典型凸优化问题的分裂收缩算法讲座

I. 凸优化及其在变分不等式框架下的邻近点算法

何 炳 生 南京大学数学系

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

江苏省研究生视觉计算与可信人工智能暑期学校

2022年7月10日

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、更新自己学术的体验.

我的几类主要研究工作的分类论文和简要介绍（附阅读建议）

1. 变分不等式的投影收缩算法(Projection-Contraction Methods)　2. 两个可分离函数的乘子交替方向法(ADMM)

3. 多个可分离函数的交替方向类算法(ADMM-Like Methods)　4. 变分不等式框架下的邻近点算法(VI & PPA)

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. 两页纸简述我职业生涯中的主要研究工作

《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（报告3 基本上是英文写成的）

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

1. 从变分不等式的投影收缩算法到凸规划的分裂收缩算法 — 我研究生涯的来龙去脉
2. 生活理念对设计优化分裂算法的帮助 — 以改造 ADMM 求解三个可分离算子问题为例
3. 凸优化的分裂收缩算法 — 变分不等式为工具的统一框架　(适合打印的 综合文本)
4. 从商业谈判的角度看一些优化方法的设计 — 从 min-max 问题的求解谈起
5. 我和乘子交替方向法(ADMM) 的20年 — 2017 年5月全国数学规划会议报告　综述版本
6. 图像处理中的凸优化问题及其相应的分裂收缩算法 — ISICDM会议报告I　报告II　报告III
7. 介绍：构造求解凸优化的分裂收缩算法—用好变分不等式和邻近点算法两大法宝
8. 线性化ALM-ADMM等方法中 的"替代"参数严重影响收敛速度—提升空间有多少?
9. 被S. Becker 誉为 Very Simple yet Powerful 的 Technique — 应用及新的进展
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
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)
12. 均困平衡的增广拉格朗日乘子法 — Balanced ALM (ArXiv: 2018.08554)
13. ADMM 类分裂收缩算法的一些最新进展　统一框架下的Balanced-ALM　便于推广的ADMM

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 $\min\{\sum_{i=1}^p\theta_i(x_i)|\sum_{i=1}^p A_ix_i=b, x_i\in\mathcal{X}_i\}$

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.

这一讲解释上述问题都可以化为一个单调变分不等式 并介绍什么是邻近点算法

凸函数的定义和基本性质

A function $f(x)$ is convex iff

$$f((1-\theta)x+\theta y) \le (1-\theta)f(x)+\theta f(y)$$

$$\forall \theta \in [0,1].$$

Properties of convex function

- $f \in \mathcal{C}^1$. f is convex iff

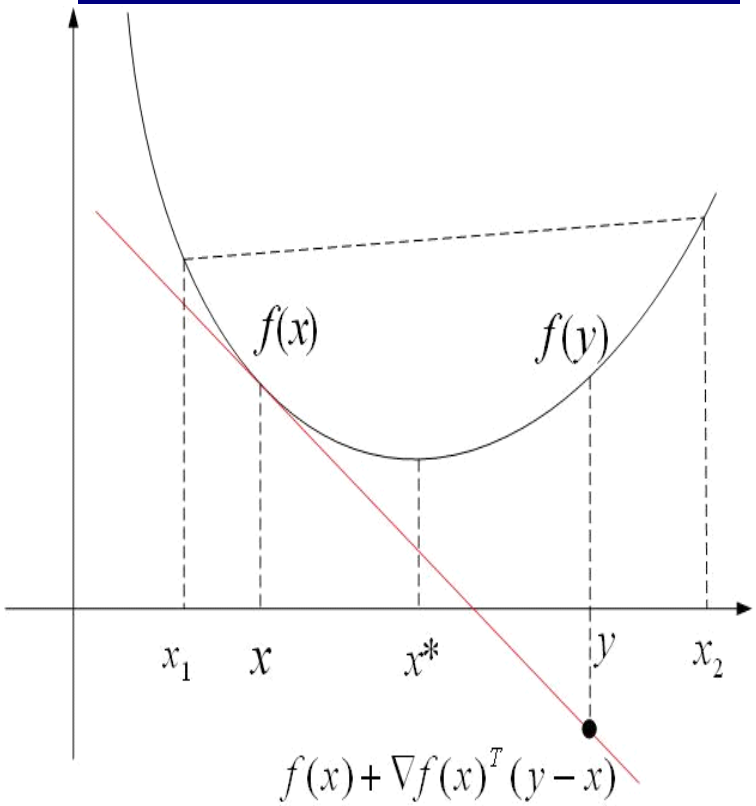
  $$f(y) - f(x) \ge \nabla f(x)^T (y-x).$$

  Thus, we have also

  $$f(x) - f(y) \ge \nabla f(y)^T (x-y).$$

- Adding above two inequalities, we get

  $$(y-x)^T(\nabla f(y) - \nabla f(x)) \ge 0.$$



Convex function

- $f \in \mathcal{C}^1$, ∇f is monotone. $f \in \mathcal{C}^2$, $\nabla^2 f(x)$ is positive semi-definite.
- Any local minimum of a convex function is a global minimum.

