智能优化算法

下一页 上一页

第3章

凸优化问题求解算法

3.2 凸优化问题：基本概念

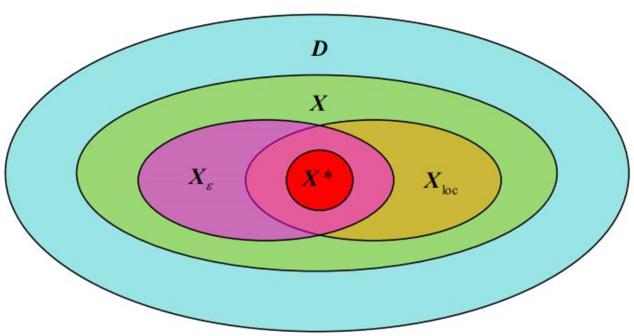
$$
\begin{aligned}
\min \quad & f_o(\boldsymbol{x}) & \\
\text{s.t.} \quad & f_i(\boldsymbol{x}) \le 0 & i = 1, \cdots, m \\
& h_j(\boldsymbol{x}) = 0 & j = 1, \cdots, p
\end{aligned}
$$

$\boldsymbol{x}$： 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） $\boldsymbol{D} \triangleq \bigcap_{i=0}^{m} \mathrm{dom} f_i \cap \bigcap_{j=0}^{p} \mathrm{dom} h_j$

可行解集（feasible set）

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

最优解集（optimal set）

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

最优值（optimal value）

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

ε 次优解（ ε -suboptimal set）

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

局部最优值（local optimal value）

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

局部最优解集（local optimal set）

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

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

3.3 无约束优化问题最优条件

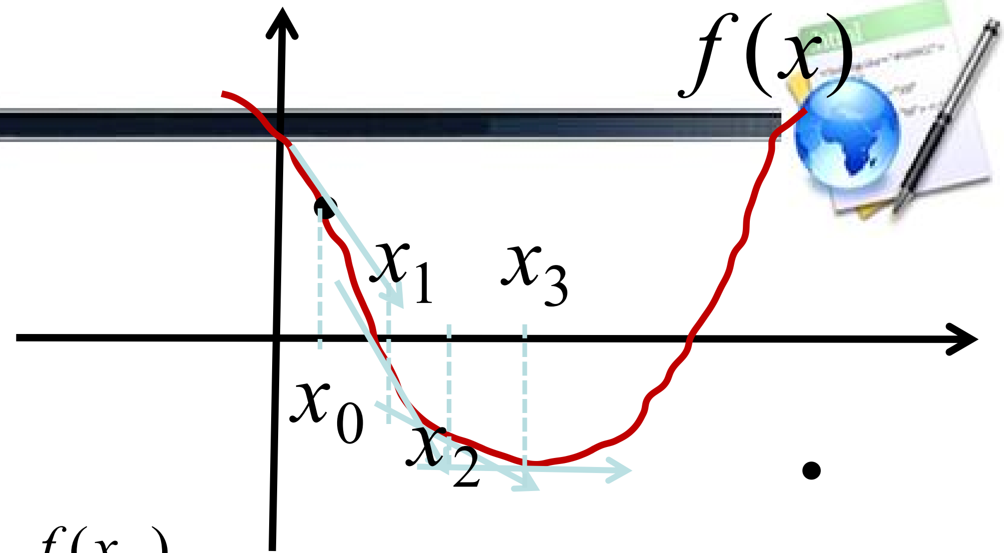
对于无约束优化问题

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

最优化条件简化为:

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

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

多维梯度下降法

$$\min_{x=(x_1,x_2\cdots,x_d)} f(x_1,x_2\cdots,x_d)=\sum_{i=1}^{n} f_i(x_1,x_2\cdots,x_d)$$

