Assignment #9

REVISED $\rightarrow$ Due Wednesday, 7 December at the start of class

Please read Chapters 13 and 15 in Nocedal & Wright. Do the following Exercises and Problems.

Exercise 13.1. (*Hint*. Pages 356–357.)

Problem P22. Fix $\alpha \in \mathbb{R}$. Consider the linear programming problem

$$\min\ \alpha x_1 - 2x_2 \qquad \text{subject to} \qquad \begin{aligned} &-3x_1 + x_2 \le 1 \\ &6x_1 - 2x_2 \le 9 \\ &x_1 \ge 0,\ x_2 \ge 0 \end{aligned}$$

(a) Sketch the feasible set with some care and note it is unbounded. For what values of α does the problem have a solution?

(b) Add slack variables to put the problem in standard form (13.1). For the particular value $\alpha = 10$, solve the problem by hand using the simplex method and a template as done in class. (*Start with a basic feasible point (vector) with $x_1 = x_2 = 0$ as in the examples done in lecture. If needed, download and print the template from online:* `bueler.github.io/M661F16/linprogtemplate.pdf`)

(c) To confirm your answer from part **(b)**, run the code `rsimpII.m`, which I posted at

`bueler.github.io/M661F16/matlab/rsimpII.m`,

You probably want to start by typing "`help rsimpII`".

Problem P23. *Recall least-squares problems from Chapter 10. It is common to minimize a sum of squares of misfits (i.e. residuals), but subject to additional "exact" requirements, giving an equality-constrained problem (e.g. as in Chapter 12). Such problems are often called "inverse modeling." This is a visualizable and finite-dimensional example.*

Consider the two sets of data

t	1	4
w	2	1

,

t	0	2	3	5	6
y	1	1	2	2	3

The first set of data with $q = 2$ points is marked by stars ($*$) in the Figure on the next page, and the second with $m = 5$ points is marked by circles ($\circ$).

Consider the problem of finding a cubic polynomial which fits the second data set as closely as possible, but which is *required* to *exactly* fit the first data set. That is, the polynomial must pass through the two stars. Using the notation of Chapter 10, let

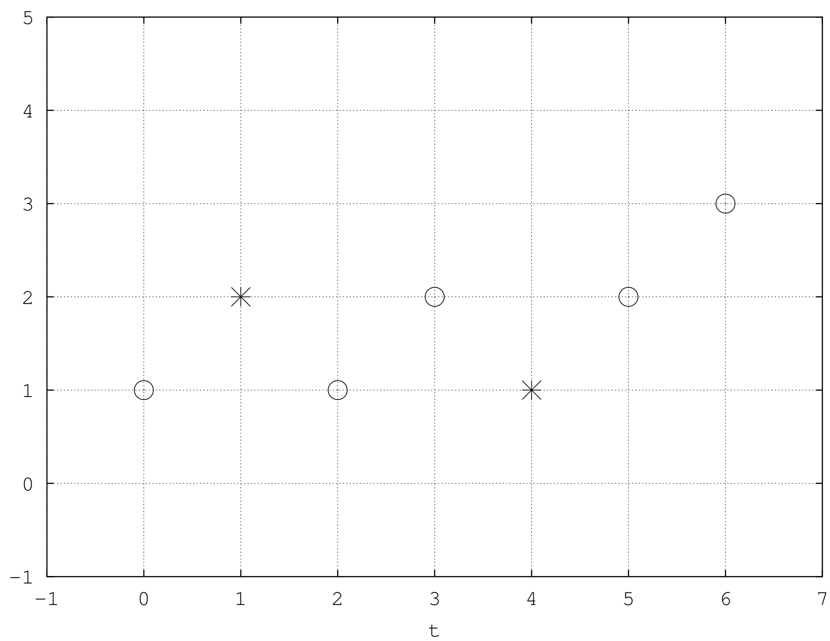
$$\phi(x; t) = x_1 + x_2\, t + x_3\, t^2 + x_4\, t^3$$

be the model, with parameters $x \in \mathbb{R}^n$ where $n = 4$. For $r_j(x) = \phi(x; t_j) - y_j$ let

$$f(x) = \frac{1}{2}\|r(x)\|^2 = \frac{1}{2}\sum_{j=1}^{m} r_j(x)^2.$$

(*Note that only the second data set is used in building* $f(x)$.) We require that the model exactly fits the first data set, so this is an equality constraint. Thus the problem is in form (1.1) = (12.1), namely

$$\min_{x \in \mathbb{R}^n} f(x) \qquad \text{subject to} \quad Ex = w. \tag{1}$$



