一类涉及矩阵范数的优化问题

作者：凯鲁嘎吉 - 博客园 http://www.cnblogs.com/kailugaji/

这篇博文主要探讨一下有关矩阵范数的优化问题，我们知道，矩阵按行或列拆开就是向量，因此矩阵范数优化问题在某种程度上可以转化为向量范数的优化，而向量拆开是一个个数值，因此向量优化问题就进一步退化为单变量优化。因此，我们从最基础的单变量优化出发，单变量优化问题搞清楚之后，再逐步拓展到向量、矩阵范数的优化上。文中涉及到向量范数、矩阵范数、软阈值算子(soft thresholding/shrinkage operator)与奇异值收缩算子(singular value shrinkage operator)等概念。这类优化问题在矩阵补全/填充、压缩感知、稀疏表达、低秩矩阵分解(比如张长青2015在ICCV会议上发表的Low-Rank Tensor Constrained Multiview Subspace Clustering中E与G的求解问题)、鲁棒主成分分析(Robust PCA)等领域有广泛的应用。尤其是矩阵的$L_{2,1}$范数与矩阵的核范数/迹范数(nuclear/trace norm)在去噪去冗余降维等方面应用最广。

1. 向量范数与矩阵范数

一类涉及矩阵范数的优化问题

➢向量的范数 $x \in \mathbf{R}^n$

- 向量的0范数

向量x中非0元素的个数

- 向量的1范数

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

- 向量的2范数

$$\|x\|_2 = (x, x)^{\frac{1}{2}} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

- 向量的p范数

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}, p \in [1, \infty)$$

- 向量的无穷范数

$$\|x\|_\infty = \max_{1 \le i \le n} |x_i|$$

➢矩阵的范数 $\mathrm{A} \in \mathbf{R}^{m \times n}$

- 矩阵的列范数 $\|\mathrm{A}\|_1 = \max_{1 \le j \le n} \sum_{i=1}^{m} |a_{ij}|$

- 矩阵的行范数 $\|\mathrm{A}\|_\infty = \max_{1 \le i \le m} \sum_{j=1}^{n} |a_{ij}|$

- 矩阵的2范数 $\|\mathrm{A}\|_2 = \sqrt{\lambda_{\max}(\mathrm{A}^{\mathrm{T}}\mathrm{A})}$

其中 $\lambda_{\max}(\mathrm{A}^{\mathrm{T}}\mathrm{A})$表示 $\mathrm{A}^{\mathrm{T}}\mathrm{A}$ 的最大特征值

- 矩阵的Frobenius范数

$$\|\mathrm{A}\|_F = \sqrt{\mathrm{tr}(\mathrm{A}^{\mathrm{T}}\mathrm{A})} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2}$$

- $L_{2,1}$范数 $\|\mathrm{A}\|_{2,1} = \sum_{i=1}^{m} \sqrt{\sum_{j=1}^{n} a_{ij}^2}$

- 核范数：矩阵的奇异值之和(用来约束低秩low-rank) $\mathrm{A} = U\Sigma V^T$

$$\|\mathrm{A}\|_* = \mathrm{tr}\left(\sqrt{\mathrm{A}^{\mathrm{T}}\mathrm{A}}\right) = \mathrm{tr}(\Sigma)$$

1

2．单变量绝对值优化问题

➢ 单变量绝对值优化问题

- 问题描述：已知x与y是数值，给定如下优化问题，y已知，α>0，求x的最优解

$$\min_x \; J(x) = \alpha|x| + \frac{1}{2}|x-y|^2$$

这里假定0/0=0，$|x|$表示x的绝对值

- x的最优解如下：

$$\hat{x} = \frac{y}{|y|}\max(|y|-\alpha,0) = \mathrm{sgn}(y)\max(|y|-\alpha,0) = soft(y,\alpha)$$

- 其中sgn()为符号函数

$$\mathrm{sgn}(y) = \frac{y}{|y|} = \begin{cases} 1, & y>0; \\ 0, & y=0; \\ -1, & y<0. \end{cases}$$

- soft()为软阈值算子(soft thresholding operator)或收缩算子(shrinkage operator)

$$soft(y,\alpha) = \mathrm{sgn}(y)\max(|y|-\alpha,0) = \begin{cases} y+\alpha, & y \le -\alpha; \\ 0, & |y| \le \alpha; \\ y-\alpha, & y \ge \alpha. \end{cases}$$

➢ 单变量绝对值优化问题

- 解：$\dfrac{\partial J(x)}{\partial x}=x-y+\alpha\dfrac{x}{|x|}=0\Rightarrow y=x+\alpha\dfrac{x}{|x|}=\dfrac{|x|+\alpha}{|x|}x$

若$y>0$，则$x\geq 0$.

