

# Continuous Optimization: Assignment 3

Due on May 14, 2024

Honglu Ma

Hiroyasu Akada

Mathivathana Ayyappan

## Exercise 1

1.  $f(x) - f(\bar{x}) = \int_0^1 \langle \nabla f(\bar{x} + t(x - \bar{x}), x - \bar{x}) \rangle dt$   
should be  $f(x) - f(\bar{x}) = \int_0^1 \langle \nabla f(\bar{x} + t(x - \bar{x}), x - \bar{x}) \rangle dt$
2. (continue from the previous equation)  $\dots \left| \int_0^1 \langle \nabla f(\bar{x} + t(x - \bar{x}) - \nabla f(\bar{x}), x - \bar{x}) \rangle dt \right|$   
should drop the absolute value sign
3. The correct equation consists the previous three errors should be

$$\begin{aligned}
 f(x) - f(\bar{x}) &= \langle \nabla f(\bar{x}), x - \bar{x} \rangle \\
 &= \int_0^1 \langle \nabla f(\bar{x} + t(x - \bar{x}), x - \bar{x}) - \nabla f(\bar{x}), x - \bar{x} \rangle dt \\
 &\leq \int_0^1 \|\nabla f(\bar{x} + t(x - \bar{x})) - \nabla f(\bar{x})\| \|x - \bar{x}\| dt
 \end{aligned}$$

Cauchy Schwarz Inequality

4. Continue with previous derivation we have

$$\begin{aligned}
 &\int_0^1 \|\nabla f(\bar{x} + t(x - \bar{x})) - \nabla f(\bar{x})\| \|x - \bar{x}\| dt \\
 &\leq \int_0^1 L \|\bar{x} + t(x - \bar{x}) - \bar{x}\| \|x - \bar{x}\| dt
 \end{aligned}$$

Lipschitz Continuous Gradient

$$\begin{aligned}
 &= \int_0^1 L \|t(x - \bar{x})\| \|x - \bar{x}\| dt \\
 &= L \|x - \bar{x}\|^2 \int_0^1 t dt \\
 &= L \|x - \bar{x}\|^2 \left( \frac{t^2}{2} \Big|_0^1 \right) \\
 &= \frac{L}{2} \|x - \bar{x}\|^2
 \end{aligned}$$

where in the original proof, there is an error at  $\dots \int_0^1 \frac{Lt}{2} \dots$

- 5-10. There are several errors in the last part of the proof. In addition to adding a half to  $Lt$ . The derivation also does not follow the correct steps of integration. Item 3 and 4 above show the correct derivation and conclude the proof.

## Exercise 2

(a)

Since  $\mathbf{Q}$  is a square matrix, we can write  $\mathbf{Q} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$  where  $\mathbf{U}$  is an orthogonal matrix and  $\mathbf{\Lambda}$  is a diagonal matrix with the eigenvalues of  $\mathbf{Q}$  i.e.  $\lambda_i$  on the diagonal. Then, we have

$$\begin{aligned}
 \langle \mathbf{x}, \mathbf{Q}\mathbf{x} \rangle &= \langle \mathbf{x}, \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top \mathbf{x} \rangle \\
 &= \mathbf{x}^\top \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top \mathbf{x} \\
 &= (\mathbf{U}^\top \mathbf{x})^\top \mathbf{\Lambda} \mathbf{U}^\top \mathbf{x} \\
 &= (\mathbf{U}\mathbf{x})^\top \mathbf{\Lambda} (\mathbf{U}\mathbf{x}) & \mathbf{U} = \mathbf{U}^\top \\
 &= \sum_{i=1}^n \lambda_i ((\mathbf{U}\mathbf{x})_i)^2 \\
 &\leq \sum_{i=1}^n \lambda_{\max}(\mathbf{Q}) ((\mathbf{U}\mathbf{x})_i)^2 & \lambda_i \leq \lambda_{\max}(\mathbf{Q}) \\
 &= \lambda_{\max}(\mathbf{Q}) \sum_{i=1}^n ((\mathbf{U}\mathbf{x})_i)^2 \\
 &= \lambda_{\max}(\mathbf{Q}) (\mathbf{U}\mathbf{x})^\top \mathbf{U}\mathbf{x} \\
 &= \lambda_{\max}(\mathbf{Q}) \mathbf{x}^\top \mathbf{U}^\top \mathbf{U}\mathbf{x} \\
 &= \lambda_{\max}(\mathbf{Q}) \mathbf{x}^\top \mathbf{x} & \mathbf{U}^\top \mathbf{U} = \mathbf{I} \\
 &= \lambda_{\max}(\mathbf{Q}) \|\mathbf{x}\|^2
 \end{aligned}$$

