The University of British Columbia

CPSC 406

Computational Optimization

Homework #4

Student name:

ALI SIAHKOOHI

Student ID:

84264167

Instructor:

Professor Ascher

Date Submitted:

November 8^{th}, 2016

Question #1

Part (a)

For a matrix we know that the condition number in 2-norm is equal to the ratio of its largest singular value to the smallest one. In other words, a $m \times n$ matrix A which $m \geq n$, has n singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$. We can write the following equation for 2-norm condition number of matrix A:

$$\kappa_2(A) = \frac{\sigma_1}{\sigma_n}. \qquad (1-1)$$

In order to confirm the inequality in the exercise we need to compute the condition number of the matrix $B = A^T A + \gamma I$. Considering the singular value decomposition of matrix A we can write:

$$A = U_{m\times m}\Sigma_{m\times n}V_{n\times n}{}^T \Rightarrow A^T A = \left(V\Sigma^T U^T\right)\left(U\Sigma V^T\right) = V\Sigma^T\Sigma V^T \qquad (1-2)$$

In the equation above, $\Sigma^T\Sigma$ is a diagonal $n \times n$ matrix which its diagonal elements are $\sigma_1{}^2, \sigma_2{}^2, \ldots, \sigma_n{}^2$. If $\bar{\Sigma} = \Sigma^T\Sigma$ then we can rewrite the equation $(1-2)$ as:

$$A^T A = V\bar{\Sigma}V^T \qquad (1-3)$$

We define a diagonal matrix with diagonal entries equal to γ:

$$\Gamma = \gamma I_{n\times n} \qquad (1-4)$$

Using the mentioned matrix and equation $(1-2)$ and $(1-3)$ we can write the singular value decomposition of matrix B:

$$\begin{aligned}
&A^T A + \gamma I = V\Sigma^T\Sigma V^T + \gamma V V^T = V\Sigma^T\Sigma V^T + V\Gamma V^T, \\
&\Rightarrow A^T A + \gamma I = V\left(\Sigma^T\Sigma V^T + \Gamma V^T\right) = V\left(\Sigma^T\Sigma + \Gamma\right)V^T, \\
&\Rightarrow A^T A + \gamma I = V\left(\Sigma^T\Sigma + \gamma I\right)V^T.
\end{aligned} \qquad (1-5)$$

So the singular values of matrix B are as follows:

$$\sigma_1{}^2 + \gamma \geq \sigma_2{}^2 + \gamma \geq \ldots \geq \sigma_n{}^2 + \gamma \qquad (1-6)$$

Therefore the 2-norm condition number of matrix B is:

$$\kappa_2(A^T A+\gamma I)=\frac{\sigma_1^{\,2}+\gamma}{\sigma_n^{\,2}+\gamma} \tag{1-7}$$

Comparing to 2-norm condition number of A, we can verify the following inequality:

$$\kappa_2(A^T A+\gamma I)=\frac{\sigma_1^{\,2}+\gamma}{\sigma_n^{\,2}+\gamma}$$

$$\kappa_2(A)^2=\frac{\sigma_1^{\,2}}{\sigma_n^{\,2}}=\frac{\sigma_1^{\,2}\left(\sigma_n^{\,2}+\gamma\right)}{\sigma_n^{\,2}\left(\sigma_n^{\,2}+\gamma\right)}=\frac{\sigma_1^{\,2}\left(\sigma_n^{\,2}+\gamma\right)-\gamma\sigma_n^{\,2}+\gamma\sigma_n^{\,2}}{\sigma_n^{\,2}\left(\sigma_n^{\,2}+\gamma\right)}=\frac{\sigma_n^{\,2}\left(\sigma_1^{\,2}+\gamma\right)+\gamma\left(\sigma_1^{\,2}-\sigma_n^{\,2}\right)}{\sigma_n^{\,2}\left(\sigma_n^{\,2}+\gamma\right)}$$

$$=\frac{\gamma\left(\sigma_1^{\,2}-\sigma_n^{\,2}\right)}{\sigma_n^{\,2}\left(\sigma_n^{\,2}+\gamma\right)}+\frac{\sigma_1^{\,2}+\gamma}{\sigma_n^{\,2}+\gamma}=\frac{\gamma\left(\sigma_1^{\,2}-\sigma_n^{\,2}\right)}{\sigma_n^{\,2}\left(\sigma_n^{\,2}+\gamma\right)}+\kappa_2(A^T A+\gamma I), \tag{1-8}$$

