一些典型凸优化问题的分裂收缩算法

基于变分不等式和邻近点算法的统一框架

南师大数科院系列讲座 2022 年元月 4 – 9 日

何 炳 生

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

三. 从增广拉格朗日乘子法(ALM)到交替方向法(ADMM)

现在我们关心如何求解两个可分离块的凸优化问题

$$\min\{\theta_1(x)+\theta_2(y)\,|\,Ax+By=b,\ x\in\mathcal{X},\ y\in\mathcal{Y}\}. \tag{0.1}$$

1 从增广拉格朗日乘子法

我们先从等式约束凸优化问题

$$\min\{\theta(u)\mid \mathcal{A}u=b,\ u\in\mathcal{U}\} \tag{1.1}$$

的增广拉格朗日乘子法谈起.

问题 (1.1) 的 Lagrange 函数的鞍点等价于变分不等式

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega, \tag{1.2a}$$

的解点, 其中

$$w=\begin{pmatrix} u\\ \lambda\end{pmatrix},\quad F(w)=\begin{pmatrix} -\mathcal{A}^T\lambda\\ \mathcal{A}u-b\end{pmatrix},\quad \Omega=\mathcal{U}\times\Lambda. \tag{1.2b}$$

我们先从等式约束凸优化问题 (1.1) 的经典的增广拉格朗日乘子法 (ALM) [10, 11] 谈起.

问题 (1.1) 的增广拉格朗日函数是在 Lagrange 函数后面加一项 $\frac{\beta}{2}\|\mathcal{A}u-b\|^2$, 其中 $\beta>0$ 是给定的常数, 也就是说

$$\mathcal{L}_\beta(u,\lambda)=\theta(u)-\lambda^T(\mathcal{A}u-b)+\frac{\beta}{2}\|\mathcal{A}u-b\|^2. \tag{1.3}$$

求解等式约束凸优化问题(1.1) 的增广拉格朗日乘子法(ALM):
对给定的 $\beta > 0$, k 次迭代从给定的 λ^k 开始, 通过

$$\begin{cases} u^{k+1} \in \arg\min\{\mathcal{L}_\beta(u,\lambda^k) \mid u \in \mathcal{U}\}, & \text{(1.4a)} \\ \lambda^{k+1} = \lambda^k - \beta(\mathcal{A}u^{k+1} - b), & \text{(1.4b)} \end{cases}$$

得到新的 λ^{k+1}.

在 (1.4) 中, u^{k+1} 是根据 λ^k 计算得来的结果. 因此, 我们称

u 为算法的中间变量 (intermidiate variable), λ 为核心变量 (essential variable).

根据优化问题的最优性条件的基本引理, 方法 (1.4) 提供的 $w^{k+1} = (u^{k+1}, \lambda^{k+1}) \in \Omega$ 满足

$$\begin{cases} \theta(u) - \theta(u^{k+1}) \;+\; (u - u^{k+1})^T\{-\mathcal{A}^T\lambda^k + \beta\mathcal{A}^T(\mathcal{A}u^{k+1} - b)\} \;\; \geq \;\; 0, \;\; \forall u \in \mathcal{U}, \\ \qquad\qquad\qquad (\lambda - \lambda^{k+1})^T\{(\mathcal{A}u^{k+1} - b) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)\} \;\; \geq \;\; 0, \;\; \forall \lambda \in \Re^m. \end{cases}$$

进一步利用关系式 $\lambda^{k+1} = \lambda^k - \beta(\mathcal{A}u^{k+1} - b)$, 上式可以改写成: $w^{k+1} \in \Omega$,

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} u - u^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \begin{pmatrix} -\mathcal{A}^T\lambda^{k+1} \\ (\mathcal{A}u^{k+1} - b) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k) \end{pmatrix} \geq 0, \quad \forall\, w \in \Omega.$$

接着利用 (1.2) 的记号, 可以把上式写成更紧凑的式子:

$$\begin{aligned} w^{k+1} \in \Omega, \;\; \theta(u) - \theta(u^{k+1}) &+ (w - w^{k+1})^T F(w^{k+1}) \\ &\geq (\lambda - \lambda^{k+1})^T \frac{1}{\beta}(\lambda^k - \lambda^{k+1}), \quad \forall w \in \Omega. \end{aligned} \tag{1.5}$$

