凸优化和单调变分不等式的收缩算法
—— 统一框架与应用

前言目录与各章提要

南京大学数学系　　何炳生
hebma@nju.edu.cn

一个科学家最大的本领就在于化复杂为简单，用简单的方法去解决复杂的问题。 —— 冯康

✠ 先贤名言，铭记在心。纵然可以没有本领，也不迷失价值标准 ✠

爱美之心，人皆有之。数学往往被人认为是枯燥的，计算更是被人看作是繁琐的。领悟了数学之美的数学工作者，他的职业生涯才可能是充满乐趣的。

✠ 世上三百六十行，应是同一道理 ✠

数学之美，不是纯数学的专利。为应用服务的最优化方法研究，同样可以追求简单与统一。只因自以为发现了其中的数学之美，研究才变得满怀激情和欲罢不能。

♣ 对凸规划和单调变分不等式的收缩算法的研究, 我们始终追求简明统一的原则。一个统一框架, 二个孪生方向, 三个基本不等式。不同方向, 竟可用相同步长。漂亮的形式, 令我沾沾自喜。但也常问自己, 或以平庸为神奇？

♣ 统一框架, 指导我们针对问题设计算法, 也帮助我们简化算法的收敛性证明。理论有了保证, 还要计算验证。尺有所短, 寸有所长。自己动手编程, 多些计算体验, 才能给初学者提些有价值的参考意见。

♣ 投影收缩算法的定义在 Springer 出版社1975 年出版的 Blum 和 Oettli 的德文专著中就有了说明, 作者均是苏黎世大学的博士, Oettli 生前曾是 Math. Programming 的编委。出于对历史的敬畏和对前人工作的尊重, 根据算法具有的收缩特征, 我们把所述方法都称为收缩方法。

♣ 从研究出发点开始, 介绍收缩算法的基本原理。分若干篇章, 对问题类型和主要方法作了梳理。谈问题, 说算法, 讲应用, 给基本算例, 也附简单程序。一条主线, 一个模式。读懂一个算法, 理解后面的篇章就不要再费多少力气。

One algorithm framework should be flexible enough to solve many problems !

凸优化和单调变分不等式的收缩算法 — 目录

第一部分: 单调变分不等式的求解方法

第二部分: 凸优化问题 $\{\min f(x) | Ax = b, x \in X\}$ 的求解方法

第三部分: 基于投影梯度的收缩算法

第八讲 基于梯度投影的凸优化收缩算法和下降算法

第九讲 * 线性约束凸优化基于对偶上升的自适应方法

第十讲 * 线性约束单调变分不等式的自适应投影收缩算法

第四部分: 凸优化问题 $\min\left\{\theta_1(x)+\theta_2(y)\,\middle|\,\begin{array}{l}Ax+By=b\\ x\in X, y\in Y\end{array}\right\}$ 的交替方向法

第十一讲 结构型优化的交替方向收缩算法

第十二讲 * 线性化的交替方向收缩算法

第十三讲 定制 PPA 算法意义下的交替方向法

第十四讲 * 定制 PPA 算法意义的线性化交替方向法

第五部分: 多个可分离算子凸优化问题带简单校正的分裂方法

第十五讲 三个可分离算子凸优化的平行分裂 ALM 乘子法

第十六讲 三个可分离算子凸优化的略有改动的交替方向法

第十七讲 * 多个可分离算子凸优化带回代的交替方向收缩算法

第十八讲 * 多个可分离算子凸优化带回代的线性化交替方向法

第六部分: 计算复杂性分析

第十九讲 L-连续的单调变分不等式投影收缩算法的收敛速率

第二十讲 交替方向法的计算复杂性和收敛速率

可只读不带 * 的篇章. 分解降低难度, 整合把握方向, 是该系列讲义的主要思想. 懂些变分不等式的基本概念, 对凸优化收缩算法的设计和收敛性分析很有帮助.

凸优化和单调变分不等式的收缩算法 — 内容提要

1. 变分不等式作为多种问题的统一表述模式

变分不等式的一般形式是: 求

$$u^* \in \Omega, \quad (u - u^*)^T F(u^*) \ge 0, \quad \forall\, u \in \Omega.$$

应用科学中, 特别是统计与机器学习中的许多问题都可以归结为形如

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

或者

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

的凸优化问题。它们的一阶必要性条件就是一个单调变分不等式, 其中

$$F(u) = \begin{pmatrix} \nabla\theta(x) - A^T\lambda \\ Ax - b \end{pmatrix} \quad \text{或} \quad F(u) = \begin{pmatrix} \nabla\theta_1(x) - A^T\lambda \\ \nabla\theta_2(y) - B^T\lambda \\ Ax + By - b \end{pmatrix}.$$

相应的 $\Omega = \mathcal{X} \times \Re^m$ 或者 $\Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m$. 在凸优化算法设计和收敛性证明中利用梯度表述的最优性条件, 就像一元函数利用导数求极值, 会带来很大的方便。经济活动中的平衡问题, 保护资源、保障供给中的调控手段, 用经济手段解决交通疏导等问题, 都可以用变分不等式来描述。

2. 三个基本不等式和变分不等式的投影收缩算法

变分不等式等价于求 $e(u) := u - P_{\Omega}[u - F(u)]$ 的一个零点. 单调变分不等式的解集 Ω^* 是闭凸集. 设 $u^* \in \Omega^*$ 是变分不等式的一个解点. 由

$$\tilde{u} = P_{\Omega}[u - F(u)]$$

给出 $\tilde{u}$. 根据变分不等式解的性质、投影以及单调算子的基本性质得出

$$\begin{array}{ll} \text{(FI1)} & (\tilde{u} - u^*)^T F(u^*) \ge 0, \\ \text{(FI2)} & (\tilde{u} - u^*)^T \big((u - \tilde{u}) - F(u)\big) \ge 0, \\ \text{(FI3)} & (\tilde{u} - u^*)^T (F(\tilde{u}) - F(u^*)) \ge 0 \end{array}$$

