# Assignment 1

Name: Soham Kamble
CWID: A20517098

Problem 1.

1.) Given: $Q \in R^{N\times N}, P \in R^{M\times M}, B \in R^{M\times N}$

To Prove:

$$(Q^{-1} + B^T P^{-1} B)^{-1} B^T P^{-1} = QB^T (BQB^T + P)^{-1}$$

Proof:

Taking LHS:

$(Q^{-1} + B^T P^{-1} B)^{-1} B^T P^{-1}$

$= [(I + B^T P^{-1} BQ)Q^{-1}]^{-1} B^T P^{-1}$

$= (Q^{-1})^{-1}(I + B^T P^{-1} BQ)^{-1} B^T P^{-1}$ $[\because \boldsymbol{(AB)^{-1} = B^{-1}A^{-1}}]$

$= Q(I + B^T P^{-1} BQ)^{-1} B^T P^{-1}$ $[\because \boldsymbol{(A^{-1})^{-1} = A}]$

$= QB^T(I + P^{-1}BQB^T)^{-1} P^{-1}$

$[\because (I + AB)^{-1}A = A(I + BA)^{-1}]$

$= QB^T[P(I + P^{-1}BQB^T)]^{-1}$ $[\because \boldsymbol{B^{-1}A^{-1} = (AB)^{-1}}]$

$= QB^T[P(I + P^{-1}BQB^T)]^{-1}$ $[\because \boldsymbol{(A^{-1})^{-1} = A}]$

$\boldsymbol{= QB^T(BQB^T + P)^{-1}}$ $[\because \boldsymbol{AA^{-1} = I}]$

LHS = RHS

Hence proved.

2.) To Prove:

$$(A + BD^{-1}C)^{-1} = A^{-1} - A^{-1}B(D + CA^{-1}B)^{-1}CA^{-1}$$

Proof:

Taking LHS:

$(A + BD^{-1}C)^{-1}$

$= [(I + BD^{-1}CA^{-1})A]^{-1}$

$= A^{-1}(I + BD^{-1}CA^{-1})^{-1}$ $[\because \boldsymbol{(AB)^{-1} = B^{-1}A^{-1}}]$

$= A^{-1}[I - BD^{-1}(I + CA^{-1}BD^{-1})^{-1}CA^{-1}]$

$[\because \boldsymbol{(I + AB)^{-1} = I - A(I + BA)^{-1}B}]$ (166 – The Matrix Cookbook)

$= A^{-1}\{I - B[(I + CA^{-1}BD^{-1})D]^{-1}CA^{-1}\}$

$[\because \boldsymbol{B^{-1}A^{-1} = (AB)^{-1}}$ & $\boldsymbol{(A^{-1})^{-1} = A}]$

$= A^{-1}[I - B(D + CA^{-1}B)^{-1}CA^{-1}]$ $[\because \boldsymbol{AA^{-1} = A^{-1}A = I}]$

$\boldsymbol{= A^{-1} - A^{-1}B(D + CA^{-1}B)^{-1}CA^{-1}}$

LHS = RHS

Hence proved.

Problem 2

1.) Given:

$x = [x_1;\ x_2;\ x_3]$

$$y = [y_1;\ y_2]$$
$$y_1 = x_1^2 - x_2$$
$$y_2 = x_3^2 + 3x_2$$

We know that:

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} \\ \frac{\partial y_1}{\partial x_3} & \frac{\partial y_2}{\partial x_3} \end{bmatrix}$$

Substituting $y_1$ & $y_2$:

$$= \begin{bmatrix} \frac{\partial (x_1^2 - x_2)}{\partial x_1} & \frac{\partial (x_3^2 + 3x_2)}{\partial x_1} \\ \frac{\partial (x_1^2 - x_2)}{\partial x_2} & \frac{\partial (x_3^2 + 3x_2)}{\partial x_2} \\ \frac{\partial (x_1^2 - x_2)}{\partial x_3} & \frac{\partial (x_3^2 + 3x_2)}{\partial x_3} \end{bmatrix}$$

$$\boldsymbol{\frac{\partial y}{\partial x} = \begin{bmatrix} 2x_1 & 0 \\ -1 & 3 \\ 0 & 2x_3 \end{bmatrix}}$$

2.) Given:

$$x = r \sin\theta \cos\emptyset$$
$$y = r \sin\theta \sin\emptyset$$
$$z = r \cos\theta$$
$$r > 0; 0 < \theta < \pi; 0 < \ \emptyset < 2\pi$$
$$x = [x; y; z]$$
$$y = [r;\ \theta;\ \emptyset]$$

We know that:

$$\frac{\partial x}{\partial y} = \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} & \frac{\partial z}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} & \frac{\partial z}{\partial \theta} \\ \frac{\partial x}{\partial \emptyset} & \frac{\partial y}{\partial \emptyset} & \frac{\partial z}{\partial \emptyset} \end{bmatrix}$$