变分不等式 (1.5) 中, 如果 $\lambda^k = \lambda^{k+1}$, 那么 w^{k+1} 就是变分不等式 (1.2) 的解. 将 (1.5) 中的任意 $w \in \Omega$ 设成任意一个固定的解点 $w^* \in \Omega^*$, 我们得到

$$(\lambda^{k+1} - \lambda^*)^T \frac{1}{\beta}(\lambda^k - \lambda^{k+1}) \geq \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1}).$$

利用前面强调过的 $(w^{k+1} - w^*)^T F(w^{k+1}) = (w^{k+1} - w^*)^T F(w^*)$, 就有

$$(\lambda^k - \lambda^*)^T \frac{1}{\beta}(\lambda^k - \lambda^{k+1}) \geq \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^*).$$

由 $w^{k+1} \in \Omega$ 和 w^* 的最优性, 根据(1.2), 有 $\theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^*) \geq 0$, 上式右端非负, 所以有

$$(\lambda^{k+1} - \lambda^*)^T(\lambda^k - \lambda^{k+1}) \geq 0. \tag{1.6}$$

令 $a = \lambda^k - \lambda^*$, $b = \lambda^{k+1} - \lambda^*$, 上面的不等式就相当于 $b^T(a - b) \geq 0$,

$$\|\lambda^{k+1} - \lambda^*\|^2 \leq \|\lambda^k - x^*\|^2 - \|\lambda^k - \lambda^{k+1}\|^2.$$

求解问题 (1.1) 的另一个方法是二次罚函数法, 简称罚函数法. 设 $\{r_k\}$ 是一个严格单调递增的正数序列. 罚函数法的第 k 次迭代通过

$$u^{k+1} \in \operatorname{argmin}\{\theta(u) + \frac{r_k}{2}\|\mathcal{A}u - b\|^2 \mid u \in \mathcal{U}\} \tag{1.7}$$

得到 u^{k+1}. 在求解子问题 (1.7) 时往往以上一步迭代的解 u^k 当做初始点. 求解问题 (1.1), 增广 Lagrange 乘子法和罚函数法每步迭代要求解的子问题难度相当, 而后者数值表现远没有前者好, 对此 Nocedal 和 Wright 在他们的专著 Numerical Optimization [9] 的第 17 章中说的很清楚.

2 交替方向法

我们要求解的问题是 (0.1). 虽然用罚函数法 (1.7) 求解 (0.1) 也可以松弛成交替极小化方法–Alternating Minimization Method (AMA):

$$\begin{cases} x^{k+1} \in \operatorname{argmin}\big\{\theta_1(x) + \frac{\beta}{2}\|Ax + By^k - b\|^2 | x \in \mathcal{X}\big\}, \\ y^{k+1} \in \operatorname{argmin}\big\{\theta_2(y) + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 | y \in \mathcal{Y}\big\}. \end{cases}$$

由于增广拉格朗日乘子法优于罚函数方法 [9]. 而乘子交替方向法(ADMM)和交替极小化方法(AMA)分别由增广拉格朗日乘子法和罚函数方法松弛而来的, 因此 ADMM 优于 AMA 已成共识.

假如用经典的增广 Lagrange 乘子法 (1.4) 求解, 方法可以描述为

求解等式约束凸优化问题 (0.1) 的增广拉格朗日乘子法(ALM):
对给定的 $\beta > 0$, k 次迭代从给定的 λ^k 开始, 通过

$$\begin{cases} (x^{k+1}, y^{k+1}) \in \operatorname{argmin}\left\{ \begin{array}{c} \theta_1(x) + \theta_2(y) - (\lambda^k)^T(Ax + By - b) \\ +\frac{\beta}{2}\|Ax + By - b\|^2 \end{array} \middle| \begin{array}{c} x \in \mathcal{X} \\ y \in \mathcal{Y} \end{array} \right\}, & \text{(2.1a)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & \text{(2.1b)} \end{cases}$$

得到新的 λ^{k+1}.