Since $\sigma_1 \ge \sigma_n$, therefore:

$$\frac{\gamma\left(\sigma_1^{\,2}-\sigma_n^{\,2}\right)}{\sigma_n^{\,2}\left(\sigma_n^{\,2}+\gamma\right)}\ge 0,$$

$$\Rightarrow \kappa_2(A)^2 \ge \kappa_2(A^T A+\gamma I). \tag{1-9}$$

The above inequality means that when using parameter γ we are conditioning the normal equations. In other words, the condition number of the conditioned system is smaller than condition number of usual normal equations.

Part (b)

$$\left(A^T A+\gamma I\right)x_\gamma = A^T b \Rightarrow A^T\left(Ax_\gamma - b\right)+\left(\gamma^{\frac{1}{2}} I\right)^T\left(\gamma^{\frac{1}{2}} I\right)=0,$$

$$\Rightarrow \nabla_{x_\gamma}\frac{1}{2}\left\|Ax_\gamma - b\right\|_2^2+\nabla_{x_\gamma}\frac{1}{2}\left\|\gamma^{\frac{1}{2}}x\right\|_2^2=0$$

$$\Rightarrow x_\gamma=\arg\min_x \frac{1}{2}\left\|Ax_\gamma - b\right\|_2^2+\frac{1}{2}\left\|\gamma^{\frac{1}{2}}x\right\|_2^2$$

$$\Rightarrow x_\gamma=\arg\min_x \frac{1}{2}\left\|\begin{pmatrix}A\\ \gamma^{\frac{1}{2}} I\end{pmatrix}x-\begin{pmatrix}b\\0\end{pmatrix}\right\|_2^2. \tag{1-10}$$

Part (c)

We know that $x=\arg\min_{\hat{x}}\frac{1}{2}\left\|A\hat{x}-b\right\|_2^2$. So for every other $\hat{x}$ the value of objective function is larger. Therefore for $\hat{x}=x_\gamma$ the same holds. In other words:

$$\frac{1}{2}\|Ax-b\|_2^2 \le \frac{1}{2}\|Ax_\gamma - b\|_2^2 \Rightarrow \|Ax-b\|_2 \le \|Ax_\gamma - b\|_2 . \qquad (1-11)$$

Same for $x_\gamma = \arg\min_{\hat{x}} \frac{1}{2}\|A\hat{x}-b\|_2^2 + \frac{\gamma}{2}\|\hat{x}\|_2^2$ applies. It means that for every other $\hat{x}$ the value of objective function is larger. In other words:

$$\frac{1}{2}\|Ax_\gamma - b\|_2^2 + \frac{\gamma}{2}\|x_\gamma\|_2^2 \le \frac{1}{2}\|Ax-b\|_2^2 + \frac{\gamma}{2}\|x\|_2^2 \qquad (1-12)$$

Using equation (1 – 11) we can conclude that:

$$\frac{1}{2}\|Ax_\gamma - b\|_2^2 + \frac{\gamma}{2}\|x_\gamma\|_2^2 \le \frac{1}{2}\|Ax-b\|_2^2 + \frac{\gamma}{2}\|x\|_2^2 \le \frac{1}{2}\|Ax_\gamma - b\|_2^2 + \frac{\gamma}{2}\|x\|_2^2,$$
$$\Rightarrow \|x_\gamma\|_2 \le \|x\|_2 . \qquad (1-13)$$

Part (d)

$$\begin{cases} \left(A^TA+\gamma I\right)x_\gamma = A^Tb, \\ A^TAx = A^Tb \Rightarrow \left(A^TA+\gamma I\right)x = A^Tb + \gamma x \end{cases} \Rightarrow \left(A^TA+\gamma I\right)\left(x - x_\gamma\right) = \gamma x \qquad (1-14)$$

