Homework 5 for "Convex Optimization" Part. 2

1500010611 汪祎非

We consider the $l1$-regularized problem

$$\min_x \frac{1}{2}\|Ax - b\|_2^2 + \mu\|x\|_1 \tag{1}$$

where $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$ and $\mu > 0$ are given. We denote $f(x) = \frac{1}{2}\|Ax - b\|_2^2 + \mu\|x\|_1$, $g(x) = \frac{1}{2}\|Ax - b\|_2^2$, $h(x) = \|x\|_1$.

1 Problem 3(c)

We consider to apply the continuation strategy and Barzila-Borwein step size. We have three parameters α, M_1, M_2 for continuation. We set $\mu_{max} = \max\{\alpha\|A^T b\|_\infty, \mu\}$. Then, we take $\mu_0 = \mu_{max}$.

We consider h_λ, the Huber penalty approximation of h:

$$h_\lambda(x) = \sum_{l=1}^{n} h_\lambda^{(l)}(x) \tag{2}$$

where

$$h_\lambda^{(l)}(x) = \begin{cases} x_l^2/(2\lambda), & |x_l| \le \lambda \\ |x_l| - \lambda/2, & |x_l| > \lambda \end{cases} \tag{3}$$

We set $\lambda_0 = 10^{-1}$ and $\lambda = 10^{-6}$. We also have an extra parameter β for the decay of λ_0. Then, we denote $f_{i,j}(x) = g(x) + \mu_i h_{\lambda_j}(x)$. The gradient of $f_{i,j}(x)$ is

$$\nabla f_{i,j}(x)_l = \begin{cases} (A^T Ax - A^T b)_l + \mu_i x_l/(\lambda_j), & |x_l| \le \lambda_j \\ (A^T Ax - A^T b)_l + \mu_i \mathrm{sign}(x_l), & |x_l| > \lambda_j \end{cases}$$

In $(k+1)$-th iteration, if $k=0$, we use the initial step size s. Otherwise, we calculate the BB step size s_k:

$$s_k = \frac{(x_k - x_{k-1})^T(x_k - x_{k-1})}{(x_k - x_{k-1})^T(\nabla f_{i,j}(x_k) - \nabla f_{i,j}(x_{k-1}))}$$

Then, we update x_{k+1} by

$$x_{k+1} = x_k - s_k \nabla f_{i,j}(x_k) \tag{sgrad-upd}$$

and set $k = k+1$.

In the case that $\mu_i > \mu$ or $\lambda_j > \lambda$, after M_1 iterations, we update both μ_{i+1} and λ_{j+1} by

$$\mu_{i+1} = \max\{\mu, \alpha \min\{\|\nabla g(x_k)\|_\infty, \mu_i\}\} \tag{4}$$

$$\lambda_{j+1} = \max\{\beta\lambda_j, \lambda\} \tag{5}$$

and set $i = i+1$, $j = j+1$. Then, we calculate $f_{i,j}(x_k)$ and $\nabla f_{i,j}(x_k)$, reset the step size $s_k = \min\{s, \lambda_j\}$, update x_{k+1} by (sgrad-upd) and set $k = k+1$. Because the Huber penalty approximation is λ_j-continuous.

If $\lambda_j = \lambda$, $\mu_i = \mu$, we stop the algorithm after M_2 iterations. The gradient method for smoothed primal problem with continuation strategy is given below:

Algorithm 1 Gradient method for smoothed primal problem with continuation strategy

Input: initial guess x_0, step size s, continuation parameter α, M_1, M_2, lambda decay parameter β.

