

学校代码: 10475

学 号: 104753140673

河 南 大 学 研 究 生 硕 士 学 位 论 文

求解 ℓ_1 - ℓ_1 -范数型优化问题的算法及其应用

Methods for ℓ_1 - ℓ_1 -Norm Minimization
Problems with Applications

专 业 名 称: 运 筹 学 与 控 制 论

专 业 代 码: 0 7 0 1 0 5

研 究 方 向: 最优化理论与算法

年 级: 二 零 一 四 级

研 究 生 姓 名: 陈 丽 雅

导师 姓名、职称: 肖 运 海 教 授

完 成 日 期: 二 零 一 七 年 四 月

论 文 主 题 词: 运筹学/数学规划/信号处理

关于学位论文独立完成和内容创新的声明

本人向河南大学提出硕士学位申请。本人郑重声明：所呈交的学位论文是本人在导师的指导下独立完成的，对所研究的课题有新的见解。据我所知，除文中特别加以说明、标注和致谢的地方外，论文中不包括其他人已经发表或撰写过的研究成果，也不包括其他人为获得任何教育、科研机构的学位或证书而使用过的材料。与我一同工作的同事对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位申请人(学位论文作者)签名:_____

20 年 月 日

关于学位论文著作权使用授权书

本人经河南大学审核批准授予硕士学位。作为学位论文的作者，本人完全了解并同意河南大学有关保留、使用学位论文的要求，即河南大学有权向国家图书馆、科研信息机构、数据收集机构和本校图书馆等提供学位论文(纸质文本和电子文本)以供公众检索、查阅。本人授权河南大学出于宣扬、展览学校学术发展和进行学术交流等目的，可以采取影印、缩印、扫描和拷贝等复制手段保存、汇编学位论文(纸质文本和电子文本)。

(涉及保密内容的学位论文在解密后使用本授权书)

学位获得者(学位论文作者)签名:_____

20 年 月 日

学位论文指导教师签名:_____

20 年 月 日

摘 要

本文研究求解基于鲁棒人脸识别的 ℓ_1 - ℓ_1 范数优化问题的原始对偶两种交替方向乘子法, 求解基于压缩传感 ℓ_1 - ℓ_1 范数优化模型的非单调谱梯度算法. 给出算法的收敛性结果, 并模拟数据验证交替方向乘子法的数值有效性.

第一章, 简单介绍鲁棒人脸识别和压缩传感问题以及相应的优化模型; 给出本文所用的优化基础知识; 列举求解凸优化问题的交替方向乘子法, 非单调谱梯度算法; 陈述本文主要贡献并列举论文中所使用的符号.

第二章, 对鲁棒人脸识别模型进行校正, 提出求解鲁棒人脸识别的 ℓ_1 - ℓ_1 -范数优化问题的非精确交替方向乘子法. 构造对偶问题模型, 设计求解对偶模型的精确交替方向乘子法. 给出两种算法的收敛性结果, 并使用模拟数据对算法效率进行测试. 试验结果表明: 基于对偶模型的交替方向乘子法具有更高的数值效率.

第三章, 求解基于压缩传感且同时含有脉冲噪音和高斯噪音的稀疏信号恢复的 ℓ_1 - ℓ_1 -范数优化模型. 通过变量代换和变量分裂, 把模型等价转化为凸二次规划模型, 然后使用非单调谱梯度算法进行求解, 给出算法的全局收敛性.

第四章, 总结本文研究成果, 指出有待进一步研究的方向.

关键词: 鲁棒人脸识别; 压缩传感; 交替方向乘子法; 非单调谱梯度算法; 全局收敛性

ABSTRACT

In this thesis, we mainly study the alternating direction method of multipliers with primal and dual form, and non-monotone spectral gradient algorithm to solve ℓ_1 - ℓ_1 norm optimization model which originated from robust face recognition and compressive sensing. We establish the algorithms' convergence, and use the simulated data to verify that the alternating direction method of multipliers is effective.

In chapter one, we simply introduce the minimization models in robust face recognition and compressive sensing. We present some preliminaries which are fundamental in the subsequent development. Furthermore, we provide the alternating direction method of multipliers and non-monotone spectral gradient algorithm for solving convex minimization problems. Finally, we point out the main contribution of this thesis and list of some notations used in the context.

In chapter two, we propose a couple of alternating direction methods of multipliers to solve the ℓ_1 - ℓ_1 minimization problem in face recognition. The algorithms take primal and dual approaches, which lie in inexact and exact versions. We establish the algorithms' convergence results, and report numerical experiments by the simulated data to show the efficiency of the algorithms. The experimental results show that the alternating direction method of multipliers based on the dual model is more effective.

In chapter three, we solve the ℓ_1 - ℓ_1 minimization model to recover sparse signal from limited measurement which contains Gaussian and pulse noises in compressive sensing. Through substituting and splitting the variables, we reformulate the model equivalently as a convex quadratic programming model. Moreover, we solve the model by non-monotone spectral gradient algorithm, and present global convergence of the proposed algorithm.

In chapter four, we conclude the thesis and list some future research topics.

Key words: Robust face recognition, Compressive sensing, Alternating direction method of multipliers, Non-monotone spectral gradient algorithm, Global convergence

目 录

摘要	I
Abstract	III
第一章 绪论	
1.1 人脸识别	1
1.1.1 人脸识别的概念和意义	1
1.1.2 鲁棒性人脸识别的模型	1
1.2 压缩传感	3
1.2.1 压缩传感的概念和意义	3
1.2.2 压缩传感的模型	3
1.3 预备知识	5
1.4 交替方向乘子法	6
1.5 非单调谱梯度算法	8
1.6 本文主要贡献	12
1.7 本文所用符号	13
第二章 基于人脸识别的ℓ_1-ℓ_1范数优化问题的原始对偶交替方向乘子法	
2.1 引言	15
2.2 算法设计	15
2.2.1 交替方向乘子法求解鲁棒性人脸识别问题	15
2.2.2 交替方向乘子法求解鲁棒性人脸识别的对偶问题	18
2.3 数据模拟	20
第三章 基于压缩传感ℓ_1-ℓ_1范数优化模型的非单调谱梯度算法	
3.1 引言	25
3.2 压缩传感模型转换	25

3.3	算法设计	27
3.4	收敛性分析	29
第四章	结论	33
参考文献	35
致 谢	39
附 录	41

第一章 绪论

运筹学是从求解某种实际问题的众多可行方案中寻找最优的方案. 广泛应用于国防, 工程, 军事等各个领域. 随着科学技术的不断进步, 运筹学的理论和算法在各个应用领域发挥着越来越重要的作用. 人脸识别主要是从动态和静态的人脸图像库中找到需要的人脸图像. 随着科学技术的发展, 人脸识别技术在计算机视觉, 神经科学和心理学等各行各业扮演着非常重要的角色. 压缩传感, 又称为压缩采样, 主要是利用信号可压缩和稀疏的特性, 在远小于Nyquist 采样率的条件下, 随机获得离散样本, 用数值算法重建信号.

交替方向乘子算法是增广拉格朗日算法的一种分裂形式, 它充分利用了目标函数可分离的性质, 将原问题转化成多个易于求解的子问题逐个求解. 由于交替方向乘子法迭代形式简单, 内存需求量低等特点, 被广泛用来求解信号处理, 机器学习, 统计分析等领域中出现的可分离凸优化问题. Barzilai-Borwein算法又称为谱梯度算法或两点步长算法, 是求解光滑优化问题的高效算法, 并被广泛用来求解稀疏信号恢复和图像处理等问题. 本论文重点研究非单调谱梯度算法和交替方向乘子法求解 ℓ_1 - ℓ_1 -范数型非光滑凸优化问题, 并分别验证算法在稀疏信号恢复和鲁棒人脸识别领域中的应用.

§1.1 人脸识别

§1.1.1 人脸识别的概念和意义

人脸识别技术于20世纪60年代诞生, 70年代初发展, 90年代被剑桥大学, 麻省理工等著名大学广泛研究^[1]. 人脸识别技术主要是提取的人脸图像的特征数据与数据库中存储的特征模板进行搜索匹配, 则把匹配成功的结果输出, 即将待识别的人脸特征与已得到的人脸特征模板进行比较, 根据相似程度对人脸的身份信息进行判断^[1]. 人脸识别需要积累采集到的大量人脸图像相关的数据, 采用高效的算法不断提高识别准确性. 人脸识别技术已广泛应用于医疗保险, 福利保障, 电子商务, 安全防务及众多企事业单位等领域.

§1.1.2 鲁棒性人脸识别的模型

随着科学技术的发展, 人脸识别技术在各行各业扮演者非常重要的角色, 成为研究活

动的主要焦点. 鲁棒人脸识别是由Wright, Yang, Ganesh等^[2]提出: 给定一个有限的训练样本集, 它的样本按照矩阵逐列存储, 即 $A_i = [v_{i1}, v_{i2}, \dots, v_{in_i}]$, $A_i \in R^{m \times n_i}$, 其中 n_i 是第 i 类训练样本的个数, m 是样本的维数. 设 $b \in R^m$ 是待识别的灰色图像, 若 b 来自于给定第 i 类样本, 则 b 可以近似写成样本的线性组合即

$$b = \alpha_{i1}v_{i1} + \alpha_{i2}v_{i2} + \dots + \alpha_{in_i}v_{in_i}, \quad (1.1)$$

其中 $\alpha_{ij} \in R$ ($j = 1, 2, \dots, n_i$)只有一个元素为1, 其它为零. 若待识别图像 b 的分类未知或在 k 个训练样本库中^[2], 即 $A = [A_1, A_2, A_3, \dots, A_k] \in R^{m \times n}$, 其中 $n = n_1 + n_2 + \dots + n_k$. 这时, b 可以表示成所有样本的线性组合即 $A\bar{x} = b$, 其中 \bar{x} 称为 b 的稀疏表示, 除了与第 i 个有关的元素之外其余为0. 稀疏表示 \bar{x} 可以求解线性方程 $Ax = b$ 得到. 但是由于 $m \ll n$, 即意味着方程组 $Ax = b$ 有无穷多个解.

在鲁棒人脸识别问题中^[2], 类别 k 越大并且每个类别的样本个数越多, 即 n 越大, 则问题的解越稀疏, 就越容易识别. 在此情况下, 可以寻找方程组 $Ax = b$ 的稀疏解:

$$\min_{x \in R^m} \{\|x\|_0 : Ax = b\}, \quad (1.2)$$

其中 $\|x\|_0$ 称作 x 的伪零范数, 即 x 中非零元素的个数. 众所周知, ℓ_0 -范数优化问题是组合问题, 属于NP-难问题. 此外, 若 b 含有脉冲噪音^[3], 即 $b = Ax + \gamma$, 其中 γ 少部分元素非零, 则上述问题等价于

$$\min_{x \in R^m} \|Ax - b\|_0. \quad (1.3)$$

结合问题(1.2)和(1.3)得

$$\min_{x \in R^m} \{\nu\|x\|_0 + \|\gamma\|_0 : Ax + \gamma = b\}, \quad (1.4)$$

其中 $\nu > 0$ 是平衡参数.

研究表明: 若 x 和 γ 是充分稀疏的, 则可以求解如下 ℓ_1 - ℓ_1 凸松弛问题:

$$\min_{x \in R^m} \{\nu\|x\|_1 + \|\gamma\|_1 : Ax + \gamma = b\}. \quad (1.5)$$

此外, 问题(1.5)同样来自于观测的数据含有脉冲噪音的压缩传感领域. 在此情况下, 大量数据表明 ℓ_1 范数正则化和 ℓ_1 数据数据拟合项更能高精度地重构原始信号. 从优化角度看问题(1.5)包含两个非光滑项, 在算法设计方面难度加大.

§1.2 压缩传感

§1.2.1 压缩传感概念和意义

随着科学技术的发展, 人们对信息的需求日益剧增, 传统的压缩传感技术要求采样速率必须达到信号宽带的两倍以上才能重建. 压缩传感利用信号可压缩或稀疏的特性, 在远小于Nyquist 条件下, 随机获得离散样本, 用一些算法进行重建信号. 它一经提出, 在图像处理, 模式识别, 信息论, 天文学等方面受到广泛关注. 压缩传感使用贪婪算法, 优化算法, 组合算法等从有限的测量值中估计或重构原始稀疏信号.

