

记号与约定

$R^m, R^{n \times n}, R_{++}$: 分别表示 m 维欧氏空间, n 阶实矩阵构成的集合, 全体正实数.

S^n : $n \times n$ 阶实对称矩阵空间.

$S(n_1, \dots, n_m)$: 分块对角矩阵空间, 且各分块的维数为 n_1, \dots, n_m .

S_+^n : 半正定矩阵锥.

K° : 锥 K 的极锥.

X_{ij} : 矩阵 X 上的 (i, j) 元素.

X^T : 矩阵 X 的转置.

X^{-1}, X^\dagger : 分别表示矩阵 X 的逆矩阵, Moore-penrose 广义逆.

$X \succeq 0 (X \succ 0)$: 矩阵 X 是半正定 (正定) 的.

$X^{\frac{1}{2}}, \sqrt{X}$: 半正定矩阵 X 的对称方根.

$\text{vec}(X)$: n 阶矩阵 X 的列叠成的 n^2 维列向量.

mat : vec 的逆算子, 表示 n^2 维列向量排成 n 阶矩阵.

$\|X\|_F$: 矩阵 $X \in R^{n \times n}$ 的 Frobenius 范数.

e : R^n 中的 n 维向量, 且 $e = (1, \dots, 1)^T$.

\bullet : S^n 中的标准内积.

\otimes : 标准 Kronecker 积.

$\text{tr}(X), \text{rank}(X)$: 矩阵 $X \in S^n$ 的迹, 矩阵 X 的秩.

A : S^n 到 R^m 上的线性算子, 且 $AX = [A_1 \bullet X, \dots, A_m \bullet X]^T$, A_i 为给定对称阵.

A^* : 线性算子 A 的伴随算子, 且有 $A^*y = \sum_{i=1}^m y_i A_i$.

I_n, I : n 阶单位矩阵.

$\text{Diag}(x_1, \dots, x_n)$: 对角元为 x_1, \dots, x_n 的 n 阶对角矩阵.

$\text{diag}(X)$: 矩阵 X 对角元构成的 n 维向量.

$\lambda_{\max}(X)$: 矩阵 X 的最大特征值.

$\lambda(X)$: 矩阵 X 实特征值组成的向量, 即 $\lambda(X) = (\lambda_1, \dots, \lambda_n)^T$.

$c = \max\{a, b\}$: c 中第 i 个分量 $c_i = \max\{a_i, b_i\}$, 其中 a, b 是 R^n 中 n 维向量.

$\arg \min\{f(x) : x \in X\}$: $f(x)$ 在 X 上的最小值解集.

$\nabla f(x)$: $f(x)$ 在点 x 的梯度.

$+\infty$: 正无穷大.

$\text{int}S, \text{ri}S, \partial S$: 分别是集合 S 的内部, 相对内部和边界.

■: 定理或命题证毕, 例子求解完毕.

福建师范大学学位论文使用授权声明

本人(姓名) 康志林 学号 2005650

专业 应用数学 所呈交的论文(论文题

目: 线性二次半定规划问题若干研究)是我个人在导师指导下

进行的研究工作及取得的研究成果。尽我所知,除了文中特别加以标注

和致谢的地方外,论文中不包含其他人已经发表或撰写过的研究成果。

本人了解福建师范大学有关保留、使用学位论文的规定,即:学校有权

保留送交的学位论文并允许论文被查阅和借阅;学校可以公布论文的全

部或部分内容;学校可以采用影印、缩印或其他复制手段保存论文。

(保密的论文在解密后应遵守此规定)

学位论文作者签名 康志林 指导教师签名 李光圣

签名日期 2008.5.29

摘要

本文主要探讨线性二次半定规划问题 (L-QSDP) 的结构特征及其求解算法, 主要由三部分组成. 第一部分, 先讨论线性二次半定规划问题的对偶性理论及其最优性条件, 进而讨论该规划问题的原始对偶内点算法, 给出了基于 NT 搜索方向唯一性的证明. 数值试验上, 对 $n = 3$ 的情况, 给出具体算例, 并在 MATLAB 7.01 上进行数值模拟, 验证了算法的可行性. 同时, 进一步探讨了该二次半定规划与半定最小二乘问题的联系, 给出了在一定条件下它们之间的转换关系. 第二部分, 拓广了半定最小二乘问题模型 (SDLS) 的定义, 提出了变量有界半定最小二乘问题 (BV-SDLS), 同时给出了任意实对称矩阵在由有界约束矩阵变量构成的闭凸集上的精确投影表示式, 并在此投影基础上, 探讨了该 (BV-SDLS) 的求解算法, 即投影拟牛顿算法, 给出了其算法框架. 最后在 MATLAB 7.01 上进行数值试验, 并与原始对偶内点算法比较, 进一步说明算法的可行性及有效性. 第三部分, 进一步考虑 (BV-SDLS) 模型的拓广形式, 探讨了一些特殊情形以及施加某种限制下的投影显式, 并考虑它相应的求解算法.

关键词: 二次半定规划 半定最小二乘 内点算法 变量有界 投影拟牛顿算法

Abstract

In this paper, we discuss the feature and solving algorithms for the Linear Quadratic Semi-definite Programming(L-QSDP). This paper mainly consists of three sections.

In the first section, we establish the duality theory and the optimality conditions in the problem of L-QSDP, and study the primal dual interior point method for this problem. Meanwhile, we give the proof of the unique solution based on the NT search direction. Furthermore, we make the MATLAB code for experiment when the number of dimension $n=3$, and give the numerical simulation in MATLAB 7.01, which verify its feasibility. Finally, we discuss the relation between the quadratic semi-definite programming and the semi-definite least squares problem, and then provide the conversion relationship between them under some conditions.

In the second section, we extend the definition of the semi-definite least squares problem(SDLS), and propose the semi-definite least squares problem with bounded variables(BV-SDLS). At the same time, we give the projection of symmetric matrix onto the closed convex set, which is composed of the boundary constrains. We also study the solving algorithm for (BV-SDLS) based on this projection, i.e. the projected quasi-newton algorithm, and give the algorithm frame for this problem. In the end, we make an experiment in MATLAB 7.01, and compare numerical results with that of the primal dual interior point method. Numerical results show that the extended approach is feasible and validity.

In the third section, we further discuss the extension of (BV-SDLS), and propose the explicit projective formula in certain particular circumstances. Moreover, we also study its solving algorithm.

Keywords: quadratic semidefinite programming, semi-definite least squares, interior point method, bounded variables, projected quasi-newton algorithm.

中文文摘

本文主要探讨二次半定规划问题的特征和求解算法,研究的理论依据主要是对称矩阵论及非线性规划的对偶理论.文章的结构安排如下.

第一章概括了目前国内外关于(二次)半定规划问题的研究情况和主要成果,指出(二次)半定规划数学模型广泛的实际应用背景及其重要的理论意义.

第二章研究一类特殊的二次半定规划问题的结构,其目标函数中的半正定自伴随线性算子 $Q(X) = \sum_{i=1}^l P_i X P_i, P_i \succeq 0$, 即

$$\begin{aligned} \min_{X \in S^n} \quad & \frac{1}{2} X \bullet \sum_{i=1}^l P_i X P_i + C \bullet X \\ \text{s.t.} \quad & AX = b \\ & X \succeq 0 \end{aligned}$$

其中 $C \in S^n, X \in S^n, AX = (A_1 \bullet X, \dots, A_m \bullet X), A_i \in S^n$, 且集合 $\{A_i | i = 1, 2, \dots, m\}$ 中各矩阵元素线性无关.同时,讨论该线性二次半定规划问题的对偶性理论及其最优性条件,在此基础上,进而考虑该规划问题的原始对偶内点算法,给出了当 X, Z 正定时,系统

$$\begin{pmatrix} H & 0 & 0 \\ -\sum_{i=1}^l (P_i \otimes P_i) & H^T & I \\ M & 0 & N \end{pmatrix} \begin{pmatrix} \text{vec}(\Delta X) \\ \Delta \lambda \\ \text{vec}(\Delta Z) \end{pmatrix} = \begin{pmatrix} r_p \\ r_d \\ r_c \end{pmatrix}$$

基于 NT 搜索方向 $(\Delta X, \Delta \lambda, \Delta Z) \in S^n \times R^m \times S^n$ 唯一性的证明,即定理 2.2.3.其中 $M = W^{-1} \otimes W^{-1}, N = I \otimes I, r_d = \text{vec}(R_d), r_c = \text{vec}(R_c)$, 且有如下表示式:

$$\left\{ \begin{aligned} [H(N \sum_{i=1}^l P_i \otimes P_i + M)^{-1} N H^T] \Delta \lambda &= r_p + H(N \sum_{i=1}^l P_i \otimes P_i + M)^{-1} (N r_d - r_c) \\ \Delta X &= \text{mat}[(N \sum_{i=1}^l P_i \otimes P_i + M)^{-1} (N H^T \Delta \lambda - N r_d + r_c)] \\ \Delta Z &= \text{mat}[r_d - H^T \Delta \lambda + \sum_{i=1}^l (P_i \otimes P_i) \text{vec}(\Delta X)] \end{aligned} \right.$$

在此基础上,给出了求解该二次半定规划问题基于 NT 方向的原始对偶内点算法框架,即算法 2.2.1.数值试验上,对 $n = 3$ 的情况,给出具体算例,并在 MATLAB 7.01 上进行数值模拟,验证了算法的可行性.同时,进一步探讨了该二次半定规划

与半定最小二乘问题的联系及其转换关系, 主要结论是命题 2.4.1, 命题 2.4.2, 命题 2.4.3.

第三章中, 我们从第二章中二次半定规划问题出发, 继续深入探讨特殊的二次半定规划问题 ($l = 1, P = I_n$), 即半定最小二乘问题 (SDLS), 并在此基础上, 提出了比 (SDLS) 更一般化的变量有上限约束半定最小二乘问题 (BV-SDLS)

$$\begin{aligned} \min_{X \in S^n} \quad & \frac{1}{2} \|X - C\|_F^2 \\ \text{s.t.} \quad & AX = b \\ & 0 \preceq X \preceq \beta I_n \end{aligned} \quad (BV - SDLS)$$

本章主要工作是, 先对半正定矩阵锥 S_+^n 进一步拓广, 定义一闭凸矩阵集合 $\Omega^\beta = \{X \in S^n \mid 0 \preceq X \preceq \beta I_n\}$, 其中 $\beta > 0$, 并给出了任意实对称矩阵 $C \in S^n$ 在该闭凸集 Ω^β 上的精确投影表示式

$$P_{\Omega^\beta}(C) = C_+^\beta = P_C \text{Diag}\{\max\{0, \min\{\beta e, \lambda(C)\}\}\} P_C^T$$

以及 C 到闭凸集 Ω^β 的半平方距离 (half-squared distance)

$$\begin{aligned} d_{\Omega^\beta}(C) &= \min_{X \in \Omega^\beta} \frac{1}{2} \|X - C\|_F^2 \\ &= \frac{1}{2} \left(\sum_{\lambda_i > \beta} (\lambda_i - \beta)^2 + \sum_{\lambda_i < 0} \lambda_i^2 \right) \end{aligned}$$

其中 λ_i 是实对称矩阵 C 的特征值. 主要结论有定理 3.2.10 及其推论 3.2.11, 推论 3.2.12. 在上述投影的基础上, 进而探讨了变量有界半定最小二乘问题 (BV-SDLS) 的求解算法, 即投影拟牛顿算法, 并给出了其算法框架, 即算法 0.0.1(文中算法 3.3.1).

算法 0.0.1. 投影拟牛顿算法

给定初始点 $y_0 \in R^m$, 初始 Hessian 逆近似矩阵 $H_0 \in S^m, 0 < \varepsilon < 1$, 通常取 $H_0 = I_n$.

Step 1: 根据定理 3.2.10 的投影显式计算:

$$\begin{aligned} X_k &= P_{\Omega^\beta}(C + A^* y_k) \\ \nabla(-\eta(y_k)) &= AX_k - b \triangleq g_k \\ \eta(y_k) &= d_{\Omega^\beta}(C + A^* y_k) - \frac{1}{2} \|C + A^* y_k\|_F^2 + \frac{1}{2} \|C\|_F^2 + y_k^T b \end{aligned}$$

Step 2: 若 $\|g_k\| < \varepsilon$, 则迭代中止. 否则计算搜索方向 $d_k = -H_k g_k$

Step 3: 沿方向 d_k 作线搜索, 确定步长 $\alpha_k > 0$, 令 $y_{k+1} = y_k + \alpha_k d_k$. 计算:

$$X_{k+1} = P_{\Omega^{\sigma}}(C + \mathcal{A}^* y_{k+1}), \nabla(-\eta(y_{k+1})) = \mathcal{A}X_{k+1} - b \triangleq g_{k+1}$$

Step 4: 校正 H_k 产生 H_{k+1} , 使得拟牛顿条件成立:

$$H_{k+1}q_k = p_k \text{ 其中 } q_k = g_{k+1} - g_k, p_k = y_{k+1} - y_k$$

Step 5: $k = k + 1$, 转 Step 2

依此算法迭代, 最终可获得对偶问题 (3.1) 的最优逼近解 y^* , 与此同时得到原问题 (BV-SDLS) 的最优解 X^* . 在算法执行上, 需要计算 $P_{\Omega^{\sigma}}, d_{\Omega^{\sigma}}$, 文中我们已明确给出其精确表达式. 值得一提的是, 该算法对于规模稍大的问题也是适用的.

最后我们在 MATLAB 7.01 上进行实验, 通过三个数值实验验证了算法的可行性, 并将其与内点算法进行比较, 进而说明了算法的有效性.

第四章中, 我们对 (BV-SDLS) 模型作进一步推广, 讨论如下二次半定规划问题:

$$\begin{aligned} \min_{X \in S^n} \quad & \frac{1}{2} \|X - C\|_F^2 \\ \text{s.t.} \quad & \mathcal{A}X = b \\ & 0 \preceq X \preceq B, B \in S_+^n \end{aligned}$$

指出在矩阵变换下可将上述问题转换为如下等价问题:

$$\begin{aligned} \min_{\tilde{X} \in S^n} \quad & \frac{1}{2} \|\tilde{X} - \tilde{C}\|_F^2 \\ \text{s.t.} \quad & \tilde{\mathcal{A}}\tilde{X} = b \\ & 0 \preceq \tilde{X} \preceq \text{Diag} \end{aligned}$$

其中 $\text{Diag} = \text{Diag}\{\mu_1, \mu_2, \dots, \mu_n\}, \mu_1, \mu_2, \dots, \mu_n$ 是矩阵 B 的 n 个特征值.

除此之外, 我们进一步考虑了几种特殊情形下任意实对称矩阵 $C \in S^n$ 在该闭凸集 $\Omega' = \{X \in S^n \mid 0 \preceq X \preceq \text{Diag}\}$ 上的投影, 主要有命题 4.1.1, 命题 4.1.2. 进一步地, 在对矩阵 C 与 B 施加某种限制 $CB = BC$ 后, 得到如下更广义的投影表达式

$$P_{\Omega}(C) = Q \text{Diag}\{\max\{0, \min\{\lambda(C), \lambda(B)\}\}\} Q^T$$

即定理 4.1.3, 其中 $\lambda(C), \lambda(B)$ 分别是矩阵 C 和 B 的特征值向量.

第 1 章 引言

1.1 历史文献介绍

半定规划 (Semi-definite Programming Problem(*SDP*)) 是定义在半正定矩阵锥上的规划, 是线性规划, 凸二次规划, 二阶锥规划的一种拓广^[1], 它之所以具有较强的生命力, 不仅在于它在系统论、控制论、特征值优化、组合优化^[2]等诸多领域具有着广泛的应用, 而且在于它为研究更广泛的凸优化问题的性质和算法提供了一种独特的方法, 且无论在理论上还是在实际中它都有行之有效的多项式时间算法. 目前, 研究较多的半定规划问题是在满足约束“对称矩阵的仿射组合半正定”的条件下, 使线性函数极小化, 其标准形式如下:

$$\begin{aligned} \min_{X \in S^n} \quad & C \bullet X \\ \text{s.t.} \quad & AX = b \\ & X \succeq 0 \end{aligned} \quad (L-SDP)$$

其中, $C, X, A_i \in S^n$, A 是定义在 S^n 到 R^m 的线性算子, $AX = (A_1 \bullet X, \dots, A_m \bullet X)^T$, $A_i \in S^n, i = 1, \dots, m$, 且集合 $\{A_i | i = 1, 2, \dots, m\}$ 线性无关. $b \in R^m$. 其相应的伴随算子为 $A^*: R^m \rightarrow S^n$, 且 $A^*y = \sum_{i=1}^m y_i A_i$. 容易发现, 这个锥约束是非线性的, 非光滑的, 凸的, 故线性半定规划是一凸优化问题. 进一步地, 如果目标函数和约束条件都是线性的, 称之为线性半定规划 (Linear Semi-definite Programming, 简记为 *L-SDP*), 其它的称之为非线性半定规划 (Non-Linear Semi-definite Programming, 简记为 *NL-SDP*). 一般情况下, 如果没有特别注明, 半定规划指的是线性半定规划.

早在 1963 年, R. Bellman 和 K. Fan^[3] 就构造了一个 *SDP* 问题, 考虑将线性规划中的向量变量用矩阵变量代替, 并给出了对偶理论以及强对偶理论成立的一个正则条件. 20 世纪 70 年代初, Donath 和 Hoffman, Cullum 和 Wolfe^[4, 5] 把一些图分割问题转化为特征值优化问题, 进而转化为标准的 *SDP* 问题. 事实上, 很多组合上的问题都可以最终松弛为半定规划问题, 比如最大割问题^{[6][7]}、适应性问题 (Satisfiability Problem(SAT))(NP-C 1971 Cook)^[8], 以及求一个图的团数 (Clique Number)(NP-hard)、色数 (Chromatic Number)(NP-hard) 等, 都可以用 *SDP* 松弛得到它的一个界值估计. 如考虑一无环、无重边的无向图 $G = (V, E), V = \{1, \dots, n\}$. (i, j) 表示顶点为 i, j 的边, w_{ij} 表示边 (i, j) 的权, 最大割问题其实是把顶点分为 V_+, V_- 两部

分, 使得 $\sum_{i \in V_+, j \in V_-} w_{ij}$ 达到最大, 事实上等价于如下规划问题

$$\begin{aligned} \max \quad & \sum \frac{1}{4} w_{ij} (1 - v_i v_j) \\ \text{s.t.} \quad & v_i \in \{-1, 1\}, i = 1, \dots, n \end{aligned}$$

当 $i \in V_+$ 时, $v_i = 1$; 当 $i \in V_-$ 时, $v_i = -1$. 令 $W = (w_{ij})_{n \times n}$, $L = \text{Diag}(We) - W$, 则上述问题可转化为

$$\begin{aligned} \max \quad & \frac{1}{4} v^T L v \\ \text{s.t.} \quad & v \in \{-1, 1\}^n \end{aligned}$$

其中 $v = (v_1, \dots, v_n)^T$. 令 $C = \frac{L}{4}$, $X = v^T v$, 则可得到如下半定规划松弛问题

$$\begin{aligned} \max_{X \in S^n} \quad & C \bullet X \\ \text{s.t.} \quad & \text{diag}(X) = e \\ & X \succeq 0 \end{aligned}$$

基于该问题的广泛性, 人们就希望能够通过分析半定规划的内部结构, 寻求一种求解 *SDP* 问题的有效算法. 1984 年, Karmarkar^[9] 提出了求解线性规划的多项式时间内点算法, 至此以后的几年里, 内点算法发展迅速并日趋成熟, 同时也为 *SDP* 的发展奠定了必要的基础. 1988 年, Nesterov 和 Nemirovskii^[10, 11] 通过引入自调和函数 (Self-concordant function) 提出了用内点算法解决凸规划问题的方法框架, 同时说明凸集上线性函数的最小化, 可借助于该集上的自调和障碍函数在 "多项式时间" 内完成. 特别指出, 线性规划, 含凸二次约束的凸二次规划, 半定规划都有明确且易于计算的自调和障碍函数, 因此它们都可以在 "多项式时间" 内完成. 由于半定规划是线性规划的拓展, 许多学者就考虑将线性规划的内点算法推广至半定规划, 1992 年, Alizadeh^[12] 直接将求解线性规划的多项式时间的内点算法推广至半定规划, 自此以后的十几年里, 诸多学者就一直热衷于研究半定规划问题的内点算法, 包括 Jarre^[13], Alizadeh, Haeberly, and Overton^[14], Kojima, Shindoh, and Hare^[15], Nesterov and Todd^[16] 等等. 值得注意的是, 半定规划的内点算法不同于线性规划, 依 *KKT* 系统给出的搜索方向并不能满足对称性, 故有必要先对 *KKT* 系统中的互补松弛条件对称化. 经过几年的发展和完善, 现今的对称化方法主要有 *AHO* 方向^[14], *KSH* 方向^[15], *NT* 方向^[16], 经验表明使用 *NT* 方向相比于 *AHO*, *KSH* 方向会更加稳健 (Robust)^[16].

可以发现, 自上世纪 90 年代以来的十几年里, 半定规划的求解更多的是考虑用内点算法, 但在一定条件下对矩阵变量的维数 n 足够大, 约束超过 7000 个的问题

时, 内点算法就显得无能为力了, 而在实际工程运用中, 经常遇到大规模的问题, 故有效解决大规模半定规划问题已经成为迫切需要解决的问题. 基于此, Christoph Helmberg 等^{[17][7]}于 1997 年将割平面算法用于线性半定规划的松弛问题, 并得到了很好的效果. 至此, 割平面算法也成为了求解线性半定规划的热门研究话题. 直至今日, 割平面算法主要有三种不同类型的方法: 谱丛方法^[18], 线性规划割平面方法^[19], 解析中心割平面算法 (ACCPM)^[20]. 2001 年, Kanzow 和 Nagel^{[21][22]}将扰动的互补松弛条件等价改写为非线性映射 $\phi: S^n \times S^n \times R_+ \rightarrow S^n: \phi(X, S, \tau) = 0$, 在此主要讨论了两种 ϕ 形式: 光滑最小化函数 (Smoothed minimum Function) 和光滑 F-B 函数 (Smoothed Fischer-Burmeister Function), 并将牛顿型方法应用于非线性系统, 数值结果显示这种方法优于内点法. 2006 年, J.Povh^[23]等提出了线性半定规划的边界点算法 (Boundary Point Method), 实际上是将增广 Lagrangian 罚函数方法应用于半定规划问题, 从半定规划的最优性条件 (KKT 系统) 出发, 但其不同于原始对偶内点算法, 在每次的循环过程中, 始终保证每次迭代的原始变量 X_k 及对偶变量 Z_k 处于半正定矩阵锥的边界上, 且恒满足 KKT 系统互补松弛条件. 当关于原始与对偶线性方程在误差允许范围内满足可行性, 该算法最终保证原始变量 X_k 和对偶变量 Z_k 趋于原问题和对偶问题的最优解 X^*, Z^* .

