图像处理中的一些凸优化问题及其相应的分裂收缩算法

何 炳 生

南方科技大学数学系　　南京大学数学系

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

感谢电子科技大学 张文星博士 提供图像素材

图像计算与数字医药国际研讨会, 2017.09.23 – 25

1 单个算子问题 PDHG $\longrightarrow$ Customized PPA

$$\begin{aligned} \min \ & \theta(x) \\ \text{s.t} \ & Ax = b \\ & x \in \mathcal{X} \end{aligned}$$

$$\min\{\theta(x) \mid Ax = b, \ x \in \mathcal{X}\} \qquad (1.1)$$

Image deblurring

Blurry can be produced by

defocus the camera's lens, the moving object, turbulence in the air, $\cdots$

Notations: $\mathbf{g}$ — observation, $\mathbf{f}$ — ideal image;

$\mathcal{U}$ — restriction on pixels, e.g., $\mathcal{U} = \{u \mid 0 \le u \le 255\}$

$\mathbf{g} = H\mathbf{f}, \quad H$ — blur matrix .

Model

$$\min \ \{\|\nabla \mathbf{f}\|_1 \ \mid \ H\mathbf{f} = \mathbf{g}, \ \ \mathbf{f} \in \mathcal{U}\}$$

original image

blurred image

restored image

Image inpainting

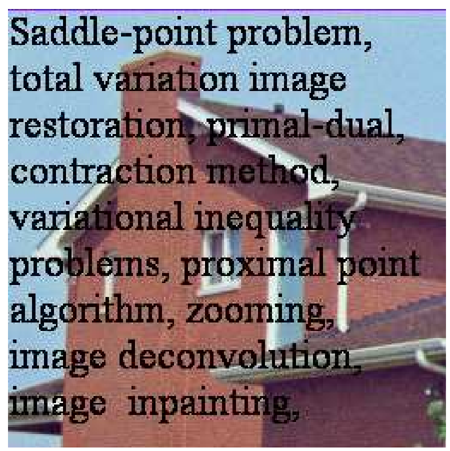
Some pixels are missing in image. Partial information of image is available

$$\mathbf{g} = S\,\mathbf{f}, \quad S \text{ — mask (missing pixels)}$$

Model

$$\min \; \{\|\nabla \mathbf{f}\|_1 \;\mid\; S\,\mathbf{f} = \mathbf{g}, \;\; \mathbf{f} \in \mathcal{U}\}$$

original image



missing pixel image

restored image

Image zooming and super-resolution

Produce a high-resolution (HR) image by its low-resolution (LR) image(s)

$$\mathbf{g} = D\,\mathbf{f}, \quad \mathbf{f} \text{ — HR image, } \mathbf{g} \text{ — LR image, } D \text{ — down-sampling}$$

Model $$\min\ \{\|\nabla \mathbf{f}\|_1 \ \mid\ D\mathbf{f} = \mathbf{g},\ \ \mathbf{f} \in \mathcal{U}\}$$

LR image

HR image

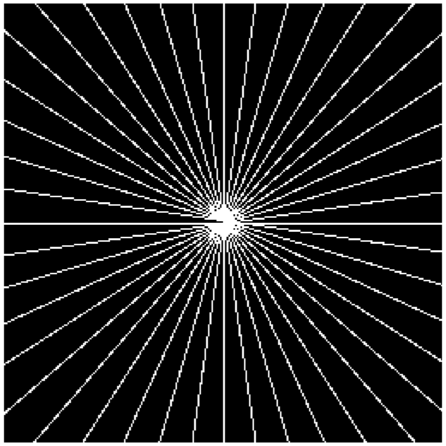
Magnetic resonance imaging (MRI)

Reconstruct a medical image by sampling its Fourier coefficients partially

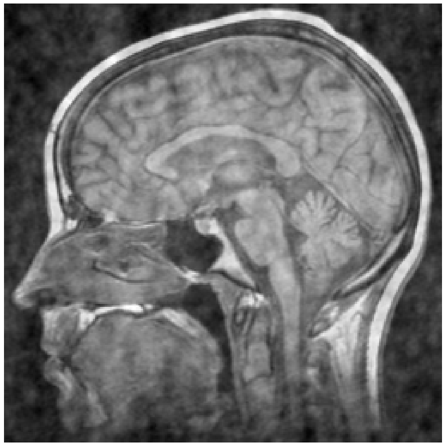
$$\mathcal{F}\mathbf{g} = P\mathcal{F}\mathbf{f}, \quad P \text{ — sampling mask, } \mathcal{F} \text{ — Fourier transform}$$