§1.2.2 压缩传感的模型

由于生成稀疏解^[4], ℓ_1 -范数优化问题同样在信号处理, 压缩传感, 机器学习方面等应用领域. 在压缩传感中, 高维稀疏信号首先被编码成一个相对较小测量值, 然后从该测量值中恢复原始信号. 设 $\bar{x} \in R^n$ 是待恢复的稀疏原始信号, $A \in R^{m \times n}$ ($m \ll n$) 是编码矩阵, $b \in R^m$ 是观测值, 满足

$$A\bar{x} = b.$$

一般来说, 原始信号 \bar{x} 可以通过求解欠定的线性方程组 $Ax = b$ 恢复. 但是该线性系统有无穷多个解, 不能确定哪个具体的解是原始信号.

在压缩传感中, 信号 \bar{x} 一般是稀疏的或者在某个简单的正交变换下是稀疏的. 设存在正交变换基 $W \in R^{n \times n}$ 满足 $\tilde{x} := W\bar{x}$, 并且 \tilde{x} 中大部分元素为零, 则其可通过如下模型求解:

$$\min_{x \in R^n} \{ \|x\|_0 : Ax = b \}. \quad (1.6)$$

从压缩传感理论^[4, 5, 6, 7, 8, 26] 可知, 可以通过求解如下凸松弛 ℓ_1 范数最小化问题恢复原始信号:

$$\min_{x \in R^n} \{ \|x\|_1 : Ax = b \}. \quad (1.7)$$

在实际应用领域, 观测值 b 通常含有噪音, 即 $Ax = b$ 只是一个近似估计. 若 b 含有高斯噪音 $\epsilon \in R^m$, 即 $Ax = b + \epsilon$, 则可增加关于 ϵ 的二次惩罚项:

$$\min_{x \in R^n} \left\{ \nu \|x\|_1 + \frac{1}{2} \|\epsilon\|_2^2 : Ax = b + \epsilon \right\}, \quad (1.8)$$

其中参数 $\nu > 0$ 是罚因子. 通过求解 ℓ_1 范数极小化问题(1.7) 和(1.8)来恢复稀疏信号的知名算法有: 基于连续技术的固定点算法FPC^[9], 稀疏重建算法GPSR^[10], 两点步长迭代收

缩算法TwIST^[11], 基于梯度算法包括坐标梯度下降算法^[12], Nesterov 的梯度算法^[13], 交替方向乘子算法YALL1^[3], Barzilai-Borwein 梯度算法^{[14][15]}. 在压缩传感中, 算子“ A ”通常不直接储存为矩阵形式, 它通常是通过随机抽取正交变换矩阵的某些行产生, 这样的方式可加快矩阵向量乘法运算.

当观测到的信号 b 的含有脉冲噪音时, 数据拟合项通常采用 ℓ_1 范数代替 ℓ_2 范数. 假设 $A\bar{x} + \gamma = b$, 其中 $\gamma \in R^m$ 是脉冲噪音, 即向量小部分元素非零. 为了恢复原始信号 \bar{x} , 可同时极小化关于变量求 x 和 γ 的优化问题:

$$\min_{x \in R^n, \gamma \in R^m} \left\{ \nu \|x\|_1 + \|\gamma\|_1 : Ax + \gamma = b \right\}. \quad (1.9)$$

模型(1.9) 包含两个非光滑项 ℓ_1 范数, 则求解难度相对加大. 由Yang & Zhang^[3]知模型(1.9)可等价变形为

$$\min_{\hat{x} \in R^{n+m}} \left\{ \|\hat{x}\|_1, : \hat{A}\hat{x} = \hat{b} \right\},$$

其中

$$\hat{A} = \frac{[A \ \nu I]}{\sqrt{1 + \nu^2}}, \quad \hat{b} = \frac{\nu b}{\sqrt{1 + \nu^2}}, \quad \hat{x} = \begin{pmatrix} \nu x \\ \gamma \end{pmatrix}. \quad (1.10)$$

Yang & Zhang^[3]把问题(1.5)等价转化为等式约束 ℓ_1 -范数优化问题, 然后使用交替方向乘子法求解. Xiao, Zhu & Wu^[16] 提出了求解问题(1.5)的原始和对偶交替方向乘子法.

§1.3 预备知识

设 \mathcal{E} 是 n 维欧氏空间, $\langle \cdot, \cdot \rangle$ 表示内积, $\|\cdot\|$ 表示诱导范数. 设 $D \subseteq \mathcal{E}$, 对任意的 $x \in D$, $y \in D$, $a \in [0, 1]$ 都有 $ax + (1-a)y \in D$ 则称集合 D 为凸集^[17]. 若对任意的 $x \in C$, $y \in C$, $a \in R$ 都有 $ax + (1-a)y \in C$ 则称集合 C 为仿射集. 包含凸集 D 最小的仿射集称作仿射包, 记作 $aff D$. 相对内部就是相对内点的集合, 记作 $ri D = \left\{ x \in aff D \mid \exists \varepsilon > 0, (x + \varepsilon B) \cap (aff D) \subset D \right\}$, 其中 $x + \varepsilon B$ 表示以 x 为圆心, 以 ε 为半径的球^[17]. 设 $A : \mathcal{E}^n \rightarrow \mathcal{E}^n$ 为单调算子, 则 $\forall \vartheta_0 \in Ax_0, \forall \vartheta_1 \in Ax_1$ 满足 $\langle \vartheta_1 - \vartheta_0, x_1 - x_0 \rangle \geq 0$. 若 $\forall (\hat{x}, \hat{\vartheta}) \in (R^n \times R^n) \notin gph A$, 存在 $(\tilde{x}, \tilde{\vartheta}) \in gph A$ 满足

$$\langle \hat{\vartheta} - \tilde{\vartheta}, \hat{x} - \tilde{x} \rangle < 0, \quad (1.11)$$

其中 $gph A := \{(x, \vartheta)\} \in gph A$, 则称 A 为极大单调算子^[18].

若函数 $f : R^n \rightarrow R$ 满足 $f(ax + (1-a)y) \leq af(x) + (1-a)f(y)$, 则称函数 f 为凸函数. 若对任意的 $x \in D$, 都有凸函数 f 满足 $-\infty < f(x) \leq +\infty$, 则称函数 f 为正常凸函数. 设 $f : \mathcal{E} \rightarrow (-\infty, +\infty]$ 是一个正常闭凸函数, x^* 是 f 在 x 点的次梯度即 $f(z) \geq f(x) \langle x^*, z - x \rangle$. f 在 x 点的所有次梯度组成的集合称作 f 在 x 点的次微分, 记作 $\partial f(x)$ ^[17]. 显然 $\partial f(x)$ 是一个非空闭凸集, 凸函数 f 在 $x \in \mathcal{E}$ 的Fenchel共轭函数为

$$f^*(y) = \sup_x \{\langle x, y \rangle - f(x)\} = -\inf_x \{f(x) - \langle x, y \rangle\}, \quad \forall y \in \mathcal{E}. \quad (1.12)$$

众所周知, 若 f 是正常函数^[17], 那么它的共轭函数 $f^*(y)$ 也是正常闭凸函数. 对任何正常闭凸函数 $f : \mathcal{E} \rightarrow (-\infty, +\infty]$, 它在 $x \in \mathcal{E}$ 的Moreau-Yosida 正则化可以写成

$$\varphi_f^\beta(x) := \min_{y \in \mathcal{E}} \left\{ f(y) + \frac{1}{2\beta} \|y - x\|^2 \right\}, \quad (1.13)$$

其中 $\beta > 0$. 对任何 $x \in \mathcal{E}$, 问题(1.13)都有唯一的最优解, 函数 f 在 x 点的临近点记作 $P_f^\beta(x)$ 即

$$P_f^\beta(x) := \arg \min_{y \in \mathcal{E}} \left\{ f(y) + \frac{1}{2\beta} \|y - x\|^2 \right\}.$$

设 $F : R^n \rightarrow R^n$ 是一个向量值函数^[17], 若

$$F(x) = Q^\top f(Qx),$$

则函数 F 是对称的, 其中 Q 是实置换矩阵, 即每一行或每一列中非零元素是1或者-1, 剩下的元素全是零.

§1.4 交替方向乘子法

本节回顾古典的交替方向乘子法和半临近交替方向乘子法^[19]. 设 $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ 是有限维实欧几里得空间, 考虑如下可分离凸优化问题:

$$\begin{aligned} \min_{y,z} \quad & f(y) + g(z) \\ \text{s.t.} \quad & \mathcal{F}^*y + \mathcal{G}^*z = c, \end{aligned} \quad (1.14)$$

其中 $f: \mathcal{Y} \rightarrow (-\infty, +\infty]$ 和 $g: \mathcal{Z} \rightarrow (-\infty, +\infty]$ 都是正常闭凸函数. $\mathcal{F}: \mathcal{X} \rightarrow \mathcal{Y}$ 和 $\mathcal{G}: \mathcal{X} \rightarrow \mathcal{Z}$ 都是线性映射, \mathcal{F}^* 和 \mathcal{G}^* 表示伴随算子, $c \in \mathcal{X}$, 易知, ∂f 和 ∂g 是极大单调算子. 存在半正定线性算子 $\Sigma_f: \mathcal{Y} \rightarrow \mathcal{Y}$ 和 $\Sigma_g: \mathcal{Z} \rightarrow \mathcal{Z}$, 对于 $y, y' \in \mathcal{Y}, z, z' \in \mathcal{Z}, u \in \partial f(y), u' \in \partial f(y'), v \in \partial g(z), v' \in \partial g(z')$, 使得如下不等式成立:

$$\langle u - u', y - y' \rangle \geq \|y - y'\|_{\Sigma_f}^2, \quad \langle v - v', z - z' \rangle \geq \|z - z'\|_{\Sigma_g}^2. \quad (1.15)$$

模型(1.14)的拉格朗日函数为

$$\mathcal{L}(y, z; x) := f(y) + g(z) - \langle x, \mathcal{F}^*y + \mathcal{G}^*z - c \rangle, \quad \forall (x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z},$$

其中 x 是拉格朗日乘子. 模型(1.14)的对偶问题如下:

$$\begin{aligned} \max_x \inf_{y,z} \mathcal{L}(y, z; x) &= \max_x \left\{ \inf_y \{f(y) - \langle x, \mathcal{F}^*y \rangle\} + \inf_z \{g(z) - \langle x, \mathcal{G}^*z \rangle\} + \langle c, x \rangle \right\} \\ &= \max_x \left\{ \inf_y \{f(y) - \langle \mathcal{F}(x), y \rangle\} + \inf_z \{g(z) - \langle \mathcal{G}(x), z \rangle\} \right. \\ &\quad \left. + \langle c, x \rangle \right\}. \end{aligned} \quad (1.16)$$

由共轭函数定义得

$$\inf_y \{f(y) - \langle \mathcal{F}(x), y \rangle\} = -f^*(\mathcal{F}(x)), \quad (1.17)$$

和

$$\inf_z \{g(z) - \langle \mathcal{G}(x), z \rangle\} = -g^*(\mathcal{G}(x)), \quad (1.18)$$

则对偶问题可转化为

$$\max_x \left\{ \langle c, x \rangle - f^*(\mathcal{F}(x)) - g^*(\mathcal{G}(x)) \right\} \quad (1.19)$$

设 $(\bar{x}, \bar{y}, \bar{z}) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ 是Karush-Kuhn-Tucker(KKT)点, 即

$$\mathcal{F}\bar{x} \in \partial f(\bar{y}), \quad \mathcal{G}\bar{x} \in \partial g(\bar{z}), \quad \mathcal{F}^*\bar{y} + \mathcal{G}^*\bar{z} = c.$$

模型(1.14)的增广拉格朗日函数为

$$\mathcal{L}_\sigma(y, z; x) := f(y) + g(z) + \langle x, \mathcal{F}^*y + \mathcal{G}^*z - c \rangle + \frac{\sigma}{2} \|\mathcal{F}^*y + \mathcal{G}^*z - c\|^2,$$

其中罚参数 $\sigma > 0$. 取初始点 $(x^0, y^0, z^0) \in \mathcal{X} \times (\text{dom } f) \times (\text{dom } g)$, 其中 dom 表示定义域. 古典交替方向乘子法迭代格式如下:

$$\begin{cases} y^{k+1} : &= \arg \min_y \mathcal{L}_\sigma(y, z^k; x^k), \\ z^{k+1} : &= \arg \min_z \mathcal{L}_\sigma(y^{k+1}, z; x^k), \\ x^{k+1} : &= x^k - \tau\sigma(\mathcal{F}^*y^{k+1} + \mathcal{G}^*z^{k+1} - c), \end{cases} \quad (1.20)$$

其中 $\tau \in (0, (1 + \sqrt{5})/2)$. 交替方向乘子法见Glowinski & Marroco^[20], Gabay & Mercier^[21]. 最早收敛性证明见Gabay & Mercier^[21], Glowinski^[22]; Fortin & Glowinski^[23]. 交替方向乘子法的发展详细介绍见[24].

