Week 7: Friday, Oct 9

Logistics

1. HW 3 has been posted. It was first posted on Wednesday night, but there were some typographical errors. The last corrections were Thursday afternoon; if you downloaded before then, make sure you have the latest version. The due date is Oct 21.

2. HW 2 solutions were posted to the CMS site on Wednesday night.

Sensitivity of $y = A^T x$

At the end of the last lecture, we started on sensitivity of least squares. After thinking about it, I decided it might be best to backtrack and start with a simpler problem: first-order perturbation analysis of rectangular matrix multiplication.

Let $A \in \mathbb{R}^{m \times n}$ with $m \geq n$ and consider

$$y = A^T b.$$

We wish to compute the relative change in y under small relative perturbations to A and b. We start, as usual, by differentiating the formula in order to get a first-order perturbation relationship:

$$\delta y = \delta A^T b + A^T \delta b.$$

Taking norms gives us a relationship between the magnitudes of the perturbations. Using the two-norm, we have

$$\|\delta y\| \leq \|\delta A\|\|b\| + \|A\|\|\delta b\|.$$

Now let us divide everything by $\|y\| = \|A^T b\|$ and rearrange to relate $\|\delta y\|/\|y\|$ to $\|\delta A\|/\|A\|$ and $\|\delta b\|/\|b\|$:

$$\frac{\|\delta y\|}{\|y\|} \leq \frac{\|A\|\|b\|}{\|A^T b\|} \frac{\|\delta A\|}{\|A\|} + \frac{\|A\|\|b\|}{\|A^T b\|} \frac{\|\delta b\|}{\|b\|}.$$

How shall we interpret the quantity $\|A\|\|b\|/\|A^T b\|$? If A were a vector rather than a matrix, the answer would be clear: this would be the formula
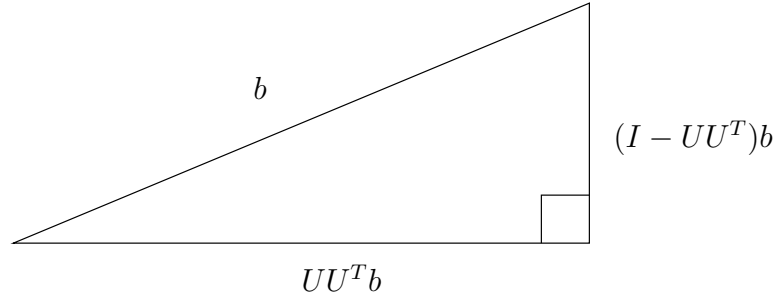
Figure 1: Right triangle consisting of b, the closest vector to b in the range space of A (which is $UU^T b$), and the difference between the two.

for the secant of the angle between two vectors. Let us explore this connection further, using the economy SVD $A = U\Sigma V^T$ (U rectangular, Σ and V square) to help with the analysis[1]. Using invariance under orthogonal transforms, we have

$$\frac{\|A\|\|b\|}{\|A^T b\|} = \frac{\|\Sigma\|\|b\|}{\|\Sigma U^T b\|} \leq \frac{\sigma_1}{\sigma_n}\frac{\|b\|}{\|U^T b\|} = \kappa(A)\sec(\theta),$$

where $\kappa(A) = \sigma_1/\sigma_n$ is the condition number for the square matrix A.

How shall we interpret $\|b\|/\|U^T b\|$? Using orthogonal invariance, note that we can write $\|U^T b\| = \|UU^T b\|$; and $UU^T b = A(A^T A)^{-1}A^T b$ is the closest vector to b in the range space of A. Put differently, we can write b as a sum of two orthogonal vectors

$$b = UU^T b + (I - UU^T)b,$$

where $UU^T b$ is the projection of b onto the range space of A and $(I - UU^T)b$ is normal to the range space of A (Figure 1). So

$$\frac{\|b\|}{\|U^T b\|} = \frac{\|b\|}{\|UU^T b\|} = \sec(\theta),$$

where θ is the smallest angle between the range space of A and the vector b. Putting things together, we have

$$\frac{\|\delta y\|}{\|y\|} \leq \kappa(A)\sec(\theta)\left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|b\|}\right).$$

---

[1] I don't think this is quite the explanation I used in class, but I'm also not sure where I put the relevant page of my original notes!

