Solutions for Mid-Term 2

ECE490 Introduction to Optimization — Fall 2018

1 Problem 1

1. (15 points) Apply Newton's method (in the pure form) to the minimization of the function $f(x) = x^3$. Write out the iteration formula and show that Newton's method achieves a linear convergence rate in this case.

2. (10 points) Suppose we apply the BFGS method $x_{k+1} = x_k - \alpha_k H_k^{-1} \nabla f(x_k)$ to minimize

$$f(x) = \frac{1}{2} x^\mathsf{T} Q x$$

where $Q = \begin{bmatrix} 5 & 1 \\ 1 & 2 \end{bmatrix}$. For simplicity, we start with $x_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\alpha_0 = 0.1$, and $H_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. What is H_1^{-1}?

Solution:

1. Newton's method iterates as

$$x_{k+1} = x_k - (\nabla^2 f(x_k))^{-1} \nabla f(x_k)$$

Since $\nabla f(x_k) = 3x_k^2$ and $\nabla^2 f(x_k) = 6x_k$, Newton's method actually iterates as

$$x_{k+1} = x_k - \left(\frac{1}{6x_k}\right) 3x_k^2 = \frac{1}{2} x_k$$

Therefore, Newton's method converges to the stationary point $x^* = 0$ at a linear rate $\rho = \frac{1}{2}$.

2. We need to calculate $s_0 = x_1 - x_0$ and $y_0 = \nabla f(x_1) - \nabla f(x_0)$. We have

$$\nabla f(x_0) = Q x_0 = \begin{bmatrix} 5 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}$$

$$x_1 = x_0 - 0.1 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 6 \\ 3 \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.7 \end{bmatrix}$$

$$\nabla f(x_1) = \begin{bmatrix} 5 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 0.4 \\ 0.7 \end{bmatrix} = \begin{bmatrix} 2.7 \\ 1.8 \end{bmatrix}$$

$$y_0 = \nabla f(x_1) - \nabla f(x_0) = \begin{bmatrix} -3.3 \\ -1.2 \end{bmatrix}$$

$$s_0 = x_1 - x_0 = \begin{bmatrix} -0.6 \\ -0.3 \end{bmatrix}$$

$$H_1^{-1} = \left(I - \frac{s_0 y_0^\intercal}{y_0^\intercal s_0} \right) H_0^{-1} \left(I - \frac{y_0 s_0^\intercal}{y_0^\intercal s_0} \right) + \frac{s_0 s_0^\intercal}{y_0^\intercal s_0} = \begin{bmatrix} 0.272 & -0.249 \\ -0.249 & 0.933 \end{bmatrix}$$

2 Problem 2

Consider the unconstrained minimization problem

$$\min_{x \in \mathbb{R}^p} \{f(x) + \mu \|x\|_1\}$$

where f is L-smooth and m-strongly convex. Suppose x^* is a point satisfying $-\nabla f(x^*) \in \partial \mu \|x^*\|_1$. Apply the proximal gradient method with a constant stepsize α.

1. (5 points) Please write down the proximal gradient method for the above problem.

2. (5 points) How to perform the proximal step for the above problem? (Hint: Shrinkage.)

3. (15 points) Since f is m-strongly convex and L-smooth, the following inequality holds for all $x \in \mathbb{R}^p$

$$\begin{bmatrix} x - x^* \\ \nabla f(x) - \nabla f(x^*) \end{bmatrix}^\mathsf{T} \begin{bmatrix} 2mLI & -(m+L)I \\ -(m+L)I & 2I \end{bmatrix} \begin{bmatrix} x - x^* \\ \nabla f(x) - \nabla f(x^*) \end{bmatrix} \leq 0$$

   Use the above inequality to show the proximal gradient method with stepsize $\alpha = \frac{1}{L}$ satisfies

$$\|x_k - x^*\| \leq \left(1 - \frac{m}{L}\right)^k \|x_0 - x^*\|$$

Solution:

1. The proximal gradient method updates as $x_{k+1} = \mathrm{prox}_{g,\alpha}(x_k - \alpha \nabla f(x_k))$ where $g(x) = \mu \|x\|_1$.