Substituting $x, y, z$:

$$= \begin{bmatrix} \frac{\partial (r\sin\theta\cos\emptyset)}{\partial r} & \frac{\partial (r\sin\theta\sin\emptyset)}{\partial r} & \frac{\partial (r\cos\theta)}{\partial r} \\ \frac{\partial (r\sin\theta\cos\emptyset)}{\partial \theta} & \frac{\partial (r\sin\theta\sin\emptyset)}{\partial \theta} & \frac{\partial (r\cos\theta)}{\partial \theta} \\ \frac{\partial (r\sin\theta\cos\emptyset)}{\partial \emptyset} & \frac{\partial (r\sin\theta\sin\emptyset)}{\partial \emptyset} & \frac{\partial (r\cos\theta)}{\partial \emptyset} \end{bmatrix}$$

$$\boldsymbol{\frac{\partial x}{\partial y} = \begin{bmatrix} \sin\theta\cos\emptyset & \sin\theta\sin\emptyset & \cos\theta \\ r\cos\theta\cos\emptyset & r\cos\theta\sin\emptyset & -r\sin\theta \\ -r\sin\theta\sin\emptyset & r\sin\theta\cos\emptyset & 0 \end{bmatrix}}$$

Problem 3.

1.) Given:

$$L(w) = \frac{1}{2}\sum_{i=1}^{n}(x_i^T w - y_i)^2$$

We know that:

$$X = [x_1, x_2, \dots, x_n]; \qquad X^T = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix}; \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\therefore X^T w - y = \begin{bmatrix} x_1^T w - y_1 \\ x_2^T w - y_2 \\ \vdots \\ x_n^T w - y_n \end{bmatrix}$$

$$\because \|x\|_2^2 = \sum_i |x_i|^2 = \sum_i x_i^2$$

$$\because \|X^T w - y\|_2^2 = \sum_i^n (x_i^T w - y_i)^2$$

The above matrix form can be rewritten as:

$$\|X^T w - y\|_2^2 = (X^T w - y)^T (X^T w - y)$$
$$= w^T X X^T w - 2w^T X y + y^T y$$

We know that:

$$\nabla_w^2 L(w) = H(w)$$
$$H(w) = \frac{\partial}{\partial w}\left(\frac{\partial}{\partial w} L(w)\right)$$
$$= \frac{\partial}{\partial w}\left(\frac{\partial}{\partial w}\frac{1}{2}(w^T X X^T w - 2w^T X y + y^T y)\right)$$
$$= \frac{\partial}{\partial w}\left(\frac{1}{2}(2XX^T w - 2Xy)\right) \qquad \left[\because \frac{\partial}{\partial x}(\boldsymbol{xBx^T}) = \boldsymbol{2Bx}\right]$$
$$\boldsymbol{H(w) = XX^T}$$

2.) To prove:

$$w^* = (XX^T)^{-1} X y$$

The First Iteration of Newton's Method is defined as:

$$x_{k+1} = x_k - \left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k)$$

We know that:

$$H(w) = \nabla^2 f(x_k)$$

The term $-\left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k)$ in the Newton's Method provides the direction of descent.

Using Newton's method, we can find:

$$w^* = w - \left(\nabla^2 L(w)\right)^{-1} \nabla L(w)$$

We know that:

$$\nabla^2 L(w) = XX^T$$
$$\nabla L(w) = XX^T w - Xy$$

Therefore:

$$w^* = w - (XX^T)^{-1}(XX^T w - Xy)$$

$w^* = w - [(XX^T)^{-1}XX^Tw - (XX^T)^{-1}Xy]$
$w^* = w - (w - (XX^T)^{-1}Xy)$
$\boldsymbol{w^* = (XX^T)^{-1}Xy}$

This shows that Newton's method finds the global minimum in a single iteration.

Problem 4.

1.) We know that:

$$\min_w L(w) = \sum_{n=1}^{N}(f(x_n; w) - t_n)^2 \qquad \text{s.t., } \|w\|_p^p \leq \gamma$$

Above equation can be written in matrix form as:

$$L(w) = \frac{1}{2}\|Xw - t_n\|_2^2 \qquad \text{s.t., } \|w\|_p^p \leq \gamma$$

Applying Lagrangian Multiplier to the above equation with $l_2$ norm:

$$L(w, \lambda) = \|Xw - t_n\|_2^2 + \lambda(\|w\|_2^2 - \gamma)$$

The constrained stationary condition tells us that:

$$\frac{\partial}{\partial w}\big(L(w, \lambda)\big) = 0$$

Therefore,

$$\frac{\partial}{\partial w}\big(L(w, \lambda)\big) = 2X^T(Xw - t_n) + 2\lambda w = 0$$
$$2X^T(Xw - t_n) + 2\lambda w = 0$$
$$X^TXw - X^Tt_n + \lambda w = 0$$
$$X^TXw + \lambda w = X^Tt_n$$
$$w(X^TX + \lambda) = X^Tt_n$$
$$w = (X^TX + \lambda)^{-1} X^Tt_n$$

The solution to obtain the minimizing $L(w)$ s.t., $\|w\|_p^p \leq \gamma$ is by optimizing the Lagrangian function.

