CS 395T: Homework 3

Brady Zhou
brady.zhou@utexas.edu

October 19, 2017

1 Low Rank Matrix Recovery (Alternating Minimization)

We are interested in minimizing the following

$$\operatorname*{argmin}_{B,C} \sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}(a_{ij} - B_i^{\intercal} C_j)^2 + \frac{\mu}{2}(||B||_F^2 + ||C||_F^2)$$

We can do this by taking partials with respect to row vectors $B_k^{\intercal}$ and column vectors C_k.

$$\frac{\partial}{\partial B_k^{\intercal}}\Big(\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}(a_{ij} - B_i^{\intercal} C_j)^2 + \frac{\mu}{2}(||B||_F^2 + ||C||_F^2)\Big) = -2\sum_{j=1}^{n} g_{kj}(a_{kj} - B_k^{\intercal} C_j)C_j^{\intercal} + \mu B_k^{\intercal}$$
$$\frac{\partial}{\partial C_k}\Big(\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}(a_{ij} - B_i^{\intercal} C_j)^2 + \frac{\mu}{2}(||B||_F^2 + ||C||_F^2)\Big) = -2\sum_{i=1}^{n} g_{ik}(a_{ik} - B_i^{\intercal} C_k)B_i + \mu C_k$$

Setting $\frac{\partial}{\partial B_k^{\intercal}} = 0$, we have the following

$$\begin{aligned}
0 &= -2\sum_{j=1}^{n} g_{kj}(a_{kj} - B_k^{\intercal} C_j)C_j + \mu B_k^{\intercal} \\
2\sum_{j=1}^{n} g_{kj}a_{kj}C_j^{\intercal} &= B_k^{\intercal}\Big(2\sum_{j=1}^{n} g_{kj}C_jC_j^{\intercal} + \mu I\Big) \\
2\sum_{j=1}^{n} g_{kj}a_{kj}C_j &= \Big(2\sum_{j=1}^{n} g_{kj}C_jC_j^{\intercal} + \mu I\Big){B_k^{\intercal}}^{\intercal} && \text{Transpose for sanity.}
\end{aligned}$$

So we can see the optimal $B_k^{\intercal}$ is the solution to $Ax = b$, where

$$A = 2\sum_{j=1}^{n} g_{kj}C_jC_j^{\intercal} + \mu I$$
$$b = 2\sum_{j=1}^{n} g_{kj}a_{kj}C_j$$

Similarly, for column vectors C_k,

$$\begin{aligned}
0 &= -2\sum_{i=1}^{n} g_{ik}(a_{ik} - B_i^{\intercal} C_k)B_i + \mu C_k \\
2\sum_{i=1}^{n} g_{ik}a_{ik}B_i &= \Big(2\sum_{i=1}^{n} g_{ik}B_iB_i^{\intercal} + \mu I\Big)C_k
\end{aligned}$$

And we have our result.

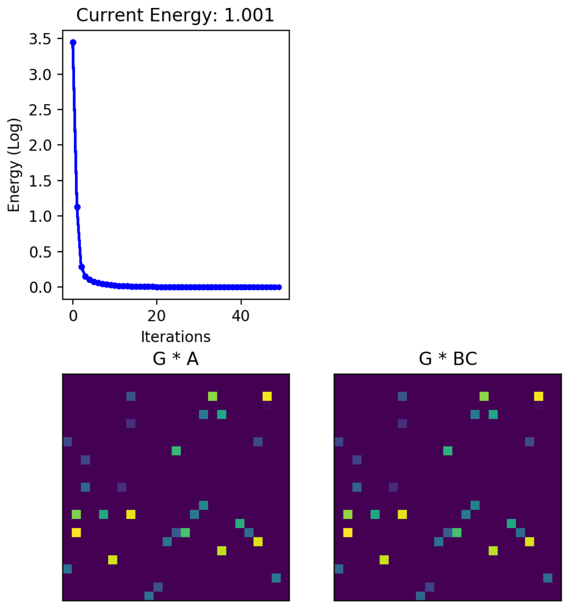

Figure 1: Results of alternating minimization.

