乘子交替方向法的一些收敛性质[1]

何 炳 生

(南京大学数学系 南京, 210093)

On the Convergence Properties of Alternating Direction Method of Multipliers

(Department of Mathematics, Nanjing University, Nanjing, 210093)

Abstract. Alternating direction method of multipliers (ADMM) is effectively applied to solving the optimization problems arising from image processing, machine learning and other applied computation areas. In this paper, by using the approach of the variational inequality and some simple identities, we give a short and brief proof of the convergence properties of ADMM, which includes the strong contractility, $O(1/t)$ ergotic, and strong pointwise iteration-complexity. The only mathematical background required of this paper is the basic concepts of optimality of the convex optimization.

Key words. Alternating direction method of multipliers, contraction, ergodic iteration-complexity, strong pointwise iteration-complexity

AMS(2000) 90C25, 90C30, 90C33

中图法分类号 O221.2, O224

1 引 言

本文讨论的两个可分离算子的线性约束凸优化问题是

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

其中 $A\in\Re^{m\times n_1}$, $B\in\Re^{m\times n_2}$, $b\in\Re^m$; $\mathcal{X}\subset\Re^{n_1}, \mathcal{Y}\subset\Re^{n_2}$ 是闭凸集; $\theta_1(x):\Re^{n_1}\to\Re$ 和 $\theta_2(y):\Re^{n_2}\to\Re$ 是 (不一定光滑的) 凸函数. 这类问题大量出现在图像处理, 机器学习等稀疏优化领域 [2]. 乘子交替方向法(Alternating Directions Method of Multipliers), 简称 ADMM, 通常称之为交替方向法, 最初由 Glowinski 等为偏微分方程数值求解在 [6, 7] 中提出, 近年被公认为求解具有可分离结构凸问题 (1.1) 的有效方法 [2], 受到广泛重视.

ADMM 方法的 k-次迭代从给定的 (y^k,λ^k) 开始, 通过

$$\begin{cases} x^{k+1}=\arg\min\{\theta_1(x)-(\lambda^k)^TAx+\frac{\beta}{2}\|Ax+By^k-b\|^2\,|\,x\in\mathcal{X}\},\\ y^{k+1}=\arg\min\{\theta_2(y)-(\lambda^k)^TBy+\frac{\beta}{2}\|Ax^{k+1}+By-b\|^2\,|\,y\in\mathcal{Y}\},\\ \lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}-b),\end{cases} \tag{1.2}$$

完成. 我们 1997 年就开始对交替方向法研究 [15]. 20 年中取得了一些进展: 在算法改进方面, 主要有交换 y 和 λ 次序, 再加延伸的"邻近点型"交替方向法 [3], 和两次校正乘子的"对称型"交替方向法 [11] 和扩展步长的算法 [12], 对算法的不精确实现也做过相应的研究 [8, 10]. 对三个和三个以上算子的问题, 直接推广的交替方向法是不能保证收敛的 [5].

[1]本文将发表于《高等学校计算数学学报》第 38 卷第 4 期，2016

我们做了一些理论上有收敛保证的求解多个可分离算子的分裂算法, 带高斯回代的交替方向法 [13] 和部分平行的正则化方法 [14].

对求解 (1.1) 的 ADMM 方法 (1.2), 本文并不讨论 (1.2) 中子问题如何具体求解, 只讨论方法的强收缩性 [9, 15], 遍历意义下和点列意义下的 $O(1/t)$ 迭代复杂性 [16, 17] 等优美的收敛性质. 为了和由 ADMM 引发的一些分裂算法在一个统一框架内讨论, 我们在关于迭代复杂性的研究 [16, 17] 中引进了辅助变量. 如果仅对 ADMM 本身, 引进辅助变量就没有必要. 本文讨论的 ADMM 性质, 其他文献中也能找到证明, 但都附加一些额外的条件 [18, 22], 有的篇幅偏长 [18], 初学者难以受益.

本文不加任何附加条件, 不用辅助变量, 只利用优化的最基本的概念和一些初等恒等式, 统一证明交替方向法 (1.2) 的强收缩性、遍历意义下和点列意义下的 $O(1/t)$ 迭代复杂性等重要性质. 为读者认识和理解交替方向法提供一个简捷的途径.

2 预备知识

我们把凸优化的最优性条件和增广 Lagrange 乘子法的主要性质作为预备知识介绍.

2.1 简单约束优化的最优性条件

交替方向法的迭代过程 (1.2) 中, 要求解的 x 和 y 子问题, 往往是一个简单的凸优化. 子问题的目标函数是两个凸函数的和, 其中一个 (例如, $\theta_1(x)$ 和 $\theta_2(y)$) 不一定光滑. 通篇, 在分析算法收敛性时, 常常用到以下的引理. 结论是显然的, 为了完整性, 我们也给出证明.

Lemma 2.1. 设 $\mathcal{X} \subset \Re^n$ 是闭凸集, $\varphi(x)$ 和 $f(x)$ 都是 $\Re^{n_1} \to \Re$ 的凸函数, 其中 $f(x)$ 在包含 $\mathcal{X}$ 的一个区域可微. 设 凸优化问题 $\min\{\varphi(x) + f(x) \,|\, x \in \mathcal{X}\}$ 的解集非空. 那么,

$$x^* = \arg\min\{\varphi(x) + f(x) \,|\, x \in \mathcal{X}\} \tag{2.3a}$$

的充分必要条件是

$$x^* \in \mathcal{X}, \quad \varphi(x) - \varphi(x^*) + (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall\, x \in \mathcal{X}. \tag{2.3b}$$

证明: 首先, 如果(2.3a) 正确, 那么对任意的 $x \in \mathcal{X}$, 我们有

$$\frac{\varphi(x_\alpha) - \varphi(x^*)}{\alpha} + \frac{f(x_\alpha) - f(x^*)}{\alpha} \geq 0, \tag{2.4}$$

