练　　习　　题

从变分不等式的投影收缩算法到ADMM为代表的凸优化的分裂收缩算法

八单元每单元两页习题

南京大学　数学学院

何 炳 生

Homepage: maths.nju.edu.cn/~hebma

目 录

说明. 需要这八单元的配套习题参考答案的, 可以微信向我索要.

主 要 内 容

在习题对应的讲义里, 我们把线性约束凸优化问题的 Lagrange 函数的鞍点归结为变分不等式

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall w \in \Omega \tag{1}$$

的解. 为求解这个变分不等式, 我们已经设计了下面的 **预测-校正的统一框架**.

[预测] 第 k 步迭代从给定的核心变量 v^k (v 也可以是 w) 开始, 求得预测点 $\tilde{w}^k$, 使得

$$\tilde{w}^k \in \Omega, \ \ \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \ \forall w \in \Omega, \tag{2}$$

成立. 当矩阵 $Q^T + Q$ 正定的时, 我们称 (2) 为一个合格的预测.

我们称 Q 为预测矩阵. 如果 (2) 中 $v^k = \tilde{v}^k$, $\tilde{w}^k$ 相当于 (1) 的 w^*, 就是问题的解.

[校正] 根据预测得到的 $\tilde{v}^k$, 给出核心变量 v 的新迭代点 v^{k+1} 的公式为

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \tag{3}$$

我们称 M 为校正矩阵. 由 (2) 预测, (3) 校正的求解问题 (1) 的迭代方法的收敛性条件.

收敛性条件 对给定的预测矩阵 Q, 要求设计的校正矩阵 M 满足:

$$\text{存在正定矩阵} \ \ H \succ 0 \ \ \text{使得} \ \ HM = Q, \ \ \text{并且} \ \ G := Q^T + Q - M^T HM \succ 0. \tag{4}$$

迭代序列满足的收敛关键不等式

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^{k+1} - v^*\|_H^2 - \|v^k - v^{k+1}\|_G^2. \tag{5}$$

对于一个确定的合格的预测, 选择符合收敛条件的校正矩阵 M 并不神秘.

合格预测 (2) 中的预测矩阵 Q 满足 $Q^T + Q \succ 0$, 我们可以选 D, 使其满足

$$0 \prec D \prec Q^T + Q, \qquad \text{然后取} \qquad M = Q^{-T} D \tag{5}$$

做 (3) 那样的校正. 对这样选择的 D 和 M, 记 $H = QD^{-}Q^T$ 就会有

$$HM = Q \quad \text{和} \quad M^T HM = D, \quad \text{因此} \quad G = Q^T + Q - D \succ 0. \tag{6}$$

我们只是在收敛性分析中需要正定的 H, 计算过程中并不要求给出显式的 H. 由于 $M = Q^{-T} D$

$$\text{校正 (3) 也可以通过} \qquad Q^T (v^{k+1} - v^k) = D(\tilde{v}^k - v^k) \qquad \text{实现.}$$

A 单元习题　凸函数 投影 变分不等式 最优性条件 等基础知识

练习题 1-2 讨论最小一模问题和最小无穷模问题跟线性变分不等式的关系

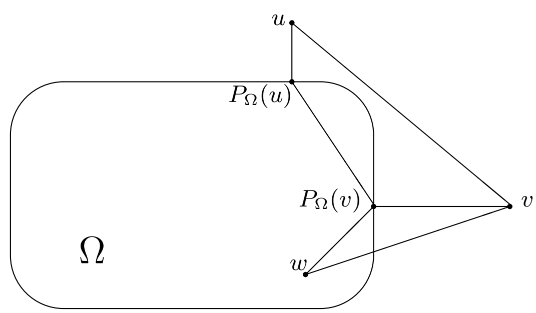
最小二乘问题 $\min \|Ax-b\|_2$ 是个线性代数问题, 有直接方法可以求解. 最小一模问题 $\min \|Ax-b\|_1$ 和最小无穷模问题 $\min \|Ax-b\|_\infty$ 却不是, 它们都具有组合问题的特性, 一般需要用迭代的方法才能求解.

1. 设 $A \in \Re^{m\times n}, b \in \Re^m$. 请分别将最小一模问题 $\min \|Ax-b\|_1$ 和最小无穷模问题 $\min \|Ax-b\|_\infty$ 写成等价的线性规划.

2. 最小一模问题 (Least absolute deviations) 是一个非光滑凸优化问题, 它的数学形式是
$$\min \|Ax-b\|_1, \tag{A.1}$$
用 e 表示每个分量都是 1 的 m-维向量. 记 $B_\infty = \{y \in \Re^m \mid -e \le y \le e\}$. 称其为无穷模下单位球.
   (a) 请说明, 对任何的 $d \in \Re^m$, $\|d\|_1 = \max\{y^T d \mid y \in B_\infty\}$.
   (b) 请将最小一模问题化成相应的单调线性变分不等式
$$u^* \in \Omega, \quad (u-u^*)^T(Mu^*+q) \ge 0, \quad \forall u \in \Omega.$$
   写出其中具体的 u, M, q 和 Ω.

练习题 3-4 讨论在变分不等式求解过程中要用到的一些投影的性质

3. 设 $\Omega \subset \Re^n$ 是闭凸集, 利用短篇 **A2** 中的引理证明
   (a) $\|P_\Omega(v) - P_\Omega(u)\| \le \|v-u\|, \qquad \forall u, v \in \Re^n.$
   (b) $\|P_\Omega(v) - w\| \le \|v-w\|, \qquad \forall v \in \Re^n, w \in \Omega.$
   (c) $\|P_\Omega(v) - w\|^2 \le \|v-w\|^2 - \|v-P_\Omega(v)\|^2,$
   $\forall v \in \Re^n, w \in \Omega.$



4. 由于变分不等式 $u^* \in \Omega,\ (u-u^*)^T F(u^*) \ge 0,\ \forall u \in \Omega$ 等价于求 $e(u,\beta)=0$ 的零点, 其中
$$e(u,\beta) = u - P_\Omega[u - \beta F(u)]$$
在求解单调变分不等式的投影收缩算法中, $\|e(u,\beta)\|$ 常被用作误差度量而需要做进一步研究.
证明对任意确定的 $u \in \Re^n$ 和 $\tilde\beta \ge \beta > 0$, 我们有
$$\|e(u,\tilde\beta)\| \ge \|e(u,\beta)\| \qquad \text{和} \qquad \frac{\|e(u,\tilde\beta)\|}{\tilde\beta} \le \frac{\|e(u,\beta)\|}{\beta}. \tag{A.2}$$
   (a) 若设 $t = \|e(x,\tilde\beta)\|/\|e(x,\beta)\|$, (A.2) 的结论就相当于要证明 t 为一元二次不等式
$$(t-1)(t-\tilde\beta/\beta) \le 0 \quad \text{的解.} \tag{A.3}$$
   (b) 利用投影的基本性质证明
$$\{e(u,\beta) - \beta F(u)\}^T\{e(u,\tilde\beta) - e(u,\beta)\} \ge 0. \tag{A.4}$$
   和
$$\{e(u,\tilde\beta) - \tilde\beta F(u)\}^T\{e(u,\beta) - e(u,\tilde\beta)\} \ge 0. \tag{A.5}$$
   (c) 利用 (A.4) 和 (A.5), 证明
$$\beta\|e(u,\tilde\beta)\|^2 - (\beta+\tilde\beta)\|e(u,\beta)\|\cdot\|e(u,\tilde\beta)\| + \tilde\beta\|e(u,\beta)\|^2 \le 0. \tag{A.6}$$
   最后证明到结论 (A.2).

练习题 5-6 讨论凸优化问题 PPA 的一些性质

5. 设 $f(x):\Re^n\to\Re$ 是凸函数, $\Omega\subset\Re^n$ 是闭凸集. f 在某个包含 Ω 的开集上可微. 根据讲义的定理 **A3**

$$x^*\in\arg\min\{f(x)\,|\,x\in\Omega\}\qquad \text{当且仅当}\qquad x^*\in\Omega,\;\; (x-x^*)^T\nabla f(x^*)\ge 0,\;\; \forall x\in\Omega, \tag{A.7}$$

证明 (A.7) 式与下面的 (A.8) 式等价

$$x^*=P_\Omega[x^*-\nabla f(x^*)]. \tag{A.8}$$

6. 邻近点算法 (Proximal Point Algorithm). 设 $\beta>0$ 是给定的常数, f 是可微凸函数. 对给定的 x^k, 由

$$x^{k+1}=\arg\min\{f(x)+\frac{1}{2\beta}\|x-x^k\|^2\,|\,x\in\Omega\}. \tag{A.9}$$

(a) 证明 (A.9) 中的 x^{k+1} 满足下面的投影方程 $x^{k+1}=P_\Omega[x^k-\beta\nabla f(x^{k+1})]$. 因此 PPA 可以看作一种隐式梯度投影法.

(b) 证明 (A.9) 产生的序列 $\{x^k\}$ 满足

$$\|x^{k+1}-x^*\|^2\le\|x^k-x^*\|^2-\|x^k-x^{k+1}\|^2,\quad \forall x^*\in\Omega^*. \tag{A.10}$$

(c) 若 $x^k=\frac{1}{k}, x^*=0$, 说明 仅有条件 (A.10) 证明算法收敛是不够的.

练习题 7-11 帮助熟悉回顾一些凸优化求解方法中需要用到的线性代数知识

7. 设 $f(x):\Re^n\to\Re$ 是可微 (不一定凸的) 函数, $A\in\Re^{m\times n}$, $b\in\Re^m$. 假设 x^* 是优化问题

$$\min\{f(x)\,|\,Ax=b,\,x\in\Re^n\}$$

的(局部)最优解, 则存在 $y^*\in\Re^m$, 使得 (x^*,y^*) 是非线性方程组

$$\left\{\begin{array}{rcl}\nabla f(x) &=& A^Ty,\\ Ax &=& b\end{array}\right. \tag{A.11}$$

的解. **提示**: 用 $C(A^T)$ 表示 A^T 的列空间, 用 $N(A)$ 表示 A 的零空间, 并利用 $\Re^n=C(A^T)\bigoplus N(A)$.

8. 判定下述矩阵的秩 r, 并将其分解成 $(6\times r)$ 和 $(r\times 6)$ 矩阵的乘积.

$$\begin{bmatrix}1&2&3&4&5&6\\2&0&0&0&0&0\\3&0&0&0&0&0\\4&0&0&0&0&0\\5&0&0&0&0&0\\6&0&0&0&0&0\end{bmatrix},\quad \begin{bmatrix}0&0&0&0&0&1\\0&0&0&0&0&2\\0&0&0&0&0&3\\0&0&0&0&0&4\\0&0&0&0&0&5\\1&2&3&4&5&6\end{bmatrix},\quad \begin{bmatrix}1&2&3&\cdots&n\\2&0&0&\cdots&0\\3&0&0&\cdots&0\\\vdots&\vdots&\vdots&&\vdots\\n&0&0&\cdots&0\end{bmatrix},\quad \begin{bmatrix}0&0&\cdots&0&1\\0&0&\cdots&0&2\\\vdots&\vdots&&\vdots&\vdots\\0&0&\cdots&0&n-1\\1&2&\cdots&n-1&n\end{bmatrix}.$$

9. 已知矩阵 $A\in\Re^{n\times n}$ 的逆矩阵 A^{-1}. $U,V\in\Re^{n\times m}$, $W\in\Re^{m\times m}$. 设 $m\le n$. 请给出

$$M=I+uv^T,\quad M=A+uv^T,\quad M=I+UV^T,\quad M=A+UWV^T\quad \text{的逆矩阵.}$$

10. 计算矩阵

$$\begin{bmatrix}I&u\\v^T&1\end{bmatrix},\qquad \begin{bmatrix}A&u\\v^T&a\end{bmatrix},\qquad \begin{bmatrix}I_n&U\\V^T&I_m\end{bmatrix},\qquad \begin{bmatrix}A&U\\V^T&W\end{bmatrix}\quad \text{的逆矩阵.}$$

11. $\log x$ 在定义域 $x>0$ 上的导数是 $\dfrac{1}{x}$. 当 $\det(X)>0$ 时, 请给出 $\nabla(\log\det(X))$ 的矩阵形式.