Model $\min \; \{\|\nabla \mathbf{f}\|_1 \;\mid\; P\mathcal{F}\mathbf{f} = \mathcal{F}\mathbf{g}\}$

medical image



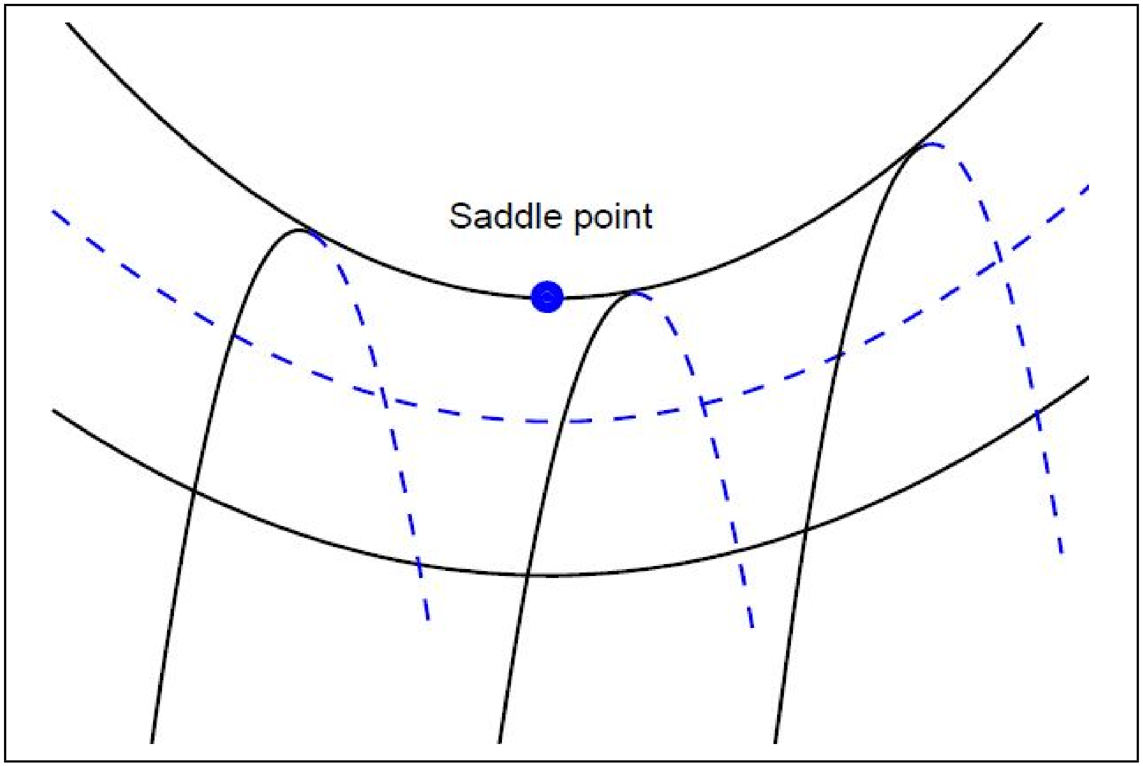
sampling mask



reconstruction

The Lagrange function of (1.1) is

$$L(x,\lambda) = \theta(x) - \lambda^T(Ax - b), \qquad (x,\lambda) \in \mathcal{X} \times \Re^m.$$



A pair of (x^*, λ^*) is called a saddle point of the Lagrange function, if

$$L_{\lambda\in\Re^m}(x^*,\lambda) \le L(x^*,\lambda^*) \le L_{x\in\mathcal{X}}(x,\lambda^*).$$

1.1 Original primal-dual hybrid gradient algorithm

For given (x^k, λ^k), produce a pare of (x^{k+1}, λ^{k+1}). First,

$$x^{k+1} = \text{Argmin}\{L(x, \lambda^k) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}, \tag{1.2a}$$

and then we obtain λ^{k+1} via

$$\lambda^{k+1} = \text{Argmax}\{L(x^{k+1}, \lambda) - \frac{s}{2}\|\lambda - \lambda^k\|^2 \mid \lambda \in \Re^m\}. \tag{1.2b}$$

However, this method is not necessarily convergent.

1.2 Proximal Point Algorithm-Classical Version

$$x^{k+1} = \text{argmin}\Big\{L(x, \lambda^k) + \frac{r}{2}\big\|x - x^k\big\|^2 \,\Big|\, x \in \mathcal{X}\Big\}. \tag{1.3a}$$

$$\lambda^{k+1} = \text{argmax}\Big\{L\big([2x^{k+1} - x^k], \lambda\big) - \frac{s}{2}\big\|\lambda - \lambda^k\big\|^2\Big\} \tag{1.3b}$$

By ignoring the constant term in the objective function, getting x^{k+1} from (1.3a) is equivalent to obtaining x^{k+1} from

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

The solution of (1.3b) is given by

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

Assumption: $\min\left\{\theta(x) + \frac{r}{2}\|x - a\|^2 \,|\, x \in \mathcal{X}\right\}$ **is simple**

Indeed, under the assumption, the sub-problem (1.3a) is simple.