半定规划的日渐成熟与发展, 为分析和解决工程及其他规划问题提供了强有力工具, 比如非凸二次约束的二次规划问题、整数规划问题. 对于非凸二次约束的二次规划问题:

$$\begin{aligned} \inf_x x^T Q_0 x & \quad \Leftrightarrow \quad \inf_x Q_0 \bullet xx^T \\ \text{s.t. } x^T Q_i x = c_i \quad \forall i & \quad \text{s.t. } Q_i \bullet xx^T = c_i \quad \forall i \end{aligned}$$

通过适当的替换 (令 $xx^T = Y$), 可以很容易得到如下 SDP 松弛问题

$$\begin{aligned} \inf_{Y \in S^n} Q_0 \bullet Y \\ \text{s.t. } Q_i \bullet Y = c_i \quad \forall i \\ Y \succeq 0 \end{aligned}$$

从而可获得原问题的一个界值估计.

我们知道, 大部分整数规划都是 NP- 完全或者 NP- 难问题, 很难找到多项式时间算法求解, 目前解决整数规划 (IP) 的通用方法主要有: 分支定界法, 割平面方法, 动态规划方法. 当然, 在利用这些通用算法求解时, 并不是仅仅孤立地使用, 为了得到更好的效果, 经常是将这些方法或者与其他方法结合起来. 在求解整数规划过程中经常涉及到最优解的定界问题, 可以通过求解 SDP 松弛问题^[24]来获得

原问题最优解的一个界值估计. 当原始问题不好求解时, 经常借助对偶问题来分析原问题, 甚至可通过对偶问题获得原问题的解. 比如, 二次 $(-1,1)$ 整数规划问题

$$\min_{x \in \{-1,1\}^n} \frac{1}{2}x^T Qx + c^T x$$

其中 $Q \in S^n, c \in R^n$. 令人惊奇的是, 其对偶问题最终可变为如下简单的线性半定规划问题:

$$\begin{aligned} & \max_{(\lambda, \tau) \in R^{m+1}} -\tau \\ & \text{s.t.} \begin{pmatrix} Q + 2\text{Diag}(\lambda) & c \\ c^T & 2\tau - 2e^T \lambda \end{pmatrix} \succeq 0 \end{aligned}$$

20 世纪 90 年代考虑的 SDP 都是线性半定规划问题 (L-SDP), 即目标函数是线性的, 随着研究的深入, 人们自然而然地要考虑非线性半定规划问题, 而线性二次半定规划 (Linear-Quadratic Semi-definite Programming, 简记为: L-QSDP) 是二次规划的推广, 同时也是最简单的一种非线性半定规划, 而且它在控制论中也有比较强的应用背景, 因此探讨二次半定规划问题的结构及求解算法就显得很有必要也很有意义. 在此之前, 先给出二次半定规划一般模型 [25]:

$$\begin{aligned} & \min_{X \in S^n} \frac{1}{2}X \bullet Q(X) + C \bullet X \\ & \text{s.t.} \quad AX = b \\ & \quad X \succeq 0 \end{aligned} \tag{QSDP}$$

其中, Q 是 S^n 到 S^n 的半正定自伴随算子 (Self-adjoint Positive Semi-definite Operator), 即对任意 $A, B \in S^n$, 有

$$Q(A) \bullet B = A \bullet Q(B), \quad Q(A) \bullet A \geq 0$$

进一步地, 若算子 $Q(X)$ 关于 X 是线性的, 则二次半定规划 (QSDP) 即为线性二次半定规划 (L-QSDP). 显然, 对 (QSDP), 若不存在二次项, 则 (QSDP) 即为线性半定规划问题 (L-SDP).

Guan [25] 考虑了如下二次半定规划:

$$\begin{aligned} & \min_{X \in S^n} \frac{1}{2}\text{vec}^T(X) \mathcal{F}^T \mathcal{F} \text{vec}(X) - b^T \mathcal{F} \text{vec}(X) + C \bullet X \\ & \text{s.t.} \quad AX = b \\ & \quad X \succeq 0 \end{aligned}$$

其中 $\mathcal{F} = (\text{vec}^T F_1, \dots, \text{vec}^T F_s)^T \in R^{s \times n^2}, F_i (i = 1, \dots, s)$ 均为 n 阶实对称矩阵. Guan 将此二次半定规划问题的最优性条件等价转化为线性变分不等式, 将用于求

解单调线性变分不等式的投影收缩算法 (PC 算法), 应用到该二次半定规划问题中来, 但是该算法解决的仍只是一些小规模问题, 当 (QSDP) 的规模比较大时, PC 算法收效甚微. 基于此, Xu^[26] 在文献 [16] 基础上将求解线性半定规划的原始对偶内点算法推广至文献 [25] 中讨论的二次半定规划, 给出了短步长路径跟踪算法, 数值试验结果也表明基于 NT 方向的内点算法优于 PC 算法. 除此之外, K.C.Toh 等^[27] 进一步探讨了一般二次半定规划的不精确原始对偶内点算法.

半定最小二乘问题是一类特殊的二次半定规划问题, 其标准形式如下:

$$\begin{aligned} \min_{X \in S^n} \quad & \frac{1}{2} \|X - C\|_F^2 \\ \text{s.t.} \quad & AX = b \\ & X \succeq 0 \end{aligned}$$

容易发现, 该规划是严格凸规划问题, 且目标函数具有平方和这种特殊形式, 从而给问题的求解带来了某种方便, 对于该问题, 除了能够运用一般求解方法 (原始对偶内点算法) 外, 还可以考虑其他一些更为简便有效算法. 2002 年, Higham^[28] 考虑了约束仿射算子为对角算子 ($A = \text{diag}$) 半定最小二乘问题的求解算法, 即校正交错投影算法, 该算法在原来交错投影运算的基础上引入 Dykstra 校正步^[29], 然后交错投影到由约束集构成的闭凸集合上, 最终保证算法收敛到全局最优解. 有关交错投影算法的描述, 可进一步查阅文献 [30]. 在半定矩阵锥投影算子 $P_{S_+^n}$ ^[31] 的基础上, 2005 年, Malick.J^[32] 只对仿射约束求 *Lagrangian* 对偶, 提出了该问题的部分对偶方法, 并利用对称锥的 Moreau 分解定理将对偶问题改写为目标函数如下相对简化的无约束优化问题

$$\begin{aligned} \eta(y) &= -d_{(S_+^n)^\circ}(C + A^*y) + \frac{1}{2} \|C\|_F^2 + y^T b \\ &= -\frac{1}{2} \|P_{S_+^n}(C + A^*y)\|_F^2 + \frac{1}{2} \|C\|_F^2 + y^T b \end{aligned}$$

这种方法可以有效地解决中等规模的 (SDLS) 问题. Boyd S, Lin Xiao^[33] 在 [32] 的基础上进一步考虑了含不等式约束的二次半定规划问题, 称之为协方差矩阵校正问题 (LSCAP):

$$\begin{aligned} \min_{X \in S^n} \quad & \frac{1}{2} \|X - C\|_F^2 \\ \text{s.t.} \quad & AX = b \\ & \mathcal{B}X \leq d \\ & X \succeq 0 \end{aligned}$$

其中 $\mathcal{B}X = (B_1 \bullet X, \dots, B_p \bullet X)^T, d \in R^p$. 易知 (LSCAP) 也是凸二次半定规划问题, 若不考虑约束中不等式, 则 (LSCAP) 即为半定最小二乘问题. 对于此问题, 当

变量个数小于 1000, 即对应于 $n = 45$ 情况, 可以由内点算法^[16]得到有效的解决. Boyd S 从问题本身出发, 经过化简, 得到不含矩阵变量的有约束对偶问题, 其约束变量个数为不等式约束的个数, 并给出了 (LSCAP) 的对偶投影次梯度算法, 同时讨论了步长选取的条件及算法收敛性问题. 之后, Houduo Qi 等^[34]进一步考虑了约束仿射算子为对角算子的半定最小二乘问题, 通过对偶性理论将规划问题转化为求解关于半正定锥投影算子的非光滑方程

$$F(y) := \mathcal{A}(C + A^*y)_+ = b$$

其中 X_+ 定义为 C 到半正定矩阵锥 S_+^n 上的尺度投影. 同时, 作者利用非光滑牛顿法^[35]求解该方程, 并由半正定锥投影算子的强半光滑性^{[36][37]}, 证明了当初始迭代点充分靠近最优点时, 非光滑牛顿法的二次收敛性.

2007 年, Igor Grubisic 等^[38]进一步强化了仿射和半正定约束, 考虑低秩 (low-rank) 矩阵校正问题, 即在约束中增加一个秩的约束.

$$\begin{aligned} \min_{X \in S^n} \quad & \frac{1}{2} \|X - C\|^2 \\ \text{s.t.} \quad & \text{rank}(X) \leq d \\ & X_{ii} = b_i, i = 1, \dots, n \\ & X \succeq 0 \end{aligned}$$

其中 $d \in \{1, \dots, n\}$, 目标函数中的 $\|\cdot\|$ 是定义在 S^n 上的哈达玛半范数 (Hadamard semi-norm), i.e.,

$$\frac{1}{2} \|X - C\|^2 = \frac{1}{2} \sum_{i < j} W_{ij} (X_{ij} - C_{ij})^2$$

其中 W 是元素非负的加权矩阵. 注意到, 当 $d < n$ 时, 约束集 $\{X \in S^n | \text{rank}(X) \leq d\}$ 是一非凸集合, 因此无法用标准的凸优化方法, 如之前介绍的内点法, 交错投影算法来解决. 事实上, 交错投影算法所产生的序列只能产生问题的一可行点, 并不能保证收敛到最优点. Igor Grubisic 等从几何优化角度出发, 对约束集进行参数化处理得到 Cholesky 流形 (Cholesky Manifold), 并赋予黎曼结构, 将原问题转化到 Cholesky 黎曼流形上的优化问题, 并利用近几年发展起来的流形上数值优化算法, 如牛顿法, 共轭梯度法进行求解, 得到了很好的数值结果.

上述这些都是国内外关于 (二次) 半定规划取得的一些成果及最新进展. 在以下章节, 作者拟对线性二次半定规划及半定最小二乘问题作进一步的讨论与拓展.

1.2 本文主要工作

线性二次半定规划 (L-QSDP) 是半正定锥上最简单的一类非线性半定规划问题, 它在控制论中有很多应用, 特别是作为 (L-QSDP) 的特例 - 半定最小二乘问题 (SDLS), 在证券, 金融风险分析^[32], 稳健性二次优化中更凸显其魅力.

考虑到二次半定规划在实际生活中的广泛应用, 本文在线性半定规划的基础上, 考虑了一类线性二次半定规划问题的对偶性理论及其最优性条件, 进而探讨该规划问题的原始对偶内点算法, 给出了基于 NT 搜索方向唯一性的证明. 同时, 进一步研究了该二次半定规划与半定最小二乘问题间的联系, 给出了它们之间的相互转换关系. 数值试验上, 对 $n=3$ 的情况, 给出具体算例, 并在 MATLAB 7.01 上进行数值模拟, 验证了算法的可行性.

半定最小二乘问题 (SDLS) 在投资组合证券风险分析、估计等方面有着广泛的应用, 而在实际问题中, 经常要求矩阵变量约束是有界约束, 即上有界及下有界. 基于此, 我们对半定最小二乘问题模型 (SDLS) 的定义进一步拓广, 提出了变量有界半定最小二乘问题 (BV-SDLS), 同时给出了任意实对称矩阵在由界约束矩阵变量构成的闭凸集上的精确投影表示式, 该式是在半正定矩阵锥上投影的推广, 并在此投影基础上, 探讨了该 (BV-SDLS) 的求解算法, 即投影拟牛顿算法, 给出了投影拟牛顿算法的求解框架. 最后在 MATLAB 7.01 上进行数值试验, 并与原始对偶内点算法比较, 进一步说明算法的可行性及有效性. 最后, 进一步考虑 (BV-SDLS) 模型的一般形式, 探讨了一些特殊情形以及施加某种限制下的投影显式, 并探讨了它相应的求解算法.

第 2 章 二次半定规划问题及其内点算法

自上世纪 90 年代以来, 内点算法就一直是解决线性半定规划 (L-SDP) 问题比较常用的算法, 而对于线性二次半定规划 (QSDP), 我们知道, 目标函数中半正定自伴随线性算子 $Q(X)$ 不同就对应着不同类型的二次半定规划, 本章中取

$$Q(X) = \sum_{i=1}^l P_i X P_i, P_i \succeq 0$$

并考虑该线性二次半定规划问题的对偶性理论及其最优性条件, 进而探讨该规划问题的原始对偶内点算法, 给出了基于 NT 搜索方向唯一性的证明. 同时, 我们进一步探讨了该二次半定规划与半定最小二乘问题的联系, 给出了它们一定条件下的相互转换关系.

在此, 将本章节中有用到的标准 Kronecker 积的性质叙述如下, 详细的可参阅文献 [16, 41, 42]:

- $Q \bullet R = \text{vec}(Q)^T \text{vec}(R)$;
- $(Q \otimes R)^T = Q^T \otimes R^T$;
- $(Q \otimes R) \text{vec}(G) = \text{vec}(RGQ^T)$;
- $(Q \otimes R)(G \otimes T) = QG \otimes RT$;
- 若 Q, R 均可逆, 则 $(Q \otimes R)^{-1} = Q^{-1} \otimes R^{-1}$;

• 若 n 阶矩阵 Q 的所有特征值的集合为 $\Lambda(Q) = \{\lambda_i | i = 1, 2, \dots, n\}$, n 阶矩阵 R 的所有特征值的集合为 $\Lambda(R) = \{\mu_j | j = 1, 2, \dots, n\}$, 则 $Q \otimes R$ 的所有特征值的集合为

$$\Lambda(Q \otimes R) = \{\lambda_i \mu_j | i, j = 1, 2, \dots, n\}.$$

2.1 二次半定规划问题及对偶理论

考虑如下二次半定规划:

$$\begin{aligned} \min_{X \in S^n} \quad & \frac{1}{2} X \bullet \sum_{i=1}^l P_i X P_i + C \bullet X \\ \text{s.t.} \quad & \mathcal{A}X = b \\ & X \succeq 0 \end{aligned} \tag{2.1}$$

其中 $P_i \in S_+^n$ ($i = 1 \cdots l$). 易验证 $Q(X) = \sum_{i=1}^l P_i X P_i$ ($P_i \succeq 0$) 是半正定自伴随线性算子. 事实上, 对于任意的 $A, B \in S^n$, 有:

$$Q(A) \bullet B = \sum_{i=1}^l P_i A P_i \bullet B = \text{tr} \left(\sum_{i=1}^l P_i A P_i B \right) = \sum_{i=1}^l \text{tr}(P_i A P_i B)$$

$$A \bullet Q(B) = A \bullet \sum_{i=1}^l P_i B P_i = \text{tr} \left(\sum_{i=1}^l A P_i B P_i \right) = \text{tr} \left(\sum_{i=1}^l P_i A P_i B \right) = \sum_{i=1}^l \text{tr}(P_i A P_i B)$$

即

$$Q(A) \bullet B = A \bullet Q(B)$$

且由于 $P_i^{\frac{1}{2}} A P_i^{\frac{1}{2}} \in S^n$ ($i = 1, \dots, l$), 则

$$\begin{aligned} Q(A) \bullet A &= \sum_{i=1}^l P_i A P_i \bullet A = \sum_{i=1}^l \text{tr}(P_i A P_i A) \\ &= \sum_{i=1}^l \text{tr}(P_i^{\frac{1}{2}} P_i^{\frac{1}{2}} A P_i^{\frac{1}{2}} P_i^{\frac{1}{2}} A) = \sum_{i=1}^l \text{tr}(P_i^{\frac{1}{2}} A P_i^{\frac{1}{2}} P_i^{\frac{1}{2}} A P_i^{\frac{1}{2}}) = \sum_{i=1}^l \| P_i^{\frac{1}{2}} A P_i^{\frac{1}{2}} \|_F^2 \geq 0 \end{aligned}$$

在后续章节中我们将发现该类型二次半定规划是半定最小二乘问题的拓广形式.

通过对目标函数进行改写, 我们有:

$$\begin{aligned} & \frac{1}{2} X \bullet \sum_{i=1}^l P_i X P_i + C \bullet X \\ &= \frac{1}{2} \text{vec}(X)^T \text{vec} \left(\sum_{i=1}^l P_i X P_i \right) + C \bullet X \\ &= \frac{1}{2} \text{vec}(X)^T \left(\sum_{i=1}^l P_i \otimes P_i \right) \text{vec}(X) + C \bullet X \end{aligned}$$

当 $P_i \succeq 0$ ($P_i \succ 0$), 由标准 Kronecker 积性质, 有 $\sum_{i=1}^l P_i \otimes P_i \succeq 0$ ($\sum_{i=1}^l P_i \otimes P_i \succ 0$), 矩阵的半正定约束虽然是非线性非光滑的, 但也是凸的, 故该二次半定规划为凸 (严格凸) 二次优化问题.

若取矩阵变量 $X = \text{Diag}(x_1, x_2, \dots, x_n)$, 令 $y = \text{vec}(X)$, 则 $X \succeq 0 \Leftrightarrow \text{vec}(X) \geq 0$. 故二次半定规划就等价于如下二次规划问题:

$$\begin{aligned} \min_{y \in \mathbb{R}^n} & \frac{1}{2} y^T \left(\sum_{i=1}^l P_i \otimes P_i \right) y + \text{vec}(C)^T y \\ \text{s.t.} & \text{vec}(A_i)^T y = b_i, i = 1, \dots, m \\ & y \geq 0 \end{aligned} \tag{2.2}$$

在讨论该规划的对偶性及最优性条件之前, 先给出以下几个有用的引理:

引理 2.1.1. ^[6] 设 $A, B \in S^n$, 若 $A \succeq 0, B \succeq 0$, 则 $A \bullet B \geq 0$, 且 $A \bullet B = 0$ 当且仅当 $AB = 0$.

引理 2.1.2. ^[6] 设 B 是 $n \times n$ 阶非奇异矩阵, 则 $A \in S_+^n$ 当且仅当 $B^T A B \in S_+^n$; $A \in S_{++}^n$ 当且仅当 $B^T A B \in S_{++}^n$.

引理 2.1.3. 若 $P_i = P_i^T, i = 1, \dots, l$, 则 $\frac{\partial(\frac{1}{2} X \bullet \sum_{i=1}^l P_i X P_i)}{\partial X} = \sum_{i=1}^l P_i X P_i$.

证明: 由 [41], $\frac{\partial(\text{tr}(X A X B))}{\partial X} = (A X B + B X A)^T$, 则

$$\begin{aligned} \frac{\partial(\frac{1}{2} X \bullet \sum_{i=1}^l P_i X P_i)}{\partial X} &= \frac{\partial(\frac{1}{2} \text{tr}(X \sum_{i=1}^l P_i X P_i))}{\partial X} \\ &= \frac{\partial(\frac{1}{2} \text{tr}(\sum_{i=1}^l X P_i X P_i))}{\partial X} \\ &= \frac{1}{2} \frac{\partial(\text{tr}(\sum_{i=1}^l X P_i X P_i))}{\partial X} \\ &= \frac{1}{2} \sum_{i=1}^l (P_i X P_i + P_i X P_i)^T \\ &= \sum_{i=1}^l P_i X P_i \end{aligned}$$

引入拉格朗日乘子 $\lambda \in R^m, Z \in S_+^n$, 则二次半定规划问题的拉格朗日函数为:

$$L(X, \lambda, Z) = \frac{1}{2} X \bullet \sum_{i=1}^l P_i X P_i + C \bullet X - \sum_{i=1}^m \lambda_i (A_i \bullet X - b_i) - X \bullet Z,$$

由引理 2.1.3 可求得函数 L 关于变量 X 的偏导数, 并令其为 0, 则有:

$$\frac{\partial L}{\partial X} = \sum_{i=1}^l P_i X P_i + C - \sum_{i=1}^m \lambda_i A_i - Z = 0, Z \succeq 0$$

利用凸规划的对偶理论, 可得原二次半定规划的对偶规划:

$$\begin{aligned} \max_{\lambda, Z} \quad & -\frac{1}{2} X \bullet \sum_{i=1}^l P_i X P_i + \sum_{i=1}^m \lambda_i b_i \\ \text{s.t.} \quad & \sum_{i=1}^l P_i X P_i + C - \sum_{i=1}^m \lambda_i A_i - Z = 0 \\ & Z \succeq 0 \end{aligned} \tag{2.3}$$

显然, 原二次半定规划 (2.1) 的对偶规划也是二次半定规划, 目标函数是拉格朗日函数在可行域上的简化. 当然, 在一定条件下, 可以把对偶规划改写为只含对偶变量 λ, Z 的规划形式. 在此, 我们考虑一般情况, 即 $P_i \geq 0 (i = 1, \dots, l)$ 的情形, 而 $P_i > 0 (i = 1, \dots, l)$ 是显然的.

引理 2.1.4. ^[41] (1) 若矩阵 $Q_i (i = 1, \dots, l)$ 相互正交, 即 $Q_i^T Q_j = 0 (i \neq j)$, 则

$$(Q_1 + Q_2 + \dots + Q_l)^\dagger = Q_1^\dagger + Q_2^\dagger + \dots + Q_l^\dagger.$$

(2) 对于任意 n 阶矩阵 Q, R , 有 $(Q \otimes R)^\dagger = Q^\dagger \otimes R^\dagger, (Q^\dagger)^T = (Q^T)^\dagger$.

引理 2.1.5. 若矩阵 $P_i^T = P_i (i = 1, \dots, l)$, 则 P_i 与 $P_j (i \neq j)$ 正交当且仅当 $P_i \otimes P_i$ 与 $P_j \otimes P_j$ 正交.