提示: 利用代数余子式的定义和性质. 设 X_{kj} 是 x_{kj} 的代数余子式. 则有

$$\sum_{j=1}^n x_{ij}X_{kj}=\left\{\begin{array}{ll}\det(X) & \text{if } i=k,\\ 0 & \text{if } i\ne k.\end{array}\right.\qquad XX^*=\det(X)I,\quad \text{其中}\quad X^*=\begin{pmatrix}X_{11}&X_{21}&\cdots&X_{n1}\\X_{12}&X_{22}&\cdots&X_{n2}\\\vdots&\vdots&&\vdots\\X_{1n}&X_{2n}&\cdots&X_{nn}\end{pmatrix}.$$

B 单元习题　线性和非线性单调变分不等式的投影收缩算法

练习题 1-4 关于单调线性变分不等式: $u^* \in \Omega, (u-u^*)^T(Mu^*+q) \ge 0,\ \forall u \in \Omega$ 的求解方法

1. 设 $\Omega \subset \Re^n$ 为闭凸集. $F(u) = Mu + q$. $M \in \Re^{n\times n}$, $M^T + M \succeq 0$, $q \in \Re^n$. 对给定的常数 $\beta > 0$, $\tilde{u}^k = P_\Omega[u^k - \beta(Mu^k + q)]$. 假如

$$u^{k+1} = u^k - \alpha(I+\beta M)^{-1}(u^k - \tilde{u}^k), \quad \alpha \in (0,2), \tag{B.1}$$

证明迭代序列 $\{u^k\}$ 满足

$$\|u^{k+1} - u^*\|_G^2 \le \|u^k - u^*\|_G^2 - \alpha(2-\alpha)\|u^k - \tilde{u}^k\|^2, \tag{B.2}$$

其中　$G = (I+\beta M^T)(I+\beta M)$.

2. 设 $\Omega \subset \Re^n$ 为闭凸集. $F(u) = Hu + q$. $H \in S_+^n$, $q \in \Re^n$. 对给定的 $u^k \in \Re^n$ 和常数 $\beta > 0$, 预测点由 $\tilde{u}^k = P_\Omega[u^k - \beta(Hu^k + q)]$ 给定. 假如

$$u^{k+1} = u^k - \gamma\alpha_k^*(u^k - \tilde{u}^k), \quad \alpha_k^* = \frac{\|u^k - \tilde{u}^k\|^2}{(u^k - \tilde{u}^k)(I+\beta H)(u^k - \tilde{u}^k)}, \quad \gamma \in (0,2), \tag{B.3}$$

那么有

$$\|u^{k+1} - u^*\|_{(I+\beta H)}^2 \le \|u^k - u^*\|_{(I+\beta H)}^2 - \gamma(2-\gamma)\alpha_k^*\|u^k - \tilde{u}^k\|^2. \tag{B.4}$$

3. 接上题. 如果投影 $\tilde{u}^k = P_\Omega[u^k - \beta(Hu^k + q)]$ 总满足

$$(u^k - \tilde{u}^k)^T\beta H(u^k - \tilde{u}^k) \le \nu\|u^k - \tilde{u}^k\|^2, \quad \nu \in (0,1) \tag{B.5}$$

那么, 直接取 $u^{k+1} = \tilde{u}^k$ 的方法是收敛的, 请给予证明.

4. 对一般的单调线性变分不等式, 用迭代法 $u^{k+1} = P_\Omega[u^k - \beta(Mu^k + q)]$ 是不能保证收敛的. 例如,

$$M = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad q = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \Omega = \{u \in \Re^2 \mid \|u\| \le 1\}. \tag{B.6}$$

试证明 (i) $u^* = (0,0)^T$ 是问题的唯一解; (ii)) $u^0 = (1,0)^T$ 为初始点, 所述方法一定不收敛.

练习题 5-6 讨论非线性单调变分不等式: $u^* \in \Omega, (u-u^*)^T F(u^*) \ge 0,\ \forall u \in \Omega$ 的求解方法

5. 设 $\Omega \subset \Re^n$ 为闭凸集. $F(u): \Re^n \to \Re^n$ 是单调算子. 对给定的 $u^k \in \Re^n$ 和常数 $\beta > 0$, 预测点由 $\tilde{u}^k = P_\Omega[u^k - \beta F(u^k)]$ 给定. 设 u^{k+1} 是非线性方程组

$$u + \beta F(u) = u^k + \beta F(u^k) - \alpha(u^k - \tilde{u}^k), \quad \alpha \in (0,2) \tag{B.7}$$

的解. 请证明

$$\|(u^{k+1}-u^*)+\beta(F(u^{k+1})-F(u^*))\|^2 \le \|(u^k-u^*)+\beta(F(u^k)-F(u^*))\|^2 - \alpha(2-\alpha)\|u^k-\tilde{u}^k\|^2. \tag{B.8}$$

说明这种求解方法的主要意义, 以及它与习题 1 之间的关系.

6. 考虑求解非线性单调变分不等式 $u^* \in \Omega, (u-u^*)^T F(u^*) \ge 0,\ \forall u \in \Omega$ 的方法.

若采用显式迭代法　$u^{k+1} = P_\Omega[u^k - \beta F(u^k)]$,　容易实现却不能保证收敛;

若采用隐式迭代法　$u^{k+1} = P_\Omega[u^k - \beta F(u^{k+1})]$,　能保证收敛却不容易实现.

Korpelevich 方法采用显式方法预测, 再用隐式方法校正, 也就是通过

$$\textbf{[Korpelevich]} \quad \begin{cases} \tilde{u}^k = P_\Omega[u^k - \beta_k F(u^k)], & \text{(B.9a)} \\ u^{k+1} = P_\Omega[u^k - \beta_k F(\tilde{u}^k)] & \text{(B.9b)} \end{cases}$$

得到新的迭代点. 如果预测投影满足

$$\beta_k\|F(u^k)-F(\tilde u^k)\|\le \nu\|u^k-\tilde u^k\|,\quad \nu\in(0,1). \tag{B.10}$$

请证明序列 $\{u^k\}$ 满足

$$\|u^{k+1}-u^*\|^2\le\|u^k-u^*\|^2-(1-\nu^2)\|u^k-\tilde u^k\|^2\quad \forall u^*\in\Omega^* \tag{B.11}$$

练习题 7-10 讨论求解单调线性变分不等式姊妹方法的一些性质

设 $\tilde u^k=P_\Omega[u^k-\beta(Mu^k+q)]$. 求解单调线性变分不等式姊妹方法生成新迭代点的方式为:

$$\text{(LVI-Alg-1)}\qquad u_I^{k+1}=u^k-\gamma\alpha_k^*(I+\beta M^T)(u^k-\tilde u^k), \tag{B.12}$$

$$\text{(LVI-Alg-2)}\qquad u_{II}^{k+1}=P_\Omega\big\{u^k-\gamma\alpha_k^*\beta[M^T(u^k-\tilde u^k)+(Mu^k+q)]\big\} \tag{B.13}$$

其中 $\alpha_k^*=\dfrac{\|u^k-\tilde u^k\|^2}{\|(I+\beta M^T)(u^k-\tilde u^k)\|^2},\quad \gamma\in(0,2).$

7. 对给定的 $u^k\in\Re^n$, s $\tilde u^k=P_\Omega[u^k-\beta(Mu^k+q)]$. 那么, 对任何 u, 不等式

$$(u-\tilde u^k)^T\beta(Mu+q)\ge(u-\tilde u^k)^T\beta\{(Mu^k+q)+M^T(u^k-\tilde u^k)\}-\|u^k-\tilde u^k\|_D^2 \tag{B.14}$$

总成立, 其中 $D=\frac12\beta(M^T+M)$ 是半正定矩阵.

8. 对给定的 $u^k\in\Re^n$, $\tilde u^k=P_\Omega[u^k-\beta(Mu^k+q)]$. 设 u^{k+1} 由方法 (B.12) 生成. 证明

$$\gamma\alpha_k^*(u-\tilde u^k)^T\beta(Mu+q)\ge\tfrac12\big(\|u-u^{k+1}\|^2-\|u-u^k\|^2\big)+\tfrac12\gamma(2-\gamma)\alpha_k^*\|u^k-\tilde u^k\|^2 \tag{B.15}$$

对任何的 $u\in\Omega$ 都成立. （提示：利用结论 (B.14)）.

9. 对给定的 $u^k\in\Omega$, $\tilde u^k=P_\Omega[u^k-\beta(Mu^k+q)]$. 设 u^{k+1} 由方法 (LVI-Alg-2) 生成. 证明

$$\gamma\alpha_k^*(u-\tilde u^k)^T\beta(Mu+q)\ge\tfrac12\big(\|u-u^{k+1}\|^2-\|u-u^k\|^2\big)+\tfrac12\gamma(2-\gamma)\alpha_k^*\|u^k-\tilde u^k\|^2 \tag{B.16}$$

对任何的 $u\in\Omega$ 都成立. （提示：利用结论 (B.14)）.

