1a. (20%) Consider the problem: \( \min_y h(y) = f(Sy) \).

Newton’s method generates a sequence according to

\[
y^{k+1} = y^k - \alpha_k \left( \nabla^2 h(y^k) \right)^{-1} \nabla h(y^k)
\]

where

\[
\nabla h(y) = S^T \nabla f(Sy)
\]

and

\[
\nabla^2 h(y) = S^T \nabla^2 f(Sy) S
\]

As such, Newton’s method in the \( y \)-space can be rewritten as

\[
Sy^{k+1} = Sy^k - \alpha_k S \left( \nabla^2 h(y^k) \right)^{-1} \nabla h(y^k)
\]

\[
= Sy^k - \alpha_k S \left( S^T \nabla^2 f(Sy^k) S \right)^{-1} S^T \nabla f(Sy^k)
\]

\[
= Sy^k - \alpha_k SS^{-1} \left( \nabla^2 f(Sy^k) \right)^{-1} \left( S^T \right)^{-1} S^T \nabla f(Sy^k)
\]

\[
= Sy^k - \alpha_k \left( \nabla^2 f(Sy^k) \right)^{-1} \nabla f(Sy^k)
\]

By replacing \( Sy^k \) with \( x^k \), we have

\[
x^{k+1} = x^k - \alpha_k \left( \nabla^2 f(x^k) \right)^{-1} \nabla f(x^k)
\]

which is Newton’s method in the \( x \)-space.

1b. (5%) The results from part (a) indicate that Newton’s method is invariant with respect to scaling. This implies that the proposed transformation will not provide any benefit with respect to convergence.
2. (35%) We are considering the following problems:

\[
P: \ \text{Minimize} \ \{f(x) : g_i(x) \leq 0, \ i = 1, 2, \ldots, p; \ x \in \mathbb{R}^n\}
\]

\[
R: \ f(x^*) = \text{Minimize} \ \{f(x) + \mu_1 g_1(x) + \cdots + \mu_p g_p(x) : \mu_i g_i(x) = 0, \ i = 1, 2, \ldots, p; \ x \in \mathbb{R}^n\}
\]

where \(x^*\) is a regular point of the constraints in \(P\).

Part 1 (Necessity): We must show that \(x^*\) solves \(P\) only if there exist nonnegative multipliers \(\mu_1 \geq 0, \mu_2 \geq 0, \ldots, \mu_p \geq 0\) such that \(x^*\) solves \(R\).

Assume that there are no nonnegative multipliers such that \(x^*\) solves \(R\). This would imply that \(x^*\) does not solve \(P\) because it would not be possible to find nonnegative multipliers that satisfied the FONC (first-order necessary conditions) for \(P\).

Part 2 (Sufficiency): Must show \(x^*\) solves \(P\) if there exist nonnegative multipliers such that \(x^*\) solves \(R\).

Let the multipliers be those that solve \(P\). Substitute these multipliers into \(R\) but ignore the constraints \(\mu_i g_i(x) = 0, \ i = 1, 2, \ldots, p\). The FONC for \(R\) (without the constraints) is that the gradient of the objective function in \(R\) be zero. This will be true at \(x^*\) since \(x^*\) solves \(P\). Thus \(x^*\) solves \(R\) for the given multipliers because complementary slackness will automatically be satisfied. (If \(x^*\) solves \(R\) without the constraints, and \(x^*\) satisfies the constraints then it also solves \(R\) because the objective function could not improve when constraints are added.

Finally, global optimality is assured because \(P\) and \(R\) (without the constraints) are convex programs and \(x^*\) is a regular point.
3. (40%) We are solving the problem Minimize $f(x) = \frac{1}{2} x^T Q x - b^T x$ using the method of steepest descent. For this method, the direction is

$$d^k = -\nabla f(x^k) = -Qx^k + b.$$

a. (10%) If we use an exact line search to find $x^{k+1} = x^k + \alpha d^k$ it is necessary to solve the following problem of a single variable, $\alpha$.

Minimize $\{ f(x^k + \alpha d^k) : \alpha \geq 0 \}$

First-order necessary conditions require

$$\frac{df(x^k + \alpha d^k)}{d\alpha} = \nabla f(x^k + \alpha d^k) d^k = 0,$$

so

$$[Q(x^k + \alpha d^k - b)]^T d^k = 0$$

Solving for $\alpha$ we get:

$$\alpha_k = -\frac{(Qx^k - b)^T d^k}{(d^k)^T Q d^k}$$

b. (10%) Perform one iteration: $\nabla f(x^0) = (-1, -1, -1) = -d^0$. Using these values we get $\alpha_0 = 0.0968$. Thus $x^1 = x^0 + \alpha_0 d^0 = (0.0968, 0.0968, 0.0968)$.

c. (10%) For steepest descent, we have

$$\frac{E(x_{k+1})}{E(x_k)} \leq \left(\frac{r-1}{r+1}\right)^2,$$

where $r = \frac{A}{a} = \frac{\text{largest eigenvalue}}{\text{smallest eigenvalue}}$.

The eigenvalues of $Q$ are 1, 5 and 25 so $r = 25/1 = 25$. Accordingly, the convergence ratio is $(24/26)^2 = 0.8521$. This value means that for $k$ large, each iteration of the steepest descent algorithm reduces the error (difference between the current objective value and the optimal value) by a factor of about 0.85.

d. (10%) In general, steepest descent only converges in the limit, even for a quadratic objective function. If all the eigenvalues are the same (a sphere), however, it will converge in one iteration.