Homework 3

Yanyao Shen

October 20, 2017

Problems are solved individually.

1 Problem 3

Derive necessary and sufficient conditions for p^* being an optimal solution to the following optimization problem:

$$\textbf{minimize}_p g^T p + \frac{1}{2} p^T B p \tag{1.1}$$

$$\textbf{subject to } p^T A p \leq d^2 \tag{1.2}$$

where B is symmetric and A is positive semidefinite (not necessarily positive definite).

Solution:
We use the following lemma from the class:

Lemma 1. *Lemma m be the quadratic function defined by*

$$m(p) = g^T p + \frac{1}{2} p^T B p, \tag{1.3}$$

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

- *m attains a minimum if and only if B is positive semi-definite and g 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.*

The proof of this Lemma is shown in the lecture notes. Back to the original problem, we conclude that the p^* is a global solution if.f. p^* is feasible and $\exists \lambda \geq 0$, such that

$$(B + \lambda A)p^* = -g, \tag{1.4}$$

$$\lambda(d - \sqrt{p^{*T} A p^*}) = 0, \tag{1.5}$$

$$B + \lambda A \succeq 0. \tag{1.6}$$

$\Leftarrow$. If $\exists \lambda \geq 0$, such that the above conditions hold. Then, p^* is a minimizer of

$$\hat{m}(p) = g^T p + \frac{1}{2} p^T (B + \lambda A) p = m(p) + \frac{\lambda}{2} p^T A p. \tag{1.7}$$

Since p^* is the minimizer, $\hat{m}(p) \geq \hat{m}(p^*)$. We have

$$m(p) \geq m(p^*) + \frac{\lambda}{2}({p^*}^T A p^* - p^T A p) \tag{1.8}$$

Since

$$\lambda(d^2 - {p^*}^T A p^*) = 0, \tag{1.9}$$

$$m(p) \geq m(p^*) + \frac{\lambda}{2}(d^2 - p^T A p), \tag{1.10}$$

Therefore, p^* is the minimizer of m.

$\Rightarrow$. If p^* is the minimizer, we consider two cases.
If ${p^*}^T A p^* < d^2$. Then, this is a locally unconstrained minimizer, by taking the first and second derivative, we have

$$Bp^* + g = 0, B \succeq 0. \tag{1.11}$$

If ${p^*}^T A p^* = d^2$. Then, on the boundary, the normal vector for m and the constraint should be aligned, that is to say, $\exists \lambda$, such that

$$Bp^* + g + \lambda A p^* = 0, \Rightarrow (B + \lambda A) p^* = -g. \tag{1.12}$$

For other p that satisfies $p^T A p = d^2$, we have:

$$m(p) \geq m(p^*) + \frac{\lambda}{2}({p^*}^T A p^* - p^T A p). \tag{1.13}$$

Since we know $p^* = (B + \lambda A)^{-1} g$, the above inequality can be converted to

$$g^T(p - p^*) + \frac{1}{2}(p^T B p - {p^*}^T B p^*) + \frac{\lambda}{2}(p^T A p - {p^*}^T A p^*) \geq 0 \tag{1.14}$$

$$\Rightarrow {p^*}^T (B + \lambda A)(p - p^*) + \frac{1}{2}(p^T B p - {p^*}^T B p^*) + \frac{\lambda}{2}(p^T A p - {p^*}^T A p^*) \geq 0 \tag{1.15}$$

$$\Rightarrow \frac{1}{2}(p - p^*)^T (B + \lambda A)(p - p^*) \geq 0. \tag{1.16}$$

Therefore, $B + \lambda A \succeq 0$.
Next, we argue that $\lambda \geq 0$, if all possible λ are negative, then, being a minimizer of $\hat{m}$ implies that p^* is a global minimizer of m, which then gives $\lambda \geq 0$ that satisfies the condition, which is contradict with the assumption.
Therefore, we have proved that the condition we give is both necessary and sufficient.

2 Problem 4

Derive a close-form solution for the 2D trust-region problem under the L1-norm:

$$\textbf{minimize}_{x,y \in \mathbb{R}} \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{2.1}$$

$$\textbf{subject to } |x| + |y| \leq d. \tag{2.2}$$

Your solution should be as a close-form of $b_1, b_2, a_{11}, a_{12}, a_{22}, d$.