Convergence Condition $rs > \|A^T A\|.$

Produce (x^{k+1}, λ^{k+1}) by using the dual-primal order:

$$\lambda^{k+1} = \text{argmax}\Big\{L\big(x^k, \lambda\big) - \frac{s}{2}\big\|\lambda - \lambda^k\big\|^2\Big\} \tag{1.4a}$$

$$x^{k+1} = \text{argmin}\Big\{L(x, (2\lambda^{k+1} - \lambda^k)) + \frac{r}{2}\big\|x - x^k\big\|^2 \,\Big|\, x \in \mathcal{X}\Big\}. \tag{1.4b}$$

1.3 Customized Proximal Point Algorithm-Extended Version

Extended PPA $\gamma = 1.5 \in (0, 2)$.

$$x^{k+1} := x^k - \gamma(x^k - x^{k+1}),$$

$$\lambda^{k+1} := \lambda^k - \gamma(\lambda^k - \lambda^{k+1}).$$

1.4 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

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

where

$$F \left(\begin{array}{c} x^{k+1} \\ y^{k+1} \end{array} \right) = \left(\begin{array}{c} \partial G(x^{k+1}) + K^T y^{k+1} \\ \partial F^*(y^{k+1}) - K x^{k+1} \end{array} \right) ,$$

and

$$M = \left[\begin{array}{cc} \mathrm{T}^{-1} & -K^T \\ -\theta K & \Sigma^{-1} \end{array} \right] . \tag{6}$$

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

[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.

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

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-

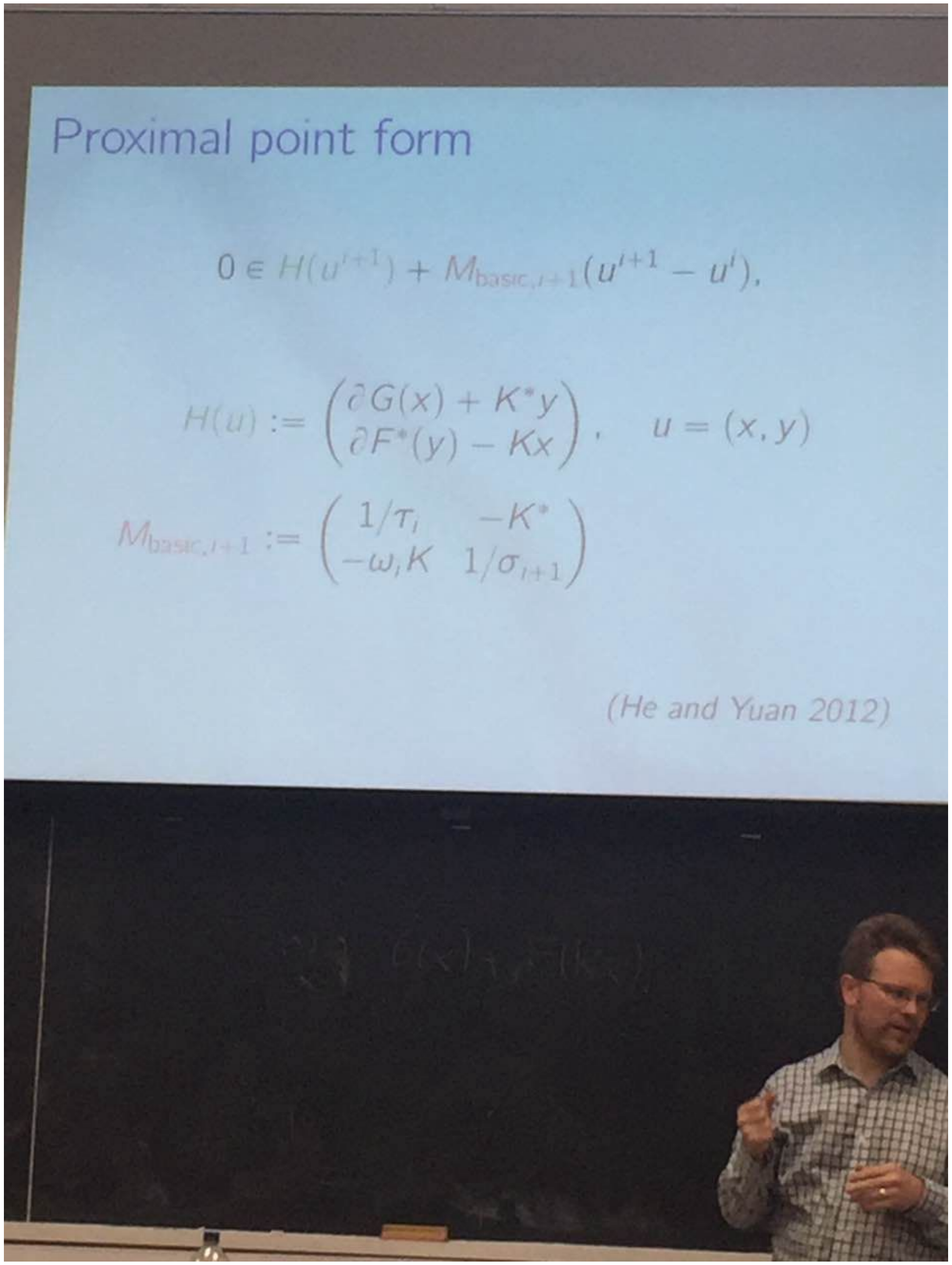
Algorithm 1: $O(1/N)$ Non-linear primal–dual algorithm

- 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. .



