从最优化方法的基本原理到凸优化的分裂收缩算法

I. 从邻近点算法到均困的ALM和ADMM方法

中学的数理基础　必要的社会实践
普通的大学数学　一般的优化原理

何 炳 生　　南京大学数学系

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

MISAI清华三亚国际数学论坛 2024年2月15日

连续优化中一些代表性数学模型

1. 鞍点问题 $\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}}\{\Phi(x,y)=\theta_1(x)-y^TAx-\theta_2(y)\}$
2. 线性约束的凸优化问题 $\min\{\theta(x)|Ax=b\,(\text{or}\geq b),\ x\in\mathcal{X}\}$
3. 结构型凸优化 $\min\{\theta_1(x)+\theta_2(y)|Ax+By=b,\ x\in\mathcal{X},\ y\in\mathcal{Y}\}$
4. 多块可分离凸优化 $\min\{\sum_{i=1}^p\theta_i(x_i)|\sum_{i=1}^pA_ix_i=b,x_i\in\mathcal{X}_i\}$

变分不等式(VI) 是瞎子爬山的数学表达形式

邻近点算法(PPA) 是步步为营 稳扎稳打的求解方法.

变分不等式和邻近点算法是分析和设计凸优化方法的两大法宝.

分裂是指迭代中子问题都通过分拆求解. 收缩算法有别于可行方向法, 又有别于下降算法, 它的迭代点离优化问题的拉格朗日函数的鞍点越来越近.

先解释上述问题如何化为一个单调变分不等式 并介绍什么是变分不等式的邻近点算法

1 Optimization problem and VI

1.1 Differential convex optimization in Form of VI

Let $\Omega \subset \Re^n$, we consider the convex minimization problem

$$\min\{f(x) \mid x \in \Omega\}. \tag{1.1}$$

What is the first-order optimal condition ?

$x^* \in \Omega^* \quad \Leftrightarrow \quad x^* \in \Omega$ and any feasible direction is not a descent one.

Optimal condition in variational inequality form

- $S_d(x^*) = \{s \in \Re^n \mid s^T \nabla f(x^*) < 0\} \;=\;$ Set of the descent directions.
- $S_f(x^*) = \{s \in \Re^n \mid s = x - x^*,\; x \in \Omega\} \;=\;$ Set of feasible directions.

$x^* \in \Omega^* \quad \Leftrightarrow \quad x^* \in \Omega$ **and** $S_f(x^*) \cap S_d(x^*) = \emptyset$.

瞎子爬山判定山顶的准则是: 所有可行方向都不再是上升方向

The optimal condition can be presented in a variational inequality (VI) form:

$$x^* \in \Omega, \quad (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall x \in \Omega. \tag{1.2}$$

若 $\Omega = R^n$, 跟任意向量 $(x - x^*)$ 内积都非负的只有零向量. 无约束优化最优点必须满足 $\nabla f(x^*) = 0$.

Substituting $\nabla f(x)$ with an operator F (from $\Re^n$ into itself), we get a classical VI.

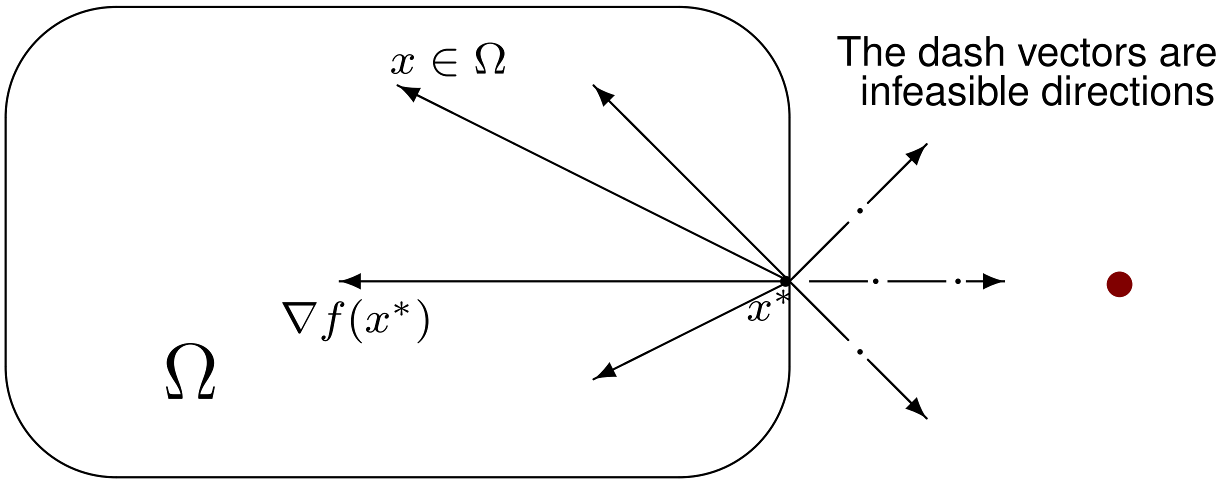


Fig. 1.1 Differential Convex Optimization and VI

Since $f(x)$ is a convex function, we have

$$\begin{array}{l} f(y) \geq f(x) + \nabla f(x)^T (y - x) \\ f(x) \geq f(y) + \nabla f(y)^T (x - y) \end{array} \quad \text{thus} \quad (x - y)^T (\nabla f(x) - \nabla f(y)) \geq 0.$$

We say the gradient ∇f of the convex function f is a monotone operator.

通篇我们需要用到的大学数学　主要是基于微积分学的一个引理

$$x^* \in \mathrm{argmin}\{\theta(x)|x \in \mathcal{X}\} \Leftrightarrow \; x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) \ge 0, \quad \forall x \in \mathcal{X};$$

$$x^* \in \mathrm{argmin}\{f(x)|x \in \mathcal{X}\} \Leftrightarrow \; x^* \in \mathcal{X}, \quad (x - x^*)^T \nabla f(x^*) \ge 0, \; \forall x \in \mathcal{X}.$$

上面的凸优化最优性条件是最基本的, 看起来合在一起就是下面的引理:

定理 1 *Let* $\mathcal{X} \subset \Re^n$ *be a closed convex set,* $\theta(x)$ *and* $f(x)$ *be convex functions and* $f(x)$ *is differentiable. Assume that the solution set of the minimization problem* $\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\}$ *is nonempty. Then,*

$$x^* \in \arg\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\} \tag{1.3a}$$

if and only if　　该凸优化最优性条件定理是我们分析的基础

$$x^* \in \mathcal{X}, \;\; \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \ge 0, \;\; \forall x \in \mathcal{X}. \tag{1.3b}$$

定理 1 把优化问题 (1.3a) 转换成了等价的变分不等式 (1.3b).

