$

**Lemma 14.21** (单层位势算子方程的表现式). 设给定  $w \in L^\infty(\Gamma)$ ，则单层位势算子方程的表现式 (*representation formula*) 是一个弱奇异表面积分

$$(Vw)(x) = \gamma_0^{int}(\tilde{V}w)(x) = \int_{\Gamma} U^*(x, y)w(y) ds_y, \quad x \in \Gamma. \quad (14.183)$$

证明. 对于给定的  $\tilde{x} \in \Omega, x \in \Gamma$ ，假定  $\exists \varepsilon > 0$  满足  $|x - \tilde{x}| < \varepsilon$ ，那么

$$\begin{aligned} & \left| \int_{\Gamma} U^*(\tilde{x}, y)w(y) ds_y - \int_{y \in \Gamma: |y-x| > \varepsilon} U^*(x, y)w(y) ds_y \right| \\ & \leq \underbrace{\left| \int_{y \in \Gamma: |y-x| > \varepsilon} [U^*(\tilde{x}, y) - U^*(x, y)] w(y) ds_y \right|}_{=: \mathcal{A}} + \underbrace{\left| \int_{y \in \Gamma: |y-x| \leq \varepsilon} U^*(\tilde{x}, y)w(y) ds_y \right|}_{=: \mathcal{B}}, \end{aligned} \quad (14.184)$$

其中

1.  $\mathcal{A} \Rightarrow$

$$\lim_{\Omega \ni \tilde{x} \rightarrow x \in \Gamma} \left| \int_{y \in \Gamma: |y-x| > \varepsilon} [U^*(\tilde{x}, y) - U^*(x, y)] w(y) ds_y \right| = 0.$$

2.  $\mathcal{B} \Rightarrow$

$$\begin{aligned} \left| \int_{y \in \Gamma: |y-x| \leq \varepsilon} U^*(\tilde{x}, y)w(y) ds_y \right| & \leq \|w\|_{L^\infty(\Gamma \cap B_\varepsilon(x))} \int_{\Gamma \cap B_\varepsilon(x)} |U^*(\tilde{x}, y)| ds_y \\ & \leq \|w\|_{L^\infty(\Gamma)} \underbrace{\int_{B_\varepsilon(x)} |U^*(\tilde{x}, y)| ds_y}_{=: \mathcal{C}}, \end{aligned}$$

3. 类似地，分  $d = 2, 3$  两种情况来讨论  $\mathcal{C}$  的值

(a)  $d = 2 \Rightarrow$  (14.178)

$$\begin{aligned} \mathcal{C} & := \int_{B_\varepsilon(x)} |U^*(\tilde{x}, y)| dy = \frac{1}{2\pi} \int_{|y-\tilde{x}| < 2\varepsilon} |\log|y-\tilde{x}|| dy \\ & = \frac{1}{2} \int_0^{2\pi} \int_0^{2\varepsilon} |\log r| r dr d\varphi \\ & = \varepsilon^2 [1 - 2 \log(2\varepsilon)]. \end{aligned}$$

(b)  $d = 3 \Rightarrow$  (14.179)

$$\begin{aligned} \mathcal{C} & := \int_{B_\varepsilon(x)} |U^*(\tilde{x}, y)| dy = \frac{1}{4\pi} \int_{|y-\tilde{x}| < 2\varepsilon} \frac{1}{|y-\tilde{x}|} dy \\ & = \frac{1}{4\pi} \int_0^{2\pi} \int_0^\pi \int_0^{2\varepsilon} \frac{1}{r} r^2 \sin\psi dr d\psi d\varphi \\ & = 2\varepsilon^2. \end{aligned}$$

取极限  $\tilde{x} \rightarrow x, \varepsilon \rightarrow 0$ ，可证得。  $\square$

类似地，我们也可以得到单层位势算子的外界迹 (exterior trace)

$$(Vw)(x) = \gamma_0^{\text{ext}}(\tilde{V}w)(x) := \lim_{\Omega^c \ni \tilde{x} \mapsto x \in \Gamma} (\tilde{V}w)(\tilde{x}), \quad x \in \Gamma. \quad (14.185)$$

单层位势的跃动关系 (jump relation) 因此可以表示为

$$[\gamma_0 \tilde{V}w] := \gamma_0^{\text{ext}}(\tilde{V}w)(x) - \gamma_0^{\text{int}}(\tilde{V}w)(x) = 0, \quad x \in \Gamma. \quad (14.186)$$

### 14.6.3 伴随双层位势

由第14.6.2可知，基于给定密度方程  $w \in H^{-\frac{1}{2}}(\Gamma)$ ，我们可以定义一个位势算子  $(\tilde{V}w) \in H^1(\Omega)$ ，作为齐次偏微分方程的解，详见 Lemma 14.20。

在此基础上，可以进一步将对应的内界共形导数定义成以下有界线性算子  $\gamma_1^{\text{int}} \tilde{V} : H^{-\frac{1}{2}}(\Gamma) \rightarrow H^{-\frac{1}{2}}(\Gamma)$ ，满足

$$\|\gamma_1^{\text{int}} \tilde{V}w\|_{H^{-\frac{1}{2}}(\Omega)} \leq c \|w\|_{H^{-\frac{1}{2}}(\Gamma)}, \quad \forall w \in H^{-\frac{1}{2}}(\Gamma).$$

于是有伴随双层位势算子 (adjoint double layer potential) 如下

**Lemma 14.22** (伴随双层位势算子). 给定密度方程  $w \in H^{-\frac{1}{2}}(\Gamma)$ ，对应内部共形导数  $\gamma_1^{\text{int}}(\tilde{V}w)(x) \in H^{-\frac{1}{2}}(\Gamma)$  的表现式为

$$\gamma_1^{\text{int}}(\tilde{V}w)(x) = \sigma(x)w(x) + (K'w)(x), \quad x \in \gamma,$$

采用内积形式可表现为

$$\langle \gamma_1^{\text{int}}(\tilde{V}w)(x), \nu \rangle_\Gamma = \langle \sigma w + K'w, \nu \rangle_\Gamma, \quad \forall \nu \in H^{\frac{1}{2}}(\Gamma),$$

其中与伴随双层位势相关的算子  $(K'w)(x)$  和  $\sigma(x)$  分别定义为

$$(K'w)(x) := \lim_{\varepsilon \rightarrow 0} \int_{y \in \Gamma: |y-x| \geq \varepsilon} \gamma_{1,x}^{\text{int}} U^*(x, y) w(y) ds_y, \quad (14.187)$$

$$\sigma(x) := \lim_{\varepsilon \rightarrow 0} \frac{1}{2(d-1)\pi} \frac{1}{\varepsilon^{d-1}} \int_{y \in \Omega: |y-x|=\varepsilon} ds_y, \quad x \in \Gamma. \quad (14.188)$$

证明. 已知给定方程  $w \in H^{-\frac{1}{2}}(\Gamma)$ ，齐次偏微分方程的解， $u(x) = (\tilde{V}w) \in H^1(\Omega)$ ，见 Lemma 14.20。

进而由格林第一恒等式(14.6)可得，对某一  $\varphi \in C^\infty(\Omega)$  我们有

$$\begin{aligned} \int_{\Gamma} u(x) \gamma_0^{\text{int}} \varphi(x) ds_x &= \int_{\Omega} \nabla_x u(x) \nabla_x \varphi(x) dx \\ &= \int_{\Omega} \nabla_x \underbrace{\int_{\Gamma} U^*(x, y) w(y) ds_y}_{=:u(x)} \nabla \varphi(x) dx \end{aligned}$$

由弱奇异表面积分 Lemma 14.21 可见

$$u(x) = (Vw) = \gamma_0^{\text{int}}(\tilde{V}w) = \int_{\Gamma} U^*(x, y) w(y) ds_y,$$

上式因此可调整为

$$\begin{aligned}\int_{\Gamma} \gamma_1^{\text{int}} u(x) \gamma_0^{\text{int}} \varphi(x) ds_x &= \int_{\Omega} \nabla_x \left[ \lim_{\varepsilon \rightarrow 0} \int_{y \in \Gamma: |y-x| \geq \varepsilon} U^*(x, y) w(y) ds_y \right] \nabla_x \varphi(x) dx \\ &= \int_{\Gamma} w(y) \underbrace{\lim_{\varepsilon \rightarrow 0} \int_{x \in \Gamma: |x-y| \geq \varepsilon} \nabla_x U^*(x, y) \nabla_x \varphi(x) dx ds_y}_{=: \mathcal{A}},\end{aligned}$$

对  $\mathcal{A}$  再一次使用格林第一恒等式(14.6)

$$\begin{aligned}\mathcal{A} &:= \int_{x \in \Gamma: |x-y| \geq \varepsilon} \nabla_x U^*(x, y) \nabla_x \varphi(x) dx \\ &= \underbrace{\int_{x \in \Omega: |x-y|=\varepsilon} \gamma_{1,x}^{\text{int}} U^*(x, y) \varphi(x) ds_x}_{=: \mathcal{B}} + \underbrace{\int_{x \in \Gamma: |x-y| \geq \varepsilon} \gamma_{1,x}^{\text{int}} U^*(x, y) \gamma_0^{\text{int}} \varphi(x) ds_x}_{=: \mathcal{C}},\end{aligned}$$

其中

1.  $\mathcal{C}$  对应伴随双层位势算子  $K'$ , 如式(14.187)。

2.  $\mathcal{B}$  计算如下

$$\begin{aligned}\mathcal{B} &:= \int_{x \in \Omega: |x-y|=\varepsilon} \gamma_{1,x}^{\text{int}} U^*(x, y) \varphi(x) ds_x \\ &= \underbrace{\int_{x \in \Omega: |x-y|=\varepsilon} \gamma_{1,x}^{\text{int}} U^*(x, y) [\varphi(x) - \varphi(y)] ds_x}_{=: \mathcal{B}_1} + \varphi(y) \underbrace{\int_{x \in \Omega: |x-y|=\varepsilon} \gamma_{1,x}^{\text{int}} U^*(x, y) ds_x}_{=: \mathcal{B}_2},\end{aligned}$$

并且

$$\|\mathcal{B}\| \leq \max_{x \in \Omega: |x-y|=\varepsilon} |\varphi(x) - \varphi(y)| \underbrace{\int_{x \in \Omega: |x-y|=\varepsilon} |\gamma_{1,x}^{\text{int}} U^*(x, y)| ds_x}_{=: \mathcal{D}}$$

(a) 对于  $\mathcal{B}_1$ , 取  $\varepsilon \rightarrow 0$  时的极值

$$\lim_{\varepsilon \rightarrow 0} |\mathcal{B}_1| = \lim_{\varepsilon \rightarrow 0} \left| \int_{x \in \Omega: |x-y|=\varepsilon} \gamma_{1,x}^{\text{int}} U^*(x, y) [\varphi(x) - \varphi(y)] ds_x \right| = 0,$$

(b) 对于  $\mathcal{B}_2$ , 设

$$n_x := \frac{y-x}{|y-x|}, x \in \Omega, \quad \varepsilon := |y-x|,$$

$$\begin{aligned}\hookrightarrow \mathcal{B}_2 &= \int_{x \in \Omega: |x-y|=\varepsilon} \gamma_{1,x}^{\text{int}} U^*(x, y) ds_x \\ &= -\frac{1}{2(d-1)\pi} \int_{x \in \Omega: |x-y|=\varepsilon} \frac{(nx, x-y)}{|x-y|^d} ds_x \\ &= \frac{1}{2(d-1)\pi} \int_{x \in \Omega: |x-y|=\varepsilon} \frac{1}{|x-y|^{d-1}} ds_x \\ &= \frac{1}{2(d-1)\pi} \frac{1}{\varepsilon^{d-1}} \int_{x \in \Omega: |x-y|=\varepsilon} ds_x\end{aligned}$$

(c) 分  $d = 2, 3$  两种情况来讨论  $\mathcal{D}$  的值

i.  $d = 2 \Rightarrow$

$$\begin{aligned}\mathcal{D} &\leq \int_{x \in \mathcal{R}^2 : |x-y|=\varepsilon} |\gamma_{1,x}^{\text{int}} U^*(x,y)| ds_x \\ &= \frac{1}{2\pi} \int_{x \in \mathcal{R}^2 : |x-y|=\varepsilon} \frac{1}{|x-y|} ds_x = 1,\end{aligned}$$

ii.  $d = 3 \Rightarrow$

$$\begin{aligned}\mathcal{D} &\leq \int_{x \in \mathcal{R}^3 : |x-y|=\varepsilon} |\gamma_{1,x}^{\text{int}} U^*(x,y)| ds_x \\ &= \frac{1}{4\pi} \int_{x \in \mathcal{R}^3 : |x-y|=\varepsilon} \frac{1}{|x-y|^2} ds_x = 1.\end{aligned}$$

3. 由此我们可得

$$\begin{aligned}&\int_{\Gamma} \gamma_1^{\text{int}} u(x) \gamma_0^{\text{int}} \varphi(x) ds_x \\ &= \int_{\Gamma} \left\{ \lim_{\varepsilon \rightarrow 0} \int_{x \in \Gamma : |x-y| \geq \varepsilon} \gamma_{1,x}^{\text{int}} U^*(x,y) \gamma_0^{\text{int}} \varphi(x) ds_x + \gamma_0^{\text{int}} \varphi(y) \sigma(y) \right\} ds_y \\ &= \left[ \int_{\Gamma} \gamma_0^{\text{int}} \varphi(x) \lim_{\varepsilon \rightarrow 0} \int_{x \in \Gamma : |x-y| \geq \varepsilon} \gamma_{1,x}^{\text{int}} U^*(x,y) w(y) ds_y ds_x \right] + \left[ \int_{\Gamma} w(y) \sigma(y) \varphi(y) ds_y \right] \\ &= \int_{\gamma} \gamma_0^{\text{int}} \varphi(x) [\sigma(x) w(x) + (K' w)(x)] ds_x.\end{aligned}$$

□

设  $x \in \gamma = \partial\Omega$  是可积区间 (或至少在  $\Gamma$  的某一段邻域中可积), 此时定义式(14.188)由

$$\sigma(x) = \frac{1}{2}, \quad \text{对于 (几乎所有) } x \in \Gamma.$$

伴随着单层位势出现的边界积分算子  $K'$  是伴随双层位势 (adjoint double layer potential)。 $K'$  线性且有界, 即

$$\|K'w\|_{H^{-\frac{1}{2}}(\Gamma)} \leq c_2^{K'} \|w\|_{H^{-\frac{1}{2}}(\Gamma)}, \quad w \in H^{-\frac{1}{2}}(\Gamma). \quad (14.189)$$

由 Lemma 14.22 进一步可知, 单层位势算子  $\tilde{V}$  在  $H^{-\frac{1}{2}}(\Gamma)$  中的外部共形导数 (exterior conormal derivative)

$$(\gamma_1 \tilde{V} w)(x) = [\sigma(x) - 1] w(x) + (K' w)(x), \quad x \in \Gamma.$$

**Lemma 14.23** (单层位势外部共形导数的跃动关系). 在  $H^{-\frac{1}{2}}(\Gamma)$  中, 单层位势  $\tilde{V}$  的外部共法导数满足跃动关系 (jump relation)

$$\gamma_1 (\tilde{V} w) := \gamma_1^{\text{ext}} (V w) - \gamma_1^{\text{int}} (\tilde{V} w) = -w(x), \quad x \in \Gamma. \quad (14.190)$$

证明. 对于  $u = \tilde{V} w, \varphi \in C_0^\infty(\mathbb{R}^d)$ ,

1. 一方面我们有

$$\begin{aligned}\int_{\mathbb{R}^d} [-\Delta u(x)] \varphi(x) dx &= \int_{\mathbb{R}^d} -\Delta_x \int_{\Gamma} U^*(x,y) w(y) ds_y \varphi(x) dx \\ &= \int_{\Gamma} w(y) \int_{\mathbb{R}^d} -\Delta_x U^*(x,y) \varphi(x) dx ds_y \\ &= \int_{\Gamma} w(y) \int_{\mathbb{R}^d} \delta_0(x-y) \varphi(x) dx ds_y \\ &= \int_{\Gamma} w(y) \gamma_0^{\text{int}} \varphi(y) ds_y.\end{aligned}$$

2. 另一方面我们有

$$\begin{aligned} \int_{\mathbb{R}^d} [-\Delta u(x)] \varphi(x) dx &= a_{\mathbb{R}^d}(u, \varphi) = a_{\Omega}(u, \varphi) + a_{\Omega^c}(u, \varphi) \\ &= \int_{\Gamma} \gamma_1^{\text{int}} u(x) \gamma_0^{\text{int}} \varphi(x) ds_x + \int_{\Gamma} \gamma_1^{\text{ext}} u(x) \gamma_0^{\text{ext}} \varphi(x) ds_x. \end{aligned}$$

3. 结合上两式可得

$$\int_{\Gamma} w(x) \varphi(x) ds_x = \int_{\Gamma} \gamma_1^{\text{int}} u(x) \gamma_0^{\text{int}} \varphi(x) ds_x - \int_{\Gamma} \gamma_1^{\text{ext}} u(x) \gamma_0^{\text{int}} \varphi(x) ds_x, \quad \forall \varphi \in C_0^{\infty}(\mathbb{R}^d).$$

对  $\|\cdot\|_{H^{\frac{1}{2}}(\Gamma)}$  取  $C_0^{\infty}(\mathbb{R}^d)$  的闭包，再使用配偶配对，可证(14.190)成立。

□

#### 14.6.4 双层位势

对于给定的密度方程  $\nu(x) \in H^{\frac{1}{2}}(\Gamma)$ ，我们将双层位势 (double layer potential) 定义如下

$$u(x) = (W\nu)(x) := \int_{\Gamma} [\gamma_{1,y}^{\text{int}} U^*(x, y)] \nu(y) dy, \quad x \in \Omega \cup \Omega^c. \quad (14.191)$$

**Lemma 14.24.**  $u(x) = (W\nu)(x), x \in \Omega \cup \Omega^c$  构成齐次偏微分方程

$$-\Delta_x u(x) = 0, \quad x \in \Omega \cup \Omega^c$$

的解。若  $\nu \in H^{\frac{1}{2}}(\Gamma)$ ，则我们有  $u \in H^1(\Omega)$ ，满足关系

$$\|u\|_{H^1(\Omega)} = \|W\nu\|_{H^1(\Omega)} \leq c \|\nu\|_{H^{\frac{1}{2}}(\Gamma)}.$$

证明. 由  $x \in \Omega \cup \Omega^c, y \in \Gamma$  可得  $x \neq y$ ，进而可以通过改动差分、积分的顺序，调整泊松方程(14.159)中的基本解  $U(x, y)$ : 对于  $\varphi \in C^{\infty}(\Omega)$  我们有

$$\begin{aligned} \langle W\nu, \varphi \rangle_{\Omega} &= \int_{\Omega} \int_{\Gamma} [\gamma_{1,y}^{\text{int}} U^*(x, y)] \nu(y) ds_y \varphi(x) dx \\ &= \int_{\Gamma} \nu(y) \gamma_{1,y}^{\text{int}} \int_{\Gamma} U^*(x, y) \varphi(x) dx ds_y \\ &= \int_{\Gamma} \nu(y) \gamma_{1,y}^{\text{int}} (\tilde{N}_0, \varphi)(y) ds_y \\ &= \langle \nu, \gamma_1^{\text{int}} (\tilde{N}_0, \varphi) \rangle_{\Gamma}. \end{aligned}$$

对于  $f \in \tilde{H}^{-1}(\Omega)$  而言，上式意味着

$$\langle W\nu, f \rangle_{\Omega} = \langle \nu, \gamma_1^{\text{int}} \tilde{N}_0 f \rangle_{\Gamma}.$$

根据 Corollary 14.6 可知， $\tilde{N}_0 f \in H^1(\Omega)$  是以下非齐次偏微分方程的解

$$-\Delta_x (\tilde{N}_0 f)(x) = f(x), \quad x \in \Omega.$$

由 Lemma 14.19 可得

$$\gamma_1^{\text{int}} \tilde{N}_0 f \in H^{-\frac{1}{2}}(\Gamma),$$

进而

$$\begin{aligned}\|W\nu\|_{H^1(\Omega)} &= \sup_{0 \neq f \in \tilde{H}^{-1}(\Omega)} \frac{\langle W\nu, f \rangle_\Omega}{\|f\|_{\tilde{H}^{-1}(\Omega)}} \\ &= \sup_{0 \neq f \in \tilde{H}^{-1}(\Omega)} \frac{\langle \nu, \gamma_1^{\text{int}} \tilde{N}_0 f \rangle_\Gamma}{\|f\|_{\tilde{H}^{-1}(\Omega)}} \\ &\leq c \|\nu\|_{H^{\frac{1}{2}}(\Gamma)}.\end{aligned}$$

□

由此可见，双层位势(14.191)定义了一个有界的线性算子  $W : H^{\frac{1}{2}}(\Gamma) \mapsto H^1(\Omega)$ 。

类似地，将内界迹算子  $\gamma_0^{\text{int}}$  作用于双层位势算子  $u = W\nu \in H^1(\Omega), \nu \in H^{\frac{1}{2}}(\Gamma)$ ，我们得到一个有界线性算子

$$\gamma_0^{\text{int}} W : H^{\frac{1}{2}}(\Gamma) \mapsto H^{\frac{1}{2}}(\Gamma),$$

满足

$$\|\gamma_0^{\text{int}} W\nu\|_{H^{\frac{1}{2}}(\Gamma)} \leq c \|\nu\|_{H^{\frac{1}{2}}(\Gamma)}, \quad \nu \in H^{\frac{1}{2}}(\Gamma).$$

**Lemma 14.25** (双层位势的跃动关系). 对于  $\nu \in H^{\frac{1}{2}}(\Gamma)$ ，可以将  $\gamma_0^{\text{int}}(W\nu)(x)$  的表现式写为

$$\gamma_0^{\text{int}}(W\nu)(x) = [-1 + \sigma(x)] \nu(x) + (K\nu)(x), \quad x \in \Gamma, \quad (14.192)$$

其中  $\sigma(x), (K\nu)(x)$  分别由伴随双层位势算子定义(14.188)(14.187)而来

$$\sigma(x) := \lim_{\varepsilon \rightarrow 0} \frac{1}{2(d-1)\pi} \frac{1}{\varepsilon^{d-1}} \int_{y \in \Omega: |y-x|=\varepsilon} ds_y, \quad (14.193)$$

$$(K\nu)(x) := \lim_{\varepsilon \rightarrow 0} \int_{y \in \Omega: |y-x| \geq \varepsilon} [\gamma_{1,y}^{\text{int}} U^*(x, y)] \nu(y) ds_y, \quad x \in \Gamma. \quad (14.194)$$

证明. 设任一给定的  $\varepsilon > 0$ ，现在要证明算子满足关系

$$(K_\varepsilon \nu) = \int_{y \in \Gamma: |y-x| \geq \varepsilon} [\gamma_{1,y}^{\text{int}} U^*(x, y)] \nu(y) ds_y.$$

来考虑极限值  $\Omega \ni \tilde{x} \rightarrow x \in \Gamma$  的情况。此时需假设  $|\tilde{x} - x| < \varepsilon$ ，根据

$$\begin{aligned}(W\nu)(\tilde{x}) - (K_\varepsilon \nu)(x) &= \int_{y \in \Gamma: |y-x| \geq \varepsilon} [\gamma_{1,y}^{\text{int}} U^*(\tilde{x}, y) - \gamma_{1,y}^{\text{int}} U^*(x, y)] \nu(y) ds_y + \int_{y \in \Gamma: |y-x| < \varepsilon} [\gamma_{1,y}^{\text{int}} U^*(\tilde{x}, y)] \nu(y) ds_y \\ &= \underbrace{\int_{y \in \Gamma: |y-x| \geq \varepsilon} [\gamma_{1,y}^{\text{int}} U^*(\tilde{x}, y) - \gamma_{1,y}^{\text{int}} U^*(x, y)] \nu(y) ds_y}_{=: \mathcal{A}} \\ &\quad + \underbrace{\int_{y \in \Gamma: |y-x| < \varepsilon} [\gamma_{1,y}^{\text{int}} U^*(\tilde{x}, y)] [\nu(y) - \nu(x)] ds_y}_{=: \mathcal{B}} \\ &\quad + \nu(x) \underbrace{\int_{y \in \Gamma: |y-x| < \varepsilon} \gamma_{1,y}^{\text{int}} U^*(\tilde{x}, y) ds_y}_{=: \mathcal{C}}.\end{aligned}$$

1.  $\mathcal{A} \Rightarrow$

$$\lim_{\Omega \ni \tilde{x} \rightarrow x \in \Gamma} \mathcal{A} = \lim_{\Omega \ni \tilde{x} \rightarrow x \in \Gamma} \int_{y \in \Gamma: |y-x| \geq \varepsilon} [\gamma_{1,y}^{\text{int}} U^*(\tilde{x}, y) - \gamma_{1,y}^{\text{int}} U^*(x, y)] \nu(y) ds_y = 0, \quad \forall \varepsilon > 0.$$

2.  $\mathcal{B} \Rightarrow$

$$\begin{aligned} |\mathcal{B}| &= \left| \int_{y \in \Gamma: |y-x| < \varepsilon} [\gamma_{1,y}^{\text{int}} U^*(\tilde{x}, y)] [\nu(y) - \nu(x)] ds_y \right| \\ &\leq \sup_{y \in \Gamma: |y-x| < \varepsilon} |\nu(x) - \nu(y)| \underbrace{\int_{y \in \Gamma: |y-x| < \varepsilon} |\gamma_{1,y}^{\text{int}} U^*(\tilde{x}, y)| ds_y}_{=: \mathcal{B}_1} \end{aligned}$$

由于

$$\int_{\Gamma} |\gamma_{1,y}^{\text{int}} U^*(\tilde{x}, y)| ds_y \leq M, \quad \tilde{x} \in \Omega,$$

因此随着  $\varepsilon \rightarrow 0$ ,  $\mathcal{B}_1 \rightarrow 0$ ,  $\mathcal{B} \rightarrow 0$ 。

3. 来考虑

$$B_{\varepsilon} = \{y \in \Omega : |y-x| < \varepsilon\},$$

那么我们有

$$\mathcal{C} = \underbrace{\int_{\partial B_{\varepsilon}(x)} \gamma_{1,y}^{\text{int}} U^*(\tilde{x}, y) ds_y}_{=: \mathcal{C}_1} - \underbrace{\int_{y \in \Omega: |y-x| = \varepsilon} \gamma_{1,y}^{\text{int}} U^*(\tilde{x}, y) ds_y}_{=: \mathcal{C}_2}.$$

(a) 利用  $u(x) = 1$  时的泊松方程表现式(14.159), 以及  $\tilde{x} \in B_{\varepsilon}(x)$ , 我们有<sup>15</sup>

$$\mathcal{C}_1 = \int_{\partial B_{\varepsilon}(x)} \gamma_{1,y}^{\text{int}} U^*(\tilde{x}, y) ds_y = -1.$$

(b) 设  $\underline{n}(y) = \frac{y-x}{\varepsilon}$ ,

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \lim_{\Omega \ni \tilde{x} \rightarrow x \in \Gamma} \mathcal{C}_2 &= \lim_{\varepsilon \rightarrow 0} \int_{y \in \Omega: |y-x| = \varepsilon} \gamma_{1,y}^{\text{int}} U^*(\tilde{x}, y) ds_y \\ &= - \lim_{\varepsilon \rightarrow 0} \frac{1}{2(d-1)\pi} \int_{y \in \Omega: |y-x| = \varepsilon} \frac{(n_y, y-x)}{|x-y|^d} ds_y \\ &= - \lim_{\varepsilon \rightarrow 0} \frac{1}{2(d-1)\pi} \frac{1}{\varepsilon^{d-1}} \int_{y \in \Omega: |y-x| = \varepsilon} ds_y \\ &= -\sigma(x). \end{aligned}$$

由此可证。  $\square$

类似地, 我们有外界迹

$$\gamma_0^{\text{ext}}(W\nu)(x) = \sigma(x)\nu(x) + (K\nu)(x), \quad x \in \Gamma. \quad (14.195)$$

结合(14.195)和(14.192), 可得双层位势的跃动关系

$$[\gamma_0 W\nu] := \gamma_0^{\text{ext}}(W\nu)(x) - \gamma_0^{\text{int}}(W\nu)(x) = \nu(x), \quad x \in \Gamma. \quad (14.196)$$

**Lemma 14.26.** 双层位势  $u(x) = (W\nu)(x)$  的共形导数  $\gamma_1 W\nu$ , 其跃动关系可以表示为

$$[\gamma_1 W\nu] := \gamma_1^{\text{ext}}(W\nu)(x) - \gamma_1^{\text{int}}(W\nu)(x) = 0, \quad x \in \Gamma. \quad (14.197)$$

<sup>15</sup>cf. 第14.6.5节。

证明. 1. 一方面, 对于双层位势  $u(x) = (W\nu)(x)$ ,  $x \in \mathbb{R}^n$ ,  $\varphi \in C_0^\infty(\mathbb{R}^d)$  我们有

$$\begin{aligned}\int_{\mathbb{R}^d} [-\Delta u(x)] \varphi(x) dx &= \int_{\mathbb{R}^d} -\Delta_x \int_\Gamma U^*(x, y) w(y) ds_y \varphi(x) dx \\ &= \int_\Gamma w(y) \gamma_{1,y}^{\text{int}} \int_{\mathbb{R}^d} -\Delta_x U^*(x, y) \varphi(x) dx ds_y \\ &= \int_\Gamma w(y) \gamma_{1,y}^{\text{int}} \int_{\mathbb{R}^d} \delta_0(x - y) \varphi(x) dx ds_y = 0.\end{aligned}$$

2. 另一方面, 根据格林第一恒等式(14.6)我们有

$$\begin{aligned}a_{\mathbb{R}^d}(u, \varphi) &= a_\Omega(u, \varphi) + a_{\Omega^c}(u, \varphi) \\ &= \int_\Gamma \gamma_1^{\text{int}} u(x) \gamma_0^{\text{int}} \varphi(x) ds_x - \int_\Gamma \gamma_1^{\text{ext}} u(x) \gamma_0^{\text{ext}} \varphi(x) ds_x.\end{aligned}$$

3. 两式联立  $\Rightarrow$

$$\int_\Gamma \gamma_1^{\text{int}} u(x) \gamma_0^{\text{int}} \varphi(x) ds_x = \int_\Gamma \gamma_1^{\text{ext}} u(x) \gamma_0^{\text{ext}} \varphi(x) ds_x.$$

对  $\|\cdot\|_{H^1(\mathbb{R}^d)}$  取  $C_0^\infty(\mathbb{R}^d)$  的闭包, 我们可得  $\gamma_0^{\text{int}} \varphi = \gamma_0^{\text{ext}} \varphi$ , 进而证得

$$\gamma_1^{\text{ext}} u - \gamma_1^{\text{int}} u = 0.$$

□

### 14.6.5 超奇异边界积分算子

由第14.6.4节关于双层位势的介绍可见, 给定密度方程  $\nu \in H^{\frac{1}{2}}(\Gamma)$ , 它的双层位势  $(W\nu)$  的余法向导数 (conormal derivative)  $\gamma_1^{\text{int}}(W\nu)(x)$  定义了一个有界算子

$$\gamma_1^{\text{int}} W : H^{\frac{1}{2}}(\Gamma) \mapsto H^{-\frac{1}{2}}(\Gamma),$$

可表示为

$$(D\nu)(x) := -\gamma_1^{\text{int}}(W\nu)(x) = -\lim_{\Omega \ni \tilde{x} \rightarrow x \in \Gamma} n_x \nabla_{\tilde{x}} (W\nu)(\tilde{x}), \quad x \in \Gamma, \quad (14.198)$$

满足有界性

$$\|D\nu\|_{H^{-\frac{1}{2}}(\Gamma)} \leq c_2^D \|\nu\|_{H^{\frac{1}{2}}(\Gamma)}, \quad \nu \in H^{\frac{1}{2}}(\Gamma). \quad (14.199)$$

具体说来, 由(14.198)可见

1. 对于  $d = 2$  的二维系统, 双层位势表示为

$$(W\nu)(\tilde{x}) = \frac{1}{2\pi} \lim_{\varepsilon \rightarrow 0} \int_{y \in \Gamma: |y-x| \geq \varepsilon} \frac{(\tilde{x} - y, n_y)}{|x - y|^2} \nu(y) ds_y, \quad \tilde{x} \in \Omega. \quad (14.200)$$

对于任一给定的  $\varepsilon > 0$ , 我们取极限值  $\Omega \ni \tilde{x} \rightarrow x \in \Gamma$ , 将(14.200)对  $\tilde{x}$  求偏导, 代入(14.198)得

$$(D_\varepsilon \nu)(x) = \frac{1}{2\pi} \lim_{\varepsilon \rightarrow 0} \int_{y \in \Gamma: |y-x| \geq \varepsilon} \left[ -\frac{(n_x, n_y)}{|x - y|^2} + 2 \frac{(x - y, n_x)(x - y, n_y)}{|x - y|^4} \right] \nu(y) ds_y. \quad (14.201)$$

2. 类似地, 对于  $d = 3$  的三维系统, 可得

$$(D_\varepsilon \nu)(x) = \frac{1}{4\pi} \int_{y \in \Gamma: |y-x| \geq \varepsilon} \left[ -\frac{(n_x, n_y)}{|x - y|^3} + 3 \frac{(y - x, n_x)(y - x, n_y)}{|x - y|^5} \right] \nu(y) ds_y. \quad (14.202)$$

然而，无论是二维系统(14.201)还是三维系统(14.202)的情况，极限  $\varepsilon \rightarrow 0, x \in \Gamma$  下的积分都不存在柯西主值 (Cauchy principal value)。我们将此类算子  $D$  称为超奇异边界积分算子 (hypersingular boundary integral operator)，作为柯西积分 (Cauchy integral) 的一个扩展<sup>16</sup>。

在不存在柯西主值的情况下，为了求得  $D$  的线性表现式，往往需引入一些正则设定，如将  $u_0(x) \equiv 1$  引入泊松方程表现式(14.159)中可得 (更详细介绍见419页第14.6.8节。)

$$1 = - \int_{\Gamma} \gamma_{1,y}^{\text{int}} U^*(\tilde{x}, y) ds_y, \quad \tilde{x} \in \Omega,$$

$$\hookrightarrow \nabla_x (W u_0)(\tilde{x}) = 0, \quad \tilde{x} \in \Omega,$$

$$\hookrightarrow (D u_0)(x) = 0, \quad x \in \Gamma. \quad (14.203)$$

进而可以将(14.198)写为

$$\begin{aligned} (D\nu)(x) &= - \lim_{\Omega \ni \tilde{x} \rightarrow x \in \Gamma} n_x \nabla_{\tilde{x}} (W\nu)(\tilde{x}) \\ &= - \lim_{\Omega \ni \tilde{x} \rightarrow x \in \Gamma} n_x \nabla_{\tilde{x}} \int_{\Gamma} \gamma_{1,y}^{\text{int}} U^*(\tilde{x}, y) [\nu(y) - \nu(x)] ds_y, \quad x \in \Gamma. \end{aligned}$$

根据上式，如果密度方程是连续的，我们就可以将超奇异边界积分算子  $D$  的表现式写为

$$(D\nu)(x) = - \int_{\Gamma} \gamma_{1,x}^{\text{int}} \gamma_{1,y}^{\text{int}} U^*(x, y) [\nu(y) - \nu(x)] ds_y, \quad x \in \Gamma. \quad (14.204)$$

下面分二维、三维空间两个情况继续展开分析。

## 二维空间

以  $d = 2$  为例。假定  $\Gamma = \partial\Omega$  是分段平滑的，满足

$$\Gamma = \cup_{k=1}^p \Gamma_k,$$

每一分段  $\Gamma_k$  都可由局部参数化表示

$$\Gamma_k := y = y(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix}, \quad t \in (t_k, t_{k+1}), \quad (14.205)$$

其中  $y_i(t), i = 1, 2$  为连续可导方程。

结合(14.204)、(14.205)我们有

$$\begin{aligned} ds_y &= \sqrt{[y'_1(t)]^2 + [y'_2(t)]^2} dt \\ n(y) &= \frac{1}{\sqrt{[y'_1(t)]^2 + [y'_2(t)]^2}} \begin{pmatrix} y'_2(t) \\ -y'_1(t) \end{pmatrix}, \quad y \in \Gamma_k, \end{aligned}$$

其中  $n(y)$  表示外部法向量 (exterior normal vector)。

二维系统中对于  $x \in \mathbb{R}^2$  的情况，某一标量方程  $\tilde{\nu}(x)$  的旋转 (rotation, 或称 curl) 可以定义为

$$\underline{\text{curl}} \tilde{\nu}(x) := \begin{pmatrix} \frac{\partial}{\partial x_2} \tilde{\nu}(x) \\ -\frac{\partial}{\partial x_1} \tilde{\nu}(x) \end{pmatrix},$$

<sup>16</sup>关于超奇异边界积分方程 (hypersingular boundary integral equation)，一个简要介绍可见 (Mason and Handscomb, 2003, Ch.9)，更多介绍可见Lifanov et al. (2003); Ang (2013)，数值近似算法的探讨可见Guiggiani et al. (1992)。

现在来看  $\tilde{\nu}(x)$  的计算：给定某一密度方程  $\nu(x), x \in \Gamma_k$ ，则我们可以考虑一个在  $\Gamma_k$  的适宜邻域上的延拓  $\tilde{\nu}$

$$\tilde{\nu}(\tilde{x}) = \nu(x), \quad \tilde{x} = x + (\tilde{x} - x, \underline{n}(x)) \underline{n}(x). \quad (14.206)$$

也即，对于  $x \in \Gamma_k$ ，我们可以定义  $\nu(x)$  的旋转  $\operatorname{curl}_{\Gamma_k} \nu(x)$  为

$$\begin{aligned} \operatorname{curl}_{\Gamma_k} \nu(x) &:= \underline{n}(x) \operatorname{curl} \tilde{\nu}(x) \\ &= n_1(x) \frac{\partial}{\partial x_2} \tilde{\nu}(x) - n_2(x) \frac{\partial}{\partial x_1} \tilde{\nu}(x), \quad x \in \Gamma_k. \end{aligned}$$

根据上式我们有

$$\begin{aligned} \int_{\Gamma_k} \operatorname{curl}_{\Gamma_k} \nu(y) ds_y &= \int_{\Gamma_k} \left[ n_1(y) \frac{\partial}{\partial y_2} \tilde{\nu}(y) - n_2(y) \frac{\partial}{\partial y_1} \tilde{\nu}(y) \right] ds_y \\ &= \int_{t_k}^{t_{k+1}} \left[ y'_2(t) \frac{\partial}{\partial y_2} \nu(y(t)) + y'_1(t) \frac{\partial}{\partial y_1} \nu(y(t)) \right] dt \\ &= \int_{t_k}^{t_{k+1}} \frac{d}{dt} \nu(y(t)) dt, \end{aligned}$$

通过该式可见， $\operatorname{curl}_{\Gamma_k} \nu$  与预先选定的延拓  $\tilde{\nu}$  无关。

**Lemma 14.27.** 设一个开放的有界区间  $\Gamma_k$ ， $\Gamma_k$  由局部参数式(14.205)所定义。对应的连续可导方程为  $y_i(t), i = 1, 2$ 。

如果  $\nu$  和  $w$  连续可导，则我们有分部积分式

$$\int_{\Gamma_k} \nu(y) \operatorname{curl}_{\Gamma_k} w(y) ds_y = - \int_{\Gamma_k} \operatorname{curl}_{\Gamma_k} \nu(y) w(y) ds_y + \nu(y(t)) w(y(t)) \Big|_{t_k}^{t_{k+1}}.$$

证明.

$$\begin{aligned} \int_{\Gamma_k} \operatorname{curl}_{\Gamma_k} [\nu(y) w(y)] ds_y &= \int_{t_k}^{t_{k+1}} \frac{d}{dt} [\nu(y(t)) w(y(t))] dt \\ &= [\nu(y(t)) w(y(t))] \Big|_{t=t_k}^{t=t_{k+1}}. \end{aligned}$$

□

若  $\nu$  是沿着闭曲线 (closed curve)  $\Gamma$  而定义的<sup>17</sup>，则我们可将  $\nu$  的旋转  $\operatorname{curl}_{\Gamma} \nu(x)$  定义为

$$\operatorname{curl}_{\Gamma} \nu(x) := \operatorname{curl}_{\Gamma_k} \nu(x), \quad x \in \Gamma_k, k = 1, \dots, p,$$

那么根据 Lemma 14.27，我们有

**Corollary 14.7.** 设  $\Gamma$  是一个分段平滑的闭曲线，如果方程  $\nu$  和  $w$  均分段连续可微，则我们有

$$\int_{\Gamma} \nu(y) \operatorname{curl}_{\Gamma} w(y) ds_y = - \int_{\Gamma} \operatorname{curl}_{\Gamma} \nu(y) w(y) ds_y + \sum_{k=1}^p \nu(y(t)) w(y(t)) \Big|_{t_k}^{t_{k+1}}, \quad (14.207)$$

进一步，如果  $\nu$  和  $w$  全局连续，则

$$\int_{\Gamma} \nu(y) \operatorname{curl}_{\Gamma} w(y) ds_y = - \int_{\Gamma} \operatorname{curl} \nu(y) w(y) ds_y. \quad (14.208)$$

证明. 略。

□

<sup>17</sup>闭曲线是指这样的一种曲线，其始点同时也是其终点。

利用分部积分法，我们可以将超奇异边界积分算子  $D$  的双线性形式，改写为单层位势算子  $V$  的双线性形式。在二维系统中，这种改写可以表示如下。

**Theorem 14.33.** 设  $\Gamma$  是一个分段平滑的封闭曲线。 $u, v \in \Gamma$  是全局连续方程。进一步，设  $u, v$  在  $\Gamma$  的某一段  $\Gamma_k$  上连续可微。那么我们有

$$\langle Du, v \rangle_{\Gamma} = -\frac{1}{2\pi} \int_{\Gamma} \operatorname{curl}_{\Gamma} v(x) \int_{\Gamma} \log|x-y| \operatorname{curl}_{\Gamma} u(y) ds_y ds_x. \quad (14.209)$$

证明。根据定义式(14.198)，超奇异边界积分算子  $D$  可以定义为双层位势算子  $W$  的法导数  $\gamma_1^{\text{int}} W$  的负数。对于  $\tilde{x} \in \Omega$ ，定义

$$w(\tilde{x}) := (Wu)(\tilde{x}) = -\frac{1}{2\pi} \int_{\Gamma} u(y) \frac{\partial}{\partial n_y} \log|\tilde{x}-y| ds_y, \quad \tilde{x} \in \Omega.$$

一方面  $\tilde{x} \in \Omega, y \in \Gamma \Rightarrow \tilde{x} \neq y$ ，另一方面

$$\frac{\partial}{\partial y_i} \log|\tilde{x}-y| = \frac{y_i - \tilde{x}_i}{|\tilde{x}-y|^2} = -\frac{\tilde{x}_i - y_i}{|\tilde{x}-y|^2} = -\frac{\partial}{\partial \tilde{x}_i} \log|\tilde{x}-y|,$$

结合这两方面可得

$$\frac{\partial}{\partial \tilde{x}_i} \left( \frac{\partial}{\partial n_y} \log|\tilde{x}-y| \right) = -n(y) \nabla_y \left( \frac{\partial}{\partial y_i} \log|\tilde{x}-y| \right).$$

由于  $\Delta_y \log|\tilde{x}-y| = 0$ ，对于  $y \neq \tilde{x}$ ，则我们有以下两个等式

$$\begin{aligned} \operatorname{curl}_{\Gamma,y} \left( \frac{\partial}{\partial y_1} \log|\tilde{x}-y| \right) &= n_1(y) \frac{\partial}{\partial y_2} \frac{\partial}{\partial y_1} \log|\tilde{x}-y| - n_2(y) \frac{\partial}{\partial y_1} \frac{\partial}{\partial y_1} \log|\tilde{x}-y| \\ &= n_1(y) \frac{\partial}{\partial y_2} \frac{\partial}{\partial y_1} \log|\tilde{x}-y| + n_2(y) \frac{\partial}{\partial y_2} \frac{\partial}{\partial y_2} \log|\tilde{x}-y| \\ &= \underline{n}(y) \nabla_y \left( \frac{\partial}{\partial y_2} \log|\tilde{x}-y| \right), \end{aligned}$$

$$\operatorname{curl}_{\Gamma,y} \left( \frac{\partial}{\partial y_2} \log|\tilde{x}-y| \right) = \underline{n}(y) \nabla_y \left( \frac{\partial}{\partial y_1} \log|\tilde{x}-y| \right).$$

因此，可以将全局连续方程  $u$  的双层位势  $w = (Wu)$  的偏导数写为如下两个分部积分形式

$$\begin{aligned} \frac{\partial}{\partial \tilde{x}_1} w(\tilde{x}) &= -\frac{1}{2\pi} \int_{\Gamma} u(y) \frac{\partial}{\partial \tilde{x}} \frac{\partial}{\partial n_y} \log|\tilde{x}-y| ds_y \\ &= \frac{1}{2\pi} \int_{\Gamma} u(y) n(y) \nabla_y \left( \frac{\partial}{\partial y_1} \log|\tilde{x}-y| \right) ds_y \\ &= -\frac{1}{2\pi} \int_{\Gamma} u(y) \operatorname{curl}_{\Gamma,y} \left( \frac{\partial}{\partial y_2} \log|\tilde{x}-y| \right) ds_y \\ &= \frac{1}{2\pi} \int_{\Gamma} \operatorname{curl}_{\Gamma} u(y) \frac{\partial}{\partial y_2} \log|\tilde{x}-y| ds_y, \end{aligned}$$

$$\frac{\partial}{\partial \tilde{x}_2} w(\tilde{x}) = -\frac{1}{2\pi} \int_{\Gamma} \operatorname{curl}_{\Gamma} u(y) \frac{\partial}{\partial y_1} \log|\tilde{x}-y| ds_y.$$

根据定义，双层位势  $w = Wu$  的法导数

$$\begin{aligned} \underline{n}(\tilde{x}) \nabla_{\tilde{x}} w(\tilde{x}) &= \frac{1}{2\pi} \int_{\Gamma} \operatorname{curl}_{\Gamma} u(y) \left[ n_1(x) \frac{\partial}{\partial y_2} \log|\tilde{x}-y| - n_2(x) \frac{\partial}{\partial y_1} \log|\tilde{x}-y| \right] ds_y \\ &= \frac{1}{2\pi} \lim_{\varepsilon \rightarrow 0} \int_{y \in \Gamma: |y-x| \geq \varepsilon} \operatorname{curl}_{\Gamma} u(y) \left[ n_1(x) \frac{\partial}{\partial y_2} \log|\tilde{x}-y| - n_2(x) \frac{\partial}{\partial y_1} \log|\tilde{x}-y| \right] ds_y \end{aligned}$$

取极限  $\Omega \ni \tilde{x} \rightarrow x \in \Gamma$ , 上式变为

$$\begin{aligned} & \frac{\partial}{\partial n_x} w(x) \\ &= \frac{1}{2\pi} \lim_{\varepsilon \rightarrow 0} \int_{y \in \Gamma: |y-x| \geq \varepsilon} \operatorname{curl}_\Gamma u(y) \left[ n_1(x) \frac{\partial}{\partial y_2} \log |x-y| - n_2(x) \frac{\partial}{\partial y_1} \log |x-y| \right] ds_y \\ &= -\frac{1}{2\pi} \lim_{\varepsilon \rightarrow 0} \int_{y \in \Gamma: |y-x| \geq \varepsilon} \operatorname{curl}_\Gamma u(y) \left[ n_1(x) \frac{\partial}{\partial x_2} \log |x-y| - n_2(x) \frac{\partial}{\partial x_1} \log |x-y| \right] ds_y \\ &= -\frac{1}{2\pi} \lim_{\varepsilon \rightarrow 0} \int_{y \in \Gamma: |y-x| \geq \varepsilon} \operatorname{curl}_\Gamma u(y) \operatorname{curl}_{\Gamma,x} \log |x-y| ds_y. \end{aligned}$$

进而

$$\begin{aligned} & \int_\Gamma \nu(x) \frac{\partial}{\partial n_x} w(x) ds_x \\ &= -\frac{1}{2\pi} \int_{x \in \Gamma} \nu(x) \lim_{\varepsilon \rightarrow 0} \int_{y \in \Gamma: |y-x| \geq \varepsilon} \operatorname{curl}_\Gamma u(y) \operatorname{curl}_{\Gamma,x} \log |x-y| ds_y ds_x \\ &= -\frac{1}{2\pi} \int_{y \in \Gamma} \operatorname{curl}_\Gamma u(y) \lim_{\varepsilon \rightarrow 0} \int_{x \in \Gamma: |x-y| \geq \varepsilon} \nu(x) \operatorname{curl}_{\Gamma,x} \log |x-y| ds_x ds_y. \end{aligned}$$

由此可证。  $\square$

### 三维空间

超奇异积分算子  $D$  的分部积分表现式也可以用于三维系统  $d = 3$  的分析中 (Dautray and Lions, 1990)。

设一个分段平滑表面  $\Gamma = \cup_{k=1}^p \Gamma_k$ , 其中每一个分段  $\Gamma_k$  都可以用参数表示如下

$$y \in \Gamma_k : y(s, t) = \begin{pmatrix} y_1(s, t) \\ y_2(s, t) \\ y_3(s, t) \end{pmatrix}, \quad (s, t) \in \tau,$$

$\tau$  指某个参考元 (reference element)。

$\mathbb{R}^3$  中某个向量方程  $\underline{\nu}$  的旋转可以定义为

$$\underline{\operatorname{curl}}\nu(x) := \nabla \times \underline{\nu}(x), \quad x \in \mathbb{R}^3.$$

如果  $u$  是个  $\Gamma_k$  上的标量方程, 其表面旋转 (surface curl) 可定义为

$$\underline{\operatorname{curl}}_{\Gamma_k} u(x) := \underline{n}(x) \times \nabla \tilde{u}(x), \quad x \in \Gamma_k,$$

其中  $\tilde{u} \in \Gamma_k$  是一个给定的  $u \in \Gamma_k$  在  $\Gamma_k$  的三维邻域上的适宜延拓。

此外我们有

$$\operatorname{curl}_{\Gamma_k} \underline{\nu}(x) := \underline{n}(x) \times \underline{\operatorname{curl}}\tilde{\nu}(x), \quad x \in \Gamma_k.$$

**Lemma 14.28.** 设  $\Gamma$  为一个  $\mathbb{R}^3$  中的分段平滑封闭利普希茨表面。假定每个表面分段  $\Gamma_k$  都是平滑的, 对应分段平滑边界曲线  $\partial\Gamma_k$ 。设  $u$  和  $\nu$  均在  $\Gamma$  中全局连续, 并且在每个分段  $\Gamma_k$  中都局部有节且平滑。那么由分部积分可得

$$\int_\Gamma \underline{\operatorname{curl}}\Gamma u(x) \underline{\nu}(x) ds_x = - \int_\Gamma u(x) \operatorname{curl}_\Gamma \underline{\nu}(x) ds_x. \quad (14.210)$$

证明. 根据乘积法则可得

$$\nabla \times [\tilde{u}(x) \underline{\nu}(x)] \equiv \nabla \tilde{u}(x) \times \underline{\nu}(x) + \tilde{u}(x) [\nabla \times \underline{\nu}(x)],$$

LHS (14.210)  $\Rightarrow$

$$\begin{aligned} \int_{\Gamma_k} \underline{\operatorname{curl}}_{\Gamma_k} u(x) \underline{\nu}(x) ds_x &= \int_{\Gamma_k} [\underline{n}(x) \times \nabla \tilde{u}(x)] \nu(x) ds_x \\ &= \int_{\Gamma_k} [\nabla \tilde{u}(x) \times \underline{\nu}(x)] n(x) ds_x \\ &= \int_{\Gamma_k} \{[\nabla \times (\tilde{u}(x) \times \underline{\nu}(x))] - [\tilde{u}(x) (\nabla \times \underline{\nu}(x))] \} n(x) ds_x \\ &= \underbrace{\int_{\partial \Gamma_k} u(x) \underline{\nu}(x) \underline{t}(x) d\sigma}_{=:0} - \int_{\Gamma_k} u(x) \operatorname{curl}_{\Gamma_k} \underline{\nu}(x) ds_x \\ &= - \int_{\Gamma_k} u(x) \operatorname{curl}_{\Gamma_k} \underline{\nu}(x) ds_x \end{aligned}$$

□

**Theorem 14.34** (超奇异积分算子的双线性表现式). 设一个分段平滑的封闭表面  $\Gamma$ 。在  $\Gamma$  中定义全局连续方程  $u$  和  $\nu$ , 且在  $\Gamma_k$  中可微。那么超奇异积分算子  $D$  的双线性表现式为

$$\langle Du, \nu \rangle_{\Gamma} = \frac{1}{4\pi} \int_{\Gamma} \int_{\Gamma} \frac{\underline{\operatorname{curl}}_{\Gamma} u(y) \underline{\operatorname{curl}}_{\Gamma} \nu(x)}{|x - y|} ds_x ds_y. \quad (14.211)$$

证明. 与二维系统的证明过程 (Theorem 14.33) 类似。根据定义式(14.198), 超奇异边界积分算子  $D$  可以定义为双层位势算子  $W$  的法导数  $\gamma_1^{\text{int}} W$  的负数。对于  $\tilde{x} \in \Omega$ , 定义

$$w(\tilde{x}) := -\frac{1}{4\pi} \int_{\Gamma} u(y) \underbrace{\frac{\partial}{\partial n_y} \frac{1}{|\tilde{x} - y|}}_{\text{核方程}} ds_y.$$

1. 求核方程的偏导数。根据

$$\frac{\partial}{\partial y_i} \frac{1}{|\tilde{x} - y|} = \frac{\tilde{x} - y_i}{|\tilde{x} - y|^3} = -\frac{y_i - \tilde{x}}{|\tilde{x} - y|^3} = -\frac{\partial}{\partial x_i} \frac{1}{|\tilde{x} - y|},$$

则

$$\frac{\partial}{\partial \tilde{x}_i} \left( \frac{\partial}{\partial y_i} \frac{1}{|\tilde{x} - y|} \right) = -\underline{n}(y) \nabla_y \left( \frac{\partial}{\partial y_i} \frac{1}{|\tilde{x} - y|} \right).$$

2. 设  $\underline{e}_i$  为  $\mathbb{R}^3$  中的第  $i$ -th 个单位向量 (unit vector)。由于  $\tilde{x} \neq y$  可得向量乘

$$\begin{aligned} \underline{\operatorname{curl}}_y \left( \underline{e}_i \times \nabla_y \frac{1}{|\tilde{x} - y|} \right) &= \nabla_y \times \left( \underline{e}_i \times \nabla_y \frac{1}{|\tilde{x} - y|} \right) \\ &= \left( \nabla_y \cdot \nabla_y \frac{1}{|\tilde{x} - y|} \right) \underline{e}_i - (\nabla_y \cdot \underline{e}_i) \nabla_y \frac{1}{|\tilde{x} - y|} \\ &= \underbrace{\Delta_y \frac{1}{|\tilde{x} - y|} \underline{e}_i}_{=:0} - \frac{\partial}{\partial y_i} \left( \nabla_y \frac{1}{|\tilde{x} - y|} \right) \\ &= -\frac{\partial}{\partial y_i} \left( \nabla_y \frac{1}{|\tilde{x} - y|} \right). \end{aligned}$$

3. 对双层位势  $w(\tilde{x})$  作关于  $\tilde{x}_i$  的偏微分, 调整微分、积分顺序得

$$\begin{aligned}\frac{\partial}{\partial \tilde{x}_i} w(\tilde{x}) &= -\frac{1}{4\pi} \int_{\Gamma} u(y) \frac{\partial}{\partial \tilde{x}_i} \left( \frac{\partial}{\partial n_y} \frac{1}{|\tilde{x}-y|} \right) ds_y \\ &= \frac{1}{4\pi} \int_{\Gamma} u(y) n_y \cdot \nabla_y \left( \frac{\partial}{\partial y_i} \frac{1}{|\tilde{x}-y|} \right) ds_y \\ &= -\frac{1}{4\pi} \int_{\Gamma} u(y) n(y) \cdot \underline{\text{curl}}_y \left( \underline{e}_i \times \nabla_y \frac{1}{|\tilde{x}-y|} \right) ds_y \\ &= -\frac{1}{4\pi} \int_{\Gamma} u(y) \cdot \underline{\text{curl}}_{\Gamma,y} \left( \underline{e}_i \times \nabla_y \frac{1}{|\tilde{x}-y|} \right) ds_y.\end{aligned}$$

根据 Lemma 14.28, 上式进一步改写为

$$\begin{aligned}\frac{\partial}{\partial \tilde{x}_i} w(\tilde{x}) &= \frac{1}{4\pi} \int_{\Gamma} \underline{\text{curl}}_{\Gamma,y} u(y) \cdot \left( \underline{e}_i \times \nabla_x \frac{1}{|\tilde{x}-y|} \right) ds_y \\ &= -\frac{1}{4\pi} \int_{\Gamma} \underline{e}_i \left( \underline{\text{curl}}_{\Gamma,y} u(y) \times \nabla_y \frac{1}{|\tilde{x}-y|} \right) ds_y.\end{aligned}$$

由此可得双层位势的梯度 (gradient)

$$\begin{aligned}\nabla_{\tilde{x}} w(\tilde{x}) &= -\frac{1}{4\pi} \int_{\Gamma} \left( \underline{\text{curl}}_{\Gamma,y} u(y) \times \nabla_y \frac{1}{|\tilde{x}-y|} \right) ds_y \\ &= \frac{1}{4\pi} \int_{\Gamma} \left( \underline{\text{curl}}_{\Gamma,y} u(y) \times \nabla_{\tilde{x}} \frac{1}{|\tilde{x}-y|} \right) ds_y.\end{aligned}$$

两侧同时乘以法向量 (normal vector)  $\underline{n}(x)$  得

$$\begin{aligned}\underline{n}(x) \cdot \nabla_{\tilde{x}} w(\tilde{x}) &= \frac{1}{4\pi} \int_{\Gamma} \left( \underline{\text{curl}}_{\Gamma,y} u(y) \times \nabla_{\tilde{x}} \frac{1}{|\tilde{x}-y|} \right) \cdot \underline{n}(x) ds_y \\ &= -\frac{1}{4\pi} \int_{\Gamma} \underline{\text{curl}}_{\Gamma,y} u(y) \cdot \left( \underline{n}(x) \times \nabla_{\tilde{x}} \frac{1}{|\tilde{x}-y|} \right) ds_y \\ &= -\frac{1}{4\pi} \lim_{\varepsilon \rightarrow 0} \int_{y \in \Gamma: |y-x| \geq \varepsilon} \underline{\text{curl}}_{\Gamma,y} u(y) \cdot \left( \underline{n}(x) \times \nabla_{\tilde{x}} \frac{1}{|\tilde{x}-y|} \right) ds_y.\end{aligned}$$

取极限  $\Omega \ni \tilde{x} \rightarrow x \in \Gamma$ , 上式变为

$$\begin{aligned}(Du) &= -\frac{1}{4\pi} \lim_{\varepsilon \rightarrow 0} \int_{y \in \Gamma: |y-x| \geq \varepsilon} \underline{\text{curl}}_{\Gamma,y} u(y) \cdot \left( \underline{n}(x) \times \nabla_x \frac{1}{|x-y|} \right) ds_y \\ &= -\frac{1}{4\pi} \lim_{\varepsilon \rightarrow 0} \int_{y \in \Gamma: |y-x| \geq \varepsilon} \underline{\text{curl}}_{\Gamma,y} u(y) \cdot \underline{\text{curl}}_{\Gamma,x} \frac{1}{|x-y|} ds_y\end{aligned}$$

4. 进而, 双线性形式

$$\begin{aligned}\langle Du, \nu \rangle_{\Gamma} &= -\frac{1}{4\pi} \int_{\Gamma} \nu(x) \lim_{\varepsilon \rightarrow 0} \int_{y \in \Gamma: |y-x| \geq \varepsilon} \underline{\text{curl}}_{\Gamma,y} u(y) \cdot \underline{\text{curl}}_{\Gamma,x} \frac{1}{|x-y|} ds_y ds_x \\ &= -\frac{1}{4\pi} \int_{\Gamma} \lim_{\varepsilon \rightarrow 0} \int_{x \in \Gamma: |x-y| \geq \varepsilon} (\nu(x) \underline{\text{curl}}_{\Gamma,y} u(y)) \underline{\text{curl}}_{\Gamma,x} \frac{1}{|x-y|} ds_x ds_y \\ &= \frac{1}{4\pi} \int_{\Gamma} \lim_{\varepsilon \rightarrow 0} \int_{x \in \Gamma: |x-y| \geq \varepsilon} \underbrace{\underline{\text{curl}}_{\Gamma,x} (\nu(x) \underline{\text{curl}}_{\Gamma,y} u(y))}_{\underline{\text{curl}}_{\Gamma,x} \nu(x) \underline{\text{curl}}_{\Gamma,y} u(y)} \frac{1}{|x-y|} ds_x ds_y,\end{aligned}$$

其中

$$\begin{aligned}\underline{\text{curl}}_{\Gamma,x} (\nu(x) \underline{\text{curl}}_{\Gamma,y} u(y)) &= \underline{n}(x) \cdot [\nabla_x \times (\nu(x) \underline{\text{curl}}_{\Gamma,y} u(y))] \\ &= \underline{n}(x) \cdot [\nabla_x \nu(x) \times \underline{\text{curl}}_{\Gamma,y} u(y)] \\ &= [n(x) \times \nabla_x \nu(x)] \cdot \underline{\text{curl}}_{\Gamma,y} u(y) \\ &= \underline{\text{curl}}_{\Gamma,x} \nu(x) \underline{\text{curl}}_{\Gamma,y} u(y).\end{aligned}$$

因此

$$\langle Du, \nu \rangle_{\Gamma} = \frac{1}{4\pi} \int_{\Gamma} \lim_{\varepsilon \rightarrow 0} \int_{x \in \Gamma: |x-y| \geq \varepsilon} \underline{\operatorname{curl}}_{\Gamma,x} \nu(x) \underline{\operatorname{curl}}_{\Gamma,y} \nu(y) \frac{1}{|x-y|} ds_x ds_y$$

对上式取极限  $\varepsilon \rightarrow 0$ , 证毕。

□

#### 14.6.6 边界积分算子之间的关系

回顾一下泊松方程的表现式(14.159)

$$u(\tilde{x}) = \int_{\Omega} U^*(\tilde{x}, y) f(y) dy + \int_{\Gamma} U^*(\tilde{x}, y) \gamma_1^{\text{int}} u(y) ds_y - \int_{\Gamma} U^*(\tilde{x}, y) \gamma_0^{\text{int}} u(y) ds_y, \quad \tilde{x} \in \Omega.$$

取极限  $\Omega \ni \tilde{x} \rightarrow x \in \Gamma$ , 由上式可以得出一系列边界和体积位势的性质。其中需要关注的是两个边界积分方程。第一个是将内界迹算子  $\gamma_0^{\text{int}}$  应用到解  $u(x), x \in \Gamma$  中

$$\gamma_0^{\text{int}} u(x) = (V \gamma_1^{\text{int}} u)(x) + (1 - \sigma(x)) \gamma_0^{\text{int}} u(x) - (K \gamma_0^{\text{int}} u)(x) + N_0 f(x). \quad (14.212)$$

第二个是将共法导数算子  $\gamma_1^{\text{int}}$  应用到解  $u(x), x \in \Gamma$  中

$$\gamma_1^{\text{int}} u(x) = \sigma(x) \gamma_1^{\text{int}} u(x) + (K' \gamma_1^{\text{int}} u)(x) + (D \gamma_0^{\text{int}} u)(x) + N_1 f(x). \quad (14.213)$$

联立(14.212)和(14.213), 构成一个边界积分方程系统

$$\begin{pmatrix} \gamma_0^{\text{int}} u(x) \\ \gamma_1^{\text{int}} u(x) \end{pmatrix} = \underbrace{\begin{pmatrix} (1-\sigma) & V \\ D & \sigma I + K' \end{pmatrix}}_{=: \mathcal{C}} \begin{pmatrix} \gamma_0^{\text{int}} u(x) \\ \gamma_1^{\text{int}} u(x) \end{pmatrix} + \begin{pmatrix} N_0 f(x) \\ N_1 f(x) \end{pmatrix}. \quad (14.214)$$

我们将系数矩阵定义为卡尔德隆投影 (Calderón projection)

$$\mathcal{C} := \begin{pmatrix} (1-\sigma) & V \\ D & \sigma I + K' \end{pmatrix}. \quad (14.215)$$

**Lemma 14.29** (卡尔德隆投影). 式(14.215)算子  $\mathcal{C}$  是一个投影, 满足  $\mathcal{C} = \mathcal{C}^2$ 。

证明. 设任一给定的  $(\psi, \varphi) \in H^{-\frac{1}{2}}(\Gamma) \times H^{\frac{1}{2}}(\Gamma)$ 。那么齐次偏微分方程的解  $u(\tilde{x}), \tilde{x} \in \Omega$  可以表示为

$$u(\tilde{x}) := (V\psi)(\tilde{x}) - (W\varphi)(\tilde{x}), \quad \tilde{x} \in \Omega.$$

由  $u$  的内界迹和共法导数, 对应  $x \in \Gamma$ , 我们有边界位势的性质

$$\begin{aligned} \gamma_0^{\text{int}} &= (V\psi)(x) + [1 - \sigma(x)] \varphi(x) - (K\varphi)(x), \\ \gamma_1^{\text{int}} &= \sigma\psi(x) + (K'\psi)(x) (D\varphi)(x). \end{aligned}$$

不难看出,  $u$  是齐次偏微分方程的解。通过  $[\gamma_0^{\text{int}}, \gamma_1^{\text{int}}], x \in \Gamma$ , 可以生成一组对应的柯西数 (Cauchy data)。这些柯西数构成边界积分方程系统(14.212)-(14.213)的解, 即

$$\begin{aligned} (V\gamma_1^{\text{int}} u)(x) &= (\sigma I + K) \gamma_0^{\text{int}} u(x), \\ (D\gamma_0^{\text{int}} u)(x) &= ((1-\sigma) I - K') \gamma_1^{\text{int}} u(x). \end{aligned} \quad (14.216)$$

上式等价于

$$\begin{pmatrix} \gamma_0^{\text{int}} u(x) \\ \gamma_1^{\text{int}} u(x) \end{pmatrix} = \begin{pmatrix} (1-\sigma) & V \\ D & \sigma I + K' \end{pmatrix} \begin{pmatrix} \gamma_0^{\text{int}} u(x) \\ \gamma_1^{\text{int}} u(x) \end{pmatrix}. \quad (14.217)$$

联系

$$\begin{pmatrix} \gamma_0^{\text{int}} u(x) \\ \gamma_1^{\text{int}} u(x) \end{pmatrix} = \begin{pmatrix} (1-\sigma) & V \\ D & \sigma I + K' \end{pmatrix} \begin{pmatrix} \varphi(x) \\ \psi(x) \end{pmatrix}$$

可见,  $\mathcal{C}$  是一个投影。根据定义, 投影矩阵, 又称幂等矩阵 (idempotent matrix), 满足  $\mathcal{C} = \mathcal{C}^2$  (Seber and Lee, 2003, p.464)。  $\square$

在得到卡尔德隆投影之后, 我们有以下有界积分算子之间的关系。

**Corollary 14.8** (有界积分算子之间的关系). 所有有界积分算子均满足以下关系

$$VD = (\sigma I + K) [(1-\sigma)I - K], \quad (14.218)$$

$$DV = (\sigma I + K') [(1-\sigma)I - K'], \quad (14.219)$$

$$VK' = KV, \quad (14.220)$$

$$K'D = DK. \quad (14.221)$$

由(14.220)可见, 双层位势算子  $K$  是对称的——并不是自伴随的对称, 而是随着单层位势算子  $V$  的对称。

此外, 边界积分方程系统中, 假定单层位势  $V$  是可逆的 (更多讨论见), 则我们也可得到牛顿位势  $N_1 f$  的表现式如下

section 6.1.1

**Lemma 14.30** (牛顿位势算子与其他算子的关系). 牛顿位势算子  $(N_1 f)(x), x \in \Gamma$  的表现式为

$$(N_1 f)(x) = ((\sigma - 1)I + K') V^{-1} (N_0 f)(x) \quad (14.222)$$

证明. 由边界积分方程系统(14.214)第一行可知, 假设  $V$  是可逆的

$$\gamma_1^{\text{int}} u(x) = V^{-1} (\sigma I + K') \gamma_0^{\text{int}} u(x) - V^{-1} (N_0 f)(x), \quad x \in \Gamma.$$

代入第二行, 可得

$$\begin{aligned} \gamma_1^{\text{int}} u(x) &= (D \gamma_0^{\text{int}} u)(x) + (\sigma I + K') \gamma_1^{\text{int}} u(x) + (N_1 f)(x) \\ &= (D \gamma_0^{\text{int}} u)(x) + (\sigma I + K') [V^{-1} (\sigma I + K) \gamma_0^{\text{int}} u(x) - V^{-1} (N_0 f)(x)] + (N_1 f)(x) \\ &= [D + (\sigma I + K') V^{-1} (\sigma I + K)] \gamma_0^{\text{int}} u(x) - (\sigma I + K') V^{-1} (N_0 f)(x) + (N_1 f)(x), \quad x \in \Gamma. \end{aligned}$$

整理可得

$$-V^{-1} (N_0 f)(x) = -(\sigma I + K') V^{-1} (N_0 f)(x) + (N_1 f)(x),$$

证毕。  $\square$

#### 14.6.7 单层位势算子的椭圆性和可逆性

单层位势算子  $V : H^{-\frac{1}{2}}(\Gamma) \mapsto H^{\frac{1}{2}}(\Gamma)$  的可逆性可由拉克斯—密格拉蒙定理 Theorem 14.14 证得。在此基础上, 进一步讨论  $V$  的  $H^{\frac{1}{2}}(\Gamma)$ -椭圆特性。

回顾, 已知方程  $u(x) = (\tilde{V} w)(x), x \in \Gamma$  是以下内界狄利克雷边界值问题的解

$$-\Delta u(x) = 0, \quad x \in \Omega,$$

$$u(x) = \gamma_0^{\text{int}} (\tilde{V} w)(x) = (V w)(x), \quad x \in \Gamma.$$

假定  $w \in H^{-\frac{1}{2}}(\Gamma)$ , 则  $u = \tilde{V}w \in H^1(\Omega)$ 。通过选取方程  $\nu \in H^1(\Omega)$ , 由(14.6)可得内界狄利克雷边界值问题的格林第一恒等式

$$a_\Omega(u, \nu) := \int_{\Omega} \nabla u(x) \nabla \nu(x) dx = \langle \gamma_1^{\text{int}} u, \gamma_0^{\text{int}} \nu \rangle_{\Gamma}. \quad (14.223)$$

代入(14.129) (Corollary 14.5), 上式改写为不等式形式

$$a_\Omega(u, u) \geq c_1^{\text{int}} \|\gamma_1^{\text{int}} u\|_{H^{-\frac{1}{2}}(\Gamma)}. \quad (14.224)$$

随后我们需要计算外界共法导数  $\gamma_1^{\text{int}} u \in H^{-\frac{1}{2}}(\Gamma)$ , 这可以通过分析  $|x| \rightarrow \infty$  时单层位势  $(\tilde{V}w)(x)$  的远端特征而得。为此, 我们需要引入子空间  $H_*^{-\frac{1}{2}}(\Gamma)$ , 子空间中的方程正交于常数

$$H_*^{-\frac{1}{2}}(\Gamma) := \left\{ w \in H^{-\frac{1}{2}}(\Gamma) : \langle w, 1 \rangle_{\Gamma} = 0 \right\}. \quad (14.225)$$

**Lemma 14.31.** 假定

$$|x - y_0| > \max \{1, 2 \operatorname{diam}(\Omega)\}, \quad y_0 \in \Omega, x \in \mathbb{R}^d.$$

设

$$w \in \begin{cases} H^{-\frac{1}{2}}(\Gamma) & d = 3, \\ H_*^{-\frac{1}{2}}(\Gamma) & d = 2. \end{cases}$$

那么  $u = \tilde{V}w$  的边界可以表示为

$$|u(x)| = \left| (\tilde{V}w)(x) \right| \leq c_1(w) \frac{1}{|x - y_0|}, \quad (14.226)$$

$$|\nabla u(x)| = \left| \nabla (\tilde{V}w)(x) \right| \leq c_2(w) \frac{1}{|x - y_0|^2}. \quad (14.227)$$

证明. 由三角不等式我们有

$$\begin{aligned} |x - y_0| &\leq |x - y| + |y - y_0| \\ &\leq |x - y| + \operatorname{diam}(\Omega) \\ &\leq |x - y| + \frac{1}{2} |x - y_0|. \end{aligned}$$

$$\hookrightarrow |x - y| \geq \frac{1}{2} |x - y_0|.$$

下面分  $d = 3, 2$  分别证明。

1. 三维空间  $d = 3 \Rightarrow$

$$\begin{aligned} \|U^*(x, \cdot)\|_{H^1(\Omega)}^2 &= \frac{1}{16\pi^2} \int_{\Omega} \frac{1}{|x - y|^2} dy + \frac{1}{16\pi^2} \int_{\Omega} \frac{1}{|x - y|^4} dy \\ &\leq \frac{1}{4\pi^2} \int_{\Omega} \frac{1}{|x - y_0|^2} dy + \frac{1}{\pi^2} \int_{\Omega} \frac{1}{|x - y_0|^4} dy \\ &\leq \frac{5}{4} \frac{|\Omega|}{\pi^2} \frac{1}{|x - y_0|^2} dy, \end{aligned}$$

证得(14.226)。

$$\begin{aligned}
|u(x)| &= |\langle \gamma_0^{\text{int}} U^*(x, \cdot), w \rangle_{\Gamma}| \\
&\leq \|\gamma_0^{\text{int}} U^*(x, \cdot)\|_{H^{\frac{1}{2}}(\Gamma)} \|w\|_{H^{-\frac{1}{2}}(\Gamma)} \\
&\leq c_T \|\gamma_0^{\text{int}} U^*(x, \cdot)\|_{H^1(\Omega)} \|w\|_{H^{-\frac{1}{2}}(\Gamma)}.
\end{aligned}$$

$$\hookrightarrow \frac{\partial}{\partial x_i} u(x) = \frac{1}{4\pi} \int_{\Gamma} \frac{y_i - x_i}{|x - y|^3} w(y) ds_y,$$

证得(14.227)。

2. 二维空间  $d = 2$  中, 沿着某个适宜的定值  $\bar{y} \in \Omega$  作泰勒展开

$$\log |y - x| \approx \log |y_0 - x| + \frac{(y - y_0, \bar{y} - x)}{|\bar{y} - x|^2}.$$

进而根据  $w \in H_*^{-\frac{1}{2}}(\Gamma)$  可得

$$u(x) = -\frac{1}{2\pi} \int_{\Gamma} \frac{(y - y_0, \bar{y} - x)}{|\bar{y} - x|^2} w(y) ds_y,$$

证得(14.226)。在此基础上证得(14.227)。

□

对于  $y_0 \in \Omega, R \geq 2 \operatorname{diam}(\Omega)$ , 定义球体

$$B_R(y_0) := \{x \in \mathbb{R}^d : |x - y_0| < R\}.$$

那么此时  $u(x) = (Vw)(x), x \in \Omega^c$  就成为以下狄利克雷边界值问题的唯一解

$$-\Delta u(x) = 0, \quad x \in B_R(y_0) \setminus \bar{\Omega}, \tag{14.228}$$

$$u(x) = \gamma_0^{\text{int}} (\tilde{V}w)(x) = (Vw)(x), \quad x \in \Gamma, \tag{14.229}$$

$$u(x) = (\tilde{V}w)(x), \quad x \in \partial B_R(y_0). \tag{14.230}$$

其中, 对于有界域  $B_R(y_0) \setminus \bar{\Omega}$  (14.228) 中的情况, 利用格林第一恒等式(14.6)可得

$$a_{B_R(y_0) \setminus \bar{\Omega}}(u, \nu) = \underbrace{-\langle \gamma_1^{\text{ext}} u, \gamma_0^{\text{ext}} \nu \rangle_{\Gamma}}_{=: \mathcal{A}} + \underbrace{\langle \gamma_1^{\text{int}} u, \gamma_0^{\text{int}} \nu \rangle_{\partial B_R(y_0)}}_{=: \mathcal{B}}. \tag{14.231}$$

1. 先来看  $\mathcal{B}$ 。设  $u = \nu$ 。由 Lemma 14.31 可得

$$\begin{aligned}
|\langle \gamma_1^{\text{int}} u, \gamma_0^{\text{int}} u \rangle_{\partial B_R(y_0)}| &\leq c_1(w) c_2(w) \int_{|x-y_0|=R} \frac{1}{|x-y_0|^3} ds_x \\
&\leq c R^{d-4}.
\end{aligned}$$

2. 再来看极限  $R \rightarrow \infty$  的情况。在外界域 (exterior domain)  $\partial B_R(y_0)$  (14.230) 中, 根据格林第一恒等式(14.6)可得

$$a_{\Omega^c}(u, u) := \int_{\Omega^c} \nabla u(x) \nabla u(x) dx = -\langle \gamma_1^{\text{ext}} u, \gamma_0^{\text{ext}} u \rangle_{\Gamma}. \tag{14.232}$$

在二维空间中，假设条件  $w \in H_*^{-\frac{1}{2}}(\Gamma)$  即可保证上述结论成立。以类似于 Corollary 14.5, (14.129) 的方法，对于(14.232)的外界狄利克雷边界值问题，满足不等式

$$a_{\Omega^c}(u, u) \geq c_1^{\text{ext}} \|\gamma_1^{\text{ext}} u\|_{H^{-\frac{1}{2}}(\Gamma)}^2. \quad (14.233)$$

**Theorem 14.35** (单层位势算子的椭圆性 (三维空间)). 设

$$w \in \begin{cases} H^{-\frac{1}{2}}(\Gamma) & d = 3, \\ H_*^{-\frac{1}{2}}(\Gamma) & d = 2. \end{cases}$$

那么有以下不等式

$$\langle Vw, w \rangle_\Gamma \geq c_1^V \|w\|_{H^{-\frac{1}{2}}(\Gamma)}^2, \quad (14.234)$$

其中常数  $c_1^V \geq 0$ 。

证明. 根据  $u = \tilde{V}w$ , 由内界、外界狄利克雷边界值问题的格林第一恒等式(14.223)、(14.232)可知

$$\begin{aligned} a_\Omega(u, u) &= \langle \gamma_1^{\text{int}} u, \gamma_0^{\text{int}} u \rangle_\Gamma, \\ a_{\Omega^c}(u, u) &= -\langle \gamma_1^{\text{ext}} u, \gamma_0^{\text{ext}} u \rangle_\Gamma. \end{aligned}$$

两式相加

$$\underbrace{a_\Omega(u, u) + a_{\Omega^c}(u, u)}_{=: \mathcal{C}} = \underbrace{\langle [\gamma_1^{\text{int}} u - \gamma_1^{\text{ext}} u], [\gamma_0^{\text{int}} u - \gamma_0^{\text{ext}} u] \rangle_\Gamma}_{=: \mathcal{A}}.$$

1. 由单层位势算子的跃动关系(14.186)可得

$$\mathcal{B} := \gamma_0^{\text{int}} u - \gamma_0^{\text{ext}} u := \gamma_0 u,$$

2. 由单层位势共法导数的跃动关系(14.190)可得

$$\mathcal{A} := \gamma_1^{\text{int}} u - \gamma_1^{\text{ext}} u := w,$$

3. 将(14.224), (14.233)代入  $\mathcal{C}$  得

$$\begin{aligned} \mathcal{C} &:= a_\Omega(u, u) + a_{\Omega^c}(u, u) \\ &\geq c_1^{\text{int}} \|\gamma_1^{\text{int}} u\|_{H^{-\frac{1}{2}}(\Gamma)} + c_1^{\text{ext}} \|\gamma_1^{\text{ext}} u\|_{H^{-\frac{1}{2}}(\Gamma)}^2 \\ &\geq \min \{c_1^{\text{int}}, c_2^{\text{ext}}\} \left[ \|\gamma_1^{\text{int}} u\|_{H^{-\frac{1}{2}}(\Gamma)} + \|\gamma_1^{\text{ext}} u\|_{H^{-\frac{1}{2}}(\Gamma)}^2 \right] \end{aligned}$$

$$\begin{aligned} \langle \mathcal{A}, \mathcal{B} \rangle_\Gamma &= \langle w, \gamma_0 u \rangle_\Gamma = \langle Vw, w \rangle_\Gamma \\ &= \mathcal{C} \geq \min \{c_1^{\text{int}}, c_2^{\text{ext}}\} \left[ \|\gamma_1^{\text{int}} u\|_{H^{-\frac{1}{2}}(\Gamma)} + \|\gamma_1^{\text{ext}} u\|_{H^{-\frac{1}{2}}(\Gamma)}^2 \right]. \end{aligned}$$

此外根据(14.23)有

$$\begin{aligned} \|w\|_{H^{-\frac{1}{2}}(\Gamma)}^2 &= \|\gamma_1^{\text{int}} u - \gamma_1^{\text{ext}} u\|_{H^{-\frac{1}{2}}(\Gamma)}^2 \\ &\leq \left[ \|\gamma_1^{\text{int}} u\|_{H^{-\frac{1}{2}}(\Gamma)} + \|\gamma_1^{\text{ext}} u\|_{H^{-\frac{1}{2}}(\Gamma)} \right]^2 \\ &\leq 2 \left[ \|\gamma_1^{\text{int}} u\|_{H^{-\frac{1}{2}}(\Gamma)}^2 + \|\gamma_1^{\text{ext}} u\|_{H^{-\frac{1}{2}}(\Gamma)}^2 \right], \end{aligned}$$

$$\begin{aligned} \langle Vw, w \rangle_\Gamma &\geq \min \{c_1^{\text{int}}, c_2^{\text{ext}}\} \left[ \|\gamma_1^{\text{int}} u\|_{H^{-\frac{1}{2}}(\Gamma)} + \|\gamma_1^{\text{ext}} u\|_{H^{-\frac{1}{2}}(\Gamma)}^2 \right] \\ &\geq \underbrace{\min \{c_1^{\text{int}}, c_2^{\text{ext}}\}}_{=: c_1^V \geq 0} \|w\|_{H^{-\frac{1}{2}}(\Gamma)}^2. \end{aligned}$$

□

对于二维系统  $d = 2$  的情况，我们也可从 Theorem 14.35 推得，单层位势  $V$  的  $H_*^{-\frac{1}{2}}(\Gamma)$ -椭圆特性。具体来说，有两种证明思路。

### 鞍点变分问题求证椭圆性

第一种证明思路是，汇总  $d = 2, 3$  的情况，构建鞍点问题

$$\langle Vt, \tau \rangle_\Gamma - \lambda \langle 1, \tau \rangle_\Gamma = 0, \quad \forall \tau \in H_*^{-\frac{1}{2}}(\Gamma), \quad (14.235)$$

$$\langle t, 1 \rangle_\Gamma = 1, \quad (14.236)$$

寻找解  $(t, \lambda) \in H_*^{-\frac{1}{2}}(\Gamma) \times \mathbb{R}$ 。

将  $t$  定义如下

$$t := \tilde{t} + \frac{1}{|\Gamma|}, \quad \tau \in H_*^{-\frac{1}{2}}(\Gamma),$$

不难看出，如果对于某一  $\tilde{t} \in H_*^{-\frac{1}{2}}(\Gamma)$ , (14.235) 的解  $t$  同时也满足关系式(14.236)。那么求解上述按点问题的关键就变为，寻找合适的  $\tilde{t}$ ，使满足如下关系

$$\langle V\tilde{t}, \tau \rangle_\Gamma = -\frac{1}{|\Gamma|} \langle V1, \tau \rangle_\Gamma, \quad \forall \tau \in H_*^{-\frac{1}{2}}(\Gamma).$$

由 Theorem 14.35 可见，上式中单层位势是  $H_*^{-\frac{1}{2}}(\Gamma)$ -椭圆的。因此该变分问题可解，解唯一。设这个唯一解为  $w_{\text{eq}}$ ，表示自然密度方程 (natural density function)。

$$w_{\text{eq}} := \tilde{t} + \frac{1}{|\Gamma|}, \quad (14.237)$$

更多自然密度的介绍见后

设  $\tau = w_{\text{eq}}$ ，则拉格朗日乘子  $\lambda$  满足

$$\lambda = \langle Vw_{\text{eq}}, w_{\text{eq}} \rangle_\Gamma.$$

在三维空间  $d = 3$  中，由 Theorem 14.35 可得拉，拉格朗日乘子严格为正  $\lambda > 0$ ，定义为  $\Gamma$  的容度 (capacity)，用  $\text{cap}_\Gamma$  表示。

在二维空间  $d = 2$  中， $\text{cap}_\Gamma$  可定义为指数形式

$$\text{cap}_\Gamma := \exp(-2\pi\lambda),$$

基于参数的基本解为

$$U^*(x, y) := \frac{1}{2\pi} \log r - \frac{1}{2\pi} \log |x - y|, \quad r \in \mathbb{R}^+,$$

对应的边界积分算子

$$(V_r w)(x) := \int_\Gamma U_r^*(x, y) w(y) ds_y, \quad x \in \Gamma.$$

由此可得

$$(V_r w)(x) = \frac{1}{2\pi} \log r + \lambda = \frac{1}{2\pi} \log \frac{r}{\text{cap}_\Gamma},$$

尤其是当  $r = 1$  时，有

$$\lambda := \frac{1}{2} \log \frac{1}{\text{cap}_\Gamma}. \quad (14.238)$$

由(14.238)可见，在二维系统  $d = 2$  中， $r = 1$  的前提下，只有当指数容度  $\text{cap}_\Gamma < 1$  时，才能确保  $\lambda > 0$ ，进而使得解存在。Hsiao and Wendland (1977) 论证了，使满足  $\text{cap}_\Gamma < 1$  的一个充分条件是假定  $\text{diam}(\Omega) < 1$ ——有时这便需要我们对给定的域  $\Omega \in \mathbb{R}^2$  作规模调节，使得直径小于 1 的假定得到满足（见418页第14.6.7节）。

**Theorem 14.36** (单层位势算子的椭圆性 (二维空间)). 二维空间  $d = 2$  中, 设  $\text{diam}(\Omega) < 1$ , 从而  $\lambda > 0$ 。单层位势  $V$  是  $H^{-\frac{1}{2}}(\Gamma)$ -椭圆的, 即

$$\langle Vw, w \rangle_{\Gamma} \geq \tilde{c}_1^V \|w\|_{H^{-\frac{1}{2}}(\Gamma)}^2, \quad \forall w \in H^{-\frac{1}{2}}(\Gamma). \quad (14.239)$$

证明. 1. 看 RHS。对于任一给定  $w \in H^{-\frac{1}{2}}(\Gamma)$ , 都可以对  $w$  作唯一的分解

$$w = \tilde{w} + \alpha w_{\text{eq}}, \quad \tilde{w} \in H_*^{-\frac{1}{2}}(\Gamma), \alpha = \langle w, 1 \rangle_{\Gamma}.$$

$$\begin{aligned} \hookrightarrow \|w\|_{H^{-\frac{1}{2}}(\Gamma)}^2 &= \|\tilde{w} + \alpha w_{\text{eq}}\|_{H^{-\frac{1}{2}}(\Gamma)}^2 \\ &\leq \left[ \|\tilde{w}\|_{H^{-\frac{1}{2}}(\Gamma)} + \alpha \|w_{\text{eq}}\|_{H^{-\frac{1}{2}}(\Gamma)} \right]^2 \\ &\leq 2 \left[ \|\tilde{w}\|_{H^{-\frac{1}{2}}(\Gamma)}^2 + \alpha^2 \|w_{\text{eq}}\|_{H^{-\frac{1}{2}}(\Gamma)}^2 \right] \\ &\leq 2 \max \left\{ 1, \|w_{\text{eq}}\|_{H^{-\frac{1}{2}}(\Gamma)}^2 \right\} \left[ \|\tilde{w}\|_{H^{-\frac{1}{2}}(\Gamma)}^2 + \alpha^2 \right]. \end{aligned}$$

2. 看 LHS。

$$\begin{aligned} \langle Vw, w \rangle_{\Gamma} &= \langle V(\tilde{w} + \alpha w_{\text{eq}}), \tilde{w} + \alpha w_{\text{eq}} \rangle_{\Gamma} \\ &= \underbrace{\langle V\tilde{w}, \tilde{w} \rangle}_{=: \mathcal{A}} + 2\alpha \underbrace{\langle \tilde{V}w_{\text{eq}}, \tilde{w} \rangle_{\Gamma}}_{=: \mathcal{B}} + \alpha^2 \underbrace{\langle Vw_{\text{eq}}, w_{\text{eq}} \rangle_{\Gamma}}_{=: \mathcal{C}}, \end{aligned}$$

其中

(a) 根据 Theorem 14.35

$$\mathcal{A} := \langle \tilde{V}w_{\text{eq}}, \tilde{w} \rangle_{\Gamma} \geq c_1^V \|\tilde{w}\|_{H^{-\frac{1}{2}}(\Gamma)}^2,$$

(b)

$$\mathcal{B} := \langle \tilde{V}w_{\text{eq}}, \tilde{w} \rangle_{\Gamma} = 0,$$

(c) 根据定义有

$$\mathcal{C} := \langle Vw_{\text{eq}}, w_{\text{eq}} \rangle_{\Gamma} = \lambda,$$

因此上式变为

$$\begin{aligned} \langle Vw, w \rangle_{\Gamma} &= \mathcal{A} + 2\alpha\mathcal{B} + \alpha^2\mathcal{C} \\ &\geq c_1^V \|\tilde{w}\|_{H^{-\frac{1}{2}}(\Gamma)}^2 + \alpha^2\lambda \\ &\geq \min\{c_1^V, \lambda\} \left[ \|\tilde{w}\|_{H^{-\frac{1}{2}}(\Gamma)}^2 + \alpha^2 \right] \end{aligned}$$

3. 联立 LHS 和 RHS, 可证。 □

第二种思路是从自然密度方程  $w_{\text{eq}} \in H^{-\frac{1}{2}}(\Gamma)$  入手。已知  $w_{\text{eq}}$  是以下带有约束条件的算子方程的解

$$\begin{aligned} (Vw_{\text{eq}})(x) &= \lambda, \quad x \in \Gamma, \\ \langle w_{\text{eq}}, 1 \rangle_{\Gamma} &= 1. \end{aligned}$$

我们首先对  $w_{\text{eq}}$  作规模调节:

$$w_{\text{eq}} := \lambda \tilde{w}_{\text{eq}},$$

新的算子方程变为

$$\begin{aligned} (V\tilde{w}_{\text{eq}})(x) &= 1, \quad x \in \Gamma, \\ \langle \tilde{w}_{\text{eq}}, 1 \rangle_{\Gamma} &= \frac{1}{\lambda}. \end{aligned} \tag{14.240}$$

求解思路略。大致说来，第一步求解边界积分方程，对应(14.240)第一行，以求得作为自然密度的解  $w_{\text{eq}}$ 。第二步，将规模调整得到的  $\tilde{w}_{\text{eq}}$  引入第二行，计算容度。

### 单层位势算子的可逆性

已知单层位势边界积分算子  $V : H^{-\frac{1}{2}}(\Gamma) \mapsto H^{\frac{1}{2}}(\Gamma)$  第一有界(14.182)，第二  $H^{-\frac{1}{2}}(\Gamma)$ -椭圆( $d = 2$  的证明见 Theorem 14.36,  $d = 3$  的证明见 Theorem 14.35)。那么可应用拉克斯—密格拉蒙定理 (Theorem 14.14) 证得单层位势  $V$  可逆，即  $V^{-1} : H^{\frac{1}{2}}(\Gamma) \mapsto H^{-\frac{1}{2}}(\Gamma)$  有界，并且满足

$$\|V^{-1}\nu\|_{H^{-\frac{1}{2}}(\Gamma)} \leq \frac{1}{c_1^V} \|\nu\|_{H^{\frac{1}{2}}(\Gamma)}, \quad \forall \nu \in H^{\frac{1}{2}}(\Gamma).$$

在此基础上，双线性内积形式的单层位势  $V$

$$\begin{aligned} \langle Vw, w_{\text{eq}} \rangle_{\Gamma} &= \langle w, Vw_{\text{eq}} \rangle_{\Gamma} \\ &= \lambda \langle w, 1 \rangle_{\Gamma} = 0, \quad \forall w \in H_*^{\frac{1}{2}}(\Gamma). \end{aligned}$$

$$\begin{aligned} \therefore Vw &\in H_*^{\frac{1}{2}}(\Gamma), \\ H_*^{\frac{1}{2}}(\Gamma) &:= \left\{ \nu \in H^{\frac{1}{2}}(\Gamma) : \langle \nu, w_{\text{eq}} \rangle_{\Gamma} = 0 \right\}. \end{aligned}$$

因此可见，单层位势  $V : H_*^{-\frac{1}{2}}(\Gamma) \mapsto H_*^{\frac{1}{2}}(\Gamma)$  是一个同构 (isomorphism)<sup>18</sup>。

### 14.6.8 超奇异边界积分算子的椭圆性

第14.6.5节介绍了超奇异边界积分算子  $D$ 。由(14.203)可得，若引入特征解 (eigensolution)  $u_0 \equiv 1, x \in \Gamma$ ，那么  $(Du_0)(x) = 0$ 。可见我们无法保证超奇异边界积分算子  $D$  在  $H^{\frac{1}{2}}(\Gamma)$  中全部都是椭圆的，而只是半椭圆 (semi-ellipticity) 的。在本节中，首先我们介绍半椭圆性。随后介绍替代方案：在  $H^{\frac{1}{2}}(\Gamma)$  中构建满足椭圆特征的子空间，两种常见的子空间如下文所示。

#### 超奇异边界积分算子的半椭圆性

**Theorem 14.37** (超奇异边界积分算子的半椭圆性). 超奇异边界积分算子  $D$  是半椭圆的，即

$$\langle D\nu, \nu \rangle_{\Gamma} \geq c_1^D \|\nu\|_{H^{\frac{1}{2}}(\Gamma)}^2, \quad \forall \nu \in H_*^{\frac{1}{2}}(\Gamma).$$

证明. 1. 对于  $\nu \in H_*^{\frac{1}{2}}(\Gamma)$ ，考虑齐次偏微分方程的解  $u(x)$  可以表示为如下双层位势

$$u(x) := -(Wu)(x), \quad x \in \Omega \cup \Omega_c.$$

对于  $x \in \Gamma$ ,  $u(x)$  的迹算子和共法导数算子以伴随双层位势的形式表示如下

$$\begin{cases} \gamma_0^{\text{int}} u(x) = (1 - \sigma(x)) \nu(x) - (K\nu)(x), \\ \gamma_1^{\text{int}} u(x) = (D\nu)(x), \quad x \in \Gamma, \end{cases}$$

<sup>18</sup> 同构的概念，可参考 (Muscat, 2014, pp.34, Definition 3.11)。

$$\begin{cases} \gamma_0^{\text{ext}} u(x) = -\sigma(x)\nu(x) - (K\nu)(x), \\ \gamma_1^{\text{ext}} u(x) = (D\nu)(x), \quad x \in \Gamma. \end{cases}$$