其中

$$x_\alpha = (1-\alpha)x^* + \alpha x, \quad \forall \alpha \in (0, 1].$$

因为 $\varphi(x)$ 是 x 的凸函数, 根据 x_α 的定义有

$$\varphi(x_\alpha) \leq (1-\alpha)\varphi(x^*) + \alpha\varphi(x),$$

并且因此

$$\varphi(x) - \varphi(x^*) \geq \frac{\varphi(x_\alpha) - \varphi(x^*)}{\alpha}, \ \ \forall \alpha \in (0, 1].$$

将此代入 (2.4) 的左边, 我们就有

$$\varphi(x) - \varphi(x^*) + \frac{f(x_\alpha) - f(x^*)}{\alpha} \geq 0, \quad \forall \alpha \in (0, 1].$$

注意到上式中 $f(x_\alpha)=f(x^*+\alpha(x-x^*))$, 由于 $f(x)$ 可微, 令 $\alpha\to 0_+$, 得到

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

这样就从 (2.3a) 得到了 (2.3b). 反过来, 因为 f 是凸函数, 有

$$f(x_\alpha)\leq(1-\alpha)f(x^*)+\alpha f(x),$$

这可以写成

$$f(x_\alpha)-f(x^*)\leq\alpha(f(x)-f(x^*)).$$

因此, 对所有的 $\alpha\in(0,1]$, 我们有

$$f(x)-f(x^*)\geq\frac{f(x_\alpha)-f(x^*)}{\alpha}=\frac{f(x^*+\alpha(x-x^*))-f(x^*)}{\alpha},$$

取 $\alpha\to 0_+$, 有

$$f(x)-f(x^*)\geq\nabla f(x^*)^T(x-x^*).$$

将此代入 (2.3b) 的左边, 得

$$x^*\in\mathcal{X},\quad \varphi(x)-\varphi(x^*)+f(x)-f(x^*)\geq 0,\quad \forall\, x\in\mathcal{X},$$

因此 (2.3a) 为真. 引理得证. □

2.2 线性约束凸优化的最优性条件

如果不考虑可分离结构, 记 $u=(x,y)$, $\theta(u)=\theta_1(x)+\theta_2(y)$, $M=(A,B)$ 和 $\mathcal{U}=\mathcal{X}\times\mathcal{Y}$, 问题 (1.1) 就可以写成

$$\min\{\theta(u)\mid Mu=b,\ u\in\mathcal{U}\}. \tag{2.5}$$

凸优化问题 (2.5) 的 Lagrange 函数是定义在 $\mathcal{U}\times\Re^m$ 上的

$$L(u,\lambda)=\theta(u)-\lambda^T(Mu-b). \tag{2.6}$$

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

$$L(u^*,\lambda)\leq L(u^*,\lambda^*)\leq L(u,\lambda^*),\ \ \forall(u,\lambda)\in\mathcal{U}\times\Re^m, \tag{2.7}$$

则称 (u^*,λ^*) 为 Lagrange 函数 (2.6) 在 $\mathcal{U}\times\Re^m$ 上的鞍点. 鞍点 (u^*,λ^*) 中的 u^* 是 (2.5) 的解, λ^* 是相应的最优 Lagrange 乘子. 因此, 求解 (2.5) 等价于求 (2.6) 的鞍点. 不等式关系 (2.7) 可以写成

$$\begin{cases} u^* &=& \arg\min\{L(u,\lambda^*)|u\in\mathcal{U}\},\\ \lambda^* &=& \arg\max\{L(u^*,\lambda)|\lambda\in\Re^m\}.\end{cases}$$

利用 (2.6) 和引理 2.1, 上述优化问题的最优性条件是

$$\begin{cases} u^*\in\mathcal{U}, & \theta(u)-\theta(u^*)+(u-u^*)^T(-M^T\lambda^*)\geq 0,\ \ \forall\, u\in\mathcal{U},\\ \lambda^*\in\Re^m, & (\lambda-\lambda^*)^T(Mu^*-b)\geq 0,\ \ \forall\,\lambda\in\Re^m.\end{cases}$$

写成更紧凑的形式, 鞍点 (u^*,λ^*) 就是变分不等式

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\geq 0,\quad \forall\, w\in\Omega, \tag{2.8a}$$

的解, 其中

$$w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -M^T\lambda \\ Mu - b \end{pmatrix} \quad \text{和} \quad \Omega = \mathcal{U} \times \Re^m. \tag{2.8b}$$

我们将变分不等式 (2.8) 的解集记为 Ω^*. 注意到, 仿射算子 F 中的矩阵是斜对称 (skew-symmetric) 的, 总有

$$(w - \bar{w})^T(F(w) - F(\bar{w})) = 0, \quad \forall\, w, \bar{w}. \tag{2.9}$$

对线性约束的凸优化问题, 我们一直主张在变分不等式框架下考虑问题, 这分析会带来极大的方便, 得到了学者们的积极响应 [20], 称之为优雅的诠释 (elegant interpretation) [4].

2.3 求解问题 (2.5) 的增广 Lagrange 乘子法

将凸优化问题 (2.5) 的等式约束的平方项 $\frac{\beta}{2}\|Mu - b\|^2$ 作为罚函数, 加到 Lagrange 函数 $L(u, \lambda)$ (见 (2.6)), 就得到相应的增广 Lagrange 函数

$$\mathcal{L}_\beta(u, \lambda) = \theta(u) - \lambda^T(Mu - b) + \frac{\beta}{2}\|Mu - b\|^2, \quad (\beta > 0 \ \text{是等式约束的罚参数}).$$

增广 Lagrange 乘子法(Augmented Lagrangian Method–简称ALM) 的 k-次迭代从一个给定的 λ^k 开始, 通过

$$\left\{ \begin{array}{lcl} u^{k+1} &=& \arg\min\bigl\{\mathcal{L}_\beta(u, \lambda^k) \bigm| u \in \mathcal{U}\bigr\}, \\ \lambda^{k+1} &=& \lambda^k - \beta(Mu^{k+1} - b), \end{array} \right. \tag{2.10}$$

完成, 为下一次迭代提供了一个新的 λ^{k+1}. 在变分不等式 (2.8) 中, u 是原始变量, λ 是对偶变量. 在算法 (2.10) 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).