上面直接用增广 Lagrange 乘子法求解, 缺点是没有利用问题问题 (0.1) 的可分离结构. 为了克服 ALM 在求解问题 (0.1) 中的缺点, 人们考虑采用乘子交替方向法 (ADMM). 主要思想是将子问题分裂成两部分. 中间变量只是 $u = (x, y)$ 中的一部分 x. 核心变量为 $v = (y, \lambda)$. 我们设迭代初始值为 $v^0 = (y^0, \lambda^0)$.

求解问题 (0.1) **的松弛的增广拉格朗日乘子法– ADMM** 选定 $\beta > 0$.
第 k 步迭代从给定的 (y^k, λ^k) 开始

1. 首先, 对于给定的(y^k, λ^k), 求 x^{k+1} 使得

$$x^{k+1} \in \arg\min\left\{ \begin{array}{c} \theta_1(x) - (\lambda^k)^T(Ax + By^k - b) \\ +\frac{\beta}{2}\|Ax + By^k - b\|^2 \end{array} \middle| x \in \mathcal{X} \right\}. \tag{2.2a}$$

2. 利用 λ^k 和已经求得的x^{k+1}, 求 y^{k+1} 使得

$$y^{k+1} \in \operatorname{argmin}\left\{ \begin{array}{c} \theta_2(y) - (\lambda^k)^T(Ax^{k+1} + By - b) \\ +\frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \end{array} \middle| y \in \mathcal{Y} \right\}. \tag{2.2b}$$

3. 最后, 更新Lagrange乘子

$$\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \tag{2.2c}$$

使用凸优化问题 (0.1) 的增广 Lagrange 函数

$$\mathcal{L}_\beta^{[2]}(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax + By - b) + \frac{\beta}{2}\|Ax + By - b\|^2, \tag{2.3}$$

乘子交替方向法可以简单地表示成:

求解问题 (0.1) **的 ADMM 方法** 选定 $\beta > 0$.
第 k 次迭代从给定的 (y^k, λ^k) 开始

$$(\text{ADMM}) \quad \left\{ \begin{array}{ll} x^{k+1} = \arg\min\big\{\mathcal{L}_\beta^{[2]}(x, y^k, \lambda^k) \;\big|\; x \in \mathcal{X}\big\}, & (2.4a) \\ y^{k+1} = \arg\min\big\{\mathcal{L}_\beta^{[2]}(x^{k+1}, y, \lambda^k) \;\big|\; y \in \mathcal{Y}, & (2.4b) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b), & (2.4c) \end{array} \right.$$

求得 $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$.

由于变动目标函数中的常数项对问题的解没有影响, 乘子交替方向法 (2.4) 的第 k 次迭代可以表述成从 (y^k, λ^k) 开始:

$$(\text{ADMM}) \quad \left\{ \begin{array}{ll} x^{k+1} \in \operatorname{argmin}\big\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k - b\|^2 \big| x \in \mathcal{X}\big\} & (2.5a) \\ y^{k+1} \in \operatorname{argmin}\big\{\theta_2(y) - y^T B^T \lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \big| y \in \mathcal{Y}\big\}, & (2.5b) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & (2.5c) \end{array} \right.$$

采用 ADMM 求解 (0.1) 的好处是子问题 (2.2a) 和 (2.2b) 分别只含有自变量 x 和 y, 使子问题得到简化. 譬如说下面的问题:

$$\min\{\frac{1}{2}\|X - C\|_F^2 \mid X \in S_+^n \cap S_B\}, \tag{2.6}$$

其中集合 S_+^n 和 S_B 分别为

$$S_+^n = \{H \in R^{n\times n} \mid H^T = H,\ H \succeq 0\}$$

和

$$S_B = \{H \in R^{n\times n} \mid H^T = H,\ H_L \le H \le H_U\}.$$

H_L 和 H_U 都是给定的对称矩阵. 问题可以转换成等价的

$$\begin{array}{rl} \min & \frac{1}{2}\|X - C\|^2 + \frac{1}{2}\|Y - C\|^2 \\ \text{s.t} & X - Y = 0, \\ & X \in S_+^n,\ Y \in S_B. \end{array} \tag{2.7}$$