10. 由 (B.15) 和 (B.16) 可以得到线性变分不等式姊妹方法收敛速率的性质. 此外顺手还有副产品

$$\|u^{k+1}-u^*\|^2\le\|u^k-u^*\|^2-\gamma(2-\gamma)\alpha_k^*\|u^k-\tilde u^k\|^2,\quad \forall u^*\in\Omega. \tag{B.17}$$

练习题 11-13 讨论求解单调非线性变分不等式姊妹方法的一些性质

设 $\tilde u^k=P_\Omega[u^k-\beta_kF(u^k)]$ 满足 $\beta_k\|F(u^k)-F(\tilde u^k)\|\le\nu\|u^k-\tilde u^k\|$. 求解单调非线性变分不等式姊妹方法生成新迭代点的方式为:

$$\text{(NVI-Alg-1)}\qquad u_I^{k+1}=u^k-\gamma\alpha_k^*d(u^k,\tilde u^k), \tag{B.18}$$

$$\text{(NVI-Alg-2)}\qquad u_{II}^{k+1}=P_\Omega\big\{u^k-\gamma\alpha_k^*\beta_kF(\tilde u^k)\big\} \tag{B.19}$$

其中 $d(u^k,\tilde u^k)=(u^k-\tilde u^k)-\beta_k\big(F(u^k)-F(\tilde u^k)\big),\quad \alpha_k^*=\dfrac{(u^k-\tilde u^k)^Td(u^k,\tilde u^k)}{\|d(u^k,\tilde u^k)\|^2},\quad \gamma\in(0,2).$

11. 对给定的 $u^k\in\Re^n$, 由投影收缩算法 (B.18) 更新的 u^{k+1} 有

$$(u-\tilde u^k)^T\gamma\alpha_k^*\beta_kF(\tilde u^k)+\frac12\big(\|u-u^k\|^2-\|u-u^{k+1}\|^2\big)\ \ge\ \frac12\gamma(2-\gamma)(\alpha_k^*)^2\|d(u^k,\tilde u^k)\|^2,\quad \forall\, u\in\Omega. \tag{B.20}$$

Hint: 利用 (A1.7), 即 $(u-\tilde u^k)^T\beta_kF(\tilde u^k)\ge(u-\tilde u^k)^Td(u^k,\tilde u^k),\quad \forall u\in\Omega.$

12. 对给定的 $u^k\in\Re^n$, 由投影收缩算法 (B.19) 更新的 u^{k+1} 有

$$(u-\tilde u^k)^T\gamma\alpha_k^*\beta_kF(\tilde u^k)+\frac12\big(\|u-u^k\|^2-\|u-u^{k+1}\|^2\big)\ \ge\ \frac12\gamma(2-\gamma)(\alpha_k^*)^2\|d(u^k,\tilde u^k)\|^2,\quad \forall\, u\in\Omega. \tag{B.21}$$

Hint: 分拆 $(u-\tilde u^k)^T\gamma\alpha_k^*\beta_kF(\tilde u^k)=(u^{k+1}-\tilde u^k)^T\gamma\alpha_k^*\beta_kF(\tilde u^k)+(u-u^{k+1})^T\gamma\alpha_k^*\beta_kF(\tilde u^k).$
然后再处理右端那两个交叉项.

13. 由 (B.20) 和 (B.21) 可以得到非线性变分不等式姊妹方法收敛速率的性质. 此外顺手还有副产品

$$\|u^{k+1}-u^*\|^2\le\|u^k-u^*\|^2-\gamma(2-\gamma)\alpha_k^*(u^k-\tilde u^k)^Td(u^k,\tilde u^k),\quad \forall\, u^*\in\Omega.$$

C 单元习题　交替方向法(ADMM)和线性化交替方向法

练习题 1-3 讨论计算步长的 ADMM 和线性化的 ADMM 以及计算步长的线性化 ADMM

1. 考虑计算步长的ADMM方法, 它的每步迭代由预测和校正两部分组成. 首先, 将短篇 **C1** 中从 (y^k,λ^k) 出发, 通过 (C1.5) 求得的 $(x^{k+1},y^{k+1},\lambda^{k+1})$ 记为 $(\tilde{x}^k,\tilde{y}^k,\tilde{\lambda}^k)$, 当作预测点, 即

$$\begin{cases} \tilde{x}^k = \text{argmin}\{\theta_1(x) - (\lambda^k)^T(Ax+By^k-b) + \frac{1}{2}\beta\|Ax+By^k-b\|^2 \,|\, x\in\mathcal{X}\} & \text{(C.1a)}\\ \tilde{y}^k = \text{argmin}\{\theta_1(y) - (\lambda^k)^T(A\tilde{x}^k+By-b) + \frac{\beta}{2}\beta\|A\tilde{x}^k+By-b\|^2 \,|\, y\in\mathcal{Y}\} & \text{(C.1b)}\\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k+B\tilde{y}^k-b). & \text{(C.1c)}\end{cases}$$

记

$$v = \begin{pmatrix} y\\ \lambda\end{pmatrix}, \qquad H = \begin{pmatrix} \beta B^TB & 0\\ 0 & \frac{1}{\beta}I_m\end{pmatrix}. \tag{C.2}$$

(a) 请写出 (C.1) 的变分不等式形式, 证明 $(\tilde{v}^k-v^*)H(v^k-\tilde{v}^k) \ge (y^k-\tilde{y}^k)^TB^T(\lambda^k-\tilde{\lambda}^k)$. 并因此

$$(v^k-v^*)H(v^k-\tilde{v}^k) \ge \|v^k-\tilde{v}^k\|_H^2 + (y^k-\tilde{y}^k)^TB^T(\lambda^k-\tilde{\lambda}^k). \tag{C.3}$$

(b) 若记 $\varphi(v^k,\tilde{v}^k) = \|v^k-\tilde{v}^k\|_H^2 + (y^k-\tilde{y}^k)^TB^T(\lambda^k-\tilde{\lambda}^k)$, 证明

$$\varphi(v^k,\tilde{v}^k) \ge \tfrac{1}{2}\|v^k-\tilde{v}^k\|_H^2.$$

(c) 设生成新迭代点 v^{k+1} 的校正为

$$v^{k+1} = v^k - \gamma\alpha_k^*(v^k-\tilde{v}^k), \quad 其中\ \alpha_k^* = \frac{\varphi(v^k,\tilde{v}^k)}{\|v^k-\tilde{v}^k\|_H^2}, \ \ \gamma\in(0,2). \tag{C.4}$$

请给出这个方法收敛的关键不等式. (取 $\alpha=1.5\in(1.2,1.8)$, 效率比经典方法有 30% 的提高).

2. ADMM 方法中, 假设只需要对 y-子问题做线性化处理. 我们研究这一类线性化 ADMM 方法.

(a) 首先, 请证明经典 ADMM 方法中 y-子问题可以表示成

$$y^{k+1} = \arg\min\big\{\theta_2(y) - y^TB^T[\lambda^k - \beta(Ax^{k+1}+By^k-b)] + \tfrac{1}{2}\beta\|B(y-y^k)\|^2 \;\big|\; y\in\mathcal{Y}\big\}. \tag{C.5}$$

(b) 上面子问题的目标函数中既有凸函数 $\theta_2(y)$, 又有非平凡的二次函数 $\frac{1}{2}\beta\|B(y-y^k)\|^2$. 所谓线性化的 ADMM, 就是用简单的二次函数 $\frac{1}{2}s\|y-y^k\|^2$ 去代替 $\frac{1}{2}\beta\|B(y-y^k)\|^2$ $(s\ge\beta\|B^TB\|)$.

证明对仅做这点改动的线性化 ADMM 有

$$\begin{aligned} &w^{k+1}\in\Omega,\ \ \theta(u)-\theta(u^{k+1}) + (w-w^{k+1})^TF(w^{k+1})\\ &\quad \ge \beta\begin{pmatrix} x-x^{k+1}\\ y-y^{k+1}\end{pmatrix}^T\begin{pmatrix} A^T\\ B^T\end{pmatrix}B(y^{k+1}-y^k) + (v-v^{k+1})^TH(v^k-v^{k+1}), \ \ \forall\, w\in\Omega, \end{aligned} \tag{C.6a}$$

其中

$$v = \begin{pmatrix} y\\ \lambda\end{pmatrix}, \quad H = \begin{pmatrix} \beta B^TB + D_B & 0\\ 0 & \frac{1}{\beta}I_m\end{pmatrix}, \quad D_B = sI_{n_2} - \beta B^TB. \tag{C.6b}$$

(c) 证明对上面的矩阵 H, 有

$$(v^{k+1}-v^*)^TH(v^k-v^{k+1}) \ge (\lambda^k-\lambda^{k+1})^TB(y^k-y^{k+1}), \ \ \forall w^*\in\Omega^*. \tag{C.7}$$

(d) 证明这样的线性化 ADMM 产生的序列 $\{v^k\}$ 对任何满足

$$(\lambda^k-\lambda^{k+1})^TB(y^k-y^{k+1}) \ge \tfrac{1}{2}\|y^k-y^{k+1}\|_{D_B}^2 - \tfrac{1}{2}\|y^{k-1}-y^k\|_{D_B}^2. \tag{C.8}$$

(e) 据此证明收敛性的关键不等式: 线性化 ADMM 产生的序列 $\{v^k\}$, 对任何 $v^*\in\mathcal{V}^*$, 满足

$$\big(\|v^{k+1}-v^*\|_H^2 + \|y^k-y^{k+1}\|_{D_B}^2\big) \le \big(\|v^k-v^*\|_H^2 + \|y^{k-1}-y^k\|_{D_B}^2\big) - \|v^k-v^{k+1}\|_H^2. \tag{C.9}$$

3. 考虑计算步长的线性化 ADMM 方法, 它的每步迭代由预测和校正两部分组成. 首先, 将短篇 **C1** 中从 (y^k,λ^k) 出发, 通过 (C1.5) 求得的 $(x^{k+1},y^{k+1},\lambda^{k+1})$ 记为 $(\tilde{x}^k,\tilde{y}^k,\tilde{\lambda}^k)$, 当作预测点, 即

$$\begin{cases} \tilde{x}^k = \text{argmin}\{\theta_1(x) - (\lambda^k)^T(Ax+By^k-b) + \frac{1}{2}\beta\|Ax+By^k-b\|^2 \,|\, x\in\mathcal{X}\} & \text{(C.10a)}\\ \tilde{y}^k = \text{argmin}\{\theta_1(y) - (\lambda^k)^T(A\tilde{x}^k+By-b) + \frac{1}{2}\beta\|A\tilde{x}^k+By-b\|^2 + \frac{1}{2}\|y-y^k\|_{D_B}^2 \,|\, y\in\mathcal{Y}\} & \text{(C.10b)}\\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k+B\tilde{y}^k-b). \qquad (D_B = sI_{n_2}-\beta B^TB). & \text{(C.10c)}\end{cases}$$

根据上一题的 (C.6), 就有

$$\tilde{w}^k \in \Omega,\ \ \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \\ \ge \beta \begin{pmatrix} x-\tilde{x}^k \\ y-\tilde{y}^k \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(\tilde{y}^k - y^k) + (v-\tilde{v}^k)^T H(v^k-\tilde{v}^k),\ \ \forall\, w\in\Omega, \tag{C.11a}$$

其中

$$v=\begin{pmatrix} y \\ \lambda \end{pmatrix},\quad H=\begin{pmatrix} \beta B^T B + D_B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix},\quad D_B = sI_{n_2} - \beta B^T B. \tag{C.11b}$$

请据此跟第 1 题那样设计计算步长的线性化 ADMM, 并给出关键收敛性质的证明.