证明: 若 P_i 与 P_j 正交, 即 $P_i P_j = 0$, 则

$$(P_i \otimes P_i)^T (P_j \otimes P_j) = (P_i \otimes P_i) (P_j \otimes P_j) = P_i P_j \otimes P_i P_j = 0.$$

反之, 若 $P_i \otimes P_i$ 与 $P_j \otimes P_j$ 正交, 则 $P_i P_j \otimes P_i P_j = (P_i \otimes P_i) (P_j \otimes P_j) = (P_i \otimes P_i)^T (P_j \otimes P_j) = 0$, 故推得 $P_i P_j = 0$. ■

现记

$$\sum_{i=1}^l P_i X P_i = Z + \sum_{i=1}^m \lambda_i A_i - C \triangleq S \quad (2.4)$$

假定关于 X 的矩阵方程 (2.4) 是有解的, 用 vec 算子作用, 有 $(\sum_{i=1}^l P_i \otimes P_i) \text{vec}(X) = \text{vec}(S)$, 则取其特殊解: $\text{vec}(X) = (\sum_{i=1}^l P_i \otimes P_i)^\dagger \text{vec}(S)$. 因此, 当二次半定规划 (2.1) 中矩阵 P_i 与 P_j 正交, 即 $P_i P_j = 0$ 时, 由引理 2.1.4, 引理 2.1.5, 易发现其对偶目标函数也具有二次半定规划 (2.1) 的表达形式. 事实上, 对偶规划 (2.3) 的目

标函数可进一步改写为:

$$\begin{aligned}
 & -\frac{1}{2}X \bullet \sum_{i=1}^l P_i X P_i + \sum_{i=1}^m \lambda_i b_i \\
 = & -\frac{1}{2} \text{vec}(X)^T \left(\sum_{i=1}^l P_i \otimes P_i \right) \text{vec}(X) + \sum_{i=1}^m \lambda_i b_i \\
 = & -\frac{1}{2} \text{vec}(S)^T \left(\sum_{i=1}^l P_i \otimes P_i \right)^\dagger \left(\sum_{i=1}^l P_i \otimes P_i \right) \left(\sum_{i=1}^l P_i \otimes P_i \right)^\dagger \text{vec}(S) + \sum_{i=1}^m \lambda_i b_i \\
 = & -\frac{1}{2} \text{vec}(S)^T \left(\sum_{i=1}^l P_i \otimes P_i \right)^\dagger \text{vec}(S) + \sum_{i=1}^m \lambda_i b_i \\
 = & -\frac{1}{2} \text{vec}(S)^T \sum_{i=1}^l (P_i \otimes P_i)^\dagger \text{vec}(S) + \sum_{i=1}^m \lambda_i b_i \\
 = & -\frac{1}{2} \text{vec}(S)^T \left(\sum_{i=1}^l P_i^\dagger \otimes P_i^\dagger \right) \text{vec}(S) + \sum_{i=1}^m \lambda_i b_i \\
 = & -\frac{1}{2} \text{vec}(S)^T \text{vec} \left(\sum_{i=1}^l P_i^\dagger S P_i^\dagger \right) + \sum_{i=1}^m \lambda_i b_i \\
 = & -\frac{1}{2} S \bullet \sum_{i=1}^l P_i^\dagger S P_i^\dagger + \sum_{i=1}^m \lambda_i b_i
 \end{aligned}$$

因而, 当二次半定规划 (2.1) 中矩阵 P_i 与 P_j 正交, 即 $P_i P_j = 0$ 时, 其对偶规划具有与原规划异常对称的表示式:

$$\begin{aligned}
 \max_{\lambda, Z} \quad & -\frac{1}{2} S \bullet \sum_{i=1}^l P_i^\dagger S P_i^\dagger + \sum_{i=1}^m \lambda_i b_i \\
 \text{s.t.} \quad & \lambda \in R^m, Z \succeq 0
 \end{aligned} \tag{2.5}$$

其中 $S = Z + \sum_{i=1}^m \lambda_i A_i - C$, P_i^\dagger 是矩阵 P_i 的 Morre - Penrose 广义逆.

设原始问题的目标函数为 $F(X)$, 对偶问题的目标函数为 $q(\lambda, Z) = \min_X L(X, \lambda, Z)$, 原问题的可行域为 $\mathcal{D} = \{X \in S^n \mid A_i \bullet X = b_i (i = 1, \dots, m), X \succeq 0\}$. 由凸规划的对偶性理论, 我们有如下弱对偶定理:

命题 2.1.6. 设 X 为二次半定规划原问题的可行解, λ, Z 为对偶问题的可行解, 则

$$\sup_{\lambda, Z} q(\lambda, Z) \leq \inf_X F(X)$$

证明: 设 Y 为原问题的可行解, X, λ, Z 为对偶问题的可行解. $F(X) = \frac{1}{2} X \bullet \sum_{i=1}^l P_i X P_i + C \bullet X$ 为凸函数, 由凸性可知, $\forall X, Y \in \mathcal{D}$, 有 $F(Y) \geq F(X) + \nabla F(X) \bullet (Y - X)$, 其中 $\nabla F(X) = \sum_{i=1}^l P_i X P_i + C$. 因为 X, λ, Z 为对偶问题的可行

解, 所以 $\nabla_X L(x, \lambda, Z) = \nabla F(X) - \sum_{i=1}^m \lambda_i A_i - Z, Z \succeq 0$.

$$\begin{aligned} F(Y) &\geq F(X) + \left(\sum_{i=1}^m \lambda_i A_i - Z\right) \bullet (Y - X) \\ &= F(X) + \sum_{i=1}^m \lambda_i (A_i \bullet Y) + Z \bullet Y - \sum_{i=1}^m \lambda_i (A_i \bullet X) - X \bullet Z \\ &\geq F(X) - \sum_{i=1}^m \lambda_i (A_i \bullet X - b_i) - X \bullet Z = L(X, \lambda, Z) \end{aligned}$$

故 $q(\lambda, Z) = \inf_X L(X, \lambda, Z) \leq F(X)$, 进而 $q^* = \sup_{\lambda, Z} q(\lambda, Z) \leq \inf_X F(X)$ ■

命题 2.1.7. 设 $\bar{X}, \bar{\lambda}, \bar{Z}$ 满足二次半定规划 (2.1) 的 KKT 条件 (2.6), 则 \bar{X} 和 $(\bar{\lambda}, \bar{Z})$ 是原问题与对偶问题的最优解, 且具有零对偶间隙.

证明: 由已知 $\bar{X}, \bar{\lambda}, \bar{Z}$ 满足二次半定规划 (2.1) 的 KKT 条件, 故 \bar{X} 是原始可行的. 又 $L(X, \bar{\lambda}, \bar{Z})$ 关于 X 是凸函数, 由 KKT 条件的第二式知, $L(X, \bar{\lambda}, \bar{Z})$ 在 \bar{X} 处的梯度为 0, 即 $\nabla_X L(\bar{X}, \bar{\lambda}, \bar{Z}) = 0$, 也即 $L(X, \bar{\lambda}, \bar{Z})$ 在 \bar{X} 处取得最小值, 因而有

$$\begin{aligned} q(\bar{\lambda}, \bar{Z}) &= L(\bar{X}, \bar{\lambda}, \bar{Z}) \\ &= \frac{1}{2} \bar{X} \bullet \sum_{i=1}^l P_i \bar{X} P_i + C \bullet \bar{X} - \sum_{i=1}^m \bar{\lambda}_i (A_i \bullet \bar{X} - b_i) - \bar{X} \bullet \bar{Z} \\ &= \frac{1}{2} \bar{X} \bullet \sum_{i=1}^l P_i \bar{X} P_i + C \bullet \bar{X} \\ &= F(\bar{X}) \end{aligned}$$

故 \bar{X} 与 $(\bar{\lambda}, \bar{Z})$ 具有零对偶间隙. 再由命题 2.1.6 知, \bar{X} 是原始最优解, $(\bar{\lambda}, \bar{Z})$ 是对偶最优解. ■

注 2.1.1. 命题 2.1.7 给出了二次半定规划的最优性条件, 为我们用原始对偶内点算法解决此二次半定规划问题 (2.1) 提供了理论依据.

2.2 原始对偶路径跟踪算法

众所周知, 在对优化问题 (特别是凸规划问题) 进行求解时, 经常从该问题的最优性条件出发, 去求关于最优性条件的非线性方程组, 因此我们有必要先了解最广泛也最经典的求解非线性方程组方法, 即牛顿法.

设 $f: R^n \rightarrow R^n$ 是连续可微向量值函数, 考虑非线性方程 $f(x) = 0$, 若给定初始点 $x_0 \in R^n$, 依

$$x^{k+1} = x^k - \nabla f(x^k)^{-1} f(x^k)$$

迭代, 并设搜索方向 $d = \Delta x^k$, 下一代点 $x^{k+1} = x^k + \Delta x^k$, 则有

$$\nabla f(x^k) \Delta x^k = -f(x^k)$$

易发现, 经过牛顿法改写的方程组是一线性方程组, 相比较于非线性方程组求解, 自然会容易得多. 如果 f 仅仅是局部 Lipschitz 函数, 而非光滑向量值函数, 显然上式是不适用的. 但是, 我们可以用 f 在点 x^k 的广义 Jacobian 矩阵 $\partial f(x^k)$ ^[35] 来代替 $\nabla f(x^k)$, 利用迭代式 $x^{k+1} = x^k - V_k^{-1} f(x^k)$, 其中 $V_k \in \partial f(x^k)$. 可以证明, 当向量值函数 f 是半光滑 (semismooth) 函数且满足某些条件下, 迭代式所产生的点列局部 (全局) 收敛于 $f(x) = 0$ 的解 [35]. 因此, 牛顿法对于求解非线性方程组是相对有效的.

进一步, 在讨论二次半定规划的内点算法之前, 先引入矩阵函数方向导数^[36] 概念. 其实, 矩阵函数的方向导数是向量值函数方向导数的推广.

定义 2.2.1.^[36] 设 F 是 S^n 到 S^n 的矩阵函数, 且 $X, H \in S^n$, 若极限 $F'(X, H) = \lim_{\tau \rightarrow 0} \frac{F(X+\tau H) - F(X)}{\tau}$ 存在, 则称 $F'(X, H)$ 是 F 在点 X 沿着 H 的方向导数.

引理 2.2.1. 矩阵函数 $F(X) = \sum_{i=1}^l P_i X P_i$ 沿 H 的方向导数为 $F'(X, H) = \sum_{i=1}^l P_i H P_i$.

证明: 由矩阵函数方向导数的定义, 有

$$\begin{aligned} F'(X, H) &= \lim_{\tau \rightarrow 0} \frac{F(X+\tau H) - F(X)}{\tau} \\ &= \lim_{\tau \rightarrow 0} \frac{\sum_{i=1}^l P_i (X+\tau H) P_i - \sum_{i=1}^l P_i X P_i}{\tau} \\ &= \lim_{\tau \rightarrow 0} \frac{\tau \sum_{i=1}^l P_i H P_i}{\tau} \\ &= \sum_{i=1}^l P_i H P_i \end{aligned}$$

因此有 $F'(X, H) = \sum_{i=1}^l P_i H P_i$. ■

由 §2.1 节出发, 我们容易得到二次半定规划 (2.1) 的 KKT 条件:

$$\begin{aligned} A_i \bullet X &= b_i, i = 1 \cdots m \\ \sum_{i=1}^l P_i X P_i + C - \sum_{i=1}^m \lambda_i A_i &= Z \\ X \bullet Z = XZ = X \bullet \left(\sum_{i=1}^l P_i X P_i \right) + C \bullet X - \sum_{i=1}^m \lambda_i b_i &= 0 \\ X = X^T \succeq 0, Z = Z^T \succeq 0 \end{aligned} \quad (2.6)$$

先对 KKT 系统的第三个方程进行松弛, 即将其改写为 $XZ = \mu I (\mu > 0)$, 得到扰动方程组:

$$\begin{aligned} A_i \bullet X &= b_i, i = 1 \cdots m \\ Z - \sum_{i=1}^l P_i X P_i - C + \sum_{i=1}^m \lambda_i A_i &= 0 \\ X \bullet Z = XZ &= \mu I \end{aligned}$$

假设原问题和对偶问题是严格可行的, 即存在点 (X, λ, Z) , 使得 $X \succ 0, Z \succ 0$. 该系统的解 $\{X(\mu), \lambda(\mu), Z(\mu), \mu > 0\}$ 称为二次半定规划的中心路径. 我们称 $X(\mu)$ 为原问题的 μ 中心, $(\lambda(\mu), Z(\mu))$ 为对偶问题的 μ 中心. 如果 $\mu \rightarrow 0$, 则可导出原问题和对偶问题的最优解.

按扰动方程组可定义一映射 $\psi_\mu : S_+^n \times R^m \times S_+^n \rightarrow R^m \times S_+^n \times R^{n \times n}$

$$\psi_\mu(X, \lambda, Z) = \begin{pmatrix} AX - b \\ Z - \sum_{i=1}^l P_i X P_i - C + \sum_{i=1}^m \lambda_i A_i \\ XZ - \mu I \end{pmatrix}$$

进一步, 设下一步迭代的搜索方向为 $(\Delta X, \Delta \lambda, \Delta Z)$, 用 *Newton* 法对非线性矩阵方程组 $\psi_\mu(X, \lambda, Z) = 0$ 进行求解, 再利用引理 2.2.1, 可得牛顿系统方程为:

$$\begin{aligned} A_i \bullet \Delta X &= -(A_i \bullet X - b_i) \triangleq r_{pi}, i = 1 \cdots m \\ \Delta Z - \sum_{i=1}^l P_i \Delta X P_i + \sum_{i=1}^m \Delta \lambda_i A_i &= -(Z - \sum_{i=1}^l P_i X P_i - C + \sum_{i=1}^m \lambda_i A_i) \triangleq R_d \\ \Delta X Z + X \Delta Z &= \mu I - XZ \end{aligned}$$

遗憾的是, 由上述系统得到的搜索方向 $(\Delta X, \Delta \lambda, \Delta Z) \in R^{n \times n} \times R^m \times R^{n \times n}$, 我们希望得到的新的搜索方向 $(\Delta X, \Delta \lambda, \Delta Z) \in S^n \times R^m \times S^n$, 因此需将牛顿系统方程进行对称化, 从而获得 $S^n \times R^m \times S^n$ 中新的搜索方向 $(\Delta X, \Delta \lambda, \Delta Z)$. 针对第三个

方程的对称化方法, 主要有 NT , AHO , KSH , HKM ^[16, 43, 44]. 在此简要叙述下这几种对称化方法的基本思想.

AHO 方法基本思想是直接对 XZ 对称化, 将 KKT 系统的第三个方程 $XZ = \mu I$ 等价地改写为 $\frac{XZ+ZX}{2} = \mu I$ ^[43]. KSH 方法和 HKM 方法的思想是相似的, 原则是不影响中心路径的定义. 对于 KSH 方法, 由于 $X^{-\frac{1}{2}}(X^{\frac{1}{2}}ZX^{\frac{1}{2}})X^{\frac{1}{2}} = \mu I = ZX$, ZX 相似于 $X^{\frac{1}{2}}ZX^{\frac{1}{2}}$, 故用 $X^{\frac{1}{2}}ZX^{\frac{1}{2}} = \mu I$ 来取代 $XZ = \mu I$. 对于 HKM 方法, 由于 $Z^{-\frac{1}{2}}(Z^{\frac{1}{2}}XZ^{\frac{1}{2}})Z^{\frac{1}{2}} = \mu I = XZ$, XZ 相似于 $Z^{\frac{1}{2}}XZ^{\frac{1}{2}}$, 故用 $Z^{\frac{1}{2}}XZ^{\frac{1}{2}} = \mu I$ 来取代 $XZ = \mu I$. 有关 KSH 和 HKM 方向的讨论, 有兴趣的读者可进一步参阅 [44].

基于 NT 方向的稳定性, 我们采用 NT ^[16] 搜索方向, 令 $W = X^{\frac{1}{2}}(X^{\frac{1}{2}}ZX^{\frac{1}{2}})^{-\frac{1}{2}}X^{\frac{1}{2}}$, 将牛顿系统的第三个方程改写为:

$$W^{-1}\Delta XW^{-1} + \Delta Z = \mu IX^{-1} - Z \triangleq R_c$$

则该二次半定规划基于 NT 的搜索方向 $(\Delta X, \Delta \lambda, \Delta Z) \in S^n \times R^m \times S^n$ 可由以下系统方程求得:

$$\begin{cases} \mathcal{A}\Delta X = r_p \\ \Delta Z - \sum_{i=1}^l P_i \Delta X P_i + \sum_{i=1}^m \Delta \lambda_i A_i = R_d \\ W^{-1}\Delta XW^{-1} + \Delta Z = R_c \end{cases}$$

其中 $\mathcal{A}\Delta X = (A_1 \bullet \Delta X, \dots, A_m \bullet \Delta X)^T$, $r_p = (r_{p1}, r_{p2}, \dots, r_{pm})^T$.

定义 $H = (\text{vec}(A_1), \dots, \text{vec}(A_m))^T \in R^{m \times n^2}$, $\mathcal{A}X = (A_1 \bullet X, \dots, A_m \bullet X)^T$, 将 vec 算子作用于上述系统, 得到:

$$\begin{cases} H \text{vec}(\Delta X) = r_p \\ \text{vec}(\Delta Z) - \sum_{i=1}^l (P_i \otimes P_i) \text{vec}(\Delta X) + H^T \text{vec}(\Delta \lambda) = r_d \\ (W^{-1} \otimes W^{-1}) \text{vec}(\Delta X) + (I \otimes I) \text{vec}(\Delta Z) = r_c \end{cases}$$

将其改写为矩阵的形式:

$$\begin{pmatrix} H & 0 & 0 \\ -\sum_{i=1}^l (P_i \otimes P_i) & H^T & I \\ M & 0 & N \end{pmatrix} \begin{pmatrix} \text{vec}(\Delta X) \\ \Delta \lambda \\ \text{vec}(\Delta Z) \end{pmatrix} = \begin{pmatrix} r_p \\ r_d \\ r_c \end{pmatrix} \quad (2.7)$$

其中 $M = W^{-1} \otimes W^{-1}$, $N = I \otimes I$, $r_d = \text{vec}(R_d)$, $r_c = \text{vec}(R_c)$.

引理 2.2.2. 若 X, Z 正定, 则 M 和 $\sum_{i=1}^l (P_i \otimes P_i) + N^{-1}M$ 都是正定的.

证明: 显然, $M = W^{-1} \otimes W^{-1}$ 对称正定. 而

$$N^{-1}M = (I \otimes I)(W^{-1} \otimes W^{-1}) = W^{-1} \otimes W^{-1} = M.$$

由于对任意的 i , 有 $P_i \succeq 0$, 故 $P_i \otimes P_i \succeq 0 (i = 1, \dots, l)$, 则 $\sum_{i=1}^l P_i \otimes P_i \succeq 0$, 因此 $\sum_{i=1}^l (P_i \otimes P_i) + N^{-1}M$ 是正定的. \blacksquare

定理 2.2.3. 若 X, Z 是正定的, 则系统 (2.7) 有唯一解 $(\Delta X, \Delta \lambda, \Delta Z) \in S^n \times R^m \times S^n$, 且有表示式

$$\left\{ \begin{array}{l} [H(N \sum_{i=1}^l P_i \otimes P_i + M)^{-1}NH^T]\Delta \lambda = r_p + H(N \sum_{i=1}^l P_i \otimes P_i + M)^{-1}(Nr_d - r_c) \\ \Delta X = \text{mat}[(N \sum_{i=1}^l P_i \otimes P_i + M)^{-1}(NH^T\Delta \lambda - Nr_d + r_c)] \\ \Delta Z = \text{mat}[r_d - H^T\Delta \lambda + \sum_{i=1}^l (P_i \otimes P_i)\text{vec}(\Delta X)] \end{array} \right.$$

证明: 令系数矩阵

$$B = \begin{pmatrix} H & 0 & 0 \\ -\sum_{i=1}^l (P_i \otimes P_i) & H^T & I \\ M & 0 & N \end{pmatrix}.$$

假设 $B \in R^{(2n^2+m) \times (2n^2+m)}$ 是奇异的, 则系统 $Bd = 0$ 有非零解 $d = (d_1, d_2, d_3)^T \in R^{n^2+m+n^2}$, 且 $d \neq 0$, 则有

$$\left\{ \begin{array}{l} Hd_1 = 0 \\ -\sum_{i=1}^l (P_i \otimes P_i)d_1 + H^T d_2 + d_3 = 0 \\ Md_1 + Nd_3 = 0 \end{array} \right.$$

第二式两边同乘以 d_1^T , 得到 $-d_1^T \sum_{i=1}^l (P_i \otimes P_i)d_1 + d_1^T H^T d_2 + d_1^T d_3 = 0$, 由 $Hd_1 = 0$ 知, $d_1^T \sum_{i=1}^l (P_i \otimes P_i)d_1 - d_1^T d_3 = 0$, 又由第三式知 $d_3 = -N^{-1}Md_1$, 故

$$d_1^T \left(\sum_{i=1}^l P_i \otimes P_i + N^{-1}M \right) d_1 = 0,$$

由引理 2.2.2, $\sum_{i=1}^l P_i \otimes P_i + N^{-1}M$ 正定, 所以 $d_1 = 0$, 则有 $d_3 = 0$. 又由第二式, 有: $H^T d_2 = 0$, 两边同时左乘以 HM , 则 $HMH^T d_2 = 0$. 由于 M 是正定, 且由 H 的定义和集合 $\{A_i, i = 1, \dots, m\}$ 线性无关知, H 是行满秩, 即有 HMH^T 是正定, 因而可推得 $d_2 = 0$, 即得到矛盾. 因此系统 (2.7) 有唯一解.

以下用块高斯消去法给出 $\Delta X, \Delta \lambda, \Delta Z$ 的表示式.

由系统 (2.7), 即

$$\begin{cases} H\text{vec}(\Delta X) = r_p & \textcircled{1} \\ -\sum_{i=1}^l (P_i \otimes P_i)\text{vec}(\Delta X) + H^T \Delta \lambda + \text{vec}(\Delta Z) = r_d & \textcircled{2} \\ M\text{vec}(\Delta X) + N\text{vec}(\Delta Z) = r_c & \textcircled{3} \end{cases}$$

③ 式 $\times (-N^{-1}) +$ ② 式, 得到:

$$-(\sum_{i=1}^l (P_i \otimes P_i) + N^{-1}M)\text{vec}(\Delta X) + H^T \Delta \lambda = r_d - N^{-1}r_c \quad \textcircled{4}$$

④ 式 $\times (N)$, 得到:

$$-(N \sum_{i=1}^l (P_i \otimes P_i) + M)\text{vec}(\Delta X) + NH^T \Delta \lambda = Nr_d - r_c \quad \textcircled{5}$$

⑤ 式 $\times H(N \sum_{i=1}^l (P_i \otimes P_i) + M)^{-1}$, 得到:

$$-H\text{vec}(\Delta X) + H(N \sum_{i=1}^l (P_i \otimes P_i) + M)^{-1}NH^T \Delta \lambda = H(N \sum_{i=1}^l (P_i \otimes P_i) + M)^{-1}(Nr_d - r_c) \quad \textcircled{6}$$

⑥ 式 $+ \textcircled{1}$ 式, 即可推得 $\Delta \lambda$ 的表示式:

$$[H(N \sum_{i=1}^l (P_i \otimes P_i) + M)^{-1}NH^T] \Delta \lambda = r_p + H(N \sum_{i=1}^l (P_i \otimes P_i) + M)^{-1}(Nr_d - r_c)$$

由 ⑥ 式, 可推得 ΔX 的表示式:

$$\Delta X = \text{mat}[(N \sum_{i=1}^l (P_i \otimes P_i) + M)^{-1}(NH^T \Delta \lambda - Nr_d + r_c)]$$

由 ② 式, 可推得 ΔZ 的表示式:

$$\Delta Z = \text{mat}[r_d - H^T \Delta \lambda + \sum_{i=1}^l (P_i \otimes P_i)\text{vec}(\Delta X)]$$

基于 (QSDP) 问题的上述分析及定理 2.2.3, 以下给出 (QSDP) 基于 NT 方向短步长 (Short Step) 路径跟踪算法框架:

算法 2.2.1. 基于 NT 方向短步长 (*Short Step*) 路径跟踪算法

给定 $L > 1$ 和一严格初始可行点 (X_0, λ_0, Z_0) , $\mu_0 = \frac{\langle X_0, Z_0 \rangle}{n}$.

