Problem Set #6

Shirley Yuan, in collaboration with Zeshun Zong, Winston, and Fiona.

Exercise 9.1

An unconstrained linear objective function is of the form $f(\mathbf{x}) = \mathbf{c}^T\mathbf{x}$, where $\mathbf{c}$ is vector coefficient. If $\mathbf{c} = \mathbf{0}$, then $f(\mathbf{x}) = 0$, which is constant. If $\mathbf{c} \neq \mathbf{0}$, by contradiction assume $\mathbf{x}^* = argminf(\mathbf{x})$. i.e., $\forall \mathbf{x} \in \mathbb{R}, \mathbf{c}^T\mathbf{x}^* \leq \mathbf{c}^T\mathbf{x}$. It follows that $\mathbf{c}^T\mathbf{x}^* < 0$, since if $\mathbf{c}^T\mathbf{x}^* \geq 0$, then $\mathbf{c}^T(-\mathbf{x})^* = -\mathbf{c}^T\mathbf{x}^* \leq 0 \leq \mathbf{c}^T\mathbf{x}^*$.
Now let $y = 2\mathbf{x}^*$, then $\mathbf{c}^T\mathbf{y}^* = 2\mathbf{c}^T\mathbf{x}^* < \mathbf{c}^T\mathbf{x}^* < 0$

Exercise 9.2

Since $\|Ax - b\|_2 \geq 0$, to minimize $\|Ax - b\|_2$ is equivalent of minimizing $\|Ax - b\|_2^2$.
Now $\|Ax - b\|_2^2 = \langle Ax - b, Ax - b\rangle = (Ax - b) * T(Ax - b) = x^TA^TAx - x^TA^Tb - b^TAx + b^T = x^TA^TAx - 2x^TA^Tb + b^T$
The last term is a constant so in the minimization problem we can drop it.
Let $f(x) = x^TA^TAx - 2x^TA^Tb$, then $Df(x) = 2x^T(A^TA)^T - 2b^TA$, and $D^2f(x) = 2A^TA$
If A is non-singular, then $D^2f(x) > 0$.
By FOC, let $Df(x) = 0$, we have $x^T(A^TA)^T = b^TA \Leftrightarrow A^TAx = A^Tb$

Exercise 9.3

Gradient decent: slow but cheap
Newton: fast but expensive
conjugate gradient: a combination of both

Exercise 9.4

"$\Leftarrow$":
Suppose $Df(x_0)^T = Qx_0 - b = \mathbf{v}$ is an eigenvector of Q, then
$alpha_0 = \frac{Df(x_0)Df(x_0)^T}{Df(x_0)QDf(x_0)^T} = \frac{V^TV}{V^TQV} = \frac{V^TV}{V^T\lambda V} = \frac{1}{\lambda}$
Now by our algorithm, $x_1 = x_0 - \alpha_0 Df(x_0)^T = x_0 - \frac{1}{\lambda}\mathbf{V}$
Observe that $Q\mathbf{x_1} = Q(x_0 - \frac{1}{\lambda}\mathbf{V}) = Qx_0 - \mathbf{V} = Qx_0 - (Qx_0 - b) = \mathbf{b}$
Hence $\mathbf{x_1} = A^{-1}b$ is a minimizer and therefore the algorithm converges in one step.
"$\Rightarrow$":
If $\mathbf{x_1} = Q^{-1}\mathbf{b}$, then $Q\mathbf{x_1} = \mathbf{b}$
Since $x_1 = x_0 - \alpha_0 Df(x_1)^T = x_0 - \alpha_0(Qx_0 - b)$
We have $Q[x_0 - \alpha(Qx_0 - b)] = \mathbf{b} \Rightarrow Qx_0 - \alpha Q^2x_0 + \alpha Qb - b = 0$
Observe that $(I - \alpha Q)(Qx_0 - b) = Qx_0 - b - \alpha Q^2x_0 + \alpha Qb = 0$
Let $Qx_0 - \mathbf{b} = \mathbf{v}$, we have $(I - \alpha Q)\mathbf{v} = 0$, so $\mathbf{v} = \alpha Q\mathbf{v} \Rightarrow Q\mathbf{v} = \frac{1}{\alpha}\mathbf{v}$
Hence $Qx_0 isaneigenvectorofQ$.