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, 首位报告人讲到, 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。

2 两个算子问题的交替方向法

$$\begin{aligned} \min \ & \theta_1(x) + \theta_2(y) \\ \text{s.t} \ & Ax + By = b \\ & x \in \mathcal{X}, y \in \mathcal{Y} \end{aligned}$$

$$\min \{\theta_1(x) + \theta_2(y) \,|\, Ax + By = b,\ x \in \mathcal{X},\ y \in \mathcal{Y}\} \quad (2.1)$$

Image decomposition

Separate the sketch (cartoon) and oscillating component (texture) of image

$$\mathbf{f} = \mathbf{u} + \mathbf{v}, \quad \mathbf{u} \text{ — cartoon part, } \mathbf{v} \text{ — texture part}$$

Model $\min \ \{\|\nabla \mathbf{u}\|_1 + \tau \|\mathbf{v}\|_{-1,\infty} \ | \ \mathbf{u} + \mathbf{v} = \mathbf{f}\}$

original image | cartoon part | texture part

Background extraction of surveillance video (I)

Considering the foreground object detection in complex environments and extract the background in surveillance video

$$D = X + Y, \quad D \text{ — original video, } X \text{ — background, } Y \text{ — foreground}$$

Model $\min \; \{\|X\|_* + \tau\|Y\|_1 \;\mid\; X + Y = D\}$

original video

foreground

background

Image denoising

Pixels are perturbed by a whole range of external and unwanted disturbances

$$\mathbf{g} = \mathbf{f} + noise$$

Model

$$\min \ \left\{\|\nabla\mathbf{f}\|_1 + \tfrac{1}{2}\|\mathbf{f}-\mathbf{g}\|_2^2\right\} \Leftrightarrow \min \ \left\{\|\mathbf{y}\|_1 + \tfrac{1}{2}\|\mathbf{f}-\mathbf{g}\|_2^2 \mid \nabla\mathbf{f}-\mathbf{y}=0\right\},$$

original image

noised image

restored image

- 2009 年春, 我的博士生杨俊锋从 Rice 联合培养回国, 介绍了用交替极小化算法 (AMA) 求解图像处理问题。

- 两个可分离算子问题罚函数方法松弛后就是交替极小化算法。

- 增广 Lagrange 乘子法优于罚函数方法(Nocedal $\&$ Wright [18]).

- 两个可分离算子问题的增广 Lagrange 乘子法松弛就是 ADMM。

Lagrange 函数

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

增广 Lagrange 函数

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

求解问题 (2.1) 的罚函数方法

$$(x^{k+1}, y^{k+1}) = \text{Argmin}\{\theta_1(x) + \theta_2(y) + \tfrac{\beta}{2}\|Ax + By - b\|^2 \big| x \in \mathcal{X},\ y \in \mathcal{Y}\}$$

求解问题 (2.1) 的增广 Lagrange 乘子法 从给定的 λ^k 开始

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

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

子问题难能度一样, 增广 Lagrange 乘子法(ALM)优于罚函数方法。

原因：迭代犹如博弈双方讨价还价, (ALM) 同时考虑了对方的感受。

共同的缺点 没有利用 x 和 y 的可分离结构！求解会无从着手。

求解问题 (2.1) 的松弛的罚函数方法 — 交替极小化方法(AMA)

从给定的 y^k 开始

$$x^{k+1} = \text{Argmin}\Big\{\theta_1(x) + \frac{\beta}{2}\|Ax + By^k - b\|^2 \Big| x \in \mathcal{X}\Big\},$$

$$y^{k+1} = \text{Argmin}\Big\{\theta_2(y) + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \Big| y \in \mathcal{Y}\Big\}.$$

