



单位代码 10006

学 号 17377234

分 类 号 TN953

北京航空航天大学
B E I H A N G U N I V E R S I T Y

毕业设计(论文)

范数球交集约束优化问题的应用研究

学 院 名 称	<u>数学科学学院</u>
专 业 名 称	<u>统计学</u>
学 生 姓 名	<u>张思泓</u>
指 导 教 师	<u>张奇业</u>

2022 年 6 月

论文封面书脊

论
文
题
目

姓
名

北
京
航
空
航
天
大
学



本人声明

我声明，本论文及其研究工作是由本人在导师指导下独立完成的，在完成论文时所利用的一切资料均已在参考文献中列出。

作者：张思泓

签字：张思泓

时间：2022 年 6 月



本科生毕业设计（论文）任务书

I、毕业设计（论文）题目：

范数球交集约束优化问题的应用研究

II、毕业设计（论文）所使用的原始资料（数据）及设计技术要求：

1. 数学分析，高等代数以及最优化理论基础
2. 会一种计算机语言，比如 Matlab 或 Python 或 R 语言

III、毕业设计（论文）工作内容：

1. 完成广义典型相关分析相关理论的学习，包括：
 - 1) 典型相关分析(CCA)的含义，模型及求解；
 - 2) 广义典型相关分析的模型(GCCA)模型及求解该模型优化问题的算法，以及算法的性质；
 - 3) 广义正则典型相关分析(RGCCA)模型及求解该模型优化问题的算法，以及该算法的性质；
 - 4) 稀疏广义典型相关分析(SGCCA)模型及求解该模型优化问题的算法，以及该算法的性质；
2. 学习范数球交集约束优化问题的算法应用：
 - 1) 学习如何将向量分别投影于三种范数球交集约束的可行域；
 - 2) 学习线搜索算法相关理论；学习线搜索准则；
 - 3) 学习梯度下降算法，并结合投影设计梯度投影法；
 - 4) 学习 BB 步长有关理论；并结合梯度投影法设计带有 BB 步长的非精确线搜索梯度投影算法；
 - 5) 将上述算法编写成 Python 程序。
3. 对比稀疏广义典型相关分析(SGCCA)应用中原始子问题算法与应用范数球交集约束优化问题的投影算法设计出的带有 BB 步长的梯度投影法在解的质量和求解效率方面的异同，写出分析报告。



IV、主要参考资料:

《Variable selection for generalized canonical correlation analysis. Biostatistics》

《Regularized Generalized Canonical Correlation Analysis (RGCCA)》

《Projections onto the Intersection of a One-Norm Ball or Sphere and a Two-Norm Ball or Sphere》

《最优化：建模、算法与理论》

数学科学学院 统计学专业类 180923 班

学生 张思泓

毕业设计（论文）时间：2022 年 1 月 5 日至 2022 年 5 月 22 日

答辩时间：年月日

成绩：

指导教师：

兼职教师或答疑教师（并指出所负责部分）：

系（教研室） 主任（签字）：

注：任务书应该附在已完成的毕业设计（论文）的首页。



范数球交集约束优化问题的应用研究

学 生：张思泓

指导教师：张奇业

摘 要

范数球交集约束优化问题是一类常见的最优化子问题。由于使用 l_1 -范数球/球壳约束(或者等价的 l_1 -范数正则化)经常能得到稀疏解, 近些年有关 l_1 -范数球和 l_2 -范数球的交集约束优化问题最为常见。事实上, 自上个世纪 80 年代以来范数球交集约束优化问题广泛存在、应用于生物学、医学、分子化学等各个领域。为方便起见, 将这些子问题分别记为:

(P1) l_1 -范数球和 l_2 -范数球交集的欧氏投影(Euclidean projection);

(P2) l_1 -范数球壳和 l_2 -范数球壳交集上的投影;

(P3) l_1 -范数球和 l_2 -范数球壳交集上的投影。

最近, Liu 和 Wang 给出了统一的框架和算法求解这三类子问题, 他们通过构造一个辅助函数将 P1, P2 和 P3 子问题的求解转化为求该辅助函数的零点, 设计了求解该零点的二分法, 从而可以统一求解上述三类子问题。而稀疏广义典型相关分析(SGCCA)被建模为具有 P3 约束的最大化问题。为此, 本文将求解上述三类子问题的统一算法应用于稀疏广义典型相关分析中, 设计了基于 Liu-Wang 算法的求解 SGCCA 问题的梯度投影法, 采用 BB 步长, 比较了三类子问题在 SGCCA 中在求解质量和效率上的不同; 进而比较了求解 P3 子问题的 Liu-Wang 算法与已有算法在求解质量和效率上的不同, 最后给出对比结果。

关键词: l_1 -范数和 l_2 -范数球交集约束, Liu-Wang 算法, 稀疏广义典型相关分析, 梯度投影法, BB 步长



Research on the Application of Norm Sphere Intersection Constraint Optimization Problem

Author : ZHANG Si-hong

Tutor : ZHANG Qi-ye

Abstract

The norm ball or sphere intersection constrained optimization problems are a kind of common optimization sub-problems. Since l_1 -norm ball/sphere constraints (or equivalent l_1 -norm regularization) are often used to obtain sparse solutions, optimization of intersection constraints of l_1 -norm ball/sphere and l_2 -norm ball/sphere is the most common problem in recent years. In fact, since the 1980s, the norm ball/sphere intersection constrained optimization problems have been widely used in biology, medicine, molecular chemistry and other fields. For convenience, these sub-problems are denoted as:

- (P1) Euclidean projection of the intersection of l_1 -norm ball and l_2 -norm ball;
- (P2) Projections on the intersection of l_1 -norm sphere and l_2 -norm sphere;
- (P3) Projections on the intersection of l_1 -norm ball and l_2 -norm sphere.

Recently, Liu and Wang presented a unified framework and algorithm for solving these three kinds of sub-problems. They transformed the solution of (P1), (P2) and (P3) sub-problems into finding the zeros of the auxiliary function by constructing an auxiliary function, and designed a dichotomy method for solving the zeros, so that the three kinds of sub-problems can be solved uniformly. Sparse generalized canonical correlation analysis (SGCCA) is modeled as a maximization problem with (P3) constraints. Therefore, in this paper, Liu-Wang's unified algorithm for solving the above three sub-problems is applied to SGCCA model, and a gradient projection method with BB step size is designed for solving SGCCA. Through several numerical experiments, the difference of solved quality and efficiency between the three sub-problems in SGCCA are compared. Furthermore, the difference between the Liu-Wang algorithm and the existing algorithm in solving the P3 sub-problem in solved quality and efficiency is also compared. The simulation results show the proposed algorithm are effective.

Key words: l_1 -norm and l_2 -norm ball/sphere intersection constraint, Liu-Wang algorithm, SGCCA, gradient projection method, BB step size



目 录

1 绪论	1
1.1 课题背景及目的	1
1.2 国内外研究状况	2
1.3 课题研究方法	3
1.3.1 研究目标	3
1.3.2 研究内容	3
1.3.3 拟解决的关键问题及难点	
1.4 论文构成及研究内容	3
2 稀疏广义典型相关分析	5
2.1 典型相关分析(CCA)	5
2.1.1 典型相关分析的阐释	5
2.1.2 典型相关分析的定义	6
2.1.3 典型相关分析的解推导	6
2.2 稀疏广义典型相关分析(SGCCA)	7
2.2.1 广义典型相关分析(GCCA)	7
2.2.2 广义正则典型相关分析(RGCCA)	9
2.2.3 稀疏广义典型相关分析(SGCCA)	10
3 基于 Liu-Wang 算法和 BB 步长的 SGCCA 问题的梯度投影法	12
3.1 算法思路	12
3.1.1 线搜索方法	12
3.1.2 梯度下降法	13
3.1.3 块坐标下降法	14
3.1.4 BB 步长	14
3.2 基于 Liu-Wang 算法和 BB 步长的 SGCCA 问题的梯度投影法	15
4 数值仿真及结果	12



4.1 算法思路仿真环境及参数配置.....	12
4.2 数据集介绍.....	12
4.3 评价指标.....	12
4.4 仿真结果.....	12
结论.....	19
致谢.....	20
参考文献.....	22



1 绪论

1.1 课题背景及目的

范数球交集约束优化问题是一类常见的最优化子问题。其一般形式可表述如下：

$$\begin{aligned} \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) \\ \text{s.t. } \|\mathbf{x}\|_p \leq s (\text{or } = s), \\ \|\mathbf{x}\|_q \leq t (\text{or } = t) \end{aligned}$$

其中 $f(\mathbf{x})$ 是定义在 \mathbf{R}^n 上满足某种性质的函数, p, q, s, t 均为正实数。由于使用 l_1 -范数球/球壳约束(或者等价的 l_1 -范数正则化)经常能得到稀疏解, 近些年有关 l_1 -范数球和 l_2 -范数球的交集约束优化问题(即上式中 $p, q = 1, 2$ 的情形)最为常见。例如, 稀疏主成分(sparse Principal Component Analysis, SPCA)^[1-4], 稀疏非负矩阵分解(sparse non-negative matrix factorization, SNNMF)^[5-6]、具有稀疏激活和稀疏连通性的监督在线自编码器^[7]、基于强迫稀疏性投影的字典学习^[8]和稀疏广义典型相关分析(Sparse Generalized Canonical Correlation Analysis, SGCCA)^[9], 等。事实上, 自上个世纪 80 年代以来范数球交集约束优化问题广泛存在、应用于生物学、医学、分子化学等各个领域。为方便起见, 将这些子问题分别记为:

(P1) l_1 -范数球和 l_2 -范数球交集的欧氏投影(Euclidean projection);

(P2) l_1 -范数球壳和 l_2 -范数球壳交集上的投影;

(P3) l_1 -范数球和 l_2 -范数球壳交集上的投影。

最近, Liu 和 Wang 给出了统一的框架和算法求解这三类子问题, 他们通过构造一个辅助函数将(P1), (P2)和(P3)子问题的求解转化为求该辅助函数的零点, 设计了求解该零点的二分法, 从而可以统一求解上述三类子问题。