练习题 4-5 讨论矩阵的 Frobenius 范数和实对称矩阵在 S_+^n 的投影

4. 设 $A=(a_{ij})\in\Re^{m\times n}$, $\|A\|_F=\left(\sum_{i=1}^m\sum_{j=1}^n a_{ij}^2\right)^{1/2}$. 证明对任意的 $m\times m$ 正交矩阵 U 和 $n\times n$ 正交矩阵 V, 有
$$\|U^T AV\|_F = \|A\|_F. \tag{C.12}$$

5. 记 $S^n=\{X\in\Re^{n\times n}\mid X^T=X\}$, $S_+^n=\{X\in\Re^{n\times n}\mid X^T=X,\ X\succeq 0\}$. 对给定的 $A\in S^n$, 请利用 $A=V\Lambda V^T$, 给出问题
$$\min\{\|X-A\|_F\,|X\in S_+^n\}\quad \text{的解}. \tag{C.13}$$

练习题 6-8 讨论用 ADMM 求解一些矩阵优化问题

6. 给定实对称矩阵 A, 考虑求解凸优化问题:
$$\min\left\{\tfrac{1}{2}\|X-A\|_F^2 \,\middle|\, X\in S_+^n\cap S_B\right\}, \tag{C.14a}$$
其中
$$S_+^n=\{H\in R^{n\times n}\mid H^T=H,\ H\succeq 0\},\quad S_B=\{H\in R^{n\times n}\mid H^T=H,\ H_L\le H\le H_U\}. \tag{C.14b}$$
H_L 和 H_U 都是给定的实对称矩阵. 直接求解问题 (C.14) 有一定困难, 因此转化成
$$\begin{array}{ll} \min & \frac{1}{2}\|X-A\|_F^2| + \frac{1}{2}\|Y-A\|_F^2 \\ & X-Y=0, \\ & X\in S_+^n,\ Y\in S_B \end{array} \tag{C.15}$$
   (a) 请描述用 ADMM 求解问题 (C.15) 的基本步骤.
   (b) 若用
```
rand('state',0);     A=rand(n,n);  A=(A'+A)-ones(n,n) + eye(n);
%%% A is symmetric and A_{ij} is in (-1,1),  A_{jj} is in (0,2)  %%
HU=ones(n,n)*0.1;  HL=-HU; for i=1:n   HU(i,i)=1;  HL(i,i)=1; end;
```
   生成矩阵 A, H_L 和 H_U, 请编程实现.
   (c) 按照第 1 题的方法, 对上面的程序稍作修改, 比较一下收敛效果.

7. 设 C 和 X 都是实对称矩阵. 证明 Tr(CX) 是 X 的凸函数, $-\log\det(X)$ 是 X 在 S_{++}^n 的凸函数. 请用矩阵形式写出它们的梯度.

8. 给定实对称矩阵 C 和常数 $\rho>0$, 考虑求解凸优化问题:
$$\min\ \{\mathrm{Tr}(CX)-\log\det(X)+\rho e^T|X|e\,|\,X\in S_{++}^n\}, \tag{C.16}$$
其中 e 是每个分量都为 1 的 n-维列向量. $|X|$ 表示把矩阵 X 的每个分量都取绝对值的矩阵.
直接求解 (C.16) 有一定困难, 因此转化成
$$\begin{array}{ll} \min & \mathrm{Tr}(CX)-\log\det(X)+\rho e^T|Y|e \\ & X-Y=0, \\ & X\in S_{++}^n,\ Y\in S^n. \end{array} \tag{C.17}$$
请描述用 ADMM 求解问题 (C.17) 的基本步骤.

D 单元习题　邻近点算法 (PPA) 求解矩阵优化问题及一些技巧

练习题 1 讨论延伸 ADMM 的性质

1. 单元 **D** 中两个短篇提到 PPA 都建议采用 $w^{k+1} := w^k - \alpha(w^k - w^{k+1}),\ \alpha = 1.5 \in (0, 2)$. 称这种方法为延伸的 PPA 方法. 我们将经典的 PPA 输出记成 $\tilde{w}^k$：

   (a) 证明 这样定义的 $\tilde{w}^k$ 满足:

   $$(w^k - w^*)^T H(w^k - \tilde{w}^k) \geq \|w^k - \tilde{w}^k\|_H^2,\ \ \forall w^* \in \Omega^*. \tag{D.1}$$

   (b) 说明采用延伸 PPA 的意义并证明方法生成的序列满足

   $$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \alpha(2-\alpha)\|w^k - \tilde{w}^k\|_H^2,\ \ \forall w^* \in \Omega^*. \tag{D.2}$$

练习题 2 介绍回顾矩阵奇异值分解的相关知识　(由于矩阵优化中矩阵核模定义的需要)

2. 设矩阵 $A = (a_{ij}) \in \Re^{m\times n}$ 的秩为 r. 矩阵 A 的奇异值分解是指存在 $m\times m$ 正交矩阵 U 和 $n\times n$ 正交矩阵 V, 使得 $A = U\Sigma V^T$ [(orthogonal)(diagonal)(orthogonal)], 其中

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \sigma_2 & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \sigma_r & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & 0 & 0 \end{pmatrix}.$$

人们记 $\|A\|_* = \sum_{i=1}^r \sigma_i$, 并称其为矩阵 A 的核模.

(a) 根据奇异值分解 $A = U\Sigma V^T$ 说明:

first	r	column of U	column space of A	$C(A)$
last	$m-r$	column of U	left nullspace of A	$N(A^T)$
first	r	column of V	row space of A^T	$C(A^T)$
last	$n-r$	column of V	nullspace of A	$N(A)$

(D.3)

(b) 设 $\sigma_1^2, \ldots, \sigma_r^2$ 是对称矩阵 A^TA 的非零特征值. $V = [v_1, v_2, \cdots, v_r, v_{r+1}, \cdots, v_n]$ 是相应的标准化特征向量. 证明

对 $j = 1, \ldots, r,\ \frac{1}{\sigma_j}Av_j$ 是 AA^T 的标准化特征向量.　$Av_j = 0,\ for\ \ j = r+1, \ldots, n$

(c) 记 $(1/\sigma_j)Av_j = u_j$, 则有

$$\begin{aligned} AV &= A[v_1, v_2, \cdots, v_r, v_{r+1}, \cdots, v_n] = [u_1, u_2, \cdots, u_r, 0, \cdots, 0] \\ &= [u_1, u_2, \cdots, u_r, 0, \cdots, 0]\begin{pmatrix} \sigma_1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \sigma_2 & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \sigma_r & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & 0 & 0 \end{pmatrix} = U\Sigma. \end{aligned} \tag{D.4}$$

请说说你补全矩阵 U 的方案, 并证明正交变换下矩阵的奇异值不变.

练习题 3 讨论用 PPA 求解一些矩阵优化问题

3. 对给定的 $C \in S^n$, 采用短篇 **D1** 中介绍的 PPA 算法求解问题

$$\min\{\tfrac{1}{2}\|X - C\|_F^2 \mid \mathrm{diag}(X) = e, X \in S_+^n\}, \tag{D.5}$$

其中 e 是每个分量都为 1 的 n-维向量.

(a) 说明为什么线性约束 $\mathrm{diag}(X) = e$ 隐含了 $\|A^TA\| = 1$, 并写出用 PPA 求解问题 (D.5) 的基本步骤.

(b) 由于 $\|A^TA\| = 1$, 只需要设 $rs > 1$ 就行. 若用

```
clear;  close all; n = 1000;     tol=1e-5;     r=2.0;    s=1.05/r;
 rand('state',0);  C=rand(n,n);   C=(C'+C)-ones(n,n) + eye(n);
```

生成矩阵 C, 并用上面给定 r 和 s, 请编程实现求解问题 (D.5).

(c) 按照第 1 题的方法, 取 $\alpha = 1.5$, 对上面的程序稍作修改, 比较一下收敛效果.

练习题 4 讨论求解 $\min\{\theta(x) \mid Ax = b,\ x \in X\}$ 的几种典型的 PPA 方法

我们已经把线性约束的凸优化问题 $\min\{\theta(x) \mid Ax = b, x \in X\}$ 的鞍点转化成变分不等式

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega \tag{D.6a}$$

其中

$$w = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ Ax - b \end{pmatrix} \quad \text{and} \qquad \Omega = \mathcal{X} \times \Re^m. \tag{D.6b}$$

4. 请写出求解变分不等式 (G.1) 的下述几类 PPA 型迭代法的变分不等式形式并指出方法的优缺点.

- Augmented Lagrangian Method (关于乘子 λ 的 PPA)

$$\begin{cases} x^{k+1} \in \arg\min\{\theta(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax - b\|^2 | x \in \mathcal{X}\} & \text{(D.7a)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} - b). & \text{(D.7b)} \end{cases}$$

- Customized Proximal Point Algorithm (关于变量 w 的 PPA)

$$\begin{cases} x^{k+1} = \arg\min\{\theta(x) - x^T A^T \lambda^k + \frac{r}{2}\|x - x^k\|^2 \,| x \in \mathcal{X}\} & \text{(D.8a)} \\ \lambda^{k+1} = \lambda - \frac{1}{s}\big(A[2x^{k+1} - x^k] - b\big). & \text{(D.8b)} \end{cases}$$

- Balanced ALM (关于变量 w 的 PPA)

$$\begin{cases} x^{k+1} = \text{argmin}\big\{\theta(x) - x^T A^T \lambda^k + \frac{r}{2}\|x - x^k\|^2 \,|\, x \in \mathcal{X}\big\}, & \text{(D.9a)} \\ \lambda^{k+1} = \lambda^k - (\frac{1}{r}AA^T + \delta I)^{-1}\big(A[2x^{k+1} - x^k] - b\big). & \text{(D.9b)} \end{cases}$$

练习题 5 讨论均困 PPA 求解优化问题需要注意的一些技巧

5. 采用“均困”方法 (D.9), 主要工作量之一是求 $(\frac{1}{r}AA^T + \delta I)^{-1}$. 假如我们处理的是运输问题的约束集合 $\{Ax = b, x \in \mathcal{X}\}$ 是

$$\sum_{j=1}^{n} x_{ij} = s_i, \;\; i = 1, \cdots, m; \quad \sum_{i=1}^{n} x_{ij} = d_j, \;\; j = 1, \cdots, n; \qquad x_{ij} \ge 0. \tag{D.10}$$

换句话说,

$$A = \begin{pmatrix} e_n^T & 0 & \cdots & 0 \\ 0 & e_n^T & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & e_n^T \\ I_n & I_n & \cdots & I_n \end{pmatrix} \in \Re^{(m+n)\times mn}, \tag{D.11}$$

其中 e_n 是每个分量都是 1 的 n-维向量. 注意到 A 的上面有 m 行.

(a) 证明 $(m+n) \times (m+n)$ 矩阵

$$AA^\top = \begin{pmatrix} nI_m & e_m e_n^T \\ e_n e_m^T & mI_n \end{pmatrix} = \begin{pmatrix} nI_m & 0 \\ 0 & mI_n \end{pmatrix} + \begin{pmatrix} 0 & e_m e_n^T \\ e_n e_m^T & 0 \end{pmatrix}. \tag{D.12}$$

上式右端的是第二部分是个秩 2 矩阵, 可以表示成 $(m+n) \times 2$ 和 $2 \times (m+n)$ 矩阵的乘积.

$$\begin{pmatrix} 0 & e_m e_n^T \\ e_n e_m^T & 0 \end{pmatrix} = \begin{pmatrix} e_m & 0 \\ 0 & e_n \end{pmatrix} \begin{pmatrix} 0 & e_n^T \\ e_m^T & 0 \end{pmatrix}. \tag{D.13}$$

(b) 关于秩二校正: 若矩阵 $C = B + UV^T$, B, C 都可逆, 则有

$$C^{-1} = B^{-1} - B^{-1}U(I + V^T B^{-1} U)^{-1} V^T B^{-1}. \tag{D.14}$$

据此给出矩阵 $(\delta I + \frac{1}{r}AA^T)^{-1}$ 的显式表达式.

