Ed Qu

Homework 4

(Dated: March 22, 2016)

QUESTION 1

According to the definition of Lipschitz continuous condition, $\forall x, y$ in the domain of the function $F(z)$

$$\|\nabla F(x) - \nabla F(y)\|_2 \leq L\|x - y\|_2 \tag{1}$$

L is the Lipschitz constant. Here, $F(x)$ is a quadratic function

$$\begin{aligned}\nabla F(x) &= \nabla\left(x^T H^T H x - 2H^T g_1 x + g_1^T g_1\right) \\ &= 2H^T H x - 2H^T g_1 \end{aligned} \tag{2}$$

Therefore, we can reformulate the Eq. 1 as

$$\begin{aligned}\|\nabla F(x) - \nabla F(y)\|_2 &= \|2H^T Hx - 2H^T Hy\|_2 \\ &\leq 2\|H^T H\|_2 \|x - y\|_2\end{aligned}$$

By comparing the Eq. 1 and the Eq. 3, we know that

$$L \leq 2\|H^T H\|_2 \tag{3}$$

Since the matrix H in this inverse problem is a sparse matrix, and Matlab can calculate only the Frobenius norm of a sparse matrix

$$\|H^T H\|_F^2 = \sum_{i=1}^{n} \|H^T H e_i\|_2^2 \tag{4}$$

Creating a vector $x = \sum_{i=1}^{n} c_i e_i$ for which $\|x\|_2 = \sum_i |c_i|^2 = 1$, the 2-norm and the Frobenius norm satisfy

$$\begin{aligned}\|H^T H\|_2^2 &= \|H^T H x\|_2^2 \\ &= \left\|\sum_{i=1}^{n} c_i H^T H e_i\right\|_2^2 \\ &\leq \left(\sum_{i=1}^{n} |c_i| \|H^T H e_i\|_2\right)^2 \\ &\leq \left(\sum_{i=1}^{n} |c_i|^2\right) \sum_{i=1}^{n} \|H^T H e_i\|_2^2 \\ &= \sum_{i=1}^{n} \|H^T H e_i\|_2^2 \\ &= \|H^T H\|_F^2\end{aligned}$$

Therefore, we can regard the $2\|H^T H\|_F$ as the Lipschitz constant,

$$L = 2\|H^T H\|_F \approx 0.0024 \tag{5}$$

QUESTION 2

We have shown in the last question that the function $F(x)$ is Lipschitz continuous. Since the $F(x)$ is second-order differentiable, it must be weakly convex.

For any strongly convex function $f(x)$, we will have

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{m}{2}\|x - y\|_2^2 \tag{6}$$

Here m must be larger than 0. Eq. 6 is equivalent to say that the minimum eigenvalue of the Hessian matrix $\nabla^2 f$ is not smaller than m.

In this question,

$$\nabla^2 F(x) = 2H^T H \tag{7}$$

Using the function " eigs() ", Matlab identifies the Hessian matrix $\nabla^2 F(x)$ as a singular matrix whose smallest eigenvalue is equal to 0. Therefore, $F(x)$ is not a strongly convex function.

QUESTION 3

A We show some reconstructed images in Fig. 1. The step size for which the image was reconstructed is shown on the top. The number of iterations is 100. When $\alpha_k = 3.5/L$, we have the best result in this group of images. As the step size decreases, the convergence rate slows down, and the image becomes blurrier. When we increase the step size to $\alpha_k = 4/L$, the algorithm diverges. We can conclude from these results that an accurate or a smaller enough Lipschitz constant can provide us a reference point for the choice of a constant step size, but sometimes the optimized choice may be larger than the minimum Lipschitz constant.

B Here we choose $\alpha_k = 3.5/L$ as an example. We show the final results in Fig. 2. After the first 10 iterations, most features of the phantom appear in the reconstructed image. In comparison with other algebra methods that we have shown in the previous work, which may converge after hundreds of steps, the gradient descent method converge pretty fast.

C We show the results in Fig. 3. As the step size increase from $0.25/L$ to $3.5/L$, the decay of the root-mean-square errors (RMSE) becomes faster. If the step size jumps over a critical point, where they do not converge anymore, the RMSE will increase to infinity as the iteration goes on.

