Homework 5

EE 263 Stanford University Summer 2019

Due: July 31, 2019

Question **7**, and **8** are **optional** questions (they will not be graded) and you do not need to submit them to gradescope. However, doing them will help improve your understanding and grasp of the material. Q8 relates to the Least-norm solution of nonlinear equations.

1. **The smoothest input that takes the state to zero.** We consider the discrete-time linear dynamical system $x(t+1) = Ax(t) + Bu(t)$, with

$$
A = \begin{bmatrix} 1.0 & 0.5 & 0.25 \\ 0.25 & 0 & 1.0 \\ 1.0 & -0.5 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 1.0 \\ 0.1 \\ 0.5 \end{bmatrix}, \qquad x(0) = \begin{bmatrix} 25 \\ 0 \\ -25 \end{bmatrix}.
$$

The goal is to choose an input sequence $u(0), u(1), \ldots, u(19)$ that yields $x(20) = 0$. Among the input sequences that yield $x(20) = 0$, we want the one that is *smoothest, i.e.*, that minimizes

$$
J_{\text{smooth}} = \left(\frac{1}{20} \sum_{t=0}^{19} (u(t) - u(t-1))^2 \right)^{1/2},
$$

where we take $u(-1) = 0$ in this formula. Explain how to solve this problem. Plot the smoothest input u_{smooth}, and give the associated value of J_{smooth}.

2. **Minimum fuel and minimum peak input solutions.** Suppose $A \in \mathbb{R}^{m \times n}$ is fat and full rank, so there are many x's that satisfy $Ax = y$. In lecture we encountered the *least-norm* solution given by $x_{\text{ln}} = A^{\mathsf{T}}(AA^{\mathsf{T}})^{-1}y$. This solution has the minimum (Euclidean) norm among all solutions of $Ax = y$. In many applications we want to minimize another norm of x (*i.e.*, measure of size of x) subject to $Ax = y$. Two common examples are the 1-norm and ∞-norm, which are defined as

$$
\|x\|_1 = \sum_{i=1}^{n} |x_i|, \qquad \|x\|_\infty = \max_{i=1,\ldots,n} |x_i|.
$$

The 1-norm, for example, is often a good measure of fuel use; the ∞-norm is the *peak* of the vector or signal x. There is no simple formula for the least 1-norm or ∞-norm solution of $Ax = y$, like there is for the least (Euclidean) norm solution. They can be computed very easily, however. (That's one of the topics of EE364.) The analysis is a bit trickier as well, since we can't just differentiate to verify that we have the minimizer. For example, how would you *know* that a solution of $Ax = y$ has minimum 1-norm? In this problem you will explore this

idea. First verify the following inequality, which is like the Cauchy-Schwarz inequality (but even easier to prove): for any $v, w \in \mathbb{R}^p$, the following inequality holds: $w^\mathsf{T} v \le \|v\|_\infty \|w\|_1$. From this inequality it follows that whenever $v \neq 0$,

$$\|w\|_1 \ge \frac{w^\mathsf{T} v}{\|v\|_\infty}.$$

Now let z be any solution of $Az = y$, and let $\lambda \in \mathbb{R}^m$ be such that $A^\mathsf{T}\lambda \neq 0$. Explain why we must have

$$\|z\|_1 \ge \frac{\lambda^\mathsf{T} y}{\|A^\mathsf{T}\lambda\|_\infty}.$$

Thus, any solution of $Az = y$ must have 1-norm at least as big as the righthand side expression. Therefore if you can find $x_\text{mf} \in \mathbb{R}^n$ (mf stands for minimum fuel) and $\lambda \in \mathbb{R}^m$ such that $Ax_\text{mf} = y$ and

$$\|x_\text{mf}\|_1 = \frac{\lambda^\mathsf{T} y}{\|A^\mathsf{T}\lambda\|_\infty},$$