E 单元习题　根据统一框架研究 ADMM 类算法及 Glowinski 方法

练习题 1 讨论交换 y, λ 更新顺序的 ADMM 并对 y-子问题线性化

1. 在交换 y, λ 更新顺序的 ADMM (见短篇 **E1** 的 (E1.4)) 的迭代公式为

$$\begin{cases} x^{k+1} \in \text{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k - b\|^2 \,|\, x \in \mathcal{X}\}, & \text{(E.1a)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^k - b), & \text{(E.1b)} \\ y^{k+1} \in \text{argmin}\{\theta_2(y) - y^T B^T \lambda^{k+1} + \frac{1}{2}\beta\|Ax^{k+1} + By - b\|^2 \,|\, y \in \mathcal{Y}\}. & \text{(E.1c)} \end{cases}$$

(a) 证明 (E.1) 中 y-子问题可以等价地表示为

$$y^{k+1} \in \text{argmin}\{\theta_2(y) - y^T B^T(2\lambda^{k+1} - \lambda^k) + \frac{1}{2}\beta\|B(y - y^k)\|^2 \,|\, y \in \mathcal{Y}\}. \tag{E.2}$$

(b) 对 y-子问题线性化就是用 $\frac{1}{2}s\|y - y^k\|^2$ 去代替 $\frac{1}{2}\beta\|B(y - y^k)\|^2$ $(s \geq \beta\|B^T B\|)$. 证明这样迭代的变分不等式形式是

$$\theta(u) - (u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \geq (v - v^{k+1})^T H(v^k - v^{k+1}), \ \ \forall w \in \Omega, \tag{E.3a}$$

其中

$$v = \begin{pmatrix} y \\ \lambda \end{pmatrix}, \quad H = \begin{pmatrix} sI_{n_2} & -B^T \\ -B & \frac{1}{\beta}I_m \end{pmatrix} \quad \text{为正定矩阵.} \tag{E.3b}$$

并因此而说明可以进一步延拓校正: $v^{k+1} := v^k - \alpha(v^k - v^{k+1}), \ \alpha \in (0, 2)$.

练习题 2 讨论在经典 ADMM 基础上对子问题目标函数加小参数正则项的方法

2. 经典的 ADMM 是从给定的 $v^k = (y^k, \lambda^k)$ 出发, 通过

$$\begin{cases} x^{k+1} \in \text{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k - b\|^2 \,|\, x \in \mathcal{X}\}, & \text{(E.4a)} \\ y^{k+1} \in \text{argmin}\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}\beta\|Ax^{k+1} + By - b\|^2 \,|\, y \in \mathcal{Y}\}, & \text{(E.4b)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & \text{(E.4c)} \end{cases}$$

(a) 对方法 (E.4) 的 x, y-子问题目标函数加上正则项, 迭代就从 $w^k = (x^k, y^k, \lambda^k)$ 出发, 由

$$\begin{cases} x^{k+1} \in \text{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k - b\|^2 + \frac{\delta_r}{2}\|x - x^k\|^2 \,|\, x \in \mathcal{X}\}, & \text{(E.5a)} \\ y^{k+1} \in \text{argmin}\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}\beta\|Ax^{k+1} + By - b\|^2 + \frac{\delta_s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}, & \text{(E.5b)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b) & \text{(E.5c)} \end{cases}$$

求得 $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$. 小参数 $\delta_r > 0$ 和 $\delta_s > 0$ 会改善子问题求解的条件.
若将算法 (E.5) 拆解成预测-校正方法, 取

$$\tilde{x}^k = x^{k+1}, \qquad \tilde{y}^k = y^{k+1}, \qquad \tilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b) \tag{E.6}$$

为预测点. 预测的变分不等式形式写成

$$\theta(u) - (\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (w - \tilde{w}^k)^T Q(w^k - \tilde{w}^k), \ \ \forall w \in \Omega \tag{E.7}$$

请给出 (E.7) 中矩阵 Q.

(b) 算法 (E.5) 按 (E.6) 预测拆解后, 其校正公式为 $w^{k+1} = w^k - M(w^k - \tilde{w}^k)$. 请写出校正矩阵 M 的具体形式.

(c) 对于这对预测-校正矩阵 Q 和 M, 请给出满足 $HM = Q$ 的正定矩阵 H. 证明上面得到的 H, M 和 Q, 满足 $G = Q^T + Q - M^T HM \succ 0$, 迭代序列满足

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \|w^k - \tilde{w}^k\|_G^2, \quad \forall w \in \Omega.$$

练习题 3 研究 Glowinski's ADMM 的关键收敛性质

Glowinski 的 ADMM 跟我们在 **C** 单元讨论的 ADMM 方法, 也就是这里的 (E.4), 差别只是 Lagrange 乘子的更新用了一个松弛因子. 具体就是

$$\begin{cases} x^{k+1} \in \operatorname{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k - b\|^2 \,|\, x \in \mathcal{X}\}, & \text{(E.8a)} \\ y^{k+1} \in \operatorname{argmin}\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}\beta\|Ax^{k+1} + By - b\|^2 \,|\, y \in \mathcal{Y}\}, & \text{(E.8b)} \\ \lambda^{k+1} = \lambda^k - \gamma\beta(Ax^{k+1} + By^{k+1} - b), \qquad \gamma \in (0, \frac{1+\sqrt{5}}{2}). & \text{(E.8c)} \end{cases}$$

(a) 根据 y-问题的最优性条件, 对每个 $k > 1$, 我们有 $y^{k+1} \in \mathcal{Y}$,

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{-B^T\lambda^k + \beta B^T(Ax^{k+1} + By^{k+1} - b)\} \geq 0, \quad \forall y \in \mathcal{Y}. \quad \text{(E.9)}$$

利用 $\lambda^k = \lambda^{k-1} - \gamma\beta(Ax^k + By^k - b)$, 证明

$$(Ax^{k+1} + By^{k+1} - b)^T B(y^k - y^{k+1}) \geq (1-\gamma)(Ax^k + By^k - b)^T B(y^k - y^{k+1}). \quad \text{(E.10)}$$

(b) 将 (E.8) 的迭代用 (E.6) 定义的预测分拆成预测和校正. 证明预测和校正分别为

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \ \forall w \in \Omega \quad \text{(E.11a)}$$

和

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \quad \text{(E.11b)}$$

其中

$$Q = \begin{pmatrix} \beta B^T B & 0 \\ -B & \frac{1}{\beta}I \end{pmatrix} \qquad \text{和} \qquad M = \begin{pmatrix} I & 0 \\ -\gamma\beta B & \gamma I \end{pmatrix}. \quad \text{(E.11c)}$$

(c) 若令 $H = \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\gamma\beta}I \end{pmatrix}$, 请证明

$$HM = Q, \qquad G = Q^T + Q - M^T HM = \begin{pmatrix} (1-\gamma)\beta B^T B & (\gamma-1)B^T \\ (\gamma-1)B & (2-\gamma)\frac{1}{\beta}I \end{pmatrix}. \quad \text{(E.12)}$$

并有

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \geq \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \quad \text{(E.13)}$$

(d) 利用 (E.12) 中的矩阵 G, 证明

$$\|v^k - \tilde{v}^k\|_G^2 = (1-\gamma)\frac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k - \beta B(y^k - \tilde{y}^k)\|^2 + \frac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2. \quad \text{(E.14)}$$

因此当 $\gamma \in (0, 1]$ 时收敛. 对 $\gamma > 1$, 利用 $\lambda^k - \tilde{\lambda}^k = \beta(Ax^{k+1} + By^k - b)$ 证明

$$\|v^k - \tilde{v}^k\|_G^2 = \beta\|Ax^{k+1} + By^k - b\|^2 + (1-\gamma)\beta\|Ax^{k+1} + By^{k+1} - b\|^2. \quad \text{(E.15)}$$

(e) 证明: 从 (E.15) 和 (E.10) 得到

$$\begin{aligned} \|v^k - \tilde{v}^k\|_G^2 \geq & (2-\gamma)\beta\|Ax^{k+1} + By^{k+1} - b\|^2 + \beta\|B(y^k - y^{k+1})\|^2 \\ & + 2\beta(1-\gamma)(Ax^k + By^k - b)^T B(y^k - y^{k+1}) \end{aligned} \quad \text{(E.16)}$$

(f) 利用 Cauchy-Schwarz 不等式, 从 (E.16) 右端最后的交叉项得到 ($p > 0$ 为任意实数)

$$\begin{aligned} & 2\beta(1-\gamma)(Ax^k + By^k - b)^T B(y^k - y^{k+1}) \\ & \quad \geq -\tfrac{\gamma-1}{p}\beta\|Ax^k + By^k - b\|^2 - (\gamma-1)p\beta\|B(y^k - y^{k+1})\|^2. \end{aligned} \quad \text{(E.17)}$$

(g) 从 (E.13), (E.16) 和 (E.17) 出发, 说明为什么取 $p = \gamma$? 从而使得

$$\begin{aligned} & \big(\|v^k - v^*\|_H^2 + \tfrac{\gamma-1}{\gamma}\|Ax^k + By^k - b\|^2\big) - \big(\|v^{k+1} - v^*\|_H^2 + \tfrac{\gamma-1}{\gamma}\beta\big(\|Ax^{k+1} + By^{k+1} - b\|^2\big) \\ & \quad \geq \tfrac{1+\gamma-\gamma^2}{\gamma}\beta\big(\|Ax^{k+1} + By^{k+1} - b\|^2 + \gamma\|B(y^k - y^{k+1})\|^2\big). \end{aligned} \quad \text{(E.18)}$$

(h) 最后得到当 $\gamma \in [1, \frac{1+\sqrt{5}}{2})$ 收敛的关键不等式

$$\begin{aligned} & \big(\|v^{k+1} - v^*\|_H^2 + \tfrac{\gamma-1}{\gamma}\beta(\|A(x^{k+1} - x^*) + B(y^{k+1} - y^*)\|^2\big) \\ & \quad \leq \big(\|v^k - v^*\|_H^2 + \tfrac{\gamma-1}{\gamma}\beta\|A(x^k - x^*) + B(y^k - y^*)\|^2\big) \\ & \qquad - \tfrac{1+\gamma-\gamma^2}{\gamma}\beta(\|Ax^{k+1} + By^{k+1} - b\|^2 + \gamma\|B(y^k - y^{k+1})\|^2). \end{aligned} \quad \text{(E.19)}$$

F 单元习题　统一框架下三个可分离块问题的 ADMM 类方法

F 单元讨论的方法中, 预测矩阵满足 Q^T+Q 正定条件下, 校正 $v^{k+1}=v^k-M(v^k-\tilde v^k)$ 在条件

$$H\succ 0,\quad HM=Q,\quad G=Q^T+Q-M^THM\succ 0$$

的条件下进行. 这单元的习题中, 校正用 $v^{k+1}=v^k-\alpha M(v^k-\tilde v^k)$ 实现, 仍然需要

$$H\succ 0,\quad 和\quad HM=Q.\quad 但是\quad G=Q^T+Q-M^THM\quad 不一定正定.$$

这时可以通过计算确定每一步的步长 α_k, 实现序列 $\{\|v^k-v^*\|_H^2\}$ 的单调下降 (序列 $\{v^k\}$ 向 v^* 收缩).