Calculate $\mu_{max} = \max\{\alpha\|A^T b\|_\infty, \mu\}$. Let $i=0$, $\mu_0 = \mu_{max}$, $s_0 = s$, $k=0$.
Update x_{k+1} by (sgrad-upd), $k = k+1$
while $\lambda_j > \lambda$ or $\mu_i > \mu$ **do**
  for $l = 1 : M_1$ **do**
    Calculate BB step size s_k, update x_{k+1} by (sgrad-upd), $k = k+1$
  end for
  $\mu_{i+1} = \max\{\mu, \alpha \min\{\|\nabla g(x_k)\|_\infty, \mu_i\}\}$, $\lambda_{j+1} = \max\{\beta\lambda_j, \lambda\}$, $i = i+1$, $j = j+1$
  Calculate $f_{i,j}(x_k)$ and $\nabla f_{i,j}(x_k)$
  Reset $s_k = \min\{s, \lambda_j\}$, update x_{k+1} by (sgrad-upd), $k = k+1$
end while
for $l = 1 : M_2$ **do**
  Calculate BB step size s_k, update x_{k+1} by (sgrad-upd), $k = k+1$
end for

In practice, we take $s = 4 \times 10^{-4}$, $\alpha = 0.1$, $M_1 = 200$, $M_2 = 300$, $\beta = 0.1$.

2 Problem 3(d)

We consider to apply the continuation strategy. Similarly, we have three parameters α, M_1, M_2 for continuation. We set $\mu_{max} = \max\{\alpha \|A^T b\|_\infty, \mu\}$. Then, we take $\mu_0 = \mu_{max}$.

We set $\lambda_0 = 10^{-3}$ and $\lambda = 10^{-6}$. We also have an extra parameter β for the decay of λ_0. Then, we denote $f_{i,j}(x) = g(x) + \mu_i h_{\lambda_j}(x)$.

We set $x_{-1} = x_0$. In $(k+1)$-th iteration, we update x_{k+1} by

$$\begin{cases} y = x_k + \dfrac{k-1}{k+2}(x_k - x_{k-1}) \\ x_{k+1} = y - s_k \nabla g(x_k) \end{cases} \tag{fsgrad-upd}$$

and set $s_{k+1} = s_k$, $k = k+1$. Here we fix the step size.

In the case that $\mu_i > \mu$ or $\lambda_j > \lambda$, after M_1 iterations, we update both μ_{i+1} and λ_{j+1} by

$$\mu_{i+1} = \max\{\mu, \alpha \min\{\|\nabla g(x_k)\|_\infty, \mu_i\}\} \tag{6}$$

$$\lambda_{j+1} = \max\{\beta\lambda_j, \lambda\} \tag{7}$$

and set $i = i+1$, $j = j+1$. Then, we reset $s_k = \min\{s, \lambda_j\}$, $x_{-1} = x_0 := x_k$ and $k = 0$.

If $\lambda_j = \lambda$, $\mu_i = \mu$, we stop the algorithm after M_2 iterations. The fast gradient method for smoothed primal problem with continuation strategy is given below:

Algorithm 2 Fast gradient method for smoothed primal problem with continuation strategy

Input: initial guess x_0, step size s, continuation parameter α, M_1, M_2, lambda decay parameter β.

Calculate $\mu_{max} = \max\{\alpha \|A^T b\|_\infty, \mu\}$. Let $i = 0$, $\mu_0 = \mu_{max}$, $s_0 = s$, $k = 0$.
Update x_{k+1} by (fsgrad-upd), $k = k+1$
while $\lambda_j > \lambda$ or $\mu_i > \mu$ **do**
  for $l = 1 : M_1$ **do**
    Update x_{k+1} by (fsgrad-upd), $s_{k+1} = s_k$, $k = k+1$
  end for
  $\mu_{i+1} = \max\{\mu, \alpha \min\{\|\nabla g(x_k)\|_\infty, \mu_i\}\}$, $\lambda_{j+1} = \max\{\beta\lambda_j, \lambda\}$, $i = i+1$, $j = j+1$
  Reset $s_k = \min\{s, \lambda_j\}$, $x_{-1} = x_0 := x_k$ and $k = 0$.
end while
for $l = 1 : M_2$ **do**
  Update x_{k+1} by (fsgrad-upd), $s_{k+1} = s_k$, $k = k+1$
end for

In practice, we take $s = 4 \times 10^{-4}$, $\alpha = 0.2$, $M_1 = 200$, $M_2 = 700$, $\beta = 0.5$.