三个基本不等式. 由基本不等式的不同组合导出方向 $d(u, \tilde{u})$ 满足

$$(u - u^*)^T d(u, \tilde{u}) \ge \varphi(u, \tilde{u}) \quad \text{并且有} \quad \varphi(u, \tilde{u}) \ge \delta \|e(u)\|^2, \quad \delta > 0.$$

基本不等式的组合提供了距离函数 $\|u - u^*\|^2$ 的下降方向 $-d(u, \tilde{u})$. 对典型的线性和非线性变分不等式, 构造到解集距离越来越近的收缩算法. 由于 $\tilde{u}$ 是由投影得来的, 这类算法称为投影收缩算法. 算例说明, 投影收缩算法效率比花费相当的外梯度算法 (Extragradient Methods) 有明显的提高。

3. 单调变分不等式收缩算法的统一框架

对给定的 u, 由一定的法则生成 $\tilde{u}\in\Omega$, 如果 $\tilde{u}=u \;\Leftrightarrow\; u\in\Omega^*$, 这样的 $\tilde{u}$ 称作检验点或者预测点。生成检验点 $\tilde{u}$ 的公式往往能写成

$$\tilde{u}\in\Omega,\quad (u'-\tilde{u})^T\{d_2(u,\tilde{u})-d_1(u,\tilde{u})\}\ge 0,\quad \forall\, u'\in\Omega.$$

或其等价形式 $\tilde{u}=P_\Omega\{\tilde{u}-[d_2(u,\tilde{u})-d_1(u,\tilde{u})]\}$.

由投影 $\tilde{u}=P_\Omega[u-\beta F(u)]$ 得到的检验点, 可以写成 $\tilde{u}=P_\Omega\{\tilde{u}-[d_2(u,\tilde{u})-d_1(u,\tilde{u})]\}$ 其中 $d_1(u,\tilde{u})=(u-\tilde{u})-\beta[(F(u)-F(\tilde{u}))]$ 和 $d_2(u,\tilde{u})=\beta F(\tilde{u})$.

这样的 $d_1(u,\tilde{u})$ 和 $d_2(u,\tilde{u})$ 称为一对孪生方向. 不难找到

$$(\tilde{u}-u^*)^T d_2(u,\tilde{u})+(u-\tilde{u})^T d_1(u,\tilde{u})$$

的下界函数 $\varphi(u,\tilde{u})\ge\delta\|u-\tilde{u}\|^2$ 满足 $\varphi(u,\tilde{u})=0 \;\Leftrightarrow\; u=\tilde{u}$. 例如：

PPA 算法从 u 得到的 $\tilde{u}$ 满足 $\tilde{u}\in\Omega,\ (u'-\tilde{u})^T\{F(\tilde{u})+(\tilde{u}-u)\}\ge 0,\ \forall\, u'\in\Omega$. $\tilde{u}$ 就是一个检验点. 记 $d_1(u,\tilde{u})=u-\tilde{u},\ d_2(u,\tilde{u})=F(\tilde{u})$ 和 $\varphi(u,\tilde{u})=\|u-\tilde{u}\|^2$ 就有 $\delta=1$.

由此对任何解点 u^*, 得到二个几乎相同的不等式：

$$(u-u^*)^T d_1(u,\tilde{u})\ge\varphi(u,\tilde{u}) \quad \text{和} \quad (u-u^*)^T d_2(u,\tilde{u})\ge\varphi(u,\tilde{u}).$$

不同的只是后者只对属于 Ω 的 u 成立. 用相同的步长 α, 不同的迭代公式

$$u^{k+1}=u^k-\alpha d_1(u^k,\tilde{u}^k) \quad \text{和} \quad u^{k+1}=P_\Omega[u^k-\alpha d_2(u^k,\tilde{u}^k)]$$

产生的序列 $\{u^k\}$ 有同样的理论性质. 算例说明了不同算法的效率差别。

4. 为线性约束凸优化定制的 PPA 算法及其应用

在 $\min\{\theta(x)+\frac{r}{2}\|x-a\|^2 \,|\, x\in\mathcal{X}\}$ 容易求解的假设下, 对凸优化问题

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

等价的变分不等式 **VI**(Ω,F) 给出了一个 G-模下的 PPA 算法, 这里

$$u=\begin{pmatrix} x\\ \lambda\end{pmatrix},\qquad F(u)=\begin{pmatrix}\nabla\theta(x)-A^T\lambda\\ Ax-b\end{pmatrix},\qquad \Omega=\mathcal{X}\times\Re^m.$$

算法的每步迭代从当前点 $u^k=(x^k,\lambda^k)$出发, 由

$$\tilde\lambda^k=\lambda^k-\tfrac{1}{s}(Ax^k-b)$$

给出 Dual 预测点$\tilde\lambda^k$. 然后, 通过求解 “简单问题”

$$\min\Big\{\theta(x)+\tfrac{r}{2}\big\|x-\big[x^k+\tfrac{1}{r}A^T(2\tilde\lambda^k-\lambda^k)\big]\big\|^2 \,\Big|\, x\in\mathcal{X}\Big\}$$

得到 Primal 预测点 $\tilde x^k$. 在 $rs>\|A^TA\|$ 的假设下, 生成的预测点 $\tilde u^k$ 满足

$$\tilde u^k\in\Omega,\quad (u-\tilde u^k)^T\{F(\tilde u^k)+G(\tilde u^k-u^k)\}\ge 0,\ \forall\, u\in\Omega,$$