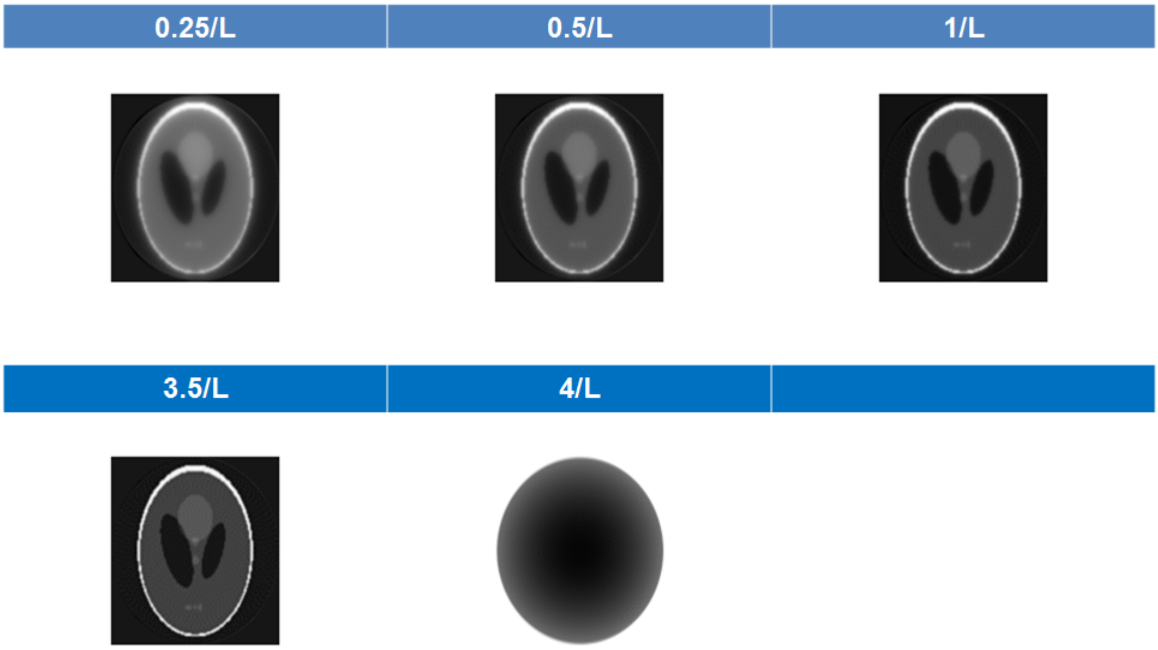

FIG. 1: Reconstructed images using constant step sizes, whose values have been shown on the top of each image.

