智能优化算法

作业1

分别计算以下 5＊5 图像的Laplace算子，梯度算子，边界采用周期边界条件

9	2	3	2	6	9	2
1	1	2	3	2	1	1
2	2	1	2	6	2	2
6	3	0	8	7	6	3
6	1	2	7	8	6	1
9	2	3	2	6	9	2
1	1	2	3	2	1	1

常用范数

- **L1范数：**

$$\|x\|_1 = |x_1| + \cdots + |x_n|$$

- **L2范数：**

$$\|x\|_2 = (x^T x)^{1/2} = (x_1^2 + \cdots + x_n^2)^{1/2}.$$

- 无穷范数：

$$\|x\|_\infty = \max\{|x_1|, \cdots, |x_n|\}$$

- **Lp范数：**

$$\|x\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}.$$

- 矩阵**F**范数：

$$\|X\|_F = (\mathrm{tr}(X^T X))^{1/2} = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} X_{ij}^2 \right)^{1/2}.$$

下一页 上一页

练习2

- 给定3*3灰度图片:

$$u = (u_{ij}), i = 1,2,3; j = 1,2,3$$

(1) $\boldsymbol{y}(\boldsymbol{u}) = \| \Delta \boldsymbol{u} \|_2^2$,

（2） $\boldsymbol{y}(\boldsymbol{u}) = \| \Delta \boldsymbol{u} \|_1$,

试将以上范数展开

$$\| x \|_1 = \sum_{\mathrm{i}} | x_i | \qquad \| x \|_2 = \sqrt{\sum_{\mathrm{i}} x_i^2}$$

$$y(u) = \| \Delta u \|_2^2 = \left(\sqrt{\sum_{ij} (\Delta u_{ij})^2}\right)^2 = \sum_{ij} (\Delta u_{ij})^2$$

$$= \sum_{ij} (-4u_{ij} + u_{ij+1} + u_{ij-1} + u_{i+1j} + u_{i-1j})^2$$

$$y(u) = \| \Delta u \|_{\mathbf{1}} = \sum_{ij} | \Delta u_{ij} |$$

$$= \sum_{ij} | -\mathbf{4} u_{ij} + u_{ij+\mathbf{1}} + u_{ij-\mathbf{1}} + u_{i+\mathbf{1}j} + u_{i-\mathbf{1}j} |$$

作业3

- 给定3*3灰度图片:

$$u = (u_{ij}), i = 1,2,3; j = 1,2,3$$

(1) $\boldsymbol{y}(\boldsymbol{u}) = \mathrm{di}\boldsymbol{v}(\nabla \boldsymbol{u}),$

（2） $\boldsymbol{y}(\boldsymbol{u}) = \boldsymbol{div}\left(\dfrac{\nabla \boldsymbol{u}}{|\nabla \boldsymbol{u}|}\right),$

试将以上公式展开，同时证明以下公式成立

$< \nabla \boldsymbol{u}, \nabla \boldsymbol{u} >= -2 < \boldsymbol{div}(\nabla \boldsymbol{u}), \boldsymbol{u} >,$ $<>$ 表示内积

$$(1)\quad y(u)=\mathrm{div}(\nabla u),$$

$$\nabla u=(u_x,u_y),\ div(\nabla u)=(u_{xx}+u_{yy})$$

$$(2)\quad y(u)=div\left(\frac{\nabla u}{|\nabla u|}\right),$$

$$\mathrm{g}(x,y)=\frac{\nabla u}{|\nabla u|}=\frac{(u_x,u_y)}{\sqrt{u_x^2+u_y^2}}=(g_1,g_2)$$

$$\mathrm{div}(g(x,y))=g_{1x}+g_{2y}$$

$$g_{1x}=\frac{\partial g_1}{\partial x}=\frac{\partial\left(\frac{u_x}{\sqrt{u_x^2+u_y^2}}\right)}{\partial x}=\frac{u_{xx}\sqrt{u_x^2+u_y^2}-u_x\frac{u_xu_{xx}+u_yu_{yx}}{\sqrt{u_x^2+u_y^2}}}{u_x^2+u_y^2}$$

下一页 上一页

$$<\nabla \boldsymbol{u}, \nabla \boldsymbol{u}>=-<\boldsymbol{div}(\nabla \boldsymbol{u}), \boldsymbol{u}>,$$ <> 表示内积