求解问题 (2.1) 的松弛的增广 Lagrange 乘子法 — ADMM

从给定的 (y^k, λ^k) 开始

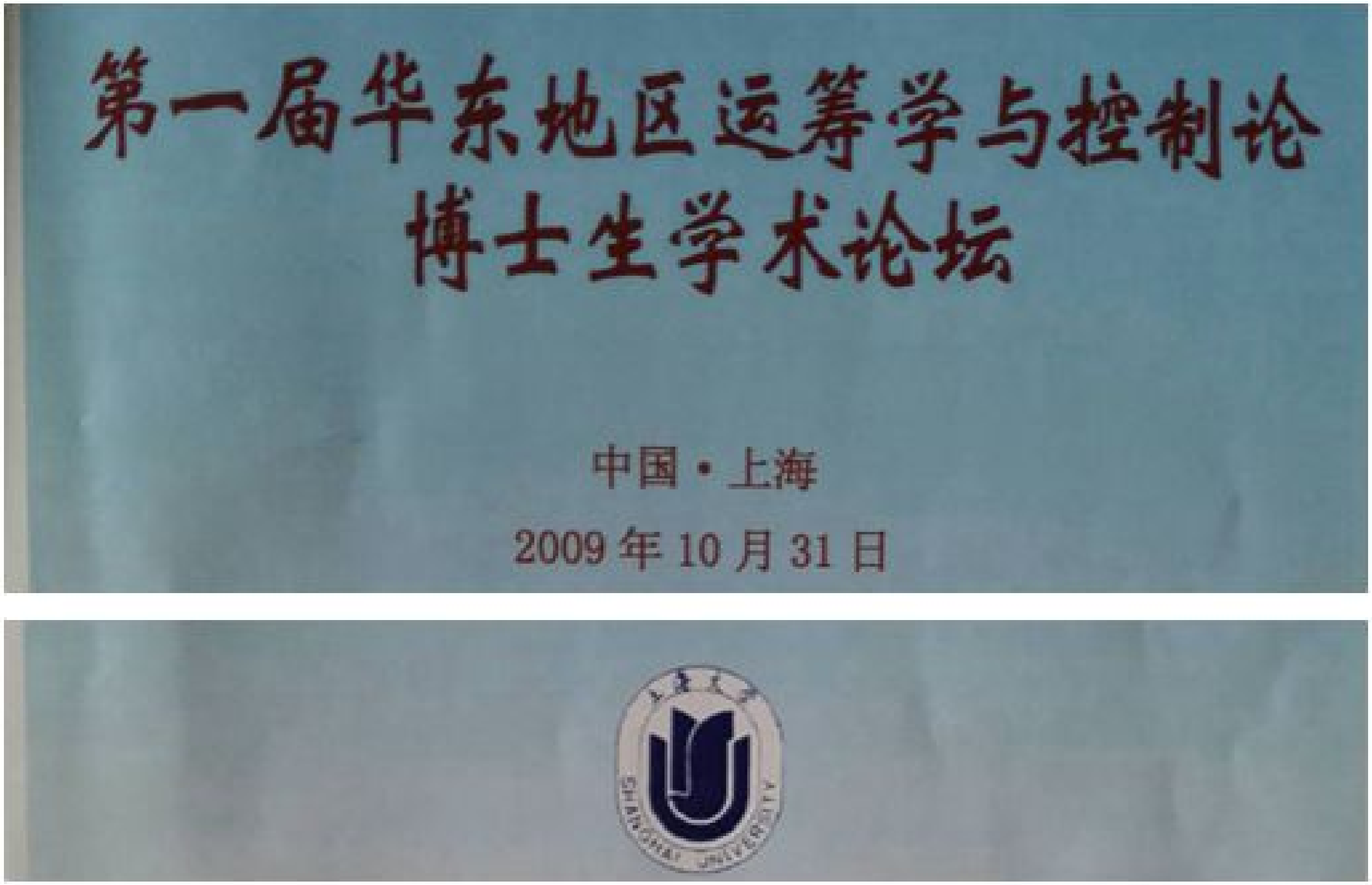
$$x^{k+1} = \text{Argmin}\Big\{\theta_1(x) - (\lambda^k)^T Ax + \frac{\beta}{2}\|Ax + By^k - b\|^2 \Big| x \in \mathcal{X}\Big\},$$

$$y^{k+1} = \text{Argmin}\Big\{\theta_2(y) - (\lambda^k)^T By + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \Big| y \in \mathcal{Y}\Big\},$$

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

都松弛, 乘子交替方向法 (ADMM) 应该优于交替极小化方法 (AMA)

我们新的一轮工作是从 2009 年开始的



2009 年 10 月华东地区运筹学与控制论博士论坛

与周青教授一起应邀在博士论坛做大会报告

大会报告 1

题　目：TBA

报告人：周　青　教授

华东师范大学数学系

Email: qzhou@math.ecnu.edu.cn

简　介：1957 年 3 月出生，获北京大学学士学位、美国加利福尼亚大学洛杉矶分校硕士和博士学位，教授、博士生指导教师。主要研究兴趣为低维拓扑、大范围分析和随机复杂网络。曾主持国家自然科学基金面上项目和重点项目，获

大会报告 2

题　目：信息技术中的凸优化问题及交替方向法求解

报告人：何炳生　教授

南京大学数学系

Email: hebma@nju.edu.cn

简　介：1981 年于南京大学数学系获理学学士，1986 年获博士学位，

题目：信息技术中的凸优化问题和交替方向法求解

摘　要：根据不完整的（或者受污染）信息求得全部（正确的）信息，在信息技术中有着广泛的应用背景。这些一度被认为是没法求解的数学问题，经过数学工作者（包括菲尔兹奖获得者和世界数学家大会一小时报告人）近几年的努力，已经证明，在一些实际问题能满足的假设条件下，通过求解相应的松弛问题，可以求得原问题的真解。这些松弛了的（光滑或非光滑）凸优化问题一般具有特殊结构、形式简单明了，问题条件也不是太坏。来源于实际的问题一般是大型的，有的问题变量是阶数很高的矩阵，需要极小化的目标函数是矩阵的核模——矩阵奇异值之和，问题规模超大。求解这些问题给数学规划工作者提供了新的用武之地的同时也提出了新的挑战。我们介绍这些问题的基本模型，说明这些问题都可以用交替方向法求解。在用数值计算结果说明交替方向法是求解这些问题的一个简单有效方法的同时提出提高算法效率方面一些值得继续研究的问题。

2009 年, 我的大会报告提请青年学者注意交替方向法

2.1 两个算子问题的 ADMM 方法的 (主要) 改进

1. ADMM in sense of PPA 交换顺序并外延 从 (y^k, λ^k) 出发.

$$\left\{\begin{array}{l}\left\{\begin{array}{lcl} x^{k+1} & = & \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \lambda^{k+1} & = & \lambda^k - \beta(Ax^{k+1} + By^k - b), \\ y^{k+1} & = & \text{Argmin}\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+1}) \,|\, y \in \mathcal{Y}\}, \end{array}\right. \quad (2.2a) \\ \left\{\begin{array}{lcl} y^{k+1} & := & y^k - \gamma(y^k - y^{k+1}), \\ \lambda^{k+1} & := & \lambda^k - \gamma(\lambda^k - \lambda^{k+1}). \end{array}\right. \quad \text{（松弛延拓）} \quad (2.2b) \end{array}\right.$$

这里 $\gamma \in (0, 2)$. 赋值号 $:=$ 表示 (2.2b) 右端的 (y^{k+1}, λ^{k+1}) 是由算法的前半部分 (2.2a) 产生的. 对多数问题, 这样往往能加快收敛速度.