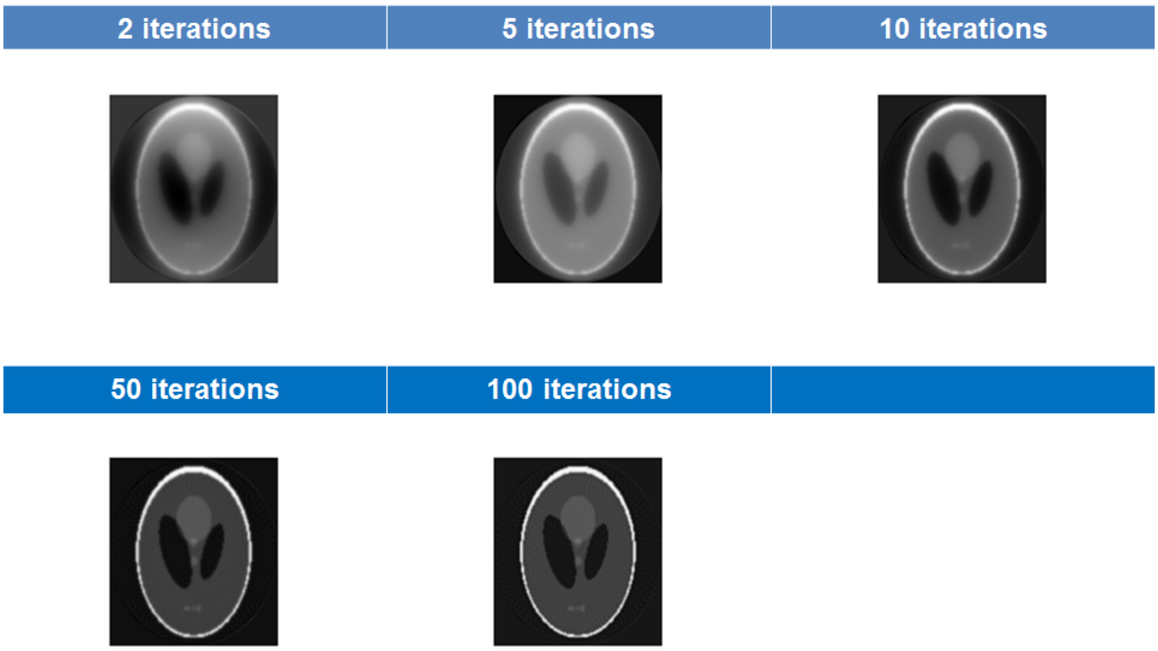

FIG. 2: Reconstructed images using a constant step size $3.5/L$ at several iteration steps. The index of these step have been shown on the top.

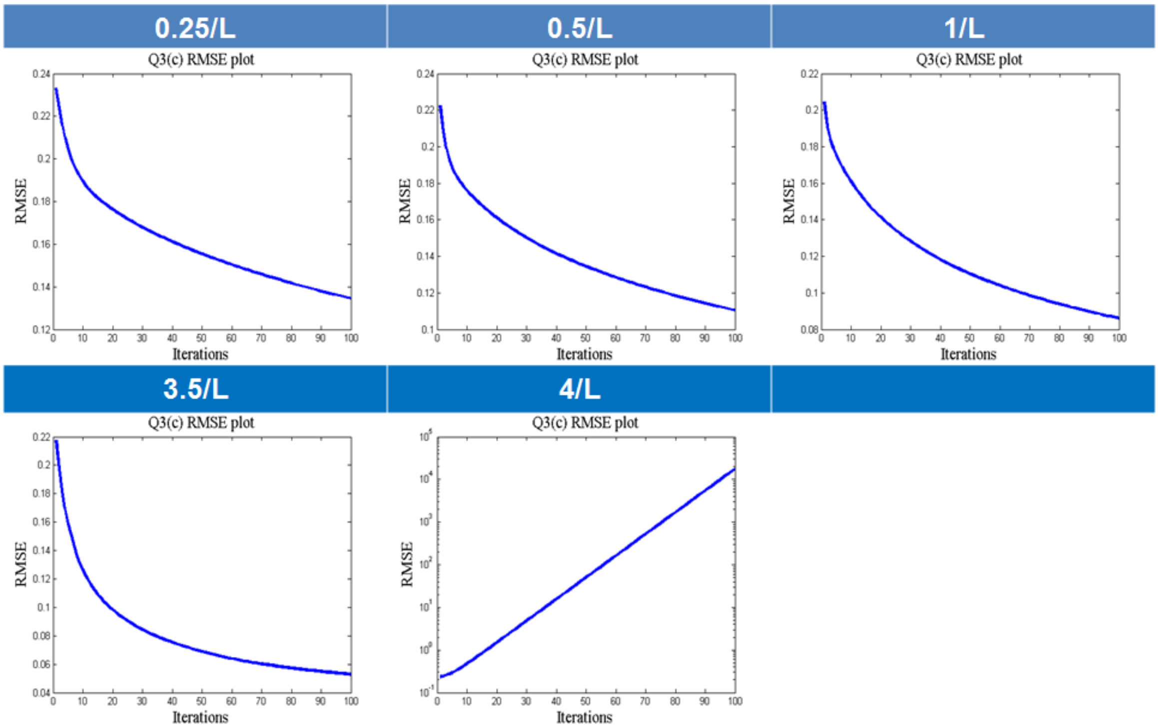

FIG. 3: Root-mean-square errors for different constant step sizes, whose value have been shown on the top.