为了计算方便, Eckstein^[25]给子问题都增加临近点项:

$$\begin{cases} y^{k+1} : &= \arg \min_y \mathcal{L}_\sigma(y, z^k; x^k) + \frac{\sigma}{2} \|y - y^k\|_{\mathcal{T}_f}^2, \\ z^{k+1} : &= \arg \min_z \mathcal{L}_\sigma(y^{k+1}, z; x^k) + \frac{\sigma}{2} \|z - z^k\|_{\mathcal{T}_g}^2, \\ x^{k+1} : &= x^k - \tau\sigma(\mathcal{F}^*y^{k+1} + \mathcal{G}^*z^{k+1} - c), \end{cases} \quad (1.21)$$

其中 $\mathcal{T}_f \succeq 0$, $\mathcal{T}_g \succeq 0$ 即为半正定矩阵. 增加临界点保证子问题的解都存在. 关于半临近交替方向乘子法的收敛性定理如下: 它的证明见文献[19, Theorem B.1].

假设 1.4.1 存在 $(\bar{y}, \bar{z}) \in \text{ri}(\text{dom}(f) \times \text{dom}(g))$ 满足 $\mathcal{F}^*\bar{u} + \mathcal{G}^*\bar{v} = c$, 其中 ri 表示相对内部, dom 表示定义域.

定理 1.4.2 [19, 定理B.1] 设 Σ_f 和 Σ_g 都是半正定算子. 假设(1.14)的解存在和假设1.4.1成立, 序列 $\{(y^k, z^k, x^k)\}$ 是由(1.21)产生. 初始点 $(y^0, z^0, x^0) \in \text{dom}(f) \times \text{dom}(g) \times \mathcal{X}$, $(y^0, z^0, x^0) \in \text{dom}(f) \times \text{dom}(g) \times \mathcal{X}$, 则有:

(1) 若 $(y^\infty, z^\infty, x^\infty)$ 是 $\{(y^k, z^k, x^k)\}$ 的聚点, 则 (y^∞, z^∞) 是问题(1.14)的解, x^∞ 是问题(1.19)的解.

(2) 若 $\sigma^{-1}\Sigma_f + \mathcal{T}_f + \mathcal{F}\mathcal{F}^*$ 和 $\sigma^{-1}\Sigma_g + \mathcal{T}_g + \mathcal{G}\mathcal{G}^*$ 正定, 算法产生的序列 $\{(y^k, z^k, x^k)\}$ 收敛到 $(y^\infty, z^\infty, x^\infty)$, 则 (y^∞, z^∞) 是问题(1.14)的解, x^∞ 是模型(1.14)对偶问题的解.

(3) 仅有 y 或者 z 项时, 那么结论(1)-(2)在 $\tau \in (0, 2)$ 时成立.

§1.5 非单调谱梯度算法

无约束优化问题为

$$\min_{x \in R^n} f(x), \quad (1.22)$$

其中 f 是 $R^n \rightarrow R$ 上连续可微函数. 若给出可行解 $x^* \in R^n$, 存在 $\varepsilon > 0$, 当 $x \neq x^*$ 时总有 $f(x) > f(x^*)$, ($x \in R^n \cap B(x^*, \varepsilon)$)成立, 则称 x^* 是问题(1.22)的局部最优解. 若对于任意的 $\varepsilon > 0$, 都有 $f(x) > f(x^*)$, $x \in R^n$, 则 x^* 为问题(1.22)的全局最优解^[28].

设 $x, d \in R^n$, 若存在数 $\bar{\alpha} > 0$, 使得

$$f(x + \alpha d) < f(x) \quad \forall \alpha \in (0, \bar{\alpha}), \quad (1.23)$$

则称 d 是函数 f 在点 x 处的一个下降方向. 从几何上下降方向 d 可解释为: 当点从 x 处出发, 沿方向 d 移动时, 函数 f 的值的变化呈单调递减趋势. 若 $\phi(\alpha) = f(x + \alpha d)$, 则方向 d 是 f 在 x 处的下降方向等价于一元函数 ϕ 在 $\alpha = 0$ 处单调递减^[28]. 求解无约束问题(1.22)的基本思想^[28]是从某个初始点 x_0 出发, 按照使目标函数值下降的原则构造点列 $\{x_k\}$, 即点列 $\{x_k\}$ 满足条件 $f(x_{k+1}) < f(x_k)$, ($k = 0, 1, \dots$). 算法的最终目标是使得点列 $\{x_k\}$ 中的某个点或某个极限点是问题(1.22)的解或稳定点^[28].

设 d_k 是 f 在 x_k 处得一个下降方向且满足

$$\nabla f(x_k)^\top d_k < 0. \quad (1.24)$$

从而, 当 α 充分小时, $f(x_k + \alpha_k d_k) < f(x_k)$. 取 $x_{k+1} = x_k + \alpha_k d_k$, 其中 $\alpha_k > 0$ 使得

$$f(x_k + \alpha_k d_k) < f(x_k). \quad (1.25)$$

上述不等式(1.25)中的数 α_k 称为步长. 确定步长的方法有很多种^[28], 最常用的线性搜索有两种, 即精确线性搜索和非精确线性搜索. 精确线性搜索要求步长 α_k 使一元函数 $\min_{\alpha > 0} f(x_k + \alpha d_k)$ 取到最小值点. 非精确线性搜索要求步长在下一个迭代点比当前迭代点有一定的下降量即可. 非精确线性搜索主要有Armijo型线性搜索, Wolfe-Powell型线性搜索, Goldstein型线性搜索和强Wolfe型线性搜索^[28].

精确线性搜索和非精确线性搜索的一个共同点是: 步长满足 $f(x_k + \alpha_k d_k) < f(x_k)$ 即算法产生的序列 $\{f(x_k)\}$ 是单调递减的. 单调性线性搜索的缺点是有时获得的步长会很小, 尤其当算法产生的方向 d_k 与负梯度方向 $-\nabla f(x_k)$ 接近垂直时这种现象更加明显. 为了克

服这种缺陷, 可以采取非单调线性搜索^[28]

$$f(x_k + ad_k) \leq \max_{0 \leq i \leq M} f(x_{k-i}) + \sigma_1 a_k \nabla f(x_k)^\top d_k, \quad (1.26)$$

其中 $M \geq 0$ 为整数, $k = 0, 1, 2, \dots, M$. 当 $M = 0$ 时, 非单调线性搜索和 Armijo 线性搜索等价.

由下降方向满足的条件知, 负梯度方向 $-\nabla f(x_k)$ 是函数 f 在 x_k 处的下降方向^[28], 则相应的下降算法就是最速下降算法. 众所周知, 最速下降算法比较简单, 但是容易产生锯齿现象, 收敛速率下降.

设 f 二次连续可微, 对任意 $x \in R^n$, $\nabla^2 f(x)$ 正定. 考虑 $f(x)$ 在 x_k 处的二次近似式

$$f(x_k) + \nabla f(x_k)^\top s + \frac{1}{2} s^\top \nabla^2 f(x_k) s \approx f(x_k + s), \quad (1.27)$$

的最小值点, 或等价地, d_k 下面关于 d 的线性方程组的解:^[28]

$$\nabla^2 f(x_k) d + \nabla f(x_k) = 0. \quad (1.28)$$

令 $d_k = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$ 得到的算法是 Newton 法^[28]. Newton 法具有二次终止性, 但是 $\nabla^2 f(x_k)$ 不正定时, 那么得到的方向 d_k 是不能保证函数值下降; 若 $\nabla^2 f(x_k)$ 奇异时, 方向 d_k 可能不存在.

为了克服 Newton 法这一缺陷, 可采用修正 Newton 法. 基本思想是用矩阵 $A_k = \nabla^2 f(x_k) + \nu_k I$ 代替 $\nabla^2 f(x_k)$, 其中 $I \in R^{n \times n}$ 是单位矩阵, 参数 ν_k 使得矩阵 A_k 正定. 克服 Newton 法这种缺陷的另一种方法是结合 Newton 法和最速下降法构造 Newton-最速下降混合型算法. 该算法的基本思想是: 当 Newton 方向不存在或 Newton 方向存在但不是 f 在 x_k 处得下降方向时, 采用最速下降方向取代 Newton 方向.

但是在修正 Newton 法中, 参数 ν_k 的选取十分重要^[28]. 若参数 ν_k 过小, 则相应的修正 Newton 方向仍不能保证是 f 在 x_k 处的下降方向. 若参数 ν_k 过大, 则会影响收敛速度. 除此之外, Newton 法及其修正形式都需要计算函数 f 的二阶导数, 拟 Newton 法可以克服 Newton 法的上述缺陷.

拟 Newton 法的基本思想是在 Newton 法的子问题中用 $\nabla^2 f(x_k)$ 的某个近似矩阵 B_k 取代 $\nabla^2 f(x_k)$. 矩阵 B_k 具有如下特征: 在某种意义下, $B_k \approx \nabla^2 f(x_k)$, 使相应算法产生的方向是 Newton 方向的近似^[28], 以保证算法具有较快的收敛速度: B_k 对称正定, 从而使得算

法产生的方向是函数 f 在 x_k 的下降方向; 且矩阵 B_k 容易计算. 由于 f 二次连续可微, 利用多元函数Taylor展开式得如下近似式:

$$\nabla f(x_k) \approx \nabla f(x_{k+1}) - \nabla^2 f(x_{k+1})(x_{k+1} - x_k)$$

用 B_{k+1} 取代 $\nabla^2 f(x_{k+1})$ 时, 上面的近似式成立等式. 令 $s_k = x_{k+1} - x_k$, $y_{k+1} = \nabla f(x_{k+1}) - \nabla f(x_k)$, 可得拟牛顿方程为 $B_{k+1}s_k = y_k$.

由于拟Newton算法每次迭代都要计算一个Hessian矩阵的逆^[28], 不太适合处理大规模问题, 这时要考虑1988年Barzilai和Borwein^[29]的Barzilai-Borwein 梯度(谱梯度)算法. 令 $B_{k+1} = \alpha_{k+1}I$, 可得

$$\alpha_{k+1}s_k = y_k. \quad (1.29)$$

使得 $\alpha_{k+1}^{(1)} = \frac{s_k^\top y_k}{\|s_k\|^2}$ 或者 $\alpha_{k+1}^{(2)} = \frac{\|y_k\|^2}{s_k^\top y_k}$.

设 $\Omega \subset R^n$ 是一个闭凸集, $\Omega = \{x \in R^n : l \leq x \leq u\}$. Birgin, Martínez 和Raydan^[34]提出求解如下的有限约束最小化问题非单调Barzilai-Borwein (谱)梯度算法:

$$\min f(x) \quad \text{s.t. } x \in \Omega. \quad (1.30)$$

非单调谱梯度算法的算法^[34] 步骤如下:

算法 1.5.1 (非单调谱梯度算法)

步0. 设 $x^0 \in \Omega$, 常数 $M \geq 1$, 给定参数 $\alpha_{max} > \alpha_{min} > 0$, 常数 $\gamma \in (0, 1)$, $0 < \sigma_1 < \sigma_2 < 1$, $\alpha_k = 1$, $c > 0$. 给定初始点 $\alpha_0 \in [\alpha_{min}, \alpha_{max}]$, 设 $k = 0$.

步1. 若 $\|\mathcal{P}_\Omega(x^k - g(x^k)) - x^k\| = 0$, 其中 $g(\cdot)$ 表示梯度, 其中 $\mathcal{P}_\Omega(x^k - g(x^k))$ 表示 $x^k - g(x^k)$ 到凸集 Ω 上的度量投影, 是 $\min_{x \in \Omega} \|x^k - g(x^k) - x\|_2^2$ 的最优解. 停止; 输出 x_k .