多元函数 $f : \Re^n \to \Re$ 的梯度 (Gradient) 和 Hassian

$$\nabla f(x) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{pmatrix}, \qquad \nabla^2 f(x) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \dfrac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

$$\nabla^2 f = \nabla\big((\nabla f)^T\big).$$

$f(x) = c^T x, c$ 是 n-维列向量

$$\nabla(c^T x) = \nabla(x^T c) = c, \qquad \nabla^2(c^T x) = 0_{n\times n}.$$

$f(x) = \frac{1}{2}\|Ax - b\|^2, A \in \Re^{m\times n}, b \in \Re^m$

$$\nabla\Big(\frac{1}{2}\|Ax - b\|^2\Big) = A^T(Ax - b) \in \Re^n, \qquad \nabla^2\Big(\frac{1}{2}\|Ax - b\|^2\Big) = (A^T A)_{n\times n}.$$

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.

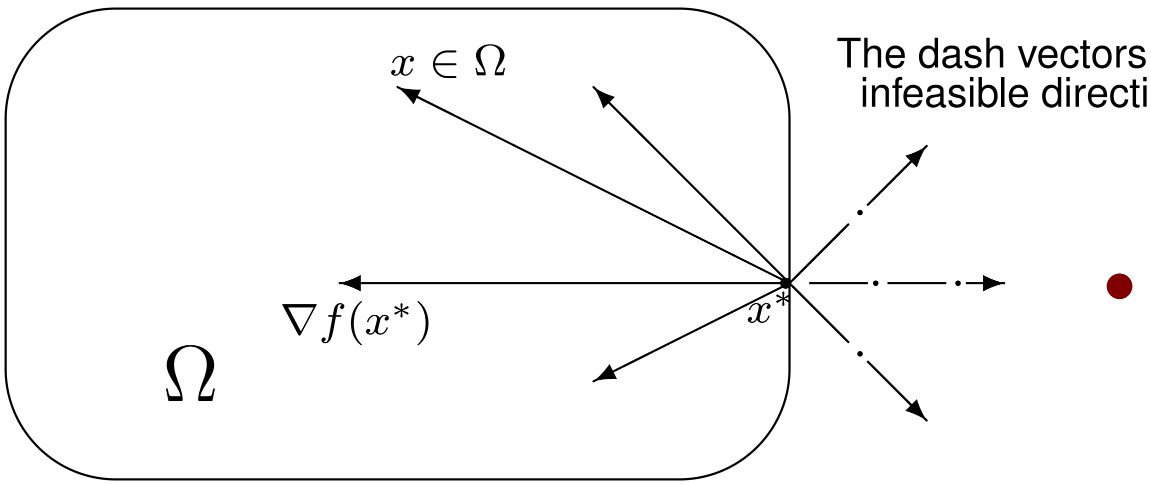


Fig. 1.1 Differential Convex Optimization and VI

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

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

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

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:

Lemma 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 凸优化最优性条件引理

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

引理把优化问题 (1.3a) 转换成了变分不等式 (1.3b). 下面给出证明.

Proof : First, if (1.3a) is true, then for any $x \in \mathcal{X}$, we have

$$\frac{\theta(x_\alpha) - \theta(x^*)}{\alpha} + \frac{f(x_\alpha) - f(x^*)}{\alpha} \geq 0, \tag{1.4}$$

where

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

Because $\theta(\cdot)$ is convex, it follows that

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

and thus

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

Substituting the last inequality in the left hand side of (1.4), we have

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

Using $f(x_\alpha) = f(x^* + \alpha(x - x^*))$ and letting $\alpha \to 0_+$, from the above inequality we get