- 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 Math., 56 (2013), 2179-2186.

2. Symmetric ADMM　对称的交替方向法

原始变量 x 和 y 本质上是平等的. 所以建议采用对称的交替方向法.

Symmetric Alternating Direction Method of Multipliers is described as

$$
\begin{cases}
x^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\[2ex]
\lambda^{k+\frac{1}{2}} = \lambda^k - \mu\beta(Ax^{k+1} + By^k - b), \\[2ex]
y^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \,|\, y \in \mathcal{Y}\}, \\[2ex]
\lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \mu\beta(Ax^{k+1} + By^{k+1} - b).
\end{cases}
\tag{2.3}
$$

wehre $\mu \in (0, 1)$ (usually $\mu = 0.9$).

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

3 多个可分离算子的凸优化问题

$$\begin{array}{rl}\min & \theta_1(x)+\theta_2(y)+\theta_3(z)\\ \text{s.t.} & Ax+By+Cz=b\\ & x\in\mathcal{X},\ y\in\mathcal{Y},\ z\in\mathcal{Z}\end{array}$$

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z)|Ax+By+Cz=b,\ x\in\mathcal{X},y\in\mathcal{Y},z\in\mathcal{Z}\}. \tag{3.1}$$

Background extraction of surveillance video (II)

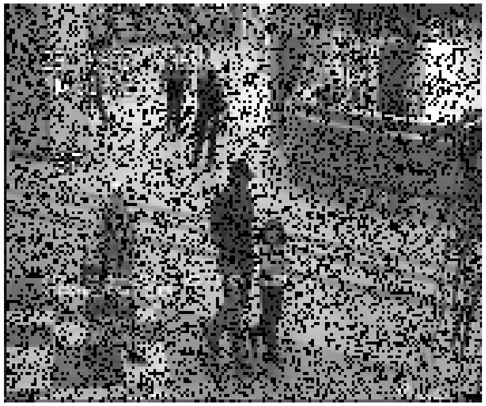
The original surveillance video has missing information and additive noise

$$P_\Omega(D) = P_\Omega(X+Y)\text{+noise}$$

P_Ω — indicating missing data, Z — noise/outliers

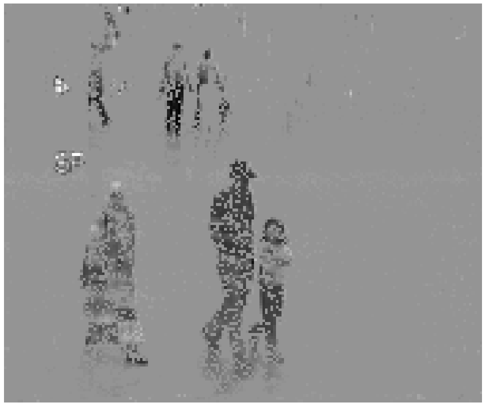
Model

$$\min\left\{\|X\|_* + \tau\|Y\|_1 + \|P_\Omega(Z)\|_F^2 \;\mid\; X+Y-Z=D\right\}$$