其中 G 是正定矩阵. 这是 G-模下 PPA 算法的基本框架. 用

$$u^{k+1}=u^k-\gamma(u^k-\tilde u^k),\quad \gamma\in(0,2)$$ 产生的序列 $\{u^k\}$ 就满足

$$\|u^{k+1}-u^*\|_G^2\le\|u^k-u^*\|_G^2-\gamma(2-\gamma)\|u^k-\tilde u^k\|_G^2$$

这样的 PPA 收缩性质. 给出了典型算例, 说明方法简单, 效果不错。

5. 线性约束凸优化问题基于松弛 PPA 的收缩算法

在 $\min\{\theta(x)+\frac{r}{2}\|x-a\|^2 \mid x\in\mathcal{X}\}$ 容易求解的假设下, 对凸优化问题

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

等价的变分不等式 **VI**(Ω,F) 给出了一个基于松弛 PPA 的收投影收缩算法. 对给定的 $u^k=(x^k,\lambda^k)$ 和 $r>0$, 通过求解

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

得到 Primal 预测点 $\tilde{x}^k$. 再用自调比法则选取适当的 s, 并用

$$\tilde{\lambda}^k=P_\Lambda\{\lambda^k-\tfrac{1}{s}(A\tilde{x}^k-b)\}$$

生成 Dual 预测点 $\tilde{\lambda}^k$. 在一定条件下, 这样得到的预测点 $\tilde{u}^k$ 满足

$$\tilde{u}^k\in\Omega,\quad (u-\tilde{u}^k)^T\{F(\tilde{u}^k)-Hd(u^k,\tilde{u}^k)\}\geq 0,\quad \forall\, u\in\Omega,$$

其中 H 是正定矩阵. 这是统一框架中的基本形式. 用收缩算法的步长准则确定 α_k, 由迭代公式 $\quad u^{k+1}=u^k-\alpha_k d(u^k,\tilde{u}^k) \quad$ 产生的序列 $\{u^k\}$ 具有 H-模下的收缩性质 ($c_0>0$ 为常数)

$$\|u^{k+1}-u^*\|_H^2\leq\|u^k-u^*\|_H^2-c_0\|u^k-\tilde{u}^k\|_H^2.$$

与前一讲的 PPA 算法比, 需多算一个步长, 但计算实践说明总花费更少。

6. 线性约束凸优化扩展问题的 PPA 和松弛 PPA 收缩算法

还是在问题 $\min\{\theta(x)+\frac{r}{2}\|x-a\|^2\,|\,x\in\mathcal{X}\}$ 容易求解的假设下, 对凸优化扩展问题

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

等价的变分不等式给出类似于前二讲的收缩算法。与前二讲的方法相比, 只是生成预测点的 Dual 部分 $\tilde{\lambda}^k$ 的公式有了改变。如果采用PPA 根据求 $\tilde{x}^k$ 和 $\tilde{\lambda}^k$ 的不同顺序, 分别由

$$\tilde{\lambda}^k=\frac{1}{s}\left\{P_{\mathcal{B}}[Ax^k-s\lambda^k]-(Ax^k-s\lambda^k)\right\}$$

和

$$\tilde{\lambda}^k=\frac{1}{s}\left\{P_{\mathcal{B}}[A(2\tilde{x}^k-x^k)-s\lambda^k]-[A(2\tilde{x}^k-x^k)-s\lambda^k]\right\}$$

给出 $\tilde{\lambda}^k$. 在其他公式全部相同的情况下, 得到相应的求解扩展问题的 PPA 方法. 当这里的 $\mathcal{B}$ 退化成 $\mathcal{B}=\{b\}$ 时, 上面两个生成 Dual 预测点 $\tilde{\lambda}^k$ 的公式就分别变成前两讲中的

$$\tilde{\lambda}^k=\lambda^k-\frac{1}{s}(Ax^k-b)\quad\text{和}\quad\tilde{\lambda}^k=\lambda^k-\frac{1}{s}(A(2\tilde{x}^k-x^k)-b).$$

按照同样的思路, 也可以将松弛 PPA 算法改造后用来求解扩展问题。

7. 基于增广 Lagrange 乘子法的 PPA 收缩算法

凸优化问题 $\min\{\theta(x) \mid Ax = b,\ x \in \mathcal{X}\}$ 的增广 Lagrange 函数

$$\mathcal{L}_A(x,\lambda) = \theta(x) - \lambda^T(Ax-b) + \tfrac{1}{2s}\|Ax-b\|^2,$$

多加了二次函数 $\frac{1}{2s}\|Ax-b\|^2$. 增广 Lagrange 乘子法中 x 只是中间变量. 对给定的 λ^k, 通过求解 x 的子问题 $\min\{\mathcal{L}_A(x,\lambda^k)|x \in \mathcal{X}\}$ 得 $\tilde{x}^k$, 再用

$$\lambda^{k+1} = \lambda^k - \tfrac{\gamma}{s}(A\tilde{x}^k - b), \quad \gamma \in (0,2)$$

产生新的迭代点 λ^{k+1}, 迭代序列 $\{\lambda^k\}$ 满足

$$\|\lambda^{k+1} - \lambda^*\|^2 \le \|\lambda^k - \lambda^*\|^2 - \gamma(2-\gamma)\|\tfrac{1}{s}(A\tilde{x}^k - b)\|^2.$$

在只有形如 $\min\{\theta(x) + \frac{r}{2}\|x-a\|^2 \mid x \in \mathcal{X}\}$ 的问题才容易求解的假设下, 对增广 Lagrange 函数的二次部分 $\frac{1}{2s}\|Ax-b\|^2$ 做线性化近似, 并加上正则项 $\frac{r}{2}\|x-x^k\|^2$. 这样迭代就要从 $u^k = (x^k, \lambda^k)$ 出发, 通过求解

$$\min\{\theta(x) + \tfrac{1}{s}((Ax^k-b) - s\lambda^k)^T Ax + \tfrac{r}{2}\|x-x^k\|^2 \mid x \in \mathcal{X}\}$$