Similar derivation can be shown for the smallest eigenvalue:  $\langle \mathbf{x}, \mathbf{Q}\mathbf{x} \rangle \geq \lambda_{\min}(\mathbf{Q}) \|\mathbf{x}\|^2$ .

(b)

Suppose  $\lambda$  is an eigenvalue of  $\mathbf{Q}$  with eigenvector  $\mathbf{v}$ . Then, we have

$$\begin{aligned}
 \mathbf{Q}\mathbf{v} &= \lambda\mathbf{v} \Rightarrow \tau\mathbf{Q}\mathbf{v} = \tau\lambda\mathbf{v} \\
 &\Rightarrow \mathbf{I}\mathbf{v} - \tau\mathbf{Q}\mathbf{v} = \mathbf{I}\mathbf{v} - \tau\lambda\mathbf{v} \\
 &\Rightarrow (\mathbf{I} - \tau\mathbf{Q})\mathbf{v} = (\mathbf{I} - \tau\text{diag}(\lambda))\mathbf{v}
 \end{aligned}$$

$\mathbf{I} - \tau\text{diag}(\lambda)$  is a matrix with same diagonal entries  $1 - \tau\lambda$

$$\Rightarrow (\mathbf{I} - \tau\mathbf{Q})\mathbf{v} = (1 - \tau\lambda)\mathbf{v}$$

Above shows that  $1 - \tau\lambda$  is an eigenvalue of  $\mathbf{I} - \tau\mathbf{Q}$ .

$$\begin{aligned}
 &\Rightarrow (\mathbf{I} - \tau\mathbf{Q})(\mathbf{I} - \tau\mathbf{Q})\mathbf{v} = (\mathbf{I} - \tau\mathbf{Q})(1 - \tau\lambda)\mathbf{v} \\
 &\Rightarrow (\mathbf{I} - \tau\mathbf{Q})(\mathbf{I} - \tau\mathbf{Q})\mathbf{v} = (1 - \tau\lambda)(\mathbf{I} - \tau\mathbf{Q})\mathbf{v} \\
 &\Rightarrow (\mathbf{I} - \tau\mathbf{Q})(\mathbf{I} - \tau\mathbf{Q})\mathbf{v} = (1 - \tau\lambda)(1 - \tau\lambda)\mathbf{v} \\
 &\Rightarrow (\mathbf{I} - \tau\mathbf{Q})^2\mathbf{v} = (1 - \tau\lambda)^2\mathbf{v}
 \end{aligned}$$

Thus  $(1 - \tau\lambda)^2$  is an eigenvalue of  $(\mathbf{I} - \tau\mathbf{Q})^2$  for each eigenvalue  $\lambda$  of  $\mathbf{Q}$ .

## Exercise 3

(a)

Projection of  $\mathbf{v}$  onto the space spanned by the columns of  $\mathbf{A}$  means that we want to find a point in the column space of  $\mathbf{A}$  that is closest to  $\mathbf{v}$  i.e. for vector  $\mathbf{p}$  in the column space we want the distance between  $\mathbf{v}$  and  $\mathbf{p}$  to be as small as possible

$$\operatorname{argmin}_p \operatorname{dist}(\mathbf{v}, \mathbf{p})$$

which is equivalent to

$$\operatorname{argmin}_p \|\mathbf{v} - \mathbf{p}\|$$

or

$$\operatorname{argmin}_p \sqrt{\langle \mathbf{v} - \mathbf{p}, \mathbf{v} - \mathbf{p} \rangle}$$

