从变分不等式的邻近点算法到广义邻近点算法

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

曲阜师范大学　2023年11月20日

华罗庚先生普及“双法”对我们的启示

- 华罗庚先生当年普及的双法— 统筹法和优选法。　　普及双法以优选法为主。

- 要“牢记把方法交给群众”。

  —华罗庚《数学工作者要大力为农业服务》

  人民日报 1960 年 10 月 30 日

- 这成为从上世纪60年代开始的近20年间，华罗庚从事数学普及工作的指导思想。

  — 王元《华罗庚》

- 随着全民族文化水平的提高，群众有了新的定义. 提供工程师们容易掌握的方法，可以 作为部分优化学者的工作目标.

能够交给“群众”的方法，应该是普通大学生能够理解，掌握的方法.

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

Department of Mathematics, Nanjing University,
Nanjing, 210093, China E-mail: hebma@nju.edu.cn

Current Research Areas:
Mathematical Programming, Numerical Optimization,
Variational Inequalities, Projection and contraction methods for VI,
ADMM-like splitting contraction methods for convex optimization

Education:
PhD: Applied Mathematics, The University of Wuerzburg, Germany, 1986
Thesis Advisor: Professor Dr. Josef Stoer
BSc: Computational Mathematics, Nanjing University, 1981

Work:
2015-2020 Professor, Dept. of Math, Southern University of Science and Technology (SUSTech)
2013-2015 Professor, School of Management Science and Engineering, Nanjing University
1997~2013 Professor, Department of Mathematics, Nanjing University
1992~1997 Associate Professor, Department of Mathematics,Nanjing University

我会尽量维护自己的主页, 不断修正、更新自己的学术体验.

My Thinkings: 1. 关门感想 2. 说说我的主要研究兴趣 — 兼谈华罗庚推广优选法对我的影响

3. 说说我的主要研究兴趣(续) --- 我们在ADMM类方法的主要工作 4. 古稀回首

5. 两页纸给出ADMM 收敛性证明 6. 两页纸简述我职业生涯中的主要研究工作

7. 两页纸描述古稀之后的研究进展: 从好不容易凑出一个方法到并不费力构造一簇算法

8. 两页PPT讲述求解线性约束凸优化问题预测-校正的广义邻近点(Generaized PPA)算法

《数学文化》2020 年/ 第 11 卷第 2 期刊登的我的自述: 四十年上下求索 一份珍贵的回忆材料

My Talks: 比较系统的知识建议阅读第3个报告. 也建议阅读最近的一些系列报告

For more systematic knowledge, it is recommended to read Talk 3, which is written in English.

19. 2023 年10月在天元数学东北中心八次课程的汇总讲义 前言与目录 I II III IV V VI VII VIII

18. 2023年1月在华南师大《华人数学家论坛》的报告 — 凸优化分裂收缩算法统一框架的最新进展

17. 2022年11月在西电《最优化前沿论坛》的报告 — 从好不容易凑出一个算法到并不费劲构造一簇算法

16. 2022 年7月南理工《计算机》方向暑期班六讲摘要 和六讲的PPT I II III IV V VI

15. 利用预测-校正统一框架构造凸优化的分裂收缩算法(由预测矩阵构造校正矩阵) (ArXiv: 2204.11522)

14. 2022 年元月南师大数科院系列报告B站视频辅助材料 A B C D E F G H I J K L

13. ADMM 类分裂收缩算法的一些最新进展 统一框架下Balanced-ALM 便于向多块推广的ADMM

12. 均困平衡的增广拉格朗日乘子法 — Balanced ALM (一类新的增广拉格朗日乘子法ArXiv: 2108.08554)

11. 一类便于向求解多块问题推广并能处理不等式约束问题的交替方向法 (ArXiv:2107.01897)

10. 瞎子爬山-步步为营—凸优化算法中的变分不等式和邻近点策 略(南京大学数学系本科生论坛上的报告)

9. 被S. Becker 誉为 Very Simple yet Powerful 的 Technique — 应用及新的进展

8. 线性化ALM 线性化ADMM 以及处理三个可分离块问题中缩小有关参数至3/4提高效率的方法

我的报告的 PDF 文件, 一般都可以在我的主页上查到.

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:

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

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

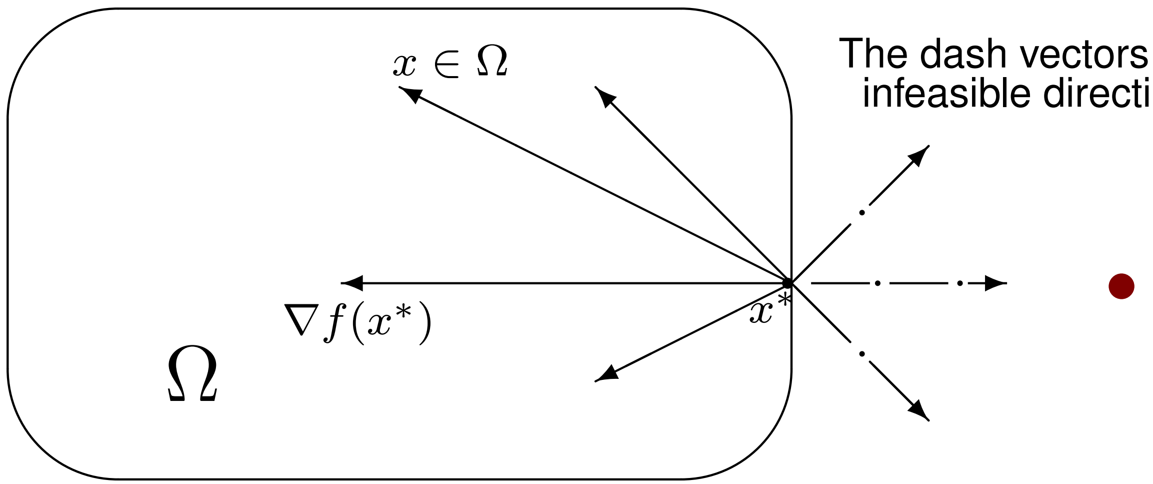


Fig. 1.1 Differential Convex Optimization and VI

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

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

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

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

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

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　　　凸优化最优性条件定理

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

定理 1 把优化问题 (1.3a) 转换成了变分不等式 (1.3b).

1.2 Linear constrained convex optimization and VI

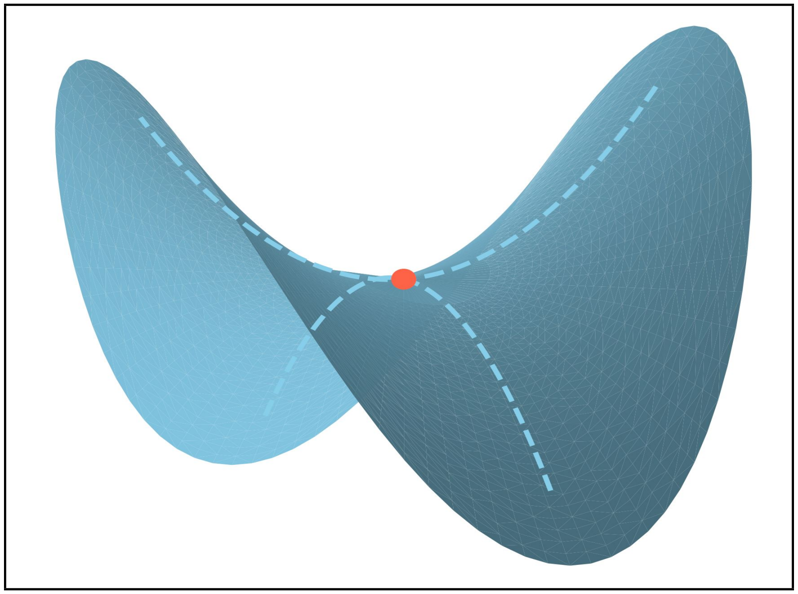
We consider the linearly constrained convex optimization problem

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

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

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

The above inequalities can be written as

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

According to the definition of $L(u,\lambda)$ (see(1.5)),

$$\begin{aligned} & L(u,\lambda^*) - L(u^*,\lambda^*) \\ = \quad & [\theta(u) - (\lambda^*)^T(\mathcal{A}u - b)] - [\theta(u^*) - (\lambda^*)^T(\mathcal{A}u^* - b)] \\ = \quad & \theta(u) - \theta(u^*) + (u - u^*)^T(-\mathcal{A}^T\lambda^*) \end{aligned}$$

it follows from (1.6a) that

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

Similarly, for (1.6b), since

$$\begin{array}{rl} & L(u^*,\lambda^*) - L(u^*,\lambda) \\ = & [\theta(u^*) - (\lambda^*)^T(\mathcal{A}u^* - b)] - [\theta(u^*) - (\lambda)^T(\mathcal{A}u^* - b)] \\ = & (\lambda - \lambda^*)^T(\mathcal{A}u^* - b), \end{array}$$

thus we have

$$\lambda^* \in \Re^m, \quad (\lambda - \lambda^*)^T(\mathcal{A}u^* - b) \ge 0, \quad \forall\, \lambda \in \Re^m. \tag{1.8}$$

Notice that the expression (1.8) (the inner product of the vector $(\mathcal{A}u^* - b)$ with any vector is nonnegative) is equivalent to

$$\mathcal{A}u^* - b = 0.$$

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

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

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

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

where

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

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

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

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

线性约束的凸优化问题, 都转换成了变分不等式. 问题归结为求一个鞍点.