当$x=0$，$J(x)=\frac{1}{2}y^2$;

当$x>0$，$y=\alpha+x\Rightarrow x=y-\alpha, J(x)=\alpha(y-\alpha)+\frac{1}{2}\alpha^2=\alpha y-\frac{1}{2}\alpha^2$.

$\frac{1}{2}y^2-\left(\alpha y-\frac{1}{2}\alpha^2\right)=\frac{1}{2}(y-\alpha)^2\geq 0$. $\therefore x=\max(y-\alpha,0)$.

若$y<0$，则$x\leq 0$.

当$x=0$，$J(x)=\frac{1}{2}y^2$;

当$x<0$，$y=x-\alpha\Rightarrow x=y+\alpha, J(x)=-\alpha(y+\alpha)+\frac{1}{2}\alpha^2=-\alpha y-\frac{1}{2}\alpha^2$.

$\frac{1}{2}y^2-\left(-\alpha y-\frac{1}{2}\alpha^2\right)=\frac{1}{2}(y+\alpha)^2\geq 0$. $\therefore x=\min(y+\alpha,0)$.

若$y=0$，则$x=0$或$x>0$或$x<0$.

当$x>0$，$x=-\alpha$，$J(x)=-\frac{1}{2}\alpha^2<0$(舍).

当$x=0$，$J(x)=0$;

当$x<0$，$x=\alpha$，$J(x)=-\frac{1}{2}\alpha^2<0$(舍).

$\therefore x=0$.

$$
\begin{aligned}
\therefore \hat{x}&=\max(y-\alpha,0)+\min(y-\alpha,0)\\
&=\frac{y}{|y|}\max(|y|-\alpha,0)\\
&=\operatorname{sgn}(y)\max(|y|-\alpha,0)\\
&=soft(y,\alpha)
\end{aligned}
$$

➢ 单变量绝对值优化问题

- 从结果引出思考

$$\hat{x} = soft(y,\alpha) = \mathrm{sgn}(y)\max(|y|-\alpha,0) = \begin{cases} y+\alpha, & y \le -\alpha; \\ 0, & |y| \le \alpha; \\ y-\alpha, & y \ge \alpha. \end{cases}$$

- 由 $y = x + \alpha \dfrac{x}{|x|} = \dfrac{|x|+\alpha}{|x|} x$ 及上述结果可知，

当|y|>α时，x与y呈正线性相关，**两者同向**。因此有x与y这两者的单位长度相等，即 $\dfrac{\partial |x|}{\partial x} = \dfrac{x}{|x|} = \dfrac{y}{|y|} \ (|y| > \alpha)$

$$\therefore \ x - y + \alpha \frac{x}{|x|} = x - y + \alpha \frac{y}{|y|} = 0$$

$$\Rightarrow \hat{x} = y - \alpha \frac{y}{|y|} = \frac{y}{|y|}(|y|-\alpha) \ (|y| > \alpha)$$