Since $\gamma > 0$ therefore $A^TA+\gamma I$ is nonsingular. Continuing from (1 – 14) we get:

$$\Rightarrow \|x - x_\gamma\|_2 = \left\|\left(A^TA+\gamma I\right)^{-1}(\gamma x)\right\|_2 \le \left\|\left(A^TA+\gamma I\right)^{-1}\right\|_2 \|\gamma x\|_2 \le \gamma \left\|\left(A^TA+\gamma I\right)^{-1}\right\|_2 \|x\|_2 \qquad (1-15)$$

Using singular values decomposition we can find $\left\|\left(A^TA+\gamma I\right)^{-1}\right\|_2$:

$$A^TA+\gamma I = V\Sigma^T\Sigma V^T + \gamma VV^T = V\begin{pmatrix} \sigma_1^{\,2} & & & & & \\ & \sigma_2^{\,2} & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & \sigma_n^{\,2} \end{pmatrix}_{n\times n} V^T + V\begin{pmatrix} \gamma & & & & & \\ & \gamma & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & \gamma \end{pmatrix}_{n\times n} V^T,$$

$$\Rightarrow \left(A^T A+\gamma I\right)^{-1} = V\begin{pmatrix} \frac{1}{\sigma_1^{\,2}+\gamma} & & & & & \\ & \frac{1}{\sigma_2^{\,2}+\gamma} & & & & \\ & & . & & & \\ & & & . & & \\ & & & & . & \\ & & & & & \frac{1}{\sigma_n^{\,2}+\gamma} \end{pmatrix}_{n\times n} V^T \Rightarrow \left\|\left(A^T A+\gamma I\right)^{-1}\right\|_2 = \frac{1}{\sigma_n^{\,2}+\gamma}. \qquad (1-16)$$

Continuing from (1 – 15) we get:

$$\Rightarrow \left\|x - x_\gamma\right\|_2 \le \gamma \left\|\left(A^T A+\gamma I\right)^{-1}\right\|_2 \left\|x\right\|_2 = \frac{\gamma}{\sigma_n^{\,2}+\gamma}\left\|x\right\|_2,$$

$$\Rightarrow \frac{\left\|x - x_\gamma\right\|_2}{\left\|x\right\|_2} \le \frac{\gamma}{\sigma_n^{\,2}+\gamma}. \qquad (1-17)$$

For the relative error to be bound below a given value ε we need to satisfy the following inequalities:

$$\frac{\left\|x - x_\gamma\right\|_2}{\left\|x\right\|_2} \le \frac{\gamma}{\sigma_n^{\,2}+\gamma} \le \varepsilon,$$

$$\Rightarrow \frac{\gamma}{\sigma_n^{\,2}+\gamma} - 1 = \frac{-\sigma_n^{\,2}}{\sigma_n^{\,2}+\gamma} \le \varepsilon - 1,$$

$$\text{Assume } \varepsilon < 1 \Rightarrow \frac{\sigma_n^{\,2}}{\sigma_n^{\,2}+\gamma} \ge 1-\varepsilon \Rightarrow \sigma_n^{\,2}+\gamma \le \frac{\sigma_n^{\,2}}{1-\varepsilon} \Rightarrow \gamma \le \frac{\sigma_n^{\,2}}{1-\varepsilon} - \sigma_n^{\,2} = \frac{\varepsilon}{1-\varepsilon}\sigma_n^{\,2},$$

$$\Rightarrow \gamma \le \frac{\varepsilon}{1-\varepsilon}\sigma_n^{\,2}. \qquad (1-18)$$

Part (e)

Using the following matrix and vector, and by reformatting the problem as a linear least squares problem, equation (1 – 10), I solved the problem using different values of γ:

$$A = \begin{pmatrix} 1 & 0 & 1 & 2 \\ 2 & 3 & 5 & 10 \\ 5 & 3 & -2 & 6 \\ 3 & 5 & 4 & 12 \\ -1 & 6 & 3 & 8 \end{pmatrix}, \quad b = \begin{pmatrix} 4 \\ -2 \\ 5 \\ -2 \\ 1 \end{pmatrix} \qquad (1-19)$$

$$x_\gamma = \arg\min_x \frac{1}{2} \left\| \begin{pmatrix} A \\ \gamma^{\frac{1}{2}} I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|_2^2 \qquad (1-20)$$