1.2 Linear constrained convex optimization and VI

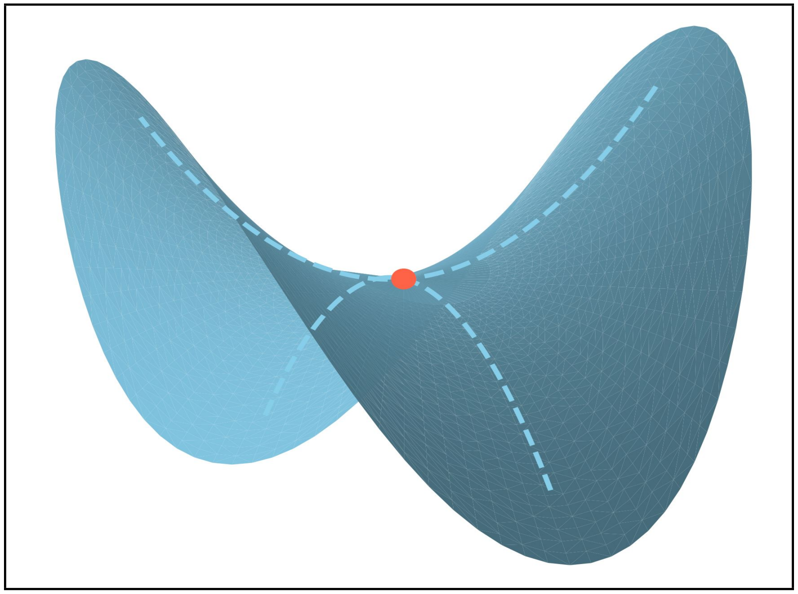
We consider the linearly constrained convex optimization problem

$$\min\{\theta(u) \mid \mathcal{A}u = b,\ u \in \mathcal{U}\}. \qquad (1.4)$$

The Lagrangian function of the problem (1.4) is

$$L(u,\lambda) = \theta(u) - \lambda^T(\mathcal{A}u - b), \qquad (1.5)$$

which is defined on $\mathcal{U} \times \Re^m$.



A pair of (u^*, λ^*) is called a saddle point of the Lagrange function (1.5), if $(u^*, \lambda^*) \in \mathcal{U} \times \Re^m$, and

$$L(u^*,\lambda) \le L(u^*,\lambda^*) \le L(u,\lambda^*), \ \ \forall (u,\lambda) \in \mathcal{U} \times \Re^m.$$

我们关心如何求解得 Lagrange 函数的鞍点.

The above inequalities can be written as

$$\begin{cases} u^* \in \mathcal{U}, & L(u,\lambda^*) - L(u^*,\lambda^*) \geq 0, \quad \forall\, u \in \mathcal{U}, & (1.6a)\\ \lambda^* \in \Re^m, & L(u^*,\lambda^*) - L(u^*,\lambda) \geq 0, \quad \forall\, \lambda \in \Re^m. & (1.6b) \end{cases}$$

According to the definition of $L(u,\lambda)$ (see(1.5)), it follows from (1.6a) that

$$u^* \in \mathcal{U}, \;\; \theta(u) - \theta(u^*) + (u-u^*)^T(-\mathcal{A}^T\lambda^*) \geq 0, \;\; \forall\, u \in \mathcal{U}. \qquad (1.7)$$

Similarly, from (1.6b), we have

$$\lambda^* \in \Re^m, \;\; (\lambda - \lambda^*)^T(\mathcal{A}u^* - b) \geq 0, \;\; \forall\, \lambda \in \Re^m. \qquad (1.8)$$

Writing (1.7) and (1.8) together, we get the following variational inequality:

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

Using a more compact form, the saddle-point can be characterized as the solution of the following VI:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \tag{1.9a}$$

where

$$w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -\mathcal{A}^T \lambda \\ \mathcal{A}u - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{U} \times \Re^m. \tag{1.9b}$$

Setting $w = (u, \lambda^*)$ and $w = (u^*, \lambda)$ in (1.9), we get (1.7) and (1.8), respectively. Because F is an affine operator and

$$F(w) = \begin{pmatrix} 0 & -\mathcal{A}^T \\ \mathcal{A} & 0 \end{pmatrix} \begin{pmatrix} u \\ \lambda \end{pmatrix} - \begin{pmatrix} 0 \\ b \end{pmatrix}.$$

The matrix is skew-symmetric, we have

$$(w - \tilde{w})^T (F(w) - F(\tilde{w})) \equiv 0.$$

线性约束的凸优化问题 (1.4), 转换成了混合变分不等式 (1.9).

Two block separable convex optimization

We consider the following structured separable convex optimization

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b,\ x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{1.10}$$

This is a special problem of (1.4) with

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \mathcal{U} = \mathcal{X} \times \mathcal{Y}, \quad \mathcal{A} = (A, B).$$

The Lagrangian function of the problem (1.10) is

$$L^{(2)}(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax + By - b).$$

The same analysis tells us that the saddle point is a solution of the following VI:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{1.11}$$

where

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \theta(u) = \theta_1(x) + \theta_2(y), \quad w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \tag{1.12a}$$

$$F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix}, \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m. \tag{1.12b}$$

The affine operator $F(w)$ has the form

$$F(w) = \begin{pmatrix} 0 & 0 & -A^T \\ 0 & 0 & -B^T \\ A & B & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ b \end{pmatrix}.$$

Again, due to the skew-symmetry, we have $(w - \tilde{w})^T(F(w) - F(\tilde{w})) \equiv 0.$

可分离线性约束凸优化问题 (1.10), 转换成了变分不等式 (1.11)–(1.12).

Convex optimization problem with three separable functions

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z) \mid Ax+By+Cz=b,\ x\in\mathcal{X}, y\in\mathcal{Y}, z\in\mathcal{Z}\},$$

is a special problem of (1.4) with three blocks. The Lagrangian function is

$$L^{(3)}(x,y,z,\lambda)=\theta_1(x)+\theta_2(y)+\theta_3(z)-\lambda^T(Ax+By+Cz-b).$$

The same analysis tells us that the saddle point is a solution of the following VI:

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega.$$

where $\theta(u)=\theta_1(x)+\theta_2(y)+\theta_3(z)$,

$$w=\begin{pmatrix} x\\ y\\ z\\ \lambda \end{pmatrix},\quad u=\begin{pmatrix} x\\ y\\ z \end{pmatrix},\quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ -C^T\lambda\\ Ax+By+Cz-b \end{pmatrix},$$

and $\qquad \Omega=\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}\times\Re^m.$

求线性约束凸优化拉格朗日函数的鞍点, 都转换成了相应的变分不等式.

2 Proximal point algorithms and its Beyond

引理 1 *Let the vectors* $a, b \in \Re^n$, $H \in \Re^{n\times n}$ *be a positive definite matrix. If* $b^T H(a-b) \geq 0$, *then we have*

$\|x\|^2 = x^T x, \quad \|x\|_H^2 = x^T H x.$

$$\|b\|_H^2 \leq \|a\|_H^2 - \|a-b\|_H^2. \tag{2.1}$$

The assertion follows from $\|a\|_H^2 = \|b+(a-b)\|_H^2 \geq \|b\|_H^2 + \|a-b\|_H^2$.