$$= \mathrm{sgn}(y)\max(|y|-\alpha,0)$$

解析解

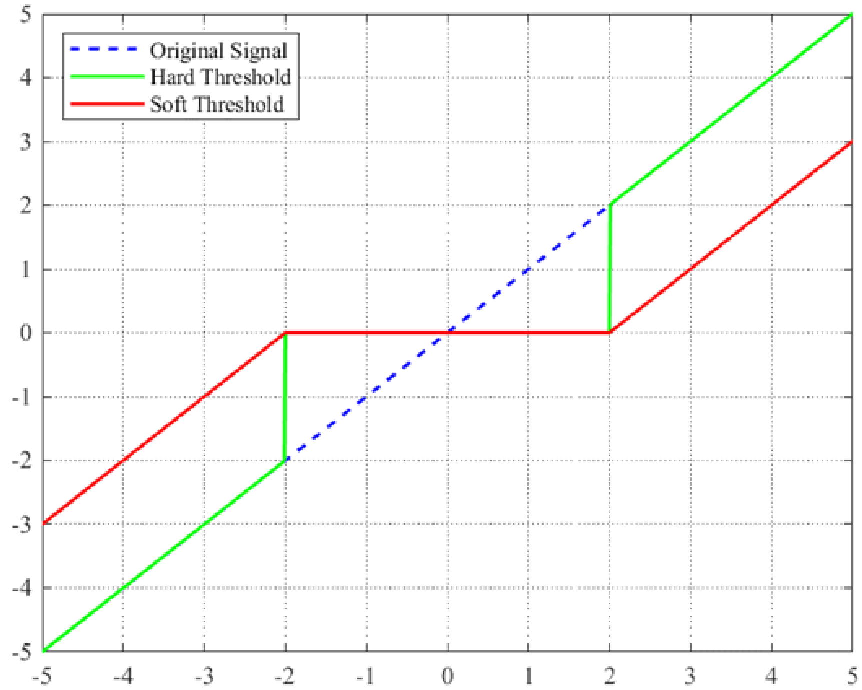
一类涉及矩阵范数的优化问题

➢ 单变量绝对值优化问题(MATLAB程序实现)

$$\hat{x} = soft(y,\alpha) = \operatorname{sgn}(y)\max(|y|-\alpha,0) = \begin{cases} y+\alpha, & y \le -\alpha; \\ 0, & |y| \le \alpha; \\ y-\alpha, & y \ge \alpha. \end{cases}$$

Flag=4

Flag=3



蓝虚线：原始信号，绿线：硬阈值函数，红线：软阈值函数。(α=2，横坐标y，纵坐标x)

5

3．向量范数优化问题

➢ 向量范数优化问题

- 问题描述：已知$\boldsymbol{x}$与$\boldsymbol{y}$是两个向量，给定如下优化问题，$\boldsymbol{y}$已知，$\alpha>0$，求$\boldsymbol{x}$的最优解

$$\min_{\boldsymbol{x}} \quad J(\boldsymbol{x}) = \alpha\|\boldsymbol{x}\| + \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{y}\|^2$$

这里假定0/0=0，$\|\boldsymbol{x}\|$表示向量$\boldsymbol{x}$的某种范数

- 解：$\dfrac{\partial J(\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{x}-\boldsymbol{y}+\alpha\dfrac{\boldsymbol{x}}{\|\boldsymbol{x}\|} = 0 \Rightarrow \boldsymbol{y} = \boldsymbol{x}+\alpha\dfrac{\boldsymbol{x}}{\|\boldsymbol{x}\|} = \dfrac{\|\boldsymbol{x}\|+\alpha}{\|\boldsymbol{x}\|}\boldsymbol{x} \Rightarrow \boxed{\boldsymbol{x} = \dfrac{\|\boldsymbol{x}\|}{\|\boldsymbol{x}\|+\alpha}\boldsymbol{y}}$

数值解

由$\dfrac{\partial\|\boldsymbol{x}\|}{\partial \boldsymbol{x}} = \dfrac{\boldsymbol{x}}{\|\boldsymbol{x}\|} = \dfrac{\boldsymbol{y}}{\|\boldsymbol{y}\|} \; (\|\boldsymbol{y}\|>\alpha) \quad \therefore\; \boldsymbol{x}-\boldsymbol{y}+\alpha\dfrac{\boldsymbol{x}}{\|\boldsymbol{x}\|} = \boldsymbol{x}-\boldsymbol{y}+\alpha\dfrac{\boldsymbol{y}}{\|\boldsymbol{y}\|} = 0$

$$\boxed{\begin{aligned}\Rightarrow \hat{\boldsymbol{x}} &= \boldsymbol{y}-\alpha\frac{\boldsymbol{y}}{\|\boldsymbol{y}\|} = \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|}(\|\boldsymbol{y}\|-\alpha) \; (\|\boldsymbol{y}\|>\alpha)\\ &= \operatorname{sgn}(\boldsymbol{y})\max(\|\boldsymbol{y}\|-\alpha, 0)\end{aligned}}$$