then x_mf is a minimum fuel solution. (Explain why.) Methods for computing x_mf and the mysterious vector λ are described in EE364. In the rest of this problem, you'll use these ideas to verify a statement made during lecture. Now consider the problem from the lecture notes of a unit mass acted on by forces $x_1, \ldots, x_{10}$ for one second each. The mass starts at position $p(0) = 0$ with zero velocity and is required to satisfy $p(10) = 1$, $\dot{p}(10) = 0$. There are, of course, many force vectors that satisfy these requirements. In the lecture notes, you can see a plot of the least (Euclidean) norm force profile. In class I stated that the minimum fuel solution is given by $x_\text{mf} = (1/9, 0, \ldots, 0, -1/9)$, i.e., an accelerating force at the beginning, 8 seconds of coasting, and a (braking) force at the end to decelerate the mass to zero velocity at $t = 10$. Prove this. *Hint:* try $\lambda = (1, -5)$. Verify that the 1-norm of x_mf is less than the 1-norm of x_ln, the (Euclidean) least-norm solution. Feel free to use matlab. There are several convenient ways to find the 1- and ∞-norm of a vector z, e.g., `norm(z,1)` and `norm(z,inf)` or `sum(abs(z))` and `max(abs(z))`. One last question, for fun: what do you think is the minimum peak force vector x_mp? How would you verify that a vector x_mp (mp for minimum peak) is a minimum ∞-norm solution of $Ax = y$? This input, by the way, is very widely used in practice. It is (basically) the input used in a disk drive to move the head from one track to another, while respecting a maximum possible current in the disk drive motor coil. *Hints:*

- The input is called *bang-bang*.

- Some people drive this way.

3. **Singularity of the KKT matrix.** This problem concerns the general norm minimization with equality constraints problem (described in the lectures notes on pages 8-13),

$$\begin{array}{ll} \text{minimize} & \|Ax - b\| \\ \text{subject to} & Cx = d \end{array}$$

where the variable is $x \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, and $C \in \mathbb{R}^{k \times n}$. We asume that C is fat ($k \le n$), i.e., the number of equality constraints is no more than the number of variables.

Using Lagrange multipliers, we found that the solution can be obtained by solving the linear equations

$$\begin{bmatrix} A^\mathsf{T}A & C^\mathsf{T} \\ C & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} A^\mathsf{T}b \\ d \end{bmatrix}$$

for x and λ. (The vector x gives the solution of the norm minimization problem above.) The matrix above, which we will call $K \in \mathbb{R}^{(n+k)\times(n+k)}$, is called the *KKT matrix* for the problem. (KKT are the initials of some of the people who came up with the optimality conditions for a more general type of problem.)

One question that arises is, when is the KKT matrix K nonsingular? The answer is: K is nonsingular if and only if C is full rank and $\operatorname{null}(A) \cap \operatorname{null}(C) = \{0\}$.

You will fill in all details of the argument below.

a) Suppose C is not full rank. Show that K is singular.

b) Suppose that there is a nonzero $u \in \operatorname{null}(A) \cap \operatorname{null}(C)$. Use this u to show that K is singular.

c) Suppose that K is singular, so there exists a nonzero vector $[u^\mathsf{T} \ v^\mathsf{T}]^\mathsf{T}$ for which

$$\begin{bmatrix} A^\mathsf{T}A & C^\mathsf{T} \\ C & 0 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = 0.$$

Write this out as two block equations, $A^\mathsf{T}Au + C^\mathsf{T}v = 0$ and $Cu = 0$. Conclude that $u \in \operatorname{null}(C)$. Multiply $A^\mathsf{T}Au + C^\mathsf{T}v = 0$ on the left by u^T, and use $Cu = 0$ to conclude that $\|Au\| = 0$, which implies $u \in \operatorname{null}(A)$. Finish the argument that leads to the conclusion that either C is not full rank, or $\operatorname{null}(A) \cap \operatorname{null}(C) \neq \{0\}$.