得到 Primal 预测点 $\tilde{x}^k$. 再由 $\tilde{\lambda}^k = \lambda^k - \frac{1}{\tau}(A\tilde{x}^k - b)$ 给出 Dual 预测点 $\tilde{\lambda}^k$. 当 $rs > \|A^TA\|$ 时, 生成的预测点 $\tilde{u}^k$ 满足统一框架中 PPA 算法的条件:

$$\tilde{u}^k \in \Omega, \quad (u - \tilde{u}^k)^T\{F(\tilde{u}^k) + G(\tilde{u}^k - u^k)\} \ge 0,\ \forall\, u' \in \Omega.$$

用 $u^{k+1} = u^k - \gamma(u^k - \tilde{u}^k)$ 产生 G-模下 $\{\|u^k - u^*\|_G\}$ 的收缩序列。

8. **基于梯度投影的凸优化收缩算法和下降算法** 讨论可微凸优化

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

的梯度算法. 对无约束凸二次规划, 有关方法相当于缩短了步长的最速下降法. 有算例说明, 最速下降法故意缩短步长, 收敛速率却有令人难以置信的数量级提高. 现实生活中有一类变分不等式

$$x^* \in \Omega, \quad (x - x^*)^T g(x^*) \geq 0, \qquad \forall\, x \in \Omega,$$

其中 $g(x)$ 是某个凸函数 $f(x)$ 的梯度. 这种变分不等式形式上等价于所讨论的约束凸优化问题, 然而 $f(x)$ 是无法提供的, 只是对给定的 x, 可以观测到 $g(x)$, 求解时只有 $g(x)$ 的信息可以使用. 采用迭代公式

$$x^{k+1} = P_\Omega[x^k - \beta_k g(x^k)],$$

要求 x^k, 步长 β_k 和新的迭代点 x^{k+1} 满足

$$(x^k - x^{k+1})^T (g(x^k) - g(x^{k+1})) \leq (\nu/\beta_k)\|x^k - x^{k+1}\|^2, \quad \nu \in (0,1).$$

这样产生的序列 $\{x^k\}$ 隐含了 (那个我们并不知道的) $f(x)$ 满足下降性质

$$f(x^{k+1}) \leq f(x^k) - \big((1-\nu)/\beta_k\big)\|x^k - x^{k+1}\|^2.$$

对第一讲提及的经济平衡互补问题求解, 说明了方法的实用性和优越性.

9. 线性约束凸优化基于对偶上升的自适应方法

凸优化问题 $\min\{f(x) \mid Ax = b,\ x \in \mathcal{X}\}$ 的 Lagrange 的对偶问题是

$$\left\{\begin{array}{rl} \max_{x,\lambda} & L(x,\lambda) = f(x) - \lambda^T(Ax-b) \\ \text{s. t} & x \in \mathcal{X},\ (x'-x)^T\nabla_x L(x,\lambda) \ge 0,\ \forall x' \in \mathcal{X}. \end{array}\right.$$

一对 (x,λ) 被说成是对偶可行当且仅当

$$x \in \mathcal{X},\ (x'-x)^T(\nabla f(x) - A^T\lambda) \ge 0,\ \forall x' \in \mathcal{X}.$$

换句话说, 对给定的 $\lambda^k \in \Re^m$, 通过求解 $\min\{L(x,\lambda^k) \,\big|\, x \in \mathcal{X}\}$ 得到 x^k, 这样的一对 (x^k,λ^k) 就是对偶可行的. 从对偶可行的 (x^k,λ^k) 开始, 令

$$\lambda^{k+1} = \lambda^k - \beta_k(Ax^k - b),$$

并通过 $x^{k+1} =$ Argmin $\{L(x,\lambda^{k+1}) \,\big|\, x \in \mathcal{X}\}$ 求得 x^{k+1}, 要求它们满足

$$(\lambda^k - \lambda^{k+1})^T A(x^k - x^{k+1}) \le (\nu/\beta_k)\|\lambda^k - \lambda^{k+1}\|^2, \quad \nu \in (0,1).$$

这样产生的序列 $\{(x^k,\lambda^k)\}$ 使对偶问题的目标函数满足上升性质

$$L(x^{k+1},\lambda^{k+1}) - L(x^k,\lambda^k) \ge \big((1-\nu)/\beta_k\big)\|\lambda^k - \lambda^{k+1}\|^2.$$

对第四、五讲提及的相关性矩阵校正问题求解并与前述方法比较, 收敛速度几乎提高了一倍. 说明对某些问题对偶上升方法是很有效的方法.

10. 线性约束单调变分不等式的自适应投影收缩算法

设 $A \in \Re^{m\times n}$, $b \in \Re^m$. 线性约束的单调变分不等式是求 $x^* \in S$, 使得

$$(x - x^*)^T f(x^*) \ge 0, \quad \forall x \in S,$$

其中 $S = \{x \in \Re^n \mid A^T x = b \ (or\ A^T x \le b), \ x \in \mathcal{X}\}$. 这类问题大量出现在交通网络分析里. 可微凸优化 $\min\{\theta(x) \,|\, x \in S\}$, 记 $f(x) = \nabla\theta(x)$, 就可以描述成这类变分不等式的特例, 对这类凸优化问题, 这一讲的方法是只用梯度的方法. 对线性约束引入 Lagrange 乘子 λ, 问题要求 (x^*, λ^*),

$$\left\{ \begin{array}{ll} x^* \in \mathcal{X}, & (x - x^*)^T\{f(x^*) + A\lambda^*\} \ge 0, \quad \forall\, x \in \mathcal{X}, \\ \lambda^* \in \Lambda, & (\lambda - \lambda^*)^T(-A^T x^* + b) \ge 0, \qquad \forall\, \lambda \in \Lambda, \end{array} \right.$$