Sensitivity of least squares

Let's now return to the sensitivity of the least squares problems. The steps are roughly the same as those in the analysis of the previous section:

1. Write down an equation for what we want (the normal equation) and then differentiate to obtain a first-order perturbation relationship.

2. Take norms in order to bound (to first order) the absolute magnitude of the perturbation to the solution in terms of the absolute magnitudes of the perturbations to the problem data.

3. Manipulate the formula to get a bound on the *relative* perturbation to the solution in terms of the relative perturbations to the problem data.

4. Interpret the coefficients in the relative perturbation formula trigonometrically.

We start with the normal equations

$$A^T A x = A^T b.$$

The first-order perturbation relation is

$$\delta A^T A x + A^T \delta A x + A^T A x = \delta A^T b + A^T \delta b,$$

which we arrange to get

$$\delta x = (A^T A)^{-1} \delta A^T (b - Ax) + (A^T A)^{-1} A^T (\delta b - \delta A x).$$

If we define $r = b - Ax$ and let $A = U\Sigma V^T$ with singular values $\sigma_1, \ldots, \sigma_n$, we have

$$\|\delta x\| \leq \|(A^T A)^{-1}\| \|\delta A\| \|r\| + \|(A^T A)^{-1} A^T\| (\|\delta b\| + \|\delta A\| \|x\|)$$

$$= \frac{\|\delta A\|}{\sigma_n^2} \|r\| + \frac{1}{\sigma_n} (\|\delta b\| + \|\delta A\| \|x\|)$$

Dividing through by $\|x\|$ and doing some algebra so that $\|\delta A\|$ and $\|\delta b\|$ only appear in ratios with $\|A\|$ and $\|b\|$, we have

$$\frac{\|\delta x\|}{\|x\|} \leq \kappa(A)^2 \frac{\|r\|}{\|A\| \|x\|} \frac{\|\delta A\|}{\|A\|} + \kappa(A) \left(\frac{\|b\|}{\|A\| \|x\|} \frac{\|\delta b\|}{\|b\|} + \frac{\|\delta A\|}{\|A\|} \right)$$
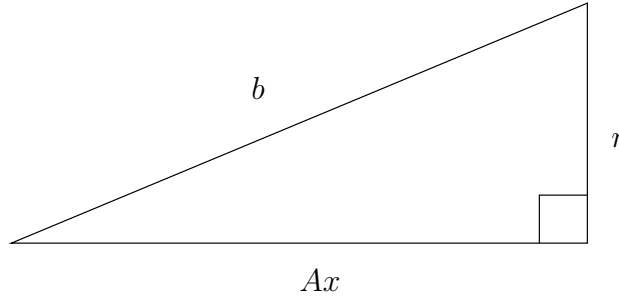
Figure 2: Right triangle consisting of b, the closest vector to b in the range space of A (which is Ax), and the difference between the two (r).

As in the previous section, we can draw an illuminating right triangle (Figure 2) whose sides are $\|Ax\| \le \|A\|\|x\|$, $\|r\|$, and $\|b\|$. In fact, this is the same triangle that we showed in Figure 1, since $Ax = UU^T b$. In terms of this triangle, we have

$$\frac{\|r\|}{\|A\|\|x\|} \le \frac{\|r\|}{\|Ax\|} = \tan(\theta),$$

$$\frac{\|b\|}{\|A\|\|x\|} \le \frac{\|b\|}{\|Ax\|} = \sec(\theta).$$

Putting everything together, we have

$$\frac{\|\delta x\|}{\|x\|} \le \left(\kappa(A)^2 \tan(\theta) + \kappa(A)\right) \frac{\|\delta A\|}{\|A\|} + \kappa(A)\sec(\theta) \frac{\|\delta b\|}{\|b\|}.$$

When the angle θ is not too large ($\kappa(A)\tan(\theta)$ small), we essentially have that small relative changes to A and b are only amplified by $\kappa(A)$. When $\tan(\theta)$ becomes larger, though, perturbations to A get amplified like the *squared* condition number. However, the book shows that the *quality* of the best fit, as measured by r, only changes like the condition number:

$$\frac{\|\delta r\|}{\|b\|} \le \frac{\|\delta b\|}{\|b\|} + 2\kappa(A)\frac{\|\delta A\|}{\|A\|}.$$

Which matters more — the change in the coefficients or the change in the quality of the fit — depends on the application in which the least squares problem arises.