Solution:

Comment on 'close-form'. Any result containing 'if' statement and 'argmax/argmin' over finite terms are considered close form, since they can be easily converted to operation of *Identity functions*.
The problem should be considered in different cases: minimizing an ellipsoid, maximizing an ellipsoid, or maximizing/minimizing a hyperbola. This depends on whether matrix A's eigenvalues are: (i) both positive; (ii) both negative; or (iii) one negative, one positive.
We denote the matrix consists of a_{ij} as A, the vector consists of b_i as b.

$A \succ 0$. Minimizing the ellipsoid. If $\|A^{-1}b\|_1 \leq d$, then we can set vector $[x, y]^T$ to be $-A^{-1}b$, which is the global minimizer. If this is not the case, then the equality of the constraint must be satisfied. If the center $-A^{-1}b$ gives positive x and positive y, then we know that $x^* + y^* = d, x^*, y^* \in [0, d]$. For negative x, positive y, or positive x, negative y, or negative x and y, we get similar condition for x^*, y^*. Since the results are symmetric, we only consider the case where the center has positive x and positive y. We now use the fact that the gradient of the minimizer must be aligned with the normal of the constraint. Let f be the function we minimize, then,

$$\frac{\partial f/\partial y}{\partial f/\partial x} = 1 \tag{2.3}$$

Or, in matrix form (notice we use x for $[x, y]^T$), we have:

$$Ax^* + b + \lambda \mathbf{1} = 0. \tag{2.4}$$

We also know that

$$x^T \mathbf{1} = d \tag{2.5}$$

Therefore, we have

$$x = A^{-1}b + \frac{d - b^T A^{-1}\mathbf{1}}{\mathbf{1}^T A^{-1}\mathbf{1}} A^{-1}\mathbf{1} \tag{2.6}$$

If x is not element-wise positive, we truncate the result to either $(0, d)$ or $(d, 0)$.

$A \prec 0$. Maximizing the ellipsoid. Just calculate the 4 corners and take the maximum.

A has one negative eigenvalue, and one positive eigenvalue We are minimizing a hyperbola. Hyperbola contains two bunch of contours, so the center would not be the optimal point, and the minimizer must be on the boundary. We just solve all possible combinations and find the minimal one. Let $r \in \{-d, d\}, s \in \{[1, 1]^T, [-1, 1]^T\}$, we calculate

$$x = A^{-1}b + \frac{r - b^T A^{-1}\mathbf{s}}{\mathbf{s}^T A^{-1}\mathbf{s}} A^{-1}\mathbf{s} \tag{2.7}$$

for all r, s (4 combinations). Then, we eliminate those outside the boundary, and compare them with the 4 corner's value and find the minimum.

A is singular. If A is zero matrix. Then, we find the one in b_1 and b_2 with larger norm, say b_1, and then set $x = -sgn(b_1)d, y = 0$. If only one eigenvalue of A is zero, then we are minimizing a parabola. There is a special case where the function can be written in the form $f = \lambda(v^T x - k)^2$, where λ is the non-zero eigenvalue (must be positive, if negative, just calculate corners), v is the corresponding eigenvector. The minimizer satisfies $x = kv + tu$, where $u^T v = 0$ and t can be any value. Since the possible solution lies in a line, we can find x by finding t to make x on the boundary. Notice that in this case, the gradient of the function is zero.
In all other cases, we only need to compute the gradient and find all possible boundary points, compare them with the corners and take the minimum. We solve the following problem:

$$(Ax + b)^T([0, 2]^T - s) = 0 \tag{2.8}$$
$$x^T s = r \tag{2.9}$$

It is possible that above equation has no solution for some s and r, and we just ignore it. Given all the results from above equation, we compare their value with the 4 corners and pick the minimum.

3 Problem 6

We consider solving the following "sub-problem" using conjugate gradient descent:

$$\textbf{minimize}_p \phi(p) := g^T p + \frac{1}{2} p^T B p \tag{3.1}$$

$$\textbf{subject to } p^T C p \le d^2. \tag{3.2}$$

Here B is a real symmetric matrix and C is positive definite matrix.
The algorithm we consider is described as follows:

- **Step I: Set $p_0 = 0$, $r_0 = -g$. Solve $C\hat{r_0} = r_0$. Set $d_0 = \hat{r_0}, i = 0$.**
- **Step II: Compute $\gamma_i = d_i^T B d_i$. If $\gamma_i > 0$, then continue with Step III. Otherwise compute $\tau > 0$ so that $(p_i + \tau d_i)^T C (p_i + \tau d_i) = d^2$, set $p = p_i + \tau d_i$ and terminate.**
- **Step III: Compute $\alpha_i = \frac{r_i^T \hat{r_i}}{\gamma_i}, p_{i+1} = p_i + \alpha_i d_i$. If $p_{i+1}^T C p_{i+1} < d^2$, then continue with step IV. Otherwise compute $\tau > 0$ so that $(p_i + \tau d_i)^T C (p_i + \tau d_i) = d^2$, set $p = p_i + \tau d_i$ and terminate.**
- **Step IV: Compute $r_{i+1} = r_i - \alpha_i B d_i$. If $\frac{r_{i+1}^T C r_{i+1}}{g^T C g} \le \epsilon^2$ then set $p = p_{i+1}$ and terminate. Otherwise continue with step V.**
- **Step V: Solve $C\hat{r_{i+1}} = r_{i+1}$. Compute $\beta_i = \frac{r_{i+1}^T \hat{r_{i+1}}}{r_i^T \hat{r_i}}$ and $d_{i+1} = \hat{r_{i+1}} + \beta_i d_i$. Set $i := i + 1$ and continue with step II.**

Let $p_j, j = 0, \cdots, i$ be the iterates generated by the algorithm described above.

- **Show that $\phi(p_j)$ is strictly decreasing and $\phi(p) \le \phi(p_i)$.**
- **Show that $p_j^T C p_j$ is strictly increasing for $j = 0, \cdots, i$ and $p^T C p \ge p_j^T C p_j$.**

Solution:
Calculating $(p_i + \tau d_i)^T C (p_i + \tau d_i)$ in terms of τ is calculating the positive root of the following quadratic function in τ:

$$\tau^2 \cdot d_i^T C d_i + 2\tau \cdot p_i^T C d_i = d^2 - p_i^T C p_i. \tag{3.3}$$

Lemma 2. *We notice that for $\gamma_i \ne 0, i = 0, \cdots, k$*

$$r_i^T d_j = r_j^T \hat{r_j}, 0 \le i \le j \le k \tag{3.4}$$

$$d_i^T C d_j = \frac{r_j^T \hat{r_j} d_i^T C d_i}{r_i^T \hat{r_i}}, 0 \le i \le j \le k \tag{3.5}$$

$$\phi(p_{i+1}) = \phi(p_i) - \frac{1}{2} \frac{(r_i^T \hat{r_i})^2}{\gamma_i}, 0 \le i < k. \tag{3.6}$$

The Lemma is proved by induction. We try to conclude the following equations are also correct:

$$r_j^T d_i = 0, j \ge i + 1 \tag{3.7}$$

$$d_j^T B d_i = 0, j \ge i + 1 \tag{3.8}$$

$$r_i^T \hat{r_{j+1}} = 0, j \ge i \tag{3.9}$$

By induction, we assume that the above equations are correct for some $j \le k$. Then, for (3.7),

$$r_{k+1}^T d_i = d_i^T r_{k+1} = d_i^T (r_k - \alpha_k B d_k) \tag{3.10}$$

$$= 0 - 0 = 0 \text{ if } k \ge i + 1 \tag{3.11}$$

$$= r_k^T (d_k - \hat{r_k}) = \beta_{k-1} r_k^T d_{k-1} = 0 \text{ if } k = i \tag{3.12}$$

For (3.8),

$$d_{k+1}^T B d_i = (\hat{r_{k+1}} + \beta_k d_k)^T B d_i \tag{3.13}$$

Notice that we can rewrite

$$\begin{aligned}
\beta_k =& \frac{r_{k+1}^T \hat{r_{k+1}}}{r_k^T \hat{r_k}} = \frac{\hat{r_{k+1}}^T r_{k+1}}{\hat{r_k}^T r_k} && (3.14)\\
=& \frac{\hat{r_{k+1}}^T (r_{k+1} - r_k)}{\hat{r_k}^T (r_k - r_{k+1})} \text{ due to (3.9)} && (3.15)\\
=& -\frac{\hat{r_{k+1}}^T B d_k}{d_k^T B d_k} \text{ due to (3.8)} && (3.16)
\end{aligned}$$

Therefore, for (3.13), when $i = k$, using (3.16), the RHS is zero, when $i < k$, RHS is also zero directly based on (3.7) and (3.8). For (3.9),

$$\begin{aligned}
r_i^T \hat{r_{k+2}} =& r_{k+2}^T C^{-1} r_i && (3.17)\\
=& (r_{k+1} - \alpha_{k+1} B d_{k+1})^T \hat{r_i} && (3.18)\\
=& (r_{k+1} - \alpha_{k+1} B d_{k+1})^T (d_i - \beta_i d_{i-1}) && (3.19)\\
=& 0 \text{ discuss for } i = k+1 \text{ and } i < k+1 && (3.20)
\end{aligned}$$

