Problem 1: Suppose $A \in \mathbb{R}^{m \times n}$ is full column rank, and $b \in \mathbb{R}^m$. Derive a closed-form solution (meaning a solution expressed involving the parameters $A$ and $b$) to the optimization problem

$$(P) \quad \min \|Ax - b\|_2 \text{ such that } x \in \mathbb{R}^n.$$ 

Be clear why your proposed solution is indeed the solution. [ Hint: $(A^T A)^{-1}$ exists since $A$ is full column rank. Another hint: it doesn’t matter if the objective function is instead $\|Ax - b\|_2^2$. ]

Solution: Indeed, because squaring is an order-preserving operation on $\mathbb{R}_{\geq 0}$, we may equivalently minimize the objective function $\|Ax - b\|_2^2 = (Ax - b)^T(Ax - b) = x^T A^T A x - 2x^T A^T b + b^T b$, which is a quadratic function having gradient $2(A^T A x - A^T b)$ and Hessian $2A^T A$. Since $A$ is full column rank, we have that $A^T A$ is invertible, and since $A^T A$ is always positive semidefinite (previous hw), therefore $A^T A$ here is positive definite, so the Hessian $2A^T A$ is positive definite, hence the objective function is strictly convex. So if the gradient at any point is zero then that point will be a strict and unique global minimizer. Indeed, setting the gradient to zero yields $A^T A x = A^T b$, hence (again because $A^T A$ is invertible) the optimal solution is $x^* = (A^T A)^{-1} A^T b$.

Problem 2: Suppose the height $Y$ that a particular species of plant grows to can be expressed as $Y = \alpha_0 + \alpha_1 t_1 + \alpha_2 t_2 + \alpha_3 t_3 + \epsilon$ where $t_1$ is the amount of nutrient 1 in the soil, $t_2$ is the amount of nutrient 2 in the soil, $t_3$ is the amount of nutrient 3 in the soil, $\alpha_0$, $\alpha_1$, $\alpha_2$, $\alpha_3$ are fixed constants, and $\epsilon$ is a random variable with mean 0. You don’t know the value of these constants $\alpha_0$, $\alpha_1$, $\alpha_2$, $\alpha_3$ and you wish to estimate them. To do this, you do six tests; for each of $i = 1, 2, 3, 4, 5, 6$ you choose levels of nutrients 1, 2, 3 to be $\tau_{i1}$, $\tau_{i2}$, $\tau_{i3}$, respectively, and with these respective levels of nutrients the plant height was, say, $y_i$ where the matrix $\tau$ and vector $y$ are as follows:

$$\tau = \begin{bmatrix} 2 & 3 & 4 \\ 6 & 7 & 2 \\ 1 & 9 & 6 \\ 4 & 8 & 3 \\ 2 & 6 & 1 \\ 7 & 1 & 3 \end{bmatrix}, \quad y = \begin{bmatrix} 5.14 \\ 8.41 \\ 7.98 \\ 7.91 \\ 5.01 \\ 6.91 \end{bmatrix}$$
(For example, in the fifth test you used 2 units of nutrient 1, 6 units of nutrient 2, and 1 unit of nutrient 3 in the soil, and the plant grew to the height 5.01.) In order to estimate \( \alpha_0, \alpha_1, \alpha_2, \alpha_3 \), you decide to select the values of \( a_0, a_1, a_2, a_3 \) which minimizes \( \sum_{i=1}^{6} (a_0 + a_1 \tau_1 + a_2 \tau_2 + a_3 \tau_3 - y_i)^2 \), and these optimal values of \( a_0, a_1, a_2, a_3 \) will be taken as estimates for \( \alpha_0, \alpha_1, \alpha_2, \alpha_3 \), respectively.

a) Express this problem as precisely the kind of problem in this homework’s Problem 1.

b) Solve this problem using your solution to Problem 1, i.e., give your estimates of \( \alpha_0, \alpha_1, \alpha_2, \alpha_3 \).

Solution: In the notation of Problem 1, the matrix \( A \) is the matrix \( \tau \) with an additional column of ones appended to its left, \( x \) is the vector of unknowns \([a_0, a_1, a_2, a_3]^T\), and \( b \) is the vector \( y \). Solving \( \min \|Ax - b\|_2^2 \) for \( x \in \mathbb{R}^4 \) is what needs to be accomplished. Computing \((A^T A)^{-1} A^T b\) yields the estimates 7985, 6163, 4181, 4630 respectively for \( \alpha_0, \alpha_1, \alpha_2, \alpha_3 \).

Problem 3: Let \( S \subset \mathbb{R}^n \) be a convex set, and suppose \( f : S \to \mathbb{R} \) is a function. Prove that \( f \) is a convex function if and only if the epigraph of \( f \) is a convex set.