$$\sum_{ij}[(u_{i+1j}-u_{ij})^2+(u_{ij+1}-u_{ij})^2]$$

$$=-\sum_{ij}(-4u_{ij}+u_{i+1j}+u_{ij+1}+u_{i-1j}+u_{ij-1})u_{ij}$$

$$(-4u_{ij}+u_{i+1j}+u_{ij+1}+u_{i-1j}+u_{ij-1})u_{ij}$$

$$+(-4u_{i+1j}+u_{i+2j}+u_{ij+2}+u_{ij}+u_{i+1j-1})u_{i+1j}$$

$$+(-4u_{i-1j}+u_{ij}+u_{i-1j+1}+u_{i-2j}+u_{i-1j-1})u_{i-1j}$$

$$+(-4u_{ij+1}+u_{i+1j+1}+u_{ij+2}+u_{ij}+u_{i-1j+1})u_{ij+1}$$

$$+(-4u_{ij-1}+u_{i+1j-1}+u_{ij}+u_{ij-2}+u_{i-1j-1})u_{ij-1}$$

$$+\sum$$

下一页 上一页

第3章

凸优化问题求解算法

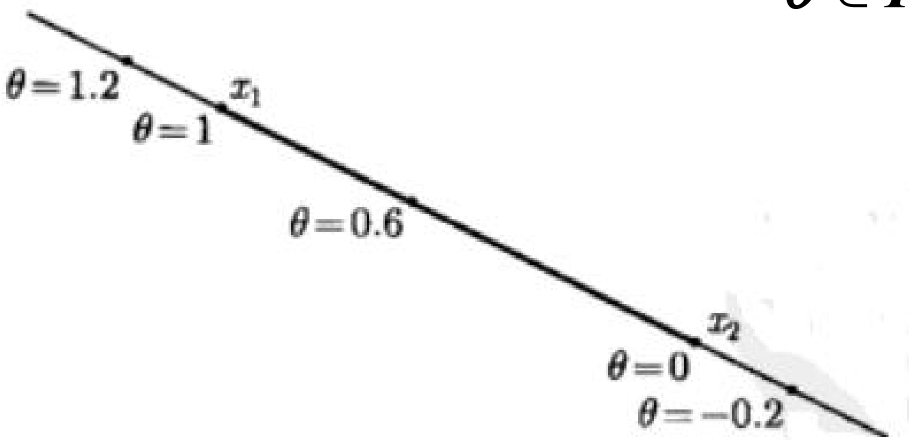
3.1 线段与直线

设 $x_1 \neq x_2$ 为 $\mathbf{R}^n$ 空间中的两个点

x_1 和 x_2 之间的（闭）**线段** 具有以下形式

$$y = \theta x_1 + (1-\theta)x_2,$$ 其中，$0 \le \theta \le 1$

$\theta \in R$为直线



仿射函数

函数 $f : \mathbf{R}^n \to \mathbf{R}^m$ 是仿射的，

如果它是一个线性函数和一个常数的和，

$$f(x) = Ax + b$$

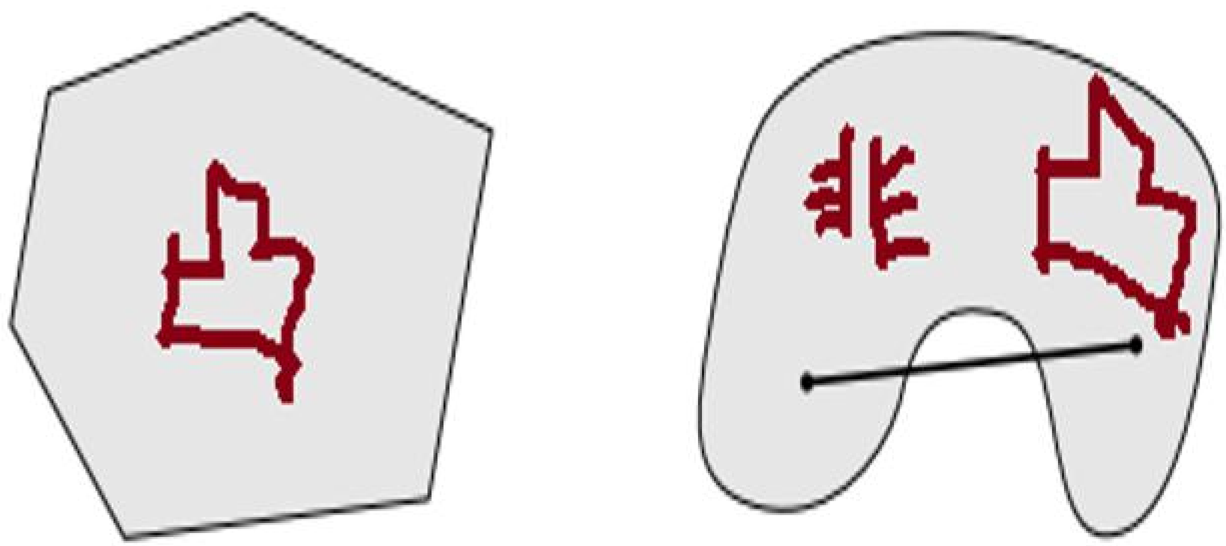
下一页 上一页

凸集

集合 C 被称为**凸集**，如果 C 中任意两点间的线段仍然在 C 中，

即对于任意 $x_1, x_2 \in C$ 和满足 $0 \leqslant \theta \leqslant 1$ 的 θ 都有

$$\theta x_1 + (1-\theta)x_2 \in C.$$



凸集例子

10. 欧几里得空间中的球：

$$B(x_c, r) = \left\{ x \,\middle|\, \| x - x_c \|_2 \le r, x_c \in \mathbf{R}^n, r \in R \right\}$$

10）证明：设 $x_1, x_2 \in B$，则 $\begin{matrix} \| x_1 - x_c \|_2 \le r \\ \| x_2 - x_c \|_2 \le r \end{matrix}$

$$\| (\theta x_1 + (1-\theta) x_2) - x_c \|_2 = \| (\theta x_1 + (1-\theta) x_2) - (\theta x_c + (1-\theta) x_c) \|_2$$

$$= \| \theta (x_1 - x_c) + (1-\theta)(x_2 - x_c) \|_2$$

$$\le \theta \| (x_1 - x_c) \|_2 + (1-\theta) \| (x_2 - x_c) \|_2$$
$$\le \theta r + (1-\theta) r = r$$

凸函数　函数 f 的定义域，用符号 dom f 描述

函数f称为凸函数，如果对于其定义域中的任意两点 x, y，都满足以下条件：

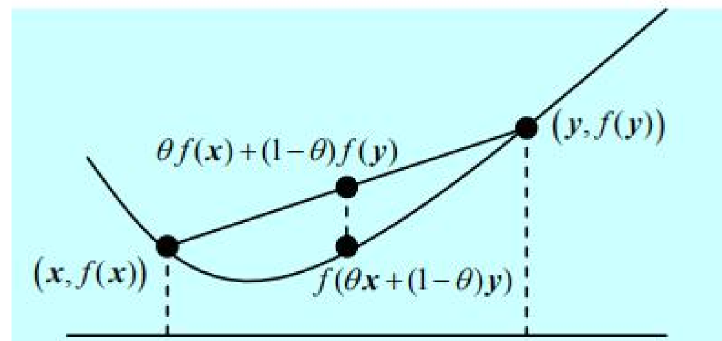
$$f(\theta x+(1-\theta)y)\le \theta f(x)+(1-\theta)f(y)$$