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 Preliminaries of PPA for Variational Inequalities

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

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w-w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{2.2}$$

混合变分不等式—简称变分不等式

PPA for VI (2.2) in H-norm (定义)

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

邻近点算法

$$w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \\ \geq (w - w^{k+1})^T H(w^k - w^{k+1}), \quad \forall\, w \in \Omega, \qquad (2.3)$$

w^{k+1} is called the proximal point of the k-th iteration for the problem (2.2).

(2.3) 是求解 VI (2.2) 的 PPA 算法的定义. 后面会用例子说明这是容易做到的.

✠ w^{k+1} **is the solution of** (2.2) **if and only if** $w^k = w^{k+1}$ ✠

Setting $w = w^*$ in (2.3), we obtain

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

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 (2.2)) 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

$$(w^{k+1} - w^*)^T H(w^k - w^{k+1}) \ge 0. \tag{2.4}$$

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

the inequality (2.4) means that $\boldsymbol{b^T H(a - b) \ge 0}$.

By using Lemma 1, we obtain

$$\|\boldsymbol{w^{k+1} - w^*}\|_{\boldsymbol{H}}^{\boldsymbol{2}} \le \|\boldsymbol{w^k - w^*}\|_{\boldsymbol{H}}^{\boldsymbol{2}} - \|\boldsymbol{w^k - w^{k+1}}\|_{\boldsymbol{H}}^{\boldsymbol{2}}. \tag{2.5}$$

We get the nice convergence property of Proximal Point Algorithm.

2.2 Variants of PPA for Variational Inequalities

Let v be a sub-vector of w. The k-th iteration begins with given v^k . v 核心变量

PPA for VI (2.2) in H-norm For given v^k and $H \succ 0$, find w^{k+1},

$$w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \\ \geq (v - v^{k+1})^T H(v^k - v^{k+1}), \quad \forall\, w \in \Omega, \qquad (2.6)$$

w^{k+1} is called the proximal point of the k-th iteration for the problem (2.2).

w^{k+1} is the solution of (2.2) if and only if $v^k = v^{k+1}$

In this case, v is called the essential variables of w. In addition, we define

$$\mathcal{V}^* = \{v^* \text{ is a subvector of } w^* \mid w^* \in \Omega^*\}.$$

Setting $w = w^*$ in (2.6), we obtain

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

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 (2.2)) we obtain

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

Thus, we have

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \geq 0. \tag{2.7}$$

By using Lemma 1, we obtain

$$\|\boldsymbol{v}^{\boldsymbol{k+1}} - \boldsymbol{v}^*\|_{\boldsymbol{H}}^{\boldsymbol{2}} \leq \|\boldsymbol{v}^{\boldsymbol{k}} - \boldsymbol{v}^*\|_{\boldsymbol{H}}^{\boldsymbol{2}} - \|\boldsymbol{v}^{\boldsymbol{k}} - \boldsymbol{v}^{\boldsymbol{k+1}}\|_{\boldsymbol{H}}^{\boldsymbol{2}}. \tag{2.8}$$

We get the nice convergence property of Proximal Point Algorithm.

The residue sequence $\{\|v^k - v^{k+1}\|_H\}$ is also monotonically no-increasing.

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

2.3 The relaxed PPA （延伸的邻近点算法）

We shall maintain our focus on the monotone variational inequality (2.2), namely,

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

The PPA form (2.6) reads as

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

Set the output of the above VI as $\tilde{w}^k$, we have

$$\begin{aligned} \tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) & \\ \geq (v - \tilde{v}^k)^T H(v^k - \tilde{v}^k), \;\; \forall\, w \in \Omega. & \end{aligned} \tag{2.1}$$

Setting $w = w^*$ in (2.1), we obtain

$$(\tilde{v}^k - v^*)^T H(v^k - \tilde{v}^k) \geq \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k). \tag{2.2}$$

Applying (see (1.9b)) the identity

$$(\tilde{w}^k - w^*)^T F(\tilde{w}^k) \equiv (\tilde{w}^k - w^*)^T F(w^*)$$

to (2.2), we obtain

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

Because RHS of the above inequality is , we have

$$(\tilde{v}^k - v^*)^T H(v^k - \tilde{v}^k) \ge 0.$$

We write it as

$$\{(v^k - v^*) - (v^k - \tilde{v}^k)\}^T H(v^k - \tilde{v}^k) \ge 0$$

and thus

$$(v^k - v^*)^T H(v^k - \tilde{v}^k) \ge \|v^k - \tilde{v}^k\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{2.3}$$

The inequality (2.3) means that $(v^k - \tilde{v}^k)$ is the ascent direction of the unknown distance function $\frac{1}{2}\|v - v^*\|_H^2$ at the point v^k.

$$\Big\langle \nabla\big(\tfrac{1}{2}\|v - v^*\|_H^2\big)\Big|_{v=v^k},\ (v^k - \tilde{v}^k)\Big\rangle \geq \|v^k - \tilde{v}^k\|_H^2, \quad \forall\, v^* \in \mathcal{V}^*.$$

The task of the algorithm is to produce a decreasing sequence $\{\|v^k - v^*\|_H^2\}$.
Set

$$v^{k+1}(\alpha) = v^k - \alpha(v^k - \tilde{v}^k) \tag{2.4}$$

which is an α dependent new iterate. It is clear we want to maximize

$$\vartheta(\alpha) = \|v^k - v^*\|_H^2 - \|v^{k+1}(\alpha) - v^*\|_H^2. \tag{2.5}$$

Note that

$$\begin{aligned}\vartheta(\alpha) &= \|v^k - v^*\|_H^2 - \|(v^k - v^*) - \alpha(v^k - \tilde{v}^k)\|_H^2 \\ &= 2\alpha(v^k - v^*)^T H(v^k - \tilde{v}^k) - \alpha^2\|v^k - \tilde{v}^k\|_H^2 \qquad (2.6)\end{aligned}$$

is a quadratic function of α.

We can not directly maximize $\vartheta(\alpha)$ in (2.6) because the coefficient of the linear term $2(v^k - v^*)^T H(v^k - \tilde{v}^k)$ contains the unknown solution v^*.

Using (2.3), from (2.6) we get

$$\vartheta(\alpha) \geq 2\alpha\|v^k - \tilde{v}^k\|_H^2 - \alpha^2\|v^k - \tilde{v}^k\|_H^2 \tag{2.7}$$