2. 由此可见, 方程  $u = -Wu$  构成内界狄利克雷边界值问题的唯一解

$$\begin{cases} -\Delta u(x) = 0 & x \in \Gamma, \\ \gamma_0^{\text{int}} u(x) = (1 - \sigma(x))\nu(x) - (K\nu)(x) & x \in \Gamma. \end{cases}$$

由格林第一恒等式(14.6)可得, 解  $u(x)$  满足如下关系

$$\int_{\Gamma} \nabla u(x) \nabla w(x) dx = \langle \gamma_1^{\text{int}} u, \gamma_0^{\text{int}} w \rangle_{\Gamma}, \quad \forall w \in H^1(\Omega).$$

3. 设一个包围  $\Omega$  的球体  $B_R(y_0)$ , 以  $y_0 \in \Omega$  为圆心, 以  $R > 2 \operatorname{diam}(\Omega)$  为半径。即  $\Omega \subset B_R(y_0)$ 。那么  $u = -Wu$  同时也是以下狄利克雷边界值问题的唯一解

$$\begin{cases} -\Delta u(x) = 0 & x \in B_R(y_0) \setminus \bar{\Omega}, \\ \gamma_0^{\text{ext}} u(x) = -\sigma(x)\nu(x) - (K\nu)(x) & x \in \Gamma = \partial\Omega, \\ \gamma_0 u(x) = \gamma_0^{\text{int}} u(x) + \gamma_0^{\text{ext}} u(x) = -(W\nu)(x) & x \in \partial B_r(y_0). \end{cases}$$

根据格林第一恒等式可得(14.6)可得, 解  $u(x)$  表示为如下双线性形式

$$\int_{B_R(y_0) \setminus \bar{\Omega}} \nabla u(x) \nabla w(x) dx = -\langle \gamma_1^{\text{ext}} u, \gamma_0^{\text{ext}} w \rangle_{\Gamma} + \langle \gamma_1 u, \gamma_0 w \rangle_{\partial B_r(y_0)}, \quad \forall w \in H^1(B_R(y_0) \setminus \bar{\Omega}).$$

(a) 对于  $x \in \Gamma$ , 根据定义有

$$u(x) = \frac{1}{2(d-1)\pi} \int_{\Gamma} \frac{(y-x, n_y)}{|x-y|^d} \nu(y) ds_y.$$

(b) 对于  $x \in \partial B_R(y_0)$ , 有

$$|u(x)| \leq c_1(\nu) R^{1-d},$$

$$|\nabla u(x)| \leq c_2(\nu) R^{-d}.$$

那么选取  $w = u = -W\nu$ , 取极限  $\lim R \rightarrow \infty$ , 可得外界域中的格林第一恒等式

$$\int_{\Omega^c} |\nabla u(x)|^2 dx = -\langle \gamma_1^{\text{ext}} u, \gamma_0^{\text{ext}} u \rangle_{\Gamma}.$$

4. 将内、外界域的格林第一恒等式相加求和, 并考虑不同边界积分算子的跃动关系, 我们有超奇异边界积分算子  $D$  的双线性形式

$$\begin{aligned} \underbrace{\langle D\nu, \nu \rangle_{\Gamma}}_{=: \mathcal{C}} &= \langle \gamma_1^{\text{int}} u, [\gamma_0^{\text{int}} - \gamma_0^{\text{ext}} u] \rangle_{\Gamma} \\ &= \langle \gamma_1^{\text{int}} u, \gamma_0^{\text{int}} u \rangle_{\Gamma} - \langle \gamma_1^{\text{int}} u, \gamma_0^{\text{ext}} u \rangle_{\Gamma} \\ &= \int_{\Omega} |\nabla u(x)|^2 dx + \int_{\Omega^c} |\nabla u(x)|^2 dx \\ &= \underbrace{|u|_{H^1(\Omega)}^2}_{=: \mathcal{B}} + \underbrace{|u|_{H^1(\Omega^c)}^2}_{=: \mathcal{A}}. \end{aligned}$$

(a) 来看  $\mathcal{A}$ 。外界域  $\Omega^c$  中, 双层位势  $u(x) = -(Wu)(x)$ , 取  $\lim|x| \rightarrow \infty$ , 有

$$c_1 \|u\|_{H^1(\Omega^c)}^2 \leq |u|_{H^1(\Omega^c)}^2.$$

(b) 来看  $\mathcal{B}$ 。

i. 由以下三点

- 对于  $\nu \in H_*^{\frac{1}{2}}(\Gamma)$ , 有自然密度  $w_{\text{eq}} \in H^{-\frac{1}{2}}(\Gamma)$ ,
- $Vw_{\text{eq}} = 1$ ,
- 有界积分算子呈对称关系(14.220) 成立

可得

$$\begin{aligned} \langle \gamma_0^{\text{int}} u, w_{\text{eq}} \rangle_\Gamma &= \left\langle \left( \frac{1}{2}I - K \right) \nu, w_{\text{eq}} \right\rangle_\Gamma \\ &= \langle \nu, w_{\text{eq}} \rangle_\Gamma - \left\langle \left( \frac{1}{2}I + K \right) \nu, w_{\text{eq}} \right\rangle_\Gamma \\ &= -\left\langle \left( \frac{1}{2}I + K \right) \nu, V^{-1} \right\rangle_\Gamma \\ &= -\left\langle V^{-1} \left( \frac{1}{2}I + K \right) \nu, 1 \right\rangle_\Gamma \\ &= -\left\langle \left( \frac{1}{2}I + K \right) V^{-1} \nu, 1 \right\rangle_\Gamma \\ &= -\left\langle V^{-1} \nu, \left( \frac{1}{2}I + K \right) \right\rangle_\Gamma \\ &= 0. \end{aligned}$$

ii. 进而, 根据索伯列夫空间中的范数等价定理(14.37)(Theorem 14.4) 可得  $H^1(\Omega)$  空间中的一个等价范

$$\|u\|_{H_*^1(\Omega)} := \left\{ [\langle \gamma_0^{\text{int}} u, w_{\text{eq}} \rangle_\Gamma]^2 + \|\nabla u\|_{L^2(\Omega)}^2 \right\}^{\frac{1}{2}}.$$

(c) 来看  $\mathcal{C}$ 。由迹定理和双层位势的跃动关系可得

$$\begin{aligned} \langle D\nu, \nu \rangle_\Gamma &\geq c \left\{ \|u\|_{H^1(\Omega)}^2 + \|u\|_{H^1(\Omega^c)}^2 \right\} \\ &\geq \tilde{c} \left\{ \|\gamma_0^{\text{int}} u\|_{H^{\frac{1}{2}}(\Gamma)}^2 + \|\gamma_0^{\text{ext}} u\|_{H^{\frac{1}{2}}(\Gamma)}^2 \right\} \\ &\geq \frac{1}{2} \tilde{c} \|\gamma_0^{\text{int}} u - \gamma_0^{\text{ext}} u\|_{H^{\frac{1}{2}}(\Gamma)}^2 \\ &= c_1^D \|\nu\|_{H_*^{\frac{1}{2}}(\Gamma)}^2, \quad \forall \nu \in H_*^{\frac{1}{2}}(\Gamma). \end{aligned}$$

由此可证得超奇异边界积分算子  $D$  的  $H_*^{\frac{1}{2}}(\Gamma)$ -椭圆性。  $\square$

### 椭圆性所处子空间的替代方案

Theorem 14.37 对超奇异边界积分算子  $D$  的  $H_*^{\frac{1}{2}}(\Gamma)$ -椭圆性的证明, 涉及到对解方程所处空间的限定: 需要将其限定在一个更适宜的子空间之内——即正交于常数的子空间。显然, 随着算子  $D$  双线性内积类型的不同, 对应正交子空间的种类也有所不同。具体说来, 根据索伯列夫空间中的范数等价定理(14.37)(Theorem 14.4), 可定义一个  $H^{\frac{1}{2}}(\Gamma)$  中的等价范

$$\|\nu\|_{H_*^{\frac{1}{2}}(\Gamma)} := \left\{ [\langle \nu, w_{\text{eq}} \rangle_\Gamma]^2 + |\nu|_{H^{\frac{1}{2}}(\Gamma)}^2 \right\}^{\frac{1}{2}},$$

其中  $w_{\text{eq}} \in H^{-\frac{1}{2}}(\Gamma)$  由自然密度的定义式 (14.240) 给出。那么

**Corollary 14.9** (超奇异边界积分算子的半椭圆性). 超奇异积分算子  $D$  是  $H^{\frac{1}{2}}(\Gamma)$ -半椭圆 (*semi-elliptic*) 的, 即

$$\langle D\nu, \nu \rangle_{\Gamma} \geq \bar{c}_1^D |\nu|_{H^{\frac{1}{2}}(\Gamma)}^2, \quad \forall \nu \in H^{\frac{1}{2}}(\Gamma). \quad (14.241)$$

Corollary 14.9 中, 对子空间  $H_*^{\frac{1}{2}}(\Gamma)$  的定义, 涉及到边界积分方程(14.240)的唯一解, 即自然密度  $w_{\text{eq}} \in H^{-\frac{1}{2}}(\Gamma)$ 。然而在实际计算过程中, 该定义往往较难直接处理。有鉴于此, 需要另外寻求替代方案。

常见的替代方案之一是定义另一个子空间  $H_{**}^{\frac{1}{2}}(\Gamma)$ , 从而简化内积结构

$$H_{**}^{\frac{1}{2}}(\Gamma) := \left\{ \nu \in H^{\frac{1}{2}}(\Gamma) : \langle \nu, 1 \rangle_{\Gamma} = 0 \right\}.$$

则由(14.241)得, 对于  $\nu \in H_{**}^{\frac{1}{2}}(\Gamma)$ , 内积可以简化为

$$\begin{aligned} \langle D\nu, \nu \rangle_{\Gamma} &\geq \bar{c}_1^D |\nu|_{H^{\frac{1}{2}}(\Gamma)}^2 \\ &= \bar{c}_1^D \left\{ |\nu|_{H^{\frac{1}{2}}(\Gamma)}^2 + [\langle \nu, 1 \rangle_{\Gamma}]^2 \right\} \\ &\geq \tilde{c}_1^D \|\nu\|_{H^{\frac{1}{2}}(\Gamma)}^2. \end{aligned} \quad (14.242)$$

若是采用子空间  $H_{**}^{\frac{1}{2}}(\Gamma)$ , 我们也可以证明算子  $D$  的  $H_{**}^{\frac{1}{2}}(\Gamma)$ -椭圆性。(证明略)。

另一个替代方案是, 考虑一个子空间  $\Gamma_0 \subset \Gamma$ 。对于某一给定的  $\nu \in \tilde{H}^{\frac{1}{2}}(\Gamma_0)$ , 定义其延拓  $\tilde{\nu} \in H^{\frac{1}{2}}(\Gamma)$  如下

$$\tilde{\nu}(x) = \begin{cases} \nu(x) & x \in \Gamma_0, \\ 0 & \text{否则.} \end{cases}$$

根据范数等价定理,  $H^{\frac{1}{2}}(\Gamma)$  上的等价范可定义为

$$\|w\|_{H^{\frac{1}{2}}(\Gamma), \Gamma_0} := \left\{ \|w\|_{L^2(\Gamma \setminus \Gamma_0)}^2 + |w|_{H^{\frac{1}{2}}(\Gamma)}^2 \right\}^{\frac{1}{2}}.$$

那么, 对于  $\nu \in \tilde{H}^{\frac{1}{2}}(\Gamma_0)$ , 内积简化为

$$\begin{aligned} \langle D\nu, \nu \rangle_{\Gamma_0} &= \langle D\tilde{\nu}, \tilde{\nu} \rangle_{\Gamma} \\ &\geq \bar{c}_1^D \left[ \|\tilde{\nu}\|_{L^2(\Gamma \setminus \Gamma_0)}^2 + |\tilde{\nu}|_{H^{\frac{1}{2}}(\Gamma)}^2 \right] \\ &= \bar{c}_1^D \|\tilde{\nu}\|_{H^{\frac{1}{2}}(\Gamma), \Gamma_0}^2 \\ &\geq \tilde{c}_1^D \|\tilde{\nu}\|_{H^{\frac{1}{2}}(\Gamma)}^2 \\ &= \tilde{c}_1^D \|\nu\|_{\tilde{H}^{\frac{1}{2}}(\Gamma_0)}^2, \end{aligned} \quad (14.243)$$

即算子  $D$  是  $\tilde{H}^{\frac{1}{2}}(\Gamma_0)$ -椭圆的。

### 14.6.9 Steklov-Poincaré 算子

在求解边界值问题的过程中, 柯西数  $\gamma_0^{\text{int}} u$ ,  $\gamma_1^{\text{int}} u$  之间的关系变得非常重要。以齐次偏微分边界方程系统为例, 即将  $f \equiv 0$  引入(14.214)

$$\begin{pmatrix} \gamma_0^{\text{int}} u(x) \\ \gamma_1^{\text{int}} u(x) \end{pmatrix} = \underbrace{\begin{pmatrix} (1-\sigma)I - K & V \\ D & \sigma I + K' \end{pmatrix}}_{=: \mathcal{C}} \begin{pmatrix} \gamma_0^{\text{int}} u(x) \\ \gamma_1^{\text{int}} u(x) \end{pmatrix}. \quad (14.244)$$

单层位势算子  $V$  可逆。由(14.244)第一行可得

$$\begin{aligned}\gamma_0^{\text{int}} u &= ((1-\sigma)I - K)\gamma_0^{\text{int}} u + V\gamma_1^{\text{int}} u, \\ \hookrightarrow \gamma_1^{\text{int}} u &= \underbrace{V^{-1}(\sigma I + K)}_{=:S} \gamma_0^{\text{int}} u.\end{aligned}\quad (14.245)$$

这构成一个由狄利克雷到诺依曼的映射 (Dirichlet to Neumann map)(Behrndt and ter Elst, 2015)。由此我们可以定义 Steklov-Poincaré 算子如下

$$S := V^{-1}(\sigma I + K) : H^{\frac{1}{2}}(\Gamma) \mapsto H^{-\frac{1}{2}}(\Gamma). \quad (14.246)$$

(14.245)代回第二行

$$\begin{aligned}\gamma_1^{\text{int}} u(x) &= (D\gamma_0^{\text{int}} u)(x) + (\sigma I + K')\gamma_1^{\text{int}} u(x) \\ &= \underbrace{[D + (\sigma I + K')V^{-1}(\sigma I + K)]}_{=:S} \gamma_0^{\text{int}} u(x), \quad x \in \Gamma,\end{aligned}\quad (14.247)$$

可见 Steklov-Poincaré 算子也可表示为如下对称形式

$$S := D + (\sigma I + K')V^{-1}(\sigma I + K) : H^{\frac{1}{2}}(\Gamma) \mapsto H^{-\frac{1}{2}}(\Gamma). \quad (14.248)$$

结合(14.245)和 (14.247)，狄利克雷到诺依曼的映射表达式表示为

$$\gamma_1^{\text{int}} u(x) = (S\gamma_0^{\text{int}} u)(x), \quad x \in \Gamma, \quad (14.249)$$

即狄利克雷到诺依曼的映射将某一给定的狄利克雷数  $\gamma_0^{\text{int}} u \in H^{\frac{1}{2}}(\Gamma)$  映射至对应的诺依曼数  $\gamma_1^{\text{int}} u \in H^{-\frac{1}{2}}(\Gamma)$ ，其中和谐方程 (harmonic function)  $u \in H^1(\Omega)$  满足  $Lu \equiv 0$ 。

由逆单层位势算子  $V^{-1}$  的  $H^{\frac{1}{2}}(\Omega)$ -椭圆属性可得

$$\begin{aligned}\langle S\nu, \nu \rangle_\Gamma &= \langle D\nu, \nu \rangle_\Gamma + \langle V^{-1}(\sigma I + K)\nu, (\sigma I + K)\nu \rangle_\Gamma \\ &\geq \langle D\nu, \nu \rangle_\Gamma, \quad \forall \nu \in H^{\frac{1}{2}}(\Gamma),\end{aligned}\quad (14.250)$$

可见 Steklov-Poincaré 算子继承了超奇异边界积分算子的椭圆特性，随着不同的子空间，可表示如下

$$\langle S\nu, \nu \rangle_\Gamma \geq c_1^D \|\nu\|_{H^{\frac{1}{2}}(\Gamma)}^2, \quad \forall \nu \in H_*^{\frac{1}{2}}(\Gamma). \quad (14.251)$$

$$\langle S\nu, \nu \rangle_\Gamma \geq \tilde{c}_1^D \|\nu\|_{H^{\frac{1}{2}}(\Gamma)}^2, \quad \forall \nu \in H_{**}^{\frac{1}{2}}(\Gamma). \quad (14.252)$$

$$\langle S\nu, \nu \rangle_{\Gamma_0} \geq \hat{c}_1 \|\nu\|_{\tilde{H}^{\frac{1}{2}}(\Gamma_0)}^2, \quad \forall \nu \in \tilde{H}^{\frac{1}{2}}(\Gamma_0). \quad (14.253)$$

### 14.6.10 双层位势的收缩属性

在证得单层位势  $V$ (以及相伴随的超奇异积分算子  $D$ ) 的椭圆性后，我们科进一步证明对应的双层位势  $\sigma I + K : H^{\frac{1}{2}}(\Gamma) \mapsto H^{\frac{1}{2}}(\Gamma)$  也是椭圆的 (Steinbach and Wendland, 2001)。

由于算子  $V : H^{-\frac{1}{2}}(\Gamma) \mapsto H^{\frac{1}{2}}(\Gamma)$  有界且  $H^{-\frac{1}{2}}(\Gamma)$  椭圆，我们可以定义一个  $H^{\frac{1}{2}}(\Gamma)$  中的等价范

$$\|u\|_{V^{-1}} := \sqrt{\langle V^{-1}u, u \rangle_\Gamma}, \quad \forall u \in H^{\frac{1}{2}}(\Gamma). \quad (14.254)$$

**Theorem 14.38** (双层位势的收缩属性). 对于  $u \in H_*^{\frac{1}{2}}(\Gamma)$  我们有

$$(1 - c_K) \|u\|_{V^{-1}} \leq \|(\sigma I + K) u\|_{V^{-1}} \leq c_K \|u\|_{V^{-1}}, \quad (14.255)$$

其中常数  $c_K$  满足

$$c_K = \frac{1}{2} + \sqrt{\frac{1}{4} - c_1^V c_1^D} < 1, \quad (14.256)$$

$c_1^V, c_1^D$  分别是单层位势  $V$  和超奇异边界积分算子  $D$  的椭圆常数。

证明. 由 Steklov-Poincaré 算子的对称表现式 (14.248) 可得

$$\begin{aligned} \|(\sigma I + K) u\|_{V^{-1}}^2 &= \langle V^{-1}(\sigma I + K), (\sigma I + K) u \rangle_\Gamma \\ &= \underbrace{\langle Su, u \rangle_\Gamma}_{=: \mathcal{A}} - \underbrace{\langle Du, u \rangle_\Gamma}_{=: \mathcal{B}}, \end{aligned}$$

1.  $\mathcal{A}$ 。设算子  $A := JV^{-1} : H^{\frac{1}{2}}(\Gamma) \mapsto H^{\frac{1}{2}}(\Gamma)$  自伴随且  $H^{\frac{1}{2}}(\Gamma)$ -椭圆。考虑  $A = A^{\frac{1}{2}} A^{\frac{1}{2}}$ , 以及  $S$  的表现式(14.246)得

$$\begin{aligned} \mathcal{A} &:= \langle Su, u \rangle_\Gamma = \langle V^{-1}(\sigma I + K) u, u \rangle_\Gamma \\ &= \langle JV^{-1}(\sigma I + K) u, u \rangle_\Gamma \\ &= \langle A^{\frac{1}{2}}(\sigma I + K) u, A^{\frac{1}{2}} u \rangle_\Gamma \\ &\leq \|A^{\frac{1}{2}}(\sigma I + K) u\|_{H^{\frac{1}{2}}(\Gamma)} \underbrace{\|A^{\frac{1}{2}} u\|_{H^{\frac{1}{2}}(\Gamma)}}_{=: \mathcal{A}_1}, \end{aligned}$$

由于

$$\begin{aligned} \|A^{\frac{1}{2}} u\|_{H^{\frac{1}{2}}(\Gamma)} &= \langle A^{\frac{1}{2}} \nu, A^{\frac{1}{2}} \nu \rangle_{H^{\frac{1}{2}}(\Gamma)} \\ &= \langle JV^{-1} \nu, \nu \rangle_{H^{\frac{1}{2}}(\Gamma)} \\ &= \langle V^{-1} \nu, \nu \rangle_\Gamma \\ &= \|\nu\|_{V^{-1}}, \end{aligned}$$

$$\hookrightarrow \mathcal{A}_1 = \|u\|_{V^{-1}}, \hookrightarrow \mathcal{A} = \langle Su, u \rangle_\Gamma \leq \|(\sigma I + K) u\|_{V^{-1}} \|u\|_{V^{-1}}.$$

2.  $\mathcal{B}$ 。已知对于  $u \in H_*^{\frac{1}{2}}(\Gamma)$ , 算子  $D$  椭圆。则逆单层位势  $V^{-1}$  的映射特征为

$$\begin{aligned} \mathcal{B} &:= \langle Du, u \rangle_\Gamma \geq c_1^D \|u\|_{H^{\frac{1}{2}}(\Gamma)}^2 \\ &\geq c_1^D c_1^V \langle V^{-1} u, u \rangle_\Gamma \\ &= c_1^D c_1^V \|u\|_{V^{-1}}. \end{aligned}$$

3.  $\therefore$

$$\begin{aligned} \|(\sigma I + K) u\|_{V^{-1}}^2 &= \mathcal{A} - \mathcal{B} \\ &\leq \underbrace{\|(\sigma I + K) u\|_{V^{-1}}}_{=: \mathcal{C}} \underbrace{\|u\|_{V^{-1}}}_{=: \mathcal{D}} - c_1^D c_1^V \|u\|_{V^{-1}}. \end{aligned}$$

定义

$$\begin{cases} \mathcal{C} := \|(\sigma I + K) u\|_{V^{-1}} & \mathcal{C} > 0, \\ \mathcal{D} := \|u\|_{V^{-1}} & \mathcal{D} > 0, \end{cases}$$

上式变为

$$\left(\frac{\mathcal{C}}{\mathcal{D}}\right)^2 - \left(\frac{\mathcal{C}}{\mathcal{D}}\right) + c_1^D c_1^V \leq 0,$$

求解得

$$\frac{1}{2} - \sqrt{\frac{1}{4} - c_1^D c_1^V} \leq \frac{\mathcal{C}}{\mathcal{D}} \leq \frac{1}{2} + \sqrt{\frac{1}{4} - c_1^D c_1^V}$$

证毕。  $\square$

双层位势  $\sigma I + K$  的收缩特性 (contraction property), 尤其是  $H_*^{1/2}(\Gamma)$  中的上界, 可以延拓至  $H^{1/2}(\Gamma)$  空间中, 并且仍然成立。

**Corollary 14.10** (双层位势收缩比率). 对于  $u \in H^{1/2}(\Gamma)$ , 双层位势  $\sigma I + K$  满足如下不等式

$$\|(\sigma I + K) u\|_{V^{-1}} \leq c_K \|u\|_{V^{-1}}, \quad (14.257)$$

其中正常数  $0 \leq c_K < 1$  表示收缩比率 (contraction rate), 其值由(14.256)给出。

证明. 对于任一给定  $u \in H^{1/2}(\Gamma)$ , 我们有

$$u = \tilde{u} + \frac{\langle u, w_{\text{eq}} \rangle_{\Gamma}}{\langle 1, w_{\text{eq}} \rangle_{\Gamma}} u_0, \quad \tilde{u} \in H_*^{1/2}(\Gamma), u_0 \equiv 1.$$

1. 已知  $(\sigma I + K) u_0 = 0$ , 代入 Theorem 14.38 我们有

$$\|(\sigma I + K) u\|_{V^{-1}} = \|(\sigma I + K) \tilde{u}\|_{V^{-1}} \leq c_K \|\tilde{u}\|_{V^{-1}}.$$

2. 此外我们有

$$\|u\|_{V^{-1}}^2 = \|\tilde{u}\|_{V^{-1}}^2 + \frac{[\langle u, w_{\text{eq}} \rangle_{\Gamma}]^2}{\langle 1, w_{\text{eq}} \rangle_{\Gamma}} \geq \|\tilde{u}\|_{V^{-1}}.$$

$\square$

**Corollary 14.11** (位移双层位势的收缩比率). 对于  $\nu \in H_*^{1/2}(\Gamma)$ , 以下不等式成立

$$(1 - c_K) \|\nu\|_{V^{-1}} \leq \|[ (1 - \sigma) - K ] \nu\|_{V^{-1}} \leq c_K \|\nu\|_{V^{-1}}, \quad (14.258)$$

其中收缩比值  $c_K$  由(14.256)给出。

证明. 1. 由三角不等式可得

$$\begin{aligned} \|\nu\|_{V^{-1}} &= \|[(1 - \sigma) I - K] \nu + (\sigma I + K) \nu\|_{V^{-1}} \\ &\leq \|[ (1 - \sigma) I - K ] \nu\|_{V^{-1}} + \|(\sigma I + K) \nu\|_{V^{-1}} \\ &\leq \|[ (1 - \sigma) I - K ] \nu\|_{V^{-1}} + c_K \|\nu\|_{V^{-1}}, \\ &\rightarrow (1 - c_K) \|\nu\|_{V^{-1}} \leq \|[ (1 - \sigma) I - K ] \nu\|_{V^{-1}}. \end{aligned}$$

2. 同时使用 Steklov-Poincaré 算子的两种表现式 (14.246), (14.248) 可得

$$\begin{aligned}
 \|[(1 - \sigma)I - K]\nu\|_{V^{-1}}^2 &= \|[I - (\sigma I + K)]\nu\|_{V^{-1}}^2 \\
 &= \|\nu\|_{V^{-1}}^2 + \|\sigma I + K\|_{V^{-1}}^2 - 2\langle V^{-1}(\sigma I + K)\nu, \nu \rangle_\Gamma \\
 &= \|\nu\|_{V^{-1}}^2 + \|\sigma I + K\|_{V^{-1}}^2 - 2\langle S\nu, \nu \rangle_\Gamma \\
 &= \|\nu\|_{V^{-1}}^2 - \|\sigma I + K\|_{V^{-1}}^2 - 2\langle D\nu, \nu \rangle_\Gamma \\
 &\leq \left[1 - (1 - c_K)^2 - 2c_1^D c_1^V\right] \|\nu\|_{V^{-1}}^2 \\
 &= c_K^2 \|\nu\|_{V^{-1}}^2. \\
 \hookrightarrow \|[1 - \sigma) - K]\nu\|_{V^{-1}} &\leq c_K \|\nu\|_{V^{-1}}.
 \end{aligned}$$

□

现在来看子空间  $H_*^{\frac{1}{2}}(\Gamma)$  的情况, 定义如 (14.225)。根据单层位势的性质  $V : H_*^{-\frac{1}{2}}(\Gamma) \mapsto H_*^{\frac{1}{2}}(\Gamma)$ , 我们可以将双层位势  $\sigma I + K : H^{\frac{1}{2}}(\Gamma) \mapsto H^{\frac{1}{2}}(\Gamma)$  的测度 Theorem 14.38, (14.255), 转变为伴随双层位势  $\sigma I + K' : H^{-\frac{1}{2}}(\Gamma) \mapsto H^{-\frac{1}{2}}(\Gamma)$  的测度。

**Corollary 14.12** (伴随双层位势的收缩属性). 对于  $w \in H_*^{-\frac{1}{2}}(\Gamma)$  的伴随双层位势, 满足

$$(1 - c_K) \|w\|_{V^{-1}} \leq \|(\sigma I + K')w\|_V \leq c_K \|w\|_V. \quad (14.259)$$

证明. 1. 已知对于  $w \in H_*^{\frac{1}{2}}(\Gamma)$ , 存在唯一确定的  $\nu \in H_*^{\frac{1}{2}}(\Gamma)$ , 满足  $\nu = Vw$  或是  $\nu = V^{-1}\nu$ 。

由有界积分算子的对称性(14.220)可得

$$\begin{aligned}
 \|(\sigma I + K')w\|_V^2 &= \langle V(\sigma I + K')V^{-1}\nu, (\sigma I + K')V^{-1}\nu \rangle_\Gamma \\
 &= \langle V^{-1}(\sigma I + K)\nu, (\sigma I + K)\nu \rangle_\Gamma \\
 &= \|(\sigma I + K)\nu\|_{V^{-1}}^2.
 \end{aligned}$$

2.

$$\|w\|_V^2 = \langle Vw, w \rangle_\Gamma = \langle V^{-1}\nu, \nu \rangle_\Gamma = \|\nu\|_{V^{-1}}^2.$$

3. 代入 Corollary 14.11, 可证。

□

与在双层位势  $(\sigma I + K)$  延拓收缩属性 (Corollary 14.10) 相类似, 我们也可以将伴随双层位势  $(\sigma I + K')$  的收缩属性延拓到  $H^{-\frac{1}{2}}(\Gamma)$  中。

**Corollary 14.13** (伴随双层位势收缩比率). 对于  $w \in H^{-\frac{1}{2}}(\Gamma)$ , 伴随双层位势  $(\sigma I + K')$  满足如下不等式

$$\|(\sigma I + K')w\|_V \leq c_K \|w\|_V. \quad (14.260)$$

证明. 略。

□

同样地

**Corollary 14.14** (位移伴随双层位势收缩比率). 对于  $w \in H_*^{-\frac{1}{2}}(\Gamma)$ , 位移伴随双层位势  $(1 - \sigma)I - K'$  满足如下不等式

$$(1 - c_K) \|w\|_V \leq \|(1 - \delta)I - K'\|_V \leq c_K \|w\|_V, \quad (14.261)$$

收缩比率  $c_K$  的值由(14.256)给出。

证明. 略。

□

### 14.6.11 边界积分算子的映射特性

至今我们分析的一系列边界积分算子的特性，都是从牛顿位势的映射  $\tilde{N}_0 : \tilde{H}^{-1}(\Omega) \mapsto H^1(\Omega)$ ，以及逆定理和对偶配对而展开的。对于利普希茨域，我们可以得到更为通用的结论。

**Theorem 14.39** (牛顿位势在利普希茨域中的映射特性). 对于  $s \in [-2, 0]$ ，牛顿位势  $\tilde{N}_0 : \tilde{H}^s(\Omega) \mapsto H^{s+1}(\Omega)$  是一个连续映射，即

$$\|\tilde{N}_0 f\|_{H^{s+2}(\Omega)} \leq c \|f\|_{\tilde{H}^s(\Omega)}, \quad \forall f \in \tilde{H}^s(\Omega).$$

证明. 1. 设  $s \in [-1, 0]$ 。牛顿位势  $f \in H^s(\Omega)$  的定义如(14.160)。 $\tilde{f} \in H^s(\mathbb{R}^d)$  是  $f$  的延拓，我们有

$$\begin{aligned} \|\tilde{f}\|_{H^s(\mathbb{R}^d)} &= \sup_{0 \neq \nu \in H^{-s}(\mathbb{R}^d)} \frac{\langle \tilde{f}, \nu \rangle_{\mathbb{R}^d}}{\|\nu\|_{H^{-s}(\mathbb{R}^d)}} \\ &\leq \sup_{0 \neq \nu \in H^{-s}(\Omega)} \frac{\langle f, \nu \rangle_{\Omega}}{\|\nu\|_{H^{-s}(\Omega)}} \\ &= \|f\|_{H^s(\Omega)}, \end{aligned}$$

由牛顿位势算子的映射定理14.31得

$$\|\tilde{N}_0 f\|_{H^{s+2}(\Omega)} \leq c \|f\|_{\tilde{H}^s(\Omega)}, \quad \forall f \in \tilde{H}^s(\Omega).$$

2. 设  $s \in [-2, -1)$ 。由于牛顿位势  $\tilde{N}_0$  是自伴随算子，根据配偶配对我们有

$$\begin{aligned} \|\tilde{N}_0 f\|_{H^{s+2}(\Omega)} &= \sup_{0 \neq g \in \tilde{H}^{-2-s}(\Omega)} \frac{\langle \tilde{N}_0 f, g \rangle_{\Omega}}{\|g\|_{\tilde{H}^{-2-s}(\Omega)}} \\ &= \sup_{0 \neq g \in \tilde{H}^{-2-s}(\Omega)} \frac{\langle f, \tilde{N}_0 g \rangle_{\Gamma}}{\|g\|_{\tilde{H}^{-2-s}(\Omega)}} \\ &\leq \|f\|_{\tilde{H}^s(\Omega)} \sup_{0 \neq g \in \tilde{H}^{-2-s}(\Omega)} \frac{\|\tilde{N}_0 g\|_{H^{-s}(\Omega)}}{\|g\|_{\tilde{H}^{-2-s}(\Omega)}} \\ &\leq c \|f\|_{\tilde{H}^s(\Omega)} \end{aligned}$$

□

利用牛顿位势在利普希茨域中的映射特性 (定理14.39)，我们可以得到单层位势  $\tilde{V}$ (14.180)的映射特性。进而，利用  $\tilde{V}$  的迹，可得边界积分算子  $V := \gamma_0^{\text{int}} \tilde{V}$  的映射特性。

**Theorem 14.40** (单层位势算子的映射特性). 对于  $|s| < \frac{1}{2}$ ，单层位势  $V : H^{-\frac{1}{2}+s} \mapsto H^{\frac{1}{2}+s}$  有界，即

$$\|Vw\|_{H^{\frac{1}{2}+s}(\Gamma)} \leq c \|w\|_{H^{-\frac{1}{2}+s}(\Gamma)}, \quad \forall w \in H^{-\frac{1}{2}+s}(\Gamma).$$

证明. 1. 对于  $\varphi \in C^\infty(\Omega)$ ，有

$$\begin{aligned} \langle \tilde{V}w, \varphi \rangle_{\Omega} &= \int_{\Omega} \varphi(x) \int_{\Gamma} U^*(x, y) w(y) ds_y dx \\ &= \int_{\Gamma} w(y) \int_{\Gamma} U^*(x, y) \varphi(x) dx ds_y \\ &= \langle w, \gamma_0^{\text{int}} \tilde{N}_0 \varphi \rangle_{\Gamma} \\ &\leq \|w\|_{H^{-\frac{1}{2}+s}(\Gamma)} \left\| \gamma_0^{\text{int}} \tilde{N}_0 \varphi \right\|_{H^{\frac{1}{2}-s}(\Gamma)}. \end{aligned}$$

根据迹定理14.10可得

$$\begin{aligned}\langle \tilde{V}w, \varphi \rangle_{\Omega} &\leq \|w\|_{H^{-\frac{1}{2}+s}(\Gamma)} \left\| \gamma_0^{\text{int}} \tilde{N}_0 \varphi \right\|_{H^{\frac{1}{2}-s}(\Gamma)} \\ &\leq c_T \|w\|_{H^{-\frac{1}{2}+s}(\Gamma)} \left\| \tilde{N}_0 \varphi \right\|_{H^{1-s}(\Gamma)}.\end{aligned}$$

2. 利用 Theorem 14.39, 可得

$$\begin{aligned}\langle \tilde{V}w, \varphi \rangle_{\Omega} &\leq c_T \|w\|_{H^{-\frac{1}{2}+s}(\Gamma)} \left\| \tilde{N}_0 \varphi \right\|_{H^{1-s}(\Gamma)} \\ &\leq c \|w\|_{H^{-\frac{1}{2}+s}(\Gamma)} \|\varphi\|_{\tilde{H}^{-1-s}(\Omega)}, \forall \varphi \in C^{\infty}(\Omega).\end{aligned}$$

3. 将  $w$  视为密度方程, 根据定义可得  $\tilde{V}w \in H^{1+s}(\Omega)$ 。对  $\tilde{V}w$  取迹, 在满足  $\frac{1}{2} + s > 0$  的情况下, 我们有  $Vw : \gamma_0^{\text{int}} \tilde{V}w \in H^{\frac{1}{2}}(\Gamma)$ 。

□

对于利普希茨域  $\Omega$  的情况, 可由 Theorem 14.40求得全部边界积分算子的映射特性。

**Theorem 14.41** (边界积分算子的映射特性). 设利普希茨域  $\Omega$ , 对应边界  $\Gamma := \partial\Omega$ 。则对于  $s \in [-\frac{1}{2}, \frac{1}{2}]$ , 以下边界积分算子都有界, 且满足映射关系 (*Costabel, 1988*)

$$\begin{aligned}V : H^{-\frac{1}{2}+s}(\Gamma) &\mapsto H^{\frac{1}{2}+s}(\Gamma), \\ K : H^{\frac{1}{2}+s}(\Gamma) &\mapsto H^{\frac{1}{2}+s}(\Gamma), \\ K' : H^{-\frac{1}{2}+s}(\Gamma) &\mapsto H^{-\frac{1}{2}+s}(\Gamma), \\ D : H^{\frac{1}{2}+s}(\Gamma) &\mapsto H^{-\frac{1}{2}+s}(\Gamma).\end{aligned}$$

证明. 1.  $|s| < \frac{1}{2}$  时单层位势  $V$  的映射特性可由 Theorem 14.40证得。 $|s| = \frac{1}{2}$  时  $V$  的映射特性, 见 *Verchota (1984); McLean (2000)*。

2. 在此基础上,  $K, K', D$  的映射特性可由共法导数算子推得。已知单层位势  $u(x) = \tilde{V}w(x), x \in \Omega$  构成齐次偏微分方程的解, 对应狄利克雷数  $\gamma_0^{\text{int}} u(x) = (Vw)(x), x \in \Gamma$ 。

已知单层位势  $V : L^2(\Gamma) \mapsto L^1(\Gamma)$  是连续的, 根据 Theorem 14.24我们有

$$\|\gamma_1^{\text{int}} u\|_{L^2(\Gamma)} \leq c \|Vw\|_{H^1(\Gamma)} \leq \tilde{c} \|w\|_{L^2(\Gamma)},$$

进而可证得伴随双层位势算子  $\gamma_1^{\text{int}} \tilde{V} = \sigma I + K' : L^2(\Gamma) \mapsto L^2(\Gamma)$  的连续性。另一方面, 由对偶配对可得

$$\|\gamma_1^{\text{int}} u\|_{L^2(\Gamma)} = \sup_{0 \neq \varphi \in H^1(\Gamma)} \frac{\langle \gamma_1^{\text{int}} u, \varphi \rangle_{\Gamma}}{\|\varphi\|_{H^1(\Gamma)}}.$$

对于任一  $\varphi \in H^1(\Gamma)$ , 狄利克雷边界值问题

$$\begin{cases} L\nu(x) = 0 & x \in \Gamma, \\ \gamma_0^{\text{int}} \nu(x) = \varphi(x) & x \in \Gamma \end{cases}$$

有唯一解  $\nu \in H^{\frac{3}{2}}(\Omega)$ 。在此基础上, 由 Theorem 14.24可得

$$\|\gamma_1^{\text{int}} \nu\|_{L^2(\Gamma)} \leq c \|\gamma_0^{\text{int}} \nu\|_{H^1(\Gamma)} \leq c \|\varphi\|_{H^1(\Gamma)}.$$

既然  $u = \tilde{V}w$  和  $\nu$  都是齐次偏微分方程的解，连续两次使用格林第一恒等式(14.6)有

$$\begin{aligned}\langle \gamma_1^{\text{int}} u, \varphi \rangle_{\Gamma} &= a(u, \nu) \\ &= a(\nu, u) = \langle \gamma_1^{\text{int}} \nu, \gamma_0^{\text{int}} u \rangle_{\Gamma} \\ &\leq \|\gamma_1^{\text{int}} \nu\|_{L^2(\Gamma)} \|\gamma_0^{\text{int}} u\|_{L^2(\Gamma)} \\ &\leq c \|\varphi\|_{H^1(\Gamma)} \|Vw\|_{L^2(\Gamma)}.\end{aligned}$$

可见，由单层位势  $V : H^{-1}(\Gamma) \mapsto L^2(\Gamma)$  的连续性，我们有

$$\|\gamma_1^{\text{int}} u\|_{H^{-1}(\Gamma)} \leq c \|Vw\|_{L^2(\Gamma)} \leq \tilde{c} \|w\|_{H^{-1}(\Gamma)},$$

进而可得共法导数算子  $\gamma_1^{\text{int}} \tilde{V} : \sigma I + K' : H^{-1}(\Gamma) \mapsto H^{-1}(\Gamma)$  的连续性。在此基础上，若满足  $s \leq \frac{1}{2}$ ，那么利用插值定理可得  $K' : H^{-\frac{1}{2}+s}(\Gamma) \mapsto H^{-\frac{1}{2}+s}(\Gamma)$ 。

3. 由于

$$\begin{aligned}\|K\nu\|_{H^{\frac{1}{2}+s}(\Gamma)} &= \sup_{0 \neq w \in H^{-\frac{1}{2}-s}(\Gamma)} \frac{\langle K\nu, w \rangle_{\Gamma}}{\|w\|_{H^{-\frac{1}{2}-s}(\Gamma)}} \\ &= \sup_{0 \neq w \in H^{-\frac{1}{2}-s}(\Gamma)} \frac{\langle \nu, K'w \rangle_{\Gamma}}{\|w\|_{H^{-\frac{1}{2}-s}(\Gamma)}} \\ &= \|\nu\|_{H^{\frac{1}{2}+s}(\Gamma)} \sup_{0 \neq w \in H^{-\frac{1}{2}-s}(\Gamma)} \frac{\|K'w\|_{H^{-\frac{1}{2}-s}(\Gamma)}}{\|w\|_{H^{-\frac{1}{2}-s}(\Gamma)}} \\ &\leq c \|\nu\|_{H^{\frac{1}{2}+s}(\Gamma)},\end{aligned}$$

进而可得对于  $|s| \leq \frac{1}{2}$ ，有  $K : H^{\frac{1}{2}+s}(\Gamma) \mapsto H^{\frac{1}{2}+s}(\Gamma)$ 。

4. 由双层位势的定义  $u(x) = (Wu)(x)$  可得，利用 Theorem 14.24，有

$$\begin{aligned}\|D\nu\|_{L^2(\Gamma)} &= \|\gamma_1^{\text{int}} u\|_{L^2(\Gamma)} \\ &\leq c \|\gamma_0^{\text{int}} u\|_{H^1(\Gamma)} \\ &= c \|[(\sigma - 1)I + K]\nu\|_{H^1(\Gamma)},\end{aligned}$$

进而可得  $D : H^1(\Gamma) \mapsto L^2(\Gamma)$ 。进一步使用配偶配对以及插值法，得  $D : H^{\frac{1}{2}}(\Gamma) \mapsto H^{-\frac{1}{2}}(\Gamma)$ 。  $\square$

如果有界域  $\Omega \subset \mathbb{R}^d$  的边界  $\gamma = \partial\Omega$  是分段平滑的，那么即使  $|s|$  延展到  $\frac{1}{2}$  之外，Theorem 14.41 也可能成立。举例来说，若  $\Omega \subset \mathbb{R}^2$  是一个有界的多边形，有  $J$  个角，内夹角  $\alpha_j$ ，可定义

$$\sigma_0 := \min_{j=1,\dots,J} \left\{ \min \left[ \frac{\pi}{\alpha_j}, \frac{\pi}{2\pi - \alpha_j} \right] \right\},$$

此时 Theorem 14.41 对于所有  $|s| < \sigma_0$  均成立 (Costabel and Stephan, 1985)；如果边界  $\Gamma$  是  $C^\infty$  的，则 Theorem 14.41 对于所有  $s \in \mathbb{R}$  都成立。

## 14.7 边界积分方程

在对一系列边界积分算子有了初步了解后 (第14.1) 节，本届来分析边界积分方程 (BIEs, boundary integral equations)。具体来说，考虑由标量齐次偏微分方程表示的边界值问题

$$(Lu)(x) = 0, \quad x \in \Gamma, \tag{14.262}$$

其中  $L$  是一个椭圆、自伴随的二阶偏微分算子。 $\Omega$  是一个有界的单连通域 (simply connected domain)，其利普希茨边界定义为  $\Gamma := \partial\Omega$ 。<sup>19</sup>

对于某个  $\tilde{x} \in \Omega$ ，齐次偏微分方程(14.262)的解  $u(\tilde{x})$  可由表现式(14.151)给出

$$u(\tilde{x}) = \int_{\Gamma} U^*(\tilde{x}, y) \gamma_1^{\text{int}} u(y) ds_y - \int_{\Gamma} \gamma_{1,y}^{\text{int}} U^*(\tilde{x}, y) \gamma_0^{\text{int}} u(y) ds_y. \quad (14.263)$$

可见为了通过(14.263)求解(14.262)，一方面需要求得基本解  $U^*(\tilde{x}, y)$ ，另一方面我们需要完整柯西数  $[\gamma_0^{\text{int}} u(\tilde{x}), \gamma_1^{\text{int}} u(\tilde{x})]$ 。前者见第14.5节。就后者而言，计算完整柯西数所需的全部信息，可通过构建适宜的边界积分方程求得。以下述边界积分方程系统为例，

$$\begin{pmatrix} \gamma_0^{\text{int}} u \\ \gamma_1^{\text{int}} u \end{pmatrix} = \begin{pmatrix} (1-\sigma)I - K & V \\ D & \sigma I + K' \end{pmatrix} \begin{pmatrix} \gamma_0^{\text{int}} u \\ \gamma_1^{\text{int}} u \end{pmatrix}. \quad (14.264)$$

上式中，如果所有边界积分算子的密度方程都恰好是柯西数  $[\gamma_0^{\text{int}} u(x), \gamma_1^{\text{int}} u(x)]$ ,  $x \in \Gamma$ ，则我们称这种求解思路为直接法。

与之相对应，另一种求解思路为间接法，是指用适宜的位势算子求解边界值问题。例如，对于给定的  $\tilde{x} \in \Omega$ ，齐次偏微分方程(14.262)的解  $u(\tilde{x})$ ，可以由单层位势求解而得

$$u(\tilde{x}) = \int_{\Gamma} U^*(\tilde{x}, y) w(y) ds_y, \quad \tilde{x} \in \Omega, \quad (14.265)$$

或由双层位势求解而得

$$u(\tilde{x}) = - \int_{\Gamma} \gamma_{1,y}^{\text{int}} U^*(\tilde{x}, y) \nu(y) ds_y, \quad \tilde{x} \in \Omega, \quad (14.266)$$

需要指出的是，间接法中的密度方程  $w(y)$  和  $\nu(y)$  往往并无明确的经济学含义。

接下来我们将分别介绍几种不同的边界积分方程，以及如何用这些边界积分方程生成对应的柯西数，来描述带有不同边界条件的边界值问题。

### 14.7.1 狄利克雷边界值问题

考虑如下狄利克雷边界值问题

$$\begin{aligned} (Lu)(x) &= 0, \quad x \in \Omega, \\ \gamma_0^{\text{int}} u(x) &= g(x), \quad x \in \Gamma. \end{aligned} \quad (14.267)$$

#### 直接法求解狄利克雷边界值问题

使用直接法求解(14.267)， $u(x)$  的表现式由(14.263)改写为

$$u(\tilde{x}) = \int_{\Gamma} U^*(\tilde{x}, y) \gamma_1^{\text{int}} u(y) ds_y - \int_{\Gamma} \gamma_{1,y}^{\text{int}} U^*(\tilde{x}, y) g(y) ds_y, \quad \tilde{x} \in \Omega. \quad (14.268)$$

求解(14.268)需要首先计算未知的诺依曼数  $\gamma_1^{\text{int}} u \in H^{-\frac{1}{2}}(\Gamma)$ 。由边界积分方程系统 (卡尔德隆系统)(14.264)的第一行我们有

$$(V\gamma_1^{\text{int}} u)(x) = \sigma(x)g(x) + (Kg)(x), \quad x \in \Gamma, \quad (14.269)$$

我们称之为第一类弗雷德霍姆边界积分方程 (the first kind of Fredholm boundary integral function)(Atkinson, 1996, 1997)。

(14.269)的解  $u(x)$ ，其唯一性可由以下方式证得

<sup>19</sup>在这里我们以齐次泊松方程，即拉普拉斯方程为例来说明。求解非齐次偏微分方程，可引入牛顿位势，将非齐次偏微分方程中的全部牛顿位势化简为表面积位势，如见Jung and Steinbach (2002); Of et al. (2010)。

1. 单层位势算子  $V : H^{-\frac{1}{2}}(\Gamma) \mapsto H^{\frac{1}{2}}(\Gamma)$  有界(14.182),
2.  $V$  是  $H^{-\frac{1}{2}}(\Gamma)$ -椭圆的。 $d = 3$  的证明见 Theorem 14.35;  $d = 2$  的证明见 Theorem 14.36。
3. 已知  $V$  有界且椭圆, 由拉克斯—密格拉蒙定理 Theorem 14.14 可得  $u(x)$  的唯一可解性。

唯一可解的  $\gamma_1^{\text{int}} u(x) \in H^{-\frac{1}{2}}(\Gamma)$  满足

$$\begin{aligned} \|\gamma_1^{\text{int}} u\|_{H^{-\frac{1}{2}}(\Gamma)} &\leq \frac{1}{c_1^V} \|(\sigma I + K) g\|_{H^{\frac{1}{2}}(\Gamma)} \\ &\leq \frac{c_2^W}{c_1^V} \|g\|_{H^{\frac{1}{2}}(\Gamma)}. \end{aligned}$$

由于边界积分方程(14.269)存在于  $H^{\frac{1}{2}}(\Gamma)$  空间中, 则我们有

$$\begin{aligned} 0 &= \|V\gamma_1^{\text{int}} u - (\sigma I + K) g\|_{H^{\frac{1}{2}}(\Gamma)} \\ &= \sup_{0 \neq \tau \in H^{-\frac{1}{2}}(\Gamma)} \frac{\langle V\gamma_1^{\text{int}} u - (\sigma I + K) g, \tau \rangle_{\Gamma}}{\|\tau\|_{H^{-\frac{1}{2}}(\Gamma)}}, \end{aligned}$$

由此可见, 除了直接求解(14.269)之外, 我们也可以考虑如下等价的变分问题: 寻找解  $\gamma_1^{\text{int}} u \in H^{-\frac{1}{2}}(\Gamma)$ , 使其满足

$$\begin{aligned} \langle V\gamma_1^{\text{int}} u, \tau \rangle_{\Gamma} &= \langle (\sigma I + K) g, \tau \rangle_{\Gamma} \\ &= \left\langle \left( \frac{1}{2} I + K \right) g, \tau \right\rangle_{\Gamma}, \quad \forall \tau \in H^{-\frac{1}{2}}(\Gamma), \end{aligned} \tag{14.270}$$

其中第二行等式是由于, 根据(14.188), 对于几乎所有  $x \in \Gamma$  都有  $\sigma(x) = \frac{1}{2}$ 。

除了根据卡尔德隆系统(14.264)的第一行, 构建(14.269)来求解狄利克雷边界值之外, 另一种方法是根据第二行计算未知的诺依曼数  $\gamma_1^{\text{int}} u \in H^{-\frac{1}{2}}(\Gamma)$ , 即

$$[(1 - \sigma) I - K'] \gamma_1^{\text{int}} u(x) = (Dg)(x), \quad x \in \Gamma, \tag{14.271}$$

我们称之为第二类弗雷德霍姆边界积分方程 (the second kind of Fredholm boundary integral function)(Atkinson, 1996, 1997)。方程的解  $\gamma_1^{\text{int}} u(x)$  由诺依曼级数 (Neumann series) 给出

$$\gamma_1^{\text{int}} u(x) = \sum_{\ell=0}^{\infty} (\sigma I + K')^{\ell} (Dg)(x), \quad x \in \Gamma. \tag{14.272}$$

已知伴随双层位势  $(\sigma I + K')$  的收缩特性可由单层位势  $V$  中等价范的形式  $\|\cdot\|_V$  来表示, 见 Corollary 14.13 式(14.260), 则诺依曼级数(14.272)在  $H^{-\frac{1}{2}}(\Gamma)$  中收敛。

### 间接法求解狄利克雷边界值问题

利用间接法, 通过单层位势  $V$ (14.265)求解狄利克雷边界值问题 (齐次偏微分方程), 前提是需要计算未知的密度方程  $w \in H^{-\frac{1}{2}}(\Gamma)$ , 其关键是求解如下边界积分方程

$$(Vw)(x) = g(x), \quad x \in \Gamma. \tag{14.273}$$

将间接法之一(14.273), 与直接法之一 (14.269)相比, 只是在 RHS 的定义上有所不同。换句话说, (14.273)解的唯一存在性, 可由(14.269)解的唯一存在性而证得。

除了利用单层位势  $V$ , 也可以利用双层位势  $(\sigma I + K)(14.266)$ 。根据双层位势的跃动关系 Lemma 14.25 式(14.192), 可建立如下边界积分方程

$$[1 - \sigma(x)] \nu(x) - (K\nu)(x) = g(x), \quad x \in \Gamma, \tag{14.274}$$

其密度方程  $\nu(x) \in H^{\frac{1}{2}}(\Gamma)$  由诺依曼级数 (Neumann series) 算得

$$\nu(x) = \sum_{\ell=0}^{\infty} (\sigma I + K')^{\ell} g(x), \quad \forall x \in \Gamma. \quad (14.275)$$

类似地, 已知双层位势  $(\sigma I + K)$  的收缩属性可由逆单层位势  $V^{-1}$  中等价范的形式  $\|\cdot\|_{V^{-1}}$  来表示, 见 Corollary 14.10 式(14.257), 则诺依曼级数(14.275)在  $H^{\frac{1}{2}}(\Gamma)$  中收敛, 进而

$$\begin{aligned} 0 &= \|(1 - \sigma)\nu - K\nu - g\|_{H^{\frac{1}{2}}(\Gamma)} \\ &= \sup_{0 \neq \tau \in H^{-\frac{1}{2}}(\Gamma)} \frac{\langle \frac{1}{2}\nu - K\nu - g, \tau \rangle_{\Gamma}}{\|\tau\|_{H^{-\frac{1}{2}}(\Gamma)}}, \end{aligned}$$

其中根据(14.188), 对于几乎所有  $x \in \Gamma$  都有  $\sigma(x) = \frac{1}{2}$ 。

因此除了直接求解边界积分方程(14.273), 也可以考虑如下等价变分问题: 寻找解  $\nu \in H^{\frac{1}{2}}(\Gamma)$ , 使满足

$$\left\langle \left( \frac{1}{2}I - K \right) \nu, \tau \right\rangle_{\Gamma} = \langle g, \tau \rangle_{\Gamma}, \quad \forall \tau \in H^{-\frac{1}{2}}(\Gamma). \quad (14.276)$$

**Lemma 14.32** (变分问题的稳定条件). 变分问题(14.276)有稳定条件

$$c_S \|\nu\|_{H^{\frac{1}{2}}(\Gamma)} \leq \sup_{0 \neq \tau \in H^{-\frac{1}{2}}(\Gamma)} \frac{\langle \left( \frac{1}{2}I - K \right) \nu, \tau \rangle_{\Gamma}}{\|\tau\|_{H^{-\frac{1}{2}}(\Gamma)}}, \quad \forall \nu \in H^{\frac{1}{2}}(\Gamma), \quad (14.277)$$

其中  $c_S$  是个正常数。

证明. 设任一  $\nu \in H^{\frac{1}{2}}(\Gamma)$ , 定义  $\tau_{\nu} := V^{-1}\nu \in H^{-\frac{1}{2}}(\Gamma)$ , 则有

$$\|\tau_{\nu}\|_{H^{-\frac{1}{2}}(\Gamma)} = \|V^{-1}\nu\|_{H^{-\frac{1}{2}}(\Gamma)} \leq \frac{1}{c_1^V} \|\nu\|_{H^{\frac{1}{2}}(\Gamma)}.$$

由双层位势的收缩比率 Corollary 14.10 式(14.257), 以及单层位势  $V$  的映射特性 Theorem 14.41, 我们有

$$\begin{aligned} \left\langle \left( \frac{1}{2}I - K \right) \nu, \tau_{\nu} \right\rangle_{\Gamma} &= \left\langle \left( \frac{1}{2}I - K \right) \nu, V^{-1}\nu \right\rangle_{\Gamma} \\ &= \langle V^{-1}\nu, \nu \rangle_{\Gamma} - \langle V^{-1} \left( \frac{1}{2}I + K \right) \nu, \nu \rangle_{\Gamma} \\ &\geq \|\nu\|_{V^{-1}}^2 - \left\| \left( \frac{1}{2}I + K \right) \nu \right\|_{V^{-1}} \|\nu\|_{V^{-1}} \\ &\geq (1 - c_K) \|\nu\|_{V^{-1}}^2 \\ &\geq (1 - c_K) \langle V^{-1}\nu, \nu \rangle_{\Gamma} \\ &\geq (1 - c_K) \frac{1}{c_2^V} \|\nu\|_{H^{\frac{1}{2}}(\Gamma)}^2 \\ &\geq (1 - c_K) \frac{c_1^V}{c_2^V} \|\nu\|_{H^{\frac{1}{2}}(\Gamma)} \|\tau_{\nu}\|_{H^{-\frac{1}{2}}(\Gamma)}, \end{aligned}$$

证毕。  $\square$

在此基础上, 变分问题(14.276)的唯一可解性, 可由 Theorem 14.16 证得。

**Remark 14.1.** 对于狄利克雷边界值问题(14.267), 我们介绍了 4 种求解方法, 对应 4 个边界积分方程。这四种解都是唯一的。随着研究对象的不同和所采用的离散化方法 (第 12.D 节) 的不同, 这些方法各有利弊。在随后的分析中, 我们以(14.270)方法为主, 作进一步分析。

### 14.7.2 諾依曼边界值问题

考慮如下諾依曼边界值問題

$$\begin{aligned} Lu(x) &= 0, \quad x \in \Gamma, \\ \gamma_1^{\text{int}} u(x) &= g(x), \quad x \in \Gamma. \end{aligned} \quad (14.278)$$

在滿足假定可解性条件(14.17)

$$\int_{\Gamma} g(x) ds_x = 0. \quad (14.279)$$

的情况下，諾依曼边界值問題 (14.278)的解可由表現式(14.263)改写为

$$u(\tilde{x}) = \int_{\Gamma} U^*(\tilde{x}, y) g(y) ds_y - \int_{\Gamma} \gamma_{1,y}^{\text{int}} U^*(\tilde{x}, y) \gamma_0^{\text{int}} u(y) ds_y, \quad \tilde{x} \in \Omega. \quad (14.280)$$

为了求解(14.280)，我们需要计算諾依曼数  $\gamma_0^{\text{int}} u \in H^{\frac{1}{2}}(\Gamma)$ 。由卡尔德隆系统(14.264)的第二行可得第一类弗雷德霍姆边界积分方程

$$(D\gamma_0^{\text{int}} u)(x) = (1 - \sigma(x)) g(x) - (K' g)(x), \quad x \in \Gamma. \quad (14.281)$$

先来证明(14.281)存在唯一解  $\gamma_0^{\text{int}} u \in H_*^{\frac{1}{2}}(\Gamma)$ 。

1. 证明可解。根据(14.203)， $u_0 \equiv 1$  是超奇异边界积分算子  $D$  的特征解，满足  $(Du_0)(x) = 0$ 。因此可得  $\ker D = \text{span}\{u_0\}$ 。根据闭值域定理 Theorem 14.15，我们需要假设如下可解条件

$$(1 - \sigma) g - K' g \in \text{Im}(D) = (\ker D)^0,$$

其中  $(\ker D)^0$  是  $v$  的正交空间，定义见(14.82)。由正交属性可得

$$\begin{aligned} \langle (1 - \sigma) g - K' g, u_0 \rangle_{\Gamma} &= \langle g, 1 \rangle_{\Gamma} - \langle (\sigma I + K') g, u_0 \rangle_{\Gamma} \\ &= \langle g, 1 \rangle_{\Gamma} - \langle g, (\sigma I + K) u_0 \rangle_{\Gamma} \\ &= 0, \end{aligned} \quad (14.282)$$

可证的边界积分方程(14.281)可解。

2. 超奇异积分算子  $D : H^{\frac{1}{2}}(\Gamma) \mapsto H^{-\frac{1}{2}}(\Gamma)$  有界(14.199)，且  $H_*^{\frac{1}{2}}(\Gamma)$ -椭圆 (Theorem 14.14)。
3. 由拉克斯—密格拉蒙定理 Theorem 14.14 可得，边界积分方程(14.281)存在唯一解  $\gamma_0^{\text{int}} u \in H_*^{\frac{1}{2}}(\Gamma)$ 。

在此基础上，求解諾依曼边界值問題(14.278)，便等价于构建一个带有约束条件的变分问题：寻找解  $\gamma_0^{\text{int}} u \in H_*^{\frac{1}{2}}(\Gamma)$ ，使满足

$$\langle D\gamma_0^{\text{int}} u, \nu \rangle_{\Gamma} = \left\langle \left( \frac{1}{2}I - K' \right) g, \nu \right\rangle_{\Gamma}, \quad \forall \nu \in H_*^{\frac{1}{2}}(\Gamma), \quad (14.283)$$

带约束条件的变分问题又等价于以下鞍点变分问题：寻找解  $(\gamma_0^{\text{int}} u, \lambda) \in H^{\frac{1}{2}}(\Gamma) \times \mathbb{R}$ ，使满足

$$\begin{aligned} \langle D\gamma_0^{\text{int}} u, \nu \rangle_{\Gamma} + \lambda \langle \nu, w_{\text{eq}} \rangle_{\Gamma} &= \left\langle \left( \frac{1}{2}I - K' \right) g, \nu \right\rangle_{\Gamma}, \\ \langle \gamma_0^{\text{int}} u, w_{\text{eq}} \rangle_{\Gamma} &= 0, \quad \forall \nu \in H^{\frac{1}{2}}(\Gamma). \end{aligned} \quad (14.284)$$

在(14.284)的第一行中，设  $\nu = u_0 \in H^{\frac{1}{2}}(\Gamma)$  作为检测方程，我们有  $Du_0 = 0$ 。进而，由正交关系(14.282)可得

$$0 = \lambda \langle 1, w_{\text{eq}} \rangle_{\Gamma} = \lambda \langle 1, V^{-1} 1 \rangle_{\Gamma}, \quad (14.285)$$

由于逆单层位势  $V^{-1}$  是椭圆的，根据上式我们有  $\lambda = 0$ 。进而鞍点变分问题(14.284)变为调整鞍点变分问题：寻找解  $(\gamma_0^{\text{int}} u, \lambda) \in H^{\frac{1}{2}}(\Gamma) \times \mathbb{R}$ ，使满足

$$\begin{aligned} \langle D\gamma_0^{\text{int}} u, \nu \rangle_{\Gamma} + \lambda \langle \nu, w_{\text{eq}} \rangle_{\Gamma} &= \left\langle \left( \frac{1}{2} I - K' \right) g, \nu \right\rangle_{\Gamma}, \\ \langle \gamma_0^{\text{int}} u, w_{\text{eq}} \rangle_{\Gamma} - \frac{\lambda}{\alpha} &= 0, \quad \forall \nu \in H^{\frac{1}{2}}(\Gamma), \end{aligned} \quad (14.286)$$

其中  $\alpha \in \mathbb{R}_+$  是某个待选参数。将第二行代回第一行以消除拉格朗日乘数  $\lambda$ ，从而得到最终的调整变分问题：寻找解  $\gamma_0^{\text{int}} u \in H^{\frac{1}{2}}(\Gamma)$ ，使满足

$$\langle D\gamma_0^{\text{int}} u, \nu \rangle_{\Gamma} + \underbrace{\alpha \langle \gamma_0^{\text{int}} u, w_{\text{eq}} \rangle_{\Gamma}}_{=: \lambda} \langle \nu, w_{\text{eq}} \rangle_{\Gamma} = \left\langle \left( \frac{1}{2} I - K' \right) g, \nu \right\rangle_{\Gamma}, \quad \forall \nu \in H^{\frac{1}{2}}(\Gamma). \quad (14.287)$$

来看调整鞍点变分问题(14.287)。定义一个调整超奇异边界积分算子  $\tilde{D} : H^{\frac{1}{2}}(\Gamma) \mapsto H^{-\frac{1}{2}}(\Gamma)$

$$\langle \tilde{D}w, \nu \rangle_{\Gamma} := \langle Dw, \nu \rangle_{\Gamma} + \alpha \langle w, w_{\text{eq}} \rangle_{\Gamma} \langle \nu, w_{\text{eq}} \rangle_{\Gamma},$$

由于

$$\begin{aligned} \langle \tilde{D}\nu, \nu \rangle_{\Gamma} &= \langle D\nu, \nu \rangle_{\Gamma} + \alpha [\langle \nu, w_{\text{eq}} \rangle_{\Gamma}]^2 \\ &\geq \bar{c}_1^D |\nu|_{H^{\frac{1}{2}}(\Gamma)}^2 + \alpha [\langle \nu, w_{\text{eq}} \rangle_{\Gamma}]^2 \\ &\geq \min \{ \bar{c}_1^D, \alpha \} \left\{ |\nu|_{H^{\frac{1}{2}}(\Gamma)}^2 + [\langle \nu, w_{\text{eq}} \rangle_{\Gamma}]^2 \right\} \\ &= \min \{ \bar{c}_1^D, \alpha \} \|\nu\|_{H_*^{\frac{1}{2}}(\Gamma)} \\ &\geq \hat{c}_1^D \|\nu\|_{H_*^{\frac{1}{2}}(\Gamma)}, \quad \forall \nu \in H^{\frac{1}{2}}(\Gamma), \end{aligned}$$

我们得  $\tilde{D}$  有界且  $H^{\frac{1}{2}}(\Gamma)$ -椭圆  $\forall w, \nu \in H^{\frac{1}{2}}(\Gamma)$ 。这意味着，在消除拉格朗日乘数  $\lambda$ ，从调整鞍点变分问题(14.286)变为(14.287)的过程中，任何给定的 RHS，在左侧都有唯一一个  $\alpha$  值与之相对应。换句话说，对于任何的诺依曼数  $\gamma_1^{\text{int}} u = g \in H^{-\frac{1}{2}}(\Gamma)$ ，(14.287)都存在唯一解  $\gamma_0^{\text{int}} u \in H^{\frac{1}{2}}(\Gamma)$ 。

现在假定有一组给定的诺依曼数  $\gamma_1^{\text{int}} u = g \in H^{-\frac{1}{2}}(\Gamma)$ ，满足可解性条件(14.279)。那么，通过引入检测方程  $\nu = u_0 \equiv 1$ ，调整鞍点变分问题(14.287)可以变为

$$\begin{aligned} \alpha \langle \gamma_0^{\text{int}} u, w_{\text{eq}} \rangle_{\Gamma} \langle 1, w_{\text{eq}} \rangle_{\Gamma} &= 0, \\ \langle 1, w_{\text{eq}} \rangle_{\Gamma} &= \langle 1, V^{-1} 1 \rangle_{\Gamma} > 0, \end{aligned}$$

因此我们有  $\gamma_0^{\text{int}} u \in H_*^{\frac{1}{2}}(\Gamma)$ 。可见调整变分问题(14.287)与原变分问题(14.284)等价。

根据 Corollary 14.9，超奇异边界积分算子是  $H_{**}^{\frac{1}{2}(\Gamma)}$ -椭圆的，则(14.281)也存在唯一的解  $\gamma_0^{\text{int}} \in H_{**}^{\frac{1}{2}(\Gamma)}$ 。这让我们可以用与上文相近的思路，构建一个子空间中的调整变分问题：寻找解  $\gamma_0^{\text{int}} \in H^{\frac{1}{2}}(\Gamma)$ ，使满足

$$\langle D\gamma_0^{\text{int}} u, \nu \rangle_{\Gamma} + \bar{\alpha} \langle \gamma_0^{\text{int}} u, 1 \rangle_{\Gamma} \langle \nu, 1 \rangle_{\Gamma} = \left\langle \left( \frac{1}{2} I - K' \right) g, \nu \right\rangle_{\Gamma}, \quad \forall \nu \in H^{\frac{1}{2}}(\Gamma), \quad (14.288)$$

同样地， $\bar{\alpha}$  是个待选参数。在此基础上，如果假定可解性条件(14.279)得到满足，那么变分问题(14.288)的解  $\gamma_0^{\text{int}} \in H_{**}^{\frac{1}{2}}(\Gamma)$ 。

定义一个调整超奇异积分算子  $\hat{D} : H^{\frac{1}{2}}(\Gamma) \mapsto H^{-\frac{1}{2}}(\Gamma)$  如下

$$\langle \hat{D}w, \nu \rangle_{\Gamma} := \langle Dw, \nu \rangle_{\Gamma} + \bar{\alpha} \langle w, 1 \rangle_{\Gamma} \langle \nu, 1 \rangle_{\Gamma}, \quad \forall w, \nu \in H^{\frac{1}{2}}(\Gamma), \quad (14.289)$$

可见  $\hat{D}$  也是有界且  $H^{\frac{1}{2}}(\Gamma)$ -椭圆的算子。

如果我们使用间接双层位势(14.266)来计算未知的密度方程  $\nu \in H_*^{\frac{1}{2}}(\Gamma)$ , 则我们可以得到超奇异边界积分算子形式的边界积分方程

$$(D\nu)(x) = g(x), \quad x \in \gamma, \quad (14.290)$$

进而我们可以采取与分析 BIE(14.281)相近似的方法, 来分析 BIE(14.290)。

如果我们使用直接法的表现式(14.265)来计算  $\nu \in H_*^{\frac{1}{2}}(\Gamma)$ , 那么代入卡尔德隆系统(14.264)第一行可得边界积分方程

$$(\sigma I + K)\gamma_0^{\text{int}}u(x) = (V\gamma_1^{\text{int}}u)(x) = (Vg)(x), \quad x \in \Gamma, \quad (14.291)$$

其中待求解为狄利克雷数  $\gamma_0^{\text{int}}u(x), x \in \Gamma$ 。(14.291)的解, 可由诺依曼级数求得

$$\gamma_0^{\text{int}}u(x) = \sum_{\ell=0}^{\infty} [(1-\sigma)I - K]^{\ell}(Vg)(x), \quad \forall x \in \Gamma. \quad (14.292)$$

诺依曼级数(14.292)的收敛性证明: 已知位移双层位势  $[(1-\sigma)I - K]$  的收缩特性可由逆单层位势  $V^{-1}$  中的等价范形式  $\|\cdot\|_{V^{-1}}$  来表示, 见 Corollary 14.10 式(14.257)。则可以在  $H^{\frac{1}{2}}(\Gamma)$  中构建基于(14.291)的变分问题: 寻找解  $\gamma_0^{\text{int}}u \in H_*^{\frac{1}{2}}(\Gamma)$ , 使满足

$$\langle (\sigma I + K)\gamma_0^{\text{int}}u, \nu \rangle_{V^{-1}} = \langle Vg, \nu \rangle_{V^{-1}}, \quad \forall \nu \in H_*^{\frac{1}{2}}(\Gamma), \quad (14.293)$$

其中已知单层位势  $V : H^{-\frac{1}{2}}(\Gamma) \mapsto H^{\frac{1}{2}}(\Gamma)$  有界且  $H^{-\frac{1}{2}}(\Gamma)$ -椭圆,  $\langle \cdot, \cdot \rangle_{V^{-1}}$  定义为  $H^{\frac{1}{2}}(\Gamma)$  中的内积形式

$$\langle w, \nu \rangle_{V^{-1}} := \langle V^{-1}w, \nu \rangle_{\Gamma}, \quad w, \nu \in H^{\frac{1}{2}}(\Gamma),$$

那么变分问题(14.293)可以调整为如下变分问题: 寻找  $\gamma_0^{\text{int}}u \in H_*^{\frac{1}{2}}(\Gamma)$ , 使满足

$$\begin{aligned} \langle S\gamma_0^{\text{int}}u, \nu \rangle_{\Gamma} &= \langle V^{-1}(\sigma I + K)\gamma_0^{\text{int}}u, \nu \rangle_{\Gamma} \\ &= \langle g, \nu \rangle_{\Gamma}, \quad \forall \nu \in H_*^{\frac{1}{2}}(\Gamma), \end{aligned} \quad (14.294)$$

其中  $S : H^{\frac{1}{2}}(\Gamma) \mapsto H^{-\frac{1}{2}}(\Gamma)$  是 Steklov-Poincaré 算子。

由于 Steklov-Poincaré 算子  $S$  和超奇异边界积分算子  $D : H^{\frac{1}{2}}(\Gamma) \mapsto H^{-\frac{1}{2}}(\Gamma)$  有同样的映射特性, 见式(14.245)和 Theorem 14.41。那么对变分问题(14.294)的唯一可解性的证明, 便类似于对变分问题(14.283)唯一可解性的证明过程。

将 Steklov-Poincaré 算子的对称型表现式(14.248)代入调整变分问题(14.294)得

$$\begin{aligned} \langle S\gamma_0^{\text{int}}u, \nu \rangle_{\Gamma} &= \langle [D + (\sigma I + K')V^{-1}(\sigma I + K)]\gamma_0^{\text{int}}u, \nu \rangle_{\Gamma} \\ &= \langle g, \nu \rangle_{\Gamma}, \quad \forall \nu \in H_*^{\frac{1}{2}}(\Gamma), \end{aligned} \quad (14.295)$$

不难看出, 在根据单层位势(14.265), 使用间接法求解诺依曼边界值问题是, 我们最终得到以下边界积分方程

$$(\sigma I + K')w(x) = g(x), \quad x \in \Gamma, \quad (14.296)$$

其中未知的密度方程  $w \in H^{-\frac{1}{2}}(\Gamma)$  可由诺依曼级数求得

$$w(x) = \sum_{\ell=0}^{\infty} [(1-\sigma)I - K']^{\ell}g(x), \quad x \in \Gamma, \quad (14.297)$$

在  $H^{-\frac{1}{2}}(\Gamma)$  空间中, 转移伴随双层位势  $[(1-\sigma)I - K']$  的收缩特性, 可由逆单层位势  $V^{-1}$  形式的等价范  $\|\cdot\|_{V^{-1}}$  Corollary (14.14) 式(14.261)求得。由此可得诺依曼级数(14.297)的收敛特性。

**Remark 14.2.** 对于诺依曼边界值问题(14.278), 我们介绍了 4 种求解方法, 对应 4 个边界积分方程。这四种解都是唯一的。随着研究对象的不同和所采用的离散化方法 (第 12.D 节) 的不同, 这些方法各有利弊。在随后的分析中, 我们以(14.288)方法为主, 作进一步分析。

### 14.7.3 混合边界值问题

实际应用中的边界值问题，常常是两种边界条件，即狄利克雷边界条件和诺依曼边界条件的混合，

$$\begin{aligned} Lu(x) &= 0, \quad x \in \Omega, \\ \gamma_0^{\text{int}} u(x) &= g_D(x), \quad x \in \Gamma_D, \\ \gamma_1^{\text{int}} u(x) &= g_N(x), \quad x \in \Gamma_N. \end{aligned} \quad (14.298)$$

根据表现式(14.262)，混合边界值问题的解  $u(\tilde{x}), \tilde{x} \in \Omega$  可以写成如下表现形式

$$\begin{aligned} u(\tilde{x}) &= \int_{\Gamma_N} U^*(\tilde{x}, y) g_N(y) ds_y + \int_{\Gamma_D} U^*(\tilde{x}, y) \gamma_1^{\text{int}} u(y) ds_y \\ &\quad - \int_{\Gamma_D} \gamma_1^{\text{int}} U^*(\tilde{x}, y) g_D(y) ds_y - \int_{\Gamma_N} \gamma_1^{\text{int}} U^*(\tilde{x}, y) \gamma_0^{\text{int}} u(y) ds_y, \end{aligned} \quad (14.299)$$

可见为了求解混合边界值问题(14.298)，需要计算的未知数有两个，分别是狄利克雷数  $\gamma_0^{\text{int}} u(x), x \in \Gamma_D$  和诺依曼数  $\gamma_1^{\text{int}} u(x), x \in \Gamma_N$ 。如前文所介绍的，狄利克雷边界值问题和诺依曼边界值问题分别有几种求解方法可供选择。对应地，可将这些方法之间搭配组合来求混合边界值问题。举两个例子。

#### 直接法

采取两种直接法的组合，构建如下对称公式：对于  $x \in \Gamma_D$ ，采用卡尔德隆系统(14.264)的第一行求解；对于  $x \in \Gamma_N$ ，采用第二行求解，即

$$\begin{aligned} \underbrace{(V\gamma_1^{\text{int}} u)(x)}_{=: (Vg_N)(x)} &= (\sigma I + K) \underbrace{\gamma_0^{\text{int}} u(x)}_{=: g_D(x)}, \quad x \in \Gamma D, \\ \underbrace{(D\gamma_0^{\text{int}} u)(x)}_{=: (Dg_D)(x)} &= [(1 - \sigma) I - K'] \underbrace{\gamma_1^{\text{int}} u(x)}_{=: g_N(x)}, \quad x \in \Gamma N. \end{aligned} \quad (14.300)$$

对于已知的  $g_D \in H^{\frac{1}{2}}(\Gamma_D)$ ,  $g_N \in H^{-\frac{1}{2}}(\Gamma_N)$ , 分别定义二者在  $\Gamma$  中的适宜的延拓  $\tilde{g}_D \in H^{\frac{1}{2}}(\Gamma_D)$ ,  $\tilde{g}_N \in H^{-\frac{1}{2}}(\Gamma_N)$ , 满足

$$\begin{aligned} \tilde{g}_D(x) &= g_D(x), \quad x \in \Gamma_D, \\ \tilde{g}_N(x) &= g_N(x), \quad x \in \Gamma_N. \end{aligned}$$

设

$$\begin{aligned} \tilde{u} &:= \gamma_0^{\text{int}} u - \tilde{g}_D \in \tilde{H}^{\frac{1}{2}}(\Gamma_N), \\ \tilde{t} &:= \gamma_1^{\text{int}} u - \tilde{g}_N \in \tilde{H}^{-\frac{1}{2}}(\Gamma_D), \end{aligned}$$

则边界积分方程系统(14.300)变为

$$\begin{aligned} (V\tilde{t})(x) - (K\tilde{u})(x) &= (\sigma I + K)\tilde{g}_D(x) - (V\tilde{g}_N)(x), \quad x \in \Gamma_D, \\ (D\tilde{u})(x) + (K'\tilde{t})(x) &= [(1 - \sigma)I - K']\tilde{g}_N(x) - (D\tilde{g}_D)(x), \quad x \in \Gamma_N. \end{aligned} \quad (14.301)$$

显然，新的边界积分方程系统(14.301)可以用一个变分问题来表示：寻找解  $(\tilde{t}, \tilde{u}) \in \tilde{H}^{-\frac{1}{2}}(\Gamma_D) \times \tilde{H}^{\frac{1}{2}}(\Gamma_N)$ ，使得  $\forall (\tau, \nu) \in \tilde{H}^{-\frac{1}{2}}(\Gamma_D) \times \tilde{H}^{\frac{1}{2}}(\Gamma_N)$ ，均满足

$$\begin{aligned} a(\tilde{t}, \tilde{u}; \tau, \nu) &= F(\tau, \nu), \\ a(\tilde{t}, \tilde{u}; \tau, \nu) &:= \langle V\tilde{t}, \tau \rangle_{\Gamma_D} - \langle K\tilde{u}, \tau \rangle_{\Gamma_D} + \langle K'\tilde{t}, \nu \rangle_{\Gamma_N} + \langle D\tilde{u}, \nu \rangle_{\Gamma_N}, \\ F(\tau, \nu) &= \left\langle \left( \frac{1}{2}I + K \right) \tilde{g}_D - (V, \tilde{g}_D), \tau \right\rangle_{\Gamma_D} + \left\langle \left( \frac{1}{2}I - K' \right) \tilde{g}_N - (D\tilde{g}_D), \nu \right\rangle_{\Gamma_N}. \end{aligned} \quad (14.302)$$

**Lemma 14.33.** 式(14.302)中, 对称边界积分算子的双线性形式  $a(.,.;.,.)$  有界, 并且  $\tilde{H}^{-\frac{1}{2}}(\Gamma_D) \times \tilde{H}^{\frac{1}{2}}(\Gamma_N)$ -椭圆, 即

$$\begin{aligned} a(t, u; \tau, \nu) &\leq c_2^A \| (t, u) \|_{\tilde{H}^{-\frac{1}{2}}(\Gamma_D) \times \tilde{H}^{\frac{1}{2}}(\Gamma_N)} \| (\tau, \nu) \|_{\tilde{H}^{-\frac{1}{2}}(\Gamma_D) \times \tilde{H}^{\frac{1}{2}}(\Gamma_N)} \\ &\quad \forall (t, u), (\tau, \nu) \in \tilde{H}^{-\frac{1}{2}}(\Gamma_D) \times \tilde{H}^{\frac{1}{2}}(\Gamma_N), \end{aligned} \quad (14.303)$$

并且

$$a(\tau, \nu; \tau, \nu) \geq \min \{c_1^V, \hat{c}_1^D\} \| (\tau, \nu) \|_{\tilde{H}^{-\frac{1}{2}}(\Gamma_D) \times \tilde{H}^{\frac{1}{2}}(\Gamma_N)}^2, \quad (14.304)$$

其中范数形式定义为

$$\| (\tau, \nu) \|_{\tilde{H}^{-\frac{1}{2}}(\Gamma_D) \times \tilde{H}^{\frac{1}{2}}(\Gamma_N)}^2 := \| \tau \|_{\tilde{H}^{-\frac{1}{2}}(\Gamma_D)}^2 + \| \nu \|_{\tilde{H}^{\frac{1}{2}}(\Gamma_N)}^2.$$

证明. 根据定义式(14.302)

$$\begin{aligned} a(\tau, \nu; \tau, \nu) &= \langle V\tau, \tau \rangle_{\Gamma_D} - \langle K\nu, \tau \rangle_{\Gamma_D} + \langle K'\tau, \nu \rangle_{\Gamma_N} + \langle D\nu, \nu \rangle_{\Gamma_N} \\ &= \langle V\tau, \tau \rangle_{\Gamma_D} + \langle D\nu, \nu \rangle_{\Gamma_N} \\ &\geq \underbrace{c_1^V \| \tau \|_{\tilde{H}^{-\frac{1}{2}}(\Gamma_D)}^2}_{=: \mathcal{A}} + \underbrace{\hat{c}_1^D \| \nu \|_{\tilde{H}^{\frac{1}{2}}(\Gamma_N)}^2}_{=: \mathcal{B}}. \end{aligned}$$

最后一行中,  $\mathcal{A}$  由单层位势算子  $V$  的椭圆性求得, 见 Theorem 14.36-14.35;  $\mathcal{B}$  由超奇异边界积分算子  $D$  的椭圆性求得, 见(14.243)。

由  $V, D$  的有界特性可得  $a(.,.;.,.)$  有界。  $\square$

**Lemma 14.34.** 变分问题(14.302)是唯一可解的。

证明.  $\forall (\tau, \nu) \in \tilde{H}^{-\frac{1}{2}}(\Gamma_D) \times \tilde{H}^{\frac{1}{2}}(\Gamma_N)$ ,  $F(\tau, \nu) =$  都有界。并且根据 Lemma 14.303,  $a(.,.;.,.)$  有界且椭圆。那么可根据拉克斯一密格拉蒙定理 Theorem 14.14 证得唯一可解性。  $\square$

### 间接映射法

利用狄利克雷到诺依曼的映射(14.249), 则混合边界值问题(14.298)变为, 寻找解  $\gamma_0^{\text{int}} \in H^{\frac{1}{2}}(\Gamma)$ , 使满足

$$\begin{aligned} \gamma_0^{\text{int}} u(x) &= g_D(x), \quad x \in \Gamma_D, \quad x \in \Gamma_D, \\ \gamma_1^{\text{int}} u(x) &= (S\gamma_0^{\text{int}} u)(x) = g_N(x), \quad x \in \Gamma_N. \end{aligned}$$

对于给定的狄利克雷数  $g_D \in H^{\frac{1}{2}}(\Gamma_D)$ , 假设存在任一延拓  $\tilde{g}_D \in H^{\frac{1}{2}}(\Gamma)$ , 则可以表示为如下变分问题: 寻找  $\tilde{u} := \gamma_0^{\text{int}} u - \tilde{g}_D \in \tilde{H}$ , 使满足

$$\langle S\tilde{u}, \nu \rangle_{\Gamma_N} = \langle g_N - S\tilde{g}_D, \nu \rangle_{\Gamma_N}, \quad \forall \nu \in \tilde{H}^{\frac{1}{2}}(\Gamma_N). \quad (14.305)$$

类似地, 由于 Steklov-Poincaré 算子  $S : H^{\frac{1}{2}}(\Gamma) \mapsto H^{-\frac{1}{2}}(\Gamma)$  有界并且  $\tilde{H}^{\frac{1}{2}}(\Gamma_N)$ -椭圆(14.253), 可以利用拉克斯一密格拉蒙定理 Theorem 14.14 证得  $\tilde{u}$  的唯一可解性。在计算得到狄利克雷数  $\gamma_0^{\text{int}} u(x)$  的值后, 我们可以通过求解狄利克雷边界值问题得到完整的诺依曼数  $\gamma_1^{\text{int}} u \in H^{-\frac{1}{2}}(\Gamma)$ 。

### 14.7.4 罗宾边界值问题

考虑如下罗宾边界值问题

$$(Lu)(x) = 0, \quad x \in \Omega, \\ \gamma_1^{\text{int}} u(x) + \kappa(x) \gamma_0^{\text{int}} u(x) = g(x), \quad x \in \Gamma.$$

利用狄利克雷到诺依曼的映射(14.249)，我们可以构建如下边界积分方程，用于求解未知的狄利克雷数  $\gamma_0^{\text{int}} u(x) \in H^{\frac{1}{2}}(\Gamma)$

$$\gamma_1^{\text{int}} u(x) = (S\gamma_0^{\text{int}} u)(x) = g(x) - \kappa(x) \gamma_0^{\text{int}} u(x), \quad x \in \Gamma.$$

因此可以构建如下变分问题：寻找解  $\gamma_0^{\text{int}} u \in H^{\frac{1}{2}}(\Gamma)$ ，使满足

$$\langle S\gamma_0^{\text{int}} u, \nu \rangle_{\Gamma} + \langle \kappa \gamma_0^{\text{int}} u, \nu \rangle_{\Gamma} = \langle g, \nu \rangle_{\Gamma}, \quad \forall \nu \in H^{\frac{1}{2}}(\Gamma). \quad (14.306)$$

已知 Steklov-Poincaré 算子  $S$  和超奇异边界积分算子  $D$  之间的关系满足不等式(14.250)

$$\langle S\nu, \nu \rangle_{\Gamma} \geq \langle D\nu, \nu \rangle_{\Gamma}, \quad \forall \nu \in H^{\frac{1}{2}}(\Gamma)$$

根据  $D$  的半椭圆，见 Corollary 14.9 式(14.241)，我们有

$$\langle D\nu, \nu \rangle_{\Gamma} \geq \bar{c}_1^D \|\nu\|_{H^{\frac{1}{2}}(\Gamma)}^2, \quad \forall \nu \in H^{\frac{1}{2}}(\Gamma),$$

此外假定  $\kappa(x) \geq \kappa_0, x \in \Gamma$ 。那么定义一个双线性形式  $a(.,.)$ ，满足性质

$$\begin{aligned} a(.,.) &:= \langle S\nu, \nu \rangle_{\Gamma} + \langle \kappa \nu, \nu \rangle_{\Gamma} \\ &\geq \langle D\nu, \nu \rangle_{\Gamma} + \kappa_0 \|\nu\|_{H^{\frac{1}{2}}(\Gamma)}^2 \\ &\geq \bar{c}_1^D \|\nu\|_{H^{\frac{1}{2}}(\Gamma)}^2 + \kappa_0 \|\nu\|_{H^{\frac{1}{2}}(\Gamma)}^2 \\ &\geq \min \{\bar{c}_1^D, \kappa_0\} \|\nu\|_{H^{\frac{1}{2}}(\Gamma)}^2, \end{aligned}$$

可见双线性形式  $a(.,.)$  是  $H^{\frac{1}{2}}(\Gamma)$ -椭圆的。

在此基础上，变分问题(14.306)的唯一可解性可由拉克斯—密格拉蒙定理 Theorem 14.14 证得。

## 14.8 近似方法

本节主要介绍含有算子方程的变分问题 (第14.3) 的常见近似求解方法。这些近似法有所不同，大体上来说都是通过有限维的协调检测空间 (conforming trial spaces)，将原本非线性方程系统作近似线性化的。

### 14.8.1 伽辽金——布博诺夫法

设一个算子  $A : X \mapsto X'$  且  $X$ -椭圆，对于所有  $\nu \in X$  满足

$$\begin{aligned} \langle A\nu, \nu \rangle &\geq c_1^A \|\nu\|_X^2, \\ \|A\nu\|_{X'} &\leq c_1^A \|\nu\|_X. \end{aligned}$$

假定对于某一给定的方程  $f \in X'$ ，我们想要求得变分问题(14.69)的解  $u \in X$

$$\langle Au, \nu \rangle = \langle f, \nu \rangle, \quad \forall \nu \in X. \quad (14.307)$$

由拉克斯—密格拉蒙定理 Theorem 14.14 可知, (14.307) 存在唯一的解  $u \in X$ , 满足

$$\|u\|_X \leq \frac{1}{c_1^A} \|f\|_{X'}.$$

考虑如下一个序列 (sequence)  $\{X_M\}$ ,  $M \in \mathbb{N}$ , 序列由一系列协调检测空间 (conforming trial spaces) 构成, 定义如下

$$X_M := \text{span } \{\varphi_k\}_{k=1}^M \subset X.$$

那么可以根据(14.307)构建伽辽金——布博诺夫变分问题 (Galerkin-Bubnov variational problem), 寻找近似解  $u_M \in X_M$ , 使满足

$$\langle Au_M, \nu_M \rangle = \langle f, \nu_M \rangle, \quad \forall u_M \in X_M, \quad (14.308)$$

其中近似解  $u_M \in X_M$  定义为

$$u_M := \sum_{k=1}^M u_k \varphi_k \in X_M. \quad (14.309)$$

为了证明伽辽金——布博诺夫变分法, 是对原变分问题(14.307)的有效近似方法, 需要分别证明以下三方面

- 近似解  $u_M \in X_M$  是唯一可解的,
- 近似解  $u_M \in X_M$  的稳定性,
- 近似解  $u_M \in X_M$  的收敛性。

### 近似解的唯一可解性

由于  $X_M \subset X$ , 选取  $\nu = \nu_M \in X_M$  代入原变分问题(14.307)中, 进而减去(14.308)可得

$$\langle A(u - u_M), \nu_M \rangle = 0, \quad \forall \nu_M \in X_M, \quad (14.310)$$

这称为伽辽金正交 (Galerkin Orthogonality)。

将近似解  $u_M$  的定义式(14.309)代入伽辽金——布博诺夫变分式(14.308)得

$$\left\langle A \sum_{k=1}^M u_k \varphi_k, \nu_M \right\rangle = \langle f, \nu_M \rangle,$$

由算子  $A$  的线性定义, 上式可以进一步调整为有限维度的变分问题

$$\sum_{k=1}^M u_k \langle A \varphi_k, \varphi_\ell \rangle = \langle f, \varphi_\ell \rangle, \quad \ell = 1, \dots, M. \quad (14.311)$$

利用伽辽金——布博诺夫近似法, 有限维变分问题(14.311), 进而原变分问题(14.307), 等价于一个线性方程系统: 求解系数向量  $\underline{u} \in \mathbb{R}^M$ , 使满足

$$A_M \underline{u} = \underline{f}, \quad (14.312)$$

其中

- $A_M \in \mathbb{R}^{M \times M}$  称为刚度矩阵 (stiffness matrix), 矩阵中的元素  $A_M [\ell, k]$  定义为

$$A_M [\ell, k] := \langle A \varphi_k, \varphi_\ell \rangle, \quad k, \ell = 1, \dots, M.$$

- 向量  $\underline{f}$  由一系列方程  $f_\ell$  构成,  $f_\ell$  定义为

$$f_\ell := \langle f, \varphi_\ell \rangle, \quad \ell = 1, \dots, M.$$

已知对于某一向量  $\nu \in \mathbb{R}^M$ , 可定义方程如下

$$\nu_M = \sum_{k=1}^M \nu_k \varphi_k \in X_M,$$

那么对于任意给定的  $\underline{u}, \underline{\nu} \in \mathbb{R}^M$ , 我们有

$$\begin{aligned} \langle A_M \underline{u}, \underline{\nu} \rangle &= \sum_{k=1}^M \sum_{\ell=1}^M A_M [\ell, k] u_k \nu_\ell \\ &= \sum_{k=1}^M \sum_{\ell=1}^M \langle A \varphi_k, \varphi_\ell \rangle u_k \nu_\ell \\ &= \left\langle A \sum_{k=1}^M u_k \varphi_k, \sum_{\ell=1}^M \nu_\ell \varphi_\ell \right\rangle \\ &= \langle Au_M, \nu_M \rangle. \end{aligned}$$

由此可见, 刚度矩阵  $A_M \in \mathbb{R}^{M \times M}$  继承了线性椭圆算子  $A : X \mapsto X'$  的所有特性, 其中尤其重要的是:

- $A$  是自伴随的  $\rightarrow A_M$  是对称的
- $A$  是  $X$ -椭圆的  $\rightarrow A_M$  是正定的

$$\begin{aligned} \langle A_M \underline{\nu}, \nu \rangle &= \langle A \nu_M, \nu_M \rangle \\ &\geq c_1^A \|\nu_M\|_X^2, \quad \forall \nu \in \mathbb{R}^M \leftrightarrow \nu_M \in X_M. \end{aligned}$$

由此我们有:  $A$  的椭圆性  $\rightarrow$  变分问题(14.307)的唯一可解性  $\rightarrow$  伽辽金——布博诺夫变分问题(14.308)的唯一可解性  $\rightarrow$  线性方程系统(14.312)的唯一可解性。

### 近似解的稳定性

近似解的稳定性可由齐亚引理 (Céa's lemma) 予以证明。

**Theorem 14.42** (齐亚引理). 设  $A : X \mapsto X'$  是一个有界且  $X$ -椭圆的算子。对于伽辽金——布博诺夫变分问题(14.308)的唯一近似解  $u_M \in X_M$  而言, 满足稳定性条件

$$\|u_M\|_X \leq \frac{1}{c_1^A} \|f\|_{X'}, \tag{14.313}$$

并且误差测度项为

$$\|u - u_M\|_X \leq \frac{c_2^A}{c_1^A} \inf_{\nu_M \in X_M} \|u - \nu_M\|_X. \tag{14.314}$$

证明. 近似解  $u_M$  的唯一可解性由第439页第14.8.1节给出。对于唯一的近似解  $u_M \in X_M$ , 由算子  $A$  的  $X$ -椭圆性有

$$c_1^A \|u_M\|_X^2 \leq \langle Au_M, u_M \rangle \langle f, u_M \rangle \leq \|f\|_{X'} \|u_M\|_X,$$

进而得到稳定条件(14.313)。

由线性算子  $A$  有界且  $X$ -椭圆, 以及伽辽金正交(14.310)可得, 对于任一  $\nu_M \in X_M$  都有

$$\begin{aligned} c_1^A \|u - u_M\|^2 &\leq \langle A(u - u_M), u - u_M \rangle \\ &= \langle A(u - u_M), u - \nu_M \rangle + \langle A(u - u_M), \nu_M - u_M \rangle \\ &= \langle A(u - u_M), u - \nu_M \rangle \\ &\leq c_2^A \|u - u_M\|_X \|u - \nu_M\|_X, \end{aligned}$$