其中 $0\le\theta\le 1$

dom$f$ 为凸集

是为了保证 $\theta x+(1-\theta)y$ 在定义域内

从几何意义上看，上述不等式意味着点 $(x,f(x))$ 和 $(y,f(y))$ 之间的线段，即从x到y的弦，在函数 f 的图像上方



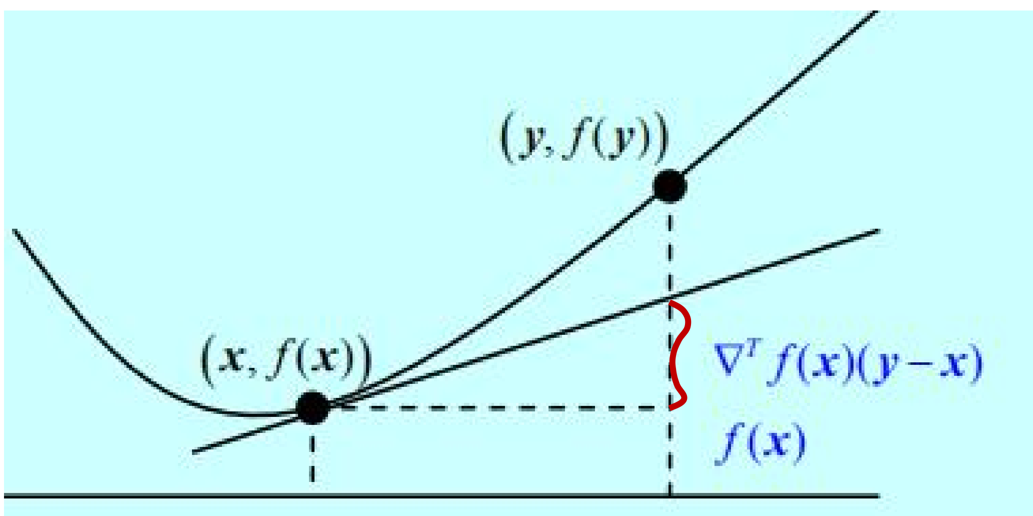
如果等号不成立，称函数为严格凸的

凸函数判定条件(参见P63页证明)

一阶条件

假设 f 可微，则函数 f 是凸函数的充要条件是 $\mathbf{dom}\, f$ 是凸集且对于任意 $x,\ y \in \mathbf{dom}\, f$，下式成立

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



性质

若 f 为凸函数，$\exists \boldsymbol{x}_0 \in \mathrm{dom} f$ ，使 $\nabla f(\boldsymbol{x}_0) = \boldsymbol{0}$ ，则

对 $\forall \boldsymbol{y} \in \mathrm{dom} f$ ， $f(\boldsymbol{y}) \ge f(\boldsymbol{x}_0) + \nabla^T f(\boldsymbol{x}_0)(\boldsymbol{y} - \boldsymbol{x}_0) = f(\boldsymbol{x}_0)$

即 $f(\boldsymbol{x}_0)$ 是 f 的最小值

定义

设向量 $\boldsymbol{x}$ 的数量值函数 $f(\boldsymbol{x})$ 二阶可微，则有

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

（Hessian 矩阵）

二阶条件

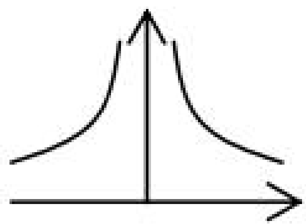
设函数 $f:\mathbf{R}^n \to \mathrm{R}$ 二阶可微，即梯度 $\nabla^2 f$ 在 $\mathrm{dom}f$ 均存在，则 f 为凸函数

$$\Longleftrightarrow$$

若 $\mathrm{dom}f$ 为凸集，且对 $\forall \boldsymbol{x} \in \mathrm{dom}f$，有 $\nabla^2 f(\boldsymbol{x})$ 半正定

Note：对于一维情况，则要求该函数的二阶偏导≥0

例：函数 $f(x)=\dfrac{1}{x^2}$



虽然 $\dfrac{\partial^2 f(x)}{\partial x^2}=6x^{-4}>0$，但 $\mathrm{dom}f$ 不是凸集，

故不能通过凸函数的二阶条件来判定凸函数

事实上，该函数确实不是凸函数

例

$$f(x)=e^{ax}, x\in \mathrm{R}$$

$$f'(x)=e^{ax}a$$

$$f''(x)=e^{ax}a^2\geq 0$$

故指数函数是凸函数

$$f(x)=\log x, x\in \mathrm{R}_{++}$$

$$f'(x)=\frac{1}{x}$$

$$f''(x)=-\frac{1}{x^2}<0$$

故对数函数是凹函数

3.2 凸优化问题：基本概念

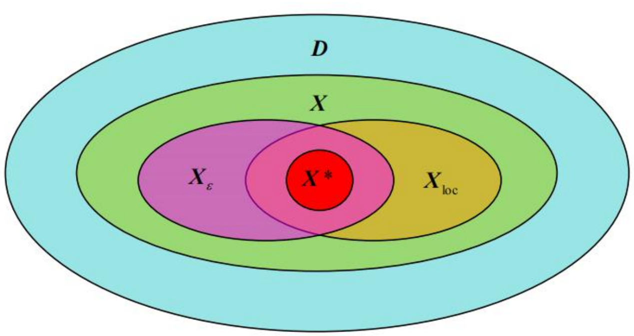
$$
\begin{aligned}
\min \quad & f_o(\boldsymbol{x}) \\
\text{s.t.} \quad & f_i(\boldsymbol{x}) \le 0 \qquad i = 1, \cdots, m \\
& h_j(\boldsymbol{x}) = 0 \qquad j = 1, \cdots, p
\end{aligned}
$$

$\boldsymbol{x}$：优化变量（optimization variable）

f_o：目标函数/损失函数（objective function / cost function）

f_i：不等式约束（inequality constraint）

h_i：等式约束（equality constraint）



域（Domain） $\boldsymbol{D} \triangleq \bigcap_{i=0}^{m} \operatorname{dom} f_i \cap \bigcap_{j=0}^{p} \operatorname{dom} h_j$