为此, 本文旨在将求解上述三类子问题的统一算法应用于 SGCCA 中, 比较三类子问题在 SGCCA 中在求解质量和效率上的不同, 进而比较该算法与已有算法在求解质量和效率上的不同。



1.2 国内外研究状况

1936年, Hotelling 提出了典型相关分析(CCA)^[11], 即“使得两组变量的线性函数的相关系数最大值”的问题, 并提出了该问题的一种解法。此即为稀疏广义典型相关分析(SGCCA)的基础。1961年, Horst 提出 CCA 的另一种算法^[12-13], 并尝试将该结果扩展到 3 个及以上的变量集合上。1971年, Kettenring 提出了多个变量集合上的典型相关分析^[14], 由此产生了 GCCA 模型。

在 GCCA 相关问题研究的早期阶段提出的方法中, 部分方法基于目标函数相关系数或协方差的函数, 如 SUMCOR, SSQCOR, SABCOR, SUMCOV, SSQCOV, SABSCOV, 或基于相关系数和协方差二者的函数。这些方法虽然纷繁, 却存在相似性。于是, 在各种模型和方法层出不穷时, 研究者开始致力于用一个框架概括这类问题, 并证明这个框架理论上的合理性, 尝试给出通用的拥有较好性质的算法。

2010年, Hardoon 提出了 SCCA 模型^[15], 即稀疏典型相关分析; 2011年, Tenenhaus 提出了 RGCCA 模型^[9], 即广义正则典型相关分析模型, 并借鉴偏最小二乘路径模型的算法给出求解 RGCCA 模型的算法。Tenenhaus 提出的 RGCCA 模型将上述提到的 SUMCOR 等作为特例包含在内, 且还包含分层模型等诸多此前提出的其他问题的模型, 具有非常强的灵活性; 同时, Tenenhaus 提出求解该稀疏广义典型相关分析的算法, 阐述了该算法拥有的性质。RGCCA 模型本质上是一个目标函数为相关系数的某些函数、约束条件为 l_2 -正则化的优化问题。

2014年 Tenenhaus 在 RGCCA 的基础上进一步提出 SGCCA 模型^[16], 即稀疏广义典型相关分析模型, 为原 RGCCA 模型优化问题的约束条件中, 增加对外权向量 l_1 -范数球体的约束, 即稀疏性的约束后得到的模型, 用于求解得到该问题的变量。Tenenhaus 同时给出了一个拥有较好性质的算法。而 SGCCA 模型本质上为 l_1 -范数球和 l_2 -范数球的交集约束优化问题中的一种, 即属于[10]中所提及的(P3)子问题。

2020年, Liu 和 Wang 给出了统一的框架和算法求解(P1), (P2), (P3)这三类子问题^[10], 他们通过构造一个辅助函数将 l_1 -范数球和 l_2 -范数球的交集约束优化问题的求解转化为求该辅助函数的零点, 设计了求解该零点的二分法。利用[10]中的算法, 可将变量投影到可行域中, 因此在求解稀疏广义典型相关分析的模型问题时可以利用该算法进行求解。



1.3 课题研究方法

1.3.1 研究目标

在学习稀疏广义典型相关分析(SGCCA)模型理论的基础上, 将[10]中求解(P1), (P2), (P3)三类子问题的统一算法应用于 SGCCA 模型求解中, 比较三类子问题在 SGCCA 中在求解质量和效率上的不同, 进而比较[10]中算法与已有算法在稀疏广义典型相关分析(SGCCA)中的求解质量和效率上的不同。

1.3.2 研究内容

- 1、进行文献调研, 学习 SGCCA 模型及相关理论和算法;
- 2、学习[10]中求解(P1), (P2), (P3)三类子问题的统一算法及相关代码;
- 3、将[10]中算法应用于 SGCCA 中, 比较三类子问题在稀疏广义典型相关分析中在求解质量和效率上的不同;
- 4、比较[10]中算法与已有算法在 SGCCA 中的求解质量和效率上的不同。

1.3.3 拟解决的关键问题及难点

- 1、对 SGCCA 模型的理论理解, 以及算法收敛性分析的理解是难点;
- 2、对 SGCCA 原始算法的复现是关键;
- 3、将[10]中算法应用于 SGCCA 中, 其中算法的实现是关键和难点。

1.4 论文构成及研究内容

本文研究主要内容为 SGCCA 有关的理论及现有算法, [10]中范数球交集约束投影算法的学习; 本文主要展示已有 SGCCA 算法与[10]中算法在求解质量、效率上的差异。

第 2 章 稀疏广义典型相关分析, 介绍有关 SGCCA 的理论, 包括 SGCCA 的模型及其由来、以及已有的用于求解该模型的算法及其性质等基本理论。

第 3 章 基于 Liu-Wang 算法和 BB 步长的 SGCCA 问题的梯度投影法, 首先简单介绍设计本文算法所需的线搜索技术、Armijo 准则、BB 步长、梯度下降法以及块坐标下降法等理论, 然后给出基于 Liu-Wang 算法和 BB 步长的 SGCCA 问题的梯度投影法的具体算法步骤, 并对比了三类子问题算法在 SGCCA 问题的求解质量和效率上的异同。

第 4 章 数值仿真与结果, 在两个数据集上比较本文算法与[16]中算法在求解效果, 展示了本文算法在求解 SGCCA 问题上的有效性。

最后是结论。



2 稀疏广义典型相关分析

2.1 典型广义相关分析(CCA)

2.1.1 典型相关分析的阐释

典型相关分析(CCA)是一种研究两组变量之间相关关系的统计分析方法,用以揭示两组变量之间的线性依赖关系^[1]。例如:神射手们肩并肩同时向靶射击,偏差部分源于独立个体误差,部分源于共同因素(如风)。这提供了一个常见的相关理论的介绍。但被讨论的常常仅有风速的水平分量相关,而有时水平分量和垂直分量的复合可能更让人感兴趣^[1]。再比如:研究个体生理与心理差异,需要详细研究有关的变量的集合间的关系:如针对同一个人,可以比较其心理测试分数与其身体测量结果。问题在于,如何确定这些数据显示出的存在于生理与心理的独立关系的数量和性质,以及如何从系统的相关多样性中提取出这些独立关系的合适特征。

2.1.2 典型相关分析的定义

设 $\mathbf{x} = (x_1, x_2, \dots, x_p)^T$ 和 $\mathbf{y} = (y_1, y_2, \dots, y_q)^T$ 是两组随机变量,且 $V(\mathbf{x}) = \Sigma_{11} (> 0)$, $V(\mathbf{y}) = \Sigma_{22} (> 0)$, $\text{Cov}(\mathbf{x}, \mathbf{y}) = \Sigma_{12}$, 即有:

$$V \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad (2.2.1)$$

其中 $\Sigma_{12} = \Sigma_{21}^T$ 。

将 \mathbf{x} 和 \mathbf{y} 分别进行线性组合,得到 $u = \mathbf{a}^T \mathbf{x}$ 和 $v = \mathbf{b}^T \mathbf{y}$, 其中 $\mathbf{a} = (a_1, a_2, \dots, a_p)^T$, $\mathbf{b} = (b_1, b_2, \dots, b_q)^T$ 皆为非零常数向量。典型相关分析定义为如下的优化问题:

$$\begin{aligned} \max_{\mathbf{a}, \mathbf{b}} \quad & \rho(\mathbf{a}^T \mathbf{x}, \mathbf{b}^T \mathbf{y}) \\ \text{s.t.} \quad & \mathbf{a}^T \Sigma_{11} \mathbf{a} = 1, \\ & \mathbf{b}^T \Sigma_{22} \mathbf{b} = 1 \end{aligned} \quad (2.2.2)$$

即使得 u 和 v 的相关系数 $\rho(\mathbf{a}^T \mathbf{x}, \mathbf{b}^T \mathbf{y})$ 达到最大值,约束条件可以避免解的不必要重复。对随机向量进行线性组合可以有效对随机向量进行降维,同时在求得上述最优优化问题的解以后可以通过对线性组合系数的观察初步断定在两个随机向量的相关关



系中，每一个随机向量中的哪些随机变量更为活跃。

在实际应用中，由于常常无法得知随机向量的协方差矩阵 $V\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ ，常常使用样本协方差阵对其进行估计。记数据矩阵为 $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)^T$ ， $Y = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_q)^T$ ，则样本协方差矩阵为

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix} \quad (2.2.3)$$

其中

$$\begin{aligned} \mathbf{S}_{11} &= \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T, \quad \mathbf{S}_{22} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T \\ \mathbf{S}_{12} &= \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{y}_i - \bar{\mathbf{y}})^T, \quad \mathbf{S}_{21} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \\ \bar{\mathbf{x}} &= \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i, \quad \bar{\mathbf{y}} = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \end{aligned}$$

2.1.3 典型相关分析的解推导

利用高等代数的知识，我们可以分析得到典型相关分析(CCA)的解。

首先有结论：

设 Σ 是 p 阶对称矩阵，记 Σ 的特征值由大到小依次为 $\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_m^2 > 0$ ，其中 m 为该矩阵的秩， λ_i 为 λ_i^2 的算术平方根， $i = 1, \dots, m$ ；相应的一组正交单位特征向量为 $\beta_1, \beta_2, \dots, \beta_m$ ，则：

$$\max \quad \mathbf{x}^T \Sigma \mathbf{x} = \lambda_1^2 \quad (\mathbf{x} = \beta_1) \quad (2.2.4)$$

回到典型相关分析。令 $\alpha = \Sigma_{11}^{-1/2} \mathbf{a}$ ， $\beta = \Sigma_{22}^{-1/2} \mathbf{b}$ ，于是典型相关分析的约束条件化为：

$$\alpha^T \alpha = 1, \beta^T \beta = 1 \quad (2.2.5)$$

目标函数为：

$$\mathbf{a}^T \Sigma_{12} \mathbf{b} = \alpha^T \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{1/2} \beta \quad (2.2.6)$$