梯度下降法：

$$\nabla_{\boldsymbol{x}} f(\boldsymbol{x})=\left[\frac{\partial f(\boldsymbol{x})}{\partial x_1},\frac{\partial f(\boldsymbol{x})}{\partial x_2},\ldots,\frac{\partial f(\boldsymbol{x})}{\partial x_d}\right]^{\top}$$

$$=\sum_{i=1}^{n}\nabla f_i(x_1,x_2\cdots,x_d)$$

$$x_i^{k+1}=x_i^k-\eta\sum_{i=1}^{n}\nabla f_i(x_1^k,x_2^k\cdots,x_d^k)$$

$$i=1,2,\cdots,d$$

下一页 上一页

作业1

给定数据集$\{(x_i, y_i)\}_{i=1}^{m}$,其中$x_i = (x_{i1}, x_{i2}, \cdots, x_{in})$

请写出使用梯度下降法求解以下优化问题的步骤

$$\min_{w} h(w) = \sum_{i=1}^{m} (w_1 x_{i1} + w_2 x_{i2} + \cdots + w_n x_{in} - y_i)^2$$

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

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

- 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

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

拉格朗日函数（Lagrangian Function）

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

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

$\boldsymbol{x}$： 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

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

性质：

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

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

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

即最优值存在下界

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

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

- 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）

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

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

对偶问题（Lagrange Dual Problem）

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

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

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

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

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

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：

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

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

作业：求以下优化问题的对偶函数，对偶问题

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

*Lagrange*函数为：$L(x,\nu)= x^T x+\nu^{\mathrm{T}}(Ax-b)$

最优条件为：$\nabla L(x,\nu)=2x+A^T\nu=0$

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

对偶函数为：$g(\nu)=\inf_x L(x,\nu)$

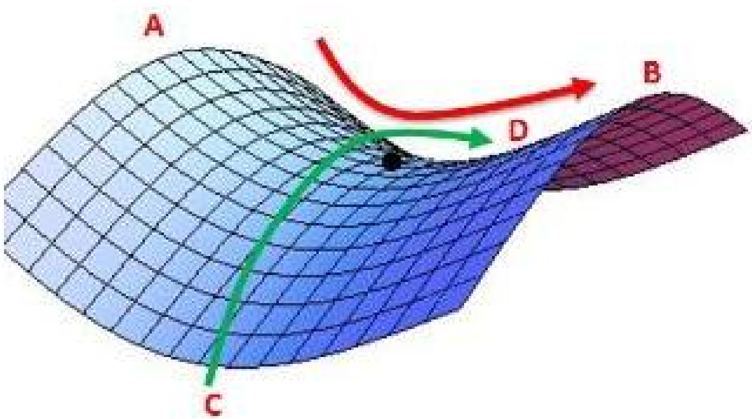
$$=(\frac{1}{2}A^T\nu)^T(\frac{1}{2}A^T\nu)^T+\nu^T(-\frac{1}{2}AA^T\nu-b)$$

$$\le \inf_x\{x^T x \mid Ax=b\}$$

对偶问题为：$\max_\nu g(\nu)=\inf_x L(x,\nu)$

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

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

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

下一页 上一页

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

拉格朗日乘子法

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

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

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

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

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：

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

增广拉格朗日函数：

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

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

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

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

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

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

下一页 上一页

多个变量交替方向乘子方法

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

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 $\min\limits_{x,z} L_r(x,z;\lambda)$