可行解集（feasible set）

$$\boldsymbol{X} = \left\{ \boldsymbol{x} \in \boldsymbol{D} \,\middle|\, \begin{matrix} f_i(\boldsymbol{x}) \le 0 & i = 1, \cdots, m \\ h_j(\boldsymbol{x}) = 0 & j = 1, \cdots, p \end{matrix} \right\}$$

ε 次优解（ε-suboptimal set）

$$\boldsymbol{X}_\varepsilon = \left\{ \boldsymbol{x} \in \boldsymbol{X} \,\middle|\, f_o(\boldsymbol{x}) = p^* + \varepsilon \right\}$$

最优值（optimal value）

$$p^* = \inf\left\{ f_o(\boldsymbol{x}) \mid \boldsymbol{x} \in \boldsymbol{X} \right\}$$

最优解集（optimal set）

$$\boldsymbol{X}^* = \left\{ \boldsymbol{x}^* \in \boldsymbol{X} \,\middle|\, f_o(\boldsymbol{x}^*) = p^* \right\}$$

局部最优值（local optimal value）

$$\exists R > 0，\quad p_{\text{loc}} = \inf\left\{ f_o(\boldsymbol{z}) \,\middle|\, \|\boldsymbol{z} - \boldsymbol{x}\|_2 \le R, \boldsymbol{x} \in \boldsymbol{X}, \boldsymbol{z} \in \boldsymbol{X} \right\}$$

局部最优解集（local optimal set）

$$\boldsymbol{X}_{\text{loc}} = \left\{ \boldsymbol{x}_{\text{loc}} \in \boldsymbol{X} \,\middle|\, f_o(\boldsymbol{x}_{\text{loc}}) = p_{\text{loc}} \right\}$$

下一页　上一页

凸优化问题

$$\begin{aligned} \min \quad & f_o(\boldsymbol{x}) \\ \text{s.t.} \quad & f_i(\boldsymbol{x}) \le 0 \qquad i = 1, \cdots, m \\ & \boldsymbol{A}_j^T \boldsymbol{x} = \boldsymbol{b}_j \qquad j = 1, \cdots, p \end{aligned}$$

1）目标函数 $f_o(\boldsymbol{x})$ 为凸函数

2）不等式约束 $f_1(\boldsymbol{x}), \cdots, f_m(\boldsymbol{x})$ 为凸函数

3）等式约束 $\boldsymbol{A}_1^T \boldsymbol{x} = \boldsymbol{b}_1, \cdots, \boldsymbol{A}_p^T \boldsymbol{x} = \boldsymbol{b}_p$ 为仿射函数

性质：凸优化问题的任意局部最优解也是全局最优解

凸优化问题解的最优条件：p132证明

$$\underset{x}{\text{minimize}}\ f_0(x)$$

$$\text{subject to}\ f_i(x) \leq 0,\ i = 1, \ldots, m$$

$$a_j^T x = b_j,\ j = 1, \ldots, p$$

x 是最优解，当且仅当 $x \in X$ 且

$$\nabla f_0(x)^T (y - x) \geqslant 0,\ \forall\, y \in X.$$

$$X = \{x \mid f_i(x) \leqslant 0,\ i = 1, \cdots, m,\ h_i(x) = 0,\ i = 1, \cdots, p\}.$$

3.3 无约束优化问题最优条件

对于无约束优化问题

$$\min_{x} f_0(x)$$

最优化条件简化为:

$$\nabla f_0(x) = 0.$$

例1 线性优化问题

例1：

$$\min_{x} \frac{1}{2} \boldsymbol{x}^T \boldsymbol{P}\boldsymbol{x} + \boldsymbol{q}^T \boldsymbol{x} + \boldsymbol{r}$$

最优性条件为：$\boldsymbol{P}\boldsymbol{x}^* + \boldsymbol{q} = 0$

如果P为正定矩阵，存在唯一解$\boldsymbol{x}^* = -\boldsymbol{P}^{-1}\boldsymbol{q}$

如果P不是正定矩阵，需讨论解的存在性

例2：最小二乘优化问题

$$\min_{x} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 = \langle \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}, \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \rangle = \boldsymbol{x}^T \boldsymbol{A}^T \boldsymbol{A}\boldsymbol{x} - 2(\boldsymbol{A}^T\boldsymbol{b})\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{b}$$

最优性条件为：$\boldsymbol{A}^T \boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{A}^T\boldsymbol{b}$

下一页 上一页

例2：非线性优化问题

$$\min_{x} f(x)^T f(x)$$

$$f(x) = \begin{bmatrix} 2(x_2 - x_1^2) \\ 1 - x_1 \end{bmatrix}$$

最优条件：

$$\nabla f(x) = \begin{bmatrix} 8(x_1^3 - x_1 x_2) + x_1 - 1 \\ 4(x_2 - x_1^2) \end{bmatrix} = 0$$