Step 1: 计算 $W_k = X_k^{\frac{1}{2}}(X_k^{\frac{1}{2}}Z_kX_k^{\frac{1}{2}})^{-\frac{1}{2}}X_k^{\frac{1}{2}}$; [16, 26]

Step 2: 根据定理 2.2.3 中 $(\Delta X, \Delta \lambda, \Delta Z)$ 的表示式求出系统 (2.7) 解 $(\Delta X_k, \Delta \lambda_k, \Delta Z_k)$;

Step 3: 选择参数 α, β 使得 $X_k + \alpha \Delta X_k, Z_k + \beta \Delta Z_k$ 正定;

Step 4: 更新迭代点及 μ_k :

$$X_{k+1} = X_k + \alpha \Delta X_k, \lambda_{k+1} = \lambda_k + \alpha \Delta \lambda_k, Z_{k+1} = Z_k + \beta \Delta Z_k, \mu_{k+1} = \frac{\langle X_{k+1}, Z_{k+1} \rangle}{n}$$

Step 5: 当 μ_{k+1} 满足 $\mu_{k+1} \leq 2^{-L}\mu_0$ 时, 迭代中止. 否则转 Step1.

由于在最优解处, 须保证 $XZ = 0$, 因此通过更新序列 $\mu_k = \frac{\langle X_k, Z_k \rangle}{n}$, 当 $\mu_k \rightarrow 0$, 就可导出原问题和对偶问题的最优解 (X^*, λ^*, Z^*) .

从二次半定规划的原始对偶内点算法的讨论中, 我们可以发现, 在寻找新的搜索方向 $(\Delta X, \Delta \lambda, \Delta Z)$ 时, 使用 *Newton* 方法来求解 *KKT* 系统需要对互补松弛条件进行对称化, 才能保证搜索方向 $(\Delta X, \Delta \lambda, \Delta Z) \in S^n \times R^m \times S^n$, 这不仅增加了计算上的复杂度, 而且让系统逐渐趋于不稳定. 基于此, Serge Kruk 等人在文献 [45] 提出了用于解决线性半定规划的 *Gauss Newton* 方向 (*GN direction*), 本质上它是将 *Gauss Newton* 方法应用于一个由最优性条件构成的最小二乘问题来获得新的搜索方向, 等价于求线性扰动最优性条件的最小二乘解. 数值求解上, 它比用 *Newton* 方法更稳定, 且避免了互补松弛条件对称化的过程. 因此, 我们是否可以考虑将求解线性半定规划的 *Gauss Newton* 方法推广至求解二次半定规划, 从而直接获得 $S^n \times R^m \times S^n$ 上对称的搜索方向 $(\Delta X, \Delta \lambda, \Delta Z)$.

2.3 数值试验

考虑 $l = 2, n = 3, m = 2$; 控制精度 $eps = 1e - 5$. 取矩阵 $C = I_n$

$$P_1 = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 3 & 0 \\ 1 & 0 & 4 \end{pmatrix}; P_2 = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 3 & 1 \\ 0 & 1 & 4 \end{pmatrix};$$

$$A_1 = \begin{pmatrix} 3 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 1 & 3 \end{pmatrix}; A_2 = \begin{pmatrix} 1 & -2 & 3 \\ -2 & 2 & -1 \\ 3 & -1 & 2 \end{pmatrix}; b = \begin{pmatrix} 8 \\ 8 \end{pmatrix};$$

在算法实现上, 需给定一严格初始可行点, 不妨取

$$X_0 = \begin{pmatrix} 5 & 1 & 2 \\ 1 & 2 & -1 \\ 2 & -1 & 31 \end{pmatrix}; \lambda_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}; Z_0 = \begin{pmatrix} 4 & -3 & -1 \\ -3 & 5 & 2 \\ -1 & 2 & 3 \end{pmatrix};$$

显然, $X_0 \succ 0, Z_0 \succ 0$, 并在 Matlab 7.0.1 上经基于 NT 方向短步长 (*Short Step*) 路径跟踪算法测试, 可得原始及对偶最优解 X_{opt}, Z_{opt} 及最优 *Lagrangian* 乘子 λ_{opt} :

$$X_{opt} = \begin{pmatrix} 1.4914 & -0.3393 & 0.5499 \\ -0.3393 & 0.5706 & -0.0402 \\ 0.5499 & -0.0402 & 0.2908 \end{pmatrix}; \lambda_{opt} = \begin{pmatrix} 4.3289 \\ 1.5675 \end{pmatrix};$$

$$Z_{opt} = \begin{pmatrix} 0.1357 & 0.0529 & -0.2111 \\ 0.0529 & 0.1532 & -0.0729 \\ -0.2111 & -0.0729 & 0.5331 \end{pmatrix};$$

及最优目标函数值 $Obj_{opt} = 24.6176$.

在图 Fig(1) 中, 很明确地显示了对偶间隙 (dual gap) 与迭代次数 k 的变化关系. 其实, 该解并不是真正意义上的最优解, 因为对偶间隙 $X_{opt} \bullet Z_{opt} = \text{trace}(X \bullet Z) = 0.1826$, 也就是说原问题与对偶问题对偶间隙并没有严格趋于零. 因此, 在算法编程测试上面临的一个问题就是如何控制参数使得对偶间隙尽可能趋向于零. 此外, 如何更好地进行参数控制及其误差控制, 这也是后续需要做的工作.

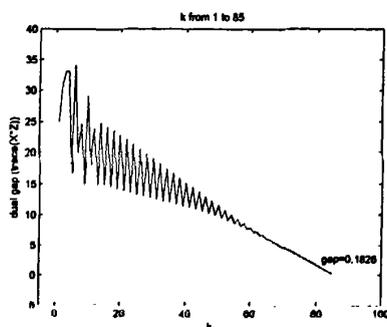


Fig (1)

图 Fig(2), Fig(3) 中显示了解矩阵 X_k, Z_k 的条件数与迭代次数 k 的变化关系, 矩阵条件数定义如下 $Cond(X) = \|X\| \|X^{-1}\|$, 当矩阵范数取 2-范数 $\|\cdot\|_2 =$

$\sqrt{\lambda_{\max}(A^T A)}$ 时, $Cond(X) = \frac{\lambda_{\max}(X)}{\lambda_{\min}(X)}$, 曲线从某一角度 (条件数) 反映了从初值点趋向最优点的条件值点的逼近路径. 在 Fig(3) 中, $Cond(Z_3) = 2.2642 \times 10^3$, $Cond(Z_5) = 1.9630 \times 10^3$, $Cond(Z_9) = 3.1580 \times 10^3$. 显然, 当 $j = 3, 5, 9$ 时, 解矩阵 Z 显得很病态.

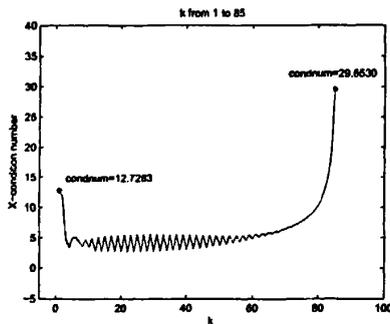


Fig (2)

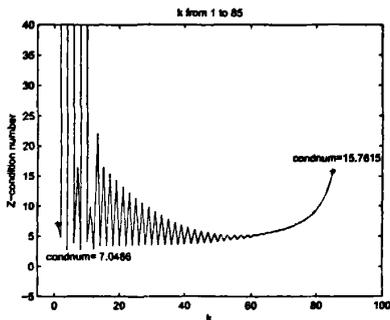


Fig (3)

2.4 与半定最小二乘问题的联系

本节主要讨论二次半定规划目标函数定量参数 l, P 取值不同的一些特殊实例, 可以发现它与半定最小二乘问题有着紧密联系. 半定最小二乘问题, 实际上是计算给定对称矩阵在由半正定矩阵锥和仿射子空间的交构成的闭凸集上的投影. 它在证

券投资风险分析和稳健性二次优化具有很强的应用背景. 其标准模型为

$$\begin{aligned} \min_{X \in S^n} \quad & \frac{1}{2} \|X - C\|_F^2 \\ \text{s.t.} \quad & \mathcal{A}X = b \\ & X \succeq 0 \end{aligned} \tag{SDLS}$$

由二次半定规划 (2.1) 出发, 易得以下结论:

命题 2.4.1. 当 $l = 1, P = I_n$ 时, 二次半定规划 (2.1) 等价于:

$$\begin{aligned} \min_{X \in S^n} \quad & \frac{1}{2} \|X - (-C)\|_F^2 \\ \text{s.t.} \quad & \mathcal{A}X = b \\ & X \succeq 0 \end{aligned} \tag{2.8}$$

证明: 事实上, 当 $l = 1, P = I_n$ 时, 二次半定规划 (2.1) 对应的目标函数为 $\frac{1}{2}X \bullet X + C \bullet X$, 添上一常数项 $\frac{1}{2}(-C) \bullet (-C)$, 得

$$\frac{1}{2}X \bullet X + C \bullet X + \frac{1}{2}(-C) \bullet (-C) = \frac{1}{2} \|X - (-C)\|_F^2$$

由此可以发现, 半定最小二乘问题 (2.8) 是二次半定规划 (2.1) 的一个特例.

命题 2.4.2. 当 $l = 1, C = 0, \det(P) \neq 0$ (更强的: $P \succ 0$) 时, 二次半定规划 (2.1) 等价于:

$$\begin{aligned} \min_{Y \in S^n} \quad & \frac{1}{2} \|Y\|_F^2 \\ \text{s.t.} \quad & \overline{A}_i \bullet Y = b_i (i = 1, \dots, m) \\ & Y \succeq 0 \end{aligned} \tag{2.9}$$

其中 $Y = P^{\frac{1}{2}} X P^{\frac{1}{2}}, \overline{A}_i = P^{-\frac{1}{2}} A_i P^{-\frac{1}{2}} (i = 1, \dots, m)$.

证明: 事实上, 当 $l = 1, C = 0$ 时, 可对目标函数进行适当改写

$$\begin{aligned} \frac{1}{2} X \bullet P X P &= \frac{1}{2} \text{tr}(X P X P) \\ &= \frac{1}{2} \text{tr}(X P X P^{\frac{1}{2}} P^{\frac{1}{2}}) \\ &= \frac{1}{2} \text{tr}(P^{\frac{1}{2}} X P X P^{\frac{1}{2}}) \\ &= \frac{1}{2} \text{tr}(P^{\frac{1}{2}} X P^{\frac{1}{2}} P^{\frac{1}{2}} X P^{\frac{1}{2}}) \\ &= \frac{1}{2} \|P^{\frac{1}{2}} X P^{\frac{1}{2}}\|_F^2 \end{aligned}$$

令 $Y = P^{\frac{1}{2}} X P^{\frac{1}{2}}$, 因为 $X \succeq 0, \det(P) \neq 0$, 由引理 2.1.2, 可知 $Y \succeq 0$. 令 $\bar{A}_i = P^{-\frac{1}{2}} A_i P^{-\frac{1}{2}} (i = 1, \dots, m)$, 不难验证 $\bar{A}_i \bullet Y = A_i \bullet X = b_i$. ■

可以发现, (2.9) 是一类更特殊、更简单的半定最小二乘问题.

在命题 2.4.2, 我们考虑了 $l = 1$ 的情况, 倘若对 l 不加限制, 我们有:

命题 2.4.3. 当 $C = 0, \det(P_i) \neq 0$ (更强的: $P_i \succ 0, i = 1, \dots, l$) 时, 二次半定规划 (2.1) 等价于:

$$\begin{aligned} \min_{Y_i \in S^n} \quad & \frac{1}{2} \sum_{i=1}^l \|Y_i\|_F^2 \\ \text{s.t.} \quad & \bar{A}_i \bullet Y_i = b_i (i = 1, \dots, m) \\ & Y_i \succeq 0 \end{aligned} \quad (2.10)$$

其中 $Y_i = P_i^{\frac{1}{2}} X P_i^{\frac{1}{2}} (i = 1, \dots, l)$.

证明: 类似于命题 2.4.2 的计算, 有:

$$\frac{1}{2} X \bullet \sum_{i=1}^l P_i X P_i = \frac{1}{2} \sum_{i=1}^l \|P_i^{\frac{1}{2}} X P_i^{\frac{1}{2}}\|_F^2$$

令 $Y_i = P_i^{\frac{1}{2}} X P_i^{\frac{1}{2}} (i = 1, \dots, l)$, 因为 $X \succeq 0, \det(P_i) \neq 0$, 同样由引理 2.1.2, 可知 $Y_i \succeq 0$.

若 $l \geq m$ 令 $\bar{A}_i = P_i^{-\frac{1}{2}} A_i P_i^{-\frac{1}{2}} (i = 1, \dots, m)$, 容易验证 $\bar{A}_i \bullet Y_i = A_i \bullet X = b_i$.

若 $l < m$, 可令 $\bar{A}_i = P_i^{-\frac{1}{2}} A_i P_i^{-\frac{1}{2}}, Y_i = P_i^{\frac{1}{2}} X P_i^{\frac{1}{2}} (i = 1, \dots, l); \bar{A}_j = P_j^{-\frac{1}{2}} A_j P_j^{-\frac{1}{2}}, Y_j = P_j^{\frac{1}{2}} X P_j^{\frac{1}{2}} (j = l+1, \dots, m), j \in \{1, \dots, l\}$. 同样易验证 $\bar{A}_i \bullet Y_i = A_i \bullet X = b_i$. ■

注 2.4.1. 针对命题 2.4.3 所讨论的二次半定规划, 转化后的规划形式上类似与半定最小二乘问题, 但并非是严格意义上的半定最小二乘问题, 因此, 我们可进一步探讨 (2.10) 解集与如下一系列简化半定最小二乘问题解集之间关系.

$$\begin{aligned} \min_{Y_1 \in S^n} \quad & \frac{1}{2} \|Y_1\|_F^2 & \min_{Y_2 \in S^n} \quad & \frac{1}{2} \|Y_2\|_F^2 & \min_{Y_l \in S^n} \quad & \frac{1}{2} \|Y_l\|_F^2 \\ \text{s.t.} \quad & \bar{A}_1 \bullet Y_1 = b_1, & \text{s.t.} \quad & \bar{A}_2 \bullet Y_2 = b_2, \dots, & \text{s.t.} \quad & \bar{A}_l \bullet Y_l = b_l \\ & Y_1 \succeq 0 & & Y_2 \succeq 0 & & Y_l \succeq 0 \end{aligned}$$

综上所述, 可以发现, 当取定二次半定规划目标函数中的定量参数 l, P, C 时, 就对应着不同类型的半定最小二乘问题. 而对于这些情况, 我们不仅可以用原始对偶内点算法来求解, 还可从半定最小二乘问题特殊结构出发, 去构造相应的算法进行求解. 在下一章节中, 我们将针对这一问题展开讨论, 并对 (SDLS) 进行必要的推广.

第 3 章 变量有界半定最小二乘问题及其求解 算法

在本章中,从二次半定规划问题出发,继续深入探讨第 2 章中二次半定规划问题的特殊形式,即半定最小二乘问题 (SDLS),并在此基础上,对 (SDLS) 模型的定义进行推广,提出了比 (SDLS) 更一般化的变量有界半定最小二乘问题 (BV-SDLS). 由于目标函数具有平方和这种特殊形式,因此给问题的求解带来了某种方便,对于这类问题,除了能够运用前面介绍的一般求解方法外,还可以考虑一些更为简便有效算法,下面我们围绕该话题展开详细的讨论.

本章主要工作是,给出了任意实对称矩阵 C 在由界约束矩阵变量构成的闭凸集 $\Omega^\beta = \{X \in S^n \mid 0 \preceq X \preceq \beta I_n\}$ 上的精确投影表示式 $P_{\Omega^\beta}(C)$,该式是在半正定矩阵锥上投影的推广,并在此投影基础上,探讨了该 (BV-SDLS) 特殊有效的求解算法,即投影拟牛顿算法,给出了投影拟牛顿算法框架.最后在 MATLAB 7.01 上进行数值试验,并与原始对偶内点算法比较,进一步说明算法的可行性及有效性.

3.1 (变量有界) 半定最小二乘问题

在第二章中,我们已详细讨论了一类二次半定规划问题的对偶性理论及其求解算法,并详细讨论了该规划与半定最小二乘问题的联系,由命题 2.4.1 知,当 $l = 1, P = I_n$ 时,二次半定规划问题可改写为以下标准的半定最小二乘问题 (SDLS):

$$\begin{aligned} \min_{X \in S^n} \quad & \frac{1}{2} \|X - C\|_F^2 \\ \text{s.t.} \quad & AX = b \\ & X \succeq 0 \end{aligned} \quad (\text{SDLS})$$

Malick, Ji^[32] 讨论了该类型问题的求解算法,并简要介绍了 (SDLS) 在仿射算子 A 取 $A = \text{diag}$ 的另一求解算法,即交错投影算法 (*Alternating Projection Method*),同时进一步解释了这种算法实际上是将最速下降梯度算法应用于 (SDLS) 的对偶问题,有关交错投影算法,可参阅 [30].

在第二章第 2.4 节中已简要介绍了半定最小二乘问题模型 (SDLS) 的结构,并指出半定最小二乘问题在证券风险分析的协方差风险矩阵估计等方面同样的广泛应用.投资组合上有一个非常重要的模型,即 Markowitz 模型 [40]. 假设 n 项风险资产 ξ_i , 它们的预期收益率为 $E(\xi_i)(i = 1, \dots, n)$, 彼此之间的协方差为 $\sigma_{ij}(i, j =$

$1, \dots, n)$. x_i 表示相应资产在组合资产中的比重, 由于在中国市场上禁止卖空, 所以 $x_i \geq 0$. 因此, 投资组合 ξ 的预期收益率和方差是 $E(\xi) = \sum_{i=1}^n x_i E(\xi_i), \sigma^2 = \sum_{i=1}^n \sum_{j=1}^n x_i x_j \sigma_{ij}$, 其中 $\sigma_{ij} = \frac{1}{m-1} \sum_{k=1}^m (\xi_{ik} - \bar{\xi}_i)(\xi_{jk} - \bar{\xi}_j)$, m 为所获取的单只股票的观测点数, ξ_{ik} 为第 i 只股票第 k 个观测点上的收益率, $\bar{\xi}_i$ 为第 i 只股票的算术平均数, 即 $\bar{\xi}_i = \frac{1}{m} \sum_{k=1}^m \xi_{ik}$.

Markowitz 模型的基本思想是如何选择资产, 使在期望收益一定的情况下最大限度的减少风险, 或在所能承受的风险水平下最大限度地增加收益, 而投资的风险和收益分别用方差和期望来度量. 借助于风险厌恶因子 ω , Markowitz 模型可表示为单目标优化问题:

$$\begin{aligned} \max \quad & (1 - \omega) \sum_{i=1}^n x_i E(\xi_i) - \omega \sum_{i=1}^n \sum_{j=1}^n x_i x_j \sigma_{ij} \\ \text{s.t.} \quad & \sum_{i=1}^n x_i = 1, x_i \geq 0 \\ & 0 \leq \omega \leq 1 \end{aligned}$$

当 $\omega = 1$, 表示投资者极端保守, 不论收益高低, 首先把风险放在第一位; 当 $\omega = 0$ 时, 表示投资者是绝对冒险者, 只要有一点点收益高的可能, 都愿意冒险去追逐高收益. 现实中的投资者大多数都位于这两者之间. 当然, 许多投资者总希望保证期望收益不低于某个固定值情况下, 使得风险尽可能的小, 对这一类型投资者, Markowitz 模型可表达为:

$$\begin{aligned} \min \quad & \sum_{i=1}^n \sum_{j=1}^n x_i x_j \sigma_{ij} \\ \text{s.t.} \quad & \sum_{i=1}^n x_i E(\xi_i) \geq \beta \\ & \sum_{i=1}^n x_i = 1, x_i \geq 0 \end{aligned}$$

令 $Q = (\sigma_{ij})_{n \times n}$, 矩阵 Q (风险矩阵) 经常被用来度量风险, 因而有

$$\begin{aligned} \min \quad & x^T Q x \\ \text{s.t.} \quad & \sum_{i=1}^n x_i E(\xi_i) \geq \beta \\ & \sum_{i=1}^n x_i = 1, x_i \geq 0 \end{aligned}$$

在证券管理中, 在 n 项资产中求得理论上的协方差矩阵 (theoretical covariance matrix) 好的近似逼近是很重要的, 真实的风险矩阵 Q 一般情况下是未知的, 样本

相关矩阵中的元素都是通过观测数据得到的. 设 \tilde{Q} 是对协方差矩阵 Q 第一次经验风险估计, \tilde{Q} 的条件数一般是很大的 (ill-conditioned), 为了获得事前风险的一个平稳估计, 需对协方差矩阵进行校正, 因而就有如下协方差矩阵校正问题 [32]:

$$\begin{aligned} \min_{X \in S^n} \quad & \frac{1}{2} \|X - \tilde{Q}\|_F^2 \\ \text{s.t.} \quad & X \succeq \alpha I_n \\ & \langle I_n, X \rangle = \text{tr}(\tilde{Q}) \\ & \langle A_i, X \rangle = \sigma_i^2 \end{aligned}$$

其中, $X \succeq \alpha I_n, \alpha > 0$, 可避免风险系数过低. σ_i^2 表示恰当投资组合的事后波动.

在实际应用中, 也经常会遇到变量有界的规划问题, 基于这样的事实, 本章对 (SDLS) 进行推广, 进一步强化半正定矩阵锥约束, 讨论如下变量有界的二次半定规划:

$$\begin{aligned} \min_{X \in S^n} \quad & \frac{1}{2} \|X - C\|_F^2 \\ \text{s.t.} \quad & AX = b \\ & 0 \preceq X \preceq \beta I_n \end{aligned} \quad (BV - SDLS)$$

其中 $\beta \in R_{++}, I_n$ 是 n 阶单位阵. 我们不妨称该问题为: 变量有界半定最小二乘问题, 并简记为: BV-SDLS.