练习题 1-2 讨论三个可分离块基于逐个向前预测的方法 分别采用计算步长和确定步长的校正

F1 中基于直接推广的 ADMM 求解三个可分离块问题的预测可以写成

$$\textbf{[预测]}\quad\left\{\begin{array}{ll}\tilde x^k\in\arg\min\big\{\theta_1(x)-x^TA^T\lambda^k+\frac12\beta\|Ax+By^k+Cz^k-b\|^2\mid x\in\mathcal X\big\}, & \text{(F.1a)}\\ \tilde y^k\in\arg\min\big\{\theta_2(y)-y^TB^T\lambda^k+\frac12\beta\|A\tilde x^k+By+Cz^k-b\|^2\mid y\in\mathcal Y\big\}, & \text{(F.1b)}\\ \tilde z^k\in\arg\min\big\{\theta_3(z)-z^TC^T\lambda^k+\frac12\beta\|A\tilde x^k+B\tilde y^k+Cz-b\|^2\mid z\in\mathcal Z\big\}, & \text{(F.1c)}\\ \tilde\lambda^k=\lambda^k-\beta\big(A\tilde x^k+By^k+Cz^k-b\big). & \text{(F.1d)}\end{array}\right.$$

预测的变分不等式形式可以写成

$$\tilde w^k\in\Omega,\quad \theta(u)-\theta(\tilde u^k)+(w-\tilde w^k)^TF(\tilde w^k)\ge(v-\tilde v^k)^TQ(v^k-\tilde v^k),\quad\forall w\in\Omega,\tag{F.2a}$$

其中

$$Q=\begin{pmatrix}\beta B^TB & 0 & 0\\ \beta C^TB & \beta C^TC & 0\\ -B & -C & \frac1\beta I_m\end{pmatrix}.\tag{F.2b}$$

1. 我们讨论基于上述逐个向前预测, 再进行校正的方法.

   (a) 若取 （请注意下面的校正矩阵 M 与短篇 **F1** 中 M 矩阵的差别）

   $$H=\begin{pmatrix}\beta B^TB & \beta B^TC & 0\\ \beta C^TB & \beta[C^TC+C^TB(B^TB)^{-1}B^TC] & 0\\ 0 & 0 & \frac1\beta I\end{pmatrix},\quad M=\begin{pmatrix}I & -(B^TB)^{-1}B^TC & 0\\ 0 & I & 0\\ -\beta B & -\beta C & I\end{pmatrix}.\tag{F.3}$$

   证明: 当 B,C 列满秩时 H 正定, $HM=Q$. 并计算 $G=Q^T+Q-M^THM$.

   (b) 采用校正

   $$v^{k+1}=v^k-\gamma\alpha_k^*M(v^k-\tilde v^k),\quad \alpha_k^*=\frac12(\|v^k-\tilde v^k\|^2_{[Q^T+Q]})/(\|M(v^k-\tilde v^k)\|_H^2),\quad\gamma\in(0,2)\tag{F.4}$$

   证明

   $$\|v^{k+1}-v^*\|_H^2\le\|v^k-v^*\|_H^2-\tfrac12\gamma(2-\gamma)\alpha_k^*\|v^k-\tilde v^k\|^2_{[Q^T+Q]}.\tag{F.5}$$

   (c) 若用 (F.3) 中的矩阵 M 和校正公式

   $$v^{k+1}=v^k-\frac12\gamma M(v^k-\tilde v^k),\quad\gamma\in(0,2)\tag{F.6}$$

   证明方法收敛的关键不等式

   $$\|v^{k+1}-v^*\|_H^2\le\|v^k-v^*\|_H^2-\tfrac14\gamma(2-\gamma)\big(\beta\|B(y^k-\tilde y^k)\|^2+\beta\|C(z^k-\tilde z^k)\|^2+\frac1\beta\|\lambda^k-\tilde\lambda^k\|^2\big).\tag{F.7}$$

2. 由于预测 (F.1) 可以从给定的 (By^k,Cz^k,λ^k) 开始, 校正也只需提供 $(By^{k+1},Cz^{k+1},\lambda^{k+1})$. 将 (F.2) 中的 w 设为 w^*, 就有 $(\tilde v^k-v^*)^TQ(v^k-\tilde v^k)\ge 0,\ \forall v^*\in\mathcal V^*$. 再做变换, 设

   $$\xi=\begin{pmatrix}By\\ Cz\\ \lambda\end{pmatrix},\qquad \mathcal Q=\begin{pmatrix}\beta I & 0 & 0\\ \beta I & \beta I & 0\\ -I & -I & \frac1\beta I\end{pmatrix},\tag{F.8}$$

   我们用 （请注意下面的校正矩阵 $\mathcal M$ 与短篇 **F1** 中矩阵 $\mathcal M$ 的差别）

   $$\xi^{k+1}=\xi^k-\alpha_k\mathcal M(\xi^k-\tilde\xi^k)\quad 进行校正, 其中\quad \mathcal M=\begin{pmatrix}I & -I & 0\\ 0 & I & 0\\ -\beta I & -\beta I & I\end{pmatrix}.\tag{F.9}$$

   请给出使得 $\mathcal{HM}=\mathcal Q$ 的正定矩阵 $\mathcal H$, 并给出相应的计算步长的校正公式和关键的收敛性质.

练习题 3-4 讨论三个可分离块优化问题基于预测中 y, z 子问题平行处理的方法

短篇 **F2** 中处理三个可分离块问题采用 y, z 子问题平行处理的方法. 可以表示成先行预测

$$
\textbf{[预测]}\quad
\begin{cases}
\tilde{x}^k \in \arg\min\big\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k + Cz^k - b\|^2 \mid x \in \mathcal{X}\big\}, & \text{(F.10a)}\\
\tilde{y}^k \in \arg\min\big\{\theta_2(y) - y^T B^T \tilde{\lambda}^k + \frac{\mu}{2}\beta\|B(y - y^k)\|^2 | y \in \mathcal{Y}\big\}, & \text{(F.10b)}\\
\tilde{z}^k \in \arg\min\big\{\theta_3(z) - z^T C^T \tilde{\lambda}^k + \frac{\mu}{2}\beta\|C(z - z^k)\|^2 | z \in \mathcal{Z}\big\}, & \text{(F.10c)}\\
\tilde{\lambda}^k = \lambda^k - \beta\big(A\tilde{x}^k + By^k + Cz^k - b\big). \qquad (\text{要求 } \mu > 2) & \text{(F.10d)}
\end{cases}
$$

然后再进行如下校正:

$$
\textbf{[校正]}\qquad v^{k+1} = v^k - M(v^k - \tilde{v}^k), \quad \text{其中} \quad M = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\beta B & -\beta C & I \end{pmatrix}. \tag{F.11}
$$

3. 如果将预测 (F.10) 改成

$$
\begin{cases}
\tilde{x}^k \in \arg\min\big\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k + Cz^k - b\|^2 \mid x \in \mathcal{X}\big\}, & \text{(F.12a)}\\
\tilde{y}^k \in \arg\min\big\{\theta_2(y) - y^T B^T \tilde{\lambda}^k + \beta\|B(y - y^k)\|^2 | y \in \mathcal{Y}\big\}, & \text{(F.12b)}\\
\tilde{z}^k \in \arg\min\big\{\theta_3(z) - z^T C^T \tilde{\lambda}^k + \beta\|C(z - z^k)\|^2 | z \in \mathcal{Z}\big\}, & \text{(F.12c)}\\
\tilde{\lambda}^k = \lambda^k - \beta\big(A\tilde{x}^k + By^k + Cz^k - b\big). & \text{(F.12d)}
\end{cases}
$$

   请给出预测的变分不等式形式

$$
\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \tag{F.13}
$$

   中的预测矩阵 Q. 若用 (F.11) 中的矩阵 M,

$$
v^{k+1} = v^k - \alpha_k M(v^k - \tilde{v}^k) \tag{F.14}
$$

   实现校正, 请给出满足 $HM = Q$ 的正定矩阵 H, 校正中步长计算公式和收敛的关键不等式.

4. 在求解三个块可分离块凸优化问题时, 若要求预测变分不等式是

$$
\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \tag{F.15a}
$$

   其中

$$
Q = \begin{pmatrix} \frac{3}{2}\beta B^T B & 0 & 0 \\ 0 & \frac{3}{2}\beta C^T C & 0 \\ -B & -C & \frac{1}{\beta} I \end{pmatrix}. \tag{F.15b}
$$

   请问如何实现这个预测.

   (a) 若用 (F.11) 中的矩阵 M, 满足 $HM = Q$ 的正定矩阵 H 是什么?

   (b) 用 (F.14) 实现校正, 请给出步长计算公式和收敛的关键不等式.

5. 求解三个块可分离块凸优化问题时, 若要求预测变分不等式是

$$
\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \tag{F.16a}
$$

   其中

$$
Q = \begin{pmatrix} \beta B^T B & 0 & B^T \\ 0 & \beta C^T C & C^T \\ -B & -C & \frac{1}{\beta} I \end{pmatrix}. \tag{F.16b}
$$

   请问如何实现这个预测. 并说明实现这样的预测关于 x, y, z 的子问题都平行求解.

   (a) 对 (F.16) 中的 Q, 请给出变换 (F.8) 下的核矩阵 $\mathcal{Q}$. 若取 $\mathcal{H} = \begin{pmatrix} \beta I & 0 & 0 \\ 0 & \beta I & 0 \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix}$,

   满足 $\mathcal{H}\mathcal{M} = \mathcal{Q}$ 的矩阵 $\mathcal{M}$ 是什么?

   (b) 对这样确定的 $\mathcal{M}$, 请给出校正 $\xi^{k+1} = \xi^k - \alpha_k \mathcal{M}(\xi^k - \tilde{\xi}^k)$ 中步长计算法则和相应的收敛关键不等式.

6. 如果上题中的 B, C 都是单位矩阵, 可以取 $M = Q^{-T}$ 和 $H = QQ^T$, 再用 (F.14) 实现校正, 请给出 步长计算公式和收敛的关键不等式.

G 单元习题　根据收敛条件的等价性构造一簇方法

练习题 1 分析比较求解 $\min\{\theta(x)\,|\,Ax=b,\ x\in\mathcal{X}\}$ 的几类预测-校正方法

我们已经把线性约束的凸优化问题　$\min\{\theta(x)\,|\,Ax=b,\ x\in X\}$ 的鞍点转化成变分不等式

$$w^*\in\Omega,\quad \theta(x)-\theta(x^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega \tag{G.1a}$$

其中

$$w=\begin{pmatrix} x\\ \lambda\end{pmatrix},\quad F(w)=\begin{pmatrix} -A^T\lambda\\ Ax-b\end{pmatrix}\quad \text{and}\qquad \Omega=\mathcal{X}\times\Re^m. \tag{G.1b}$$

1. 我们继续讨论求解变分不等式 (G.1) 的预测-校正方法. 预测一般形式是

$$\tilde{w}^k\in\Omega,\ \ \theta(x)-\theta(\tilde{x}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(w-\tilde{w}^k)^TQ(w^k-\tilde{w}^k),\ \ \forall\, w\in\Omega \tag{G.2}$$

   (a) 若取 $Q=\begin{pmatrix} rI & A^T\\ 0 & H_0\end{pmatrix}$, 其中 $H_0=(\delta I+\frac{1}{r}AA^T)$, 请问预测如何实现?
   用校正 $w^{k+1}=w^k-M(w^k-\tilde{w}^k)$ 产生新的迭代点 w^{k+1},
   你选择什么样的 M, 能使得校正容易实现, 并给出相应的 H 和 G, 验证收敛性条件.

   (b) 若取 $Q=\begin{pmatrix} rI & 0\\ -A & H_0\end{pmatrix}$, 其中 $H_0=(\delta I+\frac{1}{r}AA^T)$, 请问预测如何实现?
   用校正 $w^{k+1}=w^k-M(w^k-\tilde{w}^k)$ 产生新的迭代点 w^{k+1},
   你选择什么样的 M, 能使得校正容易实现, 并给出相应的 H 和 G, 验证收敛性条件.