2 Low Rank Matrix Recovery (Newton Trust Region)

To perform the Newton Trust Region method, we will have to calculate the Hessian of the objective function, where we are now minimizing over

$$x = \begin{bmatrix} {B_1^\intercal}^\intercal \\ {B_2^\intercal}^\intercal \\ \vdots \\ {B_n^\intercal}^\intercal \\ \hline C_1 \\ C_2 \\ \vdots \\ C_n \end{bmatrix}$$

Restating the result from Problem 1,

$$\frac{\partial f}{\partial {B_i^\intercal}^\intercal} = -2\sum_{k=1}^{n} g_{ik}(a_{ik} - B_i^\intercal C_k)C_k + \mu {B_i^\intercal}^\intercal$$

$$\frac{\partial f}{\partial C_i} = -2\sum_{k=1}^{n} g_{ki}(a_{ki} - B_k^\intercal C_i){B_k^\intercal}^\intercal + \mu C_i$$

By inspection, we can see $\frac{\partial f}{\partial B_i^{\intercal \intercal} \partial B_j^{\intercal \intercal}} = \frac{\partial f}{\partial C_i \partial C_j} = 0$ when $i \neq j$. For $i = j$, we have the following

$$\frac{\partial f}{\partial^2 B_i^{\intercal \intercal}} = 2 \sum_{k=1}^{n} g_{ik} C_k C_k^\intercal + \mu I$$
$$\frac{\partial f}{\partial^2 C_i} = 2 \sum_{k=1}^{n} g_{ki} B_k^{\intercal \intercal} B_k^\intercal + \mu I$$

And now for the cross terms,

$$\begin{aligned}\frac{\partial f}{\partial B_i^{\intercal \intercal} \partial C_j} &= \frac{\partial}{\partial C_j} \Big(-2 \sum_{k=1} g_{ik} (a_{ik} - B_i^\intercal C_k) C_k + \mu B_i^{\intercal \intercal} \Big) \\ &= 2 g_{ij} \Big(-a_{ij} I + C_j^\intercal B_i^{\intercal \intercal} I + B_i^{\intercal \intercal} C_j^\intercal \Big)\end{aligned}$$

And by symmetry of the Hessian, we have

$$\frac{\delta f}{\delta C_i \delta B_j^{\intercal \intercal}} = \Big(\frac{\delta f}{\delta B_i^{\intercal \intercal} \delta C_j} \Big)^\intercal$$

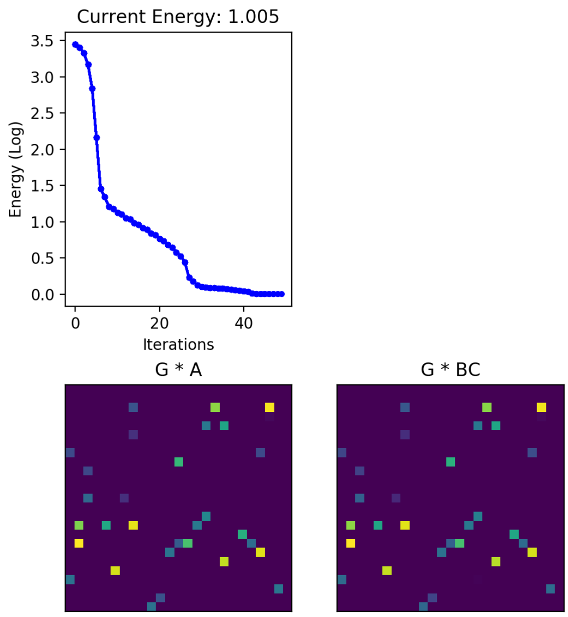

Figure 2: Results of trust region.

3 Trust Region Conditions with Arbitrary Norm

Let B be symmetric, A be positive semidefinite.