直接用 ALM 求解 (2.6) 和 (2.7) 都无从下手. 问题 (2.7) 的等式约束的 Lagrange 乘子也是一个矩阵, 我们记它为 Z. 用 ADMM (2.2) 求解 (2.7), 只要按下面的步骤来实施:

- 对给定的 Y^k 和 Z^k,

  $$X^{k+1} = \arg\min\{\frac{1}{2}\|X - C\|_F^2 - \mathrm{Tr}(Z^k X) + \frac{\beta}{2}\|X - Y^k\|_F^2 \mid X \in S_+^n\}.$$

  这样的 X^{k+1} 可以由

  $$X^{k+1} = P_{S_+^n}\Big\{\frac{1}{1+\beta}(\beta Y^k + Z^k + C)\Big\}$$

  直接得到. 如何在半正定矩阵的集合上投影, 在预备知识一章已经做了介绍.

- 有了 X^{k+1}, 用 ADMM 求 Y^{k+1} 就是

  $$Y^{k+1} = \arg\min\{\frac{1}{2}\|Y - C\|_F^2 + \mathrm{Tr}(Z^k Y) + \frac{\beta}{2}\|\tilde{X}^k - Y\|_F^2 \mid Y \in S_B\}$$

  这个 Y^{k+1} 通过

  $$Y^{k+1} = P_{S_B}\{\frac{1}{1+\beta}(\beta X^{k+1} - Z^k + C)\}$$

  得到, 矩阵在 A 在 $S_B = \{H \mid H_L \le H \le H_U\}$ 上投影,

  $$P_{S_B}(A) = \min(\max(H_L, A), H_U).$$

- 最后,

  $$z^{k+1} = z^k - \beta(X^{k+1} - Y^{k+1})$$

  就完成了一次迭代.

参考文献

[1] X.J. Cai, G.Y. Gu, B.S. He and X.M. Yuan, A proximal point algorithms revisit on the alternating direction method of multipliers, Science China Mathematics, 56 (2013), 2179-2186.

[2] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1984.

[3] Gu G Y, He B S and Yuan X M. Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: a unified approach. Comput. Optim. Appl., 2014, 59: 135-161.

[4] B. S. He, PPA-like contraction methods for convex optimization: a framework using variational inequality approach. J. Oper. Res. Soc. China **3** (2015) 391 – 420.

[5] B. S. He, M. Tao and X.M. Yuan, Alternating direction method with Gaussian back substitution for separable convex programming, *SIAM Journal on Optimization* **22**(2012), 313-340.

[6] B.S. He, M. Tao and X.M. Yuan, A splitting method for separable convex programming, *IMA Journal of Numerical Analysis*, **31**, 394-426, 2015.

[7] B.S. He and X. M. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective, SIAM Journal on Imaging Science, **5**, 119-149, 2012.

[8] B. Martinet, Regularisation, d'inéquations variationelles par approximations succesives, *Rev. Francaise d'Inform. Recherche Oper.*, **4**, 154-159, 1970.

[9] J. Nocedal and S. J. Wright, Numerical Optimization, Springer Verlag, New York, 1999.

[10] M. R. Hestenes, Multiplier and gradient methods, *JOTA* **4**, 303-320, 1969.

[11] M. J. D. Powell, A method for nonlinear constraints in minimization problems, in Optimization, R. Fletcher, ed., Academic Press, New York, NY, pp. 283-298, 1969.

[12] R.T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Cont. Optim.*, **14**, 877-898, 1976.

[13] 何炳生. 从变分不等式的统一收缩算法到凸优化的分裂收缩算法, 高等学校计算数学学报, 2016, 38: 74–96.

[14] 何炳生. 凸优化的一阶分裂算法–变分不等式为工具的统一框架, 见 http://maths.nju.edu. cn/~hebma 中的《My Talk》.

[15] 何炳生. 凸优化和单调变分不等式的收缩算法, 见 http://maths.nju.edu.cn/~hebma 中的系列讲义.