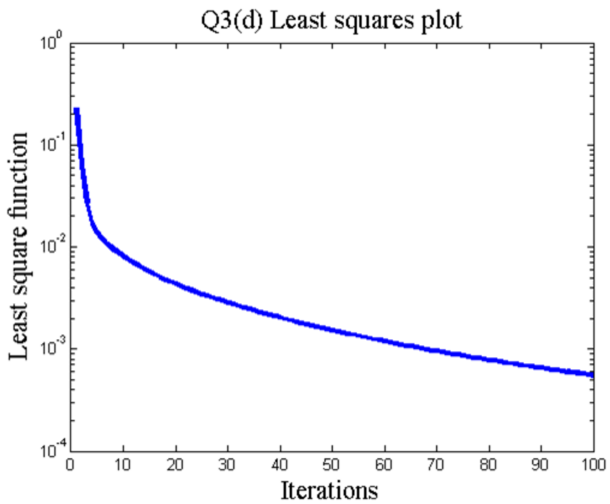

FIG. 4: The value of the least square function $F(x)$ v.s. iteration numbers.

D We show the results in Fig. 4. The value of the cost function converges to 0 as the iteration number increases. This is my expectation.

We have proved that the cost function $F(x)$ is convex, and hence it must converge to the its global minimum. For the least square expression $\|g_1 - Hx\|_2^2$, the global minimum is 0. Therefore, the cost function should converge to 0.

QUESTION 4

The backtracking line search requires three parameters and one condition to adjust the step size t_k. The first parameter is the initial guess α_{max} for the step size $t_k = \alpha_{max}$; the second one is the adjustment coefficient β; the third one is the restriction coefficient η. The condition is

$$\|g_1 - Hx\|_2^2 - \|g_1 - H(x - t_k \nabla F(x))\|_2^2 < \eta t_k \|\nabla F(x)\|_2^2 \quad (8)$$

If the logical value of this condition is true, then we update $t_k = \beta t_k$.

A Here I chose $\alpha_{max} = 7/L$, because in this question we always assume $\beta = 0.5$. According to our result in *Question 3*, we have a relatively highly efficient reconstruction algorithm when $\alpha_k = 3.5/L$. If the initial step size t_k does not fulfill the condition (Eq. 8), the step size t_k will jump to a value which provides us a efficient reconstruction algorithm after once adjustment.

B We show the results in Fig. 5. The image qualities at different iterations are pretty similar what we have in Fig. 2.

C We show the result in Fig. 6. There is no a monotonic function which can describe the relation between η and the convergence rate. When we increase η to 0.5, the convergence rate is higher than