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

Thus (1.3b) follows from (1.3a). Conversely, since f is convex, it follow that

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

and it can be rewritten as

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

Thus, we have

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

for all $\alpha \in (0, 1]$. Letting $\alpha \to 0_+$, we get

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

Substituting it in the left hand side of (1.3b), we get

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

and (1.3a) is true. The proof is complete. $\Box$

1.2 Linear constrained convex optimization and VI

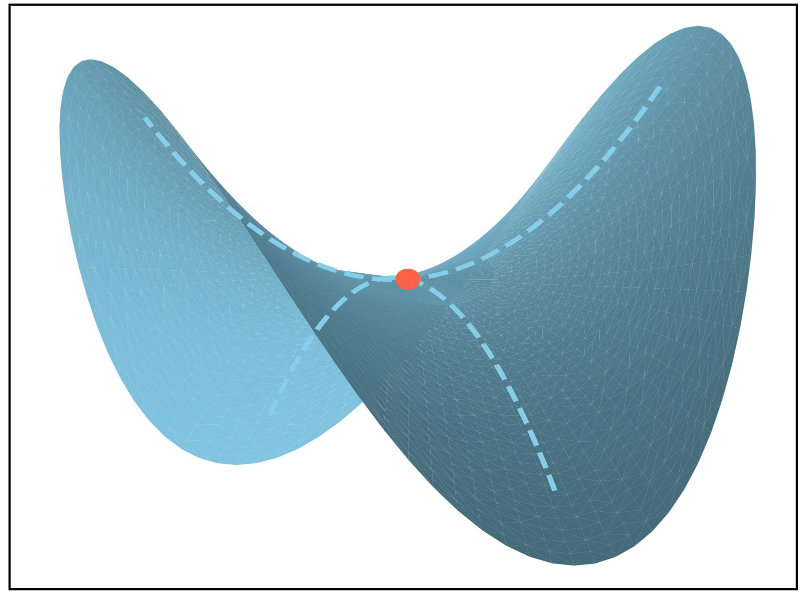
We consider the linearly constrained convex optimization problem

$$\min\{\theta(u) \mid \mathcal{A}u = b,\ u \in \mathcal{U}\}. \qquad (1.5)$$

The Lagrangian function of the problem (1.5) is

$$L(u,\lambda) = \theta(u) - \lambda^T(\mathcal{A}u - b), \qquad (1.6)$$

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



Example 1 of the problem (1.5): Finding the nearest correlation matrix

A positive semi-definite matrix, whose each diagonal element is equal 1, is called the correlation matrix. For given symmetric $n \times n$ matrix C, the mathematical form of finding the nearest correlation matrix X is

$$\min\{\tfrac{1}{2}\|X - C\|_F^2 \mid \mathrm{diag}(X) = e,\ X \in S_+^n\}, \qquad (1.7)$$

where S^n_+ is the positive semi-definite cone and e is a n-vector whose each element is equal 1. The problem (1.7) is a concrete problem of type (1.5).

Example 2 of the problem (1.5): The matrix completion problem

Let M be a given $m \times n$ matrix, Π is the elements indices set of M,

$$\Pi \subset \{(ij) | i \in \{1, \ldots, m\},\ j \in \{1, \ldots, n\}\}.$$

The mathematical form of the matrix completion problem is relaxed to

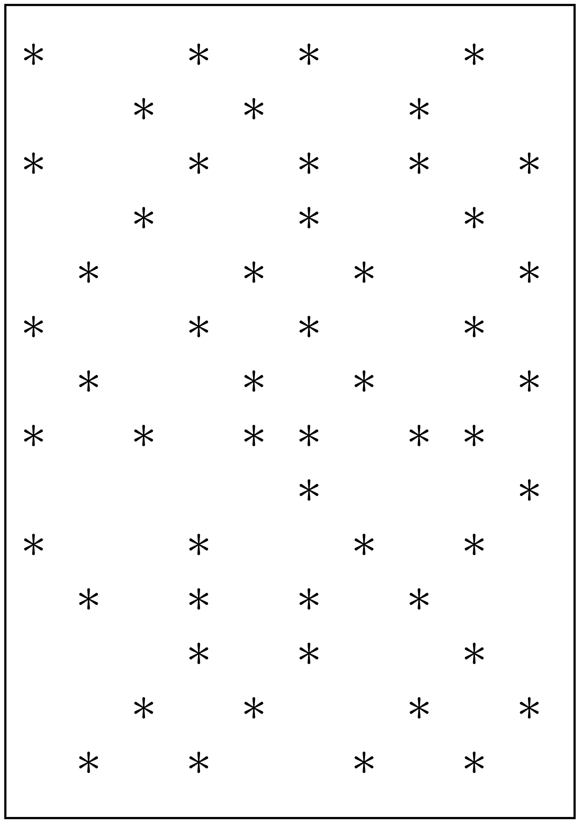
$$\min\{\|X\|_* \mid X_{ij} = M_{ij},\ (ij) \in \Pi\}, \qquad (1.8)$$

where $\|\cdot\|_*$ is the nuclear norm–the sum of the singular values of a given matrix. The problem (1.8) is a convex optimization of form (1.5). The matrix A in (1.5) for the linear constraints

$$X_{ij} = M_{ij},\ (ij) \in \Pi,$$

is a projection matrix, and thus $\|A^T A\| = 1$.