For p^* to be an optimal solution to the following problem

$$\begin{aligned} \underset{p}{\operatorname{argmin}} \quad & g^\intercal p + \frac{1}{2} p^\intercal B p \\ \text{subject to} \quad & \frac{1}{2} p^\intercal A p \leq \Delta^2 \end{aligned}$$

we have three necessary and sufficient conditions.

1. $(B + \lambda A)p = -g$
2. $\exists\, \lambda \geq 0$ such that $\lambda(\frac{1}{2}p^\intercal Ap - \Delta^2) = 0$
3. $B + \lambda A$ is symmetric positive semidefinite

Lemma 3.1. *Let $\hat{m}(p)$ be the following*

$$\operatorname*{argmin}_p p^\intercal g + \frac{1}{2}p^\intercal Bp$$

where B is positive semidefinite.

Then $\hat{m}$ attains a minimum if and only if B is positive semi-definite and $g \in Range(B)$.

Stated without proof (formal proof in lecture notes).

Proof. $\Longleftarrow$

We want to show that if $\bar{p}$ satisfies conditions $1, 2, 3$, then $\bar{p} = p^\star$.

Consider a similar problem.

$$\begin{aligned}
\hat{m}(p) &= p^\intercal g + \frac{1}{2}p^\intercal(B + \lambda A)p \\
&= m(p) + \frac{1}{2}\lambda p^\intercal Ap
\end{aligned}$$

By Lemma 3.1, we know that $\bar{p}$ is the minimizer to $\hat{m}$, that is, $\hat{m}(\bar{p}) \leq \hat{m}(p)\ \forall\ p$.

$$\begin{aligned}
\hat{m}(\hat{p}) &\leq \hat{m}(p) \\
m(\hat{p}) + \frac{\lambda}{2}\hat{p}^\intercal A\hat{p} &\leq m(p) + \frac{\lambda}{2}p^\intercal Ap \\
m(\hat{p}) &\leq m(p) + \frac{\lambda}{2}p^\intercal Ap - \frac{\lambda}{2}\hat{p}^\intercal A\hat{p} \\
&= m(p) + \frac{\lambda}{2}p^\intercal Ap - \frac{\lambda}{2}\hat{p}^\intercal A\hat{p} + \lambda\Big(\frac{1}{2}\hat{p}^\intercal A\hat{p} - \Delta^2\Big) && \text{By 2.} \\
&= m(p) + \lambda\Big(p^\intercal Ap - \Delta^2\Big) \\
&\leq m(p) && \text{Since } \frac{1}{2}p^\intercal Ap \leq \Delta^2 \text{ and } \lambda \geq 0.
\end{aligned}$$

We can see that $\bar{p}$ is the minimizer to m.

$\square$

Proof. $\Longrightarrow$

Consider the unconstrained formulation of m (Lagrangian) as

$$\operatorname*{argmin}_{p,\lambda\geq 0} \quad p^\intercal g + \frac{1}{2}p^\intercal Bp + \lambda\Big(\frac{1}{2}p^\intercal Ap - \Delta^2\Big)$$

p^* will lie at a stationary point of the Lagrangian. Taking the gradient with respect to p and setting it to 0, we have $(B + \lambda A)p = -g$. So we can see (1) is true.

Similarly, at a stationary point of the Lagrangian, either $\lambda = 0$, or $\frac{1}{2}p^\intercal Ap - \Delta^2 = 0$. Combining these two facts, we have (2).

Now we have to consider two cases.

Case 1: $\frac{1}{2}p^\intercal Ap < \Delta^2$, and $\lambda = 0$.

For this, we are on the interior and this turns into unconstrained optimization. We need to show $B + \lambda A$ is positive semi-definite. If p^* is a minimizer, then B is positive semi-definite, by Lemma 3.1, and $B + \lambda A = B$, since in this case $\lambda = 0$.