解析解

4. 矩阵$L_{2,1}$范数优化问题

➢ 矩阵$L_{2,1}$范数优化问题

- 问题描述：已知X与Y是两个矩阵(大小为N*M)，给定如下优化问题，Y已知，α>0，求X的最优解

$$\min_X \quad J(X)=\alpha\|X\|_{2,1}+\frac{1}{2}\|X-Y\|_F^2$$

这里假定0/0=0，$\|X\|$表示矩阵X的某种范数

- 解：$\|X\|_{2,1}=\sum_{i=1}^{N}\sqrt{\sum_{j=1}^{M}x_{ij}^2}=\sum_{i=1}^{N}\|X_{i\cdot}\|_2$

$$\frac{\partial J(X_{i\cdot})}{\partial X_{i\cdot}}=X_{i\cdot}-Y_{i\cdot}+\alpha\frac{\partial\|X_{i\cdot}\|_2}{\partial X_{i\cdot}}=X_{i\cdot}-Y_{i\cdot}+\alpha\frac{X_{i\cdot}}{\|X_{i\cdot}\|_2}=0\Rightarrow \boxed{X_{i\cdot}=\frac{\|X_{i\cdot}\|_2}{\|X_{i\cdot}\|_2+\alpha}Y_{i\cdot}}$$

数值解

由$\dfrac{\partial\|X_{i\cdot}\|_2}{\partial X}=\dfrac{X_{i\cdot}}{\|X_{i\cdot}\|_2}=\dfrac{Y_{i\cdot}}{\|Y_{i\cdot}\|_2}\quad(\|Y_{i\cdot}\|_2>\alpha)$

$$\Rightarrow \hat{X}_{i\cdot}=Y_{i\cdot}-\alpha\frac{Y_{i\cdot}}{\|Y_{i\cdot}\|_2}\quad(\|Y_{i\cdot}\|_2>\alpha)=\frac{Y_{i\cdot}}{\|Y_{i\cdot}\|_2}\left(\|Y_{i\cdot}\|_2-\alpha\right)\quad(\|Y_{i\cdot}\|_2>\alpha)$$

$$=\boxed{\frac{Y_{i\cdot}}{\|Y_{i\cdot}\|_2}\max(\|Y_{i\cdot}\|_2-\alpha,0)}$$

解析解

7

5．矩阵核范数优化问题

➢ 矩阵核范数优化问题

- 问题描述：已知X与Y是两个矩阵，给定如下优化问题，Y已知，α>0，求X的最优解

$$\min_X \quad J(X) = \alpha\|X\|_* + \frac{1}{2}\|X-Y\|_F^2$$

这里假定0/0=0，$\|X\|$表示矩阵X的某种范数

- 解：$\dfrac{\partial J(X)}{\partial X} = X - Y + \alpha\dfrac{\partial\|X\|_*}{\partial X} = 0$

考虑特征值分解$X = U_X\Sigma_X V_X^T$，$Y = U_Y\Sigma_Y V_Y^T$

Σ中元素非负，U与V为正交阵，$U^TU = UU^T = I$.

$$\|X\|_* = tr\left(\sqrt{X^TX}\right) = tr\left(\sqrt{(U_X\Sigma_X V_X^T)^T(U_X\Sigma_X V_X^T)}\right)$$

$$= tr\left(\sqrt{V_X\Sigma_X^2 V_X^T}\right) = tr\left(\sqrt{V_X^TV_X\Sigma_X^2}\right) = tr\left(\sqrt{\Sigma_X^2}\right) = tr(\Sigma_X)$$

(用到迹的循环性, Σ非负性, U与V正交矩阵的性质)

$$\Rightarrow \frac{\partial\|X\|_*}{\partial X} = \frac{\partial tr(\Sigma_X)}{\partial X} = \frac{tr(\partial\Sigma_X)}{\partial X}$$

$$= \frac{tr\left(U_X^T\partial X V_X\right)}{\partial X} = \frac{tr\left(V_X U_X^T\partial X\right)}{\partial X}$$

$$= (V_X U_X^T)^T = U_X V_X^T$$

$$\partial X = \partial U_X\Sigma_X V_X^T + U_X\partial\Sigma_X V_X^T + U_X\Sigma_X\partial V_X^T$$

$$\therefore U_X\partial\Sigma_X V_X^T = \partial X - \partial U_X\Sigma_X V_X^T - U_X\Sigma_X\partial V_X^T$$