For (3.4), with $i = j = k+1$,

$$r_{k+1}^T d_{k+1} = r_{k+1}^T \hat{r_{k+1}} + \beta_k r_{k+1}^T d_k = r_{k+1}^T \hat{r_{k+1}}. \tag{3.21}$$

When $i < j = k+1$,

$$r_i^T d_{k+1} = \beta_k r_i^T d_k + r_i^T \hat{r_{k+1}} = \beta_k r_k^T r_k = r_{k+1}^T r_{k+1}. \tag{3.22}$$

For (3.5), assume the ratio between $\frac{d_i^T C d_j}{r_j^T \hat{r_j}} = s$, then,

$$\begin{aligned}
d_i^T C d_{j+1} =& \beta_i d_i^T C d_j + d_i^T C \hat{r_{j+1}} && (3.23)\\
=& s r_{j+1}^T \hat{r_{j+1}} + 0 && (3.24)
\end{aligned}$$

For (3.6),

$$\begin{aligned}
\phi(p_{i+1}) - \phi(p_i) =& g^T (p_{i+1} - p_i) + \frac{1}{2} p_{i+1}^T B p_{i+1} - \frac{1}{2} p_i^T B p_i && (3.25)\\
=& \alpha_i g^T d_i + \frac{1}{2} \alpha_i^2 d_i^T B d_i + \alpha_i d_i^T B p_i && (3.26)
\end{aligned}$$

Since $g = -r_0$, the first term in (3.26) is $-\alpha_i r_i^T \hat{r}_i$. The second term is $\frac{1}{2}\alpha_i^2 \gamma_i$. For the third term, we recursively expand the term p_i, and it is easy to see that this term is 0. Therefore,

$$\phi(p_{i+1}) - \phi(p_i) = -\alpha_i r_i^T \hat{r}_i + \frac{1}{2} \alpha_i^2 \gamma_i = -\frac{1}{2} \frac{(r_i^T \hat{r}_i)^2}{\gamma_i}. \tag{3.27}$$

Now, we have proved the Lemma. Back to the original problem.
From Step III, we have:

$$p_j = \sum_{k=0}^{j-1} \alpha_k d_k, j = 0, \cdots, i \tag{3.28}$$

and

$$\alpha_j > 0, j = 0, \cdots, i-1 \tag{3.29}$$

Therefore,

$$p_j^T C d_j > 0, j = 0, \cdots, i \tag{3.30}$$

due to (3.5). Therefore, based on (3.29) and (3.30), we have:

$$p_{j+1}^T C p_{j+1} = p_j^T C p_j + 2\alpha_j p_j^T C d_j + \alpha_j^2 d_j^T C d_j \geq p_j^T C p_j. \tag{3.31}$$

This implies that $p_j^T C p_j$ is strictly increasing. If $p = p_{i+1}$, then $p_T C p \geq p_j^T C p_j$ is trivial. On the other hand, if $p = p_i + \tau d_i$, then,

$$p^T C p = p_i^T C p_i + 2\tau p_i^T C d_i + \tau^3 d_i^T C d_i > p_i^T C p_i \tag{3.32}$$

using (3.30) and the fact that $\tau > 0$.
From (3.6), we already know that $\phi(p_j)$ is strictly decreasing. Consider p now. If $p = p_{i+1}$, then the result follows directly. From (3.3), we have that:

$$r_i^T d_i = r_i^T \hat{r}_i = (Bp_i + g)^T C^{-1} (Bp_i + g) > 0, \tag{3.33}$$

using that C is positive definite, hence d_i is a descent direction for $\phi(p_i)$, i.e.,

$$-\left(\nabla \phi(p_i)\right)^T d_i = (Bp_i + g)^T d_i > 0. \tag{3.34}$$

If $\gamma > 0$, then we have

$$\phi(p_i) \geq \phi(p_i + \tau d_i) \geq \phi(p_{i+1}), \text{ for } 0 < \tau \leq \alpha_i. \tag{3.35}$$

Since $\tau \leq \alpha_i$, we have the desired result. For $\gamma_i \leq 0$, then the quadratic term is non-positive, and we have

$$\phi(p_i) \geq \phi(p_i + \tau d_i) \text{ for } 0 \leq \tau, \tag{3.36}$$

and the result follows directly.