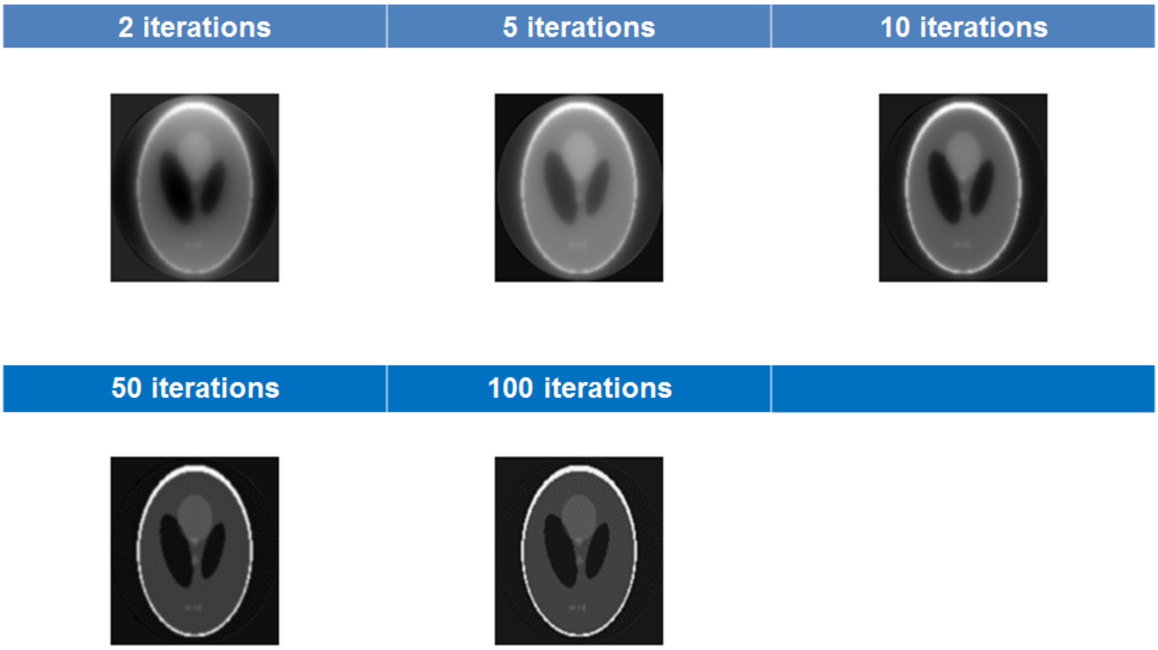

FIG. 5: Reconstructed images using backtracking line search ($\beta = 0.5$, $\eta = 10^{-4}$) at several iteration steps. The index of these step have been shown on the top.

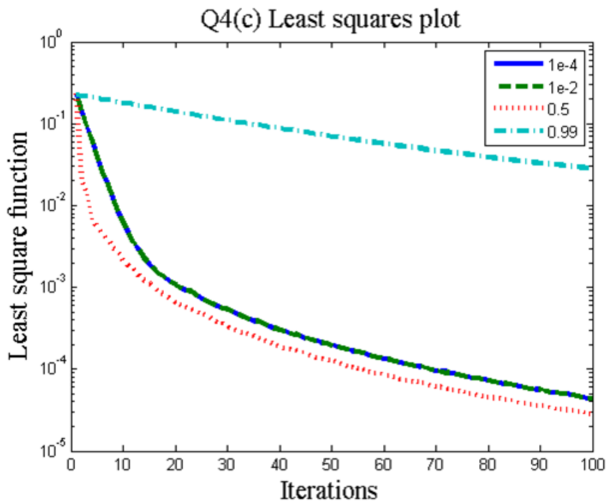

FIG. 6: The values of the least square function $F(x)$ in terms of different restriction coefficient η v.s. iteration numbers. The results for $\eta = 10^{-4}$ and for $\eta = 10^{-2}$ overlap together.

the case in which η is equal to 10^{-4} or 10^{-2}. If we further increases η to 0.99, the convergence rate in terms of the least square function $F(x)$ decreases, even lower than the cases in which η is equal to 10^{-4} or 10^{-2}.

D We show the result in Fig. 7. The result is almost same as what we have in Fig. 7 for $\alpha_k = 3.5/L$, but this result is better than choices of α_k.

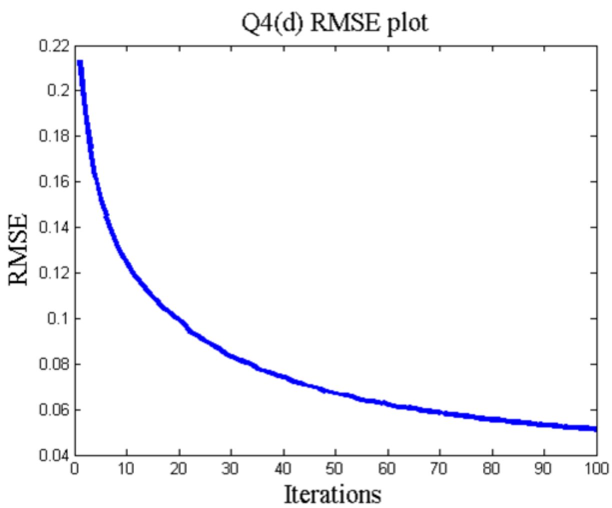

FIG. 7: Root-mean-square errors for the backtracking line search.