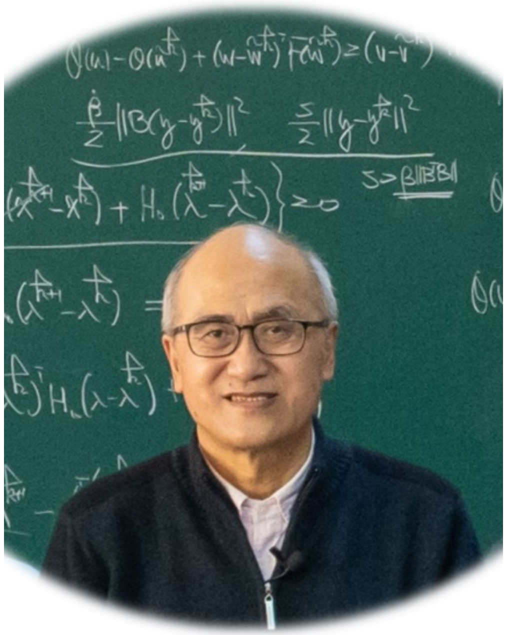
凸优化问题的一些典型问题及其求解方法

9. 多个可分离块凸优化问题的 ADMM 类算法



南京大学数学系

何炳生

Bingsheng He

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

南京师范大学数学科学学院 2022 年元月 4日 - 9日

1 ADMM with wider applications

Let us consider the general two-block separable convex optimization model

$$\min\left\{\theta_1(x)+\theta_2(y) \mid Ax+By=b \text{ (or } \geq b), x\in\mathcal{X}, y\in\mathcal{Y}\right\}. \tag{1.1}$$

The linear constraints can be a system of linear equations or linear inequalities.

We define

$$\Lambda = \begin{cases} \Re^m, & \text{if } Ax+By=b, \\ \Re^m_+, & \text{if } Ax+By\geq b. \end{cases}$$

The projection on Λ is denoted by $P_\Lambda[\cdot]$.

For such special Λ, the projection on Λ is clear !

The only difference: $P_{\Re^m}(\lambda)=\lambda, \quad P_{\Re^m_+}(\lambda)=\max\{\lambda,0\}.$

1.1 Primal-dual extension of ADMM with wider application

A Primal-Dual Extension of the ADMM for (1.1).

From (Ax^k, By^k, λ^k) to $(Ax^{k+1}, By^{k+1}, \lambda^{k+1})$:

1. (Prediction Step) With given (Ax^k, By^k, λ^k), find $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ via

$$\left\{\begin{array}{l} \tilde{x}^k \in \text{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|A(x - x^k)\|^2 \mid x \in \mathcal{X}\}, \\ \tilde{y}^k \in \text{argmin}\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}\beta\|A(\tilde{x}^k - x^k) + B(y - y^k)\|^2 \mid y \in \mathcal{Y}\}, \\ \tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(A\tilde{x}^k + B\tilde{y}^k - b\big)\big]. \end{array}\right. \tag{1.2a}$$

2. (Correction Step) Generate the new iterate $(Ax^{k+1}, By^{k+1}, \lambda^{k+1})$ with $\nu \in (0,1)$ by

$$\begin{pmatrix} Ax^{k+1} \\ By^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} Ax^k \\ By^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 \\ 0 & \nu I_m & 0 \\ -\nu\beta I_m & 0 & I_m \end{pmatrix} \begin{pmatrix} Ax^k - A\tilde{x}^k \\ By^k - B\tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \tag{1.2b}$$

这是一类预测–校正方法. 需要额外的校正, 但校正花费很小！

1.2 Dual-Primal extension of ADMM with wider application

A Dual-Primal Extension of the ADMM for (1.1).

From (Ax^k, By^k, λ^k) to $(Ax^{k+1}, By^{k+1}, \lambda^{k+1})$:

1. (Prediction Step) With given (Ax^k, By^k, λ^k), find $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ via

$$\left\{\begin{array}{l} \tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(Ax^k + By^k - b\big)\big], \\[2ex] \tilde{x}^k \in \text{argmin}\big\{\theta_1(x) - x^TA^T\tilde{\lambda}^k + \frac{1}{2}\beta\|A(x-x^k)\|^2 \mid x \in \mathcal{X}\big\}, \\[2ex] \tilde{y}^k \in \text{argmin}\big\{\theta_2(y) - y^TB^T\tilde{\lambda}^k + \frac{1}{2}\beta\|A(\tilde{x}^k-x^k) + B(y-y^k)\|^2 \mid y \in \mathcal{Y}\big\}. \end{array}\right. \tag{1.3a}$$

2. (Correction Step) Generate the new iterate $(Ax^{k+1}, By^{k+1}, \lambda^{k+1})$ with $\nu \in (0,1)$ by

$$\begin{pmatrix} Ax^{k+1} \\ By^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} Ax^k \\ By^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 \\ 0 & \nu I_m & 0 \\ -\beta I_m & -\beta I_m & I_m \end{pmatrix} \begin{pmatrix} Ax^k - A\tilde{x}^k \\ By^k - B\tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \tag{1.3b}$$

预测采用不同顺序, 校正公式也略有不同. 校正同样是花费很小的.