显然, 当 $\beta \rightarrow +\infty$ 时, 问题 (BV-SDLS) 就是标准的半定最小二乘问题 (SDLS). 因此, 问题 (BV-SDLS) 是半定最小二乘问题 (SDLS) 的推广. 进一步发现, 问题 (BV-SDLS) 是考虑变量有界约束的二次半定规划问题, 对于变量有界约束, *E.A.Papa Quiroz* 提出了针对目标函数为线性半定规划的障碍函数方法 [39], 同时给出了这种求解算法的收敛性证明. 当然, (BV-SDLS) 同样也是一般的半定规划问题, 且已有较为成熟有效的算法, 即原始对偶内点算法, 只是这种算法对于大规模问题, 即矩阵维数 n 足够大或数据矩阵异常稠密时显得有点力不从心. 在后续章节, 我们将详细讨论该 (BV-SDLS) 的另一求解算法. 在此之前, 先讨论该算法的理论基础.

3.2 闭凸矩阵集上的投影

在半正定矩阵锥 S_+^n 上的投影等价于求矩阵 $X \in S_+^n$ 使得

$$X = \arg \min_{X \in S_+^n} \|X - C\|_F \triangleq P_{S_+^n}(C)$$

我们知道, 对于任意矩阵 $C \in S^n$, 存在一正交矩阵 P_C , 使得 $C = P_C \text{Diag}\{\lambda(C)\} P_C^T$. 文献 [31, 32] 给出了 $C \in S^n$ 在半正定矩阵锥 S_+^n 上的投影, 即

$$P_{S_+^n}(C) = P_C \text{Diag}\{\max\{0, \lambda(C)\}\} P_C^T$$

为了讨论变量有界半定最小二乘问题 (BV-SDLS) 的求解算法, 现对半正定矩阵锥进行必要的扩广.

定义集合 $\Omega^\beta = \{X \in S^n \mid 0 \preceq X \preceq \beta I_n\}$, 其中 β 是任意给定, 且 $\beta > 0$. 特别地, 当 $\beta \rightarrow +\infty$ 时, 有 $\Omega^\beta = S_+^n$. 本节主要讨论对称矩阵 $C \in S^n$ 在一般集合 Ω^β 上的投影, 并给出 C 在 Ω^β 的投影表达式. 在此之前, 先给出几个有用的引理, 可参阅 [42, 47, 48].

引理 3.2.1. [42] $A, B \in S^n$, 若 $A \preceq B$, 则 $\lambda(A) \leq \lambda(B)$, $\text{tr}(A) \leq \text{tr}(B)$.

引理 3.2.2. [42] 设 A, E 均为 n 阶 Hermite 矩阵, $B = A + E$, 且 A, B, E 的特征值分别为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$; $\mu_1 \geq \mu_2 \geq \dots \geq \mu_n$; $\varepsilon_1 \geq \varepsilon_2 \geq \dots \geq \varepsilon_n$, 则对任意 $i \in \{1, \dots, n\}$, 有 $\lambda_i + \varepsilon_n \leq \mu_i \leq \lambda_i + \varepsilon_1$.

引理 3.2.3. [47] 设 C 是 R^n 中的一个非空凸集, $f: C \rightarrow R$ 是一凸函数, $g: R \rightarrow R$ 是包含集值 $\{f(x) \mid x \in C\}$ 的单调递增凸函数, 则复合函数 $h(x) = g(f(x))$ 在集合 C 上是凸函数. 此外, 如果 g 是单调递增, 且 f 是严格凸函数, 则 h 是严格凸的.

引理 3.2.4. [48] 设 Δ 是某一集合, $f(x) = \sup\{\phi(y, x) \mid y \in \Delta\}$. 假定对任意给定的 $y \in \Delta$, 函数 $\phi(y, x)$ 关于 x 是闭凸函数, 则 $f(x)$ 在定义域 $\text{dom } f = \{x \in \bigcap_{y \in \Delta} \text{dom } \phi(y, \cdot) \mid \exists \gamma: \phi(y, x) \leq \gamma, \forall y \in \Delta\}$ 是闭凸函数.

引理 3.2.5. $A \in S^n$, 当 $\beta \geq \lambda_{\max}(A)$ 时, $\beta I_n - A$ 是半正定矩阵. 其中 $\lambda_{\max}(A)$ 是 A 的最大特征值.

证明: $A \in S^n$, 则存在正交矩阵 Q , 使得 $Q^T A Q = \Lambda = \text{Diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$. 不妨设 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$, 易知 $Q^T(\beta I_n - A)Q = \beta I_n - \Lambda = \text{Diag}\{\beta - \lambda_1, \beta - \lambda_2, \dots, \beta - \lambda_n\}$. 由题设 $\beta \geq \lambda_{\max}(A) = \lambda_1$, 因此 $\beta \geq \lambda_i (i = 1, \dots, n)$. 故 $\beta - \lambda_i \geq 0$, 即 $\beta I_n - A$ 是半正定. ■

由上述引理, 容易得到以下几个命题, 这些命题是讨论对称矩阵 $C \in S^n$ 在集合 $\Omega^\beta = \{X \in S^n \mid 0 \preceq X \preceq \beta I_n\}$ 上投影的基础.

命题 3.2.6. 若 $\beta > 0, X \in S^n, 0 \preceq X \preceq \beta I_n$ 当且仅当 $0 \leq \lambda(X) \leq \beta e$.

证明: " \Rightarrow " 因为 $0 \preceq X \preceq \beta I_n$, 显然 $\lambda(X) \geq 0, \lambda(\beta I_n - X) \geq 0$. 由引理 3.2.2 可知, $\lambda(X) - \beta e \leq \lambda(X - \beta I_n) \leq \lambda(X) - \beta e$, 故 $\lambda(X - \beta I_n) = \lambda(X) - \beta e$, 即有 $\lambda(\beta I_n - X) = \beta e - \lambda(X) \geq 0$. 因此 $0 \leq \lambda(X) \leq \beta e$ 成立.

" \Leftarrow " 由引理 3.2.5 易得. ■

命题 3.2.7. 集合 $\Omega^\beta = \{X \in S^n \mid 0 \preceq X \preceq \beta I_n\}$ 是闭凸集.

证明: 任意矩阵 $X, Y \in \Omega^\beta$, 即有 $0 \preceq X \preceq \beta I_n, 0 \preceq Y \preceq \beta I_n$. 易知对任意的 $0 \leq \lambda \leq 1, \lambda X + (1 - \lambda)Y \in \Omega^\beta$. 事实上,

$$\begin{aligned} \beta I_n - [\lambda X + (1 - \lambda)Y] &= [\lambda \beta I_n + (1 - \lambda)\beta I_n] - [\lambda X + (1 - \lambda)Y] \\ &= \lambda(\beta I_n - X) + (1 - \lambda)(\beta I_n - Y) \succeq 0. \end{aligned}$$

闭性由矩阵特征值函数 $\lambda_i: S^n \rightarrow R (i = 1, \dots, n)$ 在每一点 $X \in S^n$ 是连续函数 (强半光滑 strongly semismooth [46] 函数) 与命题 3.2.6 易得. ■

注 3.2.1. 进一步地, 集合 $\Omega^\beta = \{X \in S^n \mid 0 \preceq X \preceq \beta I_n\}$ 是尖 (pointed)、实心 (solid). 事实上, $0 < \varepsilon < \beta$, 取 $X_0 = \varepsilon I_n$, 对任意 $A \in S_n$, 有

$$\varepsilon I_n + \frac{A}{\frac{2\max\{|\lambda_{\max}(A)|, |\lambda_{\min}(A)|\}}{\varepsilon}} \succ 0, (\beta - \varepsilon)I_n + \frac{A}{\frac{2\max\{|\lambda_{\max}(A)|, |\lambda_{\min}(A)|\}}{\beta - \varepsilon}} \succ 0.$$

则有 $0 \prec X_0 \prec \beta I_n$, 即 $X_0 \in \text{int } \Omega^\beta$. 因此, Ω^β 是满维 (full dimensional) 的闭凸集.

命题 3.2.8. 设 $A, B \in S^n$, 且 $A \succeq 0$, 则 $\text{tr}(AB) \leq \lambda_{\max}(B)\text{tr}(A)$.

证明: 因为 $A \succeq 0$, $\text{tr}(AB) = \text{tr}(A^{\frac{1}{2}}BA^{\frac{1}{2}})$, 由引理 3.2.5 可推得,

$$\begin{aligned} &\lambda_{\max}(B)A - A^{\frac{1}{2}}BA^{\frac{1}{2}} \\ &= \lambda_{\max}(B)A^{\frac{1}{2}}A^{\frac{1}{2}} - A^{\frac{1}{2}}BA^{\frac{1}{2}} \\ &= A^{\frac{1}{2}}(\lambda_{\max}(B)I - B)A^{\frac{1}{2}} \succeq 0 \end{aligned}$$

故由引理 3.2.1, 有 $\text{tr}(A^{\frac{1}{2}}BA^{\frac{1}{2}}) \leq \lambda_{\max}(B)\text{tr}(A)$, 因此 $\text{tr}(AB) \leq \lambda_{\max}(B)\text{tr}(A)$. ■

引理 3.2.9. (投影定理)^[47] 设 F 为 R^n 中非空闭凸集.

(a) 对任意 $x \in R^n$, 存在 F 中唯一向量使得 $\|z - x\|$ 取得最小值. 该向量称为 x 在

F 上的投影. 记为 $P_F(x)$, $P_F(x) = \arg \min_{z \in F} \|z - x\|$.

(b) 对任意 $x \in R^n, z \in F$:

$$z = P_F(x) \Leftrightarrow \forall y \in F, (y - z)'(x - z) \leq 0.$$

(c) 函数 $P_F: R^n \rightarrow F$ 是连续且是非扩张的, 即: $\|P_F(y) - P_F(x)\| \leq \|y - x\|, \forall x, y \in R^n$. (有时候将这个不等式形象地说成“投影不增大距离”)

(d) 距离函数 $d: R^n \rightarrow R, d(x, F) = \min_{z \in F} \|z - x\|$ 是关于变量 x 的凸函数.

记实对称矩阵 C 在闭凸集 $\Omega^\beta = \{X \in S^n | 0 \preceq X \preceq \beta I_n\}$ 上的投影

$$P_{\Omega^\beta}(C) = \arg \min_{X \in \Omega^\beta} \|X - C\|_F,$$

C 到闭凸集 Ω^β 的半平方距离 $d_{\Omega^\beta}(C) = \min_{X \in \Omega^\beta} \frac{1}{2} \|X - C\|_F^2$.

定理 3.2.10. 设 $C \in S^n$, 则矩阵 C 在集合 Ω^β 上的投影 $P_{\Omega^\beta}(C) = C_+^\beta$, 且

$$d_{\Omega^\beta}(C) = \frac{1}{2} \left(\sum_{\lambda_i > \beta} (\lambda_i - \beta)^2 + \sum_{\lambda_i < 0} \lambda_i^2 \right).$$

其中 $C_+^\beta = P_C \text{Diag}\{\max\{0, \min\{\beta e, \lambda(C)\}\}\} P_C^T, \lambda_i$ 是实对称矩阵 C 的特征值.

证明: 易验证 $C_+^\beta \in \Omega^\beta$. 事实上,

$$\{\max\{0, \min\{\beta e, \lambda(C)\}\}\}_i = \begin{cases} 0, & \lambda_i(C) < 0 \\ \lambda_i(C), & 0 \leq \lambda_i(C) \leq \beta \\ \beta, & \lambda_i(C) > \beta \end{cases}$$

故 $0 \leq \max\{0, \min\{\beta e, \lambda(C)\}\} \leq \beta e$. 由命题 3.2.6 易知, $0 \leq C_+^\beta \preceq \beta I_n$.

欲证明 $P_{\Omega^\beta}(C) = C_+^\beta$, 由引理 3.2.9 可知, 只须证明对任意矩阵 $X \in \Omega^\beta$, 满足

$$(X - C_+^\beta) \bullet (C - C_+^\beta) \leq 0$$

对于任意矩阵 $C \in S^n$, 存在正交矩阵 P_C , 使得 $C = P_C \text{Diag}\{\lambda_1, \dots, \lambda_n\} P_C^T$. 进一步假设 C 的这 n 个特征值中有 r 个大于 β , 不妨设为 $\lambda_1, \lambda_2, \dots, \lambda_r$, 即 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > \beta$; 有 s 个介于 0 与 β 之间, 不妨设为 $\lambda_{r+1}, \dots, \lambda_{r+s}$, 即 $\beta \geq \lambda_{r+1} \geq \dots \geq \lambda_{r+s} \geq 0$; 有 l 个小于 0, 不妨设为 $\lambda_{r+s+1}, \dots, \lambda_n$, 即 $0 > \lambda_{r+s+1} \geq \dots \geq \lambda_n$; 且 $r + s + l = n$.

依 C_+^β 的构造, 可知: $C_+^\beta = P_C \text{Diag}\{\underbrace{\beta, \dots, \beta}_r, \underbrace{\lambda_{r+1}, \dots, \lambda_{r+s}}_s, \underbrace{0, \dots, 0}_l\} P_C^T$.

故 $C - C_+^\beta = P_C \text{Diag}\{\underbrace{\lambda_1 - \beta, \dots, \lambda_r - \beta}_r, \underbrace{0, \dots, 0}_s, \underbrace{\lambda_{r+s+1}, \dots, \lambda_n}_l\} P_C^T$. 从而我们有:

$$C_+^\beta \cdot (C - C_+^\beta) = \beta(\lambda_1 - \beta) + \dots + \beta(\lambda_r - \beta) = \beta\lambda_1 + \dots + \beta\lambda_r - r\beta^2.$$

而

$$\begin{aligned} X \cdot (C - C_+^\beta) &= X \cdot P_C \text{Diag}\{\underbrace{\lambda_1 - \beta, \dots, \lambda_r - \beta}_r, \underbrace{0, \dots, 0}_s, \underbrace{\lambda_{r+s+1}, \dots, \lambda_n}_l\} P_C^T \\ &= X \cdot P_C \text{Diag}\{\underbrace{\lambda_1 - \beta, \dots, \lambda_r - \beta}_r, \underbrace{0, \dots, 0}_s, \underbrace{0, \dots, 0}_l\} P_C^T \\ &\quad + X \cdot P_C \text{Diag}\{\underbrace{0, \dots, 0}_r, \underbrace{0, \dots, 0}_s, \underbrace{\lambda_{r+s+1}, \dots, \lambda_n}_l\} P_C^T \end{aligned}$$

由命题 3.2.8 和 $-\beta \leq \lambda(-X) \leq 0, -(\lambda_{r+s+1} + \dots + \lambda_n) > 0$, 知

$$\begin{aligned} &X \cdot P_C \text{Diag}\{\underbrace{\lambda_1 - \beta, \dots, \lambda_r - \beta}_r, \underbrace{0, \dots, 0}_s, \underbrace{0, \dots, 0}_l\} P_C^T \\ &\leq \lambda_{\max}(X)(\lambda_1 + \dots + \lambda_r - r\beta) \\ &\leq \beta(\lambda_1 + \dots + \lambda_r - r\beta) \\ &= \beta(\lambda_1 + \dots + \lambda_r) - r\beta^2 \\ &X \cdot P_C \text{Diag}\{\underbrace{0, \dots, 0}_r, \underbrace{0, \dots, 0}_s, \underbrace{\lambda_{r+s+1}, \dots, \lambda_n}_l\} P_C^T \\ &= (-X) \cdot P_C \text{Diag}\{\underbrace{0, \dots, 0}_r, \underbrace{0, \dots, 0}_s, \underbrace{-\lambda_{r+s+1}, \dots, -\lambda_n}_l\} P_C^T \\ &\leq \lambda_{\max}(-X)(-\lambda_{r+s+1} - \dots - \lambda_n) \leq 0 \end{aligned}$$

因此, 我们有 $X \cdot (C - C_+^\beta) - C_+^\beta \cdot (C - C_+^\beta) \leq \beta(\lambda_1 + \dots + \lambda_r) - r\beta^2 - [\beta\lambda_1 + \dots + \beta\lambda_r - r\beta^2] = 0$. 故对 $\forall X \in \Omega^\beta$, 有 $(X - C_+^\beta) \cdot (C - C_+^\beta) \leq 0$ 成立, 即 $P_{\Omega^\beta}(C) = C_+^\beta$. 且

$$\begin{aligned} d_{\Omega^\beta}(C) &= \min_{X \in \Omega^\beta} \frac{1}{2} \|X - C\|_F^2 \\ &= \frac{1}{2} \|P_{\Omega^\beta}(C) - C\|_F^2 \\ &= \frac{1}{2} \|C_+^\beta - C\|_F^2 \\ &= \frac{1}{2} \left(\sum_{\lambda_i > \beta} (\lambda_i - \beta)^2 + \sum_{\lambda_i < 0} \lambda_i^2 \right). \end{aligned}$$

注 3.2.2.

$$\begin{aligned} C_+^\beta &= P_C \text{Diag}\{\max\{0, \min\{\beta e, \lambda(C)\}\}\} P_C^T \\ &= P_C \text{Diag}\{\min\{\max\{0, \lambda(C)\}, \beta e\}\} P_C^T \end{aligned}$$

当 $\beta \rightarrow +\infty$ 时

$$C_+^{+\infty} = P_C \text{Diag}\{\max\{0, \lambda(C)\}\} P_C^T$$

显然, 这就是文献 [32] 中给出的矩阵 $C \in S^n$ 在半正定锥 S_+^n 上的投影. 因此, 我们给出的实对称矩阵 C 在闭凸集 $\Omega^\beta = \{X \in S^n | 0 \preceq X \preceq \beta I_n\}$ 上的投影 C_+^β 是其在半正定锥 S_+^n 上投影的简易拓展.

注 3.2.3. 记

$$C_\beta = \sum_{\lambda_i > \beta} (\lambda_i - \beta) q_i q_i^T, C_+^\beta = \sum_{0 < \lambda_i \leq \beta} \lambda_i q_i q_i^T + \sum_{\lambda_i > \beta} \beta q_i q_i^T, C_- = \sum_{\lambda_i < 0} -\lambda_i q_i q_i^T$$

$C = \sum_{i=1}^n \lambda_i q_i q_i^T$ 是 C 的特征值分解, q_1, \dots, q_n 是 C 中特征值 $\lambda_1, \dots, \lambda_n$ 对应的特征向量. 显然, $C_\beta \succeq 0, 0 \preceq C_+^\beta \preceq \beta I_n, C_- \succeq 0$, 且 $C = C_\beta + C_+^\beta - C_-$. C_+^β 可由 $C + C_- - C_\beta$ 计算.

注 3.2.4. 诚如文献 [32] 所强调的, 实对称矩阵空间 S^n 的范数经常是赋予加权的 Frobenius 范数

$$\|X\|_W = \|W^{\frac{1}{2}} X W^{\frac{1}{2}}\|, W \in S_{++}^n$$

在这样的加权内积下, 同样易得矩阵 $C \in S^n$ 在 Ω^β 上的投影

$$W^{-\frac{1}{2}} (W^{\frac{1}{2}} C W^{\frac{1}{2}})_+^\beta W^{-\frac{1}{2}}$$

推论 3.2.11. 若 $X \in S(n_1, \dots, n_m)$, 即 $X = \text{Diag}\{X_1, \dots, X_m\}, X_i \in S^{n_i}, i = 1, \dots, m$. 则块对角矩阵 $C = \text{Diag}\{C_1, \dots, C_m\} \in S(n_1, \dots, n_m)$ 在闭凸集 $\{X \in S(n_1, \dots, n_m) | 0 \preceq X \preceq \text{Diag}\{\beta_1 I_1, \dots, \beta_m I_m\}\}$ 上的投影为:

$$C_+^\beta = \begin{pmatrix} (C_1)_+^{\beta_1} & & & \\ & (C_2)_+^{\beta_2} & & \\ & & \dots & \\ & & & (C_m)_+^{\beta_m} \end{pmatrix}$$

其中 $\beta_i > 0, (C_i)_+^{\beta_i} = P_{C_i} \text{Diag}\{\max\{0, \min\{\beta_i e, \lambda(C_i)\}\}\} P_{C_i}^T, i = 1, \dots, m$.