根据引理 2.1, 算法 (2.10) 提供的 $w^{k+1} = (u^{k+1}, \lambda^{k+1}) \in \Omega$ 满足

$$\left\{ \begin{array}{rcl} \theta(u) - \theta(u^{k+1}) \ + \ (u - u^{k+1})^T\{-M^T\lambda^k + \beta M^T(Mu^{k+1} - b)\} &\geq& 0, \ \ \forall u \in \mathcal{U}, \\ (\lambda - \lambda^{k+1})^T\{(Mu^{k+1} - b) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)\} &\geq& 0, \ \ \forall \lambda \in \Re^m. \end{array} \right.$$

利用关系式 $\lambda^{k+1} = \lambda^k - \beta(Mu^{k+1} - b)$ 和 (2.8) 的记号, 可以把上式写成紧凑的式子:

$$\begin{aligned} w^{k+1} \in \Omega, \quad & \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \\ & \geq (\lambda - \lambda^{k+1})^T \frac{1}{\beta}(\lambda^k - \lambda^{k+1}), \quad \forall w \in \Omega. \end{aligned} \tag{2.11}$$

上面的变分不等式中, 如果 $\lambda^k = \lambda^{k+1}$, w^{k+1} 就是变分不等式 (2.8) 的解. 将 (2.11) 中的 $w \in \Omega$ 设成一个任意的 $w^* \in \Omega^*$, 我们得到

$$(\lambda^{k+1} - \lambda^*)^T \frac{1}{\beta}(\lambda^k - \lambda^{k+1}) \geq \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1}), \ \forall w^* \in \Omega^*. \tag{2.12}$$

利用 (2.9), $(w^{k+1} - w^*)^T F(w^{k+1}) = (w^{k+1} - w^*)^T F(w^*)$. 再由 w^* 的最优性,

$$\theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^*) \geq 0,$$

推得 (2.12) 的右端非负. 最终得到

$$(\lambda^{k+1} - \lambda^*)^T(\lambda^k - \lambda^{k+1}) \geq 0. \tag{2.13}$$

在恒等式

$$\|b\|^2 = \|a\|^2 - \|a-b\|^2 - 2b^T(a-b)$$

中置 $a = \lambda^k - \lambda^*$ 和 $b = \lambda^{k+1} - \lambda^*$, 并利用 (2.13) (即 $b^T(a-b) \ge 0$), 就有

$$\|\lambda^{k+1} - \lambda^*\|^2 \le \|\lambda^k - \lambda^*\|^2 - \|\lambda^k - \lambda^{k+1}\|^2. \tag{2.14}$$

这说明增广 Lagrange 乘子法产生的迭代序列 $\{\lambda^k\}$ 有强收缩性质— 如果当前点 $\lambda^k \ne \lambda^*$, 新的迭代点 λ^{k+1} 离最优 Lagrange 乘子 λ^* 就更近. 这说明增广 Lagrange 乘子法是对偶变量的邻近点算法 (PPA) [19, 21].

3 求解凸优化问题的交替方向法 (ADMM)

前面的分析告知我们, 对可分离结构的凸优化问题 (1.1), Lagrange 函数和增广 Lagrange 函数分别是

$$L(x,y,\lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax+By-b), \tag{3.1}$$

和

$$\mathcal{L}_\beta(x,y,\lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax+By-b) + \frac{\beta}{2}\|Ax+By-b\|^2. \tag{3.2}$$

相应的变分不等式是

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w-w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \tag{3.3a}$$

其中

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \qquad \theta(u) = \theta_1(x) + \theta_2(y), \qquad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m,$$

$$w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix} \qquad \text{和} \qquad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax+By-b \end{pmatrix}. \tag{3.3b}$$

我们用 Ω^* 表示 (3.3) 的解集. 同样, 这里的仿射算子 F 有斜对称性质 (2.9).

用增广 Lagrange 乘子法 (2.10) 求解 (1.1), 迭代格式是

$$\left\{ \begin{array}{rcl} (x^{k+1}, y^{k+1}) & = & \arg\min\big\{\mathcal{L}_\beta(x,y,\lambda^k) \;\big|\; x \in \mathcal{X}, y \in \mathcal{Y}\big\}, \\ \lambda^{k+1} & = & \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \end{array} \right. \tag{3.4}$$

其中核心变量还是 λ, 迭代序列 $\{\lambda^k\}$ 具有 (2.14) 这样好的性质, 缺点是问题 (1.1) 的可分离性质没有得到利用.

既然增广 Lagrange 乘子法有很好的收敛性质. 用交替方向法 (Alternating Directions Method of Multipliers) [6, 7] 求解问题 (1.1), k-次迭代从给定的 (y^k, λ^k) 开始, 通过

$$\left\{ \begin{array}{l} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \,|\, y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b) \end{array} \right. \tag{3.5}$$

完成, 为下一次迭代提供了一个新的 (y^{k+1}, λ^{k+1}). 注意到

- 在 (3.5) 中, x-子问题和 y-子问题是利用了问题 (1.1) 的可分离性质分别求解的,
- 从 (3.4) 到 (3.5), ADMM 可以看作是松弛了的增广 Lagrange 乘子法(ALM).

ADMM 方法 (3.5) 中, x^{k+1} 是由给定的 (y^k,λ^k) 通过计算得到的, 是中间变量. 核心变量则是 $v=(y,\lambda)$. y 和 λ 分别是核心变量中的原始部分和对偶部分.