练习题 2-3 讨论三个可分离块优化问题变换下的收敛性等价条件

2. 处理多个可分离块凸优化问题的时候, 预测变分不等式

$$\tilde{w}^k\in\Omega,\quad \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(v-\tilde{v}^k)^TQ(v^k-\tilde{v}^k),\quad \forall w\in\Omega. \tag{G.3}$$

   以三个可分离块问题为例, 采用变换

$$P=\begin{pmatrix} B & 0 & 0\\ 0 & C & 0\\ 0 & 0 & I\end{pmatrix},\quad Q=P^T\mathcal{Q}P,\quad \xi=Pv=\begin{pmatrix} By\\ Cz\\ \lambda\end{pmatrix}, \tag{G.4}$$

   证明会得到

$$(\tilde{\xi}^k-\xi^*)^T\mathcal{Q}(\xi^k-\tilde{\xi}^k)\ge 0,\quad \forall\xi^*\in\Xi^*. \tag{G.5}$$

   所谓预测矩阵 Q^T+Q 本质上正定, 是指核矩阵 $\mathcal{Q}$ 满足 $\mathcal{Q}^T+\mathcal{Q}\succ 0$.
   我们在 $\mathcal{Q}^T+\mathcal{Q}\succ 0$ 的前提下给出收敛的充分条件. 首先证明:

   (a) 从 (G.5), 若有 $\mathcal{H}\succ 0$ 和 $\mathcal{H}\mathcal{M}=\mathcal{Q}$, 校正公式采用 $\xi^{k+1}=\xi^k-\mathcal{M}(\xi^k-\tilde{\xi}^k)$, 就可得到

$$(\xi^k-\xi^{k+1})^T\mathcal{H}(\tilde{\xi}^k-\xi^*)\ge 0,\quad \forall\,\xi^*\in\Xi^*. \tag{G.6}$$

   (b) 证明　（利用 $(a-b)^TH(c-d)=\frac{1}{2}(\|a-d\|_H^2-\|b-d\|_H^2)-\frac{1}{2}(\|c-a\|_H^2-\|c-b\|_H^2)$）

$$\|\xi^{k+1}-\xi^*\|_{\mathcal{H}}^2\le\|\xi^k-\xi^*\|_{\mathcal{H}}^2-\|\xi^k-\tilde{\xi}^k\|_{\mathcal{G}}^2,\quad \forall\xi^*\in\Xi^*, \tag{G.7}$$

   其中 $\mathcal{G}=\mathcal{Q}^T+\mathcal{Q}-\mathcal{M}^T\mathcal{H}\mathcal{M}$, $\|\xi^k-\tilde{\xi}^k\|_{\mathcal{G}}^2=(\xi^k-\tilde{\xi}^k)^T\mathcal{G}(\xi^k-\tilde{\xi}^k)$.

   (c) 在 $\mathcal{Q}^T+\mathcal{Q}\succ 0$, $\mathcal{H}\succ 0, \mathcal{H}\mathcal{M}=\mathcal{Q}$ 的前提下, 若有

$$\mathcal{G}=\mathcal{Q}^T+\mathcal{Q}-\mathcal{M}^T\mathcal{H}\mathcal{M}\succ 0, \tag{G.8}$$

   则由校正 $\xi^{k+1}=\xi^k-\mathcal{M}(\xi^k-\tilde{\xi}^k)$ 产生的序列是收缩的, 即 $\{\|\xi^k-\xi^*\|_{\mathcal{H}}^2\}$ 是严格单调下降的.

3. 证明在 $\mathcal{Q}^T+\mathcal{Q}$ 正定的前提下, 选取 $\mathcal{H}\succ 0$, 使得 $\mathcal{H}\mathcal{M}=\mathcal{Q}$ 和 $\mathcal{G}=\mathcal{Q}^T+\mathcal{Q}-\mathcal{M}^T\mathcal{H}\mathcal{M}\succ 0$ 等价于

   选取

$$\mathcal{D}\succ 0,\quad \mathcal{G}\succ 0,\quad 使得\quad \mathcal{D}+\mathcal{G}=\mathcal{Q}^T+\mathcal{Q}. \tag{G.9a}$$

   然后取

$$\mathcal{M}=\mathcal{Q}^{-T}\mathcal{D},\qquad \mathcal{H}=\mathcal{Q}\mathcal{D}^{-1}\mathcal{Q}^T\qquad 来完成. \tag{G.9b}$$

4. 对给定的 $\varepsilon>0$, 如果 $\tilde{w}$ 满足: $\tilde{w}\in\Omega$, $\theta(u)-\theta(\tilde{u})+(w-\tilde{w})^TF(w)\ge-\varepsilon$, $\forall w\in\{\Omega(\tilde{w})\,|\,\|w-\tilde{w}\|\le 1\}$. 我们称其为 ε 近似解. 试解释为什么这样定义 ε 近似解是合理的, 并且统一框架给出的所有分裂收缩算法都会提供一个 ε 近似解.

练习题 5 讨论三个可分离块优化问题在确定预测下借助核矩阵构造不同的校正方法

5. 在 **F** 单元的习题的 (F.10) 式中取 $\mu = 1.5$, 得到形如 (G.3) 的预测不等式, 其中

$$Q = \begin{pmatrix} \frac{3}{2}\beta B^T B & 0 & 0 \\ 0 & \frac{3}{2}\beta C^T C & 0 \\ -B & -C & \frac{1}{\beta} I \end{pmatrix}, \text{ 采用第 (2) 题中的变换, 得到 } \mathcal{Q} = \begin{pmatrix} \frac{3}{2}\beta I & 0 & 0 \\ 0 & \frac{3}{2}\beta I & 0 \\ -I & -I & \frac{1}{\beta} I \end{pmatrix}. \quad \text{(G.10)}$$

(a) 证明 $\mathcal{Q}^T + \mathcal{Q}$ 是正定的. 若记

$$\mathcal{D} = \begin{pmatrix} \beta I & 0 & 0 \\ 0 & \beta I & 0 \\ 0 & 0 & \frac{3}{4\beta} I \end{pmatrix}, \quad \text{证明条件 (G.9a) 满足.} \quad \text{(G.11)}$$

按 (G.9b) 给出矩阵 $\mathcal{M}$ 和校正公式 $\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k)$ 的具体形式.

(b) 如果将 (G.11) 中的矩阵 $\mathcal{D}$ 改成 $\mathcal{D} = \begin{pmatrix} 2\beta I & 0 & -I \\ 0 & 2\beta I & -I \\ -I & -I & \frac{5}{4\beta} I \end{pmatrix}$, 证明条件 (G.9a) 满足.

按 (G.9b) 给出矩阵 $\mathcal{M}$ 和校正公式 $\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k)$ 的具体形式.

练习题 6-7 从给定预测核矩阵出发讨论三个可分离块凸优化问题的预测校正方法

6. 对 **F** 单元讨论的三个可分离块的问题, 预测实现的变分不等式是 (G.3).

若给定预测矩阵的核矩阵

$$\mathcal{Q} = \begin{pmatrix} \beta I & 0 & -I \\ \beta I & \beta I & -I \\ -I & -I & \frac{5}{2\beta} I \end{pmatrix}, \quad \text{(G.12)}$$

(a) 证明 $\mathcal{Q}^T + \mathcal{Q}$ 正定, 给出相应的预测矩阵 Q 并说明预测该如何实现?

(b) 若取 $\mathcal{D} = \begin{pmatrix} \frac{1}{2}\beta I & 0 & 0 \\ 0 & \frac{1}{2}\beta I & 0 \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix}$, $\mathcal{G} = \begin{pmatrix} \frac{3}{2}\beta I & \beta I & -2I \\ \beta I & \frac{3}{2}\beta I & -2I \\ -2I & -2I & \frac{4}{\beta} I \end{pmatrix}$, 证明条件 (G.9a) 满足.

请根据 $\mathcal{M} = \mathcal{Q}^{-T}\mathcal{D}$ 给出校正公式 $\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k)$ 的具体形式.

(c) 将上一小题中的 $\mathcal{D}$ 和 $\mathcal{G}$ 互换, 再给出校正公式 $\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k)$ 的具体形式.

(d) 若取 $\mathcal{D} = \alpha(\mathcal{Q}^T + \mathcal{Q}), \quad \alpha \in (0,1)$, 请给出校正公式 $\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k)$ 的具体形式.

7. 同样求解三个可分离块的问题, 将 (G.12) 给出的预测矩阵的核矩阵改为

$$\mathcal{Q} = \begin{pmatrix} \beta I & 0 & I \\ \beta I & \beta I & I \\ I & I & \frac{5}{2\beta} I \end{pmatrix}. \quad \text{(G.13)}$$

(a) 证明 $\mathcal{Q}^T + \mathcal{Q}$ 正定, 给出相应的预测矩阵 Q 并说明预测该如何实现?

(b) 若取 $\mathcal{D} = \begin{pmatrix} \frac{1}{2}\beta I & 0 & 0 \\ 0 & \frac{1}{2}\beta I & 0 \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix}$, $\mathcal{G} = \begin{pmatrix} \frac{3}{2}\beta I & \beta I & 2I \\ \beta I & \frac{3}{2}\beta I & 2I \\ 2I & 2I & \frac{4}{\beta} I \end{pmatrix}$, 证明条件 (G.9a) 满足.

请根据 $\mathcal{M} = \mathcal{Q}^{-T}\mathcal{D}$ 给出校正公式 $\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k)$ 的具体形式.

(c) 将上一小题中的 $\mathcal{D}$ 和 $\mathcal{G}$ 互换, 再给出校正公式 $\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k)$ 的具体形式.

(d) 若取 $\mathcal{D} = \alpha(\mathcal{Q}^T + \mathcal{Q}), \quad \alpha \in (0,1)$, 请给出校正公式 $\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k)$ 的具体形式.

思考题 基于统一框架的算法产生的序列具有性质

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*.$$

和

$$\|v^k - v^{k+1}\|_H^2 \le \|v^{k-1} - v^k\|_H^2.$$

希望提高这类算法的效率, 还有哪些方向可以继续努力?

H 单元习题　统一框架下预测-校正的广义邻近点算法

练习题 1 讨论等式约束的多个可分离块凸优化问题基于逐个向前预测的广义 PPA 方法

线性等式约束的凸优化问题的 Lagrange 函数的鞍点等价表示的变分不等式

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega \tag{H.1}$$

第 k 步迭代的预测的变分不等式形式可以写成

$$\tilde{w}^k \in \Omega, \ \ \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \ \forall\, w \in \Omega \tag{H.2}$$

所谓预测-校正的广义 PPA 其校正由

$$v^{k+1} = v^k - \alpha M(v^k - \tilde{v}^k), \quad \alpha \in (0,2), \quad M = \tfrac{1}{2}Q^{-T}(Q^T + Q) \tag{H.3}$$

实现. 换句话说, 预测校正的广义 PPA 算法是由预测唯一确定的. 在实际计算中, 我们只要求预测矩阵的核矩阵正定.