等式两边左乘U_X^T，右乘V_X

$$\therefore U_X^TU_X\partial\Sigma_X V_X^TV_X = U_X^T\partial XV_X - U_X^T\partial U_X\Sigma_X V_X^TV_X - U_X^TU_X\Sigma_X\partial V_X^TV_X$$

$$\therefore \partial\Sigma_X = U_X^T\partial XV_X - U_X^T\partial U_X\Sigma_X - \Sigma_X\partial V_X^TV_X$$

$$\therefore tr(\partial\Sigma_X) = tr\left(U_X^T\partial XV_X - U_X^T\partial U_X\Sigma_X - \Sigma_X\partial V_X^TV_X\right)$$

$$= tr\left(U_X^T\partial XV_X\right)$$

(对角阵乘以反对称阵为零矩阵)

一类涉及矩阵范数的优化问题

➢ 矩阵核范数优化问题

- 问题描述：已知X与Y是两个矩阵，给定如下优化问题，Y已知，$\alpha>0$，求X的最优解

$$\min_{X} \quad J(X) = \alpha \|X\|_* + \frac{1}{2}\|X - Y\|_F^2$$

这里假定0/0=0，$\|X\|$表示矩阵X的某种范数

- 解：$\dfrac{\partial J(X)}{\partial X} = X - Y + \alpha \dfrac{\partial \|X\|_*}{\partial X} = X - Y + \alpha U_X V_X^T = 0 \quad \Rightarrow X = Y - \alpha U_X V_X^T$

由 $\dfrac{\partial \|X\|_*}{\partial X} = U_X V_X^T = U_Y V_Y^T \ (\Sigma_Y > \alpha I)$

$$\begin{aligned}
\Rightarrow \hat{X} &= Y - \alpha U_Y V_Y^T \ (\Sigma_Y > \alpha I) \\
&= U_Y \Sigma_Y V_Y^T - \alpha U_Y V_Y^T = U_Y(\Sigma_Y - \alpha I)V_Y^T \ (\Sigma_Y > \alpha I) \\
&= \boxed{U_Y \max(\Sigma_Y - \alpha I, 0) V_Y^T = D(Y, \alpha)}
\end{aligned}$$