利用引理 2.1, 由 (3.5) 产生的序列 $\{w^k\}$ 有如下的性质.

Lemma 3.1. 设 $\{w^k\}$ 是用(3.5) 求解 (1.1) 所产生的序列, 那么有

$$\begin{aligned}&\theta(u)-\theta(u^{k+1})+(w-w^{k+1})^TF(w)\\ &\quad\geq\ \beta(Ax^{k+1}-Ax)^T(By^k-By^{k+1})+\frac{1}{\beta}(\lambda-\lambda^{k+1})^T(\lambda^k-\lambda^{k+1}),\quad \forall w\in\Omega.\end{aligned} \tag{3.6}$$

证明: 对 ADMM 算法 (3.5) 中的 x-子问题和 y-子问题用引理 2.1, 分别有

$$x^{k+1}\in\mathcal{X},\ \theta_1(x)-\theta_1(x^{k+1})+(x-x^{k+1})^T\{-A^T\lambda^k+\beta A^T(Ax^{k+1}+By^k-b)\}\geq 0,\ \forall\, x\in\mathcal{X},$$

和

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

利用 (3.5) 中的乘子校正公式 $\lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}-b)$, 上式可以写成

$$x^{k+1}\in\mathcal{X},\ \theta_1(x)-\theta_1(x^{k+1})+(x-x^{k+1})^T\{-A^T\lambda^{k+1}+\beta A^TB(y^k-y^{k+1})\}\geq 0,\ \forall\, x\in\mathcal{X}, \tag{3.7a}$$

和

$$y^{k+1}\in\mathcal{Y},\ \theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\{-B^T\lambda^{k+1}\}\geq 0,\ \forall\, y\in\mathcal{Y}. \tag{3.7b}$$

注意到乘子校正公式 $\lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}-b)$ 本身可以写成

$$\lambda^{k+1}\in\Re^m,\quad (\lambda-\lambda^{k+1})^T\{(Ax^{k+1}+By^{k+1}-b)+\frac{1}{\beta}(\lambda^{k+1}-\lambda^k)\}\geq 0,\ \forall\,\lambda\in\Re^m. \tag{3.8}$$

将 (3.7) 和 (3.8) 加在一起, 并利用 (3.3) 中的记号, 我们得到

$$\begin{aligned}&\theta(u)-\theta(u^{k+1})+(w-w^{k+1})^TF(w^{k+1})\\ &\quad\geq\ \beta(Ax^{k+1}-Ax)^T(By^k-By^{k+1})+\frac{1}{\beta}(\lambda-\lambda^{k+1})^T(\lambda^k-\lambda^{k+1}),\ \forall w\in\Omega.\end{aligned} \tag{3.9}$$

对上式左端中的 $(w-w^{k+1})^TF(w^{k+1})$ 利用 (2.9), 有

$$(w-w^{k+1})^TF(w^{k+1})=(w-w^{k+1})^TF(w).$$

这就证明了该引理的结论 (3.6). $\Box$

对引理 3.1 的右端, 利用一些简单的恒等式, 就可以得到下面的引理.

Lemma 3.2. 设 $\{w^k\}$ 是用 (3.5) 求解 (1.1) 所产生的序列, 则有

$$\begin{aligned}&\beta(Ax^{k+1}-Ax)^T(By^k-By^{k+1})+\frac{1}{\beta}(\lambda-\lambda^{k+1})^T(\lambda^k-\lambda^{k+1})\\ &=\ \frac{1}{2}\Big(\beta\|Ax+By^{k+1}-b\|^2+\frac{1}{\beta}\|\lambda-\lambda^{k+1}\|^2\Big)-\frac{1}{2}\Big(\beta\|Ax+By^k-b\|^2+\frac{1}{\beta}\|\lambda-\lambda^k\|^2\Big)\\ &\quad+\frac{\beta}{2}\|Ax^{k+1}+By^k-b\|^2,\quad\forall w\in\Omega.\end{aligned} \tag{3.10}$$

证明： 对 (3.10) 左端的第一部分, $\beta(Ax^{k+1}-Ax)^T(By^k-By^{k+1})$, 利用恒等式

$$(a-b)^T(c-d)=\frac{1}{2}\left((\|a+c\|^2-\|a+d\|^2)+(\|b+d\|^2-\|b+c\|^2)\right)$$

可以得到

$$\begin{aligned}
&\beta(Ax^{k+1}-Ax)^T(By^k-By^{k+1}) \\
&= \beta\{(Ax^{k+1}-b)-(Ax-b)\}^T(By^k-By^{k+1}) \\
&= \frac{\beta}{2}\left\{\|Ax^{k+1}+By^k-b\|^2-\|Ax^{k+1}+By^{k+1}-b\|^2\right\} \\
&\quad +\frac{\beta}{2}\left\{\|Ax+By^{k+1}-b\|^2-\|Ax+By^k-b\|^2\right\}. \qquad (3.11)
\end{aligned}$$

对 (3.10) 左端的第二部分, $\frac{1}{\beta}(\lambda-\lambda^{k+1})^T(\lambda^k-\lambda^{k+1})$, 在恒等式

$$b^T(b-a)=\frac{1}{2}\left(\|b\|^2-\|a\|^2+\|b-a\|^2\right)$$

中设 $a=\lambda-\lambda^k$ 和 $b=\lambda-\lambda^{k+1}$, 我们有

$$\frac{1}{\beta}(\lambda-\lambda^{k+1})^T(\lambda^k-\lambda^{k+1})=\frac{1}{2\beta}\left\{\|\lambda-\lambda^{k+1}\|^2-\|\lambda-\lambda^k\|^2+\|\lambda^k-\lambda^{k+1}\|^2\right\}. \qquad (3.12)$$

将 (3.11) 和 (3.12) 相加, 并利用 $\beta\|Ax^{k+1}+By^{k+1}-b\|^2=\frac{1}{\beta}\|\lambda^k-\lambda^{k+1}\|^2$, 就得到该引理的结论. □

上面两个引理的证明非常简单, 只用到了引理 2.1 和两个初等的恒等式. 这些证明可以分别在作者的系列讲义 [9, 23] 的第 11 讲和文献 [22] 的附录中找到. 至此, 我们已经做好了证明 ADMM 主要性质的全部准备.