Exercise 9.5

Assume $Df(x_k) \neq \mathbf{0}$, so we haven't reached the minimum yet.
Since $\mathbf{x}_{k+1} - \mathbf{x}_k = -\alpha_k Df(x_k)^T$, and $\mathbf{x}_{k+2} - \mathbf{x}_{k+1} = -\alpha_{k+1}Df(x_{k+1})^T$,
we want to show $(\mathbf{x}_{k+1} - \mathbf{x}_k)^T(\mathbf{x}_{k+2} - \mathbf{x}_{k+1}) = \alpha_k\alpha_{k+1}Df(x_k)^TDf(x_{k+1})^T = 0$
i.e., $Df(x_k)^TDf(x_{k+1})^T = 0$.
Now, since $\alpha_k = argminf(x_k - \alpha Df(x_k)^T)$, and $f \in \mathbb{C}'$, by First Order Necessary Condition, we have $-Df(x_k)Df(x_{k+1})^T = 0 \Rightarrow Df(x_k)Df(x_{k+1})^T = 0$

Exercise 9.10
Observe that $Df(x) = x^T Q^T - b^T$, and $D^2 f(x) = Q > 0$,
By Newton's method, $x_1 = x_0 - Q^{-1}(Qx_0 - b) = Q^{-1}b$
Since $D^2 f(x_1) = Q > 0$ and $Df(x_1)^T = Qx_1 - b = QQ^{-1}b - b = \mathbf{0}$
$\Rightarrow$ we know that x_1 is the unique minimizer. **Exercise 9.12**
Suppose (λ_i, v_i) is an eigen-pari of A.
Observe that $Bv_i = (A\mu I)v_i = Av_i + \mu I v_i = \lambda_i v_i + \mu v_i = (\lambda_i + \mu)v_i$.
So $(\lambda_i + \mu, v_i)$ is an eigenpair of B.
Exercise 9.15
Observe that $BC(C^{-1} + DA^{-A}B) = B + BCDA^{-1}B = (A + BCD)A^{-1}B$
So, $(A + BCD)^{-1}BC = A^{-1}B(C^{-1} + DA^{-1}B)^{-1}$
Hence,

$$
\begin{aligned}
A^{-1} &= (A + BCD)^{-1}(A + BCD)A^{-1} \\
&= (A + BCD)^{-1}(1 + BCDA^{-1}) \\
&= (A + BCD)^{-1} + [(A + BCD)^{-1}BC]DA^{-1} \\
&= (A + BCD)^{-1} + A^{-1}B(C^{-1}DA^{-1}B)^{-1}DA^{-1}
\end{aligned}
$$

$\Rightarrow (A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1}DA^{-1}B)^{-1}DA^{-1}$
Exercise 9.18
Observe that

$$
\begin{aligned}
\phi_n(\alpha) &= f(\mathbf{x_k} + \alpha_k \mathbf{d_k}) \\
&= \frac{1}{2}(\mathbf{x_k} + \alpha_k \mathbf{d_k})^T Q(\mathbf{x_k} + \alpha_k \mathbf{d_k}) - \mathbf{b}^T(\mathbf{x_k} + \alpha_k \mathbf{d_k}) + c \\
&= \frac{1}{2}[\mathbf{x_k}^T Q \mathbf{x_k} + \alpha_k^2 \mathbf{d_k}^T Q \mathbf{d_k} + \alpha_k \mathbf{d_k}^T Q \mathbf{x_k} + \alpha_k \mathbf{x_k}^T Q \mathbf{d_k}] - \mathbf{b}^T \mathbf{x_k} - \alpha_k \mathbf{b}^T \mathbf{d_k} \\
\phi_k^{'}(\alpha) &= \alpha_k(\mathbf{d_k}^T Q \mathbf{d_k}) + (\frac{1}{2}\mathbf{d_k}^T Q \mathbf{x_k} + \frac{1}{2}\mathbf{x_k}^T Q \mathbf{d_k}) - \mathbf{b}^T \mathbf{d_k} \\
&= \alpha_k(\mathbf{d_k}^T Q \mathbf{d_k}) + \frac{1}{2}(Q\mathbf{x_k})^T \mathbf{d_k} + \frac{1}{2}(Q\mathbf{x_k})^T \mathbf{d_k} - \mathbf{b}^T \mathbf{d_k} \\
&= \alpha_k(\mathbf{d_k}^T Q \mathbf{d_k}) + (Q\mathbf{x_k})^T \mathbf{d_k} - \mathbf{b}^T \mathbf{d_k}
\end{aligned}
$$

Setting derivative to 0, we have

$$
\alpha_k = \frac{\mathbf{b}^T \mathbf{d_k} - (Q\mathbf{x_k})^T \mathbf{d_k}}{\mathbf{d_k}^T Q \mathbf{d_k}} = \frac{\mathbf{r_k}^T \mathbf{d_k}}{\mathbf{d_k}^T Q \mathbf{d_k}}
$$

where $\mathbf{r_k} = \mathbf{b} - Q\mathbf{x_k}$