Set

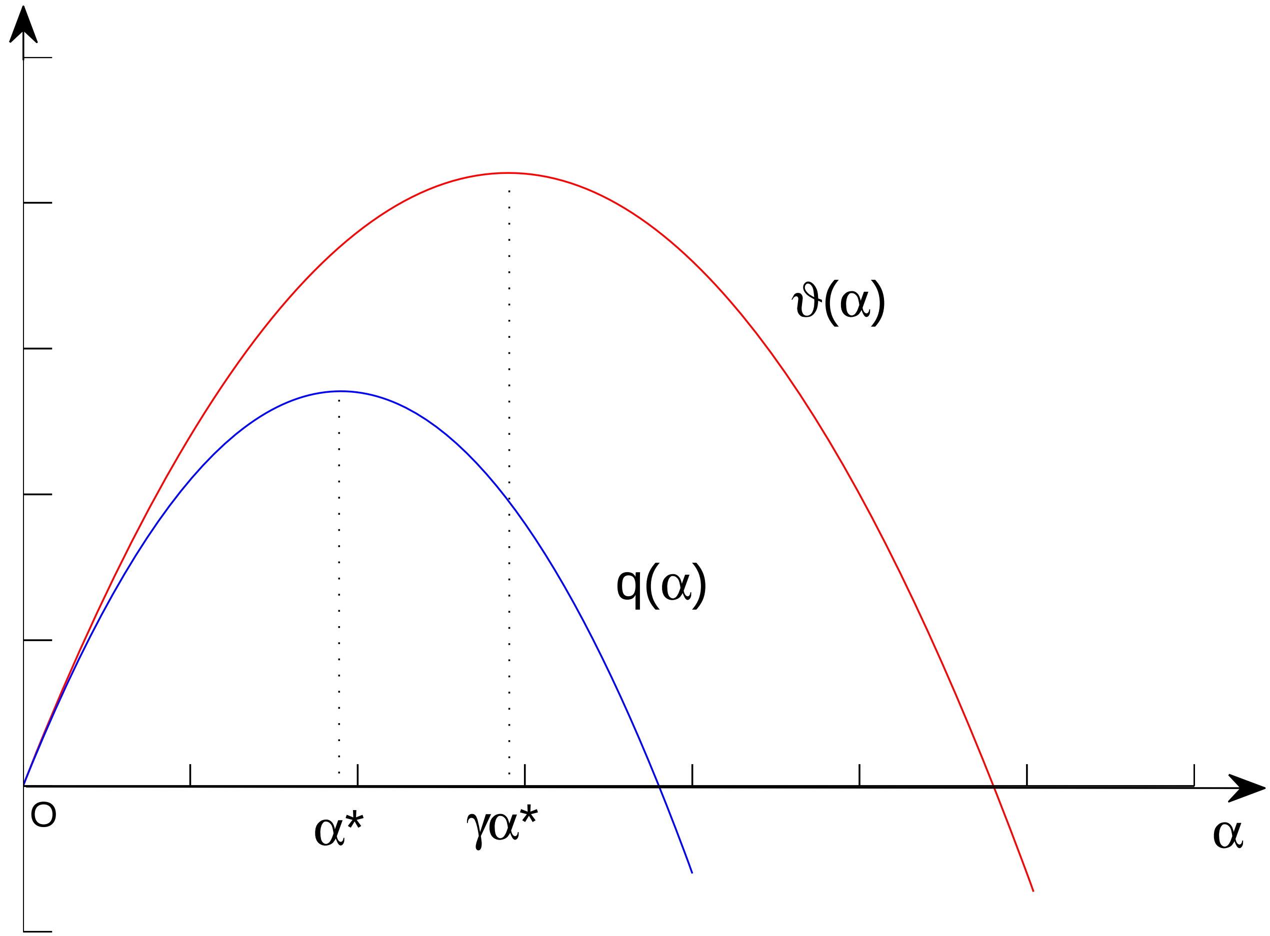
$$q(\alpha) = (2\alpha - \alpha^2)\|v^k - \tilde{v}^k\|_H^2, \tag{2.8}$$

which is a quadratic lower-bound function of $\vartheta(\alpha)$. The quadratic function $q(\alpha)$ reaches its maximum at $\alpha^* \equiv 1$.

$$v^{k+1} = v^k - \gamma(v^k - \tilde{v}^k), \quad \gamma \in (0, 2) \tag{2.9}$$

The generated sequence $\{v^k\}$ satisfies

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \gamma(2-\gamma)\|v^k - \tilde{v}^k\|_H^2. \tag{2.10}$$



取 $\gamma \in [1, 2)$ 的示意图

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 Augmented Lagrangian Method (ALM)

We consider the convex optimization, namely

$$\min\{\theta(u) \,|\, \mathcal{A}u = b,\ u \in \mathcal{U}\}. \tag{3.1}$$

The related variational inequality of the saddle point of the Lagrangian function is

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{3.2a}$$

where

$$w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -\mathcal{A}^T\lambda \\ \mathcal{A}u - b \end{pmatrix} \text{ and } \quad \Omega = \mathcal{U} \times \Re^m. \tag{3.2b}$$

Augmented Lagrangian Method

The augmented Lagrangian function of the problem (3.1) is

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

The k-th iteration of the **Augmented Lagrangian Method** [10, 12] begins with a given λ^k, obtain $w^{k+1} = (u^{k+1}, \lambda^{k+1})$ via

$$\text{(ALM)} \quad \begin{cases} u^{k+1} = \arg\min\{\mathcal{L}_\beta(u, \lambda^k) \mid u \in \mathcal{U}\}, & \text{(3.3a)} \\ \lambda^{k+1} = \lambda^k - \beta(\mathcal{A}u^{k+1} - b). & \text{(3.3b)} \end{cases}$$

In (3.3), u^{k+1} is only a computational result of (3.3a) from given λ^k, it is called the intermediate variable. In order to start the k-th iteration of ALM, we need only to have λ^k and thus we call it as the essential variable.

The subproblem (3.3a) is a problem of mathematical form

$$\min\{\theta(u) + \tfrac{\beta}{2}\|\mathcal{A}u - p^k\|^2 \mid u \in \mathcal{U}\} \tag{3.4}$$

where $\beta > 0$ is a given scalar and $p^k = b + \frac{1}{\beta}\lambda^k$.

Assumption: The solution of problem (3.4) has closed-form solution or can be efficiently computed with a high precision.

Changing the constant term in the objective function does not affect the solution of the optimization problem. Thus,

$$\begin{aligned} u^{k+1} &\in \operatorname{argmin}\{\mathcal{L}_\beta(u,\lambda^k) \mid u \in \mathcal{U}\} \\ &= \operatorname{argmin}\{\theta(u) - (\lambda^k)^T \mathcal{A}u + \tfrac{\beta}{2}\|\mathcal{A}u - b\|^2 \mid u \in \mathcal{U}\} \\ &= \operatorname{argmin}\{\theta(u) + \tfrac{\beta}{2}\|(\mathcal{A}u - b) - \tfrac{1}{\beta}\lambda^k\|^2 \mid u \in \mathcal{U}\} \end{aligned}$$

According to Lemma 1, the optimal condition of (3.3a) is $u^{k+1} \in \mathcal{U}$ and

$$\theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T\{-\mathcal{A}^T\lambda^k + \beta\mathcal{A}^T(\mathcal{A}u^{k+1} - b)\} \geq 0, \ \forall u \in \mathcal{U}.$$

Because $\lambda^k - \beta(\mathcal{A}u^{k+1} - b) = \lambda^{k+1}$, the above VI can be written as

$$u^{k+1} \in \mathcal{U}, \quad \theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T\{-\mathcal{A}^T\lambda^{k+1}\} \geq 0, \ \forall u \in \mathcal{U}. \quad (3.5)$$

The update form (3.3b) is

$$(\mathcal{A}u^{k+1} - b) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k) = 0.$$

and it is equivalent to

$$(\lambda - \lambda^{k+1})^T(\mathcal{A}u^{k+1} - b) \geq (\lambda - \lambda^{k+1})^T\frac{1}{\beta}(\lambda^k - \lambda^{k+1}), \ \ \forall \lambda \in \Re^m. \quad (3.6)$$

Combining VI's (3.5) and (3.6), we get

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} u - u^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \begin{pmatrix} -\mathcal{A}^T \lambda^{k+1} \\ \mathcal{A}u^{k+1} - b \end{pmatrix} \geq (\lambda - \lambda^{k+1})^T \frac{1}{\beta}(\lambda^k - \lambda^{k+1}),$$

for all $w = (u, \lambda) \in \Omega$. Using the notations in (3.2), we get the compact form

$$\begin{aligned} \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \\ \geq \quad (\lambda - \lambda^{k+1})^T \frac{1}{\beta}(\lambda^k - \lambda^{k+1}), \ \forall w \in \Omega. \end{aligned} \tag{3.7}$$

This is the PPA form (2.6) in which

$$v = \lambda \qquad \text{and} \qquad H = \frac{1}{\beta} I_m.$$

The related contraction inequality (2.8) becomes

$$\|\lambda^{k+1} - \lambda^*\|^2_{\frac{1}{\beta}I_m} \leq \|\lambda^k - \lambda^*\|^2_{\frac{1}{\beta}I_m} - \|\lambda^k - \lambda^{k+1}\|^2_{\frac{1}{\beta}I_m}$$

or

$$\|\lambda^{k+1} - \lambda^*\|^2 \leq \|\lambda^k - \lambda^*\|^2 - \|\lambda^k - \lambda^{k+1}\|^2. \tag{3.8}$$

The above inequality is the key for the convergence proof of the ALM.

