D-凸优化问题对应的变分不等式和因需定制的PPA算法

1 预备定理 (定理证明参见本人主页系列讲义第六讲)

定理 D. 设 $\mathcal{X} \subset \Re^n$ 是闭凸集, $\theta(x)$ 和 $f(x)$ 是凸函数, $f(x)$ 在包含 $\mathcal{X}$ 的某个开集上可微. 如果极小化问题 $\min\{\theta(x)+f(x)\,|\,x\in\mathcal{X}\}$ 的解集非空, 那么

$$x^* \in \arg\min\{\theta(x)+f(x)\,|\,x\in\mathcal{X}\}$$

当且仅当

$$x^*\in\mathcal{X},\ \ \theta(x)-\theta(x^*)+(x-x^*)^T\nabla f(x^*)\ge 0,\ \ \forall\, x\in\mathcal{X}.$$

引理 D. 设向量 $a,b\in\Re^n$, $H\in\Re^{n\times n}$ 是半正定矩阵. 如果 $b^\top H(a-b)\ge 0$, 则有

$$\|b\|_H^2 \le \|a\|_H^2 - \|a-b\|_H^2.$$

论断能从 $\|a\|_H^2 = \|b+(a-b)\|_H^2 \ = \ \|b\|_H^2 + 2b^\top H(a-b) + \|a-b\|_H^2$ 和 $b^\top H(a-b)\ge 0$ 直接推得.

2 凸优化和相对应的变分不等式

1). min−max 问题. 设 $\theta_1(x),\theta_2(y)$ 是凸函数, $\mathcal{X},\mathcal{Y}$ 是闭凸集, (x^*,y^*) 是 $\min-\max$ 问题(鞍点问题)

$$\min_x\max_y\{\Phi(x,y)=\theta_1(x)-y^TAx-\theta_2(y)\,|\,x\in\mathcal{X},y\in\mathcal{Y}\} \tag{2.1}$$

的解, 利用 $\Phi(x,y)$ 的表达式, 以上关系可以写成

$$\left\{\begin{array}{lll} x^*\in\mathcal{X}, & \theta_1(x)-\theta_1(x^*)+(x-x^*)^T(-A^Ty^*)\ge 0, & \forall\, x\in\mathcal{X},\\ y^*\in\mathcal{Y}, & \theta_2(y)-\theta_2(y^*)\ +\ (y-y^*)^T(\ Ax^*)\ \ \ge 0, & \forall\, y\in\mathcal{Y}.\end{array}\right.$$

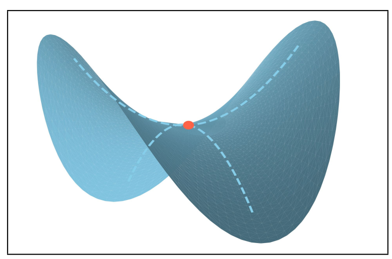
进一步, 可以写成单调变分不等式形式:

$$u^*\in\Omega,\quad \theta(u)-\theta(u^*)+(u-u^*)^TF(u^*)\ge 0,\quad \forall\, u\in\Omega, \tag{2.2}$$

其中 $u=\begin{pmatrix} x\\ y\end{pmatrix},\qquad \theta(u)=\theta_1(x)+\theta_2(y),\qquad F(u)=\begin{pmatrix} -A^Ty\\ Ax\end{pmatrix}$ and $\Omega=\mathcal{X}\times\mathcal{Y}$. (2.3)

注意到 (2.3) 中的 $F(u)$ 具有性质 $(u-\tilde u)^T(F(u)-F(\tilde u))\equiv 0$, 因此变分不等式 (2.2) 是单调的.