无法直接解方程组，

需要采用迭代方法进行求解

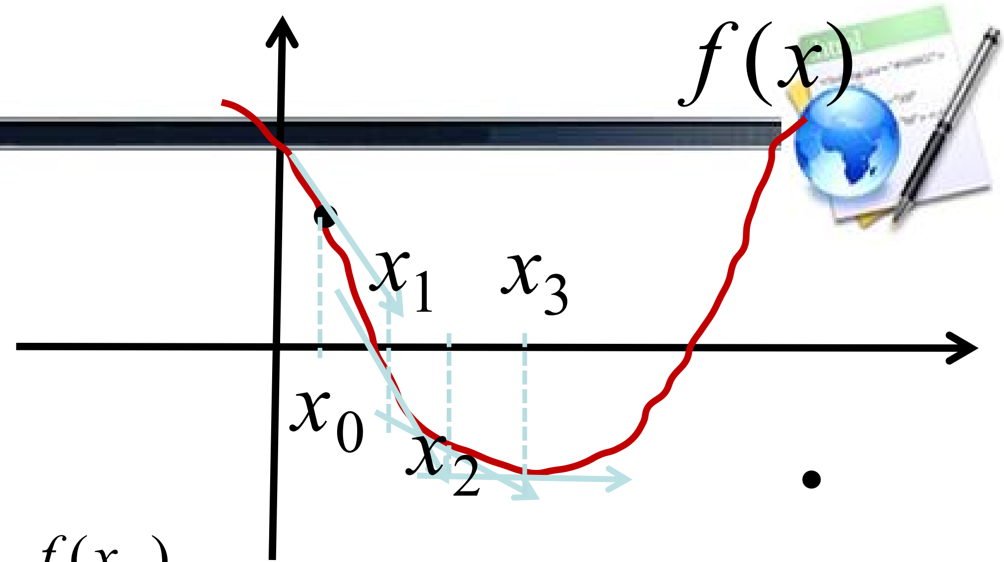
下一页 上一页

迭代方法求解无约束优化

1. 梯度下降法
2. 最速下降法
3. 牛顿迭代法

参考数值分析

梯度下降法(一阶逼近)



问题：$\min_{x} f(x)$

1. 给定初始值 x_0
2. 一阶泰勒级数 h(x)逼近 $f(x_0)$

$$h(x) = f(x_0 + \Delta x) = f(x_0) + f'(x_0)(x - x_0)$$

3. 求解满足h(x)最小的值x

$$h(x) = f(x_0 + \Delta x) = f(x_0) - f'(x_0)x_0 + f'(x_0)x$$

4. 当 x 与 $f'(x_0)$方向相反时，
   即成180度时，h(x)得到最小值

5. 即沿着负梯度方向移动，每一步的f值都减小

下一页 上一页

梯度下降算法

输入：初始点 x_0，固定步长 α

k = 1;

Repeat

$$x_k = x_{k-1} - \alpha \nabla f(x_{k-1})$$

$$k = k + 1$$

Until (满足终止条件)

步长选α取问题：

Backtracking line search

3.4 只含等式约束的问题

$$\min_{x} f_0(x)$$

$$\text{s.t.} \quad Ax = b$$

最优化条件：

即存在 $\nu \in \mathbf{R}^p$ $\nabla f_0(x) + A^T\nu = 0.$

求解方法：将等式约束问题转化为无约束问题

如：拉格朗日乘子法

3.5 Lagrangian dual,拉格朗日对偶

- 标准的优化问题

$$
\begin{aligned}
\underset{x}{\text{minimize}} \quad & f_0(x) \\
\text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \ldots, m \\
& h_j(x) = 0, \quad j = 1, \ldots, p
\end{aligned}
$$

- 拉格朗日对偶的基本思想：

处理优化问题中的约束条件

- 标准的优化问题

$$\begin{aligned}\underset{x}{\text{minimize}}\ & f_0(x)\\ \text{subject to}\ & f_i(x)\le 0,\ i=1,\dots,m\\ & h_j(x)=0,\ j=1,\dots,p\end{aligned}$$

拉格朗日函数（Lagrangian Function）

$$L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{v})=f_o(\boldsymbol{x})+\sum_{i=1}^{m}\lambda_i f_i(\boldsymbol{x})+\sum_{j=1}^{p}v_j h_j(\boldsymbol{x})$$

$$\mathrm{dom}L=\boldsymbol{D}\times\mathbf{R}^m\times\mathbf{R}^p$$

$\boldsymbol{x}$： 原变量（primal variable）

$\boldsymbol{\lambda},\boldsymbol{v}$： 对偶变量（dual variable）

λ_i,v_i： 拉格朗日乘子（Lagrange Multiplier）

下一页 上一页

拉格朗日对偶函数

$$g(\boldsymbol{\lambda},\boldsymbol{v})=\inf_{x\in D}L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{v})=\inf_{x\in D}\left(f_o(\boldsymbol{x})+\sum_{i=1}^{m}\lambda_i f_i(\boldsymbol{x})+\sum_{j=1}^{p}v_j h_j(\boldsymbol{x})\right)$$

性质：

假设$\boldsymbol{p}^*$为原问题的最优值， $f_o(\boldsymbol{x}^*)=p^*$

对$\forall\boldsymbol{\lambda}\geq 0$， $\boldsymbol{v}$，对偶函数满足以下不等式：

$$\boldsymbol{g}(\boldsymbol{\lambda},\boldsymbol{v})\leq\boldsymbol{p}^*$$

即最优值存在下界

证明：设 $\boldsymbol{x}^*$ 是最优解，则 $f_i(\boldsymbol{x}^*) \le 0$， $h_j(\boldsymbol{x}^*) = 0$，且 $f_o(\boldsymbol{x}^*) = p^*$

$$L(\boldsymbol{x}^*, \boldsymbol{\lambda}, \boldsymbol{v}) = f_o(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}^*) + \sum_{j=1}^{p} v_j h_j(\boldsymbol{x}^*) \le p^*$$

$$g(\boldsymbol{\lambda}, \boldsymbol{v}) = \inf_{\boldsymbol{x}\in \boldsymbol{D}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{v}) \le L(\boldsymbol{x}^*, \boldsymbol{\lambda}, \boldsymbol{v}) \le p^*$$

例1：

$$\min_{\boldsymbol{x}} \ \boldsymbol{x}^T\boldsymbol{x}，\text{s.t. } \boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$$

Lagrange函数为：$\boldsymbol{L}(\boldsymbol{x},\boldsymbol{v}) = \boldsymbol{x}^T\boldsymbol{x}+\boldsymbol{v}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})$

最优条件为：$\nabla \boldsymbol{L}(\boldsymbol{x},\boldsymbol{v}) = 2\boldsymbol{x}+\boldsymbol{A}^T\boldsymbol{v}=0$

$$\Rightarrow \boldsymbol{x} = -\frac{1}{2}\boldsymbol{A}^T\boldsymbol{v}$$

对偶函数为：$\boldsymbol{g}(\boldsymbol{v}) = \inf\limits_{\boldsymbol{x}} \boldsymbol{L}(\boldsymbol{x},\boldsymbol{v})$

$$=(\frac{1}{2}\boldsymbol{A}^T\boldsymbol{v})^T(\frac{1}{2}\boldsymbol{A}^T\boldsymbol{v})^T+\boldsymbol{v}^T(-\frac{1}{2}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{v}-\boldsymbol{b})$$

$$\le \inf_{\boldsymbol{x}}\{\boldsymbol{x}^T\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}\}$$