Following table shows the norm of residual and norm of the obtained solution using different values of γ:

gamma	L2_norm_of_x	L2_norm_of_residual
0	1.8182	5.025
1	0.90651	5.0286
0.001	0.94708	5.025
1e-06	0.94712	5.025
1e-12	0.94713	5.025

```
(Exercise #1 e) 2-norm of the residual using TSVD (norm(A*x_gamma-b)) = 5.025
(Exercise #1 e) 2-norm of the solution using TSVD (norm(x_gamma))= 0.947124
```

Using equations provided in this question we can rewrite two approaches using SVD decomposition:

1) Using $\left(A^T A + \gamma I\right) x_\gamma = A^T b$:

$$x_\gamma = \left(A^T A + \gamma I\right)^{-1} A^T b \rightarrow A_\lambda^{-1} = \left(A^T A + \gamma I\right)^{-1} A^T$$

$$\Rightarrow A_\lambda^{-1} = V \begin{pmatrix} \frac{\sigma_1}{\sigma_1^2 + \gamma} & & & & & \\ & \frac{\sigma_n}{\sigma_n^2 + \gamma} & & & & \\ & & . & & & \\ & & & . & & \\ & & & & . & \\ & & & & & \frac{\sigma_n}{\sigma_n^2 + \gamma} \end{pmatrix} U^T \qquad (1-21)$$

2) Using TSVD:

$$A_{TSVD}{}^{-1} = V \begin{pmatrix} \frac{1}{\sigma_1} & & & & & & \\ & \cdot & & & & & \\ & & \cdot & & & & \\ & & & \frac{1}{\sigma_r} & & & \\ & & & & 0 & & \\ & & & & & \cdot & \\ & & & & & & \cdot \\ & & & & & & & 0 \end{pmatrix} U^T \qquad (1-22)$$

Both regularization and TSVD tend to filter out the contribution to the solution corresponding to the smallest singular values! The difference is the filtering approach for TSVD is sharp. The singular values of $A_{TSVD}{}^{-1}$ go to zero instantly. On the other hand for usual regularization corresponding to a smooth filter that dampens the components corresponding to $\sigma_1 < \gamma$.

<u>When parameter r (cut off singular value) is chosen such that $\sigma_r = \gamma$, the sharp filter of the TSVD can in fact be seen as an approximation to the smooth filter of the regularization.</u>

Reference: Hansen PC. The truncatedsvd as a method for regularization. BIT Numerical Mathematics. 1987 Dec 1;27(4):534-53.

- Description for $\gamma = 0$:

Using $\gamma = 0$, we have came up with a solution with L2 norm of residual as same as TSVD and when using $\gamma > 0$. There is no surprise here because matrix A is rank deficient. So there are multiple solutions for $\arg\min_x \frac{1}{2}\|Ax - b\|_2^2$. We could expect solutions with large entries. As it can be seen in the table in previous page, L2 norm of solution $\|x\|_2$ associated with $\gamma = 0$ has the largest value.

- Description for $\gamma = 1$:

In this situation, the weird difference between L2 norm of solution of this solution and other γ's occur because the value γ is not small enough comparing to the singular values. There is two outcomes. 1) the L2 norm of the solution is smaller than other choices. 2) The fitting error (L2 norm of residuals) is bigger, because the parameter (which is big) has chosen a solution with small L2 norm in expense of larger misfit.

For other values of γ, TSVD and usual regularizing have nearly the same solution.

MATLAB code used in this question

```
clear; clc; close all

A = [1 0 1 2; 2 3 5 10; 5 3 -2 6; 3 5 4 12; -1 6 3 8];
b = [4; -2; 5; -2; 1];

k = 1;
d = [b; zeros(size(A, 2), 1)];
for gamma = [0, 10.^-[0 3 6 12]]

    B = [A; sqrt(gamma)*eye(size(A, 2))];
    x_gamma(:, k) = B\d;

    l2_res(k) = norm(A*x_gamma(:, k) - b);
    l2_sol(k) = norm(x_gamma(:, k));

    k = k+1;
end

% TSVD

[u, s, v] = svd(A);

[U, sv, tol] = svdtrunc(A, 1e-5);
z = U'*b;
y = zeros(4, 1);
y(1:3) = z./sv;
xn = v*y;
xn;
norm(A*xn - b);
norm(xn);


% Results
HW4_Q1_e = table;
HW4_Q1_e.gamma = [0, 10.^-[0 3 6 12]]';
HW4_Q1_e.L2_norm_of_x = l2_sol';
HW4_Q1_e.L2_norm_of_residual = l2_res';
HW4_Q1_e

fprintf('(Exercise #1 e) 2-norm of the residual using TSVD (norm(A*x_gamma-
b)) = %g \n',...
    norm(A*xn-b))
fprintf('(Exercise #1 e) 2-norm of the solution using TSVD (norm(x_gamma))=
%g \n',...
    norm(xn))
```

```
function [U sv tol] = svdtrunc(A, tol)

[U S] = svd(A, 'econ');
sv = diag(S);
ns = sum(sv > tol);

if ns < length(sv)
    tol = sv(ns+1);
    sv = sv(1:ns);
    U = U(:,1:ns);
else
    tol = 0;
end
```