M is low Rank, only some elements of M are known.



A pair of $(u^*, \lambda^*) \in \mathcal{U} \times \Re^m$ is called a saddle point of the Lagrange function (1.6), if

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

The above inequalities can be written as

$$\begin{cases} u^* \in \mathcal{U}, \quad L(u, \lambda^*) - L(u^*, \lambda^*) \ge 0, \;\; \forall\, u \in \mathcal{U}, & \text{(1.9a)} \\ \lambda^* \in \Re^m, \; L(u^*, \lambda^*) - L(u^*, \lambda) \ge 0, \;\; \forall\, \lambda \in \Re^m. & \text{(1.9b)} \end{cases}$$

According to the definition of $L(u, \lambda)$ (see(1.6)),

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

it follows from (1.9a) that

$$u^* \in \mathcal{U}, \;\; \theta(u) - \theta(u^*) + (u - u^*)^T(-\mathcal{A}^T\lambda^*) \ge 0, \;\; \forall\, u \in \mathcal{U}. \quad \text{(1.10)}$$

Similarly, for (1.9b), since

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

thus we have

$$\lambda^* \in \Re^m, \quad (\lambda - \lambda^*)^T(\mathcal{A}u^* - b) \geq 0, \quad \forall\, \lambda \in \Re^m. \tag{1.11}$$

Notice that the expression (1.11) (the inner product of the vector $(\mathcal{A}u^* - b)$ with any vector is nonnegative) is equivalent to

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

Writing (1.10) and (1.11) together, we get the following variational inequality:

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

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

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \qquad \text{(1.12a)}$$

where

$$w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -\mathcal{A}^T\lambda \\ \mathcal{A}u - b \end{pmatrix} \;\text{ and }\; \Omega = \mathcal{U} \times \Re^m. \qquad \text{(1.12b)}$$

Setting $w = (u, \lambda^*)$ and $w = (u^*, \lambda)$ in (1.12), respectively, we get (1.10) and (1.11). Because F is a 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 (1.5), 转换成了混合变分不等式 (1.12).

Two block separable convex optimization

We consider the following structured separable convex optimization

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b,\ x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{1.13}$$

This is a special problem of (1.5) 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 (1.13) 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:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{1.14}$$

where

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \theta(u) = \theta_1(x) + \theta_2(y), \quad w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \qquad \text{(1.15a)}$$

$$F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix}, \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m. \qquad \text{(1.15b)}$$

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 (1.13), 转换成了变分不等式 (1.14)–(1.15).

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 (1.5) 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.

从线性规划看拉格朗日函数、鞍点和变分不等式

设某饲养场要购进一批粮食作为添加饲料.

要求 这批粮食中含有淀粉和蛋白质分别有 2100 千克和 600 千克.

市场 市场上共有 3 种粮食（玉米, 小麦, 大豆） $(j = 1, 2, 3)$

可供选择，第 j 种粮食每公斤的价格为 c_j 元, 采购量为 x_j.

第 j 种粮食含 i 种营养素的量为 a_{ij} （含量比）.

含量	玉米 x_1	小麦 x_2	大豆 x_3	总需求量
淀　粉	$a_{11} = 0.50$	$a_{12} = 0.50$	$a_{13} = 0.20$	2100
蛋白质	$a_{21} = 0.10$	$a_{22} = 0.12$	$a_{23} = 0.40$	600
单价	$c_1 = 3$	$c_2 = 4$	$c_3 = 8.4$	

饲养场满足营养要求的最小支出采购计划是线性规划问题:

$$\begin{aligned} \min \quad & 3x_1 + 4x_2 + 8.4x_3 \\ \text{s. t.} \quad & 0.50x_1 + 0.50x_2 + 0.20x_3 = 2100, \\ & 0.10x_1 + 0.12x_2 + 0.40x_3 = 600, \\ & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0. \end{aligned} \tag{1.16}$$

利用记号

$$A = \begin{pmatrix} 0.50 & 0.50 & 0.20 \\ 0.10 & 0.12 & 0.40 \end{pmatrix}, \quad b = \begin{pmatrix} 2100 \\ 600 \end{pmatrix}, \quad c = \begin{pmatrix} 3 \\ 4 \\ 8.4 \end{pmatrix}. \tag{1.17}$$