拉格朗日对偶问题

- 对偶函数给出原问题最优值的一个下界，即

$$g(\lambda,\nu) \leq p^*$$

- 如何找到最优的下界？

$$\begin{aligned} &\underset{\lambda,\nu}{\text{maximize}}\ g(\lambda,\nu) \\ &\text{subject to}\ \lambda \succeq 0 \end{aligned}$$

- 最大的下界问题，**即为Lagrange对偶问题**

原问题（Primal Problem）

$$
\begin{array}{lll}
\min & f_o(\boldsymbol{x}) & \\
\text{s.t.} & f_i(\boldsymbol{x}) \le 0 & i = 1, \cdots, m \\
 & h_j(\boldsymbol{x}) = 0 & j = 1, \cdots, p
\end{array}
$$

最优值（optimal value）： $p^* = \inf \{ f_o(\boldsymbol{x}) \mid \boldsymbol{x} \in \boldsymbol{X} \}$

对偶问题（Lagrange Dual Problem）

$$
\begin{cases}
\max & g(\boldsymbol{\lambda}, \boldsymbol{v}) \\
\text{s.t.} & \boldsymbol{\lambda} \succeq \mathbf{0}
\end{cases}
$$

最优值（optimal value）： $d^* = \inf \{ g(\boldsymbol{\lambda}, \boldsymbol{v}) \mid \boldsymbol{\lambda}, \boldsymbol{v} \in \mathrm{dom} g \ \& \ \boldsymbol{\lambda} \succeq \mathbf{0} \}$

由对偶函数的性质， $d^* \le p^*$

弱对偶性、强对偶性、对偶间隙

对偶问题的最优值 d^* 与原问题的最优值 p^* 满足：$d^* \le p^*$

$p^* - d^*$——对偶间隙（Duality Gap）

$d^* \le p^*$——弱对偶性（Weak Duality）

$d^* = p^*$——强对偶性（Strong Duality）

对偶问题例子

$$\min_{x} c^T x, \quad \text{s.t.} \quad Ax = b, x \geq 0$$

Lagrange函数为：$L(x, \lambda, \nu) = c^T x - \sum \lambda_i x_i + \nu^T (Ax - b)$

$$= -b^T \nu + (c + A^T \nu - \lambda)^T x.$$

对偶函数为：$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = -b^T \nu + \inf_x (c + A^T \nu - \lambda)^T x,$

$$g(\lambda, \nu) = \begin{cases} -b^T \nu & A^T \nu - \lambda + c = 0 \\ -\infty & \text{其他情况.} \end{cases}$$

可转化为对偶问题：

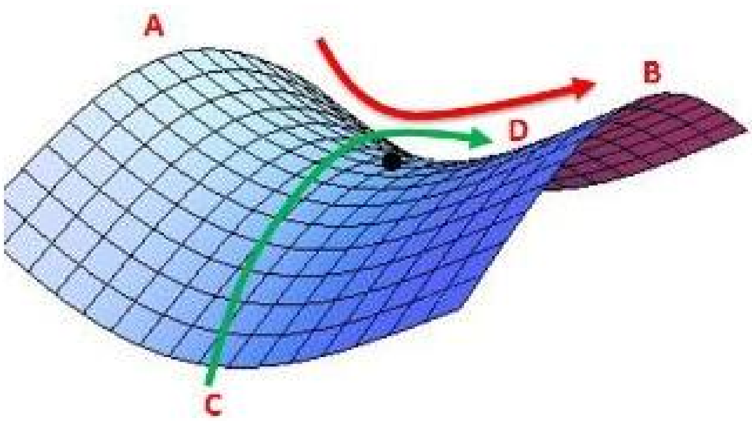
$$\max_{\lambda, \nu} -b^T \nu,$$

$$\text{s.t.} \quad A^T \nu - \lambda + c = b, \lambda \geq 0$$

下一页 上一页

3.6 鞍点

即不是最大值也不是最小值



Saddle Point- Black dot placed on the PES shows a minima along path A-B and a maxima along path C-D. It represents a transition state along path C-D which, in this case, is the reaction coordinate.

Primal-Dual最优解

若 $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{\lambda}})$ 为 $L(\boldsymbol{x}, \boldsymbol{\lambda})$ 的鞍点 $\Leftrightarrow$

对偶问题满足强对偶性，且 $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{\lambda}})$ 为 Primal-Dual 最优解

$$\inf_{\boldsymbol{x}} \sup_{\boldsymbol{\lambda} \succeq 0} L(\boldsymbol{x}, \boldsymbol{\lambda}) = \sup_{\boldsymbol{\lambda} \succeq 0} \inf_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}),$$

且 $\begin{cases} \tilde{\boldsymbol{x}} = \arg \inf_{\boldsymbol{x}} \sup_{\boldsymbol{\lambda} \succeq 0} L(\boldsymbol{x}, \boldsymbol{\lambda}) & \text{primal optimal point} \\ \tilde{\boldsymbol{\lambda}} = \arg \sup_{\boldsymbol{\lambda} \succeq 0} \inf_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}) & \text{dual optimal point} \end{cases}$

