Midterm exam

1. *Learning system dynamics from data.* In *system identification*, we are given some time series values for a discrete-time input vector signal,

$$u(1),\ u(2), \ldots, u(N) \in \mathbf{R}^m,$$

and also a discrete-time state vector signal,

$$x(1),\ x(2), \ldots, x(N) \in \mathbf{R}^n,$$

and we are asked to find matrices $A \in \mathbf{R}^{n\times n}$ and $B \in \mathbf{R}^{n\times m}$ such that we have

$$x(t+1) \approx Ax(t) + Bu(t), \quad t = 1, \ldots, N-1. \tag{1}$$

We use the symbol $\approx$ since there may be small measurement errors in the given signal data, so we don't expect to find matrices A and B for which the linear dynamical system equations hold exactly. Let's give a quantitative measure of how well the linear dynamical system model (1) holds, for a particular choice of matrices A and B. We define the RMS (root-mean-square) value of the residuals associated with our signal data and a candidate pair of matrices A, B as

$$R = \left(\frac{1}{N-1} \sum_{t=1}^{N-1} \|x(t+1) - Ax(t) - Bu(t)\|^2 \right)^{1/2}.$$

We define the RMS value of x, over the same period, as

$$S = \left(\frac{1}{N-1} \sum_{t=1}^{N-1} \|x(t+1)\|^2 \right)^{1/2}.$$

We define the *normalized residual*, denoted ρ, as $\rho = R/S$. If we have $\rho = 0.05$, for example, it means that the state equation (1) holds, roughly speaking, to within 5%. Given the signal data, we will choose the matrices A and B to minimize the RMS residual R (or, equivalently, the normalized residual ρ).

Explain how to do this. Does the method always work? If some conditions have to hold, specify them.

2. *QP with a norm constraint.* As a very short review of the prox-gradient algorithm, we can think of a composite function $F(x) = f(x) + g(x)$ where $f(x)$ is differentiable and $g(x)$ does not have to be. The prox-gradient algorithm finds x^{k+1} to improve the current x^k by applying the following two alternating steps.

   (a) *Gradient step:* $x^{k+1/2} = x^k - h^k \nabla f(x^k)$

   (b) *Prox step:* choose x^{k+1} to minimize $g(x) + \frac{1}{2h^k} \|x - x^{k+1/2}\|_2^2$

   Iteratively applying the prox-gradient steps from the initial point x^0 will eventually finds a sequence $x_0, x_1, \dots, x_k$ that converges to the optimal x^* minimizing $F(x)$. The learning rate h^k can be chosen in a variety of sophisticated ways, however we don't care about it in this problem; just assume it's constant.

   Now the problem. Use the prox-gradient algorithm to solve a class of quadratic optimization problems with a norm constraint. Given P, q, r in appropriate sizes, and a real number s, your job is to find the optimal $x \in \mathbf{R}^n$ to the following.

$$\begin{array}{ll} \underset{x}{\text{minimize}} & x^T P x + 2q^T x \\ \text{subject to} & \|x - r\|_2 \leq s \end{array}$$

   Explain how you will apply the prox-gradient algorithm to the above problem. Explicitly state the gradient step and the prox step in precise details.