Part (f)

Regarding to the reference previously, there isn't a big difference in terms of solution norm and residual norm when using these two different methods (figure (1 – 1)). But <u>the potential difference is in the computational cost of these methods</u>. Computing the singular values is a very expensive computation for large problems. And since parameter γ can be chosen such that two methods generate nearly the same solution, thus <u>for large ill-conditioned problems using regular regularization using γ is preferable</u>. In the figure below, k is the number of singular values after truncation and λ is the same as γ in this question.

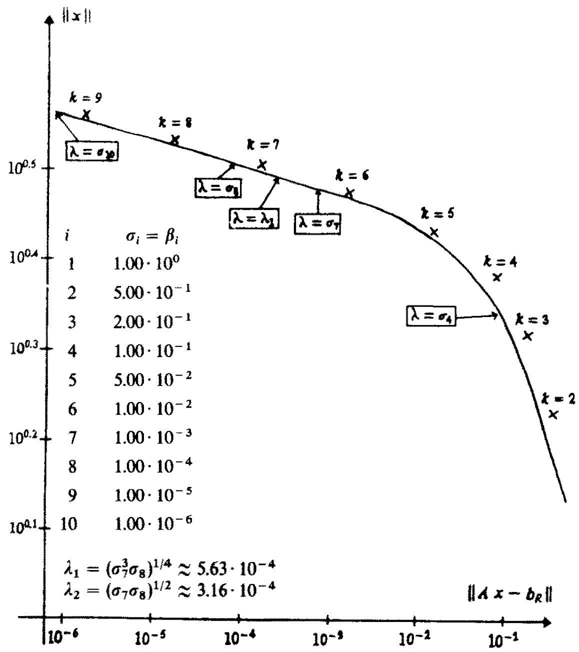

Figure (1 – 1) Hansen PC. The truncatedsvd as a method for regularization. BIT Numerical Mathematics. 1987 Dec 1;27(4):534-53.

Comparison of the TSVD and regularization methods for an ill-conditioned matrix with ill-determined numerical rank. There is no intuitive way of choosing λ or k.

Question #2
First Part of the question
First I am going to show that if we can prove the theorem for the case which B = I, then we can prove it for the original one.

$$(A^T A+\gamma R^T R)x_\gamma = A^T b \Leftrightarrow (A^T A+\gamma R^T R)\left(RR^{-1}\right)x_\gamma = A^T b,$$
$$\Leftrightarrow (A^T A+\gamma R^T R)\left(R^{-1}R\right)x_\gamma = A^T b \qquad (2-1)$$

Consider the following change of parameters (R is nonsingular therefore its inverse exists!):

$$\begin{cases} Rx = z \\ AR^{-1} = M \end{cases} \qquad (2-2)$$

Then we can rewrite the equation (2 – 1) as below:

$$\left(R^{-1}\right)^T (A^T A+\gamma R^T R)\left(R^{-1}R\right)x_\gamma = \left(R^{-1}\right)^T A^T b,$$
$$\Rightarrow (\left(R^{-1^T} A^T\right)\left(AR^{-1}\right)+\gamma R^{-1^T} R^T R R^{-1})\left(Rx_\gamma\right) = \left(R^{-1^T} A^T\right)b,$$
$$\Rightarrow (M^T M+\gamma I)z_\gamma = M^T b. \qquad (2-3)$$