步2. 非单调线性搜索

步2.1. 设 $\lambda = \alpha_k$;

步2.2. 令 $x_+ = \mathcal{P}_\Omega(x^k - \lambda g(x^k))$;

步2.3. 若

$$f(x^+) \leq \max_{0 \leq j \leq \min\{k, M-1\}} f(x^{k-j}) + \gamma \langle \nabla f(x^k), (x^+ - x^k) \rangle, \quad (1.31)$$

令 $\lambda_k = \lambda$, $x^{k+1} = x^+$, $s^k = x^{k+1} - x^k$, $y^k = \nabla f(x^{k+1}) - \nabla f(x^k)$ 转步3. 若(1.31) 不成立, 定义

$$\lambda_{new} \in [\sigma_1 \lambda, \sigma_2 \lambda], \quad (1.32)$$

令 $\lambda = \lambda_{new}$, 转步2.2.

步3. 计算 $n_k = \langle s_k, y_k \rangle$. 若 $n_k \leq 0$, 令 $\alpha^{k+1} = \alpha_{\max}$. 否则, 计算 $m_k = \langle s_k, s_k \rangle$ 和令

$$\alpha^{k+1} = \min\{\alpha_{\max}, \max\{\alpha_{\min}, \frac{m_k}{n_k}\}\}.$$

步4. 令 $k = k + 1$. 转步1.

关于非单调Barzilai-Borwein 梯度算法收敛性见文献[34, Theorems 2.1-2.2]

定理 1.5.2 [34, Theorems 2.1-2.2] 设序列 $\{x_k\}$ 是由算法1.5.1产生, 参数 α_k 由(1.29)产生, 则序列 $\{x_k\}$ 收敛到问题(1.30)的稳定点.

§1.6 本文主要贡献

求解基于人脸识别的 ℓ_1 - ℓ_1 范数优化问题的原始对偶交替方向乘子法和基于压缩传感 ℓ_1 - ℓ_1 范数优化模型的非单调谱梯度算法, 每步迭代, 给出算法的收敛性结果, 用数据模拟验证交替方向乘子法求解鲁棒人脸识别模型和对偶模型的数值有效性. 交替方向乘子法求解鲁棒人脸识别对偶模型于从时间上要远远优于求解鲁棒人脸识别模型.

第二章的主要内容是校正鲁棒人脸识别模型, 设计算法, 用非精确交替方向乘子法求解鲁棒人脸识别的 ℓ_1 - ℓ_1 -范数优化问题, 给出算法的收敛性结果; 构造鲁棒人脸识别的对偶模型, 设计算法, 用精确的交替方向乘子法求解人脸识别问题的对偶模型, 给出算法的收敛性结果; 并且用模拟数据对算法进行测试. 数值结果表明交替方向乘子法求解鲁棒人脸识别对偶模型具有更好的数值效率.

第三章的主要内容是对同时含有脉冲和高斯噪音的稀疏信号恢复的 ℓ_1 - ℓ_1 -范数优化模型进行校正, 先通过变量代换和变量分裂把 ℓ_1 - ℓ_1 -范数优化模型转化成凸二次规划模型, 然后用非单调谱梯度算法进行求解. 最后对算法进行收敛分析.

§1.7 本文所用符号

x :	实向量
A :	实矩阵 A
B :	实矩阵 B
v_{i,n_i} :	矩阵 v 的第 i 行第 n_i 列的元素
$f(x)$:	目标函数
$\partial f(x)$:	$f(x)$ 的次微分
n :	函数的维数
R :	全体实数组成的集合
R^n :	全体 n 维实向量组成的集合
A^\top :	矩阵 A 的转置
I :	单位矩阵
\forall :	任意的
$\ \cdot\ $:	向量的欧氏范数
$\ \cdot\ _1$:	矩阵的1范数
$\ \cdot\ _2$:	矩阵的2范数
$\langle \cdot, \cdot \rangle$:	向量或矩阵的内积

第二章 基于人脸识别的 ℓ_1 - ℓ_1 范数优化问题的原始对偶两种交替方向乘子法

§2.1 引言

本节是在传统的 ℓ_1 - ℓ_1 范数模型中添加自适应校正项, 对鲁棒人脸识别模型进行校正, 提出求解鲁棒人脸识别的 ℓ_1 - ℓ_1 范数优化问题的非精确交替方向乘子法; 然后构造对偶问题模型, 设计求解对偶模型的精确交替方向乘子法.

§2.2 算法设计

本节考虑带有校正项的 ℓ_1 - ℓ_1 模型^[43]

$$\min_{x \in R^n, \gamma \in R^m} \{ \nu \|x\|_1 + \|\gamma\|_1 + \nu \langle F(\tilde{x}), x \rangle : Ax + \gamma = b \}, \quad (2.1)$$

其中 $F: R^n \rightarrow R^n$, $F: (F_1(x), F_2(x), \dots, F_n(x))$ 为对称函数, 其定义为

$$F_i(x) = \begin{cases} \phi\left(\frac{x_i}{\|x\|_\infty}\right), & x \in R^n \setminus \{0\}, \\ 0, & x = 0. \end{cases}$$

其中函数 $\phi: R^n \rightarrow R$ 为

$$\phi(t) = \text{sgn}(t)(1 + \omega^\alpha) \frac{|t|^\alpha}{\|t\|^\alpha + \omega^\alpha}, \quad t \in R,$$

其中参数 $\omega > 0$, $\alpha > 0$.

§2.2.1 交替方向乘子法求解鲁棒人脸识别模型

模型(2.1)的增广拉格朗日函数为

$$\mathcal{L}_{\mathcal{A}}(x, \gamma, \lambda) = \nu \|x\|_1 + \nu \langle F(\tilde{x}), x \rangle + \|\gamma\|_1 - \langle \lambda, Ax + \gamma - b \rangle + \frac{\sigma}{2} \|Ax + \gamma - b\|_2^2, \quad (2.2)$$

其中 $\lambda \in R^m$ 是拉格朗日乘子, 罚参数 $\sigma > 0$. 设当前点是 (x^k, λ^k) , 则 γ^{k+1} 由如下公式产生

$$\begin{aligned}\gamma^{k+1} &= \arg \min_{\gamma \in R^m} \left\{ \|\gamma\|_1 - \langle \lambda^k, Ax^k + \gamma - b \rangle + \frac{\sigma}{2} \|Ax^k + \gamma - b\|_2^2 \right\} \\ &= \arg \min_{\gamma \in R^m} \left\{ \|\gamma\|_1 + \frac{\sigma}{2} \|\gamma + Ax^k - b - \lambda^k/\sigma\|_2^2 \right\} \\ &= P_{\|x\|_1}^{1/\sigma} \left(b - Ax^k + \lambda^k/\sigma \right) \\ &= \max \left\{ |b - Ax^k + \lambda^k/\sigma| - \frac{1}{\sigma}, 0 \right\} \frac{b - Ax^k + \lambda^k/\sigma}{|b - Ax^k + \lambda^k/\sigma|},\end{aligned}$$

其中 $|\cdot|$ 表示绝对值, “max” 表示最大值且定义 $0 \cdot 0/0 = 0$. 点 x^{k+1} 计算如下:

$$\begin{aligned}x^{k+1} &= \arg \min_{x \in R^n} \left\{ \nu \|x\|_1 + \nu \langle F(\tilde{x}), x \rangle - \langle \lambda^k, Ax + \gamma^{k+1} - b \rangle + \frac{\sigma}{2} \|Ax + \gamma^{k+1} - b\|_2^2 \right\} \\ &= \arg \min_{x \in R^n} \left\{ \nu \|x\|_1 + \nu \langle F(\tilde{x}), x \rangle + \frac{\sigma}{2} \|Ax + (\gamma^{k+1} - b - \lambda^k/\sigma)\|_2^2 \right\}.\end{aligned}\quad (2.3)$$

为了准确的求解(2.3), 增加临界点项:

$$x^{k+1} = \arg \min_{x \in R^n} \left\{ \nu \|x\|_1 + \nu \langle F(\tilde{x}), x \rangle + \frac{\sigma}{2} \|Ax + (\gamma^{k+1} - b - \lambda^k/\sigma)\|_2^2 + \frac{\sigma}{2} \|x - x_k\|_{\mathcal{S}}^2 \right\},$$

其中 $\mathcal{S} = \frac{1}{\rho}I - A^\top A$. 易知, 当 $\rho \leq \lambda_{\max}(A^\top A)$ 时, \mathcal{S} 半正定. 其中 $\lambda_{\max}(\cdot)$ 表示矩阵最大特征值. 设 $g^k = A^\top(Ax^k + \gamma^{k+1} - b - \lambda^k/\sigma)$, 则

$$\begin{aligned}x_{k+1} &= \arg \min_{x \in R^n} \left\{ \nu \|x\|_1 + \nu \langle F(\tilde{x}), x \rangle + \frac{\sigma}{2} \|Ax + (\gamma^{k+1} - b - \lambda^k/\sigma)\|_2^2 + \frac{\sigma}{2} \|x - x_k\|_{\mathcal{S}}^2 \right\} \\ &= \arg \min_{x \in R^n} \left\{ \nu \|x\|_1 + \frac{\sigma}{2\rho} \left\| x - \left(x_k - \rho g_k - \frac{\rho\nu}{\sigma} F(\tilde{x}) \right) \right\|_2^2 \right\} \\ &= P_{\|x\|_1}^{\rho\nu/\sigma} \left(x_k - \rho g_k - \frac{\rho\nu}{\sigma} F(\tilde{x}) \right) \\ &= \max \left\{ \left| x_k - \rho g_k - \frac{\rho\nu}{\sigma} F(\tilde{x}) \right| - \frac{\rho\nu}{\sigma}, 0 \right\} \frac{x_k - \rho g_k - \frac{\rho\nu}{\sigma} F(\tilde{x})}{\left| x_k - \rho g_k - \frac{\rho\nu}{\sigma} F(\tilde{x}) \right|}.\end{aligned}\quad (2.4)$$

令 $\gamma = \gamma^{k+1}$, $x = x^{k+1}$, 则拉格朗日乘子更新为

$$\lambda^{k+1} = \lambda^k - \tau\sigma(Ax^{k+1} + \gamma^{k+1} - b).\quad (2.5)$$

依据以上推导, 算法过程如下:

算法 2.2.1 (PADMM算法) 给定参数 $\sigma > 0$, $\nu > 0$, $\tau \in (0, (1 + \sqrt{5})/2)$, $\rho \leq \lambda_{\max}(A^\top A)$, 初始点 x^0 , λ^0 , 令 $k := 0$.

- 步1. 计算 $\gamma^{k+1} = \max \left\{ \left| b - Ax^k + \lambda^k / \sigma \right| - \frac{1}{\sigma}, 0 \right\} \frac{b - Ax^k + \lambda^k / \sigma}{\left| b - Ax^k + \lambda^k / \sigma \right|};$
- 步2. 计算 $x_{k+1} = \max \left\{ \left| x_k - \rho g_k - \frac{\rho \nu}{\sigma} F(\tilde{x}) \right| - \frac{\rho \nu}{\sigma}, 0 \right\} \frac{x_k - \rho g_k - \frac{\rho \nu}{\sigma} F(\tilde{x})}{\left| x_k - \rho g_k - \frac{\rho \nu}{\sigma} F(\tilde{x}) \right|}, g^k = A^\top (Ax^k + \gamma^{k+1} - b - \lambda^k / \sigma);$
- 步3. 计算 $\lambda^{k+1} = \lambda^k - \tau \sigma (Ax^{k+1} + \gamma^{k+1} - b);$
- 步4. 若序列 (γ^k, x^k) 满足某种收敛准则, 停止; 否则, 令 $k = k + 1$ 转步1.

由第一章定理1.4.2和假设1.4.1, 易知算法PADMM收敛性定理转化为如下定理.

定理 2.2.2 设 $\sigma > 0$, 序列 $\{(\gamma^k, x^k)\}$ 是由(PADMM算法)产生. 若 $\rho \leq \lambda_{\max}(A^\top A)$, 则序列 $\{(\gamma^k, x^k)\}$ 收敛到问题(2.1) 的解.