证明: 因为 $X = \text{Diag}\{X_1, \dots, X_m\}, X_i \in S^{n_i}, i = 1, \dots, m$, 结论易由

$$0 \preceq X \preceq \text{Diag}\{\beta_1 I_1, \dots, \beta_m I_m\} \Leftrightarrow 0 \preceq X_i \preceq \beta_i I_i, i = 1, \dots, m$$

及定理 3.2.10 推得. ■

推论 3.2.12. 对称矩阵 $C \in S^n$ 在集合 $\Omega_\alpha^\beta = \{X \in S^n | \alpha I_n \preceq X \preceq \beta I_n\}$ ($\alpha < \beta$) 上的投影

$$P_{\Omega_\alpha^\beta}(C) = P_C \text{Diag}\{\max\{\alpha e, \min\{\beta e, \lambda(C)\}\}\} P_C^T$$

3.3 投影拟牛顿算法

二次半定规划 (BV-SDLS) 中的约束条件是由仿射约束 $\{X \in S^n | \mathcal{A}X = b\}$ 与闭凸集约束 $\{X \in S^n | 0 \preceq X \preceq \beta I_n\}$ 构成. 在此只对仿射约束进行部分对偶. 设其 Lagrangian 对偶函数为 $L(X, y)$, 则:

$$\begin{aligned} L(X, y) &= \frac{1}{2} \|X - C\|_F^2 - y^T (\mathcal{A}X - b) \\ &= \frac{1}{2} \|X - C\|_F^2 - \langle \mathcal{A}^* y, X \rangle + y^T b \\ &= \frac{1}{2} \|X - C\|_F^2 - \frac{1}{2} \|\mathcal{A}^* y\|_F^2 - \langle C, \mathcal{A}^* y \rangle + y^T b \\ &= \frac{1}{2} \|X - (C + \mathcal{A}^* y)\|_F^2 - \frac{1}{2} \|C + \mathcal{A}^* y\|_F^2 + \frac{1}{2} \|C\|_F^2 + y^T b \end{aligned}$$

定义对偶函数 $\eta(y) = \min_{X \in \Omega^\beta} L(X, y)$, 由上式易知 $L(X, y)$ 在 $X(y) = P_{\Omega^\beta}(C + \mathcal{A}^* y)$ 处取得最小值. 故对偶函数

$$\begin{aligned} \eta(y) &= \min_{X \in \Omega^\beta} L(X, y) \\ &= d_{\Omega^\beta}(C + \mathcal{A}^* y) - \frac{1}{2} \|C + \mathcal{A}^* y\|_F^2 + \frac{1}{2} \|C\|_F^2 + y^T b \end{aligned}$$

由引理 3.2.9 中的 (d), 可知距离函数 $d: S^n \rightarrow R, d(Y, \Omega^\beta) = \min_{X \in \Omega^\beta} \|X - Y\|$ 是关于变量 Y 的凸函数, 且 $g(t) = t^2$ 在区间 $[0, +\infty)$ 上是单调递增凸函数, 而 $d_{\Omega^\beta}(Y) = d(Y, \Omega^\beta)^2$, 直接由引理 3.2.3 可知 $d_{\Omega^\beta}(Y) = \min_{X \in \Omega^\beta} \frac{1}{2} \|X - Y\|_F^2$ 是关于 $Y \in S^n$ 的可微凸函数, 由凸性的仿射不变性, 易知 $d_{\Omega^\beta}(C + \mathcal{A}^* y)$ 是关于 y 的可微凸函数, 且:

$$\begin{aligned} \nabla_y d_{\Omega^\beta}(C + \mathcal{A}^* y) &= \nabla_y \min_{X \in \Omega^\beta} \frac{1}{2} \|X - (C + \mathcal{A}^* y)\|_F^2 \\ &= \nabla_y \frac{1}{2} \|P_{\Omega^\beta}(C + \mathcal{A}^* y) - (C + \mathcal{A}^* y)\|_F^2 \\ &= \mathcal{A}(C + \mathcal{A}^* y - P_{\Omega^\beta}(C + \mathcal{A}^* y)). \end{aligned}$$

命题 3.3.1. $\eta(y)$ 是可微的凹函数, 且 $\nabla_y \eta(y)$ 是 Lipschitz 连续的. $\nabla_y \eta(y)$ 的表示式为

$$\nabla_y \eta(y) = -\mathcal{A}X(y) + b.$$

证明: 事实上, 由于 $L(X, y) = \frac{1}{2}\|X - C\|_F^2 - y^T(\mathcal{A}X - b)$ 关于 y 是凹函数, Ω^β 是有界闭凸集, 由 $\eta(y) = \min_{X \in \Omega^\beta} L(X, y)$ 的构造及引理 3.2.4, 可知 $\eta(y)$ 关于 y 是凹函数. 又 $\nabla_y \frac{1}{2}\|C + \mathcal{A}^*y\|_F^2 = \mathcal{A}(C + \mathcal{A}^*y)$, 且 \mathcal{A} 是有界线性算子, 故

$$\begin{aligned} \nabla_y \eta(y) &= \nabla_y d_{\Omega^\beta}(C + \mathcal{A}^*y) - \nabla_y \frac{1}{2}\|C + \mathcal{A}^*y\|_F^2 + b \\ &= \mathcal{A}(C + \mathcal{A}^*y - P_{\Omega^\beta}(C + \mathcal{A}^*y)) - \mathcal{A}(C + \mathcal{A}^*y) + b \\ &= -\mathcal{A}(P_{\Omega^\beta}(C + \mathcal{A}^*y)) + b \\ &= -\mathcal{A}X(y) + b. \end{aligned}$$

对任意 $x, y \in R^m$, 有

$$\begin{aligned} \|\nabla \eta(x) - \nabla \eta(y)\| &= \|\mathcal{A}(P_{\Omega^\beta}(C + \mathcal{A}^*x)) + b - (\mathcal{A}(P_{\Omega^\beta}(C + \mathcal{A}^*y)) + b)\| \\ &\leq \|\mathcal{A}\| \|\mathcal{A}^*x - \mathcal{A}^*y\| \\ &\leq \|\mathcal{A}\| \|\mathcal{A}^*\| \|x - y\|. \end{aligned}$$

■

由此可知, 变量有界半定最小二乘问题 (BV-SDLS) 的对偶问题 (dual problem):

$$\begin{aligned} \max \quad & \eta(y) \\ \text{s.t.} \quad & y \in R^m \end{aligned} \tag{3.1}$$

其中, 目标函数 $\eta(y) = d_{\Omega^\beta}(C + \mathcal{A}^*y) - \frac{1}{2}\|C + \mathcal{A}^*y\|_F^2 + \frac{1}{2}\|C\|_F^2 + y^T b$.

注 3.3.1. 显然, 对偶问题 (3.1) 是对函数 $\eta(y)$ 求最大值的无约束优化问题.

命题 3.3.2. 假设 y^* 是对偶问题 (3.1) 的最优解, 则 $X^* = P_{\Omega^\beta}(C + \mathcal{A}^*y^*)$ 是原问题 (BV-SDLS) 的最优解.

证明: 函数 $\eta(y)$ 是可微的凹函数, 则 $\eta(y)$ 在 y^* 的梯度 $\nabla \eta(y^*) = 0$, 则有 $\nabla_y \eta(y) = -\mathcal{A}(P_{\Omega^\beta}(C + \mathcal{A}^*y^*)) + b = 0$, 即 $\mathcal{A}X(y^*) = b$ 成立, 因此 $X(y^*)$ 是原问题 (BV-SDLS) 的可行解. 进一步地,

$$\begin{aligned} \eta(y^*) &= L(X(y^*), y^*) \\ &= \frac{1}{2}\|X(y^*) - C\|_F^2 - y^{*T}(\mathcal{A}X(y^*) - b) \\ &= \frac{1}{2}\|X(y^*) - C\|_F^2 \end{aligned}$$

由弱对偶定理可知, $\eta(y^*) \leq \frac{1}{2}\|X(y^*) - C\|_F^2$, 即 $\eta(y^*)$ 是变量有界半定最小二乘问题 (BV-SDLS) 目标函数的下界. 因此, 在 $X(y^*)$ 处可获得原问题 (BV-SDLS) 的最优解, 且 $X^* = P_{\Omega^\rho}(C + \mathcal{A}^*y^*)$.

3.2 投影拟牛顿算法

由命题 3.3.2 知, 从对偶问题 (3.1) 的最优解出发可获得变量有界半定最小二乘问题 (BV-SDLS) 的最优解, 而对偶问题 (3.1) 是无约束优化问题, 可用一系列无约束优化迭代算法求解, 如牛顿法、最速下降算法、共轭梯度法等^[50], 由于拟牛顿算法是一类公认的解决无约束优化问题较为有效的算法, 因此算法的流程以拟牛顿算法为理论框架.

基于此, 在算法实现上, 从对偶问题 (3.1) 出发, 等价于求目标函数为 $-\eta(y)$ 的无约束最小化问题:

$$\begin{aligned} \min \quad & -\eta(y) \\ \text{s.t.} \quad & y \in R^m \end{aligned} \quad (3.2)$$

因此, 在 Malick.J^[32] 给出的算法的基础上进一步改写, 可给出如下 (BV-SDLS) 问题模型的算法求解框架:

算法 3.3.1. 投影拟牛顿算法给定初始点 $y_0 \in R^m$, 初始 Hessian 逆近似矩阵 $H_0 \in S^m, 0 < \varepsilon < 1$, 通常取 $H_0 = I_n$.

Step 1: 根据定理 3.2.10 的投影显式计算:

$$\begin{aligned} X_k &= P_{\Omega^\rho}(C + \mathcal{A}^*y_k) \\ \nabla(-\eta(y_k)) &= \mathcal{A}X_k - b \triangleq g_k \\ \eta(y_k) &= d_{\Omega^\rho}(C + \mathcal{A}^*y_k) - \frac{1}{2}\|C + \mathcal{A}^*y_k\|_F^2 + \frac{1}{2}\|C\|_F^2 + y_k^T b \end{aligned}$$

Step 2: 若 $\|g_k\| < \varepsilon$, 则迭代中止. 否则计算搜索方向 $d_k = -H_k g_k$

Step 3: 沿方向 d_k 作线搜索, 确定步长 $\alpha_k > 0$, 令 $y_{k+1} = y_k + \alpha_k d_k$. 计算:

$$X_{k+1} = P_{\Omega^\rho}(C + \mathcal{A}^*y_{k+1}), \nabla(-\eta(y_{k+1})) = \mathcal{A}X_{k+1} - b \triangleq g_{k+1}$$

Step 4: 校正 H_k 产生 H_{k+1} , 使得拟牛顿条件成立:

$$H_{k+1}q_k = p_k. \text{ 其中 } q_k = g_{k+1} - g_k, p_k = y_{k+1} - y_k$$

Step 5: $k = k + 1$, 转 Step 2

依此算法迭代, 最终可获得对偶问题 (3.1) 的最优逼近解 y^* , 与此同时可获得变量有界半定最小二乘问题 (BV-SDLS) 的最优解 X^* . 在算法执行上, 需要计算 P_{Ω^s} , d_{Ω^s} , 具体表示式在第二节定理 3.2.10 中已明确给出. 在 *Hessian* 逆矩阵校正上有不同的选择方法, 主要有秩 1 校正, *DFP* 校正, *BFGS* 校正等^[49]. 在后续算法编程测试上, 我们采用了 *DFP* 校正.

3.3 交错投影算法

变量有界半定最小二乘问题 (BV-SDLS) 同样可由交错投影算法进行求解, 换句话说, 可以将用于解决特殊半定最小二乘问题 (SDLS)($A = \text{diag}$) 的交错投影算法直接推广到问题 (BV-SDLS). 当然, 运用此方法的前提是能够明确给出在由约束条件构成闭凸集的投影显式. 为了完整性, 我们简要描述下交错投影算法的基本思想.

假定 C, D 是 R^n 中的闭凸集, P_C, P_D 是在 C, D 上的投影. 选择一初始点 $x_0 \in C$, 然后交错投影到闭凸集 C, D , 即:

$$\begin{aligned} y_k &= P_D(x_k) \\ x_{k+1} &= P_C(y_k) \\ k &= 0, 1, 2, \dots \end{aligned}$$

由此迭代可产生序列 $\{x_k\}, \{y_k\}$, 且 $\{x_k\} \subseteq C, \{y_k\} \subseteq D$. 如果 $C \cap D \neq \emptyset$, 则序列 $\{x_k\}, \{y_k\}$ 收敛到点 $x^* \in C \cap D$. 换句话说, 交错投影算法最终可在 $C \cap D (C \cap D \neq \emptyset)$ 上找到一公共点. 即 $x_k \in C$ 满足: $\text{dist}(x_k, D) \rightarrow 0 (k \rightarrow +\infty)$; $y_k \in D$, 满足: $\text{dist}(y_k, C) \rightarrow 0 (k \rightarrow +\infty)$.

定义集合 $U = \{X \in S^n | AX = b\}, V = \{X \in S^n | 0 \preceq X \preceq \beta I_n\}$. 则变量有界半定最小二乘问题 (BV-SDLS) 可以改写为

$$\begin{aligned} \min_{X \in S^n} \quad & \frac{1}{2} \|X - C\|_F^2 \\ \text{s.t.} \quad & X \in U \cap V \end{aligned}$$

本质上, 交错投影算法的主要步骤就是重复以下交错投影运算:

$$X \leftarrow P_V(P_U(X))$$

由上述分析, 单纯的交错投影只能保证最终迭代序列收敛到 U 与 V 的交点, 而无法保证收敛到优化问题的最优解. 因此, *Dykstra*^[29] 对其做了相应的改进

校正交错投影算法^[29, 32]

矩阵 $C \in S^n$, 初始 $\Delta S_0 = 0, Y_0 = C$, 对 $k = 1, 2, \dots$

Begin

$R_k = Y_{k-1} - \Delta S_{k-1}$ % Dijkstra's correction step

$X_k = P_V(R_k)$ % Projection onto V

$\Delta S_k = X_k - R_k$

$Y_k = P_U(X_k)$ % Projection onto U

End

注 3.3.2. 从算法框架看, 需要明确矩阵 R_k 在 V 上的投影 $P_V(R_k)$ 以及矩阵 X_k 在 U 上的投影 $P_U(X_k)$. 而对于任意矩阵 $C \in S^n$, C 在 V 上的投影显式在本文的定理 3.2.10 中已给出. 而 C 在仿射集 U 上的投影 $P_U(C) = C - \sum_{i=1}^m u_i A_i^{[30]}$, 其中 u_i 满足法方程 $Gu = (A_1 \bullet C - b_1, \dots, A_m \bullet C - b_m)^T$, 矩阵 $G \in S^m$, 且 $G_{ij} = A_i \bullet A_j (i, j = 1, \dots, m)$. 因此, 从理论上说该算法对于解决变量有界半定最小二乘问题 (BV-SDLS) 是切实可行的.

3.4 数值测试

在以上两节中, 我们已经从理论上阐述了投影拟牛顿算法用于求解变量有界半定最小二乘问题 (BV-SDLS) 的理论基础, 本节主要通过实例的编程测试 (实验一 ~ 实验三) 进一步说明算法执行的可行性, 并将其与内点算法得出的结果进行比较.

众所周知, 拟牛顿方法主要通过利用先前迭代的梯度信息逐步建立逼近逆 Hessian 矩阵. 在拟牛顿算法的实现上, 每一次迭代需要对 Hessian 逆矩阵进行校正, 在此我们采用 DFP^[50] 校正公式:

$$H_{k+1} = H_k + \frac{p_k p_k^T}{p_k^T q_k} - \frac{H_k q_k q_k^T H_k}{q_k^T H_k q_k}$$

其中 $q_k = g_{k+1} - g_k, p_k = y_{k+1} - y_k$. DFP 方法是一个实际上广为采用的方法, 它在理论分析和实际应用中都起了很大的作用. 此外, 在实际编程测试中, 经常出现相邻两次迭代导数的差 q_k 为零的情况, 从而使得校正公式的分母为零, 以致程序无法继续进行. 因此, 有必要对拟牛顿算法作进一步改进^[51], 按拟牛顿法的收敛原理只要以一个对称正定矩阵为初始修正矩阵, 算法就会收敛. 因而, 在计算校正公式前, 先对 q_k 进行一次判断, 如果 $\|q_k\|$ 小于一定的值 $\varepsilon_0 = 10^{-3}$, 就以单位矩阵 I_m 为下次迭代的校正矩阵.

步长 α 的选择可以采用精确一维搜索方法, 但实际测试中, 精确一维搜索往往需要花费很大的工作量, 特别是当迭代点远离问题的解时, 精确地求解一个一维子问题通常不是十分有效的. 此外, 在实际中, 很多最优化方法 (牛顿法, 拟牛顿法等), 其收敛速率并不依赖于精确一维搜索. 为了考虑问题的方便, 实验一中我们分析了两种简单的步长选取方式: 定步长 $\alpha_k = 1$ 以及与迭代次数 k 相关的动态步长 $\alpha_k = \frac{k}{k+1}$, 并简要分析了这两种步长的优劣, 在实验二, 三, 的测试中, 直接取 $\alpha_k = \frac{k}{k+1}$.

以下我们对变量有界半定最小二乘问题 (BV-SDLS) 用投影拟牛顿算法测试, 并将其与用内点算法得到的结果进行比较, 而内点算法的测试我们主要基于 *Sedumi*^{[52],[53]} 环境, 直接利用 Lofberg 其相对成熟的软件包 *Yalmip*^{[54],[55]}. *SeDuMi* 是 Self Dual Minimization 的简称, 可以用于求解一些半正定规划问题. *Yalmip* 最初是用于解决 SDP 和 LMIs 的相关问题, 最近几年它的应用越来越广泛, 特别是 *Yalmip* 3 版本, 不仅支持线性规划 (LP), 二次规划 (QP), 二阶锥规划 (SOCP), 半定规划 (SDP), 而且也支持混合整数规划, 含双线性矩阵不等式 (BMI) 的半定规划, 多参数线性规划和二次规划等等. *Yalmip* 最大的特点是会自动识别, 分析用户所定义的问题, 并在分析的基础上选择合适的求解程序. 若没有合适的求解程序可用, *Yalmip* 会尝试着对问题进行转换以便能够合理地解决问题.

实验一说明了投影拟牛顿算法的可行性, 并分析了两种步长选取的优劣. 实验二说明了投影拟牛顿算法在解决问题 (BV-SDLS) 明显优于内点算法, 从而进一步显示了投影拟牛顿算法的有效性. 实验三显示了投影拟牛顿算法迭代次数及 *cpu* 运行时间随矩阵 C 维数增加的变化趋势.

在实际测试中, 取约束中 $\beta = 3$, 初始点 $y_0 = 2e \in R^m$, 算法中止条件:

$$\|\nabla\eta(y_k)\| = \|AX(y_k) - b\| \leq \epsilon = 10^{-5}$$

所有的测试均基于 Windows XP, Intel CPU 1.7GHz, 512M 内存, MATLAB 7.01 的测试环境.

实验一: 令 $n = 3; m = 2$. 取

$$C = \begin{pmatrix} 2 & 1 & 3 \\ 1 & 5 & 2 \\ 3 & 2 & 2 \end{pmatrix}; A_1 = \begin{pmatrix} 3 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 1 & 3 \end{pmatrix}; A_2 = \begin{pmatrix} 1 & -2 & 3 \\ -2 & 2 & -1 \\ 3 & -1 & 2 \end{pmatrix}; b = \begin{pmatrix} 1 \\ 2 \end{pmatrix};$$

取步长 $\alpha_k = \frac{k}{k+1}$, 经投影拟牛顿算法测试得原问题和对偶问题最优解 X_{quasi}, y_{quasi}

$$X_{quasi} = \begin{pmatrix} 0.04736532 & -0.16930559 & 0.00487167 \\ -0.16930559 & 0.60517661 & -0.01741361 \\ 0.00487167 & -0.01741361 & 0.00050106 \end{pmatrix}; y_{quasi} = \begin{pmatrix} -2.54773147 \\ 0.11580902 \end{pmatrix};$$

及对偶目标函数最优值 $\eta_{quasi} = 27.97065234$.

若运用原始对偶内点算法 (Sedumi+Yalmip) 编程求解, 可得最优解 $X_{interior}$ 及目标函数的最优值 $obj_{interior}$

$$X_{interior} = \begin{pmatrix} 0.04736616 & -0.16930699 & 0.00487131 \\ -0.16930699 & 0.60517581 & -0.01741217 \\ 0.00487131 & -0.01741217 & 0.00050098 \end{pmatrix}; obj_{interior} = 27.97065399$$

可以发现, 在控制误差 $\epsilon = 10^{-5}$ 下, 两种算法的数值误差是微乎其微的, 同时也验证了投影拟牛顿算法求解变量有界半定最小二乘问题 (BV-SDLS) 的可行性. 进一步, 当步长 α_k 取常步长 $\alpha_k = 1$ 和动态步长 $\alpha_k = \frac{k}{k+1}$ 时, 对偶函数值 $\eta(y_k)$ 与迭代次数 k 的关系 (Fig 3.1) 及仿射约束剩余值 $\|AX_k - b\|$ 与迭代次数 k 关系 (Fig 3.2) 如下:

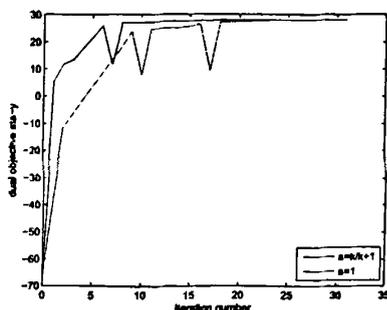


Fig 3.1

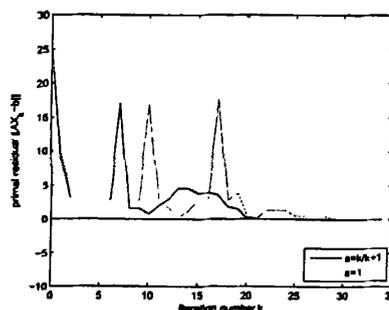


Fig 3.2

由图中不难发现, 当 α_k 取动态步长 $\alpha_k = \frac{k}{k+1}$ 时, 收敛效果会优于定步长 $\alpha_k = 1$, 且原始约束剩余值 $\|AX - b\|$ 在 24 次迭代后减小至 4×10^{-4} . 基于此, 在后续实验二, 三中, 取步长 $\alpha_k = \frac{k}{k+1}$.

实验二:

取 $C = I_n$, 分别考虑 C 的维数 n 取自 5、15、25、35、45, 并取仿射约束中的 A, b 为 $A = \text{diag}$, $b = e$, 即 $\{X \in S^n | X_{ii} = 1, i = 1, \dots, n\}$. 当矩阵 C 遍历上

述五种维数不同单位矩阵时, 分别经投影拟牛顿算法与原始对偶内点算法测试, 可得以下数据:

method	矩阵 C	I_5	I_{15}	I_{25}	I_{35}	I_{45}	I_{100}
$P - quasi$	iter	9	9	9	9	9	9
	cputime(s)	0.2344	0.2656	0.4375	0.4688	0.7031	3.6719
interior	iter	11	14	16	16	17	
	cputime (s)	1.8750	4.5938	16.9688	74.3281	280.0313	

$P - quasi$: 投影拟牛顿算法; interior: 内点算法;

iter: 迭代次数; cputime: cpu运行时间;

容易发现, 对于 C 取 $C = I_n$ 的变量有界半定最小二乘问题 (BV-SDLS), 使用投影拟牛顿算法所需的迭代次数明显少于原始对偶内点算法, 在 cpu 运行时间上, 内点算法呈几何级数上升. 此外, 一个显著的区别是 $n = 45$ 时, 投影拟牛顿算法在 0.7031 秒后即可得到最优解 $X = I_n$, 而内点算法却要花费 280.0313 秒. 特别地, 当 $n \geq 100$ 时, 内点算法的测试显得无能为力, 而投影拟牛顿算法在同样的迭代次数下及少许的 cpu 时间即可得出结论, 进而从另一角度显示了其优越性. Fig 3.3 直观地显示了 C 取遍 $I_n (n = 5, 15, 25, 35, 45)$ 对偶函数 $\eta(y_k)$ 与迭代次数 k 的关系.

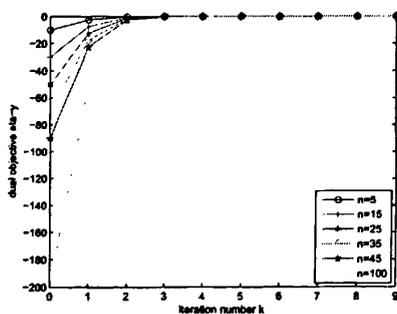


Fig 3.3

实验三:

C 是由计算机随机生成的稠密矩阵, 且 C 中元素满足 $C_{ii} = 1, C_{ij} \in [-1, 1]$. 为便于算法的实现, 仿射约束取实验二中的约束, 特别取 $m = \lfloor \sqrt{n} \rfloor$. 采用投影拟牛顿算法进行测试:

随机矩阵 $C_{n \times n}$	迭代次数	cpu 运行时间 (s)
100 × 100	22	2.1094
200 × 200	29	27.1563
300 × 300	39	137.4219
400 × 400	46	380.2656
500 × 500	47	886.1250
800 × 800	50	4.2294×10^3
1000 × 1000	57	9.3918×10^3
1200 × 1200	62	1.7008×10^4

可以发现, 当稠密矩阵 C 的维数 n 逐渐增大时, 迭代次数及 cpu 时间呈递增趋势. 特别地, 当 $n \geq 500$ 时, 算法运行占用的 cpu 时间增幅较大, 而当矩阵维数 $n \geq 1000$ 时, 算法运行的 cpu 时间较长, 需在 2.6 h (9.3918×10^3 s) 后得该变量有界半定最小二乘问题 (BV-SDLS) 最优解.