2.1 Proximal point algorithms for convex optimization

Convex Optimization Now, let us consider the *simple* convex optimization

$$\min\{\theta(x) + f(x) \mid x \in \mathcal{X}\}, \tag{2.2}$$

where $\theta(x)$ and $f(x)$ are convex but $\theta(x)$ is not necessary smooth, $\mathcal{X}$ is a closed convex set. For solving (2.2), the k-th iteration of the proximal point algorithm (abbreviated to PPA) [27, 29] begins with a given x^k, offers the new iterate x^{k+1} via the recursion

邻近点算法

$$x^{k+1} = \operatorname{argmin}\{\theta(x) + f(x) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}. \tag{2.3}$$

Since x^{k+1} is the optimal solution of (2.3), it follows from Theorem 1 that

$$\theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T \{\nabla f(x^{k+1}) + r(x^{k+1} - x^k)\} \geq 0, \ \forall x \in \mathcal{X}. \tag{2.4}$$

Setting $x = x^*$ in the above inequality, it follows that

$$(x^{k+1} - x^*)^T r(x^k - x^{k+1}) \geq \theta(x^{k+1}) - \theta(x^*) + (x^{k+1} - x^*)^T \nabla f(x^{k+1}).$$

Because f isconvex, $(x^{k+1} - x^*)^T \nabla f(x^{k+1}) \geq (x^{k+1} - x^*)^T \nabla f(x^*)$, it follows that

$$\begin{aligned} &\theta(x^{k+1}) - \theta(x^*) + (x^{k+1} - x^*)^T \nabla f(x^{k+1}) \\ \geq \quad &\theta(x^{k+1}) - \theta(x^*) + (x^{k+1} - x^*)^T \nabla f(x^*) \geq 0 \end{aligned}$$

and consequently,

$$(x^{k+1} - x^*)^T (x^k - x^{k+1}) \geq 0. \tag{2.5}$$

Let $a = x^k - x^*$ and $b = x^{k+1} - x^*$ and using Lemma 1, we obtain

PPA 算法的收缩性质 $$\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\|^2 \leq \|\boldsymbol{x}^k - \boldsymbol{x}^*\|^2 - \|\boldsymbol{x}^k - \boldsymbol{x}^{k+1}\|^2, \tag{2.6}$$

which is the nice convergence property of Proximal Point Algorithm.

We write the problem (2.2) and its PPA (2.3) in VI form

For the optimization problem (2.2) , namely, $\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\}$,
the equivalent variational inequality form is

$$x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \ge 0, \quad \forall\, x \in \mathcal{X}. \tag{2.7}$$

For solving the problem (2.2), the PPA is

$$x^{k+1} = \text{Argmin}\{\theta(x) + f(x) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}.$$

variational inequality form of the k-th iteration of the PPA (see (2.4)) is:

$$\begin{aligned} x^{k+1} \in \mathcal{X}, \quad \theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T \nabla f(x^{k+1}) & \\ \ge (x - x^{k+1})^T r(x^k - x^{k+1}), \quad \forall\, x \in \mathcal{X}. & \end{aligned} \tag{2.8}$$

PPA 通过求解一系列的 (2.3), 求得 (2.2) 的解, 采用的是步步为营的策略.

The solution of (2.8) is Proximal Point, it has the contraction property (2.6).

2.2 Preliminaries of PPA for Variational Inequalities

The optimal condition of the linearly constrained convex optimization is characterized as a mixed monotone variational inequality:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega. \tag{2.9}$$

$\boldsymbol{u}$ 往往指自变量，向量 $\boldsymbol{w}$ 包含自变量 $\boldsymbol{u}$ 和对偶变量 $\boldsymbol{\lambda}$.

PPA for VI (2.9) in H-norm (定义)

For given w^k and $H \succ 0$, find w^{k+1} such that

邻近点算法

$$\begin{aligned} w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) &+ (w - w^{k+1})^T F(w^{k+1}) \\ &\ge (w - w^{k+1})^T H(w^k - w^{k+1}), \quad \forall\, w \in \Omega, \end{aligned} \tag{2.10}$$

w^{k+1} is called the proximal point of the k-th iteration for the problem (2.9).

✠ w^{k+1} **is the solution of** (2.9) **if and only if** $w^k = w^{k+1}$ ✠

(2.10) 是求解 VI (2.9) 的 PPA 算法的定义. H 可以是适当的分块矩阵, 当然 H 首先要是对称矩阵。后面将会有大量的例子说明：可以通过构造适当的正定矩阵 H, 然后求解一些小型的凸优化问题就能实现 (2.10)。

Setting $w = w^*$ in (2.10), we obtain

$$(w^{k+1} - w^*)^T H (w^k - w^{k+1}) \ge \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1}).$$

Note that (see the structure of $F(w)$ in (1.9b))

$$(w^{k+1} - w^*)^T F(w^{k+1}) = (w^{k+1} - w^*)^T F(w^*),$$

and consequently (by using (2.9)) we obtain

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

Thus, we have

$$(w^{k+1} - w^*)^T H (w^k - w^{k+1}) \ge 0. \qquad (2.11)$$

By setting $a = w^k - w^*$ and $b = w^{k+1} - w^*$,

the inequality (2.11) means that $b^T H(a-b) \geq 0$.

By using Lemma 1, we obtain

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2. \tag{2.12}$$

We get the nice convergence property of Proximal Point Algorithm.

$\|w^k - w^{k+1}\|^2 \leq \|w^{k-1} - w^k\|^2$, 即序列 $\{\|w^k - w^{k+1}\|_H\}$ 是单调不增的.

以上的预备知识. 要求读者理解（或者是先承认）优化问题拉格朗日函数的鞍点和变分不等式 (VI) 解点的等价的关系, 以及 PPA 算法的定义及收缩性质.

3 从原始-对偶混合梯度法到按需定制的邻近点算法

We consider the $\min - \max$ problem (*e. g.* 图像处理中的 ROF Model [4, 30])

$$\min_x \max_y \{\Phi(x,y) = \theta_1(x) - y^T Ax - \theta_2(y) \,|\, x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{3.1}$$

Let (x^*, y^*) be the solution of (3.1), then we have

$$\begin{cases} x^* \in \mathcal{X}, \;\; \Phi(x,y^*) - \Phi(x^*,y^*) \ge 0, \;\; \forall\, x \in \mathcal{X}, & \text{(3.2a)} \\ y^* \in \mathcal{Y}, \;\; \Phi(x^*,y^*) - \Phi(x^*,y) \ge 0, \;\; \forall\, y \in \mathcal{Y}. & \text{(3.2b)} \end{cases}$$

Using the notation of $\Phi(x,y)$, it can be written as

$$\begin{cases} x^* \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^*) + (x - x^*)^T(-A^T y^*) \ge 0, \quad \forall\, x \in \mathcal{X}, \\ y^* \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^*) \;+\; (y - y^*)^T(\;Ax^*) \;\; \ge 0, \quad \forall\, y \in \mathcal{Y}. \end{cases}$$

Furthermore, it can be written as a variational inequality in the compact form:

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

where

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

Since $F(u) = \begin{pmatrix} -A^T y \\ Ax \end{pmatrix} = \begin{pmatrix} 0 & -A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$, we have

$$(u - v)^T (F(u) - F(v)) \equiv 0.$$

3.1 求解鞍点问题的 原始-对偶混合梯度法 PDHG [33]

For given (x^k, y^k), PDHG [33] produces a pair of (x^{k+1}, y^{k+1}): First,

$$x^{k+1} = \text{argmin}\{\Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2 \,|\, x \in \mathcal{X}\}, \tag{3.4a}$$

and then we obtain y^{k+1} via

$$y^{k+1} = \text{argmax}\{\Phi(x^{k+1}, y) - \frac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}. \tag{3.4b}$$

Ignoring the constant term in the objective function, the subproblems (3.4) are reduced to

$$\begin{cases} x^{k+1} = \operatorname{argmin}\{\theta_1(x) - x^T A^T y^k + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}, & \text{(3.5a)} \\ y^{k+1} = \operatorname{argmin}\{\theta_2(y) + y^T A x^{k+1} + \frac{s}{2}\|y - y^k\|^2 \mid y \in \mathcal{Y}\}. & \text{(3.5b)} \end{cases}$$

According to Theorem 1, the optimality condition of (3.5a) is $x^{k+1} \in \mathcal{X}$ and

$$\theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T\{-A^T y^k + r(x^{k+1} - x^k)\} \ge 0, \ \forall\, x \in \mathcal{X}. \quad (3.6)$$

Similarly, from (3.5b) we get $y \in \mathcal{Y}$ and

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{A x^{k+1} + s(y^{k+1} - y^k)\} \ge 0, \ \forall\, y \in \mathcal{Y}. \quad (3.7)$$

Combining (3.6) and (3.7), we have

$$\begin{aligned} u^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) &+ (u - u^{k+1})^T F(u^{k+1}) \\ &\ge (u - u^{k+1})^T Q(u^k - u^{k+1}), \ \ \forall\, u \in \Omega. \end{aligned} \quad (3.8)$$

where $\Omega = \mathcal{X} \times \mathcal{Y}$ and

$$Q = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix} \qquad \text{is not symmetric.}$$

It does not be the PPA form (2.10), and we can not expect its convergence.