3.7 交替方向乘子方法

Alternating Direction Method of Multipliers

简称：**ADMM**

用于求解带等式约束的凸优化问题

解决多个优化变量可分问题

下一页 上一页

拉格朗日乘子法

$$\min_{x} f(x), \qquad \text{s.t.} \quad Ax = b$$

Lagrangian函数： $$L(x,\lambda) = f(x) + \lambda(Ax - b)$$

对偶函数： $$g(\lambda) = \inf_{x} L(x,\lambda)$$

对偶问题： $$\max_{\lambda} g(\lambda) = \max_{\lambda} \min_{x} L(x,\lambda)$$

求解方法：交替求解

Dual ascent algorithm:交替求解 (x^k, λ^k)

$$x^{k+1} = \arg\min_x L(x, \lambda^k)$$

$$\lambda^{k+1} = \arg\max_\lambda L(x^{k+1}, \lambda)$$

对偶变量求解：梯度方法

$$\lambda^{k+1} = \arg\max_\lambda g(\lambda) = \arg\max_\lambda < \lambda, Ax^{k+1} - b >$$

λ与$Ax^{k+1} - b$同方向时达到最大

迭代格式为：

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

下一页 上一页

交替求解 (x^k, λ^k)

$$x^{k+1} = \arg\min_{x} L(x, \lambda^k)$$

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

增广拉格朗日函数

为了提高对偶上升方法的**鲁棒性**和放松目标函数的强凸性，引入了增广拉格朗日函数

传统的拉格朗日函数：

$$\boldsymbol{L}(\boldsymbol{x},\boldsymbol{\lambda}) = \boldsymbol{f}(\boldsymbol{x}) + < \boldsymbol{\lambda},(\boldsymbol{Ax}-\boldsymbol{b}) >$$

增广拉格朗日函数：

$$\boldsymbol{L}_{\boldsymbol{r}}(\boldsymbol{x},\boldsymbol{\lambda}) = \boldsymbol{f}(\boldsymbol{x}) + < \boldsymbol{\lambda},(\boldsymbol{Ax}-\boldsymbol{b}) > + \frac{\boldsymbol{r}}{2}\| \boldsymbol{Ax}-\boldsymbol{b} \|_2^2 \text{，} \quad \boldsymbol{r} > 0$$

注意区分二者的区别

下一页 上一页

$$\min_{x} f(x), \qquad \text{s.t.} \quad Ax = b$$

augmented Lagrangian:
增广拉格朗日函数

$$L_r(x,\lambda) = f(x) + <\lambda, (Ax-b)> + \frac{r}{2}\|Ax-b\|_2^2, \quad r>0$$

乘子方法：$\max_{\lambda}\min_{x} L_r(x,\lambda)$

$$x^{k+1} = \arg\min_{x} L_r(x,\lambda^k)$$
$$\lambda^{k+1} = \lambda^{k+1} + r(Ax^{k+1} - b)$$

交替方向乘子方法

$$\min_{x} f(x)+g(z), \quad \text{s.t.} \quad Ax+Bz=c$$

augmented Lagrangian

增广拉格朗日函数：

$$L_r(x,z,\lambda)=f(x)+g(z)+<\lambda,(Ax+Bz-c)>+\frac{r}{2}\|Ax+Bz-c\|_2^2$$

ADMM算法：$\max\limits_{\lambda}\min\limits_{x,z} L_r(x,z;\lambda)$

$$(x^{k+1},z^{k+1})=\arg\min_{x,z} L_r(x,z,\lambda^k)$$

$$=\arg\min_{x,z} f(x)+g(z)<\lambda^k,(Ax+Bz-c)>+\frac{r}{2}\|Ax+Bz-c\|_2^2$$

$$\lambda^{k+1}=\arg\max_{\lambda} L_r(x^{k+1},z^{k+1},\lambda)$$

$$=\arg\max_{\lambda}<\lambda,(Ax^{k+1}+Bz^{k+1}-c)>$$

进一步分裂 $\min_{x,z} L_r(x,z;\lambda)$

ADMM算法： $\max_{\lambda}\min_{x,z} L_r(x,z;\lambda)$

固定z,λ,更新x

$$x^{k+1}=\arg\min_{x} L_r(x,z^k,\lambda^k)=f(x)+<\lambda^k,(Ax)>+\frac{r}{2}\|Ax+Bz^k-c\|_2^2$$

固定x,λ,更新z

$$z^{k+1}=\arg\min_{z} L_r(x^{k+1},z,\lambda^k)=g(z)+<\lambda^k,(Bz-c)>+\frac{r}{2}\|Ax^{k+1}+Bz-c\|_2^2$$

固定z,x,更新λ

$$\lambda^{k+1}=\arg\max_{\lambda} L_r(x^{k+1},z^{k+1},\lambda)=\arg\max_{\lambda}<\lambda,(Ax^{k+1}+Bz^{k+1}-c)>$$

ADMM算法应用实例

- 求解下列最优化问题

$$\min_{x,y}\ (x-1)^2+(y-2)^2$$

$$\begin{aligned} s.t.\ \ & 0\le x\le 3, \\ & 1\le y\le 4, \\ & 2x+3y=5 \end{aligned}$$

- 增广拉格朗日函数为：

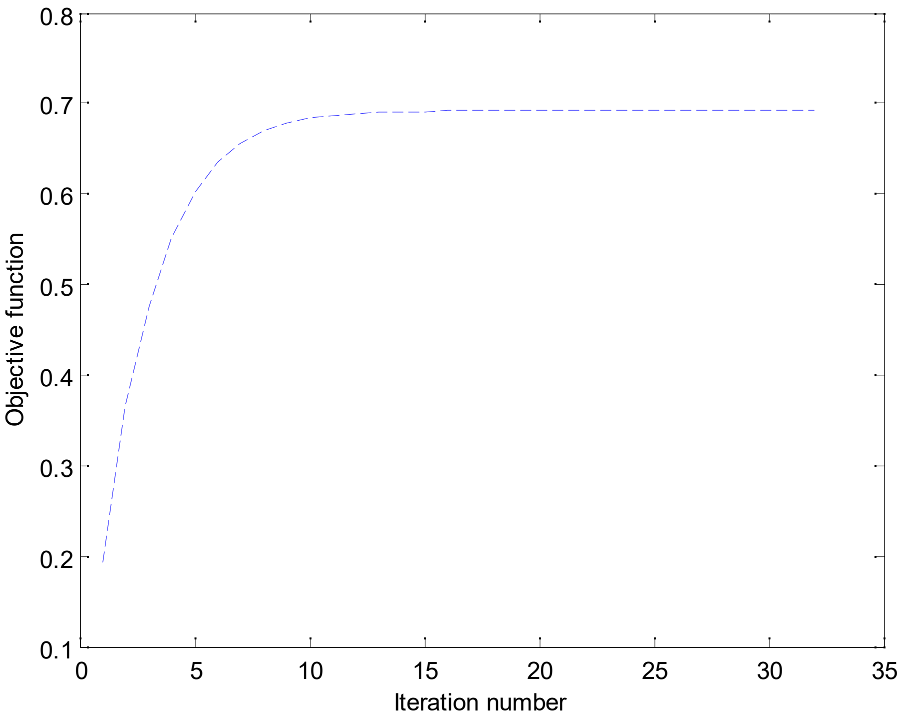
$$L_\rho(x,y,\lambda)=(x-1)^2+(y-2)^2+\lambda(2x+3y-5)+\frac{\rho}{2}(2x+3y-5)^2$$