进而得到误差测度项(14.314)。  $\square$

### 近似解的收敛性

近似解的收敛性是指, 随着  $M \rightarrow \infty$ , 近似解  $u_M \rightarrow u$ , 这与检测空间  $X_M$  的近似属性有关

$$\lim_{M \rightarrow \infty} \inf_{\nu_M \in X_M} \|\nu - \nu_M\|_X = 0, \quad \forall \nu \in X. \quad (14.315)$$

在构建协调检测空间的序列  $\{X_M\}_{M \in \mathbb{N}} \subset X$  的过程中, 需要确保每个检测空间  $X_M$  均满足近似属性(14.315): 其方法之一是利用局域多项式作基方程构建序列, 第九、十分别讨论有限元、有界元情况下的基方程。此外, 随着近似解的类型不同, 我们还要考虑其他一些近似属性, 见下。

作 reference

### 14.8.2 线性形近似

在构建经济学模型时, 变分问题(14.307)右侧的  $f \in X'$  有时表示为以下线性形式  $f = Bg$ , 其中  $g \in Y$  为预先给定的方程, 有界线性算子  $B : Y \mapsto X'$  满足

$$\|Bg\|_{X'} \leq c_2^B \|g\|_Y, \quad \forall g \in Y.$$

对应的, 变分问题调整为: 寻找解  $u \in X$ , 使满足

$$\langle Au, \nu \rangle = \langle Bg, \nu \rangle, \quad \forall \nu \in X. \quad (14.316)$$

利用伽辽金——布博诺夫法(14.308)构建近似变分问题: 寻找唯一的近似解  $u_M \in X_M$ , 使满足

$$\langle Au_M, \nu_M \rangle = \langle Bg, \nu_M \rangle, \quad \forall \nu_M \in X_M. \quad (14.317)$$

随后的任务是生成类似(14.312)的线性方程系统。通过测度算子  $B : Y \mapsto X'$  或伴随算子  $B' : X \mapsto Y'$  来计算

$$f_\ell = \langle Bg, \varphi_\ell \rangle = \langle g, B'\varphi_\ell \rangle, \quad \ell = 1, \dots, M,$$

用  $g_N$  对上式中的  $g$  作近似替代

$$g_N = \sum_{i=1}^N g_i \psi_i \in Y_N = \text{span}\{\psi_i\}_{i=1}^N \subset Y, \quad i = 1, \dots, N.$$

这样我们有了一个新的扰动变分问题 (perturbed variational problem): 寻找近似解  $\tilde{u}_M \in X_M$ , 使满足

$$\langle A\tilde{u}_M, \nu_M \rangle = \langle Bg_N, \nu_M \rangle, \quad \forall \nu_M \in X_M. \quad (14.318)$$

该问题等价于如下线性方程系统

$$A_M \tilde{u}_M = B_N g, \quad (14.319)$$

其中

- 矩阵  $A_M, B_N$  中的元素分别为

$$\begin{aligned} A_M [\ell, k] &= \langle A\varphi_k, \varphi_\ell \rangle, \quad k, \ell = 1, \dots, M, \\ B_N [\ell, i] &= \langle B\psi_i, \varphi_\ell \rangle, \quad i = 1, \dots, N. \end{aligned}$$

- 向量  $\underline{g}$  中包含  $g_N$ , 是对  $g$  的近似。

借助线性方程系统(14.319), 我们可以计算矩阵  $B_N$ , 并且  $B_N$  不受给定的近似方程  $g_N$  的影响。

根据有界线性算子  $A$  的  $X$ -椭圆性, 可得  $A_M$  是正定矩阵, 进而(14.319)唯一可解, 进而扰动变分问题(14.318)唯一可解。

唯一近似解  $u_M \in X_M$  对唯一解  $u \in X$  的近似精度 (误差), 主要受到(14.319)中  $g_N$  对  $g$  的近似精度 (误差) 的影响, 误差测度见以下斯特朗第一引理 (Strang's First Lemma)。

**Theorem 14.43** (斯特朗第一引理). 设一个线性算子  $A : X \mapsto X'$  有界且  $X$ -椭圆。设  $u \in X$  是连续变分问题(14.316)的唯一解, 那么对于扰动变分问题(14.318)的唯一近似解  $\tilde{u}_M \in X_M$ , 近似的误差测度为

$$\|u - \tilde{u}_M\|_X \leq \frac{1}{c_1^A} \left\{ c_2^A \inf_{\nu_M \in X_M} \|u - \nu_M\|_X + c_2^B \|g - g_N\|_X \right\}. \quad (14.320)$$

证明. 将两个变分问题(14.317), (14.318)相减得

$$\langle A(u_M - \tilde{u}_M), \nu_M \rangle = \langle B(g - g_N), \nu_M \rangle, \quad \forall \nu_M \in X_M.$$

现在定义检测方程为  $\nu_M := u_M - \tilde{u}_M \in X_M$ , 再考虑到  $A$  和  $B$  的有界性, 以及  $A$  的  $X$ -椭圆性, 上式变为

$$\begin{aligned} c_1^A \|u_M - \tilde{u}_M\|_X^2 &\leq \langle A(u_M - \tilde{u}_M), (u_M - \tilde{u}_M) \rangle \\ &= \langle B(g - g_N), (u_M - \tilde{u}_M) \rangle \\ &\leq \|B(g - g_N)\|_{X'} \|u_M - \tilde{u}_M\|_X \\ &\leq c_2^B \|g - g_N\|_Y \|u_M - \tilde{u}_M\|_X, \\ \hookrightarrow \|u_M - \tilde{u}_M\|_X &\leq \frac{c_2^B}{c_1^A} \|g - g_N\|_Y, \end{aligned}$$

进一步应用三角不等式, 结合齐亚引理 Theorem 14.42式(14.314)有

$$\begin{aligned} \|u - \tilde{u}_M\|_X &\leq \|u - u_M\|_X + \|u_M - \tilde{u}_M\|_X \\ &\leq \frac{c_2^A}{c_1^A} \inf_{\nu_M \in X_M} \|u - \nu_M\|_X + \frac{c_2^B}{c_1^A} \|g - g_N\|_Y, \end{aligned}$$

证毕。  $\square$

### 14.8.3 算子的近似

除了对变分问题(14.307)右侧的近似之外, 在进行求积分的数值近似计算时, 有时我们也要考虑对给定算子作近似。举例来说, 将伽辽金变分问题(14.308)转化为如下扰动变分问题: 寻求解  $\tilde{u}_M \in X_M$ , 使满足

$$\langle \tilde{A}\tilde{u}_M, \nu_M \rangle = \langle f, \nu_M \rangle, \quad \forall \nu_M \in X_M, \quad (14.321)$$

其中近似算子  $\tilde{A}: X \mapsto X'$  是个有界的线性算子，满足

$$\|\tilde{A}\nu\|_{X'} \leq \tilde{c}_2^A \|\nu\|_X, \quad \forall \nu \in X. \quad (14.322)$$

扰动变分问题(14.321)与伽辽金变分问题(14.308)相减，得伽辽金正交条件

$$\langle Au_M - \tilde{A}\tilde{u}_M, \nu_M \rangle = 0, \quad \forall \nu_M \in X_M. \quad (14.323)$$

为了确保变分问题(14.321)唯一可解，我们假定近似算子  $\tilde{A}$  的稳定性。在此基础上，唯一近似解  $\tilde{u}_M \in X_M$  的误差测度项见斯特朗第二引理 (Strang's Second Lemma)。

**Theorem 14.44** (斯特朗第二引理). 假设近似算子  $\tilde{A}: X \mapsto X'$  线性，有界， $X$ -椭圆，即

$$\langle \tilde{A}\nu_M, \nu_M \rangle \geq \tilde{c}_1^A \|\nu_M\|_X^2. \quad (14.324)$$

那么扰动变分问题(14.321)存在唯一解  $\tilde{u}_M \in X_M$ ，满足误差测度项

$$\|u_M - \tilde{u}_M\|_X \leq \left[ 1 + \frac{1}{\tilde{c}_1^A} (c_2^A + \tilde{c}_2^A) \right] \frac{c_2^A}{\tilde{c}_1^A} \inf_{\nu_M \in X_M} \|u - \nu_M\|_X + \frac{1}{\tilde{c}_1^A} \left\| (A - \tilde{A}) u \right\|_{X'}. \quad (14.325)$$

证明. 第一步。由假设条件近似算子  $\tilde{A}$  的  $X$ -椭圆性可得，对应的刚度矩阵  $\tilde{A}_M$  正定，进而扰动变分问题(14.321)存在唯一解  $\tilde{u}_M \in X_M$ 。

第二步。对于伽辽金变分问题(14.308)的唯一解  $u_M \in X_M$ ，由  $\tilde{A}$  的  $X_M$ -椭圆性，以及正交条件(14.323)可得

$$\begin{aligned} \tilde{c}_1^A \|u_M - \tilde{u}_M\|_X^2 &\leq \langle \tilde{A}(u_M - \tilde{u}_M), u_M - \tilde{u}_M \rangle \\ &= \langle (\tilde{A} - A)u_M, u_M - \tilde{u}_M \rangle + \langle Au_M - \tilde{A}u_M, u_M - \tilde{u}_M \rangle \\ &= \langle (\tilde{A} - A)u_M, u_M - \tilde{u}_M \rangle \\ &\leq \left\| (\tilde{A} - A)u_M \right\|_{X'} \|u_M - \tilde{u}_M\|_X, \\ &\rightarrow \|u_M - \tilde{u}_M\|_X \leq \frac{1}{\tilde{c}_1^A} \left\| (\tilde{A} - A)u_M \right\|_{X'}. \end{aligned}$$

由于两个算子  $A, \tilde{A}: X \mapsto X'$  均有界，

$$\begin{aligned} \left\| (A - \tilde{A})u_M \right\|_{X'} &\leq \left\| (A - \tilde{A})u \right\|_{X'} + \left\| (A - \tilde{A})(u - u_M) \right\|_{X'} \\ &\leq \left\| (A - \tilde{A})u \right\|_{X'} + [c_2^A + \tilde{c}_2^A] \|u - u_M\|_X. \end{aligned}$$

第三步。在此基础上

$$\begin{aligned} \|u_M - \tilde{u}_M\|_X &\leq \|u - u_M\|_X + \|u_M - \tilde{u}_M\|_X \\ &\leq \|u - u_M\|_X + \frac{1}{\tilde{c}_1^A} \left\| (\tilde{A} - A)u_M \right\|_{X'} \\ &\leq \|u - u_M\|_X + \frac{1}{\tilde{c}_1^A} \left\| (\tilde{A} - A)u \right\|_{X'} + \frac{c_2^A + \tilde{c}_2^A}{\tilde{c}_1^A} \|u - u_M\|_X \\ &\leq \left[ 1 + \frac{1}{\tilde{c}_1^A} (c_2^A + \tilde{c}_2^A) \right] \frac{c_2^A}{\tilde{c}_1^A} \inf_{\nu_M \in X_M} \|u - \nu_M\|_X + \frac{1}{\tilde{c}_1^A} \left\| (A - \tilde{A})u \right\|_{X'}. \end{aligned}$$

□

#### 14.8.4 伽辽金——佩特洛夫法

设一个有界线性算子  $B : X \mapsto \Pi'$  满足(连续)稳定性条件

$$c_S \|\nu\| \leq \sup_{0 \neq q \in \Pi} \frac{\langle B\nu, q \rangle}{\|q\|_\Pi}, \forall \nu \in (\ker B)^\top \subset X, \quad (14.326)$$

那么根据定理14.16, 对于某一给定的  $g \in \text{Im}_X(B)$ , 算子方程  $Bu = g$  有唯一解  $u \in (\ker B)^\top$ , 满足

$$\langle Bu, q \rangle = \langle g, q \rangle_\Pi, \quad \forall q \in \Pi.$$

对于  $M \in \mathbb{R}$ , 引入两个协调检测空间序列

$$\begin{aligned} X_M &= \text{span } \{\varphi_k\}_{k=1}^M \subset (\ker B)^\top, \\ \Pi_M &= \text{span } \{\psi_k\}_{k=1}^M \subset \Pi. \end{aligned}$$

则伽辽金——佩特洛夫变分问题(Galerkin-Petrov variational problem)可以定义如下: 寻找近似解  $u_M \in X_M$ ,  $u_M$  的定义如(14.309), 使满足

$$\langle Bu_M, q_M \rangle = \langle g, q_M \rangle, \quad \forall q_M \in \Pi_M. \quad (14.327)$$

与伽辽金——布博诺夫法(14.307)相比, 在伽辽金——佩特洛夫法(14.327)中有两组检测空间。由  $\Pi_M \subset \Pi$  可得伽辽金正交条件

$$\langle B(u - u_M), q_M \rangle = 0, \quad \forall q_M \in \Pi_M. \quad (14.328)$$

那么伽辽金——佩特洛夫变分问题(14.327)等价于如下线性方程系统

$$B_M \underline{u}_M = g, \quad (14.329)$$

- 刚度矩阵  $B_M$  中的元素定义为

$$B_M [\ell, k] = \langle B\varphi_k, \psi_\ell \rangle, \quad k, \ell = 1, \dots, M,$$

- 向量  $\underline{g}$  由  $g_\ell$  组成

$$g_\ell = \langle g, \psi_\ell \rangle, \quad \ell = 1, \dots, M.$$

假定线性方程系统(14.329)满足离散的稳定条件(我们将在下文中证明这一假设, 见 Lemma 14.35)

$$\tilde{c}_S \|\nu_M\|_X \leq \sup_{0 \neq q_M \in \Pi_M} \frac{\langle B\nu_M, q_M \rangle}{\|q_M\|_\Pi}, \quad \forall \nu_M \in X_M, \quad (14.330)$$

那么可得系统存在唯一解, 见如下定理。

**Theorem 14.45** (伽辽金——佩特洛夫变分问题的唯一近似解). 设算子方程  $Bu = g$  存在唯一解  $u \in (\ker B)^\top$ , 则伽辽金——佩特洛夫变分问题(14.327)有唯一近似解  $u_M \in X_M$ 。

假定离散稳定条件(14.330)成立, 则误差测度项表示为

$$\|u - u_M\|_X \leq \left(1 + \frac{c_2^B}{\tilde{c}_B}\right) \inf_{\nu_M \in X_M} \|u - \nu_M\|_X. \quad (14.331)$$

证明. 1. 对于任一给定  $\nu \in (\ker B)^\top$ , 以下变分问题都存在唯一确定的近似解  $\nu_M = P_M \nu \in X_M$ , 满足

$$\langle B\nu_M, q_M \rangle = \langle B\nu, q_M \rangle, \quad \forall q_M \in \Pi_M.$$

2. 由离散稳定性条件(14.330)可得, 唯一近似解  $\nu_M \in X_M$  满足

$$\begin{aligned}\tilde{c}_S \|\nu_M\|_X &\leq \sup_{0 \neq q_M \in \Pi_M} \frac{\langle B\nu_M, q_M \rangle}{\|q_M\|_\Pi} \\ &= \sup_{0 \neq q_M \in \Pi_M} \frac{\langle B\nu, q_M \rangle}{\|q_M\|_\Pi} \\ &\leq c_2^B \|\nu\|_X,\end{aligned}$$

由此可见, 对于任一  $\nu \in (\ker B)^\top$ , 都有唯一的  $\nu_M = P_M \nu \in \Pi_M$  与之相对应, 满足

$$\|P_M \nu\|_X \leq \frac{c_2^B}{\tilde{c}_S^B} \|\nu\|_X.$$

3. 一方面伽辽金——佩特洛夫变分问题(14.327)有唯一解  $u_M \in X_M$ , 据此可得  $u_M = p_M u$ 。另一方面, 对于所有  $\nu_M$  都有  $\nu_M \in P_M \nu_M$ 。因此, 对于某一任意的  $\nu_M \in X_M$ , 都有

$$\begin{aligned}\|u - u_M\|_X &= \|(u - \nu_M) + (\nu_M - u_M)\|_X \\ &= \|(u - \nu_M) - P_M(u - \nu_M)\|_X \\ &\leq \|u - \nu_M\|_X + \|P_M(u - \nu_M)\|_X \\ &\leq \left(1 + \frac{c_2^B}{\tilde{c}_S}\right) \|u - \nu_M\|_X \\ &\leq \left(1 + \frac{c_2^B}{\tilde{c}_S}\right) \inf_{\nu_M \in X_M} \|u - \nu_M\|_X.\end{aligned}$$

□

随着  $M \rightarrow \infty$ , 伽辽金——佩特洛夫法(14.327)近似解  $u_M \rightarrow u$ , 收敛属性的证明过程同伽辽金——布博诺夫法, 见第441页第14.8.1节。

对伽辽金——佩特洛夫法(14.327)近似解  $u_M \in X_M$  的离散稳定性条件(14.330)的证明, 可见Fortin (1977)。

**Lemma 14.35** (离散稳定性条件). 设一个有界线性算子  $B : X \mapsto \Pi'$ , 满足连续稳定条件(14.326)。如果存在一个有界的投影算子  $R_M : \Pi \mapsto \Pi_M$ , 满足

$$\langle B\nu_M, q - R_M q \rangle = 0, \quad \forall \nu_M \in X_M,$$

$$\|R_M q\|_\Pi \leq c_R \|q\|_\Pi, \quad \forall q \in \Pi,$$

则离散稳定条件(14.330)成立, 对应  $\tilde{c}_S = \frac{c_S}{c_R}$ 。

证明. 由连续稳定条件(14.326)可得, 对于某一  $q_N \in \Pi_N \subset \Pi$  有

$$\begin{aligned}c_S \|\nu_M\|_X &\leq \sup_{0 \neq q \in \Pi} \frac{\langle B\nu_M, q \rangle}{\|q\|_\Pi} \\ &= \sup_{0 \neq q \in \Pi} \frac{\langle B\nu_M, R_M q \rangle}{\|q\|_\Pi} \\ &\leq c_R \sup_{0 \neq q \in \Pi} \frac{\langle B\nu_M, R_M q \rangle}{\|R_M q\|_\Pi} \\ &\leq c_R \sup_{0 \neq q_M \in \Pi_M} \frac{\langle B\nu_M, q_M \rangle}{\|q_M\|_\Pi}, \\ &\hookrightarrow \underbrace{\frac{c_R}{c_S}}_{=: c_S} \|\nu_M\|_X \leq \frac{\langle B\nu_M, q_M \rangle}{\|q_M\|_\Pi}\end{aligned}$$

□

### 14.8.5 混合边界值问题

#### 混合算子方程的鞍点变分问题

现在来看关于混合算子方程的鞍点变分问题(14.93a)-(14.93b): 寻找解  $(u, p) \in X \times \Pi$ , 使满足

$$\begin{aligned} \langle Au, \nu \rangle + \langle B\nu, p \rangle &= \langle f, \nu \rangle, \\ \langle Bu, q \rangle &= \langle g, q \rangle, \quad \forall (\nu, q) \in X \times \Pi. \end{aligned} \tag{14.332}$$

假定两个有界线性算子  $A : X \mapsto X'$ ,  $B : X \mapsto \Pi'$ 。假定  $A$  是  $X$ -椭圆的。假定连续稳定条件(14.326)成立, 那么 Theorem 14.19 的所有前提条件均得到满足, 混合算子方程的鞍点变分问题(14.332)存在唯一的解  $(u, p) \in X \times \Pi$ 。

#### 伽辽金变分问题

对于  $M, N \in \mathbb{N}$ , 定义两个协调检测空间

$$\begin{aligned} X_M &= \text{span } \{\varphi_k\}_{k=1}^M \subset X, \\ \Pi_N &= \text{span } \{\psi_i\}_{i=1}^N \subset \Pi. \end{aligned}$$

则鞍点变分问题(14.332)可以表示为如下伽辽金变分问题: 寻求解  $(u_M, p_N) \in X_M \times \Pi_N$ , 使满足

$$\begin{aligned} \langle Au_M, \nu_M \rangle + \langle B\nu_M, p_N \rangle &= \langle f, \nu_M \rangle, \\ \langle Bu_M, q_N \rangle &= \langle g, q_N \rangle, \quad \forall (\nu_M, q_N) \in X_M \times \Pi_N. \end{aligned} \tag{14.333}$$

#### 线性方程系统

定义矩阵  $A_M, B_N$ , 其中元素

$$\begin{aligned} A_M [\ell, k] &= \langle A\varphi_k, \varphi_\ell \rangle, \quad k, \ell = 1, \dots, M, \\ B_N [j, k] &= \langle B\varphi_k, \psi_j \rangle, \quad j = 1, \dots, N. \end{aligned}$$

定义向量  $\underline{f}, \underline{g}$ , 向量中的元素  $f, g$  分别为

$$\begin{aligned} f_\ell &= \langle f, \varphi_\ell \rangle, \\ g_j &= \langle g, \psi_j \rangle. \end{aligned}$$

则伽辽金变分问题(14.333)等价于下述线性方程系统

$$\begin{pmatrix} A_M & B_N^\top \\ B_N & 0 \end{pmatrix} \begin{pmatrix} \underline{u} \\ \underline{p} \end{pmatrix} = \begin{pmatrix} \underline{f} \\ \underline{g} \end{pmatrix}. \tag{14.334}$$

#### 舒尔补系统

先来看线性方程系统(14.334)唯一可解性的证明。首先, 唯一可解性的必要条件是  $M \geq N$ 。设系统矩阵  $K$  的维度  $\dim K = M + N$ , 根据

$$\text{range } A_M \leq M, \quad \text{range } B_N \leq \min \{M, N\}$$

可得

$$\text{range } K \leq M + \min \{M, N\}.$$

若  $M < N$ , 会出现以下情况

$$\text{range } K \leq 2M < M + N = \dim K,$$

换句话说, 线性方程系统(14.334)不可解。为了使解存在, 必要条件是要审慎定义检测空间  $X_M$  和  $\Pi_N$ , 使  $M \geq N$ , 换句话说, 检测空间  $X_M$  需要大于另一个检测空间  $\Pi_N$ 。

在满足必要条件  $M \geq N$  后, 伽辽金变分问题(14.333)唯一可解性的证明, 可由 Theorem 14.20 证得。

如果  $A_M$  是可逆的, 则可将线性方程系统(14.334)转换成舒尔补系统 (Schur complement system), 见第11.A.1节。

由于协调检测空间  $X_M \subset X$ , 由有界线性算子  $A$  的  $X$ -椭圆性可得, 刚度矩阵  $A_M$  是正定的, 即

$$\langle A_M \underline{\nu}, \underline{\nu} \rangle = \langle A \nu_M, \nu_M \rangle \geq c_1^A \|\nu_M\|_X^2 > 0, \quad \forall \underline{\nu} \in \mathbb{R}^M \Leftrightarrow \nu_M \in X_M,$$

因此  $A_M$  可逆。

线性方程系统(14.334)转换成舒尔补系统 (Schur complement system) 如下

$$B_N A_M^{-1} B_N^\top \underline{p} = B_N A_M^{-1} f - g, \quad (14.335)$$

### 唯一可解性

关于舒尔补系统(14.335)唯一可解性的证明。

假定系统满足离散稳定条件

$$\tilde{c}_S \|q_N\|_{\Pi} \leq \sup_{0 \neq \nu_M \in X_M} \frac{\langle B \nu_M, q_N \rangle}{\|\nu_M\|_X}, \quad \forall q_N \in \Pi_N, \quad (14.336)$$

上式又称 Babuška-Brezzi-Ladyshenskaya (BBL) 条件。通常来说我们无法直接从 Theorem 14.18 式 14.94 直接推得离散稳定 (BBL) 条件(14.336)的成立, 一个间接的证明方法见下文 Lemma 14.37 节。

**Lemma 14.36** (舒尔补系统的唯一可解性). 设两个有界线性算子  $A : X \mapsto X'$ ,  $B : X \mapsto \Pi'$ , 且  $A$  是  $X$ -椭圆的。对于协调检测空间  $X_M \subset X$ ,  $\Pi_N \subset \Pi$ , 假定离散稳定条件(14.336)成立, 那么在舒尔补系统(14.335)中, 定义舒尔补矩阵

$$S_N := B_N A_M^{-1} B_N^\top,$$

则  $S_N$  是正定矩阵, 满足

$$(S_N \underline{q}, \underline{q}) \geq c_1^{S_N} \|q_N\|_{\Pi}^2, \quad \forall \underline{q} \in \Pi_N \Leftrightarrow q_N \in \Pi_N. \quad (14.337)$$

证明. 对于任意给定的定值  $\underline{q} \in \mathbb{R}^N$ , 定义  $\bar{u} := A_M^{-1} B_N^\top \underline{q}$ , 即对于方程  $q_N \in \Pi_N$ ,  $\bar{u}_M \in X_M$ , 我们有

$$\langle A \bar{u}_M, \nu_M \rangle = \langle B \nu_M, q_N \rangle, \quad \forall \nu_M \in X_M.$$

由有界线性算子  $A$  的  $X$ -椭圆性, 可得

$$\begin{aligned} c_1^A \|\bar{u}_M\|_X^2 &\leq \langle A \bar{u}_M, \bar{u}_M \rangle \\ &= \langle B \bar{u}_M, q_N \rangle \\ &= (B_N \bar{u}, \underline{q}) \\ &= \left( \underbrace{B_N A_M^{-1} B_N^\top}_{=: S_N} \underline{q}, \underline{q} \right). \end{aligned}$$

另一方面, 由离散稳定性条件(14.336)可得

$$\begin{aligned}
c_S \|q_N\|_\Pi &\leq \sup_{\underline{0} \neq \underline{\nu}_M \in \mathbb{R}^M} \frac{\langle B\nu_M, q_N \rangle}{\|\nu_M\|_X} \\
&= \sup_{\underline{0} \neq \underline{\nu}_M \in \mathbb{R}^M} \frac{\langle A\bar{u}_M, \nu_M \rangle}{\|\nu_M\|_X} \\
&\leq c_2^A \|\bar{u}_M\|_X, \\
\hookrightarrow \|q_N\|_\Pi^2 &\leq \left(\frac{c_2^A}{c_S}\right)^2 \|\bar{u}_M\|_X^2 \\
&\leq \left(\frac{1}{c_1^A}\right) \left(\frac{c_2^A}{c_S}\right)^2 (B_N A_M^{-1} B_N^\top \underline{q}, \underline{q}), \\
\hookrightarrow \underbrace{\left(B_N A_M^{-1} B_N^\top \underline{q}, \underline{q}\right)}_{=: S_N} &\geq \underbrace{c_1^A \left(\frac{c_S}{c_2^A}\right)^2}_{=: c_1^{S_N}} \|q_N\|_\Pi^2.
\end{aligned}$$

□

由舒尔补系统(14.335)唯一可解, 可得线性方程系统(14.334)唯一可解。

### 稳定性测度

线性方程系统(14.334)唯一解的稳定性测度, 见下定理。

**Theorem 14.46** (唯一解的稳定性测度). 设有界线性算子  $A : X \mapsto X'$ ,  $B : X \mapsto \Pi$ ,  $A$  是  $X$ -椭圆的。对于协调检测空间  $X_M \subset X$ ,  $\Pi_N \subset \Pi$ , 假设离散稳定条件(14.336)成立。则鞍点变分问题(14.333)存在唯一近似解  $(u_M, p_N) \in X_M \times \Pi_N$ , 对应的稳定性测度如下

$$\|p_N\|_\Pi \leq \frac{1}{c_1^{S_N}} \frac{c_2^B}{c_1^A} \|f\|_{X'} + \frac{1}{c_1^{S_N}} \|g\|_{\Pi'}, \quad (14.338)$$

$$\|u_M\|_X \leq \left(1 + \frac{c_2^B}{c_1^{S_N}} \frac{c_2^B}{c_1^A}\right) \|f\|_{X'} + \frac{1}{c_1^{S_N}} \frac{c_2^B}{c_1^A} \|g\|_{\Pi'}. \quad (14.339)$$

证明. 设线性方程系统(14.334)和伽辽金变分问题(14.333)的唯一(近似)解分别为

$$(\underline{u}, \underline{p}) \in \mathbb{R}^M \times \mathbb{R}^N \leftrightarrow (u_M, p_N) \in X_M \times \Pi_N.$$

1. 由 Lemma 14.36 可得

$$\begin{aligned}
c_1^{S_N} \|p_N\|_\Pi^2 &\leq (S_N \underline{p}, \underline{p}) \\
&= (B_N A_M^{-1} B_N^\top \underline{p}, \underline{p}) \\
&= (B_N A_M^{-1} \underline{f} - \underline{g}, \underline{p}) \\
&= \langle B\bar{u}_M - g, p_N \rangle \\
&\leq [c_2^B \|\bar{u}_M\|_X + \|g\|_{\Pi'}] \|p_N\|_\Pi,
\end{aligned}$$

$$\hookrightarrow \|p_N\|_\Pi \leq \frac{1}{c_1^{S_N}} [c_2^B \|\bar{u}_M\|_X + \|g\|_{\Pi'}].$$

在上式中,  $\bar{u} = A_M^{-1} \underline{f} \in \mathbb{R}^M \leftrightarrow \bar{u}_M \in X_M$  是如下变分问题的唯一解

$$\langle A\bar{u}_M, \nu_M \rangle = \langle f, \nu_M \rangle, \quad \forall \nu_M \in X_M,$$

由  $A$  的  $X$ -椭圆性有

$$\|\bar{u}_M\|_X \leq \frac{1}{c_1^A} \|f\|_{X'},$$

则我们有(14.338)成立。

2.

$$\begin{aligned} c_1^A \|u_M\|_X^2 &\leq \langle Au_M, u_M \rangle \\ &= \langle f, u_M \rangle - \langle Bu_M, p_N \rangle \\ &\leq [\|f\|_{X'} + c_2^B \|p_N\|_\Pi] \|u_M\|_X, \\ \hookrightarrow \|u_M\|_X &\leq \frac{1}{c_1^A} [\|f\|_{X'} + c_2^B \|p_N\|_\Pi], \end{aligned}$$

将(14.338)代入上式，我们有(14.339)成立。

□

### 误差测度项

由唯一近似解  $(u_M, p_N) \in X_M \times \Pi_N$  的稳定性测度 (14.339)-(14.338) 可进一步求得误差测度项，见如下定理。

**Theorem 14.47** (误差测度项). 设 Theorem 14.46 的所有假设条件都成立。则鞍点变分问题(14.333)的唯一近似解  $(u_M, p_N) \in X_M \times \Pi_N$ , 对应的误差测度项为

$$\|u - u_M\|_X + \|p - p_N\|_\Pi \leq c \left\{ \inf_{\nu_M \in X_M} \|u - \nu_M\|_X + \inf_{q_N \in \Pi_N} \|p - q_N\|_\Pi \right\}. \quad (14.340)$$

证明. 利用协调测试空间  $X_M \subset X, \Pi_N \subset \Pi$ , 将伽辽金变分式(14.333)与连续鞍点变分式(14.332)相减, 可得伽辽金正交

$$\begin{aligned} \langle A(u - u_M), \nu_M \rangle + \langle B\nu_M, p - p_N \rangle &= 0, \\ \langle B(u - u_M), q_N \rangle &= 0, \quad \forall (\nu_M, q_N) \in X_M \times \Pi_N. \end{aligned}$$

那么, 对于任意给定的  $(\bar{u}_M, \bar{p}_N) \in X_M \times \Pi_N$ , 我们有

$$\begin{aligned} \langle A(\bar{u}_M - u_M), \nu_M \rangle + \langle B\nu_M, \bar{p}_N - p_N \rangle &= \langle A(\bar{u}_M - u), \nu_M \rangle + \langle B'(\bar{p}_N - p), \nu_M \rangle, \\ \langle B(\bar{u}_M - u_M), q_N \rangle &= \langle B(\bar{u}_M - u), q_N \rangle. \end{aligned}$$

同样地, 由 Theorem (14.46) 可得, 存在唯一解  $(\bar{u}_M - u_M, \bar{p}_N - p_N) \in X_M \times \Pi_N$ , 以及对应的稳定性测度。那么对于任意  $(\bar{u}_M, \bar{p}_N) \in X_M \times \Pi_N$ , 我们有

$$\begin{aligned} \|\bar{p}_N - p_N\|_\Pi &\leq c_1 \|A(\bar{u}_M - u_M) + B'(\bar{p}_N - p_N)\|_{X'} + c_2 \|B(\bar{u}_M - u)\|_{\Pi'}, \\ \|\bar{u}_M - u_M\|_X &\leq c_3 \|A(\bar{u}_M - u_M) + B'(\bar{p}_N - p_N)\|_{X'} + c_4 \|B(\bar{u}_M - u)\|_{\Pi'}. \end{aligned}$$

根据有界线性算子  $A, B, B'$  的映射属性, 利用三角不等式关系可得, 对于任意  $(\bar{u}_M, \bar{p}_N) \in X_M \times \Pi_N$  都有

$$\begin{aligned} \|\bar{p}_N - p_N\|_\Pi &\leq \|p - \bar{p}_N\|_\Pi + \|\bar{p}_N - p_N\|_\Pi \\ &\leq (1 + c_1 c_2^B) \|p - \bar{p}_N\|_\Pi + (c_1 c_2^A + c_2 c_2^B) \|u - \bar{u}_M\|_X. \end{aligned}$$

采用类似的方式, 我们将也可求得  $\|u - u_M\|$  的不等式关系。二者相加, 证得(14.340)。 □

### 离散稳定条件

采用如(14.35)类似的思路, 利用Fortin (1977), 可证得离散稳定条件(14.336)。

**Lemma 14.37** (离散稳定条件的证明). 设有界线性算子  $B : X \mapsto \Pi'$ , 满足连续稳定性条件

Theorem 14.18式14.94。

如果存在一个有界的投影算子  $P_M : X \mapsto X_M$ , 满足

$$\begin{aligned}\langle B(\nu - P_M \nu), q_N \rangle &= 0, \quad \forall q_N \in \Pi_N, \\ \|P_M \nu\|_X &\leq c_p \|\nu\|_X, \quad \forall \nu \in X,\end{aligned}$$

那么离散稳定条件(14.336)成立, 对应  $\tilde{c}_S = \frac{c_S}{c_P}$ 。

### 14.8.6 强制算子

对于算子方程  $Au = f, u \in X, f \in X'$  的近似解, 如果线性有界算子  $A : X \mapsto X'$  不满足  $X$ -椭圆性, 那么上述近似解法可能不适用。一个替代的弱假定是  $A$  的强制性 (coercive operator), 即存在一个紧凑算子 (compact operator)  $C : X \mapsto X'$ , 满足 Gårding 不等式 Definition 14.42式(14.109)

$$\langle (A + C)\nu, \nu \rangle \geq C_1^A \|\nu\|_X^2, \quad \forall \nu \in X.$$

设一个由有限维检测空间组成的序列。对应地, 强制算子方程可以改写为如下伽辽金变分问题: 寻找解  $u_M \in X_M$ , 使满足

$$\langle Au_M, \nu_M \rangle = \langle f, \nu_M \rangle, \quad \forall \nu \in X_M. \quad (14.341)$$

注意: (14.341)与(14.307)的方程形式一致, 但不同之处在于线性有界算子算子  $A$  的特性从椭圆性扩展到更一般的强制性上来, 从而在计算数值解的稳定性和误差测度项时有所不同。

**Theorem 14.48** (齐亚引理). 设有界线性强制算子  $A : X \mapsto X_M$ , 假定满足离散稳定性条件。离散稳定性条件的证明见 Theorem 14.49。

$$c_S \|w_M\|_X \leq \sup_{\nu_M \in X_M, \|\nu_M\|_X > 0} \frac{\langle Aw_M, \nu_M \rangle}{\|\nu_M\|_X}, \quad \forall w_M \in X_M, \quad (14.342)$$

那么伽辽金变分问题(14.341)有唯一解  $u_M \in X_M$ , 对应的稳定性测度和误差项测度分别如下

$$\|u_M\| \leq \frac{1}{c_S} \|f\|_{X'}, \quad (14.343)$$

$$\|u - u_M\|_X \leq \left(1 + \frac{c_2^A}{c_S}\right) \inf_{\nu_M \in X_M} \|u - u_M\|_X. \quad (14.344)$$

证明. 1. 考虑一个齐次线性方程系统如下

$$A_M \bar{w} = 0,$$

对应的齐次变分问题为, 寻找解  $\bar{w}_M \in X_M$ , 使满足

$$\langle A_M \bar{w}_M, \nu_M \rangle = 0, \quad \forall \nu_M \in X_M.$$

根据离散稳定性条件(14.342)可得

$$c_S \|\bar{w}_M\|_X \leq \sup_{\nu_M \in X_M, \|\nu_M\|_X > 0} \frac{\langle A \bar{w}_M, \nu_M \rangle}{\|\nu_M\|_X}, \quad \forall \bar{w}_M \in X_M,$$

由此可见

$$\bar{w}_M = 0 \leftrightarrow \bar{w}_M \in X_M,$$

因此我们有: 线性方程系统  $A_M \underline{u} = f$ , 以及伽辽金变分问题(14.341)都是唯一可解的。

2. 设这个唯一解为  $\underline{u} \in \mathbb{R}^M \leftrightarrow u_M \in X_M$ 。再一次根据根据离散稳定性条件(14.342)有

$$\begin{aligned} c_S \|u_M\|_X &\leq \sup_{\nu_M \in X_M, \|\nu_M\|_X > 0} \frac{\langle Au_M, \nu_M \rangle}{\|\nu_M\|_X} \\ &= \sup_{\nu_M \in X_M, \|\nu_M\|_X > 0} \frac{\langle f, \nu_M \rangle}{\|\nu_M\|_X}, \end{aligned}$$

证得稳定性测度(14.343)。

3. 对于任意  $w \in X$ , 定义如下伽辽金变分问题, 使得存在唯一近似解  $w_M \in X_M$  满足

$$\langle Aw_M, \nu_M \rangle = \langle Aw, \nu_M \rangle, \quad \forall \nu_M \in X_M.$$

根据上式, 定义一个投影算子  $G_M : X \mapsto X_M$ , 使  $w_M = G_M w$ , 满足

$$\begin{aligned} \|G_M w\|_X &= \|w_M\|_X \leq \frac{1}{c_S} \|f\|_{X'} \\ &= \frac{1}{c_S} \|Aw\|_{X'} \leq \frac{c_2^A}{c_S} \|w\|_X. \end{aligned}$$

因此伽辽金变分问题(14.341)的唯一解为  $u_M = G_M u$ 。

4. 由投影算子  $G_M$  的性质可得

$$G_M \nu_M = \nu_M, \forall \nu_M \in X_M,$$

那么

$$\begin{aligned} \|u - u_M\|_X &= \|u - \nu_M + G_M \nu_M - u_M\|_X \\ &\leq \|u - \nu_M\|_X + \|G_M(u - \nu_M)\|_X \\ &\leq \left(1 + \frac{c_2^A}{c_S}\right) \|u - \nu_M\|_X, \quad \forall \nu_M \in X_M, \end{aligned}$$

由此证得误差项测度(14.344)。

□

现在我们回头来看离散稳定性条件(14.342)的证明。根据前文的介绍可知, 若  $A$  是椭圆算子, 则有下式成立

$$\begin{aligned} c_1^A \|\nu_M\|_X &\leq \frac{\langle Aw_M, w_M \rangle}{\|w_M\|_X} \\ &\leq \sup_{\nu_M \in X_M, \|\nu_M\|_X > 0} \frac{\langle Aw_M, \nu_M \rangle}{\|\nu_M\|_X}, \end{aligned}$$

那么若  $A$  是强制算子, 离散稳定性条件的证明见下定理。

**Theorem 14.49** (强制算子的离散稳定性条件). 设一个有界线性算子  $A : X \mapsto X'$  是强制且单射 (injective) 的。设  $X_M \subset X$  是一个密集 (dense) 协调检测空间序列。那么, 存在一个指数  $M_0 \in \mathbb{N}$ , 使得对于  $M > M_0$  的情况, 离散稳定性条件(14.342)成立。

证明. 1. 稳定性、误差测度、收敛性。

设任意给定且固定的  $w_M \in X_M$ 。对于强制算子  $A : X \mapsto X'$ , 对应存在一个紧凑算子  $C : X \mapsto X'$ , 使得有界算子  $D = A + C : X \mapsto X'$  是  $X$ -椭圆的。对应地, 构建如下变分问题, 其中  $\bar{\nu} := D^{-1}Cw_M \in X$  是唯一解

$$\langle D\bar{\nu}, \nu \rangle = \langle Cw_M, \nu \rangle, \quad \forall \nu \in X.$$

此外, 定义伽辽金变分问题, 唯一近似解  $\nu_M \in X_M$  满足

$$\langle D\bar{\nu}_M, \nu_M \rangle = \langle Cw_M, \nu_M \rangle, \quad \forall \nu_M \in X_M.$$

两个变分问题联立, 可得伽辽金正交条件

$$\langle D(\bar{\nu} - \bar{\nu}_M), \nu_M \rangle = 0, \quad \forall \nu_M \in X.$$

根据算子  $D$  的  $X$ -椭圆特性, 对伽辽金变分问题应用齐亚引理 Theorem 14.42, 可得稳定性测度

$$\|\bar{\nu}_M\|_X \leq \frac{1}{c_1^D} \|Cw_M\|_{X'} \leq \frac{c_2^C}{c_1^D} \|w_M\|_X.$$

$$\begin{aligned} \hookrightarrow \|w_M - \bar{\nu}_M\|_X &\leq \|w_M\|_X - \|\bar{\nu}_M\|_X \\ &\leq \left(1 + \frac{c_2^C}{c_1^D}\right) \|w_M\|_X. \end{aligned}$$

进而我们有误差测度项

$$\|\bar{\nu} - \bar{\nu}_M\|_X \leq \frac{c_2^D}{c_1^D} \inf_{\nu_M \in X_M} \|\bar{\nu} - \nu_M\|_X,$$

即检测空间  $X_M$  的近似属性成立, 因此收敛性(14.315)成立。

2. 设一个检测方程  $\nu_M = w_M - \bar{\nu}_M \in X_M$ 。我们有

$$\begin{aligned} \langle Aw_M, w_M - \bar{\nu}_M \rangle &= \langle Aw_M, w_M - \bar{\nu} \rangle + \langle Aw_M, \bar{\nu} - \bar{\nu}_M \rangle \\ &= \underbrace{\langle Aw_M, w_M - D^{-1}Cw_M \rangle}_{=: \mathcal{A}} + \underbrace{\langle Aw_M, \bar{\nu} - \bar{\nu}_M \rangle}_{=: \mathcal{B}}, \end{aligned}$$

- 由于  $A : X \mapsto X'$  有界且可逆

$$\begin{aligned} \mathcal{A} &:= \langle Aw_M, w_M - D^{-1}Cw_M \rangle = \langle Aw_M, D^{-1}(D - C)w_M \rangle \\ &= \langle Aw_M, D^{-1}Aw_M \rangle \\ &\geq c_1^D \|Aw_M\|_{X'}^2 \\ &\geq c_1^D c_A \|w_M\|_X^2. \end{aligned}$$

- 由于伽辽金正交条件

$$\begin{aligned} |\mathcal{B}| &:= |\langle Aw_M, \bar{\nu} - \bar{\nu}_M \rangle| = |\langle (D - C)w_M, \bar{\nu} - \bar{\nu}_M \rangle| \\ &= |\langle Dw_M, \bar{\nu} - \bar{\nu}_M \rangle - \langle Cw_M, \bar{\nu} - \bar{\nu}_M \rangle| \\ &= \left| \underbrace{\langle w_M, D(\bar{\nu} - \bar{\nu}_M) \rangle}_{=0} - \langle w_M - C(\bar{\nu} - \bar{\nu}_M) \rangle \right| \\ &= |\langle w_M, C(\bar{\nu} - \bar{\nu}_M) \rangle| \\ &\leq \|w_M\|_X \|C(\bar{\nu} - \bar{\nu}_M)\|_{X'}. \end{aligned}$$

3. 由于算子  $C$  紧凑, 那么存在一个子序列  $\{\bar{\nu}_M\}_{M \in \mathbb{N}}$ , 满足

$$\lim_{M \rightarrow \infty} \frac{\|C(\bar{\nu} - \bar{\nu}_M)\|_{X'}}{\|\bar{\nu}\|_X} = 0,$$

因此存在一个指数  $M_0 \in \mathbb{N}$ , 使得满足

$$\langle Aw_M, w_M - \bar{v}_M \rangle \geq \frac{1}{2} c_1^D c_A \left(1 + \frac{c_2^D}{c_1^D}\right)^{-1} \|w_M\| \|w_M - \bar{v}_M\|_X, \quad \forall M \geq M_0,$$

因此我们有离散稳定性条件(14.342)成立。

□

## 14.9 有限元

第14.4节介绍了如何将边界积分问题转换为变分问题, 本节进一步介绍求得变分问题近似解的方法。首先介绍如何设定合适的有限维检测空间, 然后证明索伯列夫空间的一些近似特性。出于简化的考虑, 在这里我们只考虑用低阶多项式来构建基方程; 更一般形式的有限元近似方法, 可参考 Cheng and Cheng (2005); Brenner and Scott (2008)。

### 14.9.1 参考元

设一个有界域  $\Omega \subset \mathbb{R}^d, d = 1, 2, 3$ , 域的边界可能是多边形 (polygonal,  $d = 2$ ) 或多面体 (polyhedra,  $d = 3$ )。将某一个  $\bar{\Omega}$  分解为含有有限个网格 (mesh)  $\tau_\ell$  的序列  $\{\mathcal{T}_N\}_{N \in \mathbb{N}}$ , 分解 (decomposition) 可写作

$$\bar{\Omega} = \overline{\mathcal{T}_N} = \bigcup_{\ell=1}^N \bar{\tau}_\ell. \quad (14.345)$$

#### 基本概念

在最简单的例子中, 每一个网格对应一个区间 (interval,  $d = 1$ ), 三角形 (triangle,  $d = 2$ ), 或多面体 (tetrahedron,  $d = 3$ )。 $\mathcal{T}_N$  中全部结点 (nodes) 的集合用  $\{x_k\}_{k=1}^M$  来表示。对于  $d = 2, 3$  的情况, 边 (edge) 用  $\{k_j\}_{j=1}^K$  来表示。分解  $\tau_\ell$  和对应结点  $x_k$  的关系, 可参考图14.1。

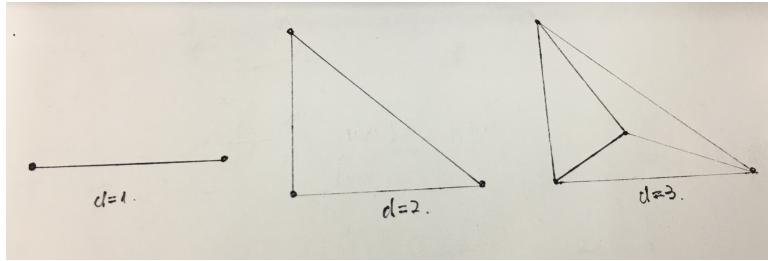


图 14.1: 有限元  $\tau_\ell$  及其结点  $x_k$

定义以下三个集合

- $I(k)$  表示全部网格  $\tau_\ell$  的集合

$$I(k) := \{\ell \in \mathbb{N} : x_k \in \bar{\tau}_\ell\}, \quad k = 1, \dots, M,$$

- $J(\ell)$  表示网格  $\bar{\tau}_\ell$  中全部结点  $x_k$  的集合

$$J(\ell) := \{k \in \mathbb{N} : x_k \in \bar{\tau}_\ell\}, \quad \ell = 1, \dots, N,$$

可见  $\dim J(\ell) = d + 1$ ,

- $K(j)$  表示网格  $\bar{\tau}_\ell$  中全部边  $k_j$  的集合

$$K(j) := \{\ell \in \mathbb{N} : k_j \in \bar{\tau}_\ell\}, \quad j = 1, \dots, K.$$

如果有限元分解(14.345)中, 所有两个相邻元之间都有共享的结点 ( $d=1,2,3$ ), 边 ( $d=2,3$ ), 三角形 ( $d=3$ ), 那么我们称这种分解为容许分解 (admissible decomposition)。以  $d = 2$  为例, 容许和不容许分解的例子见图14.2。

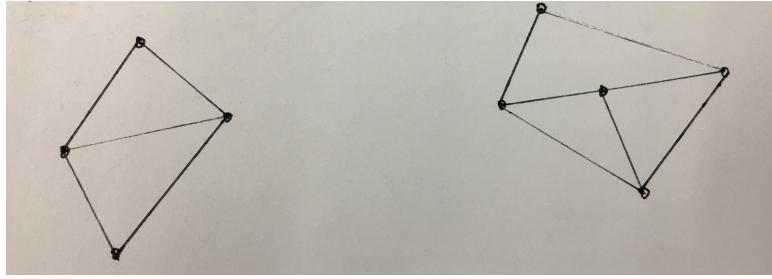


图 14.2: 容许分解, 以及不容许分解的三角分解 ( $d = 2$ )

在这里我们只考虑可计算域  $\Omega$  中容许分解的情况。对于某一有限元  $\tau_\ell$ , 作如下定义

- $\Delta_\ell$  表示体积 (volume)

$$\Delta_\ell := \int_{\tau_\ell} dx,$$

- $h_\ell$  表示局部网格尺寸 (local mesh size)

$$h_\ell := \Delta_\ell^{\frac{1}{d}},$$

- $d_\ell$  表示 (最大) 直径, 即有限元  $\tau_\ell$  中最长的一条边的长度

$$d_\ell := \sup_{x,y \in \tau_\ell} |x - y|,$$

显然对于  $d = 1$ , 我们有  $\Delta_\ell = h_\ell = d_\ell$ ,

- 对应  $d_\ell, r_\ell$  表示 (可包住) 有限元  $\tau_\ell$  的最大圆 ( $d = 2$ ) 或球体 ( $d = 3$ ) 的半径。

经容许分解(14.345)后得到的有限元  $\tau_\ell$ , 如果直径  $d_\ell$  由半径  $r_\ell$  约束而一致有界 (uniformly bounded), 即满足

$$d_\ell \leq c_F r_\ell, \quad \ell = 1, \dots, N,$$

那么我们称  $\tau_\ell$  为正则型有限元 (regular finite element), 其中常数  $c_F$  与  $T_N$ :

- 对于  $d = 2$  的情况

$$\begin{aligned} \pi r_\ell^2 &\leq \Delta_\ell = h_\ell^2 \leq d_\ell^2 \leq c_F^2 r_\ell^2, \\ \hookrightarrow \pi^{\frac{1}{2}} r_\ell &\leq h_\ell \leq d_\ell \leq c_F r_\ell, \end{aligned}$$

- 对于  $d = 3$  的情况

$$\begin{aligned} \frac{4}{3}\pi r_\ell^3 &\leq \Delta_\ell = h_\ell^3 \leq d_\ell^3 = c_F^3 r_\ell^3, \\ \hookrightarrow \left(\frac{4}{3}\pi\right)^{\frac{1}{3}} r_\ell &\leq h_\ell \leq d_\ell \leq c_F r_\ell. \end{aligned}$$

我们将最大、最小的局部网格尺寸分别定义为  $h_{\max}, h_{\min}$

$$h_{\max} := \max_{\ell=1,\dots,N} h_\ell,$$

$$h_{\min} := \min_{\ell=1,\dots,N} h_\ell,$$

对应地将全局网格尺寸 (global mesh size) 定义为

$$h = h_{\max}.$$

如果  $h_{\max}$  和  $h_{\min}$  的比值小于等于一个全常数  $c_G \geq 0$ , 即以下比值有界

$$\frac{h_{\max}}{h_{\min}} \leq c_G, \quad (14.346)$$

那么我们称对应的分解  $\mathcal{T}_N$  是全局拟一致的 (globally quasi-uniform) 的;  $c_G \geq 1$  的值与  $N \in \mathbb{N}$  无关。

对应地, 如果对于任何两个临近的有限元  $\tau_\ell$  和  $\tau_j$  都有

$$\frac{h_\ell}{h_j} \leq c_L, \ell, j = 1, \dots, N,$$

那么我们称这样的分解  $\mathcal{T}_N$  为局部拟一致 (locally quasi-uniform)。对于某一组相邻的有限元  $\bar{\tau}_\ell$  和  $\bar{\tau}_j$ , 如果  $\bar{\tau}_\ell \cap \bar{\tau}_j$  包括不少于 1 个结点、1 条边或 1 个三角形, 那么我们称二者为相邻有限元 (neighboring)。

### 一维空间的局部参数化

$d = 1$  时, 每个有限元  $\tau_\ell$  都可以用局部参数化 (local parameterization) 的形式予以描述。尤其是对于  $x \in \tau_\ell$ ,  $\ell_1, \ell_2 \in J(\ell)$ , 我们有

$$x = x_{\ell_1} + \xi(x_{\ell_2} - x_{\ell_1}) = x_{\ell_1} + \xi h_\ell, \quad \xi \in (0, 1),$$

其中我们将  $\tau := (0, 1)$  称为参考元 (reference element)。

考虑这样一个方程  $\nu(x)$ ,  $x \in \tau_\ell$ , 由上面的定义可得

$$\nu(x) = \nu(x_{\ell_1} + \xi h_\ell) =: \tilde{\nu}_\ell(\xi), \quad \forall \xi \in \tau,$$

即对于  $x \in \tau_\ell$ , 我们可以将方程  $\nu(x)$  识别为参考元中的方程  $\tilde{\nu}_\ell(\xi)$ ,  $\xi \in \tau$ , 对应范数

$$\begin{aligned} \|\nu\|_{L^2(\tau_\ell)}^2 &= \int_{\tau_\ell} (\nu(x))^2 dx = \int_{\tau} |\tilde{\nu}_\ell(\xi)|^2 h_\ell d\xi \\ &= h_\ell \|\tilde{\nu}_\ell(\xi)\|_{L^2(\tau)}^2. \end{aligned}$$

我们经常需要计算识别方程的导数, 根据链式法则(14.60)可得

#### 1. 一阶导数

$$\begin{aligned} \frac{d}{d\xi} \tilde{\nu}_\ell(\xi) &= h_\ell \frac{d}{dx} \nu(x), \\ \hookrightarrow \frac{d}{dx} \nu(x) &= h_\ell^{-1} \frac{d}{d\xi} \tilde{\nu}_\ell(\xi), \quad x \in \tau_\ell, \xi \in \tau. \end{aligned}$$

#### 2. 推广至高阶导数如 $m \in \mathbb{N}$ 的情况

$$\frac{d^m}{dx^m} \nu(x) = h_\ell^{-m} \frac{d^m}{d\xi^m} \tilde{\nu}_\ell(\xi), \quad x \in \tau_\ell, \xi \in \tau,$$

对应局部范数

$$\left\| \frac{d^m}{dx^m} \nu \right\|_{L^2(\tau_\ell)}^2 = h_\ell^{1-2m} \left\| \frac{d^m}{d\xi^m} \tilde{\nu}_\ell \right\|_{L^2(\tau)}^2, \quad m \in \mathbb{N}_0. \quad (14.347)$$

### 二维空间的局部参数化

$d = 2$  时, 有限元  $\tau_\ell$  和参考元  $\tau$  的关系, 见图14.3。

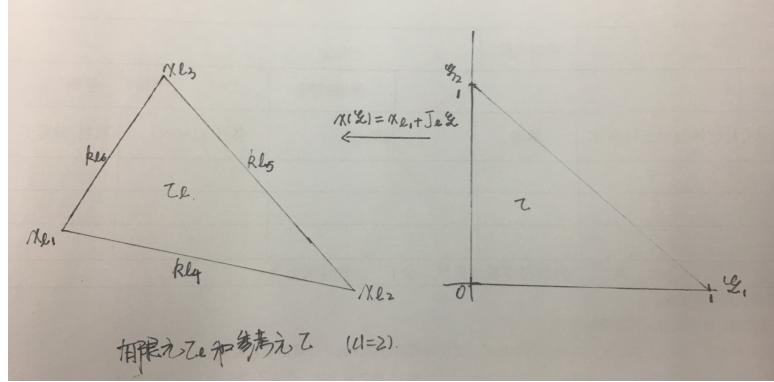


图 14.3: 有限元  $\tau_\ell$  和参考元  $\tau$  ( $d = 2$ )

参考元  $\tau$  由下面三角形给出

$$\tau = \{\xi \in \mathbb{R}^2 : 0 \leq \xi_1 \leq 1, 0 \leq \xi_2 \leq 1 - \xi_1\}. \quad (14.348)$$

对  $x \in \tau_\ell$  作局部参数化

$$x = x_{\ell_1} + \sum_{i=1}^2 \xi_i (x_{\ell_{i+1}} - x_{\ell_1}) = x_{\ell_1} + J_\ell \xi, \quad \xi \in \tau,$$

其中  $J_\ell$  是个雅各比矩阵

$$J_\ell := \begin{pmatrix} x_{\ell_2,1} - x_{\ell_1,1} & x_{\ell_3,1} - x_{\ell_1,1} \\ x_{\ell_2,2} - x_{\ell_1,2} & x_{\ell_3,2} - x_{\ell_1,2} \end{pmatrix}.$$

由此我们得有限元  $\tau_\ell$  的面积, 同时也是体积

$$\begin{aligned} \Delta_\ell &= \int_{\tau_\ell} ds_x = \int_{\tau} |\det J_\ell| d\xi \\ &= |\det J_\ell| \int_0^1 \int_0^{1-\xi_1} d\xi_2 d\xi_1 = \frac{1}{2} |\det J_\ell|, \\ \hookrightarrow |\det J_\ell| &= 2\Delta_\ell. \end{aligned} \quad (14.349)$$

$\nu(x)$ ,  $x \in \tau_\ell$  对应的识别方程可表示为

$$\nu(x) = \nu(x_{\ell_1} + J_\ell \xi) =: \tilde{\nu}_\ell(\xi), \quad \xi \in \tau.$$

继续应用链式法则

$$\begin{aligned} \nabla_\xi \tilde{\nu}_\ell(\xi) &= J_\ell^\top \nabla_x \nu(x), \\ \hookrightarrow \nabla_x \nu(x) &= J_\ell^{-\top} \nabla_\xi \tilde{\nu}_\ell(\xi). \end{aligned}$$

等价范数的测度由下引理给出。

**Lemma 14.38** (识别方程等价范的测度 ( $d = 2$ )).  $d = 2, m \in \mathbb{N}_0$  的情况下, 识别方程的等价范满足不等式关系

$$\frac{1}{c_m} (2\Delta_\ell)^{1-m} \|\nabla_\xi^m \tilde{\nu}_\ell\|_{L^2(\tau)}^2 \leq \|\nabla_x^m \nu\|_{L^2(\tau_\ell)}^2 \leq c_m (2\Delta_\ell)^{1-m} \|\nabla_\xi^m \tilde{\nu}_\ell\|_{L^2(\tau)}^2, \quad (14.350)$$

其中常数  $c_m = \left(\frac{c_F^2}{\pi}\right)^m$ 。

证明. 对于  $m \in \mathbb{N}_0$ , 分三种情况来分别证明。

1.  $m = 0$  时。

$$\begin{aligned}\|\nu\|_{L^2(\tau_\ell)}^2 &= \int_{\tau_\ell} |\nu(x)|^2 dx \\ &= \int_\tau |\tilde{\nu}_\ell(\xi)|^2 |\det J_\ell| d\xi \\ &= 2\Delta_\ell \|\tilde{\nu}_\ell(\xi)\|_{L^2(\tau)}^2,\end{aligned}$$

可证得(14.350)成立。

2.  $m = 1$  时。

$$\begin{aligned}\|\nabla_x \nu\|_{L^2(\tau_\ell)}^2 &= \int_{\tau_\ell} |\nabla_x \nu(x)|^2 dx \\ &= \int_\tau |J_\ell^{-\top} \nabla_\xi \tilde{\nu}_\ell(\xi)|^2 |\det J_\ell| d\xi \\ &= 2\Delta_\ell \int_\tau (J_\ell^{-\top} \nabla_\xi \tilde{\nu}_\ell(\xi), \nabla_\xi \tilde{\nu}_\ell(\xi)) d\xi \\ &= 2\Delta_\ell \int_\tau (J_\ell^{-1} J_\ell^{-\top} \nabla_\xi \tilde{\nu}_\ell(\xi), \nabla_\xi \tilde{\nu}_\ell(\xi)) d\xi \\ &\leq 2\Delta_\ell \lambda_{\max}(J_\ell^{-1} J_\ell^{-\top}) \int_\tau |\nabla_\xi \tilde{\nu}_\ell(\xi)| d\xi \\ &= 2\Delta_\ell \lambda_{\max}(J_\ell^{-1} J_\ell^{-\top}) \|\nabla_\xi \tilde{\nu}_\ell(\xi)\|_{L^2(\tau)}^2,\end{aligned}$$

以及类似地

$$\|\nabla_x \nu\|_{L^2(\tau_\ell)}^2 \geq 2\Delta_\ell \lambda_{\min}(J_\ell^{-1} J_\ell^{-\top}) \|\nabla_\xi \tilde{\nu}_\ell(\xi)\|_{L^2(\tau)}^2.$$

以上两个不等式都需要计算矩阵  $(J_\ell^{-1} J_\ell^{-\top})$  的特征值, 这可以从  $(J_\ell^\top J_\ell)$  开始

$$J_\ell^\top J_\ell = \begin{pmatrix} a^2 & ab \cos \alpha \\ ab \cos \alpha & b^2 \end{pmatrix},$$

其中

$$\begin{aligned}a &:= |x_{\ell_2} - x_{\ell_1}|, \\ b &:= |x_{\ell_3} - x_{\ell_1}|, \\ \alpha &:= \angle(x_{\ell_3} - x_{\ell_1}, x_{\ell_2} - x_{\ell_1}).\end{aligned}$$

$(J_\ell^{-1} J_\ell^{-\top})$  的特征值为<sup>20</sup>

$$\lambda_{1,2} = \frac{1}{2} \left[ a^2 + b^2 \pm \left( (a^2 - b^2)^2 + 4a^2 b^2 (\cos \alpha)^2 \right)^{\frac{1}{2}} \right]$$

- 对应的上限特征值  $\lambda_1$  有

$$\frac{1}{2} (a^2 + b^2) \leq \lambda_1 \leq a^2 + b^2,$$

- 两个特征值的乘积, 根据(14.349)有

$$\lambda_1 \lambda_2 = \det |J_\ell^\top J_\ell| = \det |J_\ell|^2 = 4\Delta_\ell^2.$$

---

<sup>20</sup>善用 Mathematica!

- 因此对应的下限特征值  $\lambda_2$  有

$$\lambda_2 = \frac{4\Delta_\ell^2}{\lambda_1} \geq \frac{4\Delta_\ell^2}{a^2 + b^2}.$$

- 由此我们有

$$\frac{4\Delta_\ell^2}{a^2 + b^2} \leq \lambda_{\min}(J_\ell^\top J_\ell) \leq \lambda_{\max}(J_\ell^\top J_\ell) \leq a^2 + b^2.$$

- 此外根据定义我们有

$$a^2 + b^2 \leq 2d_\ell^2 \leq 2c_F^2 r_\ell^2 \leq \frac{2c_F^2}{\pi} \Delta_\ell,$$

代入上式，整理得

$$\frac{2\pi}{c_F^2} \Delta_\ell \leq \lambda_{\min}(J_\ell^\top J_\ell) \leq \lambda_{\max}(J_\ell^\top J_\ell) \leq \frac{2c_F^2}{\pi} \Delta_\ell.$$

所以， $J_\ell^\top J_\ell$  的逆矩阵  $J_\ell^{-1} J_\ell^{-\top}$  的特征值满足

$$\begin{aligned} \frac{\pi}{c_F^2} (2\Delta_\ell)^{-1} &\leq \lambda_{\min}(J_\ell^{-1} J_\ell^{-\top}) \leq \lambda_{\max}(J_\ell^{-1} J_\ell^{-\top}) \leq \frac{c_F^2}{\pi} (2\Delta_\ell)^{-1}, \\ \hookrightarrow \left(\frac{1}{c_m}\right) (2\Delta_\ell)^{-1} (2\Delta_\ell)^{-1} &\leq \lambda_{\min}(J_\ell^{-1} J_\ell^{-\top}) \leq \lambda_{\max}(J_\ell^{-1} J_\ell^{-\top}) \leq c_m (2\Delta_\ell)^{-1}, \end{aligned}$$

因此证得(14.350)。

3.  $m > 1$  时，可通过递归重复上述步骤证得(14.350)。

□

### 三维空间的局部参数化

$d = 3$  时，参考元  $\tau$  表现为四面体 (tetrahedron) 形式

$$\tau = \{\xi \in \mathbb{R}^3 : 0 \leq \xi_1 \leq 1, 0 \leq \xi_2 \leq 1 - \xi_1, 0 \leq \xi_3 \leq 1 - \xi_1 - \xi_2\}.$$

对  $x \in \tau_\ell$  作局部参数化

$$x = x_{\ell_1} + \sum_{i=1}^3 \xi_i (x_{\ell_{i+1}} - x_{\ell_1}) = x_{\ell_1} + J_\ell \xi, \quad \xi \in \tau,$$

其中雅各比矩阵

$$J_\ell = \begin{pmatrix} x_{\ell_2,1} - x_{\ell_1,1} & x_{\ell_3,1} - x_{\ell_1,1} & x_{\ell_4,1} - x_{\ell_1,1} \\ x_{\ell_2,2} - x_{\ell_1,2} & x_{\ell_3,2} - x_{\ell_1,2} & x_{\ell_4,2} - x_{\ell_1,2} \\ x_{\ell_2,3} - x_{\ell_1,3} & x_{\ell_3,3} - x_{\ell_1,3} & x_{\ell_4,3} - x_{\ell_1,3} \end{pmatrix}.$$

有限元的体积

$$\begin{aligned} \Delta_\ell &= \int_{\tau_\ell} ds_x = \int_{\tau} |\det J_\ell| d\xi \\ &= |\det J_\ell| \int_0^1 \int_0^{1-\xi_1} \int_0^{1-\xi_1-\xi_2} d\xi_3 d\xi_2 d\xi_1 \\ &= \frac{1}{6} |\det J_\ell|, \\ \hookrightarrow |\det J_\ell| &= 6\Delta_\ell. \end{aligned} \tag{14.351}$$

用识别方程  $\tilde{\nu}_\ell(\xi), \xi \in \tau$  来识别  $\nu(x), x \in \tau_\ell$

$$\nu(x) = \nu(x_{\ell_1} + J_\ell \xi) =: \tilde{\nu}_\ell(\xi), \quad \xi \in \tau.$$

根据链式法则，可得导数

$$\begin{aligned}\nabla_\xi \tilde{\nu}_\ell(\xi) &= J_\ell^\top \nabla_x \nu(x), \\ \hookrightarrow \nabla_x \nu(x) &= J_\ell^\top \nabla_\xi \tilde{\nu}_\ell(\xi).\end{aligned}$$

等价范数的测度见下引理。

**Lemma 14.39** (识别方程等价范的测度 (d=3)).  $d = 3, m \in \mathbb{N}_0$  的情况下，识别方程的等价范满足不等式关系

$$c_1 \Delta_\ell h_\ell^{-2m} \|\nabla_\xi^m \tilde{\nu}_\ell\|_{L^2(\tau)}^2 \leq \|\nabla_x^m \nu\|_{L^2(\tau_\ell)}^2 \leq c_2 \Delta_\ell h_\ell^{-2m} \|\nabla_\xi^m \tilde{\nu}_\ell\|_{L^2(\tau)}^2 \quad (14.352)$$

其中常数  $(c_1, c_2) > 0$ ，可能与  $m$  和  $c_F$  有关。

证明. 分三种情况来证明。

1.  $m = 0$  时。

$$\begin{aligned}\|\nu(x)\|_{L^2(\tau_\ell)}^2 &= \int_{\tau_\ell} |\nu(x)|^2 dx \\ &= \int_\tau |\tilde{\nu}_\ell(\xi)|^2 |\det J_\ell| d\xi \\ &= 6\Delta_\ell \|\tilde{\nu}_\ell(\xi)\|_{L^2(\tau)}^2,\end{aligned}$$

可证得(14.352)成立。

2.  $m = 1$  时。

$$\begin{aligned}\|\nabla_x \nu\|_{L^2(\tau_\ell)}^2 &= \int_{\tau_\ell} |\nabla_x \nu(x)|^2 dx \\ &= \int_\tau |\det J_\ell| |J_\ell^{-\top} \nabla_\xi \tilde{\nu}_\ell(\xi)|^2 d\xi \\ &= |\det J_\ell| \int_\tau (J_\ell^{-\top} \nabla_\xi \tilde{\nu}_\ell(\xi), J_\ell \nabla_\xi \tilde{\nu}_\ell(\xi)) d\xi \\ &= 6\Delta_\ell \int_\tau (J_\ell^{-1} J_\ell^{-\top} \nabla_\xi \tilde{\nu}_\ell(\xi), \nabla_\xi \tilde{\nu}_\ell(\xi)) d\xi \\ \hookrightarrow \|\nabla_x \nu\|_{L^2(\tau_\ell)}^2 &\begin{cases} \geq 6\Delta_\ell \lambda_{\min}(J_\ell^{-1} J_\ell^{-\top}) \|\nabla_\xi \tilde{\nu}_\ell\|_{L^2(\tau)}^2 \\ \leq 6\Delta_\ell \lambda_{\max}(J_\ell^{-1} J_\ell^{-\top}) \|\nabla_\xi \tilde{\nu}_\ell\|_{L^2(\tau)}^2 \end{cases}\end{aligned}$$

计算  $(J_\ell^\top J_\ell)$  矩阵的特征值

$$J_\ell^\top J_\ell = \begin{pmatrix} a^2 & ab \cos \alpha & ac \cos \beta \\ ab \cos \alpha & b^2 & bc \cos \gamma \\ ac \cos \beta & bc \cos \gamma & c^2 \end{pmatrix},$$

其中

$$\begin{aligned} a &:= |x_{\ell_2} - x_{\ell_1}|, \\ b &:= |x_{\ell_3} - x_{\ell_1}|, \\ b &:= |x_{\ell_4} - x_{\ell_1}|, \\ \alpha &:= \angle(x_{\ell_2} - x_{\ell_1}, x_{\ell_3} - x_{\ell_1}), \\ \beta &:= \angle(x_{\ell_2} - x_{\ell_1}, x_{\ell_4} - x_{\ell_1}), \\ \gamma &:= \angle(x_{\ell_3} - x_{\ell_1}, x_{\ell_4} - x_{\ell_1}), \end{aligned}$$

对应的三个特征值  $\lambda_i > 0$ ,  $i = 1, 2, 3$ , 满足

$$\begin{aligned} \sum_{i=1}^3 \lambda_i &= \text{trace}(J_{\ell}^{\top} J_{\ell}) = a^2 + b^2 + c^2, \\ \prod_{i=1}^3 \lambda_i &= \det(J_{\ell}^{\top} J_{\ell}) = |\det J_{\ell}|^2 = 36\Delta_{\ell} \end{aligned}$$

- 最大特征值的上限

$$\lambda_{\max}(J_{\ell}^{\top} J_{\ell}) \leq a^2 + b^2 + c^2.$$

- 因此可得最小特征值的下限

$$\lambda_{\min}(J_{\ell}^{\top} J_{\ell}) \geq \frac{\prod_{i=1}^3 \lambda_i}{\lambda_{\max}^2(J_{\ell}^{\top} J_{\ell})} \geq \frac{36\Delta_{\ell}^2}{(a^2 + b^2 + c^2)^2}$$

- 因此我们有

$$\frac{36\Delta_{\ell}^2}{(a^2 + b^2 + c^2)^2} \leq \lambda_{\min}(J_{\ell}^{\top} J_{\ell}) \leq \lambda_{\max}(J_{\ell}^{\top} J_{\ell}) \leq a^2 + b^2 + c^2.$$

由于假定  $\tau_{\ell}$  是正则型有限元 (regular shape), 则所有边的长度对应

$$a^2 + b^2 + c^2 \leq 3d_{\ell}^2 \leq 3c_F^2 r_{\ell}^2 \leq 3 \left(\frac{4}{3}\pi\right)^{-\frac{2}{3}} c_F^2 h_{\ell}^2.$$

代入上式可得

$$\frac{4}{c_F^4} \left(\frac{4}{3}\pi\right)^{\frac{4}{3}} h_{\ell}^2 \leq \lambda_{\min}(J_{\ell}^{\top} J_{\ell}) \leq \lambda_{\max}(J_{\ell}^{\top} J_{\ell}) \leq 3 \left(\frac{4}{3}\pi\right)^{-\frac{2}{3}} c_F^2 h_{\ell}^2.$$

可证得(14.352)成立。

3.  $m > 1$  时, 可通过递归重复上述步骤证得(14.352)。

□

### 识别方程等价范的测度不等式

综上, 识别方程等价范的测度不等式,  $d = 1$  时为(14.347),  $d = 2$  时为 Lemma 14.38 式(14.350),  $d = 3$  时为 Lemma 14.39 式(14.352)。那么可以得到如下定理

**Theorem 14.50** (识别方程等价范的测度不等式). 设一个正则型的容许分解  $\mathcal{T}_N$  中的有限元  $\tau_{\ell} \subset \mathbb{R}^d$ 。如果  $\nu$  充分平滑, 对于  $m \in \mathbb{N}_0$  我们有

$$c_1 \Delta_{\ell} h_{\ell}^{-2m} \|\nabla_{\xi}^m \tilde{\nu}_{\ell}\|_{L^2(\tau)}^2 \leq \|\nabla_x^m \nu\|_{L^2(\tau_{\ell})}^2 \leq c_2 \Delta_{\ell} h_{\ell}^{-2m} \|\nabla_{\xi}^m \tilde{\nu}_{\ell}\|_{L^2(\tau)}^2,$$

其中常数  $(c_1, c_2) > 0$ , 并且可能与  $m$  和  $c_F$  有关。

### 14.9.2 形式方程

基于分解元  $\mathcal{T}_N$ (14.345)，现在来分析检测空间 (trial spaces)。检测空间由分段多项式构成，这些多项式组成不同形式的基方程 (base functions)。基方程与全局自由度相关，不过更是在局部定义的：是针对有限元  $\tau_\ell$ ，通过选取合适的形式方程 (form function) 而作局部定义。

考虑这样一个参考元  $\tau$ ，它可以是一个区间 ( $d = 1$ )，三角形 ( $d = 2$ ) 或是四面体 ( $d = 3$ )。

#### 常数型形式方程

最简单的形式方程可以是个常数

$$\psi_1^0(\xi) = 1, \quad \xi \in \tau,$$

对应地，如果有限元  $\tau_\ell$  中的某个方程  $\nu_h(x)$ ,  $x \in \tau_\ell$  是个常数，那么  $\nu$  的表现形式可写为

$$\nu_h(x) = \nu_h(x_{\ell_1} + J_\ell \xi) = \nu_\ell \psi_1^0(\xi), \quad x \in \tau_\ell, \xi \in \tau,$$

其中系数  $\nu_\ell$  反映  $\nu_h \in \tau_\ell$  在参考元  $\tau$  中对应的值。由此可得

$$\|\nu_h\|_{L^2(\tau_\ell)}^2 = \Delta_\ell \nu_\ell^2. \quad (14.353)$$

#### 线性形式方程

如果  $\nu_h(x)$ ,  $x \in \tau_\ell$  是个线性方程，那么  $\nu_h$  可由参考元  $\tau$  中结点的值  $\tilde{\nu}_k$  所唯一决定

$$\tilde{\nu}_h(\xi) = \sum_{k=1}^{d+1} \tilde{\nu}_k \psi_k^1(\xi), \quad \xi \in \tau, \quad (14.354)$$

其中  $\psi_k^1$  的值如下

- $d = 1$  时结点数  $k = 2$

$$\begin{cases} \psi_1^1 &:= 1 - \xi, \\ \psi_2^1 &:= \xi. \end{cases}$$

- $d = 2$  时结点数  $k = 3$

$$\begin{cases} \psi_1^1 &:= 1 - \xi_1 - \xi_2, \\ \psi_2^1 &:= \xi_1, \\ \psi_3^1 &:= \xi_2. \end{cases}$$

- $d = 3$  时结点数  $k = 4$

$$\begin{cases} \psi_1^1 &:= 1 - \xi_1 - \xi_2 - \xi_3, \\ \psi_2^1 &:= \xi_1, \\ \psi_3^1 &:= \xi_2, \\ \psi_4^1 &:= \xi_3. \end{cases}$$

这样一来我们有：设任意一个有限元  $\tau_\ell$ ，对应结点  $x_{\ell_k}$ ,  $\ell_k \in J(\ell)$ 。 $\tau_\ell$  中的线性方程  $\nu_h(x)$ ,  $x \in \tau_\ell$  可以写为

$$\nu_h(x) = \nu_h(x_{\ell_1} + J_\ell \xi) = \sum_{k=1}^{d+1} \nu_{\ell_k} \psi_k^1(\xi), \quad x \in \tau_\ell, \xi \in \tau. \quad (14.355)$$

与常数形式方程的范数(14.353)类似，线性形式方程  $\nu_h(x)$ ,  $x \in \tau_\ell$  的范数  $\|\nu_h\|_{L^2(\tau_\ell)}$  也可以用结点的值  $\nu_\ell(\xi)$ ,  $\xi \in \tau$  来表示，见如下引理。

**Lemma 14.40** (线性形式方程的范数). 设线性方程  $\nu_h(x), x \in \tau_\ell$  如(14.355)所示。那么我们有范数不等式关系

$$\frac{\Delta_\ell}{(d+1)(d+2)} \sum_{k=1}^{d+1} \nu_{\ell_k}^2 \leq \|\nu_h\|_{L^2(\tau_\ell)}^2 \leq \frac{\Delta_\ell}{(d+1)} \sum_{k=1}^{d+1} \nu_{\ell_k}^2 \quad (14.356)$$

证明. 线性方程  $\nu_h(x) \in L^2(\tau_\ell)$  的范数

$$\begin{aligned} \|\nu_h\|_{L^2(\tau_\ell)}^2 &= \langle \nu_h, \nu_h \rangle_{L^2(\tau_\ell)} \\ &= \sum_{i=1}^{d+1} \sum_{j=1}^{d+1} \nu_i \nu_j \int_\tau \psi_i(\xi) \psi_j(\xi) |\det J_\ell| d\xi \\ &= (G_\ell \underline{\nu}^\ell, \underline{\nu}^\ell), \end{aligned}$$

其中  $G_\ell$  是局部质量矩阵 (local mass matrix)

$$G_\ell = \frac{\Delta_\ell}{(d+1)(d+2)} \underbrace{(I_{d+1} + e_{d+1} e_{d+1}^\top)}_{=: \mathcal{A}}, \quad e_{d+1} = \underline{1} \in \mathbb{R}^{d+1}.$$

提取  $G_\ell$  矩阵  $\mathcal{A}$ , 计算特征值, Mathematica 中代码如下

```
d = 3 (*三维系统d=3。改为2或1, 对应二维、一维系统*)
ee = Table[1, d + 1, d + 1]
et = Transpose[ee]
ii = IdentityMatrix[d + 1]
new = ii + ee*et
Eigenvalues[new]
```

可得  $\lambda_1[\mathcal{A}] = 1, \lambda_2[\mathcal{A}] = \dots = \lambda_{d+1}[\mathcal{A}] = 1$ 。进而证得(14.356)。  $\square$

许多应用中需要将  $\nu_h(x)$  的斜率与  $\nu_h(x)$  自身关联起来, 见下引理。

**Lemma 14.41** (方程范数与方程斜率的范数). 设线性方程  $\nu_h$  如(14.355)所给定。则以下局部逆不等式关系成立

$$\|\nabla_x \nu_h\|_{L^2(\tau_\ell)} \leq c_I h_\ell^{-1} \|\nu_h\|_{L^2(\tau_\ell)}, \quad (14.357)$$

其中常数  $c_I > 0$ 。

证明. 分以下几步来证明。

第一步。由 Theorem 14.50 可得  $\nu_h(x), x \in \tau_\ell$  方程斜率的范, 与形式方程  $\tilde{\nu}_\ell(\xi), \xi \in \tau$  方程斜率范的关系

$$\|\nabla_x \nu_h\|_{L^2(\tau_\ell)}^2 \leq c_2 \Delta_\ell h_\ell^{-2} \|\nabla_\xi \tilde{\nu}_\ell\|_{L^2(\tau)}^2.$$

那么下面需要计算线性方程

$$\tilde{\nu}_\ell(\xi) = \sum_{k=1}^{d+1} \nu_{\ell_k} \psi_k^1(\xi)$$

的斜率以及范数。分  $d = 1, 2, 3$  三种情况。

- $d = 1$  时,

$$\nabla_\xi \tilde{\nu}_\ell = \nu_{\ell_2} - \nu_{\ell_1},$$

$$\|\nabla_\xi \tilde{\nu}\|_{L^2(\tau)}^2 = (\nu_{\ell_2} - \nu_{\ell_1})^2 \leq 2 [\nu_{\ell_2}^2 + \nu_{\ell_1}^2] \leq 4 \|\nu_h\|_{L^\infty(\tau_\ell)}^2.$$

- $d = 2$  时,

$$\nabla_{\xi} \tilde{\nu}_{\ell} = \begin{pmatrix} \nu_{\ell_2} - \nu_{\ell_1} \\ \nu_{\ell_3} - \nu_{\ell_1} \end{pmatrix},$$

$$\begin{aligned} \|\nabla_{\xi} \tilde{\nu}_{\ell}\|_{L^2(\tau)}^2 &= \frac{1}{2} \left[ (\nu_{\ell_2} - \nu_{\ell_1})^2 + (\nu_{\ell_3} - \nu_{\ell_1})^2 \right] \\ &\leq \frac{1}{2} [2\nu_{\ell_2}^2 + 2\nu_{\ell_3}^2 + 4\nu_{\ell_4}^2] \\ &\leq 4 \|\nu_h\|_{L^\infty(\tau_{\ell})}. \end{aligned}$$

- $d = 3$  时

$$\nabla_{\xi} \tilde{\nu}_{\ell} = \begin{pmatrix} \nu_{\ell_2} - \nu_{\ell_1} \\ \nu_{\ell_3} - \nu_{\ell_1} \\ \nu_{\ell_4} - \nu_{\ell_1} \end{pmatrix},$$

$$\begin{aligned} \|\nabla_{\xi} \tilde{\nu}_{\ell}\|_{L^2(\tau)}^2 &= \frac{1}{6} \left[ (\nu_{\ell_2} - \nu_{\ell_1})^2 + (\nu_{\ell_3} - \nu_{\ell_1})^2 + (\nu_{\ell_4} - \nu_{\ell_1})^2 \right] \\ &\leq \frac{1}{6} [2\nu_{\ell_2}^2 + 2\nu_{\ell_3}^2 + 2\nu_{\ell_4}^2 + 6\nu_{\ell_1}^2] \\ &\leq 4 \|\nu_h\|_{L^\infty(\tau_{\ell})}^2. \end{aligned}$$

汇总可得

$$\begin{aligned} \|\nabla_x \nu_h\|_{L^2(\tau_{\ell})}^2 &\leq c_2 \Delta_{\ell} h_{\ell}^{-2} \|\nabla_{\xi} \tilde{\nu}_{\ell}\|_{L^2(\tau)}^2 \\ &\leq 4c_2 \Delta_{\ell} h_{\ell}^2 \|\nu_h\|_{L^\infty(\tau_{\ell})}^2, \end{aligned}$$

$$\hookrightarrow \|\nabla_x \nu_h\|_{L^2(\tau_{\ell})} \leq c_I h_{\ell}^{-1} \|\nu_h\|_{L^2(\tau_{\ell})}.$$

□

## 二次形式方程

更高幂次的非线性局部形式方程, 可由相应的低幂次形式方程逐级地推定义而得。以二次形式方程为例, 定义如下:

- $d = 1$  时,

$$\begin{cases} \psi_1^2(\xi) := 1 - \xi, \\ \psi_2^2(\xi) := \xi, \\ \psi_3^2(\xi) := 4\xi(1 - \xi). \end{cases}$$

- $d = 2$  时,

$$\begin{cases} \psi_1^2(\xi) := 1 - \xi_1 - \xi_2, \\ \psi_2^2(\xi) := \xi_1, \\ \psi_3^2(\xi) := \xi_2, \\ \psi_4^2(\xi) := 4\xi_1(1 - \xi_1 - \xi_2), \\ \psi_5^2(\xi) := 4\xi_1\xi_2, \\ \psi_6^2(\xi) := 4\xi_2(1 - \xi_1 - \xi_2). \end{cases}$$

- $d = 3$  时,

$$\begin{cases} \psi_1^2(\xi) := 1 - \xi_1 - \xi_2 - \xi_3, \\ \psi_2^2(\xi) := \xi_1, \\ \psi_3^2(\xi) := \xi_2, \\ \psi_4^2(\xi) := \xi_3, \\ \psi_5^2(\xi) := 4\xi_1(1 - \xi_1 - \xi_2 - \xi_3), \\ \psi_6^2(\xi) := 4\xi_1\xi_2, \\ \psi_7^2(\xi) := 4\xi_2(1 - \xi_1 - \xi_2 - \xi_3), \\ \psi_8^2(\xi) := 4\xi_3(1 - \xi_1 - \xi_2 - \xi_3), \\ \psi_9^2(\xi) := 4\xi_3\xi_1, \\ \psi_{10}^2(\xi) := 4\xi_2\xi_3. \end{cases}$$

注意, 线性形式方程与结点  $x_k \in \tau_\ell$  的自由度有关, 而非线性形式方程则与边的中值点  $x_{k_j}^*$  有关。若方程  $\nu_h$  在  $\tau_\ell$  中是二次形式的, 那么可以写作

$$\begin{aligned} \nu_h(x) &= \nu_h(x_{\ell_1} + J_\ell \xi) \\ &= \sum_{k=1}^{\frac{1}{2}(d+1)(d+2)} \nu_{\ell_k} \psi_k^2(\xi), \quad x \in \tau_\ell, \xi \in \tau. \end{aligned} \tag{14.358}$$

类似于 Lemma 14.40, 二次形式方程的范数可写作

$$\begin{aligned} \|\nu_h\|_{L^2(\tau_\ell)}^2 &= \sum_{i,j=1}^{\frac{1}{2}(d+1)(d+2)} \nu_i \nu_j \int_\tau \psi_i^2(\xi) \psi_j^2(\xi) |\det J_\ell| d\xi \\ &= (G_\ell \underline{\nu}^\ell, \underline{\nu}^\ell), \end{aligned}$$

- $d = 1$  时, 局部质量矩阵  $G_\ell$  为

$$G_\ell = \begin{pmatrix} \frac{1}{3} & \frac{1}{6} & \frac{1}{3} \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{8}{15} \end{pmatrix},$$

对应特征值计算得

$$\begin{aligned} \lambda_1 &= \frac{\Delta_\ell}{6}, \\ \lambda_2 &= \Delta_\ell \left[ \frac{31}{60} + \frac{\sqrt{89}}{20} \right] \\ \lambda_3 &= \Delta_\ell \left[ \frac{31}{60} - \frac{\sqrt{89}}{20} \right]. \end{aligned}$$

- $d = 2, 3$  时,

- 等价范数的估计与线性形式方程情况下的 Lemma 14.40 式(14.356)一致

$$c_1 \Delta_\ell \sum_{k=1}^{\frac{1}{2}(d+1)(d+2)} \nu_{\ell_k}^2 \leq \|\nu_h\|_{L^2(\tau_\ell)}^2 \leq c_2 \Delta_\ell \sum_{k=1}^{\frac{1}{2}(d+1)(d+2)} \nu_{\ell_k}^2, \tag{14.359}$$

其中二次形式方程  $\nu_h$  的定义如(14.358)。

- 逆不等式也与线性形式下的情况 Lemma 14.41 式(14.357)相一致。

### 14.9.3 检测空间

对于二阶偏微分方程的边界值问题，为了计算近似解，需要我们构建检测空间  $\mathcal{S}_h^1(\mathcal{T}_N)$ ，空间中包括局部分段线性方程和全局连续方程。

对于容许分解(14.345)来说，在分解产生的结点  $x_k$  处， $\mathcal{S}_h^1(\mathcal{T}_N)$  便由结块方程 (nodal function) 所唯一确定， $\nu_k = \nu_h(x_k)$ 。通过这样的方式，可以将有限元  $\tau_\ell$  用局部形式方程予以表现。

全局检测空间  $\mathcal{S}_h^1(\mathcal{T}_N)$  的维数  $\dim \mathcal{S}_h^1(\mathcal{T}_N)$  也因此等于分解后的结点数。对应地， $\mathcal{S}_h^1(\mathcal{T}_N)$  的基  $\psi_k^1(x)$  如图所示，可以写为

$$\psi_k^1(x) := \begin{cases} 1 & x = x_k, \\ 0 & x = x_\ell \neq x_k, \\ \text{线性} & \text{其他情况.} \end{cases}$$

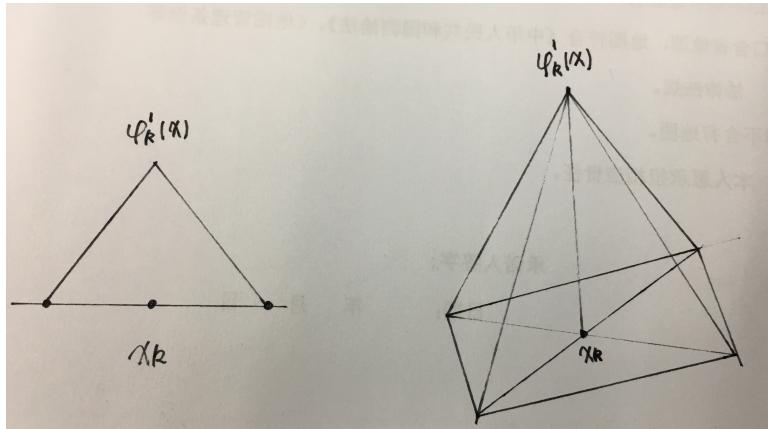


图 14.4: 检测空间  $\mathcal{S}_h^1(\mathcal{T}_N)$  的基  $\psi_k^1(x)$

分段线性方程  $\nu_h(x) \in \mathcal{S}_h^1(\mathcal{T}_N)$  可以表示为

$$\nu_h(x) = \sum_{k=1}^M \nu_k \psi_k^1(x).$$

进而检测空间中的范数见下引理

**Lemma 14.42** (光谱等价范数不等式).  $\nu_h(x) \in \mathcal{S}_h^1(\mathcal{T}_N)$  在分解中的光谱等价范数不等式为

$$\frac{1}{(d+1)(d+2)} \sum_{k=1}^M \left( \sum_{\ell \in I(k)} \Delta_\ell \right) \nu_k^2 \leq \|\nu_h\|_{L^2(\mathcal{T}_N)} \leq \frac{1}{(d+1)} \sum_{k=1}^M \left( \sum_{\ell \in I(k)} \Delta_\ell \right) \nu_k^2. \quad (14.360)$$

证明. 由 Lemma 14.40 可得

$$\begin{aligned} \|\nu_h\|_{L^2(\mathcal{T}_N)}^2 &= \sum_{\ell=1}^N \|\nu_h\|_{L^2(\mathcal{T}_\ell)}^2 \\ &\leq \sum_{\ell=1}^N \frac{\Delta_\ell}{d+1} \sum_{k=1}^{d+1} \nu_{\ell_k}^2 \\ &= \frac{1}{d+1} \sum_{k=1}^M \left( \sum_{\ell \in I(k)} \Delta_\ell \right) \nu_k^2 \end{aligned}$$

为上限界。采用类似的方法可得下限界。证得(14.360)。  $\square$

**Lemma 14.43** (逆等价范数不等式).  $\nu_h(x) \in \mathcal{S}_h^1(\mathcal{T}_N)$  在分解中的逆等价范数不等式为

$$\|\nabla_x \nu_h\|_{L^2(\mathcal{T}_N)}^2 \leq c_I \sum_{\ell=1}^N h_\ell^{-2} \|\nu_h\|_{L^2(\tau_\ell)}^2. \quad (14.361)$$

如果分解是全局拟一致的, 见(14.346), 那么我们有

$$\|\nabla_x \nu_h\|_{L^2(\mathcal{T}_N)} \leq c h^{-1} \|\nu_h\|_{(\mathcal{T}_N)}. \quad (14.362)$$

证明. 可由 Lemma 14.41 直接求得。  $\square$

为了求得检测空间  $\mathcal{S}_h^1(\mathcal{T}_N)$  的近似特性, 首先要定义相应的插值算子和投影算子, 然后计算误差测度项。

### 检测空间的插值算子

设  $\nu \in C(\mathcal{T}_N)$  是一个全局连续的方程。在含有分段线性方程的空间中定义插值如下

$$I_h \nu(x) := \sum_{k=1}^M \nu(x_k) \varphi_k(x) \in \mathcal{S}_h^1(\mathcal{T}_N). \quad (14.363)$$

**Lemma 14.44** (分段线性插值的局部误差测度项). 设给定的  $\nu|_{\tau_\ell} \in H^2(\tau_\ell)$ 。则我们有局部误差测度项

$$\|\nu - I_h \nu\|_{L^2(\tau_\ell)} \leq c h_\ell^2 |\nu|_{h^2(\tau_\ell)}. \quad (14.364)$$

证明. 1. 根据识别方程等价范的测度不等式 Theorem 14.50 可得

$$\|\nu - I_h \nu\|_{L^2(\tau_\ell)} \leq c \Delta_\ell \|\tilde{\nu}_\ell - I_\tau \tilde{\nu}_\ell\|_{L^2(\tau)},$$

其中  $I_\tau : H^2(\tau) \mapsto L^2(\tau)$  是相应参考元  $\tau$  的线性插值算子, 满足

$$\|I_\tau \tilde{\nu}_\ell\|_{L^2(\tau)} \leq \text{meas}(\tau) \|\tilde{\nu}_\ell\|_{L^\infty(\tau)},$$

$\text{meas}(\tau)$  表示  $\tau$  的测度 (measure)。由索伯列夫嵌入定理 (第339页14.2.4节) 可得

$$\|\tilde{\nu}_\ell\|_{L^\infty(\tau)} \leq c \|\tilde{\nu}_\ell\|_{L^2(\tau)},$$

可见  $I_\tau : H^2(\tau) \mapsto L^2(\tau)$  有界。

2. 对于任意且固定的方程  $\omega \in L^2(\tau)$ , 可定义一个线性泛函如下

$$f(u) := \int_\tau [(I - I_\tau) u(\xi)] \omega(\xi) d\xi.$$

如果给定方程  $u \in H^2(\tau)$ , 那么我们有

$$\begin{aligned} |f(u)| &= \left| \int_\tau [(I - I_\tau) u(\xi)] \omega(\xi) d\xi \right| \\ &\leq \|(I - I_\tau) u\|_{L^2(\tau)} \|\omega\|_{L^2(\tau)} \\ &\leq c \|u\|_{H^2(\tau)} \|\omega\|_{L^2(\tau)}, \end{aligned}$$

即线性泛函  $f$  有界。

对于任一线性方程  $q \in \mathcal{P}_1(\tau)$ , 我们有  $I_\tau q = q$ , 因此

$$f(q) = 0, \quad \forall q \in \mathcal{P}_1(\tau).$$

可见 Bramble-Hilbert 引理 Lemma 14.1 的全部前提条件均得到满足。根据该引理可得

$$|f(u)| \leq \tilde{c} \|\omega\|_{L^2(\tau)} |u|_{H^2(\tau)}.$$

3. 定义  $u(\xi)$  和  $\omega(\xi)$  的值如下

$$\begin{cases} u := \tilde{\nu}_\ell, \\ \omega := (I - I_\ell) \tilde{\nu}_\ell, \end{cases}$$