4 交替方向法的收敛性质

交替方向法的一些重要性质, 诸如强收缩性质 [1], 遍历意义下和强点列意义下的 $O(1/t)$ 迭代复杂性 [16, 17], 很容易从引理 3.1 和引理 3.2 推得. 由 (3.6) 和 (3.10), 我们有

$$\begin{aligned}
&\theta(u)-\theta(u^{k+1})+(w-w^{k+1})^TF(w) \\
&\ge \frac{1}{2}\Big(\beta\|Ax+By^{k+1}-b\|^2+\frac{1}{\beta}\|\lambda-\lambda^{k+1}\|^2\Big) \\
&\quad -\frac{1}{2}\Big(\beta\|Ax+By^k-b\|^2+\frac{1}{\beta}\|\lambda-\lambda^k\|^2\Big) \\
&\quad +\frac{1}{2}\beta\|Ax^{k+1}+By^k-b\|^2, \quad \forall w\in\Omega. \qquad (4.1)
\end{aligned}$$

4.1 强收缩性质

与 ALM (2.10) 的强收缩性质 (2.14) 类似, ADMM (3.5) 的相应性质表述为以下的定理:

Theorem 4.1. 设 $\{w^k\}$ 是用交替方向法 (3.5) 求解 (1.1) 所产生的序列, 则有

$$\|v^{k+1}-v^*\|_H^2\le\|v^k-v^*\|_H^2-\|v^k-v^{k+1}\|_H^2,\ \ \forall\, v^*\in\mathcal{V}^*, \qquad (4.2)$$

其中

$$v = \begin{pmatrix} y \\ \lambda \end{pmatrix}, \quad H = \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix} \quad \text{和} \quad \mathcal{V}^* = \{(y^*, \lambda^*) \mid (x^*, y^*, \lambda^*) \in \Omega^*\}. \tag{4.3}$$

假如 B 是列满秩矩阵, $\{v^k\}$ 收敛于解集 $\mathcal{V}^*$ 中的一点 v^∞.

证明. 在 (4.1) 中令 $w = w^*$ 并利用 $Ax^* - b = -By^*$, 我们有

$$\begin{aligned} &\Big(\beta\|By^k - By^*\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^*\|^2\Big) - \Big(\beta\|By^{k+1} - By^*\|^2 + \frac{1}{\beta}\|\lambda^{k+1} - \lambda^*\|^2\Big) \\ &\quad \geq \quad \beta\|Ax^{k+1} + By^k - b\|^2 + 2\big\{\theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^*)\big\}. \end{aligned}$$

由于 $\theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^*) \geq 0$, 就有

$$\begin{aligned} &\Big(\beta\|By^{k+1} - By^*\|^2 + \frac{1}{\beta}\|\lambda^{k+1} - \lambda^*\|^2\Big) \\ &\quad \leq \quad \Big(\beta\|By^k - By^*\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^*\|^2\Big) - \beta\|Ax^{k+1} + By^k - b\|^2. \end{aligned} \tag{4.4}$$

如果我们对 (4.4) 的右端的 $\beta\|Ax^{k+1} + By^k - b\|^2$ 证明了

$$\beta\|Ax^{k+1} + By^k - b\|^2 \geq \beta\|B(y^k - y^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2, \tag{4.5}$$

再利用 (4.3) 中的记号, 就完成了对 (4.2) 的证明. 下面, 我们证明 (4.5) 成立. 首先, (3.7b) 式表示

$$y^{k+1} \in \mathcal{Y},\ \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{-B^T\lambda^{k+1}\} \geq 0,\ \forall\, y \in \mathcal{Y}. \tag{4.6}$$

以 $k-1$ 置换 (4.6) 中的 k, 就有

$$y^k \in \mathcal{Y},\ \ \theta_2(y) - \theta_2(y^k) + (y - y^k)^T\{-B^T\lambda^k\} \geq 0,\ \ \forall\, y \in \mathcal{Y}. \tag{4.7}$$

将 (4.6) 和 (4.7) 中的任意的 $y \in \mathcal{Y}$ 分别设成 y^k 和 y^{k+1}, 然后将两式相加, 有

$$(\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}) \geq 0.$$

利用 $Ax^{k+1} + By^{k+1} - b = \frac{1}{\beta}(\lambda^k - \lambda^{k+1})$ 和上式, 就有

$$\begin{aligned} \beta\|Ax^{k+1} + By^k - b\|^2 &= \beta\|(Ax^{k+1} + By^{k+1} - b) + B(y^k - y^{k+1})\|^2 \\ &= \beta\|\frac{1}{\beta}(\lambda^k - \lambda^{k+1}) + B(y^k - y^{k+1})\|^2 \\ &\geq \beta\|B(y^k - y^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2, \end{aligned}$$

不等式 (4.5) 成立, 结论 (4.2) 得证.