4. **Portfolio selection with sector neutrality constraints.** We consider the problem of selecting a portfolio composed of n assets. We let $x_i \in \mathbb{R}$ denote the investment (say, in dollars) in asset i, with $x_i < 0$ meaning that we hold a short position in asset i. We normalize our total portfolio as $\mathbf{1}^\mathsf{T}x = 1$, where $\mathbf{1}$ is the vector with all entries 1. (With normalization, the x_i are sometimes called *portfolio weights*.)

The portfolio (mean) return is given by $r = \mu^\mathsf{T}x$, where $\mu \in \mathbb{R}^n$ is a vector of asset (mean) returns. We want to choose x so that r is large, while avoiding risk exposure, which we explain next.

First we explain the idea of *sector exposure*. We have a list of k economic sectors (such as manufacturing, energy, transportation, defense, ...). A matrix $F \in \mathbb{R}^{k\times n}$, called the *factor loading matrix*, relates the portfolio x to the *factor exposures*, given as $R^{\text{fact}} = Fx \in \mathbb{R}^k$. The number R_i^{fact} is the portfolio risk exposure to the ith economic sector. If R_i^{fact} is large (in magnitude) our portfolio is exposed to risk from changes in that sector; if it is small, we are less exposed to risk from that sector. If $R_i^{\text{fact}} = 0$, we say that the portfolio is *neutral* with respect to sector i.

Another type of risk exposure is due to fluctuations in the returns of the individual assets. The *idiosyncratic risk* is given by

$$R^{\text{id}} = \sum_{i=1}^{n} \sigma_i^2 x_i^2,$$

where $\sigma_i > 0$ are the standard deviations of the asset returns. (You can take the formula above as a definition; you do not need to understand the statistical interpretation.)

We will choose the portfolio weights x so as to maximize $r - \lambda R^{\mathrm{id}}$, which is called the *risk-adjusted return*, subject to neutrality with respect to all sectors, *i.e.*, $R^{\mathrm{fact}} = 0$. Of course we also have the normalization constraint $\mathbf{1}^{\mathsf{T}} x = 1$. The parameter λ, which is positive, is called the *risk aversion parameter*. The (known) data in this problem are $\mu \in \mathbb{R}^n$, $F \in \mathbb{R}^{k \times n}$, $\sigma = (\sigma_1, \ldots, \sigma_n) \in \mathbb{R}^n$, and $\lambda \in \mathbb{R}$.

a) Explain how to find x, using methods from the course. You are welcome (even encouraged) to express your solution in terms of block matrices, formed from the given data.

b) Using the data given in `sector_neutral_portfolio_data.m`, find the optimal portfolio. Report the associated values of r (the return), and R^{id} (the idiosyncratic risk). Verify that $\mathbf{1}^{\mathsf{T}} x = 1$ (or very close) and $R^{\mathrm{fact}} = 0$ (or very small).

5. **A simple population model.** We consider a certain population of fish (say) each (yearly) season. $x(t) \in \mathbb{R}^3$ will describe the population of fish at year $t \in \mathbb{Z}$, as follows:

- $x_1(t)$ denotes the number of fish less than one year old

- $x_2(t)$ denotes the number of fish between one and two years old

- $x_3(t)$ denotes the number of fish between two and three years

(We will ignore the fact that these numbers are integers.) The population evolves from year t to year $t + 1$ as follows.

- The number of fish less than one year old in the next year $(t + 1)$ is equal to the total number of offspring born during the current year. Fish that are less than one year old in the current year (t) bear no offspring. Fish that are between one and two years old in the current year (t) bear an average of 2 offspring each. Fish that are between two and three years old in the current year (t) bear an average of 1 offspring each.

- 40% of the fish less than one year old in the current year (t) die; the remaining 60% live on to be between one and two years old in the next year $(t + 1)$. fish, and 50% of the old fish.

- 30% of the one-to-two year old fish in the current year die, and 70% live on to be two-to-three year old fish in the next year.

- All of the two-to-three year old fish in the current year die.