可以写成线性规划问题

$$\begin{aligned} \min \quad & c^T x \\ & Ax = b \\ & x \geq 0. \end{aligned} \tag{1.18}$$

设有一家（虚拟的）公司, 它能供应饲养场所需的淀粉和蛋白质

为了推销产品, 什么样的广告词最有说服力？公司营销员说：

> 买了我们的淀粉和蛋白质去配制含同量营养成分的玉米, 小麦和大豆,
> 配制价不高于市场上玉米, 小麦和大豆的价格. 饲养场主尽管放心！

为了最大赢利, 营养素公司该怎样去确定营养素的价格？

设营养素公司对这 2 种营养素将设定的价格分别为 y_1, y_2.

含量	玉米	小麦	大豆	价格
淀　粉	$a_{11} = 0.50$	$a_{12} = 0.50$	$a_{13} = 0.20$	y_1
蛋白质	$a_{21} = 0.10$	$a_{22} = 0.12$	$a_{23} = 0.40$	y_2
单价	$c_1 = 3.0$	$c_2 = 4.0$	$c_3 = 8.4$	

营养素公司实现承诺的前提下实现最大盈利的价格方案是线性规划问题

$$
\begin{array}{llll}
\max & 2100y_1 + 600y_2 & & \\
\text{配制玉米} & 0.50y_1 + 0.10y_2 & \le & 3.0 \\
\text{配制小麦} & 0.50y_1 + 0.12y_2 & \le & 4.0 \\
\text{配制大豆} & 0.20y_1 + 0.40y_2 & \le & 8.4
\end{array}
$$

利用记号 (1.17), 就有

$$
\begin{array}{ll}
\max & b^T y \\
 & A^T y \le c.
\end{array}
\qquad (1.19)
$$

买方要实现采购要求, 卖方会实现价格承诺

$$x^T(c - A^T y) \ge 0, \quad x^T c - x^T A^T y \ge 0, \quad c^T x - y^T A x \ge 0.$$

$$c^T x \ge y^T A x = y^T b = b^T y.$$

买方要化的钱 $\ge$ 卖方能挣的钱. $c^T x \ge b^T y$

买方要化的钱 $=$ 卖方能挣的钱. $c^T x^* = b^T y^*$

买卖双方都达到了最优解! $(x^*)^T(c - A^T y^*) = 0$

对任意的 $x \geq 0$, 由于 $c - A^T y^* \geq 0$, 所以 $\underline{x^T(c - A^T y^*)} \geq 0$.

又由于 $(x^*)^T(c - A^T y^*) = 0$, 所以 $(x - x^*)^T(c - A^T y^*) \geq 0$. 因此

$$\begin{cases} x^* \geq 0, \quad c^T x - c^T x^* + (x - x^*)^T(-A^T y^*) \geq 0, \quad \forall x \geq 0, \\ Ax^* - b = 0. \end{cases} \tag{1.20}$$

含量	玉米 4000	小麦 0	大豆 500	总需求量
淀　粉	$a_{11} = 0.50$	$a_{12} = 0.50$	$a_{13} = 0.20$	2100
蛋白质	$a_{21} = 0.10$	$a_{22} = 0.12$	$a_{23} = 0.40$	600
单价	$c_1 = 3$	$c_2 = 4$	$c_3 = 8.4$	

$$\boxed{4000 \times 3.0 + 500 \times 8.4 = 12000 + 4200 = 16200}$$

含量	玉米	小麦	大豆	价格
淀　粉	$a_{11}=0.50$	$a_{12}=0.50$	$a_{13}=0.20$	$y_1=2$
蛋白质	$a_{21}=0.10$	$a_{22}=0.12$	$a_{23}=0.40$	$y_2=20$
单价	$c_1=3.0$	$c_2=4.0$	$c_3=8.4$	

$$\boxed{2100\times 2+600\times 20=4200+12000=16200}$$

满足 $c^Tx^*=b^Ty^*$ 的 $x^*=\begin{pmatrix}4000\\0\\500\end{pmatrix}$ 和 $y^*=\begin{pmatrix}2\\20\end{pmatrix}$ 分别是原始和对偶问题的最优解.

线性规划问题的 Lagrange 函数是定义在 $\Re^n_+\times\Re^m$ 上的

$$L(x,y)=c^Tx-y^T(Ax-b).$$

如果 $x^*\in\Re^n_+$ 和 $y^*\in\Re^m$ 满足

$$L_{y\in\Re^m}(x^*,y)\le L(x^*,y^*)\le L_{x\in\Re^n_+}(x,y^*),$$

则称 (x^*,y^*) 是 Lagrange 函数 $L(x,y)$ 在 $\Re^n_+\times\Re^m$ 上的一个鞍点.