3.2 Customized Proximal Point Algorithm-Classical Version

通常, 我们把这种凑成的邻近点算法称为“按需定制的邻近点算法”.

If we change the non-symmetric matrix Q to a symmetric matrix H such that

$$Q = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix} \quad \Rightarrow \quad H = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix},$$

then the variational inequality (3.8) will become the following desirable form:

$$\theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T\{F(u^{k+1}) + H(u^{k+1} - u^k)\} \geq 0, \ \forall u \in \Omega.$$

For this purpose, we need only to change (3.7) in PDHG, namely,

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{Ax^{k+1} + s(y^{k+1} - y^k)\} \geq 0, \ \forall\, y \in \mathcal{Y}.$$

to

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{Ax^{k+1} + A(x^{k+1} - x^k) + s(y^{k+1} - y^k)\} \geq 0, \ \forall\, y \in \mathcal{Y}.$$

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{A[2x^{k+1} - x^k] + s(y^{k+1} - y^k)\} \geq 0. \quad (3.9)$$

Thus, for given (x^k, y^k), producing a proximal point (x^{k+1}, y^{k+1}) via (3.4a) and (3.9) can be summarized as:

$$x^{k+1} = \text{argmin}\Big\{\Phi(x, y^k) + \frac{r}{2}\big\|x - x^k\big\|^2 \;\Big|\; x \in \mathcal{X}\Big\}. \tag{3.10a}$$

$$y^{k+1} = \text{argmax}\Big\{\Phi\big([2x^{k+1} - x^k], y\big) - \frac{s}{2}\big\|y - y^k\big\|^2\Big\} \tag{3.10b}$$

By ignoring the constant term in the objective function, getting x^{k+1} from (3.10a) is equivalent to obtaining x^{k+1} from

$$x^{k+1} = \text{argmin}\Big\{\theta_1(x) + \frac{r}{2}\big\|x - \big[x^k + \frac{1}{r}A^T y^k\big]\big\|^2 \;\Big|\; x \in \mathcal{X}\Big\}.$$

The solution of (3.10b) is given by

$$y^{k+1} = \text{argmin}\Big\{\theta_2(y) + \frac{s}{2}\big\|y - \big[y^k + \frac{1}{s}A(2x^{k+1} - x^k)\big]\big\|^2 \;\Big|\; y \in \mathcal{Y}\Big\}.$$

According to the assumption, there is no difficulty to solve (3.10a)-(3.10b).

In the case that $rs > \|A^T A\|$, the matrix

$$H = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix} \quad \text{is positive definite.}$$

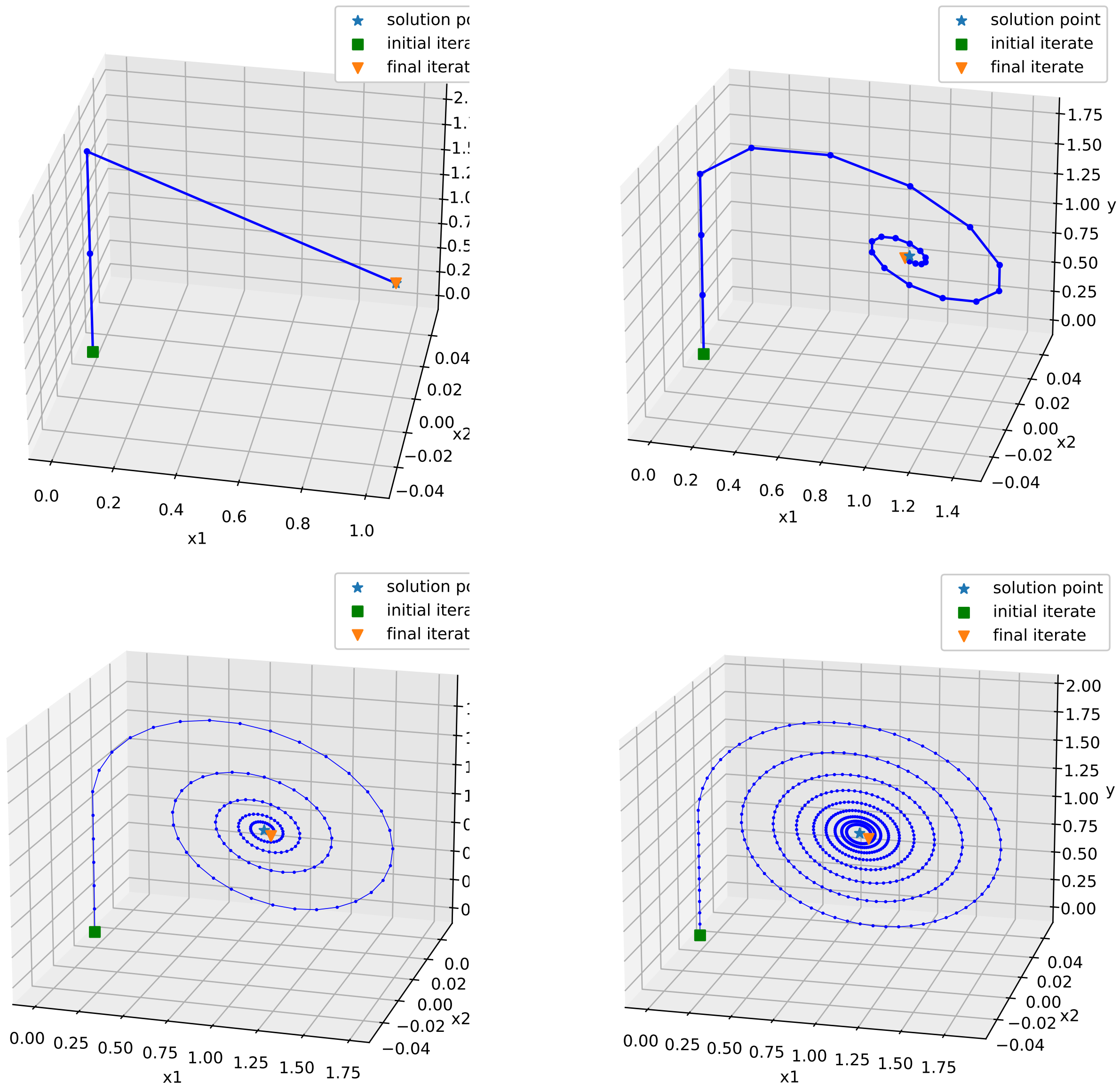
定理 2 *The sequence* $\{u^k = (x^k, y^k)\}$ *generated by the customized PPA* (3.10) *satisfies*

$$\|u^{k+1} - u^*\|_H^2 \le \|u^k - u^*\|_H^2 - \|u^k - u^{k+1}\|_H^2. \tag{3.11}$$