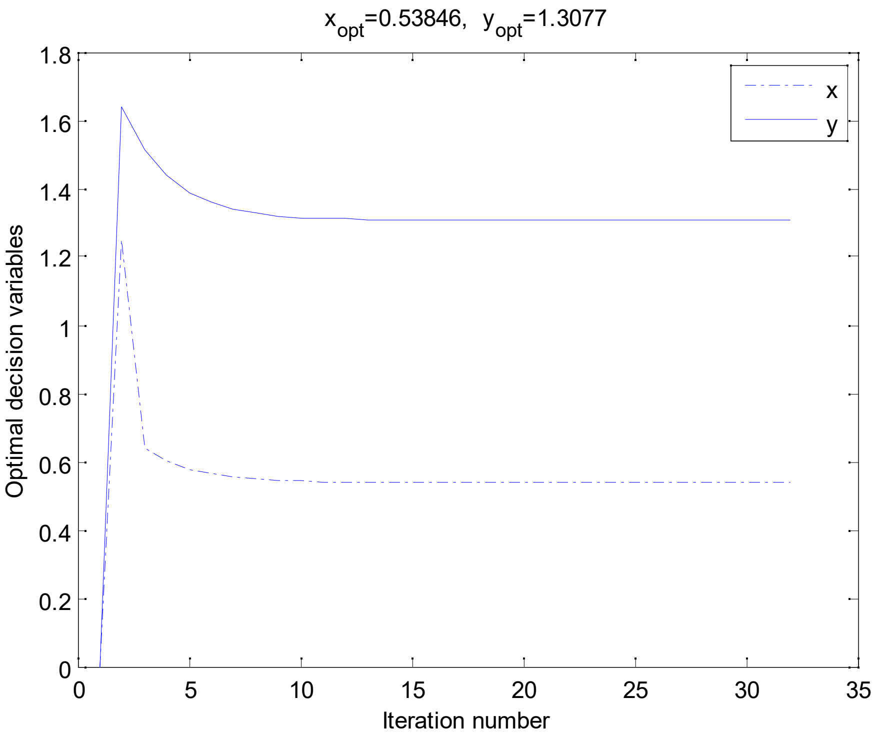
- 求解步骤如下

- 1） $x^k = y^k = \lambda^k = 0$
- 2） $x^{k+1} = \arg\min\limits_{0\le x\le 3} L_\rho(x, y^k, \lambda^k),$
- 3） $y^{k+1} = \arg\min\limits_{1\le y\le 4} L_\rho(x^{k+1}, y, \lambda^k),$
- 4） $\lambda^{k+1} = \lambda^k + \rho(2x^{k+1} + 3y^{k+1} - 5)$

  $if \left\| s^{k+1} \right\|_2 \le \varepsilon^{pri}\ and \left\| \lambda^{k+1} \right\|_2 \le \varepsilon^{dual},\ stop.\ Else,\ goto\ 2)$

- MATLAB仿真结果如下：





3.4 ADMM 在图像恢复中的应用

经典稀疏优化问题：lasso模型

$$\min_{x} \quad \frac{1}{2}\|Ax-b\|_2^2+\lambda\|x\|_1$$

L1范数为稀疏范数，具有良好的保边性质

L1范数在0处不可微，

以上无约束问题直接取梯度求解，会出现退化

可通过引入新变量，分离出该不可微项

$$\begin{cases} \min\limits_{\boldsymbol{x},\boldsymbol{y}} & \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2^2+\lambda\|\boldsymbol{y}\|_1 \\ \text{s.t.} & \boldsymbol{x}=\boldsymbol{y} \end{cases}$$

增广拉格朗日函数

$$L_c(\boldsymbol{x},\boldsymbol{y},\boldsymbol{v})=\frac{1}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2^2+\lambda\|\boldsymbol{y}\|_1+\boldsymbol{v}^T(\boldsymbol{x}-\boldsymbol{y})+\frac{c}{2}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2$$

① $\boldsymbol{x}$ 的子问题

$$\boldsymbol{x}^{k+1}=\arg\min_{\boldsymbol{x}}\frac{1}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2^2+\left(\boldsymbol{v}^k\right)^T\boldsymbol{x}+\frac{c}{2}\|\boldsymbol{x}-\boldsymbol{y}^k\|_2^2$$

② $\boldsymbol{y}$ 的子问题

$$\boldsymbol{y}^{k+1}=\arg\min_{\boldsymbol{y}}\lambda\|\boldsymbol{y}\|_1-\left(\boldsymbol{v}^k\right)^T\boldsymbol{y}+\frac{c}{2}\|\boldsymbol{y}-\boldsymbol{x}^{k+1}\|_2^2$$

③ $\boldsymbol{v}$ 的子问题

$$\boldsymbol{v}^{k+1}=\boldsymbol{v}^k+c\left(\boldsymbol{x}^{k+1}-\boldsymbol{y}^{k+1}\right)$$