So it means that if we can prove the theorem for equation (2 – 3), means with B = I and any matrix M, we can prove the original theorem. So hereon I am going to concentrate on solving equation (2 – 3) as it is indicated in equation (2 – 4).

$$(A^T A+\gamma I)x_\gamma = A^T b \qquad (2-4)$$

Consider the singular decomposition of A as below:

$$A_{m\times n} = U_{m\times m}\Sigma_{m\times n}V^T{}_{n\times n} \qquad (2-5)$$

Starting from equation (2 – 4) (it’s worth mentioning that $(A^T A+\gamma I)$ is nonsingular when $\gamma > 0$ so we can discuss about its inverse):

$$(A^T A+\gamma I)x_\gamma = A^T b \Rightarrow x_\gamma = (A^T A+\gamma I)^{-1}A^T b \qquad (2-6)$$

In terms of singular value decomposition we can rewrite matrix $(A^T A+\gamma I)^{-1}A^T$ as below:

$$(A^TA+\gamma I)^{-1}A^T=\left(V\Sigma^T\Sigma V^T+\gamma VV^T\right)^{-1}V\Sigma^TU^T$$

$$=\left(V\begin{pmatrix}\sigma_1{}^2 &&&&&& \\ & . &&&&& \\ && . &&&& \\ &&& \sigma_m{}^2 &&& \\ &&&& 0 && \\ &&&&& . & \\ &&&&&& . \\ &&&&&&& 0\end{pmatrix}_{n\times n}V^T+V\begin{pmatrix}\gamma &&& \\ & . && \\ && . & \\ &&& \gamma\end{pmatrix}_{n\times n}V^T\right)^{-1}V\Sigma^TU^T$$

$$=V\begin{pmatrix}\frac{1}{\sigma_1{}^2+\gamma} &&&&&& \\ & . &&&&& \\ && . &&&& \\ &&& \frac{1}{\sigma_m{}^2+\gamma} &&& \\ &&&& \frac{1}{\gamma} && \\ &&&&& . & \\ &&&&&& . \\ &&&&&&& \frac{1}{\gamma}\end{pmatrix}_{n\times n}\Sigma^T{}_{n\times m}U^T=V\begin{pmatrix}\frac{\sigma_1}{\sigma_1{}^2+\gamma} &&& \\ & . && \\ && . & \\ &&& \frac{\sigma_m}{\sigma_m{}^2+\gamma} \\ && 0 & \end{pmatrix}_{n\times m}U^T$$

$$=V\Sigma^T{}_{n\times m}\begin{pmatrix}\frac{1}{\sigma_1{}^2+\gamma} &&& \\ & . && \\ && . & \\ &&& \frac{1}{\sigma_m{}^2+\gamma}\end{pmatrix}_{m\times m}U^T \qquad (2-7)$$

Define:

$$\begin{pmatrix}\frac{1}{\sigma_1{}^2+\gamma} &&& \\ & . && \\ && . & \\ &&& \frac{1}{\sigma_m{}^2+\gamma}\end{pmatrix}_{m\times m}=\begin{pmatrix}\frac{1}{\sqrt{\sigma_1{}^2+\gamma}} &&& \\ & . && \\ && . & \\ &&& \frac{1}{\sqrt{\sigma_m{}^2+\gamma}}\end{pmatrix}_{m\times m}\begin{pmatrix}\frac{1}{\sqrt{\sigma_1{}^2+\gamma}} &&& \\ & . && \\ && . & \\ &&& \frac{1}{\sqrt{\sigma_m{}^2+\gamma}}\end{pmatrix}_{m\times m}$$

$$=\Gamma_{m\times m}\Gamma_{m\times m} \qquad (2-8)$$

Using equation (2 – 8) and continuing from equation (2 – 7):

$$\Rightarrow (A^TA+\gamma I)^{-1}A^T = V\Sigma^T{}_{n\times m}\begin{pmatrix} \frac{1}{\sigma_1{}^2+\gamma} & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \frac{1}{\sigma_m{}^2+\gamma}\end{pmatrix}_{m\times m} U^T = V\Sigma^T\Gamma\Gamma U^T = V\Sigma^T U^T U\Gamma\Gamma^T U^T =$$

$$V\Sigma^T U^T\left(U\Gamma^{-1}\Gamma^{-1}U^T\right)^{-1}, \qquad (2-9)$$

Define:

$$\Lambda_{m\times n} = \begin{pmatrix}\Gamma^{-1} & 0\end{pmatrix}_{m\times n} \Rightarrow \Lambda\Lambda^T = \Gamma^{-2}, \qquad (2-10)$$