4 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 [3, 14])

$$\min_x \max_y \{\Phi(x,y) = \theta_1(x) - y^T Ax - \theta_2(y) \,|\, x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{4.1}$$

Let (x^*, y^*) be the solution of (4.1), then we have

$$\left\{ \begin{array}{ll} x^* \in \mathcal{X}, \;\; \Phi(x,y^*) - \Phi(x^*,y^*) \ge 0, \;\; \forall\, x \in \mathcal{X}, & \text{(4.2a)} \\ y^* \in \mathcal{Y}, \;\; \Phi(x^*,y^*) - \Phi(x^*,y) \ge 0, \;\; \forall\, y \in \mathcal{Y}. & \text{(4.2b)} \end{array} \right.$$

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

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

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

$$u^* \in \Omega, \quad \theta(u) - \theta(u^*) + (u-u^*)^T F(u^*) \ge 0, \; \forall\, u \in \Omega, \tag{4.3}$$

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

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

For the convex optimization problem $\min\{\theta(x) \mid Ax = b, x \in \mathcal{X}\}$, whose Lagrangian function is $L(x, y) = \theta(x) - y^T(Ax - b)$, we can rewrite it as

$$L(x, y) = \theta(x) - y^T Ax - (-b^T y),$$

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

Find the saddle point of the Lagrangian function is a special $\min - \max$ problem (4.1) whose $\theta_1(x) = \theta(x), \quad \theta_2(y) = -b^T y$ and $\mathcal{Y} = \Re^m$.

4.1 求解鞍点问题的 原始-对偶混合梯度法 PDHG [16]

For given (x^k, y^k), PDHG [16] produces a pair of (x^{k+1}, y^{k+1}). First,

$$x^{k+1} = \text{argmin}\{\Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2 \,|\, x \in \mathcal{X}\}, \tag{4.4a}$$

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

$$y^{k+1} = \text{argmax}\{\Phi(x^{k+1}, y) - \frac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}. \tag{4.4b}$$

Ignoring the constant term in the objective function, the subproblems (4.4) are reduced to

$$\left\{ \begin{array}{ll} x^{k+1} = \text{argmin}\{\theta_1(x) - x^T A^T y^k + \frac{r}{2}\|x - x^k\|^2 \,|\, x \in \mathcal{X}\}, & (4.5a) \\ y^{k+1} = \text{argmin}\{\theta_2(y) + y^T A x^{k+1} + \frac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}. & (4.5b) \end{array} \right.$$

According to Lemma 1, the optimality condition of (4.5a) is $x^{k+1} \in \mathcal{X}$ and

$$\theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T\{-A^T y^k + r(x^{k+1} - x^k)\} \ge 0, \;\; \forall\, x \in \mathcal{X}. \tag{4.6}$$

Similarly, from (4.5b) we get $y \in \mathcal{Y}$ and

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{Ax^{k+1} + s(y^{k+1} - y^k)\} \ge 0, \;\; \forall\, y \in \mathcal{Y}. \tag{4.7}$$

Combining (4.6) and (4.7), we have $(x^{k+1}, y^{k+1}) \in \mathcal{X} \times \mathcal{Y}$,

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T y^{k+1} \\ Ax^{k+1} \end{pmatrix} + \begin{pmatrix} r(x^{k+1} - x^k) + A^T(y^{k+1} - y^k) \\ s(y^{k+1} - y^k) \end{pmatrix} \right\} \ge 0, \quad \forall (x, y) \in \Omega.$$

The compact form is $u^{k+1} \in \Omega$,

$$\begin{aligned} u^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T F(u^{k+1}) \\ \ge (u - u^{k+1})^T Q(u^k - u^{k+1}), \quad \forall\, u \in \Omega. \end{aligned} \qquad (4.8)$$

where

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

It does not be the PPA form (2.3), and we can not expect its convergence.

The following example of linear programming indicates the original PDHG (4.4) is not necessary convergent.

Consider a pair of the primal-dual linear programming:

$$\text{(Primal)} \quad \begin{array}{ll} \min & c^T x \\ \text{s. t.} & Ax = b \\ & x \geq 0. \end{array} \qquad \text{(Dual)} \quad \begin{array}{ll} \max & b^T y \\ \text{s. t.} & A^T y \leq c. \end{array}$$

We take the following example

$$\text{(P)} \quad \begin{array}{ll} \min & x_1 + 2x_2 \\ \text{s. t.} & x_1 + x_2 = 1 \\ & x_1, \ x_2 \geq 0. \end{array} \qquad \text{(D)} \quad \begin{array}{ll} \max & y \\ \text{s. t.} & \begin{bmatrix} 1 \\ 1 \end{bmatrix} y \leq \begin{bmatrix} 1 \\ 2 \end{bmatrix} \end{array}$$

where $A = [1, 1], \ b = 1, c = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and the vector $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

Note that its Lagrange function is

$$L(x,y) = c^T x - y^T (Ax - b) \tag{4.9}$$

which defined on $\Re^2_+ \times \Re$. $x^* = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $y^* = 1$. is the unique saddle point of the Lagrange function.

For solving the $\min$-$\max$ problem (4.9), by using (4.4), the iterative formula is

$$\begin{cases} x^{k+1} = \arg\min\{c^T x - x^T A^T y^k + \frac{r}{2}\|x - x^k\|^2 | x \geq 0\} \\ \qquad\;\; = \arg\min\{\frac{r}{2}\|x - [x^k + \frac{1}{r}(A^T y^k - c)]\|^2 | x \geq 0\} \\ \qquad\;\; = P_{\Re^n_+}[x^k + \frac{1}{r}(A^T y^k - c)] \\ \qquad\;\; = \max\{[x^k + \frac{1}{r}(A^T y^k - c)], 0\}, \\ y^{k+1} = y^k - \frac{1}{s}(Ax^{k+1} - b). \end{cases}$$

We use $(x_1^0, x_2^0; y^0) = (0,0;0)$ as the start point. For this example, the method is not convergent.

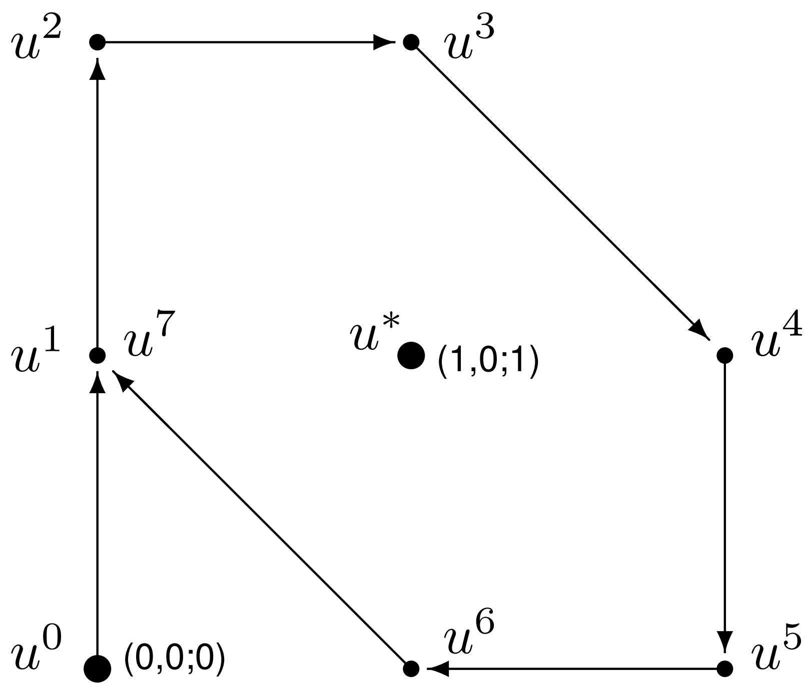
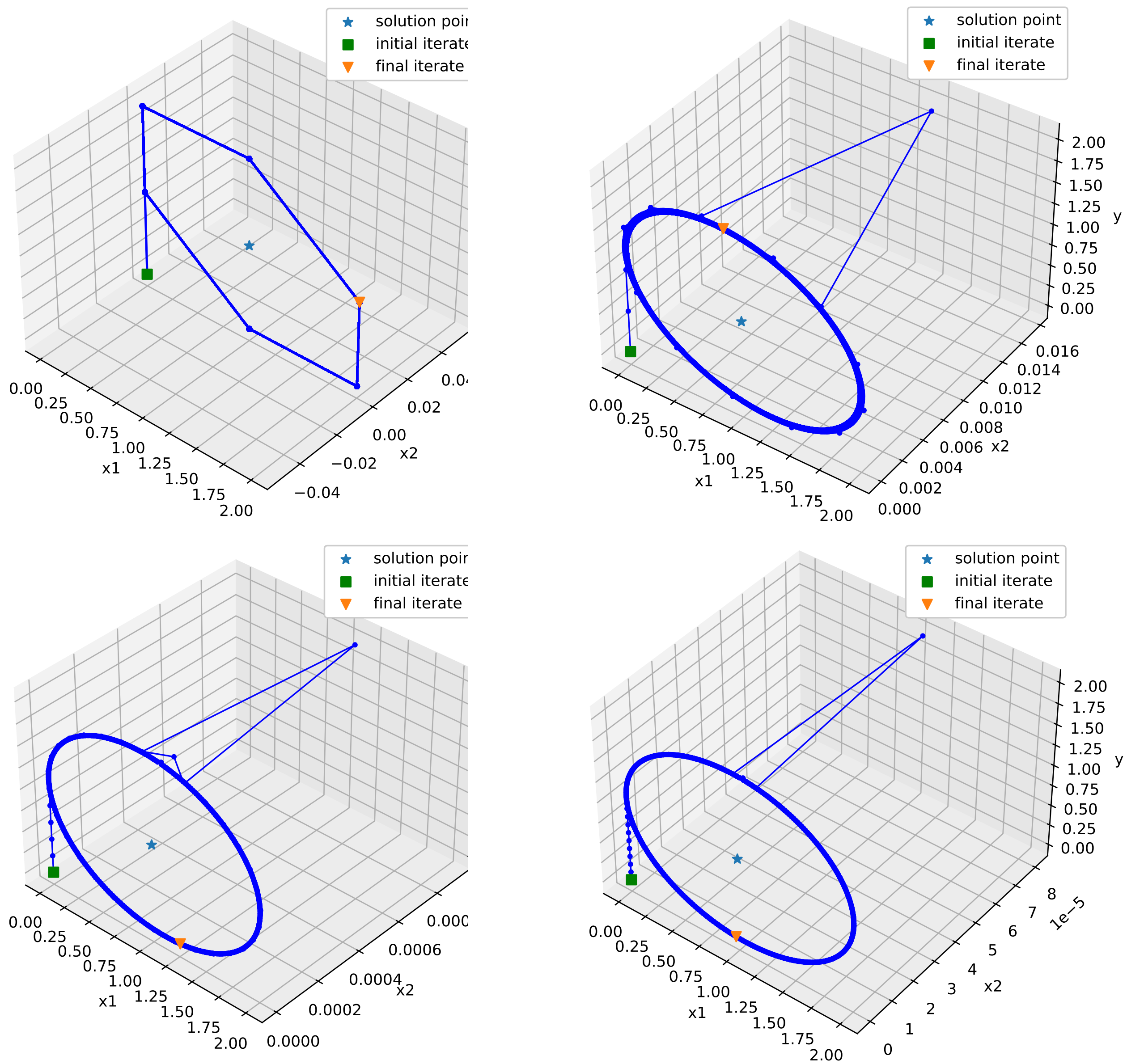