1. 对 **H** 单元讨论的等式约束的多块问题, 假定从给定的 $(A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$ 开始预测. 通过

$$\left\{\begin{array}{l}
\tilde{x}_1^k \in \arg\min\big\{\theta_1(x_1) - x_1^T A_1^T \lambda^k + \frac{1}{2}\beta\|A_1x_1 + \sum_{j=2}^p A_jx_j^k - b\|^2 \mid x_1 \in \mathcal{X}_1\big\};\\
\tilde{x}_2^k \in \arg\min\big\{\theta_2(x_2) - x_2^T A_2^T \lambda^k + \frac{1}{2}\beta\|A_1\tilde{x}_1^k + A_2x_2 + \sum_{j=3}^p A_jx_j^k - b\|^2 \mid x_2 \in \mathcal{X}_2\big\};\\
\qquad\quad \vdots\\
\tilde{x}_i^k \in \arg\min\big\{\theta_i(x_i) - x_i^T A_i^T \lambda^k + \frac{1}{2}\beta\|\sum_{j=1}^{i-1} A_j\tilde{x}_j^k + A_ix_i + \sum_{j=i+1}^p A_jx_j^k - b\|^2 \mid x_i \in \mathcal{X}_i\big\};\\
\qquad\quad \vdots\\
\tilde{x}_p^k \in \arg\min\big\{\theta_p(x_p) - x_p^T A_p^T \lambda^k + \frac{1}{2}\beta\|\sum_{j=1}^{p-1} A_j\tilde{x}_j^k + A_px_p - b\|^2 \mid x_p \in \mathcal{X}_p\big\};\\
\tilde{\lambda}^k = \lambda^k - \beta\big(A_1\tilde{x}_1^k + \sum_{j=2}^p A_jx_j^k - b\big)
\end{array}\right. \tag{H.4}$$

求得 $\tilde{w}^k$. 请给出相应的形如 (H.2) 的预测变分不等式形式.

(a) 证明由 (H.4) 给出的预测可以写成: 从给定的 $(A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$ 开始

$$\left\{\begin{array}{l}
\tilde{x}_1^k \in \arg\min\big\{\theta_1(x_1) - x_1^T A_1^T \lambda^k + \frac{1}{2}\beta\|A_1x_1 + \sum_{j=2}^p A_jx_j^k - b\|^2 \mid x_1 \in \mathcal{X}_1\big\};\\
\tilde{\lambda}^k = \lambda^k - \beta\big(A_1\tilde{x}_1^k + \sum_{j=2}^p A_jx_j^k - b\big)\\
\tilde{x}_2^k \in \arg\min\big\{\theta_2(x_2) - x_2^T A_2^T \tilde{\lambda}^k + \frac{1}{2}\beta\|A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\big\};\\
\qquad\quad \vdots\\
\tilde{x}_i^k \in \arg\min\big\{\theta_i(x_i) - x_i^T A_i^T \tilde{\lambda}^k + \frac{1}{2}\beta\|\sum_{j=1}^{i-1} A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2 \mid x_i \in \mathcal{X}_i\big\};\\
\qquad\quad \vdots\\
\tilde{x}_p^k \in \arg\min\big\{\theta_p(x_p) - x_p^T A_p^T \tilde{\lambda}^k + \frac{1}{2}\beta\|\sum_{j=1}^{p-1} A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2 \mid x_p \in \mathcal{X}_p\big\}.
\end{array}\right. \tag{H.5}$$

并据此写出预测的紧凑的变分不等式表示.

(b) 若记 $\xi = (A_2x_2^k, \cdots, A_px_p^k, \lambda^k)^T$, 请给出相应的预测矩阵 Q 的核矩阵 $\mathcal{Q}$.

(c) 讨论分块矩阵

$$\mathcal{L} = \begin{pmatrix} I & 0 & \cdots & 0 \\ I & I & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ I & \cdots & I & I \end{pmatrix}, \quad \mathcal{I} = \begin{pmatrix} I & 0 & \cdots & 0 \\ 0 & I & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I \end{pmatrix} \quad \text{和} \quad \mathcal{E} = \begin{pmatrix} I \\ I \\ \vdots \\ I \end{pmatrix} \tag{H.6}$$

的一些性质. 计算

$$\mathcal{L}^{-1}, \qquad \mathcal{L}^{-T}\mathcal{E}, \qquad \mathcal{E}^T\mathcal{L}^{-T}\mathcal{E},$$

并用它们表示核矩阵 $\mathcal{Q}$, $\mathcal{Q}^{-T}$ 和 $\frac{1}{2}\mathcal{Q}^{-T}(\mathcal{Q}^T + \mathcal{Q})$.

(d) 给出相应变换意义下广义 PPA 的校正公式, 为下一次迭代提供 $(A_2x_2^{k+1}, \cdots, A_px_p^{k+1}, \lambda^{k+1})$.

练习题 2 在练习题 1 的基础上对预测子问题加了适当的正则项

2. 对 **H** 单元讨论的等式约束的多块问题, 假设从给定的 $(A_2x_2^k,\cdots,A_px_p^k,\lambda^k)$ 开始, 做与 (H.5) 略有不同的预测. 对 $i=2,\dots,p$, 在x_i-子问题的目标函数中加一项 $\delta\frac{\beta}{2}\|A_i(x_i-x_i^k)\|^2$.

   具体说来, 从给定的 $(A_2x_2^k,\cdots,A_px_p^k,\lambda^k)$ 出发, 通过

$$\left\{\begin{array}{l}
\tilde{x}_1^k \in \arg\min\bigl\{\theta_1(x_1)-x_1^TA_1^T\lambda^k+\frac{1}{2}\beta\|A_1x_1+\sum_{j=2}^pA_jx_j^k-b\|^2 \mid x_1\in\mathcal{X}_1\bigr\};\\
\tilde{\lambda}^k=\lambda^k-\beta\bigl(A_1\tilde{x}_1^k+\sum_{j=2}^pA_jx_j^k-b\bigr)\\
\tilde{x}_2^k \in \arg\min\left\{\begin{array}{l}\theta_2(x_2)-x_2^TA_2^T\tilde{\lambda}^k+\frac{1}{2}\beta\|A_2(x_2-x_2^k)\|^2\\ +\frac{1}{2}\delta\beta\|A_2(x_2-x_2^k)\|^2\end{array}\middle| x_2\in\mathcal{X}_2\right\};\\
\qquad\qquad\vdots\\
\tilde{x}_i^k \in \arg\min\left\{\begin{array}{l}\theta_i(x_i)-x_i^TA_i^T\tilde{\lambda}^k+\frac{1}{2}\beta\|\sum_{j=1}^{i-1}A_j(\tilde{x}_j^k-x_j^k)+A_i(x_i-x_i^k)\|^2\\ +\frac{1}{2}\delta\beta\|A_i(x_i-x_i^k)\|^2\end{array}\middle| x_i\in\mathcal{X}_i\right\};\\
\qquad\qquad\vdots\\
\tilde{x}_p^k \in \arg\min\left\{\begin{array}{l}\theta_p(x_p)-x_p^TA_p^T\tilde{\lambda}^k+\frac{1}{2}\beta\|\sum_{j=1}^{p-1}A_j(\tilde{x}_j^k-x_j^k)+A_p(x_p-x_p^k)\|^2\\ +\frac{1}{2}\delta\beta\|A_p(x_p-x_p^k)\|^2\end{array}\middle| x_p\in\mathcal{X}_p\right\}.
\end{array}\right. \tag{H.7}$$

   求得 $\tilde{w}^k$. 请给出相应的形如 (H.2) 的预测变分不等式形式.

   (a) 若记 $\xi^T=(A_2x_2^k,\cdots,A_px_p^k,\lambda^k)$, 请给出相应的预测矩阵 Q 的核矩阵 $\mathcal{Q}$.

   (b) 用 (H.6) 中分块矩阵 $\mathcal{L},\mathcal{I}$ 和 $\mathcal{E}$ 表示由预测 (H.7) 产生的核矩阵 $\mathcal{Q}$, 以及 $\mathcal{Q}^{-T}$ 和 $\frac{1}{2}\mathcal{Q}^{-T}(\mathcal{Q}^T+\mathcal{Q})$.

   (c) 给出相应变换意义下广义 PPA 的校正公式, 为下一次迭代提供 $(A_2x_2^{k+1},\cdots,A_px_p^{k+1},\lambda^{k+1})$.

   (d) 对由 (H.7) 生成的预测, 不用广义 PPA, 而取

$$\mathcal{D}=\begin{pmatrix}\beta\mathcal{I} & 0\\ 0 & \frac{1}{\beta}I\end{pmatrix},\quad 计算\quad \mathcal{G}=\mathcal{Q}^T+\mathcal{Q}-\mathcal{D}\quad 并证明\quad \mathcal{G}\succ 0.\quad 按照\ \mathcal{M}=\mathcal{Q}^{-T}\mathcal{D}$$

   给出校正矩阵 $\mathcal{M}$ 和为下一次迭代提供 $(A_2x_2^{k+1},\cdots,A_px_p^{k+1},\lambda^{k+1})$ 的具体校正公式.

练习题 3 从给定的预测核矩阵 $\mathcal{G}$ 出发构造等式约束的多个可分离块凸优化问题的求解方法

3. 利用 (H.6) 中的记号, 若设预测的核矩阵为

$$\mathcal{Q}=\begin{pmatrix}\beta\mathcal{L} & -\mathcal{E}\\ -\mathcal{E}^T & \frac{5}{2\beta}I\end{pmatrix},\quad 证明\ \mathcal{Q}^T+\mathcal{Q}\ 是正定的. \tag{H.8}$$

   (a) 要实现由 (H.8) 给出的核矩阵 $\mathcal{Q}$ 那样的预测, 具体实现的方式是什么？

   (b) 对核矩阵 $\mathcal{Q}$, 给出 $\mathcal{Q}^{-T}$.

$$\mathcal{Q}^T=\begin{pmatrix}\beta\mathcal{L}^T & -\mathcal{E}\\ -\mathcal{E}^T & \frac{5}{2\beta}I\end{pmatrix}+\begin{pmatrix}0 & -\mathcal{E}\\ 0 & 0\end{pmatrix}=\begin{pmatrix}\beta\mathcal{L}^T & 0\\ -\mathcal{E}^T & \frac{5}{2\beta}I\end{pmatrix}+\begin{pmatrix}\mathcal{E}\\ 0\end{pmatrix}\begin{pmatrix}0 & -I\end{pmatrix}$$

   请给出右端第一个矩阵的逆. 并根据广义“秩一校正公式”给出 $\mathcal{Q}^{-T}$.

   (c) 给出相应变换意义下广义 PPA 的校正公式, 为下一次迭代提供 $(A_2x_2^{k+1},\cdots,A_px_p^{k+1},\lambda^{k+1})$.

   (d) 如果取 $\mu\in(0,1)$,

$$\mathcal{D}=\begin{pmatrix}\nu\beta\mathcal{I} & 0\\ 0 & \frac{1}{\beta}I\end{pmatrix},\quad 写出\quad \mathcal{G}=\mathcal{Q}^T+\mathcal{Q}-\mathcal{D} \tag{H.9}$$

   的分块矩阵表达式并证明矩阵 $\mathcal{G}\succ 0$.

   (e) 给出变换意义下校正 $\xi^{k+1}=\xi^k-\mathcal{M}(\xi^k-\tilde{\xi}^k)$ 中的矩阵 $\mathcal{M}$.
   和为下一次迭代提供 $(A_2x_2^{k+1},\cdots,A_px_p^{k+1},\lambda^{k+1})$ 的具体表达式.