§2.2.2 交替方向法乘子法求解鲁棒人脸识别对偶模型

本节主要是用交替方向乘子法 (ADMM) 求解(2.1)的对偶模型. 问题(2.1) 的拉格朗日函数为

$$\mathcal{L}(x, \gamma; \lambda) = \nu \|x\|_1 + \|\gamma\|_1 + \nu \langle F(\tilde{x}), x \rangle - \langle \lambda, Ax + \gamma - b \rangle, \quad (2.6)$$

其中 $\lambda \in R^m$ 是拉格朗日乘子. 问题(2.1)的对偶问题为

$$\begin{aligned} \max_{\lambda \in R^m} \inf_{x \in R^n, \gamma \in R^m} \mathcal{L}(x, \gamma; \lambda) &= \max_{\lambda \in R^m} \inf_{x \in R^n, \gamma \in R^m} \left\{ \nu \|x\|_1 + \|\gamma\|_1 + \nu \langle F(\tilde{x}), x \rangle - \langle \lambda, Ax + \gamma - b \rangle \right\} \\ &= \max_{\lambda \in R^m} \left\{ \inf_{\gamma \in R^m} \{ \|\gamma\|_1 - \langle \lambda, \gamma \rangle \} \right. \\ &\quad \left. + \inf_{x \in R^n} \{ \nu \|x\|_1 + \nu \langle F(\tilde{x}), x \rangle - \langle \lambda, Ax \rangle \} + \langle y, b \rangle \right\} \\ &= \max_{\lambda \in R^m} \left\{ \langle \lambda, b \rangle : \|\lambda\|_\infty \leq 1, \|\nu F(\tilde{x}) - A^\top \lambda\|_\infty \leq \nu \right\}. \end{aligned} \quad (2.7)$$

引入变量 $z \in R^n$, 则(2.7) 可以写成

$$\max_{\lambda \in R^m, z \in R^n} \left\{ \langle \lambda, b \rangle : z = \nu F(\tilde{x}) - A^\top \lambda, \|\lambda\|_\infty \leq 1, \|z\|_\infty \leq \nu \right\},$$

或

$$\min_{\lambda \in R^m, z \in R^n} \left\{ -\langle \lambda, b \rangle + \delta_{\mathcal{B}_1^\infty}(\lambda) + \delta_{\mathcal{B}_\nu^\infty}(z) : z = \nu F(\tilde{x}) - A^\top \lambda \right\}, \quad (2.8)$$

其中 $\mathcal{B}_1^\infty = \{\lambda \in R^m : \|\lambda\|_\infty \leq 1\}$, $\mathcal{B}_\nu^\infty = \{z \in R^n : \|z\|_\infty \leq \nu\}$, $\delta_{\mathcal{B}_1^\infty}(\lambda)$ 表示 λ 在 \mathcal{B}_1^∞ 上的示性函数.

定义问题(2.8)的增广拉格朗日函数为

$$\mathcal{L}_{\mathcal{A}}(z, \lambda; x) = \begin{cases} -\langle \lambda, b \rangle + \langle x, \nu F(\tilde{x}) - A^\top \lambda - z \rangle + \frac{\sigma}{2} \|\nu F(\tilde{x}) - A^\top \lambda - z\|_2^2, & \text{若 } \lambda \in \mathcal{B}_1^\infty, z \in \mathcal{B}_\nu^\infty, \\ +\infty, & \text{否则,} \end{cases}$$

其中罚参数 $\sigma > 0$. 交替方向乘子法的迭代格式如下:

$$\begin{cases} z^{k+1} = \arg \min_z \mathcal{L}_{\mathcal{A}}(z, \lambda^k; x^k), \\ \lambda^{k+1} = \arg \min_\lambda \mathcal{L}_{\mathcal{A}}(z^{k+1}, \lambda; x^k), \\ x^{k+1} = x^k - \tau \sigma (\nu F(\tilde{x}) - A^\top \lambda^{k+1} - z^{k+1}), \end{cases}$$

其中步长为 $\tau \in (0, (1 + \sqrt{5})/2)$.

首先令 $\lambda = \lambda^k$, $x = x^k$, 则 z 的子问题将由 $\mathcal{L}_{\mathcal{A}}$ 的梯度等于零可得

$$\nabla_z \mathcal{L}_{\mathcal{A}}(z, \lambda^k; x^k) = -x^k - \sigma (\nu F(\tilde{x}) - A^\top \lambda^k - z) = 0.$$

令 $z \in \mathcal{B}_\nu^\infty$ 可得

$$z^{k+1} = \mathcal{P}_{\mathcal{B}_\nu^\infty}(\nu F(\tilde{x}) - A^\top \lambda^k + x^k / \sigma),$$

其中 $\mathcal{P}_{\mathcal{B}_\nu^\infty}(\cdot)$ 表示闭凸集 \mathcal{B}_ν^∞ 上的投影. 计算关于 λ 的子问题如下:

$$\nabla_\lambda \mathcal{L}_A(z^{k+1}, \lambda; x^k) = -b - Ax^k - \sigma A(\nu F(\tilde{x}) - A^\top \lambda - z^{k+1}) = 0.$$

令 $\lambda \in \mathcal{B}_1^\infty$, 得

$$\lambda^{k+1} = \mathcal{P}_{\mathcal{B}_1^\infty}((\sigma AA^\top)^{-1}[b + Ax^k + \sigma A(\nu F(\tilde{x}) - z^{k+1})]).$$

算法 2.2.3 (*DADMM算法*) 给定参数 $\sigma > 0$, $\nu > 0$, $\tau \in (0, (1 + \sqrt{5})/2)$, 初始点 x^0 , λ^0 , 令 $k := 0$.

步1. 计算 $z^{k+1} = \mathcal{P}_{\mathcal{B}_\nu^\infty}(\nu F(\tilde{x}) - A^\top \lambda^k + x^k / \sigma)$;

步2. 计算 $\lambda^{k+1} = \mathcal{P}_{\mathcal{B}_1^\infty}((\sigma AA^\top)^{-1}[b + Ax^k + \sigma A(\nu F(\tilde{x}) - z^{k+1})])$;

步3. 计算 $x^{k+1} = x^k - \tau \sigma (\nu F(\tilde{x}) - A^\top \lambda^{k+1} - z^{k+1})$;

步4. 若序列 (z^k, λ^k) 满足某种收敛准则, 停止; 否则, 令 $k = k + 1$ 转**步1**.

由第一章定理1.4.2和假设1.4.1, 易知(DADMM算法)收敛性定理转化为如下定理.

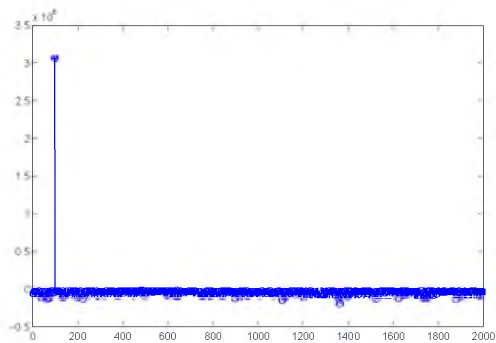
定理 2.2.4 令 $\sigma > 0$, 给定初始点 (z^0, λ^0) , 由(DADMM算法)产生的序列 $\{(z^k, \lambda^k)\}$ 收敛到问题(2.8)的解.

§2.3 数据模拟

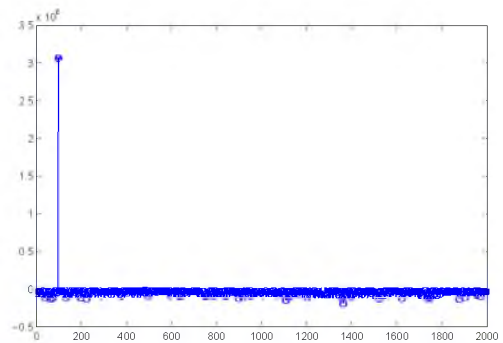
本节主要对交替方向乘子算法求解人脸识别模型和对偶模型进行数据模拟. 所有程序代码在Matlab R2014a进行编写, 在个人计算机上进行测试. 计算机的配置是Intel(R) Core(TM) i3-4000M CPU 2.40GHz, 4GB内存, Windows 8操作系统. 模拟的是400个人的 人脸图像, 其中每个人脸图像复制5遍添加噪音所得的数据库. 其中 $n = 500 \times 4 = 2000$, $m = 300$. b 是从这2000张人脸图像数据库中取出的一张人脸图像, 现在要用数据模拟这张人脸图像就是数据库中的图像, 用图像的形式显示出来.

主要用数据模拟方法说明交替方向乘子法求解人脸识别模型和对偶模型具有可行性和有效性. 设置的数值参数有: 精度 $tol = 1 \times 10^{-5}$, 最大迭代次数设置为500次, $\tau = 1.0$, $\nu = 2.0$, $\mu = 2.0$, $\sigma_{max} = 1$, $\sigma = 1$, $\rho = 0.05(\lambda_{max} A^T A)$, 当 $|x_{k+1} - x_k| \leq tol$ 时, 算法终止.

本实验模拟了数据库中10组值, 即当取数据库中的第100, 200, 400, 600, 800, 1000, 1200, 1400, 1500, 1900张人脸图像时, 分别比较交替方向乘子法求解鲁棒人脸识别模型与对偶模型的图像和时间. 以下图(a) 表示由交替方向乘子法求解鲁棒人脸识别模型(PADMM), 图(b)表示交替方向乘子法求解鲁棒人脸识别对偶模型(DADMM), 横轴代表数据库中图像的位置; 纵轴表示 x 即为图像的稀疏表示, 除了与该图像有关的元素之外全部趋于0.