2. Denote $h_k = x_k - \alpha \nabla f(x_k)$. Suppose the j-th entry of h_k is h_k^j. Then the j-th entry of x_{k+1} is updated using the shrinkage operator as follows

$$x_{k+1}^j = \begin{cases} h_k^j - \mu\alpha & \text{if } h_k^j \geq \mu\alpha \\ 0 & \text{if } -\mu\alpha < h_k^j < \mu\alpha \\ h_k^j + \mu\alpha & \text{if } h_k^j \leq -\mu\alpha \end{cases}$$

3. We can rewrite the proximal gradient method as $x_{k+1} = x_k - \alpha u_k - \alpha r_k$ where $u_k = \nabla f(x_k)$ and $r_k \in \partial g(x_{k+1})$. Since f is m-strongly convex and L-smooth, we have

$$\begin{bmatrix} x_k - x^* \\ \nabla f(x_k) - \nabla f(x^*) \\ r_k - r^* \end{bmatrix}^\mathsf{T} \begin{bmatrix} 2mLI & -(L+m)I & 0 \\ -(L+m)I & 2I & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_k - x^* \\ \nabla f(x_k) - \nabla f(x^*) \\ r_k - r^* \end{bmatrix} \leq 0$$

   Since g is convex, we directly have $(r_k - r^*)^\mathsf{T}(x_{k+1} - x^*) \geq 0$. This can be rewritten as a quadratic inequality:

$$\begin{bmatrix} x_k - x^* \\ \nabla f(x_k) - \nabla f(x^*) \\ r_k - r^* \end{bmatrix}^\mathsf{T} \begin{bmatrix} 0 & 0 & -I \\ 0 & 0 & \alpha I \\ -I & \alpha I & 2\alpha I \end{bmatrix} \begin{bmatrix} x_k - x^* \\ \nabla f(x_k) - \nabla f(x^*) \\ r_k - r^* \end{bmatrix} \leq 0$$

Therefore, we have $\|x_{k+1} - x^*\|^2 \le \rho^2 \|x_k - x^*\|^2$ if there exists non-negative λ_1 and λ_2 such that

$$\begin{bmatrix} 1-\rho^2 & -\alpha & -\alpha \\ -\alpha & \alpha^2 & \alpha^2 \\ -\alpha & \alpha^2 & \alpha^2 \end{bmatrix} \le \lambda_1 \begin{bmatrix} 2mL & -(L+m) & 0 \\ -(m+L) & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & \alpha \\ -1 & \alpha & 2\alpha \end{bmatrix}$$

For $\alpha = \frac{1}{L}$, we can choose $\lambda_1 = \frac{1}{L^2}$ and $\lambda_2 = \frac{1}{L}$ to satisfy the above inequality and show the desired convergence rate.

3 Problems 3

1. (10 points) Consider the constrained minimization problem

$$\begin{aligned} \text{minimize} \quad & x_1 + x_2 + x_3 + x_4 \\ \text{subject to} \quad & x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1. \end{aligned}$$

   Determine all the local mins for the above problem.

2. (15 points) Consider minimizing the function $f(x) = \frac{\mu}{2}\|x\|^2 + \frac{1}{2}\|Ax - b\|^2$ where $A \in \mathbb{R}^{n\times p}$, $b \in \mathbb{R}^p$, $\mu \in \mathbb{R}$, and $x \in \mathbb{R}^p$. This is the so-called ridge regression problem where x is the decision variable. This problem can be rewritten as

$$\begin{aligned} \text{minimize} \quad & \tfrac{\mu}{2}\|x\|^2 + \tfrac{1}{2}\|y - b\|^2 \\ \text{subject to} \quad & Ax = y \end{aligned}$$

   Write out the Lagrangian $L(x, y, \lambda)$ for the above problem and calculate the related dual function $D(\lambda) = \min_{x,y} L(x, y, \lambda)$.

Solution:

1. First, we write out the Lagrangian as

$$L(x, \lambda) = x_1 + x_2 + x_3 + x_4 + \lambda(x_1^2 + x_2^2 + x_3^2 + x_4^2 - 1)$$

   Applying $\nabla_x L(x^*, \lambda^*) = 0$, we have

$$\begin{aligned} 1 + 2\lambda^* x_1^* = 0, \\ 1 + 2\lambda^* x_2^* = 0, \\ 1 + 2\lambda^* x_3^* = 0, \\ 1 + 2\lambda^* x_4^* = 0. \end{aligned}$$

   We also have $(x_1^*)^2 + (x_2^*)^2 + (x_3^*)^2 + (x_4^*)^2 = 1$. There are two possible solutions: (i) $x_1^* = x_2^* = x_3^* = x_4^* = \frac{1}{2}$, $\lambda^* = -1$; (ii) $x_1^* = x_2^* = x_3^* = x_4^* = -\frac{1}{2}$, $\lambda^* = 1$. Notice $\nabla^2_{xx} L(x^*, \lambda^*) = 2\lambda^* I$ which is negative definite for $\lambda^* = -1$ and positive definite for $\lambda^* = 1$. So the only local min is $x_1^* = x_2^* = x_3^* = x_4^* = -\frac{1}{2}$.