Using equation (2 – 10) and continuing from equation (2 – 9) we would get:

$$\Rightarrow (A^TA+\gamma I)^{-1}A^T = V\Sigma^T U^T\left(U\Gamma^{-1}\Gamma^{-1}U^T\right)^{-1} = V\Sigma^T U^T\left(U\Lambda\Lambda^T U^T\right)^{-1} = V\Sigma^T U^T\left(U\Lambda V^T V\Lambda^T U^T\right)^{-1}$$

$$= V\Sigma^T U^T\left(\left(U\Lambda V^T\right)\left(U\Lambda V^T\right)^T\right)^{-1}, \qquad (2-11)$$

If we define $C = U\Lambda V^T$, then C is a matrix same size as A, with singular values λ_i slightly different from singular values of A σ_i as shown below:

$$\lambda_i = \sqrt{\sigma_i{}^2+\gamma} \qquad (2-12)$$

$$\lim_{\gamma\to 0}\lambda_i = \lim_{\gamma\to 0}\sqrt{\sigma_i{}^2+\gamma} = \sigma_i \Rightarrow \lim_{\gamma\to 0}\Lambda = \Sigma, \qquad (2-13)$$

$$\Rightarrow (A^TA+\gamma I)^{-1}A^T = V\Sigma^T U^T\left(CC^T\right)^{-1} = A^T\left(CC^T\right)^{-1}, \qquad (2-14)$$

Since all the equations written in equations (2 – 7) to (2 – 14) hold as $\gamma \to 0$ therefore we can conclude that:

$$\lim_{\gamma\to 0}(A^TA+\gamma I)^{-1}A^T=\lim_{\gamma\to 0}A^T\left(CC^T\right)^{-1}=\lim_{\gamma\to 0}A^T\left(\left(U\Lambda V^T\right)\left(U\Lambda V^T\right)^T\right)^{-1}$$
$$=A^T\left(\left(U\Sigma V^T\right)\left(U\Sigma V^T\right)^T\right)^{-1}=A^T\left(AA^T\right)^{-1},$$
$$\Rightarrow \lim_{\gamma\to 0}(A^TA+\gamma I)^{-1}A^T=A^T\left(AA^T\right)^{-1}. \qquad (2-15)$$

Based on equation (2 – 15) and equation (2 – 2) we can conclude:

$$\Rightarrow \lim_{\gamma\to 0}(M^TM+\gamma I)^{-1}M^T=M^T\left(MM^T\right)^{-1}$$
$$\Rightarrow \lim_{\gamma\to 0}z_\gamma=z^*=M^T\left(MM^T\right)^{-1}b,$$
$$\begin{cases}Rx=z\\ AR^{-1}=M\end{cases}\Rightarrow Rx^*=\left(AR^{-1}\right)^T\left(AR^{-1}\left(AR^{-1}\right)^T\right)^{-1}b,$$
$$\Rightarrow Rx^*=\left(R^{-1}\right)^TA^T\left(AR^{-1}\left(R^{-1}\right)^TA^T\right)^{-1}b,$$
$$\Rightarrow x^*=R^{-1}\left(R^{-1}\right)^TA^T\left(AR^{-1}\left(R^{-1}\right)^TA^T\right)^{-1}b,$$
$$\Rightarrow x^*=B^{-1}A^T\left(AB^{-1}A^T\right)^{-1}b. \qquad (2-16)$$

The derivation above proves that the limit solution of the problem exists and it’s unique.

Another way of looking at the uniqueness of the solution is to reformulate the original problem as below:

$$(A^TA+\gamma R^TR)x_\gamma=A^Tb\Rightarrow\left\|\begin{pmatrix}A\\ \gamma^{\frac{1}{2}}R\end{pmatrix}x-\begin{pmatrix}b\\0\end{pmatrix}\right\|_2^2=0. \qquad (2-17)$$

The reformulation shows that the optimization problem is in fact a least squares problem. Therefore it has a unique solution.

