Optimization for AI Assignment 3

Joshua Chen

October 21, 2017

Problem 3

The following proof follows directly Nocedal and Wright, pg. 89. We begin by proving a lemma, proven in a class lecture, and also in Nocedal and Wright.

Lemma. *Let m be the quadratic function defined by*

$$m(p) = g^T p + \frac{1}{2} p^T B p$$

where B is any symmetric matrix. Then the following statements are true.

1. *m attains a minimum $\iff$ B is positive semidefinite and g is in the range of B If B is positive semidefinite, then every p satisfying $Bp = -g$ is a global minimizer of m*
2. *m has a unique minimizer $\iff$ B is positive definite.*

Proof. We prove each of the two claims in turn.

1. ($\Leftarrow$): Since g is in the range of B, there is a p with $Bp = -g$.
   For all $w \in \mathbb{R}^n$ we have

$$\begin{aligned} m(p+w) &= g^T(p+w) + \frac{1}{2}(p+w)^T B(p+w) \\ &= (g^T + \frac{1}{2} p^T Bp) + g^T w + (Bp)^T w + \frac{1}{2} w^T Bw \\ &= m(p) + \frac{1}{2} w^T Bw + (Bp+g)^T w \\ &\geq m(p) \end{aligned} \tag{1}$$

   since B is positive semidefinite. Hence p is a minimizer of m.

   ($\Rightarrow$): Let p be a minimizer of m. Since $\nabla m(p) = Bp + g = 0$, we have that g is in the range of B. Also we have that $\nabla^2 m(p) = B$ is positive semidefinite, giving the result.

2. ($\Leftarrow$): The same argument as (i) suffices with the additional point that $w^T Bw > 0$ whenever $w \neq 0$ ($\Rightarrow$): We proceed as in (i) to deducate the B is positive semidefinite. If B is not positive definite, there is a vector $w \neq 0$ such that $Bw = 0$ hence from (1), we have $m(p+w) = m(p)$ so the minimizer is not unique, giving a contradiction.

□

We now state and prove a theorem in Nocedal and Wright which directly gives us the required result, substituting Δ used by Nocedal and Wright for d in our problem, and I used by Nocedal and Wright for a more general positive semidefinite A, in our problem. I make the note that since A is not specified in our problem to be symmetric the notation used below $\|\cdot\|_A$ is a bit of an abuse of notation, however it gets the point across $\|\cdot\|_A := (\cdot, A\cdot)$

Theorem. *The vector p^* is a global solution of the trust region problem*

$$\min_{p\in\mathbb{R}^n} m(p) = f + g^T p + \frac{1}{2}p^T Bp. \text{ s.t. } \|p\|_A \leq \Delta \tag{2}$$

if and only if p^* is feasible and there is a scalar $\lambda \geq 0$ such that the following conditions are satisfied:

$$(B+\lambda A)^* p = -g \tag{3a}$$

$$\lambda(\Delta - \|p^*\|_A) = 0 \tag{3b}$$

$$(B+\lambda A) \quad \text{is positive semidefinite} \tag{3c}$$

Proof. (Theorem)

Assume first that there is a $\lambda \geq 0$ such that the conditions (3) are satisfied.

The lemma implies that p^* is a global minimum of the quadratic function

$$\hat{m}(p) = g^T p + \frac{1}{2}p^T(B+\lambda A)p \;\; = \;\; m(p) + \frac{\lambda}{2}p^T Ap$$

Since $\hat{m}(p) \geq \hat{m}(p^*)$ we have

$$m(p) \geq m(p^*) + \frac{\lambda}{2}((p^*)^T Ap^* - p^T Ap) \tag{4}$$

Because $\lambda(\Delta - \|p^*\|_A) = 0$ and therefore $\lambda(\Delta^2 - (p^*)^T Ap^*) = 0$, we have

$$m(p) \geq m(p^*) + \frac{\lambda}{2}(\Delta^2 - p^T Ap)$$

Hence, from $\lambda \geq 0$, we have $m(p) \geq m(p^*)$ for all p with $\|p\|_A \leq \Delta$. Therefore p^* is a global minimizer of (2).

For the converse, we assume that p^* is a global solution of (2) and show that there is a $\lambda \geq 0$ that satisfies (3).

In the case $\|p^*\|_A < \Delta$, p^* is an unconstrained minimizer of m, and so

$$\nabla m(p^*) \;\; = \;\; Bp^* + g = 0, \qquad\qquad \nabla^2 m(p) = B \text{ is positive semidefinite}$$

and so all the properties (3) hold for $\lambda = 0$.

Assume for the remainder of the proof that $\|p^*\|_A = \Delta$. Then (3b) is immediately satisfied, and p^* also solves the constrained problem

$$\min m(p) \qquad \text{subject to } \|p\|_A = \Delta.$$

By applying optimality conditions for constrained optimization to this problem, we find that there is a λ such that the Lagrangian function defined by

$$\mathcal{L}(p,\lambda) = m(p) + \frac{\lambda}{2}(p^T Ap - \Delta^2)$$

has a stationary point at p^*. By setting $\nabla_p \mathcal{L}(p^*, \lambda)$ to zero, we obtain

$$Bp^* + g + \lambda Ap^* = 0 \qquad \Rightarrow (B - \lambda AI)p^* = -g \tag{5}$$

so that (3a) holds. Since $m(p) \geq m(p^*)$ for any p with $p^T Ap = (p^*)^T Ap^* = \Delta^2$, we have for such vectors p that

$$m(p) \geq m(p^*) + \frac{\lambda}{2}((p^*)^T Ap^* - p^T Ap)$$

If we substitute the above expression for g from (5) into this expression, we obtain after some rearrangement that

$$\frac{1}{2}(p-p^*)^T(B+\lambda AI)(p-p^*) \geq 0. \tag{6}$$

Since the set of directions

$$\left\{ w : w = \pm\frac{p-p^*}{\|p-p^*\|} \text{ for some } p \text{ with } \|p\|_A = \Delta \right\} \tag{7}$$

is dense on the unit sphere, (6) suffices to prove (3c).

It remains to show that $\lambda \geq 0$. Because (3a) and (3c) are satisfied by p^*, we have from the lemma that p^* minimizes $\hat{m}$, so (4) holds. Suppose that there are only negative values of λ that satisfy (3a) and (3c). Then we have from (4) that $m(p) \geq m(p^*)$ whenever $\|p\|_A \geq \|p^*\|_A = \Delta$. Since we already know that p^* minimizes m for $\|p\|_A \leq \Delta$ it follows that m is in fact a global, unconstrained minimizer of m. From the lemma it follows that $Bp = -g$ and B is positive semidefinite. Therefore conditions (3a) and (3c) are satisfied by $\lambda = 0$, which contradicts our assumption that only negative values of λ can satisfy the conditions. We conclude that $\lambda \geq 0$ completing the proof. $\square$

See Numerical Optimization by Nocedal and Wright for any missing details.