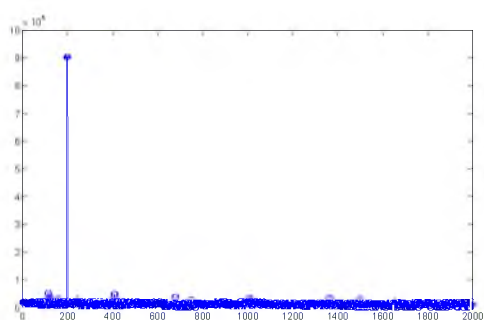


(a) PADMM算法数值结果

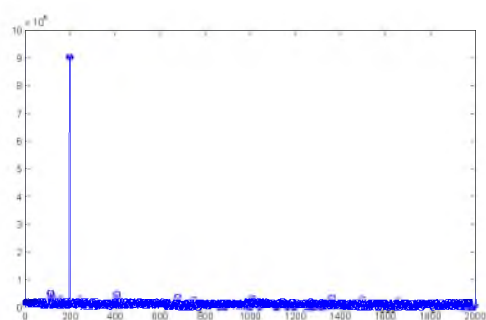


(b) DADMM算法数值结果

图 2.1 模拟的是数据库中第100张人脸图像

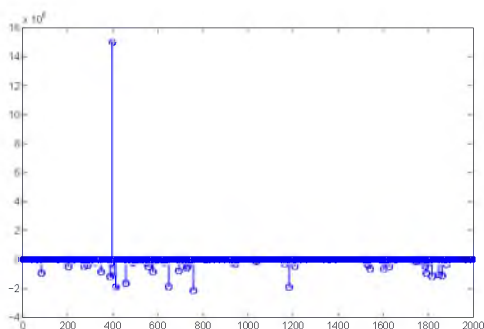


(a) PADMM算法数值结果

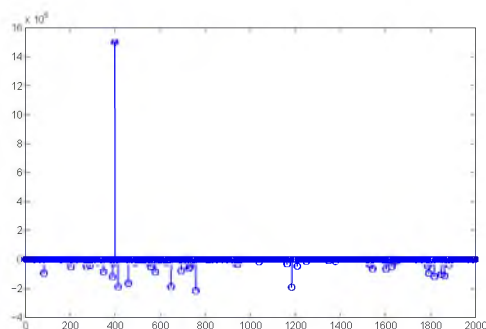


(b) DADMM算法数值结果

图 2.2 模拟的是数据库中第200张人脸图像

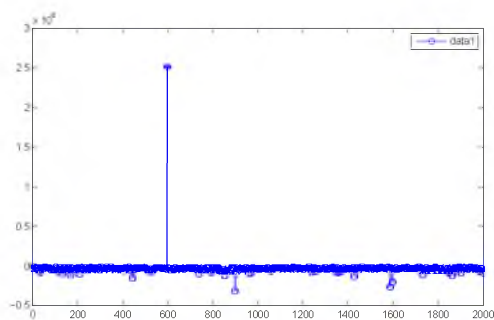


(a) PADMM算法数值结果

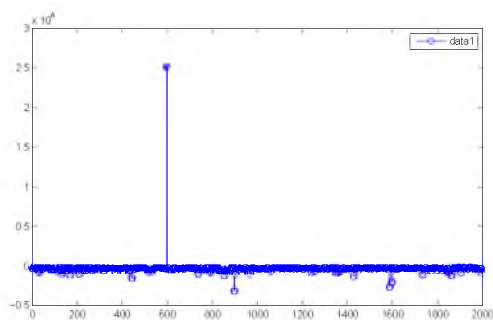


(b) DADMM算法数值结果

图 2.3 模拟的是数据库中第400张人脸图像

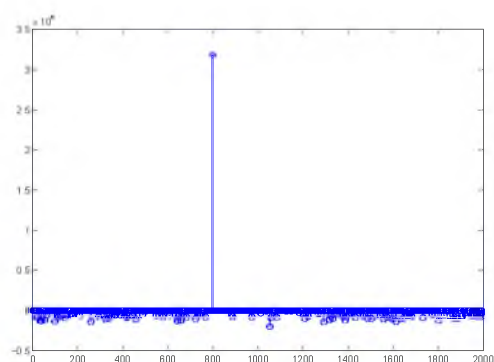


(a) PADMM算法数值结果

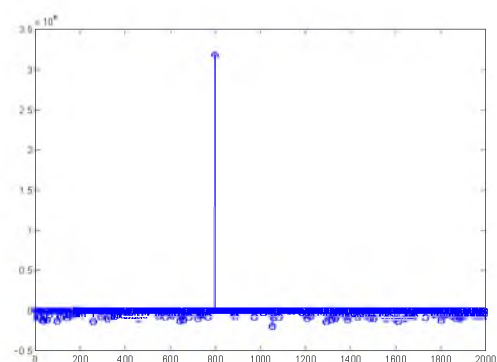


(b) DADMM算法数值结果

图 2.4 模拟的是数据库中第600张人脸图像

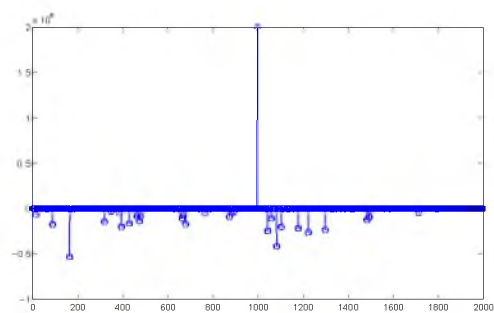


(a) PADMM算法数值结果

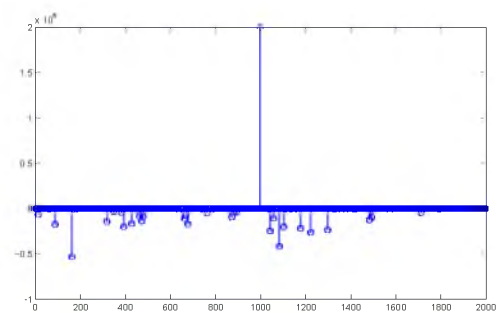


(b) DADMM算法数值结果

图 2.5 模拟的是数据库中第800张人脸图像

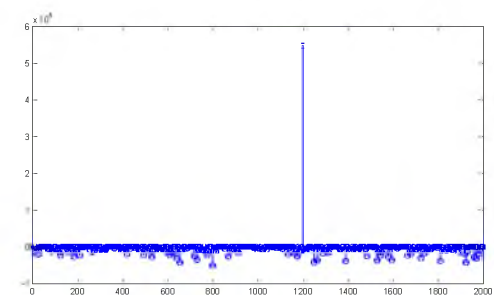


(a) PADMM算法数值结果

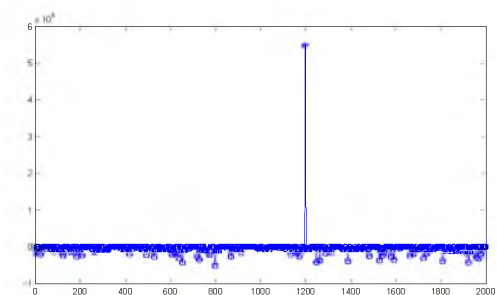


(b) DADMM算法数值结果

图 2.6 模拟的是数据库中第1000张人脸图像

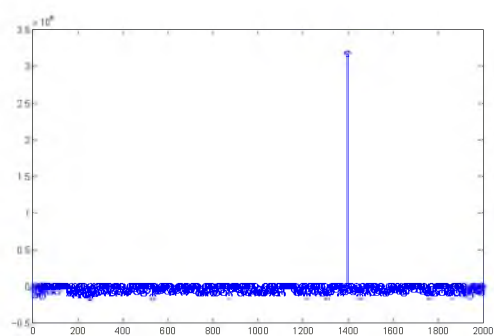


(a) PADMM算法数值结果

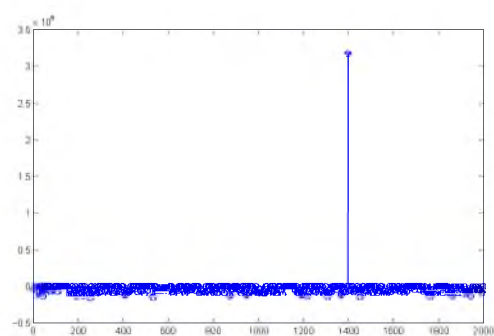


(b) DADMM算法数值结果

图 2.7 模拟的是数据库中第1200张人脸图像

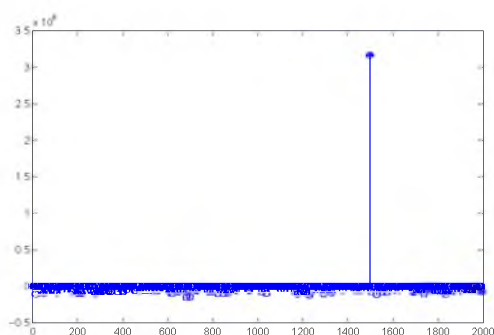


(a) PADMM算法数值结果

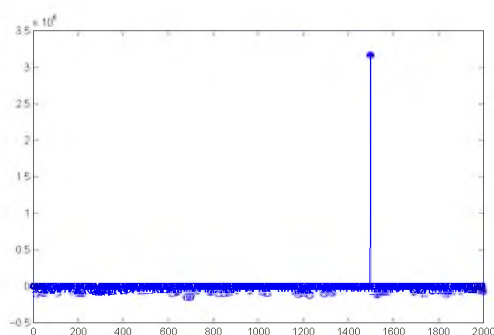


(b) DADMM算法数值结果

图 2.8 模拟的是数据库中第1400张人脸图像

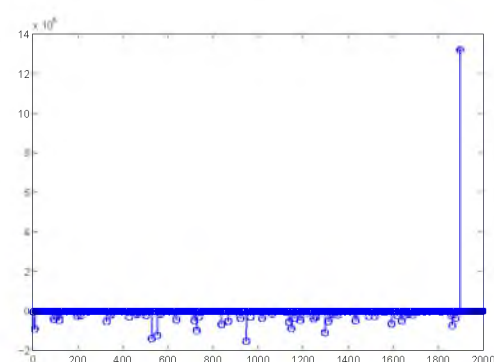


(a) PADMM算法数值结果

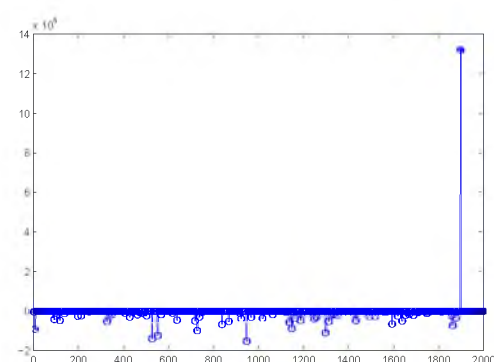


(b) DADMM算法数值结果

图 2.9 模拟的是数据库中第1500张人脸图像



(a) PADMM算法数值结果



(b) DADMM算法数值结果

图 2.10 模拟的是数据库中第1900张人脸图像

由以上结果可以看出: 用交替方向乘子法求解鲁棒人脸识别模型和对偶模型都能够

很清晰地从数图像数据库中找到所取的那张人脸图像.

表2-1 PADMM 与DADMM的计算时间(CPU)

所取数据库中图像位置	PADMM	DADMM
100	1.923	0.063
200	0.811	0.063
400	0.874	0.047
600	0.828	0.062
800	1.951	0.078
1000	0.841	0.063
1200	0.843	0.046
1400	0.842	0.048
1500	0.904	0.047
1900	0.840	0.031

由以上结果可知: 交替方向乘子法能够准确快速求解求解鲁棒人脸识别模型和对偶模型, 且基于对偶模型的交替方向乘子法具有更快的效率.

第三章 基于压缩传感 ℓ_1 - ℓ_1 范数优化模型的非单调谱梯度算法

§3.1 引言

本节主要研究原始信号含有脉冲噪音和自适应的高斯噪音的 ℓ_1 - ℓ_1 以及 ℓ_1 - ℓ_1 范数拟合项数据的优化问题. 主要考虑校正项的添加情况, 是为了得到更高稀疏的解. 求解基于压缩传感且同时含有脉冲噪音和高斯噪音的稀疏信号恢复的 ℓ_1 - ℓ_1 -范数优化模型. 通过变量代换和变量分裂, 把问题模型等价转化为凸二次规划模型, 然后使用非单调谱(Barzilai-Borwein)梯度算法进行求解.

§3.2 压缩传感模型的转换

设观测的信号 b 同时含有脉冲噪音和高斯噪音为

$$A\bar{x} + \gamma = b + \epsilon,$$

其中 $\gamma \in R^m$ 是脉冲噪音, $\epsilon \in R^m$ 是高斯噪音. 首先, 为了处理脉冲噪音 γ , 需要对 ℓ_1 范数增加惩罚项, 因此对(1.9)增加校正项为

$$\min_{x \in R^n, \gamma \in R^m, \epsilon \in R^m} \left\{ \nu \|x\|_1 + \|\gamma\|_1 + \frac{\tau}{2} \|\epsilon\|_2^2 + \nu \langle F(\tilde{x}), x \rangle : Ax + \gamma = b + \epsilon \right\}, \quad (3.1)$$

其中 $F : R^n \rightarrow R^n$, $F : (F_1(x), F_2(x), \dots, F_n(x))$ 为对称函数, 其定义为

$$F_i(x) = \begin{cases} \phi\left(\frac{x_i}{\|x\|_\infty}\right), & x \in R^n \setminus \{0\}, \\ 0, & x = 0, \end{cases}$$

其中

$$\phi(t) = \text{sgn}(t)(1 + \omega^\alpha) \frac{|t|^\alpha}{|t|^\alpha + \omega^\alpha}, \quad t \in R,$$

参数 $\omega > 0$, $\alpha > 0$. 上述校正函数见Miao, Pan 和Sun^[43].

令

$$\hat{A} = [A \ \nu I], \quad \hat{b} = \nu b, \quad \hat{\epsilon} = \nu \epsilon, \quad \hat{x} = \begin{pmatrix} \nu x \\ \gamma \end{pmatrix}, \quad \hat{F} = \begin{pmatrix} F(\tilde{x}) \\ 0 \end{pmatrix}, \quad (3.2)$$

则模型(3.1)转化为

$$\min_{\hat{x} \in R^{n+m}} \left\{ \|\hat{x}\|_1 + \frac{\tau}{2\nu^2} \|\hat{\epsilon}\|_2^2 + \langle \hat{F}, \hat{x} \rangle : \hat{A}\hat{x} = \hat{b} + \hat{\epsilon} \right\}. \quad (3.3)$$

显然, $\hat{\epsilon}$ 也是高斯噪音, 问题(3.3) 可以考虑用最小二乘法求解:

$$\min_{\hat{x} \in R^{n+m}} \left\{ \|\hat{x}\|_1 + \hat{F}^\top \hat{x} + \frac{\sigma}{2} \|\hat{A}\hat{x} - \hat{b}\|_2^2 \right\}, \quad (3.4)$$

其中 $\sigma = \tau/\nu^2 > 0$ 是罚参数.