(b)

In case of  $m = 1$ , the subspace spanned by  $\mathbf{u}$  is just a line  $c\mathbf{u}$  for some constant  $c \in \mathbb{R}$ . We have

$$\operatorname{argmin}_c \sqrt{\langle \mathbf{v} - c\mathbf{u}, \mathbf{v} - c\mathbf{u} \rangle}$$

which by definition of the inner product, is equivalent to

$$\operatorname{argmin}_c (\mathbf{v} - c\mathbf{u})^\top \mathbf{Q}(\mathbf{v} - c\mathbf{u})$$

To find minimum, we take derivative of function

$$f(c) = (\mathbf{v} - c\mathbf{u})^\top \mathbf{Q}(\mathbf{v} - c\mathbf{u}) = \mathbf{v}^\top \mathbf{Q}\mathbf{v} - 2(\mathbf{v}^\top \mathbf{Q}\mathbf{u})c + (\mathbf{u}^\top \mathbf{Q}\mathbf{u})c^2$$

We have

$$f'(c) = 2(\mathbf{u}^\top \mathbf{Q}\mathbf{u})c - 2(\mathbf{v}^\top \mathbf{Q}\mathbf{u})$$

and the second derivative

$$f''(c) = 2(\mathbf{u}^\top \mathbf{Q}\mathbf{u}) \geq 0$$

which indicate function  $f$  is convex now we just have to solve  $f'(c) = 0$

$$\begin{aligned} 2(\mathbf{u}^\top \mathbf{Q}\mathbf{u})c - 2(\mathbf{v}^\top \mathbf{Q}\mathbf{u}) &= 0 \\ c &= \frac{\mathbf{v}^\top \mathbf{Q}\mathbf{u}}{\mathbf{u}^\top \mathbf{Q}\mathbf{u}} \end{aligned}$$

and the projection  $\mathbf{p} = \frac{\mathbf{v}^\top \mathbf{Q}\mathbf{u}}{\mathbf{u}^\top \mathbf{Q}\mathbf{u}} \mathbf{u}$

(c)

In case of  $m > 1$ , we have a vector  $\hat{\mathbf{c}} \in \mathbb{R}^m$  defined by

$$\hat{\mathbf{c}} := \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix}$$

where the projection  $\mathbf{p} = \mathbf{A}\hat{\mathbf{c}}$ . We have

$$\underset{\hat{\mathbf{c}}}{\operatorname{argmin}} (\mathbf{v} - \mathbf{A}\hat{\mathbf{c}})^\top \mathbf{Q}(\mathbf{v} - \mathbf{A}\hat{\mathbf{c}})$$

To find the minimum, similarly, we define a function

$$g(\hat{\mathbf{c}}) = \mathbf{v}^\top \mathbf{Q}\mathbf{v} - 2\mathbf{v}^\top \mathbf{Q}\mathbf{A}\hat{\mathbf{c}} + \hat{\mathbf{c}}^\top \mathbf{A}^\top \mathbf{Q}\mathbf{A}\hat{\mathbf{c}}$$

We calculate its gradient

$$\nabla g(\hat{\mathbf{c}}) = -2\mathbf{v}^\top \mathbf{Q}\mathbf{A} + 2\mathbf{A}^\top \mathbf{Q}\mathbf{A}\hat{\mathbf{c}}$$

and its Hessian

$$\nabla^2 g(\hat{\mathbf{c}}) = 2\mathbf{A}^\top \mathbf{Q}\mathbf{A} = \mathbf{H}$$