假如 B 列满秩, 则矩阵 H 正定. 不等式 (4.2) 表示序列 $\{v^k\}$ 是有界序列，因此它有收敛子列 $\{v^{k_j}\}$, 设其收敛于 v^∞. 又根据 (4.2), 有 $\sum_{k=0}^\infty \|v^k - v^{k+1}\|_H^2 \leq \|v^0 - v^*\|_H^2$ 和 $\lim_{k\to\infty}\|v^k - v^{k+1}\|_H^2 = 0$, 因此 $\lim_{j\to\infty}\|v^{k_j} - v^{k_j+1}\|_H^2 = 0$ 也成立. 令 $j \to \infty$, (3.9) 式说明 v^∞ 就是 (3.3) 的解. 这同时说明 $\{v^k\}$ 的任何一个聚点都是 (3.3) 的解. 再根据 (4.2), $\{v^k\}$ 不可能有多于一个的聚点, 这就证明了 $\{v^k\}$ 收敛于 $v^\infty \in \mathcal{V}^*$. $\square$

没有矩阵 B 列满秩的假设, 就会有 $\lim_{k\to\infty}(By^k, \lambda^k) = (By^*, \lambda^*)$, 其中 $(y^*, z^*) \in \mathcal{V}^*$. 有了前一节的引理, 证明定理 4.1 的关键不等式是 (4.5). 不等式 (4.5) 的证明也可以在作者的系列讲义中英文版 [9, 23] 的第 11 讲的引理 3.2 中找到.

4.2 遍历意义下的 $O(1/t)$ 迭代复杂性

根据 (3.3), 如果 $\tilde{w}$ 满足

$$\tilde{w} \in \Omega, \ \ \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(\tilde{w}) \ge 0, \ \ \forall\, w \in \Omega,$$

$\tilde{w}$ 就是变分不等式 (3.3) 的解. 利用 $(w - \tilde{w})^T F(\tilde{w}) = (w - \tilde{w})^T F(w)$ (见(2.9)), $\tilde{w}$ 就是变分不等式 (3.3) 的解也可以表述成

$$\tilde{w} \in \Omega, \ \ \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \ge 0, \ \ \forall\, w \in \Omega.$$

因此, 可以利用上式来刻画变分不等式 (3.3) 的近似解. 具体说来, 对任给的 $\epsilon > 0$, 如果

$$\tilde{w} \in \Omega, \ \ \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \ge -\epsilon, \ \ \forall\, w \in \mathcal{D}_{(\tilde{w})},$$

其中

$$\mathcal{D}_{(\tilde{w})} = \{w \in \Omega \,|\, \|w - \tilde{w}\| \le 1\},$$

那么 $\tilde{w} \in \Omega$ 就称为变分不等式 (3.3) 的 ϵ-近似解. 以下我们证明 ADMM (3.5) 经过 t 次迭代, 就可以找到一个 $\tilde{w} \in \Omega$, 使得

$$\tilde{w} \in \Omega \quad \text{and} \quad \sup_{w \in \mathcal{D}_{(\tilde{w})}} \big\{\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w)\big\} \le \epsilon = O(\frac{1}{t}). \tag{4.8}$$

相关的定理在 [16] 通过引进辅助变量已经得到证明. 这里我们不用任何辅助变量, 直接用由引理 3.1 和引理 3.2 得到的 (4.1) 式加以证明.

Theorem 4.2. 设 $\{w^k\}$ 是由 *ADMM* (3.5) 求解 (3.3) 所产生的序列. 对任意的正整数 $t > 0$, 我们有

$$\theta(u_t) - \theta(u) + (w_t - w)^T F(w) \le \frac{1}{2(t+1)}\big(\beta\|Ax + By^0 - b\|^2 + \frac{1}{\beta}\|\lambda - \lambda^0\|^2\big), \ \forall w \in \Omega \tag{4.9}$$

其中

$$w_t = \frac{1}{t+1}\Big(\sum_{k=0}^{t} w^{k+1}\Big). \tag{4.10}$$

证明. 首先, 由 (4.1), 对任意的 $k > 0$,

$$\begin{aligned}
&\theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w) \\
&\ \ \ge \ \ \frac{1}{2}\Big(\beta\|Ax + By^{k+1} - b\|^2 + \frac{1}{\beta}\|\lambda - \lambda^{k+1}\|^2\Big) - \frac{1}{2}\Big(\beta\|Ax + By^k - b\|^2 + \frac{1}{\beta}\|\lambda - \lambda^k\|^2\Big).
\end{aligned}$$

上式可以改写成

$$\begin{aligned}
&\theta(u^{k+1}) - \theta(u) + (w^{k+1} - w)^T F(w) + \frac{1}{2}\Big(\beta\|Ax + By^{k+1} - b\|^2 + \frac{1}{\beta}\|\lambda - \lambda^{k+1}\|^2\Big) \\
&\ \ \le \ \ \frac{1}{2}\Big(\beta\|Ax + By^k - b\|^2 + \frac{1}{\beta}\|\lambda - \lambda^k\|^2\Big).
\end{aligned} \tag{4.11}$$

上面的不等式对 $k = 0, 1, 2, \ldots, t$ 连加, 我们得到

$$\begin{aligned}
&\sum_{k=0}^{t}\theta(u^{k+1}) - (t+1)\theta(u) + (\sum_{k=0}^{t} w^{k+1} - (t+1)w)^T F(w) \\
&\ \ \le \ \ \frac{1}{2}\Big(\beta\|Ax + By^0 - b\|^2 + \frac{1}{\beta}\|\lambda - \lambda^0\|^2\Big), \quad \forall\, w \in \Omega.
\end{aligned}$$

因此就有

$$\begin{aligned}&\frac{1}{t+1}\big(\sum_{k=0}^{t}\theta(u^{k+1})\big)-\theta(u)+(\tilde{w}_t-w)^TF(w)\\ \le\quad &\frac{1}{2(t+1)}\Big(\beta\|Ax+By^0-b\|^2+\frac{1}{\beta}\|\lambda-\lambda^0\|^2\Big).\end{aligned}\tag{4.12}$$

由于 $\theta(u)$ 是凸函数并且 $u_t=\frac{1}{t+1}\big(\sum_{k=0}^{t}u^{k+1}\big)$, 我们有

$$\theta(u_t)\le\frac{1}{t+1}\big(\sum_{k=0}^{t}\theta(u^{k+1})\big).$$

代入 (4.12) 的左端, 就完成了定理的证明. □

ADMM (3.5) 经过 t 次迭代, 由 (4.10) 定义的 $\tilde{w}_t$ 满足

$$\tilde{w}\in\Omega\quad\text{and}\quad\sup_{w\in\mathcal{D}_{(\tilde{w})}}\big\{\theta(\tilde{u})-\theta(u)+(\tilde{w}-w)^TF(w)\big\}\le\frac{d}{2t}=O(\frac{1}{t}),$$