Fig. 2.1 The sequence generated by PDHG Method with $r = s = 1$

$$u^0 = (0, 0; 0)$$
$$u^1 = (0, 0; 1)$$
$$u^2 = (0, 0; 2)$$
$$u^3 = (1, 0; 2)$$
$$u^4 = (2, 0; 1)$$
$$u^5 = (2, 0; 0)$$
$$u^6 = (1, 0; 0)$$
$$u^7 = (0, 0; 1)$$

$$\boldsymbol{u^{k+6} = u^k}$$



对 $r = s = 1, 2, 5, 10$, PDHG 方法都不收敛

4.2 Customized Proximal Point Algorithm-Classical Version

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 (4.8) 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 (4.7) 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

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

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{A[2x^{k+1} - x^k] + s(y^{k+1} - y^k)\} \geq 0. \quad (4.10)$$

Thus, for given (x^k, y^k), producing a proximal point (x^{k+1}, y^{k+1}) via (4.4a) and (4.10) can be summarized as:

$$x^{k+1} = \text{argmin}\Big\{\Phi(x, y^k) + \frac{r}{2}\big\|x - x^k\big\|^2 \;\Big|\; x \in \mathcal{X}\Big\}. \tag{4.11a}$$

$$y^{k+1} = \text{argmax}\Big\{\Phi\big([2x^{k+1} - x^k], y\big) - \frac{s}{2}\big\|y - y^k\big\|^2\Big\} \tag{4.11b}$$

By ignoring the constant term in the objective function, getting x^{k+1} from (4.11a) 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 (4.11b) 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 (4.11a)-(4.11b).

In the case that $rs > \|A^TA\|$, 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* (4.11) *satisfies*

$$\|u^{k+1} - u^*\|_H^2 \le \|u^k - u^*\|_H^2 - \|u^k - u^{k+1}\|_H^2. \tag{4.12}$$

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

the iterative scheme is

$$x^{k+1} = \text{argmin}\Big\{\theta(x) + \frac{r}{2}\Big\|x - \big[x^k + \frac{1}{r}A^Ty^k\big]\Big\|^2 \,\Big|\, x \in \mathcal{X}\Big\}. \tag{4.13a}$$

$$y^{k+1} = y^k - \frac{1}{s}\big[A(2x^{k+1} - x^k) - b\big]. \tag{4.13b}$$

For solving the $\min$-$\max$ problem (4.9), by using (4.11), the iterative formula is

$$\begin{cases} x^{k+1} = \max\{[x^k + \frac{1}{r}(A^T y^k - c)], 0\}, \\ y^{k+1} = y^k - \frac{1}{s}[A(2x^{k+1} - x^k) - b]. \end{cases}$$

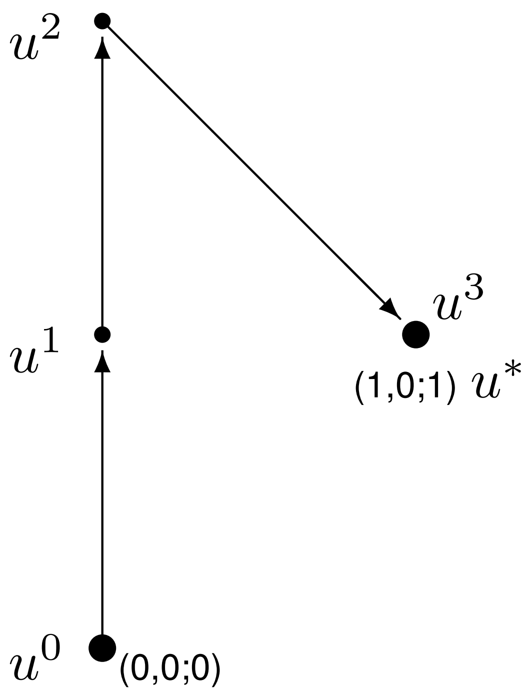
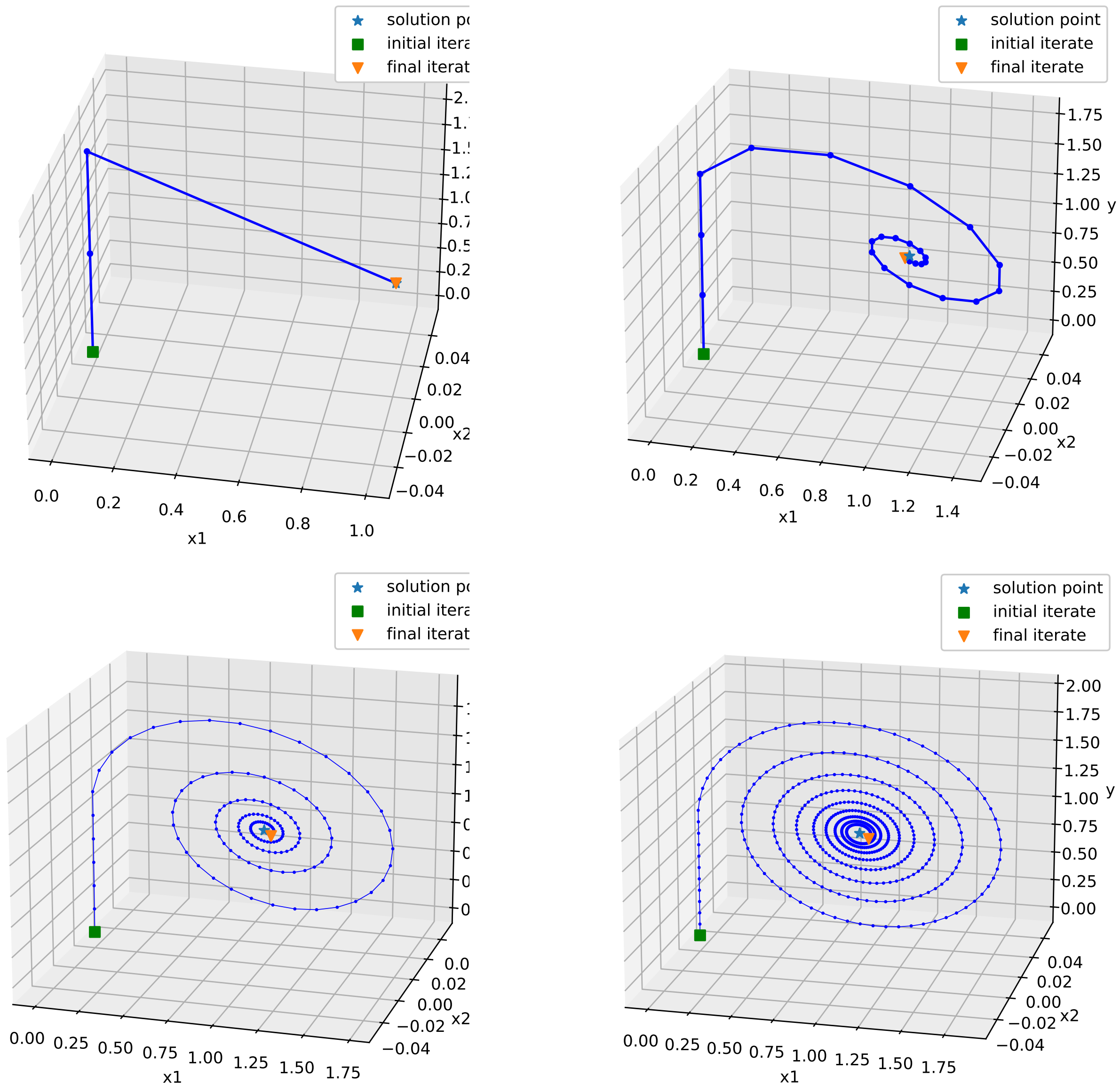