3 Problem 3(e)

Suppose the step size in $(k+1)$-th iteration is s_k. We denote $\text{prox}_{s_k\mu h}(x) = \arg\min_z \frac{1}{2}\|z-x\|_2^2 + sh(z)$ the proximal gradient of $f(x)$ at x. Because $h(x) = \|x\|_1$, we have explicit expression for $\text{prox}_{s_k\mu h}(x)$, namely $\text{prox}_{s_k\mu h}(x) = S_{s_k\mu}(x)$, where S_λ is the soft thresholding operator.

We consider to apply the continuation strategy and Barzila-Borwein step size. We have three parameters $\alpha, \epsilon_1, \epsilon_2$ for continuation. We set $\mu_{max} = \max\{\alpha\|A^T b\|_\infty, \mu\}$. Then, we take $\mu_0 = \mu_{max}$ and define $f_i = g(x) + \mu_i h(x)$. We set $x_{-1} = x_0$. In $(k+1)$-th iteration, if $k = 0$, we use the initial step size s. Otherwise, we calculate the BB step size s_k:

$$s_k = \frac{(x_k - x_{k-1})^T(x_k - x_{k-1})}{(x_k - x_{k-1})^T(\nabla g(x_k) - \nabla g(x_{k-1}))}$$

Then, we update x_{k+1} by

$$x_{k+1} = \text{prox}_{s_k\mu_i}(x_k - s_k\nabla g(x_k)) = S_{s_k\mu_i}\left(x_k - s_k A^T(Ax_k - b)\right) \tag{prox-upd}$$

and set $k = k+1$.

If $\mu_i > \mu$ and we find that $|\frac{f_i(x_k)-f_i(x_{k-1})}{f_i(x_{k-1})}| < \epsilon_1$, then we update μ_{i+1} by

$$\mu_{i+1} = \max\{\mu, \alpha\min\{\|\nabla g(x_k)\|_\infty, \mu_i\}\} \tag{mu-upd}$$

and set $i = i+1$. Then, we update x_{k+1} using the initial step size s by (prox-upd) and set $k = k+1$.

If $\mu_i = \mu$, we stop the algorithm when we find that $|\frac{f_i(x_k)-f_i(x_{k-1})}{f_i(x_{k-1})}| < \epsilon_2$. The proximal gradient method with continuation strategy is given below:

In practice, we take $s = 5\times 10^{-4}, \alpha = 0.5, \epsilon_1 = 10^{-6}, \epsilon_2 = 10^{-8}$.

4 Problem 3(f)

Suppose the step size is fixed to be s. We apply the same continuation strategy. We have three parameters $\alpha, \epsilon_1, \epsilon_2$ for continuation. We set $\mu_{max} = \max\{\alpha\|A^T b\|_\infty, \mu\}$. In $(k+1)$-th iteration. Then, we update x_{k+1} in the following way:

$$\begin{cases} y = x_k + \dfrac{k-1}{k+2}(x_k - x_{k-1}) \\ x_{k+1} = \text{prox}_{sh}(y - s_k\nabla g(x_k)) = S_{sh}(y - sA^T(Ax_k - b)) \end{cases} \tag{FISTA-upd}$$

Algorithm 3 Proximal gradient method with continuation strategy

Input: initial guess x_0, step size s, continuation parameter $\alpha, \epsilon_1, \epsilon_2$.

Calculate $\mu_{max} = \max\{\alpha\|A^T b\|_\infty, \mu\}$. Let $i = 0$, $\mu_0 = \mu_{max}$, $s_0 = s$, $k = 0$.
Update x_{k+1} by (prox-upd), $k = k + 1$
while $\mu_i > \mu$ **do**
  while $|\frac{f_i(x_k)-f_i(x_{k-1})}{f_i(x_{k-1})}| \geq \epsilon_1$ **do**
    Calculate BB step size s_k, update x_{k+1} by (prox-upd), $k = k + 1$
  end while
  $\mu_{i+1} = \max\{\mu, \alpha \min\{\|\nabla g(x_k)\|_\infty, \mu_i\}\}$, $i = i + 1$
  Reset $s_k = s$, update x_{k+1} by (prox-upd), $k = k + 1$
end while
while $|\frac{f_i(x_k)-f_i(x_{k-1})}{f_i(x_{k-1})}| \geq \epsilon_2$ **do**
  Calculate BB step size s_k, update x_{k+1} by (prox-upd), $k = k + 1$
end while

