1. Let $x \in \mathbb{R}$ be an unknown scalar variable. Consider the optimization problem

$$\text{minimize} \quad \|x\mathbb{1} - a\|_2^2,$$

in which $a \in \mathbb{R}^n$ is a given vector and $\mathbb{1} \in \mathbb{R}^n$ is the column vector of all ones, $\mathbb{1} := [\,1 \ \ldots \ 1\,]^T$.

   (a) Find the optimal value of x using variational methods, i.e., set $\partial J/\partial x = 0$ where $J(x) := \|x\mathbb{1} - a\|_2^2$.

   (b) Find the optimal value of x using the least-squares framework developed in the beginning of the semester. Fully simplify your result (i.e., do not just write the formula for least-squares).

This problem can be interpreted as finding the 'constant vector' $x\mathbb{1} = [\,x \ \ldots \ x\,]^T$ that best approximates a given vector a.

2. Consider the space of $n \times n$ matrices with the inner product $\langle X, Y\rangle := \text{trace}(X^TY)$. This inner product gives rise to the Frobenius matrix norm $\|\cdot\|_F$ by taking the inner product of a matrix with itself,

$$\|X\|_F^2 := \text{trace}(X^TX).$$

Let $x \in \mathbb{R}$ be an unknown scalar variable. Find the optimal value of x for the problem

$$\text{minimize} \quad \|xI - A\|_F^2,$$

in which $A \in \mathbb{R}^{n\times n}$ is a given matrix and I is the identity matrix.

This problem can be interpreted as finding the 'constant matrix' xI that best approximates a given matrix A.

3. Solve the regularized least-squares problem defined as

$$\text{minimize} \quad \|Ax - y\|_W^2 + (x - x_0)^T P\,(x - x_0),$$

in which x_0 is a given vector, W and P are both known symmetric positive definite matrices, and $\|\cdot\|_W^2$ is the weighted 2-norm,

$$\|z\|_W^2 := z^TWz.$$

This problem can be interpreted as follows. The parameters P and x_0 allow us to incorporate additional *a priori* knowledge into the formulation of the problem. Different choices for P will indicate how confident we are about the closeness of the optimal solution x^* to a given vector x_0; a 'large' P indicates a high degree of confidence that x_0 is a good guess for the optimal solution, while a 'small' P indicates a high degree of uncertainty in the initial guess x_0. Clearly, as $P \to 0$, we recover the standard weighted least-squares problem.

4. Consider the function $J(X) = \text{trace}(X^{-1})$, where X is a positive definite matrix. Find the first-order variation δJ and the gradient ∇J. Find the second-order variation $\delta^2 J$. (You do not need to find the Hessian $\nabla^2 J$.)

Hint: Replace X with $X + \delta X$ and use a Taylor expansion of $J(X+\delta X)$ to find the first-order variation δJ and the gradient ∇J at X. You will need the identity $(XY)^{-1} = Y^{-1}X^{-1}$. Also, it helps to think of the procedure as a generalization, to the case of matrices, of writing a Taylor expansion for $f(x) = 1/x$,

$$\begin{aligned} \frac{1}{x + \delta x} &= \frac{1}{x}\,\frac{1}{1 + \dfrac{\delta x}{x}} \\ &= \frac{1}{x}\Big(1 - \frac{\delta x}{x} + (\frac{\delta x}{x})^2 + \text{h.o.t.}\Big). \end{aligned}$$

5. (Extra problem; will not be graded.) Let J be a scalar-valued function of the matrix $X \in \mathbb{R}^{m\times n}$, and suppose that after an excursion δX in the variable X the first-order variation of $J(X + \delta X)$ can be expressed as

$$\delta J(X) = \text{trace}([f(X)]^T\delta X),$$

where f is a matrix-valued function of X. Clearly, if we require that $\delta J(X) = 0$ for *all possible* excursions $\delta X \in \mathbb{R}^{m\times n}$, then we obtain the equation $f(X) = 0$ as a necessary condition for optimality.

Now consider a scenario in which the variable X, and therefore its excursions δX, are restricted to a subspace $\mathcal{S}$ of $\mathbb{R}^{m\times n}$ that dictates the zero entries of X. For example, let $\mathcal{S}$ be the subspace of all $m\times n$ matrices whose corner elements are all zero, i.e., if $X\in\mathcal{S}$ then $X_{11}=X_{1n}=X_{m1}=X_{mn}=0$. Since X is restricted to $\mathcal{S}$ then so are all its excursions δX, and thus $(\delta X)_{11}=(\delta X)_{1n}=(\delta X)_{m1}=(\delta X)_{mn}=0$ for every allowable δX.

Returning again to the equation $\delta J(X)=\mathrm{trace}([f(X)]^T\delta X)$, we now require that $\delta J(X)=0$ for all excursions $\delta X\in\mathcal{S}$. Use this to derive necessary conditions for optimality in terms of $f(X)$.

This problem can be interpreted as pertaining to network design subject to communication constraints. For example, in distributed control systems, the entries X_{ij} and X_{ji} of the matrix X may represent the amount/strength of communication between nodes i and j. If, for some reason, certain nodes are not allowed or are unable to communicate to each other, then the corresponding elements of the matrix X have to be zero. The optimal X, that minimizes a performance objective J, is then sought subject to such communication and architectural constraints.