MATH 6490. Nonlinear Optimization in Machine Learning.
Assignment 3.

There are 3 problems, each problem is worth 5 points, total is 15 points.

1. Let $f(x) = \dfrac{1}{2}x^T Qx - b^T x + c$ as we introduced in Theorem 4.1 in the Lecture Notes. Here $Q \succ 0$ is a positively definite symmetric matrix such that its eigenvalues satisfy $0 < m = \lambda_n \le \lambda_{n-1} \le ... \le \lambda_1 = L < \infty$, where m and L are the strong–convexity constant for f and the Lipschitz constant for ∇f. Show that $\nabla f = Qx - b$. Then show that $\min\limits_x f(x) = f(x^*)$ where x^* is the unique solution of the linear equation $Qx^* - b = 0$.

2. Following problem 1. Let the dimension $n = 2$. Consider the matrix

$$T = \begin{bmatrix} (1+\beta)(I-\alpha Q) & -\beta(I-\alpha Q) \\ I & 0 \end{bmatrix}$$

that we introduced in Theorem 4.1 in the Lecture Notes. According to standard linear algebra, there exist an orthogonal matrix U of size 2×2 such that $U^T QU = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \equiv \Lambda$. Define the permutation matrix

$$\Pi = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} .$$

Show that

$$\Pi \begin{bmatrix} U & 0 \\ 0 & U \end{bmatrix}^T T \begin{bmatrix} U & 0 \\ 0 & U \end{bmatrix} \Pi^T = \mathrm{diag}(T_1, T_2) ,$$

where each block $T_i = \begin{bmatrix} (1+\beta)(1-\alpha\lambda_i) & -\beta(1-\alpha\lambda_i) \\ 1 & 0 \end{bmatrix}$ for $i = 1, 2$.

(Hint: One first easily sees that

$$\begin{bmatrix} U & 0 \\ 0 & U \end{bmatrix}^T T \begin{bmatrix} U & 0 \\ 0 & U \end{bmatrix} = \begin{bmatrix} (1+\beta)(I-\alpha\Lambda) & -\beta(I-\alpha\Lambda) \\ I & 0 \end{bmatrix} \equiv R .$$

The fact that $\Pi R \Pi^T = \mathrm{diag}(T_1, ..., T_n)$ follows from standard row and column transformations. Indeed we have

$$
\begin{aligned}
&\begin{bmatrix}
(1+\beta)(1-\alpha\lambda_1) & 0 & -\beta(1-\alpha\lambda_1) & 0\\
0 & (1+\beta)(1-\alpha\lambda_2) & 0 & -\beta(1-\alpha\lambda_2)\\
1 & 0 & 0 & 0\\
0 & 1 & 0 & 0
\end{bmatrix}\\
\xrightarrow{\text{permute column 2 and 3}}
&\begin{bmatrix}
(1+\beta)(1-\alpha\lambda_1) & -\beta(1-\alpha\lambda_1) & 0 & 0\\
0 & 0 & (1+\beta)(1-\alpha\lambda_2) & -\beta(1-\alpha\lambda_2)\\
1 & 0 & 0 & 0\\
0 & 0 & 1 & 0
\end{bmatrix}\\
\xrightarrow{\text{permute row 2 and 3}}
&\begin{bmatrix}
(1+\beta)(1-\alpha\lambda_1) & -\beta(1-\alpha\lambda_1) & 0 & 0\\
1 & 0 & 0 & 0\\
0 & 0 & (1+\beta)(1-\alpha\lambda_2) & -\beta(1-\alpha\lambda_2)\\
0 & 0 & 1 & 0
\end{bmatrix},
\end{aligned}
$$

as desired.)

3. Show that if the sequence $\rho_k \geq 0$, $\rho_0 = 0$ is defined recursively by

$$\rho_k^2 = \frac{(1-\rho_k^2)^2}{(1-\rho_{k-1}^2)^2} \ ,$$

then

$$1 - \rho_k^2 \leq \frac{2}{k+2} \ .$$

This yields the convergence rate

$$f(x^k) - f^* \leq \frac{C}{(k+1)^2}\|x^k - x^*\|^2$$

as we demonstrated in the Nesterov's optimal algorithm Algorithm 4.3 in the Lecture Notes.

(Hint: Do this by induction. For $k = 0$ we have $1 - \rho_0^2 = 1 - 0 = 1 \leq \frac{2}{0+2}$. Suppose for some k we have $1 - \rho_k^2 \leq \frac{2}{k+2}$. We aim to show that $1 - \rho_{k+1}^2 \leq \frac{2}{k+3}$. Suppose this is not the case, so that $1 - \rho_{k+1}^2 > \frac{2}{k+3}$. Then $\rho_{k+1}^2 < 1 - \frac{2}{k+3} = \frac{k+1}{k+3}$, i.e., $\frac{1}{\rho_{k+1}^2} > \frac{k+3}{k+1}$. Then using the recursive relation we see that

$$\frac{4}{(k+2)^2} \geq (1-\rho_k^2)^2 = \frac{(1-\rho_{k+1}^2)^2}{\rho_{k+1}^2} > \left(\frac{2}{k+3}\right)^2 \frac{k+3}{k+1} = \frac{4}{(k+1)(k+3)} \ .$$

This leads to $(k+1)(k+3) > (k+2)^2$, which is same as $3 > 4$, that is a contradiction.)