又由柯西不等式 $(\mathbf{x}^T \mathbf{y})^2 \leq (\mathbf{x}^T \mathbf{x})(\mathbf{y}^T \mathbf{y})$ 知：



$$\begin{aligned}
(\mathbf{a}^T \Sigma_{12} \mathbf{b})^2 &= (\boldsymbol{\alpha}^T \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2} \boldsymbol{\beta})^2 \\
&\leq (\boldsymbol{\alpha}^T \boldsymbol{\alpha}) [(\Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2} \boldsymbol{\beta})^T (\Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2} \boldsymbol{\beta})] \\
&= \boldsymbol{\beta}^T \Sigma_{22}^{-1/2} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1/2} \boldsymbol{\beta}
\end{aligned} \tag{2.2.7}$$

矩阵 $\Sigma_{22}^{-1/2} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1/2}$ 为实对称阵，其特征值从大到小依次为 $\rho_1 \geq \rho_2 \geq \dots \geq \rho_m > 0$ ，其中 m 为矩阵 Σ_{12} 的秩，与矩阵 $\Sigma_{22}^{-1/2} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1/2}$ 的秩相同， ρ_i 为 ρ_i^2 的算术平方根， $i = 1, \dots, m$ ；相应的单位正交特征向量为 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_m$ ，则由上述结论可知，当 $\boldsymbol{\beta} = \boldsymbol{\beta}_1$ 时，目标函数取得最大值 ρ_1 。为满足约束条件 $\boldsymbol{\alpha}^T \boldsymbol{\alpha} = 1$ 且 $\boldsymbol{\beta}^T \boldsymbol{\beta} = 1$ ，此时 $\boldsymbol{\alpha} = \frac{1}{\rho_1} \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2} \boldsymbol{\beta}_1$ ，且同时可使柯西不等式的等号成立。记 $\frac{1}{\rho_1} \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2} \boldsymbol{\beta}_1 = \boldsymbol{\alpha}_1$ ，因此，当 $\mathbf{a} = \Sigma_{11}^{-1/2} \boldsymbol{\alpha}_1$ 且 $\mathbf{b} = \Sigma_{22}^{-1/2} \boldsymbol{\beta}_1$ 时，上述典型相关分析有最大值 ρ_1 。

2.2 稀疏广义典型相关分析(SGCCA)

2.2.1 广义典型相关分析(GCCA)

广义典型相关分析由典型相关分析^[9](CCA)拓展而来，这种拓展具体体现在：

1. 数据集合数量：从 2 个数据集合拓展到 3 个及以上数据集合；
2. 目标函数：从相关系数拓展到可为相关系数本身，或相关系数的绝对值，或相关系数的平方。

广义典型相关分析(GCCA)考虑变量集合间的连接形成的网络。记 J 个定义在同一总体的 p_j 维随机列向量为 $\mathbf{x}_j = (x_{j1}, x_{j2}, \dots, x_{jp_j})^T$ ，以及 p_j 维非随机的列向量 $\boldsymbol{\alpha}_j = (\alpha_{j1}, \alpha_{j2}, \dots, \alpha_{jp_j})^T$ 。通过定义矩阵 $\mathbf{C} = c_{jk}$ 来考虑由随机向量间连接形成的网络：

当 \mathbf{x}_j 与 \mathbf{x}_k 相连时 $c_{jk} = 1$ ，否则 $c_{jk} = 0$ 。考虑两个线性组件 $\eta_j = \sum_{h=1}^{p_j} \alpha_{jh} x_{jh} = \boldsymbol{\alpha}_j^T \mathbf{x}_j$ ，

$\eta_k = \sum_{h=1}^{p_k} \alpha_{kh} x_{kh} = \boldsymbol{\alpha}_k^T \mathbf{x}_k$ ，两个随机变量 η_j 与 η_k 的相关系数为：

$$\rho(\boldsymbol{\alpha}_j^T \mathbf{x}_j, \boldsymbol{\alpha}_k^T \mathbf{x}_k) = \frac{\boldsymbol{\alpha}_j^T \Sigma_{jk} \boldsymbol{\alpha}_k}{(\boldsymbol{\alpha}_j^T \Sigma_{jj} \boldsymbol{\alpha}_j)^{1/2} (\boldsymbol{\alpha}_k^T \Sigma_{kk} \boldsymbol{\alpha}_k)^{1/2}} \tag{2.3.1}$$

其中： $\Sigma_{jj} = E(\mathbf{x}_j \mathbf{x}_j^T)$ ， $\Sigma_{kk} = E(\mathbf{x}_k \mathbf{x}_k^T)$ 且 $\Sigma_{jk} = E(\mathbf{x}_j \mathbf{x}_k^T)$ 。所有的 Σ_{jj} 均满秩，

$j=1, \dots, J$ 。

广义典型相关分析(GCCA)的模型如下:

$$\begin{aligned} \max_{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_J} \quad & \sum_{j,k=1; j \neq k}^J c_{jk} g(\rho(\boldsymbol{\alpha}_j^T \mathbf{x}_j, \boldsymbol{\alpha}_k^T \mathbf{x}_k)) \\ \text{s.t.} \quad & \text{Var}(\boldsymbol{\alpha}_j^T \mathbf{x}_j) = 1, \quad j = 1, \dots, J \end{aligned} \quad (2.3.2)$$

其中函数 $g(x) = x$, $g(x) = |x|$ 或 $g(x) = x^2$, 依据研究者的需要取其中一种。一般来说, $g(x) = |x|$ 与 $g(x) = x^2$ 一般给出相近的结果。采用先前已给出的符号, 该优化问题与下述问题等价:

$$\begin{aligned} \max_{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_J} \quad & \sum_{j,k=1; j \neq k}^J c_{jk} g(\boldsymbol{\alpha}_j^T \boldsymbol{\Sigma}_{jk} \boldsymbol{\alpha}_k) \\ \text{s.t.} \quad & \boldsymbol{\alpha}_j^T \boldsymbol{\Sigma}_{jj} \boldsymbol{\alpha}_j = 1, \quad j = 1, \dots, J \end{aligned} \quad (2.3.3)$$

Tenchaus 在给出广义典型相关分析(GCCA)的模型时, 同时给出了原始的算法。

广义典型相关分析(GCCA)的算法:

A. 初始化:

A1. 随机选取 J 个向量 $\tilde{\boldsymbol{\alpha}}_1^0, \tilde{\boldsymbol{\alpha}}_2^0, \dots, \tilde{\boldsymbol{\alpha}}_J^0$ 。

A2. 计算 $\tilde{\boldsymbol{\alpha}}_1^0, \tilde{\boldsymbol{\alpha}}_2^0, \dots, \tilde{\boldsymbol{\alpha}}_J^0$:

$$\tilde{\boldsymbol{\alpha}}_j^0 = [(\tilde{\boldsymbol{\alpha}}_j^0)^T \boldsymbol{\Sigma}_{jj}^{-1} \tilde{\boldsymbol{\alpha}}_j^0]^{-1/2} \boldsymbol{\Sigma}_{jj}^{-1} \tilde{\boldsymbol{\alpha}}_j^0$$

对迭代指标 $s = 0, 1, \dots$ (直至收敛) 循环:

对向量下标 $j = 1, 2, \dots, J$ 循环:

B. 计算中间向量

$$\begin{aligned} \boldsymbol{\nu}_j^s = & \sum_{k < j} c_{jk} w [\text{Cov}((\boldsymbol{\alpha}_j^s)^T \mathbf{x}_j, (\boldsymbol{\alpha}_k^{s+1})^T \mathbf{x}_k)] (\boldsymbol{\alpha}_k^{s+1})^T \mathbf{x}_k \\ & + \sum_{k > j} c_{jk} w [\text{Cov}((\boldsymbol{\alpha}_j^s)^T \mathbf{x}_j, (\boldsymbol{\alpha}_k^s)^T \mathbf{x}_k)] (\boldsymbol{\alpha}_k^s)^T \mathbf{x}_k \end{aligned}$$

其中 $w(x) = 1$ 当 $g(x) = x$, $w(x) = x$ 当 $g(x) = x^2$, $w(x) = \text{sign}(x)$ 当 $g(x) = |x|$ 。

C. 计算 $\boldsymbol{\alpha}_j^{s+1}$

$$\boldsymbol{\alpha}_j^{s+1} = [\text{Cov}(\mathbf{x}_j, \boldsymbol{\nu}_j^s)^T \boldsymbol{\Sigma}_{jj}^{-1} \text{Cov}(\mathbf{x}_j, \boldsymbol{\nu}_j^s)]^{-1/2} \boldsymbol{\Sigma}_{jj}^{-1} \text{Cov}(\mathbf{x}_j, \boldsymbol{\nu}_j^s)$$

结束

结束

且证明了: 该算法保证每次得到 $\boldsymbol{\alpha}_j^{s+1}$ 后将其带入目标函数, 目标函数值单调递增。



2.2.2 广义正则典型相关分析(RGCCA)

广义正则典型相关分析(RGCCA)由广义典型相关分析(GCCA)得到。广义正则典型相关分析(RGCCA)的模型^[9]如下:

考虑 n 个样本上测得的 J 个中心化变量块 $\mathbf{X}_1, \dots, \mathbf{X}_J$ 。 $\mathbf{X}_j, j=1, 2, \dots, J$ 的一行表示随机向量 \mathbf{x}_j^T 的一个实现; \mathbf{X}_j 的一列 \mathbf{x}_{jh} 表示一个在 n 个样本上观测的变量, \mathbf{x}_{jhi} 表示变量 \mathbf{x}_{jh} 在第 i 个样本上的取值。 $\mathbf{C} = \{c_{jk}\}$ 是用于描述块间关系网的矩阵, 若块 $j: \mathbf{X}_j$ 与块 $k: \mathbf{X}_k$ 相连时 $c_{jk}=1$, 否则 $c_{jk}=0$ 。取样本协方差阵为 $\mathbf{S}_{jj} = \frac{1}{n} \mathbf{X}_j^T \mathbf{X}_j$, $\mathbf{S}_{jk} = \frac{1}{n} \mathbf{X}_j^T \mathbf{X}_k$ 。在多重共线性或样本个数小于变量个数 ($n < p_j$) 时, 样本协方差矩阵是原 Σ_{jj} 的一个相当差的估计。矩阵中特征值等于或接近 0 的个数, 样本协方差阵比真正的协方差阵多得多。广义正则典型相关分析(GGCCA)考虑到该情况, 采用一类用单位阵 \mathbf{I} 和样本协方差阵 \mathbf{S}_{jj} 的线性组合 $\{\hat{\mathbf{S}}_{jj} = \tau_j \mathbf{I} + (1 - \tau_j) \mathbf{S}_{jj}, 0 \leq \tau_j \leq 1\}$ 来估计协方差矩阵 Σ_{jj} 。 $\hat{\mathbf{S}}_{jj}$ 称为 Σ_{jj} 的收缩估计, 而 τ_j 为收缩常数。记 α_j, η_j, ν_j 的样本分别为 $\mathbf{a}_j, \mathbf{y}_j, \mathbf{z}_j$ 。此时广义正则典型相关分析(RGCCA)的优化问题为:

$$\begin{aligned} \max_{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_J} \quad & \sum_{j,k=1; j \neq k}^J c_{jk} g(\text{cov}(\mathbf{X}_j \mathbf{a}_j, \mathbf{X}_k \mathbf{a}_k)) \\ \text{s.t.} \quad & (1 - \tau_j) \text{var}(\mathbf{X}_j \mathbf{a}_j) + \tau_j \|\mathbf{a}_j\|_2^2 = 1, \quad j=1, \dots, J \end{aligned} \quad (2.3.4)$$

仿照广义典型相关分析的算法, 给出广义正则典型相关分析的算法:

广义正则典型相关分析(RGCCA)的算法:

B. 初始化:

A1. 随机选取 J 个向量 $\tilde{\mathbf{a}}_1^0, \tilde{\mathbf{a}}_2^0, \dots, \tilde{\mathbf{a}}_J^0$ 。

A2. 计算 $\tilde{\mathbf{a}}_1^0, \tilde{\mathbf{a}}_2^0, \dots, \tilde{\mathbf{a}}_J^0$:

$$\mathbf{a}_j^0 = [(\tilde{\mathbf{a}}_j^0)^T [\tau_j \mathbf{I} + (1 - \tau_j) \frac{1}{n} \mathbf{X}_j^T \mathbf{X}_j]^{-1} \tilde{\mathbf{a}}_j^0]^{1/2} [\tau_j \mathbf{I} + (1 - \tau_j) \frac{1}{n} \mathbf{X}_j^T \mathbf{X}_j]^{-1} \tilde{\mathbf{a}}_j^0$$

对迭代指标 $s = 0, 1, \dots$ (直至收敛) 循环:

对向量下标 $j = 1, 2, \dots, J$ 循环:

B. 计算中间向量



$$\mathbf{z}_j^s = \sum_{k < j} c_{jk} w [\text{Cov}((\mathbf{X}_j \mathbf{a}_j^s, \mathbf{X}_k \mathbf{a}_k^{s+1}))] \mathbf{X}_k \mathbf{a}_k^{s+1} \\ + \sum_{k > j} c_{jk} w [\text{Cov}(\mathbf{X}_j \mathbf{a}_j^s, \mathbf{X}_k \mathbf{a}_k^s)] \mathbf{X}_k \mathbf{a}_k^s$$

其中 $w(x) = 1$ 当 $g(x) = x$, $w(x) = x$ 当 $g(x) = x^2$, $w(x) = \text{sign}(x)$ 当 $g(x) = |x|$ 。

C. 计算 \mathbf{a}_j^{s+1}

$$\mathbf{a}_j^{s+1} = [(\mathbf{z}_j^s)^T \mathbf{X}_j [\tau_j \mathbf{I} + (1 - \tau_j) \frac{1}{n} \mathbf{X}_j^T \mathbf{z}_j^s]^{-1} \mathbf{X}_j] j^T \mathbf{z}_j^s]^{-1/2} \\ \times [\tau_j \mathbf{I} + (1 - \tau_j) \frac{1}{n} \mathbf{X}_j^T \mathbf{X}_j]^{-1} \mathbf{X}_j^T \mathbf{z}_j^s$$

结束
结束

Tenchaus 同样证明, 该算法使得每次得到 \mathbf{a}_j^{s+1} 时目标函数值单调递增。

2.2.3 稀疏广义典型相关分析(SGCCA)

稀疏广义典型相关分析(SGCCA)由广义正则典型相关分析(RGCCA)得到, 是解决广义正则典型相关分析时进行稀疏变量选择优化问题。

稀疏广义典型相关分析(SGCCA)的模型^[11]如下:

$$\begin{aligned} & \underset{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_J}{\text{argmax}} \quad \sum_{j, k=1; j \neq k}^J c_{jk} g(\text{cov}(\mathbf{X}_j \mathbf{a}_j, \mathbf{X}_k \mathbf{a}_k)) \\ & \text{subject to} \quad \|\mathbf{a}_j\|_2 = 1 \text{ and } \|\mathbf{a}_j\|_1 \leq s_j, \quad j = 1, \dots, J \end{aligned} \quad (2.3.5)$$

观察该优化问题的约束条件, 可知该约束为 l_1 -范数球和 l_2 -范数球壳的交集。因此, 除了使用 Tenchaus 提出的算法对稀疏广义典型相关分析(SGCCA)进行求解外, 还可以利用梯度投影法, 结合 Liu-Wang 给出的向量在 l_1 -范数球和 l_2 -范数球壳的交集的投影的结论, 给出新的算法。

在提出稀疏广义典型相关分析(SGCCA)时, 给出的用于求解该模型的优化问题的算法如下:

稀疏广义典型相关分析(SGCCA)的算法:



需要的数据:

J 个数据块 $\mathbf{X}_1, \dots, \mathbf{X}_J$, J 个 l_1 -范数球约束 s_1, \dots, s_J , 用以表示块间连接网的矩阵 \mathbf{C} , 选取 $g(x) = x$ 或 $g(x) = x^2$ 或 $g(x) = |x|$;

随机选取的 J 个单位向量 $\mathbf{a}_1^0, \mathbf{a}_2^0, \dots, \mathbf{a}_J^0$ 。

对迭代指标 $s = 0, 1, \dots$ 循环:

对向量下标 $j = 1, 2, \dots, J$ 循环:

A. 计算中间向量

$$\mathbf{z}_j^s = \sum_{k=1}^{j-1} c_{jk} w[\text{Cov}((\mathbf{X}_j \mathbf{a}_j^s, \mathbf{X}_k \mathbf{a}_k^{s+1}))] \mathbf{X}_k \mathbf{a}_k^{s+1} \\ + \sum_{k=j+1}^J c_{jk} w[\text{Cov}(\mathbf{X}_j \mathbf{a}_j^s, \mathbf{X}_k \mathbf{a}_k^s)] \mathbf{X}_k \mathbf{a}_k^s$$

其中 $w(x) = 1$ 当 $g(x) = x$, $w(x) = x$ 当 $g(x) = x^2$, $w(x) = \text{sign}(x)$ 当 $g(x) = |x|$ 。

B. 计算 \mathbf{a}_j^{s+1}

$$\mathbf{a}_j^{s+1} \leftarrow \frac{S\left(\frac{1}{n} \mathbf{X}_j^T \mathbf{z}_j^s, \lambda_{1j}\right)}{\|S\left(\frac{1}{n} \mathbf{X}_j^T \mathbf{z}_j^s, \lambda_{1j}\right)\|_2}$$

其中 S 是如下定义的软阈值算子:

$$S(a, \lambda) = \text{sign}(a) \max(0, |a| - \lambda)$$

当 $\|\mathbf{a}_j^{s+1}\|_1 \leq s_j$ 时, $\lambda_{1j} = 0$; 否则选择 λ_{1j} 使得 $\|\mathbf{a}_j^s\|_1 = s_j$

s_j 满足: $0 \leq s_j \leq \sqrt{p_j}$

结束

直到目标函数值 $h(\mathbf{a}_1, \dots, \mathbf{a}_J) = \sum_{j,k=1; j \neq k}^J c_{jk} g(\text{Cov}(\mathbf{X}_j \mathbf{a}_j, \mathbf{X}_k \mathbf{a}_k))$ 差:

$$h(\mathbf{a}_1^{s+1}, \dots, \mathbf{a}_J^{s+1}) - h(\mathbf{a}_1^s, \dots, \mathbf{a}_J^s) \leq \varepsilon$$

结束, 返回 $\mathbf{a}_1^{s+1}, \dots, \mathbf{a}_J^{s+1}$



3 基于 Liu-Wang 算法和 BB 步长的 SGCCA 问题的梯度投影法

3.1 算法思路

对于有约束的优化问题，本算法采用的思路是首先将其视为无约束优化问题，每一次求解一个变量时采用梯度下降法，初始步长选取 BB 步长；为满足约束条件，再将每一次迭代的变量投影到可行域中；而求解多个变量采用坐标下降法。为适应该算法，只需考虑目标函数的倒数，得到满足条件的解后再对目标函数的值取倒数即可。下面分别介绍该算法用到的技术。

下考虑无约束优化问题：

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) \quad (2.4.1)$$

其中 $f(\mathbf{x})$ 是 $\mathbb{R}^n \rightarrow \mathbb{R}$ 的函数。

3.1.1 线搜索方法

线搜索类算法^[22]的过程大致为：给定当前迭代点 \mathbf{x}^k ，通过某些方法确定该点的下降方向 \mathbf{d}^k 和步长 α_k ，位移到下一个迭代点

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha_k \mathbf{d}^k \quad (2.4.2)$$