For the minimization problem $\min\{\theta(x) \mid Ax = b, x \in \mathcal{X}\}$,

the iterative scheme is

$$\begin{cases} x^{k+1} = \text{argmin}\Big\{\theta(x) + \frac{r}{2}\Big\|x - \big[x^k + \frac{1}{r}A^T y^k\big]\Big\|^2 \,\Big|\, x \in \mathcal{X}\Big\}. & \text{(3.12a)} \\ y^{k+1} = y^k - \frac{1}{s}\big[A(2x^{k+1} - x^k) - b\big]. & \text{(3.12b)} \end{cases}$$



对 $r = s = 1, 2, 5, 10$, C-PPA 方法都收敛. 参数越大, 步子越保守, 收敛越慢

3.3 Simplicity recognition

Frame of VI is recognized by some Researcher in Image Science

Diagonal preconditioning for first order primal-dual algorithms in convex optimization*

Thomas Pock
Institute for Computer Graphics and Vision
Graz University of Technology
pock@icg.tugraz.at

Antonin Chambolle
CMAP & CNRS
École Polytechnique
antonin.chambolle@cmap.polytechnique.fr

- T. Pock and A. Chambolle, IEEE ICCV, 1762-1769, 2011
- A. Chambolle, T. Pock, A first-order primal-dual algorithms for convex problem with applications to imaging, J. Math. Imaging Vison, 40, 120-145, 2011.

preconditioned algorithm. In very recent work [10], it has been shown that the iterates (2) can be written in form of a proximal point algorithm [14], which greatly simplifies the convergence analysis.

From the optimality conditions of the iterates (4) and the convexity of G and F^* it follows that for any $(x,y) \in X \times Y$ the iterates x^{k+1} and y^{k+1} satisfy

$$\left\langle \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}, F\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} + M \begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix} \right\rangle \geq 0 , \tag{5}$$

where

$$F\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} = \begin{pmatrix} \partial G(x^{k+1}) + K^T y^{k+1} \\ \partial F^*(y^{k+1}) - K x^{k+1} \end{pmatrix}$$

and

$$M = \begin{bmatrix} \mathrm{T}^{-1} & -K^T \\ -\theta K & \Sigma^{-1} \end{bmatrix} . \tag{6}$$

It is easy to check, that the variational inequality (5) now takes the form of a proximal point algorithm [10, 14, 16].

作者 C-P 说到我们的 PPA 解释极大地简化了收敛性分析.

我们依然认为, 只有当左边 (6) 式的矩阵 M 对称正定, 才是收敛的 PPA 方法.

否则, 就像我们前面给出的例子, 方法是不一定收敛的.

我们已经证明: 由 CP 方法演译得来的矩阵 M, 当 $\theta = 0$, 方法不能保证收敛.

对(6)式中 $\theta \in (0,1)$ 的 CP 方法, 收敛性没有证明, 还是一个 Open Problem.

[9] L. Ford and D. Fulkerson. *Flows in Networks*. Princeton University Press, Princeton, New Jersey, 1962.
[10] B. He and X. Yuan. Convergence analysis of primal-dual algorithms for total variation image restoration. Technical report, Nanjing University, China, 2010.

Later, the Reference [10] is published in SIAM J. Imaging Science [19].

Math. Program., Ser. A
DOI 10.1007/s10107-015-0957-3

CrossMark

FULL LENGTH PAPER

On the ergodic convergence rates of a first-order primal–dual algorithm

Antonin Chambolle[1] · Thomas Pock[2,3]

The paper published by Chambolle and Pock in Math. Progr. uses the VI framework

1 Introduction

In this work we revisit a first-order primal–dual algorithm which was introduced in [15, 26] and its accelerated variants which were studied in [5]. We derive new estimates for the rate of convergence. In particular, exploiting a proximal-point interpretation due to [16], we are able to give a very elementary proof of an ergodic $O(1/N)$ rate of convergence (where N is the number of iterations), which also generalizes to non-

Algorithm 1: $O(1/N)$ Non-linear primal–dual algorithm

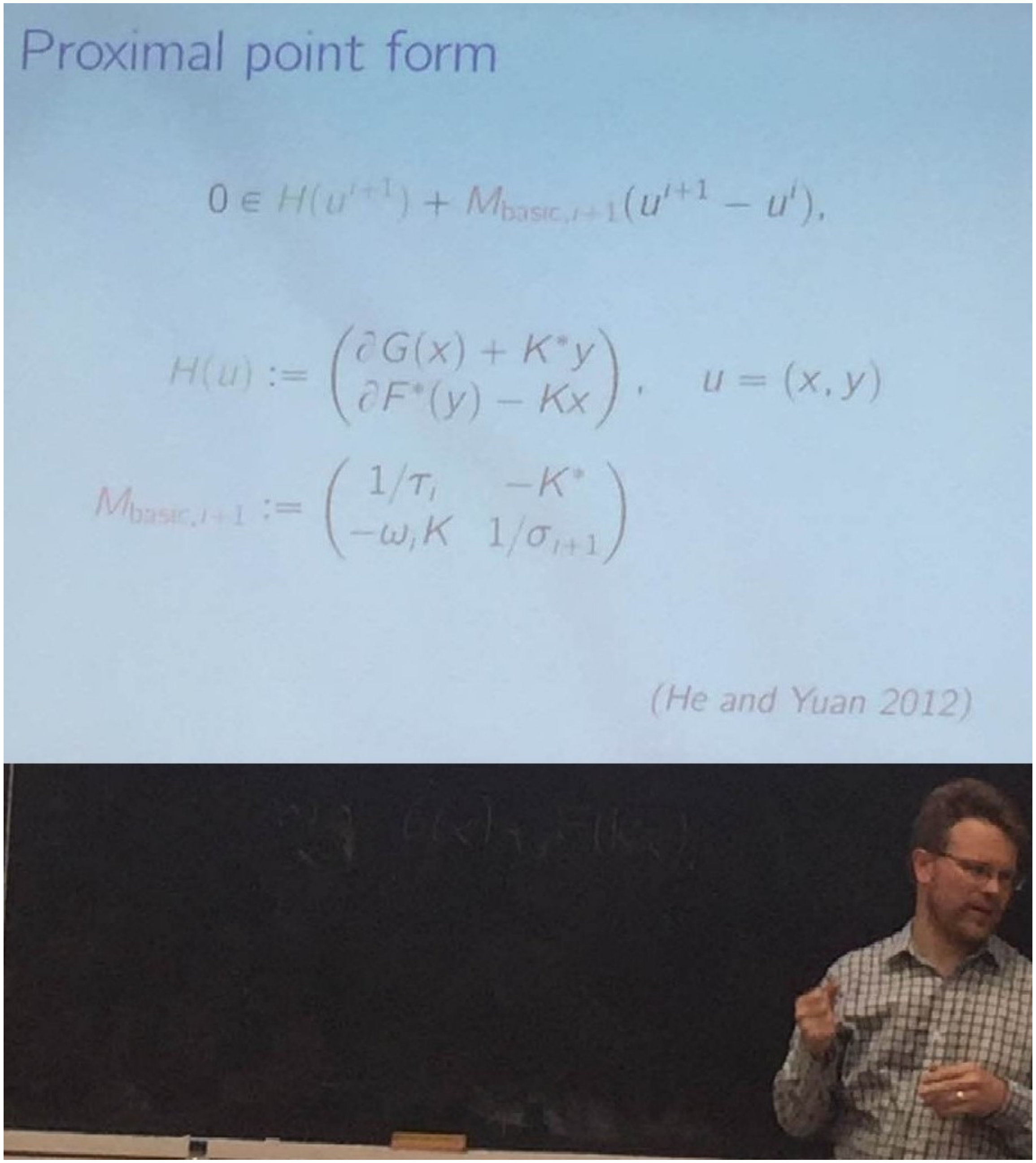
- Input: Operator norm $L := \|K\|$, Lipschitz constant L_f of ∇f, and Bregman distance functions D_x and D_y.
- Initialization: Choose $(x^0, y^0) \in \mathcal{X} \times \mathcal{Y}$, $\tau, \sigma > 0$
- Iterations: For each $n \geq 0$ let

$$(x^{n+1}, y^{n+1}) = \mathcal{PD}_{\tau,\sigma}(x^n, y^n, 2x^{n+1} - x^n, y^n) \tag{11}$$