由Figueiredo, Nowak 和Wright^[15]知, 令 $\hat{x} \in R^{m+n}$, 得

$$\hat{x} = u - v, \quad u \geq 0, \quad v \geq 0, \quad (3.5)$$

其中 $u \in R^{m+n}$, $v \in R^{m+n}$, $u_i = (x_i)_+$, $v_i = (-x_i)_+$, $i = 1, 2, \dots, m+n$. 由(3.5)计算得 $\|\hat{x}\|_1 = e_{m+n}^\top u + e_{m+n}^\top v$, 其中 e_{m+n} 是所有元素都是1的向量. 因此 ℓ_1 - ℓ_1 范数优化问题(3.4)可转化为

$$\min_{u,v} \quad \frac{\sigma}{2} \|\hat{b} - \hat{A}(u - v)\|^2 + e_{m+n}^\top u + e_{m+n}^\top v + \hat{F}^\top (u - v) \quad (3.6)$$

$$\text{s.t.} \quad u \geq 0, \quad v \geq 0. \quad (3.7)$$

由文献[15]知, 问题(3.6) 可以转化为如下凸二次规划问题:

$$\min \quad \frac{\sigma}{2} z^\top H z + c^\top z \equiv f(z) \quad (3.8)$$

$$\text{s.t.} \quad z \geq 0,$$

其中

$$z = \begin{bmatrix} u \\ v \end{bmatrix}, \quad d = \sigma \hat{A}^\top \hat{b} - \hat{F}, \quad c = e_{2(m+n)} + \begin{bmatrix} -d \\ d \end{bmatrix}, \quad H = \begin{bmatrix} \hat{A}^\top \hat{A} & -\hat{A}^\top \hat{A} \\ -\hat{A}^\top \hat{A} & \hat{A}^\top \hat{A} \end{bmatrix}. \quad (3.9)$$

易知 H 是半正定矩阵, 因此它可以由算法1.5.1很容易求解.

§3.3 算法设计

由计算知 H 是一个半正定矩阵, 因此它可以由算法1.5.1很容易求解. 由问题(1.30)知, 问题(3.8)目标函数 $f(z)$ 是一个凸二次函数, 它的梯度 $\nabla f(z) = \sigma H z + c$. 定义域限制到一个非负的凸平面 $\Omega = \{z : z \geq 0\}$, 令 $l = 0$, $u = +\infty$ 上. 算法设计如下:

算法 3.3.1 (非单调谱梯度算法)

步0. 令 $\tilde{x} \in R^n$, $b \in R^m$, $A \in R^{m \times n}$. 给定参数 $\nu > 0$, $\tau > 0$, $\alpha > 0$, $\alpha_k = 1$, $\omega > 0$. 给定常数 $M \geq 1$, 参数 $\alpha_{\max} > \alpha_{\min} > 0$, 常数 $\gamma \in (0, 1)$ 和 $0 < \sigma_1 < \sigma_2 < 1$. 给定初始点 $\hat{x}_0 = (\tilde{x}; 0) \in R^{m+n}$ 和 $\alpha_0 \in [\alpha_{\min}, \alpha_{\max}]$. 计算 $\sigma = \tau/\nu^2$. 脉冲噪音 $\epsilon_1 > 0$ 和高斯噪音 $\epsilon_2 > 0$. 令 $j = 0$, $k = 0$.

步1. 由(3.2) 计算 \hat{A} , \hat{b} .

步2. 由(3.2) 计算 \hat{F} ; 令 $u = (\hat{x}_k)_+$, $v = (-\hat{x}_k)_+$, 计算 z_k ; c , H 由(3.9)产生. 计算 $\nabla f(z_k) = \sigma H z_k + c$.

步3. 若 $\|\mathcal{P}_\Omega(z_k - \nabla f(z_k)) - z_k\| \leq \epsilon_1$, 其中 $\mathcal{P}_\Omega(z_k - \nabla f(z_k))$ 表示 $z_k - \nabla f(z_k)$ 到 Ω 上的度量投影, 是 $\min_{z \in \Omega} \|(z_k - \nabla f(z_k)) - z\|_2^2$ 的最优解, 停止, 转**步5**; 否则, 继续.

步4. 非单调线性搜索

步4.1. 令 $\lambda = \alpha_k$;

步4.2. 令 $z_+ = P_\Omega(z_k - \lambda \nabla f(z_k))$;

步4.3. 若

$$f(z_+) \leq \max_{0 \leq j \leq \min\{k, M-1\}} f(z_{k-j}) + \gamma \langle \nabla f(z_k), (z_+ - z_k) \rangle, \quad (3.10)$$

令 $\lambda_k = \lambda$, $z_{k+1} = z_+$, $s_k = z_{k+1} - z_k$, $y_k = \nabla f(z_{k+1}) - \nabla f(z_k)$, 转**步5**.

若(3.10) 不成立, 定义

$$\lambda_{new} \in [\sigma_1 \lambda, \sigma_2 \lambda]. \quad (3.11)$$

令 $\lambda = \lambda_{new}$, 转**步4.2**.

步5. 计算 $n_k = \langle s_k, y_k \rangle$. 若 $n_k \leq 0$, 令 $\alpha^{k+1} = \alpha_{\max}$; 否则, 计算 $m_k = \langle s_k, s_k \rangle$, 令

$$\alpha_{k+1} = \min \left\{ \alpha_{\max}, \max \left\{ \alpha_{\min}, \frac{m_k}{n_k} \right\} \right\},$$

$k = k + 1$, 执行步3.

步6. 令 $j = j + 1$. 由(3.5)和(3.9)计算 \hat{x}_j , 变量值是 z_k ; 由(3.2)计算 x_j 和 γ_j .

步7. 若 $\|x_j - x_{j-1}\|_2 \leq \epsilon_2$, 停止, 输出 x_j ; 否则, 令 $\tilde{x} = x_j$, $\hat{x}_0 = \hat{x}_j$, $k = 0$, 转步2.

由第一章中的定理1.5.2可得本算法的收敛性定理如下:

定理 3.3.2 [34, Theorems 2.1-2.2] 设序列 $\{z_k\}$ 是由算法(3.3.1)生成的, 参数 α_k 由(1.29)产生, 序列 $\{z_k\}$ 收敛到约束函数(3.8)的稳定点.

§3.4 收敛性分析

本节主要是对算法3.3.1求解目标函数(3.8)进行理论分析. 首先定义 $g_t(z) = [P_\Omega(z - tg(z)) - z]$, 其中 $z \in \Omega$, $t > 0$, $g(z) = \nabla f(z)$ 表示梯度. 当 $t = \alpha_k$ 时, 即 g_{α_k} 称作谱(Barzilai-Borwein)梯度; 当 $t = 1$ 时, 即 $g_1(z)$ 是连续的, 且 $\|g_1(z)\|_\infty$ 既是算法3.3.1的终止条件也是目标函数(3.8)的一阶稳定点条件. 为了对算法作理论性分析. 首先给出两个引理.

引理 3.4.1 由文献[35]和[45]知, 对所有的 $z \in \Omega$, $t \in (0, \alpha_{max}]$, 满足下面两条性质

- (1) $\langle g(z), g_t(z) \rangle \leq -\frac{1}{t} \|g_t(z)\|^2 \leq -\frac{1}{\alpha_{max}} \|g_t(z)\|^2$,
- (2) $g_t(\bar{z}) = 0$ 当且仅当 \bar{z} 是函数(3.8)的稳定点.

引理 3.4.2 见文献[35]引理2.3.1和定理2.3.3.

- (1) 对所有 $z \in \Omega$, $z \in R^n$, 设函数 $h: [0, \infty) \rightarrow R$ 为

$$h(s) = \frac{\|P_\Omega(z + sz) - z\|}{s}, s > 0, \quad (3.12)$$

单调非增.

- (2) 对所有的 $z_k \in \Omega$, 存在 $s_z > 0$, 当 $t \in [0, s_z]$ 时, 满足

$$f(P_\Omega(z - tg(z))) - f(z) \leq \gamma \langle g(z), g_t(z) \rangle. \quad (3.13)$$

由上述引理3.4.1和引理3.4.2, 易知算法的收敛性结果; 为了保证论文的完整性和易读性, 如下列出定理3.4.1的证明过程.

定理 3.4.1 [34, Theorems 2.1-2.2] 设序列 $\{z_k\}$ 是由算法(3.3.1)生成的, 参数 α_k 由(1.29)产生, 序列 $\{z_k\}$ 收敛到约束函数(3.8)的稳定点.

证明: 由引理3.4.2知, $\forall \lambda \in [0, \min\{s_{z_k}, \alpha_{min}\}]$, 满足

$$f(P_\Omega(z_k - \lambda g(z_k))) - \max_{0 \leq j \leq M-1} f(z_{k-j}) \leq f(P_\Omega(z_k - \lambda g(z_k))) - f(z_k) \leq \gamma \langle g(z_k), g_\lambda(z_k) \rangle. \quad (3.14)$$

设 \bar{z} 是序列 $\{z_k\}$ 的聚点, 即序列 $\{z_k\}$ 收敛到 \bar{z} . 下面只需证 \bar{z} 是目标函数(3.8)的稳定点即可.

- (1) 若 $\inf \lambda_k = 0$, 则存在序列 $\{z_k\}_K$ 满足 $\lim_{k \in K} \lambda_k = 0$; 令 $k \geq \bar{k}$, $0 < \sigma_1 < \rho_k < \sigma_2$,

$\psi_k = \lambda_k / \rho_k > 0$. 由线搜索(3.10) 得

$$\begin{aligned} f(P_\Omega(z_k - \psi_k g(z_k))) &> \max_{0 \leq j \leq M-1} f(z_{k-j}) + \gamma \langle g(z_k), P_\Omega(z_k - \psi_k g(z_k)) - z_k \rangle \\ &\geq f(z_k) + \gamma \langle g(z_k), P_\Omega(z_k - \psi_k g(z_k)) - z_k \rangle. \end{aligned} \quad (3.15)$$

由(3.15)得

$$f(P_\Omega(z_k - \psi_k g(z_k))) - f(z_k) > \gamma \langle g(z_k), g_{\psi_k}(z_k) \rangle. \quad (3.16)$$

由中值定理可得

$$f(P_\Omega(z_k - \psi_k g(z_k))) - f(z_k) = \langle g(z_k), g_{\psi_k}(z_k) \rangle + \langle g(\xi_k) - g(z_k), g_{\psi_k}(z_k) \rangle \quad (3.17)$$

其中 ξ_k 介于点 z_k 与点 $P_\Omega(z_k - \psi_k g(z_k))$ 之间. $\forall k \in K$, 由(3.16)和(3.17)得

$$(1 - \gamma) \langle g(z_k), g_{\psi_k}(z_k) \rangle > \langle g(z_k) - g(\xi_k), g_{\psi_k}(z_k) \rangle \quad (3.18)$$

由引理3.4.1和引理3.4.2得

$$\begin{aligned} \langle g(z_k), g_{\psi_k}(z_k) \rangle &\leq -\frac{1}{\psi_k} \|g_{\psi_k}(z_k)\|_2^2 \\ &\leq -\frac{1}{\alpha_k} \|g_{\alpha_k}(z_k)\|_2 \|g_{\psi_k}(z_k)\|_2 \end{aligned} \quad (3.19)$$

由不等式(3.17)和(3.19)且运用Cauchy-Schwartz不等式得

$$\begin{aligned} \frac{(1 - \gamma)}{\alpha_k} \|g_{\alpha_k}(z_k)\|_2 \|g_{\psi_k}(z_k)\|_2 &< \langle g(\xi_k) - g(z_k), g_{\psi_k}(z_k) \rangle \\ &\leq \|g(\xi_k) - g(z_k)\|_2 \|g_{\psi_k}(z_k)\|_2. \end{aligned} \quad (3.20)$$

当 $\|g_{\psi_k}(z_k)\|_2 \neq 0$ 时, 得

$$\frac{(1 - \gamma)}{\alpha_k} \|g_{\alpha_k}(z_k)\|_2 < \|g(\xi_k) - g(z_k)\|_2 \quad (3.21)$$

令 $\psi_k \rightarrow 0$, $z_k \rightarrow \bar{z}$, $k \rightarrow \infty$, $k \in K$, 得 $\xi_k \rightarrow \bar{z}$. 令 $\bar{K} \subset K$, 使 α_k 收敛到 $\bar{\alpha} \in [\alpha_{min}, \alpha_{max}]$.

对不等式(3.21)取极限且令 $k \rightarrow \infty$, $k \in \bar{K}$, 得

$$\|g_{\bar{\alpha}}(\bar{z})\|_2 \leq 0 \quad (3.22)$$

由范数的性质得 $\|g_{\bar{\alpha}}(\bar{z})\|_2 = 0$. 所以 \bar{z} 是目标函数(3.8) 的稳定点.