其中 \mathbf{d}^k 具体要满足：

$$(\mathbf{d}^k)^T \nabla f(\mathbf{x}^k) < 0 \quad (2.4.3)$$

该式将保证每次位移到下一个迭代点时，目标函数值会变小。由于可运用不同的方法选取下降方向 \mathbf{d}^k 和步长 α_k ，因此产生了不同的线搜索算法。

构造辅助函数 $\phi(\alpha) = f(\mathbf{x}^k + \alpha \mathbf{d}^k)$ ，其含义是将目标函数限制在射线 $\mathbf{x}^k + \alpha \mathbf{d}^k$ 上。线搜索的目标是选取合适的步长 α_k 使得该辅助函数在 α_k 处的取值 $\phi(\alpha_k)$ 尽可能地小。若取步长 α_k 使得辅助函数在此时有最小值，则称该线搜索算法为精确线搜索算法；反之，若选取步长 α_k 不能达到该辅助函数的最小值，但消耗较小的计算量即可达到一个理想的目标函数值下降，则称该线搜索算法为非精确线搜索算法。本算法即采用非精确线搜索算法。为明确非精确线搜索算法的何种步长 α_k 为理想的，以及保证算法的收敛性，需介绍线搜索准则。此处介绍本算法所采用的 Armijo 准则。



Armijo 准则^[23]是一个常用的线搜索准则，其目的是保证每一步迭代充分下降。设 \mathbf{d}^k 是点 \mathbf{x}^k 处的下降方向，若：

$$f(\mathbf{x}^k + \alpha \mathbf{d}^k) \leq f(\mathbf{x}^k) + c_1 \alpha \nabla f(\mathbf{x}^k)^T \mathbf{d}^k \quad (2.4.4)$$

其中 c_1 是一个常数；则称步长 α 满足 Armijo 准则。

3.1.2 梯度下降法

梯度下降法^[22]的本质是利用目标函数的一阶导数信息选取每次迭代的下降方向，即选择负梯度方向为下降方向。梯度下降法是梯度类算法中最基本的算法。

对于光滑函数 $f(\mathbf{x})$ ，在迭代点 \mathbf{x}^k 处的泰勒展开为：

$$\phi(\alpha) = f(\mathbf{x}^k) + \alpha \nabla f(\mathbf{x}^k)^T \mathbf{d}^k + \mathcal{O}(\alpha^2 \|\mathbf{x}^k\|^2) \quad (2.4.5)$$

由柯西不等式，当 α 足够小的时候，取 $\mathbf{d}^k = -\nabla f(\mathbf{x}^k)$ 时，该函数下降最快。梯度法即取下降方向 \mathbf{d}^k 为负梯度方向的算法。

梯度下降法的算法流程如下：

梯度下降算法：

对优化问题： $\min_{\mathbf{x}} f(\mathbf{x})$ ：

A. 初始化：

合理选取出发点 \mathbf{x}^0 ，以及此处的函数值 $f(\mathbf{x}^0)$

B. 迭代主体： 对迭代指标 $itr = 1, 2, \dots$ ：

计算迭代点 \mathbf{x}^k 处的下降方向 $\mathbf{d}^k = -\nabla f(\mathbf{x}^k)$

选取合适的步长 α_k

计算下一迭代点 $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha_k \mathbf{d}^k$ 以及该处的函数值 $f(\mathbf{x}^{k+1})$

若：

(函数值差值绝对值 $|f(\mathbf{x}^{k+1}) - f(\mathbf{x}^k)|$ 小于函数值的停机条件 ftol

且

自变量差的范数 $\|\mathbf{x}^{k+1} - \mathbf{x}^k\|$ 小于自变量的停机条件 xt看)

或

(梯度的范数 $\|\nabla f(\mathbf{x}^{k+1})\|$ 小于梯度的停机条件 gtol)

结束

否则：

位移到下一迭代点

返回：自变量及其函数值。

3.1.3 块坐标下降法



由于稀疏广义典型相关分析(SGCCA)有多个自变量,而梯度下降法针对的问题仅有1组自变量,因此为使用梯度下降法,需要每次将问题分解为共 J 组只有一个自变量的优化问题。由于稀疏广义典型相关分析(SGCCA)中自变量的约束条件是分别将各个自变量约束在范数球、范数球壳内,因此可以轻松的尝试进行分解。

在迭代主体部分,对自变量的下标依次进行仅以该下标的自变量向量为自变量,其余下标的自变量为常量,目标函数不变的梯度下降法,即块坐标下降法;对所有自变量下标运用完梯度下降法后,观察上一次迭代后与本次迭代后得到的函数值的差值,若该差值小于指定的停机条件则结束迭代。块坐标下降法的算法流程如下:

块坐标下降算法:

对优化问题: $\min_{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_J} f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_J)$:

A. 初始化:

合理选出发点 $\mathbf{x}_1^0, \mathbf{x}_2^0, \dots, \mathbf{x}_J^0$, 以及此处的函数值 $f(\mathbf{x}_1^0, \mathbf{x}_2^0, \dots, \mathbf{x}_J^0)$

B. 迭代主体: 对迭代指标 $s = 1, 2, \dots$:

对下标 $i = 1, 2, \dots, J$:

将 $\mathbf{x}_1^{s+1}, \mathbf{x}_2^{s+1}, \dots, \mathbf{x}_{i-1}^{s+1}, \mathbf{x}_{i+1}^s, \dots, \mathbf{x}_J^s$ 视为已知的常量,

梯度下降法求 \mathbf{x}_i^{s+1} 。

若:

函数值差值绝对值 $|f(\mathbf{x}_1^{s+1}, \mathbf{x}_2^{s+1}, \dots, \mathbf{x}_J^{s+1}) - f(\mathbf{x}_1^s, \mathbf{x}_2^s, \dots, \mathbf{x}_J^s)|$ 小于函数值停机条件 ftol

结束

返回: 自变量及其函数值。

3.1.4 BB 步长

BB 步长是一种初始步长,它属于非精确线搜索的步长,因此使用 BB 步长的时候,需要结合线搜索准则设计算法。

考虑无约束优化问题 $\min_{\mathbf{x}} f(\mathbf{x})$ 其中目标函数 $f(\mathbf{x})$ 可微。梯度下降法的迭代格式

$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha_k \mathbf{d}^k$ 。BB 步长即步长 α_k 。记 $\mathbf{s}^k = \mathbf{x}^{k+1} - \mathbf{x}^k$, $\mathbf{y}^k = \nabla f(\mathbf{x}^{k+1}) - \nabla f(\mathbf{x}^k)$,

定义两种 BB 步长 $\frac{(\mathbf{s}^k)^T \mathbf{s}^k}{(\mathbf{s}^k)^T \mathbf{y}^k}$ 和 $\frac{(\mathbf{s}^k)^T \mathbf{y}^k}{(\mathbf{y}^k)^T \mathbf{y}^k}$ 。经验表明,采用 BB 步长的梯度法往往比采用

固定步长的梯度法得到更好的结果。

结合 Armijo 线搜索准则和 BB 步长的梯度下降法流程如下:



基于 Armijo 线搜索准则和 BB 步长的梯度下降法:

对优化问题: $\min_{\mathbf{x}} f(\mathbf{x})$:

C. 初始化:

合理选取出发点 \mathbf{x}^0 , 以及此处的函数值 $f(\mathbf{x}^0)$

D. 迭代主体: 对迭代指标 $itr = 1, 2, \dots$:

计算迭代点 \mathbf{x}^k 处的下降方向 $\mathbf{d}^k = -\nabla f(\mathbf{x}^k)$

选取初始步长为 BB 步长 $\alpha_k = \frac{(\mathbf{s}^k)^T \mathbf{s}^k}{(\mathbf{s}^k)^T \mathbf{y}^k}$ 或 $\alpha_k = \frac{(\mathbf{s}^k)^T \mathbf{y}^k}{(\mathbf{y}^k)^T \mathbf{y}^k}$

计算下一迭代点 $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha_k \mathbf{d}^k$ 以及该处的函数值 $f(\mathbf{x}^{k+1})$

对步长收缩次数 $nls = 1, 2, \dots, 10$:

若此 α_k 满足 $f(\mathbf{x}^k + \alpha_k \mathbf{d}^k) \leq f(\mathbf{x}^k) - \rho \alpha_k \|\nabla f(\mathbf{x}^k)\|_2$ 或步长收缩次数大于等于 10 次(其中 ρ 为线搜索准则中下降量参数, 取 $1e-4$): 退出循环

否则: 步长 α_k 收缩为 $\eta \alpha_k$ (其中 η 为收缩系数, 取 0.2), 线搜索次数增加一次。

若:

(函数值差值绝对值 $|f(\mathbf{x}^{k+1}) - f(\mathbf{x}^k)|$ 小于函数值的停机条件 $ftol$

且

自变量差的范数 $\|\mathbf{x}^{k+1} - \mathbf{x}^k\|$ 小于自变量的停机条件 $xtol$)

或

(梯度的范数 $\|\nabla f(\mathbf{x}^{k+1})\|$ 小于梯度的停机条件 $gtol$)

结束

否则:

位移到下一迭代点

返回: 自变量及其函数值。

3.2 基于 Liu-Wang 算法和 BB 步长的 SGCCA 问题的梯度投影法

前面讨论了无约束条件的梯度下降法, 但 SGCCA 模型是有约束的优化问题, 因此实际算法采用梯度投影法, 在梯度下降的每一步都需要将该点投影到可行域中。

于是我们根据坐标下降法, 结合 Armijo 线搜索准则, 设计了基于 Liu-Wang 算法和 BB 步长的 SGCCA 问题的梯度投影法, 其算法步骤如下:

基于 Liu-Wang 算法和 BB 步长的 SGCCA 问题的梯度投影法:



对优化问题: $\min_{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_J} f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_J)$:

A. 初始化:

随机初始化单位向量 $\mathbf{x}_1^0, \mathbf{x}_2^0, \dots, \mathbf{x}_J^0$;

计算此处的函数值 $f(\mathbf{x}_1^0, \mathbf{x}_2^0, \dots, \mathbf{x}_J^0)$;

B. 迭代主体: 对迭代指标 $s = 1, 2, \dots$:

对下标 $i = 1, 2, \dots, J$:

将 $\mathbf{x}_1^{s+1}, \mathbf{x}_2^{s+1}, \dots, \mathbf{x}_{i-1}^{s+1}, \mathbf{x}_{i+1}^s, \dots, \mathbf{x}_J^s$ 视为已知的常量, 带入目标函数中;