Express the population dynamics as an autonomous linear system with state $x(t)$, *i.e.*, in the form $x(t + 1) = Ax(t)$. **Remark:** this example is silly, but more sophisticated population dynamics models are very useful and widely used.

6. Controlling a system using the initial conditions. Consider the mechanical system shown below:
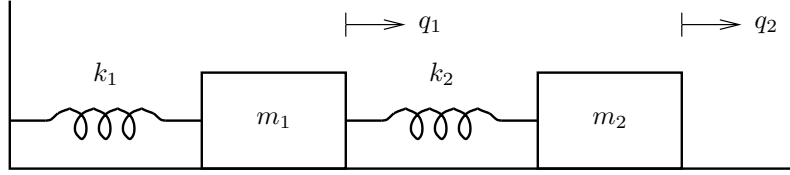
Here q_i give the displacements of the masses, m_i are the values of the masses, and k_i are the spring stiffnesses, respectively. The dynamics of this system are

$$\dot{x} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\frac{k_1+k_2}{m_1} & \frac{k_2}{m_1} & 0 & 0 \\ \frac{k_2}{m_2} & -\frac{k_2}{m_2} & 0 & 0 \end{bmatrix} x$$

where the state is given by

$$x = \begin{bmatrix} q_1 \\ q_2 \\ \dot{q}_1 \\ \dot{q}_2 \end{bmatrix}.$$

Immediately before $t = 0$, you are able to apply a strong impulsive force α_i to mass i, which results in initial condition

$$x(0) = \begin{bmatrix} 0 \\ 0 \\ \alpha_1/m_1 \\ \alpha_2/m_2 \end{bmatrix}.$$

(i.e., each mass starts with zero position and a velocity determined by the impulsive forces.) This problem concerns selection of the impulsive forces α_1 and α_2. For parts a–c below, the parameter values are

$$m_1 = m_2 = 1, \quad k_1 = k_2 = 1.$$

Consider the following specifications:

a) $q_2(10) = 2$

b) $q_1(10) = 1$, $q_2(10) = 2$

c) $q_1(10) = 1$, $q_2(10) = 2$, $\dot{q}_1(10) = 0$, $\dot{q}_2(10) = 0$

d) $q_2(10) = 2$ when the parameters have the values used above (*i.e.*, $m_1 = m_2 = 1$, $k_1 = k_2 = 1$), and *also*, $q_2(10) = 2$ when the parameters have the values $m_1 = 1$, $m_2 = 1.3$, $k_1 = k_2 = 1$.

Determine whether each of these specifications is feasible or not (*i.e.*, whether there exist α_1, $\alpha_2 \in \mathbb{R}$ that make the specification hold). If the specification is feasible, find the particular α_1, α_2 that satisfy the specification and minimize $\alpha_1^2 + \alpha_2^2$. If the specification is infeasible, find the particular α_1, α_2 that come closest, in a least-squares sense, to satisfying the specification. (For example, if you cannot find α_1, α_2 that satisfy $q_1(10) = 1$, $q_2(10) = 2$, then find α_i that minimize $(q_1(10) - 1)^2 + (q_2(10) - 2)^2$.) Be sure to be very clear about which alternative holds for each specification.

7. **Robust input design.** We are given a system, which we know follows $y = Ax$, with $A \in \mathbb{R}^{m \times n}$. Our goal is to choose the input $x \in \mathbb{R}^n$ so that $y \approx y^{\text{des}}$, where $y^{\text{des}} \in \mathbb{R}^m$ is a given target outcome. We'll assume that $m \leq n$, *i.e.*, we have more degrees of freedom in our choice of input than specifications for the outcome. If we knew A, we could use standard EE263 methods to choose x. The catch here, though, is that we don't know A exactly; it varies a bit, say, day to day. But we do have some possible values of A,

$$A^{(1)}, \ldots, A^{(K)},$$

which might, for example, be obtained by measurements of A taken on different days. We now define $y^{(i)} = A^{(i)}x$, for $i = 1, \ldots, K$. Our goal is to choose x so that $y^{(i)} \approx y^{\text{des}}$, for $i = 1, \ldots, K$.