2 p-block separable convex optimization problems

In the following we consider the multiple-block convex optimization:

$$\min\Big\{\sum_{i=1}^{p}\theta_i(x_i)\;\Big|\;\sum_{i=1}^{p}A_ix_i = b \;(\text{or} \geq b), \quad x_i\in\mathcal{X}_i\Big\}. \tag{2.1}$$

The Lagrangian function is

$$L(x_1,\ldots,x_p,\lambda) = \sum_{i=1}^{p}\theta_i(x_i) - \lambda^T\Big(\sum_{i=1}^{p}A_ix_i - b\Big),$$

which is defined on $\Omega = \prod_{i=1}^{p}\mathcal{X}_i\times\Lambda,$ where

$$\Lambda = \begin{cases} \Re^m, & \text{if } \sum_{i=1}^{p}A_ix_i = b, \\ \Re^m_+, & \text{if } \sum_{i=1}^{p}A_ix_i \geq b. \end{cases}$$

Let $(x_1^*,\ldots,x_p^*,\lambda^*)\in\Omega$ be a saddle point of the Lagrangian function, then

$$L_{\lambda\in\Lambda}(x_1^*,\ldots,x_p^*,\lambda)\le L(x_1^*,\ldots,x_p^*,\lambda^*)\le L_{x_i\in\mathcal{X}_i}(x_1,\ldots,x_p,\lambda^*).$$

The optimality condition of (2.1) can be written as the following VI:

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

where

$$w=\begin{pmatrix} x_1\\ \vdots\\ x_p\\ \lambda\end{pmatrix},\quad x=\begin{pmatrix} x_1\\ \vdots\\ x_p\end{pmatrix},\quad F(w)=\begin{pmatrix} -A_1^T\lambda\\ \vdots\\ -A_p^T\lambda\\ \sum_{i=1}^p A_ix_i-b\end{pmatrix}, \tag{2.2b}$$

and

$$\theta(x)=\sum_{i=1}^p\theta_i(x_i),\qquad \Omega=\prod_{i=1}^p\mathcal{X}_i\times\Lambda.$$

Again, we denote by Ω^* the solution set of the VI (2.2).

2.1 Primal-dual extension of the ADMM for p-block Problems

A Primal-Dual Extension of the ADMM for (2.1) **Prediction Step** .

From $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$ to $(A_1x_1^{k+1}, A_2x_2^{k+1}, \cdots, A_px_p^{k+1}, \lambda^{k+1})$:

With given $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$, find $\tilde{w}^k \in \Omega$ via

$$\left\{\begin{array}{l}
\tilde{x}_1^k \in \arg\min\big\{\theta_1(x_1) - x_1^T A_1^T \lambda^k + \frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\big\}; \\
\tilde{x}_2^k \in \arg\min\big\{\theta_2(x_2) - x_2^T A_2^T \lambda^k + \frac{\beta}{2}\|A_1(\tilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\big\}; \\
\qquad\vdots \\
\tilde{x}_i^k \in \arg\min_{x_i \in \mathcal{X}_i}\big\{\theta_i(x_i) - x_i^T A_i^T \lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1} A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2\big\}; \\
\qquad\vdots \\
\tilde{x}_p^k \in \arg\min_{x_p \in \mathcal{X}_p}\big\{\theta_p(x_p) - x_p^T A_p^T \lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{p-1} A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2\big\}; \\
\tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(\sum_{j=1}^{p} A_j\tilde{x}_j^k - b\big)\big].
\end{array}\right. \tag{2.3}$$

预测先原始再对偶. 对可分离的原始变量子问题逐一按序求解.

A Primal-Dual Extension of the ADMM for (2.1) **Correction Step** .

From $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$ to $(A_1x_1^{k+1}, A_2x_2^{k+1}, \cdots, A_px_p^{k+1}, \lambda^{k+1})$:

Generate the new iterate $(A_1x_1^{k+1}, A_2x_2^{k+1}, \cdots, A_px_p^{k+1}, \lambda^{k+1})$ with $\nu \in (0,1)$ by

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \\ \lambda^{k} \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 & \cdots & 0 \\ 0 & \nu I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\nu I_m & 0 \\ 0 & \cdots & 0 & \nu I_m & 0 \\ -\nu\beta I_m & 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \tag{2.4}$$

对照一下就可以发现, §2.1 中的方法, 就是 §1.1方法的直接推广.