Lagrange 乘子的意义就是“影子价格”

把鞍点的这两个不等式写出来就是

$$\begin{cases} x^* \in \Re^m_+, & L(x,y^*) - L(x^*,y^*) \ge 0, & \forall x \in \Re^m_+; \\ y^* \in \Re^m, & L(x^*,y^*) - L(x^*,y) \ge 0, & \forall y \in \Re^m. \end{cases}$$

把 $L(x,y)$ 的具体形式放进去, 得到

$$\begin{cases} x^* \in \Re^m_+, & c^Tx - c^Tx^* + (x-x^*)^T(-A^Ty^*) \ge 0, & \forall x \in \Re^m_+; \\ y^* \in \Re^m, & (y-y^*)^T(Ax^*-b) \ge 0, & \forall y \in \Re^m. \end{cases}$$

这就是变分不等式 (1.12) 的一个具体形式.

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w-w^*)^T F(w^*) \ge 0, \quad \forall w \in \Omega,$$

其中

$$w = \begin{pmatrix} x \\ y \end{pmatrix}, \qquad F(w) = \begin{pmatrix} -A^Ty \\ Ax-b \end{pmatrix}, \qquad \Omega = \Re^n_+ \times \Re^m.$$

我们用线性规划讲述了优化问题拉格朗日函数的鞍点和变分不等式解点的等价关系.

2 Proximal point algorithms and its Beyond

Lemma 2 *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) [8, 10] begins with a given x^k, offers the new iterate x^{k+1} via the recursion

邻近点算法

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

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

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

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,

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

Let $a = x^k - x^*$ and $b = x^{k+1} - x^*$ and using Lemma 2, we obtain

PPA 算法的收缩性质

$$\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\|^2 \leq \|\boldsymbol{x}^k - \boldsymbol{x}^*\|^2 - \|\boldsymbol{x}^k - \boldsymbol{x}^{k+1}\|^2, \quad (2.6)$$

which is the nice convergence property of Proximal Point Algorithm.

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

Proof. Replacing $k+1$ in (2.4) with k, we get

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

Let $x = x^{k+1}$ in the above inequality, it follows that

$$\theta(x^{k+1}) - \theta(x^k) + (x^{k+1} - x^k)^T \{\nabla f(x^k) + r(x^k - x^{k-1})\} \ge 0. \qquad (2.7)$$

Setting $x = x^k$ in (2.4), we become

$$\theta(x^k) - \theta(x^{k+1}) + (x^k - x^{k+1})^T \{\nabla f(x^{k+1}) + r(x^{k+1} - x^k)\} \ge 0. \qquad (2.8)$$

Adding (2.7) and (2.8) and using $(x^k - x^{k+1})^T [\nabla f(x^k) - \nabla f(x^{k+1})] \ge 0$, we get

$$(x^k - x^{k+1})^T \{(x^{k-1} - x^k) - (x^k - x^{k+1})\} \ge 0. \qquad (2.9)$$

Setting $a = x^{k-1} - x^k$ and $b = x^k - x^{k+1}$ in (2.9) and using (2.1), we obtain

$$\|\boldsymbol{x}^k - \boldsymbol{x}^{k+1}\|^2 \le \|\boldsymbol{x}^{k-1} - \boldsymbol{x}^k\|^2 - \|(\boldsymbol{x}^{k-1} - \boldsymbol{x}^k) - (\boldsymbol{x}^k - \boldsymbol{x}^{k+1})\|^2. \qquad (2.10)$$

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

For the optimization problem (2.2) , namely, $\min\{\theta(x) + f(x) \mid x \in \mathcal{X}\}$,

the equivalent variational inequality form is

$$x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall x \in \mathcal{X}. \tag{2.11}$$

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:

$$\begin{aligned} x^{k+1} \in \mathcal{X}, \quad \theta(x) - \theta(x^{k+1}) &+ (x - x^{k+1})^T \nabla f(x^{k+1}) \\ &\geq (x - x^{k+1})^T r(x^k - x^{k+1}), \quad \forall x \in \mathcal{X}. \end{aligned} \tag{2.12}$$

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 (2.12) 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: 变分不等式

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega. \tag{2.13}$$

PPA for VI (2.13) **in H-norm (定义)** For given w^k and $H \succ 0$, find w^{k+1},

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

邻近点算法

$$\ge (w - w^{k+1})^T H(w^k - w^{k+1}), \quad \forall\, w \in \Omega, \tag{2.14}$$

w^{k+1} is called the proximal point of the k-th iteration for the problem (2.13).