If $\mu_i > \mu$ and we find that $|\frac{f(x_k)-f(x_{k-1})}{f(x_{k-1})}| < \epsilon_1$, then we update μ_{i+1} by

$$\mu_{i+1} = \max\{\mu, \alpha \min\{\|\nabla g(x_k)\|_\infty, \mu_i\}\} \tag{mu-upd}$$

and set $i = i + 1$. Then, we reset our algorithm, starting from x_k. We set $x_{-1} = x_0 := x_k$, $k = 0$. We update x_{k+1} by (FISTA-upd) and set $k = k + 1$.

If $\mu_i = \mu$, we stop the algorithm when we find that $|\frac{f(x_k)-f(x_{k-1})}{f(x_{k-1})}| < \epsilon_2$. The proximal gradient method with continuation strategy is given below:

Algorithm 4 Fast proximal gradient method with continuation strategy

Input: initial guess x_0, step size s, continuation parameter $\alpha, \epsilon_1, \epsilon_2$.

Calculate $\mu_{max} = \max\{\alpha\|A^T b\|_\infty, \mu\}$. Let $i = 0$, $\mu_0 = \mu_{max}$, $x_{-1} = x_0$, $k = 0$.
Update x_{k+1} by (FISTA-upd), $k = k + 1$
while $\mu_i > \mu$ **do**
  while $|\frac{f(x_k)-f(x_{k-1})}{f(x_{k-1})}| \geq \epsilon_1$ **do**
    Update x_{k+1} by (FISTA-upd), $k = k + 1$
  end while
  $\mu_{i+1} = \max\{\mu, \alpha \min\{\|\nabla g(x_k)\|_\infty, \mu_i\}\}$, $i = i + 1$
  Reset $x_0 = x_{-1} =: x_k$, $k = 0$, update x_{k+1} by (FISTA-upd), $k = k + 1$
end while
while $|\frac{f(x_k)-f(x_{k-1})}{f(x_{k-1})}| \geq \epsilon_2$ **do**
  Update x_{k+1} by (FISTA-upd), $k = k + 1$
end while

In practice, we take $s = 5 \times 10^{-4}$, $\alpha = 0.5$, $\epsilon_1 = 10^{-6}$, $\epsilon_2 = 10^{-8}$.

5 Numerical result

The whole test program is named *Test_hw05_02.m*. This time we run all algorithms mentioned before with same A and b. Suppose the value of objective function from cvx mosek is f_c, the value from proposed method is f_p, then the value of *objval to cvx mosek* is $\frac{f_p - f_c}{f_c}$. The numerical result is given in the following tables:

Table 1 Random seed is 2. The cpu time of cvx mosek is 1.13

Method	cpu time	objval to cvx mosek	error to cvx mosek
grad smoothed	0.81	-6.55×10^{-8}	1.17×10^{-6}
fast grad smoothed	1.37	-5.84×10^{-8}	1.11×10^{-6}
prox grad	0.23	-1.87×10^{-6}	3.27×10^{-6}
fast prox grad	1.00	-1.71×10^{-6}	2.01×10^{-6}

Table 2 Random seed is 7. The cpu time of cvx mosek is 1.15

Method	cpu time	objval to cvx mosek	error to cvx mosek
grad smoothed	0.88	-1.54×10^{-6}	1.63×10^{-6}
fast grad smoothed	1.39	-1.57×10^{-6}	1.74×10^{-6}
prox grad	0.21	-3.09×10^{-6}	3.19×10^{-6}
fast prox grad	0.97	-3.00×10^{-6}	2.53×10^{-6}

Table 3 Random seed is 9. The cpu time of cvx mosek is 1.09

Method	cpu time	objval to cvx mosek	error to cvx mosek
grad smoothed	1.01	-3.01×10^{-7}	9.98×10^{-7}
fast grad smoothed	1.48	-3.26×10^{-7}	1.01×10^{-6}
prox grad	0.19	-2.06×10^{-6}	2.65×10^{-6}
fast prox grad	0.70	-1.97×10^{-6}	1.98×10^{-6}