进而

$$\begin{aligned} \|\omega\|_{L^2(\tau)}^2 &= \int_\tau (\omega(\xi))^2 d\xi = \int_\tau [(I - I_\tau) \tilde{\nu}_\ell(\xi)] \omega(\xi) d\xi = \int_\tau [(I - I_\tau) u(\xi)] \omega(\xi) d\xi = |f(\tilde{\nu}_\ell)| \\ &\leq \tilde{c} \|\omega\|_{L^2(\tau)} |\tilde{\nu}_\ell|_{H^2(\tau)} = \tilde{c} \|(I - I_\tau) \tilde{\nu}_\ell\|_{L^2(\tau)} |\tilde{\nu}_\ell|_{H^2(\tau)}, \\ &\hookrightarrow \|(I - I_\tau) \tilde{\nu}_\ell\|_{L^2(\tau)} \leq \tilde{c} |\tilde{\nu}_\ell|_{H^2(\tau)}. \end{aligned}$$

4. 代入 Theorem 14.50 可得

$$\begin{aligned} \|\nu - I_h \nu\|_{L^2(\tau)} &\leq c \Delta_\ell \|\tilde{\nu}_\ell\|_{H^2(\tau)} \\ &\leq \hat{c} h_\ell^2 |\nu|_{H^2(\tau_\ell)}. \end{aligned}$$

□

由 Lemma 14.44 可得分段线性插值的全局误差项

$$\|\nu - I_h \nu\|_{L^2(\mathcal{T}_N)}^2 \leq c \sum_{\ell=1}^{\ell} h_\ell^4 |\nu|_{H^2(\tau_\ell)}^2, \quad (14.365)$$

以及

$$\|\nu - I_h \nu\|_{H^1(\mathcal{T}_N)}^2 \leq c \sum_{\ell=1}^{\ell} h_\ell^2 |\nu|_{H^2(\tau_\ell)}^2. \quad (14.366)$$

### 检测空间的投影算子

为了应用插值算子执行上述运算, 需要假定待差值的方程  $\nu \in C(\mathcal{T}_N)$  是全局连续的, 这个假设过于强硬, 有时较难满足。一个替代方案是以下若假设:



# 第十五章 误差分析

无论用哪种方法近似求解 DSGE 模型，最后一个步骤很可能都是评估数值近似解的误差，即原方程和近似方程之间的差值。由于原方程往往未知（不然也没必要做“近似”了），直接分析误差往往很难。但不同文献已经提出了一系列间接估计误差的方法，这里介绍其中之一 (Aruoba et al., 2006)，即重点关注两个评价指标，分别是

- $\chi^2$  误差检测 (Den Haan and Marcet, 1994),
- 欧拉等式误差检测 Judd (1992)

## 15.1 理论分析

有为数不多的研究从理论上讨论了近似误差的（上下）边界及其影响：

1. 误差项的上界研究方面，Santos and Vigo-Aguiar (1998) 对一个模型做价值方程迭代的近似求解，进而推算了误差的上界。在此基础上，Santos and Rust (2006) 进一步做了策略方程的迭代并推算误差上界。Santos and Peralta-Alva (2005) 进一步提出了一个一般条件，在该条件下，对模型短的模拟误差，随着近似方程越来越接近于原来的未知方程，而逐渐收敛至零。Fernández-Villaverde et al. (2006) 在 Santos and Peralta-Alva (2005) 的基础上，探讨了似然方程估计下误差项的收敛。Stachurski and Martin (2008) 计算了状态变量的遍历分布的密度，进而探讨误差的上界。
2. 误差项的下界研究方面，Judd et al. (2014) 指出研究近似误差下界的重要性，并讨论了测量方法。Kogan and Mitra (2014) 基于 Krusell and Smith Jr (1998) 的不完全市场模型，对存在异质化个体的经济系统，作数值近似求解并分析误差。

Peralta-Alva and Santos (2014) 对关于近似误差边界的理论研究做了综述。总的说来，该领域仍需展开进一步深入的讨论。

## 15.2 初步评估

在作深入的  $\xi^2$ -误差检测、欧拉等式误差检测等之前，我们可以先做一系列初步的评估，如

1. 检验计算误差是否具有一系列（理论上应当具有的）性质，如决策方程的凹性 (concavity)，单调性 (monotonicity) 等。
2. 检验关于决策方程、冲击响应方程等一系列误差式的形式和结构，和模型的基本结论，是否与理论假定相矛盾。
3. 稳健性检验，调整参数值，观察近似解的变化幅度。

通常来说，这一系列检查能为我们提供关于近似解精确度的很多信息，其信息量的丰富程度甚至超过任何“正式”的“检测”。

### 15.3 chi 方误差检测

Den Haan and Marcet (1994) 提出  $\chi^2$ -误差检测，其主要思路如下。假定有一个模型，对应均衡条件

$$f(y_t) = E_t [\phi(y_{t+1}, y_{t+2}, \dots)],$$

其中向量  $y_t$  是模型的  $n \times 1$  个状态变量，算子  $f : \mathcal{R}^n \mapsto \mathcal{R}^n$ 。算子  $\phi : \mathcal{R}^n \times \mathcal{R}^\infty \mapsto \mathcal{R}^m$ ， $f$  和  $\phi$  的形式均设为已知。

定义全体的矩 (population moment)  $u_{t+1}$  如下

$$u_{t+1} = \phi(y_{t+1}, y_{t+2}, \dots) - f(y_t), \quad (15.1)$$

那么以下条件成立

$$E_t (u_{t+1} \otimes h(x_t)) = 0, \quad (15.2)$$

其中  $\otimes$  表示张量乘， $h : \mathcal{R}^k \mapsto \mathcal{R}^q$  是任意方程， $x_t$  是在  $t$  期的任意可测度变量的向量。

截取一段样本长度为  $T$  的向量序列  $\{y_t\}_{t=1:T}$ ，如果我们对 DSGE 模型用某一种方法做近似求解，表示为  $\{y_t^j\}_{t=1:T}$ ，其中上角标  $j$  表示近似展开的阶数 (扰动法)，或基方程的数量 (投影法) 等<sup>1</sup>，那么我们可以寻找对应的  $\{u_{t+1}^j, x_t^j\}_{t=1:T}$ ，并根据(15.2)，将样本距  $B_T^j$  表示如下

$$B_T^j = \frac{1}{T} \sum_{t=1}^T u_{t+1}^j \otimes h(x_t^j). \quad (15.3)$$

如果这种近似方法是有效的，那么我们会得到样本矩的收敛

$$\lim_{T \rightarrow \infty} B_T^j \rightarrow 0.$$

定义一个新的统计量  $\mathcal{B}$

$$\mathcal{B} (B_T^j)^\top (A_T^j)^{-1} (B_T^j), \quad A_T^j := \sum_{t=-\infty}^{\infty} E_t [(u_{t+1} \otimes h(x_t)) (u_{t+1} \otimes h(x_t))^\top], \quad (15.4)$$

在零假设即全体矩假设(15.2)成立的前提下， $\mathcal{B}$  收敛到一个自由度为  $qm$  的  $\chi^2$  分布。这成为一个检验近似方法是否精确的标准： $\mathcal{B}$  的值越是显著高于 0，样本误差越高，近似方法的精度越低。

需要注意的是，假定其他条件不变，随着  $T$  逐渐变大， $\mathcal{B}$  越来越容易否定零假设。为了解决这个问题，Den Haan and Marcet (1990) 建议重复若干次上述模拟过程，在得到的一组分布中，取上限和下限的各 5% 处所对应的值：若近似方法是有效的，那么这两个值应当比平均值大或小 5%。

### 15.4 欧拉等式误差检测

Judd (1992) 提出一种标准化欧拉等式误差检测法，用于反映近似方法的精确度。其核心思路是检验处于 DSGE 模型核心的欧拉等式的近似决策式，在何种程度上接近于原未知方程。

以第12.6节的随机 NCGT 模型为例，欧拉等式表示为

$$u_c'(c_t, \ell_t) = \beta E_t \left\{ u_c'(c_{t+1}, \ell_{t+1} R_{t+1}) \right\}, \quad (15.5)$$

<sup>1</sup> 如  $c^j(k_t, z_t)$  是对原消费决策式  $c(k_t, z_t)$  的  $j$  阶扰动或者  $j$  个基的近似方程， $k_t, z_t$  是决策变量。

消费的边际效用  $u'_c(c_t, \ell_t)$  满足

$$u'_c(c_t, \ell_t) = \frac{[c_t^\tau (1 - \ell_t)^{1-\tau}]^{-\eta}}{c_t}, \quad (15.6)$$

资本的总回报率  $R_t$  满足

$$R_{t+1} = 1 + \alpha \exp[z_{t+1}] k_{t+1}^{\alpha-1} \ell_{t+1}^{1-\alpha} - \delta, \quad (15.7)$$

结合(15.6)-(15.7), 将(15.5)改写为

$$1 - \frac{1}{c_t} \left\{ u'_c \left( \beta E_t \left\{ u'_c(c_{t+1}, \ell_{t+1}) R_{t+1} \right\}, \ell_t \right)^{-1} \right\} = 0. \quad (15.8)$$

设消费, 劳动力供应的决策方程, 以及资本积累的运动法则分别为

$$\begin{aligned} c_t &= c(k_t, z_t), \\ \ell_t &= \ell(k_t, z_t), \\ k_{t+1} &= k(k_t, z_t), \end{aligned}$$

(15.8)进一步改写为

$$\begin{aligned} 1 - \frac{1}{c(k_t, z_t)} \left\{ u'_c \left( \beta E_t \left\{ u'_c(c(k_t, z_t), z_{t+1}), \ell(k(k_t, z_t), z_{t+1}) \underbrace{R_{t+1}(k_t, z_t, z_{t+1})}_{\text{ }} \right\}, \ell(k_t, z_t) \right)^{-1} \right\} &= 0, \\ R_{t+1}(k_t, z_t, z_{t+1}) &= 1 + \alpha \exp[z_{t+1}] k(k_t, z_t)^{\alpha-1} \ell(k(k_t, z_t), z_{t+1})^{1-\alpha} - \delta, \quad \forall k_t, z_t. \end{aligned} \quad (15.9)$$

在原决策方程无法直接求得的情况下, 假定采取某种近似求解方法, 分别对应

$$\begin{aligned} c^j(k_t, z_t) &\sim c(k_t, z_t), \\ \ell^j(k_t, z_t) &\sim \ell(k_t, z_t), \\ k^j(k_t, z_t) &\sim k(k_t, z_t), \end{aligned}$$

若该近似方法是有效的, 这意味着这组替代值返回(15.9), 等号仍然成立。基于这种思路, 可以定义欧拉等式误差  $EEE(k_t, z_t)$  为

$$\begin{aligned} EEE(k_t, z_t) &:= 1 - \\ &\frac{1}{c^j(k_t, z_t)} \left\{ u'_c \left( \beta E_t \left\{ u'_c(c^j(k_t, z_t), z_{t+1}), \ell^j(k^j(k_t, z_t), z_{t+1}) \underbrace{R_{t+1}^j(k_t, z_t, z_{t+1})}_{\text{ }} \right\}, \ell^j(k_t, z_t) \right)^{-1} \right\}, \end{aligned} \quad (15.10)$$

其中

$$R_{t+1}^j(k_t, z_t, z_{t+1}) = 1 + \alpha \exp[z_{t+1}] k^j(k_t, z_t)^{\alpha-1} \ell(k^j(k_t, z_t), z_{t+1})^{1-\alpha} - \delta.$$

有三点值得注意。

1. 欧拉等式误差取决于状态变量  $k_t$  和  $z_t$  的值。

若是采用扰动法做近似, 越是接近展开点, 误差越小; 反之误差越大。与之相对, 若是采用投影法做近似, 误差是沿着全部值域  $\Omega$  均匀分布的, 在这种情况下, 应当关注  $EEE$  的均值, 或是  $\max EEE \in \Omega$ 。

2.  $EEE(k_t, z_t)$  的单位是消费量，有实际经济学意义的：基于近似的策略方程，近似精度的差异导致实际行为与“最优”决策之间的差距有大有小 (Judd and Guu, 1993)。

例如  $EEE(k_t, z_t) = 0.01$  意味着行为人每 100 元的经济活动会有 1 元是错误的；若  $EEE(k_t, z_t) = 1E^{-6}$ ，那么每 100 万元的经济活动才会产生 1 元的错误决策。

3.  $EEE$  的重要性还在于，通过模型可以看出，在满足一些特定条件的前提下

- (a) 决策方程的近似误差等比于  $EEE$ ，
- (b) 福利水平变化的误差等比于  $EEE$ ，
- (c)  $EEE$  中的常数项与模型设定有关 (Santos, 2000)。

在上例中我们用线性代数转换来计算  $EEE$ ，然而需要指出的是，在一些 DSGE 模型中，我们很难利用显性转换，求得（以消费量或其他自然经济变量为单位的） $EEE$ 。另一方面， $EEE$  的均值可以使简单的算是平均，或是用某些状态变量的遍历分布 (ergodic distribution) 的估计，后者的问题在于，有时可能无法获得这个遍历分布——它来自模型的解，而研究过程中这个解恰恰是我们所不知道的。对此 Aruoba et al. (2006) 提出了一些可能的解决办法。

图15.1绘出第12.6节随机 NCGT 模型的投影法欧拉等式误差  $\log_{10} \|EEE(k_t, z_t)\|$ ，横轴单位  $k_t$ ，纵轴单位  $\log_{10} \|\cdot\|$ ，5 条不同颜色的曲线分别对应不同的生产率水平  $z_t$ 。这再一次印证了前文的观点，即切比雪夫配点投影方法近似求解 DSGE 模型，可以达到很高的近似精度。

图 15.1: 欧拉等式误差检测

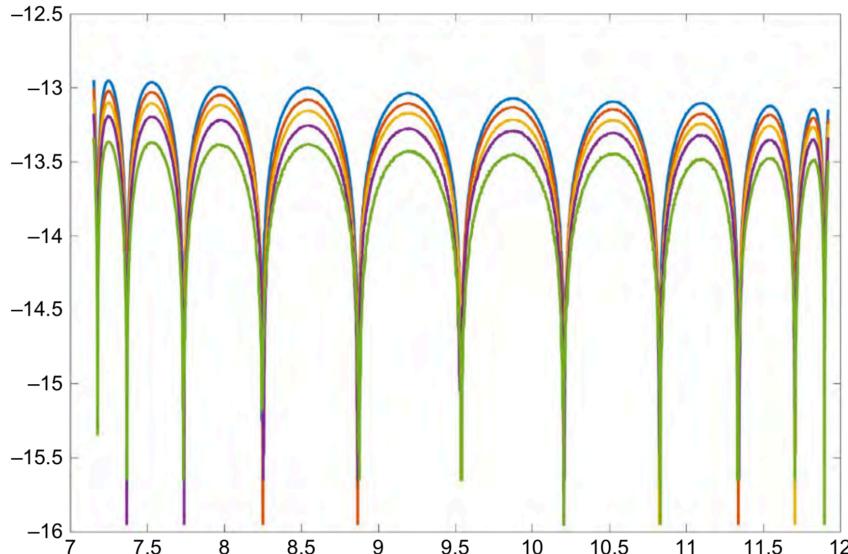
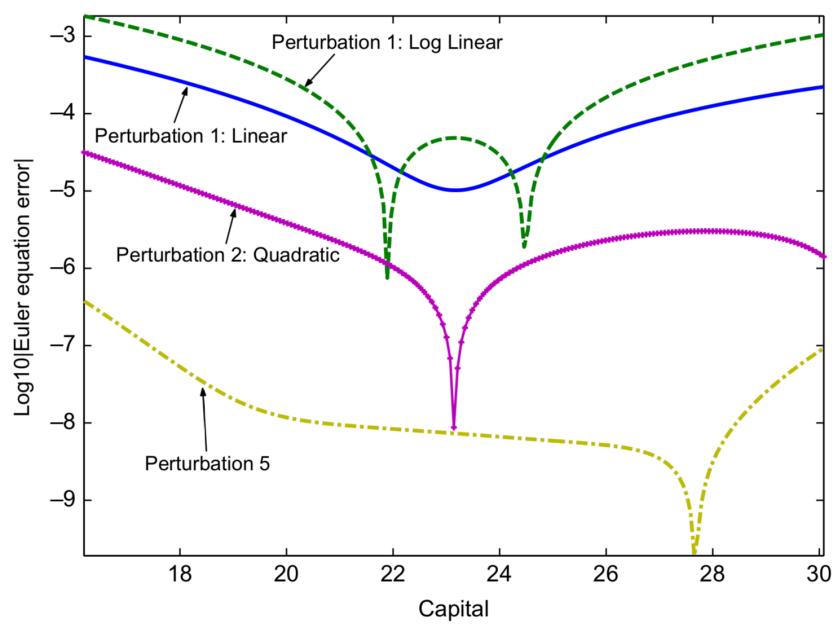


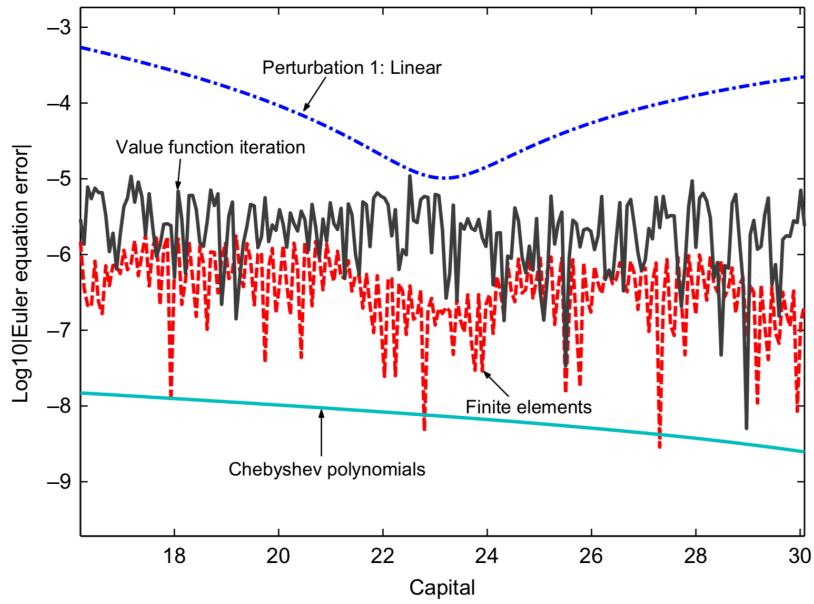
图15.2列出了几种不同近似方法下  $\log_{10} EEE(k_t, z_t)$  的比较，来自 (Aruoba et al., 2006) 的随机 NCGT 模型（与本文第12.6节的模型基本类似，除了几处餐厨校准和生产率  $z_t$  的处理）。两张图均设  $z_t = 0$ ，以及资本沿着稳态值  $k = 23.14$  的 70% 和 130% 之间截取  $\log_{10} EEE$  展开分析。

1. 图15.3(a)分别画出 1 阶扰动（对数线性）、1 阶扰动（线性）、2 阶和 5 阶扰动的  $\log_{10} EEE$ 。  
不难看出
  - (a) 在稳态点  $k = 23.14$  附近，误差值明显低，越远离稳态点的  $k_t$  取值，对应的误差越大，
  - (b) 从 1 阶到 2 阶近似，精度有明显提升，

图 15.2: 欧拉等式误差检测的比较



(a) 扰动法



(b) 投影法

- (c) 5 阶近似的精度很高，并且几乎是全局最优的——即便  $k_t$  取值远离稳态值的 30% 以上，精度仍保持在一个较高的水平上。
2. 图15.3(b)分别画出价值方程迭代 ( $25,000 \times 40 = 1,000,000$  个网格点，25,000 和 40 分别表示  $k_t$  和  $z_t$ )、有限元 (71 个有限元)、切比雪夫多项式 (11 个多项式，如第12.6节模型设定) 的  $\log_{10} EEE$ ，此外绘出 1 阶扰动 (线性) 的误差作为比较。不难看出
    - (a) 比起局部的扰动法来，全局性的价值方程迭代法和投影法的精度更高。
    - (b) 价值方程迭代、有限元和切比雪夫多项式对应的  $EEE$  较难直观理解，因为他们分别依赖于求解过程中设定的网格点数量、有限元划分、切比雪夫多项式设定。但三者之中，切比雪夫多项式方法提供了最精确的全局解，并且程序运算的时间也是三者中最低的：这与模型设定有关，随机 NCGT 模型的决策法则总的来说表现良好，非常适合用光谱基方程来做近似。

欧拉方程误差的检测已经成为 DSGE 模型近似求解过程中判断解法精度的标准检测手段之一，这是由于它所能提供的一系列非常有价值的数据。但需要指出的是，它也存在先天不足，主要表现在很难对近似误差随时间的累积过程作出详细说明。例如，Santos and Peralta-Alva (2005) 介绍了如何理解欧拉方程误差对模型近似矩的影响。因此，欧拉方程误差检测可以作为  $\chi^2$ -检测的有益补充，但不能完全替代后者。

## 15.5 误差值的改进

在求得近似解的误差后，可以决定是否进一步提高近似解的精度。这涉及到取舍权衡问题：理论上来讲，精度越高越好，然而实际操作过程中，追求精度的提升往往意味着更长的编程时间，更久的程序运行时间。以图15.3(b)为例，在用有限元法作近似求解的过程中，我们可以使用一些现代的科学计算库比如 GNU 多重精度运算库 (GNU Multiple Precision Arithmetic Library, GMP)，来在硬件 (CPU、内存等) 和时间允许的范围内，追求给定精度的计算。在研究过程中做灵活掌握：先从最简单的近似方法入手，如果计算精度未达到要求，就进一步调整近似算法。

如果研究目标对精度没有特别高的要求，那么可以做如下选择

1. 若采用扰动法，可以进一步提高扰动的阶数。
2. 若采用投影法，可增加基方程中元素的数量。
3. 此外也可以做变量变换或改变求解方法，以使得代求解系统进一步线性化。

# **第三部分**

## **模型估计**



# 第十六章 将 DSGE 模型和实际数据联系起来

第一部分讨论了在参数已经被赋值的前提下，如何近似求解 DSGE 模型。这一部分讨论如何基于实际观测到的数据，为 DSGE 模型的参数复制，并判断 DSGE 模型在何种程度上再现了经济现实。具体说来，大概有以下几个问题：

- 如何基于观测到的（宏观）经济数据，估计 DSGE 模型中的参数值？
- 含有估计参数的 DSGE 模型（以下简称估计 DSGE 模型），可以在多大程度上反映观测数据的核心特征？
- 估计 DSGE 模型对于解释以下经济现象有何启发？如经济周期波动的成因、外甥冲击的影响、宏观经济政策变化的效果、时间序列数据的未来走势等。
- 在存在不确定性的情况下，如何更有效度量参数，和制定更恰当的 DSGE 模型的量化政策建议？

为了回答这些问题，我们首先建立一个典型的 DSGE 模型（第16.1节），随后讨论该模型的性质（第16.2节）。DSGE 模型生成的模拟时间序列数据（以下简称模拟数据），包括全体的矩，协方差，光谱，冲击响应方程等，可以被视为是实际观测到的数据（以下简称观测数据）中特定样本同样具有的特征，二者的比较见第16.3节。宏观经济学时间序列中含有的一些趋势中，有些能为 DSGE 模型所反映出来，有些则不能，相关讨论见第16.4节。

在掌握宏观计量经济学基础知识（PHD 最初几年学习）的基础上，关于 DSGE 模型估计的基础教科书可参考如Canova (2011); DeJong and Dave (2007) 等。近年来 DSGE 估计方面的研究进展很快，下文将介绍一系列“标准方法”，如

- DSGE 模型的识别
- 对识别稳健 (identification-robust) 的频率学派推断
- 贝叶斯分析中的马尔科夫链——蒙特卡洛方法 (Marcov Chain Monte Carlo, MCMC)(Herbst and Schorfheide, 2015)
- 此外还讨论了经济计量推断中误设定问题的后果，和频率学派的若干方法等。

## 16.1 典型 DSGE 模型

整个第三部分基于一个对数线性化形式（可参考 Remark 11.1）的新凯恩斯主义 DSGE 模型（New Keynesian-DSGE）展开讨论。这个模型来自Del Negro and Schorfheide (2008)，做了一些参数简化（Smets and Wouters, 2003; Christiano et al., 2005），以下简称典型 DSGE 模型。

模型分为 5 个部门，分别是家庭部门，中间产品生产部门，最终产品生产部门，负责制定货币政策的中央银行，和负责制定财务政策的公共部门。4 个外生冲击分别是技术进步冲击  $z_t$ ，劳动——休闲偏好冲击  $\phi_t$ ，加成定价格 (price markup) 冲击  $\lambda_t$ ，货币政策冲击  $\epsilon_{R,t}$ 。

最终产品生产率水平  $Z_t$  设为外生，是一个带漂移 (drift) 的随机游走 (random walk) 过程

$$\begin{aligned}\log Z_t &= \log Z_{t-1} + \log \gamma + z_t, \\ z_t &= \rho_z z_{t-1} + \sigma_z \epsilon_{z,t},\end{aligned}\tag{16.1}$$

生产率过程  $Z_t$  使得产出  $X_t$  和实际工资  $w_t$  呈现随机趋势过程 (stochastic trend process)，因此需要去趋势 (第11.3.2节)，作如下定义：

$$x_t \equiv \frac{X_t}{Z_t}, \quad w_t \equiv \frac{W_t}{Z_t}.\tag{16.2}$$

定义如下稳定状态 (第3.A节)

$$\bar{x}_t = x^*, \quad \bar{w} = \bar{lsh} = \frac{1}{1+\lambda}, \quad \bar{\pi} = \pi^*, \quad \bar{R} = \pi^* \frac{\gamma}{\beta},\tag{16.3}$$

其中

- 系数  $\lambda$  表示中间产品垄断生产者的稳态加成率，
- 系数  $\beta$  表示家庭的时间贴现，
- 系数  $\gamma$  是技术进步的稳态增长率，
- 稳定状态下，市场工资率等于全部收入中劳动者的分成  $\bar{w} = \bar{lsh}$ ，
- 自由参数  $\pi^*$  可理解为中央银行设定的目标通胀率，
- 自由参数  $x^*$  原则上可以由家庭部门效用函数最大化决策中，对劳动的负效率和休闲的效用之间的权衡所计算得到，
- 均衡条件下产品的使用  $\bar{x} = \bar{c}$ 。

### 16.1.1 对数线性化均衡条件

定义一组变量，用上角标表示其距离稳态值的对数偏离

$$\begin{aligned}\hat{x}_t &= \log \left( \frac{x_t}{\bar{x}} \right), \\ \hat{w}_t &= \log \left( \frac{w_t}{\bar{w}_t} \right), \\ \hat{\pi}_t &= \log \left( \frac{\pi_t}{\bar{\pi}} \right), \\ \hat{R}_t &= \log \left( \frac{R_t}{\bar{R}} \right),\end{aligned}\tag{16.4}$$

由此可得典型 DSGE 模型的一系列均衡条件。

家庭部门跨期消费的欧拉等式

$$\hat{x}_t = E_t \left[ \hat{x}_{t+1} + \hat{\pi}_t + z_{t+1} - \hat{R}_t \right],\tag{16.5}$$

上式中出现对未来期技术进步冲击的期望  $z_{t+1}$  是因为欧拉等式反映产出 (消费) 距其稳态的对数线性化偏离，而产出的跨期变化受到随机技术趋势  $Z_t$  的影响。

假定不存在工资粘性 (wage rigidities, 第3.2.2节, 更详细内容可参考Erceg et al. (2000))。根据欧拉等式(16.5)可得劳动力供应的决定式

$$\hat{w}_t = (1 + \nu) \hat{x}_t + \phi_t, \quad (16.6)$$

其中  $\hat{x}_t \propto \hat{L}_t$  见(16.9)。 $1/(1+\nu)$  是劳动力供应的 Frisch 弹性系数 (第3.B节)。外生偏好冲击  $\phi_t$  影响劳动力供应, 满足如下 AR(1) 过程

$$\phi_t = \rho_\phi \phi_{t-1} + \sigma_\phi \epsilon_{\phi,t}. \quad (16.7)$$

中间产品生产者  $j$ , 从家庭部门雇佣劳动力生产差异化中间产品  $j$ , 对应线性生产技术

$$X_t(j) = Z_t L_t(j), \quad (16.8)$$

作去趋势处理, 并沿稳态值做对数线性化, 得生产函数

$$\hat{x}_t(j) = \hat{L}_t(j). \quad (16.9)$$

中间产品种类的差异化导致  $j$  的垄断竞争地位。垄断定价策略使得经济系统中存在名义价格粘性 (nominal price rigidity), 可用 Calvo 定价机制来反映 (Calvo, 1983)。每一期  $t$ ,  $j$  有  $\xi_p$  概率不能调整价格,  $1 - \xi_p$  概率可以调整价格。在可以调整价格的情况下,  $j$  依据前向型定价策略, 追求他在未来一段时间  $t+s$  之内的期望利润之和的最大化, 直到  $t+s+1$  期他能再次调整价格为止。

最终产品生产者购买各种中间产品, 将它们组合起来从事最终产品的生产, 生产技术表现为 CES 形式, 市场形式表现为完全竞争。

中间品部门和最终产品部门的最优决策组合起来, 构成新凯恩斯主义菲利普斯曲线 (New Keynesian Philips Curve, NKPC, 参考笔记第30页)

$$\hat{\pi}_t = \beta E_t \hat{\pi}_{t+1} + \kappa_p (\hat{w}_t + \lambda_t), \quad \kappa_p \equiv \frac{1 - \zeta_p}{\zeta_p} (1 - \zeta_p \beta), \quad (16.10)$$

其中对价格加成的外部冲击  $\lambda_t$  设满足 AR(1) 过程

$$\lambda_t = \rho_\lambda \lambda_{t-1} + \sigma_\lambda \epsilon_{\lambda,t}. \quad (16.11)$$

经济体中总量约束条件的计算, 具体来说, 首先将总产出和产出的使用联系起来

$$\hat{x}_t = \hat{c}_t, \quad (16.12)$$

其次将劳动力的总需求和总产出联系起来。基于这样的总资源约束条件, 可得

$$\widehat{lsh}_t = \hat{w}_t. \quad (16.13)$$

中央银行根据市场上反馈的信息, 设定当期名义利率

$$\hat{R}_t = \psi \hat{\pi}_t + \sigma_R \epsilon_{R,t}, \quad \psi = \frac{1}{\beta}, \quad (16.14)$$

这也是做了相当程度的简化。讨论货币政策的 DSGE 模型常常要引入利率的平滑机制, 以及央行在制定货币政策是常常也需要考虑其他一些反映实际经济运行状况的指标如实际产出、潜在产出等 (第2.6.2节), 又称泰勒法则 (Taylor's rule), 如见第2.7节。外甥货币政策冲击  $\epsilon_{R,t}$  用于反映无法提前预知的系统性偏差。假定  $\psi = \frac{1}{\beta}$  时为了确保这样一组含有理性预期的线性方程系统, 有且只有一个稳定解 (理性期望模型的求解, 见第十章。)。

公共部门的主要工作是决定债务水平和一揽子税等, 使得政府的预算约束条件得到满足。本模型中假定稳定状态下债务为 0, 总税收等于总转移支付, 因而在均衡条件中可以不考虑。

### 16.1.2 模型求解

我们从整个经济系统中选取四个变量作为核心状态变量，进行系统缩减，以减轻计算负担（第12.5.5节）。由前述对数线性化的系统结构可见，四个运动法则方程  $\{\hat{x}_t, \hat{lsh}_t, \hat{\pi}_t, \hat{R}_t\}$  构成的缩减系统也是线性的。以下依次说明。

#### 产出的运动法则

将泰勒法则（16.14）代入欧拉等式（16.5）以消除名义利率  $\hat{R}_t$

$$\hat{x}_t = E_t \hat{x}_{t+1} - \underbrace{\left( \frac{1}{\beta} \hat{\pi}_t - E_t \hat{\pi}_{t+1} \right)}_{=: \mathcal{A}} - \sigma_R \epsilon_{R,t} + E_t z_{t+1}, \quad (16.15)$$

根据 NKPC（16.10）和劳动力供应的决策式（16.6）有

$$\mathcal{A} = \frac{1}{\beta} \hat{\pi}_t - E_t \hat{\pi}_{t+1} = \frac{\kappa_p}{\beta} (\hat{w}_t + \lambda_t) = \frac{\kappa_p}{\beta} [(1 + \nu) \hat{x}_t + \phi_t + \lambda_t], \quad (16.16)$$

（16.15）因此变为

$$\begin{aligned} \hat{x}_t &= E_t \hat{x}_{t+1} - \left\{ \frac{\kappa_p}{\beta} [(1 + \nu) \hat{x}_t + \phi_t + \lambda_t] + \sigma_R \epsilon_{R,t} \right\} + E_t z_{t+1}, \\ &\hookrightarrow \underbrace{\left[ 1 + \frac{\kappa_p}{\beta} (1 + \nu) \right]}_{\psi_p} \hat{x}_t = E_t \hat{x}_{t+1} - \frac{\kappa_p}{\beta} (\phi_t + \lambda_t) - \sigma_R \epsilon_{R,t} + E_t z_{t+1}, \\ &\hookrightarrow \hat{x}_t = \psi_p E_t \hat{x}_{t+1} - \frac{\kappa_p \psi_p}{\beta} (\phi_t + \lambda_t) - \psi_p \sigma_R \epsilon_{R,t} + \psi_p E_t z_{t+1}, \quad \psi_p := \left[ 1 + \frac{\kappa_p}{\beta} (1 + \nu) \right] \in [0, 1], \end{aligned} \quad (16.17)$$

不难看出，产出的运动法则表示为一个含有期望的线性差分方程。由此我们将  $\hat{x}_t$  表示为线性决策形式（decision rule）

$$\begin{aligned} \hat{x}_t &:= \hat{x}_t(\phi_t, \lambda_t, z_t, \epsilon_{R,t}) = x_\phi \phi_t + x_\lambda \lambda_t + x_z z_t + x_{\epsilon_R} \epsilon_{R,t}, \\ x_\phi &= -\frac{\psi_p \kappa_p}{\beta} \frac{1}{1 - \psi_p \rho_\phi}, \\ x_\lambda &= -\frac{\psi_p \kappa_p}{\beta} \frac{1}{1 - \psi_p \rho_\lambda}, \\ x_z &= \frac{\rho_z \psi_p}{1 - \psi_p \rho_z}, \\ x_{\epsilon_R} &= -\psi_p \sigma_R, \end{aligned} \quad (16.18)$$

作为泛函形式  $E_t \mathcal{H}(\hat{x}_t(\phi_t, \lambda_t, z_t, \epsilon_{R,t})) = 0$  的解：

$$E_t \left[ \hat{x}(\phi_t, \lambda_t, z_t, \epsilon_{R,t}) - \psi_p \hat{x} \begin{pmatrix} \rho_\phi \phi_t + \sigma_\phi \epsilon_{\phi,t+1}, \\ \rho_\lambda \lambda_t + \sigma_\lambda \epsilon_{\lambda,t+1}, \\ \rho_z z_t + \sigma_z \epsilon_{z,t+1}, \\ \epsilon_{R,t+1} \end{pmatrix} + \frac{\psi_p \kappa_p}{\beta} (\phi_t + \lambda_t) + \psi_p \sigma_R \epsilon_{R,t} - \psi_p E_t z_{t+1} \right] = 0. \quad (16.19)$$

#### 劳动收入份额的运动法则

将劳动力供应的决定式（16.6）和（16.13）可得

$$\hat{lsh}_t = [(1 + \nu) \hat{x}_t] + \phi_t,$$

引入(16.18), 用  $\hat{x}_t(\phi_t, \lambda_t, z_t, \epsilon_{R,t})$  替代掉  $\hat{x}_t$ , 可得劳动收入份额的线性决策式

$$\begin{aligned}\widehat{lsh}_t &:= \widehat{lsh}_t(\phi_t, \lambda_t, z_t, \epsilon_{R,t}) \\ &= [(1+\nu)x_\phi + 1]\phi_t + (1+\nu)x_\lambda\lambda_t + (1+\nu)x_zz_t + (1+\nu)x_{\epsilon_R}\epsilon_{R,t}.\end{aligned}\quad (16.20)$$

### 通货膨胀率的运动法则

(16.20)代回 NKPC (16.10), 用收入份额的线性决策式  $\widehat{lsh}_t(\phi_t, \lambda_t, z_t, \epsilon_{R,t})$  替代  $\hat{w}_t$ 。将通货膨胀率  $\hat{\pi}_t$  表示为近似线性形式  $\hat{\pi}_t(\phi_t, \lambda_t, z_t, \epsilon_{R,t})$ , 作为泛函  $E_t \mathcal{H} \hat{\pi}_t(\phi_t, \lambda_t, z_t, \epsilon_{R,t}) = 0$  的解

$$E_t \left[ \hat{\pi}(\phi_t, \lambda_t, z_t, \epsilon_{R,t}) - \beta \hat{\pi} \begin{pmatrix} \rho_\phi \phi_t + \sigma_\phi \epsilon_{\phi,t+1}, \\ \rho_\lambda \lambda_t + \sigma_\lambda \epsilon_{\lambda,t+1}, \\ \rho_z z_t + \sigma_z \epsilon_{z,t+1}, \\ \epsilon_{R,t+1} \end{pmatrix} - \kappa_p \underbrace{\widehat{lsh}(\phi_t, \lambda_t, z_t, \epsilon_{R,t})}_{(16.20)} - \kappa_p \lambda_t \right], \quad (16.21)$$

进而可得通货膨胀率的线性决策式

$$\begin{aligned}\hat{\pi}_t &= \hat{\pi}(\phi_t, \lambda_t, z_t, \epsilon_{R,t}) = \frac{\kappa_p}{1-\beta\rho_\phi} [1 + (1+\nu)x_\phi] \phi_t + \frac{\kappa_p}{1-\beta\rho_\lambda} [1 + (1+\nu)x_\lambda] \lambda_t \\ &\quad + \frac{\kappa_p}{1-\beta\rho_z} (1+\nu)x_z z_t + \kappa_p (1+\nu)x_{\epsilon_R} \epsilon_{R,t}\end{aligned}\quad (16.22)$$

### 利率的运动法则

将(16.22)代入泰勒法则(16.14), 用线性近似决策式  $\hat{\pi}_t(\phi_t, \lambda_t, z_t, \epsilon_{R,t})$  替代  $\hat{\pi}_t$ , 可得

$$\begin{aligned}\hat{R}_t &= \frac{\kappa_p}{\beta} \frac{1}{1-\beta\rho_\phi} [1 + (1+\nu)x_\phi] \phi_t + \frac{\kappa_p}{\beta} \frac{1}{1-\beta\rho_\lambda} [1 + (1+\nu)x_\lambda] \lambda_t \\ &\quad + \frac{\kappa_p}{\beta} \frac{1}{1-\beta\rho_z} (1+\nu)x_z z_t + \left[ \frac{\kappa_p}{\beta} (1+\nu)x_{\epsilon_R} + \sigma_R \right] \epsilon_{R,t}.\end{aligned}\quad (16.23)$$

### 外生冲击

外生冲击  $\{\epsilon_{z_t}, \phi_t, \lambda_t\}$  的运动法则分别由 (16.1), (16.7), (16.11) 给出, 假定  $\{\epsilon_{R,t}, \epsilon_{z,t}, \epsilon_{\phi,t}, \epsilon_{\lambda,t}\}$  都是鞅差序列 (Martingale Difference Sequence, MDS, Definition 16.1)。

**Definition 16.1** (鞅差序列). 概率论中, 假定一个定义在概率空间  $(\omega, F, P)$  中的随机序列  $X = \{x(n)\}_{n \geq 0}$ , 其中  $F = \{F(n)\}$  是一个上升  $\sigma$  代数族 (Definition 14.4)。如果以下条件成立, 那么我们称  $X$  为一个鞅差序列 (Martingale Difference Sequence, MDS):

1.  $X$  是  $F$ -适应的, 即对任一  $n \geq 0$ ,  $X(n)$  都是  $F(n)$ -可测度的,
2. 对于任一  $n \geq 0$ ,  $X(n)$  都可积,
3. 对任一  $n \geq 0$ , 满足  $E[X(n+1)|F(n)] = 0$ , a.s., 其中 a.s. 表示几乎总是 (almost surely)。

### 16.1.3 状态——空间表现式

为了更好的比较模拟数据与观测数据，需要对变量做处理。测量方程 (measurement equation) 表现为以下形式

$$\begin{aligned} \log\left(\frac{X_t}{X_{t-1}}\right) &= \hat{x}_t - \hat{x}_{t-1} + z_t + \log\gamma, \\ \log(lsh_t) &= \widehat{lsh}_t + \log(lsh), \\ \log\pi_t &= \hat{\pi}_t + \log\pi^*, \\ \log R_t &= \hat{R}_t + \log\left(\frac{\pi^*\gamma}{\beta}\right). \end{aligned} \quad (16.24)$$

通过测量方程的转换，我们可以将 DSGE 模型的模拟数据表示为状态空间模型的形式 (Shumway and Stoffer, 2017, Ch. 6)。具体说来，定义一组状态变量组成的向量  $s_t$ ，和由 13 个参数构成的参数向量  $\theta$

$$\begin{aligned} \underbrace{s_t}_{[n_s \times 1]} &\equiv [\phi_t, \lambda_t, z_t, \epsilon_{R,t}, \hat{x}_{t-1}]^\top, \\ \underbrace{\theta}_{[13 \times 1]} &\equiv [\beta, \gamma, \lambda, \pi^*, \xi_p, \nu, \rho_\phi, \rho_\lambda, \rho_z, \sigma_\phi, \sigma_\lambda, \sigma_z, \sigma_R]^\top. \end{aligned} \quad (16.25)$$

这样，可以将状态转移方程 (state transition equation) 表示如下

$$s_t = \phi_1(\theta) s_{t-1} + \phi_\epsilon(\theta) \epsilon_t, \quad (16.26)$$

其中

- 预测误差向量  $\epsilon_t$

$$\underbrace{\epsilon_t}_{[n_\epsilon \times 1]} \equiv [\epsilon_{\phi,t}, \epsilon_{\lambda,t}, \epsilon_{z,t}, \epsilon_{R,t}]^\top,$$

- $\phi_1(\theta)$  和  $\phi_\epsilon(\theta)$  是系数矩阵，矩阵中的元素值反映以下几组关系
  - $\phi_t \leftrightarrow \phi_{t-1}$ , (16.7)
  - $\lambda_t \leftrightarrow \lambda_{t-1}$ , (16.11)
  - $z_t \leftrightarrow z_{t-1}$ , (16.1)
  - $\epsilon_{R,t} \equiv \epsilon_{R,t}$
  - $\hat{x}_{t-1}$  的决定式, (16.17)。

另一方面，对于观测数据，定义为向量  $y_t$

$$\underbrace{y_t}_{[n_y \times 1]} \equiv \underbrace{M_y^\top}_{[n_y \times 4]} \left[ \log\left(\frac{X_t}{X_{t-1}}\right), \log lsh_t, \log\pi_t, \log R_t \right]^\top, \quad (16.27)$$

其中  $M_y^\top$  是一个选择矩阵 (selection matrix)，用于选择系数矩阵  $\psi_0(\theta)$ ,  $\psi_1(\theta)$  中的行。这样，将观测数据的测量方程表示为

$$y_t = \psi_0(\theta) + \psi_1(\theta) s_t, \quad (16.28)$$

其中  $s_t$  由(16.26)定义，其值由四条运动法则组成的近似线性系统所给出，(16.18), (16.20), (16.22), (16.23)；系数矩阵  $\psi_0(\theta)$ ,  $\psi_1(\theta)$  可由(16.24)计算求得。

### 从 DSGE 模型到状态空间模型

这样，典型 DSGE 模型所对应的状态——空间模型为如下表达式：

$$\begin{aligned} y_t &= \psi_0(\theta) + \psi_1(\theta) s_t, \\ s_t &= \phi_1(\theta) s_{t-1} + \phi_\epsilon(\theta) \epsilon_t, \end{aligned}$$

其中系数矩阵分别为

$$\begin{aligned} \psi_0(\theta) &= M_y^\top \begin{pmatrix} \log \gamma \\ \log(lsh) \\ \log \pi^* \\ \log\left(\frac{\gamma \pi^*}{\beta}\right) \end{pmatrix}, \\ \psi_1(\theta) &= M_y^\top \begin{pmatrix} x_\phi & x_\lambda & x_z + 1 & x_{\epsilon_R} & -1 \\ 1 + (1 + \nu) x_\phi & (1 + \nu) x_\lambda & (1 + \nu) x_z & (1 + \nu) x_{\epsilon_R} & 0 \\ \frac{\kappa_p}{1 - \beta \rho_\phi} [1 + (1 + \nu) x_\phi] & \frac{\kappa_p}{1 - \beta \rho_\lambda} [1 + (1 + \nu) x_\lambda] & \frac{\kappa_p}{1 - \beta \rho_z} (1 + \nu) x_z & \kappa_p (1 + \nu) x_{\epsilon_R} & 0 \\ \frac{\kappa_p}{\beta(1 - \beta \rho_\phi)} [1 + (1 + \nu) x_\phi] & \frac{\kappa_p}{\beta(1 - \beta \rho_\lambda)} [1 + (1 + \nu) x_\lambda] & \frac{\kappa_p}{\beta(1 - \beta \rho_z)} (1 + \nu) x_z & \frac{\kappa_p}{\beta} (1 + \nu) x_{\epsilon_R} + \sigma_R & 0 \end{pmatrix}, \\ \phi_1(\theta) &= \begin{pmatrix} \rho_\phi & 0 & 0 & 0 & 0 \\ 0 & \rho_\lambda & 0 & 0 & 0 \\ 0 & 0 & \rho_z & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ x_\phi & x_\lambda & x_z & x_{\epsilon_R} & 0 \end{pmatrix}, \\ \phi_\epsilon(\theta) &= \begin{pmatrix} \sigma_\phi & 0 & 0 & 0 \\ 0 & \sigma_\lambda & 0 & 0 \\ 0 & 0 & \sigma_z & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \end{aligned}$$

其中  $M_y^\top$  是一个  $[n_y \times 4]$  的选择矩阵，用于选取  $\psi_0(\theta)$  和  $\psi_1(\theta)$  中的特定行。系数  $\{x_\phi, x_\lambda, x_z, x_{\epsilon_R}\}$  的计算方法见(16.18)。

由(16.28) (16.26)所构成的典型 DSGE 模型的状态——空间表现形式，称为一系列后续计量经济分析的基础。它描述了在给定模型参数  $\theta$  的基础上，观测数据  $y_t$  和状态变量  $s_t$  的联合分布概率

$$p(Y_{1:T}, S_{1:T} | \theta) = \int \left[ \prod_{t=1}^T p(y_t | s_t, \theta) p(s_t | s_{t-1}, \theta) \right] p(S_0 | \theta) dS_0, \quad (16.29)$$

其中  $Y_{1:T} = \{y_1, \dots, y_T\}$ ,  $S_{1:T} = \{s_1, \dots, s_T\}$ 。状态变量集合  $S_{1:T}$  中，常常有一部分信息无法直接观测到，直接利用(16.29)计算联合分布概率的思路很可能不可行。对此，常见的处理方法是，基于实际观测数据，计算边际分布概率  $p(Y_{1:T} | \theta)$

$$p(Y_{1:T} | \theta) = \int p(Y_{1:T}, S_{1:T} | \theta) dS_{1:T}. \quad (16.30)$$

边际分布概率  $p(Y_{1:T} | \theta)$  是一个关于  $\theta$  的方程，有时也称似然方程 (likelihood function，见第十七章)，它在计量推断过程中的作用非常重要。

**Remark 16.1** (“状态”变量). 来看以下两组变量，第一组是  $\{\phi_t, \lambda_t, z_t, \epsilon_{R,t}\}$ ，第二组是  $s_t = \{\phi_t, \lambda_t, z_t, \epsilon_{R,t}, \hat{x}_{t-1}\}$ 。前者是经济学意义上的状态变量，与经济系统中行为人的跨期最优决策有关。后者是计量经济学意义上的状态变量，用于将 DSGE 模型的近似解投进状态——空间表现式(16.28) (16.26)中去。

笔记打出来之后  
是在 sec 10，做  
一个 ref

1. 前者在我们的简单 DSGE 模型设定中，都假定是外生的。但一个更复杂的 DSGE 模型中， $s_t$  完全可以包括其他一些内生变量如实物资本存量等。
2. 测量方程(16.28) (16.26)中产出的增速  $\hat{x}_{t-1}$  也可以用产出水平  $x_{t-1}, x_{t-2}$  来表示，如果用产出水平代替增速，就需要在参数向量中加入稳态产出  $x^*$ ，在  $s_t$  中减去  $\hat{x}_{t-1}$ ，加入  $\log Z_t/\gamma^t$ ，以及在测量方程中产出的一项上，考虑到确定性趋势  $(\log \gamma)t$  的作用。
3. 也可以在测量方程(16.26)中加入额外的测量误差项。
4. 如果在前一阶段，对 DSGE 模型是作高阶扰动或者投影法的近似求解的，那么在这一阶段，随着状态变量  $s_t$  的定义越是精确，就越可能出现状态转移方程(16.28)和测量方程(16.26)是非线性的情况。

## 16.2 模型的性质

一旦我们对状态转移方程(16.28)含有新息 (innovation) 的误差向量  $\epsilon_t$  的分布信息明确了 (这反映在系数矩阵  $\phi_\epsilon(\theta)$  中)，就可以最终确定 DSGE 模型中诸多变量的概率分步。由此，如果进一步将  $\epsilon_t$  的设为满足独立同分布

$$\epsilon_t \sim i.i.d. \mathcal{N}(0, I),$$

其中  $I$  是个单位矩阵，那么，基于这样的概率结构，我们很容易得 DSGE 模型的一组特征如下，这组特征有助于我们更好估计参数向量  $\theta$  的具体数值，以及模型的拟合度。本节介绍这些特征。

在介绍之前，出于简化模型的考虑，我们先对  $\theta$  系数值作校准，暂时赋予定值，见表 16.1。

表 16.1: 典型 DSGE 模型的参数校准

参数	值	参数	值
$\beta$	1/1.01	$\gamma$	$\exp(0.005)$
$\lambda$	0.15	$\pi^*$	$\exp(0.005)$
$\xi_p$	0.65	$\nu$	0
$\rho_\phi$	0.94	$\rho_\lambda$	0.88
$\rho_z$	0.13	$\sigma_\lambda$	0.01
$\sigma_\phi$	0.01	$\sigma_R$	0.01
$\sigma_z$	0.01		

### 16.2.1 自协方差，预测误差方差

DSGE 模型能在多大程度上解释实际观测数据中的周期性波动？这就需要做模型精确度的分析，值得关注的特征是自协方差、预测误差方差。如前文所述，DSGE 模型关注的主要对象时经济系统中的周期性波动。因此，模型中的方差 (variances)、自相关 (auto correlations)、互相关 (cross correlations) 等信息应当予以重点关注。

### 协方差

对于近似线性的 DSGE 模型，对应状态——空间表现形式(16.28) (16.26)，其自协方差方程 (autocovariance)、跨协方差方程 (crosscovariance) 可表示为

$$\begin{aligned}\Gamma_{yy}(h) &= E[y_t y_{t-h}], \\ \Gamma_{ss}(h) &= E[s_t s_{t-h}], \\ \Gamma_{ys}(h) &= E[y_t s_{t-h}^\top],\end{aligned}\tag{16.31}$$

此外也假定预测误差满足

$$E[\epsilon_t \epsilon_t^\top] = I.\tag{16.32}$$

那么，可以将  $s_t$  的协方差矩阵表示为如下李雅普诺夫方程 (Lyapunov function) 的解<sup>1</sup>

$$\Gamma_{ss}(0) = \phi_1 \Gamma_{ss}(0) \phi_1^\top + \phi_\epsilon \phi_\epsilon^\top\tag{16.33}$$

**Algorithm 16.1.** 在 Mathematica 中，先计算  $\phi_\epsilon \phi_\epsilon^\top$  的值，随后用

```
In[3]:= DiscreteLyapunovSolve[a, c]
```

来求解。Matlab 中对应的命令是 *dylap*。生成的  $\Gamma_{ss}(0)$  为

$$\Gamma_{ss}(0) = \begin{pmatrix} -0.000859107 & 0. & 0. & 0. & 0.000619325 \\ 0. & -0.000443262 & 0. & 0. & 0.000242601 \\ 0. & 0. & -0.000101719 & 0. & 1.6265930319175398E-6 \\ 0. & 0. & 0. & -1. & 0. \\ 0.000619325 & 0.000242601 & 1.62659303191754E-6 & 0. & -0.000748026 \end{pmatrix}$$

基于  $\Gamma_{ss}(0)$ ，可进一步计算  $\Gamma_{ss}(h)$ ,  $h > 0$

$$\Gamma_{ss}(h) = \phi_1^h \Gamma_{ss}(0).\tag{16.34}$$

进一步，由测量方程(16.28)可得  $\Gamma_{yy}(h)$ ,  $\Gamma_{ys}(h)$  的值

$$\Gamma_{yy}(h) = \psi_1 \Gamma_{ss}(h) \psi_1^\top,\tag{16.35}$$

$$\Gamma_{ys}(h) = \psi_1 \Gamma_{ss}(h) \psi_1^\top.\tag{16.36}$$

有了这一系列协方差矩阵后，便可进一步计算相关和自相关矩阵。

### 预测误差

另一种求解思路是，将状态向量  $s_t$  理解为一组从 0 到无穷阶移动平均 (moving average, MA) 过程的和

$$\begin{aligned}y_t &= \psi_0 + \psi_1 s_t \\ &= \psi_0 + \psi_1 \sum_{s=0}^{\infty} \phi_1^s \psi_1 \epsilon_{t-s},\end{aligned}\tag{16.37}$$

---

<sup>1</sup> 将表16.1的校准参数值赋给状态——空间表现式 (16.28)中的系数矩阵  $\phi_1(\theta)$  后，可得特征根 (善用 Mathematica!)  $Eigen = \{0.94, 0.88, 0.13, 0, 0\}$ ，即  $|Eigen| < 1$ 。这表示  $s_t$  的 VAR(1) 运动式是协方差平稳的。

在此基础上，可以计算如过去  $t-h$  期对当前  $t$  期的预测误差，定义为  $e_{t|t-h}$

$$e_{t|t-h} = y_t - E_{t-h}(y_t) = \psi_1 \sum_{s=0}^{h-1} \psi_1^s \phi_\epsilon \epsilon_{t-s}, \quad (16.38)$$

又称  $h$  步前向预测误差 (h-stepahead forecast error)。

那么可建立  $h$  步前向预测误差的协方差矩阵

$$E [e_{t|t-h} e_{t|t-h}^\top] = \psi_1 \left( \sum_{s=0}^{h-1} \phi_1^s \phi_\epsilon \phi_\epsilon^\top \phi_1^{s,\top} \right) \psi_1^\top. \quad (16.39)$$

对(16.39)取  $h \rightarrow \infty$  的极限，有

$$\lim_{h \rightarrow \infty} E [e_{t|t-h} e_{t|t-h}^\top] \approx \Gamma_{ss}(0). \quad (16.40)$$

那么在假定(16.32)成立的前提下，可以进一步计算预测误差的协方差矩阵（协方阵）的分解如下。设  $I^{(j)}$  表示单位矩阵  $I$  的对角元素中，除了第  $j$  个元素之外的元素值都为 0。那么  $I$  是全部  $I^{(j)}$  之和

$$I = \sum_{j=1}^{n_\epsilon} I^{(j)}, \quad j = 1, \dots, n_\epsilon, \quad (16.41)$$

那么编号为  $j$  的外生冲击，对观测数据的预测误差的影响表现为  $e_{t|t-h}^{(j)}$

$$e_{t|t-h}^{(j)} = \psi_1 \sum_{s=0}^{h-1} \phi_1^s \phi_\epsilon I^{(j)} \epsilon_{t-s}. \quad (16.42)$$

第  $j$  个外生冲击的  $h$  步前向预测误差的协方差矩阵因此为

$$E [e_{t|t-h}^{(j)} e_{t|t-h}^{(j),\top}] = \psi_1 \left( \sum_{s=0}^{h-1} \phi_1^s \phi_\epsilon I^{(j)} \phi_\epsilon^\top \phi_1^{s,\top} \right) \psi_1^\top. \quad (16.43)$$

那么， $h$  期第  $j$  个外生冲击，对  $t$  期第  $i$  个观测数据  $y_{i,t}$  的预测误差的影响（方差），表示为如下比值关系，定义为预测误差方差分解 (forecast error variance decomposition, FEVD)

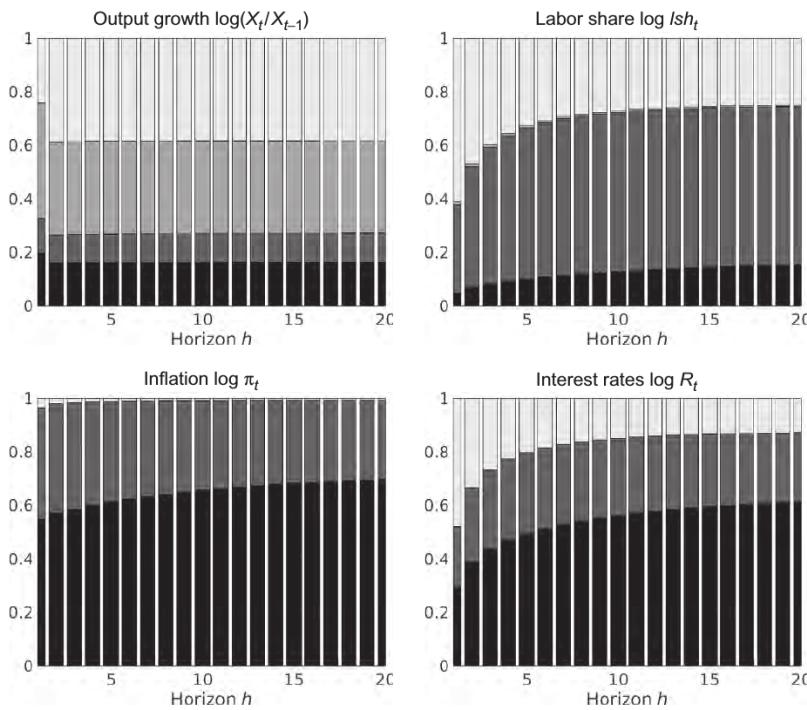
$$FEVD(i, j, h) = \frac{[e_{t|t-h}^{(j)}]_{ii}}{[E [e_{t|t-h} e_{t|t-h}^\top]]_{ii}} = \frac{\left[ \psi_1 \left( \sum_{s=0}^{h-1} \phi_1^s \phi_\epsilon I^{(j)} \phi_\epsilon^\top \phi_1^{s,\top} \right) \psi_1^\top \right]_{ii}}{\left[ \psi_1 \left( \sum_{s=0}^{h-1} \phi_1^s \phi_\epsilon \phi_\epsilon^\top \phi_1^{s,\top} \right) \psi_1^\top \right]_{ii}}, \quad (16.44)$$

其中  $[ \cdot ]_{ij}$  表示矩阵的第  $(i, j)$  元素。

图16.1给出了4个冲击  $\{\phi_t, \lambda_t, z_t, \epsilon_{R,t}\}$  对观测变量  $\{\log(X_t/X_{t-1}), \log lsh_t, \log \pi_t, \log R_t\}$  的预测误差及其分解。每一单条（值为标准化的1）表示一个累积预测误差的方差分解，单条内分为四种颜色，从下倒上由深到浅依次表示偏好冲击  $\phi_t$ ，价格加成冲击  $\lambda_t$ ，技术冲击  $z_t$ ，和货币政策冲击  $\epsilon_{R,t}$ 。基于表16.1给出的校准参数值，不难看出

- 产出增速的波动绝大多数归因于技术冲击  $z_t$  和货币政策冲击  $\epsilon_{R,t}$ ，
- 劳动收入的波动主要归因于价格加成的冲击  $\lambda_t$ ，并且从长期来看影响更为显著，
- 通货膨胀率和利率的波动收到偏好冲击  $\phi_t$  和价格加成冲击  $\lambda_t$  的强烈影响。

图 16.1: 预测误差的方差分解



### 16.2.2 光谱分析

第16.2.1节主要是从精确度的层面入手，分析预测误差的成因及影响。此外我们也可以从频率分析的角度入手<sup>2</sup>。先来看一个简单的线性周期模型，设  $\{y\}_t$  是个矢量时间序列周期过程，满足

$$\begin{aligned} y_t &= 2 \sum_{j=1}^m a_j \cos(\omega_j t + \theta_j) \\ &= 2 \sum_{j=1}^m a_j [\cos \theta_j \cos(\omega_j t) - \sin \theta_j \sin(\omega_j t)], \quad \theta_j \sim i.i.d. U[-\pi, \pi], 0 \leq \omega_j \leq \omega_{j+1} \leq \pi. \end{aligned} \tag{16.45}$$

随机变量  $\theta_j$  的导致  $y_t$  发生相位移动 (phase shift)。为了简化表述，假定初始相位移动发生在无限久远之前。这样，(16.45)将实值域中的  $y_t$  理解为傅里叶域中一系列弦波曲线之和，这些波以不同的频率彼此区分。

$\omega_j$  的单位往往与  $t$  的单位有关，后者在 DSGE 模型中设定，如以季度为时间单位，含有 32 个季度的数据，那么  $\omega_j = \frac{2\pi}{32}$ 。一个常见的经济周期模型，可能会包含 8 到 23 个季度数据，对应的  $\omega_j \in [0.196, 0.785]$ 。

对于  $\omega \in [-\pi, \pi]$ ，时间序列  $y_t$  (16.45) 在傅里叶域中的光谱分布，可以表示为

$$\mathcal{F}_{yy}(\omega) = \sum_{j=-m}^m \mathbb{E} \left[ A(\omega_j) \overline{A(\omega_j)} \right] \mathbb{I}\{\omega_j \leq \omega\}, \tag{16.46}$$

- $\omega_{-j} = -\omega_j, i = \sqrt{-1}$ ,

---

<sup>2</sup>参考第12.G节。在时间序列分析中，频率域技术的应用有着悠久的历史，经典参考文献如 (Koopmans, 1995; Diebold et al., 2010)。

- 基方程为  $A(\omega_j)$  及其复共轭 (complex conjugate, 第16.A.2节)  $\overline{A(\omega_j)}$  的卷积的平均值。

$$A(\omega_j) = \begin{cases} a_j (\cos \theta_{|j|} + i \sin \theta_{|j|}) & (j > 0), \\ a_j (\cos \theta_{|j|} - i \sin \theta_{|j|}) & (j < 0), \end{cases}$$

- $\mathbb{I}(\cdot)$  是一个指示方程 (indicator function)，满足

$$\mathbb{I}\{\omega_j \leq \omega\} = \begin{cases} 1 & \text{如果 } \{\omega_j \leq \omega\} == \text{TRUE}, \\ 0 & \text{否则} \end{cases}. \quad (16.47)$$

如果  $F_{yy}(\omega)$  对  $\omega$  可导，那么其一阶导数就是光谱密度 (spectral density) 方程  $f_{yy}(\omega)$

$$f_{yy}(\omega) = \frac{d}{d\omega} F_{yy}(\omega). \quad (16.48)$$

某个过程的光谱密度方程若是  $f_{yy}(\omega)$  (16.48)，那么该过程的协方差可以表示为  $\Gamma_{yy}(h)$

$$\Gamma_{yy}(h) = \int_{(-\pi, \pi]} \exp(ih\omega) f_{yy}(\omega) d\omega. \quad (16.49)$$

那么，线性周期方程  $y_t$  (16.45) 的自协方差  $\Gamma_{yy}(h)$  为

$$\Gamma_{yy}(h) = \sum_{j=-m}^m \mathbb{E}[A(\omega_j) \overline{A(\omega_j)}] \exp(ih\omega). \quad (16.50)$$

在给定光谱密度  $f_{yy}(\omega)$  的情况下，我们可以算得自协方差  $\Gamma_{yy}(h)$ 。反之亦然，已知  $\Gamma_{yy}(h)$ ，可求得光谱密度  $f_{yy}(\omega)$

$$f_{yy}(\omega) = \frac{d}{d\omega} F_{yy}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \Gamma_{yy}(h) \exp(-i\omega h). \quad (16.51)$$

结合自协方差数列  $\Gamma_{yy}(h)$  (16.49) 和光谱密度  $f_{yy}(\omega)$  (16.51) 不难看出，二者含有完全相同的信息 (帕塞瓦尔定理, Theorem 12.63)。

那么，回到典型 DSGE 模型的状态——空间表现式(16.28) (16.26) 中，状态向量  $s_t$  的自协方差  $\Gamma_{ss}(h)$  由(16.34)计算而得， $\Gamma_{ss}(0)$  由(16.33)计算而得。将它代入(16.51)可得状态向量  $s_t$  的光谱密度方程  $f_{ss}(\omega)$

$$\begin{aligned} f_{ss}(\omega) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \Gamma_{ss}(h) \exp(-i\omega h) \\ &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \phi_1^h \Gamma_{ss}(0) \exp(-i\omega h) \\ &= \frac{1}{2\pi} [I - \phi_1^\top \exp(i\omega)]^{-1} \phi_\epsilon \phi_\epsilon^\top [I - \phi_1 \exp(-i\omega)]^{-1}. \end{aligned} \quad (16.52)$$

进而，第  $j$  个冲击对光谱密度  $f_{ss}(\omega)$  (16.52) 的影响，记为  $f_{ss}^{(j)}(\omega)$ ，计算如下

$$f_{ss}^{(j)}(\omega) = \frac{1}{2\pi} [I - \phi_1^\top \exp(i\omega)]^{-1} \phi_\epsilon I^{(j)} \phi_\epsilon^\top [I - \phi_1 \exp(-i\omega)]^{-1}, \quad (16.53)$$

其中  $I^{(j)}$  的定义见(16.41)。

观测数据向量  $y_t$  的光谱密度方程  $f_{yy}(\omega)$

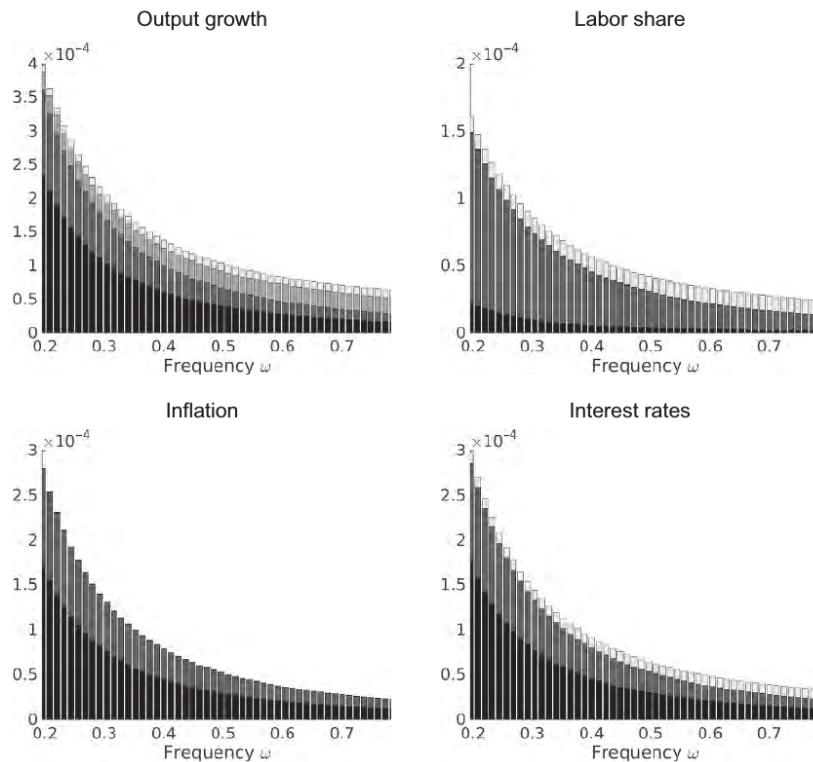
$$f_{yy}(\omega) = \psi_1 f_{ss}(\omega) \psi_1^\top, \quad (16.54)$$

第  $j$  个冲击对光谱密度  $f_{yy}(\omega)$  (16.54) 的影响, 记为  $f_{yy}^{(j)}(\omega)$ , 计算如下

$$f_{yy}^{(j)}(\omega) = \psi_1 f_{ss}^j(\omega) \psi_1^\top. \quad (16.55)$$

需要注意的是, 光谱密度方程  $f_{yy}(\omega)$  是一个方阵, 对角线元素  $f_{yy}^{(j)}$  反映第  $j$  个冲击的影响, 典型 DSGE 模型中观测数据向量中含有 4 个变量的时间序列, 分别是产出增速、劳动力收入份额、通货膨胀和利率, 进而基于表 16.1 提供的参数校准, 图 16.2 绘出 4 种冲击对 4 个观测变量的光谱密度分解, 颜色由深到浅分别表示  $\phi_t, \lambda_t, z_t, \epsilon_{R,t}$ , 4 种冲击加总构成累积光谱密度, 表现为图中的每条竖块。由于 4 个冲击彼此不相关, 且均设为符合 AR(1) 过程, 那么不难看出, 在初始频率处光谱密度值最高 (冲击的影响最大), 随着频率的增大而逐渐衰减。

图 16.2: 预测误差的方差分解



### 16.2.3 冲击响应方程

研究外生冲击对经济系统的影响, 一个重要的工具是冲击响应方程 (impulse-response functions, IRFs)。DSGE 模型中的冲击响应方程可以定义为以下两个条件期望之差

$$IRF(i, j, h | s_{t-1}) = E[y_{i,t+h} | s_{t-1}, \epsilon_{j,t} = 1] - E[y_{i,t+h} | s_{t-1}], \quad (16.56)$$

两个条件期望

- 相同之处:

- 都与上一期状态  $s_{t-1}$  有关, 从而可以向前追溯到初始状态,
- 都与冲击  $\epsilon_t$  对随后  $t+1, \dots, t+h$  期的持续影响 (积分) 有关,

- 不同之处

- 前者还包括条件  $\epsilon_{j,t} = 1$ , 即  $t$  期第  $j$  个冲击出现,
- 后者则假定  $\epsilon_{j,t} = 0$ , 即  $t$  期第  $j$  个冲击不发生。

对于状态——空间形式表现的典型 DSGE 模型系统(16.28) (16.26), 由它的线性性质和分布性质  $E[\epsilon_{t+h}|s_{t-1}] = 0, h = 0, 1, \dots$  可得, 第  $j$  个冲击的响应方程根据(16.56)写为

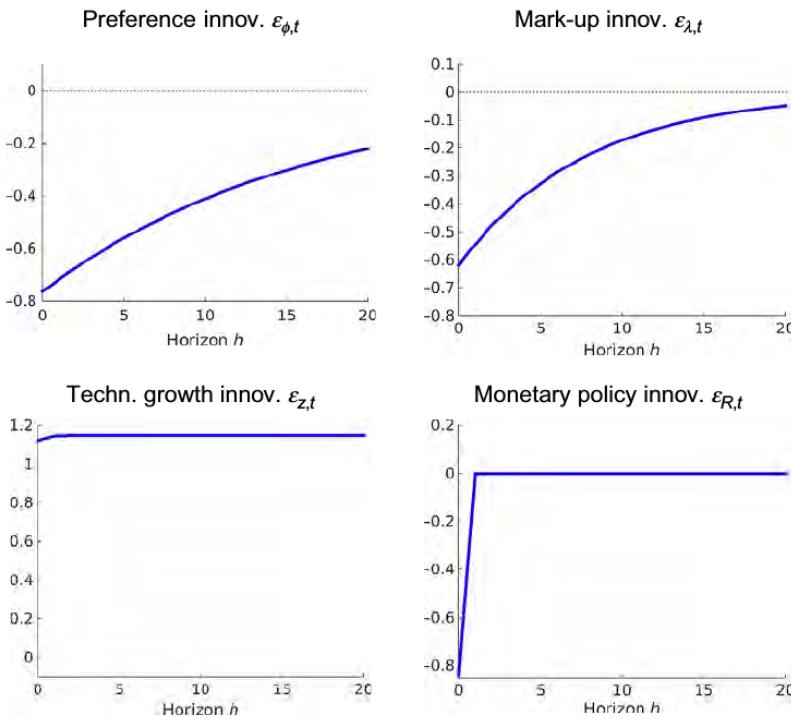
$$IRF(\cdot, j, h) = \frac{\partial}{\partial \epsilon_{j,t}} \psi_1 = \psi_1 \phi_1^h [\phi_\epsilon]_{j,j}, \quad (16.57)$$

上式中比起定义式(16.56)来省略掉了条件  $|s_{t-1}$ , 是由于分布性质的假设。 $[\cdot]_j$  表示矩阵的第  $j$  列。

图16.3, 以模型中内生变量产出增速  $100 \times \log X_{t+h}/X_t$  为例, 依次绘出  $t$  期 4 种外生冲击的新息 (innovation)  $\epsilon_{\phi,t}, \epsilon_{\lambda,t}, \epsilon_{z,t}, \epsilon_{R,t}$  在随后的  $t+h, h = 0, 1, \dots, 20$  期的影响。其中产出增速的线性近似决策式由(16.18)给出, 外生冲击的 4 个 AR(1) 过程由前节依次给定。不难看出

- 偏好冲击和价格加成的新息冲击产生负面影响, 产出随着时间变化逐渐恢复到稳态水平。
- 技术冲击的新息产生正影响, 并且正影响是持续性的。
- 货币冲击新息的影响只持续 1 期。

图 16.3: 产出的冲击响应方程



#### 16.2.4 条件矩的限制

DSGE 模型中的跨期最优决策, 都是以条件矩限制 (conditional moment restrictions) 的形式出现的。例如 NKPC 式(16.10)意味着以下条件矩限制需要得到满足

$$E_{t-1} \left[ \hat{\pi}_{t-1} - \beta \hat{\pi}_t - \kappa_p (\widehat{lsh}_{t-1} + \lambda_{t-1}) \right] = 0. \quad (16.58)$$

有条件的矩限制，可以用如下方法转换为一组无条件的矩限制所构成的向量。将全部过去时段的信息  $\{y_\tau, s_\tau, \epsilon_\tau\}_{\tau=-\infty}^t$  定义为一个 sigma 代数族  $\mathcal{F}_t$ ，进而定义一个  $\mathcal{F}_t$ —可测度的随机变量的向量  $\tilde{\mathcal{Z}}_t$ (即是说  $\tilde{\mathcal{Z}}_t$  的值可基于全部已有新息  $\{y_\tau, s_\tau, \epsilon_\tau\}_{\tau=-\infty}^t$  而确定)。那么对于每一个已经确定了的  $\tilde{\mathcal{Z}}_{t-1}$ ，都有以下等式成立

$$\begin{aligned} & E \left[ \tilde{\mathcal{Z}}_{t-1} \left( \hat{\pi}_{t-1} - \beta \hat{\pi}_t - \kappa_p (\hat{lsh}_{t-1} + \lambda_{t-1}) \right) \right] \\ &= E \left[ \tilde{\mathcal{Z}}_{t-1} E_{t-1} \left[ \left( \hat{\pi}_{t-1} - \beta \hat{\pi}_t - \kappa_p (\hat{lsh}_{t-1} + \lambda_{t-1}) \right) \right] \right] = 0, \end{aligned} \quad (16.59)$$

其中  $E_{t-1}[\cdot] \equiv E[\cdot | \mathcal{F}_{t-1}]$ 。需要指出的是，NKPC 的矩条件中包括一个隐含变量 (latent variable) 即当期价格加成冲击  $\lambda_{t-1}$ ，是作决策时并不掌握的信息。这使得若直接利用(16.59)来估计目标方程，可能会存在问题。

作为替代方案，可以考虑跨期消费的欧拉等式(16.5)，将  $t$  切换成  $t-1$

$$E_{t-1} [\hat{x}_t - \hat{x}_{t-1}] + E_{t-1} [z_t] + \hat{\pi}_{t-1} - \hat{R}_{t-1} = 0,$$

由(16.24)得到一组替换条件

$$\begin{aligned} E_{t-1} [\hat{x}_t - \hat{x}_{t-1}] + E_{t-1} [z_t] &= E_{t-1} \log \left( \frac{X_t}{X_{t-1}} \right) - \log \gamma, \\ \hat{\pi}_{t-1} &= \log \pi_{t-1} - \log \pi^*, \\ \hat{R}_{t-1} &= \log R_{t-1} - \log \left( \frac{\pi^* \gamma}{\beta} \right), \end{aligned}$$

代回上式，得到消费欧拉等式的条件矩

$$E_{t-1} \left[ \log \left( \frac{X_t}{X_{t-1}} \right) + \log \pi_{t-1} - \log R_{t-1} - \log \left( \frac{1}{\beta} \right) \right] = 0, \quad (16.60)$$

稳态通货膨胀率  $\pi^*$  和稳态技术进步率  $\gamma$  被消掉； $\beta$  参数的值在研究中被赋予；条件矩限制中只剩下了观测数据，而不再有隐含变量。

类似地，由(16.14)，将  $t$  切换为  $t-1$ ，有

$$E_{t-1} [\hat{R}_t - \psi \cdot \hat{\pi}_t - \sigma_R \epsilon_{R,t}] = 0,$$

根据  $\epsilon_{R,t}$  是个鞅差序列 MDS 的假定 (Definition 16.1)，我们有

$$E_{t-1} [\epsilon_{R,t}] = E [\epsilon_{R,t}] = 0,$$

利用(16.24)替换  $\hat{R}_t$  和  $\hat{\pi}_t$ ，那么上式进一步调整为泰勒法则的条件矩

$$E_{t-1} \left[ \log R_t - \frac{1}{\beta} \log \pi_t + \left( \frac{1-\beta}{\beta} \right) \log \pi^* - \log \left( \frac{\gamma}{\beta} \right) \right] = 0, \quad (16.61)$$

式中同样消除了隐含变量。

在此基础上，对于消费欧拉等式的条件矩(16.60)和泰勒法则的条件矩(16.61)，都可遵循(16.59)的思路，利用一个  $\mathcal{F}_{t-1}$ —可测的随机向量  $\tilde{\mathcal{Z}}_{t-1}$ ，转换为无条件矩限制。

对于第16.1节的典型 DSGE 模型，具有定义良好的特点，从而在转换成状态——空间系统之后，可以直接用解析法求出自协方差矩阵、光谱密度、冲击响应方程等。利用这些信息，在对参数向量  $\theta$  赋值后，可进一步做数值计算。

对于一些特殊线性结构的 DSGE 模型，如果采用某些特定类型的扰动法去求近似解，也存在一些对应的解析方法去求解析矩条件，但适用范围有限 (Andreasen et al., 2016)。对于更一般的非

线性 DSGE 模型，我们往往要用蒙特卡洛法 (Monte Carlo method) 作数值模拟来求系统的矩条件。例如，对一个 DSGE 模型的状态——空间表现式，首先基于解析模型设定中的分布信息，绘制初始状态向量  $s_0$  和新息矩阵  $\epsilon_t$ ，然后通过仿真生成一组模拟数据  $Y_{1:T}^* = \{\cdot\}$ ，满足

$$\frac{1}{T} \sum_{t=1}^T y_t^* \xrightarrow{s.s.} E[y_t], \quad (16.62)$$

其中 *a.s.* 表示几乎总是 (almost surely)<sup>3</sup>。

蒙特卡洛模拟法的适用性前提之一是，根据 DSGE 模型的设定，(模拟) 时间序列数据应当是严格平稳和严格遍历的。蒙特卡洛法的突出不足表现在，近似可能会有误差。

记敲到键盘  
之后，做一个  
ference：如何通  
对矩条件的仿  
近似来估计参  
向量  $\theta$

## 16.3 经验验证

结合前文提到的典型 DSGE 模型及其状态——空间表达形式，本节做经验验证。具体说来，截取美国 1984Q1-2007Q4 的季度经济数据作样本 (对应从大调整时期之后, the Great Moderation 到大衰退时间之前, the Great Recession period)，模拟全体矩。

数据下载自 FRED 数据库 (Federal Reserve Bank of St. Louis, <https://fred.stlouisfed.org/>)，结合模型实际需要，包括以下诸项 (表 16.1)：

- 实际总产出 (seasonally adjusted GDP at the annual rate, 2009 dollars). GDPC96. 对季度产出取 log 后相减。
- 劳动收入份额 (seasonally adjusted GDP at the annual rate)，计算方法为  $\log(\frac{COE}{GDP})$ ，COE 表示 compensation of employees.
- 通货膨胀率。implicit price deflator. GDPDEF. 对季度数据取 log 后相减。
- 回报率。Effective Federal Funds Rate. GDPDEF. 是月度数据并且未经 seasonally adjusted。将每个季度的三个月度数据取平均数，化为季度数据。

	$\log(X_t/X_{t-1})$	$\log(lsh_t)$	$\pi_t$	$R_t$
1984-01-01	0.0197	-0.7846	0.0106	0.0969
1984-04-01	0.0174	-0.7866	0.0084	0.1056
1984-07-01	0.0098	-0.7836	0.0081	0.1139
1984-10-01	0.0079	-0.7808	0.0067	0.0927
1985-01-01	0.0099	-0.7843	0.0114	0.0848
1985-04-01	0.0091	-0.7844	0.0062	0.0792
1985-07-01	0.0154	-0.7891	0.0058	0.0790
1985-10-01	0.0075	-0.7820	0.0058	0.0810
1986-01-01	0.0092	-0.7823	0.0049	0.0783
1986-04-01	0.0046	-0.7821	0.0040	0.0692
1986-07-01	0.0100	-0.7820	0.0040	0.0621
1986-10-01	0.0052	-0.7737	0.0056	0.0627
1987-01-01	0.0070	-0.7689	0.0072	0.0622
1987-04-01	0.0112	-0.7703	0.0067	0.0665

<sup>3</sup> 我们说一个随机变量数列  $X_T$  “几乎总是” 收敛到一个有界的随机变量  $X$ ，如果  $X_T$  不收敛至  $X$  的轨迹的概率为 0。

1987-07-01	0.0090	-0.7688	0.0072	0.0684
1987-10-01	0.0164	-0.7673	0.0082	0.0692
1988-01-01	0.0056	-0.7653	0.0078	0.0666
1988-04-01	0.0131	-0.7651	0.0096	0.0716
1988-07-01	0.0058	-0.7659	0.0117	0.0798
1988-10-01	0.0132	-0.7694	0.0080	0.0847
1989-01-01	0.0100	-0.7763	0.0109	0.0944
1989-04-01	0.0078	-0.7860	0.0103	0.0973
1989-07-01	0.0074	-0.7896	0.0072	0.0908
1989-10-01	0.0021	-0.7822	0.0069	0.0861
1990-01-01	0.0109	-0.7842	0.0110	0.0825
1990-04-01	0.0039	-0.7803	0.0103	0.0824
1990-07-01	0.0002	-0.7775	0.0089	0.0816
1990-10-01	-0.0086	-0.7747	0.0075	0.0774
1991-01-01	-0.0047	-0.7790	0.0099	0.0643
1991-04-01	0.0077	-0.7844	0.0068	0.0586
1991-07-01	0.0048	-0.7875	0.0073	0.0564
1992-01-01	0.0117	-0.7829	0.0043	0.0402
1992-10-01	0.0044	-0.7873	0.0054	0.0482
1992-04-01	0.0110	-0.7869	0.0064	0.0377
1992-07-01	0.0097	-0.7935	0.0047	0.0326
1992-10-01	0.0100	-0.7955	0.0067	0.0304
1993-01-01	0.0019	-0.8035	0.0057	0.0304
1993-07-01	0.0049	-0.8018	0.0060	0.0306
1993-04-01	0.0059	-0.8000	0.0061	0.0300
1993-10-01	0.0133	-0.8052	0.0052	0.0299
1994-01-01	0.0098	-0.8156	0.0048	0.0321
1994-04-01	0.0136	-0.8117	0.0050	0.0394
1994-07-01	0.0059	-0.8127	0.0054	0.0449
1994-10-01	0.0113	-0.8133	0.0055	0.0517
1995-01-01	0.0034	-0.8076	0.0057	0.0581
1995-04-01	0.0035	-0.8052	0.0044	0.0602
1995-07-01	0.0085	-0.8059	0.0047	0.0580
1995-10-01	0.0071	-0.8064	0.0049	0.0572
1996-01-01	0.0065	-0.8048	0.0054	0.0536
1996-04-01	0.0173	-0.8076	0.0038	0.0524
1996-07-01	0.0092	-0.8042	0.0028	0.0531
1996-10-01	0.0105	-0.8044	0.0051	0.0528
1997-01-01	0.0076	-0.7984	0.0062	0.0528
1997-04-01	0.0150	-0.8007	0.0027	0.0552
1997-07-01	0.0127	-0.7993	0.0036	0.0553
1997-10-01	0.0077	-0.7883	0.0033	0.0551
1998-01-01	0.0098	-0.7786	0.0016	0.0552

1998-04-01	0.0097	-0.7734	0.0021	0.0550
1998-07-01	0.0130	-0.7732	0.0037	0.0553
1998-10-01	0.0163	-0.7760	0.0031	0.0486
1999-01-01	0.0080	-0.7696	0.0050	0.0473
1999-04-01	0.0082	-0.7716	0.0033	0.0475
1999-07-01	0.0125	-0.7740	0.0036	0.0509
1999-10-01	0.0172	-0.7718	0.0046	0.0531
2000-01-01	0.0029	-0.7456	0.0076	0.0568
2000-04-01	0.0187	-0.7642	0.0057	0.0627
2000-07-01	0.0012	-0.7527	0.0065	0.0652
2000-10-01	0.0057	-0.7586	0.0054	0.0647
2001-01-01	-0.0028	-0.7447	0.0063	0.0559
2001-04-01	0.0053	-0.7609	0.0070	0.0433
2001-07-01	-0.0032	-0.7664	0.0033	0.0350
2001-10-01	0.0028	-0.7726	0.0030	0.0213
2002-01-01	0.0092	-0.7813	0.0032	0.0173
2002-04-01	0.0055	-0.7808	0.0037	0.0175
2002-07-01	0.0049	-0.7881	0.0044	0.0174
2002-10-01	0.0006	-0.7920	0.0054	0.0144
2003-01-01	0.0052	-0.8013	0.0061	0.0125
2003-04-01	0.0092	-0.7998	0.0032	0.0125
2003-07-01	0.0166	-0.8093	0.0055	0.0102
2003-10-01	0.0116	-0.8094	0.0047	0.0100
2004-01-01	0.0057	-0.8181	0.0087	0.0100
2004-04-01	0.0073	-0.8148	0.0087	0.0101
2004-07-01	0.0090	-0.8101	0.0061	0.0143
2004-10-01	0.0086	-0.8186	0.0070	0.0195
2005-01-01	0.0106	-0.8294	0.0092	0.0247
2005-04-01	0.0052	-0.8312	0.0072	0.0294
2005-07-01	0.0084	-0.8317	0.0093	0.0346
2005-10-01	0.0057	-0.8322	0.0076	0.0398
2006-01-01	0.0119	-0.8240	0.0078	0.0446
2006-04-01	0.0030	-0.8285	0.0080	0.0491
2006-07-01	0.0009	-0.8288	0.0070	0.0525
2006-10-01	0.0078	-0.8210	0.0035	0.0525
2007-01-01	0.0006	-0.8058	0.0112	0.0526
2007-04-01	0.0076	-0.8153	0.0056	0.0525
2007-07-01	0.0067	-0.8220	0.0035	0.0507
2007-10-01	0.0036	-0.8183	0.0043	0.0450

数据来源：FRED database. 整理后放在./data/20180403-empirics.xlsx 中。

表 16.1: 用于经验验证的观测数据集

### 16.3.1 自协方差

我们计算样本的协方差  $\widehat{\Gamma}_{yy}(h)$ , 作为全体自协方差  $\Gamma_{yy}(h)$  的近似

$$\widehat{\Gamma}_{yy}(h) = \frac{1}{T} \sum_{t=h}^T (y_t - \hat{\mu}_y)(y_{t+h} - \hat{\mu}_y)^\top, \quad \hat{\mu}_y = \frac{1}{T} \sum_{t=1}^T y_t. \quad (16.63)$$

在满足一定的正则条件的前提下, 根据强大数法则 (strong law of large numbers, SSLN, 第16.D.2节) 和中心极限定理 (central limit theorem, CLT, 第16.B.7节) 样本自协方差,  $\widehat{\Gamma}_{yy}(h)$  收敛至全体自协方差  $\Gamma_{yy}(h)$ , 这些正则条件包括

- 向量过程  $y_t$  的协方差是平稳的,
- $y_t$  的序列相关的衰减速度足够快,
- $y_t$  的 (至少一系列) 高阶矩是有界的, 等。

如果研究目标包括计算自协方差矩阵的数列<sup>4</sup>, 一个比较有效率的求解方法是, 首先建立一个辅助模型并作参数估计。第二步, 将这个辅助模型的估计参数转换为自协方差数列的参数值。辅助模型有许多种类, 一个较为合适的种类是线性向量自回归模型 (Vector Autoregressions, VARs), 一个简单的 VAR(1) 可表示如下

$$y_t = \phi_1 y_{t-1} + \phi_0 + u_t, \quad u_t \sim i.i.d. \mathbb{N}(0, \Sigma), \quad (16.64)$$

其中全体的 OLS 系数  $\phi_1$  向量, 和残差方差  $\Sigma$  无法直接计算求得, 可用样本的对应值  $\hat{\phi}_1, \hat{\Sigma}$  予以近似, 计算方法如下

$$\hat{\phi}_1 = \widehat{\Gamma}_{yy}(1) \widehat{\Gamma}_{yy}^{-1}(0) + \mathcal{O}_p(T^{-1}), \quad (16.65)$$

$$\hat{\Sigma} = \widehat{\Gamma}_{yy}(0) - \widehat{\Gamma}_{yy}(1) \widehat{\Gamma}_{yy}^{-1}(0) \widehat{\Gamma}_{yy}^\top(1) + \mathcal{O}_p(T^{-1}), \quad (16.66)$$

其中  $\mathcal{O}_p(T^{-1})$  用于反映 OLS 定义式的系数估计方法, 和样本自协方差求和估计式(16.63)之间的微小不一致——对于一组随机变量的数列  $X_t$ , 如果随着  $T \mapsto \infty$ ,  $TX_T$  随机有界, 我们将数列  $X_t$  表示为  $\mathcal{O}_p(T^{-1})$ 。

VAR(1) 过程的自协方差、跨协方差方程, 类似于(16.31), 分别表示为  $\widehat{\Gamma}_{yy}^V(h)$ ,  $\widehat{\Gamma}_{ss}^V(h)$ ,  $\widehat{\Gamma}_{ys}^V(h)$ , 并且  $\widehat{\Gamma}_{yy}(0)$  和  $\widehat{\Gamma}_{yy}(h)$  的计算方式, 类似于(16.33)(16.34); 将  $\hat{\phi}_1$  的近似估计(16.65)插入到估计过程中去, 有

$$\widehat{\Gamma}_{yy}^V(0) = \widehat{\Gamma}_{yy}(0) + \mathcal{O}_p(T^{-1}), \quad (16.67)$$

$$\widehat{\Gamma}_{yy}^V(h) = (\hat{\phi}_1)^h \widehat{\Gamma}_{yy}^V(0) + \mathcal{O}_p(T^{-1}) = [\widehat{\Gamma}_{yy}(1) \widehat{\Gamma}_{yy}^{-1}(0)]^h \widehat{\Gamma}_{yy}(0) + \mathcal{O}_p(T^{-1}). \quad (16.68)$$

来比较  $\widehat{\Gamma}_{yy}^V(h)$  和  $\widehat{\Gamma}_{yy}(h)$ :

$$\widehat{\Gamma}_{yy}^V(h) = \begin{cases} \widehat{\Gamma}_{yy}(h) + \mathcal{O}_p(T^{-1}) & h = 0, \\ \widehat{\Gamma}_{yy}(h) + \mathcal{O}_p(T^{-1}) & h = 1, \\ [\widehat{\Gamma}_{yy}(1) \widehat{\Gamma}_{yy}^{-1}(0)]^h \widehat{\Gamma}_{yy}(0) + \mathcal{O}_p(T^{-1}) & h > 1, \end{cases}$$

换句话说, 当  $h = 0, 1$  时两个自协方差的估计矩阵相等 (除了  $\mathcal{O}_p(T^{-1})$  的部分); 而当  $h > 1$  时二者不相等。[Schorfheide \(2005\)](#) 指出, 如果实际时间序列数据能够较好契合 VAR(1) 特征, 那么插入法的估计  $\widehat{\Gamma}_{yy}^V(h)$  要比 OLS 估计  $\widehat{\Gamma}_{yy}(h)$  更有效。

<sup>4</sup>Matlab 中可用 xcov 程序计算自协方差、跨协方差。

实际研究过程中，有时时间序列数据并不符合  $VAR(1)$  特征。就需要在  $VAR(p)$ ,  $p > 1$  框架下分析协自协方差

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \phi_0 + u_t, \quad u_t \sim i.i.d. \mathbb{N}(0, \Sigma) \quad (16.69)$$

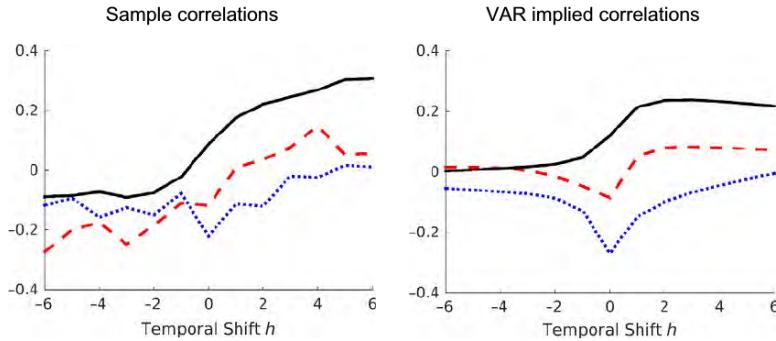
$p$  值的选取标准，可通过贝叶斯信息量 (Bayesian information criterion, BIC)方法予以估计。此外还有如赤池信息量 (Akaike information criterion, AIC)等方法。一个直观的处理方法是改写成友矩阵 (companion matrix) 形式，即将  $VAR(p)$  时间序列向量  $y_t$  理解为一个关于堆栈向量  $\tilde{y}_t = [y_t^\top, y_{t-1}^\top, \dots, y_{t-p+1}^\top]^\top$  的  $VAR(1)$  过程：

$$\begin{aligned} \tilde{y}_t &= \tilde{\phi}_1 \tilde{y}_{t-1} + \tilde{\phi}_0 + \tilde{u}_t, \quad \tilde{u}_t \sim i.i.d. \mathbb{N}(0, \tilde{\Sigma}), \\ \tilde{\phi}_1 &= \begin{pmatrix} \phi_1 & \dots & \phi_{p-1} & \phi_p \\ I_{n \times n} & \dots & 0_{n \times n} & 0_{n \times n} \\ \vdots & \ddots & \vdots & \vdots \\ 0_{n \times n} & \dots & I_{n \times n} & 0_{n \times n} \end{pmatrix}, \quad \tilde{\phi}_0 = \begin{pmatrix} \phi_0 \\ 0_{n(p-n) \times 1} \end{pmatrix}, \\ \tilde{\epsilon}_t &= \begin{pmatrix} \epsilon_t \\ 0_{n(p-n) \times 1} \end{pmatrix}, \quad \tilde{\Sigma} = \begin{pmatrix} \Sigma & 0_{n \times n(p-1)} \\ 0_{n(p-1) \times n} & 0_{n(p-1) \times n(p-1)} \end{pmatrix}, \end{aligned} \quad (16.70)$$

