CS395T Homework 3

Name: Zaiwei Zhang
Date: 10/15/2017

Programming

1. (Problem 1 and 2) First, I use the Trust Region method to solve the problem. The algorithm is stated very clearly in the lecture notes. For finding p_k, solving the sub-problem, I used the Cauchy Point method, and I used the exact Hessian to calculate p_k. The method has also been stated in the lecture notes, I will put the gradient and hessian matrix here:

$$\nabla g_B = \mu B + \sum_{i=1}^{n}\sum_{j=1}^{n} 2G_{ij}(A_{ij} - (e_i^T B)(Ce_j))(-e_i(Ce_j)^T) \tag{1}$$

$$\nabla g_C = \mu C + \sum_{i=1}^{n}\sum_{j=1}^{n} 2G_{ij}(A_{ij} - (e_i^T B)(Ce_j))(-(e_i^T B)^T e_j^T) \tag{2}$$

Now, the partial gradients above are still matrices. Let's reshape them to vectors, concatenate them and use that vector as the gradient in the pk calculation. For B_k in Cauchy Point method, I used the exact hessian and the components are listed in the following:

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial B^2} & \frac{\partial^2 f}{\partial B \partial C} \\ \frac{\partial^2 f}{\partial C^2} & \frac{\partial^2 f}{\partial C \partial B} \end{pmatrix}, H \in R^{2nr \times 2nr} \tag{3}$$

Each hessian component in the equation above has $R^{nr \times nr}$ dimension.

$$\frac{\partial^2 f}{\partial B^2} = \mu I + \sum_{i=1}^{n}\sum_{j=1}^{n} 2G_{ij}(-e_i(Ce_j)^T)(-e_i(Ce_j)^T)^T \tag{4}$$

$$\frac{\partial^2 f}{\partial C^2} = \mu I + \sum_{i=1}^{n}\sum_{j=1}^{n} 2G_{ij}(-(e_i^T B)^T e_j^T)(-(e_i^T B)^T e_j^T)^T \tag{5}$$

$$\tag{6}$$

$-e_i(Ce_j)^T$ and $-(e_i^T B)^T e_j^T$ are flattened to vectors since we are considering B and C together as a vector of unknown variables.

$$\frac{\partial^2 f}{\partial B \partial C} = \sum_{i=1}^{n}\sum_{j=1}^{n} 2G_{ij}(A_{ij} - (e_i^T B)(Ce_j)) * \begin{pmatrix} I_{r\times r} & ... & I_{r\times r} \\ ... & ... & ... \\ I_{r\times r} & ... & I_{r\times r} \end{pmatrix} + \sum_{i=1}^{n}\sum_{j=1}^{n} 2G_{ij}(-e_i(Ce_j)^T)(-(e_i^T B)^T e_j^T)^T \tag{7}$$

$$\frac{\partial^2 f}{\partial C \partial B} = \sum_{i=1}^{n}\sum_{j=1}^{n} 2G_{ij}(A_{ij} - (e_i^T B)(Ce_j)) * \begin{pmatrix} I_{r\times r} & ... & I_{r\times r} \\ ... & ... & ... \\ I_{r\times r} & ... & I_{r\times r} \end{pmatrix} + \sum_{i=1}^{n}\sum_{j=1}^{n} 2G_{ij}(-(e_i^T B)^T e_j^T)(-e_i(Ce_j)^T)^T \tag{8}$$

The rest of the implementation details are straight forward, which is just following the algorithms listed in Lecture 8 notes.