调用梯度投影法函数(函数 1)求解 \mathbf{x}_i^{s+1}

结束

计算函数值差的绝对值 $|f(\mathbf{x}_1^{s+1}, \mathbf{x}_2^{s+1}, \dots, \mathbf{x}_J^{s+1}) - f(\mathbf{x}_1^s, \mathbf{x}_2^s, \dots, \mathbf{x}_J^s)|$

若

此绝对值小于给定的停机条件, 则结束循环。

否则

继续迭代。

返回自变量及此时的函数值。

函数 1: 基于 BB 步长的梯度投影法算法

j 表示该梯度投影法求解的自变量的下标;

A. 初始化

待求解的变量记为 \mathbf{x} ; 其余自变量 $\mathbf{x}_1^{s+1}, \mathbf{x}_2^{s+1}, \dots, \mathbf{x}_{j-1}^{s+1}, \mathbf{x}_j^s, \mathbf{x}_{j+1}^s, \dots, \mathbf{x}_J^s$ 视为已知;

初始化 \mathbf{x} 的出发点为 $\mathbf{x}^0 = \mathbf{x}_j^s$;

该梯度投影法的目标函数为: $F(\mathbf{x}) = f(\mathbf{x}_1^{s+1}, \dots, \mathbf{x}_{j-1}^{s+1}, \mathbf{x}, \mathbf{x}_{j+1}^s, \dots, \mathbf{x}_J^s)$

可行域为: $\|\mathbf{x}\|_1 \leq s_j, \|\mathbf{x}\|_2 = 1$

B. 投影 (分为 P1\|P2\|P3 三种投影情况)

利用 Liu-Wang 的算法, 将 \mathbf{x}^0 投影到可行域内得到 \mathbf{w} , 将出发点更新为投影点 $\mathbf{x}^0 = \mathbf{w}$; 计算此时的函数值

C. 迭代主体

对迭代指标 $k = 1, 2, \dots$:

上一步得到的自变量为 \mathbf{x}^k ,

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha_k \mathbf{d}^k = \mathbf{x}^k - \tau \nabla f(\mathbf{x}^k);$$

a. 投影 (分为 P1\|P2\|P3 三种子问题情况, 与上述投影保持一致)

利用 Liu-Wang 的算法将 \mathbf{x}^{k+1} 投影到可行域内得到 \mathbf{w} ;

更新 \mathbf{x}^{k+1} 为 $\mathbf{x}^{k+1} = \mathbf{w}$;

$$\text{记 } \mathbf{s} = \mathbf{x}^{k+1} - \mathbf{x}^k, \mathbf{y} = \nabla f(\mathbf{x}^{k+1}) - \nabla f(\mathbf{x}^k)$$

b. 判断是否收敛:

若:

满足停机条件:

返回 \mathbf{x}^{k+1}



结束

c. 计算 BB 步长

当 $\mathbf{s}^T \mathbf{y} \neq 0$ 时:

选取初始步长为 BB 步长 $\alpha_k = \frac{(\mathbf{s}^k)^T \mathbf{s}^k}{(\mathbf{s}^k)^T \mathbf{y}^k}$ 或 $\alpha_k = \frac{(\mathbf{s}^k)^T \mathbf{y}^k}{(\mathbf{y}^k)^T \mathbf{y}^k}$;

否则将步长选取为默认步长 τ

将步长限制在 $1\text{e-}20$ 至 $1\text{e}20$ 之间



4 数值仿真及结果

4.1 仿真环境与参数配置

本数值仿真实验的cpu型号为 11th Gen Intel(R) Core(TM) i5-1135G7 @ 2.40GHz 2.42 GHz, 软件配置使用了 python3.8 版本, 调用了相关库: numpy, math, matplotlib, time, pandas, sklearn。

4.2 数据集介绍

为观察基于 Liu-Wang 算法和 BB 步长的 SGCCA 问题的梯度投影法应用于稀疏广义典型相关分析(SGCCA)的性质并和已有的求解稀疏广义典型相关分析算法相对比, 设计一个数据集如下:

数据集 1: 设计一个拥有 3 个变量块 $\mathbf{X}_j, j=1, 2, 3$ 的数据集合。其中:

$$\mathbf{X}_j = \mathbf{u}_j \mathbf{w}_j^T + \mathbf{E}_j, \quad j=1, 2, 3 \quad (2.5.1)$$

其中 $\mathbf{u}_1, \mathbf{u}_2$ 和 \mathbf{u}_3 是 50 维的向量, 服从均值为零, 相关系数分别为 $\text{cor}(\mathbf{u}_1, \mathbf{u}_2) = 0$, $\text{cor}(\mathbf{u}_1, \mathbf{u}_3) = 0.7$ 且 $\text{cor}(\mathbf{u}_2, \mathbf{u}_3) = 0.7$ 。 \mathbf{w}_1 是 200 维的向量, \mathbf{w}_2 是 500 维的向量, \mathbf{w}_3 是 700 维的向量。 $\mathbf{w}_1, \mathbf{w}_2$ 和 \mathbf{w}_3 仅前 75 维非零, 且每个元素都独立同分布, 服从区间 $[-0.3, -0.2] \cup [0.2, 0.3]$ 上的均匀分布。 \mathbf{E}_j 是 50 行 p_j 列的矩阵, 用以表示微小扰动。 \mathbf{E}_j 每个元素均独立同分布, 服从均值为 0, 方差为 0.2 的正态分布。

设计块间连接关系如下: 块 1 与块 3 相连接, 块 2 与块 3 相连接, 而块 1 与块 2 不连接。这是由于块的构造时相关系数决定的。因此用于表示块间连接关系的矩阵 \mathbf{C} 为

$$\mathbf{C} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \quad (2.5.2)$$

由于设计的数据集合较为规律, 且无法从之观察到应用于实际的更为复杂的数据集合时的运行效果, 因此将程序运用于一个实际的数据集合上。

数据集 2: 数据集合来源于 53 名患有儿童脑瘤的病人, 共包含 3 个数据块。第一个数据块为 15702 个特征的基因数据, 第二个数据块为 1229 个特征的染色体数据, 第三个数据块表示肿瘤的 3 个位置。为探究这三者间的关系, 设计块间连接关系的矩阵如



下:

$$\mathbf{C} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \quad (2.5.3)$$

4.3 评价指标

设计两个指标: 敏感度与准确度, 用以考察算法的性能。

敏感度为程序解得的向量 \mathbf{a}_j 中非零元素所在的位置上 \mathbf{w}_j 相应的也非零的位置总数除以 \mathbf{w}_j 非零元素总个数。

准确度设计为程序解得的向量 \mathbf{a}_j 中零元素所在的位置上 \mathbf{w}_j 相应的也为零的位置总数除以 \mathbf{w}_j 零元素总个数。这两个指标指示了该算法能否按照稀疏广义标准相关分析原本的目的得到想要的 \mathbf{a}_j 。

4.4 仿真结果

对上述数据 1 运行已有的用于解决 SGCCA 的优化问题(2.3.5)的算法与基于 Liu-Wang 算法和 BB 步长的 SGCCA 问题的梯度投影法, 属于 P3 子问题情况, 观察算法的结果如下:

首先观察基于 Liu-Wang 算法和 BB 步长的 SGCCA 问题的梯度投影法。在 10000 次数值实验中, 该算法每次都能使得目标函数取得最大值(停机条件为函数值差值小于 $1e-16$), 且在一万次运行后, 几乎得到相同的解。

得到运行结果后敏感性与准确性指标的结果如图所示:

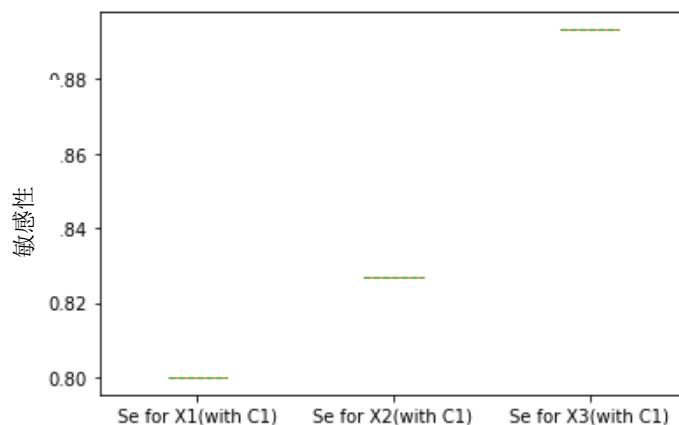


图 3.1 数据集合敏感性

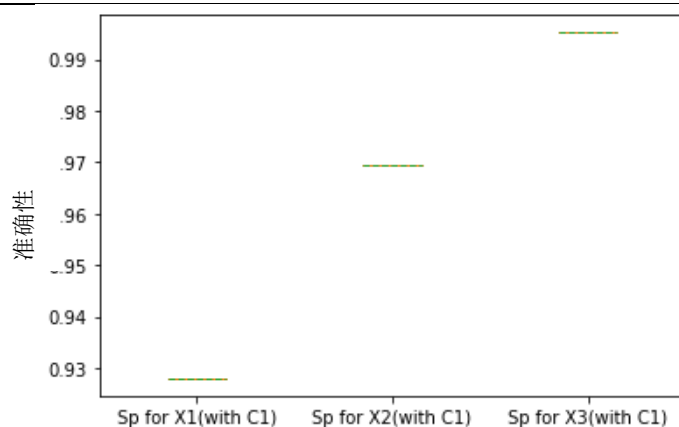


图 3.2 数据集合准确性

可以看出, 该算法求解得到的结果对所有的 \mathbf{a}_j 都得到了较高的敏感度与准确度, 均达到 80%以上; 而 \mathbf{a}_j 的向量维数越大, 敏感性越高, 准确性越低。