这样一来，估计的步骤如下：首先利用类似于(16.67)(16.68)的思路，将  $y_t$  替换为  $\tilde{y}_t$ ，算得  $\tilde{y}_t$  的自协方差  $\hat{\Gamma}_{\tilde{y}\tilde{y}}(0), \hat{\Gamma}_{\tilde{y}\tilde{y}}(h)$ 。随后根据  $y_t = M^\top \tilde{y}_t$ ，还原  $y_t$  的自协方差矩阵，其中  $M^\top$  是一个选择矩阵 (selection matrix)

$$M^\top = [I_n, 0_{n \times n(p-1)}].$$

图 16.1: 抽样观测数据的跨相关



图(16.1)画出了基于抽样观测数据 (表16.1) 的产出增速相对于利率 (黑色实线)，通货膨胀率 (红色段线)，劳动收入份额 (蓝色点线) 的跨相关 (cross correlation) 值<sup>5</sup>，横轴表示跨越的时间期  $h$ 。左图使用样本自协方差矩阵  $\hat{\Gamma}_{yy}(h)$ (16.63)测算。右图使用  $VAR(1)$  测算而得<sup>6</sup>，其中其中根据 BIC 检验确定  $p = 1$ ，根据(16.63)确定自协方差  $\hat{\Gamma}_{yy}(h)$ 。两组跨相关性质接近，但量化值不同，这是因为  $VAR$  模型更“吝啬”一些，导致  $VAR$  生成的相关系数更加平滑。

### 16.3.2 光谱分析

光谱分析的目标是近似计算光谱密度方程。

<sup>5</sup>Matlab 中用 `xcov` 计算，跨协方差就是跨相关方程。

<sup>6</sup>Matlab 中用 `varm` 来作  $VAR$  分析。

### 样本平滑周期图法

直观上来看，对于常见的  $VAR(1)$  数列，我们似乎很容易用样本的周期图 (periodogram) 来做光谱密度的近似估计，由(16.51)得

$$\begin{aligned}\hat{f}_{yy}(\omega) &= \frac{1}{2\pi} \sum_{h=-T+1}^{T-1} \widehat{\Gamma}(h) \\ &= \frac{1}{2\pi} \left\{ \widehat{\Gamma}_{yy}(0) + \sum_{h=1}^{T-1} [\widehat{\Gamma}_{yy}(h) + \widehat{\Gamma}_{yy}^\top(h)] \cos(\omega h) \right\},\end{aligned}\quad (16.71)$$

然而需要指出的是，尽管样本周期图对应的光谱密度  $\hat{f}_{yy}(\omega)$  是对全体密度  $f_{yy}(\omega)$  的渐进无偏估计，但它并不适合用作近似估计，因为它不一致<sup>7</sup>。

为了求得一致估计，需要沿着临近频率 (adjacent frequencies) 对样本周期图作平滑。基于基准频率 (base frequency)  $\omega = \frac{2\pi}{T}$ ，定义一组基本频率 (fundamental frequencies)  $\{\omega_j\}$ ，是基准频率的整数倍：

$$\omega_j = j \frac{2\pi}{T}, \quad j = 1, \dots, \frac{T-1}{2},$$

那么可以将平滑的周期图表示为  $\bar{f}_{yy}(\omega)$ ，满足如下关系

$$\bar{f}_{yy}(\omega) = \frac{\pi}{\lambda \frac{T-1}{2}} \sum_{j=1}^{\frac{T-1}{2}} \mathcal{K}\left(\frac{\omega_j - \omega}{\lambda}\right) \hat{f}_{yy}(\omega_j), \quad (16.72)$$

其中  $\mathcal{K}(\cdot)$  是一个核方程 (kernel)，满足  $\int \mathcal{K}(x) dx = 1$ 。核方程的具体形式有多种，一个简单的定义式可以表示为

$$\mathcal{K}\left(\frac{\omega_j - \omega}{\lambda}\right) \hat{f}_{yy}(\omega_j) := \mathbb{I}\left\{-\frac{1}{2} < \frac{\omega_j - \omega}{\lambda} < \frac{1}{2}\right\} = \mathbb{I}\{\omega_j \in \mathbb{B}(\omega|\lambda)\}, \quad (16.73)$$

其中  $\mathbb{I}\{\cdot\}$  是一个指示方程 (indicator function) 见(16.47)， $\mathbb{B}(\omega|\lambda) = \lambda T (2\pi)$  构成一个频率带 (frequency band)。关于核方程的简要介绍，可参考第16.B节。

如果有  $\lim_{T \rightarrow \infty} \lambda \rightarrow 0$ ，即随着  $T \rightarrow \infty$ ，频率的带宽缩减至 0，那么对应地， $\mathbb{B}(\omega|\lambda)$  中  $\omega_j$  的数量增大到无穷，平滑周期图算子因而是一致估计算子。

在经验研究中常常使用高斯核方程如  $\exp(i\omega t)$  作基方程 (第16.B.8节)，从而  $\mathcal{K}(x)$  的值等于一个标准正态分布的随机变量的密度分布方程。

### VAR(p) 的插入估值法

若 BIC 检验结果表明观测数据更适合用  $VAR(p)$ ,  $p > 1$  模型如(16.69)，那么可利用上文提到的插入法来近似估计光谱密度。具体说来，定义全体的系数向量  $\phi := [\phi_0, \phi_1, \dots, \phi_p]^\top$ ，选择矩阵  $M(z) = [Iz, \dots, Iz^p]$ 。设样本的近似系数向量  $\hat{\phi}$  是对  $\phi$  的一个估计。

对应于  $VAR(1)$  下的(16.52)， $VAR(p)$  用插入法估计的光谱分布密度  $\hat{f}_{yy}^V(\omega)$  可表示为

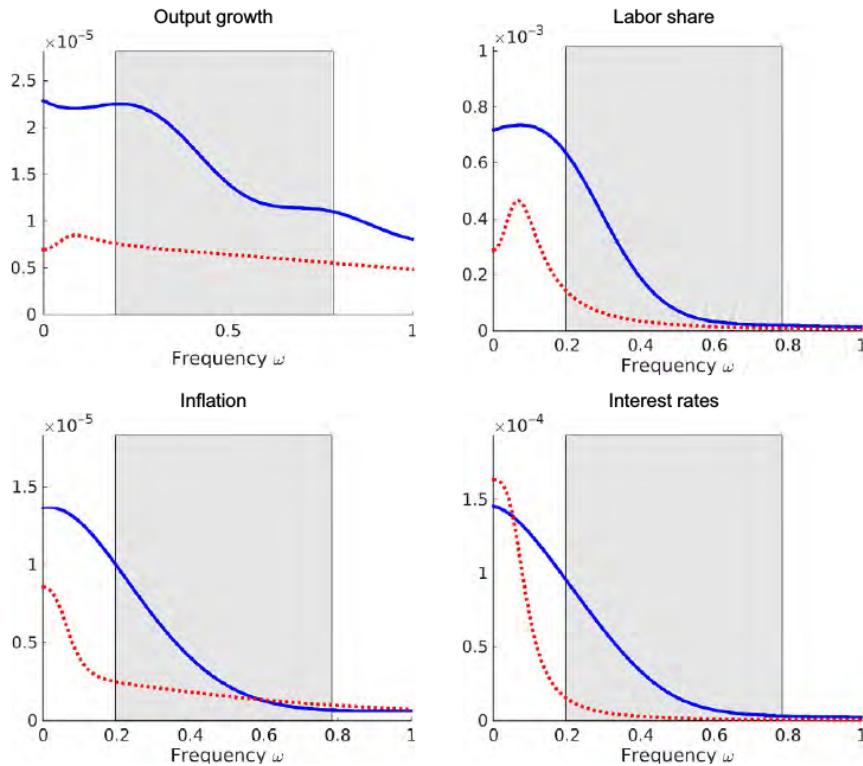
$$\hat{f}_{yy}^V(\omega) = \frac{1}{2\pi} \left[ I - \hat{\phi}^\top M^\top (\exp(-i\omega)) \right]^{-1} \hat{\Sigma} \left[ I - M(\exp(-i\omega)) \hat{\phi} \right]^{-1}, \quad (16.74)$$

(16.74)是对(16.52)的一般化处理，将  $VAR(1)$  扩展到了  $VAR(p)$  的情况。

图16.2依次绘出产出增速、劳动力收入份额、通货膨胀率和利率的光谱密度。阴影的矩形对应经济周期的频率  $[0.196, 0.785]$ 。红色点线表示根据  $VAR(1)$  (16.52)近似估计算得的密度分布，蓝色实线是根据高斯核方程  $G(\cdot; \sigma = 0.15)$  (16.72)算得的平滑周期图。

<sup>7</sup> 一致性 (consistency) 是指随着  $T \rightarrow \infty$ ，估计值的偏差 (biase) 和方差都收敛至 0，可参考 (Koopmans, 1995, Sec. 8.2)。

图 16.2: 抽样数据的光谱分布密度



1. 红色点线的光谱密度分布:

- (a) 左上图: 由于产出增速的自相关接近于 0, 使得光谱密度相对平缓。其余 3 张图中, 三个观测数列在低频  $\omega$  处的光谱密度较高, 这反映了实值域中三个时间序列具有较高的持久性。
- (b) 劳动收入份额的光谱密度表现出明显的驼峰形, 通胀和利率的光谱密度随着频率  $\omega$  的增加而单调递减。

2. 蓝色实线的平滑周期图分布:

- (a) 对  $\bar{f}_{yy}(\omega)$  的估计值取决于带宽值的设定。我们在高斯方程中设  $\sigma = 0.15$ , 大致说来在傅里叶域中, 这对应于将样本周期图以约 0.6 的频带值做平均。

### 16.3.3 冲击响应方程

VAR(p) (16.69)有时也被称作缩减式 VAR(reduced-form VAR), 这是由于新息向量  $u_t \sim i.i.d. N(0, \Sigma)$  并没有明确的结构性含义, 它仅仅是前一时期的预测误差。

根据 DSGE 模型的设定, 所谓冲击响应机制, 是指某些新息冲击以结构性的方式作用于经济系统, 使得经济系统做出响应, 表现为同期系统模拟数据和现实观测数据之间的误差, 即预测误差。可见, 为了将基于 VAR 的冲击——响应(观测数据)和 DSGE 模型的冲击——响应(模拟数据)联系起来, 就需要将预测误差引入到一个结构新息向量  $\epsilon_t$  中, 设满足关系

$$u_t = \phi_\epsilon \epsilon_t = \Sigma_{tr} \Omega \epsilon_t, \quad (16.75)$$

其中

- 对预测误差  $u_t$  的协方差矩阵  $\Sigma$  作 Cholesky 因式分解 (Cholesky factorization, 第16.C.5节), 所生成的唯一的 (对角元素非负) 下三角矩阵, 定义为  $\Sigma_{tr}$ ,
- $\Omega$  是一个  $n \times n$  正交矩阵 (orthogonal matrix, 第16.C.2节), 满足

$$\Omega \Omega^\top = I.$$

引入  $\Omega$  的正交矩阵设定, 是为了确保在状态——空间表现形式中, 新息系数矩阵  $\phi_\epsilon$  可以写为如下关于  $\Omega, \Sigma$  的形式

$$\phi_\epsilon \phi_\epsilon^\top = \Sigma_{tr} \Omega \Omega^\top \Sigma^\top = \Sigma. \quad (16.76)$$

这样的向量自回归模型, 我们称为结构向量自回归模型 SVARs (structural vector autoregressions)。根据这样的架构设定, 预测误差  $u_t$  的协方差矩阵  $\Sigma$ , 与正交矩阵  $\Omega$  的具体选取无关。这意味着我们无法通过实际数据求得  $\Omega$ 。大多数基于 SVARs 的研究文献仅仅对  $\Omega$  的设定做出必要的限定 (restrictions), 或称之为识别 (identification)。随着冲击类型的不同 (如技术进步, 货币政策, 公共支出等), 附着在冲击上的新息也具有不同的性质, 在构建模型时对应的限定/识别策略也有所区别, 更详尽的综述可参考如Cochrane (1994); Christiano, Lawrence J et al. (1999); Stock and Watson (2001) 等。

在 (利用缩减式 VAR) 算得样本系数矩阵  $\hat{\phi}$  和  $\hat{\Sigma}$  的近似估计, 以及明确对  $\Omega$  中一列或多列的识别机制之后, 我们可以将样本近似下的冲击响应方程表示如下

$$\widehat{IRF}^V(\cdot, j, h) = C_h(\hat{\phi}) \hat{\Sigma}_{tr} [\Omega]_{\cdot j}, \quad (16.77)$$

其中

- $C_h(\hat{\phi})$  是移动平均系数矩阵 (moving average coefficient matrix), 可以通过 VAR 表现式(16.70)而算得

$$C_h(\hat{\phi}) = M^\top \tilde{\phi}_1^h M, \quad M^\top \equiv [I_n, 0_{n \times n(p-1)}]. \quad (16.78)$$

- $[\mathcal{A}]_{\cdot j}$  表示矩阵  $\mathcal{A}$  的第  $j$  列。

以下对 IRFs 的冲击——响应机制做一个响应区间的分析, 即我们关注的是响应的方向 (符号为正或为负), 而非某一特定的数值, 又称符号限制条件 (sign restrictions)。这方面的研究可参考如Faust (1998); Canova and Nicoló (2002); Uhlig (2005) 等, 它们将矩阵  $\Omega$  限定在一个有限的集合区间  $\mathcal{O}(\phi, \sigma)$  内, 对应的 IRFs 需要满足该限定条件: 只有在集合区间内的 IRFs 才被识别。

通过使用对数形式的产出增速、劳动收入份额, 通货膨胀率和利率的 VAR(1), 我们可以将这个限定条件设为: 面对一个紧缩型货币政策的点冲击, 模型状态变量的响应表现为, 在随后 4 个季度中通货膨胀率降低, 利率上升。此外还做了一个不影响通用性的假定: 将货币政策的冲击设为向量  $q$ (带宽), 排在  $\Omega$  的第一列, 用于描述冲击的效果。利用缩减式 VAR 估计得到的  $\hat{\phi}, \hat{\Sigma}$ , 我们列出满足符号限定条件的 IRFs 集合, 以确定单位长度向量  $q$  的集合, 作为带宽。

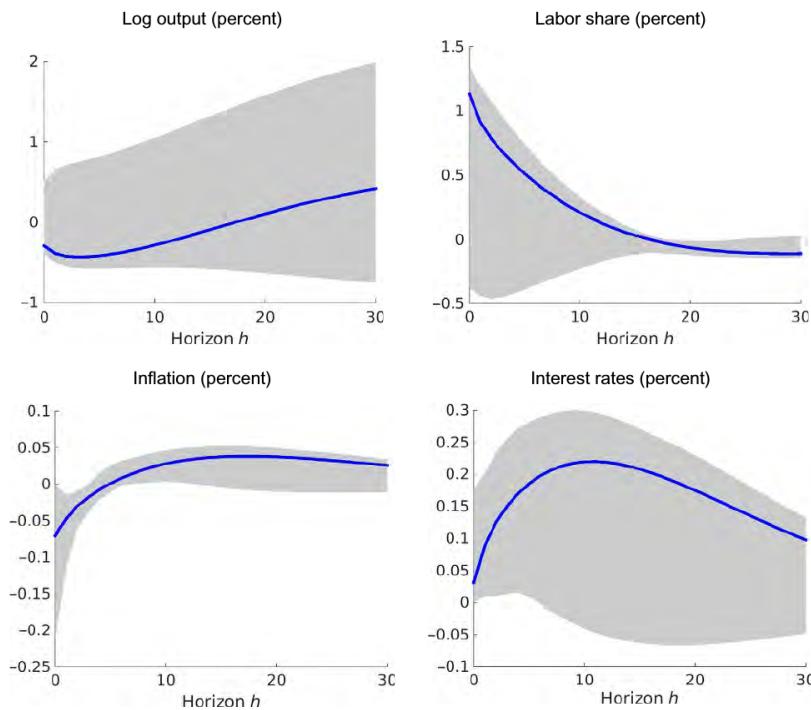
如图16.3所示,  $t = 0$  时刻货币政策的点冲击, 在随后  $t + h$  时刻产出增速、劳动收入份额、通货膨胀率、利率的已识别响应的区间用阴影表示。需要指出的是, 4 个状态变量的响应区间不同, 尤其是产出增速和劳动收入份额的响应符号 (方向) 不够清晰, 这是由于模拟过程中注入到系统里的货币政策冲击的符号限定不够强。

#### 16.3.4 条件矩的限制

根据 DSGE 模型的均衡条件, 可以推得一组条件矩限制 (第16.2.4节), 这组条件矩限制是基于全体的理想状况给出的; 基于实际观测到的抽样样本的平均数, 可以对全体矩限制制作样本矩限制的

图 16.3: 冲击响应方程

紧缩性货币政策冲击表现为 1 单位的标准差。通货膨胀和利率的响应未作年化处理。带宽 (阴影) 表示 IRFs 逐点估计的识别集合区间。此外在选取适当的  $q$  (例如设  $q$  为一个关于缩减式参数  $\hat{\phi}, \hat{\Sigma}$  的方程) 的情况下, 4 个状态变量的 IRFs 收缩为一条蓝色曲线。



近似。但需要指出的是, 这组近似矩并不是完全无条件的, 例如 NKPC 得到的矩(16.59)中含有隐含变量——当期价格加成的冲击  $\lambda_t$ , 从而使得进一步从事矩分析变得困难。

对全体矩限制的样本近似常常采用广义矩估计法 (generalized method of moments, GMM), 我们将在随后进一步介绍。

进键盘之后,  
一个 ref

## 16.4 处理趋势

趋势 (trend) 是宏观经济时间序列数据的一个重要特征。在第16.1节典型 DSGE 模型的设定中, 就含有这样一种随机趋势, 它由生产率过程  $\log Z_t$  产生, 表现为一个含有漂移的随机游走。尽管生产率中的趋势导致模型中的其他变量如消费、产出、实际工资等也含有趋势, 但根据这样的模型设定, 消费产出比、劳动收入份额等都应当是平稳的。

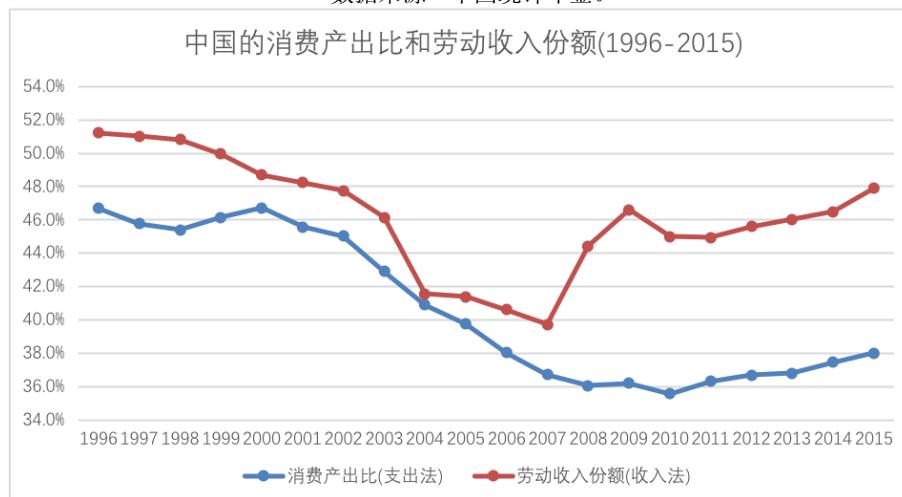
然而实际观察到的经济数据与模型的假设不符, 如图 16.1 所示, 1996-2015 年间中国的这两组数据并不平稳。实际观测数据中含有的趋势并未反映在 DSGE 模型中, 导致观测数据和模拟数据之间产生了一阶不一致。

大多数 DSGE 模型在实际应用过程中都暗含着反事实趋势, 这是由于模型中往往包含一些共趋势 (co-trending) 的限制条件 (Hatanaka and Yamada, 2012), 例如一条最的平衡增长路径 (balanced growth path, BGP), 根据这样的路径设定, 模型假定

- 产出、消费、投资、资本存量、实际工资等经济变量表现出相同的运动趋势,
- 资本回报率平稳。

图 16.1: 中国的劳动收入份额和消费产出比

数据来源: 中国统计年鉴。



这些假设条件，或多或少都与实际观测到的数据不匹配。

经济学家已提出许多种方法，致力于缩小两组数据之间的不匹配，如

1. 对观测到的每一个变量时间序列分别作去适宜的去趋势处理，进而比较（去趋势后的）观测数据与模拟数据；
2. 用同一种趋势滤波对观测数据和模拟数据作处理，进而比较；
3. 构建混合模型，如Canova et al. (2014) 模型一方面含有灵活的非结构性的趋势成分，另一方面利用结构 DSGE 模型取描述围绕缩减式趋势的上下波动；
4. 将更多现实中的趋势直接引入到 DSGE 模型的结构中。

从宏观经济学研究的角度上来说，这 4 种方法中，越是接近方法 4，就越值得提倡。



# 附录

## 16.A 欧拉公式 (复分析)

### 16.A.1 欧拉公式和欧拉恒等式

复分析意义上的欧拉公式 (Euler formula) 可以表示为

$$\exp[i\theta] = \cos \theta + i \sin \theta, \quad (16.79)$$

其中  $i$  是虚数 (imaginary number), 满足  $i = \sqrt{-1}$ 。欧拉公式的证明, 见第16.A.2节。

欧拉公式的性质, 可以从如下引理开始介绍。

**Lemma 16.1** ( $i$  的幂是周期方程). 现在来看  $f(x) = i^x, x = 0, 1, \dots$  的值

$$\begin{pmatrix} i^0 = 1, & i^4 = 1, & i^8 = 1, \\ i^1 = i, & i^5 = i, & i^9 = i, \\ i^2 = -1, & i^6 = -1, & i^{10} = -1, \\ i^3 = -i, & i^7 = -i, & i^{11} = -i, \\ \vdots & \vdots & \vdots \end{pmatrix}$$

不难看出,  $i^x, x = 0, 1, \dots$  每隔 4 个数重复一次, 因此  $f(x)$  是一个周期方程。

规律: 对于任意整数  $x$  求  $f(x)$ , 可以将  $x$  除以 4, 看余数:

$$f(x) = i^n = \begin{cases} 1 & \frac{n}{4} \text{ 余数是 } 0 \\ i & \frac{n}{4} \text{ 余数是 } 1 \\ -1 & \frac{n}{4} \text{ 余数是 } 2 \\ -i & \frac{n}{4} \text{ 余数是 } 3, \end{cases}$$

或者表示为

$$f(x) = i^x = i^{x \bmod 4}, \forall x \in \mathbb{Z}. \quad (16.80)$$

其中  $\mathbb{Z}_+$  表示全部正整数集合。 $x \bmod 4$  表示  $x$  除以 4 之后的余数。

**Lemma 16.2** (正弦和余弦的导数). 现在设  $f(x)$  是一个正弦函数, 满足  $f(x) = \sin x$ 。那么它的  $n$  次导数, 如图 16.A.1 所示。

不难看出, 三角函数如正弦的  $n$  次导数, 也是一个以 4 个数为一个周期的周期方程, 可以表

图 16.A.1: 正弦函数的导数

$$\begin{array}{c}
 \frac{d}{dx} \sin x = \cos x \\
 \downarrow \\
 \frac{d}{dx} \cos x = -\sin x \\
 \downarrow \\
 \frac{d}{dx} -\sin x = -\cos x \\
 \downarrow \\
 \frac{d}{dx} -\cos x = \sin x
 \end{array}$$

示为

$$f^{(n)}(x) = \frac{d^n \sin x}{dx^n} = \frac{d^{n \bmod 4} \sin x}{dx} = \begin{cases} \sin x & n \bmod 4 = 0, \\ \cos x & n \bmod 4 = 1, \\ -\sin x & n \bmod 4 = 2, \\ -\cos x & n \bmod 4 = 3, \end{cases} \quad \forall n \in \mathbb{Z}_+. \quad (16.81)$$

余弦函数具有类似的性质。此外，对于任意  $n \in \mathbb{Z}_+$ ，正弦或者余弦函数的  $n$  阶导数都存在，我们称这种方程为无限可导方程 (indefinitely differentiable function)。

**Lemma 16.3** (指数方程的导数). 指数方程的导数是指数方程本身，

$$\frac{d^n}{dx^n} \exp [x] = \exp [x], \forall x, n.$$

**Lemma 16.4** (泰勒——麦克劳林级数). 设  $f(x)$  是一个连续可导方程。对  $f(x)$  沿着  $x = a$  做  $n$  阶的泰勒级数展开，可以得到一个多项式相加的形式，我们称之为泰勒级数 (Taylor series):

$$f(x) = \frac{f(a)}{0!} + \frac{f'(a)}{1!} (x-a) + \frac{f''(a)}{2!} (x-a)^2 + \dots + \frac{f^{(n)}(a)}{n!} (x-a)^n. \quad (16.82)$$

不难看出， $n$  的值越大，近似的精度越高，如图(16.A.2)所示。当  $n \rightarrow \infty$  时，近似是完全精确的。

当  $a = 0$  时，泰勒级数变为沿着  $x = a = 0$  的点作  $n$  阶泰勒展开，所生成的级数称麦克劳林级数 (Maclaurin series):

$$f(x) = \frac{f(0)}{0!} + \frac{f'(0)}{1!} (x) + \frac{f''(0)}{2!} (x)^2 + \dots + \frac{f^{(n)}(0)}{n!} (x)^n. \quad (16.83)$$

**Lemma 16.5** (指数方程的麦克劳林级数 (幂级数)). 现在设  $f(x) = \exp x$ 。对应的麦克劳林级数展开由(16.83)变为

$$\begin{aligned}
 f(x) = \exp x &= \frac{f(0)}{0!} + \frac{f'(0)}{1!} (x) + \frac{f''(0)}{2!} (x)^2 + \dots + \frac{f^{(n)}(0)}{n!} (x)^n \\
 &= 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots \\
 &= \sum_{n=0}^{\infty} \frac{x^n}{n!}.
 \end{aligned} \quad (16.84)$$

又称幂级数 (power series)。

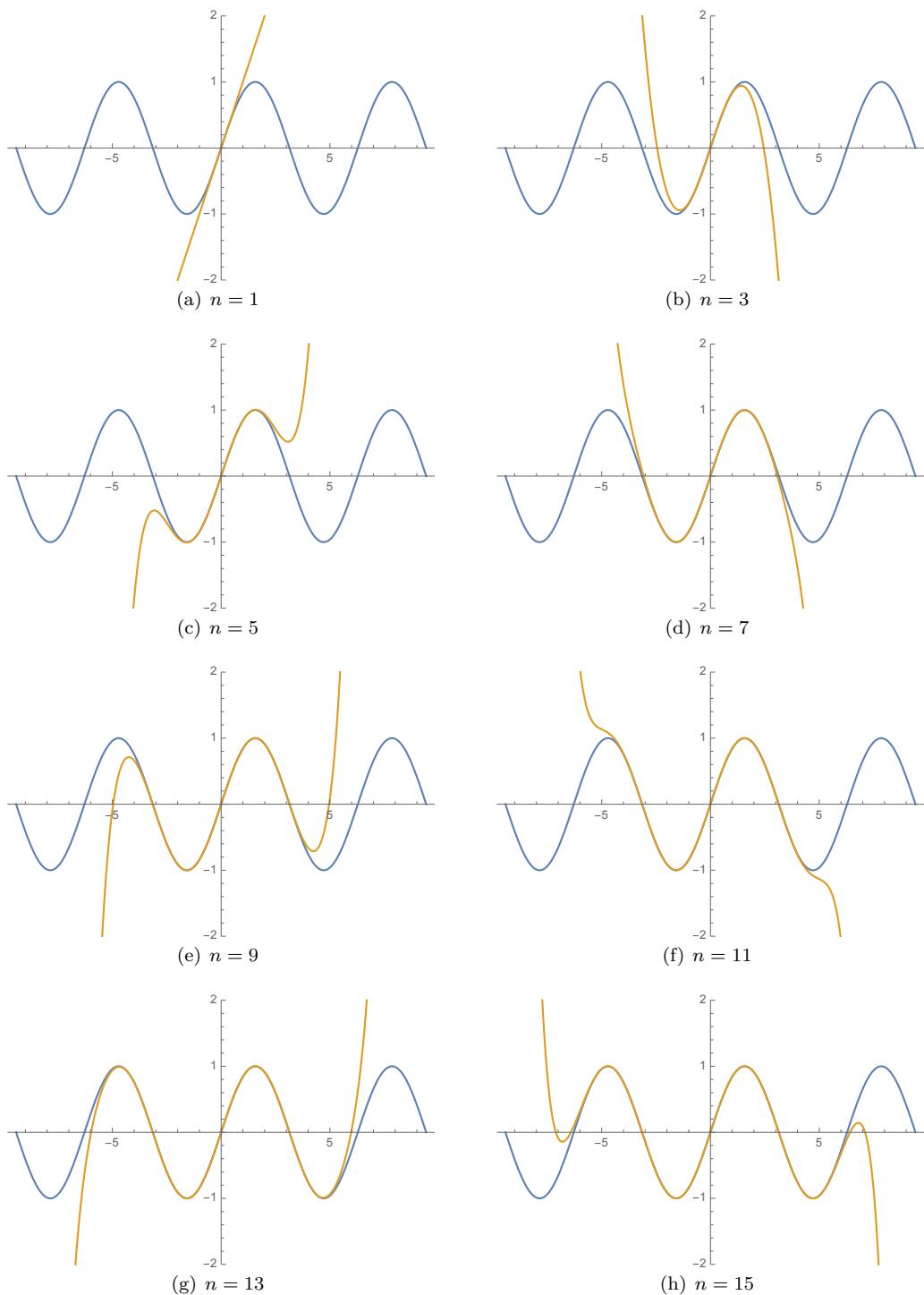


图 16.A.2: 麦克劳林级数的示例: 对  $\sin(x)$ (蓝色曲线) 沿着  $x_0 = 0$  作  $n$  阶泰勒级数展开(红色曲线)。

**Lemma 16.6** (正弦的麦克劳林级数). 现在设  $f(x) = \sin x$ , 对应的麦克劳林级数展开由(16.83)变为

$$\begin{aligned} f(x) = \sin x &= \frac{f(0)}{0!} + \frac{f'(0)}{1!}(x) + \frac{f''(0)}{2!}(x)^2 + \dots + \frac{f^{(n)}(0)}{n!}(x)^n \\ &= \sin 0 + \frac{\cos 0}{1!}x + \frac{-\sin 0}{2!}x^2 + \frac{-\cos 0}{3!}x^3 + \dots \\ &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} - \dots \\ &= \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}. \end{aligned} \quad (16.85)$$

**Lemma 16.7** (余弦的麦克劳林级数). 类似地, 若设  $f(x) = \cos(x)$ , 有余弦的麦克劳林级数

$$f(x) = \cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \dots = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}. \quad (16.86)$$

**Lemma 16.8** (欧拉恒等式). 根据指数方程的麦克劳林级数 (幂级数) 引理 (Lemma 16.5), 将式(16.84)中全部  $x$  都替换为  $ix$ , 则有

$$\begin{aligned} \exp[ix] &= \sum_{n=0}^{\infty} \frac{(ix)^n}{n!} = 1 + \frac{(i)x}{1!} + \frac{(-1)x^2}{2!} + \frac{(-i)x^3}{3!} + \frac{x^4}{4!} + \dots \\ &= 1 + i\frac{x}{1!} - \frac{x^2}{2!} - i\frac{x^3}{3!} + \frac{x^4}{4!} + i\frac{x^5}{5!} - \frac{x^6}{6!} - i\frac{x^7}{7!} + \dots \\ &= \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots\right) + i\left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots\right) \\ &= \cos x + i \sin x. \end{aligned} \quad (16.87)$$

又称欧拉恒等式 (*Euler identity*)。

## 16.A.2 欧拉公式的作用

欧拉公式、欧拉恒等式有什么用途?

### 复数的坐标形式

一个复数 (complex number) 可以表示为一个实部和一个虚部相加的形式, 例如对于复数  $z$  我们有

$$z = \Re(z) + i\Im(z) = x + iy, \quad x, y \in \mathbb{R}, i = \sqrt{-1}.$$

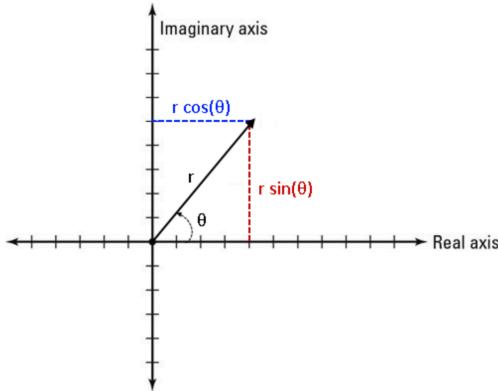
以实部为横轴, 以虚部为纵轴, 可以将  $z$  表示为复平面 (complex plane) 中的  $(x, y)$  点。我们常用模 (magnitude) 和角 (angle), 对应  $\{r, \theta\}$ , 来在复平面中表示  $z$ , 如图(16.A.3)所示。

由三角函数关系可得  $x = r \cos \theta, y = r \sin \theta$ , 那么结合欧拉恒等式 (Lemma 16.8) 可得

$$\begin{aligned} z &= x + iy = r \cos \theta + ir \sin \theta = r(\cos \theta + i \sin \theta) = r \exp(i\theta), \\ &\hookrightarrow x + iy = r \exp(i\theta), \quad x = r \cos \theta, y = r \sin \theta, \end{aligned} \quad (16.88)$$

就是欧拉公式, 将一个复数  $z$  写为关于夹角  $\theta$  和模  $r$  的三角方程

图 16.A.3: 正弦函数的导数



### 复数的乘除

复数的乘除法如

$$(a + bi)(c + di) = ac + adi + bci - bd = (ac - bd) + (ad + bc)i.$$

另一种算法时，假定要将两个指数方程相乘，满足

$$a \exp(i\theta) \cdot b \exp(i\theta\phi) = ab \exp[i(\theta + \phi)],$$

即是说，两个复数相乘，就是分别将二者的模和夹角相乘。上式可以进一步写为

$$a \exp(i\theta) \cdot b \exp(i\theta\phi) = ab [\cos(\theta + \phi) + i \sin(\theta + \phi)].$$

除法的操作类似。

### 对复数的取幂

假设我们要计算  $(a + bi)^n$  的值。常见的简化方法之一是借助复数的形式

$$(a + bi)^n = [r \exp(i\theta)]^n = r^n \exp(i\theta n) = r^n [\cos(\theta n) + i \sin(\theta n)], \quad \forall n \in \mathbb{Z}_+, \quad (16.89)$$

将一个复数的求幂运算 (exponentiation) 转换为幂指数的形式，又称棣莫弗定理 (De Moivre Theorem)。

前面介绍了如何将一个复数表示为幂指数形式，现在来看如何将一个实数表示为一个复数。已知欧拉公式 (16.79)

$$\exp(ix) = \cos x + i \sin x,$$

将其中的  $x$  替换为  $x \ln b$

$$\exp(ix \ln b) = \cos(x \ln b) + i \sin(x \ln b) = b^{ix},$$

那么，若对某个复数  $y + ix$  作幂，求  $b^{y+ix}$  的值，可计算如下

$$b^{y+ix} = b^y b^{ix} = b^y [\cos(x \ln b) + i \sin(x \ln b)]. \quad (16.90)$$

举例：

$$5^{3+2i} = 5^3 5^{2i} = 5^3 [\cos(2 \ln 5) + i \sin(2 \ln 5)] \approx -124.63 - 9.65i.$$

### 复值夹角

根据欧拉公式，可以进一步引出复值夹角 (complex angle) 和三角方程的概念。先来看复值夹角。具体说来，前面我们所涉及到的正弦、余弦方程如  $\sin(x)$ ，是对实数  $x$  的计算。如果要对一个复数求正弦余弦，该怎么操作？

对欧拉恒等式(16.87)作调整，用  $-x$  替代  $x$ ，有

$$\begin{aligned}\exp(ix) &= \cos x + i \sin(x), \text{ 原方程} \\ \exp(-ix) &= \cos x - i \sin(x) \text{ 复共轭},\end{aligned}\quad (16.91)$$

将一个复数的虚部变更符号（例如由正变为负），所组成的新的复数，称为原复数的复共轭 (complex conjugate)。

现在对欧拉恒等式(16.87)及其复共轭(16.91)相加和相减，得

$$\exp(ix) - \exp(-ix) = 2i \sin x, \Leftrightarrow \sin x = \frac{\exp(ix) - \exp(-ix)}{2i}, \quad (16.92)$$

$$\exp(ix) + \exp(-ix) = 2 \cos x, \Leftrightarrow \cos x = \frac{\exp(ix) + \exp(-ix)}{2}, \quad (16.93)$$

即，我们得到了一组用指数（和虚数）表示的正弦、余弦方程。

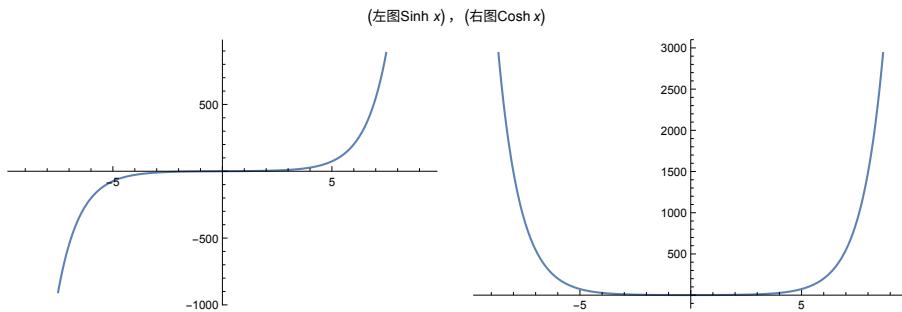
调整(16.92)-(16.93)，用  $ix$  替换原式中的  $x$ ，并且分子分母同时乘以  $i$  得

$$\sin(ix) = \frac{\exp(i^2 x) - \exp(-i^2 x)}{2i^2} \cdot \frac{i}{i} = \frac{i[\exp(x) - \exp(-x)]}{2} \equiv i \sinh x, \quad (16.94)$$

$$\cos(ix) = \frac{\exp(i^2 x) + \exp(-i^2 x)}{2} \cdot \frac{i}{i} = \frac{[\exp(x) + \exp(-x)]}{2} \equiv \cosh x, \quad (16.95)$$

其中  $\sinh(x), \cosh(x)$  分别称双曲线正弦方程和双曲线余弦方程 (hyperbolic sine/cosine function)，如图 16.A.4。值得注意的是右图：任何虚数的余弦都是一个实数，原本的虚部消失不见。

图 16.A.4: 双曲正弦/余弦方程,  $x \in [-3\pi, 3\pi]$



在此基础上，我们可以回答本节初提出的问题，如何计算一个复数的正弦/余弦，如  $\sin(a+bi)$ 。已知  $\sin(a+b) = \sin a \cos b + \cos a \sin b$ ，并且  $\sinh x, \cosh x$  的定义如(16.94)-(16.95)所示。那么，用  $bi$  替换原式中的  $b$ ，可得复数的正弦计算式

$$\begin{aligned}\sin(a+bi) &= \sin a \cos(ib) + \cos a \sin(ib) \\ &= \sin a \cosh b + i \cos a \sinh b.\end{aligned}\quad (16.96)$$

余弦:  $\cos(a + b) = \cos a \cos b - \sin a \sin b$ 。用  $bi$  替换  $b$

$$\begin{aligned}\cos(a + bi) &= \cos a \cos(ib) - \sin a \sin(ib) \\ &= \cos a \cosh b - i \sin a \sinh b.\end{aligned}\tag{16.97}$$

举例:

$$\cos(3 + 4i) = \cos(3) \cosh(4) - i \sin(3) \sinh(4) \approx -27.03 + 3.85i.$$

## 16.B 高斯核方程

### 16.B.1 高斯核

一个常见的高斯核方程 (Gaussian kernel function), 随着维度的不同, 可以表示如下

$$G^{(1)}(x; \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right),\tag{16.98}$$

$$G^{(2)}(x, y; \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right),\tag{16.99}$$

⋮

$$G^{(N)}(\vec{x}; \sigma) = \frac{1}{(\sqrt{2\pi}\sigma)^N} \exp\left(-\frac{|\vec{x}|^2}{2\sigma^2}\right),\tag{16.100}$$

其中

- $\sigma > 0$  表示高斯核的带宽 (bandwidth), 统计学中也将上式称为高斯概率密度方程 (Gaussian probability density function)——对应地, 称  $\sigma$  为标准差,  $\sigma^2$  为方差。
- $\{x\}, \{x, y\}$  分别表示 1 维和 2 维空间参数。 $N$  维空间参数往往用向量  $\vec{x}$  进行简化表示。
- 空间参数和带宽参数之间的分号是常见的表达方式, 提醒人们两者的含义不同。指数项中乘以  $1/2$  也是为了表述方便。

### 16.B.2 标准化

以 1 维高斯核方程(16.98)为例, 前面的常数项  $\frac{1}{\sqrt{2\pi}\sigma}$  是标准化常数, 引入这个常数项是由于对  $\exp\left(-\frac{x^2}{2\sigma^2}\right)$  沿着  $x \in (-\infty, \infty)$  求积, 其值不等于 1

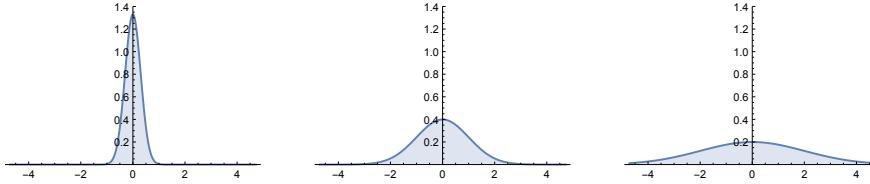
$$\int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx = \sqrt{2\pi}\sigma \neq 1.$$

在作标准化处理后, 高斯核成为一个标准化的核方程, 即满足

$$\int_{-\infty}^{\infty} G^{(1)}(x; \sigma) dx = 1, \forall \sigma,$$

另一方面, 增加带宽  $\sigma$  的值会显著降低  $G^{(1)}(x; \sigma)$  的振幅 (amplitude), 例如  $\sigma$  取 0.3, 1, 2 三组不同的值, 对应高斯方程(16.98)在  $[-1.5\pi, 1.5\pi]$  的形状, 见图16.B.1, 图中取下下方阴影部分的面积均为 1。

图 16.B.1: 高斯核方程的振幅随带宽的增加而显著降低



### 16.B.3 瀑布特征

将两个高斯核作卷积 (convolution)，可以得到一个新的高斯核方程，新方程具有与原方程相似的特性：

$$G_1(\vec{x}_1; \sigma_1) G_2(\vec{x}_2; \sigma_2) = G_{new}(\vec{x}; \sqrt{\sigma_1^2 + \sigma_2^2}), \quad (16.101)$$

我们称这种特性为自相似性 (self-similarity)，高斯核方程就是这样一种自相似方程。

高斯方程的卷积是线性运算，那么高斯方程卷积的卷积也是线性运算，以此类推，这称为瀑布特性 (cascade property)。

### 16.B.4 带宽参数

由于  $\sigma_1^2 + \sigma_2^2$  较难直接计算，我们常常用  $t \equiv 2\sigma^2$  的方式来做一个近似参数化处理。对于  $N$  维高斯核方程而言，有

$$G^{(N)}(\vec{x}; t) = \frac{1}{(\pi t)^{\frac{N}{2}}} \exp\left(-\frac{x^2}{t}\right).$$

如果  $N = 3$  即向量  $\vec{x}$  中包括三组数据  $\{x_1, x_2, x_3\}$ ，那么

$$\frac{\partial \mathcal{L}}{\partial t} = \sum_{i=1}^3 \frac{\partial^2 L}{\partial x_i^2},$$

即方差。

为了更好了解高斯核方程的自相似特征，我们引入一个无量纲 (dimensionless) 参数空间  $\tilde{x} = \frac{x}{\sqrt{2}\sigma}$  来对  $x$  轴做“参数化”处理，对应的高斯核方程为

$$G^{(N)}(\tilde{x}; \sigma) = \frac{1}{(\sigma\sqrt{2\pi})^N} \exp(-\tilde{x}^2),$$

$$G^{(N)}(\tilde{x}; t) = \frac{1}{(\pi t)^{\frac{N}{2}}} \exp(-\tilde{x}^2),$$

两式等价。换句话说，沿着空间轴 ( $x$  轴) 以  $\sigma$  的倍数为步长单位行走，其中所有的核方程都有相同的大小，又称带宽。

需要注意的是，带宽相同的高斯核，其振幅未必相同——这是由标准化计算流程所决定的。

在新的空间坐标系  $\tilde{x}$  中漫步，每一步的步长等于  $\sqrt{2}\sigma$ 。对应地，我们称之为自然高斯核方程 (natural)  $G^{(N)}(\tilde{x}; \sigma)$ ，对应的新坐标  $\tilde{x} = \frac{x}{\sqrt{2}\sigma}$  称为自然坐标。通过这种处理方法，可以从空间坐标系中去除方差  $\sigma$ ，从而使得高斯核方程  $G^{(N)}$  彼此相似，区别只在于其 (卷积  $\tilde{x}$  中的内方差)。

高斯核方程的值域是  $(-\infty, \infty)$ ，但在现实应用中， $|x| > \sigma$  的高斯核往往小到可以忽略不计。(如 Mathematica 计算  $\sigma = 1, x = 5\sigma$  的高斯核)。

### 16.B.5 狄拉克方程

高斯核方程是一个半局部 (semi-local) 方程。之所以不是全局的方程，是由于其卷积内部仍然存在着方差  $\sigma$ ，从而使得空间里存在一个高斯加权范围 (Gaussian weighted extend)。在下一节我们将提出一个通用方程来将其全局化，在此之前，本节先介绍一下狄拉克方程。

当将方差  $\sigma$  取其极限值  $\lim \sigma \rightarrow 0$  时，高斯核方程变成 Delta 方程的一种特殊形式，又称德尔塔-狄拉克方程 (Delta-Dirac function)  $\delta(\cdot)$ ：

$$\begin{aligned}\delta(\tilde{x}) &:= \lim_{\sigma \rightarrow 0} G(\tilde{x}; \sigma) = \lim_{\sigma \rightarrow 0} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\tilde{x}^2), \quad \tilde{x} = \frac{x}{\sqrt{2}\sigma} \\ &= \lim_{\sigma \rightarrow 0} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) \\ &= \begin{cases} \infty & x = 0, \\ 0 & x \neq 0. \end{cases}\end{aligned}\tag{16.102}$$

在数学上，狄拉克方程  $\delta(\cdot)$  常称为取样方程。例如，对  $f(x)$  取  $x = a$  时的值， $a$  是个常数。设  $f(x)$  在  $x = a$  处连续，可表示为如下求积形式

$$\int_{-\infty}^{\infty} \delta(x - a) f(x) dx = f(a).$$

狄拉克方程的导数可定义如下

$$\begin{aligned}\int_{-\infty}^{\infty} \delta'(x) f(x) dx &= -f'(0), \\ \int_{-\infty}^{\infty} \delta''(x) f(x) dx &= -f''(0).\end{aligned}\tag{16.103}$$

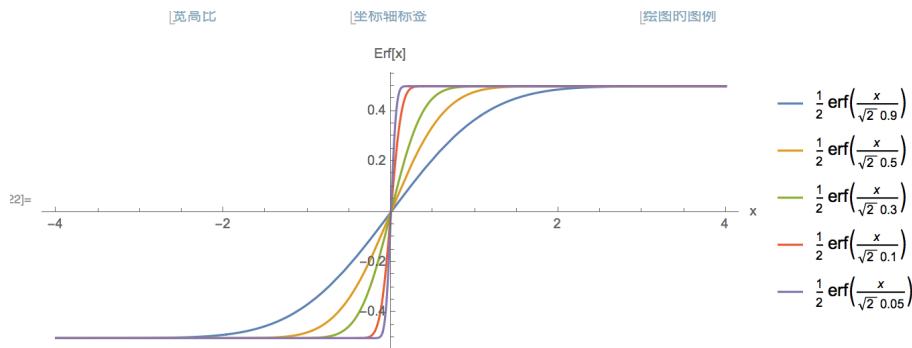
将全部  $(-\infty, x]$  区间中的高斯核加总，称为累积高斯方程 (culmulative Gaussian kernel function)，又称误差方程 (error function)

$$err(x; \sigma) := \int_0^x \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{y^2}{2\sigma^2}\right) dy,\tag{16.104}$$

在 Mathematica 中运行程序，返回值  $\frac{1}{2} Erf\left[\frac{x}{\sqrt{2}\sigma}\right]$ 。 $Erf[]$  是 Mathematica 内建的高斯误差计算方程，前面乘以 1/2 是因为求积区间只是  $[0, x]$ ，是  $[-x, x]$  的一半。

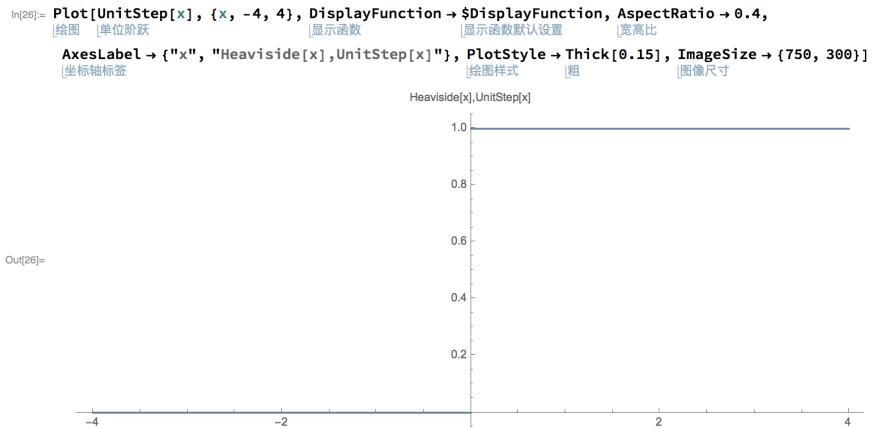
现在逐渐减少  $\sigma$  的值， $1.0 \rightarrow 0.1$ 。不难看出，随着  $\sigma$  越来越接近于 0，在  $x = 0$  附近就越会出现较大的跃动，如图(16.B.2)。

图 16.B.2: 误差方程 (累积高斯方程) 随  $\sigma$  值的变化



取  $\lim \sigma \rightarrow 0$ ，内部方差趋近于 0，此时的误差方程我们成为 Heaviside 单位阶跃方程 (unit step function)，又称黑维塞方程 (Heaviside function)。黑维塞方程的导数也是一个狄拉克方程。

图 16.B.3: 黑维塞方程 (单位阶跃方程)



## 16.B.6 高斯核和二项式系数的关系

高斯核方程也常用于多项式的幂，如

$$(x+y)^{20} = x^{20} + 20x^{19}y + 190x^{18}y^2 + 1140x^{17}y^3 + 4845x^{16}y^4 + 15504x^{15}y^5 + 38760x^{14}y^6 + 77520x^{13}y^7 + 125970x^{12}y^8 + 167960x^{11}y^9 + 184756x^{10}y^{10} + 167960x^9y^{11} + 125970x^8y^{12} + 77520x^7y^{13} + 38760x^6y^{14} + 15504x^5y^{15} + 4845x^4y^{16} + 1140x^3y^{17} + 190x^2y^{18} + 20xy^{19} + y^{20},$$

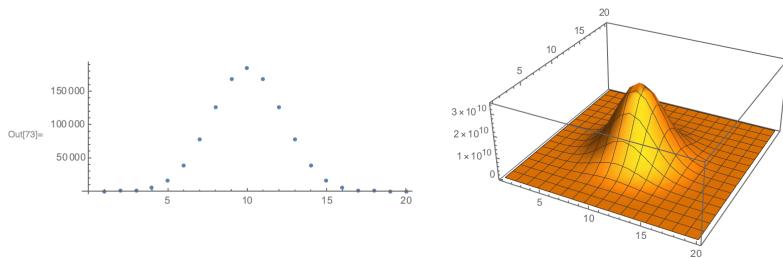
RHS 中各项前的系数，称二项式系数 (binomial coefficients)。二项式系数常常表示为

$$\binom{n}{m}$$

的形式，在 Mathematica 中用 `Binomial[n, m]` 来生成，见图 16.B.4，从左到右分别是一维和二维二项式系数的示例。

图 16.B.4: 二项式系数

```
In[71]:= binomial1 = ListPlot[Table[Binomial[20, n], {n, 1, 20}], PlotStyle -> {PointSize[0.015]}, AspectRatio -> 0.4];
binomial2 = ListPlot3D[Table[Binomial[20, n] Binomial[20, m], {n, 1, 20}, {m, 1, 20}], PlotRange -> All];
GraphicsGrid[{{binomial1, binomial2}}, ImageSize -> 750]
Out[73]=
```



### 16.B.7 中心极限定理

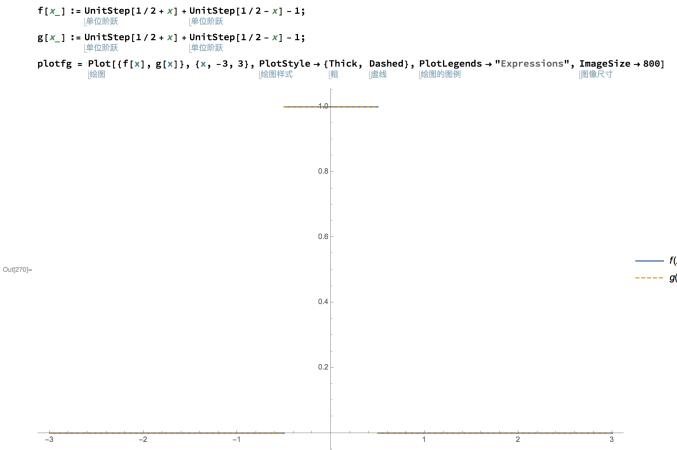
设两个块方程  $f(x), g(x)$ , 均由两个黑维塞方程相加而得, 定义为

$$f(x) := \text{UnitStep}\left(\frac{1}{2} + x\right) + \text{UnitStep}\left(\frac{1}{2} - x\right) - 1,$$

$$g(x) := \text{UnitStep}\left(\frac{1}{2} + x\right) - \text{UnitStep}\left(\frac{1}{2} - x\right) - 1,$$

见图(16.B.5), 不难看出, 两个黑维塞方程相加减还是黑维塞方程。

图 16.B.5: 黑维塞方程相加减

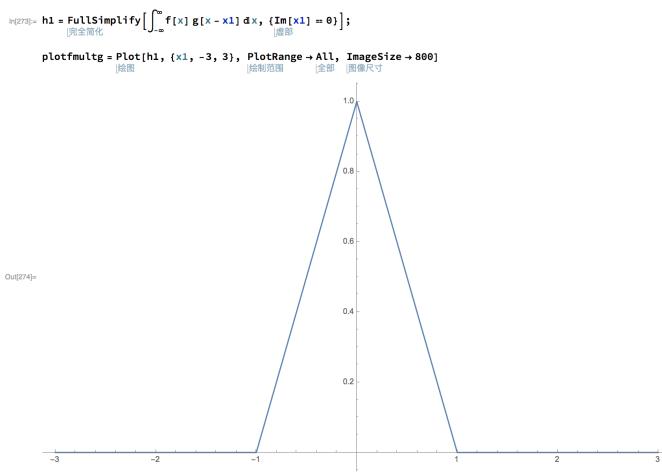


现在求卷积  $\langle f(x), g(x) \rangle$

$$\langle f(x), g(x) \rangle = \int_{-\infty}^{\infty} f(x) g(x - x_1) dx,$$

见图16.B.6, 不难看出, 两个块方程的卷积是一个三角方程。

图 16.B.6: 两个块方程的卷积是一个三角方程

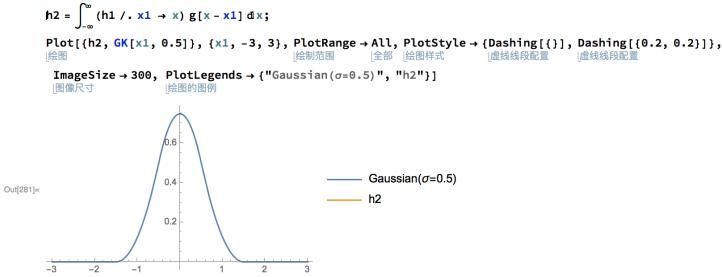


若对卷积再做一次卷积  $\langle \langle f(x), g(x) \rangle, g(x) \rangle$

$$\langle \langle f(x), g(x) \rangle, g(x) \rangle = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} f(x) g(x - x_1) dx \right\} g(x - x_1) dx,$$

如图16.B.7，可见对块方程，用它和自身再做一次卷积，二次卷积的结果看起来就很接近一个高斯核方程了。

图 16.B.7: 两个块方程卷积的卷积接近一个高斯分布



随着我们进一步作第  $3, 4, \dots, N$  次卷积， $N \rightarrow \infty$  时的结果就是一个高斯核方程。这便是中心极限定理 (central limit theorem, CLT)。

## 16.B.8 各向同性和各向异性

前面提到的高斯方程都具有各项同性 (isotropy) 的特征，意味着如果方差相同，那么方程向各个方向的变化是相同的。以 2 维高斯方程(16.99)为例，设  $\sigma = 1$ ，则 (讨论 1 维没有意义)

$$\frac{\partial}{\partial x} G^{(2)}(x, y; 1) = \frac{x}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right),$$

$$\frac{\partial}{\partial y} G^{(2)}(x, y; 1) = \frac{y}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right),$$

高斯核对  $x, y$  的偏导表现为图16.B.8中的箭头，沿着各个方向的变化幅度相同，呈圆形；每个箭头所指的方向，构成一个梯度向量 (gradient vector)。类似地，保持方差不变，3 维高斯方程的导数表现为一个球形。

另一方面，如果方差不同，那么  $N > 1$  维高斯方程在各个维度中的变化不同，我们称之为各向异性 (anisotropy)，如图16.B.9。

### 高斯核方程的傅里叶变换

实际研究中常常将实值域中的求积问题变换为傅里叶域中来计算，因为后者更方便。更多傅里叶变换的介绍见第12.G节。

傅里叶变换  $\mathcal{F}$  常用来表示，对应基方程  $\exp(i\omega x)$ ，常用的傅里叶变换以及逆傅里叶变换  $\mathcal{F}^{-1}$  可表示为

$$\mathcal{F}(f(x)) = F(\omega) \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{\infty} f(x) \exp(i\omega x) dx = \frac{1}{2\sqrt{\pi}} \exp\left(-\frac{1}{2}\sigma^2\omega^2\right), \quad (16.105)$$

$$\mathcal{F}^{-1}(F(\omega)) = \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{\infty} F(\omega) \exp(-i\omega x) d\omega. \quad (16.106)$$

在 Mathematica 中，可以自己写出  $\mathcal{F}, \mathcal{F}^{-1}$  的程序，或者用自带程序，见图16.B.10。

不难看出，对高斯核方程的傅里叶变换还是一个傅里叶变换，只是频率变化为  $\omega$ 。这一特性为高斯核方程所独有。原域中的带宽 (方差) 系数  $\sigma$ ，在新的傅里叶变换中表现为和频率  $\omega$  相乘的形式出现。 $\sigma$  值越大，方差越大，高斯核越宽，其对应的傅里叶域中的核越窄， $\omega$  越小，如下图16.B.11，从左至右分别对应  $\sigma = 1, 2, 3$ ，从上到下分别是原域中的高斯核方程，以及对应的傅里叶变换。

图 16.B.8: 2 维空间中高斯方程的各项同性

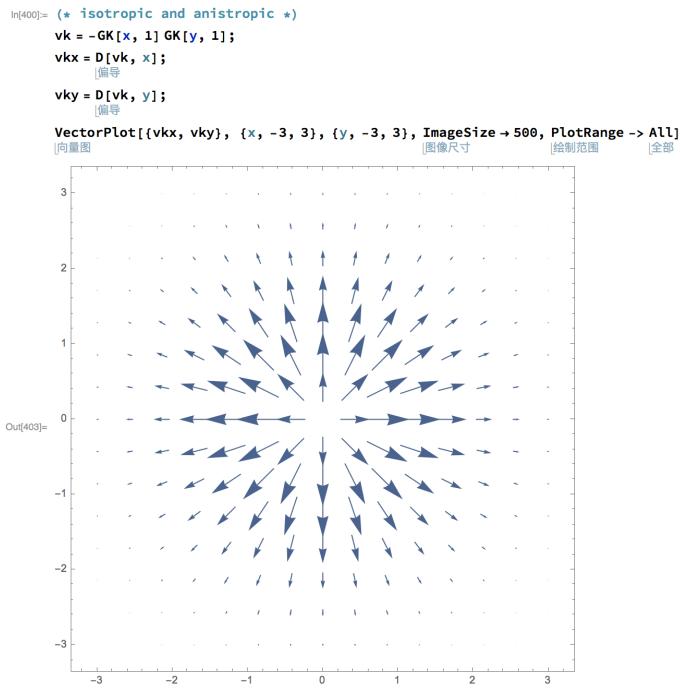


图 16.B.9: 2 维空间中高斯方程的各项异性

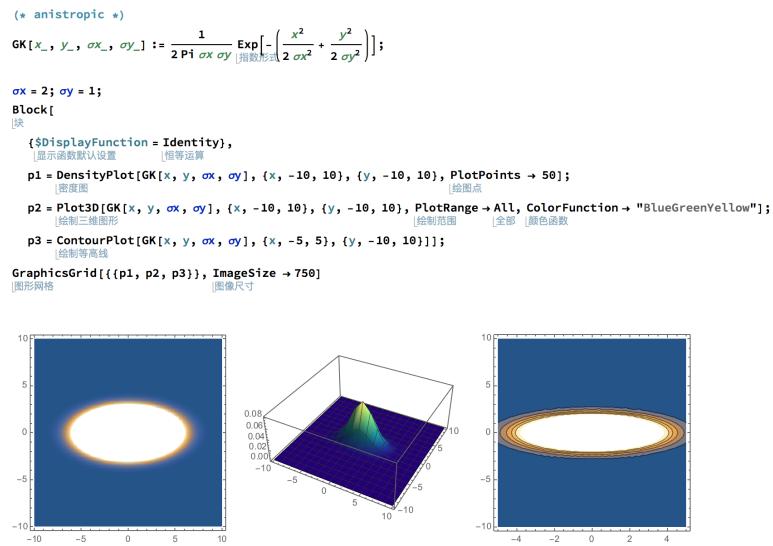


图 16.B.10: 傅里叶变换、逆傅里叶变换的 Mathematica 程序

```
(* Fourier transform *)
(* 自己写出傅里叶变换的程序 *)
σ = .;
FGK[ω_, σ_] := Simplify[ $\frac{1}{\sqrt{2 \pi}} \int_{-\infty}^{\infty} \frac{1}{\sigma \sqrt{2 \pi}} \exp\left[-\frac{x^2}{2 \sigma^2}\right] \exp[i \omega x] dx, \{\sigma > 0, Im[\sigma] == 0\}]$ ]
FGK[ω, σ]

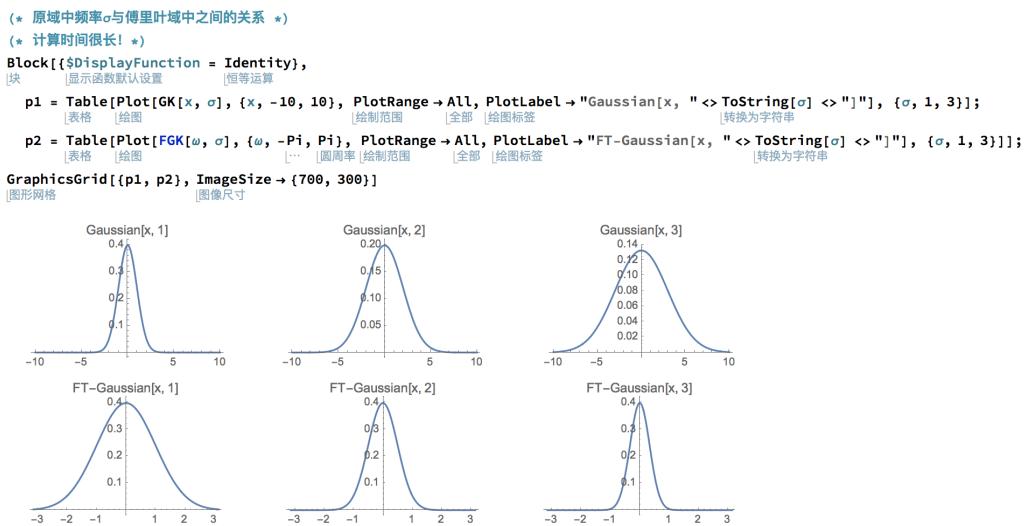
$$\frac{e^{-\frac{1}{2} \sigma^2 \omega^2}}{\sqrt{2 \pi}}$$


(* 或者利用Mathematica自带的傅里叶变换程序 *)
FullSimplify[FourierTransform[GK[x, σ], x, ω], σ > 0]
完全简化          [傅立叶变换]

$$\frac{e^{-\frac{1}{2} \sigma^2 \omega^2}}{\sqrt{2 \pi}}$$

```

图 16.B.11: 傅里叶变换、逆傅里叶变换的 Mathematica 程序



对高斯核方程  $G(x; \sigma)$  的傅里叶变换  $\mathcal{F}g(x; \sigma)$ , 有许多种不同的称谓, 若  $g(x; \sigma)$  是信号, 那么  $\mathcal{F}g(x; \sigma)$  可称为光谱 (spectrum), 起到滤波的作用, 又称带宽滤波 (bandpass filter, BPF)。

## 16.C 数值线性代数基础

### 16.C.1 矩阵结构与算法复杂度

以如下线性方程的求解为例

$$\underbrace{A}_{(n \times n)} \underbrace{x}_{(n \times 1)} = \underbrace{b}_{(1 \times 1)}, \quad (16.107)$$

假设  $A$  是非奇异矩阵 (nonsingular matrix, Proposition 11.4), 以确保对于所有  $b$  的值, 方程都有唯一解  $x = A^{-1}b$ 。 $A$  表示系数矩阵。RHS 的  $b$  设为一个常数 (随后我们会讨论更复杂一些的情况, 比如  $b$  是一个向量)。

数值近似所需的时间通常以  $n^3$  计。本节介绍的线性方程系统的通用求解方法, 在对实时性要求不高, 或者  $n$  值不太大 (如  $< 10^3$ ) 的情况下, 通常是有用的。

系数矩阵的结构问题。在通常情况下, 我们可以设系数矩阵  $A$  具有一些特殊形式, 以提高数值求解的效率 (速度), 尤其是在  $n$  值较大时, 效率的提升可能会很明显。 $A$  可供选择的形式有很多, 较简单的包括密集矩阵, 和稀疏形式的矩阵等<sup>8</sup>, 稀疏形如

- 带状矩阵 (band matrix),
- 块对角矩阵 (block diagonal matrix),
- 稀疏矩阵 (sparse matrix), 由 0 和非零元素共同构成的矩阵等。

下面以密集形矩阵为例, 介绍通用解法。在介绍之前, 先来看一下算法复杂度 (求解效率) 的判定标准。

### 利用 flops 指标作复杂度分析

数值近似计算过程中常用浮点计算数 flop(floating points operations, flops)作为测度计算效率的指标。例如求解

$$m^3 + 3m^3n + mn + 4mn^2 + 5m + 22$$

所需的浮点计算数量。需要指出的是, 在统计 flops 时常常只统计幂次较高的项, 即含有最高幂次或者主要部分的项, 的浮点运算次数, 而忽略余下的部分。在上例的数值求解过程中, 统计 flops 只需要考虑  $m^3 + 3m^2n + 4mn^2$  的部分即可; 如果  $m \ll n$ , 那么只需计算  $4mn^2$  部分即可。

现在来看基础矩阵——向量运算的成本。以计算内积  $\langle x, y \rangle$  为例, 向量  $x, y \in \mathbb{R}^n$ : 第一步需要执行  $x_i \times y_i \forall i \in n$ , 共需  $(n)$  flops。然后将各项加总, 需  $(n - 1)$  flops, 一共  $(2n - 1)$  flops。若只统计首项,  $2n$  flops。或者更简单些, 只考虑幂次,  $n$  flops。

### 标量——矩阵乘积形式

$$\alpha x, \alpha \in \mathbb{R}, x \in \mathbb{R}^n, n \text{ flops.}$$

$$x + y, n \text{ flops.}$$

如果  $x$  和  $y$  是稀疏矩阵, 即均含有一些非零元素, 那么计算速度可能更快。例如若  $x$  中的非零元素数量是  $N$ , 则  $\langle x, y \rangle = x^\top y$ , 有  $2N$  flops。

---

<sup>8</sup>较为复杂的矩阵形式暂不讨论, 如 Toeplitz, Hankel, Circulant matrix 等。

### 矩阵——向量乘积形式

以计算下式为例

$$\underbrace{\begin{matrix} y \\ \hline (m \times 1) \end{matrix}}_{(m \times 1)} = \underbrace{\begin{matrix} A \\ \hline (m \times n) \end{matrix}}_{(m \times n)} \underbrace{\begin{matrix} x \\ \hline (n \times 1) \end{matrix}}_{(n \times 1)},$$

$2mn$  flops。

$A$  的特殊形式可以提升计算速度，如

- $A$  是对角矩阵，那么  $Ax$  计算需  $n$  flops，
- $A$  是系数矩阵，其中包括  $N$  个非零元素，那么  $Ax$  计算需  $2N$  flops。

如果设  $A$  的秩满足  $\text{rank}(A) \leq \min\{m, n\}$ ，并且  $A$  可以分解为

$$\underbrace{\begin{matrix} A \\ \hline (m \times n) \end{matrix}}_{(m \times n)} = \underbrace{\begin{matrix} U \\ \hline (m \times p) \end{matrix}}_{(m \times p)} \underbrace{\begin{matrix} V \\ \hline (p \times n) \end{matrix}}_{(p \times n)},$$

那么可以首先计算  $Vx$  ( $2pn$  flops)，再计算  $U(Vx)$  ( $2mp$  flops)，总共  $2(m+n)p << 2mn$  flops。

### 矩阵——矩阵乘积形式

来看

$$\underbrace{\begin{matrix} C \\ \hline (m \times p) \end{matrix}}_{(m \times p)} = \underbrace{\begin{matrix} A \\ \hline (m \times n) \end{matrix}}_{(m \times n)} \underbrace{\begin{matrix} B \\ \hline (n \times p) \end{matrix}}_{(n \times p)},$$

$2mnp$  flops。

## 16.C.2 矩阵结构与线性系统求解

先从较简单的线性系统(16.107)求解开始。

### 对角矩阵

设  $A$  是对角矩阵 (diagonal matrix)，非奇异，即矩阵中元素满足  $a_{ii} \neq 0 \forall i \in n$ 。那么有

$$ax = b \Leftrightarrow a_{ii}x_i = b_i, i = 1, \dots, n,$$

对应的解  $\{x_i\}_{i \in n}$  满足

$$x_i = \frac{b_i}{a_{ii}}, \quad i = 1, \dots, n,$$

$n$  flops。

### 下三角矩阵

若  $A \in \mathbb{R}^{n \times n}$  满足  $a_{i,j} \forall j > i, i, j = 1, \dots, n$ ，我们称之为下三角矩阵 (lower triangular matrix)。若下三角矩阵中的所有非零元素的值都为 1，则称为单位下三角矩阵 (unit lower triangular matrix)，例如一个  $5 \times 5$  的单位下三角矩阵

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

当且仅当下三角矩阵  $A$  的全部对角元素不为 0, 即  $a_{ii} \neq 0 \forall i = 1, \dots, n$  时,  $A$  才是一个非奇异矩阵。

非奇异下三角矩阵  $A$  代回原式, 有

$$\begin{pmatrix} a_{11} & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & 0 \\ \vdots & \ddots & \ddots & 0 \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

求解, 依次有

$$\begin{aligned} x_1 &= \frac{b_1}{a_{11}}, \quad (1 \text{ flops}) \\ x_2 &= \frac{b_2 - a_{21}x_1}{a_{22}}, \quad (3 \text{ flops}) \\ x_3 &= \frac{b_3 - a_{31}x_1 - a_{32}x_2}{a_{33}}, \quad (5 \text{ flops}) \\ &\vdots \\ x_n &= \frac{b_n - a_{n1}x_1 - a_{n2}x_2 - \dots - a_{n,n-1}x_{n-1}}{a_{nn}}, \quad (2n-1 \text{ flops}) \end{aligned}$$

这种求解思路称向前替换 (forward substitution), 即在每一个计算  $x_i$  的过程中, 将已求解得到的  $x_1, x_2, \dots, x_{i-1}$  作为已知量代入。总共需要计算

$$\Sigma = 1 + 3 + 5 + \dots + (2n-1) = n^2 \text{ flops.}$$

如果  $A$  除了下三角、非奇异之外还有其他更多一些特征, 我们可以进一步优化算法, 提高求解效率, 使得计算数低于  $n^2$ 。例如若  $A$  还是稀疏的或带状的, 每一行中有不超过  $k$  个元素, 则向前替代需要  $(2k+1)$  flops, 总计  $(2k+1)n$  flops, 以首项记为  $(2kn)$  次。

### 上三角矩阵

若  $A \in \mathbb{R}^{n \times n}$  的转置  $A^\top$  是个下三角矩阵, 满足  $a_{i,j} = 0 \forall i > j, i, j = 1, \dots, n$ , 我们称之为上三角矩阵 (upper triangular matrix)。若上三角矩阵中的所有非零元素的值都为 1, 则称为单位上三角矩阵 (unit upper triangular matrix)。

对应的求解过程为向后替代 (backward substitution)

$$\begin{aligned} x_n &= \frac{b_n}{a_{n-1}}, \\ x_{n-1} &= \frac{b_{n-1} - a_{n-1,n}x_n}{a_{n-1,n-1}}, \\ x_{n-2} &= \frac{b_{n-2} - a_{n-2,n-1}x_{n-1} - a_{n-2,n}x_n}{a_{n-2,n-2}}, \\ &\vdots \\ x_1 &= \frac{b_1 - a_{12}x_2 - a_{13}x_3 - \dots - a_{1n}x_n}{a_{11}}, \end{aligned}$$

也是  $n^2$  flops。若  $A$  稀疏, 每一行最多有  $k$  个非零元素, 那么  $(2kn)$  flops。

### 正交矩阵

$A \in \mathbb{R}^{n \times n}$  若满足如下条件, 我们称之为正交矩阵 (orthogonal matrix)

$$A^\top A = I, \quad \text{即 } A^{-1} = A^\top.$$

此时线性系统  $Ax = b$  为

$$x = A^{-1}b = A^\top b, \quad (2n^2) \text{ flops.}$$

若正交矩阵还具有更多属性如

$$A = I - 2uu^\top, \quad \|u\|_2 = 1,$$

那么

$$x = A^{-1}b = (I - 2uu^\top)^\top = b - 2u^\top bu,$$

### 计算步骤

- 算  $u^\top b$ , ( $n$ ) flops,
- 算  $u^\top bu$ , ( $n$ ) flops,
- 算  $-2u^\top bu$ , ( $n$ ) flops,
- 算  $b - 2u^\top bu$ , ( $n$ ) flops,

$\Sigma = 4n$  flops。

### 置换矩阵

置换矩阵 (permutation matrix)是指只由 0 和 1 组成的方块矩阵, 每一行和每一列都有且只有一个 1, 其余均为 0。在线性代数中, 每个  $n$  元置换矩阵都代表一个对  $n$  个元素的置换, 这  $n$  个元素常常是  $n$  维空间的基。当一个矩阵乘以一个置换矩阵时, 得到的是原矩阵的行 (若置换矩阵在左) 或列 (若置换矩阵在右) 经过置换后的矩阵。具体说来, 每个  $n$  元置换都对应唯一的一个置换矩阵。设  $\pi$  为一个  $n$  元置换  $\pi : \{1, \dots, n\} \mapsto \{1, n\}$ , 对应映射图

$$\begin{pmatrix} 1 & 2 & \dots & n \\ \pi(1) & \pi(2) & \dots & \pi(n) \end{pmatrix},$$

那么它的  $n \times n$  置换矩阵  $P_\pi$  是指, 在第  $i$  行上只有  $\pi(i)$  位置上的系数是 1, 其余均为 0:

$$P_\pi = \begin{pmatrix} e_{\pi(1)} \\ e_{\pi(2)} \\ \vdots \\ e_{\pi(n)} \end{pmatrix}, \quad (16.108)$$

其中  $e_j$  表示正则基中的第  $j$  个, 也即第  $j$  行左起第  $j$  个元素是 1, 其余均为 0。

此外设一个单位矩阵 (identity matrix)  $I \in \mathbb{R}^{n \times n} = [e_1, e_2, e_3, \dots, e_n]^\top$ , 因此置换矩阵也可以看做是对单位矩阵的某些行和列交换后所得到的矩阵。

置换矩阵具有一些性质如下

1. 设 2 个  $n$  元置换  $\pi$  和  $\sigma$ , 分别对应 2 个  $n \times n$  置换矩阵  $P_\pi, P_\sigma$ , 那么

$$P_\pi P_\sigma = P_{\pi \sigma}.$$

2. 一个置换矩阵必然是正交矩阵, 满足

$$P_\pi P_\pi^\top = I,$$

并且它的逆矩阵也是一个置换矩阵

$$P_\pi^{-1} = P_{\pi^{-1}} = P_\pi^\top.$$

左乘: 置换矩阵  $P_\pi$  左乘一个列向量  $g$ , 得到的是  $g$  的系数经置换后的新向量

$$P_\pi g = \begin{pmatrix} e_{\pi(1)} \\ e_{\pi(2)} \\ \vdots \\ e_{\pi(n)} \end{pmatrix} \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \end{pmatrix} = \begin{pmatrix} g_{\pi(1)} \\ g_{\pi(2)} \\ \vdots \\ g_{\pi(n)} \end{pmatrix}.$$

右乘: 置换矩阵  $P_\pi$  右乘一个行向量  $h$ , 得到的是  $h$  的系数经置换后的新向量

$$h P_\pi = \left( h_1, h_2, \dots, h_n \right) \begin{pmatrix} e_{\pi(1)} \\ e_{\pi(2)} \\ \vdots \\ e_{\pi(n)} \end{pmatrix} = \left( h_{\pi^{-1}(1)}, h_{\pi^{-1}(2)}, \dots, h_{\pi^{-1}(n)} \right).$$

例如  $\pi = (1, 4, 2, 5, 3)$ , 对应映射

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 2 & 5 & 3 \end{pmatrix},$$

建立置换矩阵

$$P_\pi = \begin{pmatrix} e_{\pi(1)} \\ e_{\pi(2)} \\ e_{\pi(3)} \\ e_{\pi(4)} \\ e_{\pi(5)} \end{pmatrix} = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

对于一个给定的向量  $g = [g_1, g_2, g_3, g_4, g_5]^\top$ , 我们有

$$P_\pi g = \begin{pmatrix} e_{\pi(1)} \\ e_{\pi(2)} \\ e_{\pi(3)} \\ e_{\pi(4)} \\ e_{\pi(5)} \end{pmatrix} \begin{pmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \\ g_5 \end{pmatrix} = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{pmatrix} \begin{pmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \\ g_5 \end{pmatrix} = \begin{pmatrix} g_1 \\ g_4 \\ g_2 \\ g_5 \\ g_3 \end{pmatrix}.$$

回到本节  $Ax = b$  的求解问题。如果  $A$  是一个转置矩阵, 那么求解方法为, 对向量  $b$  做  $\pi^{-1}$  的置换:

$$x = A_{\pi^{-1}} b = A^T b, \quad (0) flops.$$

虽然仍然涉及到浮点数的复制粘贴, 但计为 0。

### 16.C.3 矩阵分解

在实际求解线性系统的过程中，有时很难第一眼就看出系数矩阵  $A$  具有哪些特殊结构。这时可以尝试将它分解为一组非奇异矩阵的乘

$$\begin{aligned} A &= A_1 A_2 \dots A_k, \\ \hookrightarrow Ax = b &\Leftrightarrow x = A^{-1}b = A_k^{-1} A_{k-1}^{-1} \dots A_1^{-1} b. \end{aligned} \quad (16.109)$$

可以从右向左求解

$$\begin{aligned} z_1 &= A_1^{-1}b, \\ z_2 &= A_2^{-1}z_1 = A_2^{-1}A_1^{-1}b, \\ &\vdots \\ z_i &= A_i^{-1}z_{i-1} = A_i^{-1}A_{i-1}^{-1}\dots A_1^{-1}b, \end{aligned} \quad (16.110)$$

在此基础上进一步计算系统的解  $x$

$$x = A_k^{-1}z_{k-1} = A_k^{-1}A_{k-1}^{-1}\dots A_1^{-1}b.$$

这种方法称为矩阵的因子分解 (factorization)。如果  $A_i$  具有较好的结构，比如是 (上或下) 对角矩阵、置换矩阵等，那么分解法可用于高效求解线性系统  $Ax = b$ 。有时为了记忆方便，对于将系数矩阵做分解  $A = A_1 \dots A_i$  的计算过程 (16.109) 称为分解步骤 (factorization step)，对应  $f$  flops。对于求解  $z_i = A_i^{-1}z_{i-1}$  (16.110) 的步骤称为求解步骤 (solution step)，对应  $s$  flops。通常来说，求解步骤所需成本远高于分解步骤， $f >> s$ 。

现在作个扩展， $b$  不再是个常数而是个向量，对应线性系统

$$\begin{aligned} AX &= B, \\ A &\in \mathbb{R}^{n \times n}, \\ X &= [x_1, x_2, \dots, x_m] \in \mathbb{R}^{n \times m}, \\ B &= [b_1, b_2, \dots, b_m] \in \mathbb{R}^{n \times m}. \end{aligned}$$

求解方法类似： $A$  的分解步骤需要  $f$  flops；随后对  $1, \dots, m$  分别求解  $A_i^{-1}b_i$ ，一共需要  $(ms)$  flops， $\Sigma = (f + ms)$  flops，并且往往  $f >> s$ 。

下面介绍一些常见的因子分解法，如 LU 分解，Cholesky 分解， $LDL^\top$  分解等。

### 16.C.4 LU 因子分解

任意一个非奇异矩阵  $A \in \mathbb{R}^{n \times n}$  都可以分解为 PLU 的形式

$$A = P L U, \quad (16.111)$$

其中  $P \in \mathbb{R}^{n \times n}$  是置换矩阵， $L, U \in \mathbb{R}^{n \times n}$  分别是上、下三角矩阵，这种方法称为  $A$  的 LU 分解 (LU factorization)。结合转置矩阵的特征，上式可改写为

$$P^\top A = L U, \quad (16.112)$$

数值计算求解 LU 分解的标准算法是高斯部分主元消除法 (Gaussian elimination of partial pivoting)，又称高斯行主元消除法 (Gaussian elimination of row pivoting)，消耗  $\frac{2}{3}n^3$  flops。

### 利用 LU 分解法求解线性方程系统

LU 分解 + 分解步和求解步，是线性方程系统如  $Ax = b$  的标准求解方法。

**Algorithm 16.2** (利用 LU 分解法求解线性方程系统). 给定一个线性方程系统  $Ax = b$ ，其中系数矩阵  $A$  非奇异， $b$  是个常数，经典的求解步骤如下：

1. LU 因式分解，将  $A$  分解为  $A = P L U$ ,  $(\frac{2}{3}n^3)$  flops。

2. 置换，求解  $Pz_1 = b$ ,  $(0)$  flops。

3. 向前替换，求解  $Lz_2 = z_1$ ,  $(n^2)$  flops。

4. 向后替换，求解  $Ux = z_2$ ,  $(n^2)$  flops。

总共消耗  $(\frac{2}{3}n^3 + 3n^2)$  flops。

如果  $b \in \mathbb{R}^n$  是一个向量，整个系统变成一个由  $m$  个方程组成的线性系统  $Ax_i = b_i, i = 1, \dots, m$ ，计算方法类似，消耗成本  $(\frac{2}{3}n^3 + 2mn^2)$  flops。

如果非奇异矩阵  $A$  还具有其他一些额外的性质，如带状或系数，则可通过优化算法提高数值计算的效率。

若  $A$  是带状矩阵 (band matrix)，指

$$a_{ij} = 0, \quad \text{如果 } |i - j| > k, \quad (16.113)$$

其中  $k < n - 1$ ，称为  $A$  的带宽 (bandwidth)。对于  $k \ll n$  的情况，计算消耗  $(4nk^2)$  flops。

若  $A$  是稀疏矩阵，那么其 LU 因子分解中常常同时既含有行置换，也含有列置换，即

$$A = P_1 L U P_2,$$

其中  $P_1, P_2$  都是置换矩阵。这时采用 LU 因子分解会大大提升计算效率，L 和 U 的稀疏程度取决于  $P_1$  和  $P_2$  的具体形式。

### 16.C.5 Cholesky 因子分解

如果  $A \in \mathbb{R}^{n \times n}$  是对称且正半定的矩阵，那么可以因子分解为

$$A = L L^\top, \quad (16.114)$$

其中 L 是下三角矩阵，非奇异，且对角元素值为正。这称为  $A$  的 Cholesky 因子分解 (Cholesky factorization)。当不对  $A$  作任何结构优化时，消耗  $(\frac{1}{3}n^3)$  flops。

**Algorithm 16.3** (利用 Cholesky 分解求解正定线性方程系统). 给定一组线性方程系统  $Ax = b$ ，其中系数矩阵  $A$  对称，正半定。那么利用 Cholesky 分解求解系统的算法如下

1. Cholesky 因子分解， $A = L L^\top$ ,  $(\frac{1}{3}n^3)$  flops。

2. 向前替代，求解  $Lz_1 = b$ ,  $(n^2)$  flops。

3. 向后替代，求解  $L^\top x = z_1$ ,  $(n^2)$  flops。

$$\Sigma = \frac{1}{3}n^3 + 2n^2 \approx \frac{1}{3}n^3 \text{ flops.}$$

若  $A$  还具有进一步的特殊结构，则计算效率可进一步提高。如果  $A$  对称、正半定且是个带宽为  $k$  的带状矩阵，计算消耗  $(nk^2)$  flops。

如果  $A$  对称、正半定且是个稀疏矩阵  $A = P L L^\top P^\top$ ，那么等价于

$$P^\top A P = L L^\top,$$

换句话说， $(P^\top A P)$  的 Cholesky 因子分解是  $L L^\top$ 。

原则上来说，我们可以选择任意形式的置换矩阵  $P$ ，每 1 个  $P$  都有唯一的  $L$  相对应，但不同  $P$  的选择会大大影响数值求解的效率。对此，常常需要采用不同的启发式方法来选取最合适、或者更合适的置换矩阵类型。例如当  $A$  是一个稀疏型矩阵

$$A = \begin{pmatrix} 1 & u^\top \\ u & D \end{pmatrix},$$

其中  $D \in \mathbb{R}^{n \times n}$  是正对角矩阵， $u \in \mathbb{R}^n$ 。

$A$  正定要求满足  $\det A > 0$

$$\begin{aligned} \det A &= D - u u^\top > 0, \\ \rightarrow D &> u u^\top, \\ \rightarrow u^\top D^{-1} u &< 1, \end{aligned}$$

对  $A$  的 Cholesky 因子分解可写作

$$\begin{pmatrix} 1 & u^\top \\ u & D \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ u & L \end{pmatrix} \begin{pmatrix} 1 & u^\top \\ 0 & L^\top \end{pmatrix}, \quad (16.115)$$

其中  $L$  是下三角矩阵，满足

$$L L^\top = D - u u^\top.$$

对于一般形式的  $u$  而言，矩阵  $(D - u u^\top)$  是密集矩阵，因此可以判定  $L$  也是密集矩阵。那么，尽管  $A$  可能是非常稀疏的（比如大多数行只有两个非零元素），其 Cholesky 因子  $L$  却仍是密集的。

另一方面，假设我们将  $A$  的第 1 行、第 1 列置换到最后 1 行、最后 1 列，重新排序后做 Cholesky 因子分解，可得

$$\begin{pmatrix} D & u \\ u^\top & 1 \end{pmatrix} = \begin{pmatrix} D^{\frac{1}{2}} & 0 \\ u^\top D^{-\frac{1}{2}} & (1 - u^\top D^{-1} u)^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} D^{\frac{1}{2}} & D^{-\frac{1}{2}} u \\ 0 & (1 - u^\top D^{-1} u)^{\frac{1}{2}} \end{pmatrix} \quad (16.116)$$

由(16.115) (16.116)可见，Cholesky 因子  $L$  的对角块是  $(1, 1)$ ，因此也是密集的。

这个例子表明，重新排序可以极大改变 Cholesky 因子的稀疏性。很显然，“最好”的置换方案，或者退一步说，所有“理想”的排序方法，都是将密集的行和列置换到末尾。所谓“理想”的方法，是指较为复杂的稀疏型矩阵也许很难找到最优的置换方案，即，使得  $L$  中的 0 元素个数达到最大，但还是有一些启发式方案可以追求次优。

在系数矩阵作 Cholesky 因子分解过程中， $P$  的选取往往只根据  $A$  的稀疏结构进行，而不取决于  $A$  中非零元素的具体数值<sup>9</sup>。而一旦确定了  $P$ ，我们就可以算得唯一的  $L$  的系数结构，也不取决于  $A$  中非零元素的具体数值。这成为  $A$  的符号因子分解 (symbolic factorization step)，作为求解稀疏矩阵系统的第一步。第二步是数值因子分解，即计算  $L$  中非零元素的值。常见的计算 Cholesky 因子分解的软件常常将这两个步骤的计算过程分别处理，这在大多数情况下是有必要的：第一步的计算成本往往高于第二步。

<sup>9</sup>与之相比较，对稀疏矩阵作  $LU$  分解的过程中， $P$  的选取则不只取决于  $A$  的稀疏结构，还取决于  $A$  中非零元素的具体值。

### 16.C.6 $LDL^\top$ 因子分解

每个非奇异对称矩阵  $A$  都可因式分解为如下形式

$$A = P L D L^\top P^\top, \quad (16.117)$$

其中  $P$  是置换矩阵,  $L$  是对角元素均为正的下三角矩阵,  $D$  是块对角矩阵, 对角块由  $1 \times 1$  和  $2 \times 2$  的非奇异对角块构成。这称为  $A$  的  $LDL^\top$  因子分解 ( $LDL^\top$  factorization)。

Cholesky 分解可以被视为是一种特殊形式的  $LDL^\top$  分解:  $P = I$ ,  $D = I$ 。

**Algorithm 16.4** (利用  $LDL^\top$  分解求解线性方程系统). 给定一个线性方程  $Ax = b$ , 其中  $A$  非奇异。 $LDL^\top$  法求解步骤如下

1.  $LDL^\top$  因子分解,  $A = P L D L^\top P^\top$ ,  $(\frac{1}{3}n^3)$  flops。
2. 置换, 求解  $Pz_1 = b$ , (0) flops。
3. 向前替换, 求解  $Lz_2 = z_1$ ,  $(n^2)$  flops。
4. 对角块元素求解, 求解  $Dz_3 = z_2$ , (n) flops。
5. 向后替换, 求解  $L^\top z_4 = z_3$ ,  $(n^2)$  flops。
6. 置换, 求解  $P^\top x = z_4$ , (0) flops。

当  $A$  具有一些特殊结构, 如带状矩阵或系数矩阵时, 可以进一步提高计算效率。调整过程与 Cholesky 因子分解的过程相似, 只是要额外多考虑一个对角块  $D$ 。此外,  $LDL^\top$  因子分解过程中, 在选取  $P$  时, 除了要 (像稀疏 Choleky 分解那样) 考虑稀疏矩阵的结构特征之外, 还要考虑矩阵  $A$  中非零元素的值。

### 16.C.7 块消除与 Schur 补

#### 变量块的高斯消元

求解线性方程系统  $Ax = b$  的过程中, 一种常用的方法是, 首先利用一定方法小区部分变量元素, 将剩下的变量元素组成一个更小的线性方程系统, 以便提高数值计算的效率。

若  $A$  是一个没有明显规律的密集矩阵, 这种方法很难带来计算效率的提升。然而若  $A$  中的某一部分子矩阵含有可消除变量, 易于作因子分解 (例如是快对角或带状结构的), 那么应用这种方法会带来效率的显著提升。

假定我们将  $x \in \mathbb{R}^n$  分为以下 2 个块 (子向量)

$$\underbrace{\begin{pmatrix} x \\ ((n_1+n_2) \times 1) \end{pmatrix}}_{(n_1+n_2) \times 1} = \begin{pmatrix} \underbrace{x_1}_{(n_1 \times 1)} \\ \underbrace{x_2}_{(n_2 \times 1)} \end{pmatrix},$$

相应地, 待求解线性系统变为

$$Ax = b \iff \begin{pmatrix} \underbrace{A_{11}}_{(n_1 \times n_1)} & \underbrace{A_{12}}_{(n_1 \times n_2)} \\ \underbrace{A_{21}}_{(n_2 \times n_1)} & \underbrace{A_{22}}_{(n_2 \times n_2)} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}. \quad (16.118)$$

若假定子矩阵  $A_{11}$  可逆，则可以采用如下方法，从线性方程中消去  $A_{11}$ ：由第一行方程有

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 &= b_1, \\ \hookrightarrow x_1 &= \frac{b_1 - A_{12}x_2}{A_{11}}. \end{aligned} \quad (16.119)$$

代入第二行方程替换掉其中的  $x_1$

$$\begin{aligned} A_{21}x_1 + A_{22}x_2 &= b_2, \\ \hookrightarrow \underbrace{(A_{22} - A_{21}A_{11}^{-1}A_{12})}_{=: \mathcal{S}} x_2 &= b_2 - A_{21}A_{11}^{-1}b_1, \end{aligned} \quad (16.120)$$

(16.120) 称作线性系统的缩减式 (reduced equation)。

(16.119) (16.120) 等价于原系统(16.118)。缩减式中的矩阵  $\mathcal{S}$  称为  $A$  中的块  $A_{11}$  的 Schur 补 (Schur complement, 第11.A节)。并且当且仅当  $A$  是非奇异矩阵时， $\mathcal{S}$  才是非奇异矩阵。

根据 Schur 补(16.119) (16.120)，我们可以利用高斯消元法 (Gaussian elimination) 来求解线性系统，步骤如下：

**Algorithm 16.5** (高斯消元法求解线性系统). 给定非奇异的线性方程系统(16.118)，其中分块  $A_{11}$  非奇异、可逆，系统求解步骤如下：

1. 计算  $A_{11}^{-1}A_{12}$  和  $A_{11}^{-1}b_1$ ，
2. 计算 Schur 补  $\mathcal{S} = A_{22} - A_{21}A_{11}^{-1}$  和  $\tilde{b} = b_2 - A_{21}A_{11}^{-1}b_1$ ，
3. 利用  $\mathcal{S}x_2 = \tilde{b}$  求解  $x_2$ ，
4. 利用  $A_{11}x_1 = b_1 - A_{12}x_2$  求解  $x_1$ 。