Solution: Recall that the definition of the epigraph of \( f \) is all points \([z^T] \in \mathbb{R}^{n+1}\) such that \( z \in S \) and \( a \in \mathbb{R} \) satisfy \( f(z) \leq a \). If the epigraph of \( f \) is a convex set then for any \( x, y \in S \) and any \( \lambda \in [0, 1] \) we have—because \([f(x)] \in \mathbb{R}^{n+1}\) and \([f(y)] \in \mathbb{R}^{n+1}\) are in the epigraph—that \( \lambda [f(x)] + (1 - \lambda) [f(y)] = [\lambda x + (1 - \lambda) y] \) is in the epigraph, which exactly asserts by the definition of epigraph that \( f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y) \), which means that \( f \) is a convex function. Conversely, suppose \( f \) is a convex function. Then for any \([z^T] \) and \([y^T] \) in the epigraph of \( f \) and for any \( \lambda \in [0, 1] \) we have that \( f(x) \leq a, f(y) \leq b \) and, by convexity of \( f \), \( f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y) \) \( \leq \lambda a + (1 - \lambda) b \), thus \([\lambda x + (1 - \lambda) y] \) is in the epigraph, which means that the epigraph is a convex set.

Problem 4: Suppose you are given \( A \in \mathbb{R}^{n \times n} \) symmetric positive definite, \( b \in \mathbb{R}^n \), and \( c \in \mathbb{R} \), and you are considering the function \( f : \mathbb{R}^n \to \mathbb{R} \) defined by \( f(x) = \frac{1}{2} x^T A x + b^T x + c \). Now, suppose you are given a particular point \( \bar{x} \in \mathbb{R}^n \) and a nonzero direction \( d \in \mathbb{R}^n \). Derive the optimal solution to the one-dimensional subproblem

\[
\min f(\bar{x} + \alpha d) \quad \text{such that} \quad \alpha \in \mathbb{R}.
\]

Hint: The answer will be that the optimal value \( \alpha \) is \( \alpha^* = -\frac{(A \bar{x} + b)^T d}{d^T A d} \).

Solution: This one-dimensional subproblem is to minimize (with respect to one real variable \( \alpha \)
the function
\[
  f(\bar{x} + \alpha d) = \frac{1}{2} (\bar{x} + \alpha d)^T A(\bar{x} + \alpha d) + b^T (\bar{x} + \alpha d) + c
\]
which has first derivative \( d^T Ad \cdot \alpha + \bar{x}^T A^T d + b^T d \) and second derivative \( d^T Ad \); Note that the second derivative is always positive (hence this one-dimensional function is strictly convex) and that the first derivative is zero when \( \alpha = -\frac{(Ax + b)^T d}{d^T Ad} \), hence this point is the global minimizer for the one-dimensional subproblem.

**Problem 5:** Suppose you want to minimize the function \( f : \mathbb{R}^2 \rightarrow \mathbb{R} \) given by, for all \( x \in \mathbb{R}^2 \),
\[
f(x) = \frac{1}{2} x^T A x + b^T x + c
\]
where \( A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \), \( b = \begin{bmatrix} 5 \\ 4 \end{bmatrix} \), and \( c = 10 \).

**a)** Solve this optimization problem exactly using first and second order conditions.

**b)** Show that for any initial guess \( x^{(0)} \), steepest descent method generates \( x^{(1)} \), \( x^{(2)} \), \( x^{(3)} \), \( x^{(4)} \), ... , where, for all \( k \), \( x^{(k+1)} = x^{(k)} + d^T d_k d \) with \( d = -(Ax^{(i)} + b) \). Hint: Use Problem 4 directly.

**c)** On MATLAB, run twenty iterations of steepest descent method starting from point \( x^{(0)} = \begin{bmatrix} 9 \\ 8 \end{bmatrix} \).

**d)** Plot the sequence of steepest descent iterates from part c in \( \mathbb{R}^2 \) (and zoom in as necessary to follow the iterates). What do you notice?

**Solution:** The optimal point is \( x^{(*)} = \begin{bmatrix} -5 \\ -4 \end{bmatrix} \); this can be seen by setting the gradient of \( f \) to zero, and noting that the Hessian of \( f \) is positive definite. Steepest descent will zig-zag in steps-orthogonal-to-previous-step. Sample MATLAB code for part c and d:

```matlab
function [x]=solvequadsteep(A,b,startx,iter)
    x=startx;
    store=zeros(2,iter+1);
    store(:,1)=x;
    for i=1:iter
        d=-(A*x+b);
        alpha=(d'*d)/(d'*A*d);
        x=x+alpha*d;
        store(:,i+1)=x;
    end
    close
    hold on
    plot(store(1,:),store(2,:),'b*')
    plot(store(1,:),store(2,:),'g')
end
```