2. First, we write out the Lagrangian as

$$L(x, y, \lambda) = \frac{\mu}{2}\|x\|^2 + \frac{1}{2}\|y - b\|^2 + \lambda^\mathsf{T}(Ax - y)$$

   Minimizing L over x and y with fixed λ is a positive definite quadratic minimization problem. We can just set the derivatives to 0 and obtain

$$\begin{aligned} \nabla_x L(x^*, y^*, \lambda) = 0 &\to \mu x^* + A^\mathsf{T}\lambda = 0 \to x^* = -\frac{1}{\mu}A^\mathsf{T}\lambda \\ \nabla_y L(x^*, y^*, \lambda) = 0 &\to y^* = b + \lambda \end{aligned}$$

   Therefore, the dual function can be calculated as

$$D(\lambda) = \frac{\mu\lambda^\mathsf{T}AA^\mathsf{T}\lambda}{2\mu^2} + \frac{1}{2}\lambda^\mathsf{T}\lambda + \lambda^\mathsf{T}\left(\frac{-AA^\mathsf{T}\lambda}{\mu} - b - \lambda\right) = -\frac{1}{2\mu}\lambda^\mathsf{T}AA^\mathsf{T}\lambda - \frac{1}{2}\|\lambda\|^2 - \lambda^\mathsf{T}b$$

4 Problems 4

1. (10 points) What is ADMM? Write down the ADMM update rule for the following problem:

$$\begin{aligned} \text{minimize} \quad & f(x) + g(y) \\ \text{subject to} \quad & Ax + By = c \end{aligned}$$

2. (15 points) Consider the problem $\min_x\{\mu\|x\|_1 + \sum_{i=1}^n \frac{1}{2}(a_i^\top x - b_i)^2\}$ where $a_i \in \mathbb{R}^p$, $b_i \in \mathbb{R}$, $\mu \in \mathbb{R}$, and $x \in \mathbb{R}^p$. This problem can be rewritten as

$$\begin{aligned} \text{minimize} \quad & \mu\|y\|_1 + \textstyle\sum_{i=1}^n f_i(x^i) \\ \text{subject to} \quad & x^i - y = 0, \forall i \in \{1, 2, \dots, n\} \end{aligned}$$

where $f_i(x^i) = \frac{1}{2}(a_i^\top x^i - b_i)^2$, and $x^i \in \mathbb{R}^p$ is a vector having the same dimension as a_i. The augmented Lagrangian is given by

$$L_\rho = \mu\|y\|_1 + \sum_{i=1}^n \left\{ f_i(x^i) + (\lambda^i)^\top (x^i - y) + \frac{\rho}{2}\|x^i - y\|^2 \right\}$$

Your task is to write out the ADMM update formula for the above problem. Specifically, express x_{k+1}^i, y_{k+1}, and λ_{k+1}^i as functions of x_k^i, y_k, λ_k^i, a_i, b_i, μ, and ρ. (Hint: use the shrinkage operator for the update of y_{k+1}.)

Solution:

1. ADMM is the alternating direction method of multipliers. It iterates as

$$\begin{aligned} x_{k+1} &= \arg\min_x L_\rho(x, y_k, \lambda_k) \\ y_{k+1} &= \arg\min_y L_\rho(x_{k+1}, y, \lambda_k) \\ \lambda_{k+1} &= \lambda_k + \rho(Ax_{k+1} + By_{k+1} - c) \end{aligned}$$

where L_ρ is the augmented Lagrangian defined as

$$L_\rho(x, y, \lambda) = f(x) + g(y) + \lambda^\top (Ax + By - c) + \frac{\rho}{2}\|Ax + By - c\|^2.$$

2. By definition, ADMM iterates as

$$\begin{aligned} x_{k+1}^i &= \arg\min_{x^i} \left\{ f_i(x^i) + (\lambda_k^i)^\top (x^i - y_k) + \frac{\rho}{2}\|x^i - y_k\|^2 \right\} \\ y_{k+1} &= \arg\min_y \left\{ \mu\|y\|_1 + \sum_{i=1}^n \left(-(\lambda_k^i)^\top y + \frac{\rho}{2}\|x^i - y\|^2 \right) \right\} \\ \lambda_{k+1}^i &= \lambda_k^i + \rho(x_{k+1}^i - y_{k+1}) \end{aligned}$$

Since $f_i(x^i) = \frac{1}{2}(a_i^\mathsf{T} x^i - b_i)^2$, we eventually have

$$\begin{aligned}
x_{k+1}^i &= (a_i a_i^\mathsf{T} + \rho I)^{-1}(a_i b_i + \rho y_k - \lambda_k^i) \\
y_{k+1} &= S_{\mu/(\rho n)}\left(\frac{1}{n}\sum_{i=1}^{n}(x_{k+1}^i + \lambda_k^i/\rho)\right) \\
\lambda_{k+1}^i &= \lambda_k^i + \rho(x_{k+1}^i - y_{k+1})
\end{aligned}$$

where $S_{\mu/(\rho n)}$ is the shrinkage operator that shrinks every value between $-\mu/(\rho n)$ and $\mu/(\rho n)$ to 0.

[space for Problem 4.]