总的来说, 该算法的优势是每次都收敛到想要的最大值, 且得到预期的结果; 但是由于该算法仅在初始化时随机产生初始向量, 算法本身缺乏随机性, 因此对这个增加了 \mathbf{E}_j 的数据集而言, 算法结果完全将 \mathbf{E}_j 视为数据集本质的一部分, 而得到相应的结果, 是不太妥当的。另外, 由于算法本身随机性的缺乏, 仅仅采用了最基础的梯度下降法, 算法运行缓慢, 平均耗时较长。

对比已有的用于求解 SGCCA 的算法结果, 该已有算法运算量较小, 且运行时间较短, 但运行结果每次都有差别, 并非次次都能收敛到想要的最大值; 敏感度与准确度指标因此不同, 但总的来说较高。

将约束条件修改为 P1 子问题类型: l_1 -范数球和 l_2 -范数球交集上的投影, 重新编写程序中投影函数部分, 得到相关运行结果如下:

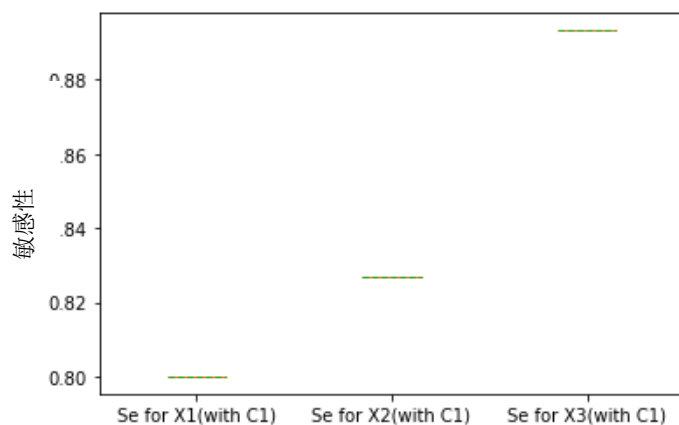


图 3.3 数据集合敏感性

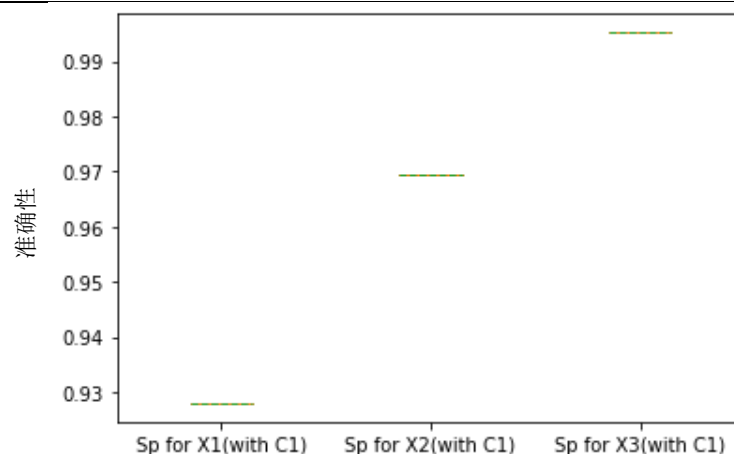


图 3.4 数据集合准确性

和 P3 子问题作为约束条件(原模型的约束条件)相比, 目标函数值均为 15.48855541347, 在误差限为 $1e-8$ 的停机条件下, 目标函数值并未增大或减小。但是 P2 子问题的运行时间平均为 1.594s, 较 P3 子问题的平均运行时间 2.051s 小。P1 子问题情况下算法的其他性质较 P3 子问题的情况相似, 但结果的准确性指标相对更高, 敏感性指标相对更低。

将约束条件修改为 P2 子问题类型: l_1 -范数球壳和 l_2 -范数球壳交集上的投影, 重新编写程序中投影函数部分, 得到相关运行结果如下:

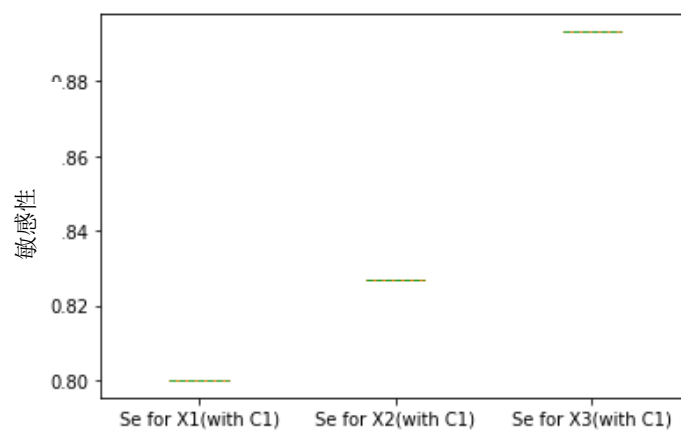


图 3.5 数据集合敏感性

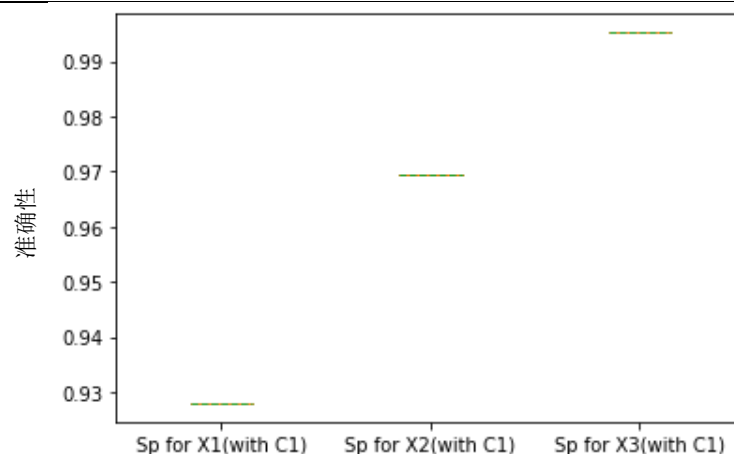


图 3.6 数据集合准确性

和 P3 子问题作为约束条件(原模型的约束条件)相比, 目标函数值均为 15.48855541347, 在误差限为 $1e-8$ 的停机条件下, 目标函数值并未增大或减小。但是 P2 子问题的运行时间平均为 1.497s, 较 P3 子问题的平均运行时间 3.069s 小。P2 子问题情况下算法的其他性质较 P3 子问题的情况相似, 但结果的准确性指标相对更高, 敏感性指标相对更低。

同时, 我们可以发现(P1)(P2)得到了几乎相同的结果。

对数据集合 2 再运行基于 Liu-Wang 算法和 BB 步长的 SGCCA 问题的梯度投影法时, 发现并不是每次都能得到预期的最大值, 仅有约 30%的情况得到系列结果中最大的那个结果, 运行效果大大降低。分析可能的原因如下:

1. 该数据中第三个块的行数大于列数, 可能因此问题的病态;
2. 第一个块特征非常多, 中心化后数据差值变小, 可能影响程序运行准确性;
3. 由于用于数据存储时仅存储到小数点后 8 位, 中心化后可能存在较大误差;

但对于运行得到目标函数取值为多次运行的最大值的的结果, 其 a_j 的敏感性和准确性依然较高, 保有量好的性质。但更换数据集合后, 程序保留了运行缓慢的特征。

可能的改进方案:

1. 将简单的梯度下降法替换为带有随机性的随机梯度下降方法, 以减少运行时间;
2. 对大型的数据集合, 计算、存储数据至小数点后 16 位;



结论

稀疏广义典型相关分析原模型属于范数球交集约束优化问题的 P3 子问题；约束优化问题的将范数球交集约束优化问题应用于稀疏广义典型相关分析时，设计的基于 Liu-Wang 算法和 BB 步长的 SGCCA 问题的梯度投影法在简单数据集合上运行时，收敛较好，结果缺乏随机性但性质较好，但耗时较长；应用于复杂的大型数据集合时效果较差，并非次次都能得到预期的目标函数最大值，但凡是得到最大值的结果性质较好。

修改稀疏广义典型相关分析模型的约束条件为 P1 子问题情形时，较 P3 子问题而言，目标函数值并未增大或减小，运行时间较短，准确性指标较高而敏感性指标较低。

修改稀疏广义典型相关分析模型的约束条件为 P2 子问题情形时，较 P3 子问题而言，目标函数值并未增大或减小，运行时间最短，准确性指标较高而敏感性指标较低。P1 子问题的情形和 P2 子问题的情形得到相似的结果，因此若需要追求准确性，则可以采用 P2 子问题的情形。



致谢

我今年已二十三有余，和同年级一同毕业的同学比，既大了一岁又并不比身边的人更为优异。小学六年，中学六年，而我本科也用了五年才勉强修完学业。我既不聪慧，期间多受到许多人帮助和关怀；因为受到如此多的帮助和关怀，支持着我完成了本次毕业设计。

首先当感谢我毕业设计的指导老师张奇业老师。最开始选题时，我并未学习过最优化的理论内容，更未在计算机上实现过相关算法。张奇业老师悉心指导，实施督促我学习有关理论，定时整理学习结果；同时指导我定期参加老师的讨论班，聆听他人的学习报告开拓知识面，学习更多相关知识，同时熟悉观摩学习优秀学长学姐如何学习研究有关问题。张奇业老师费心收集文献，并悉心指导我如何阅读文献；老师还编辑相关文档，教我学习如何使用 Latex 编写我的学习成果，同时在我提交学习成果后，还会一一教会我修改文档，十分费心。正因有张奇业老师如此悉心指导，我才能学习到有关知识，完成毕业设计相关工作。