Fig. 2.2 The sequence generated by C-PPA Method with $r = s = 1$

$$u^0 = (0,0;0)$$
$$u^1 = (0,0;1)$$
$$u^2 = (0,0;2)$$
$$u^3 = (1,0;1)$$
$$\boldsymbol{u^3 = u^*}.$$



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

Besides (4.11), (x^{k+1}, y^{k+1}) can be produced by using the dual-primal order:

$$y^{k+1} = \text{argmax}\Big\{\Phi\big(x^k, y\big) - \frac{s}{2}\big\|y - y^k\big\|^2\Big\} \qquad \text{(4.14a)}$$

$$x^{k+1} = \text{argmin}\Big\{\Phi(x, (2y^{k+1} - y^k)) + \frac{r}{2}\big\|x - x^k\big\|^2 \,\Big|\, x \in \mathcal{X}\Big\}. \qquad \text{(4.14b)}$$

By using the notation of u, $F(u)$ and Ω in (4.3), we get $u^{k+1} \in \Omega$ and

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

where

$$H = \begin{pmatrix} rI_n & -A^T \\ -A & sI_m \end{pmatrix}.$$

Note that in the primal-dual order,

$$H = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix}.$$

In the both cases, $rs > \|A^T A\|$, the matrix H is positive definite.

Remark We use CP-PPA to solve linearly constrained convex optimization.

If the equality constraints $Ax = b$ is changed to $Ax \ge b$, namely,

$$\min\{\theta(x) \mid Ax = b,\; x \in \mathcal{X}\} \quad \Rightarrow \quad \min\{\theta(x) \mid Ax \ge b,\; x \in \mathcal{X}\}.$$

In this case, the Lagrange multiplier y should be nonnegative. $\Omega = \mathcal{X} \times \Re^m_+$.
We need only to make a slight change in the algorithms.

In the primal-dual order (4.11b), it needs to change the update dual update form

$$y^{k+1} = y^k - \tfrac{1}{s}\big(A(2x^{k+1} - x^k) - b\big) \quad \Rightarrow \quad y^{k+1} = \big[y^k - \tfrac{1}{s}\big(A(2x^{k+1} - x^k) - b\big)\big]_+$$

In the dual-primal order (4.14a), it needs to change the update dual update form

$$y^{k+1} = y^k - \tfrac{1}{s}\big(Ax^k - b\big) \qquad \Rightarrow \qquad y^{k+1} = \big[y^k - \tfrac{1}{s}\big(Ax^k - b\big)\big]_+$$

4.3 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

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

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.

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

对 $\theta \in (0, 1)$, 收敛性没有证明, 至今还是一个 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 [**?**].

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

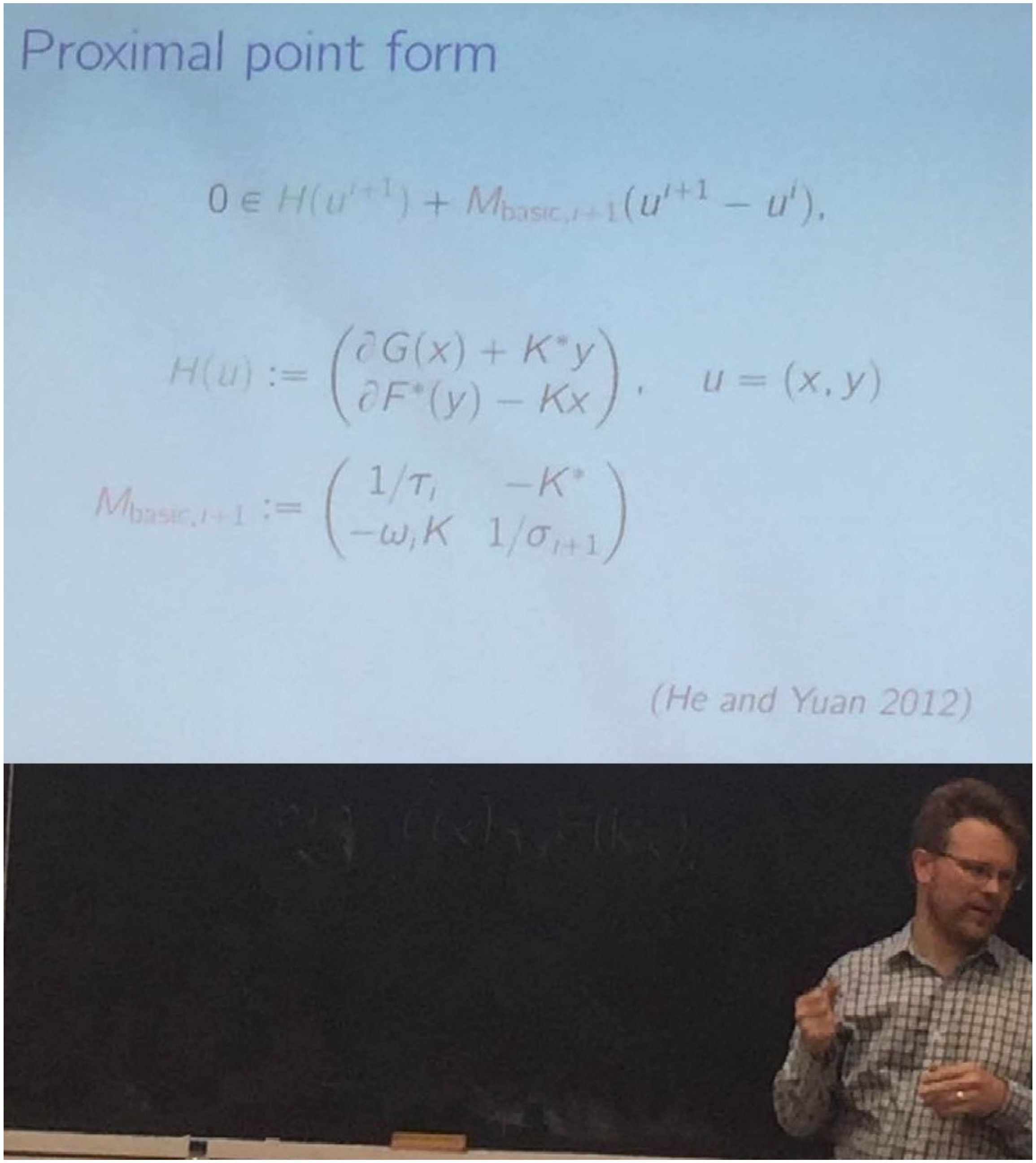
- Input: Operator norm $L := \|K\|$, Lipschitz constant L_f of ∇f, and Bregman distance functions D_x and D_y.
- Initialization: Choose $(x^0, y^0) \in \mathcal{X} \times \mathcal{Y}$, $\tau, \sigma > 0$
- Iterations: For each $n \geq 0$ let

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

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。

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.