We know $w$, therefore:

$$\|w\|_2^2 = [(X^TX + \lambda)^{-1} X^Tt_n]^2 \leq \gamma$$
$$t_n^TX(X^TX + \lambda)^{-2}X^Tt_n \leq \gamma$$

We can apply $SVD$ decomposition on $X$ to solve further,

$$X = USV^T,$$
$$t_n^TUSV^T(X^TX + \lambda)^{-2}(USV^T)^Tt_n \leq \gamma$$
$$t_n^TUSV^T(VS^2V^T + \lambda VV^T)^{-2}VSU^Tt_n \leq \gamma \qquad [\because \boldsymbol{A^TA = VS^2V^T}]$$
$$t_n^TUSV^TV(S^2 + \lambda)^{-2}V^TVSU^Tt_n \leq \gamma$$
$$t_n^TUS(S^2 + \lambda)^{-2}SU^Tt_n \leq \gamma$$

Assuming $z = U^Ty$ and converting the above equation into summation form:

$$\sum_{i=1}^{N} z_i^2 \frac{s_i^2}{(s_i^2 + \lambda)^2} \leq \gamma$$

2.) From the above equation we can say that:

i.) If $\sum_{i=1}^{N}(\frac{z_i}{s_i})^2 \leq \gamma$

No non-zero $\lambda$ would satisfy $\sum_{i=1}^{N} z_i^2 \frac{s_i^2}{(s_i^2+\lambda)^2} \leq \gamma$ in equality.

ii.) If $\sum_{i=1}^{N}(\frac{z_i}{s_i})^2 > \gamma$

$\lambda$ must be greater than 0.

From the equations above we can say that $\lambda$ is a monotonically decreasing function of $\gamma$. Therefore, choosing the right value of $\gamma$ is always equivalent to choosing the right $\lambda$ and, Solving:

$$\min_w L(w) = \min_w \|Xw - t_n\|_2^2 + \lambda\|w\|_2^2$$
$$\sum_{n=1}^{N}(f(x_n; w) - t_n)^2 + \lambda\|w\|_2^2$$

Without constraints.
Hence proved.

Problem 5

1.) We know that the gradient iterative process is:

$$x^{(k+1)} = x^{(k)} - \alpha\nabla f\left(x^{(k)}\right) = y$$

Substituting $x^{(k+1)} = x^{(k)} - \alpha\nabla f\left(x^{(k)}\right) = y$ in

$$f(x) + \nabla f(x)^T(y - x) + \frac{1}{2}L\|y - x\|_2^2 \text{ we get:}$$

$$= f(x^{k+1}) \leq f(x^k) + \nabla f(x^k)^T(x^{k+1} - x^k) + \frac{1}{2}L\|x^{k+1} - x^k\|_2^2$$
$$= f(x^k) + \nabla f(x^k)^T\left(x^{(k)} - \alpha\nabla f\left(x^{(k)}\right) - x^{(k)}\right) + \frac{1}{2}L\left\|x^{(k)} - \alpha\nabla f\left(x^{(k)}\right) - x^{(k)}\right\|_2^2$$
$$= f(x^k) - \nabla f(x^k)^T\left(\alpha\nabla f\left(x^{(k)}\right)\right) + \frac{1}{2}L\left\|\alpha\nabla f\left(x^{(k)}\right)\right\|_2^2$$
$$= f(x^k) - \alpha\|\nabla f(x^k)\|_2^2 + \frac{1}{2}L\alpha^2\left\|\nabla f\left(x^{(k)}\right)\right\|_2^2$$
$$= f(x^{k+1}) \leq f(x^k) - (1 - \frac{1}{2}L\alpha)\alpha\left\|\nabla f\left(x^{(k)}\right)\right\|_2^2 \tag{1}$$

We know that:

$$0 < \alpha \leq \frac{1}{L}$$

Therefore,

$$-\left(1 - \frac{1}{2}L\alpha\right) = (\frac{1}{2}L\alpha - 1)$$
$$\leq \frac{1}{2}(L)\left(\frac{1}{L}\right) - 1$$
$$\leq -\frac{1}{2}$$

On substituting in (1),

$$\boldsymbol{f\left(x^{k+1}\right) \leq f\left(x^k\right) - \frac{1}{2}\alpha\left\|\nabla f\left(x^k\right)\right\|_2^2} \tag{2}$$

From the above we can see that $\frac{1}{2}\alpha\|\nabla f(x^k)\|_2^2$ is always positive ($\because A^2 \geq 0$) implying that $f(x)$ is always a strictly decreasing function.