(a) Explain why $f(x) = \frac{1}{2}\|Jx - y\|^2$ where y is from the second data set and

$$J = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \\ \_\_ & \_\_ & \_\_ & \_\_ \\ \_\_ & \_\_ & \_\_ & \_\_ \end{bmatrix} \in \mathbb{R}^{m \times n};$$

please fill in the remaining entries of the matrix. (*Your answer should start by* defining J, *and only then computing the entries.*) Then compute, using the formula for $\phi(x;t)$ and the first set of data, a specific matrix $E \in \mathbb{R}^{q \times n}$ and vector $w \in \mathbb{R}^q$ for the constraints in problem (1).

(b) Consider the Lagrangian for problem (1),

$$\mathcal{L}_1(x, \lambda) = \frac{1}{2}\|Jx - y\|^2 - \lambda^\top (Ex - w),$$

with $\lambda \in \mathbb{R}^q$. Show that the KKT conditions (12.34) for problem (1) can be written "blockwise" as

$$\begin{bmatrix} J^\top J & -E^\top \\ -E & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} J^\top y \\ -w \end{bmatrix}. \tag{2}$$

The matrix A_1 on the left in (2) has size $(n+q) \times (n+q)$.

Show A_1 is symmetric but that it is not SPD. (*This should be answered theoretically, though it may be confirmed numerically. Find a nonzero vector* $z \in \mathbb{R}^{n+q}$ *for which* $z^\top A_1 z = 0$.)

Also, using MATLAB, compute $\text{cond}(A_1)$.[1]

(c) *It turns out that the condition number in part* **(b)** *is larger than necessary.*

We reformulate (1) as

$$\min_{r \in \mathbb{R}^m} \frac{1}{2}\|r\|^2 \qquad \text{subject to} \quad Ex = w \quad \text{and} \quad r = Jx - y. \tag{3}$$

Note $r \in \mathbb{R}^m$ is now a *variable*, not a function. There is no need to confirm that (3) is equivalent to (1); it should be obvious. The question we address, by looking at condition numbers, is *why* you would transform the problem this way.

Define a new Lagrangian

$$\mathcal{L}_2(r, \mu, \lambda, x) = \frac{1}{2}\|r\|^2 - \lambda^\top (Ex - w) - \mu^\top (Jx - y - r),$$

with $r \in \mathbb{R}^m, \mu \in \mathbb{R}^m, \lambda \in \mathbb{R}^q, x \in \mathbb{R}^n$.

Show that the KKT conditions for problem (3) can be written as

$$\begin{bmatrix} I & 0 & -J \\ 0 & 0 & E \\ -J^\top & E^\top & 0 \end{bmatrix} \begin{bmatrix} r \\ \lambda \\ x \end{bmatrix} = \begin{bmatrix} -y \\ w \\ 0 \end{bmatrix}. \tag{4}$$

(Oddly enough, you eliminate the "extraneous" multipliers μ in writing this down!) The matrix A_2 on the left in (4) has size $N \times N$ where $N = m + q + n$, and thus it might be much bigger than A_1 in (2), but it is rather sparse. Again A_2 is symmetric but not SPD; there is no need to prove this.

Using MATLAB, compute $\text{cond}(A_2)$.

(d) Now use MATLAB to implement both (2) and (4) to solve the problem posed at the beginning. Confirm that the solutions x and λ are the same. (*Don't show me a lot of numbers. Show norms of differences of vectors that should be the same.*) Then plot the result on top of the data, so that you generate a Figure like the one above but showing both the original data and the solution.

---

[1] This condition number, even on such a small problem, is large enough to cause several digits of error in solving (2) numerically. In bigger problems of this least-squares-with-constraints type, the loss of accuracy coming from an ill-conditioned system matrix can be catastrophic when using formulation (2).