5 ALM in PPA-sense

The methods introduced in this section are recently published in [17].

根据预设正定矩阵 构造 PPA 算法. 方法可以在 [17] 中查到.

The convex optimization problem,

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

is translated to the equivalent variational inequality :

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, u \in \Omega, \qquad \text{(5.1a)}$$

where

$$w = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ Ax - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{X} \times \Re^m. \qquad \text{(5.1b)}$$

5.1 Relaxed PPA in Primal-Dual Order

Relaxed PPA for the variational inequality (5.1) : Find $\tilde{w}^k \in \Omega$, such that

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (w - \tilde{w}^k)^T H(w^k - \tilde{w}^k), \ \forall w \in \Omega, \quad \text{(5.2a)}$$

where

$$H = \begin{pmatrix} \beta A^T A + \delta I_n & A^T \\ A & \frac{1}{\beta} I_m \end{pmatrix}. \quad \text{(5.2b)}$$

The concrete formula of (5.2) is

The underline part is $F(\tilde{w}^k)$:

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

$$\begin{cases} \theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T \\ \qquad \{\underline{-A^T\tilde{\lambda}^k} + \boldsymbol{(\beta A^T A + \delta I_n)}(\tilde{x}^k - x^k) + \boldsymbol{A^T}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \\ \qquad (\underline{A\tilde{x}^k - b}) \quad + \boldsymbol{A}(\tilde{x}^k - x^k) \quad + \quad \boldsymbol{(1/\beta)}(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases} \quad \text{(5.3)}$$

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

How to implement the prediction? To get $\tilde{w}^k$ which satisfies (5.3),

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

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

Then, we use the form

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

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

5.2 Relaxed PPA in Dual-Primal Order

Relaxed PPA for the variational inequality (5.1) : Find $\tilde{w}^k \in \Omega$, such that

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (w - \tilde{w}^k)^T H(w^k - \tilde{w}^k), \ \forall w \in \Omega, \tag{5.4a}$$

where

$$H = \begin{pmatrix} \beta A^T A + \delta I_n & -A^T \\ -A & \frac{1}{\beta} I_m \end{pmatrix}, \quad \text{(a small } \delta > 0\text{, say } \delta = 0.05\text{).} \tag{5.4b}$$

Then, we use the form

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

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

The underline part is $F(\tilde{w}^k)$:

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

The concrete form of (5.4) is

$$\begin{cases} \theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T \\ \qquad \{\underline{-A^T\tilde{\lambda}^k} + \boldsymbol{(\beta A^T A + \delta I_{n_2})}(\tilde{x}^k - x^k) \boldsymbol{- A^T}(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \\ \qquad (\underline{A\tilde{x}^k - b}) \quad \boldsymbol{-A}(\tilde{x}^k - x^k) \quad + \quad \boldsymbol{(1/\beta)}\, (\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

$$\begin{cases} \theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T \\ \qquad \{-A^T(2\tilde{\lambda}^k - \lambda^k) + (\beta A^T A + \delta I_{n_2})(\tilde{x}^k - x^k)\} \ge 0, \\ \qquad (Ax^k - b) \quad + \quad (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

Implementation of (5.4) is (Dual-Primal)

$$\begin{cases} \tilde{\lambda}^k = \lambda^k - \beta(Ax^k - b), & \text{(5.5a)} \\ \tilde{x}^k = \text{Argmin}\left\{ \begin{array}{c} \theta(x) - x^T A^T[2\tilde{\lambda}^k - \lambda^k] + \\ \frac{1}{2}(x - x^k)^T(\underline{\beta A^T A + \delta I_n})(x - x^k) \end{array} \,\middle|\, x \in \mathcal{X} \right\}. & \text{(5.5b)} \end{cases}$$

5.3 PPA in Primal-Dual Order

Relaxed PPA for the variational inequality (5.1) :

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (w - \tilde{w}^k)^T H(w^k - \tilde{w}^k), \ \forall w \in \Omega, \tag{5.6a}$$

where

$$H = \begin{pmatrix} \delta I_n & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix}. \tag{5.6b}$$

Then, we use the form

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

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

The underline part is $F(\tilde{w}^k)$:

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

The concrete form of (5.6) is

$$\left\{ \begin{array}{r} \theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T\{\underline{-A^T\tilde{\lambda}^k} + \boldsymbol{\delta I_n}(\tilde{x}^k - x^k)\} \ge 0, \\ (\underline{A\tilde{x}^k - b}) + \boldsymbol{(1/\beta)}\,(\tilde{\lambda}^k - \lambda^k) = 0. \end{array} \right.$$

Using $\boxed{\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k - b) = [\lambda^k - \beta(Ax^k - b)] - \beta A(\tilde{x}^k - x^k)}$

$$\left\{ \begin{array}{c} \theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T \left\{ \begin{array}{c} -A^T[\lambda^k - \beta(Ax^k - b)] \\ +(\delta I_n + A^TA)(\tilde{x}^k - x^k) \end{array} \right\} \ge 0, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k - b). \end{array} \right.$$

Implementation

$$\left\{ \begin{array}{l} \tilde{x}^k = \text{Argmin}\left\{ \begin{array}{l} \theta(x) - x^TA^T[\lambda^k - \beta(Ax^k - b)] + \\ \frac{1}{2}(x - x^k)^T(\underline{\beta A^TA + \delta I_n})(x - x^k) \end{array} \middle| x \in \mathcal{X} \right\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k - b). \end{array} \right.$$

5.4 Balanced ALM [8]

Relaxed PPA for the variational inequality (5.1) : Find $\tilde{w}^k \in \Omega$, such that

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (w - \tilde{w}^k)^T H(w^k - \tilde{w}^k), \ \forall w \in \Omega, \tag{5.8a}$$

where

$$H = \begin{pmatrix} rI_n & A^T \\ A & \frac{1}{r}AA^T + \delta I_m \end{pmatrix} \quad \text{is positive definite.} \tag{5.8b}$$

Then, we use the form

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

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

The underline part is $F(\tilde{w}^k)$:

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

The concrete form of (5.8) is

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

It can written as

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

Thus, the predictor $\tilde{w}^k$ in balanced ALM (5.8) is implemented by

$$\begin{cases} \tilde{x}^k = \arg\min\{\theta(x) - x^TA^T\lambda^k + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}, & \text{(5.9a)} \\ \tilde{\lambda}^k = \arg\min\Big\{\lambda^T\big(A[2\tilde{x}^k - x^k] - b\big) + \frac{1}{2}\big\|\lambda - \lambda^k\big\|^2_{(\frac{1}{r}AA^T + \delta I_m)}\Big\}. & \text{(5.9b)} \end{cases}$$

Remark. $\tilde{\lambda}^k$ in (5.9b) is the solution of the following system of linear equations:

$$H_0(\lambda - \lambda^k) + \big(A[2\tilde{x}^k - x^k] - b\big) = 0, \tag{5.10}$$

where

$$H_0 = \frac{1}{r}AA^T + \delta I_m. \tag{5.11}$$

Because the matrix H_0 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 (5.9a) 中的二次项式平凡的, 降低了问题求解的难度.
- λ-子问题 (5.9b) 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 (see (5.11)) 做一次 Cholesky 分解.

6 平行求解子问题的 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).
根据 PPA 算法的要求, 设计的右端矩阵为对称正定.