2). 线性约束的凸优化问题



考虑线性约束的凸优化问题

$$\min\{\theta(u)\mid \mathcal{A}u=b,\ u\in\mathcal{U}\}. \tag{2.4}$$

问题 (2.4) 的 Lagrange 函数是定义在 $\mathcal{U}\times\Re^m$ 上的

$$L(u,\lambda)=\theta(u)-\lambda^T(\mathcal{A}u-b). \tag{2.5}$$

如果一对 $(u^*,\lambda^*)\in\mathcal{U}\times\Re^m$ 满足

$$L(u,\lambda^*)\ge L(u^*,\lambda^*)\ge L(u^*,\lambda),\quad (u^*,\lambda^*)\in\mathcal{U}\times\Re^m,$$

则称作 Lagrange 函数 (2.5) 的鞍点. 上面的关系式可以写成

$$\left\{\begin{array}{ll} u^*\in\mathcal{U}, & L(u,\lambda^*)-L(u^*,\lambda^*)\ge 0,\ \ \forall\, u\in\mathcal{U}, \qquad (2.6a)\\ \lambda^*\in\Re^m, & L(u^*,\lambda^*)-L(u^*,\lambda)\ge 0,\ \ \forall\,\lambda\in\Re^m. \qquad (2.6b)\end{array}\right.$$

利用更紧致的格式, 鞍点可以表示成下述变分不等式

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega \tag{2.7}$$

的解, 其中 $w=\begin{pmatrix} u\\ \lambda\end{pmatrix},\qquad F(w)=\begin{pmatrix} -\mathcal{A}^T\lambda\\ \mathcal{A}u-b\end{pmatrix}$ and $\Omega=\mathcal{U}\times\Re^m.$ (2.8)

注意到这里 $F(w)$ 具有性质 $(w-\tilde w)^T(F(w)-F(\tilde w))\equiv 0$. 凸优化问题 (2.4) 转换成了变分不等式 (2.7).

问题 (2.1) 和 (2.4) 分别转换成了变分不等式 (2.2) 和 (2.7). 算法都聚焦于求解相应的变分不等式.

3 单调变分不等式的邻近点 (PPA) 算法

定义 (PPA 算法的 k次迭代) 设 $\boldsymbol{H}$ 为正定矩阵. 从给定的 w^k 开始, 求得 w^{k+1}, 使得

$$w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \geq (w - w^{k+1})^T H(w^k - w^{k+1}), \quad \forall\, w \in \Omega. \tag{3.9}$$

将(3.9) 中的 w 设为 w^*, 并利用$(w^{k+1} - w^*)^T F(w^{k+1}) = (w^{k+1} - w^*)^T F(w^*)$, 我们得到

$$(w^{k+1} - w^*)^T H(w^k - w^{k+1}) \geq \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^*) \geq 0. \tag{3.10}$$

在引理 D 中设 $a = (w^k - w^*)$ 和 $b = (w^{k+1} - w^*)$, 从不等式 (3.10) 得到

(PPA 收敛的收缩不等式) $$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2, \quad \forall w^* \in \Omega^*. \tag{3.11}$$

用 PPA 算法求解 VI, 建议采用进一步松弛的方法 $w^{k+1} := w^k - \alpha(w^k - w^{k+1})$, $\alpha = 1.5 \in (0, 2)$.

4 求解变分不等式的因需定制的 PPA (C-PPA) 和均困的 PPA (B-PPA)

1). 求解 VI (2.2) 的 C-PPA

从给定的 u^k 出发, 求得满足 (3.9) 的 u^{k+1}, 其中

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad w = u, \quad F(w) = \begin{pmatrix} -A^T y \\ Ax \end{pmatrix}, \quad H = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix}, \quad rs > \|A^T A\|. \tag{4.12}$$

对由 (4.12) 给出的 F 和 H, PPA 算法 (3.9) 的具体格式是 $\boxed{\text{下波纹线部分是 } F(w^{k+1})}$

$$\begin{cases} \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \{ \underline{-A^T y^{k+1}} + rI_n(x^{k+1} - x^k) + A^T(y^{k+1} - y^k) \} \geq 0, \ \forall x \in \mathcal{X}, \\ \theta_2(y) - \theta_2(y^{k+1}) \ + \ (y - y^{k+1})^T \{ \underline{Ax^{k+1}} + A(x^{k+1} - x^k) + sI_m(y^{k+1} - y^k) \} \geq 0, \ \forall y \in \mathcal{Y}. \end{cases}$$

根据定理 D, 完成上面的第 k-次 PPA 迭代, 可以通过求解下面的子问题完成:

$$\begin{cases} x^{k+1} = \text{argmin}\{ \theta_1(x) - x^T A^T y^k + \frac{r}{2}\|x - x^k\|^2 \,|\, x \in \mathcal{X} \}, & \text{(4.13a)} \\ y^{k+1} = \text{argmin}\{ \theta_2(y) + y^T A[2x^{k+1} - x^k] + \frac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y} \}. & \text{(4.13b)} \end{cases}$$