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

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

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

固定z,x，更新λ

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

下一页 上一页

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：

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

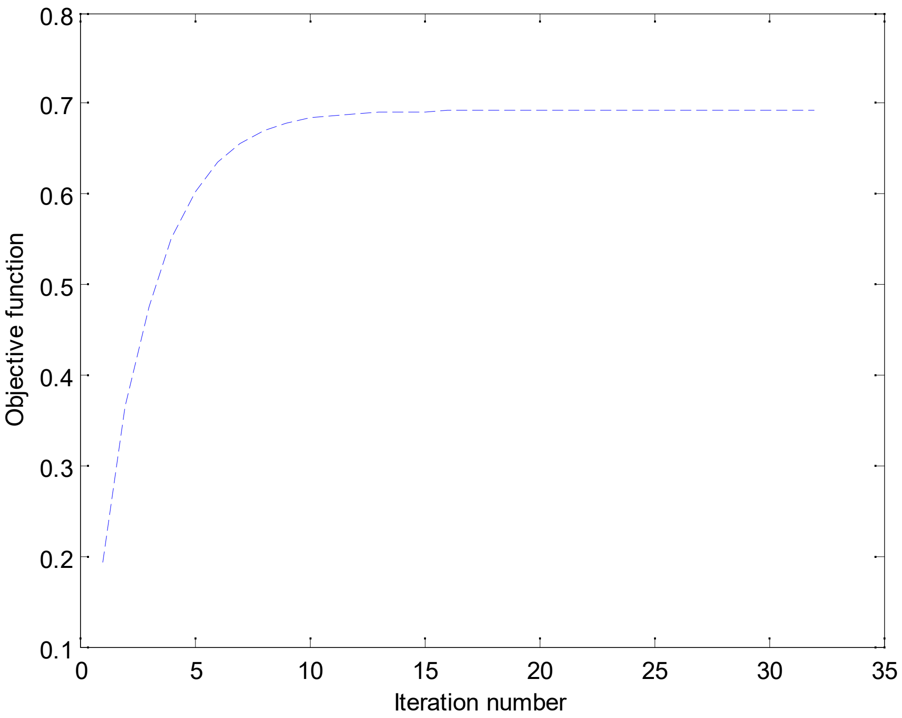
下一页 上一页

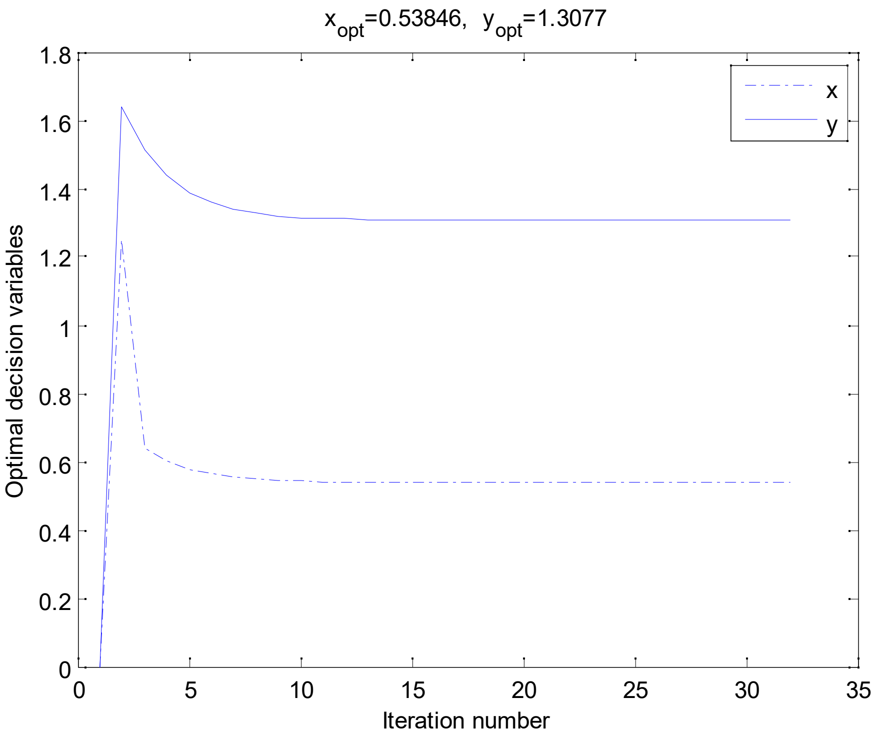
- 求解步骤如下

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





3.3 ADMM 在图像恢复中的应用

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

梯度下降法求解

$$\nabla f(x)=A^T(Ax-b)+2\lambda x$$

$$\begin{aligned} x^{k+1} &= x^k-\alpha\nabla f(x) \\ &= x^k-\alpha[A^T(Ax-b)+\lambda x^k] \end{aligned}$$

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

梯度下降法求解

$$\nabla f(x)=A^T(Ax-b)+\lambda\frac{x}{|x|}$$

在0处不可微
求解算法不稳定

$$x^{k+1}=x^k-\alpha\nabla f(x)$$

$$=x^k-\alpha[A^T(Ax^k-b)+\lambda\frac{x^k}{|x^k|}]$$

可通过引入新变量，分离出该不可微项

如何分离出该不可微项？？？？

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

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

转换该等式约束优化问题

增广拉格朗日函数

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

下一页 上一页

① $\boldsymbol{x}$ 的子问题

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

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

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

作业2

- 请写出使用ADMM算法求解以下优化问题的算法详细步骤

$$\min_{\boldsymbol{u}} \|\nabla \boldsymbol{u}\|_1 + \alpha \|\boldsymbol{u} - \boldsymbol{u}^0\|_2^2$$