The elegant interpretation in [16] shows that by writing the algorithm in this form

♣ 该文的文献 [16] 是我们发表在 SIAM J. Imaging Science 上的文章.

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.



2017 年 7 月, 南方科技大学数学系的一位副主任去英国访问. 在他参加的一个学术会议上, 首位报告人讲: 用 He and Yuan 提出的邻近点形式 (PPF), 处理图像问题。

见到一幅幻灯片介绍我们的工作, 我的同事抢拍了一张照片发给我。

这也说明, 只有简单的思想才容易得到传播, 被人接受。

University of Colorado **Boulder** Technical Report, Department of Applied Mathematics

The Chen-Teboulle algorithm is the proximal point algorithm

Stephen Becker *

November 22, 2011; posted August 13, 2019

Abstract

We revisit the
on the step-size p

Recent works such as [HY12] have proposed a very simple yet powerful technique for analyzing optimization methods.

1 Background

Recent works such as [HY12] have proposed a very simple yet powerful technique for analyzing optimization methods. The idea consists simply of working with a different norm in the *product* Hilbert space. We fix an inner product $\langle x, y\rangle$ on $\mathcal{H} \times \mathcal{H}^*$. Instead of defining the norm to be the induced norm, we define the primal norm as follows (and this induces the dual norm)

$$\|x\|_V = \sqrt{\langle Vx, x\rangle} = \sqrt{\langle x, x\rangle_V}, \quad \|y\|_V^* = \|y\|_{V^{-1}} = \sqrt{\langle y, V^{-1}y\rangle} = \sqrt{\langle y, y\rangle_{V^{-1}}}$$

for any Hermitian positive definite $V \in \mathcal{B}(\mathcal{H}, \mathcal{H})$; we write this condition as $V \succ 0$. For finite dimensional spaces $\mathcal{H}$, this means that V is a positive definite matrix.

4 单块的问题按需设计邻近点算法

根据预设正定矩阵 构造 PPA 算法. 许多相应的方法可以在 [12] 中查到.

The convex optimization problem,

$$\min\{\theta(x) \mid Ax = b,\, x \in \mathcal{X}\}$$

is translated to the equivalent variational inequality :

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, u \in \Omega, \tag{4.1a}$$

where

$$w = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ Ax - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{X} \times \Re^m. \tag{4.1b}$$

4.1 PPA in Primal-Dual Order

PPA for the variational inequality (4.1) : Find $w^{k+1} \in \Omega$, such that

$$\theta(x) - \theta(x^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \geq (w - w^{k+1})^T H(w^k - w^{k+1}), \ \forall w \in \Omega, \tag{4.2a}$$

where

$$H = \begin{pmatrix} \beta A^T A + \delta I_n & A^T \\ A & \frac{1}{\beta} I_m \end{pmatrix}. \tag{4.2b}$$

The concrete formula of (4.2) is

The underline part is $F(w^{k+1})$:

$$\boldsymbol{F(w) = \begin{pmatrix} -A^T\lambda \\ Ax - b \end{pmatrix}}$$

$$\begin{cases} \theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T \\ \qquad \{\underline{-A^T\tilde{\lambda}^k} + \boldsymbol{(\beta A^T A + \delta I_n)}(x^{k+1} - x^k) + \boldsymbol{A^T}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \\ \qquad (\underline{Ax^{k+1} - b}) \quad + \boldsymbol{A}(x^{k+1} - x^k) \quad + \quad \boldsymbol{(1/\beta)}(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases} \tag{4.3}$$

把 (4.3) 中的 “微小型” 变分不等式整理一下, 便是:

$$\begin{cases} \theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T\{-A^T\lambda^k + (\beta A^T A + \delta I_n)(x^{k+1} - x^k)\} \geq 0, \\ \big(A[2x^{k+1} - x^k] - b\big) \quad + \quad (1/\beta)(\lambda^{k+1} - \lambda^k) = 0. \end{cases}$$

How to implement the prediction? To get $\tilde{w}^k$ which satisfies (4.3),

we need only use the following procedure: (Primal-Dual)

$$\begin{cases} x^{k+1} = \text{Argmin}\left\{ \begin{array}{c} \theta(x) - x^T A^T \lambda^k \\ + \frac{1}{2}\beta\|A(x - x^k)\|^2 + \frac{1}{2}\delta\|x - x^k\|^2 \end{array} \middle| x \in \mathcal{X} \right\}, \\ \lambda^{k+1} = \lambda^k - \beta\big(A[2x^{k+1} - x^k] - b\big). \end{cases} \tag{4.4}$$

Then, we use the form

$$w^{k+1} := w^k - \alpha(w^k - w^{k+1}), \quad \alpha \in (0, 2)$$

to update the new iterate w^{k+1}.

在 (4.4) 的 x 子问题的目标函数中, 既有非线性函数 $\theta(x)$, 又有非平凡的二次函数, 有时会给求解带来不小的困难!

4.2 均困的 ALM (Balanced ALM) [22]

什么叫均困的ALM, 那是让(4.4)中的x子问题的目标函数只有非线性函数$\theta(x)$和平凡的二次函数$\frac{r}{2}\|x-x^k\|^2$. 把部分困难转移到变量λ的校正.

PPA for the variational inequality (4.1) : Find $w^{k+1} \in \Omega$, such that

$$\theta(x)-\theta(x^{k+1})+(w-w^{k+1})^T F(w^{k+1}) \geq (w-w^{k+1})^T H(w^k-w^{k+1}), \tag{4.5a}$$

for all $w \in \Omega$, where

$$H=\begin{pmatrix} rI_n & A^T \\ A & \frac{1}{r}AA^T+\delta I_m \end{pmatrix} \quad \text{is positive definite.} \tag{4.5b}$$

Then, we use the form

$$w^{k+1}=w^k-\alpha(w^k-w^{k+1}), \quad \alpha \in (0,2)$$

to update the new iterate w^{k+1}.