observed video

foreground



background

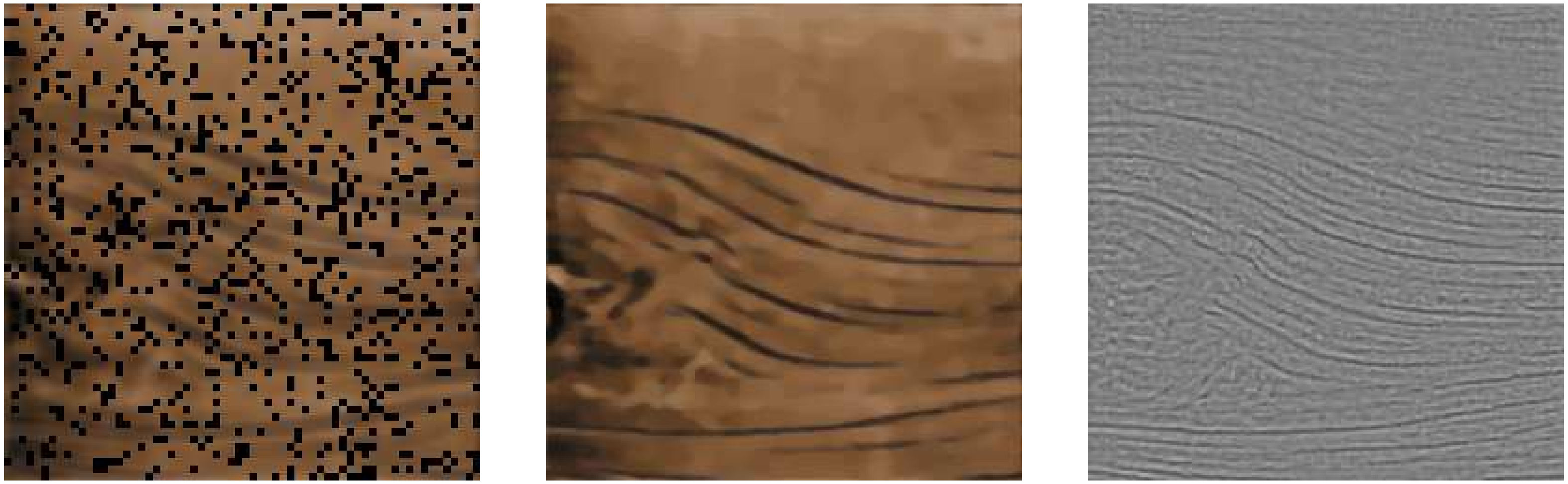
Image decomposition with degradations

The target image for decomposition contains degradations, e.g., blur, missing pixels, $\cdots$

$$\mathbf{f} = K(\mathbf{u} + \text{div}\,\mathbf{v}) + \mathbf{z}, \quad K \text{ — degradation operator}, \quad \mathbf{z} \text{ — noise/outlier}$$

Model

$$\min \left\{ \|\nabla \mathbf{u}\|_1 + \tau \|\mathbf{v}\|_\infty + \|\mathbf{z}\|_2^2 \;\middle|\; K(\mathbf{u} + \text{div}\,\mathbf{v}) + \mathbf{z} = \mathbf{f} \right\}$$



target image cartoon texture

Face recognition

Remove shadows and specularities from face images caused by varying illuminations.

这个问题的 Lagrange 函数是

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

增广 Lagrange 函数是

$$\begin{aligned}\mathcal{L}_\beta^3(x,y,z,\lambda) &= \theta_1(x)+\theta_2(y)+\theta_3(z)-\lambda^T(Ax+By+Cz-b)\\ &\quad +\frac{\beta}{2}\|Ax+By+Cz-b\|^2.\end{aligned}$$

直接推广的交替方向法

$$\left\{\begin{aligned} x^{k+1} &= \arg\min\left\{\mathcal{L}_\beta^3(x,y^k,z^k,\lambda^k)\;\middle|\; x\in\mathcal{X}\right\},\\ y^{k+1} &= \arg\min\left\{\mathcal{L}_\beta^3(x^{k+1},y,z^k,\lambda^k)\;\middle|\; y\in\mathcal{Y}\right\},\\ z^{k+1} &= \arg\min\left\{\mathcal{L}_\beta^3(x^{k+1},y^{k+1},z,\lambda^k)\;\middle|\; z\in\mathcal{Z}\right\},\\ \lambda^{k+1} &= \lambda^k-\beta(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b).\end{aligned}\right. \tag{3.2}$$

对 $m\ge 3$, 直接推广的交替方向法不能保证收敛.

3.1 带高斯回代的 ADMM 方法

以 (3.2) 提供的 (y^{k+1}, z^{k+1}) 为预测, 取 $\alpha \in (0,1)$, 校正公式为

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \end{pmatrix} := \begin{pmatrix} y^{k} \\ z^{k} \end{pmatrix} - \alpha \begin{pmatrix} I & -(B^TB)^{-1}B^TC \\ 0 & I \end{pmatrix} \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix}. \quad (3.3)$$

由于为下一步迭代只要准备 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$, 我们只要做

$$\begin{pmatrix} By^{k+1} \\ Cz^{k+1} \end{pmatrix} := \begin{pmatrix} By^{k} \\ Cz^{k} \end{pmatrix} - \alpha \begin{pmatrix} I & -I \\ 0 & I \end{pmatrix} \begin{pmatrix} B(y^k - y^{k+1}) \\ C(z^k - z^{k+1}) \end{pmatrix}.$$

- 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.

对 y 和 z, 有先后, 不公平, 那就要做找补, 调整

3.2 ADMM + Prox-Parallel Splitting ALM

简单地强制 y z 平等不能保证方法收敛

$$\left\{\begin{array}{lcl} x^{k+1} & = & \arg\min\left\{\mathcal{L}_\beta^3(x,y^k,z^k,\lambda^k)\mid x\in\mathcal{X}\right\},\\ y^{k+1} & = & \arg\min\left\{\mathcal{L}_\beta^3(x^{k+1},y,z^k,\lambda^k)\mid y\in\mathcal{Y}\right\},\\ z^{k+1} & = & \arg\min\left\{\mathcal{L}_\beta^3(x^{k+1},y^k,z,\lambda^k)\mid z\in\mathcal{Z}\right\},\\ \lambda^{k+1} & = & \lambda^k-\beta(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b).\end{array}\right.$$

y, z 子问题平行, 如果不想做后处理, 就给它们俩预先都加个正则项

$$\left\{\begin{array}{l} x^{k+1}=\arg\min\left\{\mathcal{L}_\beta^3(x,y^k,z^k,\lambda^k)\mid x\in\mathcal{X}\right\}, \qquad\qquad (\tau>1)\\ y^{k+1}=\arg\min\left\{\mathcal{L}_\beta^3(x^{k+1},y,z^k,\lambda^k)+\frac{\tau}{2}\beta\|B(y-y^k)\|^2\big|y\in\mathcal{Y}\right\},\\ z^{k+1}=\arg\min\left\{\mathcal{L}_\beta^3(x^{k+1},y^k,z,\lambda^k)+\frac{\tau}{2}\beta\|C(z-z^k)\|^2\big|z\in\mathcal{Z}\right\},\\ \lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b).\end{array}\right.$$