校正非常简单, 工作量也很小. 把校正公式分开来写就是:

$\boldsymbol{Ax_i^{k+1}, i = 1, \dots, p}$

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 & 0 \\ 0 & \nu I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -\nu I_m \\ 0 & \cdots & 0 & \nu I_m \end{pmatrix} \begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \end{pmatrix}, \tag{2.5}$$

$\boldsymbol{\lambda^{k+1}}$

$$\lambda^{k+1} = \tilde{\lambda}^k + \nu\beta(A_1x_1^k - A_1\tilde{x}_1^k). \tag{2.6}$$

还能说校正不简单! ?

2.2 Dual-primal extension of the ADMM for (2.1)

A Dual-Primal Extension of the ADMM for (2.1) **Prediction Step .**

From $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$ to $(A_1x_1^{k+1}, A_2x_2^{k+1}, \cdots, A_px_p^{k+1}, \lambda^{k+1})$:

With given $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$, find $\tilde{w}^k \in \Omega$ via

$$\left\{\begin{array}{l}
\tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(\sum_{j=1}^p A_jx_j^k - b\big)\big] \\[2mm]
\tilde{x}_1^k \in \arg\min\{\theta_1(x_1) - x_1^TA_1^T\tilde{\lambda}^k + \frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\}; \\[2mm]
\tilde{x}_2^k \in \arg\min\{\theta_2(x_2) - x_2^TA_2^T\tilde{\lambda}^k + \frac{\beta}{2}\|A_1(\tilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\}; \\[2mm]
\qquad\qquad \vdots \\[2mm]
\tilde{x}_i^k \in \arg\min_{x_i\in\mathcal{X}_i}\{\theta_i(x_i) - x_i^TA_i^T\tilde{\lambda}^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1}A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2\}; \\[2mm]
\qquad\qquad \vdots \\[2mm]
\tilde{x}_p^k \in \mathrm{argmin}_{x_p\in\mathcal{X}_p}\{\theta_p(x_p) - x_p^TA_p^T\tilde{\lambda}^k + \frac{\beta}{2}\|\sum_{j=1}^{p-1}A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2\}.
\end{array}\right. \tag{2.7}$$

预测先对偶再原始. 对可分离的原始变量子问题逐一按序求解.

A Dual-Primal Extension of the ADMM for (2.1) **Correction Step** .

From $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$ to $(A_1x_1^{k+1}, A_2x_2^{k+1}, \cdots, A_px_p^{k+1}, \lambda^{k+1})$:

Generate the new iterate $(A_1x_1^{k+1}, A_2x_2^{k+1}, \cdots, A_px_p^{k+1}, \lambda^{k+1})$ with $\nu \in (0,1)$ by

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \\ \lambda^{k} \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 & \cdots & 0 \\ 0 & \nu I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\nu I_m & 0 \\ 0 & \cdots & 0 & \nu I_m & 0 \\ -\beta I_m & -\beta I_m & \cdots & -\beta I_m & I_m \end{pmatrix} \begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \tag{2.8}$$

对照一下就可以发现, §2.2 中的方法, 就是 §1.2 方法的直接推广.

校正工作量很小. 把校正公式分开来写就是:

Ax_i^{k+1} $(i = 1, \ldots, p)$ **The correction form of the primal parts are equal.**

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 & 0 \\ 0 & \nu I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -\nu I_m \\ 0 & \cdots & 0 & \nu I_m \end{pmatrix} \begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \end{pmatrix}, \tag{2.9}$$

λ^{k+1} **The correction form of the dual parts are slightly different.**

$$\lambda^{k+1} = \tilde{\lambda}^k + \beta \sum_{i=1}^{p} (A_ix_i^k - A_i\tilde{x}_i^k). \tag{2.10}$$

两种不同方法的

$\lambda^{k+1} = \tilde{\lambda}^k + \nu\beta(A_1x_1^k - A_1\tilde{x}_1^k) \;\Rightarrow\; \lambda^{k+1} = \tilde{\lambda}^k + \beta\sum_{i=1}^{p}(A_ix_i^k - A_i\tilde{x}_i^k).$

References

[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] D. Gabay, Applications of the method of multipliers to variational inequalities, *Augmented Lagrange Methods: Applications to the Solution of Boundary-valued Problems*, edited by M. Fortin and R. Glowinski, North Holland, Amsterdam, The Netherlands, 1983, pp. 299–331.

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

[4] G.Y. Gu, B.S. He and X.M. Yuan, Customized Proximal point algorithms for linearly constrained convex minimization and saddle-point problem: a unified Approach, Comput. Optim. Appl., 59(2014), 135-161.

[5] B.S. He, H. Yang, and S.L. Wang, Alternating directions method with self-adaptive penalty parameters for monotone variational inequalities, *JOTA* **23**(2000), 349–368.

[6] B. S. He and X. M. Yuan, On the $O(1/t)$ convergence rate of the alternating direction method, *SIAM J. Numerical Analysis* **50**(2012), 700-709.

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

[8] B.S. He and X.M. Yuan, Balanced Augmented Lagrangian Method for Convex Optimization. manusript, 2021. arXiv:2108.08554