其中

$$d:=\sup\{\beta\|Ax+By^0-b\|^2+\frac{1}{\beta}\|\lambda-\lambda^0\|^2\,|\,w\in\mathcal{D}_{(\tilde{w})}\},$$

$v^0=(y^0,\lambda^0)$ 是初始迭代点. 这就意味着 $\tilde{w}_t$ 是 (3.3) 的 $O(1/t)$ 近似解. 我们把这样的收敛速率称为遍历意义下的 $O(1/t)$ 迭代复杂性.

4.3 点列意义下的 $O(1/t)$ 迭代复杂性

在定理 4.1 中, 我们证明了交替方向法 (3.5) 所产生的序列满足

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

因而, 有

$$\sum_{k=0}^{\infty}\|v^k-v^{k+1}\|_H^2\le\|v^0-v^*\|_H^2,\ \ \forall\,v^*\in\mathcal{V}^*.\tag{4.13}$$

点列意义下的 $O(1/t)$ 收敛速率是要对任意的正整数 $t>0$, 都有

$$\|v^t-v^{t+1}\|_H^2\le\frac{1}{t}\|v^0-v^*\|_H^2,\ \ \forall\,v^*\in\mathcal{V}^*.\tag{4.14}$$

上式左端用二次项 $\|v^t-v^{t+1}\|_H^2$, 是因为类比可微无约束优化问题 $\min_x f(x)$ 中, 如果 f 二次连续可微, 在最优点附近有 $f(x)-f(x^*)=O(\|x-x^*\|^2)$. 证明 (4.14), 利用 (4.13), 只要像 [17] 中那样证明序列 $\{\|v^k-v^{k+1}\|_H^2\}$ 是单调不增的. 下面我们直接用引理 3.1 证明 $\{\|v^k-v^{k+1}\|_H^2\}$ 是单调不增.

Theorem 4.3. 设 $\{w^k\}$ 是用交替方向法 (3.5) 求解 (3.3) 所产生的序列, 则有

$$\|v^k-v^{k+1}\|_H^2\le\|v^{k-1}-v^k\|_H^2,\ \ \forall\,k\ge1.\tag{4.15}$$

证明. 将 (3.6) 的右端改写, 得到需要的形式:

$$
\begin{aligned}
&\beta(Ax^{k+1}-Ax)^T(By^k-By^{k+1})+\frac{1}{\beta}(\lambda-\lambda^{k+1})^T(\lambda^k-\lambda^{k+1}) \\
= \;& \beta\{(Ax^{k+1}-Ax)+(By^{k+1}-By)\}^T(By^k-By^{k+1}) \\
&+\beta(By-By^{k+1})^T(By^k-By^{k+1})+\frac{1}{\beta}(\lambda-\lambda^{k+1})^T(\lambda^{k+1}-\lambda^k),\ \forall w\in\Omega.
\end{aligned} \tag{4.16}
$$

利用 (4.3) 中的记号, (3.6) 就有

$$
\begin{aligned}
&\theta(u)-\theta(u^{k+1})+(w-w^{k+1})^TF(w) \\
&\geq\ \beta\{(Ax^{k+1}-Ax)+(By^{k+1}-By)\}^TB(y^k-y^{k+1}) \\
&\quad +(v-v^{k+1})^TH(v^k-v^{k+1}),\quad \forall w\in\Omega.
\end{aligned} \tag{4.17}
$$

在上式中取 $w=w^k$, 利用 $\beta(Ax^{k+1}+By^{k+1}-b)=(\lambda^k-\lambda^{k+1})$ 和 $\beta(Ax^k+By^k-b)=(\lambda^{k-1}-\lambda^k)$, 就有

$$
\begin{aligned}
&\theta(u^k)-\theta(u^{k+1})+(w^k-w^{k+1})^TF(w^k) \\
&\geq\ \{(\lambda^k-\lambda^{k+1})-(\lambda^{k-1}-\lambda^k)\}^TB(y^k-y^{k+1}) \\
&\quad +(v^k-v^{k+1})^TH(v^k-v^{k+1}),\quad \forall w\in\Omega.
\end{aligned} \tag{4.18}
$$

将 (4.17) 中的 k 置换成 $k-1$, 有

$$
\begin{aligned}
&\theta(u)-\theta(u^k)+(w-w^k)^TF(w) \\
&\geq\ \beta\{(Ax^k-Ax)+(By^k-By)\}^TB(y^{k-1}-y^k) \\
&\quad +(v-v^k)^TH(v^{k-1}-v^k),\quad \forall w\in\Omega.
\end{aligned} \tag{4.19}
$$

在上式中取 $w=w^{k+1}$, 同样利用 $\beta(Ax^{k+1}+By^{k+1}-b)=(\lambda^k-\lambda^{k+1})$, $\beta(Ax^k+By^k-b)=(\lambda^{k-1}-\lambda^k)$ 和 (4.3) 中的记号有

$$
\begin{aligned}
&\theta(u^{k+1})-\theta(u^k)+(w^{k+1}-w^k)^TF(w^{k+1}) \\
&\geq\ \{(\lambda^{k-1}-\lambda^k)-(\lambda^k-\lambda^{k+1})\}^TB(y^{k-1}-y^k) \\
&\quad +(v^{k+1}-v^k)^TH(v^{k-1}-v^k),\quad \forall w\in\Omega.
\end{aligned} \tag{4.20}
$$