其中 $\Lambda = \Re^m$ $(A^T x = b)$ 或者 $\Lambda = \Re^m_+$ $(A^T x \le b)$. 它的紧凑形式是

$$u^* \in \Omega, \qquad (u' - u^*)^T F(u^*) \ge 0, \qquad \forall\, u' \in \Omega.$$

不同于简单投影 $\tilde{u}^k = P_\Omega[u^k - \beta_k F(u^k)]$, 我们用 Gauss-Seidel 型的投影

$$\tilde{\lambda}^k = P_\Lambda\{\lambda^k + \tfrac{\beta_k}{\tau}(A^T x^k - b)\}, \quad \tilde{x}^k = P_{\mathcal{X}}\big\{x^k - \beta_k\big[f(x^k) + A\tilde{\lambda}^k\big]\big\},$$

得到预测点 $\tilde{u}^k = (\tilde{x}^k, \tilde{\lambda}^k)$, 在第二个投影中用了刚刚更新的 $\tilde{\lambda}^k$. 进而根据第三讲变分不等式收缩算法的统一框架, 构造收缩算法. 由于预测点由投影得来, 迭代序列 $\{u^k\}$ 又向解集收缩, 算法还称为投影收缩算法.

11. **结构型优化的交替方向法** 从这讲开始的四讲考虑结构型优化问题

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

它等价于变分不等式 $w^*\in\Omega,\ \ (w-w^*)^TF(w^*)\geq 0,\ \forall\, w\in\Omega$. 其中

$$w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix}\quad \text{and}\quad F(w)=\begin{pmatrix}\nabla\theta_1(x)-A^T\lambda\\ \nabla\theta_2(y)-B^T\lambda\\ Ax+By-b\end{pmatrix},\quad \Omega=\mathcal{X}\times\mathcal{Y}\times\Re^m.$$

交替方向法中 x 是隐式变量, 迭代从给定的 $v^k=(y^k,\lambda^k)$ 出发, 依次求解关于 x 和 y 的子问题. 若记

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

和 $\tilde{\lambda}^k=\lambda^k-\beta(A\tilde{x}^k+B\tilde{y}^k-b)$. 以 $v^{k+1}=\tilde{v}^k$ 迭代得的序列 $\{v^k\}$ 满足

$$\|v^{k+1}-v^*\|_M^2\leq\|v^k-v^*\|_M^2-\|v^k-v^{k+1}\|_M^2.$$

介绍了 Ye-Yuan 通过计算步长由 $v^{k+1}=v^k-\alpha_k(v^k-\tilde{v}^k)$ 产生新迭代点的交替方向收缩算法. 算例说明收缩意义下最优步长使方法收敛更快.

12. 线性化交替方向收缩算法

结构型优化问题 $\min\{\theta_1(x)+\theta_2(y)\,|\,Ax+By=b\ x\in\mathcal{X},\ y\in\mathcal{Y}\}$ 中, 矩阵 B 往往是数量矩阵, 矩阵 A 是一般形式. 这对求解 x 子问题带来了困难. 对经典的 ADM 方法中 x 子问题中的二次函数 $\frac{\beta}{2}\|Ax+By^k-b\|^2$ 线性化, 并加上正则项 $\frac{r}{2}\|x-x^k\|^2$. 这样迭代就从 $w^k=(x^k,y^k,\lambda^k)$ 出发. 记 $q=(Ax^k+By^k-b)-\frac{1}{\beta}\lambda^k$, 依次求解 x 和 y 的子问题

$$\tilde{x}^k=\text{Argmin}\Big\{\theta_1(x)+\beta q^TAx+\frac{r}{2}\|x-x^k\|^2\,|\,x\in\mathcal{X}\Big\}$$

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

并令 $\tilde{\lambda}^k=\lambda^k-\beta(A\tilde{x}^k+B\tilde{y}^k-b)$. 在 $rI\succeq\beta A^TA$ 条件满足的情况下, 以 $w^{k+1}=\tilde{w}^k$ 迭代得的序列 $\{w^k\}$ 满足

$$\|w^{k+1}-w^*\|_G^2\le\|w^k-w^*\|_G^2-\|w^k-w^{k+1}\|_G^2.$$

也可以由自调比法则选取 $r>0$ 使得 $\beta\|A^TA(x^k-\tilde{x}^k)\|<r\|x^k-\tilde{x}^k\|$. 得到收缩方向 $-d(w^k,\tilde{w}^k)$. 直接由 $w^{k+1}=w^k-d(w^k,\tilde{w}^k)$ 或通过计算步长 $w^{k+1}=w^k-\alpha_kd(w^k,\tilde{w}^k)$ 决定新的迭代点. 构造基于线性化交替方向的收缩算法.