事实上，我们可以将高斯消元法理解为一种对块的分解——求解法，例如对  $A$  作因子分解

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} A_{11} & 0 \\ A_{21} & \mathcal{S} \end{pmatrix} \begin{pmatrix} I & A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix},$$

可见这是块形式的 LU 因子分解。那么求解系统(16.118)的思路为：

1. 作块的向前替换，求解  $z_1, z_2$

$$\begin{pmatrix} A_{11} & 0 \\ A_{21} & \mathcal{S} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix},$$

有

$$\begin{aligned} z_1 &= A_{11}^{-1}b_1, \\ z_2 &= \mathcal{S}^{-1}(b_2 - A_{21}z_1). \end{aligned}$$

2. 利用块的向后替换，求解  $x_1, x_2$

$$\begin{pmatrix} I & A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix},$$

有

$$\begin{aligned} x_1 &= z_1 - A_{11}^{-1}A_{12}z_2, \\ x_2 &= z_2. \end{aligned}$$

眼下利用因子分解——求解法计算线性系统，主要采取这种块分解的思路和步骤，块的大小常由计算机程序来自动作“次优”选取。

此外，如上例所示，当线性方程系统中含有一些密集的行、列时，高斯消元法和因子分解法之间其实密切相关。在高斯消元过程中也可以进一步使用因子分解法，如将  $A_{11}$  和  $\mathcal{S}$  分解为

$$\begin{aligned} A_{11} &= P_1 L_1 U_1 P_2, \\ \mathcal{S} &= P_3 L_2 U_2. \end{aligned}$$

系数矩阵  $A$  可以改写为

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} P_1 & 0 \\ 0 & P_3 \end{pmatrix} \begin{pmatrix} L_1 & 0 \\ P_3^\top A_{21} P_2^\top U_1^{-1} & L_2 \end{pmatrix} \begin{pmatrix} U_1 & L_1^{-1} P_1^T A_{12} \\ 0 & U_2 \end{pmatrix} \begin{pmatrix} P_2 & 0 \\ 0 & I \end{pmatrix}.$$

由  $A$  的对称性，可得  $A_{11}$  和  $\mathcal{S}$  的对称性，因此可以对  $A_{11}$  和  $\mathcal{S}$  作对称的因式分解。此外，对称结构也使得其他一些操作成为可能，以进一步提高矩阵乘等计算的效率。通常说来，对称结构系统的求解速度是非对称结构系统的一半。

进一步，如果  $A$  对称且正定，那么  $A_{11}$  和  $\mathcal{S}$  也对称且正定，进而可用 Cholesky 因子分解法。

除了  $A$  本身的特殊结构之外，若其分块  $A_{12}, A_{21}, A_{22}$  以及  $\mathcal{S}$  也具有一些特定的结构，那么我们也可以对这些分块再做一次 Schur 补——因子分解——求解，以进一步提高数值计算的效率。举例说明。

例：块三角矩阵和线性方程系统。设  $A_{12} = 0$ ，那么线性系统  $Ax = b$  呈块下三角结构

$$\begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix},$$

此时  $\mathcal{S} = A_{22}$ ，根据高斯消元法，作块的向前消元，有

$$\begin{aligned} x_1 &= A_{11}^{-1} b_1, \\ x_2 &= A_{22}^{-1} (b_2 - A_{21} x_1). \end{aligned}$$

例：块对角矩阵和线性方程系统。设  $A_{11}$  是对角矩阵，最大不超过  $\ell \times \ell$ ， $A_{12}, A_{21}, A_{22}$  是带宽为  $k$  的带状矩阵。由此可得  $A_{11}^{-1}$  也是对角矩阵，大小与  $A_{11}$  相同。那么  $A_{11}^{-1} A_{12}$  也是带状矩阵，带宽为  $k + \ell$ 。进而  $\mathcal{S} = A_{22} - A_{21} A_{11}^{-1} A_{12}$  也是一个带状矩阵，带宽为  $2k + \ell$ 。这意味着更高的计算效率。

### 16.C.8 矩阵求逆引理

通过前面的介绍不难看出，高斯消元法主要是指，对原线性系统消除一部分元素变量后，得到一个规模更小的新的线性系统，再结合原系统的 Schur 补等信息，对新系统进行求解计算。把这个思路颠倒过来看，如果得到了一个 Schur 补矩阵，那么我们可以引入新的元素变量，重构原方程系统并求解。大多数情况下这么做没有好处：规模越大的系统，数值求解速度越慢。但也有例外的情况，如果比起缩减系统来，恢复出的原系统具有某些特殊结构的话，对原系统做数值求解可以进一步提升计算效率。例如线性方程系统

$$(A + BC)x = b, \quad (16.121)$$

其中  $A \in \mathbb{R}^{n \times n}$  非奇异， $B \in \mathbb{R}^{n \times p}$ ， $C \in \mathbb{R}^{p \times n}$ 。

我们可以引入一个新变量  $y \equiv Cx$ ，将(16.121)改写为

$$Ax + By = b, \iff \begin{pmatrix} A & B \\ C & -I \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix}. \quad (16.122)$$

小系统(16.121)中的系数矩阵  $(A + BC)$ , 是大系统(16.122)中的 Schur 补  $\mathcal{S}$ , 二者等价。在有些情况下, 大系统(16.122)的计算效率比小系统(16.121)高, 例如  $A, B, C$  单个矩阵都较为稀疏, 但集中在一起  $(A + BC)$  较为密集。

一方面在大系统(16.122)中消除新定义的变量  $y = Cx$  后, 又会回复到小系统(16.121)。另一方面, 也可以在大系统(16.122)中消除原有变量  $x$

$$\begin{aligned} Ax + By &= b, \\ x &= A^{-1}(b - By), \\ y &= Cx = C((b - By)) = CA^{-1}b - CA^{-1}By, \\ (I + CA^{-1}B)y &= CA^{-1}b, \\ y &= (I + CA^{-1}B)^{-1}CA^{-1}b, \end{aligned}$$

进而

$$\begin{aligned} x &= A^{-1}(b - By) \\ &= A^{-1}b - A^{-1}B((I + CA^{-1}B)^{-1}CA^{-1}b) \\ &= (A^{-1} - A^{-1}B(I + CA^{-1}B)^{-1}CA^{-1})b. \end{aligned} \tag{16.123}$$

对于给定的  $b$ , 将  $b = (A + BC)x$  代回上式, 进一步得

$$(A + BC)^{-1} = A^{-1} - A^{-1}B(I + CA^{-1}B)^{-1}CA^{-1}, \tag{16.124}$$

这被称为是矩阵求逆引理 (matrix inversion lemma), 或 Sherman-Woodbury-Morrison 公式 (Sherman-Woodbury-Morrison formula), 简称 Woodbury 公式。

Woodbury 公式的应用有很多, 其中之一如, 在  $p$  值相对较小的情况下, 若我们已有一个相对有效的方法求解  $Au = v$ , 则可利用 Woodbury 公式求解线性系统  $(A + BC)x = b$ 。

下面来看若  $A$  具有特殊结构的情况, 假定  $A$  是个对角矩阵, 对角元素非零, 那么求解(16.121)有两种方法。

1. 直观方法是, 首先定义并计算  $D = A + BC$ , 然后求解  $Dx = b$ , 如果  $(BC)$  密集, 那么  $D$  的计算消耗  $(2pn^2)$  flops, 对  $D$  的 LU 因子分解消耗  $\frac{2}{3}n^3$  flops。总计  $2pn^2 + \frac{2}{3}n^3$  flops。
2. 也可以利用 Woodbury 公式来达到更高的求解效率, 通过(16.123)从右向左计算来求得  $x$ :
  - (a)  $z = A^{-1}b$ , 由于  $A$  是对角矩阵, 计算消耗  $n$  flops,
  - (b)  $E = I + CA^{-1}B$ ,  $2p^2n$  flops,
  - (c)  $Ew = Cz$ , 一个由  $p$  个变量构成的  $p$  个线性方程的系统。计算  $Ew$  和  $Cz$  分别消耗  $\frac{2}{3}p^3$  和  $2pn$  flops, 总计  $\frac{2}{3}p^3$  flops,
  - (d)  $x = z - A^{-1}bw$ , 其中矩阵向量乘  $Bw$  消耗  $2pn$  flops,
  - (e) 总计消耗  $wp^2n + \frac{2}{3}p^3$  flops。

不难看出, 利用 Woodbury 公式求解的效率更高。

Woodbury 公式的另一个重要应用是,  $A$  是非奇异稀疏矩阵, 并且  $B$  和  $C$  都是密集矩阵的情况。类似地, 求解(16.121)也有两种方法。

1. LU 分解法: 生成密集矩阵  $D = A + BC$ , 然后对  $D$  做 LU 因子分解, 再求解系统(16.121), 共消耗  $\frac{2}{3}n^3 + 2pn^2$  flops。

2. 也可以利用 Woodbury 公式，对其中的密集矩阵做 LU 分解，尤其是将  $A$  分解如下

- (a)  $A = P_1 L U P_2$ , 消耗  $f$  flops,
- (b)  $P_1 L U P_2 x = d$ , 消耗  $s$  flops,
- (c) 在此基础上对系统(16.121)从右向左分步计算,
- (d) 总计算消耗  $f + ps + wp^2n + \frac{2}{3}p^3$  flops。

若  $f \ll \frac{2}{3}n^3$ ,  $s \ll 2n^2$ , 那么利用 Woodbury 公式求解的效率更高。

### 扩展系统法

比起  $n$  个变量、 $n$  个方程的小系统(16.121)来,  $n+p$  个变量、 $n+p$  个方程的新系统(16.122)又称为扩展系统 (augmented system), 对应的求解方法称扩展系统法 (augmented system approach)。如果  $A$  非常稀疏并且  $p$  值相对较小, 那么我们可以对(16.123)利用直接做稀疏 LU 分解, 其求解效率将大大高于求解密集扩展系统(16.122)。具体说来

1. 对稀疏矩阵  $A$  作稀疏 LU 因子分解

$$A = P_1 L U P_2, \quad (16.125)$$

2. 对(16.123)中的密集矩阵  $(I + CA^{-1}B)$  作密集 LU 因子分解

$$(I + CA^{-1}B) = P_3 \tilde{L} \tilde{U},$$

3. 扩展系统(16.122)中的系数矩阵因此改写为

$$\begin{pmatrix} A & B \\ C & -I \end{pmatrix} = \begin{pmatrix} P_1 & 0 \\ 0 & P_3 \end{pmatrix} \begin{pmatrix} L & 0 \\ P_3^\top C P_2^\top U^{-1} & -\tilde{L} \end{pmatrix} \begin{pmatrix} U & L^{-1} P_1^\top B \\ 0 & \tilde{U} \end{pmatrix} \begin{pmatrix} P_2 & 0 \\ 0 & I \end{pmatrix}, \quad (16.126)$$

4. 将上式代回扩展系统(16.122)中, 求解。

可以证明, 扩展系统法在计算效率上等价于 Woodbury 公式法, 证明过程略。

值得指出的是, 在利用稀疏 LU 因子分解  $A$  的(16.125)过程中, 我们无法控制具体置换矩阵的形式, 这导致很可能在计算机实际求解过程中, 并不会用到(16.126)形式的因式分解, 导致计算效率降低。这是扩展系统法的不足。然而即便如此, 扩展系统法仍然是一种有竞争力的求解方法, 比起 Woodbury 公式法来说, 它更容易实施, 并且其数值解也更加稳定。

### 低秩修正矩阵

设  $A \in \mathbb{R}^{n \times n}$  非奇异。 $u, \nu \in \mathbb{R}^n$  满足  $1 + \nu^\top A u \neq 0$ 。则我们将目标扩展为, 求解由两组线性方程共同构成的系统

$$\begin{cases} Ax = b, \\ (A + u\nu^\top) \tilde{x} = b, \end{cases} \quad (16.127)$$

系统中下半段的解  $\tilde{x}$  称为  $x$  的秩一修正矩阵 (rank-one update matrix)。一旦求得  $x$  的解之后, 利用 Woodbury 公式, 我们可以很容易的计算  $\tilde{x}$ :

$$\begin{aligned} \tilde{x} &= (A + u\nu^\top)^{-1} b \\ &= \left( A^{-1} - \frac{1}{1 + \nu^\top A^{-1} u} A^{-1} u \nu^\top A^{-1} \right) b \\ &= x - \frac{\nu^\top x}{1 + \nu^\top A^{-1} u} A^{-1} u. \end{aligned}$$

算法如下

**Algorithm 16.6** (秩一修正的计算).

对  $A$  作因子分解,

计算  $x = A^{-1}b$ ,

计算  $w = A^{-1}u$ ,

计算  $\tilde{x} = x - \frac{\nu^\top x}{1+\nu^\top w}w$ ,

总共消耗  $(f + 2s)$  flops。

### 16.C.9 求解欠定线性系统

欠定线性方程系统 (underdetermined linear equation system) 是指诸如

$$Ax = b, \quad A \in \mathbb{R}^{p \times n}, p < n. \quad (16.128)$$

假设  $\text{rank } A = p$ , 从而对于任意给定的  $b$ , 系统至少存在一个解  $x$ 。

在一些情况下, 我们只需要找到欠定系统的一个特定解  $\hat{x}$  就可以了。然而在某些情况下, 我们需要找到全部的解:

$$\{x | Ax = b\} = \{Fz + \hat{x} | z \in \mathbb{R}^{n-p}\}, \quad (16.129)$$

其中矩阵  $F$  的列是  $A$  的零空间 (null space) 所对应的的基。

如果  $A$  的前  $p$  列彼此不相关, 那么求解欠定线性系统(16.129)的方法如下: 将  $A$  分解为两个块, 前  $n \times p$  设为  $A_1$ , 则系统改写为

$$Ax = \begin{pmatrix} \underbrace{A_1}_{(p \times p)} & \underbrace{A_2}_{(p \times (n-p))} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = A_1 x_1 + A_2 x_2 = b,$$

其中  $A_1$  非奇异。由此可得  $x_1 \in \mathbb{R}^n$  的值

$$x_1 = A_1^{-1}b - A_1^{-1}A_2x_2,$$

在此基础上求解系统: 取  $\hat{x}_2 = 0, \hat{x}_1 = A_1^{-1}b$ , 进行数值运算, 计算成本等同于求解一个由  $p$  个线性方程  $a_1 \hat{x}_1 = b$  构成的系统, 总共消耗  $(f + s)$  flops, 其中  $f$  用于对  $A_1$  作因式分解,  $s$  用于求解  $A_1 x = d$ 。

另一种求解思路是, 设  $x_2 \in \mathbb{R}^{n-p}$  作为自由参数, 将  $Ax = b$  的通解表示为

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -A_1^{-1}A_2 \\ I \end{pmatrix} x_2 + \begin{pmatrix} A_1^{-1}b \\ 0 \end{pmatrix},$$

由此可得(16.129)中的参数  $F$  和近似解  $\hat{x}$

$$F = \begin{pmatrix} -A_1^{-1}A_2 \\ I \end{pmatrix}, \quad \hat{x} = \begin{pmatrix} A_1^{-1}b \\ 0 \end{pmatrix}, \quad \Sigma = f + s(n - p + 1) \text{ flops.}$$

现在进一步放宽前提假设条件, 设  $A$  的前  $p$  列不必一定不相关。由于  $\text{rank } A = p$ , 我们从  $A$  的  $n$  列中选出不相关的  $p$  列, 置换到前面来, 进而用前述方法做因式分解 + 求解。换句话说, 建立一个置换矩阵  $P$ , 使得  $\tilde{A} = AP$  矩阵的前  $p$  列不相关, 进而对  $\tilde{A}$  做分解

$$\tilde{A} = AP = [A_1 A_2],$$

其中  $A_1$  可逆。对应地，定义  $\tilde{x} = P^T x$ ，将(16.129)改写为

$$\tilde{A}\tilde{x} = b, \quad (16.130)$$

进而  $\tilde{x}$  为

$$\tilde{x} = \begin{pmatrix} -A_1^{-1}A_2 \\ I \end{pmatrix} \tilde{x}_2 + \begin{pmatrix} A_1^{-1}b \\ 0 \end{pmatrix}.$$

在此基础上，求得通解  $x$

$$x = P\tilde{x} = P \begin{pmatrix} -A_1^{-1}A_2 \\ I \end{pmatrix} z + P \begin{pmatrix} A_1^{-1}b \\ 0 \end{pmatrix},$$

其中  $z \in \mathbb{R}^{n-p}$  是自由参数。

当我们能在  $A$  中识别出一个非奇异或是容易取逆的分块 (如一个对角元素均不为零的对角矩阵) 时，这个方法能提供较高的计算效率。

### 用 QR 因子分解法求解欠定系统

如果  $C \in \mathbb{R}^{n \times p}$ ,  $p \leq n$ ,  $\text{rank } C = p$ , 可以对  $C$  作因子分解如下

$$C = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} \begin{pmatrix} R \\ 0 \end{pmatrix}, \quad (16.131)$$

其中  $Q_1 \in \mathbb{R}^{n \times p}$ ,  $Q_2 \in \mathbb{R}^{n \times (n-p)}$  满足

$$\begin{aligned} Q_1^\top Q_1 &= I, \\ Q_2^\top Q_2 &= I, \\ Q_1^\top Q_2 &= 0, \end{aligned}$$

$R \in \mathbb{R}^{p \times p}$  是一个对角元素非零的三交矩阵。这称为对  $C$  的 QR 因子分解。共消耗  $2p^2(n - \frac{p}{3})$  flops。

QZ 分解法可以用来求解欠定线性方程系统(16.128)，设

$$A^\top = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} \begin{pmatrix} R \\ 0 \end{pmatrix}$$

是一个对  $A^\top$  的 QR 因子分解，代回(16.128)得

$$A\hat{x} = R^\top Q_1^\top Q_1 R^{-\top} b = b,$$

此外  $Q_2$  的列由  $A$  的零空间的基构成，那么系统的完全解为

$$\{x = \hat{x} + Q_2 z \mid z \in \mathbb{R}^{n-p}\}.$$

对于欠定线性方程系统，最常见的求解方法是 QR 因子分解。这种方法的一个不足是，对稀疏矩阵难以处理：哪怕  $C$  非常稀疏， $Q$  也往往是密集的。

### 用 LU 因子分解法求解欠定系统

对于稀疏系数矩阵的欠定线性系统，常用 LU 因子分解法来求解。设  $C \in \mathbb{R}^{n \times p}$ ,  $\text{rank } C = p$ , 那么可以将  $C$  作 LU 因子分解如下

$$C = P L U,$$

其中置换矩阵  $P \in \mathbb{R}^{n \times n}$ ,  $L \in \mathbb{R}^{n \times p}$ ,  $U \in \mathbb{R}^{p \times n}$  分别为单位下三角矩阵和非奇异上三角矩阵。在未对  $C$  的结构做进一步明确之前，计算消耗为  $\frac{2}{3}p^3 + p^2(n - p)$  flops。

若  $C$  是系数矩阵，那么  $C$  的 LU 因子分解常常包含行、列置换，即可将  $C$  因子分解为

$$C = P_1 L U P_2, \quad P_1 \in \mathbb{R}^{n \times n}, P_2 \in \mathbb{R}^{p \times p}.$$

对稀疏矩阵作 LU 分解，比起对密集矩阵的 LU 分解，计算效率提升显著。

用稀疏矩阵的 LU 分解来求解欠定方程系统(16.128)，设  $A^\top$  可因式分解为

$$A^\top = P L U,$$

那么将  $L$  分块

$$L = \begin{pmatrix} \underbrace{L_1}_{(p \times p)} \\ \underbrace{L_2}_{((n-p) \times p)} \end{pmatrix}.$$

不难证得，解的形式为(16.129)，其中参数  $F, \hat{x}$  为

$$F = P \begin{pmatrix} -L_1^{-\top} L_2^\top \\ I \end{pmatrix}, \quad \hat{x} = P \begin{pmatrix} L_1^{-\top} U^{-\top} b \\ 0 \end{pmatrix}.$$

## 16.D 大数法则

介绍弱大数法则、强大数法则和 (强) 均匀大数法则。

### 16.D.1 弱大数法则

大数法则 (law of large numbers, LLN) 可以追溯至概率论的萌芽时期：

- 伯努利 (James Bernoulli) 在 1700 年前后的随机抛硬币实验中发现。但直到死后，才在 1713 年的遗著中出版，为后人所知。
- 1800 年泊松 (Poisson) 将扩展为彼此独立事件下的抛硬币问题。
- 1866 年切比雪夫 (Chebyshev) 提出切比雪夫不等式，进一步扩展到含有独立随机变量的任意过程及其二阶矩。
- 切比雪夫的学生马尔科夫 (Markov) 进一步扩展到相互关联的随机变量的情况。

我们先来介绍马尔科夫不等式。

**Theorem 16.1** (马尔科夫不等式). 对一组连续的非负随机变量  $X$  有

$$P(X \geq t) \leq \frac{EX}{t}, \quad \forall t > 0. \tag{16.132}$$

证明. 连续版的证明:

$$\begin{aligned}
 EX &= \int_{-\infty}^{\infty} xf_X(x) dx \\
 &= \int_0^{\infty} xf_X(x) dx \quad (\text{由于 } X \text{ 非负}) \\
 &\geq \int_t^{\infty} xf_X(x) dx \\
 &\geq \int_t^{\infty} tf_X(x) dx \quad (\text{在求积域中 } x > t) \\
 &= t \int_t^{\infty} f_X(x) dx \\
 &= tP(X \geq t), \quad \forall t > 0
 \end{aligned}$$

$$\hookrightarrow P(X \geq t) \leq \frac{EX}{t}, \quad \forall t > 0.$$

离散版的证明。对于  $t > 0$ , 有

$$X \geq X1_{[X \geq t]} \geq t1_{[X \geq t]},$$

进而有期望值的单调性有

$$EX \geq tP(X \geq t).$$

□

由马尔科夫不等式可以得到两个重要结论, 分别是切比雪夫不等式和切尔诺夫边界。

**Theorem 16.2** (切比雪夫不等式). 如果  $X$  是非负随机变量, 那么有

$$P(|X - EX| \geq t) \leq \frac{Var(X)}{t^2}, \quad \forall t > 0. \quad (16.133)$$

证明. 简化表述, 定义  $Y = (X - EX)^2$ 。由  $X$  的假设条件可得  $Y$  也是一个非负随机变量。那么可将马尔科夫不等式(16.132)应用到  $Y$  上

$$P(Y \geq t^2) \leq \frac{EY}{t^2}, \quad \forall t > 0.$$

此外我们有

$$\begin{aligned}
 EY &= E(X - EX)^2 = Var(X), \\
 P(Y \geq t^2) &= P((X - EX)^2 \geq t^2) = P(|X - EX| \geq t),
 \end{aligned}$$

代回上式中, 有

$$P(|X - EX| \geq t) \leq \frac{Var(X)}{t^2}, \quad \forall t > 0.$$

□

由切比雪夫不等式可得, 随机变量  $X$  和它期望值  $EX$  之间的距离不可能无限大, 其上界由方差  $Var(X)$  予以限制。这就引出切尔诺夫边界的概念。

**Theorem 16.3** (切尔诺夫法). 切尔诺夫法 (*Chernoff method*) 将马尔科夫不等式(16.132)中的非负随机变量  $X$  替换为  $\exp(sX)$ , 可得

$$P(\exp(sX) \geq \exp(st)) \leq \frac{E \exp(sX)}{\exp(st)}. \quad (16.134)$$

如果  $X$  是由  $n$  个 *i.i.d.* 随机变量  $X_1, \dots, X_n$  组成的向量, 那么样本均值  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ , 总体均值  $EX$ , 方差  $\frac{Var(X)}{n}$ 。由切比雪夫不等式(16.133)得

$$P(|\bar{X}_n - EX| \geq \epsilon) = \frac{Var(X)}{n\epsilon^2}, \quad \forall \epsilon > 0. \quad (16.135)$$

进而我们得到弱大数法则 (weak law of large numbers): (16.135) 在  $n \rightarrow \infty$  的情况下变为

$$\left[ \lim_{n \rightarrow \infty} P(|\bar{X}_n - EX| \geq \epsilon) = 0, \quad \forall \epsilon > 0, \right] \quad (16.136)$$

随着样本数不断增大直至正无穷, 样本的期望值以概率 1 逐渐收敛至全体的期望值。

回到掷硬币的问题, 即二进制随机变量  $Bin(p)$ , 满足

$$P(X = 1) = 1 - P(X = 0) \equiv p, \quad (16.137)$$

其对应的切比雪夫边界 (Chebyshev bound) 为

$$P(|\bar{X}_n - p| \geq \epsilon) \leq \frac{p(1-p)}{n\epsilon^2}, \quad \forall \epsilon > 0,$$

也就是说在观察到的样本中, 事件  $X = 1$ (头像朝上) 发生的概率, 逐渐收敛到全体的真实概率  $EX = 1$ (头像朝上)。

## 16.D.2 强大数法则

弱大数法则用切比雪夫不等式和切尔诺夫方法来确定概率的上下界, 其本质上是利用观测到随机变量  $X$  和全体期望值  $EX$  之间的方差  $Var(x)$  来描述边界的。这种方法并不理想: 方差相对于  $X$  的变化是有限的。

我们用矩生成方程 (moment generating function) 来进行扩展, 定义为  $M_X(\lambda)$

$$M_X(\lambda) = E[\exp(\lambda X)], \quad (16.138)$$

不难看出, 随着  $\lambda$  值的变化,  $M_X(\lambda)$  甚至可以无限大。

利用 MGF 可以计算切尔诺夫边界。

**Theorem 16.4** (切尔诺夫边界). 非负的随机变量  $X$ , 对于任意  $t \geq 0$ , 其分布的上、下界分别为

$$P(X \geq EX + t) \leq \min_{\{\lambda \geq 0\}} E\{\exp[\lambda(X - EX)]\} \exp(-\lambda t) = \min_{\{\lambda \geq 0\}} M_{X-EX}(\lambda) \exp(-\lambda t), \quad (16.139)$$

$$P(X \leq EX + t) \leq \min_{\{\lambda \geq 0\}} E\{\exp[\lambda(EX - X)]\} \exp(-\lambda t) = \min_{\{\lambda \geq 0\}} M_{EX-X}(\lambda) \exp(-\lambda t). \quad (16.140)$$

证明. 对切尔诺夫上界(16.139)的证明。对于任意  $\lambda > 0$ , 已知

$$\{X \geq EX + t\} \iff \{\exp(\lambda X) \geq \exp[\lambda(EX + t)]\} \iff \{\exp[\lambda(X - EX)] \geq \exp(\lambda t)\},$$

其对应的马尔科夫不等式由(16.132)写为

$$P(X - EX \geq t) = P(\exp[\lambda(X - EX)] \geq \exp(\lambda t)).$$

切尔诺夫下界(16.140)也可用类似的方法证得。 □

这样，在给定 MGF 具体形式的基础上，我们可以根据切尔诺夫边界 (Theorem 16.4) 计算出概率分布的上、下界。最常见、也是最适合的形式为

$$M_X(\lambda) = E[\exp(\lambda X)] \leq \exp\left(\frac{C^2\lambda^2}{2}\right), \quad \forall \lambda \in \mathbb{R}, \quad (16.141)$$

其中常数  $C \in \mathbb{R}$  是一个与  $X$  的分布有关的常数。举两个例子。一个例子是正态分布  $X \sim \mathcal{N}(0, \sigma^2)$ ，对应 MGF

$$M_X(\lambda) = E[\exp(\lambda X)] \leq \exp\left(\frac{\lambda^2\sigma^2}{2}\right).$$

另一个例子是拉德马赫随机变量 (Radmacher random variables)，又称随机符号变量 (random sign variable)。

**Proposition 16.1** (拉德马赫随机变量的切尔诺夫边界). 设拉德马赫随机变量，或称符号随机变量  $S$  满足关系

$$S = \begin{cases} 1 & \text{概率 } \frac{1}{2}, \\ -1 & \text{概率 } \frac{1}{2}, \end{cases}$$

那么它的 MGF 可以表示为

$$M_S(\lambda) = E[\exp(\lambda S)] \leq \exp\left(\frac{\lambda^2}{2}\right), \quad \forall \lambda \in \mathbb{R}, \quad (16.142)$$

在此基础上，其切尔诺夫边界为

$$P(X \geq t) \leq \exp\left(-\frac{t^2}{2n}\right). \quad (16.143)$$

证明. 首先来看  $S$  的 MGF 计算过程。对  $\exp(x)$  作泰勒级数展开

$$\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

$S$  的取值定义，有

$$ES^k = \begin{cases} 0 & k \text{ 是奇数,} \\ 1 & k \text{ 是偶数.} \end{cases}$$

对应地有  $S$  的 MGF

$$\begin{aligned} M_S(\lambda) &= E[\exp(\lambda S)] \\ &= \sum_{k=0}^{\infty} \frac{\lambda^k E[S^k]}{k!} \\ &= \sum_{k=0,2,4,\dots}^{\infty} \frac{\lambda^k}{k!} \\ &= \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!}. \end{aligned}$$

由于

$$(2k)! \geq (2^k k!)^2, \quad \forall k = 0, 1, 2, \dots$$

我们进一步有

$$\begin{aligned} M_s(\lambda) &= E \exp(\lambda S) = \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!} \\ &\leq \sum_{k=0}^{\infty} \frac{(\lambda^2)^k}{2^k k!} \\ &\leq \sum_{k=0}^{\infty} \left(\frac{\lambda^2}{2}\right)^k \frac{1}{k!} \\ &= \exp\left(\frac{\lambda^2}{2}\right), \end{aligned}$$

即(16.142)成立。

在此基础上，假定  $S$  是一组随机符号变量的和

$$S = \sum_{i=1}^n S_i, \quad S_i \in \{\pm 1\},$$

由此我们可得其期望值  $ES = 0$ ，代入切尔诺夫边界条件(16.139)可得

$$\begin{aligned} P(S \geq t) &\leq \min_{\{\lambda \geq 0\}} \exp(-\lambda t) E[\exp(\lambda S)] \\ &= \min_{\{\lambda \geq 0\}} \exp(-\lambda t) E[\exp(\lambda S_1)^n] \\ &= \min_{\{\lambda \geq 0\}} \exp(-\lambda t) \exp\left(\frac{n\lambda^2}{2}\right), \end{aligned}$$

最小化求解，对应最优系数值  $\lambda^*$

$$\lambda^* = \operatorname{argmin}_{\{\lambda \geq 0\}} \left( \frac{n\lambda^2}{2} - \lambda t \right) = \frac{t}{n},$$

代回上式可得

$$P(S \geq t) \leq \exp\left(-\frac{t^2}{2n}\right),$$

即(16.143)成立。  $\square$

若进一步假设  $t = (2n \log \frac{1}{\delta})^{\frac{1}{2}}$ ，有

$$P\left(S \geq \left(2n \log \frac{1}{\delta}\right)^{\frac{1}{2}}\right) \leq \delta,$$

换句话说存在相当高的概率满足  $S = \sum_{i=1}^n S_i = \mathcal{O}(\sqrt{n})$ ：根据切尔诺夫边界， $n$  个独立随机符号之和的分布基本上绝不大于  $\mathcal{O}(\sqrt{n})$ 。

现在将变量形式从随机符号变量扩展到一般情况：当一组有界的随机变量之和太大或太小时，其和的概率的边界可用霍夫丁不等式 (Hoeffding inequality) 计算得到。在介绍霍夫丁不等式之前，首先需要介绍霍夫丁引理 (Hoeffding lemma)。

**Lemma 16.9** (霍夫丁引理). 设  $X \in [a, b]$  是一个有界的随机变量，那么  $(X - EX)$  的 MGF 的期望值有界，边界为

$$EM_{X-EX}(\lambda) = E\{\exp[\lambda(X - EX)]\} \leq \exp\left[\frac{\lambda^2(b-a)^2}{8}\right], \quad \forall \lambda \in \mathbb{R}, \quad (16.144)$$

称为霍夫丁引理 (Hoeffding lemma)。

证明。我们作一个弱形式的证明（在弱形式证明中，RHS 分母由 8 变为 2）。证明需要两处背景知识，一是随机符号的 MGF 边界（Proposition 16.1），二是延森不等式。

延森不等式（Jensen inequality）是指，如果  $f : \mathbb{R} \mapsto \mathbb{R}$  是一个凸方程（convex function），那么有

$$f[EX] \leq Ef(x). \quad (16.145)$$

（延森不等式的证明略）<sup>10</sup>。

在此基础上，我们用对称化（symmetrization）这一概率论中的常用技术来给出证明。首先设  $X'$  是一个与  $X$  有相同分布的独立拷贝，满足

$$X' \in [a, b], \quad E_{X'} X' = E_X X, \quad \text{并且 } X \text{ 和 } X' \text{ 彼此独立.}$$

那么我们有

$$\begin{aligned} E_X \{\exp[\lambda(X - E_X X)]\} &= E_X \{\exp[\lambda(X - E_{X'} X')]\} \\ &\leq E_X E_{X'} \{\exp[\lambda(X - X')]\}, \end{aligned}$$

最后一行由严森不等式推得。上式简化为

$$E_X \{\exp[\lambda(X - EX)]\} \leq \underbrace{E_{X, X'} \{\exp[\lambda(X - X')]\}}_{=: \mathcal{A}}. \quad (16.146)$$

由  $X$  和  $X'$  的性质，根据对称化， $(X - X')$  围绕着 0 对称。如果  $S \in \{-1, 1\}$  是一个随机符号变量，那么  $S(X - X')$  的分布与  $(X - X')$  相同。则 RHS 进一步改写为

$$\begin{aligned} \mathcal{A} &:= E_{X, X'} \{\exp[\lambda(X - X')]\} = E_{X, X', S} \exp[\lambda S(X - X')] \\ &= E_{X, X'} \left\{ \underbrace{E_S [\exp[\lambda S(X - X')]}_{=: \mathcal{B}} | X, X' \right\}, \end{aligned}$$

根据随机符号变量的 MGF 不等式（16.142），上式 RHS 中  $\mathcal{B}$  满足

$$\mathcal{B} := E_S [\exp[\lambda S(X - X')] | X, X'] \leq \exp \left[ \frac{\lambda^2}{2} (X - X')^2 \right],$$

即（16.146）改写为

$$E_X \{\exp[\lambda(X - EX)]\} \leq E_{X, X'} \exp \left[ \frac{\lambda^2}{2} (X - X')^2 \right]. \quad (16.147)$$

此外根据假定条件有

$$\{|X - X'| \leq (b - a)\} \iff \{(X - X')^2 \leq (b - a)^2\},$$

代回（16.147）得

$$\begin{aligned} E_X \{\exp[\lambda(X - EX)]\} &\leq E_{X, X'} \exp \left[ \frac{\lambda^2}{2} (X - X')^2 \right] \\ &= \exp \left[ \frac{\lambda^2 (b - a)^2}{2} \right], \end{aligned}$$

证得霍夫丁引理（16.144）成立。 □

<sup>10</sup>一个最简单的记忆延森不等式的方法是考虑  $f(t) = t^2$ ，此外结合

$$E[X] = 0 \implies f[EX] = 0,$$

，通常来说  $E x^2 > 0$ 。 $f(t) = \exp(t)$  和  $f(t) = \exp(-t)$  都是凸方程。

在此基础上，来看霍夫丁不等式。

**Theorem 16.5** (霍夫丁不等式). 设一组独立有界随机变量  $Z_1, \dots, Z_n$ , 边界  $Z_i \in [a, b], \forall i = 1, \dots, n, -\infty < a < b < \infty$ 。那么其分布的边界满足

$$P\left(\frac{1}{n} \sum_{i=1}^n (X_i - EX_i) \geq t\right) \leq \exp\left[-\frac{2nt^2}{(b-a)^2}\right], \quad (16.148)$$

$$P\left(\frac{1}{n} \sum_{i=1}^n (X_i - EX_i) \leq -t\right) \leq \exp\left[-\frac{2nt^2}{(b-a)^2}\right], \quad \forall t > 0, \quad (16.149)$$

称霍夫丁不等式 (*hoeffding inequality*)。

证明. 来证明上界不等式条件(16.148)。结合切尔诺夫上边界条件(16.139)和霍夫丁引理式(16.144)有

$$\begin{aligned} P\left(\frac{1}{n} \sum_{i=1}^n (X_i - EX_i) \geq t\right) &= P\left(\sum_{i=1}^n (X_i - EX_i) \geq nt\right) \\ &\leq \min_{\{\lambda \geq 0\}} \exp(-\lambda nt) E\left[\exp\sum_{i=1}^n (X_i - EX_i)\right] \\ &= \min_{\{\lambda \geq 0\}} \exp(-\lambda nt) \prod_{i=1}^n E[\exp(X_i - EX_i)] \\ &\leq \min_{\{\lambda \geq 0\}} \exp(-\lambda nt) \left\{ \prod_{i=1}^n \exp\left[\frac{\lambda^2(b-a)^2}{8}\right] \right\} \\ &\leq \underbrace{\min_{\{\lambda \geq 0\}} \exp\left\{\frac{n\lambda^2(b-a)^2}{8} - \lambda nt\right\}}_{=: \mathcal{A}}. \end{aligned}$$

对  $\mathcal{A}$ , 求最小化时的系数  $\lambda^*$

$$\begin{aligned} \lambda^* &= \operatorname{argmin} \left\{ \frac{n\lambda^2(b-a)^2}{8} - \lambda nt \right\} = \frac{4t}{(b-a)^2}, \\ \hookrightarrow \mathcal{A} &= \exp\left[\frac{-2nt^2}{(b-a)^2}\right], \end{aligned}$$

因此我们有

$$P\left(\frac{1}{n} \sum_{i=1}^n (X_i - EX_i) \geq t\right) \leq \exp\left[\frac{-2nt^2}{(b-a)^2}\right],$$

上边界条件(16.148)成立。采用类似的思路, 也可证得下边界条件(16.149)成立。  $\square$

### 16.D.3 强均匀大数法则

大数法则的严谨脉络, 基本与概率论的范式变化相一致。

- 在概率论的早期阶段, 弱大数法则主要有伯努利、泊松及其相关改进如切比雪夫、马尔科夫等, 基本反映早期概率论的核心思想。
- 随后发展起来的强大数法则如 Borel, Cantelli, Kolmogorov 等, 反映了概率论进一步公理化、正规化, 在 20 世纪上半叶逐渐成为测度论的一部分。

- 概率论的更新进展，以强一致大数法则为代表，它是 1970 年代开始，概率论与统计学习相交融的产物，如万普尼克 (Vapnik)、泽凡尼杰斯 (Chervonekis) 等。

本节对强一致大数法则作介绍。先来对霍夫丁不等式 (Theorem 16.5) 做更为一般的表述，把它写成一个关于独立随机变量的方程，满足差分有界 (bounded differences)：

**Definition 16.2** (差分有界). 设一个  $n$  维标准向量空间  $\mathbb{R}^n$  中有一个标准向量 (*canonical vector*) 的集合  $S \subset \mathbb{R}^n$ ，集合中  $e_i \in \mathbb{R}^n$  表示第  $i$  个标准向量，向量  $e_i$  中第  $i$  个元素的值为 1，其余均为 0。对于一个方程  $h : S \mapsto \mathbb{R}$ ，如果满足

$$|h(x) - h(x + te_i)| \leq c_i, \quad i = 1, \dots, n, \forall x \in S, \forall t \in \mathbb{R}, (x + te_i) \in S, \quad (16.150)$$

那么我们称方程  $h$  的差分有界 (*bounded differences*)，即是说，随着  $x$  的变化， $h$  值的变化不会超过  $c_i$ 。

由差分有界可进一步引出 McDiarmid 不等式

**Theorem 16.6** (McDiarmid 不等式). 设方程  $h$  差分有界，则有

$$P(|h(X_1, X_2, \dots, X_n) - Eh| \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right). \quad (16.151)$$

证明. 设

$$h - Eh = \sum_{i=1}^n Z_i,$$

其中

$$Z_i(X_1, \dots, X_i) = E[h|X_1, \dots, X_i] - E[h|X_1, \dots, X_{i-1}].$$

这一组  $\{Z_i\}$  的均值为 0，(几乎总是) 有界，对应区间  $[L_i, U_i]$ 。由此可见，上下边界  $L_i, U_i$  都之和  $X_1, \dots, X_{i-1}$  有关，并且根据  $h$  差分有界的假定我们有  $U_i - L_i \leq c_i$ 。那么利用切尔诺夫法 Theorem 16.3 和霍夫丁引理 Lemma 16.9 可得，

$$\begin{aligned} P(h - Eh \geq t) &= P(\exp[s(h - Eh)] \geq \exp(st)) \\ &\stackrel{\min_{\{s \geq 0\}}}{\leq} \frac{E \exp[s(h - Eh)]}{\exp(st)} \\ &= \min_{\{s \geq 0\}} \exp(-st) E \left[ \exp \left( s \sum_{i=1}^n Z_i \right) \right] \\ &= \min_{\{s \geq 0\}} \exp(-st) E \left\{ s \sum_{i=1}^{n-1} Z_i E[\exp(sZ_n|X_1, \dots, X_{n-1})] \right\} \\ &= \min_{\{s \geq 0\}} \exp(-st) \exp \left[ s^2 \sum_{i=1}^n \frac{c_i^2}{8} \right] \end{aligned} \quad (16.152)$$

求解最小化

$$s^* = \operatorname{argmin}_{\{s \geq 0\}} \left[ s^2 \sum_{i=1}^n \frac{c_i^2}{8} - st \right] = \frac{4t}{\sum_{i=1}^n c_i},$$

代回(16.152)

$$P(h - Eh \geq t) \leq \exp(-s^*t) \exp \left[ s^{*,2} \sum_{i=1}^n \frac{c_i^2}{8} \right] = \exp \left\{ -\frac{2t^2}{\sum_{i=1}^n c_i} \right\},$$

即 McDiarmid 不等式(16.151)成立。  $\square$

**Corollary 16.1.** 当  $h = \sum_{i=1}^n X_i$  时，McDiarmid 不等式(16.151)变为一种特殊情况即霍夫丁不等式(16.148)-(16.149)。



# **第四部分**

## **待分类**



# 第十七章 最大似然估计

The maximum-likelihood procedure in any problem is what you are most likely to do if you don't know any statistics.

Harrison H. Barrett  
Foundations of Image Science

## 17.1 线性模型

设一组含有  $n$  个观察数据的随机变量  $\{y_i\}$ ,  $i = 1, \dots, n$ 。假定其中每个观察符合正态分布  $y_i \sim \mathcal{N}(\mu_i, \sigma^2)$ 。由设定可见, 对于  $i, j \in [1, n]$ ,  $i \neq j$ , 均值也许不同, 但方差相同。

正态分布 (normal distribution)又称高斯分布 (Gaussian distribution), 如  $y_i$  的概率密度方程 (probability density function, PDF)  $f(y_i)$  定义为

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{(y_i - \mu_i)^2}{\sigma^2}\right], \quad (17.1)$$

均值为 0 方差为 1 的高斯分布, 见图17.1所示。

```
1 # 正态分布的概率密度方程(均值=0, 方差=1)
2 x <- seq(-4,4,length=100)
3 hx <- dnorm(x)
4 plot(hx ~ x, type="l", lty = 1, col="blue",
5      xlab="y", ylab="Density",
6      main = "Probability Density Function (normal distribution)")
```

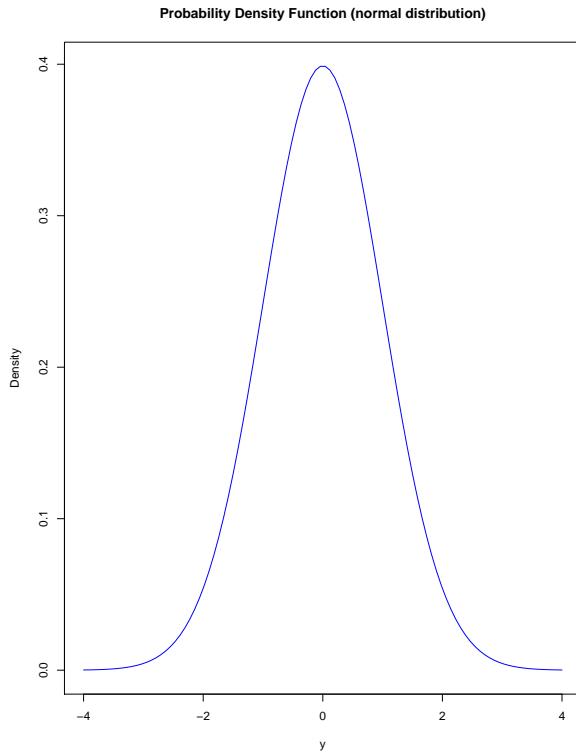
现在引入额外的假定, 对于全部  $i, j \in [0, n]$ ,  $i \neq j$ , 设观测数据  $y_i$  和  $y_j$  相互独 (mutually independent), 即  $\text{cov}(y_i, y_j) = 0$ 。这使得我们能够计算观测数据集中全部数据的联合分布 (joint distribution), 即  $\prod_{i=1}^n f(y_i)$ , 进而勾勒出似然方程, 用于进一步的估计和检测。

把  $\{y_i\}_{i=1}^n$  表示为一个  $(n \times 1)$  的列向量  $Y$ , 对应均值  $EY = \mu$ , 方差协方差矩阵  $\text{var}(Y) = \sigma^2 I$ , 其中  $I$  是单位矩阵。列向量  $\mu = \{\mu_i\}_{i=1}^n$ 。由于  $n$  个观测变量  $y_i$  彼此不相关且方差相同,  $\text{var}(Y)$  满足以下特征: 对角元素全是  $\sigma^2$ , 非对角元素全是 0。进而  $Y$  呈多元正态分布 (multivariate normal distribution)

$$Y \sim \mathcal{N}_n(\mu, \sigma^2 I). \quad (17.2)$$

来看(17.2)的模型。假定  $Y = y_1, y_2, \dots, y_n$  与某组预测变量  $X = x_1, x_2, \dots, x_n$  有关, 进一步

图 17.1: 正态分布的概率密度方程



说, 第  $i$  个观测变量  $y_i$  的期望值  $\mu_i$ , 与  $x_{i1}, x_{i2}, \dots, x_{ip}$  有关, 呈线性关系, 满足

$$\mu_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \iff \mu_i = x_i^\top \beta, \quad (17.3)$$

其中  $x_i^\top$  是由  $p$  个预测变量  $x_{i1}, x_{i2}, \dots, x_{ip}$  组成的行向量。待求解的未知系数  $\beta = \beta_1, \dots, \beta_p$  称回归系数。将全部  $n$  个(17.3)加总可得

$$\underset{(n \times 1)}{\mu} = \underset{(n \times p)}{X} \underset{(p \times 1)}{\beta}, \quad (17.4)$$

我们常将解释变量的矩阵  $X$  称为模型矩阵 (model matrix)或设计矩阵 (design matrix)。对应地,  $X\beta$  称线性预测子 (linear predictor)。

最简单的线性模型可以假定每个观测数据的期望值都相同  $\mu_i = \mu \forall i$ , 称零模型 (null model)。另一个极端是  $\mu_i \neq \mu_j \forall i \neq j$ , 称饱和模型 (saturated model), 此时观测数据量  $n$  越大, 待估计的线性系数  $\beta$  数量就越多 ( $p \times n$ )。

零模型和饱和模型是两个极端。现实应用中常取折中, 致力于分析导致线性预测子  $X\beta$  产生结构差异的系统性原因, 进而分析观测数据  $y$  和均值  $\mu$  之间的非结构性差异 (或称随机差异), 用误差项来表示。

## 17.2 参数估计

来看模型  $\mu_i = x_i^\top \beta$ 。问题: 如何估计参数  $\beta, \sigma^2$ ?

### 17.2.1 回归系数的估计

我们的求解思路是，建立似然方程 (likelihood function, LHF)，选取使 (对数) 似然方程最大化的参数值。如果观测数据之间互相独立，那么 LHF 是一组正态 PDF (17.1) 的乘积

$$\log L(\beta, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\sigma^2}, \quad (17.5)$$

$\mu_i$  如(17.3)所定义。RHS 中可定义残差平方和 (residual sum of squares, RSS)

$$RSS(\beta) = \sum_{i=1}^n (y_i - \mu_i)^2 = (y - X\beta)^\top (y - X\beta). \quad (17.6)$$

不难看出，在给定  $\sigma^2$  值不变的情况下，最佳系数  $\hat{\beta}$  的选取符合

$$\hat{\beta} = \underset{\beta}{\operatorname{argmax}} \log L(\beta, \sigma^2) = \underset{\beta}{\operatorname{argmin}} \frac{RSS(\beta)}{\sigma^2}, \quad (17.7)$$

也即，我们的目标是选取合适的  $\beta = \hat{\beta}$ ，使对应的模拟值  $\mu_i$  尽可能接近实际观测值  $y_i$ 。

求解  $\operatorname{argmin} RSS(\beta)$  等价于求解  $RSS(\hat{\beta}) = 0$ ，即

$$y - X\hat{\beta} = 0 \Rightarrow y = X\hat{\beta} \Rightarrow X^\top y = X^\top X\hat{\beta}$$

如果模型矩阵  $X$  列满秩，那么  $X^\top X$  也是满秩的，因而  $X^\top X$  可逆。由此可得线性系数  $\hat{\beta}$  的 OLS 估计式 (同时也是 MLE 估计式)

$$\hat{\beta} = (X^\top X)^{-1} X^\top y. \quad (17.8)$$

我们将这个方程称为正规方程 (normal equation)。

反之，如果  $X$  不是列满秩的，可以计算  $(X^\top X)^\dagger$ ，称为伪逆矩阵 (第11.B节)。但比起这种相对复杂的计算来，还是直接删除  $X$  中的冗余列更加方便。当前大多数主流统计软件都足够只能，可以自动识别并删除冗余列。

求解正规方程(17.8)需要借助数值方法。数值方法有多种，常见的如

- 从  $(X^\top X)$  入手，方法如高斯消元法 (第16.C.7节)、Cholesky 分解 (第16.C.5节) 等。
- 对模型矩阵  $X$  做因子分解，如
  - Householder reflections
  - Givens rotations
  - Gram-Schmidt 正交 (第12.B.2节)，等

结合大多数主流数值计算软件，可以执行上述数值运算。

基于(17.7)，利用最小化 RSS 方法测得的系数  $\hat{\beta}$ ，是一个不依赖于  $\sigma^2$  的值——方差值是事先给定的。因此我们称  $\hat{\beta}$  为最大似然法的全局最大解 (global maximum)。

对于零模型的情况： $X$  是一组 1 构成的向量； $(X^\top X) = n$  和  $X^\top y = \sum_{i=1}^n y_i$  是两个标量； $\hat{\beta} = \bar{y}$  是样本的均值。这就是说，计算出的样本均值，可堪称是在线性模型中做最大似然估计的一个最简单的例子。

关于 MLE  $\hat{\beta}$ ，有以下几个有趣的性质。

#### 1. BLUE

- (a) 如果模型设定正确, 即从(弱的)意义上说, 给定  $x_i$  的情况下, 观测  $y_i$  的期望值  $\mu_i$  就等于  $x_i^\top \beta$ , 此时的 OLS 估计  $\hat{\beta}$  是个无偏估计 (unbiased estimator), OLS 估计  $\hat{\beta}$  的期望值就等于真实参数值  $\beta$

$$E\hat{\beta} = \beta. \quad (17.9)$$

- (b) 如果观测数据之间彼此不相关  $\text{cov}(y_i, y_j) = 0, \forall i \neq j$ , 并且同方差  $\sigma_i^2 = \sigma_j^2 = \sigma^2$ 。那么, 一方面根据(17.6)-(17.7),  $\hat{\beta}$  是一个关于  $y$  的线性方程, 另一方面根据假定, 观测数据集合  $Y$  的方差协方差矩阵  $\text{var}(Y) = \sigma^2 I$ , 那么 OLS 估计  $\hat{\beta}$  的方差协方差矩阵为

$$\text{var}(\hat{\beta}) = (X^\top X)^{-1} \sigma^2, \quad (17.10)$$

因此, 基于观测数据, 构建线性方程模型所做的全部无偏 OLS 估计  $\{\beta\}$  中, MLE 估计  $\hat{\beta}$  是最佳线性无偏估计 (best linear unbiased estimator, BLUE)。对于某一给定的观测样本, 由于再没有其他无偏估计的方差会低于  $\text{var}(\hat{\beta})$ , 我们称 OLS 估计  $\hat{\beta}$  是有效估计 (efficient estimator)。

2. OLS 估计  $\hat{\beta}$  在大样本观测集合中的抽样分布, 接近于多元正态分布, 其均值和方差如(17.2)所定义, 满足

$$\hat{\beta} \sim \mathcal{N}_p\left(\beta, (X^\top X)^{-1} \sigma^2\right).$$

3. 将前两个性质代入零模型, 可见样本均值  $\bar{y}$  是对  $\mu$  的无偏估计;  $\bar{y}$  的方差为  $\frac{\sigma^2}{n}$ , 在大样本下近似正态分布。

4. 前三条性质的生成, 依赖于对观测数据的均值、方差协方差的最高到二阶矩条件的假设, 包括

$$EY = X\beta, \quad \text{var}(Y) = \sigma^2 I.$$

5. 基于观测数据的联合正态分布假定, 所测得的 OLS 估计  $\hat{\beta}$  也是 MLE。如果  $Y \sim \mathcal{N}_p(X\beta, \sigma^2 I)$ , 那么  $\hat{\beta}$  的样本分布恰好也是多元正态分布, 对应的、方差也是

$$\hat{\beta} \sim \mathcal{N}_p\left(\beta, (X^\top X)^{-1} \sigma^2\right).$$

需要指出的是, 上属性质虽然重要, 但不应被过分夸大: 只有在小样本数据的统计推断中, 才需要假定观测数据是正态分布的。而真正重要的假设其实是观测数据之间彼此不相关, 同方差: 这些对于大样本数据下的统计推断至关重要。

### 17.2.2 方差的估计

将求得的 OLS 估计  $\hat{\beta}$  (17.8)代回 log LHF (17.5), 可得一个关于方差  $\sigma^2$  的最大(对数)似然方程, 又称描述似然方程 (profile likelihood function)

$$\log L(\sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2} \frac{\text{RSS}(\hat{\beta})}{\sigma^2}. \quad (17.11)$$

类似地,  $\hat{\sigma}^2 = \underset{\sigma^2}{\text{argmax}} \log L(\sigma^2)$ , 等价于求解  $\frac{\partial \log L(\sigma^2)}{\partial \sigma^2} = 0$ , 求得的值为方差的 MLE

$$\hat{\sigma}^2 = \frac{\text{RSS}(\hat{\beta})}{n},$$

需要指出的是， $\hat{\sigma}^2$  是有偏估计；我们可以将分母用  $n - p$  代替  $n$  来变为无偏估计（类似于在估计方差时用  $n - 1$  代替  $n$ ）。

在零模型中，方差  $\sigma^2$  的估计是样本方差：这是由于  $\hat{\beta} = \bar{y}$ ,  $\text{RSS} = \sum_{i=1}^n (y_i - \bar{y})^2$ 。在正态分布的假定条件下，比值  $(\frac{\text{RSS}}{\hat{\sigma}^2})$  呈  $\chi^2$  分布 (DoF n-p)，并且与线性系数估计值  $\hat{\beta}$  无关。

小提示：使用  $\chi^2$  分布作为 LHF 来估计  $\sigma^2$ ，比起使用高斯分步来同时估计  $\beta, \sigma^2$  来，会得到无偏估计。

## 17.3 假设检验

如何对回归系数向量估计  $\hat{\beta}$  作假设检验？具体来说，这个问题可以分为两种情况

- 对向量  $\beta$  中的某个系数  $\beta_i$  作显著程度检验，
- 对几个、甚至全部系数作显著程度检验，

本节介绍一些常见的检测方法，尤其是

- 基于 MLE 的抽样分布的 Wald 检验，
- 似然率检验。

### 17.3.1 Wald 检验

假设我们只检测系数向量  $\beta$  中某一系数  $\beta_j$  的显著性，如

$$H_0 : \beta_j = 0,$$

如果零假设成立，那么其 MLE  $\hat{\beta}_j$  的分布情况因此可写为  $\sim \mathcal{N}\left(0, \text{var}(\hat{\beta}_j)\right)$ ，其中方差  $\text{var}(\hat{\beta}_j)$  是方差协方差矩阵  $\text{var}(\hat{\beta})$  (17.10) 中第  $j$  个对角元素的值。那么可以考虑如下比值，定义为 Wald 统计量  $t$

$$t_j = \frac{\hat{\beta}_j}{\sqrt{\text{var}(\hat{\beta}_j)}}, \quad (17.12)$$

计算  $t_j$  的难度在于，(17.10) 中全体方差  $\sigma^2$  常常未知。在实际应用中，常常将  $\sigma^2$  替换为无偏估计  $\hat{\sigma}^2$

$$\hat{\sigma}^2 = \frac{\text{RSS}(\hat{\beta})}{n - p}. \quad (17.13)$$

假设观测数据符合正态分布， $\hat{\sigma}^2$  依照(17.13) 计算，那么在零假设下  $\hat{\sigma}^2$  呈学生  $t$  分布 (student's t distribution), DoF n-p。

具体来说，若观测数据的二阶弱假设条件（均值、方差、方差协方差）得到满足，那么  $t$  值 (17.12) 在大样本下近似为标准正态分布。这为大样本数据下的近似推断打下良好基础。

一些研究中，不考虑数据样本量的大小，直接将(17.12) 作为学生  $t$  统计，而且这值得进一步讨论。

1. 若样本量大，假定条件“观测数据是否符合正态分布”的确不重要；
2. 若样本量相对较小，那么正态分布的假定是否成立，对 Wald 检测是否有效会产生决定性影响。对于规模适中的观测样本，应当审慎使用  $t$  检测：Student's t 值随着自由度趋近于无穷大而收敛至标准正态分布<sup>1</sup>。

<sup>1</sup> 如双边 95% 的关键  $t$  值是 2.09 (20 DoF), 1.98 (100 DoF)，而标准正态分布的关键  $t$  值是 1.96 (100 DoF)。

$t$  值可用于描述一个系数估计的置信区间 (confidence interval)，例如，某个真实系数值  $\beta_j$  以  $100 \times (1 - \alpha)\%$  的置信率落在统计区间中

$$\hat{\beta}_j \pm t_{\frac{1-\alpha}{2}, n-p} \sqrt{\text{var}(\hat{\beta}_j)}. \quad (17.14)$$

其中  $t_{\frac{1-\alpha}{2}, n-p}$  表示  $n - p$  DoF,  $\alpha$  样本大小的学生  $t$  双边分布的关键值。

Wald 检验也可用于检测一组系数的联合置信区间。将系数向量  $\beta$  分解为两块

$$\beta^\top = \begin{pmatrix} \beta_1^\top & \beta_2^\top \end{pmatrix}, \quad \{1 \times p\} \quad \{1 \times p_1\} \quad \{1 \times p_2\}$$

建立零假设

$$H_0 : \beta_2 = 0.$$

在求得 MLE  $\hat{\beta}$  后，将 Wald 检验统计量表示为二次形式

$$W = \hat{\beta}_2^\top \left( \text{var}(\hat{\beta}_2) \right)^{-1} \hat{\beta}_2, \quad (17.15)$$

如前所述，我们用  $\frac{\text{var}(\hat{\beta})}{n-p}$  近似替代全  $\sigma^2$ ，用于测算方差协方差矩阵  $\text{var}(\hat{\beta}_2)$ 。若  $p_2 = 1$  即  $\beta_2$  只有一个系数，那么(17.15)回复到 (17.13)的形式。

根据渐进理论 (asymptotic theory)，在零假设  $H_0$  下，大样本 MLE  $\hat{\beta}_2$  符合多元正态分布，均值向量为 0，方差协方差矩阵为  $\text{var}(\hat{\beta}_2)$ 。由此可得，二次形 Wald 统计  $W$  (17.15) 在大样本下也是一个  $\chi^2$  分布，对应  $p_2$  DoF。无论  $\sigma^2$  是实现给定的值，还是利用 RSS 作近似估计，上述结论都成立。

如果我们持有更强形式的假定，即观测数据呈正态分布，那么检验的结果也是更强形式的。此时若  $\sigma^2$  值事先给定， $W$  就恰好是  $\chi^2$  分布 ( $p_2$  DoF)。若用  $\frac{\text{RSS}(\hat{\beta})}{n-p}$  来近似估计  $\sigma^2$  (对应  $p_2$  DoF)，则  $\frac{W}{p_2}$  呈一个对应于  $p_2$  和 DoF  $n - p$  的  $F$  分布。

值得注意的是，给定  $p$  值不变，随着  $n \rightarrow \infty$ ,  $(n - p) \rightarrow \infty$ ,  $F_{p_2, n-p} \times p_2 \rightarrow \chi^2$  ( $p_2$  DoF)。这意味着在大样本情况下，将  $W$  统计量视作  $\chi^2$  分布，或是将  $W/p_2$  视作  $F$  分布，二者并无本质区别。

### 17.3.2 似然率检验

还来看多系数联合显著水平检验的例子，对应零假设

$$H_0 : \beta_2 = 0.$$

我们可以将模型矩阵  $X$  相应地分解为

$$X = \begin{pmatrix} X_1 & X_2 \end{pmatrix}, \quad \{p \times 1\} \quad \{p_1 \times 1\} \quad \{p_2 \times 1\}$$

如果零假设成立，这意味着后面的  $p_2$  个解释变量  $X_2$  对观测数据  $Y$  无影响。

可以建立似然率检验 (likelihood ratio test, LHR) 指标，对零假设做检验，分两步走

1. 分别构建两个嵌入模型并作拟合。

(a) 小模型：只考虑前  $p_1$  个解释变量  $X_1$ 。

(b) 大模型：用全部  $p = p_1 + p_2$  个解释变量  $X$ 。

2. 比较两个模型的最大(对数)似然方程。

先来看只考虑  $X_1$  的小模型。基于给定的方差  $\sigma^2$ , 对应似然方程(17.5), 可得最大似然方程

$$\max \log L(\beta_1) = c - \frac{1}{2} \frac{\text{RSS}(X_1)}{\sigma^2}, \quad c = -\frac{n}{2} \log(2\pi\sigma^2), \quad (17.16)$$

其中常数  $c$  的值取决于  $n$  和  $\sigma^2$ 。

对应地, 考虑  $X = (X_1, X_2)$  的大模型, 最大似然方程

$$\max \log L(\beta_1, \beta_2) = c - \frac{1}{2} \frac{\text{RSS}(X_1 + X_2)}{\sigma^2}. \quad (17.17)$$

将两个最大似然方程(17.16), (17.17)相减, 定义为  $\log \lambda$ , 作为似然率量 (likelihood ratio criterion, LHRC)

$$-2 \log \lambda = \max \log L(\beta_1) - \max \log L(X_1, X_2) = \frac{\text{RSS}(X_1) - \text{RSS}(X_1 + X_2)}{\sigma^2}. \quad (17.18)$$

关于 LHRC, 有两点说明

1. 通过两个最大似然方程的差值, 该指标反映在引入额外的  $p_2$  个解释变量  $X_2$  后, RSS 的变化情况。通常来说, 如果  $\Delta \text{RSS} > 0$ , 那么  $X_2$  对观测数据可能是有显著影响的。
2. 对  $\Delta \text{RSS}$  除以方差  $\sigma^2$  是作单位标准化处理, 表示残差平方和的变化是以总方差的单位计。如果这个比值(最大似然率)超过了事先预期值, 就表示  $X_2$  的确是显著的, 零假设不成立。

这就涉及到如何设定“事先预期”值, 它与样本的分布有关。根据大样本定律可知, 随着  $n \rightarrow \infty$ , 样本分布逐渐收敛至  $\chi^2(p_2 \text{ DoF})$ 。对于  $\chi^2$  分布, 已知期望值和方差分别为  $\nu$  和  $2\nu$ , 那么事先预期的设定可以使: 每增加额外 1 个变量  $p_2 \rightarrow p_2 + 1$ , 会导致  $\Delta \text{RSS}$  减少  $\sigma^2$  个单位, 对应标准化后的单位 1。具体说来, 若 RSS 的下降幅度超过 95% 百分位 (percentile) 的参考分布值, 就表示为误差的缩减超出预期。

类似地, LHRC (17.18) 的计算难度在于  $\sigma^2$  可能未知。可以用用大模型  $\text{RSS}(X_1 + X_2)$  的方差估计  $\hat{\sigma}^2$  作近似替代, 计算方法为

$$\hat{\sigma}^2 = \frac{\text{RSS}(X_1 + X_2)}{n - p},$$

代回(17.18), 算得的大样本下 LHRC 分布依旧是  $\chi^2$  分布 ( $p_2$  DoF)。

若观测样本符合正态分布这一更强假定, 结果也相应更强。如果  $\sigma^2$  是已知的, 那么 LHRC  $-2 \log \lambda$  就恰好是  $\chi^2$  分布 ( $p_2$  DoF)。如果  $\sigma^2$  未知, 是用估计出的  $\hat{\sigma}^2$  作近似替代, 那么对应的 LHRC 除以  $p_2$ , 即  $\frac{-2 \log \lambda}{n-p}$  就恰好是一个  $F_{p_2, n-p}$  分布, 满足

$$F = \frac{\frac{1}{p_2} [\text{RSS}(X_1) - \text{RSS}(X_1 + X_2)]}{\frac{1}{n-p} \text{RSS}(X_1 + X_2)}, \quad (17.19)$$

分子表示在每一单位自由度减少, 导致 RSS 的减少幅度; 分母是平均 RSS, 反映模型总体的噪声情况。通常来说, 如果测得的  $F$  统计值高于  $F$  分布在 95% 下对应  $p_2, n - p$  的关键值, 那么我们可以拒绝零假设  $H_0$ ,  $X_2$  的确对观测数据产生影响。

### 17.3.3 Anova 表

在实际应用中, 到底应该用基于大样本分布下最大似然估计的 Wald 检验, 还是基于最大(对数)似然估计比较的似然率检验呢? 答案是, 随着  $n$  值逐渐增大, 两种检测方法渐进等价。若模型

是线性的，则回答更为明确：两种检测完全等价。详细证明过程略，我们可以在一些具体应用中涉及相关讨论。

值得指出的是，两种检验的（渐进）等价基于模型的线性结构假定。对于非线性模型如 logistic 模型、泊松回归模型等，两种检测是有区别的。通常说来，对线性模型的检测而言，我们可以提出以下建议：

- 对单个系数可用 Wald 检验 (17.12)，
- 多个系数的检验，或者进一步对多个嵌入模型的比较，可用 LHRC 的  $F$  检验 (17.19)。

测算  $F$  统计所需的计算，常收录在 anova 表中 (analysis of variance table, anova)。表格将整体 RSS 分为三个部分，小模型  $X_1$  的 RSS、加入  $X_2$  后大模型的 RSS、剩余 RSS。每个 RSS 项后还附有 DoF，以及平均值  $\frac{\text{RSS}}{\text{DoF}}$ ，如表17.1所示。

表 17.1: anova 表

	RSS	DoF
$X_1^a$	$\text{RSS}(\phi) - \text{RSS}(X_1)$	$p_1 - 1$
$X_2 X_1^b$	$\text{RSS}(X_1) - \text{RSS}(X_1 + X_2)$	$p_2$
残差	$\text{RSS}(\phi)$	$n - p$
全部 <sup>c</sup>	$\text{RSS}(X_1 + X_2)$	$n - 1$

<sup>a</sup>  $\phi$  表示零模型

<sup>b</sup> 有时我们需要更多的细节，那么可以将  $p_2$  个  $X_2$  分拆后一个一个加入到模型中，分别计算 RSS 和自由度

<sup>c</sup> 有时也称不相关 RSS，即  $= \sum_{i=1}^n y_i^2$

一系列信息如简单相关系数、偏相关系数、复相关系数 (coefficients of simple, partial, multiple correlation) 都可基于 anova 表中 RSS 和 DoF 的信息而计算获得。

几乎全部主流统计软件都支持数值计算 anova。

### 17.3.4 简单线性回归

先来看一个线性解释变量  $x$  对  $y$  的连续影响。若已知从观测数据来看，哪怕  $x$  是定值，对应的观测数据  $y$  也有所不同，那么可将  $\{y_i\}$  理解为一组随机变量的实现

$$y_i \sim \mathcal{N}(\mu_i, \sigma^2), \quad (17.20)$$

其中均值  $\mu_i = E[y_i]$  取决于解释变量  $x_i$ ，方差  $\sigma^2$  是个常数。最简单的思路可以假设一个线性方程，从而得到一个简单线性模型 (simple linear regression model, SLM)

$$\mu_i = \alpha + \beta x_i, \quad (17.21)$$

$\alpha$  常称为常数项或截距，反映  $x_i = 0$  时，期望  $\mu_i$  的取值。 $\beta$  称为斜率，表示随着  $x_i$  的增加，期望  $\mu_i$  的变化幅度。

SLM 可以理解为一个一般线性模型 (第17.1节) 的特例：模型矩阵  $X$  包括两列，一列是常数项  $\alpha$ ，一列是解释变量  $x_i$ 。可对参数做估计，计算标准差 (第17.2节)，并执行假设检验 (第17.3节)。

在 SLRM 中,  $\alpha$  和  $\beta$  的估计为

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}, \quad (17.22)$$

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}. \quad (17.23)$$

拟合的直线(17.22)反映解释变量期望值和观测数据期望值之间的关系, 对应斜率(17.23)反映协方差  $\text{cov}(x_i, y_i)$  和方差  $\text{var}(x_i)$  之间的比值。

[找个回归跑一下, 做个说明。]

其中我们有决定系数 (coefficient of determination)  $R^2$

$$R^2 = 1 - \frac{\text{RSS}(x)}{\text{RSS}(\phi)},$$

对应地我们将  $R$  称为皮尔逊线性相关系数 (Person's linear correlation coefficient)

$$R = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (x_i - \bar{x})^2}} \quad (17.24)$$

现在, 假设我们用两个解释变量向量  $x_1, x_2$  来解释观测向量  $y$ , 其中第  $i$  个元素分别为  $x_{i1}, x_{i2}, y_i, i = 1, \dots, n$ 。 $y_i$  设为  $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$ 。那么我们有多元线性回归模型 (multivariate regression model, MLRM)

$$\mu_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2}, \quad (17.25)$$

构成一个三维空间。 $\alpha$  是常数,  $\beta_1, \beta_2$  分别是  $x_{i1}, x_{i2}$  的斜率。求解方法、参数估计和假设检验与 SLRM 相类似。

## 17.4 回归诊断

通常来说, 统计建模的工作分为三个阶段: 模型构建、用模型拟合数据、模型诊断。在前两个阶段工作的基础上, 模型诊断主要是指重新调整模型, 使其能够更好地解释现实。本节介绍相关方法。

### 17.4.1 拟合值与残差

诊断模型用的指标是残差  $r_i$ , 它是观测数据  $y_i$  和模拟数据  $\hat{y}_i = X_i^\top \beta$  之间的差

$$r_i = y_i - \hat{y}_i, i = 1, \dots, n. \quad (17.26)$$

如(17.4)所示, 将全部  $i$  个  $\hat{y}_i$  汇总在一起, 有矩阵形式

$$\hat{y} = X\hat{\beta}, \quad \hat{\beta} = \underset{\beta}{\operatorname{argmin}} \text{RSS}(\beta)$$

其中  $X$  是模型矩阵,  $\text{RSS}(\beta)$  的计算方式见(17.6)。在此基础上我们有

$$\hat{y} = Hy, \quad H = X(X^\top X)^{-1}X, \quad (17.27)$$

其中  $H : y \mapsto \hat{y}$  又称帽子矩阵 (hat matrix)。

根据这样的定义我们有, 模拟值  $\hat{y}$  的均值  $\mu$ , 方差协方差矩阵  $\text{var}(\hat{y})$  分别满足

$$\mu = E\hat{y}, \quad \text{var}(\hat{y}) = H\sigma^2.$$

在此基础上，残差矩阵  $r$  满足

$$r = y - \hat{y},$$

其中  $y, \hat{y}$  分别表示观测值向量和模拟值向量。引入(17.27)替代  $\hat{y}$ ，上式进一步变为

$$r = (I - H)\sigma^2. \quad (17.28)$$

假设观测数据满足常见的二阶条件，那么我们有残差的均值和方差协方差矩阵

$$Y \sim \mathcal{N}(0, \sigma^2) \implies \begin{cases} Er = 0 \\ \text{var}(r) = (I - H), \end{cases}$$

进而第  $i$  个残差  $r_i$  的方差

$$\text{var}(r_i) = (1 - h_{ii})\sigma^2, \quad (17.29)$$

其中  $h_{ii}$  是帽子矩阵  $H$  中第  $i$  个对角元素。

由(17.29)可看出一个重要性质：尽管全部观测数据的方差相同  $\sigma^2$ ，但具体每个观测残差的方差  $\text{var}(r_i)$  却可能不相等，其原因在于模拟值  $\hat{y}_i$  的精度取决于协变量  $x_i$  的值，而后者并不相同。为了解决这个问题，随后提出了标准化残差的概念，见第17.4.2节。

来看  $h_{ii}$  的取值区间。假定模型的截距项不为 0，那么我们有  $h_{ii} \in [\frac{1}{n}, \frac{1}{r}]$ ，其中  $n \ni i$  表示全部观测数据的数量， $r$  表示第  $i$  个观测值的复制的数量，或者说，全部协变量值等于  $x_i$  的观测样本  $i$  的数量。在 SLM 中，有

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (17.30)$$

可见当  $x_i$  恰好等于其均值  $\bar{x}$  时， $h_{ii}$  有最小值  $\frac{1}{n}$ 。

值得指出的是，观测值  $y_i$  越是接近（远离）其均值  $\bar{y}$ ，对应的模拟值  $\hat{y}_i$  的方差  $\text{var}(\hat{y}_i)$  就越小（大）。但通过(17.29)-(17.30)可见， $x_i$  越是接近（远离） $\bar{x}$ ， $h_{ii}$  越小（大），残差的方差  $\text{var}(r_i)$  就越大（小）。

### 17.4.2 标准化残差

根据残差  $r$  的定义(17.29)可见，在比较不同观测数据的残差  $r_i, r_j, i, j = 1, \dots, n$  时，需要注意可能存在  $\text{var}(r_i) \neq \text{var}(r_j)$  的情况，导致不同观测数据之间难于直接比较。因此可以调整如下，将  $r_i$  扩展为标准化残差  $s_i$  (standardized residual)

$$s_i = \frac{r_i}{\sqrt{1 - h_{ii}}\hat{\sigma}}, \quad (17.31)$$

其中  $\hat{\sigma}$  是基于 RSS 的标准差估计。

标准化残差  $s_i$  可用于检验观测样本中是否存在不规则观测值 (anomalous observation) 或称离群值 (outlier)。通常来说，若某个观测值  $y_i$  的标准化残差绝对值  $|s_i| > 2$ ，就需要对它额外关注：它可能是离群值（但并非绝对，详见第17.4.4节讨论）。

### 17.4.3 Jack-knifed 残差

标准化残差  $s_i$  的不足在于，计算式(17.31)中还依赖于标准差估计  $\hat{\sigma}$ ：它可能受到离群值的影响，而离群值很可能不易检测出来。可进一步调整如下：将  $r_i$  用一个误差方差估计  $\hat{\sigma}_{(i)}$  替代  $\hat{\sigma}$  作标准化处理，称之为 Jack-knifed 残差 (Jack-Knifed residual)或学生化残差 (studentized residual)，用  $t_i$  表示

$$t_i = \frac{r_i}{\sqrt{1 - h_{ii}}\hat{\sigma}_{(i)}}, \quad (17.32)$$

$\hat{\sigma}_{(i)}$  的计算方法：值通过“遗漏”第  $i$  项观察元素，对剩下部分做模型拟合，基于  $n - p - 1$  DoF 下 RSS 测得的标准差。需要注意的是，拟合值  $\hat{y}$  和帽子矩阵  $H$  的计算仍然基于含有第  $i$  观测样本。

**Definition 17.1** (标准化预测残差). 如果用不含有  $i$  的观测数据，不只计算 *Jack-knifed* 残差 (17.32)，还计算标准残差，那么标准残差会发生什么变化？用  $\hat{\beta}_{(i)}$  表示回归系数的估计，建立估计模型

$$\hat{y}_{(i)} = \mathbf{x}_i^\top \hat{\beta}_{(i)},$$

相应地我们将

$$y_i - \hat{y}_{(i)}$$

称预测残差 (*predictive residual*)。预测残差的方差为

$$\text{var}(y_i - \hat{y}_{(i)}) = \left(1 + \mathbf{x}_i^\top \left(\mathbf{X}_{(i)}^\top \mathbf{X}_{(i)}\right)^{-1} \mathbf{x}_i\right) \sigma_{(i)}^2,$$

由于第  $i$  项被剔除在回归分析之外，因此  $y_i$  和  $\hat{y}_{(i)}$  不相关。 $\mathbf{X}_{(i)}$  是模型矩阵  $X$  去除掉第  $i$  行后剩余的部分。 $\hat{\sigma}_{(i)}^2$  是去掉第  $i$  项观测元素后，对余下部分构建模型，计算 RSS，进而算得。用  $\hat{\sigma}_{(i)}^2$  近似未知的  $\sigma^2$ ，可得标准化预测残差 (*standardized predictive residual*)

$$t_i = \frac{y_i - \hat{y}_{(i)}}{\sqrt{\text{var}(y_i - \hat{y}_{(i)})}}. \quad (17.33)$$

不难看出，标准化预测残差 (17.33) 和 *Jack-knifed* 残差 (17.32) 等价，可以互换。

*Jack-knifed* 残差  $t_i$  的计算式 (17.32) 较为复杂。在求得标准化残差  $s_i$  (17.31) 的基础上，可将  $t_i$  计算简化为 (Cook and Weisberg, 1982)

$$t_i = s_i \left( \frac{n-p-1}{n-p-s_i^2} \right)^{\frac{1}{2}}, \quad (17.34)$$

不难看出， $t_i$  是一个关于  $s_i$  的单调方程，那么序列  $\{t_i\}_{i=1}^n$  的排序，就等价于  $\{s_i\}_{i=1}^n$  的排序。

#### 17.4.4 离群值的检测

$s_i$  或  $t_i$  的值大于 2，暗示着  $y_i \in Y, i = 1, \dots, n$  可能是离群值。仅仅“可能”不够精确，我们可以利用  $t_i$  作进一步诊断。对于原有的模型  $\mu_i = \mathbf{X}_i^\top \beta$  作扩展，加入一个哑变量 (dummy variable)  $z_i$ ，从而允许第  $i$  个观测数据出现位移

$$\mu_i = \mathbf{X}_i^\top \beta + \gamma z_i, \quad z_i = \begin{cases} 1 & \text{是第 } i \text{ 个观测} \\ 0 & \text{其他,} \end{cases}$$

系数  $\gamma$  用于表示观测  $y_i$  可以在多大程度上偏离 (由协变量  $X_i$  和回归系数  $\beta$  共同描述的) 期望值  $\mu_i$ 。

建立一个零假设用于检测  $y_i$  是不是离群值，即  $y_i$  是否与其他  $n - 1$  个观测  $\{y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n\}$  有相同的行为模式：

$$H_0 : \gamma = 0.$$

对系数  $\gamma$  作近似估计  $\hat{\gamma}$ 。对  $H_0$  作 Wald 检验，算得的  $t$  就是 *Jack-knifed* 残差  $t_i$  值，自由度  $(n - p - 1)$

$$t_i = \frac{\hat{\gamma}}{\sqrt{\text{var}(\hat{\gamma})}}. \quad (17.35)$$

这一设定允许  $\mu_i$  拥有不同于其他  $n - 1$  个观测均值的特征，用估计系数  $\hat{\gamma}$  表示。我们事实上是在用除  $i$  之外的  $n - 1$  个观测来估计  $\hat{\beta}$ ，用全部  $n$  个观测来估计  $\hat{\gamma}$ ，从这个意义上讲  $\gamma$  描述的信息与预测残差 (Definition 17.1) 相近。

在利用 Jack-knifed 残差(17.35)进行离群值检验的过程中，需要注意显著水平的值。在观测数据中，如果我们是事先怀疑  $i$  为离群值，那么该检验有效。如果  $i$  是在事后确认的，即我们在浏览观测数据之后，结合实际情况“怀疑” $i$  为离群值，那么名义显著水平、进而离群值检验是无效的：这是由于我们实际上进行了两次检验。

如果检验基于 5% 显著水平展开，这意味着大约每 20 次检验中会有 1 次结果显著。现在假设要作一组  $k$  次检验，一个控制整体显著水平的方法时，每次检验过程中都使用显著水平  $\frac{\alpha}{k}\%$ ，从而根据邦费罗尼不等式 (Bonferroni inequality)(Shumway and Stoffer, 2017, Ch.4)，整体显著程度不高于  $\alpha$ 。需要指出的是，这是一个非常保守的方法：真实的显著水平很可能远远低于  $\alpha$ 。

#### 17.4.5 影响力和杠杆

回到对帽子矩阵  $H : y \mapsto \hat{y}$  (17.27) 及其对角元素  $h_{ii}$  的分析上来。由(17.29)可见，第  $i$  项观测的残差的方差  $\text{var}(r_i)$  等于  $\sigma^2$  和  $(1 - h_{ii})$  的乘积，由定义可得  $\lim_{\{h_{ii} \rightarrow 1\}} \text{var}(r_i) \rightarrow 0$ ，即模拟值  $\hat{y}$  逼近观测值  $y_i$ 。从这个角度来说， $h_{ii}$  称为观测  $y_i$  的杠杆 (leverage)，或称潜在影响因子 (potential influence)。杠杆不能随意取值：设  $p = \text{trace}(H)$ ，相应地平均杠杆值为  $\frac{p}{n}$ 。具体到观测  $y_i$ ，若  $h_{ii} \geq \frac{p}{n}$ ，我们称  $y_i$  有高杠杆，或称潜在影响因子较大。

#### 17.4.6 真实影响因子和库克距离

根据影响因子的定义，无论如何模拟数据  $\hat{y}_i$  总是能够向  $y_i$  逐渐靠拢的，而现实观察中并非总是如此。理论和现实不匹配。为了解决这个问题，就需要调整模型，分析“真实”影响因子，一种测度方法是库克距离 (Cook's distance)  $D_i$

$$D_i = \frac{(\hat{\beta}_{(i)} - \hat{\beta})^\top (\widehat{\text{var}}(\hat{\beta}))^{-1} (\hat{\beta}_{(i)} - \hat{\beta})^\top}{p}, \quad (17.36)$$

$\hat{\beta}$  表示基于全部  $n$  个观测数据的系数估计， $\hat{\beta}_{(i)}$  表示剔除  $i$  后利用余下  $(n - 1)$  个观测数据的系数估计。不难看出，库克距离也描述是欧几里得空间中，全部  $n$  个模拟值  $\hat{y}$  与去除  $i$  后  $(n - 1)$  个估计值  $\hat{y}_{(i)}$  之间的差或称距离。已知  $(\text{var}(\hat{\beta}))^{-1} = \frac{X^\top X}{\sigma^2}$ ,  $\hat{y}_{(i)} = X\hat{\beta}_{(i)}$ , (17.36)因此等价于

$$D_i = \frac{\sum_{j=1}^n (\hat{y}_{(i)j} - \hat{y}_{ij})^2}{p \hat{\sigma}^2}. \quad (17.37)$$

测算库克距离所需的计算量比较大；可用简化算法表示如下

$$D_i = s_i^2 \frac{h_{ii}}{(1 - h_{ii}) p}, \quad (17.38)$$

其中标准化残差  $s_i$  由(17.31)算得。简化计算式的库克距离将标准化残差和杠杆结合在一起。

$D_i$  的值越接近于 1， $y_i$  的实际影响因子越显著。为了说明这一点，来看(17.37)，

$$D_i = \frac{W}{p}, \quad W := s_i^2 \frac{h_{ii}}{1 - h_{ii}},$$

其中 Wald 统计量  $W$  采取(17.15)的形式，为利用零假设  $H_0 : \beta = \beta_0$  检验系数估计  $\hat{\beta}_{(i)}$  而得。基于第17.3.1的介绍可知，当自由度较高时， $W/p$  呈  $F$  分布，进而库克距离  $D_i$  等于这个零假设的

$F$  统计量的中位值。如果删除观测  $y_i$  会使  $F$  统计量由 0 向中位值变化，那么我们说  $i$  是影响显著的<sup>2</sup>。此时较为明智的策略是将  $y_i$  剔除，利用余下  $(n - 1)$  个观测重新估计，查看系数估计值的变化情况。

## 17.5 数据变换

第17.4节介绍了一些常见的回归检测方法。如果经过检测发现，现有模型对于解释实际观测数据不够理想，除了调整系数估计之外，有时也可以对观测数据和/或解释变量作变换。

### 17.5.1 变换观测数据

变换  $\{y_i\}_{i=1}^n$  的目标是使观测数据线性、同方差。常用的方法是方差稳定变换 (variance-stabilizing transformation, VST)。某观测集  $Y$  方差  $\text{var}(Y)$  的分布已知，与均值有关，那么可以设一个转换方程  $h(Y)$ ，使其方差  $\text{var}(h)$  近似为常数。观测数据的类型不同 (均值、方差等)，可设定不同形式的转换方程，例如表17.1。

表 17.1: 变换观测数据的转换方程

$Y$ 的分布	均值——方差	转换方程 $h(Y)$
泊松分布	$\sigma^2 = \mu$	$\sqrt{Y}$
二项式分布	$\sigma^2 \propto \mu(1 - \mu)$	$(\sin \sqrt{Y})^{-1}$
对数正态*	$\sigma^2 \propto \mu^2$	$\log Y$

<sup>a</sup> 指对数正态线性化转换，见第11.5.4节。

例如，假设观测  $y_i$  看起来符合泊松分布，那么作 VST 如下：

$$E[\sqrt{y_i}|x_i] = x_i^\top \beta, \quad \text{var}(\sqrt{y_i}|x_i) = \sigma^2.$$

值得注意的是，我们并不总是那么幸运，能找到合适的转换方程  $h(Y)$ ，或者即便能，系数估计  $\hat{\beta}$  的含义也难以解释。

### 17.5.2 Box-Cox 变换

Box and Cox (1964) 提出一种对非负观测变量的变换方法，可用于一系列特殊形式如倒数、对数、平方根等，称为 Box-Cox 变换 (Box-Cox transformation)。具体说来，假设观测数据  $y$  符合线性和正态分布并且方差是个常数，可建立转换方程  $h(y, \lambda) \sim \mathcal{N}(X^\top \beta, \sigma^2)$ ，进而我们有

$$y^{(\lambda)} = h(y, \lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log y & \lambda = 0. \end{cases} \quad (17.39)$$

可以利用前文介绍的最大似然法估计系数  $\lambda, \beta, \sigma^2$ 。从估计结果来看， $\hat{\lambda}$  常常在  $[-2, 2]$  之间，一些特殊值对应倒数 (-1)、对数 (0)、平方根 (1/2)、原值 (1)、平方 (2) 等。

<sup>2</sup>这等价于在一个新的置信区间中重新作点估计，新置信区间是原置信区间的 50%。

对数似然方程为

$$\log L(\beta, \sigma^2, \lambda) = \underbrace{-\frac{n}{2} \log(2\pi\sigma^2)}_{=: \mathcal{B}} - \frac{1}{2} \sum_{i=1}^n \frac{(y_i^{(\lambda)} - \mu_i)^2}{\sigma^2} + (\lambda - 1) \underbrace{\sum_{i=1}^n \log(y_i)}_{=: \mathcal{A}}, \quad (17.40)$$

其中  $\mathcal{A}$  来自雅各比变换  $\frac{\partial y^{(\lambda)}}{\partial \lambda}$ ,  $\mathcal{B}$  是常规的似然方程。在假定  $\lambda$  值事先给定的情况下, 可以用最大似然估计法求得系数  $\hat{\beta}, \hat{\sigma}^2$ 。代回(17.40), 可得描述似然方程

$$\log L(\lambda) = c - \frac{n}{2} \log \text{RSS}(y^{(\lambda)}) + (\lambda - 1) \sum_{i=1}^n \log(y_i), \quad (17.41)$$

其中  $c = \frac{n}{2} \log(\frac{2\pi}{n})$  是个与  $\lambda$  无关的常数。

基于描述似然方程(17.41)作最大似然估计所需的计算量较大, 常用以下方法进行简化: 定义观测数据变换

$$z^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda \tilde{y}^{\lambda-1}} & \lambda \neq 0 \\ \log(y) \tilde{y} & \lambda = 0, \end{cases}$$

其中  $\tilde{y}$  是  $y$  的几何平均, 满足

$$\tilde{y} = \exp\left[\frac{\sum_{i=1}^n \log y_i}{n}\right]. \quad (17.42)$$

在此基础上可将描述似然方程写为

$$\log L(\lambda) = c - \frac{n}{2} \log \text{RSS}(z^{(\lambda)}), \quad (17.43)$$

其中  $\text{RSS}(z^{(\lambda)})$  是用  $x$  解释  $z^{(\lambda)}$  算得的 RSS。这种变换方法的好处在于, 我们可以通过 RSS 的值直接比较基于不同  $\lambda$  取值的模型。

在实际应用过程中, 我们计算不同  $\lambda$  值下的描述似然方程(17.41)或(17.43), 目的是

- 寻找使得  $\log L(\lambda)$  最大时的  $\lambda$  值,
- 或计算当  $\log L(\lambda)$  在最大值附近的  $\lambda$  值,
- 或计算当  $\lambda$  取一系列特殊值如  $-1, 0, 1/2, 1, 2$  等时的  $\log L(\lambda)$  值,

例如设

$$\hat{\lambda} = \underset{\lambda}{\operatorname{argmax}} \log L(\lambda),$$

通过构建 LHRC

$$\chi^2 = 2 [\log L(\hat{\lambda}) - \log L(\lambda_0)],$$

一个在大样本下近似  $\chi^2$  分布 (1 DoF) 的信息量, 来检测零假设

$$H_0 : \lambda = \lambda_0,$$

其中  $\lambda_0$  可以是任意定值。

此外也可将  $\lambda$  的基于似然方程的置信区间, 定义为一组值的集合, 这一组值可以在上述检测中通过, 即是说, 使

$$2 [\log L(\lambda) - \log L(\lambda_0)]$$

在 MLE 下, 处于分布  $\chi^2_{1-\alpha, 1}$  值之内的全部  $\hat{\lambda}$  值的集合。对这组  $\lambda$  值集合的识别需要使用数值计算方法。

Box-Cox 变换主要面对观测数据非负的情况而设计。对于观测数据大多为正、偶尔少量为 0 或复的情况, 可对原有数据加上一个常数  $\alpha$  后再进行变换。 $\alpha$  的取值可以通过最大似然估计方法来确定, 或是直接赋予一个较小的正值, 如 0.5 或 1(取决于观测数据的实际特点)。

### 17.5.3 阿特金森法

Box-Cox 变换为拟合一个线性序列提供了一种研究思路。在此基础上, Atkinson (1985) 提出了简化方案, 来检验一组观察数据是否需要变换。据图说来, 在模型中加入一个辅助变量  $a$

$$a_i = y_i \left[ \log \left( \frac{y_i}{\tilde{y}} \right) - 1 \right], \quad (17.44)$$

其中  $\tilde{y}$  是  $y$  的几何平均值, 如(17.42)。

那么建立包括  $a$  的扩展模型, 对应系数  $\gamma$ 。如果 MLE  $\hat{\gamma}$  显著, 那么  $a$  对模型有影响, 需要对观测数据  $y$  作 Box-Cox 变换。对应地, 可根据  $\hat{\lambda} = 1 - \hat{\gamma}$  计算  $\lambda$  的值, 这是由于, 假设真实模型满足

$$z^{(\lambda)} = X\beta + \epsilon,$$

那么将 LHS 的  $z^{(\lambda)}$  围绕  $\lambda = 1$  作 1 阶段泰勒级数展开

$$z^{(\lambda)} \approx z^{(1)} + (\lambda - 1) \frac{dz^{(\lambda)}}{d\lambda} \Big|_{\lambda=1},$$

其中

$$\frac{dz^{(\lambda)}}{d\lambda} \Big|_{\lambda=1} = a + \log \tilde{y} + 1,$$

$a$  的值由(17.44)给出。可见  $\frac{dz^{(\lambda)}}{d\lambda} \Big|_{\lambda=1}$  的值与  $\lambda$  无关, 因而可归并到常数项中。

此外根据定义式, 可得  $z^{(1)} = y - 1$ , 进而

$$y \approx X\beta + (1 - \lambda)a + \epsilon,$$

因此辅助变量  $a$  对应的系数是  $1 - \lambda$ 。

## 17.6 简介 (待归并)

设一个  $M \times 1$  的向量  $g$ , 描述一组随机数据。对应地, 设一个  $K \times 1$  的参数向量  $\theta$ , 用来描述这组随机数据  $g$ 。那么, 概率密度方程 (probability density function, PDF)  $pr(g|\theta)$  用于表示, 在给定  $\theta$  的情况下, 随机数据  $g$  的抽样分布 (sampling distribution)。一个常见的 PDF 是高斯概率密度方程 (Gaussian PDF), 可见第16.B.1节。

反之, 若在现实中已经观测到了  $g$ , 则我们可将  $pr(g|\theta)$  看成是一个关于参数  $\theta$  的方程, 称为似然方程 (likelihood function), 定义为  $L(\theta|g)$

$$L(\theta|g) = pr(g|\theta),$$

需要注意的是,  $L(\theta|g)$  并不是  $\theta$  的概率密度方程。

我们将对  $\theta$  的估计命名为  $\hat{\theta}$ 。 $\hat{\theta}$  是个关于  $g$  的方程: 由于  $g$  是随机数据向量, 导致  $\hat{\theta}(g)$  也是个关于  $g$  的随机变量向量。本质上来说, 参数估计反映的是一个映射过程

$$\begin{cases} \theta \mapsto g \mapsto \hat{\theta}, \\ pr(\theta) \mapsto pr(g|\theta) \mapsto pr(\hat{\theta}|\theta). \end{cases}$$

从这个角度来说,  $\hat{\theta}$  是一个关于  $\theta$  的随机变量。

## 17.7 近似的测度 (待归并)

估计值  $\hat{\theta}$  相比于真实值  $\theta$  往往存在误差。误差常常来自于两个方面：

- 随机误差，又称 precision：如估计偏差，噪音等。常用方差 (variance) 予以量化描述。
- 系统误差，又称 accuracy：如估计偏误，校准误差，模型误设定等。常用偏误 (bias) 或均方误差 (MSE, EMSE) 予以量化描述。

两种误差均可由估计值的条件分布  $pr(\hat{\theta}|\theta)$  予以量化。

### 17.7.1 估计的偏误：标量参数

以标量参数  $\theta$  及其 (随机) 估计值  $\hat{\theta}$  为例，估计的条件均值  $\bar{\theta}$  可由似然方程计算而得

$$\bar{\theta} = \langle \hat{\theta}(g) \rangle_{g|\theta} = \int d^{M_g} pr(g|\theta) \hat{\theta}(g), \quad (17.45)$$

其中  $\langle \cdot \rangle_{g|\theta}$  表示在给定的  $\theta$  下，各个随机估计值  $\hat{\theta}(g)$  的平均值。