(2)当 $\inf \lambda_k \geq \rho > 0$ 时, 若 \bar{z} 不是目标函数(3.8) 的稳定点, 则 $\forall \lambda \in (0, \alpha_{max})$ 满足 $\|g_\lambda(\bar{z})\|_2 > 0$. 由函数的连续性和紧致性知, 存在 $\delta > 0$, 使得 $\|g_\lambda(\bar{z})\| \geq \delta > 0$, 其

中 $\lambda \in [\rho, \alpha_{max}]$. 由文献^[41]定理中第一部分证明可得到一个单调非增的序列 $\{f(x_{l(k)})\}$, 其中 $l(k)$ 是一个常数满足 $k - \min\{k, M - 1\} \leq l(k) < k$. 令

$$f(z_{l(k)}) = \max_{0 \leq j \leq \min\{k, M-1\}} f(z_{k-j}) \quad (3.23)$$

由(3.10)知, 当 $k > M - 1$ 时, 得

$$f(z_{l(k)}) \leq f(z_{l(l(k)-1)}) + \gamma \langle g(z_{l(k)-1}), g_{\lambda_{l(k)-1}}(z_{l(k)-1}) \rangle \quad (3.24)$$

由连续性, 对于 $k > \bar{k}$ 充分大时, 得 $g_{\lambda}(\bar{z}_k) \geq \frac{\delta}{2}$. 由引理3.4.1知

$$\begin{aligned} f(z_{l(k)}) &\leq f(z_{l(l(k)-1)}) - \frac{\gamma}{\alpha_{max}} \|g_{\lambda_{l(k)-1}}(z_{l(k)-1})\|_2^2 \\ &\leq f(z_{l(l(k)-1)}) - \frac{\gamma\delta}{4\alpha_{max}}. \end{aligned} \quad (3.25)$$

当 $k \rightarrow \infty$ 时, $f(z_{l(k)}) \rightarrow -\infty$ 这与目标函数(3.8)连续相矛盾; 所以函数 $f(z_k)$ 收敛到 $f(\bar{z})$, 即 \bar{z} 是目标函数 $f(z)$ 的稳定点.

第四章 结论

本文在求解 ℓ_1 - ℓ_1 -范数优化问题算法和应用方面做了一系列的研究工作. 提出带有校正项的鲁棒人脸识别模型和压缩传感模型. 用非精确的交替方向乘子法求解鲁棒人脸识别模型; 构造鲁棒人脸识别的对偶模型, 用精确的交替方向乘子法求解; 分析它们的收敛性, 用模拟数据验证它们的数值有效性. 数值结果表明用交替方向乘子法求解鲁棒人脸识别的对偶模型要远远优于交替方向乘子法求解鲁棒人脸识别模型. 用非单调谱梯度算法求解压缩传感模型, 设计算法且对算法进行收敛性分析. 需要探讨的问题如下:

- (1): 本论文提出求解压缩传感模型的非单调谱梯度算法与其它知名算法的效率比较, 值得去进一步研究和测试.
- (2): 鲁棒人脸识别模型用广义交替方向乘子法求解.
- (3): 鲁棒人脸识别模型用非单调谱梯度算法求解.

参考文献

- [1] 池燕玲. 基于深度学习的人脸识别方法的研究[D]. 福建, 福建师范大学, 2015.
- [2] Wright J, et al. Robust face recognition via sparse representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2): 210-227.
- [3] Yang J F, Zhang Y. Alternation direction algorithm for ℓ_1 -problem in compressive sensing[J]. SIAM Journal on Scientific Computing, 2011, 33(1): 250-278.
- [4] Candès E, Romberg J, Tao T. Stable signal recovery from incomplete and inaccurate information[J]. Communications on Pure and Applied Mathematics, 2005, 59(8): 1207-1233.
- [5] Candès E, Romberg J, Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information[J]. IEEE Transactions on Information Theory, 2006, 52(2): 489-509.
- [6] Candès E, Romberg J. Quantitative robust uncertainty principles and optimally sparse decompositions[J]. Foundations of Computational Mathematics, 2006, 6: 227-254.
- [7] Candès E, Tao T. Near optimal signal recovery from random projections: universal encoding strategies[J]. IEEE Transactions on Information Theory, 2006, 52(12): 5406-5425.
- [8] Donoho D L. Compressed sensing[J]. IEEE Transaction on Information Theory, 2006, 52(4): 1289-1306.
- [9] Hale E T, Yin W, Zhang Y. Fixed-point continuation for ℓ_1 -minimization: methodology and convergence[J]. SIAM Journal on Optimization, 2008, 19(3): 1107-1130.
- [10] Wright S J, Nowak R, Figueiredo M. Sparse reconstruction by separable approximation[J]. IEEE Transactions on Signal Processing, 2009, 57(7): 2479 - 2493.
- [11] Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems[J]. SIAM Journal on Imaging Sciences, 2009, 2(1): 183-202.
- [12] Yun S, Toh K C. A coordinate gradient descent method for ℓ_1 -regularized convex minimization[J]. Computational Optimization and Application, 2011, 48(2): 273-307.

- [13] Becker S, Bobin E J, et al. NESTA: a fast and accurate first-order method for sparse recovery[J]. SIAM Journal on Imaging Sciences, 2009, 4(1): 1-39.
- [14] Ewout V D B, Friedlander M P. Probing the pareto frontier for basis pursuit solutions[J]. SIAM Journal on Scientific Computing, 2008, 31(2): 890-912.
- [15] Figueiredo M, Nowak R, Wright S J. Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems[J]. IEEE Journal of Selected Topics in Signal Processing, 2007, 1(4): 586-597.
- [16] Xiao Y, Zhu H, Wu S Y. Primal and dual alternating direction algorithms for ℓ_1 - ℓ_1 -norm minimization problems in compressive sensing[J]. Computation Optimization and Application, 2013, 54(2): 441-459.
- [17] Rockafellar R T. Convex analysis[J]. Princeton University Press, New Jersey, 1970.
- [18] Rockafellar R T, Wets R J B. Variational Analysis[M]. Springer, New York, 1998.
- [19] Fazel M, Pong T K, Sun D, et al. Hankel matrix rank minimization with applications to system identification and realization[J]. SIAM Journal on Matrix Analysis and Applications, 2013, 34(3): 946-977.
- [20] Glowinski R, Marrocco A. Sur l'approximation, par éléments finis d'ordre 1, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires[J]. Journal of Equine Veterinary Science, 1975, 2(R-2):41-76.
- [21] Gabay D, Mercier B. A dual algorithm for the solution of nonlinear variational problems via finite element approximation[J]. Computers Mathematics with Applications, 1976, 2(1): 17-40.
- [22] Glowinski R. Lectures on numerical methods for non-linear variational problems[J]. Published for the Tata Institute of Fundamental Research, Springer-Verlag 1980, 52(3): 739.
- [23] Fortin M, Glowinski R. Augmented lagrangian methods: applications to the numerical solution of boundary-value problems[J]. Studies in Mathematics and Its Applications, 1983, 15(3): 165-193.
- [24] Glowinski R. On Alternating Direction Methods of Multipliers: a Historical Perspective[M]. Modeling, Simulation and Optimization for Science and Technology. Springer Netherlands, 2014: 59-82.

-
- [25] Eckstein J. Some saddle-function splitting methods for convex programming[J]. Optimization Methods and Software, 1994, 4(1): 75-83.
- [26] Donoho D L. For most large underdetermined systems of linear equations, the minimal ℓ_1 -norm solution is also the sparsest solution[J]. Communications on Pure and Applied Mathematics, 2006, 59: 907-934.
- [27] Bai M, Zhang X J, Shao Q. Adaptive corrected procedure for TVL1 image deblurring under impulsive noise[J]. Inverse Problems, 2016, 32(8): 085004.
- [28] 李董辉, 童小娇, 万中. 数值最优化算法与理论[M]. 科学出版社, 北京, 2006.
- [29] Barzilai J, Borwein J M. Two point step size gradient method[J]. IMA Journal of Numerical Analysis, 1988, 8(1): 141-148.
- [30] Ahonen T, Hadid A, Pietikainen M. Face description with Local binary patterns: application to face recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(12): 2037-2041.
- [31] Cao S, Xiao Y, Zhu H. Linearized alternating directions method for ℓ_1 -norm inequality constrained ℓ_1 -norm minimization[J]. Applied Numerical Mathematics, 2014, 85: 142-153.
- [32] Xiao Y, Zhu H. Primal and dual alternating direction algorithms for $\ell_1 - \ell_1$ -norm problems in compressive sensing[J]. Computational Optimization and Applications, 2013, 54(2): 441-459.
- [33] Zhou Z, Wagner A, Mobahi H, et al. Face recognition with contiguous using markov random fields[J]. IEEE Conference on Computer Vision, 2009, 30(2): 1050-1057.
- [34] Birgin E G, Martínez J M, Raydan M. Nonmonotone spectral projected gradient methods on convex sets[J]. SIAM Journal on Optimization, 10 (2000): 1196-1121.
- [35] Bertsekas D P. Nonlinear programming[J]. Athena Scientific Belmont, MA, 1995.
- [36] Eckstein J, Bertsekas D P. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators[J]. Mathematical Programming, 1992, 55(3): 293-318.
- [37] Xu M H, Wu T. A class of linearized proximal alternating direction methods[J]. Journal of Optimization Theory and Applications, 2011, 151(2): 321-337.

- [38] Anderson T W. An introduction to multivariate statistical analysis[J]. Biometrics, 1958, 41(3): 981-982.
- [39] Muirhead R. Aspects of multivariate statistical theory wiley series in Probability and mathematical statistics[J]. Journal of Multivariate Analysis, 1982.
- [40] Masao Fukushima, 林贵华. 非线性最优化基础[M]. 科学出版社, 北京, 2011.
- [41] Grippo L, Lampariello F, Lucidi S. A non-monotone line search technique for Newton's method[J]. SIAM Journal on Numerical Analysis, 1986, 23(4): 707-716.
- [42] Muirhead R J. Developments in Eigenvalue Estimation[M]. Springer Netherlands, 1987, 5: 277-288.
- [43] Miao W, Pan S, Sun D F. A rank-corrected procedure for matrix completion with fixed basis coefficients[J]. Mathematical Programming, 2016, 159(1): 289-338.
- [44] Raydan M. The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem[J]. SIAM Journal on Optimization, 1997, 7(1): 26-33.
- [45] Gafni E M, Bertsekas D P. Two-metric projection methods for constrained optimization[J]. SIAM Journal on Control and Optimization, 1982, 22(6): 936-964.

致 谢

在本论文完成之际, 谨此向我的导师肖运海教授致以衷心的感谢和崇高的敬意! 本论文是在肖教授的悉心指导下完成的, 他严谨的科学态度, 对学术精益求精的精神都深深感染和激励着我, 他每周积极组织和按时参加讨论班, 积极分享自己的学术想法. 在遇到学习困难时, 他总是不厌其烦的为我讲解, 论文修改的时候不放弃每一个细节. 肖老师不仅在学习上给了我很大的帮助而且在生活中还教会我很多为人处世的道理.

感谢河南大学数学与统计学院的每一位帮助我的老师, 感谢我硕士队的战友们: 史战文, 吴灿, 丁彦昀, 丁雯雯, 李培丽, 王双月等在这三年来对我的帮助, 关怀和指导; 感谢已经毕业的师兄师姐: 杨冠雨, 曹姝晗, 朱红等, 谢谢你们的帮助和关怀! 因为你们, 我的研究生生活变得丰富多彩! 感谢我的父母, 你们永远是我前进路上的不懈动力!

最后, 我要感谢河南大学的教育和培养, 让我更加有信心去实现自己的价值, 让我能够明确今后的发展方向. 他们的帮助对我来说是一笔无价的财富, 再次感谢他们!

陈 丽 雅

2017 年 5 月

附录 攻读硕士学位期间完成的论文及获奖情况

研究生期间完成的论文如下:

(一)Yunhai Xiao, Can Wu, Liya Chen; Alternating Direction Method of Multipliers for Robust Face Recognition with Adaptive Corrections Terms, 待投.

(二)Yunhai Xiao, Wenwen Ding, Liya Chen; Nomonotone Barzilai-Borwein Gradient Method for Large Sparse Signal Recovery with Adaptive Corrections Terms, 待投.

研究生期间获奖情况:

- (1): 2014.11 获河南大学研究生三等学业奖学金;
- (2): 2015.09 获第十二届全国研究生数学建模成功参赛奖;
- (3): 2015.11 获河南大学研究生二等学业奖学金;
- (4): 2016.03 获河南大学研究生讲课大赛优秀奖;
- (5): 2016.11 获河南大学研究生二等学业奖学金;
- (6): 2017.01 被评为国培计划”优秀顶岗实习生”荣誉称号;