2). 求解 (2.7) 的 C-PPA 和 B-PPA

从给定的 w^k 出发, 求得满足 (3.9) 的 w^{k+1}, 其中

$$F(w) = \begin{pmatrix} -\mathcal{A}^T \lambda \\ \mathcal{A}u - b \end{pmatrix}, \ H = H_1 = \begin{pmatrix} \beta \mathcal{A}^T \mathcal{A} + \delta I_n & \mathcal{A}^T \\ \mathcal{A} & \frac{1}{\beta} I_m \end{pmatrix} \text{ 或 } H = H_2 = \begin{pmatrix} rI_n & \mathcal{A}^T \\ \mathcal{A} & \frac{1}{r}\mathcal{A}\mathcal{A}^T + \delta I_m \end{pmatrix}. \tag{4.14}$$

因需定制的 C-PPA. 利用 $H = H_1$ (见(4.14)) 的 PPA (3.9) 的具体格式是 $\boxed{\text{下波纹线部分是 } F(w^{k+1})}$

$$\begin{cases} \theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T \{ \underline{-\mathcal{A}^T \lambda^{k+1}} + \ (\beta \mathcal{A}^T \mathcal{A} + \delta I_n)(u^{k+1} - u^k) + \mathcal{A}^T(\lambda^{k+1} - \lambda^k) \} \geq 0, \\ \qquad \underline{(\mathcal{A}u^{k+1} - b)} \qquad + \mathcal{A}(u^{k+1} - u^k) \quad + \quad (1/\beta)(\lambda^{k+1} - \lambda^k) = 0. \end{cases}$$

根据定理 D, 完成上面的第 k-次 PPA 迭代, 可以通过求解下面的子问题完成:

$$\begin{cases} u^{k+1} = \text{argmin}\big\{ \theta(u) - u^T \mathcal{A}^T \lambda^k + \frac{1}{2}(u - u^k)^T (\beta \mathcal{A}^T \mathcal{A} + \delta I_n)(u - u^k) \,|\, u \in \mathcal{U} \big\}, & \text{(4.15a)} \\ \lambda^{k+1} = \lambda^k - \beta\big(\mathcal{A}[2u^{k+1} - u^k] - b \big). & \text{(4.15b)} \end{cases}$$

均(分)困(难)的 B-PPA. 利用 $H = H_2$ (见(4.14)) 的 PPA (3.9) 的具体格式是 $\boxed{\text{下波纹线部分是 } F(w^{k+1})}$

$$\begin{cases} \theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T \{ \underline{-\mathcal{A}^T \lambda^{k+1}} + \ rI_n(u^{k+1} - u^k) \quad + \quad \mathcal{A}^T \quad (\lambda^{k+1} - \lambda^k) \} \geq 0, \\ \qquad \underline{(\mathcal{A}u^{k+1} - b)} + \mathcal{A}(u^{k+1} - u^k) + (\frac{1}{r}\mathcal{A}\mathcal{A}^T + \delta I_m)(\lambda^{k+1} - \lambda^k) = 0. \end{cases}$$

根据定理 D, 完成上面的第 k-次 PPA 迭代, 可以通过求解下面的子问题完成:

$$\begin{cases} u^{k+1} = \text{argmin}\big\{ \theta(u) - u^T \mathcal{A}^T \lambda^k + \frac{r}{2}\|u - u^k\|^2 \,|\, u \in \mathcal{U} \big\}, \ \boxed{\text{原始变量子问题比 (4.15a)简单}} & \text{(4.16a)} \\ \lambda^{k+1} = \lambda^k - (\frac{1}{r}\mathcal{A}\mathcal{A}^T + \delta I_m)^{-1} \big(\mathcal{A}[2u^{k+1} - u^k] - b \big). \ \boxed{\text{求解中只需做一次 Cholesky 分解}} & \text{(4.16b)} \end{cases}$$

跟 (4.15) 相比, 算法 (4.16) 降低了 原始变量子问题的迭代难度, 而适当增加了对偶变量更新的工作量.