如果已知估计  $\hat{\theta}$  的条件概率密度  $pr(\hat{\theta}|\theta)$ ，那么(17.45)可进一步改写为

$$\bar{\theta} = \int d\hat{\theta} pr(\hat{\theta}|\theta) \hat{\theta}. \quad (17.46)$$

#### 偏误

我们将  $\bar{\theta}$  相对于  $\theta$  的偏差称为关于真实值的条件偏误 (conditional bias on true parameter)，定义为  $b(\theta)$

$$b(\theta) = \bar{\theta} - \theta. \quad (17.47)$$

如果某个估计  $\bar{\theta}$ ，对于任意  $\theta$  都满足  $b(\theta) = 0$ ，那我们称之为无偏估计 (unbiased estimate)。

#### 可测度性

可测度性 (estimability) 或可识别性 (identifiability)。如果无论随机数据集  $g$  的取值如何，都存在关于参数  $\theta$  的无偏估计  $\hat{\theta}$ ，满足  $b(\theta) = 0 \quad \forall \theta$ ，那我们称参数  $\theta$  是可估计的 (estimable) 或可识别的 (identifiable)，表示为  $\bar{\theta}$

$$\bar{\bar{\theta}} = \int d\theta pr(\theta) \bar{\theta} = \int d\theta pr(\theta) \int d^{M_g} pr(g|\theta) \hat{\theta}(g). \quad (17.48)$$

#### 测度的方差和均方误差

某一测度  $\hat{\theta}$  相对于测度均值  $\bar{\theta}$  的波动，我们用方差  $var(\hat{\theta})$  来表示

$$var(\hat{\theta}) = \sigma_{\hat{\theta}}^2 = \left\langle |\hat{\theta}(g) - \bar{\theta}|^2 \right\rangle_{g|\theta}. \quad (17.49)$$

某一测度  $\hat{\theta}$  相对于真实值  $\theta$  的波动，我们用均方误差 (mean square error, MSE) 来表示

$$MSE(\theta) = \left\langle |\hat{\theta} - \theta|^2 \right\rangle_{g|\theta}. \quad (17.50)$$

考虑到  $\theta$  随  $g$  而具有的随机性特征，将均方误差再对  $\theta$  取平均值，得整体均方误差 (ensemble mean square error, EMSE)

$$EMSE(\theta) = \left\langle \left\langle |\hat{\theta} - \theta|^2 \right\rangle_{g|\theta} \right\rangle_{\theta}. \quad (17.51)$$

### 17.7.2 估计的偏误：向量参数

设一个  $P$  维的参数向量  $\theta$ , 对应估计值  $\hat{\theta}$ 。我们将平均值  $\bar{\theta}$ (有时也写作  $\langle \theta \rangle$ ) 定义如下

$$\bar{\theta}(g) = \int d^{M_g} pr(g|\theta) \hat{\theta}(g) = \int d^{P_{\hat{\theta}}} pr(\hat{\theta}|\theta) \hat{\theta}(g). \quad (17.52)$$

#### 平均偏误

平均偏误  $\bar{b}(\theta)$  定义为

$$\bar{b}(\theta) = \langle b(\theta) \rangle_{\theta} \quad (17.53)$$

#### 测度的方差和均方误差

我们将  $P$  维随机估计向量  $\hat{\theta}$  中, 第  $p$  个元素的均值表示为  $\langle \hat{\theta}_p \rangle_p = \bar{\hat{\theta}}_p$ 。对应的方差为

$$\begin{aligned} var(\hat{\theta}_p) &= \left\langle \left[ \hat{\theta}_p - \langle \hat{\theta} \rangle \right] \left[ \hat{\theta}_p - \langle \hat{\theta} \rangle \right]^T \right\rangle_{g|\theta} \\ &= \int_{\infty} d^{M_g} \left| \hat{\theta}_p(g) - \langle \hat{\theta}_p(g) \rangle \right|^2 pr(g|\theta) \\ &= \int_{\infty} d^{P_{\hat{\theta}}} \left| \hat{\theta}_p - \langle \hat{\theta}_p \rangle \right|^2 pr(\hat{\theta}|\theta). \end{aligned} \quad (17.54)$$

将所有  $p \in P$  个方差汇总在一起, 得到整体的方差协方差矩阵, 定义为  $K_{\hat{\theta}}$

$$K_{\hat{\theta}} = \left\langle (\hat{\theta} - \bar{\theta}) (\hat{\theta} - \bar{\theta})^{\dagger} \right\rangle = \left\langle \Delta \hat{\theta} \Delta \hat{\theta}^{\dagger} \right\rangle. \quad (17.55)$$

#### 参数向量的均方误差

$$\begin{aligned} MSE &= \left\langle \left\| \hat{\theta} - \theta \right\|^2 \right\rangle_{g|\theta} \\ &= \int_{\infty} d^{M_g} \left\| \hat{\theta}(g) - \theta \right\|^2 pr(g|\theta) \\ &= \text{trace}(K_{\hat{\theta}}) + \text{trace}(b(\theta)b(\theta)^{\dagger}) \end{aligned} \quad (17.56)$$

## 17.8 费雪信息矩阵和克拉美罗下界 (待归并)

参数估计的方差协方差矩阵  $K$ , 存在一些限定条件。

### 17.8.1 标量参数

任意标量参数的无偏估计, 其方差都必须满足克拉美罗下界 (Cramér-Rao lower bound, CRLB)

$$var(\hat{\theta}) \geq \underbrace{\frac{1}{\left\langle \left[ \frac{\partial}{\partial \theta} \log pr(g|\theta) \right]^2 \right\rangle_{g|\theta}}}_{=: F}, \quad (17.57)$$

其中我们将分母的部分称为费雪信息矩阵 (Fischer information matrix)。

若估计是有偏的, 克拉美罗下界为

$$var(\hat{\theta}) \geq \frac{\left( \frac{db(\theta)}{d\theta} + 1 \right)^2}{\left\langle \left[ \frac{\partial}{\partial \theta} \log pr(g|\theta) \right]^2 \right\rangle_{g|\theta}} \quad (17.58)$$

### 17.8.2 向量参数

对于  $P$  维向量参数  $\theta$ , 费雪信息矩阵  $F$  成为  $P \times P$  维的埃尔米特矩阵 (Hermitian matrix), 其中第  $jk$  个元素  $F_{jk}$  表示为

$$\begin{aligned} F_{jk} &= \left\langle \left[ \frac{\partial}{\partial \theta_j} \log pr(g|\theta) \right] \left[ \frac{\partial}{\partial \theta_k} \log pr(g|\theta) \right] \right\rangle_{g|\theta} \\ &= \int_{\infty} d^{M_g} pr(g|\theta) \left[ \frac{1}{pr(g|\theta)} \frac{\partial}{\partial \theta_j} pr(g|\theta) \right] \left[ \frac{1}{pr(g|\theta)} \frac{\partial}{\partial \theta_k} pr(g|\theta) \right]. \end{aligned} \quad (17.59)$$

在得到费雪信息矩阵  $F$  的基础上, 来看 (参数向量  $\theta$  的无偏估计向量  $\hat{\theta}$  的) 方差协方差矩阵  $K_{\hat{\theta}}$ 。 $n \times n$  矩阵  $K_{\hat{\theta}}$  中的第  $nn$  个元素, 等于  $\hat{\theta}$  向量中第  $nn$  个元素的方差, 并且也满足克拉美罗下界

$$[K_{\hat{\theta}}]_{nn} = \text{var}(\hat{\theta}_{nn}) \geq [F^{-1}]_{nn}, \quad (17.60)$$

这就是说, 为了计算无偏估计向量  $\hat{\theta}$  中第  $n$  个元素方差的下界 (或方差协方差矩阵  $K_{\hat{\theta}}$  中第  $nn$  元素的下界), 我们需要首先计算费雪信息矩阵的逆矩阵, 提取其中的第  $nn$  个元素。

一个满足克拉美罗下界条件的无偏估计是有效估计 (efficient)。

## 17.9 最大似然估计 (待归并)

前面讨论了一个参数估计  $\hat{\theta}$  所应当具有的特性。随后的问题就是, 如何找到这个估计值? 常见的方法之一是最大似然估计 (maximum likelihood estimation)

$$\hat{\theta}_{ML} \equiv \underset{\theta}{\operatorname{argmax}} pr(g|\theta), \quad (17.61)$$

有时也用对数似然方程来表示

$$\hat{\theta}_{ML} \equiv \underset{\theta}{\operatorname{argmax}} \log pr(g|\theta), \quad (17.62)$$

即是说, 选择合适的参数值  $\hat{\theta}_{ML}$ , 使得在  $\hat{\theta}_{ML}$  下可能实际观测到数据集合  $g$  的概率最大。

没有其他额外限定条件的情况下,  $\hat{\theta}_{ML}$  常取值于 (对数) 似然方程对  $\theta$  的梯度 (gradient), 即一阶导数等于 0 的点

$$\underbrace{\nabla_{\theta} \log pr(g|\theta)}_{=:s(g)} = 0, \leftrightarrow \theta = \hat{\theta}_{ML}, \quad (17.63)$$

我们将这个表示斜率的随机向量  $s(g)$  称为 score, 用于描述 (对数) 似然方程对参数变化的敏感程度, 因此有时也称敏感方程 (sensitivity function)。

$s(g)$  的方差协方差矩阵就等于费雪信息矩阵的逆。

### 17.9.1 最大似然估计: to be or not to be?

1. 选择最大似然估计法的理由

- (a) 有效。如果的确存在一个有效的最大似然估计值。
- (b) 渐进有效。随着观测数据的数量增加, 估计效果逐渐提升。
- (c) 渐进无偏。
- (d) 渐进一致。
- (e) 便于计算。相对来说。

- (f) 有助于严谨地强制使得估计结果与实际观测数据相一致。
- (g) 无需先验信息  $pr(\theta)$  设定。与之相反地，参考贝叶斯估计。

## 2. 不选择最大似然估计法的理由

- (a) (严谨地强制使得估计结果与实际观测数据相一致)。

现实中观测到的数据常常是充满噪音的。强行严谨一致容易导致计算出的最大似然估计值，哪怕是无偏的，也随之充满噪音。

- (b) (无需先验信息  $pr(\theta)$  设定)。

现实中总是存在着一些先验信息  $pr(\theta)$ ，我们应当充分利用这些信息，哪怕他们有时并非完全无偏<sup>3</sup>。

## 17.10 伯努利实验：最大似然估计法（待归并）

现在我们从伯努利实验 (Bernoulli trials)入手，介绍如何应用最大似然估计法求解相关问题。基于本节的求解思路，我们随后在 python 环境中编写程序求解，见第17.10.1节。

手头一枚一元硬币，现在我们连续掷 10 次，有 7 次头朝上。我们想要利用最大似然法，估计每一次投掷得到头朝上的概率。

用  $p \in (0, 1)$  来表示每次投掷得到头朝上的概率（参数的实际值）。那么连续投掷  $N$  次，在这  $N$  次中得到  $n \in \mathbb{N}$  次头朝上的概率可以表示为

$$pr(n|p) = \binom{N}{n} p^n (1-p)^{N-n}, \quad (17.64)$$

其中二项式系数满足

$$\binom{N}{n} = \frac{N!}{n! (N-n)!}. \quad (17.65)$$

随机变量  $n$  的均值和方差分别为

$$\langle n \rangle = Np, \quad (17.66)$$

$$\sigma^2 = Np(1-p). \quad (17.67)$$

在上一组实验中我们观测到的数据是投掷  $N$  次得到  $n$  次头（如 10 次得到 7 次头）。我们想要知道如果反复重复这组实验，出现  $N/n$  的最大概率  $p$  是多少？可用最大似然法进行估计：改变  $p$  的值，使得  $pr(n|p)$  达到最大。

对  $pr(n|p)$  取对数，然后对  $p$  求导，得 score 方程  $s(n)$

$$\begin{aligned} \log pr(n|p) &= \text{constant} + n \log p + (N-n) \log(1-p), \\ s(n) &= \frac{d}{dp} \log pr(n|p) = \frac{n}{p} + \frac{N-n}{1-p}. \end{aligned} \quad (17.68)$$

设  $s(n) = 0$ ，对应的  $p = \hat{p}_{ML}$  即为  $p$  的最大似然估计

$$\hat{p}_{ML} = \frac{n}{N}, \quad (17.69)$$

---

<sup>3</sup>一种将先验  $pr(\theta)$  与最大似然估计相结合的方法称为加权似然法 (weighted likelihood)，如

$$\hat{\theta}_{WL} \equiv \underset{\theta}{\operatorname{argmax}} pr(g|\theta) pr(\theta) = \underset{\theta}{\operatorname{argmax}} pr(\theta|g).$$

分子表示观测到头的次数，分母表示总的投掷次数。

最大似然估计  $\hat{p}_{ML}$  的均值和方差，由(17.66)-(17.67)得

$$\langle \hat{p}_{ML} \rangle = \frac{\langle n \rangle}{N}, \quad (17.70)$$

$$\text{var}(\hat{p}_{ML}) = \frac{\text{var}(n)}{N^2} = \frac{p(1-p)}{N} \quad (17.71)$$

据此可得估计的费雪信息矩阵

$$F = \left\langle \left[ \frac{n}{p} - \frac{N-n}{1-p} \right]^2 \right\rangle = \frac{\langle [n-Np]^2 \rangle}{p^2(1-p)^2} = \frac{\text{var}(n)}{p^2(1-p)^2} = \frac{N}{p(1-p)}, \quad (17.72)$$

因此我们有

$$F^{-1} = \text{var}(\hat{p}_{ML}), \quad (17.73)$$

并且最大似然估计  $\hat{p}_{ML}$  是有效的。

### 17.10.1 伯努利实验：最大似然估计的 Python 程序实现

伯努利实验（频率法）在 Python 环境下的实现

伯努利实验(Bernoulli trials)是指重复若干次(设为  $T$  次)某一项实验(或称迭代 iteration)，每次实验(iteration)的结果可能是二者之一：成功(0)或者失败(1)。假定我们现在来检测成功的概率，可以基于频率学派的思路，使用最大似然估计法来实现(maximum likelihood, MLB)。

In [2]: # 初始设定

```
%matplotlib inline
from __future__ import division
import numpy as np
import pandas as pd
import statsmodels.api as sm
import sympy as sp
import pymc3
import matplotlib.pyplot as plt
import matplotlib.gridspec as gridspec
from mpl_toolkits.mplot3d import Axes3D
from scipy import stats
from scipy.special import gamma
from sympy.interactive import printing
printing.init_printing()
```

#### 模型构建

设一个伯努利实验

$$y \sim \text{Bernoulli}(\theta) = \text{Binomial}(1, \theta). \quad (17.74)$$

我们将概率密度方程，或称边际似然方程，写为

$$p(y|\theta) = \theta^y (1-\theta)^{1-y} = \begin{cases} \theta & y=1 \\ 1-\theta & y=0. \end{cases} \quad (17.75)$$

设  $\theta = 0.3$ , 随机进行 100 次试验。

In [3]: # 模拟数据

```
np.random.seed(123)

nobs = 100
theta = 0.3
# numpy.random.binomial(n, p, size=None)
# https://docs.scipy.org/doc/scipy-1.13.0/reference/generated/numpy.random.binomial.html
Y = np.random.binomial(1, theta, nobs)
```

In [4]: # 画图

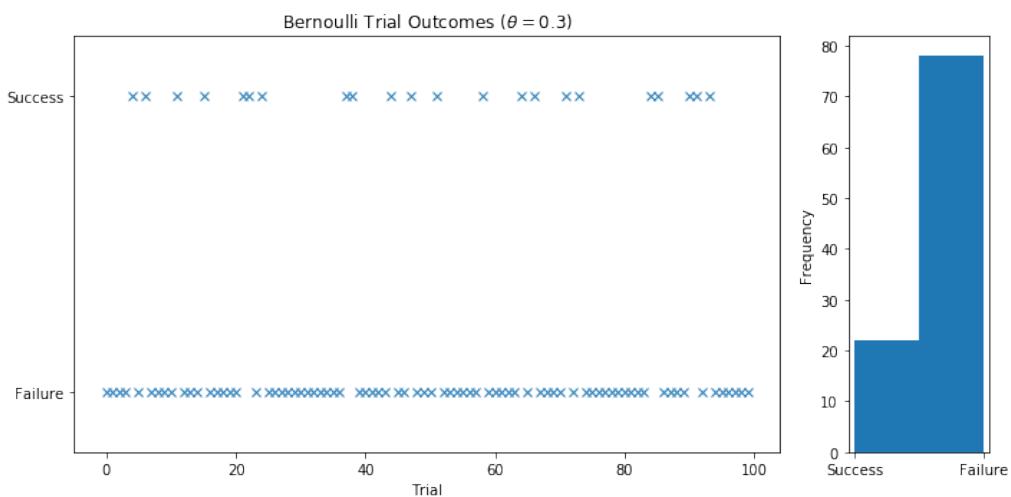
```
fig = plt.figure(figsize=(10,5))
gs = gridspec.GridSpec(1, 2, width_ratios = [5,1])
ax1 = fig.add_subplot(gs[0])
ax2 = fig.add_subplot(gs[1])

ax1.plot(range(nobs), Y, "x")
ax2.hist(-Y, bins=2)

ax1.yaxis.set(ticks=(0,1), ticklabels=("Failure", "Success"))
ax2.xaxis.set(ticks=(-1,0), ticklabels=("Success", "Failure"))

ax1.set(title=r"Bernoulli Trial Outcomes $(\theta = 0.3)$", xlabel="Trial", ylim=(-0.2, 1.2))
ax2.set(ylabel="Frequency")

fig.tight_layout()
```



### 似然方程

现在假定从随机变量 100 个  $y$  中，随机提取  $T$  个值，构成一个样本

$$Y = (y_1, \dots, y_T)^\top \quad (17.76)$$

对应的联合似然方程为

$$p(Y|\theta) = \prod_{i=1}^T \theta^{y_i} (1-\theta)^{1-y_i} = \theta^s (1-\theta)^{T-s}, \quad s = \sum_i y_i, \quad (17.77)$$

$s$  就是观测到的成功的次数（对应  $y = 1$ ）； $T - s$  是观测到的失败的次数 ( $y = 0$ )。

随后的目标就是，如何将似然方程最大化？

### 最大似然估计：解析法

一种较为直观的方法是解析法。

- 似然方程  $L(\theta; Y)$

$$L(\theta; Y) = \theta^s (1-\theta)^{T-s}. \quad (17.78)$$

- 对数似然方程  $\log(L(\theta; Y))$

$$\log L(\theta; Y) = s \log \theta + (T - s) \log (1 - \theta). \quad (17.79)$$

- score 矩阵（一阶导数）

$$\frac{\partial \log L(\theta; Y)}{\partial \theta} = \frac{s}{\theta} - \frac{T-s}{1-\theta} = s \theta^{-1} - (T-s) (1-\theta)^{-1}. \quad (17.80)$$

- 黑塞矩阵（二阶导数）

$$H(\theta) = \frac{\partial^2 \log L(\theta; Y)}{\partial \theta^2} = -s \theta^{-2} - (T-s) (1-\theta)^{-2}. \quad (17.81)$$

- 信息矩阵

$$\begin{aligned} I[\theta] &= -E[H(\theta)] = -E[-s \theta^{-2} - (T-s) (1-\theta)^{-2}] \\ &= (\theta T) \theta^{-2} + (T-\theta T) (1-\theta)^{-2} = \frac{T}{\theta(1-\theta)}. \end{aligned} \quad (17.82)$$

- 方差协方差矩阵（逆信息矩阵）

$$I[\theta]^{-1} = \frac{\theta(1-\theta)}{T}. \quad (17.83)$$

```
In [5]: t, T, s = sp.symbols("theta, T, s")
```

```
# 构建符号方程
likelihood = (t ** s) * ((1 - t)**(T - s))
loglike = s * sp.log(t) + (T - s) * sp.log(1-t)
score = (s/t) - ((T-s)/(1-t))
```

```

hessian = - s/(t**2) - (T - s)/((1-t)**2)
information = T/(t * (1-t))

# 利用 sympy.lambdify 将符号方程转换为 numpy.lambda
# sympy.utilities.lambdify.ImplementedFunction(symfunc, implementation)
# http://docs.sympy.org/latest/modules/utilities/lambdify.html
_likelihood = sp.lambdify((t,T,s), likelihood, modules='numpy')
_loglike = sp.lambdify((t,T,s), loglike, modules='numpy')
_score = sp.lambdify((t,T,s), score, modules='numpy')
_hessian = sp.lambdify((t,T,s), hessian, modules='numpy')

```

对数似然方程  $\log L(\theta; Y)$  对  $\theta$  求偏导，值为 0 处对应的  $\theta$  值定义为  $\hat{\theta}_{MLE}$

$$\frac{\partial \log L(\theta; Y)}{\partial \theta} \Big|_{\hat{\theta}_{MLE}} = 0, \\ \hookrightarrow \hat{\theta}_{MLE} = \frac{s}{T}. \quad (17.84)$$

对应地

$$H\left(\hat{\theta}_{MLE}\right) = -\frac{T^2}{s} - \frac{T^2}{T-s} < 0, \Leftrightarrow T > s, \quad (17.85)$$

二阶导数小于 0 确保最大似然条件成立。

In [6]: # 计算极大似然估计系数  $\theta$  的解析值

```

theta_hat_analytic = Y.sum() / nobs
var_analytic = (theta_hat_analytic * (1 - theta_hat_analytic)) / nobs

print('Analytic MLE Results: %.2f (%.10f)' % (theta_hat_analytic, var_analytic ** 0.5))

Analytic MLE Results: 0.22 (0.0414246304)

```

### 最大似然估计：数值法

在很多情况下，我们无法通过解析法求得最大化的解。这时需要数值近似求解，利用 `scipy.optimize` 最优算法。以下几点需要注意。

- 对似然方程的负数求最小化，达到最大化的效果。
- 最优算法常常是在不受约束的参数空间  $\mathbb{R}^k$  中进行的 ( $k$  表示参数数量即维度)。如果实际研究对象的参数值是受限的，如本例  $\theta \in (0, 1)$ ，则需要将受限参数首先转换为无约束空间中的参数  $(0, 1) \ni \theta \rightarrow \varphi \in \mathbb{R}$ ，然后使用最优算法求解。对应的转换矩阵

$$\theta = g(\varphi) = \frac{\exp(\varphi)}{1 + \exp(\varphi)}, \quad \varphi \in \mathbb{R}. \quad (17.86)$$

- 只有得到了黑塞矩阵，根据逆黑塞矩阵的负数才能计算方差协方差矩阵。而黑塞矩阵的解析形往往无法求得——可以利用 `scipy.optimize.fmin_bfgs` 数值近似黑塞矩阵。

- 利用数值近似黑塞矩阵算得的方差协方差矩阵  $Cov(\varphi)$ , 需要转换为  $Cov(\theta)$

$$Cov(\theta) = \left. \frac{\partial g(\varphi)}{\partial \varphi} \right|_{\varphi_{MLE}} \times Cov(\varphi) \times \left. \frac{\partial g(\varphi)}{\partial \varphi} \right|^{\top}_{\varphi_{MLE}} \quad (17.87)$$

转换矩阵的偏导数（斜率）为

$$\frac{\partial g(\varphi)}{\partial \varphi} = \frac{\partial \left[ \exp(\varphi) (1 + \exp(\varphi))^{-1} \right]}{\partial \varphi} = \frac{\exp(\varphi)}{(1 + \exp(\varphi))^2} \quad (17.88)$$

无约束参数  $g(\varphi)$  构成的模型系统

- 逆转换方程

$$\begin{aligned} \theta &= g(\varphi) = \frac{\exp(\varphi)}{1 + \exp(\varphi)}, \\ 1 + \exp(\varphi) \theta &= \exp(\varphi), \\ \theta &= \exp(\varphi) (1 - \theta), \\ \varphi &= \log \left( \frac{\theta}{1 - \theta} \right) = g^{-1}(\theta) \end{aligned} \quad (17.89)$$

- 似然方程  $L(\theta; Y)$

$$L(g(\varphi); Y) = g(\varphi)^s (1 - g(\varphi))^{T-s} \quad (17.90)$$

- 对数似然方程  $\log(g(\varphi); Y)$

$$\log L(g(\varphi); Y) = s \log g(\varphi) + (T - s) \log (1 - g(\varphi)) \quad (17.91)$$

- score 矩阵（一阶导数）

$$\begin{aligned} \frac{\partial \log L(g(\varphi); Y)}{g(\varphi)} &= \frac{\partial \log L(g(\varphi); Y)}{g(\varphi)} \frac{\partial g(\varphi)}{\partial \varphi} \\ &= \left( \frac{s}{g(\varphi)} - \frac{T - s}{1 - g(\varphi)} \right) \frac{\exp(\varphi)}{(1 + \exp(\varphi))^2} \\ &= \left( \frac{s}{\frac{\exp(\varphi)}{1 + \exp(\varphi)}} - \frac{T - s}{1 - \frac{\exp(\varphi)}{1 + \exp(\varphi)}} \right) \frac{\exp(\varphi)}{(1 + \exp(\varphi))^2} \\ &= s - T \frac{\exp(\varphi)}{1 + \exp(\varphi)} = s - T g(\varphi). \end{aligned} \quad (17.92)$$

- 黑塞矩阵（二阶导数）

$$H(\varphi) = \frac{\partial^2 \log L(g(\varphi); Y)}{\partial g(\varphi)^2} = -T. \quad (17.93)$$

- 信息矩阵

$$I[\varphi] = -E[H(\varphi)] = T. \quad (17.94)$$

- 方差协方差矩阵

$$I[\varphi]^{-1} = \frac{1}{T}. \quad (17.95)$$

### Delta 近似法

进而我们有 Delta 近似法

$$\begin{aligned} I_{Delta}[\hat{\theta}_{MLE}]^{-1} &= \left. \frac{\partial g(\varphi)}{\partial \varphi} \right|_{\varphi_{MLE}} \times I[\varphi]^{-1} \times \left. \frac{\partial g(\varphi)}{\partial \varphi} \right|^{\top} \Big|_{\varphi_{MLE}} \\ &= \left\{ \frac{\exp \left[ \log \left( \frac{\hat{\theta}_{MLE}}{1-\hat{\theta}_{MLE}} \right) \right]}{\left( 1 + \exp \left[ \log \left( \frac{\hat{\theta}_{MLE}}{1-\hat{\theta}_{MLE}} \right) \right] \right)^2} \right\}^2 \frac{1}{T} = \left[ \left( \frac{\hat{\theta}_{MLE}}{1-\hat{\theta}_{MLE}} \right) \left( 1 + \frac{\hat{\theta}_{MLE}}{1-\hat{\theta}_{MLE}} \right)^{-2} \right]^2 \frac{1}{T} \\ &= \left[ \frac{\hat{\theta}_{MLE}}{1-\hat{\theta}_{MLE}} \left( 1 - \hat{\theta}_{MLE} \right)^2 \right]^2 \frac{1}{T} = \left[ \hat{\theta}_{MLE} \left( 1 - \hat{\theta}_{MLE} \right) \right]^2 \frac{1}{T} \\ &= I[\hat{\theta}_{MLE}]^{-1} \left[ \hat{\theta}_{MLE} \left( 1 - \hat{\theta}_{MLE} \right) \right] \end{aligned} \quad (17.96)$$

Delta 近似容易导致方差协方差矩阵被低估。可以写出误差项

$$I[\theta]^{-1} - I_{Delta}[\theta]^{-1} = \frac{\theta(1-\theta) - [\theta(1-\theta)]^2}{T}, \quad (17.97)$$

或者误差项的偏移（相对误差）

$$\frac{|I[\theta]^{-1} - I_{Delta}[\theta]^{-1}|}{I[\theta]^{-1}} = \frac{\frac{\theta(1-\theta) - [\theta(1-\theta)]^2}{T}}{\frac{\theta(1-\theta)}{T}} = 1 - \theta(1-\theta) \quad (17.98)$$

In [7]: `fig = plt.figure(figsize = (12,4))`

```
# 误差项
ax = fig.add_subplot(121, projection="3d")
delta = 0.025
T_range = np.arange(0,100,10)
# numpy.linspace(start, stop, num=50, endpoint=True, retstep=False, dtype=None) [source]
# https://docs.scipy.org/doc/numpy-1.13.0/reference/generated/numpy.linspace.html
TH_range = np.linspace(0,1)
# numpy.meshgrid(*xi, **kwargs) [source]
# https://docs.scipy.org/doc/numpy-1.13.0/reference/generated/numpy.meshgrid.html
# [X, Y] = meshgrid(x, y) 将向量 x 和 y 定义的区域转换成矩阵 X 和 Y,
# 这两个矩阵可以用来表示 mesh 和 surf 的三维空间点以及两个变量的赋值。
# 其中矩阵 X 的行向量是向量 x 的简单复制，而矩阵 Y 的列向量是向量 y 的简单复制。
T_grid, TH_grid = np.meshgrid(T_range, TH_range)
# 误差项的测算
E_grid = (TH_grid * (1 - TH_grid) - (TH_grid * (1-TH_grid)) ** 2)/T_grid
ax.plot_wireframe(T_grid, TH_grid, E_grid, color="#348ABD")
```

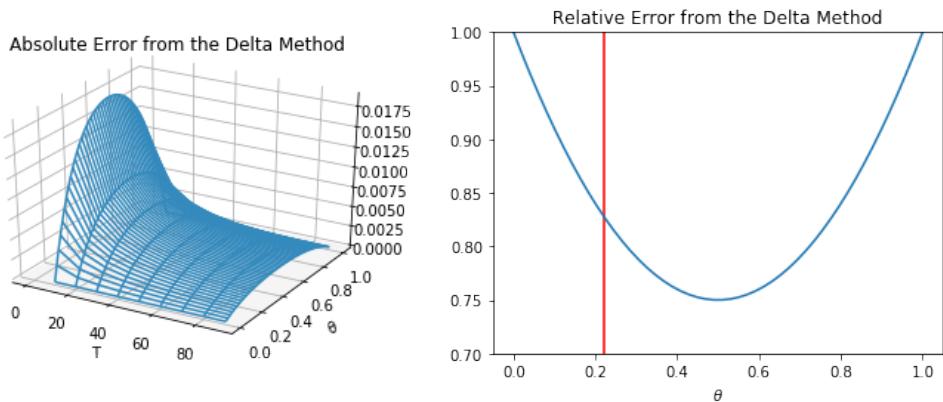
```

ax.set(
    title = "Absolute Error from the Delta Method",
    xlabel= "T",
    ylabel= r"\theta"
)
ax.xaxis.set(pane_color = (1,1,1))
ax.yaxis.set(pane_color = (1,1,1))

# 误差项的偏移 (相对误差)
ax = fig.add_subplot(122)
X = np.linspace(0,1)
ax.plot(X, 1 - X * (1-X))
ax.set(
    title = "Relative Error from the Delta Method",
    xlabel = r"\theta",
    ylim = (0.7, 1)
)
ax.vlines(theta_hat_analytic, 0.7, 1, "r")

```

Out[7]: <matplotlib.collections.LineCollection at 0x125ab96a0>



- 左图可以看出，绝对误差项随着样本数量  $T$  的增加而逐渐增大，相对误差的变化则与  $T$  无关。
- 右图可以看出，在最大似然估计的解析解 ( $\theta = 0.22$ ) 处，相对误差的值约为 82%。

### 最大似然估计：算法举例

```

In [8]: # 载入
from scipy import optimize

# 构建转换方程
transform = lambda phi: np.exp(phi) / (1 + np.exp(phi))

```

```
## 转换方程的斜率（一阶导数）
transform_grad = lambda phi: np.exp(phi) / ((1 + np.exp(phi)) **2)

# 构建优化所调用的方程。为负，是因为我们使用最小化的算法。
f = lambda params: -_loglike(
    transform(params[0]), nobs, Y.sum()
)
fprime = lambda params: np.matrix(
    -_score(
        transform(params[0]), nobs, Y.sum()
    )
)
fhess = lambda params: np.matrix(
    -_hessian(
        transform(params[0]), nobs, Y.sum()
    )
)
```

scipy.optimize 可以使用一系列算法，这里以两种方法为例作简要介绍。

第一种方法叫做 牛顿共轭梯度法 (Newton Conjugate Gradient Method, Newton-CG method)。

In [9]: # 利用牛顿共轭梯度法作最大似然估计

```
#%timeit res_ncg = optimize.fmin_ncg(f, np.array([0.5]), fprime, fhess=fhess, disp=False)
print ("Timing: 10 loops, best of 3: 4.89 ms per loop")

res_ncg = optimize.fmin_ncg (f, np.array([0.5]), fprime, fhess = fhess)[0]
var_ncg = np.linalg.inv(
    -_hessian(transform(res_ncg), nobs, Y.sum())
)[0,0]
#var_ncg = np.linalg.inv(-_hessian(transform(res_ncg), nobs, Y.sum()))[0,0]
print ("Newton-CG MLE estimate: %.2f (%.10f)" % (transform(res_ncg), var_ncg**0.5))

Timing: 10 loops, best of 3: 4.89 ms per loop
Optimization terminated successfully.

    Current function value: 52.690796
    Iterations: 35
    Function evaluations: 51
    Gradient evaluations: 85
    Hessian evaluations: 35
Newton-CG MLE estimate: 0.22 (0.0414252425)
```

第二种方法叫做 BFGS 法(Broyden–Fletcher–Goldfarb–Shanno algorithm, BFGS method)。

In [10]: # 利用 BFGS 法作最大似然估计

```
#%timeit res_bfgs = optimize.fmin_bfgs(f, np.array([0.5]), fprime=fprime, disp=False)
```

```

print("Timing: 1000 loops, best of 3: 1.85 ms per loop")

res_bfgs_full = optimize.fmin_bfgs(f, np.array([0.5]), fprime = fprime, full_output = True)
res_bfgs = res_bfgs_full[0]
est_var_bfgs = res_bfgs_full[3]

# 计算信息矩阵；计算信息矩阵的逆矩阵，得到的方差协方差矩阵
var_bfgs = np.linalg.inv(
    -_hessian(transform(res_bfgs), nobs, Y.sum())
)[0,0]
# 根据 (解析)delta 方法得到的方差协方差矩阵
var_bfgs_delta = (transform_grad(res_bfgs)**2 * est_var_bfgs)[0,0]

print(
    "BFGS MLE estimate: %.2f (%.10f)" % (transform(res_bfgs), var_bfgs**0.5)
)
print(
    "BFGS MLE estimate (delta) : %.2f (%.10f)" % (transform(res_bfgs), var_bfgs_delta ** 0.5)
)

Timing: 1000 loops, best of 3: 1.85 ms per loop
Optimization terminated successfully.
    Current function value: 52.690796
    Iterations: 6
    Function evaluations: 7
    Gradient evaluations: 7
BFGS MLE estimate: 0.22 (0.0414246303)
BFGS MLE estimate (delta) : 0.22 (0.0171611895)

```

进而我们有误差项和误差项的偏移

```

In [11]: absolute_error = var_bfgs - var_bfgs_delta
          relative_error = (var_bfgs - var_bfgs_delta) / var_bfgs

          print(absolute_error)
          print(relative_error)

0.00142149357495
0.828376209448

```

第三种方法是牛顿拉夫森法 (Newton-Raphson method)。比起前两种方法来，牛顿拉夫森法的数值计算速度更快，但稳健性较低。python 的 statsmodels 包提供了这种算法的程序。

```

In [20]: # 利用 BFGS 法作最大似然估计
          # 导入 Statsmodels 包

```

```

from statsmodels.base.model import LikelihoodModel

class Bernoulli(LikelihoodModel):
    _loglike = lambda self, theta, T, s: _loglike(theta, T, s)
    _score = lambda self, theta, T, s: _score(theta, T, s)
    _hessian = lambda self, theta, T, s: _hessian(theta, T, s)

    def __init__(self, endog, exog=None, **kwargs):
        # numpy.asarray(a, dtype=None, order=None)[source]
        # https://docs.scipy.org/doc/scipy-1.13.0/reference/generated/numpy.asarray.html
        endog = np.asarray(endog)
        super(Bernoulli, self).__init__(endog, exog, **kwargs)
        self.T = self.endog.shape[0]
        self.s = self.endog.sum()

    def loglike(self, params):
        """
        Joint log-likelihood for Bernoulli trials
        """
        return self._loglike(params[0], self.T, self.s)

    def score(self, params):
        """
        Gradient of the joint log-likelihood for Bernoulli trials
        """
        return self._score(params[0], self.T, self.s)

    def hessian(self, params):
        """
        Hessian of the joint log-likelihood for Bernoulli trials
        """
        return np.matrix(self._hessian(params[0], self.T, self.s))

In [26]: mod = Bernoulli(Y)
#%timeit res = mod.fit(start_params=[0.5], disp=False)
print("1000 loops, best of 3: 277 µs per loop")
# 最大似然估计
res = mod.fit(start_params=[0.5])
print ("Statsmodels MLE Estimate: %.2f (%.10f)" % (res.params[0,0], res.bse[0]))
```

1000 loops, best of 3: 277 µs per loop

Optimization terminated successfully.

Current function value: 0.526908

Iterations 2

Statsmodels MLE Estimate: 0.22 (0.0414246304)



# 第十八章 准最大似然估计

对于一组随机变量，不再假设一个关于这组随机变量均值的方程，而是采取更通用的形式，假定它们的联合分布（密度方程）是正确设定的，在此基础上可以进行最大似然估计（第十七章）。传统最大似然估计法假设，事先设定的密度方程就等于真实的密度方程，从而根据这样的假设，设定误差是不存在的。但是从实际观测的数据来看，“正确设定”的情况很少见——如何在可能存在模型误设定的情况下，更有效地进行统计推断？将 MLE 作一个一般化扩展，就有了准最大似然估计（quasi-maximum likelihood estimation, QMLE），将假定条件放宽到，认为事先设定的密度方程在多数情况下只是对真实值的一种近似。这里以 White (1994) 为例介绍 QMLE，更多相关讨论可参考如 White (1982); Amemiya (1985); Gourieroux et al. (1993); Chen and Swanson (2013)。

## 18.1 Kullback-Leibler 信息准则

### 18.1.1 信息

假设在一项随机实验中， $A$  事件发生的概率是  $0 \leq p \leq 1$ 。我们关注于“事件  $A$  会发生”这一信号，其价值或重要性，对于不同  $p$  值来说是不同的：当  $p$  值较小时，信号的价值高；反之当  $p$  值很大时，信号的价值低。那么，信号“事件  $A$  会发生”所含有的信息量（information content），应当理解为一个随  $p$  而递减的方程  $\tau(p)$ ，其形式可以设定如下

$$\tau(p) = \log\left(\frac{1}{p}\right),$$

需要注意的是，“事件  $A$  不会发生”的信息量  $\tau(1-p)$ ，将不同于  $\tau(p)$ ，除了当  $p = 0.5$  时

$$\tau(1-p) = \begin{cases} \neq \tau(p) & p \neq 0.5 \\ = \tau(p) & p = 0.5. \end{cases} \quad (18.1)$$

从这个角度，我们可以将  $\tau(p)$  理解为，在知道了“事件  $A$  会以概率  $P(A) = p$  发生”时的“惊讶”程度。将全部  $p \in [0, 1]$  下的惊讶作加权平均，可得期望信息（information） $I$ 。 $I$  可理解为事件  $A$  的熵（entropy）。

现在来看信号“事件  $A$  发生的概率从  $p$  变为  $q$ ”的价值： $|q - p|$  大时价值高，小时价值低，其信息含量之差

$$\tau(p) - \tau(q) = \log\left(\frac{q}{p}\right) = \begin{cases} > 0 & q > p \\ < 0, & q < p. \end{cases}$$

现在将  $A$  扩展到一组  $n$  个事件  $A_1, \dots, A_n$  的情况，其中  $A_i, i = 1, \dots, n$  的信息量为  $\log(q_i/p_i)$ ，那么全部信息的加权期望值为

$$I = \sum_{i=1}^n q_i \log\left(\frac{q_i}{p_i}\right).$$

这样我们可以进一步讨论密度方程的信息含量了。

### 18.1.2 Kullback-Leibler 信息准则

假定对于一组随机变量，存在一个未知的真实密度方程  $g(\xi)$ ，和一个猜测的密度方程  $f(\xi)$ 。由于  $g$  未知，我们将  $g$  相对于  $f$  的 Kullback-Leibler 信息准则 (Kullback-Leibler information criterion, KLIC)  $KL(g : f)$  定义为

$$KL(g : f) := \int_{\mathbb{R}} \log \left( \frac{g(\xi)}{f(\xi)} \right) g(\xi) d\xi, \quad (18.2)$$

用于描述当知道当  $g(\xi)$  是随机变量  $z$  的真实密度之后的期望“惊讶”程度。

**Theorem 18.1** (KLIC 非负).  $KLIC$  (18.2) 满足

$$KL(g : f) \geq 0 \quad (18.3)$$

并且当且仅当  $g \stackrel{a.e.}{=} f$  时，等号成立<sup>1</sup>。

证明. 已知

$$\begin{aligned} \log(1+x) &< x \quad \forall x > -1, \\ \rightarrow \log \left( \frac{g}{f} \right) &= -\log \left( 1 + \frac{f-g}{g} \right) > 1 - \frac{f}{g}, \end{aligned}$$

进而

$$\int_{\mathbb{R}} \log \left( \frac{g}{f} \right) g d\xi > \int_{\mathbb{R}} \log \left( 1 - \frac{f}{g} \right) g d\xi = \int_{\mathbb{R}} (g-f) d\xi = 0.$$

显然， $g \stackrel{a.e.}{=} f$  时， $KL(g : f) = 0$ 。反之，如果已知  $KL(g : f) = 0$ ，那么假设条件  $g \stackrel{a.e.}{=} f$  成立。□

然而值得注意的是，KLIC 第一不适合用作无方向的公制测量单位，这是因为通常来说  $KL(g : f) \neq KL(f : g)$ ，第二不满足三角不等式关系。因此只适合将 KLIC 看做是一个测度密度方程  $f$  和  $g$  的接近程度的简便工具。

设  $\{z_t\} \in \mathbb{R}^{\nu}$  为一个定义在概率空间  $(\Omega, \mathbb{F}, \mathbb{P}_0)$  中的实值随机变量数列<sup>2</sup>。在不产生混淆的情况下，我们将随机变量向量，及其实现 (realization) 表示为相同的符号  $z$ ，用

$$z^T = (z_1, z_2, \dots, z_T)$$

表示  $\{z_t\}$  生成的全部信息集合。向量  $z_t \in \mathbb{R}^{\nu \times 1}$  包括两部分，分别是标量  $y_t$  和向量  $\omega_t \in \mathbb{R}^{(\nu-1) \times 1}$ 。那么对应于  $\mathbb{P}_0$ ，存在一个关于  $z^T$  的联合概率密度方程 (joint probability density function, JPDF)  $g^T(z^T)$ ，满足

$$g^T(z^T) = g(z_1) \prod_{t=2}^T \frac{g^t(z^t)}{g^{t-1}(z^{t-1})} = g(z_1) \prod_{t=2}^T g_t(z_t | z^{t-1}), \quad (18.4)$$

其中

- $g^t$  是 JPDF 表示  $t$  个随机变量  $z_1, z_2, \dots, z_t$  的 JPDF，
- $g_t$  是 PDF，表示基于全部过去信息  $z_1, \dots, z_{t-1}$  基础上，现有随机事件  $z_t$  的密度方程，
- $g(z_1)$  是初始无条件 PDF，

<sup>1</sup> a.e.是几乎处处 (almost everywhere) 的缩写，这里指的  $g = f$  除了一个勒贝格测度为 0 的集合之外，都成立。

<sup>2</sup>  $\Omega$  是样本空间，指所有可能结果 (输出) 的集合； $\mathbb{F}$  是事件集合，单独某一事件或一组事件组成，含有 0 个或更多输出； $\mathbb{P}_0$  是将事件和事件发生概率联系在一起的方程集合。

- 因而我们可以将  $g^T$  理解为一个描述随机行为  $z^T$  的机制，又称  $Z^T$  的数据生成过程 (data generating process, DGP)。

由于真实的 DGP  $g^T$  未知，我们要猜测条件 PDF  $f_t(z_t|z^{t-1};\theta)$  用作  $g^T$  的近似，其中参数向量  $\theta \in \Theta \subseteq \mathbb{R}^k$ 。在此基础上，猜测的 JPDF  $f^T$  为

$$f^T(z^T; \theta) = f(z_1) \prod_{t=2}^T f_t(z_t|z^{t-1}; \theta), \quad (18.5)$$

类似地， $f(z_1)$  也是初始无条件 PDF。上式称为关于  $g^T$  的准似然方程 (quasi-likelihood function)，“准”是因为，基于这样的设定， $f^T$  (18.5) 可以与  $g^T(z^T; \theta)$  (18.4) 不一致。

为了表述方便，将非条件密度写成条件密度的形式  $f(z_1), g(z_1)$ ；进而可以将  $f^T(z^T; \theta), g^T(z^T)$  表示为全部有条件 PDF 的乘积。显然，猜测 JPDF  $f^T$  越是接近真实 DGP  $g^T$ ，信号的价值越高。 $g^T$  相对于  $f^T$  的 KLIC 写为

$$KL(g^T : f^T; \theta) = \int_{\mathbb{R}^T} \log \left( \frac{g^T(\xi^T)}{f^T(\xi^T; \theta)} \right) g^T(\xi^T) d(\xi^T). \quad (18.6)$$

那么可以考虑如下优化方案，对于 DGP  $g^T$ ，选择近似方程  $f^T$  使  $KL(g^T : f^T; \theta)$  最小化，即生成的“惊讶”水平最低。由于  $g^T$  与系数  $\theta$  无关，最小化(18.6)的问题等价于最大化如下问题

$$\int_{\mathbb{R}^T} [\log f^T(\xi^T; \theta)] g^T(\xi^T) d(\xi^T) = E \log f^T(z^T; \theta), \quad (18.7)$$

$E$  表示期望值。最大化(18.7)的问题又进一步等价于最大化条件 JPDF 的平均期望值：汇总  $f_t(z_t|z^{t-1}; \theta) \forall t = 1, \dots, T$ ，有

$$\bar{L}_T(\theta) = \frac{1}{T} E [\log f^T(z^T; \theta)] = \frac{1}{T} E \left[ \sum_{t=1}^T \log f_t(z_t|z^{t-1}; \theta) \right]. \quad (18.8)$$

对(18.8)作最大似然处理，算得最优参数  $\theta^*$

$$\theta^* = \underset{\theta}{\operatorname{argmax}} \bar{L}_T(\theta), \quad (18.9)$$

同时也是(18.8)、(18.7)的最优参数。

**Definition 18.1** (正确设定的识别条件). 如果  $\exists! \theta_0 \in \Theta$ , <sup>3</sup> 满足

$$f^T(z^T; \theta_0) = g^T(z^T) \quad \forall t \in T,$$

那么我们说  $\{f_t(z^t|z^{t-1}; \theta)\}_{t=1}^T$  对  $\{z_t\}_{t=1}^T$  整体上是正确设定的 (*correct specification in its entirety for  $\{z_t\}_{t=1}^T$* )，此时最优系数被识别为  $\theta_0$ ，满足 KLIC 的最小化

$$KL(g^T : f^T; \theta_0) = 0.$$

根据这样的识别条件可得， $\theta^* = \theta_0$ 。

尽管道理上来说是这样，可实际操作过程中(18.9)不易估计：这是由于  $\bar{L}_T(\theta)$  (18.8)的计算涉及到未知 DGP  $g^T$ ，和取期望值的计算。对此，我们常常采取近似的替代方案，将  $\bar{L}_T(\theta)$  替换为一个样本  $\bar{L}_T(\theta)$  下的 QLHF  $L_T(z^T; \theta)$

$$L_T(z^T; \theta) := \frac{1}{T} \sum_{t=1}^T \log f_t(z_t|z^{t-1}; \theta), \quad (18.10)$$

---

<sup>3</sup> $\exists!$  表示存在且存在唯一一个，有时也表示为  $\exists_{=1}$ 。

对应地，最优估计值

$$\tilde{\theta}_T = \operatorname{argmax}_{\theta} L_T(z^T; \theta),$$

称为  $\theta$  的准最大似然估计 (quasi maximum likelihood estimator, QMLE)。加入前缀“准”(quasi)是为了表明，估计的解  $\tilde{\theta}$  来自于一个可能存在误设定问题的对数似然方程  $\log f_t$ 。 $\{f_t\}$  可能并非关于  $\{z_t\}$  整体设定正确，此时  $f^T \neq g^T$ ,  $\tilde{\theta}_T \neq \theta$ ,  $KL > 0$ 。反之若  $\{f_t\}$  设定正确，那么  $f^T = g^T$ ,  $\tilde{\theta}_T = \theta$ ,  $KL = 0$ ，此时准最大似然估计等价于前述的表准最大似然估计。

在现实应用中，为  $z^T$  设置完整的概率模型可能是一项艰巨的任务，涉及到太多随机变量：例如  $T$  个随机变量的向量  $z_t$ ，每个  $z_t$  中都含有  $\nu$  个随机变量。这就需要作以适当简化。对于经济学来说，往往只关心其中的一部分核心内容，例如对随机变量  $y_t$  建模，假定  $y_t$  与前定变量 (predetermined variable, 如 Klein (2000) 的定义) 的向量  $x_t$  有关， $x_t$  中包含元素  $(\omega_t, z^{t-1})$ 。这样一来，模型中只需要考虑  $y_t$  的条件行为：除非  $\omega_t$  还有其他明确描述，条件 PDF  $g_t(y_t|x_t)$  提供了关于  $\{z_t\}$  的一些关键 (而非全部) 信息。在此基础上，可以构建 QLHF  $f_t(y_t|x_t; \theta)$  来近似  $g_t(y_t|x_t)$ ，KLIC 如(18.6)所示。对于全部  $t = 1, \dots, T$ ，可定义平均 KILC 为  $\overline{KL}_T(\{g_t : f_t\}; \theta)$

$$\overline{KL}_T(\{g_t : f_t\}; \theta) := \frac{1}{T} \sum_{t=1}^T KL(g_t : f_t; \theta), \quad (18.11)$$

对应平均条件 JPDF 的平均期望

$$\overline{L}_T(\theta) = \frac{1}{T} \sum_{t=1}^T E[\log f_t(y_t|x_t; \theta)]. \quad (18.12)$$

为了简化表述，设  $y^T = (y_1, \dots, y_T)$ ,  $x^T = (x_1, \dots, x_T)$ 。则优化条件为，利用最大似然估计法求解 QLME  $\theta^*$

$$\theta^* = \operatorname{argmax}_{\theta} \overline{L}_T(\theta),$$

同时也是  $\operatorname{argmin}_{\theta} \overline{KL}_T(\{g_t : f_t\}; \theta)$  的值。根据 Definition (18.1)，识别条件为：如果  $\exists! \theta_0 \in \Theta$ ，满足

$$f_t(y_t|x_t; \theta_0) = g(y_t|x_t) \quad \forall t,$$

那么  $\{f_t\}$  对  $\{y_t|x_t\}$  整体设定正确。进而有  $\overline{KL}_T(\{g_t : f_t\}; \theta_0) = 0$ ，以及  $\theta^* = \theta_0$ 。

和前述情况一样， $\overline{L}_T(\theta)$  (18.12) 无法直接计算，同样地，采用类似于(18.10)的方式，用样本 QLHF  $L_T(y^T, x^T; \theta)$  作近似

$$L_T(y^T, x^T; \theta) := \frac{1}{T} \sum_{t=1}^T \log f_t(y_t|x_t; \theta). \quad (18.13)$$

基于(18.13)作准最大似然估计求得 QMLE  $\tilde{\theta}_T$

$$\tilde{\theta}_T = \operatorname{argmax}_{\theta} L_T(y^T, x^T; \theta).$$

若  $\{f_t\}$  对于  $\{y_t|x_t\}$  完整正确设定，则 QMLE  $\tilde{\theta}_T$  可视为等价于标准 MLE。

现在来考虑另一个情况， $y_t$  具有某些特殊性质，我们猜测该特性表现为条件正态分布，均值和方差分别为  $\mu_t(x_t; \theta)$  和  $\sigma^2$ ，满足

$$y_t|x_t \sim \mathcal{N}(\mu_t(x_t; \beta), \sigma^2).$$

根据高斯分布的性质可以算得

$$\theta = (\beta^\top \sigma^2)^\top.$$

首先列出 QLHF 的计算式及其样本近似

$$\bar{L}(\theta) \approx L_T(y^T, x^T; \theta) = \frac{1}{T} \sum_{t=1}^T \log f(y_t | x_t; \theta),$$

进而 QLME  $\tilde{\beta}_T$

$$\tilde{\beta}_T = \underset{\beta}{\operatorname{argmax}} L_T(y^T, x^T; \theta)$$

同时也有

$$\tilde{\beta}_T = \underset{\beta}{\operatorname{argmin}} \frac{1}{T} \sum_{t=1}^T [y_t - \mu_t(x_t; \beta)]^\top [y_t - \mu_t(x_t; \beta)]$$

关系成立，即它同时也是非线性最小二乘法 (nonlinear regression, NLS) 的估计。由此可见，NLS 可以看做是 QMLE 在某些假定条件下的特例，这个假定条件是：观测数据表现出有条件的正态分布，以及有条件的同方差。识别条件为，设  $\exists! \theta_0$ ，满足

$$\mu_t(x_t; \theta_0) = E(y_t | x_t)$$

时，我们称  $\{\mu_t\}$  对条件均值  $\{E(y_t | x_t)\}$  整体正确设定。

除此而外，也可以考虑一个更加灵活的设定方式，如

$$y_t | x_t \sim \mathcal{N}(\mu_t(x_t; \beta), h(x_t; \alpha)),$$

可以帮助我们进一步分析、估计条件方差。

## 18.2 QMLE 的渐进性质

通常来说，QLHF  $L_T(\cdot; \theta)$  是一个关于系数向量  $\theta$  的非线性方程系统。非线性意味着我们往往无法直接求得解析解 QMLE  $\tilde{\theta}_T$ ，而是要靠非线性优化算法作数值近似。那么在数值近似之前，就需要探讨 QMLE 的金斯特征。本节中，我们假定 QLHF  $L_T(\cdot; \theta)$  在紧凑参数空间  $\Theta$  中的概率为 1，1 次、2 次连续可导，积分和微分运算可交换顺序，并且识别条件成立： $\exists! \theta^* \in \Theta$ ，使得 KLIC (18.2) 或 (18.6) 最大化。要求  $\theta^* \in \Theta$  是为了确保平均值扩展 (18.14) 和随后 QLHF  $L_T$  的渐进分析是有效的。

以关于  $z^T$  的 QLHF  $L_T(z^T; \tilde{\theta}_T)$  为例，由于可能存在误设定问题，QLME  $\tilde{\theta}_T$  可能不同于  $\theta^*$ 。将一阶矩条件  $\nabla L_T(z^T; \tilde{\theta}_T)$  针对  $\theta^*$  作平均值扩展

$$\nabla L_T(z^T; \tilde{\theta}_T) \approx \nabla L_T(z^T; \theta^*) + \nabla^2 L_T(z^T; \theta_T^\dagger)(\tilde{\theta}_T - \theta^*), \quad (18.14)$$

其中  $\theta_T^\dagger \in \Theta$  是个在  $\tilde{\theta}_T$  和  $\theta^*$  之间的参数向量。通过选取合适的  $\tilde{\theta}_T$  值，使得 (18.14) LHS = 0，进而我们有一阶矩条件和二阶矩条件的关系

$$\nabla L_T(z^T; \theta^*) = -\nabla^2 L_T(z^T; \theta_T^\dagger)(\tilde{\theta}_T - \theta^*).$$

假定  $\nabla^2 L_T(z^T; \theta_T^\dagger)$  可逆，即要求 QLHF  $L_T$  在给定的  $\theta^*$  系数下为局部二次形式。集合全部  $\theta_T^\dagger \in \Theta$  的取值，将二阶矩的期望值设为期望海森矩阵 (expected Hessian matrix)

$$H_T(\theta) := E[\nabla^2 L_T(z^T; \theta)]. \quad (18.15)$$

如果  $\nabla^2 L_T(z^T; \theta_T^\dagger)$  遵循弱均匀大数法则 (WULLN, 第16.D节), 那么如下条件在参数空间  $\Theta$  中均匀成立

$$\nabla^2 L_T(z^T; \theta) - \underbrace{E[\nabla^2 L_T(z^T; \theta)]}_{=: H_T(\theta)} \xrightarrow{\mathbb{P}} 0.$$

因此可将(18.14)改写为

$$\sqrt{T}(\tilde{\theta}_T - \theta^*) = -(H_T(\theta^*))^{-1} \sqrt{T} \nabla L_T(z^T; \theta^*) + \mathcal{O}_p(1), \quad (18.16)$$

就是说, QMLE 系数距其真实系数的标准化距离  $\sqrt{T}(\tilde{\theta}_T - \theta^*)$  的渐进分布, 就事实上取决于标准化 core 方程  $\sqrt{T} \nabla L_T(z^T; \theta^*)$  的渐进分布。

将标准化 socre 方程的方差协方差矩阵设为  $B_T(\theta)$

$$B_T(\theta) = \text{var}\left(\sqrt{T} \nabla L_T(z^T; \theta^*)\right), \quad (18.17)$$

又称信息矩阵 (information matrix)。

若假定  $\log f_t(z_t | z^{t-1})$  满足中心极限定理 (CLT, 第16.B.7), 可得

$$(B_T(\theta^*))^{-\frac{1}{2}} \sqrt{T} \nabla L_T(z^T; \theta^*) - \underbrace{E[\nabla L_T(z^T; \theta^*)]}_{=: \mathcal{A}} \xrightarrow{D} \mathcal{N}(0, I_k). \quad (18.18)$$

**Proposition 18.1.** (18.18)中的  $\mathcal{A} = 0$ 。

证明. 根据微分和积分顺序可以互换的假定, 有

$$E[\nabla L_t(z^T; \theta)] = \nabla E[L_T(z^T; \theta)] \nabla \bar{L}_T(\theta), \quad (18.19)$$

其中 RHS 是(18.8)的一阶导数。

根据 QMLE  $\theta^*$  的性质, 以下关系依次成立

$$\begin{aligned} \theta^* &= \underset{\theta}{\operatorname{argmin}} KL(z^T; \theta), \\ &\hookrightarrow KL(z^T; \theta^*) = 0, \\ &\hookrightarrow \nabla \bar{L}_T(\theta^*) = 0, \end{aligned}$$

进一步代入关于  $\theta^*$  的关系式(18.19)可得

$$\mathcal{A} = E[\nabla L_t(z^T; \theta^*)] = 0.$$

□

(18.18)改写为

$$(B_T(\theta^*))^{-\frac{1}{2}} \sqrt{T} \nabla L_T(z^T; \theta^*) \xrightarrow{D} \mathcal{N}(0, I_k)$$

进一步代回(18.16), 改写成

$$\sqrt{T}(\tilde{\theta}_T - \theta^*) = -(H_T(\theta^*))^{-1} B_T(\theta^*)^{\frac{1}{2}} \underbrace{\left[B_T(\theta^*)^{-\frac{1}{2}} \sqrt{T} \nabla L_T(z^T; \theta^*)\right]}_{\xrightarrow{D} \mathcal{N}(0, I_k)} + \mathcal{O}_p(1),$$

表现为渐进的正态分布特征。由此可得以下结论

**Theorem 18.2** (QMLE 的渐进正态分布特征). 若(18.14), (18.16), (18.18)均成立, 那么有 QMLE 系数距其真实系数的标准化距离  $\sqrt{T}(\tilde{\theta}_T - \theta^*)$  表现为渐进正态分布, 满足

$$C_T(\theta^*)^{-\frac{1}{2}} \sqrt{T}(\tilde{\theta}_T - \theta^*) \xrightarrow{D} \mathcal{N}(0, I_k), \quad (18.20)$$

其中方差协方差矩阵  $C_T(\theta^*)$  满足关系

$$C_T(\theta^*) = [H_T(\theta^*)]^{-1} B_T(\theta^*) [H_T(\theta^*)]^{-1}, \quad (18.21)$$

$B_T(\theta^*)$ ,  $H_T(\theta^*)$  分别表示标准化 score 方程的方差协方差矩阵 (信息矩阵) 和期望海森矩阵, 由(18.17)和(18.15)可得

$$\begin{aligned} B_T(\theta^*) &= \text{var}\left(\sqrt{T}\nabla L_T(z^T; \theta^*)\right), \\ H_T(\theta^*) &= E\nabla^2 L_T(z^T; \theta^*). \end{aligned}$$

再来看关于  $\{y_t|x_t\}_{t=1}^T$  的 QLHF  $L_T(y^T, x^T; \theta)$  的情况。类似于, 我们可得 QMLE  $\tilde{\theta}_T$  的渐进正态分布特征

**Corollary 18.1** (QMLE 的渐进正态分布特征). QMLE 系数距其真实系数的标准化距离  $\sqrt{T}(\tilde{\theta}_T - \theta^*)$  表现为渐进正态分布, 满足

$$C_T(\theta^*)^{-\frac{1}{2}} \sqrt{T}(\tilde{\theta}_T - \theta^*) \xrightarrow{D} \mathcal{N}(0, I_k),$$

其中

$$\begin{aligned} C_T(\theta^*) &= [H_T(\theta^*)]^{-1} B_T(\theta^*) [H_T(\theta^*)]^{-1}, \\ B_T(\theta^*) &= \text{var}\left(\sqrt{T}\nabla L_T(y^T, x^T; \theta^*)\right), \\ H_T(\theta^*) &= E\nabla^2 L_T(y^T, x^T; \theta^*). \end{aligned}$$

## 18.3 信息矩阵等式

准最大似然估计的一个重要结论是信息矩阵等式 (information matrix equality, IME), 是说在特定条件下, 信息矩阵  $B_T(\theta)$  (18.17) 等于海森矩阵期望值(18.15)的负数  $-EH_T(\theta)$ 。IME 可以简化方差协方差矩阵(18.21)  $C_T(\theta)$  的计算。

### 18.3.1 一种设定形式下的 IME

先来看设定形式  $\{z^T\}_{t=1}^T$ 。将 score 方程定义为  $s_t(z^t; \theta)$ , 反映对数 PDF 的一阶变化

$$s_t(z^t; \theta) = \nabla \log f_t(z_t | z^{t-1}; \theta),$$

从而

$$\nabla f_t(z_t | z^{t-1}; \theta) = s_t(z^t; \theta) f_t(z_t | z^{t-1}; \theta),$$

汇总全部  $t = 1, \dots, T$  的 score 方程, 有加权平均

$$s^T(z^T; \theta) = \frac{1}{T} \sum_{t=1}^T s_t(z^t; \theta). \quad (18.22)$$

进而有

$$\begin{aligned} s^T(z^T; \theta) f^T(z^T; \theta) &= \left[ \frac{1}{T} \sum_{t=1}^T s_t(z^t; \theta) \right] \left[ \prod_{\tau=1}^T f^\tau(z_\tau | z^{\tau-1}; \theta) \right] \\ &= \frac{1}{T} \sum_{t=1}^T \left[ \nabla \log f_t(z_t | z^{t-1}; \theta) \prod_{\tau=1, \tau \neq t}^T f^\tau(z_\tau | z^{\tau-1}; \theta) \right] \\ &= \frac{1}{T} \nabla f^T(z^T; \theta) \end{aligned} \quad (18.23)$$

根据 PDF 的定义

$$\int_{\mathbb{R}} f^T(z^T; \theta) dz^T = 1,$$

可得导数的积分必为 0,

$$\begin{aligned} \int_{\mathbb{R}} \nabla f^T(z^T; \theta) dz^T &= 0, \\ \int_{\mathbb{R}} \nabla^2 f^T(z^T; \theta) dz^T &= 0. \end{aligned}$$

设积分微分顺序可以互换, 则

$$\frac{1}{T} \int_{\mathbb{R}} \nabla f^T(z^T; \theta) dz^T = \int_{\mathbb{R}} s^T(z^T; \theta) dz^T = 0, \quad (18.24)$$

$$\frac{1}{T} \int_{\mathbb{R}} \nabla^2 f^T(z^T; \theta) dz^T = \int_{\mathbb{R}} \left\{ \nabla s^T(z^T; \theta) + T \left[ s^T(z^T; \theta) s^T(z^T; \theta)^\top \right] \right\} f^T(z^T; \theta) dz^T = 0. \quad (18.25)$$

进一步的分析需要分两种情况来讨论。第一种情况, 设  $\{f_t(z^t | z^{t-1}; \theta)\}_{t=1}^T$  对  $\{z_t\}_{t=1}^T$  整体正确定。那么在真实(但未知的)PDF  $g^T(z^T)$  对应的系数  $\theta_0$  下, 对近似 PDF  $f^T(z^T; \theta_0)$  取期望值, 由(18.24)-(18.25)可得

$$E[s^T(z^T; \theta_0)] = 0, \quad (18.26)$$

$$\underbrace{E[\nabla s^T(z^T; \theta_0)]}_{=: \mathcal{A}} + \underbrace{T E[s^T(z^T; \theta_0) s^T(z^T; \theta_0)^\top]}_{=: \mathcal{B}} = 0, \quad (18.27)$$

这便成为信息矩阵等式。

**Theorem 18.3** (信息矩阵等式). 假定  $\exists \theta_0 \in \Theta$  满足

$$f_t(z_t | z^{t-1}; \theta_0) = g_t(z_t | z^{t-1}; \theta_0),$$

那么可得

$$\begin{aligned} H_T(\theta_0) &= -B_T(\theta_0), \\ H_t(\theta_0) &= E \nabla^2 L_T(z^T; \theta_0), \\ B_T(\theta_0) &= \text{var}(\sqrt{T} \nabla L_T(z^T; \theta_0)). \end{aligned} \quad (18.28)$$

**Corollary 18.2** (QLME 的克拉美罗下界 (CRLB)). 假定 Theorem 18.3 成立, 那么计算标准化距离  $\sqrt{T}(\tilde{\theta}_T - \theta_0)$  的所需的方差协方差矩阵  $C_T(\theta_0)$  (17.57), 可简化为

$$C_T(\theta_0) \geq B_T(\theta_0)^{-1} = -H_T(\theta_0)^{-1}, \quad (18.29)$$

QMLE  $\theta_0$  逐渐逼近克拉美罗下界 (Cramér-Rao lower bound, CRLB)。

第二种情况，如果  $\{f_t(z^t|z^{t-1};\theta)\}_{t=1}^T$  对  $\{z_t\}_{t=1}^T$  整体并非正确设定，那么 score 方程很可能与真实 PDF  $g^T$  无关，换句话说，哪怕系数取值的确是  $\theta^*$ ，我们也无法保证 score 方程的期望均值是 0，(18.26)-(18.27) 改写为

$$E[s^T(z^T;\theta_0)] \neq 0, \quad (18.30)$$

$$\underbrace{E[\nabla s^T(z^T;\theta_0)]}_{=:A} + \underbrace{T E[s^T(z^T;\theta_0)s^T(z^T;\theta_0)^\top]}_{=:B} \neq 0, \quad (18.31)$$

在这种情况下，信息矩阵等式(18.28)不成立。

### 18.3.2 另一种 score 方程定义下的 IME

来看另一种设定形式  $\{y_t|x_t\}_{t=1}^T$ 。类似于(18.23)，有

$$\nabla f_t(y_t|x_t;\theta) = s_t(y_t|x_t;\theta) f_t(y_t|x_t;\theta), \quad (18.32)$$

$$\nabla f^T(y^T|x^T;\theta) = s^T(y^T,x^T;\theta) f^T(y^T|x^T;\theta), \quad (18.33)$$

并且有

$$\int_{\mathbb{R}} f_t(y_t|x_t;\theta) dy_t = 1.$$

那么类似地，

$$\begin{aligned} \int_{\mathbb{R}} s_t(y_t,x_t,\theta_t) f_t(y_t|x_t;\theta) dy_t &= 0, \\ \int_{\mathbb{R}} [\nabla s_t(y_t,x_t;\theta) + s_t(y_t,x_t;\theta)(y_t,x_t;\theta)^\top] f_t(y_t|x_t;\theta) dy_t &= 0. \end{aligned}$$

如果  $\{f_t\}_{t=1}^T$  关于  $\{y_t|x_t\}_{t=1}^T$  整体正确设定，那么

$$E[s_t(y_t,x_t;\theta)|x_t] = 0,$$

进而根据迭代期望法则 (law of iterated expectations, LIE, 第18.3.4节) 可得

$$E[s_t(y_t,x_t;\theta_0)] = 0, \quad (18.34)$$

因此可见 score 方程的平均值是正确设定的。

根据条件方差的分解 (Lemma 18.1)，(18.34) 改写为

$$E[\nabla s_t(y_t,s_t;\theta_0)|x_t] + E[s_t(y_t,s_t;\theta_0)s_t(y_t,s_t;\theta_0)^\top|x_t] = 0,$$

并且这个等式对  $t = 1, \dots, T$  均成立。加总可得

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T E[\nabla s_t(y_t,s_t;\theta_0)] + \frac{1}{T} \sum_{t=1}^T E[s_t(y_t,s_t;\theta_0)s_t(y_t,s_t;\theta_0)^\top] \\ = H_t(\theta_0) + \frac{1}{T} \sum_{t=1}^T E[s_t(y_t,s_t;\theta_0)s_t(y_t,s_t;\theta_0)^\top] = 0, \end{aligned} \quad (18.35)$$

在满足特定前提假设的条件下，上式等价于 IME。

### 18.3.3 动态误设定

需要指出的是, (18.35)并不必然总是等价于 IME。例如, 考虑一组对  $\{y_t|x_t\}_{t=1}^T$  整体正确设定了的  $\{f_t(y_t|x_t; \theta)\}_{t=1}^T$ 。如果改设定对  $\{y_t|\omega_t, z^{t-1}\}_{t=1}^T$  而言并非正确设定, 那么我们称之为动态误设定 (dynamic misspecification):  $\exists \theta_0 \in \Theta$ , 使得满足  $f_t(y_t|x_t; \theta_0) = g_t(y_t|\omega_t, z^{t-1})$ , 换句话说,  $\{\omega_t, z^{t-1}\}$  所包含的信息, 无法由  $\{x_t\}$  所完全表现出来。

在不存在动态误设定的情况下, 根据定义我们有

$$E[s_t(y_t, x_t; \theta_0)|x_t] = E[s_t(y_t, x_t; \theta_0)|\omega_t, z^{t-1}]$$

进而

$$\begin{aligned} B_T(\theta_0) &= \frac{1}{T} E \left[ \left( \sum_{t=1}^T s_t(y_t, x_t; \theta_0) \right) \left( \sum_{t=1}^T s_t(y_t, x_t; \theta_0) \right)^\top \right] \\ &= \frac{1}{T} \sum_{t=1}^T E[(s_t(y_t, x_t; \theta_0))(s_t(y_t, x_t; \theta_0))^\top] \\ &\quad + \underbrace{\frac{1}{T} \sum_{\tau=1}^{T-1} \sum_{t=\tau+1}^T E[s_{t-\tau}(y_{t-\tau}, x_{t-\tau}; \theta_0) s_t(y_t, x_t; \theta_0)^\top]}_{=: \mathcal{A}_1} \\ &\quad + \underbrace{\frac{1}{T} \sum_{\tau=1}^{T-1} \sum_{t=\tau+1}^T E[s_t(y_t, x_t; \theta_0) s_{t+\tau}(y_{t+\tau}, x_{t+\tau}; \theta_0)^\top]}_{=: \mathcal{A}_2} \\ &= \frac{1}{T} \sum_{t=1}^T E[s_t(y_t, x_t; \theta_0) s_t(y_t, x_t; \theta_0)^\top], \end{aligned} \tag{18.36}$$

根据 LIE (18.40) 和条件方差分解 (18.42) 可得

$$\mathcal{A}_1 = \mathcal{A}_2 = 0.$$

另一方面, 在可能存在动态误设定的情况下, (18.36) 的等号不成立: score 方程的协方差可能不为 0。这表明, 个别信息矩阵的平均值  $\text{var}(s_t(y_t, x_t; \theta_0))$  可以不是信息矩阵。

**Theorem 18.4** (信息矩阵等式). 假设  $\{y_t|x_t\}_{t=1}^T$  的设定形式, 并且  $\exists \theta_0 \in \Theta$  满足  $f_t(y_t|x_t; \theta) = g_t(y_t|x_t)$ , 并且其中不存在动态误设定的情况。那么我们有以下 IME 成立

$$\begin{aligned} H_T(\theta_0) &= -B_T(\theta_0), \\ H_T(\theta_0) &= E[\nabla^2 L_T(z^T; \theta_0)] = \frac{1}{T} \sum_{t=1}^T E[\nabla s_t(y_t, x_t; \theta_0)], \\ B_T(\theta_0) &= \text{var}(\sqrt{T} \nabla L_T(z^T; \theta)) = \frac{1}{T} \sum_{t=1}^T E[s_t(y_t, x_t; \theta_0) s_t(y_t, x_t; \theta_0)^\top]. \end{aligned} \tag{18.37}$$

**Corollary 18.3** (QLME 的克拉美罗下界 (CRLB)). 假定 Theorem (18.4) 成立, 那么计算标准化距离  $\sqrt{T}(\tilde{\theta}_T - \theta_0)$  的所需的方差协方差矩阵  $C_T(\theta_0)$  (17.57), 可简化为

$$C_T(\theta_0) \geq B_T(\theta_0)^{-1} = -H_T(\theta_0)^{-1}, \tag{18.38}$$

QMLE  $\theta_0$  逐渐逼近克拉美罗下界 (Cramér-Rao lower bound, CRLB)。

### 18.3.4 迭代期望法则和条件方差分解

迭代期望法则 (law of iterated expectations, LIE), 可简单表示为 (Wooldridge, 2002, pp. 29)

$$E[y] = E_x [E_y(y|x)]. \quad (18.39)$$

现在将  $x \in \mathbb{R}^M$  理解为一个离散向量, 向量中包含随机变量  $c_1, \dots, c_M$ , 对应概率  $p_1, \dots, p_M$ , 则 LIE 是指

$$\begin{aligned} E[y] &= p_1 E[y|x=c_1] + p_2 E[y|x=c_2] + \dots + p_M E[y|x=c_M] \\ &= \sum_{j=1}^M p_j E[y|x=c_j], \end{aligned} \quad (18.40)$$

可见  $E_y[y]$  就是一组  $E_y[y|x]$  的加权平均, 权重  $p_j$  是  $x$  值取  $c_j$  的概率, 可以表示为

$$E[y] = E_x [E_y(y|x)], \quad (18.41)$$

即加权平均的加权平均。

**Lemma 18.1** (条件方差的分解). 给定  $x$  后  $y$  的条件方差  $\text{var}(y|x)$ , 可分解为两部分之和, 一是条件均值的方差, 二是围绕条件均值的期望方差

$$\text{var}(y|x) = \text{var}(E(y|x)) + E_x [\text{var}(y|x)]. \quad (18.42)$$

证明.

$$\begin{aligned} \text{var}(y|x) &= E[(y - E(y))^2] = E\left[\left(\underbrace{y - E(y|x)}_{=0} + E(y|x) - E(y)\right)^2\right] \\ &= E[(y - E(y|x))^2] + E[(E(y|x) - E(y))^2] + 2\underbrace{E[(y - E(y|x))(E(y|x) - E(y))]}_{=: \mathcal{C}}. \end{aligned}$$

已知

$$\begin{aligned} E[y|x] &= E[(y - E(y|x))|x] = E(y|x) - E(y|x) = 0, \\ \hookrightarrow \mathcal{C} &= 0, \end{aligned}$$

进一步引入 LIE 关系(18.40), 上式改写为

$$\begin{aligned} \text{var}(y|x) &= E_x \left\{ E[(y - E(y|x))^2] \right\} + E \left\{ E[(E(y|x) - E(y))^2] \right\} \\ &= E_x [\text{var}(y|x)] + \text{var}_x [E(y|x)], \end{aligned}$$

证毕。 □

## 18.4 一个例子

考虑这样的设定

$$y_t | x_t \sim \mathcal{N}(x_t^\top \beta, \sigma^2), \quad \forall t = 1, \dots, T. \quad (18.43)$$

根据高斯正态分布的性质, 设系数矩阵  $\theta$  满足

$$\theta = (\beta^\top, \sigma^2)^\top.$$

对数 LHF  $\log f_t(y_t|x_t; \theta)$  满足

$$\log f_t(y_t|x_t; \theta) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma^2) - \frac{(y_t - x_t^\top \beta)^2}{2\sigma^2}$$

将全部  $t = 1, \dots, T$  个对数 LHF 加权求和, 设为样本的平均 LHF  $L_T(y^T, x^T; \theta)$

$$L_T(y^T, x^T; \theta) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma^2) - \frac{1}{T} \sum_{t=1}^T \frac{(y_t - x_t^\top \beta)^2}{2\sigma^2}.$$

进而可计算 1、2 次导数

$$\nabla L_T(y^T, x^T; \theta) = \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} \frac{x_t(y_t - x_t^\top \beta)}{\sigma^2} \\ -\frac{1}{2\sigma^2} + \frac{(y_t - x_t^\top \beta)^2}{2(\sigma^2)^2} \end{pmatrix}, \quad (18.44)$$

$$\nabla^2 L_T(y^T, x^T; \theta) = \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} -\frac{x_t x_t^\top}{\sigma^2} & -\frac{x_t(y_t - x_t^\top \beta)}{(\sigma^2)^2} \\ -\frac{(y_t - x_t^\top \beta) x_t^\top}{(\sigma^2)^2} & \frac{1}{2(\sigma^2)^2} - \frac{(y_t - x_t^\top \beta)^2}{(\sigma^2)^3} \end{pmatrix}, \quad (18.45)$$

(18.44) 是 socre 方程矩阵, 据此可进行准最大似然估计:

- $\beta$  的 QMLE  $\hat{\beta}_T$ :

$$\hat{\beta}_T = \underset{\beta}{\operatorname{argmax}} L_T(y^T, x^T; \theta) = \underset{\beta}{\arg} \left\{ \nabla L_T(y^T, x^T, \theta) = 0 \right\},$$

- $\sigma^2$  的 QMLE  $\hat{\sigma}^2$  就是 OLE 残差的平均值:

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T (y_t - x_t^\top \beta)^2.$$

假设上述设定对  $\{y_t|x_t\}_{t=1}^T$  整体设定正确, 那么  $\exists \theta_0 = (\beta_0^\top \sigma_0^2)^\top$ , 使得在给定  $x_t$  的情况下,  $y_t$  的真实分布  $g^T(y^T, x^T; \theta_0)$  满足

$$g^T(y^T, x^T) \sim \mathcal{N}(x_t^\top \beta_0, \sigma_0^2),$$

据此可计算海森矩阵  $H_T(\theta)$ :

- 对均值取期望, 得

$$\begin{aligned} E[x_t(y_t - x_t^\top \beta)] &= E(x_t x_t^\top)(\beta_0 - \beta) \\ &= 0 \quad \text{如果 } \beta = \beta_0. \end{aligned} \quad (18.46)$$

- 对方差取期望, 得

$$\begin{aligned} E[(y_t - x_t^\top \beta)^2] &= [(y_t - x_t^\top \beta_0 + x_t^\top \beta_0 - x_t^\top \beta)^2] \\ &= E[(y_t - x_t^\top \beta_0)^2] + E[(x_t^\top \beta_0 - x_t^\top \beta)^2] + \underbrace{2E[(y_t - x_t^\top \beta_0)(x_t^\top \beta_0 - x_t^\top \beta)]}_{=0} \\ &= \sigma_0^2 + E[(x_t^\top \beta_0 - x_t^\top \beta)^2] \\ &= \sigma_0^2 \quad \text{如果 } \beta = \beta_0. \end{aligned} \quad (18.47)$$

- 将(18.46)-(18.47)代回(18.45), 算得海森矩阵  $H_T(\theta)$

$$H_T(\theta) = E[\nabla^2 L_T(\theta)] = \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} -\frac{E[x_t x_t^\top]}{\sigma^2} & -\frac{E(x_t x_t^\top)(\beta_0 - \beta)}{(\sigma^2)^2} \\ -\frac{(\beta_0 - \beta)^\top E(x_t x_t^\top)}{(\sigma^2)^2} & \frac{1}{2(\sigma^2)^2} - \frac{\sigma_0^2 + E[(x_t^\top \beta_0 - x_t^\top \beta)^2]}{(\sigma^2)^3} \end{pmatrix}, \quad (18.48)$$

对于  $\theta = \theta_0 \iff (\beta^\top, \sigma^2)^\top = (\beta_0^\top, \sigma_0^2)^\top$ , 代入上式, 有

$$H_T(\theta_0) = \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} -\frac{E(x_t x_t^\top)}{\sigma_0^2} & 0 \\ 0 & -\frac{1}{2(\sigma_0^2)^2} \end{pmatrix}. \quad (18.49)$$

再来看信息矩阵  $B_T(\theta)$  的计算。若假设不存在动态误设定, 那么

$$B_T(\theta) = \frac{1}{T} \sum_{t=1}^T E \begin{pmatrix} \frac{(y_t - x_t^\top \beta)^2 x_t x_t^\top}{(\sigma^2)^2} & -\frac{x_t(y_t - x_t^\top \beta)}{2(\sigma^2)^2} + \frac{x_t(y_t - x_t^\top \beta)^3}{2(\sigma^2)^3} \\ \frac{(y_t - x_t^\top \beta)x_t^\top}{2(\sigma^2)^2} + \frac{(y_t - x_t^\top \beta)^3 x_t^\top}{2(\sigma^2)^3} & \frac{1}{4(\sigma^2)^2} - \frac{(y_t - x_t^\top \beta)^2}{2(\sigma^2)^3} + \frac{(y_t - x_t^\top \beta)^4}{4(\sigma^2)^4} \end{pmatrix}. \quad (18.50)$$

根据假定,  $y_t$  满足条件正态分布, 那么它的 3、4 阶条件中心距 (central moment formula) 分别为

central moment formula

•

$$\begin{aligned} E[(y_t - x_t^\top \beta)^3] &= -3\sigma_0^2 E[(x_t^\top \beta_0 - x_t^\top \beta)] - E[(x_t^\top \beta_0 - x_t^\top \beta)^3] \\ &= 0 \quad \text{如果 } \beta_0 = \beta, \end{aligned}$$

•

$$\begin{aligned} E[(y_t - x_t^\top \beta)^4] &= 3(\sigma_0^2)^2 + 6(\sigma_0^2) E[(x_t^\top \beta_0 - x_t^\top \beta)^2] + E[(x_t^\top \beta_0 - x_t^\top \beta)^3] \\ &= 3(\sigma_0^2)^2 \quad \text{如果 } \beta_0 = \beta. \end{aligned}$$

对于  $\theta = \theta_0 \iff (\beta^\top, \sigma^2)^\top = (\beta_0^\top, \sigma_0^2)^\top$ , 代回(18.50), 有  $B_T(\theta_0)$

$$B_T(\theta_0) = \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} \frac{E[x_t x_t^\top]}{\sigma_0^2} & 0 \\ 0 & \frac{1}{2(\sigma_0^2)^2} \end{pmatrix} \quad (18.51)$$

结合(18.49), (18.51)可见

$$B_T(\theta_0) = -H_T(\theta_0),$$

即 IME 成立。

设  $H_T(\tilde{\theta}_T)$  是一个关于  $H_T(\theta_0)$  的一致估计,  $\tilde{\theta}_T = (\tilde{\beta}_T^\top \tilde{\sigma}_T^2)^\top$ , 进而

$$H_T(\tilde{\theta}_T) = - \begin{pmatrix} -\frac{\sum_{t=1}^T (x_t x_t^\top)}{T \tilde{\sigma}_T^2} & 0 \\ 0 & -\frac{1}{2(\tilde{\sigma}_T^2)^2} \end{pmatrix}. \quad (18.52)$$

在此基础上由 IME 条件可得,  $B_T(\theta_0)$  的一致估计  $B_T(\tilde{\theta}_T)$  可由下式算得

$$B_T(\tilde{\theta}_T) = -H_T(\tilde{\theta}_T).$$

值得注意的是,  $H_T(\tilde{\theta}_T)$  的左上分块矩阵是对  $\tilde{\beta}_T$  系数的方差协方差矩阵的标准估计: 若存在动态误设定的情况, 则  $B_T(\theta)$  的估计值不同于上式, 这导致 IME 不成立。

## 18.5 异方差问题

哪怕模型对  $E(y_t|x_t)$  设定正确，不存在动态误设定的情况，仍然可能由于其他一些条件矩的误设定问题，如条件异方差 (heteroskedasticity)，而使得 IME 不成立。举例来说，在第18.4节，设定(18.43)的基础上，额外假定分布方差服从 DGP 以下过程

$$y_t|x_t \sim \mathcal{N}(x_t^\top \beta, h(x_t^\top \gamma_0)), \quad (18.53)$$

即条件条件方差  $h(x_t^\top \gamma_0)$  随着  $x_t$  而变化，存在条件异方差问题。

这个设定，尽管就其条件均值  $(x_t^\top \beta)$  而言是设定正确的，但却并不是对  $\{y_t|x_t\}_{t=1}^T$  整体的正确设定，因为它存在条件异方差。这样一来，根据上一节例子的思路，海森矩阵  $H_T(\theta)$  的左上分块值由(18.48)变为

$$-\frac{1}{T\sigma^2} \sum_{t=1}^T E(x_t x_t^\top), \quad (18.54)$$

而信息矩阵  $B_T(\theta)$  的左上分块值由 (18.50)仍然是

$$\frac{1}{T} \sum_{t=1}^T \frac{E[(y_t - x_t^\top \beta)^2]}{\sigma^2} \frac{E[x_t x_t^\top]}{\sigma^2} \quad (18.55)$$

由于当前设定对条件均值是正确的，对条件方差是不正确的，可以将对应的 KLIC 最小估计设为  $\theta^* = (\beta_0^\top, \sigma^{*,2})^\top$ 。在  $\theta = \theta^*$  处的海森矩阵分块(18.54)和信息矩阵分块(18.55)分别为

$$\begin{aligned} &-\frac{1}{T\sigma^{*,2}} \sum_{t=1}^T E(x_t x_t^\top), \\ &\frac{1}{T(\sigma^{*,2})^2} \sum_{t=1}^T E(h(x_t^\top \gamma_0) x_t x_t^\top), \end{aligned}$$

显然二者不相等：尽管条件方差设定“正确”，但整体上的 IME 不成立。

对  $H_T(\theta^*), B_T(\theta^*)$  的一致估计  $H_T(\tilde{\theta}_T), B_T(\tilde{\theta}_T)$  分别为

$$\begin{aligned} H_T(\tilde{\theta}_T) &= \begin{pmatrix} -\frac{\sum_{t=1}^T x_t x_t^\top}{T\tilde{\sigma}_T^2} & 0 \\ 0 & \frac{1}{2(\tilde{\sigma}_T^2)^2} \end{pmatrix}, \\ B_T(\tilde{\theta}_T) &= \begin{pmatrix} \frac{\sum_{t=1}^T \hat{e}_t^2 x_t x_t^\top}{T(\tilde{\sigma}_T^2)^2} & 0 \\ 0 & -\frac{1}{4(\tilde{\sigma}_T^2)^2} + \frac{\sum_{t=1}^T \hat{e}_t^4}{T^4 (\tilde{\sigma}_T^2)^4} \end{pmatrix}, \end{aligned}$$

其中  $\hat{e}_t$  是 OLS 估计的残差。 $H_T(\tilde{\theta}_T), B_T(\tilde{\theta}_T)$  都是块对角矩阵，因此  $H_T(\tilde{\theta}_T)^{-1}, B_T(\tilde{\theta}_T) H_T(\tilde{\theta}_T)^{-1}$  矩阵的左上分块是

$$\left( \frac{1}{T} \sum_{t=1}^T x_t x_t^\top \right)^{-1} \left( \frac{1}{T} \sum_{t=1}^T \hat{e}_t x_t x_t^\top \right) \left( \frac{1}{T} \sum_{t=1}^T x_t x_t^\top \right)^{-1},$$

是 OLS 估计  $\hat{\beta}_T$  的方差协方差矩阵的 Eicker-White estimator。

虑一下，要不  
做个 section?