We can show that  $\mathbf{H}$  is positive semi-definite. Suppose  $\lambda$  is an eigenvalue of  $\mathbf{H}$  and with corresponding eigenvector  $\mathbf{e}$

$$\begin{aligned} \mathbf{H}\mathbf{e} &= \lambda\mathbf{e} \\ 2\mathbf{A}^\top \mathbf{Q}\mathbf{A}\mathbf{e} &= \lambda\mathbf{e} \\ 2\mathbf{e}^\top \mathbf{A}^\top \mathbf{Q}\mathbf{A}\mathbf{e} &= \lambda\mathbf{e}^\top \mathbf{e} \\ \lambda &= \frac{2(\mathbf{A}\mathbf{e})^\top \mathbf{Q}\mathbf{A}\mathbf{e}}{\mathbf{e}^\top \mathbf{e}} \geq 0 \end{aligned}$$

Thus we can get the minimizer  $\hat{\mathbf{c}}$  by solving  $\nabla g(\hat{\mathbf{c}}) = 0$

$$\begin{aligned} -2\mathbf{v}^\top \mathbf{Q}\mathbf{A} + 2\mathbf{A}^\top \mathbf{Q}\mathbf{A}\hat{\mathbf{c}} &= 0 \\ \mathbf{A}^\top \mathbf{Q}\mathbf{A}\hat{\mathbf{c}} &= \mathbf{v}^\top \mathbf{Q}\mathbf{A} \end{aligned}$$

## Exercise 4

(a)

Define  $\tilde{A}$  and  $\tilde{f}$  accordingly:

$$\tilde{A} = \begin{pmatrix} A \\ \sqrt{\mu}D \end{pmatrix}, \quad \tilde{f} = \begin{pmatrix} f \\ 0 \end{pmatrix} \quad \text{Note the 0 in } \tilde{f} \text{ denotes a zero vector in } \mathbb{R}^{2N}$$

where  $\tilde{A} \in \mathbb{R}^{3N \times N}$  and  $\tilde{f} \in \mathbb{R}^{3N}$ .

We can verify that the reconstruction is equivalent to the original problem.

$$\begin{aligned} \frac{1}{2} \|\tilde{A}u - \tilde{f}\|^2 &= \frac{1}{2} (\|\tilde{A}u\|^2 - 2(\tilde{A}u)^\top \tilde{f} + \|\tilde{f}\|^2) \\ &= \frac{1}{2} (\|Au\|^2 + \|\sqrt{\mu}Du\|^2 - 2(Au)^\top f + \|f\|^2 + 0) \\ &= \frac{1}{2} (\|Au\|^2 + \mu\|Du\|^2 - 2(Au)^\top f + \|f\|^2) \\ &= \frac{1}{2} (\|Au\|^2 - 2(Au)^\top f + \|f\|^2) + \frac{\mu}{2} \|Du\|^2 \\ &= \frac{1}{2} \|Au - f\|^2 + \frac{\mu}{2} \|Du\|^2 \end{aligned}$$

(b)

We have the following objective function:

$$\begin{aligned}
 L(u) &= \frac{1}{2} \|\tilde{A}u - \tilde{f}\|^2 \\
 &= \frac{1}{2} \|\tilde{A}u\|^2 - \tilde{f}^\top \tilde{A}u + \frac{1}{2} \|\tilde{f}\|^2 \\
 &= \frac{1}{2} \langle u, \tilde{A}^\top \tilde{A}u \rangle + \langle -A^\top f, u \rangle + \frac{1}{2} \|f\|^2
 \end{aligned}$$

with corresponding standard form  $L(u) = \frac{1}{2} \langle u, Qu \rangle + \langle b, u \rangle + c$  where  $Q = \tilde{A}^\top \tilde{A} = A^\top A + \mu D^\top D$ ,  $b = -A^\top f$  and  $c = \frac{1}{2} \|f\|^2$  and its gradient:

$$\nabla L(u) = Qu + b = (A^\top A + \mu D^\top D)u - A^\top f$$

We can find the optimal step size for exact line search by solving the following minimization problem:

$$\min_{\tau_k > 0} L(u^{(k)} - \tau_k \nabla L(u^{(k)}))$$

where the derivation is similar to Exercise 1, Assignment 2. In the end, we get:

$$\tau_k = \frac{\nabla L(u^{(k)})^\top \nabla L(u^{(k)})}{\nabla L(u^{(k)})^\top Q \nabla L(u^{(k)})}$$