(2.14) 是求解 VI (2.13) 的 PPA 算法的定义. 第二讲就会用例子说明这是容易做到的.

✠ w^{k+1} **is the solution of** (2.13) **if and only if** $w^k = w^{k+1}$ ✠

Setting $w = w^*$ in (2.14), 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 (1.12b))

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

and consequently (by using (2.13)) we obtain

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

Thus, we have

$$(w^{k+1} - w^*)^T H(w^k - w^{k+1}) \geq 0. \tag{2.15}$$

By setting $\boldsymbol{a = w^k - w^*}$ and $\boldsymbol{b = w^{k+1} - w^*}$,
the inequality (2.15) means that $\boldsymbol{b^T H(a - b) \geq 0}$.

By using Lemma 2, we obtain

$$\boldsymbol{\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2}. \tag{2.16}$$

We get the nice convergence property of Proximal Point Algorithm.

2.3 Variants of PPA for Variational Inequalities

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

PPA for VI (2.13) **in H-norm** For given v^k and $H \succ 0$, find w^{k+1},

$$w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \\ \geq (v - v^{k+1})^T H(v^k - v^{k+1}), \quad \forall\, w \in \Omega, \qquad (2.17)$$

w^{k+1} is called the proximal point of the k-th iteration for the problem (2.13).

✠ w^{k+1} **is the solution of** (2.13) **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 (2.17), 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 (1.12b))

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

and consequently (by using (2.13)) 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

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \geq 0. \tag{2.18}$$

By using Lemma 2, we obtain

$$\|\boldsymbol{v^{k+1} - v^*}\|_{\boldsymbol{H}}^{\boldsymbol{2}} \leq \|\boldsymbol{v^k - v^*}\|_{\boldsymbol{H}}^{\boldsymbol{2}} - \|\boldsymbol{v^k - v^{k+1}}\|_{\boldsymbol{H}}^{\boldsymbol{2}}. \tag{2.19}$$

We get the nice convergence property of Proximal Point Algorithm.

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

$$\|v^k - v^{k+1}\|_H^2 \leq \|v^{k-1} - v^k\|_H^2 - \|(v^{k-1} - v^k) - (v^k - v^{k+1})\|_H^2.$$

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** [7, 9] 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,

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

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

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

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

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

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

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

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

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

This is the PPA form (2.17) in which

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

The related contraction inequality (2.19) 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 The relaxed PPA （延伸的邻近点算法）

We shall maintain our focus on the monotone variational inequality (2.13), namely,

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

The PPA form (2.17) reads as

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

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

$$\begin{aligned} \tilde{w}^k \in \Omega, \;\; \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) & \\ \geq (v - \tilde{v}^k)^T H(v^k - \tilde{v}^k), \;\; \forall\, w \in \Omega. & \end{aligned} \tag{4.1}$$

Setting $w = w^*$ in (4.1), we obtain

$$(\tilde{v}^k - v^*)^T H(v^k - \tilde{v}^k) \geq \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k). \tag{4.2}$$

Applying (see (1.12b)) the identity

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

to (4.2), we obtain

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

Because RHS of the above inequality is , we have

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

We write it as

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

and thus

$$(v^k - v^*)^T H(v^k - \tilde{v}^k) \geq \|v^k - \tilde{v}^k\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \qquad (4.3)$$

The inequality (4.3) 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.

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

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

Set

$$v^{k+1}(\alpha) = v^k - \alpha(v^k - \tilde{v}^k) \tag{4.4}$$

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

$$\vartheta(\alpha) = \|v^k - v^*\|_H^2 - \|v^{k+1}(\alpha) - v^*\|_H^2. \tag{4.5}$$

Note that

$$\begin{aligned} \vartheta(\alpha) &= \|v^k - v^*\|_H^2 - \|(v^k - v^*) - \alpha(v^k - \tilde{v}^k)\|_H^2 \\ &= 2\alpha(v^k - v^*)^T H(v^k - \tilde{v}^k) - \alpha^2\|v^k - \tilde{v}^k\|_H^2 \end{aligned} \tag{4.6}$$

is a quadratic function of α.

We can not directly maximize $\vartheta(\alpha)$ in (4.6) because the coefficient of the linear term $2(v^k - v^*)^T H(v^k - \tilde{v}^k)$ contains the unknown solution v^*.