## **第五部分**

### **附录**



# Appendices



# 第一章 后凯恩斯主义经济学的一个小综述

主要参考自 Hart and Kriesler (2015)。

## A.1 Introduction

## A.2 PK 的主要方法论

## A.3 PK 下的市场结构与定价

### A.3.1 PK 下的市场结构

### A.3.2 PK 下的定价策略

markup pricing principle 的一系列变体，如

- Kalecki (1937) 考虑到企业往往面临内部/外部的融资需求，(其他企业所定的)price markup 会影响到企业的现金流，进而使得金融部门，企业融资行为和企业 (家庭) 投资决策这三者之间存在重大关联，可见货币和金融的确会对实体经济产生影响 (Ball, 1964; Eichner, 1973; Wood, 1975; Harcourt and Kenyon, 1976)。
- (企业自身所定的)price markup，也会被理解为企业自己在多变、激烈和难以预测的市场竞争中寻求立足、发展的关键决策之一。

因此价格更多反映企业的利益决策而非其所在产业或市场的条件；价格是基于企业战略而非基于成本制定的；这一价格反映市场的非均衡状态，对于整个经济体动态发展起到重要作用。

PK 视野下，企业不是追求短期利润最大化的行为人，而是在面临巨大不确定性条件下的长期决策制定者。企业的产量和定价决策反映了寡头垄断的市场条件，企业自身及其竞争对手的行为，以及融资情况都是该条件的重要组成部分。

## A.4 PK 下的宏观经济

MS 更多遵循萨伊定律，即供应会产生它所对应的需求，导致就业和产出都不受需求的限制。PK 经济体更多强调有效需求，货币非中性，不确定情况下的决策，以及主观并且变化迅速的期望，这使得 PK 对萨伊定律提出挑战。值得指出的是，该挑战并不依赖于对市场完全竞争程度或者“粘性”的假定。由此 PK 坚决反对任何将凯恩斯经济学与一般均衡框架融合在一起的尝试，即反对新古典主义综合。PK 也不同于 New Keynesian，后者致力于将一系列摩擦 (frictions) 引入到 New Classical 体系当中 Akerlof (2007)。在 PK 看来，这两种思路忽略了发达资本主义经济体中最本质的不稳定因素，从而仍然使市场运作暴露在重大风险中。

延续 Keynes 和 Kalecki 的思路，PK 强调真实产出和就业的根本决定因素是有效需求的水平，真实产出和就业的波动主要是由投资支出的变化所导致的，投资支出受到对周边情形的期望（“动物精神”）的影响。对于“储蓄增加导致投资增加”的传统观点，PK 持相反意见：只要能够寻找到合适的融资渠道，投资通过乘数效应可以增加收入，进而提高储蓄。

PK 在几个方面修正了 Keynes 的理论体系，尤其是在对金融市场、制度的描述上。在 Keynes 看来，不确定性是人们持有货币以追求保值的最关键原因，持有货币的行为将人们过去已经发生了的经济行动和不确定性的未来连接在一起，进一步扩大了不确定性的风险。如 Kaldor (1985) 对货币主义者的批判，货币供应的内生性被过分夸大了：随着货币供应增加，金融机构可以放出更多贷款，导致存款增加和/或对金融财产的购买增加。这一系列借入和借出的决策取决于融资成本以及对未来的期望——金融财产的规模和组成结构直接取决于 balance sheet 中放贷者和潜在借款人的立场。他们的乐观或悲观心态，会影响到金融市场整体的行为，进而放大、波及对实体经济运转情况的看法。从这个意义上讲，金融不稳定性假设认为资本主义经济体中，金融部门和实体经济部门的不稳定性是彼此相关和不可避免的 (Minsky, 2015)，导致中央银行和财政稳定政策的努力注定无效，“有效市场”及其相关资产定价模型也需要被否定。

PK 不再关注于供应-需求互动决定的均衡价格和产品数量，而是关注产业结构和 markup 定价原则，这使得对产品相对价格的分析不再有意义：在宏观经济层面上，产出就业和总物价均发生波动。excess capacity 的存在意味着即便需求增加了，产出也会跟着调整以满足需求，从而不会出现通货膨胀的压力。当逼近 full capacity utilisation 的情况下，通货膨胀的压力也主要来自 mark-up pricing 机制，而不是原材料、工资等要素成本。就业水平主要取决于有效需求水平，而不是实际工资。实际工资不取决于市场，而是取决于以下两方面因素，第一如上文所说的企业定价策略，第二是围绕全部国民收入，在劳动收入和资本收入之间的讨价还价，这就涉及到随时间变化的制度性要素。在 PK 看来，通货膨胀是国民收入中劳动收入和资本收入比重分配不恰当所导致的，也只有重新调整比重分配才能缓解通货膨胀——这呼吁一种 permanent income policy 的出台。

PK 对经济增长和经济发展尤为关注，在 PK 看来它们是资本主义经济发展动态的最核心要素。然而在 PK 看来，它们都不是平稳和连续的。

## A.5 PK 视野下的经济政策

与主流经济学衍生出的经济政策相比，根据 PK 所得的经济增侧往往有些反直觉。总的说来，PK 认为市场运作在短期和长期都存在失灵和低效率情况，持干预市场运作的立场。

# 第二章 Keynesian, New Keynesian and New Classical Economics

主要参考自 Greenwald and Stiglitz (1987)。

简要说来。New Classical Economics 学派认为，macro-theory 缺乏 micro 基础，需要从一系列微观行为准则（理性、最大化的企业和个人行为）推导出经济总体层面上的动态；承认动态对于了解宏观行为的重要性；在分析动态行为时，期望扮演核心角色，如理性预期的形成。“理性预期学派”的问题并不在于其假定之一，理性的预期是不是可能，而在于另一个古典主义假定，市场出清。根据这样的假定，理性预期模型会得到如下结果：没有失业，政府宏观政策是不必要的。

另一大流派 New Keynesian Economics 学派，认为现实世界中真实存在的一系列经济现象，如失业、信贷配给、经济周期等，都是不能由标准 micro-theory 所解决的。因此致力于开发 micro 理论，来解释 macro 现象。NK 学派的核心工作就是从 micro 和 macro 层面了解不充分信息和不完全市场及其后果。市场失灵中的表现之一是失业。与 New Classical 相比，NK 同样追求建立一个 single 解释框架，但 NK 的框架致力于解释失业，而不像 New Classical 学派那样忽视甚至否定失业的存在。

## B.1 Keynesian 的四个核心

NK 从 traditional Keynesian 经济学发展而来。Keynes 对经济运行有着显著不同于 standard neo-classical 理论的看法，比如企业行为决策并不是基于理性的计算，而是“动物精神”。他用图形的形式表示他的观点，但当他试图构建理论模型去描述这一观点的时候，他更多回到了传统的 neo classical 框架中——这很有可能是因为他受过严格的 neo-classical 训练，束缚了他表达自己全新思想的能力。Hicks 等后人将 Keynes 的观点进一步模型化系统化理论化，成为理解失业、经济周期等问题的强有力工具。

Keynes 的观点中，如下四点对于我们构建一个解释失业和经济周期波动的模型来说至关重要。

### B.1.1 失业与有效工资理论

第一点。该通用模型必须能够解释失业的持续存在，以及一系列关键经济变量的周期性变化。为了解释失业，就需要发展出一个关于劳动力市场的理论，Keynes 从传统经济学思想（如 fixed price 学派）中借鉴了（名义，实际）工资粘性的假定，但 Keynesian 从经验和理论两个层面批评了 Fixed price 学派的工资粘性假定是不完善的。经验层面，大萧条中工资下跌了  $1/3$ ，经历通货膨胀的国家即使调整了实际工资，仍然出现失业情况。理论层面，Fixed Price 学派没能对工资粘性做出解释，没能提供制度上的设计。

根据 Keynes 的理论，工资并不需要是完全粘性的，而仅仅需要确保工资不会跌落到完全市场出清的水平，如有效工资理论 (efficiency wage theories(Stiglitz, 1984, 1987))。有效工资理论的假

定如下：

- 对异质劳动者属性特征的信息掌握是不充分的，
- 劳动者的工作表现是无法充分监控的，
- 无法与劳动者签订可以完全反映他劳动成果的工作合同。

根据这样的假定，有效工资理论认为：

- 劳动力的质量及生产率，企业的利润都可能随着工资的增加而增加，
- 工资增加可能导致劳动者周转率降低，
- 由于企业需要承担部分的劳动者周转成本，工资上升在某种程度上可以导致企业利润的上升。

然而当失业存在，工资仍可能不会降低，因为企业认识到如果他们降低员工工资，会导致生产率降低，周转率上升，利润降低。这是指充分竞争市场下，企业只在一定范围内是工资制定者的情况。如果瓦尔拉斯工资 (Walrasian wage)——劳动力需求等于劳动力供应情况下对应的工资水平——相当低，企业有可能有动机提高自己员工的工资以提高利润，对应的有效工资 (efficiency wage) 就是企业利润最大化条件下的工资水平。显然外部经济情况的不同会导致有效工资的不同，使得工资并不是完全刚性的；实际工资并不会跌落到市场出清那么低的水平上。这就肯定了政策干预的重要性，如失业救济等会影响到市场上的均衡工资水平。

### B.1.2 价格变化与经济波动

第二点。该通用模型不仅能够解释失业的存在，还要能够解释失业等核心变量的波动。分为两个问题。

#### 1. 导致经济体出现大幅度波动的冲击，是怎么产生的？

- 战争、油价波动等外生原因。
- 内部机制。
  - 对投资、尤其是存货的需求变动。然而在理论上，对于 concave 生产函数、低工资、低利率的情况，应当存在生产平滑效果：存货应当降低而不是扩大经济波动。
  - 储蓄的变化。储蓄有助于稳定消费。在储蓄不足的情况下，implicit contracts 提供的保障机制起到稳定收入、进而稳定消费的作用。因此 implicit contracts 应当是降低而不是扩大经济波动。

Keynes 正确地强调了投资变化对理解经济波动的重要性。但他用动物精神 (unexplained changes in expectations) 来解释投资的变化。这种解释不能让人完全满意：会导致经济周期理论具有无法解释甚至是非理性的特征。

#### 2. 根据 typical micro 模型，利率、工资等价格的变化有助于缓解需求或供应的波动，进而如存货一样起到稳定经济波动的效果。因此大的外部冲击 (影响需求或供应) 只可能导致经济体中均衡价格的小幅度变化。然而为什么工资、利率等价格往往是刚性的？

Keynes 不仅需要解释为什么对投资的需求曲线 shifted，还需要解释为什么利率调整无法有效缓解投资需求变化所产生的波动。Keynes 的回答不能让人完全满意：他认为对投资需求的降低会导致利率的降低，然而真实世界中当出现经济衰退，利率有时上升；甚至有时利率的变化不会影响到公司的经济决策。

Keynes 只是指出工资可以保持不变，但未能解释工资为什么会保持不变。有效工资理论解释了为什么工资不会跌落到市场出清条件那么低的水平。[Akerlof and Yellen \(1985\)](#) 的 near-rational 行为模型解释了为什么企业可以调整工资而往往倾向于不这么做。类似地，一个资本市场理论解释了为什么利率不会降低到那么低的水平 ([Stiglitz and Weiss, 1981, 1983, 2009](#))，甚至常常是变动很小的。同时，有效工资理论也强调了一个企业制定的工资往往和其他企业制定的工资之间存在互动关系。

总之，上述理论对工资，利率，和价格刚性做出了解释，它们互相作用，放大了对经济系统的冲击，导致在波动往往很剧烈而不是很缓和 (multiplier effects)。

### B.1.3 储蓄与投资，信贷配给

Keynes 敏锐注意到了家庭部门的储蓄和生产部门的投资是不同的，要加以区分。然而在构建模型时，Keynes 又回到 standard neo-classical 模型的老路上，储蓄如何转化为投资，不做进一步分析；投资由利率决定，因此不存在信贷配给 (credit rationing)。这使得尽管 Keynes 强调资本市场的不充分性特征，却也无法做深入的探讨。

### B.1.4 供应与技术进步

在 Marshallian 的框架中，经济的波动是由供应波动和需求波动两方面共同作用的结果。Keynes 基于对大萧条的分析，将精力投入到分析导致需求波动的诸因素。然而对于供应端的波动，以及技术进步，他并没有做深入解释。

## B.2 New Keynesian 经济学的四个核心

NK 是在 Keynes 思想的基础上发展起来的，但也注意到了 Keynes 过度依赖于 neoclassical 理论分析框架，以及虽然认识到了资本市场是不完善的，但没能充分认识到不完善所带来的信息成本损耗及其经济效果。NK 的核心有以下几点。

### B.2.1 有效工资

有效工资模型，用于解释非刚性工资为什么不会完全跌落到市场出清条件下的低水平。

### B.2.2 资产配给

资本市场是不充分竞争的，表现在企业经理人与潜在投资者之间的信息不对称，信息不对称导致资产配给问题。资产配给问题的重要性在于，如果企业想要获得更多资本用于扩大生产，就需要从市场上借入更多资金，承担更大的投资风险。风险还表现在，在缺乏期货市场的情况下，企业当期筹资用于扩大生产，在未来时间里用增加的产出换取利润抵付，这就存在风险。在分析企业经营决策时，就必须考虑企业的风险承受意愿。

### B.2.3 信贷配给

有意愿承担风险扩大生产的企业，还面临借不到钱的问题，即信贷配给问题。与 neoclassical 的分析框架不同，资本品提供者在面临供不应求的情况下，可能会选择不提高利率，其逻辑类似于失业（劳动力供大于求）情况下企业的工资定价：提高利率可能会使资本品所有者的期望回报率降低，其原因之一是选择效应 (selection effects，申请借贷人的集合向不利于借出者的方向变化)，其

原因之一是激励效应 (incentive effects, 借入者为了偿还更高的利率, 会倾向于采取更激进、风险更高的生产形式)。

#### B.2.4 货币政策

NK 中货币政策的重要性, 比起表现在个人持有货币维持账目平衡来, 更多体现在信贷的可获得性上。信息不对称导致若商业银行决定缩紧发放贷款, 需要资金的生产者很难找到其他合适的放贷人。此外商业银行的放贷决策也类似于企业生产者的生产决策, 充满不确定性和风险。中央银行可以通过货币政策影响商业银行的放贷意愿。

#### B.2.5 小结

总之, NK 提供一套解释经济运行的通用理论, 该理论基于微观经济学原理而建立。与 traditional Keynesian 理论相比, NK

1. 弥补了传统 K 理论的不一致性, 如
  - (a) 内部不一致, 如期望的形成机制,
  - (b) 预测和实际观察情况之间的不一致,
2. 填补了传统 K 理论的空白, 如传统 K 仅仅假定工资是刚性的, NK 则至少在一定程度上解释了工资为什么是刚性的。

NK 进一步探讨了

1. 失业问题, 如有效工资理论
2. 经济周期波动问题, 其基本逻辑是: 对经济系统的冲击影响到企业实际利用资本的存量。
  - (a) 信贷市场问题, 企业面对不充分的信贷市场, 信息不对称, 融资有风险; 即使信贷市场是完全的 (企业能以合理的利率得到自己所需的全部融资), 企业融资数额仍然受自己风险承受能力的限制。
  - (b) 根据借款契约的束缚, 企业借贷从事生产活动需要承诺的义务是固定的, 随着 working capital 的降低, 借贷风险 (破产的概率) 升高, 生产活动的规模降低, 并且需要过了一段时间才能使 working capital 缓慢回升到正常水平。

这使得 NK 不仅可以解释总量层面上的冲击 (如货币冲击导致的物价暴跌) 如何影响总量层面上的经济系统, 还可以解释部门层面上的冲击 (如未提前预知的 demand shift, 或操作价格的石油联盟等) 如何影响总量层面上的经济系统。从而生产意愿成为关于 working capital 的 concave 函数, 并且 working capital 的再分配也会产生总量效果。

### B.3 凯恩斯的不足

NK 能够解释分析一些现象, 这些现象都是 traditional Keynesian 理论所很难解释的, 如

1. 为什么经济衰退中的企业不降低产品价格, 即 markup 的周期运动模式,
2. 投资和存货的周期性变化特征,
3. 失业者在愿意降低工资重新求职的情况下, 为什么仍然很难再就业,

4. 为什么一个 unanticipated 工资-物价下降，通过降低企业的 working capital 存量，会导致经济衰退恶化而非好转，

下面做详细说明。

Keynes 最重要的错误在于他的企业理论，以及对货币如何影响经济活动的解释上。这两点不足都是由于他未能充分理解资本市场的本质。

### B.3.1 债券和股票的区别

Keynes 未能正确区分长期债券和长期股票，而将二者统称为长期资产。然而区别的确存在：

1. 风险性。即便不考虑破产清算问题，二者在风险性上有本质区别：在经济衰退中，债券的价值上升，股票的价值下降。因此在个人理财账户中，二者是互补品而非替代品。
2. 企业承诺。借入债券/借贷的企业，需要承诺在指定日期偿还；而对于股票，企业则不需要受此承诺约束。从这个角度来说，对于企业和投资者而言，二者远非完全替代品。尤其在经济衰退中，需要融资的企业很少借助股票渠道，见 Greenwald et al. (1984) 的 adverse selection model。

### B.3.2 需求和供应

Keynes 从需求端对经济波动的解释，有助于回答“为什么企业不改变定价策略，降价促销”，但该解释进一步导致下个问题的出现：为什么一个开放经济体存在 Keynesian 失业的情况？只需要它改变汇率，就可以产生对该经济体商品近乎无限的需求。

NK 不对供应和需求作明确的区分。在有明确需求的情况下，企业倾向于扩大生产（供应），此时需求引导生产；在需求不够明确的情况下，企业未必会扩大生产，而会基于供应曲线从事经济活动。这样，NK 可以解释为什么在经济周期中，企业愿意生产的产品数量是波动的。此外 NK 还可以结识为什么在经济衰退时期，（可以调整自己产品价格的）企业倾向于基于生产成本，做一个更高的 price markup。在不充分竞争和不完全信息的市场上，企业在制定产品价格时，面临现在和未来利润的权衡：今天降价导致今天利润下降，以及明天销量上升、利润上升。衰退时期企业产品定价更高的原因在于，生产企业面对更高的资本成本（并不是指市场利率），以及更严重的资产配给问题。

### B.3.3 投资的决定因素

Keynes 认为在给定期望水平的情况下，投资的最主要决定因素是利率。不论这里的“利率”应当是指实际利率还是名义利率<sup>1</sup>，通过对现实世界的观察能够发现，实际利率相对于其他经济变量，波动相对较小。NK 模型中在解释投资时，作为解释变量的（实际）利率不应当是个常数或近似常数。

此外 NK 认为，投资还受到特定时段信贷可获得程度的影响。所谓特定时段，是指货币政策会对经济活动产生影响的时期。在现实世界中，当面临经济衰退的时候，就算银行仍然愿意以现行利率将款项借出给有良好资质的企业，但很可能愿意借入的企业是不足的，这导致货币政策失效。

Keynesian-neoclassical 理论无法解释存货波动的问题，即为什么存货往往激化而不是缓和经济波动。NK 可以解释这个问题，认为资本的有效成本增加意味着企业倾向于在经济衰退时期减少库存。资本有效成本的增加是由于 equity rationing 和/或 working capital 供应量的减少所导致的。

---

<sup>1</sup> 目前经济学界通常认为，实际利率更为恰当。

### B.3.4 货币政策

在 Keynes 的分析框架中，是不可能通过货币政策影响经济运行的。从最简单的思路出发，可行的方案分三步走：

1. 政府采取改变货币供应量的举措；

NK：至少对完成交易这一目的而言，存在着货币的替代物，可能使得政府改变货币供应的举动无法完成既定目标。相当多的交易可以不依赖于货币完成，仅靠信用足矣，这使得许多基于 cash-in-advance 的模型无法较好解释现实。此外交易和收入之间的关系较弱：大多数交易体现为财产的交换，而经济周期波动往往伴随着财富的变化，进而财产分配的变化。

2. 在给定个人货币需求函数（假定个人对货币的需求取决于利率和收入）的前提下，利率发生变化；

NK：既然对货币的需求往往是基于财产的，那么显然财富而非收入成为个人经济行为的关注要素。在交易目的考虑之外，短期债券可以充当货币的完全替代品，此时持有货币的机会成本表现为短期货币利率。但影响投资行为的利率必须是实际利率<sup>2</sup>。

3. 利率变化导致投资变化。

## B.4 小结

方法论。

一系列 facts：市场是很复杂的。经济模型应当致力于描述市场的核心特征，而非完美复制市场中的一切。企业和个人今天的行为决策受到昨天决策的影响，基于对未来的期望。人们对未来的期望是非理性的。市场的确存在但是不完全的。价格的确调整。失业率较高时，工资会下降。

这些 facts 给经济模型构建者带来挑战。无法建立这样一个动态模型来完美反映未来的一切 facts，因为未来是不可完全预知的。构建模型应当致力于解决其中一些核心问题（而非全部），比如失业。

政策。长久以来关于在出现失业时政府应当做些什么，存在较大争议：a) 自由放任，什么都不做，b) 致力于降低工资以提升就业，c) 使用货币政策，d) 扩大政府支出。Keynes 理论的成功在于它为持有 d) 立场的人提供了理论依据。New Classical 理论为持有 a) 立场的人提供依据

市场效率。NK 认同 K 的观点，不充分竞争、不完全信息导致市场失灵，从而导致失业的确是资本主义经济面临的核心难题之一。

---

<sup>2</sup> 此外，Cash Management Accounts 方面的最新研究表明，完全可以通过提供 interest bearing "money" 来替代直接持有货币的方案，此时经济个体的最优决策就变成了考虑他所拟持有债务的 maturity structure。

## 参考文献

- Adda J, Cooper R. Dynamic Economics: Quantitative Methods and Applications. London: MIT Press, 2003.
- Adolfson M, Las en S, Lind J, Villani M. Evaluating an estimated new Keynesian small open economy model. *Journal of Economic Dynamics and Control* 2008a;32(8):2690–2721.
- Adolfson M, Laséen S, Lindé J, Villani M. Bayesian estimation of an open economy DSGE model with incomplete pass-through. *Journal of International Economics* 2007;72(2):481–511.
- Adolfson M, Laséen S, Lindé J, Villani M. Evaluating an estimated new Keynesian small open economy model. *Journal of Economic Dynamics and Control* 2008b;32(8):2690–2721.
- Agranovich MS. Sobolev Spaces, Their Generalizations and Elliptic Problems in Smooth and Lipschitz Domains. Springer Monographs in Mathematics. Cham: Springer International Publishing, 2015.
- Akerlof GA. The Missing Motivation in Macroeconomics. *The American Economic Review* 2007;97(1):3–36.
- Akerlof GA, Yellen JL. A Near-Rational Model of the Business Cycle, With Wage and Price Inertia. *The Quarterly Journal of Economics* 1985;100:823.
- Aldrich EM, Kung H. Computational Methods for Production-Based Asset Pricing Models with Recursive Utility. Technical Report 87; 2011.
- Altig D, Christiano LJ, Eichenbaum M, Lindé J. Firm-specific capital, nominal rigidities and the business cycle. *Review of Economic Dynamics* 2011;14(2):225–247.
- Alvarez F, Jermann UJ. Using Asset Prices to Measure the Cost of Business Cycles. *Journal of Political Economy* 2004;112(6):1223–1256.
- Amemiya T. Advanced Econometrics. Harvard University Press, 1985.
- An S, Schorfheide F. Bayesian analysis of DSGE models - Rejoinder. *Econometric Reviews* 2007;26(2-4):211–219.
- Ancona A. Elliptic operators, conormal derivatives and positive parts of functions (with an appendix by Haïm Brezis). *Journal of Functional Analysis* 2009;257(7):2124–2158.
- Anderson EW, Hansen LP, McGrattan ER, Sargent TJ. On the Mechanics of Forming and Estimating Dynamic Linear Economies. In: Amman HM, Rust J, editors. *Handbook of Computational Economics*. Amsterdam: North-Holland; 1996. p. 171–252.

- Anderson GS. A Reliable and Computationally Efficient Algorithm for Imposing the Saddle Point Property in Dynamic Models. IFAC Proceedings Volumes 1998;31(16):355–366.
- Anderson GS, Moore G. A linear algebraic procedure for solving linear perfect foresight models. Economics Letters 1985;17(3):247–252.
- Andreasen MM, Fernández-Villaverde J, Rubio-Ramírez JF. The Pruned State-Space System for Non-Linear DSGE Models: Theory and Empirical Applications. Technical Report 2016-07; 2016.
- Ang WT. Hypersingular Integral Equations in Fracture Analysis . Cambridge: Woodhead Publishing, 2013.
- Angrist JD, Pischke JS. Mostly Harmless Econometrics. An Empiricist's Companion. Princeton University Press, 2008.
- Aruoba SB, Fernández-Villaverde J, Rubio-Ramírez JF. Comparing Solution Methods for Dynamic Equilibrium economies. Journal of Economic Dynamics and Control 2006;30(12):2477–2508.
- Atkinson AC. Plots, Transformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis. Clarendon Press, 1985.
- Atkinson KE. The Numerical Solution of Boundary Integral Equations. University of Iowa, 1996.
- Atkinson KE. The Numerical Solution of Integral Equations of the Second Kind. Cambridge University Bridge, 1997.
- Babuška I. Error-Bounds for Finite Element Method. Numerische Mathematik 1971;16(4):322–333.
- Babuška I. The Finite Element Method with Lagrangian Multipliers. Numerische Mathematik 1973;20(3):179–192.
- Babuška I, Suri M. The P and H-P Versions of the Finite Element Method, Basic Principles and Properties. SIAM Review 1994;36(4):578–632.
- Balke NS, Canova F, Milani F, Wynne MA. DSGE Models in Macroeconomics: Estimation, Evaluation, and New Developments. Emerald Group Publishing Limited, 2012.
- Ball RJ. Inflation and the Theory of Money 1964;.
- Bapat RB. Linear Algebra and Linear Models. Springer Science & Business Media, 2012.
- Barthelmann V, Novak E, Ritter K. High Dimensional Polynomial Interpolation on Sparse Grids. Advances in Computational Mathematics 2000;12(4):273–288.
- Bastani H, Guerrieri L. Advances in Automatic Differentiation. In: Bischof CH, Bücker HM, Hovland P, Naumann U, Utke J, editors. Advances in Automatic Differentiation. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. p. 303–313.
- Basu S. Intermediate Goods and Business Cycles: Implications for Productivity and Welfare. The American Economic Review 1995;85(3):512–531.
- Bayoumi T, Faruqee H, Laxton D, Karam PD, Rebucci A, Lee J, Hunt B, Tchakarov I. GEM: A New International Macroeconomic Model. International Monetary Fund, 2004.

- Behrndt J, ter Elst AFM. Dirichlet-to-Neumann Maps on Bounded Lipschitz Domains. *Journal of Differential Equations* 2015;259(11):5903–5926.
- Bellman R. Dynamic Programming. Princeton University Press, 1957.
- Bender CM, Orszag SA. Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory. New York: Springer, 2013.
- Benhabib J, Farmer REA. Indeterminacy and Sunspots in Macroeconomics. In: *Handbook of Macroeconomics*. Elsevier; 1999. p. 387–448.
- Benigno P, Woodford M. Optimal Monetary and Fiscal Policy: A Linear-Quadratic Approach. *NBER Macroeconomics Annual* 2003;18:271–333.
- Bernanke BS. **Inflation Expectations and Inflation Forecasting**. 2007.
- Bernanke BS, Boivin J, Eliasz P. Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach. *The Quarterly Journal of Economics* 2005;:387–422.
- Bernanke BS, Gertler M, Gilchrist S. The financial accelerator in a quantitative business cycle framework. In: *Handbook of Macroeconomics*. Elsevier; 1999. p. 1341–1393.
- Binder M, Pesaran MH. Multivariate Rational Expectations Models and Macroeconomic Modelling: A Review and Some New Results. In: Pesaran MH, Wickens MR, editors. *Handbook of Applied Econometrics: Macroeconomics*. Oxford: Faculty of Economics, University of Cambridge; 1995. p. 139–187.
- Blanchard OJ, Kahn CM. The Solution of Linear Difference Models under Rational Expectations. *Econometrica* 1980;48(5):1305–1311.
- Bloom N. The Impact of Uncertainty Shocks. *Econometrica* 2009;77:623–685.
- Bocola L. The Pass-Through of Sovereign Risk. *Journal of Political Economy* 2016;124(4):879–926.
- Bogachev VI. Measure Theory Volume I. Springer, 2007a.
- Bogachev VI. Measure Theory Volume II. Springer, 2007b.
- Boldrin M, Christiano LJ, Fisher JDM. Habit Persistence, Asset Returns, and the Business Cycle. *The American Economic Review* 2001;91(1):149–166.
- Bouakez H, Cardia E, Ruge-Murcia FJ. Habit Formation and the Persistence of Monetary Shocks. *Journal of Monetary Economics* 2005;52(6):1073–1088.
- Box GE, Cox DR. An analysis of transformations. *Journal of the Royal Statistical Society Series B (Methodological)* 1964;:211–252.
- Boyd JP. Chebyshev and Fourier Spectral Methods. 2nd ed. Dover Publications, 2001.
- Boyd JP, Petschek R. The Relationships Between Chebyshev, Legendre and Jacobi Polynomials: The Generic Superiority of Chebyshev Polynomials and Three Important Exceptions. *Journal of Scientific Computing* 2013;59(1):1–27.

- Boyd S, Vandenberghe L. Convex Optimization. Cambridge University Press, 2004.
- Bramble JH. The Lagrange Multiplier Method for Dirichlet's Problem. *Mathematics of Computation* 1981;37(155):1–11.
- Brenner SC, Scott RL. The Mathematical Theory of Finite Element Methods. volume 15 of *Texts in Applied Mathematics*. New York, NY: Springer, 2008.
- Brubakk L, Husebø TA, Maih J, Olsen K, Østnor M. Finding NEMO: Documentation of the Norwegian economy model. Technical Report; 2006.
- Brumm J, Scheidegger S. Using Adaptive Sparse Grids to Solve High-Dimensional Dynamic Models. *Econometrica* 2017;85(5):1575–1612.
- Buiter WH. The unfortunate uselessness of most 'state of the art' academic monetary economics. MPRA paper 2009;.
- Bungartz HJ, Griebel M. Sparse Grids. *Acta Numerica* 2004;13:147–269.
- Caldara D, Fernández-Villaverde J, Rubio-Ramírez JF, Yao W. Computing DSGE Models with Recursive Preferences and Stochastic Volatility. *Review of Economic Dynamics* 2012;15(2):188–206.
- Callahan JJ. Advanced Calculus: A Geometric View. Undergraduate Texts in Mathematics. New York, NY: Springer New York, 2010.
- Calvo GA. Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics* 1983;12(3):383–398.
- Canova F. Bridging cyclical DSGE models and the raw data. Manuscript 2009;.
- Canova F. Methods for Applied Macroeconomic Research. Princeton University Press, 2011.
- Canova F, Ferroni F, Matthes C. Choosing the Variables to Estimate Singular DSGE Models. *Journal of Applied Econometrics* 2014;29(7):1099–1117.
- Canova F, Nicoló GD. Monetary disturbances matter for business fluctuations in the G-7. *Journal of Monetary Economics* 2002;49(6):1131–1159.
- Carroll CD, Overland J, Weil DN. Comparison Utility in a Growth Model. *Journal of Economic Growth* 1997;2(4):339–367.
- Carroll CD, Overland J, Weil DN. Saving and Growth with Habit Formation. *The American Economic Review* 2000;90(3):341–355.
- Chadha JS, Corrado L. Macro-prudential policy on liquidity: What does a DSGE model tell us? *Journal of Economics and Business* 2012;64(1):37–62.
- Chadha JS, Corrado L, Holly S. Reconnecting money to inflation: the role of the external finance premium. Cambridge Working Papers in Economics 2008;0852.

- Chen X, Swanson NR. Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis. Essays in Honor of Halbert L. White Jr. Springer Science & Business Media, 2013.
- Cheng AHD, Cheng DT. Heritage and early history of the boundary element method. *Engineering Analysis with Boundary Elements* 2005;29(3):268–302.
- Cho JO, Cooley TF, Kim HSE. Business cycle uncertainty and economic welfare. *Review of Economic Dynamics* 2015;18(2):185–200.
- Christiano LJ. Solving Dynamic Equilibrium Models by a Method of Undetermined Coefficients. *Computational Economics* 2002;20(1):21–55.
- Christiano LJ, Eichenbaum M. Liquidity Effects and the Monetary Transmission Mechanism. *The American Economic Review* 1992;82(2):346–353.
- Christiano LJ, Fisher JDM. Algorithms for solving dynamic models with occasionally binding constraints. *Journal of Economic Dynamics and Control* 2000;24(8):1179–1232.
- Christiano LJ, Martin Eichenbaum , Evans CL. Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. *Journal of Political Economy* 2005;113(1):1–45.
- Christiano LJ, Trabandt M, Walentin K. Chapter 7 - DSGE Models for Monetary Policy Analysis. In: Friedman BM, editor. *Handbook of Monetary Economics*. Elsevier; 2010. p. 285–367 T2 –.
- Christiano, Lawrence J , Eichenbaum, Martin , Evans, Charles L . Chapter 2 Monetary policy shocks: What have we learned and to what end? *Handbook of Macroeconomics* 1999;1:65–148.
- Christoffel K, Coenen G, Warne A. The New Area-wide Model of the Euro Area. *European Central Bank Working Paper Series* 2008;:122.
- Ciarlet PG. The Finite Element Method for Elliptic Problems. 2nd ed. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104), 2002.
- Clough RW. The Finite Element Method in Plane Stress Analysis. In: Proceedings of the nd ASCE Conference on Electronic Computation. American Society of Civil Engineers; 1960. p. 1–35.
- Clough RW, Wilson EL. Early finite element research at Berkeley. In: Fifth US National Conference on Computational Mechanics. Berkeley; 1999. .
- Cochrane JH. Shocks. *Carnegie-Rochester Conference Series on Public Policy* 1994;41:295–364.
- Cochrane JH. Asset Pricing. (Revised Edition). Princeton University Press, 2000.
- Constantinides GM. Habit Formation: A Resolution of the Equity Premium Puzzle. *Journal of Political Economy* 1990;98(3):519–543.
- Cook DR, Weisberg S. Residual and Influence in Regression. Chapman and Hall, 1982.
- Cooley TF. Frontiers of Business Cycle Research. Princeton University Press, 1995.

- Costabel M. Boundary Integral Operators on Lipschitz Domains: Elementary Results. *SIAM Journal on Mathematical Analysis* 1988;19(3):613–626.
- Costabel M, Stephan E. Boundary integral equations for mixed boundary value problems in polygonal domains and Galerkin approximation. *Banach Center Publications* 1985;15(1):175–251.
- Dautray R, Lions JL. Mathematical Analysis and Numerical Methods for Science and Technology: Volume 4 Integral Equations and Numerical Methods. Berlin, Heidelberg: Springer, 1990.
- De Walque G, Smets F, Wouters R. Firm-specific Production Factors in a DSGE Model with Taylor Price Setting. 2006.
- DeJong DN, Dave C. Structural Macroeconometrics. Princeton University Press, 2007.
- Del Negro M, Schorfheide F. Forming priors for DSGE models (and how it affects the assessment of nominal rigidities). *Journal of Monetary Economics* 2008;55(7):1191–1208.
- Del Negro M, Schorfheide F. Bayesian macroeconomics. *The Oxford handbook of Bayesian ...* 2011;.
- Delvos FJ. d-Variate Boolean interpolation. *Journal of Approximation Theory* 1982;34(2):99–114.
- Demkowicz L. Computing with HP-Adaptive Finite Elements: Volume I One and Two Dimensional Elliptic and Maxwell Problems, 2006.
- Demkowicz L, Kurtz J, Pardo D, Paszenski M, Rachowicz W, Zdunek A. Computing with hp-Adaptive Finite Elements: Volume II Frontiers Three Dimensional Elliptic and Maxwell Problems with Applications. CRC Press, 2007.
- Den Haan WJ, De Wind J. Nonlinear and stable perturbation-based approximations. *Journal of Economic Dynamics and Control* 2012;36(10):1477–1497.
- Den Haan WJ, Marcet A. Solving the Stochastic Growth Model by Parameterizing Expectations. *Journal of Business & Economic Statistics* 1990;8.
- Den Haan WJ, Marcet A. Accuracy in Simulations. *The Review of Economic Studies* 1994;61(1):3–17.
- Di Nezza E, Palatucci G, Valdinoci E. Hitchhiker's guide to the fractional Sobolev spaces. *Bulletin des Sciences Mathématiques* 2012;136(5):521–573.
- Díaz-Giménez J. Linear-Quadratic Approximations: An Introduction. In: Computational Methods for the Study of Dynamic Economies. OUP Oxford; 1999. p. 13–29.
- Diebold FX, Kilian L, Nerlove M. Time series analysis. In: Macroeconometrics and Time Series Analysis. London: Palgrave Macmillan, London; 2010. p. 317–342.
- Dixit AK, Stiglitz JE. Monopolistic Competition and Optimum Product Diversity. *The American Economic Review* 1977;67(3):297–308.

- Edge RM, Kiley MT, LaForte JP. A comparison of forecast performance between federal reserve staff forecasts, simple reduced-form models, and a DSGE model. *Journal of Applied Econometrics* 2010;25(4):720–754.
- Eichner AS. A Theory of the Determination of the Mark-up Under Oligopoly. *The Economic Journal* 1973;83(332):1184–1200.
- Erceg CJ, Guerrieri L, Gust C. SIGMA: A New Open Economy Model for Policy Analysis. *International Journal of Central Banking* 2006;.
- Erceg CJ, Henderson DW, Levin AT. Optimal monetary policy with staggered wage and price contracts. *Journal of Monetary Economics* 2000;46(2):281–313.
- Fagan G, Henry Jrm, Mestre R. An area-wide model for the euro area. *Economic Modelling* 2005;22(1):39–59.
- Faust J. The robustness of identified VAR conclusions about money. *Carnegie-Rochester Conference Series on Public Policy* 1998;49:207–244.
- Fernald J. A Quarterly, Utilization-Adjusted Series on Total Factor Productivity. Technical Report; Federal Reserve Bank of San Francisco; 2009.
- Fernández-Villaverde J. The econometrics of DSGE models. *SERIEs* 2010;1(1-2):3–49.
- Fernández-Villaverde J, Gordon G, Guerrón-Quintana PA, Rubio-Ramírez JF. Nonlinear Adventures at the Zero Lower Bound. *Journal of Economic Dynamics and Control* 2015;57:182–204.
- Fernández-Villaverde J, Gurerrón-Quintana P, Rubio-Ramírez JF, Uribe M. Risk Matters: The Real Effects of Volatility Shocks. *American Economic Review* 2011;101:2530–2561.
- Fernández-Villaverde J, Levintal O. Solution Methods for Models with Rare Disasters. NBER Working Papers 2016;.
- Fernández-Villaverde J, Rubio-Ramírez JF. Comparing Dynamic Equilibrium Models to Data: A Bayesian Approach. *Journal of Econometrics* 2004;123(1):153–187.
- Fernández-Villaverde J, Rubio-Ramírez JF. Solving DSGE models with perturbation methods and a change of variables. *Journal of Economic Dynamics and Control* 2006;30(12):2509–2531.
- Fernández-Villaverde J, Rubio-Ramírez JF. Estimating Macroeconomic Models: A Likelihood Approach. *The Review of Economic Studies* 2007;74(4):1059–1087.
- Fernández-Villaverde J, Rubio-Ramírez JF, Santos MS. Convergence Properties of the Likelihood of Computed Dynamic Models. *Econometrica* 2006;74(1):93–119.
- Fernández-Villaverde J, Rubio-Ramírez JF, Schorfheide F. Solution and Estimation Methods for DSGE Models. In: Taylor JB, Uhlig H, editors. *Handbook of Macroeconomics* Vol. 2. Elsevier; 2016. p. 527–724.
- Fornberg B. *A Practical Guide to Pseudospectral Methods* 1996;.
- Fortin M. An Analysis of the Convergence of Mixed Finite Element Methods. *RAIRO Analyse numérique* 1977;4:341–354.

- Fries TP, Belytschko T. The extended/generalized finite element method: An overview of the method and its applications. *International Journal for Numerical Methods in Engineering* 2010;84(3):253–304.
- Frisch R. New methods of measuring marginal utility. volume 3. Tübingen: Mohr, 1932.
- Frisch R. A Complete Scheme for Computing All Direct and Cross Demand Elasticities in a Model with Many Sectors. *Econometrica* 1959;27(2):177–196.
- Fuhrer JC. Habit Formation in Consumption and Its Implications for Monetary-Policy Models. *The American Economic Review* 2000;90(3):367–390.
- Galí J. Monetary Policy, inflation, and the Business Cycle: An introduction to the new Keynesian Framework. Princeton University Press, 2008.
- Galor O. Discrete Dynamical Systems. Springer Science & Business Media, 2007.
- Gaspar J, Judd KL. Solving Large-Scale Rational-Expectations Models. *Macroeconomic Dynamics* 1997;1(1):45–75.
- Gilchrist S. Comment on: Banking and interest rates in monetary policy analysis: A quantitative exploration. *Journal of Monetary Economics* 2007;54(5):1508–1514.
- Glaser A, Liu X, Rokhlin V. A Fast Algorithm for the Calculation of the Roots of Special Functions. *SIAM Journal on Scientific Computing* 2007;29(4):1420–1438.
- Goodfriend M. Narrow Money, Broad Money, and the Transmission of Monetary Policy . In: Models and Monetary Policy: Research in the Tradition of Dale Henderson, Richard Porter, and Peter Tinsley. Board of Governors of the Federal Reserve System, 2005. .
- Goodfriend M, McCallum BT. Banking and interest rates in monetary policy analysis: A quantitative exploration. *Journal of Monetary Economics* 2007;54(5):1480–1507.
- Gordon G. Computing Dynamic Heterogeneous-Agent Economies: Tracking the Distribution. PIER Working Paper 2011;.
- Gourieroux C, Monfort A, Renault E. Indirect inference. *Journal of Applied Econometrics* 1993;8(S1):S85–S118.
- Gradshteyn IS, Ryzhik IM. Table of Integrals, Series, and Products. Academic Press, 2014.
- Greenwald B, Stiglitz JE. Keynesian, New Keynesian and New Classical Economics. *Oxford Economic Papers* 1987;39(1):119–133.
- Greenwald B, Stiglitz JE, Weiss A. Informational Imperfections in the Capital Market and Macroeconomic Fluctuations. *The American Economic Review* 1984;74(2):194–199.
- Greenwood J, Hercowitz Z, Huffman GW. Investment, Capacity Utilization, and the Real Business Cycle. *The American Economic Review* 1988;78(3):402–417.
- Guerrón-Quintana PA. Refinements on macroeconomic modeling: The role of non-separability and heterogeneous labor supply. *Journal of Economic Dynamics and Control* 2008;32(11):3613–3630.

- Guiggiani M, Krishnasamy G, Rudolphi TJ, Rizzo FJ. A General Algorithm for the Numerical Solution of Hypersingular Boundary Integral Equations. *Journal of Applied Mechanics* 1992;59(3):604–614.
- Hansen GD. Indivisible labor and the business cycle. *Journal of Monetary Economics* 1985;16(3):309–327.
- Hansen LP, Heaton JC, Li N. Consumption Strikes Back? Measuring Long-Run Risk. *Journal of Political Economy* 2008;116(2):260–302.
- Hansen LP, Sargent TJ. Misspecification in Recursive Macroeconomic Theory. Book Manuscript, 2004.
- Hansen LP, Sargent TJ. Recursive Models of Dynamic Linear Economies. Princeton University Press, 2013.
- Harcourt GC, Kenyon P. PRICING AND THE INVESTMENT DECISION. *Kyklos* 1976;29(3):449–477.
- Harrison R, Nikolov K, Quinn M, Ramsay G, Scott A, Thomas R. The Bank of England Quarterly Model. Technical Report; 2005.
- Hart N, Kriesler P. Post-Keynesian Economics . Technical Report 2015 ECON 12; UNSW Business School; 2015.
- Hatanaka M, Yamada H. Co-trending: A Statistical System Analysis of Economic Trends. Springer Science & Business Media, 2012.
- Hayashi F. Tobin's Marginal q and Average q: A Neoclassical Interpretation. *Econometrica* 1982;50(1):213–224.
- Heer B, Maussner A. Dynamic General Equilibrium Modeling. Computational Methods and Applications. Berlin, Heidelberg: Springer Science & Business Media, 2009.
- Heinonen J. Lectures on Lipschitz Analysis. University of Jyväskylä, 2005.
- Herbst EP, Schorfheide F. Bayesian Estimation of DSGE Models. Princeton University Press, 2015.
- Hörmander L. The Analysis of Linear Partial Differential Operators I: Distribution Theory and Fourier Analysis. Vienna: Springer, 1983a.
- Hörmander L. The Analysis of Linear Partial Differential Operators II: Differential Operators with Constant Coefficients. Classics in Mathematics. Berlin, Heidelberg: Springer, 1983b.
- Hörmander L. The Analysis of Linear Partial Differential Operators III: Pseudo-Differential Operators. Classics in Mathematics. Berlin, Heidelberg: Springer, 1994a.
- Hörmander L. The Analysis of Linear Partial Differential Operators IV: Fourier Integral Operators. Classics in Mathematics. Berlin, Heidelberg: Springer, 1994b.
- Hsiao GC, Wendland WL. A Finite Element Method for Some Integral Equations of the First Kind. *Journal of Mathematical Analysis and Applications* 1977;58:449–481.

- Hughes TJ. *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*. Dover Publications, 2000.
- Inada KI. On a Two-Sector Model of Economic Growth: Comments and a Generalization. *Review of Economic Studies* 1963;30(2):119–127.
- Jarrow R, Protter P. **Structural** Versus Reduced Form Models: a New Information Based Perspective. *Journal of Investment Management* 2004;2(2):1–10.
- Jin HH, Judd KL. Perturbation methods for general dynamic stochastic models . Bank of England, 2002.
- Jovanović BS, Süli E. Analysis of Finite Difference Schemes: For Linear Partial Differential Equations with Generalized Solutions. volume 46. London: Springer, 2014.
- Judd KL. Projection methods for solving aggregate growth models. *Journal of Economic Theory* 1992;58(2):410–452.
- Judd KL. Numerical Methods in Economics. MIT Press, 1998.
- Judd KL, Guu SM. Perturbation Solution Methods for Economic Growth Models. In: *Economic and Financial Modeling with Mathematica®*. New York, NY: Springer, New York, NY; 1993. p. 80–103.
- Judd KL, Guu SM. Asymptotic methods for asset market equilibrium analysis. *Economic Theory* 2001;18(1):127–157.
- Judd KL, Guu SM. Asymptotic Methods for Asset Market Equilibrium Analysis. *Economic Theory* 2003;18(1):127–157.
- Judd KL, Maliar L, Maliar S. How to Solve Dynamic Stochastic Models Computing Expectations Just Once. NBER Working Paper 2011a;.
- Judd KL, Maliar L, Maliar S. Numerically Stable and Accurate Stochastic Simulation Approaches for Solving Dynamic Economic Models. *Quantitative Economics* 2011b;2(2):173–210.
- Judd KL, Maliar L, Maliar S, Valero R. Smolyak Method for Solving Dynamic Economic Models: Lagrange Interpolation, Anisotropic Grid and Adaptive Domain. *Journal of Economic Dynamics and ...* 2014;44(1):92–123.
- Jung M, Steinbach O. A Finite Element-Boundary Element Algorithm for Inhomogeneous Boundary Value Problems. *Computing* 2002;68(1):1–17.
- Kaldor N. The Scourge of Monetarism 1985;.
- Kalecki M. The Principle of Increasing Risk. *Economica* 1937;4(16):440–447.
- Kashyap AK, Rajan R, Stein JC. Banks as Liquidity Providers: An Explanation for the Coexistence of Lending and Deposit-Taking. *Journal of Finance* 2002;57(1):33–73.
- Kim J, Kim S, Schaumburg E, Sims CA. Calculating and using second-order accurate solutions of discrete time dynamic equilibrium models. *Journal of Economic Dynamics and Control* 2008;32(11):3397–3414.

- Kim J, Kim SH. Spurious welfare reversals in international business cycle models. *Journal of International Economics* 2003;60(2):471–500.
- Kimball MS. Precautionary Saving in the Small and in the Large. *Econometrica* 1990;58(1):53–73.
- King RG, Plosser CI, Rebelo ST. Production, growth and business cycles I. The basic neoclassical model. *Journal of Monetary Economics* 1988a;21(2):195–232.
- King RG, Plosser CI, Rebelo ST. Production, Growth and Business Cycles: II. New Directions. *Journal of Monetary Economics* 1988b;21(2):309–341.
- King RG, Plosser CI, Rebelo ST. Production, Growth and Business Cycles: Technical Appendix. *Computational Economics* 2002;20(1-2):87–116.
- King RG, Rebelo ST. Resuscitating Real Business Cycles. In: *Handbook of Macroeconomics*. Elsevier; 1999. p. 85.
- King RG, Watson MW. The Solution of Singular Linear Difference Systems under Rational Expectations. *International Economic Review* 1998;39(4):1015–1026.
- King RG, Watson MW. System Reduction and Solution Algorithms for Singular Linear Difference Systems under Rational Expectations. *Computational Economics* 2002;20(1-2):57–86.
- Klein P. Using the generalized Schur form to solve a multivariate linear rational expectations model. *Journal of Economic Dynamics and Control* 2000;24(10):1405–1423.
- Kocherlakota NR. Implications of Efficient Risk Sharing without Commitment. *Review of Economic Studies* 1996;63(4):595–609.
- Kogan L, Mitra I. Accuracy Verification for Numerical Solutions of Equilibrium Models; 2014.
- Koop G, Pesaran MH, Potter SM. Impulse Response Analysis in Nonlinear Multivariate Models. *Journal of Econometrics* 1996;74(1):119–147.
- Koopmans LH. The Spectral Analysis of Time Series. San Diego: Academic Press, 1995.
- Kopecky KA, Suen RM. Finite State Markov-Chain Approximations to Highly Persistent Processes. *Review of Economic Dynamics* 2010;13(3):701–714.
- Krueger D, Kubler F. Computing Equilibrium in OLG Models with Stochastic Production. *Journal of Economic Dynamics and Control* 2004;28(7):1411–1436.
- Krusell P, Mukoyama T, Rogerson R, Şahin A. Aggregate implications of indivisible labor, incomplete markets, and labor market frictions. *Journal of Monetary Economics* 2008;55(5):961–979.
- Krusell P, Mukoyama T, Rogerson R, Şahin A. A three state model of worker flows in general equilibrium. *Journal of Economic Theory* 2011;146(3):1107–1133.
- Krusell P, Smith Jr AA. Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy* 1998;106(5):867–896.
- Kydland FE, Prescott EC. Time to build and aggregate fluctuations. *Econometrica: Journal of the Econometric Society* 1982;50(6):1345.

- Ladyzhenskaya . Linear and Quasilinear Elliptic Equations. Academic Press, 1968.
- Lanczos C. Trigonometric Interpolation of Empirical and Analytical Functions. Studies in Applied Mathematics 1938;17(1-4):123–199.
- Leeper EM. Equilibria under ‘active’ and ‘passive’ monetary and fiscal policies. Journal of Monetary Economics 1991;27(1):129–147.
- Leland HE. Saving and Uncertainty: The Precautionary Demand for Saving. The Quarterly Journal of Economics 1968;82(3):465–473.
- Levintal O. Fifth-Order Perturbation Solution to DSGE Models. Journal of Economic Dynamics and Control 2017a;80:1–16.
- Levintal O. Taylor Projection: A New Solution Method to Dynamic General Equilibrium Models. SSRN Electronic Journal 2017b;.
- Lifanov IK, Poltavskii LN, Vainikko MGM. Hypersingular Integral Equations and Their Applications. CRC Press, 2003.
- Ljungqvist L, Sargent TJ. Recursive Macroeconomic Theory. MIT Press, 2004.
- Lubik TA, Schorfheide F. Computing Sunspot Equilibria in Linear Rational Expectations Models. Journal of Economic Dynamics and Control 2003;28(2):273–285.
- Lucas RE. Econometric policy evaluation: A critique. Carnegie-Rochester Conference Series on Public Policy 1976;1:19–46.
- Lucas RE, Prescott EC. Investment Under Uncertainty. Econometrica 1971;39(5):659–681.
- Maliar L, Maliar S. Merging Simulation and Projection approaches to Solve High-dimensional Problems with an Application to a New Keynesian Model. Quantitative Economics 2015;6(1):1–47.
- Maliar L, Maliar S, Villemot S. Taking Perturbation to the Accuracy Frontier: A Hybrid of Local and Global Solutions. Computational Economics 2012;42(3):307–325.
- Maliar S, Maliar L, Judd KL. Solving the multi-country real business cycle model using ergodic set methods. Journal of Economic Dynamics and Control 2011;35(2):207–228.
- Malin BA, Krueger D, Kubler F. Solving the Multi-country Real Business Cycle Model Using a Smolyak-collocation Method. Journal of Economic Dynamics and Control 2011;35(2):229–239.
- Marcet A, Marshall DA. Solving Nonlinear Rational Expectations Models by Parameterized Expectations: Convergence to Stationary Solutions. Federal Reserve Bank of Minneapolis Working Paper 1994;.
- Marcet A, Sargent TJ. Convergence of Least-Squares Learning in Environments with Hidden State Variables and Private Information. Journal of Political Economy 1989;97(6):1306–1322.
- Mas-Colell A, Whinston MD, Green JR. Microeconomic Theory. Oxford University Press, 1995.
- Mason JC, Handscomb DC. Chebyshev Polynomials. CRC, 2003.

- Mazya V. Sobolev Spaces in Mathematics I: Sobolev Type Inequalities. Springer, 2009a.
- Mazya V. Sobolev Spaces in Mathematics II: Applications in Analysis and Partial Differential Equations, 2009b.
- McCallum BT. On non-uniqueness in rational expectations models. *Journal of Monetary Economics* 1983;11(2):139–168.
- McGrattan ER. Solving the Stochastic Growth Model with a Finite Element Method. *Journal of Economic Dynamics and Control* 1996;20(1-3):19–42.
- McLean W. Strongly Elliptic Systems and Boundary Integral Equations. Cambridge University Press, 2000.
- Medina JP, Soto C. The Chilean business cycles through the lens of a stochastic general equilibrium model. *Central Bank of Chile Working Papers* 2007;457.
- Mikhailov SE. About traces, extensions and co-normal derivative operators on Lipschitz domains. In: *Integral Methods in Science and Engineering: Theoretical and Practical Aspects*. Birkhäuser; 2006. p. 149–160.
- Mikhailov SE. Traces, extensions, co-normal derivatives and solution regularity of elliptic systems with smooth and non-smooth coefficients. Preprint in Brunel University Research Archive BURA 2009;;1–28.
- Milani F. The Modeling of Expectations in Empirical DSGE Models: A Survey. *Advances in Econometrics* 2012;28:3–38.
- Minsky HP. Can "It" Happen Again?: Essays on Instability and Finance 2015;.
- Moll VH. Special Integrals of Gradshteyn and Ryzhik: the Proofs - Volume I. Taylor & Francis Group, 2015.
- Moll VH. Special Integrals of Gradshteyn and Ryzhik: the Proofs - Volume II, 2016.
- Mulligan CB. Aggregate implications of indivisible labor. *Advances in Macroeconomics* 2001;.
- Murchison S, Rennison A. ToTEM: the Bank of Canada's New Quarterly Projection Model . Technical Report 97; 2006.
- Muscat J. Functional Analysis: An Introduction to Metric Spaces, Hilbert Spaces, and Banach Algebras. Cham: Springer, 2014.
- Neri U. Singular Integrals. volume 200 of *Lecture Notes in Mathematics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1971.
- Nobile F, Tempone R, Webster CG. A Sparse Grid Stochastic Collocation Method for Partial Differential Equations with Random Input Data. *SIAM Journal on Numerical Analysis* 2008;46(5):2309–2345.
- Obstfeld M, Rogoff K. Exchange Rate Dynamics Redux. *Journal of Political Economy* 1995;103(3):624–660.

- Of G, Steinbach O, Urthaler P. Fast Evaluation of Volume Potentials in Boundary Element Methods. *SIAM Journal on Scientific Computing* 2010;32(2):585–602.
- Palais RS. A simple proof of the Banach contraction principle. *Journal of Fixed Point Theory and Applications* 2007;2(2):221–223.
- Peralta-Alva A, Santos MS. Analysis of Numerical Errors. In: Schmedders K, Judd KL, editors. *Handbook of Computational Economics*. Elsevier; 2014. p. 517–556.
- Phaneuf L, Sims ER, Victor JG. Inflation, Output, and Markup Dynamics with Forward-Looking Wage and Price Setters. *NBER Working Paper* 2015;.
- Prescott EC. Theory ahead of business-cycle measurement. *Carnegie-Rochester Conference Series on ...* 1986;25 IS -:11–44.
- Preston B, Roca M. Incomplete Markets, Heterogeneity and Macroeconomic Dynamics. *NBER Working Papers* 2007;.
- Privault N. *Understanding Markov Chains: Examples and Applications*. Springer, 2013.
- Rabanal P, Rubio-Ramírez JF. Comparing New Keynesian Models of the Business Cycle: A Bayesian Approach. *Journal of Monetary Economics* 2005;52(6):1151–1166.
- Renardy M, Rogers RC. *An Introduction to Partial Differential Equations*. 2nd ed. Springer, 2004.
- Rogerson R. Indivisible labor, lotteries and equilibrium. *Journal of Monetary Economics* 1988;21(1):3–16.
- Rogerson R, Wallenius J. Micro and macro elasticities in a life cycle model with taxes. *Monetary Policy and Capital Accumulation* *Monetary Policy and Capital* 2009a;144(6):2277–2292.
- Rogerson R, Wallenius J. Micro and macro elasticities in a life cycle model with taxes. *Journal of Economic Theory* 2009b;144(6):2277–2292.
- Rouwenhorst KG. Asset Pricing Implications of Equilibrium Business Cycle Models. In: *Frontiers of Business Cycle Research*. Princeton University Press; 1995. p. 294–330.
- Rubio-Ramírez JF, Fernández-Villaverde J. How Structural Are Structural Parameters? In: *NBER Macroeconomics Annual 2007, Volume 22*. 2008. p. 83–137.
- Ruge-Murcia FJ. Methods to Estimate Dynamic Stochastic General Equilibrium Models. *Journal of Economic Dynamics and Control* 2007;31(8):2599–2636.
- Samuelson PA. The Fundamental Approximation Theorem of Portfolio Analysis in terms of Means, Variances and Higher Moments. *Review of Economic Studies* 1970;37(4):537–542.
- Sandmo A. The Effect of Uncertainty on Saving Decisions. *The Review of Economic Studies* 1970;37(3):353.
- Santos MS. Differentiability and Comparative Analysis in Discrete-time Infinite-horizon Optimization. *Journal of Economic Theory* 1992;57(1):222–229.

- Santos MS. Accuracy of Numerical Solutions Using the Euler Equation Residuals. *Econometrica* 2000;68(6):1377–1402.
- Santos MS, Peralta-Alva A. Accuracy of Simulations for Stochastic Dynamic Models. *Econometrica* 2005;73(6):1939–1976.
- Santos MS, Rust J. Convergence Properties of Policy Iteration. *SIAM Journal on Control and Optimization* 2006;42(6):2094–2115.
- Santos MS, Vigo-Aguiar J. Analysis of a Numerical Dynamic Programming Algorithm Applied to Economic Models. *Econometrica* 1998;66(2):409.
- Schmitt-Grohé S, Uribe M. Solving Dynamic General Equilibrium Models Using a Second-Order Approximation to the Policy Function. *Journal of Economic Dynamics and Control* 2004;28(4):755–775.
- Schorfheide F. VAR Forecasting Under Misspecification. *Journal of Econometrics* 2005;128(1):99–136.
- Schorfheide F. Estimation and Evaluation of DSGE Models: Progress and Challenges. Federal Reserve Bank of Philadelphia Working Paper 2011;
- Seber GAF, Lee AJ. Linear Regression Analysis. John Wiley & Sons, 2003.
- Shen J, Tang T, Wang LL. Spectral Methods: Algorithms, Analysis and Applications. Springer Science & Business Media, 2011.
- Shumway RH, Stoffer DS. Time Series Analysis and Its Applications. Springer Texts in Statistics. Cham: Springer International Publishing, 2017.
- Sikorski K. Optimal solution of nonlinear equations. *Journal of Complexity* 1985;1(2):197–209.
- Simmonds JG, Mann JE. A First Look at Perturbation Theory. Dover Publications, 2013.
- Simon HA. Dynamic Programming Under Uncertainty with a Quadratic Criterion Function. *Econometrica* 1956;24(1):74.
- Sims CA. Interpreting the macroeconomic time series facts: The effects of monetary policy. *European Economic Review* 1992;.
- Sims CA. Solving Linear Rational Expectations Models. *Computational Economics* 2002;20(1-2):1–20.
- Sims CA, Goldfeld SM, Sachs JD. Policy Analysis with Econometric Models. *Brookings Papers on Economic Activity* 1982;1982(1):107.
- Smets F, Wouters R. An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area. *Journal of the European Economic Association* 2003;1(5):1123–1175.
- Smolyak SA. Quadrature and Interpolation Formulas for Tensor Products of Certain Classes of Functions. *Dokl Akad Nauk SSSR* 1963;148(5):1042–1045.

- Söderlind P. Solution and estimation of RE macromodels with optimal policy. *European Economic Review* 1999;43(4-6):813–823.
- Šolín P, Segeth K, Dolezel I. Higher-Order Finite Element Methods. CRC Press, 2003.
- Stachurski J. Economic Dynamics 2008;:1–392.
- Stachurski J, Martin V. Computing the Distributions of Economic Models via Simulation. *Econometrica* 2008;76(2):443–450.
- Steinbach O, Wendland WL. On C. Neumann's Method for Second-Order Elliptic Systems in Domains with Non-smooth Boundaries. *Journal of Mathematical Analysis and Applications* 2001;262(2):733–748.
- Stiglitz J, Weiss A. Credit rationing and collateral. In: Edwards J, Franks J, Mayer C, Schaefer S, editors. Recent developments in corporate finance. Cambridge: Cambridge University Press; 2009. p. 101–135.
- Stiglitz JE. Theories of Wage Rigidity. NBER Working Papers 1984;.
- Stiglitz JE. The Causes and Consequences of The Dependence of Quality on Price. *Journal of Economic Literature* 1987;25(1):1–48.
- Stiglitz JE, Weiss A. Credit Rationing in Markets with Imperfect Information. *The American Economic Review* 1981;71(3):393–410.
- Stiglitz JE, Weiss A. Incentive Effects of Terminations: Applications to the Credit and Labor Markets. *The American Economic Review* 1983;73(5):912–927.
- Stock JH, Watson MW. Vector Autoregressions. *Journal of Economic Perspectives* 2001;15(4):101–115.
- Stokey N, Lucas RE, Prescott EC. Recursive Methods in Economic Dynamics, 1989.
- Sveen T, Weinke L. New perspectives on capital, sticky prices, and the Taylor principle. *Monetary Policy and Capital Accumulation* 2005;123(1):21–39.
- Svensson LEO. Chapter 22 - Inflation Targeting. In: Woodford M, editor. *Handbook of Monetary Economics*. Elsevier; 2010. p. 1237–1302 T2 –.
- Swanson ET, Anderson GS, Levin AT. Higher-Order Perturbation Solutions to Dynamic, Discrete-Time Rational Expectations Models. *SSRN Electronic Journal* 2006;:1–31.
- Tartar L. An Introduction to Sobolev Spaces and Interpolation Spaces. Springer, 2007.
- Tauchen G. Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions. *Economics Letters* 1986;20(2):177–181.
- Tauchen G, Hussey R. Quadrature-Based Methods for Obtaining Approximate Solutions to Non-linear Asset Pricing Models. *Econometrica* 1991;59(2):371.
- Theil H. A Note on Certainty Equivalence in Dynamic Planning. *Econometrica* 1957;25(2):346.

- Thompson JF, Warsi ZU, Mastin CW. Numerical Grid Generation: Foundations and Applications. North-Holland, 1985.
- Tovar Mora CE. DSGE Models and Central Banks. Economics: The Open-Access, Open-Assessment E-Journal 2009;3(2009-16):1–31.
- Uhlig H. A Toolkit for Analyzing Nonlinear Dynamic Stochastic Models Easily. In: Computational Methods for the Study of Dynamic Economies. OUP Oxford; 1999. p. 30–61.
- Uhlig H. What are the effects of monetary policy on output? Results from an agnostic identification procedure. Journal of Monetary Economics 2005;52(2):381–419.
- Vainikko G. Multidimensional Weakly Singular Integral Equations. volume 1549 of *Lecture Notes in Mathematics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1993.
- Verchota G. Layer potentials and regularity for the Dirichlet problem for Laplace's equation in Lipschitz domains. Journal of Functional Analysis 1984;59(3):572–611.
- Vladimirov VS. Equations of Mathematical Physics. New York: Marcel Dekker, 1971.
- White H. Maximum Likelihood Estimation of Misspecified Models. Econometrica 1982;50(1):1.
- White H. Estimation, Inference and Specification Analysis. Cambridge University Press, 1994.
- Wood A. A Theory of Profits 1975;.
- Woodford M. Interest and Prices. Foundations of a Theory of Monetary Policy. Princeton University Press, 2011.
- Wooldridge JM. Econometric Analysis of Cross Section and Panel Data. MIT Press, 2002.
- Wu HH, Wu S. Various Proofs of the Cauchy-Schwarz Inequality. OCTOGON MATHEMATICAL MAGAZINE 2011;17(1):221–229.
- Yosida K. Functional Analysis. Springer, 1978.
- Yun T. Nominal price rigidity, money supply endogeneity, and business cycles. Journal of Monetary Economics 1996;37(2):345–370.



# 插图

2.1 数值模拟: $\{\pi^\#, \nu^p\}$ vs $\pi$ 。模拟过程中设参数值 $\phi = 0.25$ , $\epsilon = 10$ , $\beta = 0.99$ 。 . . . . .	32
6.1 CIA DSGE 模型结构图 . . . . .	101
7.1 DSGE 模型的工作流程 . . . . .	116
8.1 不同 $m$ 取值下的包络定理 . . . . .	129
12.1 单项式方程 . . . . .	203
12.2 切比雪夫多项式 $T_n(x)$ . . . . .	205
12.3 切比雪夫多项式的根 . . . . .	206
12.4 切比雪夫多项式的根 . . . . .	206
12.1 张量模拟 . . . . .	210
12.1 kinks . . . . .	211
12.2 三角方程形式 . . . . .	212
12.1 随机 NCGT 模型的投影法近似求解 . . . . .	221
12.1 稀疏网格 . . . . .	224
12.2 稀疏网格 G(5,3) . . . . .	225
12.C.1 正交多项式满足的关系 . . . . .	245
12.C.2 正态分布 . . . . .	246
12.C.3 埃米特多项式 . . . . .	247
12.C.4 伽玛分布 . . . . .	249
12.C.5 贝塔分布 . . . . .	254
12.C.6 切比雪夫插值近似 . . . . .	265
12.F.1 矩形法则 ( $N = 4$ ) . . . . .	280
12.F.2 梯形法则 ( $N = 4$ ) . . . . .	280
12.F.3 3 阶牛顿——寇特斯法则, 对应于区间中的 4 个配点 . . . . .	282
12.F.4 龙格现象 ( $N = 6$ ) . . . . .	283
12.F.5 比较矩形法则和梯形法则下, 近似误差随 $N$ 值增大的收敛情况 . . . . .	284
12.F.6 适应性求积 . . . . .	287
12.F.7 非周期方程的 $T$ -周期扩展 $t \in [0, 1]$ . . . . .	290
12.F.8 $f(t)$ 和 $g(\theta)$ . . . . .	292
12.F.9 $f(x), x \in [-1, 1]$ . . . . .	297
12.F.10 求积法则误差的衰减 . . . . .	299
12.G.1 傅里叶变换 . . . . .	300

12.G.2 洛伦兹方程 ( $\alpha = 2$ ) . . . . .	303
12.G.3 $f(t) = \cos^2(3t)$ , $t \in (-2, 2)$ . . . . .	308
12.G.4 锯齿波 . . . . .	309
12.G.5 吉布斯现象 . . . . .	310
14.1 有限元 $\tau_\ell$ 及其结点 $x_k$ . . . . .	453
14.2 容许分解, 以及不容许分解的三角分解 ( $d = 2$ ) . . . . .	454
14.3 有限元 $\tau_\ell$ 和参考元 $\tau(d = 2)$ . . . . .	456
14.4 检测空间 $\mathcal{S}_h^1(\mathcal{T}_N)$ 的基 $\psi_k^1(x)$ . . . . .	465
15.1 欧拉等式误差检测 . . . . .	472
15.2 欧拉等式误差检测的比较 . . . . .	473
16.1 预测误差的方差分解 . . . . .	487
16.2 预测误差的方差分解 . . . . .	489
16.3 产出的冲击响应方程 . . . . .	490
16.1 抽样观测数据的跨相关 . . . . .	496
16.2 抽样数据的光谱分布密度 . . . . .	498
16.3 冲击响应方程 . . . . .	500
16.1 中国的劳动收入份额和消费产出比 . . . . .	501
16.A.1 正弦函数的导数 . . . . .	504
16.A.2 麦克劳林级数的示例: 对 $\sin(x)$ (蓝色曲线) 沿着 $x_0 = 0$ 作 $n$ 阶泰勒级数展开(红色曲线). . . . .	505
16.A.3 正弦函数的导数 . . . . .	507
16.A.4 双曲正弦/余弦方程, $x \in [-3\pi, 3\pi]$ . . . . .	508
16.B.1 高斯核方程的振幅随带宽的增加而显著降低 . . . . .	510
16.B.2 误差方程(累积高斯方程)随 $\sigma$ 值的变化 . . . . .	511
16.B.3 黑维塞方程(单位阶跃方程) . . . . .	512
16.B.4 二项式系数 . . . . .	512
16.B.5 黑维塞方程相加减 . . . . .	513
16.B.6 两个块方程的卷积是一个三角方程 . . . . .	513
16.B.7 两个块方程卷积的卷积接近一个高斯分布 . . . . .	514
16.B.8 2 维空间中高斯方程的各项同性 . . . . .	515
16.B.9 2 维空间中高斯方程的各项异性 . . . . .	515
16.B.10 傅里叶变换、逆傅里叶变换的 Mathematica 程序 . . . . .	516
16.B.11 傅里叶变换、逆傅里叶变换的 Mathematica 程序 . . . . .	516
17.1 正态分布的概率密度方程 . . . . .	544

# 表格

8.1 偏好冲击的转移矩阵 . . . . .	127
12.1 网格结点的数量比较 . . . . .	225
12.C.1 经典正交多项式 . . . . .	244
12.G.1 傅里叶分析的常见概念界定 . . . . .	299
12.G.2 原方程域和傅里叶变换方程域 . . . . .	301
12.G.3 雅各比方程的泊松求和 . . . . .	314
16.1 典型 DSGE 模型的参数校准 . . . . .	484
16.1 用于经验验证的观测数据集 . . . . .	494
17.1 anova 表 . . . . .	550
17.1 变换观测数据的转换方程 . . . . .	555

表格

表格

# Todo list

curse of dimensionality, 可由映射法得到一定程度上的缓解, 做一个 reference.	118
加入一个 reference.	120
Karl Whelan 对 forward-looking 和 backward-looking 的描述	120
to be finished...	149
补一个 reference	162
补一个 reference.	166
补一个 ref	167
补一个 ref, 见下文的范例。	170
reference	179
reference	182
下文有注释, 对应距离方程其实是三角方程的一种。到时把讲义内容补充到这里来。	200
reference	200
reference 和 reference	201
reference	204
reference	210
见下文, 补 reference。	211
reference	213
做一个 reference	218
做一个 ref	221
笔记里有各向异性的介绍, 敲到键盘里之后, 在这处做一个 ref.	228
这部分还需要做进一步的说明。	250
弱条件补一个 reference	272
补一个 reference	273
解的存在性和唯一性	273
解的存在性和唯一性	274
补一个描述: Hilbert space	274
reference	289
补一个 measurable function 的词条	328
section 6.1.1	413
更多自然密度的介绍见后	417
作 reference	441
笔记打出来之后, 是在 sec 10, 做一个 ref	483
下文给出估计方法	484
笔记敲到键盘里之后, 做一个 reference: 如何通过对矩条件的仿真近似来估计参数向量 $\theta$	492
敲进键盘之后, 做一个 ref	500

---

■ central moment formula . . . . .	585
■ 考慮一下，要不要做个 section? . . . . .	586

# 索引

- adaptive quadrature ..... 适应性求积, 286  
adjoint operator ..... 伴随算子, 364  
Akaike information criterion(AIC) ..... 赤池信息量, 496  
analysis of variance table (anova) ..... anova 表, 550  
angular frequency ..... 角频率, 298, 300, 301  
anisotropy ..... 各向异性, 228, 514  
axiomatic set theory ..... 公理集合论, 324
- Babuška-Brezzi-Ladyshenskaya condition ..... BBL 条件, 447  
Banach space ..... 巴拿赫空间, 331, 363  
bandpass filter (BPF) ..... 带宽滤波, 517  
bandwidth (matrix) ..... 带宽 (矩阵的), 523  
base function ..... 基方程, 461  
basis  
    Chebyshev polynomial ..... 切比雪夫多项式 (基), 203  
    global ..... 全局基, 202  
    Jacobi polynomial ..... 雅各比多项式 (基), 203  
    local ..... 局部基, 202  
    monomial ..... 单项式 (基), 202  
    trigonometric series ..... 三角序列 (基), 203  
    unidimensional ..... 单维基, 202  
basis function ..... 基方程, 199  
Bayesian information criterion(BIC) ..... 贝叶斯信息量, 496  
Bellman equation ..... 贝尔曼等式, 123, 218  
Bernoulli trials ..... 伯努利实验, 561  
Bessel function ..... 贝塞尔方程, 390  
best linear unbiased estimator (BLUE) ..... 最佳线性无偏估计, 546  
Beta distribution ..... 贝塔分布, 253  
Beta integral ..... 贝塔积分, 233  
BFMG algorithm ..... BFMG 算法, 569  
bijection ..... 双射, 276  
binomial coefficients ..... 二项式系数, 512  
binomial theorem ..... 二项式定理, 232  
Bonferroni inequality ..... 邦费罗尼不等式, 554  
boundary integral equations ..... 边界积分等式, 387

- boundary value problem ..... 边界值问题, 319  
 bounded differences ..... 差分有界, 539  
 Box-Cox transformation ..... Box-Cox 变换, 555  
 Bramble-Hilbert Lemma ..... Bramble-Hilbert 引理, 347
- Calderón projection ..... 卡尔德隆投影, 412, 430  
 capacity ..... 容度, 417  
 cardinality ..... 势, 224, 325  
 cartesian product ..... 笛卡尔乘积, 223  
 Cauchy principal value ..... 柯西主值, 406  
 Cauchy sequence ..... 柯西序列, 276  
 Cauchy-Schwarz inequality ..... 柯西—施瓦茨不等式, 334  
 central limit theorem (CLT) ..... 中心极限定理, 514  
 chain rule ..... 链式法则, 354  
 Chebyshev bound ..... 切比雪夫边界, 534  
 Chebyshev distance ..... 切比雪夫距离, 327  
 Chernoff method ..... 切尔诺夫法, 534  
 Christoffel numbers ..... 克里斯托费尔数, 243  
 Christoffel-Darboux relation ..... 克里斯托费尔—达布关系, 240  
 Chu-Vandermonde summation ..... 朱世杰—范德蒙德求和公式, 234, 255  
 Clenshaw-Curtis Rule ..... 克伦肖—柯蒂斯求积法则, 223, 286, 291  
 closed curve ..... 闭曲线, 407  
 closed range theorem ..... 闭值域定理, 364  
 coefficient of determination ..... 决定系数, 551  
 coercive form ..... 强制形式, 274  
 collocation  
     ..... 配点, 204, 215  
     orthogonal ..... 正交配点, 204, 215  
 compact operator ..... 紧凑算子, 373, 450  
 complement ..... 补集, 324  
 completeness  
     Euclidean space ..... 完备性 (欧几里得空间), 332  
     complex angle ..... 复值夹角, 508  
     complex conjugate ..... 复共轭, 488, 508  
     complex modulus ..... 复数绝对值, 333  
     complex number  
         imaginary part ..... 虚部 (复数), 276  
         real part ..... 实部 (复数), 276  
     complex plane ..... 复数平面, 333  
     conditional bias ..... 条件偏误, 558  
     conditional density ..... 条件密度, 267  
     confidence interval ..... 置信区间, 548  
     conforming trial spaces ..... 协调检测空间, 439  
     conjugate ..... 共轭复数, 333

- continuous inverse ..... 连续逆, 276  
contraction mapping ..... 收缩映射, 362  
convergence  
    absolute ..... 绝对收敛, 332  
convolution ..... 卷积, 313  
Cook's distance ..... 库克距离, 554  
countable additivity ..... 可数可加性, 325  
cover(topology) ..... 覆盖(拓扑学), 324  
Cramér-Rao lower Bound ..... 克拉美罗下界, 559, 580, 582  
curse of dimensionality ..... 维数灾难, 207, 210  
Céa lemma, ..... 齐亚引理, 277, 440
- data generating process (DGP) ..... 数据生成过程, 575  
De Moivre Theorem ..... 棣莫弗定理, 507  
decomposition  
    ..... 分解(有限元), 453  
    admissible ..... 容许分解, 454  
Delta Dirac function ..... 狄拉克方程, 215, 304, 511  
design matrix ..... 设计矩阵, 544  
differentiability classification ..... 方程微分的类, 335  
Dirac distribution ..... 狄拉克分布, 349  
Dirichlet boundary value condition ..... 狄利克雷边界值条件, 321, 375  
Dirichlet condition  
    homogeneous 齐次狄利克雷条件, 275  
Dirichlet to Neumann map ..... 狄利克雷到诺依曼的映射, 423  
distance function ..... 距离方程, 200  
distribution  
    regular ..... 正则分布, 349  
    singular ..... 奇异分布, 349  
    stationary ..... 平稳分布, 208  
distribution ..... 分布, 349  
divergence theorem ..... 高斯散度定理, 319  
dual space ..... 对偶空间, 333, 339, 355  
duality pairing ..... 对偶配对, 355, 394  
dynamic programming ..... 动态规划, 121
- edge ..... 边, 453  
efficient estimator ..... 有效估计, 546  
ellipticity  
    semi ..... 半椭圆性, 419, 422  
embedding theorem of Sobolev space ..... 索伯列夫空间的嵌入定理, 340  
embeddings ..... 嵌入, 340  
ensemble mean square error (EMSE) ..... 整体均方误差, 558  
entropy ..... 熵, 573

- Envelope Theorem ..... 包络定理, 121  
 equivalence of norms ..... 范数等价, 327, 343, 355  
 ergodic set ..... 遍历集, 222  
 error function ..... 误差方程, 511  
 Euclidean norm ..... 欧几里得范数, 327  
 Euler formula (complex analysis) ..... 欧拉公式, 503  
 Euler identity ..... 欧拉恒等式, 506  
 Euler integral representation ..... 欧拉积分表达式, 233  
 Euler transformation formula ..... 欧拉转换公式, 234  
 exponentiation (complex analysis) ..... 取幂 (复分析), 507  
 extended real number line ..... 扩展实数线, 325  
 exterior algebra ..... 外代数, 319  
 exterior normal vector ..... 外部法向量, 406  
  
 factorization  
     Cholesky ..... Cholesky 因式分解, 499  
     family of sets ..... 集族, 324  
     finite element  
         neighboring ..... 相邻有限元, 455  
         regular ..... 正则型有限元, 454, 460  
     finite elements method ..... 有限元法, 211  
     Fischer information matrix ..... 费雪信息矩阵, 559  
     flops  
         (floating points operations) ..... 浮点计算数, 517  
         form function ..... 形式方程, 461  
         Fourier cosine series ..... 傅里叶余弦级数, 308  
         Fourier series ..... 傅里叶, 306  
         Fourier series ..... 傅里叶级数, 299, 337  
         Fourier sine series ..... 傅里叶正弦级数, 308  
         Fourier synthesis ..... 傅里叶综合, 310  
         Fourier transform ..... 傅里叶变换, 299, 350  
             Discrete ..... 离散傅里叶变换, 299  
             inverse ..... 逆傅里叶变换, 350  
             rotational symmetries ..... 傅里叶变换的旋转对称, 350  
             Semidiscrete ..... 半离散傅里叶变换, 299  
     Fredholm alternative theorem ..... 弗雷德霍姆二择一定理, 373  
     Fredholm boundary integral function  
         first kind ..... 第一类弗雷德霍姆边界积分方程, 430  
         second kind ..... 第二类弗雷德霍姆边界积分方程, 431  
     frequency band ..... 频率带, 497  
     full width at half maximum(FWHM) ..... 半峰全宽, 303  
     fundamental solution ..... 基本解, 387  
  
     Galerkin method ..... 伽辽金近似法, 215, 272

- Galerkin Orthogonality ..... 伽辽金正交, 439  
 Galerkin-Bubnov method ..... 伽辽金——布博诺夫法, 439  
 Galerkin-Petrov method ..... 伽辽金——佩特洛夫法, 444  
 gamma distribution ..... 伽玛分布, 249  
 Gauss quadrature formula ..... 高斯求积公式, 243  
 Gauss summation formula ..... 高斯求和公式, 233  
 Gauss theorem ..... 高斯散度定理, 319  
 Gauss-Hermite quadrature ..... 高斯——埃米特求积, 270  
 Gaussian distribution ..... 高斯分布, 543  
 Gaussian elimination ..... 高斯消元法, 526  
 Gaussian elimination ..... 高斯消元, 522  
 Gaussian function ..... 高斯方程, 303  
 generalized method of moments (GMM) ..... 广义矩估计法, 500  
 generalized partial derivative ..... 广义偏导数, 336  
 generating function  
     Hermite polynomial ..... 母方程 (埃米特多项式), 248  
     Jacobi polynomial ..... 母方程 (雅各比多项式), 257  
     Laguerre polynomial ..... 母方程 (拉盖尔多项式), 251  
     Legendre polynomial ..... 母方程 (勒让德多项式), 259  
     the first kind Chebyshev polynomial ..... 母方程 (第一类切比雪夫多项式), 262  
     the second kind Chebyshev polynomial ..... 母方程 (第二类切比雪夫多项式), 262  
 Gibbs phenomenon ..... 吉布斯现象, 309  
 Gram-Schmidt orthogonality process ..... 格拉姆-施密特正交化过程, 237, 318  
 Green identities  
     first 格林第一恒等式, 320, 374  
     second 格林第二恒等式, 320, 386  
 grid ..... 网格, 211  
 grid distance ..... 网格距离, 327  
 Gårding inequality ..... Gårding 不等式, 373  
 h-refinement finite elements ..... h——改进 (有限元分析), 213  
 Hankel moment matrix ..... 汉克尔矩阵, 238  
 hat matrix ..... 帽子矩阵, 551  
 Heaviside function ..... 黑维塞方程, 511  
 Hermitian symmetric ..... 埃米特对称, 333  
 Hessian matrix ..... 海森矩阵, 577  
 heteroskedasticity ..... , 586  
 Hilbert space ..... 希尔伯特空间, 274, 331  
 Hoeffding inequality ..... 霍夫丁不等式, 538  
 Hoeffding lemma ..... 霍夫丁引理, 536  
 hp-adaptive finite elements ..... hp——有限元法, 213  
 hyperbolic function  
     cosine ..... 双曲线余弦方程, 508

- sine ..... 双曲线正弦方程, 508  
 hypergeometric  
     function ..... 超几何方程, 231  
     function, generalized ..... 广义超几何方程, 232  
     series ..... 超几何序列, 231  
     series, monic ..... 首一超几何序列, 231  
 hypergeometric function  
     Jacobi polynomial ..... 超几何方程 (雅各比多项式), 254  
     Laguerre polynomial ..... 超几何方程 (拉盖尔多项式), 250  
 Hölder condition ..... 霍德尓条件, 338  
 Hölder continuous ..... 霍德尓连续, 324  
 Hölder inequality  
     generalization ..... 广义霍德尓不等式, 339  
 Hölder inequality ..... 霍德尓不等式, 339  
 Hölder space ..... 霍德尓空间, 323  
 Hölder's inequality ..... 霍德尓不等式, 330  
  
 iamge ..... 像, 364  
 impulse-response functions (IRFs) ..... 冲击响应方程, 489  
 Inada condition ..... 稻田条件, 122  
 indefinitely differentiable function ..... 无限可导方程, 504  
 index family ..... 索引族, 323  
 indicator function ..... 指示方程, 488, 497  
 inf-sup condition ..... 上极限—下极限条件, 275  
 infor mation  
     ..... 信息, 573  
 information matrix ..... 信息矩阵, 578  
 information matrix equality (IME) ..... 信息矩阵等式, 579  
 injection ..... 内射, 373  
 injection ..... 单射, 276  
 inner product  
     duality ..... 双内积, 274  
 inner product ..... 内积, 237, 332  
 inner product space ..... 内积空间, 332  
 integrable singular points ..... 可积奇异点, 285  
 integration by parts ..... 分部积分, 320, 336  
 integration by parts ..... 分部积分公式, 335  
 interior boundary trace ..... 内界迹, 357  
 interpolation  
     Chebyshev ..... 切比雪夫插值, 204, 263  
     Lagrange ..... 拉格朗日插值法, 243  
 interpolation theorem ..... 插值定理, 353  
 intersection ..... 交集, 325  
 isotropy, 514

- Jack-knifed residual ..... Jack-knifed 残差, 552  
 Jacobian function ..... 雅各比方程, 313  
 Jensen inequality ..... 延森不等式, 537  
 kernel ..... 核, 363, 497  
 kernel function  
   Gaussian ..... 高斯核方程, 509  
   Gaussian, cumulative ..... 累积高斯核方程, 511  
 Korn inequality ..... 科恩不等式, 275  
 Kronecker product ..... 克罗内克乘积, 237  
 Kullback-Leibler information criterion (KLIC) ..... Kullback-Leibler 信息准则, 574  
 L-infinity norm, 327  
 Lagrange interpolation ..... 拉格朗日插值, 264  
 Lamé elasticity parameters ..... 拉梅弹性参数, 275  
 Laplace function ..... 拉普拉斯方程, 391  
 Laplace operator ..... 拉普拉斯算子, 387  
 latent variable ..... 隐含变量, 491  
 law of iterated expectations (LIE) ..... 迭代期望法则, 583  
 law of large numbers (LLN)  
   ..... 大数法则, 532  
   weak(WLLN) ..... 弱大数法则, 534  
 Lax-Milgram theorem ..... 拉克斯—密格拉蒙定理, 275, 362, 371  
 Lebesgue integrable function ..... 勒贝格可积方程, 324  
 Lebesgue space  
   weighted ..... 加权勒贝格空间, 331  
 Leibniz rule ..... 莱布尼兹法则, 245  
 Lesbegue space ..... 勒贝格空间, 324  
 Levenberg-Marquardt algorithm ..... 莱文贝格—马夸特方法, 214  
 leverage ..... 杠杆, 554  
 likelihood function (LHF) ..... 似然方程, 545, 557  
 likelihood ratio criterion (LHRC) ..... 似然率量, 549  
 linear frequency ..... 线频率, 301  
 linearization ..... 线性化, 162  
 log ..... 对数线性化, 162  
 lognormal ..... 对数正态线性化, 183  
 Lipschitz continuous function ..... 利普希茨连续方程, 323  
 Lipschitz domain ..... 利普希茨域, 323  
 Lipschitz hypograph ..... 利普希茨亚图, 323  
 Lorentzian function ..... 洛伦兹方程, 302  
 LP-norm ..... LP 范数, 327  
 Lyapunov function ..... 李雅普诺夫方程, 485  
 Manhattan distance ..... 曼哈顿距离, 327

- Markov chain  
 finite ..... 有限马尔科夫链, 208
- Martingale Difference Sequence (MDS) ..... 鞍差序列, 481
- mass matrix  
 local ..... 局部质量矩阵, 462
- matrix  
 band ..... 带状矩阵, 523  
 diagonal ..... 对角矩阵, 518  
 identity ..... 单位矩阵, 241, 520  
 infinite Jacobi ..... 无限雅各比矩阵, 240  
 leading principle submatrix ..... 领先主子矩阵, 240  
 orthogonal ..... 正交矩阵, 499, 520  
 permutation ..... 置换矩阵, 520  
 rank-one update ..... 秩一修正矩阵, 529  
 triangular, lower ..... 下三角矩阵, 518  
 triangular, unit lower ..... 下三角单位矩阵, 518  
 triangular, unit upper ..... 上三角单位矩阵, 519  
 triangular, upper ..... 上三角矩阵, 519  
 tridiagonal ..... 三角对角矩阵, 240
- matrix factorization  
 $LDL^\top$  .....  $LDL^\top$  分解, 525  
 Cholesky ..... Cholesky 分解, 523  
 LU ..... LU 分解, 522  
 QR ..... QR 分解, 531
- matrix inversion lemma ..... 矩阵求逆引理, 528
- maximum likelihood estimator (MLE)  
 quasi ..... 准最大似然估计, 576
- maximum norm ..... 最大范数, 327
- mean square error (MSE) ..... 均方误差, 558
- measurable sets ..... 可测集合, 325
- measurable space ..... 可测空间, 325
- measure  
 counting ..... 计数测度, 326  
 measure ..... 测度, 325
- measurement equation ..... 测量方程, 482
- mesh ..... 网格, 453
- mesh size  
 global ..... 全局网格尺寸, 455  
 local ..... 局部网格尺寸, 454
- metric function ..... 度量方程, 200
- Minkowski inequality  
 integral ..... 闵可夫斯基积分不等式, 330  
 Minkowski inequality ..... 闵可夫斯基不等式, 329
- model matrix ..... 模型矩阵, 544

- moment ..... 矩, 237  
 moment generating function (MGF) ..... 矩生成方程, 534  
 Monte Carlo method ..... 蒙特卡洛法, 492  
  
 natural density ..... 自然密度, 417  
 nestedness of Chebyshev nodes ..... 嵌套性 (切比雪夫结点), 224  
 Neumann boundary value condition ..... 诺依曼边界值条件, 321, 381  
 Neumann series ..... 诺依曼级数, 431, 432  
 New Keynesian Philips Curve (NKPC) ..... 新凯恩斯主义菲利普斯曲线, 30, 479  
 Newton algorithm ..... 牛顿算法, 214  
 Newton Conjugate Gradient Method (Newton-CG method) ..... 牛顿共轭梯度法, 569  
 Newton-Raphson Algorithm ..... 牛顿拉夫森算法, 570  
 Newton—Cotes Rule ..... 牛顿—寇特斯法则, 279  
 node ..... 结点, 211, 453  
 nonlinear regression (NLS) ..... 非线性最小二乘法, 577  
 norm ..... 范数, 237, 275, 323, 326  
 norm equivalence theorem of Sobolev space, 342  
 normal distribution  
     ..... 正态分布, 543  
     multivariate ..... 多元正态分布, 543  
 normal distribution ..... 正态分布, 245  
 normal equation ..... 正规方程, 545  
 normed vector space ..... 赋范向量空间, 327  
 null model ..... 零模型, 544  
 null space ..... 零空间, 363, 530  
 numerical integration ..... 数值积分, 279  
 numerical quadrature ..... 数值求积, 279  
  
 orthogonal complement ..... 正交补, 364  
 orthogonality condition  
     Hermite polynomial ..... 正交条件 (埃米特多项式), 246  
     Jacobi polynomial ..... 正交条件 (雅各比多项式), 256  
     Laguerre polynomial ..... 正交条件 (拉盖尔多项式), 250  
     Legendre polynomial ..... 正交关系 (勒让德多项式), 258  
     the first kind Chebishev polynomial ..... 正交条件 (第一类切比雪夫多项式), 261  
     the second kind Chebishev polynomial ..... 正交条件 (第二类切比雪夫多项式), 261  
 Ostrogradsky-Gauss theorem ..... 奥斯特罗格拉德斯基—高斯定理, 319  
 outlier ..... 离群值, 552  
  
 p-norm ..... p 范数, 327  
 p-refinement finite elements ..... p—改进 (有限元分析), 213  
 Paley-Wiener theorems ..... 佩利—维纳诸定理, 305  
 parameterization  
     local ..... 局部参数化, 455  
 Parseval theorem ..... 帕塞瓦尔定理, 337

- Pascal triangle identity ..... 帕斯卡三角, 245  
 periodogram ..... 周期图, 497  
 periodic function ..... 周期方程, 305  
 Person's linear correlation coefficient ..... 皮尔逊线性相关系数, 551  
 Pfaff's transformation formula ..... 法夫转换公式, 234  
 Pfaff-Saalschütz summation formula ..... 法夫——萨尔舒茨求和公式, 235  
 piecewise linear approximation ..... 分段线性近似, 213  
 piecewise smooth boundary ..... 分段平滑边界, 324  
 Plancherel's theorem ..... 普朗歇尔定理, 311  
 Pochhammer symbol ..... 伯赫哈默尔符号, 231, 233  
 Poincaré inequality ..... 庞加莱不等式, 346  
 Poisson equation ..... 泊松方程, 391  
 Poisson summation formula ..... 泊松求和式, 311  
 polar coordinate system ..... 极坐标系, 350  
 polynomial  
     Chebyshev, first kind ..... 第一类切比雪夫多项式, 260  
     Chebyshev, second kind ..... 第二类切比雪夫多项式, 260  
     complete ..... 完全多项式, 210  
     Hermite ..... 埃米特多项式, 245  
     Jacobi ..... 雅各比多项式, 253  
     Lagrange ..... 拉格朗日多项式, 273  
     Laguerre ..... 拉盖尔多项式, 249  
     monic orthogonal ..... 首一正交多项式, 237  
     orthonormal ..... 标准正交多项式, 237  
 potential  
     adjoint double layer ..... 伴随双层位势, 401  
     adjoint double layer ..... 伴随双层位势算子, 399  
     Bessel ..... 贝塞尔位势, 352  
     double layer ..... 双层位势算子, 402  
     Newton ..... 牛顿位势, 391  
     single layer ..... 单层位势, 397  
 potential influence ..... 潜在影响因子, 554  
 power series ..... 幂级数, 504  
 powerset ..... 幂集, 324  
 probability density function (PDF)  
     ..... 概率密度方程, 557  
     Gaussian ..... 高斯概率密度方程, 509  
     joint ..... 联合概率密度方程, 574  
 probability density function (PDF) ..... 概率密度方程, 543  
 profile likelihood function ..... 描述似然方程, 546  
 projection method ..... 映射法, 199  
 pseudospectral ..... 伪光谱, 215  
 quadrature points ..... 求积点, 279

- quadrature rule ..... 求积法则, 279  
 quadrature weights ..... 求积权重, 279  
 quasi-uniform  
     globally ..... 全局拟一致, 455, 466  
 quasi-likelihood function (QLHF) ..... 准似然方程, 575  
 quasi-maximum likelihood estimation (QMLE) ..... 准最大似然估计, 573  
 quasi-uniform  
     locally ..... 局部拟一致, 455  
 quotient space ..... 商空间, 327
- r-refinement finite elements ..... r—改进 (有限元分析), 213  
 Radmacher random variable ..... 拉德马赫随机变量, 535  
 random sign variable ..... 随机符号变量, 535  
 random walk ..... 随机游走, 478  
 range ..... 值域, 364  
 rapidly decreasing function ..... 急减方程, 350  
 rapidly decreasing space ..... 急减空间, 350  
 rectangular (quadrature) rule ..... 矩形 (求积) 法则, 279  
 reference element ..... 参考元, 409, 455  
 residual function (projection method) , 199  
 residual sum of squares (RSS) ..... 残差平方和, 545  
 Riemann-Stieltjes integral ..... 黎曼—斯蒂尔杰斯积分, 236  
 Riesz map ..... 里兹映射, 361  
 Riesz representation theorem ..... 里兹表现定理, 359, 386  
 Riesz-Schauder theorem ..... Riesz-Schauder 定理, 373  
 Robin boundary value condition ..... 罗宾边界值条件, 321, 385  
 Rodrigues formula  
     Hermite polynomial ..... 罗德里格斯公式 (埃米特多项式), 245  
     Jacoby polynomial ..... 罗德里格斯公式 (雅各比多项式), 253  
     Laguerre polynomial ..... 罗德里格斯公式 (拉盖尔多项式), 249  
     Legendre polynomial ..... 罗德里格斯共识 (勒让德多项式), 257  
     the first kind Chebyshev polynomial ..... 罗德里格斯公式 (第一类切比雪夫多项式), 261  
     the second kind Chebyshev polynomial ..... 罗德里格斯公式 (第二类切比雪夫多项式), 261  
 rotation ..... 旋转, 406, 409  
 Runge phenomenon ..... 龙格现象, 282
- saturated model ..... 饱和模型, 544  
 Schur complement ..... Schur 补, 526  
 Schwarz inequality ..... 施瓦茨不等式, 346  
 score (maximum likelihood) ..... 评分 (最大似然估计), 560  
 second order linear differential equation  
     Hermite polynomial ..... 二阶线性微分方程 (埃米特多项式), 248  
     Jacobi Polynomial ..... 二阶线性微分方程 (雅各比多项式), 257  
     Laguerre polynomial ..... 二阶线性微分方程 (拉盖尔多项式), 253

- selected points ..... 选点, 215  
 selection matrix ..... 选择矩阵, 496  
 self-adjoint operator, 319  
 self-similarity ..... 自相似性, 510  
 seminorm  
     ..... 半范数, 326  
     trivial ..... 平凡半范数, 327  
 seminormed vector space ..... 半赋范向量空间, 330  
 sensitivity function (maximum likelihood) ..... 敏感方程 (最大似然估计), 560  
 sequence space  
     countably infinite dimensional ..... 可数无限维数列空间, 328  
 Sherman-Woodbury-Morrison formula ..... Sherman-Woodbury-Morrison 公式, 528  
 sigma  
     additivity ..... sigma 可加, 325  
     algebra ..... sigma 代数, 324  
     finite ..... sigma 有限, 326  
 Simpson's (quadrature) rule ..... 辛普森 (求积) 法则, 282  
 sinc function  
     normalized ..... 标准化 sinc 方程, 389  
     unnormalized ..... 非标准化 sinc 方程, 389  
 sine integral ..... 正弦积分, 389  
 singular integral  
     weak ..... 弱奇异积分, 395, 398  
 Smolyak algorithm ..... Smolyak 算法, 221  
 Sobolev space  
     ..... 索伯列夫空间, 272  
 Sobolev-Slobodeckij  
     norm ..... Sobolev-Slobodeckij 范数, 338  
     seminorm ..... Sobolev-Slobodeckij 半范数, 338  
     space ..... Sobolev-Slobodeckij 空间, 338  
 Sobolev-Slobodeckij norm ..... Sobolev-Slobodeckij 范数, 355  
 sparse grid ..... 稀疏网格, 221  
 spectral density ..... 光谱密度, 488  
 spectral elements ..... 光谱元, 213  
 spectral method ..... 光谱法, 202  
 spectral method ..... 有限元法, 202  
 spherical coordinate system ..... 三维坐标系, 351, 393  
 standardized predictive residual ..... 标准化预测残差, 553  
 standardized residual ..... 标准化残差, 552  
 Steklov-Poincaré operator ..... Steklov-Poincaré 算子, 423  
 stochastic trend process ..... 随机趋势过程, 478  
 stiffness matrix ..... 刚度矩阵, 439  
 Stokes equations ..... 斯托克斯公式, 275  
 Stone-Weierstrass Theorem ..... 斯通-魏尔斯特拉斯定理, 202

- Strang's First Lemma ..... 斯特朗第一引理, 442  
 Strang's Second Lemma ..... 斯特朗第二引理, 443  
 studentized residual ..... 学生化残差, 552  
 subadditivity ..... 次可加, 326  
 support ..... 支撑, 323  
 surface integral ..... 表面积分, 395, 398  
 symmetrization (probaility theory) ..... 对称化 (概率论), 537
- taxicab geometry ..... 出租车距离, 327  
 Taylor series ..... 泰勒级数, 504  
 Taylor-Maclaurin series ..... 麦克劳林级数, 504  
 telescoping sum ..... 裂项和, 241  
 tempered distribution space ..... 缓增分布空间, 350, 387  
 tensor ..... 张量, 209  
 test space ..... 试探空间, 349  
 three-term recurrence relation  
 Hermite polynomial ..... 三项递推关系 (埃米特多项式), 247  
 Laguerre polynomial ..... 三项递推关系 (拉盖尔多项式), 251  
 Legendre polynomial ..... 三项递推关系 (勒让德多项式), 259  
 monic polynomial ..... 三项递推关系 (首一正交多项式), 239  
 orthonormal polynomial ..... 三项递推关系 (标准正交多项式), 239  
 the first kind Chebishev polynomial ..... 三项递推关系 (第一类切比雪夫多项式), 261  
 the second kind Chebishev polynomial ..... 三项递推关系 (第二类切比雪夫多项式), 261  
 trace ..... 迹, 320, 374  
 transition matrix ..... 转移矩阵, 208  
 transition probability ..... 转移概率, 208  
 trapezoidal (quadrature) rule ..... 梯形 (求积) 法则, 280  
 trial space ..... 检测空间, 461  
 triangle inequality ..... 三角不等式, 329  
 trigonometric integral ..... 三角积分, 389  
 triple 三元组, 325  
 truncation  
 Chebyshev ..... 切比雪夫截断, 204, 266  
 truncation ..... 截断, 294
- unbiased estimate ..... 无偏估计, 558  
 underdetermined system ..... 欠定系统, 530  
 uniform cone condition ..... 均匀法锥条件, 353  
 uniform convergence ..... 均匀收敛, 215  
 uniform distribution ..... 均匀分布, 257  
 uniform norm ..... 均匀范数, 327  
 union ..... 并集, 324  
 unit step function ..... 单位阶跃方程, 511
- value function ..... 价值方程, 123

- VAR  
    (vector autoregression) ..... 向量自回归, 495  
    reduced form ..... 缩减式向量自回归, 498  
    SVAR ..... 结构向量自回归, 499  
variance-stabilizing transformation (VST) ..... 方差稳定变换, 555  
vector space  
    normed ..... 赋范向量空间, 329  
vector space ..... 向量空间, 326  
  
wavenumber ..... 波数, 301  
weight function ..... 权重方程, 236  
weighted residual methods ..... 加权残差法, 199  
Woodbury formula ..... Woodbury 公式, 528  
  
zero vector ..... 零向量, 326  
ZFC axioms ..... ZFC 集合论, 324