注 3.4.1. 特征值分解很大程度上占用了算法执行的 cpu 计算时间. 在投影拟牛顿算法的实现中, 每次迭代都必须对矩阵 $C + A^*y$ 进行特征值分解, 这就限制了 (BV-SDLS) 问题的维数 n 不宜太大.

3.5 小结

在本章节中, 我们考虑了实对称矩阵 C 在闭凸集 $\Omega^\beta = \{X \in S^n | 0 \preceq X \preceq \beta I_n\}$ 上的投影, 给出了其投影显示, 并进一步探讨了如下变量有界半定最小二乘问题 (BV-SDLS) 的求解算法.

$$\begin{aligned} \min_{X \in S^n} \quad & \frac{1}{2} \|X - C\|_F^2 \\ \text{s.t.} \quad & AX = b \\ & 0 \preceq X \preceq \beta I_n \end{aligned}$$

Matlab7.0.1 上的数值试验说明算法执行的可行性及有效性. 进一步地, 在推论 3.2.12, 即

$$P_{\alpha I_n \preceq X \preceq \beta I_n}(C) = P_C \text{Diag}\{\max\{\alpha e, \min\{\beta e, \lambda(C)\}\}\} P_C^T$$

保证下, 该算法同样可有效地求解如下变量上下有界半定最小二乘问题:

$$\begin{aligned} \min_{X \in S^n} \quad & \frac{1}{2} \|X - C\|_F^2 \\ \text{s.t.} \quad & AX = b \\ & \alpha I_n \preceq X \preceq \beta I_n \\ & 0 < \alpha < \beta \end{aligned}$$

当然, 算法的复杂度分析以及算法程序实现的优化等都有待进一步研究.

第 4 章 BV-SDLS 进一步推广

在第三章, 所讨论的 (BV-SDLS) 问题是含如下约束 $0 \preceq X \preceq \beta I_n$ 的二次半定规划问题, 其中 $\beta > 0, I_n$ 是单位矩阵. 易发现, 右端变量的界约束 βI_n 等价于有界变量约束 $0 \preceq X \preceq B$ 的上限约束矩阵 B 的 n 个特征值等于一个常数 $\beta (\beta \geq 0)$. 本章我们将对 (BV-SDLS) 作进一步推广, 对约束矩阵 B 不加任何限制, 讨论如下二次半定规划问题:

$$\begin{aligned} \min_{X \in S^n} \quad & \frac{1}{2} \|X - C\|_F^2 \\ \text{s.t.} \quad & \mathcal{A}X = b \\ & 0 \preceq X \preceq B, B \in S_+^n \end{aligned} \quad (4.1)$$

4.1 问题的分析与讨论

很自然地, 我们会考虑将第三章的投影拟牛顿算法应用到该问题上来, 对于任意矩阵 $B \in S_+^n$, 先定义矩阵集合 $\Omega = \{X \in S^n | 0 \preceq X \preceq B\}$, 为应用投影拟牛顿算法, 因此我们希望对称矩阵 $C \in S^n$ 在 $\Omega = \{X \in S^n | 0 \preceq X \preceq B\}$ 的投影仍然有形如 $P_\Omega(C)$ 的表示式:

$$P_\Omega(C) = P_C \text{Diag}\{\max\{0, \min\{\lambda(B), \lambda(C)\}\}\} P_C^T$$

遗憾的是, 按此表示并不能准确得到 $C \in S^n$ 在 $\Omega = \{X \in S^n | 0 \preceq X \preceq B\}$ 的投影, 考虑如下例子

$$C = \begin{pmatrix} 1 & -2 & 3 \\ -2 & 2 & -1 \\ 3 & -1 & 2 \end{pmatrix}; B = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 3 & 0 \\ 1 & 0 & 4 \end{pmatrix};$$

依照上述投影公式, C 在 $\Omega = \{X \in S^n | 0 \preceq X \preceq B\}$ 的投影矩阵为:

$$P_\Omega(C) = \begin{pmatrix} 1.6344 & -1.2296 & 1.7297 \\ -1.2296 & 1.8132 & -0.8672 \\ 1.7297 & -0.8672 & 2.0427 \end{pmatrix}$$

但是, 易知矩阵 $B - P_\Omega(C)$ 的特征值向量 $\lambda(B - P_\Omega(C)) = (-0.9406, 1.9262, 2.5241)^T$, i.e. 并不能保证 $P_\Omega(C) \preceq B$ 成立. 因此, C 在 $\Omega = \{X \in S^n | 0 \preceq X \preceq B\}$ 的投影显式并不能简单地从 C 在闭凸集 Ω^β 上的投影推广而来.

实际上, 讨论问题 (4.1) 的求解算法可转化考虑如下问题

$$\begin{aligned} \min_{\tilde{X} \in S^n} \quad & \frac{1}{2} \|\tilde{X} - C\|_F^2 \\ \text{s.t.} \quad & A\tilde{X} = b \\ & 0 \preceq \tilde{X} \preceq \text{Diag}\{\mu_1, \mu_2, \dots, \mu_n\} \end{aligned} \quad (4.2)$$

其中, 对角矩阵

$$\text{Diag}\{\mu_1, \mu_2, \dots, \mu_n\} = \begin{pmatrix} \mu_1 & & & \\ & \mu_2 & & \\ & & \ddots & \\ & & & \mu_n \end{pmatrix}$$

对角元素 $\mu_i \geq 0, i = 1, \dots, n$, 且存在指标 i, j , 使得 $\mu_i \neq \mu_j$, 并满足 $\mu_1 \leq \mu_2 \leq \dots \leq \mu_n$. 事实上, 由于 $B \in S_+^n$, 故存在正交矩阵 Q , 使得 $B = Q \text{Diag}\{\mu_1, \mu_2, \dots, \mu_n\} Q^T$, 且 $\mu_1 \leq \mu_2 \leq \dots \leq \mu_n$. 令 $X = Q\tilde{X}Q^T, \tilde{A}_i = Q^T A_i Q (i = 1, \dots, m), C = Q\tilde{C}Q^T, \tilde{A}\tilde{X} = (\tilde{A}_1 \bullet \tilde{X}, \dots, \tilde{A}_m \bullet \tilde{X})$, 则问题 (4.1) 可转化为形如 (4.2) 的形式:

$$\begin{aligned} \min_{\tilde{X} \in S^n} \quad & \frac{1}{2} \|\tilde{X} - \tilde{C}\|_F^2 \\ \text{s.t.} \quad & \tilde{A}\tilde{X} = b \\ & 0 \preceq \tilde{X} \preceq \text{Diag}\{\mu_1, \mu_2, \dots, \mu_n\} \end{aligned}$$

基于以上分析, 倘若可以得到 $C \in S^n$ 在该闭凸集

$$\{X \in S^n \mid 0 \preceq X \preceq \text{Diag}\{\mu_1, \mu_2, \dots, \mu_n\}\}$$

上的投影, 则本文第三章中讨论的投影拟牛顿算法与交错投影算法均可推广到变量有界半定最小二乘问题 (4.2), 并在此基础上, 从而获得原问题 (4.1) 的求解算法.

定义集合 $\Omega' = \{X \in S^n \mid 0 \preceq X \preceq \text{Diag}\{\mu_1, \mu_2, \dots, \mu_n\}\}$, 向量 $v = (\mu_1, \mu_2, \dots, \mu_n)^T$, 其中 $\mu_i \geq 0, i = 1, \dots, n$, 且存在下指标 i, j , 使得 $\mu_i \neq \mu_j, \mu_1 \leq \mu_2 \leq \dots \leq \mu_n$. 以下考虑 $C \in S^n$ 在闭凸集 Ω' 的投影.

特别地, 当矩阵 $C \in S^n$ 取某些特殊形式时, 容易得到投影显式的一些结论.

命题 4.1.1. 若目标矩阵 $C = \rho I_n, \rho \in R$, 则 $C \in S^n$ 在 Ω' 的投影

$$P_{\Omega'}(C) = \text{Diag}\{\max\{0, \min\{\rho e, v\}\}\}$$

若 $0 \leq \rho \leq \min\{v\}$, 则 $P_{\Omega'}(C) = \rho I_n$.

证明: 显然.

倘若对角矩阵 $C \in S^n$ 中的对角元素不是恒等的, 即存在 i, j , 使得 $\rho_i \neq \rho_j$, 则容易得到如下命题:

命题 4.1.2. 若目标矩阵 $C = \text{Diag}\{\rho_1, \rho_2, \dots, \rho_n\}$, 且存在 i, j , 使得 $\rho_i \neq \rho_j$, 令 $\rho = (\rho_1, \rho_2, \dots, \rho_n)^T$, 则 $C \in S^n$ 在 Ω' 的投影

$$P_{\Omega'}(C) = \text{Diag}\{\max\{0, \min\{\rho, v\}\}\}$$

证明: $C = \text{Diag}\{\rho_1, \rho_2, \dots, \rho_n\}$, 对任意 $X \in \Omega'$, 有:

$$\begin{aligned} \|C - X\|_F^2 &= \sum_{i \neq j} x_{ij}^2 + \sum_{i=1}^n (\rho_i - x_{ii})^2 \\ &\geq \sum_{i=1}^n (\rho_i - x_{ii})^2 \\ &= \sum_{\rho_i < 0} (\rho_i - x_{ii})^2 + \sum_{0 \leq \rho_i \leq \mu_i} (\rho_i - x_{ii})^2 + \sum_{\rho_i > \mu_i} (\rho_i - x_{ii})^2 \end{aligned}$$

由于 $0 \leq X \leq \text{Diag}\{\mu_1, \mu_2, \dots, \mu_n\}$, 故 $x_{ii} \geq 0, \mu_i - x_{ii} \geq 0$, 即 $0 \leq x_{ii} \leq \mu_i$. 取 $X = \text{Diag}(x_{ii})$, 其中

$$x_{ii} = \begin{cases} 0, & \rho_i < 0 \\ \rho_i, & 0 \leq \rho_i \leq \mu_i \\ \mu_i, & \rho_i > \mu_i \end{cases}$$

就可使目标函数 $\|C - X\|_F^2$ 达到最小值:

$$\sum_{\rho_i < 0} \rho_i^2 + \sum_{\rho_i > \mu_i} (\rho_i - \mu_i)^2$$

即 $P_{\Omega^{\beta}}(C) = \text{Diag}(x_{ii}) = \text{Diag}\{\max\{0, \min\{\rho, v\}\}\}$, 且有

$$d_{\Omega^{\beta}}(C) = \frac{1}{2} \left(\sum_{\rho_i < 0} \rho_i^2 + \sum_{\rho_i > \mu_i} (\rho_i - \mu_i)^2 \right)$$

显然, 当命题 4.1.2 中的目标矩阵 C 中对角元素相等时, 即为命题 4.1.1, 因此命题 4.1.1 是命题 4.1.2 的特殊情形.

由此受到启发, 对于一般的闭凸集 $\Omega = \{X \in S^n | 0 \preceq X \preceq B\}$, 只要对矩阵 $C \in S^n$ 与 $B \in S_+^n$ 施加某种限制, 可得如下定理.

定理 4.1.3. 设 $C, B \in S^n$ 且 $CB = BC$, 则对称矩阵 $C \in S^n$ 在闭凸集 $\Omega = \{X \in S^n | 0 \preceq X \preceq B\}$ 上的投影 $P_\Omega(C)$ 是唯一的, 且有如下表示式

$$P_\Omega(C) = Q \text{Diag}\{\max\{0, \min\{\lambda(C), \lambda(B)\}\}\} Q^T$$

C 到集合 Ω 的半平方距离

$$d_\Omega(C) = \frac{1}{2} \left(\sum_{\lambda_i < \mu_i} \lambda_i^2 + \sum_{\lambda_i > \mu_i} (\lambda_i - \mu_i)^2 \right)$$

其中 $\lambda_i, \mu_i (i = 1, \dots, n)$ 分别是矩阵 C 和 B 的特征值.

证明: 因为 $C, B \in S^n$ 且 $CB = BC$, 则对称矩阵 C 与 B 可同时对角化, 即存在正交矩阵 $Q (QQ^T = I_n)$, 使得

$$Q^T C Q = \text{Diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\} \triangleq \Lambda$$

$$Q^T B Q = \text{Diag}\{\mu_1, \mu_2, \dots, \mu_n\} \triangleq \Xi$$

其中 λ_i 是 $C \in S^n$ 的 n 个特征值, 且 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$.

对任意 $X \in \Omega$, 令 $Y = Q^T X Q$. 则有

$$\begin{aligned} \|C - X\|_F^2 &= \|\Lambda - Y\|_F^2 \\ &= \sum_{i \neq j} y_{ij}^2 + \sum_{i=1}^n (\lambda_i - y_{ii})^2 \\ &\geq \sum_{i=1}^n (\lambda_i - y_{ii})^2 \\ &= \sum_{\lambda_i < 0} (\lambda_i - y_{ii})^2 + \sum_{0 \leq \lambda_i \leq \mu_i} (\lambda_i - y_{ii})^2 + \sum_{\lambda_i > \mu_i} (\lambda_i - y_{ii})^2 \\ &\geq \sum_{\lambda_i < 0} \lambda_i^2 + \sum_{\lambda_i > \mu_i} (\lambda_i - \mu_i)^2 \end{aligned}$$

因为 $0 \preceq X \preceq B$, 容易发现 $0 \preceq Y \preceq \Xi$, 由引理 2.1.2, 有 $y_{ii} \geq 0, \mu_i - y_{ii} \geq 0$, i.e. $0 \leq y_{ii} \leq \mu_i$. 取矩阵 $Y = \text{Diag}(y_{ii})$, 其中

$$y_{ii} = \begin{cases} 0, & \lambda_i < 0 \\ \lambda_i, & 0 \leq \lambda_i \leq \mu_i \\ \mu_i, & \lambda_i > \mu_i \end{cases}$$

就可使目标函数 $\|C - X\|_F^2$ 达到下界 $\sum_{\lambda_i < 0} \lambda_i^2 + \sum_{\lambda_i > \mu_i} (\lambda_i - \mu_i)^2$. 基于以上的分析, 可以得到 $X_{\min} = Q \text{Diag}(y_{ii}) Q^T$, 即 $C \in S^n$ 在 Ω 上的投影

$$P_\Omega(C) = Q \text{Diag}\{\max\{0, \min\{\lambda(C), \lambda(B)\}\}\} Q^T$$

其中 $\lambda(C) = (\lambda_1, \dots, \lambda_n)^T, \lambda(B) = (\mu_1, \dots, \mu_n)^T$, 且 C 到集合 Ω 的半平方距离

$$\begin{aligned} d_{\Omega}(C) &= \min_{X \in \Omega} \frac{1}{2} \|X - C\|_F^2 \\ &= \frac{1}{2} \|P_{\Omega}(C) - C\|_F^2 \\ &= \frac{1}{2} \left(\sum_{\lambda_i < 0} \lambda_i^2 + \sum_{\lambda_i > \mu_i} (\lambda_i - \mu_i)^2 \right). \end{aligned}$$

推论 4.1.4. 若矩阵 $B = \beta I_n$, 其中 $\beta > 0$, 则 $C \in S^n$ 在 $\Omega = \{X \in S^n | 0 \preceq X \preceq \beta I_n\}$ 上的投影

$$P_{\Omega}(C) = P_C \text{Diag}\{\max\{0, \min\{\lambda(C), \beta e\}\}\} P_C^T$$

其中 I_n 是一单位矩阵, “ e ” 是 R^n 中的单位向量且对应分量为 1, 即 $e = (1, \dots, 1)^T$

证明: 易由 $C(\beta I_n) = (\beta I_n)C$ 和定理 4.1.3 得到.

由此可以发现, 第三章中的定理 3.2.10 可看作是定理 4.1.3 的一个推论.

注 4.1.1. 对于 $\Omega = \{X \in S^n | 0 \preceq X \preceq B\}$, 当 $CB = BC$, 由上可知 $C \in S^n$ 在 Ω 上的投影

$$\begin{aligned} P_{\Omega}(C) &= Q \text{Diag}\{\max\{0, \min\{\lambda(C), \lambda(B)\}\}\} Q^T \\ &= Q \text{Diag}\{\min\{\max\{0, \lambda(C)\}, \lambda(B)\}\} Q^T \end{aligned}$$

倘若不存在着上限约束矩阵 B (此时可将矩阵 B 的所有特征值看作 $+\infty$), 则对称矩阵 $C \in S^n$ 在闭凸集 Ω 上的投影即为

$$P_{\Omega}(C) = P_C \text{Diag}\{\max\{0, \lambda(C)\}\} P_C^T$$

显然, 此时该投影显式即为半正定矩阵锥上的投影.

上述考虑的只是对称矩阵 C 与 B 在满足 $CB = BC$ 情况下, 利用对称矩阵 $C \in S^n$ 在闭矩阵凸集 Ω 上的投影算子 $P_{\Omega}(C)$ 探讨问题 (4.1) 的求解算法. 当 C 与 B 不满足交换性时, 第三章中介绍的投影拟牛顿算法是不可行的. 因此, 我们有必要进一步深入探讨 (4.1) 的其他求解算法, 这也是后续所要要做的工作.

结论

本文主要研究以下两类特殊线性二次半定规划的特征, 相互之间的联系及其求解算法:

$$\begin{aligned} \min_{X \in S^n} \quad & \frac{1}{2} X \bullet \sum_{i=1}^l P_i X P_i + C \bullet X \\ \text{s.t.} \quad & \mathcal{A}X = b \\ & X \succeq 0 \end{aligned}$$

$$\begin{aligned} \min_{X \in S^n} \quad & \frac{1}{2} \|X - C\|_F^2 \\ \text{s.t.} \quad & \mathcal{A}X = b \\ & 0 \preceq X \preceq B \quad (B = \beta I_n) \end{aligned}$$

其中 $P_i \succeq 0$ ($i = 1 \cdots l$), $\beta > 0$, I_n 是 n 阶单位阵. 文中主要以对称矩阵论为依据, 运用非线性规划的对偶理论及 Lagrange 乘子法得到的成果有:

1. 讨论了二次半定规划的对偶性理论及其 KKT 条件, 同时给出了基于 NT 搜索方向唯一性的证明, 并在此基础上给出了二次半定规划的原始对偶内点算法, 最主要的成果是第 2 章第 2 节的定理 2.2.3.

2. 讨论该类二次半定规划问题与半定最小二乘问题之间的关系, 给出了它们在一定条件下的转换关系, 主要结果是命题 2.4.1, 命题 2.4.2, 命题 2.4.3.

3. 实对称矩阵 $C \in S^n$ 在有界闭凸集 $\Omega^\beta = \{X \in S^n | 0 \preceq X \preceq \beta I_n\}$ 上的投影精确表达式, 主要结论是定理 3.2.10 及推论 3.2.11, 推论 3.2.12.

4. 利用实对称矩阵在 Ω^β 上的投影, 通过一系列转换得到变量有界半定最小二乘问题 (BV-SDLS) 的求解算法, 主要结论是第 3 章第 3 节的算法 3.3.1.

5. 进一步讨论了矩阵 $C \in S^n$ 在有界闭凸集 $\Omega = \{X \in S^n | 0 \preceq X \preceq B\}$ 上的投影表达式, 主要结论有命题 4.1.1, 命题 4.1.2, 定理 4.1.3.

6. 进一步的研究设想:

▷ 考虑是否可以将用于线性半定规划的 Gauss-Newton 方向直接应用到二次半定规划中来, 避免对松弛条件对称化过程, 同时也加强数值稳定性.

▷ 考虑算法 2.2.1 的复杂度 (多项式时间?).

▷ 考虑投影算子 $P_{\Omega^\beta}(C)$ 的性质, 包括半光滑性 (semismooth) 等.

▷ 考虑问题 (4.1) 在 $CB \neq BC$ 下的求解算法.