Primal-Dual Order

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\geq(w-\tilde{w}^k)^TH(w^k-\tilde{w}^k),\ \forall w\in\Omega, \tag{6.1a}$$

where

$$H=\begin{pmatrix}\beta A^TA+\delta I_{n_1} & 0 & A^T\\ 0 & \beta B^TB+\delta I_{n_2} & B^T\\ A & B & \frac{2}{\beta}I_m\end{pmatrix}. \tag{6.1b}$$

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 (6.1) is

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

After simple organization, we obtain

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

In fact, the prediction can be arranged by

$$\left\{\begin{array}{ll} \tilde{x}^k = \arg\min\left\{ \begin{array}{c} \theta_1(x) - x^T A^T \lambda^k \\ +\frac{1}{2}\beta\|A(x-x^k)\|^2 + \frac{1}{2}\delta\|x-x^k\|^2 \end{array} \;\middle|\; x\in\mathcal{X} \right\} & \text{(6.2a)} \\[2ex] \tilde{y}^k = \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T \lambda^k \\ +\frac{1}{2}\beta\|B(y-y^k)\|^2 + \frac{1}{2}\delta\|y-y^k\|^2 \end{array} \;\middle|\; y\in\mathcal{Y} \right\} & \text{(6.2b)} \\[2ex] \tilde{\lambda}^k = \lambda^k - \frac{1}{2}\beta\big[2(A\tilde{x}^k + B\tilde{y}^k - b) - (Ax^k + By^k - b)\big] & \text{(6.2c)} \end{array}\right.$$

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

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

Dual-Primal Order

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (w - \tilde{w}^k)^T H(w^k - \tilde{w}^k), \ \forall w \in \Omega, \tag{6.3a}$$

where

$$H = \begin{pmatrix} \beta A^T A + \delta I_{n_1} & 0 & -A^T \\ 0 & \beta B^T B + \delta I_{n_2} & -B^T \\ -A & -B & \frac{2}{\beta} I_m \end{pmatrix}. \tag{6.3b}$$

The both matrices

$$H = \begin{pmatrix} \beta A^T A + \delta I_{n_1} & -A^T \\ -A & \frac{1}{\beta} I_m \end{pmatrix} \succ 0, \quad \begin{pmatrix} \beta B^T B + \delta I_{n_2} & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \succ 0.$$

The concrete form of (6.3) is

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \\ \qquad \{\underline{-A^T\tilde{\lambda}^k} + \boldsymbol{(\beta A^T A + \delta I_{n_1})}(\tilde{x}^k - x^k)\boldsymbol{-A^T}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \\ \qquad \{\underline{-B^T\tilde{\lambda}^k} + \boldsymbol{(\beta B^T B + \delta I_{n_2})}(\tilde{y}^k - y^k)\boldsymbol{-B^T}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \\ (\underline{A\tilde{x}^k + B\tilde{y}^k - b})\boldsymbol{-A}(\tilde{x}^k - x^k)\boldsymbol{-B}(\tilde{y}^k - y^k) + \boldsymbol{(2/\beta)}\,(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

经整理归并一下得到

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T(2\tilde{\lambda}^k - \lambda^k) \\ \qquad\qquad + (\beta A^T A + \delta I_{n_1})(\tilde{x}^k - x^k)\} \geq 0, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{-B^T(2\tilde{\lambda}^k - \lambda^k) \\ \qquad\qquad + (\beta B^T B + \delta I_{n_2})(\tilde{y}^k - y^k)\} \geq 0, \\ (Ax^k + By^k - b) + (2/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

In fact, the prediction can be arranged by

$$
\left\{
\begin{array}{l}
\tilde{\lambda}^k = \lambda^k - \frac{1}{2}\beta(Ax^k + By^k - b), \qquad \text{(6.4a)} \\[2ex]
\tilde{x}^k \in \arg\min\left\{ \begin{array}{c} \theta_1(x) - x^T A^T [2\tilde{\lambda}^k - \lambda^k] \\ +\frac{1}{2}\beta\|A(x - x^k)\|^2 + \frac{1}{2}\delta\|x - x^k\|^2 \end{array} \middle| x \in \mathcal{X} \right\} \qquad \text{(6.4b)} \\[2ex]
\tilde{y}^k \in \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T [2\tilde{\lambda}^k - \lambda^k] \\ +\frac{1}{2}\beta\|B(y - y^k)\|^2 + \frac{1}{2}\delta\|y - y^k\|^2 \end{array} \middle| y \in \mathcal{Y} \right\}. \qquad \text{(6.4c)}
\end{array}
\right.
$$

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

我们关于 ADMM 的研究, 始于 1997 年, 第一篇 ADMM 方面的论文发表于 1998 年. 这一讲中 §4-§6 介绍的 ADMM 类方法, 可以从 [17] 中找到.

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

7 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 x-子问题是比较简单的.

Primal-Dual Order

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \geq (w-\tilde{w}^k)^T H(w^k-\tilde{w}^k), \ \forall w\in\Omega, \quad (7.1a)$$

where

$$H=\begin{pmatrix} \beta A^TA+\delta I_{n_1} & 0 & A^T \\ 0 & sI_{n_2} & B^T \\ A & B & (\frac{1}{\beta}+\delta)I_m+\frac{1}{s}BB^T \end{pmatrix}. \quad (7.1b)$$

The both matrices

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

The concrete form of (7.1) is

$$\left\{\begin{array}{l}\theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T \\ \qquad\qquad \{\underline{-A^T\tilde{\lambda}^k}+\boldsymbol{(\beta A^TA+\delta I_{n_1})}(\tilde{x}^k-x^k)\boldsymbol{+A^T}(\tilde{\lambda}^k-\lambda^k)\}\ge 0,\\ \theta_2(y)-\theta_2(\tilde{y}^k)+(y-\tilde{y}^k)^T \\ \qquad\qquad \{\underline{-B^T\tilde{\lambda}^k}+\boldsymbol{sI_{n_2}}(\tilde{y}^k-y^k)\boldsymbol{+B^T}(\tilde{\lambda}^k-\lambda^k)\}\ge 0,\\ (\underline{A\tilde{x}^k+B\tilde{y}^k-b})\boldsymbol{+A}(\tilde{x}^k-x^k)\boldsymbol{+B}(\tilde{y}^k-y^k)\\ \qquad\qquad\qquad\qquad +\boldsymbol{\Big((\tfrac{1}{\beta}+\delta)I_m+\tfrac{1}{s}BB^T\Big)}(\tilde{\lambda}^k-\lambda^k)=0.\end{array}\right.$$

After simple organization, we obtain

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

In fact, the prediction can be arranged by

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

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

Dual-Primal Order

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (w - \tilde{w}^k)^T H(w^k - \tilde{w}^k), \ \forall w \in \Omega, \quad \text{(7.2a)}$$

where

$$H = \begin{pmatrix} \beta A^T A + \delta I_{n_1} & 0 & -A^T \\ 0 & sI_{n_2} & -B^T \\ -A & -B & (\frac{1}{\beta} + \delta) I_m + \frac{1}{s} B B^T \end{pmatrix}. \quad \text{(7.2b)}$$

The both matrices

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

The concrete form of (7.2) is

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \\ \qquad\qquad \{\underline{-A^T\tilde{\lambda}^k} + \boldsymbol{(\beta A^T A + \delta I_{n_1})}(\tilde{x}^k - x^k)\boldsymbol{-A^T}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \\ \qquad\qquad \{\underline{-B^T\tilde{\lambda}^k} + \boldsymbol{sI_{n_2}}(\tilde{y}^k - y^k)\boldsymbol{-B^T}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \\ (\underline{A\tilde{x}^k + B\tilde{y}^k - b})\boldsymbol{-A}(\tilde{x}^k - x^k)\boldsymbol{-B}(\tilde{y}^k - y^k) \\ \qquad\qquad\qquad + \boldsymbol{\Big((\frac{1}{\beta} + \delta)I_m + \frac{1}{s}BB^T\Big)}(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

After simple organization, we obtain

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T(2\tilde{\lambda}^k - \lambda^k) + (\beta A^T A + \delta I_{n_1})(\tilde{x}^k - x^k)\} \geq 0, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{-B^T(2\tilde{\lambda}^k - \lambda^k) + s(\tilde{y}^k - y^k)\} \geq 0, \\ (Ax^k + By^k - b) + \Big((\frac{1}{\beta} + \delta)I_m + \frac{1}{s}BB^T\Big)(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

In fact, the prediction can be arranged by

$$\left\{\begin{array}{l} \tilde{\lambda}^k = \lambda^k - \big((\frac{1}{\beta}+\delta)I_m + \frac{1}{s}BB^T\big)^{-1}(Ax^k + By^k - b) \\ \tilde{x}^k = \arg\min\left\{ \begin{array}{c} \theta_1(x) - x^T A^T(2\tilde{\lambda}^k - \lambda^k) \\ +\frac{1}{2}\beta\|A(x-x^k)\|^2 + \frac{1}{2}\delta\|x-x^k\|^2 \end{array} \middle| x \in \mathcal{X} \right\} \\ \tilde{y}^k = \arg\min\big\{\theta_2(y) - y^T B^T(2\tilde{\lambda}^k - \lambda^k) + \frac{1}{2}s\|y-y^k\|^2 \,\big|\, y \in \mathcal{Y}\big\} \end{array}\right.$$

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

References

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

[2] 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.

[3] 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.

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

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

[6] 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.

[7] B.S. He and Y. Shen, On the convergence rate of customized proximal point algorithm for convex optimization and saddle-point problem (in Chinese). Sci Sin Math, 2012, 42(5): 515 – 525, doi: 10.1360/012011-1049

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

[9] 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.

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

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

[12] 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.

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

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

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

[16] 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.

[17] 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.