Second part of the question:
We should form the KKT conditions for this constrained optimization problem:

$$L=\frac{1}{2}x^TBx-\lambda^T(Ax-b)$$
$$\Rightarrow\frac{\partial L}{\partial x}=\frac{1}{2}(x^TB+x^TB^T)-\lambda^TA=0$$
$$\Rightarrow x^TB-\lambda^TA=0$$
$$\Rightarrow Bx-A^T\lambda=0\Rightarrow x=B^{-1}A^T\lambda, \qquad (2-18)$$

The other condition is:

$$Ax = b \Rightarrow AB^{-1}A^{T}\lambda = b \tag{2 – 19}$$

Since A is full row rank and B in nonsingular therefore matrix $AB^{-1}A^{T}$ is also nonsingular. Then we can conclude that:

$$\Rightarrow \lambda = \left(AB^{-1}A^{T}\right)^{-1} b \tag{2 – 20}$$

Using equation (2 – 17) we can compute the only and unique solution of the constrained optimization problem:

$$\Rightarrow x = B^{-1}A^{T}\left(AB^{-1}A^{T}\right)^{-1} b. \tag{2 – 21}$$

Therefore throughout this question we have shown that the regularized least norm least squares problem is equivalent to the constrained optimization problem in the question as $\gamma \to 0$.

Question #3

Part (a)

I have used the following parameters and transformations:

$$\begin{cases} z_1 = x_1 - x_2 \\ z_2 = x_3 - x_4 \\ z_3 = x_5 - x_6 \\ s_1 = -5 + 4z_1 - 2z_2 + 6z_3 \\ s_2 = 9 - 7z_1 - 3z_2 - 5z_3 \\ s_3 = 2 + z_1 \end{cases} \tag{3 – 1}$$

I have used $x_i,\ i = 1,\ldots,6$ for imposing a nonnegative constraint on z_i 's. I have also used $s_i,\ i = 1,2,3$ for turning inequality to equality constraints.

By inserting equation (3 – 1) into the original problem we can get the standard primal form:

$$\min_{\substack{x_i, i=1,..,6 \\ s_j, j=1,2,3}} f = x_1 - x_2 - 5x_3 + 5x_4 - 6x_5 + 6x_6$$

$$\text{s.t.} \quad \begin{cases} 4x_1 - 4x_1 - 2x_1 + 2x_1 + 6x_1 - 6x_1 - s_1 = 5 \\ 3x_1 - 3x_2 + 4x_3 - 4x_4 - 9x_5 + 9x_6 = 3 \\ 7x_1 - 7x_2 + 3x_3 - 3x_4 + 5x_5 - 5x_6 + s_2 = 9 \\ -x_1 + x_2 + s_3 = 2 \\ x_i \ge 0, \quad i = 1,\ldots,6. \\ s_j \ge 0, \quad j = 1,2,3. \end{cases} \tag{3 – 2}$$

Part (b)

I have used lpm.m MATLAB code from the course's webpage.

```
Output:
(Exercise #3 b) z_1 = 1.15217
(Exercise #3 b) z_2 = 0.130435
(Exercise #3 b) z_3 = 0.108696
(Exercise #3 b) function value f(z) = -0.152174
```

MATLAB code used in this question (I haven't showed the m-files in course webpage)

```
Clear; clc; close all
```

```
A = [4 -4 -2 2 6 -6 -1 0 0;
    3 -3 4 -4 -9 9 0 0 0;
    7 -7 3 -3 5 -5 0 1 0;
    -1 1 0 0 0 0 0 0 1];

b = [5; 3; 9; 2];
c = [1; -1; -5; 5; -6; 6; 0; 0; 0];

[x,gap,nbas] = lpm (A,b,c);

z = zeros(3, 1);
z(1) = x(1) - x(2);
z(2) = x(3) - x(4);
z(3) = x(5) - x(6);

f = @(z) z(1) - 5*z(2) - 6*z(3);

fprintf('(Exercise #3 b) z_1 = %g\n', z(1));
fprintf('(Exercise #3 b) z_2 = %g\n', z(2));
fprintf('(Exercise #3 b) z_3 = %g\n', z(3));
fprintf('(Exercise #3 b) function value f(z) = %g\n', f(z));
```