附录

附录 1:(QSDP) 问题原始对偶路径跟踪算法 *Matlab* 7.0.1 程序:

```
% 严格初始可行点 X,lambda,Z
X=[5,1,2;1,2,-1;2,-1,3];
lambda=[1,1];
Z=[4,-3,-1;-3,5,2;-1,2,3];
system(X,lambda,Z)

function p=system(X,lambda,Z)
disp(' —primal dual path following method for QSDP— ')
% Initializations
m=2;n=3;
P1=[2,0,1;0,3,0;1,0,4]; P2=[2,1,0;1,3,1;0,1,4]; C=eye(3);
A1=[3,1,2;1,2,1;2,1,3]; A2=[1,-2,3;-2,2,-1;3,-1,2]; b=[8,8];
alpha=1;beta=1;eps=1e-5;t0=cputime;
mu=trace(X*Z)/n;j=1;g=zeros(1,1000);g(1)=trace(X*Z);l=zeros(100);l(1)=cond(X,2);
while mu > eps
% Calculate the NT direction  $W = X^{1/2}[(X^{1/2}ZX^{1/2})^{-1/2}]X^{1/2}$ ;
[L,I]=cholp(X);
[R,J]=cholp(Z);
[U,D,V]=svd(R'*L);
G=L*V*inv(rootm(D,2));
W=G*G'; % NT direction
% Solve the system equation's solution;
P=inv(W);M=kron(P,P);
e=eye(n);N=kron(e,e);
Rc=mu*inv(X)-Z; rc=vec(Rc);
Rd=-(Z-P1*X*P1-P2*X*P2-C+lambda(1)*A1+lambda(2)*A2); rd=vec(Rd);
```

```

r1=-(trace(A1*X)-b(1)); r2=-(trace(A2*X)-b(2));
rp=[r1,r2]';
H=[vec(A1),vec(A2)]';
T=H*inv(N*(kron(P1,P1)+kron(P2,P2))+M)*N*H';
[E,F]=cholp(T);
y=inv(E')*(rp+H*inv(N*(kron(P1,P1)+kron(P2,P2))+M)*(N*rd-rc));
% Caculate new search direction (Delta_x, Delta_lambda, Delta_z);
Delta_lambda=inv(E)*y;
Delta_x=mat(inv(N*(kron(P1,P1)+kron(P2,P2))+M)*(N*H'*Delta_lambda-N*rd+rc));
Delta_z=mat(rd-H'*Delta_lambda+(kron(P1,P1)+kron(P2,P2))*vec(Delta_x));
% Choose alpha > 0, beta > 0 such that the matrix X + alpha * Delta_x, Z + beta *
Delta_z are positive definite
x=eig(X+alpha*Delta_x);
z=eig(Z+beta*Delta_z);
for i=1:n
    while x(i)<=0
        alpha=alpha-0.001;
        x=eig(X+alpha*Delta_x);
    end
    while z(i)<=0
        beta=beta-0.001;
        z=eig(Z+beta*Delta_z);
    end
end
X=X + alpha*Delta_x;
lambda=lambda + alpha*Delta_lambda;
Z=Z + beta*Delta_z;
mu=mu-0.1;
j=j+1;

```

```

residual=trace(X*Z)
g(j)=residual;
condnum=cond(X,2)
l(j)=condnum;
end
mu=trace(X*Z)/n;
Xopt = X    % optimal solution
lambdaopt = lambda
Zopt = Z
x=1:j;
plot(x,g(1:j))    % plot(x,l(1:j))
axis([-5 100 -5 40])
xlabel('k')
ylabel('trace(X*Z)')    % ylabel('X- 条件数')
Objective=1/2*trace(X*(P1*X*P1+P2*X*P2)) + trace(C*X)
timeused = cputime - t0

```

附录 2: 求解 (BV-SDLS) 问题的 MATLAB7.0.1 程序

▷ (BV-SDLS) 投影拟牛顿算法 *Matlab* 程序:

% 数据初始化

```
clc;clear;
```

```
n = 1500; % 矩阵维数 n
```

```
m = fix(sqrt(n)) % 仿射约束个数 m, fix: 朝 0 方向取整
```

```
% m = n;
```

% 随机生成矩阵 C

```
C = rand(n) + (-rand(n));
```

```
C = 1/2*(C+C') - diag(diag(1/2*(C+C')))+eye(n); % Cij ∈ [-1, 1], Cii = 1
```

```
% C = eye(n);
```

% 生成仿射约束 $\mathcal{A} = \text{diag}$, $b = e$, 即 $\{X \in S^n | X_{ii} = 1, i = 1, \dots, m\}$

```

for i=1:m
    A(:, :, i) = zeros(n);
    A(i, i, i) = 1;
end
b = ones(m, 1);
t = 3;

eps = 1e - 5; % 控制误差
eps0 = 1e - 3; % 控制梯度差
y = 2 * ones(m, 1); % 初始迭代点
H0 = eye(m);
H = H0; % 初始校正矩阵
t0 = cputime; % 初始 cpu - time
f = zeros(1000, 1);
% 计算  $Q = C + A^*y$  在  $0 \leq X \leq t * I_n$  上的投影矩阵  $X$ 
Q = C;
for i = 1 : m
    Q = Q + y(i) * A(:, :, i); %  $Q = C + A^*y$ 
end
Q;
[U, V] = eig(Q);
v = diag(V);
w = zeros(n, 1);
for i = 1 : n
    if v(i) < 0
        w(i) = 0;
    elseif v(i) > t
        w(i) = t;
    else

```

```

        w(i) = v(i);
    end
end
w;
X = U * diag(w) * U';
l = zeros(m, 1);
for i = 1 : m
    l(i) = trace(A(:, :, i) * X);
end
g = l - b; % 梯度方向
g0 = g;
k = 0; % 初始迭代 k

d_F = 0;
for i = 1 : n
    if v(i) < 0
        d_F = d_F + 0.5 * v(i)^2;
    end
    if v(i) > t
        d_F = d_F + 0.5 * (v(i) - t)^2;
    end
end
end
d_F;
eta_y0 = d_F - 0.5 * trace(Q * Q) + 0.5 * trace(C * C) + y' * b;
f0 = eta_y0; % 初始点的对偶目标函数值

while norm(g0) > eps
    k = k + 1
    d = -H * g0; % 搜索方向

```

```

    a = k/(k + 1); % 步长 a
    % a = 1;
    y = y + a * d;
% 更新  $Q = C + \mathcal{A}^*y$  在  $0 \preceq X \preceq t * I_n$  上的投影矩阵  $X$ 
    Q = C;
for i = 1 : m
    Q = Q + y(i) * A(:, :, i); %  $Q = C + \mathcal{A}^*y$ 
end
    Q;
    [U, V] = eig(Q);
    v = diag(V);
    w = zeros(n, 1);
for i = 1 : n
    if v(i) < 0
        w(i) = 0;
    elseif v(i) > t
        w(i) = t;
    else
        w(i) = v(i);
    end
end
    w;
    X = U * diag(w) * U';
    l = zeros(m, 1);
for i = 1 : m
    l(i) = trace(A(:, :, i) * X);
end
    g = l - b; % 梯度方向
% 判断相邻两迭代点的梯度值  $g$  之差是否接近 0

```

```

    p = a * d;
    q = g - g0;
if norm(q) > eps0
    r = p' * q;
    s = q' * H * q;
    H = H + (p * p')/r - (H * q * q' * H)/s; % DFP 校正公式
    g0 = g;
else
    H = H0; % 以单位矩阵为下次迭代修正矩阵
    g0 = g;
end
    X;
% 计算  $d(C + \mathcal{A}^*y)$ 
    d_F = 0; % 初始化
for i = 1 : n
    if v(i) < 0
        d_F = d_F + 0.5 * v(i)^2;
    end
    if v(i) > t
        d_F = d_F + 0.5 * (v(i) - t)^2;
    end
end
    end
    d_F;
    eta_y = d_F - 0.5 * trace(Q * Q) + 0.5 * trace(C * C) + y' * b;
    f(k) = eta_y;
end
    X ; % primal optimal solution
    y; % dual optimal solution
    eta_y = d_F - 0.5 * trace(Q * Q) + 0.5 * trace(C * C) + y' * b % 对偶目标函数值

```

```

objprimal = 1/2 * trace((X - C) * (X - C)) % 原始目标函数值
timeused = cputime - t0 % cpu time
x = 0 : k;
fk = [f0, f(1 : k)];
plot(x, fk)
xlabel('iteration number');
ylabel('dual objective eta-y');
▷ (BV-SDLS) 模型内点算法 Matlab7.0.1 程序 (基于 Sedumi, Yalmip):
clear;
n = 3; m = 2;
I = eye(n);
C = [2, 1, 3; 1, 5, 2; 3, 2, 2];
A(:, :, 1) = [3, 1, 2; 1, 2, 1; 2, 1, 3];
A(:, :, 2) = [1, -2, 3; -2, 2, -1; 3, -1, 2];
b = [1, 2]; t = 3;
% C = eye(n);
% for i = 1 : m
%   A(:, :, i) = zeros(n);
%   A(i, i, i) = 1;
% end
% b = ones(m, 1);
% code
t0 = cputime;
X = sdpvar(n, n);
objective = 1/2 * trace((X - C) * (X - C));
F = set(0 < X < t * I);
for i = 1 : m
    F = F + set(trace(A(:, :, i) * X) == b(i));
end

```

```

solvesdp(F,objective);
optX=double(X)
optobj=double(objective)
timeused=cputime-t0

```

% Results of the experiment one:

SeDuMi 1.1 by AdvOL, 2005 and Jos F. Sturm, 1998,2001-2003.

Alg = 2: xz-corrector, theta = 0.250, beta = 0.500

Put 2 free variables in a quadratic cone

eqs m = 7, order n = 11, dim = 30, blocks = 5

nnz(A) = 44 + 0, nnz(ADA) = 49, nnz(L) = 28

it : b*y gap delta rate t/tP* t/tD* feas cg cg prec

0 : 6.53E-001 0.000

1 : -1.16E+001 6.44E-002 0.000 0.0985 0.9900 0.9900 0.48 1 1 2.4E+000

2 : -1.99E+001 2.95E-002 0.078 0.4577 0.9000 0.9000 1.64 1 1 9.8E-001

3 : -2.40E+001 1.34E-002 0.000 0.4542 0.9000 0.9000 0.98 1 1 3.6E-001

4 : -2.74E+001 1.27E-003 0.281 0.0950 0.9900 0.9900 0.90 1 1 3.7E-002

5 : -2.79E+001 1.08E-004 0.000 0.0853 0.9900 0.9900 0.88 1 1 3.4E-003

6 : -2.80E+001 9.93E-006 0.000 0.0916 0.9900 0.9900 0.90 1 1 3.3E-004

7 : -2.80E+001 2.36E-006 0.000 0.2380 0.9000 0.9000 1.02 1 1 7.6E-005

8 : -2.80E+001 1.42E-007 0.000 0.0600 0.9900 0.9900 1.01 1 1 4.5E-006

9 : -2.80E+001 2.59E-008 0.000 0.1824 0.9000 0.9000 1.00 1 1 8.3E-007

10 : -2.80E+001 6.71E-009 0.000 0.2596 0.9000 0.9000 1.00 1 1 2.1E-007

11 : -2.80E+001 5.36E-010 0.000 0.0799 0.9900 0.9900 1.00 1 1 1.7E-008

12 : -2.80E+001 4.40E-011 0.463 0.0821 0.9900 0.9900 1.00 2 2 1.4E-009

13 : -2.80E+001 9.26E-012 0.000 0.2105 0.9000 0.9000 1.00 2 2 3.0E-010

iter seconds digits c*x b*y

13 1.2 Inf -2.7970653995e+001 -2.7970653994e+001

$|Ax - b| = 3.7e-010$, $[Ay - c]_+ = 3.6E-010$, $|x| = 1.0e+001$, $|y| = 2.8e+001$

Detailed timing (sec)

Pre IPM Post

1.719E-001 1.188E+000 3.125E-002

Max-norms: $\|b\|=1$, $\|c\| = 3.150000e+001$,

Cholesky $|add|=0$, $|skip| = 0$, $\|L.L\| = 17.4438$.

$$opt_x = \begin{pmatrix} 0.04736616 & -0.16930699 & 0.00487131 \\ -0.16930699 & 0.60517581 & -0.01741217 \\ 0.00487131 & -0.01741217 & 0.00050098 \end{pmatrix}$$

$opt_{obj} = 27.97065399$

$time_{used} = 2.31250000$

参考文献

- [1] Robert M.Freund and Brian Anthony. *Introduction to semidefinite programming(SDP)*. May 4-6,2004.
- [2] Helmbert.C. *Semi-definite Programming for Combinatorial Optimization.[A]*. Konrad Zurse-Zentrum fur in Formationstechnik[C], Berlin. 2000.
- [3] R.Bellman and K.Fan. *On systems of linear inequalities in Hermitian matrix variables*. In V.L.Klee,editor,convexity, Volume 7 of Proceedings of symposia in Pure Mathematics, American:1-11.
- [4] W.E.Donath and A.J.Hoffman. *Lower bounds for the partitioning of graphs*. IBM Journal of Research and Development, 1973, 17:420-425.
- [5] J.Cullum,W.e.Donath and Wolf. *The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices*. Math.Programming Study, 1975, 3:35-55.
- [6] C.HELMBERG. *Semidefinite Programming for Combinatorial Optimization[M]*. Germany:Zuse Institute Berlin, 2000.
- [7] 王新辉, 刘三阳, 刘红卫. 半定规划的割平面算法及其应用. 西安电子科技大学学报 (自然科学版).2004,31(1).
- [8] M.F. Anjos. *An improved semidefinite programming relaxation for the satisfiability problem* . Math. Progr. A, 2005,102(3): 589-608.
- [9] N.Karmarker. *A new polynomial time algorithm for linear Programming*. Combinatorica, 1984,4(4):373-395.
- [10] Y.E.Nesterov AND A.S Nemirovskii. *A general approach to the design of optimal methods for smooth convex functions minimization*. Ekonomika I Matem.Metody,1988(24): 509-517.
- [11] Y.E.Nesterov AND A.S Nemirovskii. *Self-Concordant Functions and Polynomial Time Methods in Convex Programming*. Centr Econ & Math Inst.,USSR Acad.Sci.Moscow, USSR, 1989.

- [12] F.Alizadeh. *Optimization over the positive definite cone:interior point methods and combinatorial applications*. In Panos Pardalos,editor,Advances in Optimization and Parallel Computing. North-Holland, 1992.
- [13] F.Jarre. *An interior point method for minimizing the maximum eigenvalue of a linear combination of matrices*. SIAM J.Optim., 1993(31),1360-1377.
- [14] F.Alizadeh J-P.A.Haeberly,and M.L.Overton. *Primal-dual interior point methods for semidefinite programming*. Manuscript presented at the Math.Programming Symposium.Ann. Arbor, MI, 1994.
- [15] M.Kojima,S.Shindoh,And S.Hare. *Interior point methods for the monotone semidefinite linear complementarity problem in symmetric matrices*. SIAM J,Optim.,7(1997).86-125.
- [16] Todd M J, Toh K C, Tutuncu.R.H. *On the Nesterov-Todd direction in SDP[J]*. SIAM Journal of Optimization,1998,8(3):769-796.available from <http://legacy.orie.cornell.edu/techreports/TR1154.pdf>.
- [17] Christoph Helmberg,Robert Weismantel. *Cutting Plane Algorithms for Semidefinite Relaxations*. Konrad-Zuse-Zentrum fur Informationstechnik Berlin Takustrabe 7, D-14195 Berlin, 1997.
- [18] Helmberg and F.Rendl. *A Spectral Bundle Method for Semidefinite Programming*. SIAM Journal on Optimization, ,2000(10),673-696.
- [19] Krishnan and J.E.Mitchell. *A linear Programming Approach to Semidefinite Programming Problems*. Technical Report,Dept. of Mathematical Sciences,Rensselaer Polytechnic Institute,2001.
- [20] Oskoorouchi. *The Analytic Center Cutting Plane Method with Semidefinite Cuts*. Ph.D. thesis,Faculty of Management,McGill University,Montreal, Canada, May 2002.
- [21] C.Kanzow and C.Nagel.*Some practical aspects of a newton-type method for semidefinite programs*. 2001,Germany. available from <http://www.mathematik.uni-wuerzburg.de/kanzow/paper/SDPStruct.pdf>.
- [22] C.Kanzow and C.Nagel.*Semidefinite programs:New Search directions,smoothing-Type methods and numerical results*. SIAM J.optim,2002(13),1-23.

- [23] J.Povh,Novo Mesto,F.Rendl and A.Wiegele,Klagenfurt. *A Boundary point method to solve semidefinite programs*. Linear ALgebra and Applications, Computing,2006(78):277-286.
- [24] Alain Billionnet .Sourour Elloumi. *Using a Mixed Integer Quadratic Programming Solver for the Unconstrained Quadratic 0-1 Problem*[J]. Math Program,2007,Ser.A 109:55-68.
- [25] Xiu Cui Guan, Zai Yun Diao. *The quadratic semi-definite programming problem and its projection and contraction algorithms*[J]. Numerical Mathematics A Journal of Chinese,University, 2002, 24(2): 97-108.
- [26] Xu feng-min, Xu cheng-xian. *Primal-dual algorithm for quadratic semidefinite programming*[J]. Chinese Journal of Engineering Mathematics.2006,23(4): 590-598.
- [27] K.C.Toh,R.H.Tutuncu,and M.J.Todd. *Inexact primal dual path following algorithms for a special class of convex quadratic SDP and related problem*.2006.available from http://www.optimization-online.org/DB_HTML/2006/05/1399.html
- [28] Higham. *Computing a nearest symmetric correlation matrix-a problem from finance*[J]. IMA Journal of Numerical Analysis,2002,22(30):329-343.
- [29] R.L.Dykstra. *An algorithm for restricted least-square regression*[J]. Journal of the American Statistical Association,1983(78):837-842.
- [30] Stephen Boyd and Jon Dattorro. *Alternating Projections*.Autumn[J],2003.available from http://www.stanford.edu/class/ee392o/alt_proj.pdf
- [31] N.Higham. *Computing a nearest symmetric positive semidefinite matrix*[J]. Linear Algebra and its Applications,1988:103-118.
- [32] Malick J. *A dual approach to semidefinite least-squares problems*[J]. SIAM J Matrix Anal Appl, 2005(26):272-284.
- [33] Boyd S, Lin Xiao. *Least-squares covariance matrix adjustment*[J]. SIAM J Matrix Anal Appl, 2005(27):532-546.
- [34] Houduo Qi and Defeng Sun. *A quadratically convergent method for computing the nearest correlation matrix* .SIAM J. MATRIX ANAL. APPL, 2006(28):360-385 .

- [35] Liqun Qi and Jie Sun. *A nonsmooth version of Newton's method*. Mathematical Programming.1993(58):353-367.
- [36] Defeng Sun, Jie Sun. *Semismooth matrix valued functions[J]*. Mathematics of Operations Research,2002,27(1):150-169.
- [37] Jerome Malick and Hristo S.Sendov. *Clarke Generalized Jacobian of the Projection onto the Cone of Positive Semidefinite Matrices*. Set-Valued Analysis.2006(14):273-293.
- [38] Igor Grubisic and Raoul Pietersz. *Efficient rank reduction of correlation matrices*. Linear Algebra and its Applications , 2007(422):629-653.
- [39] E.A.Papa Quiroz and P.Roberto Oliveira. *A new barrier for a class of semidefinite problems*. RAIRO-Operations Research,2006(40):303-323.
- [40] 王海莉, 刑志栋. 对传统 Markowitz 模型的一种修正 [J]. 西北大学学报 (自然科学网络版), Vol.3, No.6, 2005.
- [41] 张贤达. 矩阵分析与应用 [M]. 北京: 清华大学出版社, 2004.
- [42] 戴华. 矩阵论 [M]. 北京: 科学出版社, 2000.
- [43] Farid Alizadeh, Jean-Pierre A.Haerberly, and Michael L.Overton. *Primal dual interior point methods for semidefinite programming: convergence rates, stability and numerical results*. SIAM J.OPTIM,1998(8):746-768.
- [44] Renato D.C. MONTEIRO. *Primal dual path following algorithms for semidefinite programming*. SIAM J.OPTIM,1997(7):663-678.
- [45] Kruk S, Muramatsu M, Rendl F et al. *The Gauss-Newton direction in semidefinite programming[J]*. Research Report CORR, Canada, 1998. available from <http://citeseer.ist.psu.edu/69912.html>
- [46] Alexander Shapiro. *On differentiability of symmetric matrix valued functions*. 2002. available from http://www.optimization-online.org/DB_HTML/2002/07/499.html
- [47] Dimitri P. Bertsekas, Angelia Nedic and Asuman E. Ozdaglar. *Convex analysis and optimization[M]*. Beijing: Tsinghua University Press, 2006.

- [48] Yu.Nesterov. *Introductory Lectures on Convex Programming[M]*.Volume I:Basic course, 1996.
- [49] 袁亚湘, 孙文瑜. 最优化理论与方法 [M]. 北京: 科学出版社. 2005.
- [50] 陈宝林. 最优化理论与算法 [M]. 北京: 清华大学出版社 [M], 2003.
- [51] 郑志勇. 拟牛顿法计算实现的一个改进 [J]. 北京理工大学数学系, 北京. available from <http://f.thec.cn/Aris/wendang/niniudunfa>
- [52] JOS F. Sturm. *Using sedumi 1.02,a matlab toolbox for optimization over symmetric cones*. Optimization methods and software, 1999:625-653.
- [53] <http://sedumi.mcmaster.ca>
- [54] Johan Lofberg. *YALMIP:A toolbox for modeling and optimization in MATLAB*.Computer Aided Control Systems Design,2004:284-289.
- [55] <http://control.ee.ethz.ch/~joloef/wiki/pmwiki.php>

攻读学位期间的承担的科研任务及主要成果

1. 康志林, 张圣贵. 一类二次半定规划问题及其内点算法. 福建师范大学学报, 2008,1, vol.24, No.1:1-6.
2. 林岚, 康志林, 甘萌雨, 叶金玉. 基于航空口岸的台胞大陆旅游流空间场效应分析. 地理研究, 2007,3, vol.26, No.2:403-413.
3. 康志林, 张圣贵. 变量有界半定最小二乘问题. 已投《高等学校计算数学学报》.
4. 康志林. An Extended Projective Formula and Its Application.
5. 获奖情况:

06 年第三届全国研究生数学建模竞赛国家一等奖;

校长嘉奖令 (闽师办 [2007] 第 29 号).

06-07 学年校研究生一等奖学金、校三好研究生标兵.

06-07 学年校优秀共青团干部.

致谢

长安山依旧，闽江水长流。转眼间，我已在这长安山脚下度过了七个年头。七年，一段短暂而又漫长的旅程，从本科到研究生，七年的光阴助我成长，让我从青涩走向成熟。特别是三年的研究生生涯，让我走上了不一样的人生旅程。

在这宝贵的三年研究生生活中，我要特别感谢我的硕士生导师——张圣贵教授。张老师治学严谨，学识渊博，为我们营造了很好的精神氛围；宽以待人，乐观向上的处世态度更是给我们留下了深刻印象。课堂内，他是一位良师；课堂外，他是一位益友。三年的研究生生涯，不仅让我获得了基本的研究方法，更让我学会了乐观地面对生活。在此向张圣贵老师表示由衷的感谢。

衷心的感谢作者的父母，是他们莫大的支持才得以顺利完成学业。也感谢林建伟，张艳梅同学，长期的共同讨论进一步开拓了作者的数学视野。特别感谢林惠玲师姐，郑开杰师兄在平时的学习交流中给予的指导与帮助。感谢郑寒凝，徐凌，黄银珠三位师妹在课余之外所带来的别样欢乐。

作者感谢钟怀杰，潘日晶，陈清华，肖蓬，苏维纲等任课老师，感谢作者的全体同学。

作者向文献的所有知识产权所有者表示衷心的感谢。

康志林

2008年3月

于福建师范大学数学与计算机科学学院



个人简历

基本情况

康志林, 男, 汉族, 福建永春人, 1982年9月出生.

教育状况

2001年9月至2005年7月, 就读于福建师范大学数学与计算机科学学院, 本科, 专业: 数学与应用数学. 获学士学位.

2005年9月至2008年7月, 就读于福建师范大学数学与计算机科学学院, 硕士研究生, 专业: 应用数学, 研究方向: 最优化理论与算法, 指导老师: 张圣贵教授.

社会经历

2006.2-2006.6 任福建师范大学数学与计算机科学学院 2005级数学与应用数学专业本科《高等代数》辅导老师.

2007.2-2007.6 任福建师范大学数学与计算机科学学院 2006级数学与应用数学专业本科《高等代数》辅导老师.

2007.7.29-2007.8.18 到山东曲阜师范大学日照校区参加“全国优化理论与应用”暑期班学习.

2006.12 到上海同济大学参加“第三届全国研究生数学建模竞赛”交流会暨颁奖大会.