上述做法相当于:

$$
\begin{cases}
x^{k+1} = \text{Argmin}\{\theta_1(x) + \frac{\beta}{2}\|Ax + By^k + Cz^k - b - \frac{1}{\beta}\lambda^k\|^2 \,|\, x \in \mathcal{X}\}, \\
\qquad \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b) \\
y^{k+1} = \text{Argmin}\{\theta_2(y) - (\lambda^{k+\frac{1}{2}})^T By + \frac{\mu\beta}{2}\|B(y - y^k)\|^2 \,|\, y \in \mathcal{Y}\}, \\
z^{k+1} = \text{Argmin}\{\theta_3(z) - (\lambda^{k+\frac{1}{2}})^T Cz + \frac{\mu\beta}{2}\|C(z - z^k)\|^2 \,|\, z \in \mathcal{Z}\}, \\
\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b),
\end{cases}
\tag{3.4}
$$

其中 $\mu > 2$. 例如, 可以取 $\mu = 2.01$.

- B. He, M. Tao and X. Yuan, A splitting method for separable convex programming. IMA J. Numerical Analysis, 31(2015), 394-426.

太自由, 又不校正, 就加正则项, 不忘自己昨天的承诺.

This method is accepted by Osher's research group

- E. Esser, M. Möller, S. Osher, G. Sapiro and J. Xin, A convex model for non-negative matrix factorization and dimensionality reduction on physical space, IEEE Trans. Imag. Process., 21(7), 3239-3252, 2012.

IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 21, NO. 7, JULY 2012 3239

A Convex Model for Nonnegative Matrix Factorization and Dimensionality Reduction on Physical Space

Ernie Esser, Michael Möller, Stanley Osher, Guillermo Sapiro, *Senior Member, IEEE*, and Jack Xin

$$\min_{T\geq 0, V_j \in D_j, e \in E} \zeta \sum_i \max_j (T_{i,j}) + \langle R_w \sigma C_w, T \rangle$$

$$\text{such that} \quad YT - X_s = V - X_s \mathrm{diag}(e). \quad (15)$$

Since the convex functional for the extended model (15) is slightly more complicated, it is convenient to use a variant of ADMM that allows the functional to be split into more than two parts. The method proposed by He *et al.* in [34] is appropriate for this application. Again, introduce a new variable Z

Using the ADMM-like method in [34], a saddle point of the augmented Lagrangian can be found by iteratively solving the subproblems with parameters $\delta > 0$ and $\mu > 2$, shown in the

tion refinement step. Due to the different algorithm used to solve the extended model, there is an additional numerical parameter μ, which for this application must be greater than two according to [34]. We set μ equal to 2.01. There are also model parame-

[33] E. Candes, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis,” 2009 [Online]. Available: http://arxiv.org/PS cache/arxiv/pdf/0912/0912.3599v1.pdf

[34] B. He, M. Tao, and X. Yuan, “A splitting method for separate convex programming with linking linear constraints,” Tech. Rep., 2011 [Online]. Available: http://www.optimization-online.org/DB_FILE/2010/06/ 2665.pdf

说明

- 上面提到的方法都与邻近点算法 (PPA) 有关.
- 所有的 (ADMM) 类分裂算法都源于增广 Lagrange 乘子法.
- ADMM 类方法只是对应用中出现一些实际问题有较好的计算效果. 说到底, 只是一阶方法.

References

[1] 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.

[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] C. H. Chen, B. S. He, Y. Y. Ye and X. M. Yuan, *The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent*, to appear in *Mathematical Programming, Series A*.

[5] E. Esser, M. Möller, S. Osher, G. Sapiro and J. Xin, A convex model for non-negative matrix factorization and dimensionality reduction on physical space, IEEE Trans. Imag. Process., 21(7), 3239-3252, 2012.

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

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

[8] D. Hallac, Ch. Wong, S. Diamond, A. Sharang, R. Sosič, S. Boyd and J. Leskovec, SnapVX: A Network-Based Convex Optimization Solver, Journal of Machine Learning Research 18 (2017) 1-5.

[9] B. S. He, H. Liu, Z.R. Wang and X.M. Yuan, A strictly contractive Peaceman-Rachford splitting method for

convex programming, *SIAM Journal on Optimization* **24**(2014), 1011-1040.

[10] 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.

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

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

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

[14] 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.

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

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

[17] A. Nemirovski. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM J. Optim.* **15** (2004), 229–251.

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

[19] 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.

Thank you very much for your attention ！

Thank you very much for reading ！