解析解

奇异值阈值/收缩算子

9

6. MATLAB程序

这个程序主要集中于单变量绝对值优化与向量的2范数优化问题。flag=1是向量的优化，其中我们用两种方法进行求解，解析解与数值解，从结果可以看出，在数值解拥有足够的迭代次数时，两者结果大体一致。flag=2是单变量绝对值优化问题，可以直观看出变量x与最优值之间的函数图像。flag=3与flag=4是软阈值函数图像，结果见PPT第5页。

```
clear
clc
flag=2;  % flag=1, 向量; flag=2, 单变量(一个数值); flag=3, MATLAB自带的软阈值函数;  flag=4, 自定义软阈值函数
%% 向量的情况 2D shrinkage formula
% 最小化问题：z(x)=alpha*||x||+(beta/2)*||x-y||^2
% alpha=0.5;
% beta=1;
% 用的2范数
if flag==1
    % 法1：解析解
    % 当||y||>alpha，x=(||y||-alpha/beta).*(y/||y||)
    y=0.1:0.1:1;
    num=length(y);
    yy=norm(y, 2);
    x_1=((yy-0.5)/yy).*y;
    z_1=0.5*norm(x_1, 2)+0.5.*(x_1-y)*(x_1-y)';
    % 法2：数值解--只要重复次数够大，能逼近解析解的结果
    % x=(||y||/((alpha/beta)+||y||)).*y
    N=50; %重复迭代N次
    x_2=rand(1, num);
    for i=1:N
        xx=norm(x_2, 2);
        x_2=(xx/(0.5+xx)).*y;
    end
    z_2=0.5*norm(x_2, 2)+0.5.*(x_2-y)*(x_2-y)';
    % 结果
    x_result=[x_1; x_2] % x的结果, 第一行是方法1的结果, 第二行是方法2的结果
    z=[z_1; z_2] % 最小化问题z的结果, 第一行是方法1的结果, 第二行是方法2的结果

%% 单变量的情况 1D shrinkage formula
elseif flag==2
    % z(x)=(beta/2)*(x-y)^2+alpha*|x|
    % alpha=0.5, beta=1
    % 当|y|>alpha, x=(|y|-alpha)*(y/|y|)
    x=0.1:0.1:1;
    y=1;
    z=0.5.*(x-y).^2+0.5.*abs(x);
    [z_min, index]=min(z);
```

```
    plot(x, z);
    hold on
    plot(x(index), z_min, 'ro');
    xlabel('x');
    ylabel('z');
    x_result=x(index)
    z=z_min
%% MATLAB自带的软阈值函数  1D shrinkage formula
elseif flag==3
    x=-5:0.01:5;
    thr=2;
    %  z(x)=(beta/2)*(x-y)^2+alpha*|x|
    % thr=alpha/beta
    ysoft=wthresh(x,'s',thr); % 软阈值
    ythard = wthresh(x,'h',thr); % 硬阈值
    plot(x, x, 'b--', x, ythard, 'g-', x, ysoft, 'r-', 'LineWidth',1.3);
    legend('Original Signal','Hard Threshold','Soft Threshold', 'Location', 'northwest'); %图例的设置
    % 设置网格线
    grid on;
    set(gca, 'FontName','Times New Roman');
    set(gca, 'GridLineStyle', ':');
    set(gca, 'GridAlpha', 1);
    saveas(gcf,sprintf('matlab_shrinkage.jpg'),'bmp'); %保存图片
%% 自定义软阈值函数  1D shrinkage formula
elseif flag==4
    x=-5:0.01:5;
    thr=2;
    y=Soft_Threshold(x,thr);
    plot(x, y, 'r-', 'LineWidth',1.3)
    title('Soft Threshold');
    % 设置网格线
    grid on;
    set(gca, 'FontName','Times New Roman');
    set(gca, 'GridLineStyle', ':');
    set(gca, 'GridAlpha', 1);
    saveas(gcf,sprintf('soft_shrinkage.jpg'),'bmp'); %保存图片
end
% 自定义软阈值函数
```

```
%  z(x)=(beta/2)*(x-y)^2+alpha*|x|
% thr=alpha/beta
function y=Soft_Threshold(x,thr)
    y=sign(x).*max(abs(x) - thr,0);
end
```

当flag=1时，给出向量最优化问题，x_result与z的第一行是解析解的结果，第二行是数值解的结果，可以看到在一定精度范围内两者结果一致。(前提是数值解需要一定的迭代次数)

```
x_result =

    0.0745    0.1490    0.2236    0.2981    0.3726    0.4471    0.5216    0.5961    0.6707    0.7452
    0.0745    0.1490    0.2236    0.2981    0.3726    0.4471    0.5216    0.5961    0.6707    0.7452


z =

    0.8561
    0.8561
```