The underline part is $F(w^{k+1})$:

$$\boldsymbol{F(w) = \begin{pmatrix} -A^T\lambda \\ Ax - b \end{pmatrix}}$$

The concrete form of (4.5) is

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

It can written as

$$\begin{cases} x^{k+1} \in \mathcal{X}, \quad \theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T\{-A^T\lambda^k + r(x^{k+1} - x^k)\} \geq 0, \\ \qquad A[(2x^{k+1} - x^k) - b] + (\frac{1}{r}AA^T + \delta I_m)(\lambda^{k+1} - \lambda^k) = 0. \end{cases}$$

Thus, the w^{k+1} in balanced ALM (4.5) is implemented by

$$\begin{cases} x^{k+1} = \arg\min\{\theta(x) - x^TA^T\lambda^k + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}, & \text{(4.6a)} \\ \lambda^{k+1} = \arg\min\Big\{\lambda^T\big(A[2x^{k+1} - x^k] - b\big) + \frac{1}{2}\big\|\lambda - \lambda^k\big\|^2_{(\frac{1}{r}AA^T + \delta I_m)}\Big\}. & \text{(4.6b)} \end{cases}$$

Remark. λ^{k+1} in (4.6b) is the solution of the following system of linear equations:

$$(\frac{1}{r}AA^T + \delta I_m)(\lambda - \lambda^k) + \big(A[2x^{k+1} - x^k] - b\big) = 0. \tag{4.7}$$

Because the matrix

$$H_0 = (\frac{1}{r}AA^T + \delta I_m)$$

is positive definite, there are efficient algorithms in literature for solving such a systems of linear equations.

- 均困的增广拉格朗日乘子法, x-子问题 (4.6a) 中的二次项式平凡的, 降低了问题求解的难度.
- λ-子问题 (4.6b) 要求解一个系数矩阵正定的线性方程组. 注意到, 在整个迭代过程中, 我们只要对矩阵 H_0 做一次 Cholesky 分解.

$$H_0 = LL^T, \qquad LL^T(\lambda - \lambda^k) = b - A[2x^{k+1} - x^k].$$

5 求解两可分离块的 PPA 算法

两块可分离凸优化问题

$$\min\{\theta_1(x)+\theta_2(y)|Ax+By=b, x\in\mathcal{X}, y\in\mathcal{Y}\} \tag{5.1}$$

转换成变分不等式

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

其中

$$w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix},\quad u=\begin{pmatrix} x\\ y\end{pmatrix},\quad \theta(u)=\theta_1(x)+\theta_2(y),$$

$$F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ Ax+By-b\end{pmatrix},\qquad \text{和} \qquad \Omega=\mathcal{X}\times\mathcal{Y}\times\Re^m.$$

问题 (5.1) 的增广拉格朗日函数是

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

ADMM 的 k 次迭代从给定的 $v^k = (y^k, \lambda^k)$ 开始, 通过

$$\left\{\begin{array}{lcl} x^{k+1} & \in & \arg\min\left\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k - b\|^2 \;\middle|\; x \in \mathcal{X}\right\}, \\ y^{k+1} & \in & \arg\min\left\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}\beta\|Ax^{k+1} + By - b\|^2 \;\middle|\; y \in \mathcal{Y}\right\}, \\ \lambda^{k+1} & = & \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b) \end{array}\right. \tag{5.2}$$

求得 $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$. 由于 ADMM 方法的 k 步迭代, x^{k+1} 是根据给定的 (y^k, λ^k) 算出来的, 我们称 x 为中间变量, 称 $v = (y, \lambda)$ 为核心变量.

5.1 平行求解子问题的方法

求解两个可分离块问题 (1.10) 相应的变分不等式 (1.11)-(1.12).
根据 PPA 算法的要求, 设计的右端矩阵为对称正定.

$$\theta(u)-\theta(\tilde{u}^k)+(w-w^k)^T F(w^{k+1}) \ge (w-w^{k+1})^T H(w^k-w^{k+1}), \;\; \forall w \in \Omega, \tag{5.3a}$$

where

$$H = \begin{pmatrix} \beta A^TA + \delta I_{n_1} & 0 & A^T \\ 0 & \beta B^TB + \delta I_{n_2} & B^T \\ A & B & \frac{2}{\beta} I_m \end{pmatrix}. \tag{5.3b}$$

The both matrices

$$\begin{pmatrix} \beta A^TA + \delta I_{n_1} & A^T \\ A & \frac{1}{\beta} I_m \end{pmatrix} \succ 0, \qquad \begin{pmatrix} \beta B^TB + \delta I_{n_2} & B^T \\ B & \frac{1}{\beta} I_m \end{pmatrix} \succ 0.$$

The concrete form of (5.3) is

$$\left\{\begin{array}{l} \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \\ \qquad\qquad \{\underline{-A^T\lambda^{k+1}} + (\boldsymbol{\beta A^T A + \delta I_{n_1}})(x^{k+1} - x^k) \boldsymbol{+A^T}(\lambda^{k+1} - \lambda^k)\} \ge 0, \\ \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \\ \qquad\qquad \{\underline{-B^T\lambda^{k+1}} + (\boldsymbol{\beta B^T B + \delta I_{n_2}})(y^{k+1} - y^k) \boldsymbol{+B^T}(\lambda^{k+1} - \lambda^k)\} \ge 0, \\ (\underline{Ax^{k+1} + By^{k+1} - b}) \boldsymbol{+A}(x^{k+1} - x^k) \boldsymbol{+B}(y^{k+1} - y^k) + \boldsymbol{(2/\beta)}\,(\lambda^{k+1} - \lambda^k) = 0. \end{array}\right.$$

After simple organization, we obtain

$$\left\{\begin{array}{l} \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T\{-A^T\lambda^k + (\beta A^T A + \delta I_{n_1})(x^{k+1} - x^k)\} \ge 0, \\ \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{-B^T\lambda^k + (\beta B^T B + \delta I_{n_2})(y^{k+1} - y^k)\} \ge 0, \\ [2(Ax^{k+1} + By^{k+1} - b) - (Ax^k + By^k - b)] + (2/\beta)(\lambda^{k+1} - \lambda^k) = 0. \end{array}\right.$$

In fact, the prediction can be arranged by

$$\begin{cases} x^{k+1} = \arg\min\left\{ \begin{array}{c} \theta_1(x) - x^T A^T \lambda^k \\ +\frac{1}{2}\beta\|A(x-x^k)\|^2 + \frac{1}{2}\delta\|x-x^k\|^2 \end{array} \middle| x \in \mathcal{X} \right\} & (5.4a) \\ y^{k+1} = \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T \lambda^k \\ +\frac{1}{2}\beta\|B(y-y^k)\|^2 + \frac{1}{2}\delta\|y-y^k\|^2 \end{array} \middle| y \in \mathcal{Y} \right\} & (5.4b) \\ \lambda^{k+1} = \lambda^k - \frac{1}{2}\beta\big[2(Ax^{k+1} + By^{k+1} - b) - (Ax^k + By^k - b)\big] & (5.4c) \end{cases}$$

$$\begin{cases} x^{k+1} = \arg\min\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}(x-x^k)^T(\beta A^T A + \delta I_{n_1})(x-x^k) | x \in \mathcal{X}\} \\ y^{k+1} = \arg\min\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}(y-y^k)^T(\beta B^T B + \delta I_{n_2})(y-y^k) | y \in \mathcal{Y}\} \\ \lambda^{k+1} = \lambda^k - \frac{1}{2}\beta\big[2(Ax^{k+1} + By^{k+1} - b) - (Ax^k + By^k - b)\big] \end{cases}$$

利用变分不等式 (VI) 和邻近点算法 (PPA), 更自由地设计 ADMM 类分裂收缩算法

5.2 均困的 PPA 算法

求解两个可分离块问题 (1.10) 相应的变分不等式 (1.11)-(1.12).
假设 ADMM 中 y-子问题求解比较简单，而 x-子问题必须简化.