13. 定制 PPA 算法意义下的交替方向法

同样考虑求解 $\min\{\theta_1(x)+\theta_2(y)\mid Ax+By=b\ x\in\mathcal{X},\ y\in\mathcal{Y}\}$ 的交替方向法, x 是中间变量. 迭代从给定的 $v^k=(y^k,\lambda^k)$ 出发, 求得

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

后, 就令 $\tilde{\lambda}^k=\lambda^k-\beta(A\tilde{x}^k+By^k-b)$. 最后求得

$$\tilde{y}^k=\text{Argmin}\Big\{\theta_2(y)+\tfrac{\beta}{2}\|(A\tilde{x}^k+By-b)-\tfrac{1}{\beta}\tilde{\lambda}^k\|^2\mid y\in\mathcal{Y}\Big\}.$$

与经典的交替方向法(见第十一讲)相比, 只是用了一个新的顺序 $(\tilde{x}^k,\tilde{\lambda}^k,\tilde{y}^k)$ 产生预测点. 这样产生的预测点的 $\tilde{v}^k$ 部分, 有 PPA 性质

$$(v^k-v^*)^TM(v^k-\tilde{v}^k)\ge\|v^k-\tilde{v}^k\|_M^2.$$

用基于 ADM 得到的预测点 $\tilde{v}^k$, 用 PPA 迭代式

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

产生新的迭代点 v^{k+1}. 这样的 ADM-PPA 的序列 $\{v^k\}$ 具有性质

$$\|v^{k+1}-v^*\|_M^2\le\|v^k-v^*\|_M^2-\gamma(2-\gamma)\|v^k-\tilde{v}^k\|_M^2.$$

对相关性矩阵校正和图像恢复中一些典型算例的试验说明, 达到同样的精度, 采用 $\gamma=1.5$ 的 PPA 算法, 比经典的交替方向法一般节省时间$1/3$.

14. 定制 PPA 算法意义的线性化交替方向法

相对前一讲的 ADM-PPA 方法, 就像第十二讲相对于第十一讲中经典的 ADM 方法. 同样针对 $\min\{\theta_1(x)+\theta_2(y)\,|\,Ax+By=b\; x\in\mathcal{X},\; y\in\mathcal{Y}\}$ 中矩阵 A 是一般形式的困难, 线性化处理交替方向法 x 子问题中的二次函数 $\frac{\beta}{2}\|Ax+By^k-b\|^2$, 并加上正则项 $\frac{r}{2}\|x-x^k\|^2$. 从 $w^k=(x^k,y^k,\lambda^k)$ 出发, 记 $q=(Ax^k+By^k-b)-\frac{1}{\beta}\lambda^k$, 求得

$$\tilde{x}^k=\text{Argmin}\big\{\theta_1(x)+\beta x^TA^Tq+\tfrac{r}{2}\|x-x^k\|^2\,|\,x\in\mathcal{X}\big\}$$

后, 令 $\tilde{\lambda}^k=\lambda^k-\beta(A\tilde{x}^k+By^k-b)$. 然后再求

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

在 $rI\succeq\beta A^TA$ 条件满足的情况下, 这样产生的预测点 $\tilde{w}^k$ 有 PPA 性质

$$(w^k-w^*)^TG(w^k-\tilde{w}^k)\ge\|w^k-\tilde{w}^k\|_G^2.$$

用公式

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

更新迭代点, 无需计算步长, 得的序列 $\{w^k\}$ 满足性质

$$\|w^{k+1}-v^*\|_G^2\le\|w^k-w^*\|_G^2-\gamma(2-\gamma)\|w^k-\tilde{w}^k\|_G^2.$$

也可以由自调比法则选取 $r>0$ 使得 $\beta\|A^TA(x^k-\tilde{x}^k)\|<r\|x^k-\tilde{x}^k\|$.
构造基于线性化交替方向的收缩算法.

15. **三个可分离算子凸优化的平行分裂ALM 方法** 对问题

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

考虑平行分分裂的增广 Lagrange 乘子收缩算法. 从 $w^k=(x^k,y^k,z^k,\lambda^k)$ 出发, 平行地求解关于 x,y 和 z 的子问题：

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

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

$$\tilde{z}^k=\text{Argmin}\big\{\theta_3(z)+\tfrac{\beta}{2}\|(Ax^k+By^k+Cz-b)-\tfrac{1}{\beta}\lambda^k\|^2\,|\,z\in\mathcal{Z}\big\}.$$

然后令 $\tilde{\lambda}^k=\lambda^k-\beta(A\tilde{x}^k+B\tilde{y}^k+C\tilde{z}^k-b)$. 这样生成的 $\tilde{w}^k$ 有性质

$$(w^k-w^*)^TM(w^k-\tilde{w}^k)\ge\tfrac{2-\sqrt{3}}{2}\|w^k-\tilde{w}^k\|_M^2.$$

其中 $M=\text{diag}(\beta A^TA,\beta B^TB,\beta C^TC,\frac{1}{\beta}I_m)$ 是半正定矩阵. 以

$$w^{k+1}=w^k-\gamma\alpha_k^*(w^k-\tilde{w}^k),\qquad \alpha_k^*\equiv\tfrac{2-\sqrt{3}}{2},$$

生成的的序列 $\{w^k\}$ 具有收缩性质:

$$\|w^{k+1}-w^*\|_M^2\le\|w^k-w^*\|_M^2-\gamma(2-\gamma)(\alpha_k^*)^2\|w^k-\tilde{w}^k\|_M^2.$$

由上式容易得到 $\lim_{k\to\infty}\|w^k-\tilde{w}^k\|_M^2=0$ 和方法的总体收敛性.

16. **三个可分离算子凸优化的略有改动的交替方向法** 对问题

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

从 $v^k=(y^k,z^k,\lambda^k)$ 出发, 逐个求解关于 x,y 和 z 的子问题：

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

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

$$z^{k+1}=\text{Argmin}\big\{\theta_3(z)+\tfrac{\beta}{2}\|(Ax^{k+1}+y^{k+1}+z-b)-\tfrac{1}{\beta}\lambda^k\|^2\,|\,z\in\mathcal{Z}\big\}.$$

然后校正Lagrange 乘子

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

这样的方法是将处理两个可分离算子的交替方向法, 直接推广到对三个可分离算子问题的求解. 这种直接推广的方法有没有收敛性, 至今还没有结论. 如果对直接推广方法得到的 $(y^{k+1},z^{k+1},\lambda^{k+1})$ 略作变动,

仅对 y 做些简单再加工 $y^{k+1}:=y^{k+1}+(z^k-z^{k+1})$,

z^{k+1} 和 λ^{k+1} 都保持不变. 新迭代点 $v^{k+1}=(y^{k+1},z^{k+1},\lambda^{k+1})$.