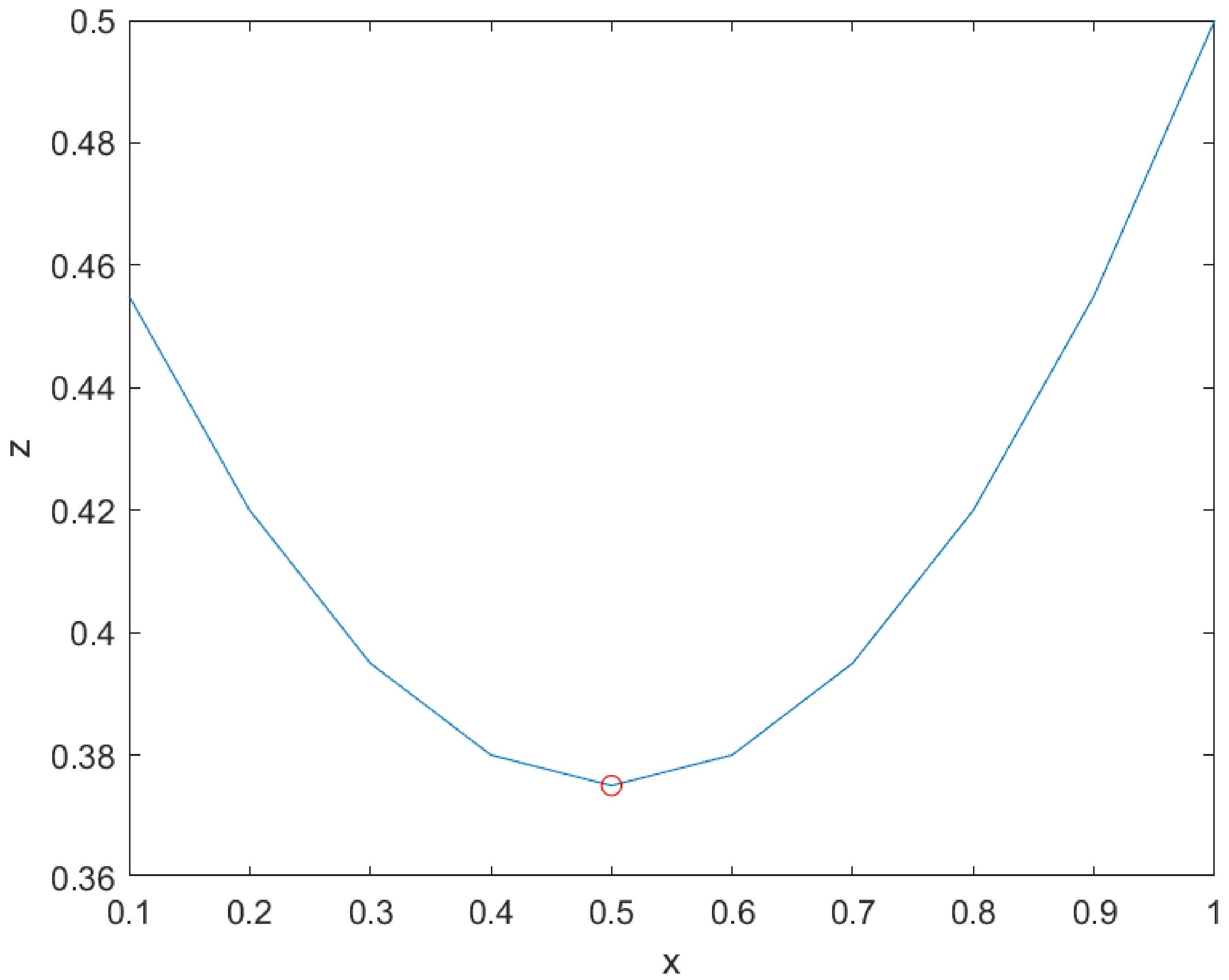
当flag=2时，可以从图中看出变量x与目标函数之间的曲线图。

```
x_result =

    0.5000


z =

    0.3750
```



7. 参考文献

[1] Yang, Junfeng & Yin, Wotao & Zhang, Yin & Wang, Yilun. (2009). A Fast Algorithm for Edge-Preserving Variational Multichannel Image Restoration. SIAM J. Imaging Sciences. 2. 569-592.

[2] Liu, Jun & Ji, Shuiwang & Ye, Jieping. (2009). Multi-Task Feature Learning Via Efficient l2, 1-Norm Minimization. Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence, UAI 2009. 339-348.

[3] Peng, Yali & Sehdev, Paramjit & Liu, Shigang & Li, Jun & Wang, Xili. (2018). l2,1-norm minimization based negative label relaxation linear regression for feature selection. Pattern Recognition Letters. 116.

[4] Cai, Jian-Feng & Candès, Emmanuel & Shen, Zuowei. (2010). A Singular Value Thresholding Algorithm for Matrix Completion. SIAM Journal on Optimization. 20. 1956-1982.

[5] 最优化之Robust PCA - 博客园 - quarryman

[6] linear algebra - Derivative of the nuclear norm - Mathematics Stack Exchange

[7] 2010 Proximal gradient method (PPT)

[8] Toh, Kim-Chuan & Yun, Sangwoon. (2010). An Accelerated Proximal Gradient Algorithm for Nuclear Norm Regularized Least Squares Problems. Pacific Journal of Optimization. 6.

[9] ECE236C - Optimization Methods for Large-Scale Systems http://www.seas.ucla.edu/~vandenbe/ee236c.html

[10] 潘少华, 文再文. 低秩稀疏矩阵优化问题的模型与算法[J]. 运筹学学报, 2020(3):1-26.