将 (4.18) 和 (4.20) 相加, 并利用 $(w^k-w^{k+1})^T\{F(w^k)-F(w^{k+1})\}\equiv 0$ (见 (2.9)) 得到

$$
\begin{aligned}
&(v^k-v^{k+1})^TH\{(v^{k-1}-v^k)-(v^k-v^{k+1})\} \\
&\geq\ \{(\lambda^{k-1}-\lambda^k)-(\lambda^k-\lambda^{k+1})\}^TB\{(y^{k-1}-y^k)-(y^k-y^{k+1})\}.
\end{aligned} \tag{4.21}
$$

在恒等式 $\|a\|_H^2-\|b\|_H^2=\|a-b\|_H^2+2b^TH(a-b)$ 中置 $a=(v^{k-1}-v^k)$, $b=(v^k-v^{k+1})$, 有

$$
\begin{aligned}
&\|v^{k-1}-v^k\|_H^2-\|v^k-v^{k+1}\|_H^2 \\
&=\ \|(v^{k-1}-v^k)-(v^k-v^{k+1})\|_H^2 \\
&\quad +2(v^k-v^{k+1})^TH\{(v^{k-1}-v^k)-(v^k-v^{k+1})\}.
\end{aligned} \tag{4.22}
$$

将 (4.21) 代入 (4.22) 的交叉项, 我们得到

$$
\begin{aligned}
&\|v^{k-1}-v^k\|_H^2-\|v^k-v^{k+1}\|_H^2 \\
&\geq\ \|(v^{k-1}-v^k)-(v^k-v^{k+1})\|_H^2 \\
&\quad +2\{(\lambda^{k-1}-\lambda^k)-(\lambda^k-\lambda^{k+1})\}^TB\{(y^{k-1}-y^k)-(y^k-y^{k+1})\}.
\end{aligned} \tag{4.23}
$$

注意到, 根据 H 的结构 (见(4.3)), (4.23) 式的右端等于

$$\left\|\sqrt{\beta}B[(y^{k-1}-y^k)-(y^k-y^{k+1})]+\frac{1}{\sqrt{\beta}}[(\lambda^{k-1}-\lambda^k)-(\lambda^k-\lambda^{k+1})]\right\|^2,$$

非负, 因此定理的结论得证. $\Box$

5 结论

交替方向法 (ADMM) 的应用范围越来越广, 人们对 ADMM 的主要性质也愈感兴趣. 这篇短文证明了 ADMM 的主要收敛性质, 所用的工具非常初等, 证明也特别简单. 虽然交替方向法有许多进展, 变分不等式总是分析的有力工具 [4], 收敛性证明基本上可以采用本文展示的套路, 容易掌握.

References

[1] Blum E and Oettli W. Mathematische Optimierung, Econometrics and Operations Research XX, Springer Verlag, 1975.

[2] Boyd S, Parikh N, Chu E, Peleato B and Eckstein J. Distributed optimization and statistical learning via the alternating direction method of multipliers. Foun. Trends Mach. Learn., 2010, 3: 1-122.

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

[4] Chambolle A and Pock T. On the ergodic convergence rates of a first-order primal – dual algorithm, Mathematical Programming, 2016, 159: 253 – 287.

[5] Chen C H, He B S, Ye Y Y and Yuan X M, The direct extension of ADMM for multi-block convex minimization problems is not necessary convergent. Mathematical Programming, 2016, 155: 57-79.

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

[7] Glowinski R and Marrocco A. Sur l'approximation par éléments finis d'ordre un et la résolution par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires, Revue Fr. Autom. Inform. Rech. Opér., Anal. Numér. 1975, 2: 41–76.

[8] Gu G Y, He B S and Yang J F. Inexact Alternating-Direction-Based Contraction Methods for Separable Linearly Constrained Convex Optimization. JOTA, 2014, 2014: 105-129.

[9] He B S. Alternating direction methods of multipliers for separable convex programming. Lectures of "Contraction Methods for Convex Optimization and Monotone Variational Inequalities", No. 11, 2009, http://maths.nju.edu.cn/~hebma

[10] He B S, Liao L Z, Han D R and Yang H. A new inexact alternating directions method for monotone variational inequalities. Math. Progr., 2002, 92: 103–118.

[11] He B S, Liu H, Wang Z R and Yuan X M. A strictly contractive Peaceman-Rachford splitting method for convex programming. SIAM Journal on Optimization, 2014, 24: 1011-1040.

[12] He B S, Ma F and Yuan X M. Convergence Study on the Symmetric Version of ADMM with Larger Step Sizes. to appaer in SIAM Journal on Image Science, 2016.

[13] He B S, Tao M and Yuan X M. Alternating Direction Method with Gaussian Back Substitution for Separable Convex Programming. SIAM J. Optim., 2012, 22: 313-340.

[14] He B S, Tao M and Yuan X M. A splitting method for separable convex programming. IMA J. Numerical Analysis, 2015, 31:394-426.

[15] He B S and Yang H. Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities. Operations research letters 1998, 23: 151-161.

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

[17] B.S. He and X.M. Yuan, On non-ergodic convergence rate of Douglas-Rachford alternating directions method of multipliers, Numerische Mathematik, 130 (2015) 567-577.

[18] Monteiro R D C and Svaiter B F. Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. SIAM J. Optim., 2013, 23:475 – 507.

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

[20] Pock T and Chambolle A. Diagonal preconditioning for first order primal-dual algorithms in convex optimization. IEEE International Conference on Computer Vision, 2012, pp. 1762-1769.

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

[22] Wang H H and Banerjee A. Online Alternating Direction Method, Dept of Computer Science & Engg, University of Minnesota, Twin Cities, Appearing in Proceddings of the 29^{th} International Conference on Machine Learning, Edinburgh, Scotland, UK, 2012.

[23] 何炳生, 结构型优化的交替方向法, 《凸优化和单调变分不等式的收缩算法》第11 讲, 2009, http://maths.nju.edu.cn/~hebma