其次感谢我的母亲和我的姥姥。这五年来，我们相互扶持，共度难关，一同欢乐。高考以来，虽然我已经成年，但母亲一直在经济上全力支持我，十分宽裕。我遇到困难，母亲也从不指责我，不厌其烦地帮我疏导情绪，之后又设身处地为我出谋划策。不论何时打给母亲的电话，母亲总是把我的事情清清楚楚的记在心上。母亲在哪，那里就是家。我和母亲同面对生活的苦与甜，母亲除了作为我的母亲受苦更多担当更重以外，我更是为母亲平添了太多心理的烦恼和实实在在的困难。然而母亲永远都如此细心耐心的爱我，无论多困难都使我享受家的温暖和安全，使我能栖息在家中舔舐伤口，整理行囊，恢复精力，一次次出发，一次次跌倒后再爬起。我的母亲对待亲朋好友十分宽顺温和，有强烈的同理心，朋友很多也受人尊敬喜爱，我常常因为母亲而收到母亲朋友的关爱。尽管母亲谦虚，但我心里，母亲是我天下第一的母亲，而即使抛开母亲的身份，我对其人格和智慧更是倍加尊敬，更加因此感到有这样的母亲的幸运。我的姥姥极其爱护我偏袒我，在我心里如同英雄一般又有实力又有思想，是我最钦佩的人之一。当听说我和母亲遇到困难需要帮助时，她正在田间开荒种田，享受晚年，但姥姥甚至顾不得地里悉心照顾好几个月的庄稼和蔬果，也顾不得收拾家里，拎着最必要最少的行李就由我的表哥送上火车，一个人千里迢迢来到异乡全力支持我的母亲和我。姥姥不熟悉住所周围的路，也听不太懂附近人晦涩的方言，在异乡并无在老家如此多亲近的亲友，也无法放开了想种地便种地，生活处处不便，但姥姥从未对母亲或者对我有过一次抱怨，反而全方位的支持着我和母亲。我生病时，心情也低落，姥姥每天和我谈心，一遍又一遍的开导我。姥姥乐观积极，且善明事理，又热心肠，像太阳一般散发温暖明媚的光芒，亲戚多把姥姥当作家族主事的人来对待敬爱。我从姥姥那里学到很多，更是被那种坚强、积极、自律的性格感染着。姥姥是我的榜样。

感谢我的高中班主任马勇老师。马老师对待班上的每一个人都如此细心，对同学们的照顾甚至多于对自己孩子的照顾。我从认识马老师开始，就总是叨扰马老师指导我；马老师除了对同学们的学业十分关心，也十分关心同学们的身体健康、心理状态、性格、三观，方方面面事无巨细。我能有成长和进步，全然离不开马老师细心耐心的指导。即使毕业了，马老师也十分关切各个同学的状态和动向，盼着每个人都能有所成就、幸福快乐。高中时就多多使马老师费心，我毕业后遇到困难，马老师仍关心我，尽可能的帮助我；我每每想到马老师对我们的关怀，总觉得自己辜负了这样难得的一位老师的心和期望。但我不论是希望能达到老师的期望还是面对自己的未来，我都一



定更加严于律己，做一个品行端正、对社会有益的人；不论再遇到何种困难，我定会乐观面对；任命运曲折，风雨摧残，我也视其为雨露滋润，傲然笑着面对，不负老师谆谆教诲。

感谢我的朋友们。感谢夏丝的爸爸妈妈和夏丝。大一新生开学，我不懂事，急着想证明自己能够自理自立，执意不让母亲送我到学校报到，约了夏丝一同去报道。夏丝的母亲和夏丝和我坐着绿皮火车，我又坐着夏丝妈妈朋友的车来到学校，一路多受照顾/后来，我又总是受到照顾，被介绍着参加各种活动，参加实践队，又多受到夏丝父亲的照顾。我在这些活动里拓宽了眼界，因此认识了好多优秀可爱的朋友们，获得了太多快乐的回忆。后来，受多个事情的打击，我情绪崩溃，嚎啕大哭的时候，夏丝安慰我，为我加油鼓劲。我实在没帮上过夏丝什么，但多多受到照顾，作为朋友十分惭愧。感谢轩轩。我们从认识起，我就没怎么和你讨论过什么有趣的娱乐的内容，多半是数学的乐趣，学业的事宜、诗词文章的美妙和未来的志向。我最喜欢的是军训时刚认识你那会的日子，后来我们一起上数学兴趣班的日子，还有那份虽然模糊了具体的事情但一同喜欢着数学、一起盼望着未来的心情。高考前大家都很忙，你却愿意带着有些自暴自弃的我自习，与我谈论着日常。在本科阶段，我很难想象你遇到的困难和压抑，但只要我找到你，你却总是热情的给我回应与帮助。我总是对和你相识这件事感到开心，同时也偶尔难以置信。我能认识你，成为你的朋友，也倍感幸运。感谢若凡。若凡是我遇到的最果敢坚强、明辨是非，又温柔可爱又聪明、坦率的人。我需要帮助，若凡尽力帮我。我约若凡一起玩，总是向若凡抱怨生活里无聊的琐事，但若凡总也不嫌我毫无长进，反而次次都愿意帮我出谋划策。每次见到若凡的时候，若凡总是明媚地微笑着，温暖了我的心。我不感谢缘分，而要感谢若凡与我共度的时光，支持着我成长起来，一路向前。感谢子墨。之前中学时，只是球友和同窗，大概知道子墨非常努力也很聪明。本科期间，因为自己总是叫嚷着要到贵校读书，子墨就非常热心的帮助我，分享学习资料，给出中肯建议，敲打我的心浮气躁，又斥责我不应该畏难。不仅仅十分优秀，子墨总是非常冷静、坚定，我之分敬佩，也是我学习的小目标。如此说，是因为作为朋友，我怎肯甘于人后，定要奋勇向前，来日和子墨一较高下哈哈。感谢老蒋。作为朋友，老蒋对我非常照顾，许我怎么任性怎么闹都行，过去也总是听我一次又一次倒苦水，承受我的负能量和我的各方面的压力。感谢我收到的你一切的照顾和关心，帮我挨过最最困难的日子。感谢哥。这些年来都常常关心我，鼓励我。哥一直相信我一定没问题，只需要更为自信。我虽然还是很难对自己很有自信，但我听了你的话更加相信一定会越来越好。感谢诗麒麟，造诣大哥，我给你们填了许多麻烦，但大家对我都很好。诗麒麟总是陪着我，督促我学习，分享日常、爱好，是我本科生活遇到的最好的朋友之一。而造诣说话从来没有废话，对我一直很严格，也总能一针见血戳破我的把戏，督促我好好学习天天向上。感谢春晖、和其他一起打过乒乓球的小伙伴们。虽然只是打球，但那段打球的日子是最灰暗的日子的第一个转折点，自那以后，我知道了原来有这么多这么优秀又有趣的小伙伴，原来大学还可以过得如此开心。春晖善于思考，也喜欢交流，很认可我。感谢一起去社会实践的小伙伴们，虽然每天都很忙碌，但有幸认识大家，收获一段永难忘怀的经历，实在是幸运。



参考文献

- [1] I. T. Jolliffe, N. T. Trendafilov, and M. Uddin. A modified principal component technique based on the lasso. *Journal of computational and Graphical Statistics*, 2003, 12(3): 531-547.
- [2] R. Luss and M. Teboulle. Convex approximations to sparse pca via lagrangian duality. *Operations Research Letters*, 2011, 39(1): 57-61.
- [3] D. M. Witten, R. Tibshirani, and T. Hastie. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*, 2009, 10(3): 515-534.
- [4] R. Luss and M. Teboulle. Conditional gradient algorithms for rank-one matrix approximations with a sparsity constraint. *SIAM Review*, 2013, 55(1): 65-98.
- [5] P O. Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, 11 2004, 5: 1457-1469.
- [6] V. K Potluru, S. M Plis, J. Le Roux, Barak A Pearlmutter, Vince D Calhoun, and Thomas P Hayes. Block coordinate descent for sparse nmf. *arXiv preprint arXiv:2013, 1301.3527*.
- [7] M. Thom and G. Palm. Sparse activity and sparse connectivity in super-vised learning. *Journal of Machine Learning Research (JMLR)*, 2013 04, 14: 1091-1143.
- [8] M. Thom, M. Rapp, and G. Palm. Efficient dictionary learning with sparseness-enforcing projections. *International Journal of Computer Vision*, 2015, 114(2-3): 168–194.
- [9] A. Tenenhaus and M. Tenenhaus. Regularized generalized canonical correlation analysis. *Psychometrika*, 2011 04, 76(2): 257-284.
- [10] Hongying Liu, Hao Wang, and Mengmeng Song. Projections onto the intersection of a one-norm ball or sphere and a two-norm ball or sphere. *Journal of Optimization Theory and Applications*, 2020, 187: 520-534.
- [11] H. Hotelling. Relations between two sets of variates. *Biometrika*, 1936, 28, 321-277.
- [12] P. Horst. Relations among m sets of measures. *Psychometrika*, 1961, 26, 129-149.
- [13] P. Horst. Factorial analysis of data matrices. New York, Holt, Rinehart and Winston, 1965.
- [14] J. R. Kettenring(1971). Canonical analysis of several sets of variables. *Biometrika*, 58, 433-451.



-
- [15] Hardoon, D.R., Shawe-Taylor, J. Sparse canonical correlation analysis. *Mach Learn.* 2011, 83, 331-353.
 - [16] A. Tenenhaus, C. Philippe, V. Guillemot, Kim-Anh K.-A. Le Cao, J. Grill, and V Frouin. Variable selection for generalized canonical correlation analysis. *Biostatistics*, 2014, 15(3): 569–583.
 - [17] Karl Pearson F.R.S. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1901, 2(11): 559-572.
 - [18] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 1933, 24(6):417.
 - [19] Christian D Sigg and Joachim M Buhmann. Expectation-maximization for sparse and non-negative pca. *Proceedings of the 25th international conference on Machine learning*, 2008, pages 960–967.
 - [20] David R. Hardoon, Sandor Szedmak, and Shawe-Taylor John. Canonical correlation analysis: An overview with application to learning methods. *Technical Report CSD-TR-03-02*, 2003 11, 1-39.
 - [21] A. Tenenhaus and M. Tenenhaus. Regularized generalized canonical correlation analysis for multiblock or multigroup data analysis. *European Journal of Operational Research*, 2014 11, 238: 391-403.
 - [22] 刘浩洋,户将,李勇锋,文再文. 最优化: 建模、算法与理论, 高等教育出版社. 北京, 2020
 - [23] 刘红英,夏勇,周水生. 数学规划基础, 航空航天大学出版社. 北京, 2012.