For alternating minimization, we first fix C and then optimize B, and then fix C and optimize B. When optimizing B and C, we first calculate the partial derivative and then let them equal to zero. Then, we have a closed form solution on B and C in each iteration. We have the following two equations:

$$\mu B + \sum_{i=1}^{n}\sum_{j=1}^{n} 2G_{ij}(e_i^T B)(Ce_j)(e_i(Ce_j)^T) = \sum_{i=1}^{n}\sum_{j=1}^{n} 2G_{ij}A_{ij}(e_i(Ce_j)^T) \tag{9}$$

$$\mu C + \sum_{i=1}^{n}\sum_{j=1}^{n} 2G_{ij}(e_i^T B)(Ce_j)((e_i^T B)^T e_j^T) = \sum_{i=1}^{n}\sum_{j=1}^{n} 2G_{ij}A_{ij}((e_i^T B)^T e_j^T) \tag{10}$$

For each B and C, we have $n \times r$ variables and we have $n \times r$ equations based on (9) and (10). Therefore, in each iteration, we can solve a linear system to get the optimal solution.

Theory

1. (Problem 3) We want to prove the following: Vector p^* is the global optimal solution if and only if p^* is feasible and there is a scalar $\lambda \geq 0$ such that the following conditions are satisfied:

$$(B + \lambda A)p^* = -g \tag{11}$$
$$\lambda(d^2 - (p^*)^T A p^*) = 0 \tag{12}$$
$$(B + \lambda A) \text{ is PSD} \tag{13}$$

Let's say that $m(p) = g^T p + \frac{1}{2} p^T B p$. From the lecture note, we know that for the sub-problem $m(p)$ with B is any symmetric metric, the following lemmas are true:

- m attains a minimum if and only if B is positive semi-definite and is in the range of B. If B is positive semi-definite, then every p satisfying $Bp = -g$ is a global minimizer of m.
- m has a unique minimizer if and only if B is positive definite.

Now, let's begin the proof. Assume there is $\lambda \geq 0$ such that the conditions in the theorem are satisfied. The lemma above implies that p^* is a global minimum of the quadratic function, we have:

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

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

$$m(p) \geq m(p^*) + \frac{\lambda}{2}((p^*)^T A p^* - p^T A p) \tag{14}$$

Since $\lambda(d^2 - (p^*)^T A p^*) = 0$, we have

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

Since $\lambda \geq 0$, we know that $m(p) \geq m(p^*)$ for all p with $p^* A p \leq d^2$.

In the other direction, we assume that p^* us a global solution. In the case $(p^*)^T A p^* \leq d^2$, p^* is an unconstrained minimizer of m, and we have:

$$\nabla m(p^*) = (B + \lambda A)p^* + g = 0, \nabla^2 m(p^*) = B + \lambda A, \text{which is PSD}$$

, so (13) holds for $\lambda \geq 0$.

Then, assume that $(p^*)^T A p^* = d^2$, then we are solving the following constrained optimization problem:

$$\min m(p), \text{subject to} (p^*)^T A p^* = d^2$$

We find a λ and the Lagrangian function defined by:

$$L(p, \lambda) = m(p) + \frac{\lambda}{2}((p^*)^T A p^* - d^2)$$

has a stationary point at p^*. By setting $\nabla_p L^*(P^*, \lambda) = 0$, we have:

$$Bp^* + g + \lambda A p^* = 0 \rightarrow (B + \lambda A)p^* = -g$$

Thus, (11) holds. Then, since $\hat{m}(p) \geq \hat{m}(p^*)$, we have:

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

Then, substitute $(B + \lambda A)p^* = -g$, we obtain that:

$$\frac{1}{2}(p - p^*)^T (B + \lambda A)(p - p^*) \geq 0$$

Since the set of directions $p - p^*$ is dense, so $B + \lambda A$ must be positive semidefinite.

We still need to show that $\lambda \geq 0$. Because (11) and (13) are satisfied by p^*, from the lemma, we know that p minimizes $L(p, \lambda)$, so (14) holds. Suppose that there are only negative values of λ that satisfy (11) and (13). Then we have from (14) that $m(p) \geq m(p^*)$ whenever $\|p\| \geq \|p^*\|$. Since we already know that p^* minimizes m for $(p^*)^T A p^* \leq d^2$, 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 (11) and (13) are satisfied by $\lambda \geq 0$, which contradicts our assumption that only negative values of $\lambda \geq 0$ can satisfy the conditions. We conclude that $\lambda \geq 0$, completing the proof.