Case 2: $\frac{1}{2}p^\intercal Ap = \Delta^2$, and $\lambda > 0$.

$$\begin{aligned}
m(p^*) &\leq m(p) \\
g^\intercal p^* + \frac{1}{2}{p^*}^\intercal Bp^* &\leq g^\intercal p + \frac{1}{2}p^\intercal Bp \\
g^\intercal(p^* - p) &\leq \frac{1}{2}p^\intercal Bp - \frac{1}{2}{p^*}^\intercal Bp^* \\
-{p^*}^\intercal(B + \lambda A)^\intercal(p^* - p) &\leq \frac{1}{2}p^\intercal Bp - \frac{1}{2}{p^*}^\intercal Bp^* \\
(p - p^*)^\intercal(B + \lambda A)p^* &\leq \frac{1}{2}p^\intercal Bp - \frac{1}{2}{p^*}^\intercal Bp^* \\
(p - p^*)^\intercal(B + \lambda A)p^* &\leq \frac{1}{2}p^\intercal Bp - \frac{1}{2}{p^*}^\intercal Bp^* + \frac{\lambda}{2}p^\intercal Ap - \frac{\lambda}{2}{p^*}^\intercal Ap^* && \text{Since } \frac{1}{2}p^\intercal Ap = \frac{1}{2}{p^*}^\intercal Ap = \Delta^2 \\
(p - p^*)^\intercal(B + \lambda A)p^* &\leq \frac{1}{2}p^\intercal(B + \lambda A)p - \frac{1}{2}{p^*}^\intercal(B + \lambda A)p^* \\
0 &\leq \frac{1}{2}p^\intercal(B + \lambda A)p - p^\intercal(B + \lambda A)p^* + \frac{1}{2}{p^*}^\intercal(B + \lambda A)p^* \\
0 &\leq \frac{1}{2}(p - p^*)^\intercal(B + \lambda A)(p - p^*)
\end{aligned}$$

Finally, we must show that $p - p^*$ is dense, that is, $\forall\ u \in \mathbb{R}^n$, we can find $p - p^* = \alpha u$ such that $\frac{1}{2}p^\intercal Ap = \Delta^2$. This is equivalent to showing there exists an α such that $\frac{1}{2}(\alpha u + p^*)^\intercal A(\alpha u + p^*) = \Delta^2$.

$$\begin{aligned}
\frac{1}{2}(\alpha u + p^*)^\intercal A(\alpha u + p^*) &= \Delta^2 \\
\frac{1}{2}\alpha^2 u^\intercal Au + \alpha u^\intercal Ap^* + \frac{1}{2}{p^*}^\intercal Ap^* &= \Delta^2 \\
\frac{1}{2}\alpha^2 u^\intercal Au + \alpha u^\intercal Ap^* &= 0 \\
\frac{1}{2}\alpha u^\intercal Au + u^\intercal Ap^* &= 0
\end{aligned}$$

We need to show this term is 0 by picking α, and now we have three subcases, depending on properties of u.

Subcase 1: $u \in Null(A)$.

Clearly the term is 0, since $Au = 0$ by definition. So all values of α satisfy this.

Subcase 2: $u^\intercal Ap^* \neq 0$.

We can simply solve $\alpha = \frac{-2u^\intercal Ap^*}{u^\intercal Au}$.

Subcase 3: $u^\intercal Ap^* = 0$.

We need to show that if this is the case, that $u^\intercal(B + \lambda A)u \geq 0$. Since the space is continuous, for any u, we can define a function $\hat{u}(\epsilon) = u + \epsilon p^*$.

From the two subcases above, we know $\lim_{\epsilon \to 0^-} \hat{u}(\epsilon)^\intercal(B + \lambda A)\hat{u}(\epsilon) = \lim_{\epsilon \to 0^+} \hat{u}(\epsilon)^\intercal(B + \lambda A)\hat{u}(\epsilon) \geq 0$. By continuity, we have our result that $u^\intercal(B + \lambda A)u \geq 0$.

□