这样的方法, 是对交替方向法的直接推广做了很小的改动. 我们证明这个方法是收敛的, 并在遍历意义下有 $O(1/t)$ 的收敛速率.

17. 多可分离算子凸优化带回代的交替方向收缩算法

考虑含多个可分离算子的最优化问题

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

在这讲的方法中, x_1 只是中间变量. 迭代从 $v^k = (x_2^k, \ldots, x_m^k, \lambda^k)$ 出发.

迭代的第一阶段是依次对 $i = 1, 2, \ldots, m$ 用交替方向法求解 x_i 的子问题

$$\min\Big\{\theta_i(x) + \frac{\beta}{2}\|(\sum_{j=1}^{i-1} A_j \tilde{x}_j^k + A_i x_i + \sum_{j=i+1}^{m} A_j \tilde{x}_j^k - b) - \frac{1}{\beta}\lambda^k\|^2\,|\,x \in \mathcal{X}\Big\}$$

得到 $\tilde{x}_i^k$, 最后令 $\tilde{\lambda}^k = \lambda^k - \beta(\sum_{j=1}^m A_j \tilde{x}_j^k - b)$. 完成预测.

迭代的第二阶段 是通过求解方程组 (其中 H 对称正定块对角, α_k^* 可计算)

$$H^{-1}M^T(v^{k+1} - v^k) = \alpha_k^*\big(\tilde{v}^k - v^k\big)$$

得到新的 v^{k+1}, 由于其中 $H^{-1}M^T$ 是对角元为 1 的上三角矩阵, 这个过程被称之为高斯回代(**Gaussian back substitution**). 迭代序列 $\{v^k\}$ 具有

$$\|v^{k+1} - v^*\|_G^2 \le \|v^k - v^*\|_G^2 - \frac{1}{4}(\|v^k - \tilde{v}^k\|_H^2 + \|v^k - \tilde{v}^k\|_Q^2).$$

这样的收缩性质, 其中 $G = MH^{-1}M^T$, Q 也是半正定矩阵.

18. 多可分离算子凸优化带回代的线性化交替方向法

还是考虑含多个可分离算子的最优化问题

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

设上一讲的方法中 x_i-子问题因矩阵 A_i 的一般性求解有困难. 将二次函数在 x_i^k 处展开, 加正则项 $\frac{r}{2}\|x_i-x_i^k\|^2$. $w^k=(x_1^k,\ldots,x_m^k,\lambda^k)$ 参与运算.

迭代的第一阶段依次对 $i=1,2,\ldots,m$ 用交替方向法求解 x_i 的子问题

$$\tilde{x}_i^k=\arg\min\left\{\theta_i(x_i)+q_i^TA_ix_i+\frac{r_i}{2}\|x_i-x_i^k\|^2\,\Big|\,x_i\in\mathcal{X}_i\right\}$$

其中 $q_i=\beta(\sum_{j=1}^{i-1}A_j\tilde{x}_j^k+\sum_{j=i}^{m}A_jx_j^k-b)$. 因子 r_i 满足 $r_i\|x_i^k-\tilde{x}_i^k\|^2\geq$ $\geq\beta\|A_i(x_i^k-\tilde{x}_i^k)\|^2$. 最后令 $\tilde{\lambda}^k=\lambda^k-\beta(\sum_{j=1}^{m}A_j\tilde{x}_j^k-b)$. 完成预测.

迭代的第二阶段是通过求解方程组 (其中 H 是分块数量矩阵, α_k^* 可计算)

$$H^{-1}M^T(w^{k+1}-w^k)=\alpha_k^*\big(\tilde{w}^k-w^k\big)$$

得到新的 w^{k+1}, 由于其中 $H^{-1}M^T$ 是对角元为 **1** 的上三角矩阵, 即用高斯回代(**Gaussian back substitution**) 得到 w^{k+1}. 迭代序列 $\{w^k\}$ 具有

$$\|w^{k+1}-w^*\|_G^2\leq\|w^k-w^*\|_G^2-\frac{1}{4}(\|w^k-\tilde{w}^k\|_H^2+\|w^k-\tilde{w}^k\|_Q^2).$$

这样的收缩性质, 其中 $G=MH^{-1}M^T$, Q 也是半正定矩阵.

19. Lipschitz 连续单调变分不等式投影收缩算法的收敛速率

变分不等式 $u^* \in \Omega,\ (u - u^*)^T F(u^*) \ge 0,\ \forall\, u \in \Omega$ 的解集可以表示成

$$\Omega^* = \bigcap_{u\in\Omega} \big\{ \tilde{u} \in \Omega : (u - \tilde{u})^T F(u) \ge 0 \big\}.$$

See Theorem 2.3.2 in: F. Facchinei and J.S. Pang, Finite-Dimensional Variational Inequalities and Complementarity problems, Vol. I. Springer Series in Operations Research. (2003). pp. 159

这隐含了 $\tilde{u}$ 是 VI(Ω, F) 的 ϵ 近似解, 如果它满足

$$\tilde{u} \in \Omega \quad \text{and} \quad \inf_{u\in\Omega}\big\{ (u - \tilde{u})^T F(u) \big\} \ge -\epsilon.$$

对给定的 $\epsilon > 0$, 若 $\tilde{u}$ 满足

$$\tilde{u} \in \Omega \quad \text{and} \quad \sup_{u\in\Omega}\Big\{ \frac{1}{\|\tilde{u} - u\|} (\tilde{u} - u)^T F(u) \Big\} \le \epsilon,$$