Using (4.3), from (4.6) we get

$$\vartheta(\alpha) \geq 2\alpha\|v^k - \tilde{v}^k\|_H^2 - \alpha^2\|v^k - \tilde{v}^k\|_H^2 \tag{4.7}$$

Set

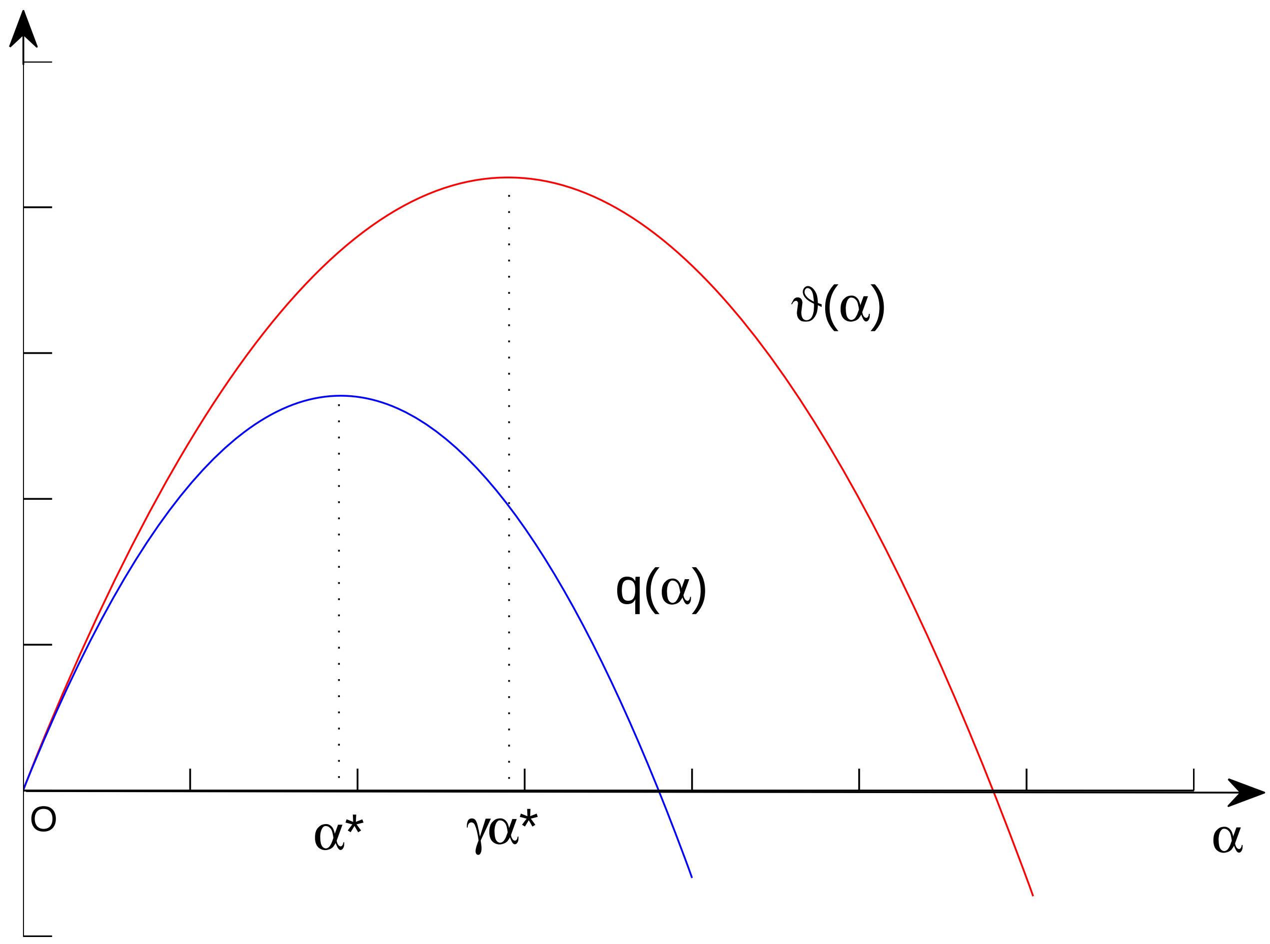
$$q(\alpha) = (2\alpha - \alpha^2)\|v^k - \tilde{v}^k\|_H^2, \tag{4.8}$$

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

$$v^{k+1} = v^k - \gamma(v^k - \tilde{v}^k), \quad \gamma \in (0, 2) \tag{4.9}$$

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

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|^2 - \gamma(2-\gamma)\|v^k - \tilde{v}^k\|_H^2. \tag{4.10}$$



取 $\gamma \in [1, 2)$ 的示意图

这一讲是预备知识. 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.

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

[4] 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.

[5] 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

[6] 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.

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

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

[9] 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.

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