Primal-Dual Order

$$\theta(u)-\theta(u^{k+1})+(w-w^{k+1})^T F(w^{k+1}) \ge (w-w^{k+1})^T H(w^k-w^{k+1}), \ \forall w \in \Omega, \tag{5.5a}$$

where

$$H = \begin{pmatrix} rI_{n_1} & 0 & A^T \\ 0 & \beta B^T B + \delta I_{n_2} & B^T \\ A & B & (\frac{1}{\beta}+\delta)I_m + \frac{1}{r}AA^T \end{pmatrix}. \tag{5.5b}$$

The both matrices

$$\begin{pmatrix} rI_{n_1} & A^T \\ A & \delta I_m + \frac{1}{r}AA^T \end{pmatrix} \succ 0, \qquad \begin{pmatrix} \beta B^T B + \delta I_{n_2} & B^T \\ B & \frac{1}{\beta}I_m \end{pmatrix} \succ 0.$$

The concrete form of (5.5) is

$$\begin{cases} \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \\ \qquad \{\underline{-A^T\lambda^{k+1}} + \boldsymbol{r}(x^{k+1} - x^k) \boldsymbol{+ A^T}(\lambda^{k+1} - \lambda^k)\} \geq 0, \\ \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \\ \qquad \{\underline{-B^T\lambda^{k+1}} + (\boldsymbol{\beta B^T B + \delta I_{n_2}})(y^{k+1} - y^k) \boldsymbol{+ B^T}(\lambda^{k+1} - \lambda^k)\} \geq 0, \\ (\underline{Ax^{k+1} + By^{k+1} - b}) \boldsymbol{+ A}(x^{k+1} - x^k) \boldsymbol{+ B}(y^{k+1} - y^k) \\ \qquad + \Big((\boldsymbol{\tfrac{1}{\beta} + \delta) I_m + \tfrac{1}{r} A A^T} \Big) (\lambda^{k+1} - \lambda^k) = 0. \end{cases}$$

After simple organization, we obtain

$$\begin{cases} \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T\{-A^T\lambda^k + r(x^{k+1} - x^k)\} \geq 0, \\ \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{-B^T\lambda^k + (\beta B^T B + \delta_{n_2}(y^{k+1} - y^k)\} \geq 0, \\ [2(Ax^{k+1} + By^{k+1} - b) - (Ax^k + By^k - b)] \\ \qquad + \big((\tfrac{1}{\beta} + \delta) I_m + \tfrac{1}{r} A A^T\big)(\lambda^{k+1} - \lambda^k) = 0. \end{cases}$$

In fact, the prediction can be arranged by

$$\begin{cases} x^{k+1} = \arg\min\Big\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2} r \|x - x^k\|^2 \,\Big|\, x \in \mathcal{X}\Big\} \\ y^{k+1} = \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T \lambda^k \\ + \frac{1}{2}\beta \|B(y - y^k)\|^2 + \frac{1}{2}\delta \|y - y^k\|^2 \end{array} \,\middle|\, y \in \mathcal{Y} \right\} \\ \left(\begin{array}{c} \text{通过求解} \\ \text{线性方程组} \end{array}\right) \quad \begin{array}{l} \Big((\frac{1}{\beta} + \delta) I_m + \frac{1}{r} A A^T\Big)(\lambda - \lambda^k) \\ \quad = \big[2(Ax^{k+1} + By^{k+1} - b) - (Ax^k + By^k - b)\big] \end{array} \\ \qquad \text{求得 } \lambda^{k+1}. \end{cases}$$

经典的变分不等式的PPA算法, 都可以做如下延伸:

$$w^{k+1} := w^k - \alpha(w^k - w^{k+1}), \qquad \alpha \in (0, 2).$$

References

[1] S. Becker, The Chen-Teboulle algorithm is the proximal point algorithm, manuscript, 2011, arXiv: 1908.03633[math.OC].

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, Foundations and Trends in Machine Learning Vol. 3, No. 1 (2010) 1 – 122.

[3] 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.

[4] A. Chambolle, T. Pock, A first-order primal-dual algorithms for convex problem with applications to imaging, J. Math. Imaging Vison, 40, 120-145, 2011.

[5] A. Chambolle and T Pock, On the ergodic convergence rates of a first-order primal – dual algorithm， Math. Program., A 159 (2016) 253-287.

[6] R. Fletcher, Practical mathods of Optimization, Second Edition, JOHN WILEY & SONS, 1987.

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

[8] T. Goldstein, B. O'Donoghue, S. Setzer and R. Baraniuk, Fast Alternating Direction Optimization Methods, SIAM J. Imaging Science, Vol. 7, No. 3, pp. 1588 – 1623, 2014.

[9] 何炳生, 从变分不等式的投影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 38 (2016), 74-96.

[10] 何炳生, 我和乘子交替方向法20年, 运筹学学报 22 (2018), 1-31.

[11] B. S. He, 何炳生, 凸优化和单调变分不等式收缩算法的统一框架, 中国科学 48 (2018), 255-272, doi: 10.1360/N012017-00034

[12] 何炳生, 利用统一框架设计凸优化的分裂收缩算法, 高等学校计算数学学报, 2022, 44: 1-35.

[13] B. S. He, H. Liu, Z. R. Wang and X. M. Yuan, A strictly Peaceman-Rachford splitting method for convex programming, SIAM J. Optim. 24 (2014), 1011-1040.

[14] 何炳生, 申远, 求解凸规划及鞍点问题定制的PPA 算法及其收敛速率, 中国科学 2012, 42(5): 515 – 525, doi: 10.1360/012011-1049

[15] 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.

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

[17] B. S. He, S. J. Xu and X. M. Yuan, Extensions of ADMM for separable convex optimization problems with linear equality or inequality constraints, arXiv:2107.01897v2[math.OC].

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

[19] 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 (2012) 119-149.

[20] 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.

[21] B. S. He and X. M. Yuan, A class of ADMM-based algorithms for three-block separable convex programming, Comput. Optim. Appl. 70 (2018), 791-826.

[22] B. S. He and X. M. Yuan, Balanced Augmented Lagrangian Method for Convex Programming, arXiv 2108.08554 [math.OC]

[23] B. S. He and X. M. Yuan, On construction of splitting contraction algorithms in a prediction- correction framework for separable convex optimization, arXiv 2204.11522 [math.OC]

[24] B.S. He, X.M. Yuan and W.X. Zhang, A customized proximal point algorithm for convex minimization with linear constraints, Comput. Optim. Appl., 56(2013), 559-572.

[25] B.S. He and J. Zhou A Modified Alternating Direction Methodfor Convex Minimization Problems, Applied Mathematics Letters 13 (2000) 123-130.

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

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

[28] 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.

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

[30] L. Rudin, S. Osher, and E. Fatemi, Nonlinear total variation based noise removal algorithms, Phys. D, 60 (1992), pp. 227 – 238

[31] Shengjie Xu, A dual-primal balanced augmented Lagrangian method for linearly constrained convex programming, manusript, 2021. arXiv:2109.02106 [math.OC]

[32] X. Q. Zhang, M. Burger and S. Osher, A unified primal-dual algorithm framework based on Bregman iteration, J. Sci. Comput., 2010, 46: 20 – 46.

[33] M. Zhu and T. F. Chan, An Efficient Primal-Dual Hybrid Gradient Algorithm for Total Variation Image Restoration, CAM Report 08-34, UCLA, Los Angeles, CA, 2008.