则 $\tilde{u}$ 被称为遍历意义下的 ϵ 近似解. 第三讲中提及的单调变分不等式的投影收缩算法, 预测点 $\tilde{u}^k$ 和步长 α_k 满足

$$\tilde{u}^k = P_\Omega\{\tilde{u}^k - [\beta_k F(\tilde{u}^k) - d(u^k, \tilde{u}^k)]\}, \quad \alpha_k = \frac{(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2} \ge \frac{1}{2}.$$

对用不同 (的孪生) 方向, 相同步长产生新迭代点

$$u^{k+1} = u^k - \gamma \alpha_k d(u^k, \tilde{u}^k) \quad \text{和} \quad u^{k+1} = P_\Omega[u^k - \gamma \alpha_k \beta_k F(\tilde{u}^k)]$$

的投影收缩算法, 在算子 F Lipschitz 连续的假设下, 都能经过 $O(1/\epsilon)$ 次迭代得到遍历意义下的 ϵ 近似解. 算法因此具有 $O(1/k)$ 收敛速率.

20. **交替方向法的计算复杂性和收敛速率** 结构型优化问题

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

的交替方向法中 x 是隐式变量, 迭代从给定的 $v^k=(y^k,\lambda^k)$ 出发, 依次求解关于 x 和 y 的子问题. 分别是

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

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

和 $\lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}-b)$. 所得的序列 $\{v^k\}$ 满足

$$\|v^{k+1}-v^*\|_H^2\le\|v^k-v^*\|_H^2-\|v^k-v^{k+1}\|_H^2,\quad H=\begin{pmatrix}\beta B^TB & 0\\ 0 & \frac{1}{\beta}I_m\end{pmatrix}.$$

这样的收缩性质. 在变分不等式的框架下证明了交替方向法具有 $O(1/k)$ 收敛速率以外, 同时证明了残量序列 $\{\|v^k-v^{k+1}\|_H^2\}$ 具有单调不增性

$$\|v^k-v^{k+1}\|_H^2\le\|v^{k-1}-v^k\|_H^2.$$

由收缩性和单调不增性, 经过 k 次迭代, 就能是误差 $\|v^k-v^{k+1}\|_H^2$ 达到

$$\|v^k-v^{k+1}\|_H^2\le\frac{1}{k}\|v^0-v^*\|_H^2$$ 这样 $O(1/k)$ 的精度.

动因与说明 与这个系列讲义相关的研究工作, 主要包括以下四个方面：投影收缩算法、交替方向法、多元分裂算法以及变分不等式框架下的松弛 PPA 算法. 促使我整理这些讲义的动因, 是这些工作分别得到了不同学科的一些著名学者引用.

一．投影收缩算法：投影收缩算法方面的论文, 得到包括 UC Berkeley 计算机系 Jordan 教授在内的学者引用, 相关内容被写进他们的附录. 这类成果近年也被一些北美名校 (宾习法尼亚大学,多伦多大学, UC Berkeley, 哥伦比亚大学) 的博士们在语音识别、光纤网络、机器学习等的博士论文中提到. Jordan 教授是美国科学院和工程院院士.

二．交替方向法：最优化中采用分裂算法, 就像现实生活中大工程需要分包. 从招收第一批博士生开始, 在收缩算法框架下研究交替方向法就是我们的一个主要课题. 新应用领域的发现, 使交替方向法渐得广泛认可. 我们多年前发表的这类方法中的一个自调比准则, 被 Stanford 大学电子工程系 Boyd 教授在 2010 年的一篇综述文章中称为一个简单而有效的公式 (A simple scheme that often works well), 对我们的分析依据也作了简要介绍. Boyd 教授是 2006 年世界数学家大会邀请报告人.

三．多个可分离算子的分裂算法：交替方向法处理的是含两个可分离算子的问题. 对多于两个可分离算子的问题, 同样需要易于实现的分裂方法. 基于到鞍点越来越近的收缩思想, 我们给出了一些解决此类问题的方法. 这类工作, 已经被 UCLA 数学系教授 Osher 的课题组在降维问题上应用. 他们在文章中说到, The method proposed by He, Tao and Yuan is appropriate for this application. Osher 教授是美国科学院院士, 2010 年世界数学家大会一小时邀请报告人.

四．变分不等式框架下的松弛 PPA 算法：在变分不等式框架中构造算法的优越性, 得到越来越多的学者认可. 我们 2012 年在 SIAM Journal Image Science 发表的文章, 初稿就被一些欧洲的学者在他们最新的研究中引用. 例如, Pock and Chambolle 的文章中说到, 用 He and Yuan 的 PPA 形式, 极大地简化了收敛性分析 (which greatly simplified the convergence analysis). Chambolle 是在图像学领域有很大影响的学者.

五．一阶算法的收敛速率：好方法需要有好的理论结果保证. 2005 年以来, Nemirovski 和 Nestrov 等最优化理论与方法领域的世界数学家大会邀请报告人对一阶算法收敛速率表现出的浓厚兴趣, 激发了我们对这个系列讲义中主要方法收敛速率的研究热情. 最后两讲, 我们用简单的方法和简短的篇幅, 分别论证了投影收缩算法和交替方向法都具有 $O(1/k)$ 的收敛速率. 换句话说, 对给定的 $\epsilon > 0$, 经过 $O(1/\epsilon)$ 次迭代, 就能得到遍历意义下的 ϵ 近似解.

授人以鱼不如授人以渔。一个好的优化方法, 应该是容易被工程师们掌握, 让人用来自己解决问题的方法。

✠ 为相关学科所用, 恰是我们从事优化方法研究的本源追求 ✠