2.) Since $f$ is convex,
$f(y) \geq f(x)^T(y - x)$
For the optimal value $x^* = y$ & $x = x^k$

$$f(x^*) \geq f(x^k) + \nabla f(x^k)^T(x^* - x^k)$$
$$f(x^k) \leq f(x^*) + \nabla f(x^k)^T(x^k - x^*) \tag{3}$$

Now substituting $f(x^k)$ in (2) we get:

$$f(x^{k+1}) \leq f(x^k) - \frac{1}{2}\alpha\|\nabla f(x^k)\|_2^2$$
$$f(x^{k+1}) \leq [f(x^*) + \nabla f(x^k)^T(x^k - x^*)] - \frac{1}{2}\alpha\|\nabla f(x^k)\|_2^2$$
$$f(x^{k+1}) - f(x^*) \leq \nabla f(x^k)^T(x^k - x^*) - \frac{1}{2}\alpha\|\nabla f(x^k)\|_2^2$$
$$f(x^{k+1}) - f(x^*) \leq \frac{1}{2\alpha}[2\alpha\nabla f(x^k)^T(x^k - x^*) - \alpha^2\|\nabla f(x^k)\|_2^2]$$
$$(x^{k+1}) - f(x^*) \leq \frac{1}{2\alpha}[2\alpha\nabla f(x^k)^T(x^k - x^*) - \alpha^2\|\nabla f(x^k)\|_2^2 + \|x^k - x^*\|_2^2 - \|x^k - x^*\|_2^2]$$

Using $-(A-B)^2 = -(A^2 + B^2 - 2AB) = -A^2 - B^2 + 2AB$

Where $A = \alpha\nabla f(x^k)^T$ and $B = x^k - x^*$

$$f(x^{k+1}) - f(x^*) \leq \frac{1}{2\alpha}[-\|x^k - x^* - \alpha\nabla f(x^k)\|_2^2 + \|x^k - x^*\|_2^2]$$
$$f(x^{k+1}) - f(x^*) \leq \frac{1}{2\alpha}[\|x^k - x^*\|_2^2 - \|x^k - \alpha\nabla f(x^k) - x^*\|_2^2]$$

We know that the iterative process of gradient descent equation is:

$$x^{k+1} = x^k - \alpha\nabla f(x^k)$$
$$x^{k+1} - x^k = -\alpha\nabla f(x^k)$$

Therefore

$$f(x^{k+1}) - f(x^*) \leq \frac{1}{2\alpha}[\|x^k - x^*\|_2^2 - \|x^k + x^{k+1} - x^k - x^*\|_2^2]$$

Which finally derives:

$$\boldsymbol{f(x^{k+1}) - f(x^*) \leq \frac{1}{2\alpha}[\|x^k - x^*\|_2^2 - \|x^{k+1} - x^*\|_2^2]}$$

The above equation is correct for all $x^{k+1}$ on every iteration of the gradient descent.

3.) By applying summation on both sides, we get:

$$\sum_{k=1}^{K}(f(x^{k+1}) - f(x^*)) \leq \sum_{k+1}^{K}\frac{1}{2\alpha}(\|x^k - x^*\|_2^2 - \|x^{k+1} - x^*\|_2^2) \quad (4)$$

Above RHS is in the form of telescoping summation.

We know that a Summation S is known as telescoping if:

$$S = \sum_{i=1}^{n-1}(a_i - a_{i+1})$$
$$= (a_1 - a_2) + (a_2 - a_3) + \cdots + (a_{n-1} - a_n)$$
$$= (a_1 - a_n)$$

The subsequent terms in the above summation cancels out therefore applying the above theorem to (4) We get:

$$\sum_{k=1}^{K} f(x^k) - f(x^*) \leq \frac{1}{2\alpha}(\|x^0 - x^*\|_2^2 - \|x^K - x^*\|_2^2)$$
$$\boldsymbol{\sum_{k=1}^{K} f(x^k) - f(x^*) \leq \frac{1}{2\alpha}(\|x^0 - x^*\|_2^2)}$$

Since we know that $f$ is a strictly decreasing function we can derive the following:

$$f(x^k) - f(x^*) \leq \frac{1}{K}\sum_{k=1}^{K} f(x^k) - f(x^*)$$
$$\leq \frac{\|x^0 - x^*\|_2^2}{2\alpha K}$$

Hence proved.