We will consider two different methods to choose x.

- *Least norm method.* Define $\bar{A} = (1/K) \sum_{i=1}^{K} A^{(i)}$. Choose x^{ln} to be the least-norm solution of $\bar{A}x = y^{\text{des}}$. (You can assume that $\bar{A}$ is full rank.)

- *Mean-square error minimization method.* Choose x^{mmse} to minimize the mean-square error

$$\frac{1}{K} \sum_{i=1}^{K} \|y^{(i)} - y^{\text{des}}\|^2.$$

a) Give formulas for x^{ln} and x^{mmse}, in terms of y^{des} and $A^{(1)}, \ldots, A^{(K)}$. You can make any needed rank assumptions about matrices that come up, but please state them explicitly.

b) Find x^{ln} and x^{mmse} for the problem with data given in `rob_inp_des_data.m`. Running this M-file will define `ydes` and the matrices $A^{(i)}$ (given as a 3 dimensional array; for example, `A(:,:,13)` is $A^{(13)}$). Also included in the data file (commented out) is code to produce scatter plots of your results. Write down the values of x^{ln} and x^{mmse} you found. Produce and submit scatter plots of $y^{(i)}$ for x^{ln} and x^{mmse}. Use the code we provided as a template for your plots.

8. **Least-norm solution of nonlinear equations.** Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is a function, and $y \in \mathbb{R}^m$ is a vector, where $m < n$ (*i.e.*, x has larger dimension than y). We say that $x \in \mathbb{R}^n$ is a least-norm solution of $f(x) = y$ if for any $z \in \mathbb{R}^n$ that satisfies $f(z) = y$, we have $\|z\| \geq \|x\|$. When the function f is linear or affine (*i.e.*, linear plus a constant), the equations $f(x) = y$ are linear, and we know how to find the least-norm solution for such problems. In general, however, it is an extremely difficult problem to compute a least-norm solution to a set of nonlinear equations. There are, however, some good heuristic iterative methods that work well when the function f is not too far from affine, *i.e.*, its nonlinear terms are small compared to its linear and constant part. You may assume that you have a starting guess, which we'll call $x^{(0)}$. This guess doesn't necessarily satisfy the equations $f(x) = y$.

a) Suggest an iterative method for (approximately) solving the nonlinear least-norm problem, starting from the initial guess $x^{(0)}$. Use the notation $x^{(k)}$ to denote the kth iteration of your method. Explain clearly how you obtain $x^{(k+1)}$ from $x^{(k)}$. If you need to make any assumptions about rank of some matrix, do so. (You don't have to worry about

what happens if the matrix is not full rank.) Your method should have the property that $f(x^{(k)})$ converges to y as k increases. (In particular, we don't need to have the iterates satsfiy the nonlinear equations exactly.) Suggest a name for the method you invent. Your method should not be complicated or require a long explanation. You do not have to prove that the method converges, or that when it converges, it converges to a least-norm solution. All you have to do is suggest a sensible, simple method that ought to work well when f is not too nonlinear, and the starting guess $x^{(0)}$ is good.

b) Now we consider a specific example, with the function $f : \mathbb{R}^5 \to \mathbb{R}^2$ given by

$$f_1(x) = 2x_1 - 3x_3 + x_5 + 0.1x_1x_2 - 0.5x_2x_5,$$
$$f_2(x) = -x_2 + x_3 - x_4 + x_5 - 0.6x_1x_4 + 0.3x_3x_4.$$

Note that each component of f consists of a linear part, and also a quadratic part. Use the method you invented in part a to find the least-norm solution of

$$f(x) = y = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

(We repeat that you do not have to prove that the solution you found is really the least-norm one.) As initial guess, you can use the least-norm solution of the linear equations resulting if you ignore the quadratic terms in f. Make sure to turn in your matlab code as well as to identify the least-norm x you find, its norm, and the equation residual, *i.e.*, $f(x) - y$ (which should be very small).