2019-09-23

1 Error analysis for linear systems

We now discuss the sensitivity of linear systems to perturbations. This is relevant for two reasons:

1. Our standard recipe for getting an error bound for a computed solution in the presence of roundoff is to combine a backward error analysis (involving only features of the algorithm) with a sensitivity analysis (involving only features of the problem).

2. Even without rounding error, it is important to understand the sensitivity of a problem to the input variables if the inputs are in any way inaccurate (e.g. because they come from measurements).

We describe several different bounds that are useful in different contexts.

1.1 First-order analysis

We begin with a discussion of the first-order sensitivity analysis of the system

$$Ax = b.$$

Using our favored variational notation, we have the following relation between perturbations to A and b and perturbations to x:

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

or, assuming A is invertible,

$$\delta x = A^{-1}(\delta b - \delta A\, x).$$

We are interested in relative error, so we divide through by $\|x\|$:

$$\frac{\|\delta x\|}{\|x\|} \leq \frac{\|A^{-1}\delta b\|}{\|x\|} + \frac{\|A^{-1}\delta A\, x\|}{\|x\|}$$

The first term is bounded by

$$\frac{\|A^{-1}\delta b\|}{\|x\|} \leq \frac{\|A^{-1}\|\|\delta b\|}{\|x\|} = \kappa(A)\frac{\|\delta b\|}{\|A\|\|x\|} \leq \kappa(A)\frac{\|\delta b\|}{\|b\|}$$

and the second term is bounded by

$$\frac{\|A^{-1}\delta A\, x\|}{\|x\|} \leq \frac{\|A^{-1}\|\|\delta A\|\|x\|}{\|x\|} = \kappa(A)\frac{\|\delta A\|}{\|A\|}$$

Putting everything together, we have

$$\frac{\|\delta x\|}{\|x\|} \leq \kappa(A)\left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|b\|}\right),$$

That is, the relative error in x is (to first order) bounded by the condition number times the relative errors in A and b.

1.2 Beyond first order

What if we want to go beyond the first-order error analysis? Suppose that

$$Ax = b \quad \text{and} \quad \hat{A}\hat{x} = \hat{b}.$$

Then (analogous to our previous manipulations),

$$(\hat{A} - A)\hat{x} + A(\hat{x} - x) = \hat{b} - b$$

from which we have

$$\hat{x} - x = A^{-1}\left((\hat{b} - b) - E\hat{x}\right),$$

where $E \equiv \hat{A} - A$. Following the same algebra as before, we have

$$\frac{\|\hat{x} - x\|}{\|x\|} \leq \kappa(A)\left(\frac{\|E\|}{\|A\|}\frac{\|\hat{x}\|}{\|x\|} + \frac{\|\hat{b} - b\|}{\|b\|}\right).$$

Assuming $\|A^{-1}\|\|E\| < 1$, a little additional algebra (left as an exercise to the student) yields

$$\frac{\|\hat{x} - x\|}{\|x\|} \leq \frac{\kappa(A)}{1 - \|A^{-1}\|\|E\|}\left(\frac{\|E\|}{\|A\|} + \frac{\|\hat{b} - b\|}{\|b\|}\right).$$

Is this an important improvement on the first order bound? Perhaps not, for two reasons:

- One typically cares about the order of magnitude of possible error, not the exact bound, and
- The first-order bound and the "true" bound only disagree when both are probably pretty bad. When our house is in flames, our first priority is not to gauge whether the garage will catch as well; rather, we want to call the firefighters to put it out!

1.3 Componentwise relative bounds

What if we have more control over the perturbations than a simple bound on the norms? For example, we might have a componentwise perturbation bound

$$|\delta A| < \epsilon_A |A| \quad |\delta b| < \epsilon_b |b|,$$

and neglecting $O(\epsilon^2)$ terms, we obtain

$$|\delta x| \leq |A^{-1}| \left(\epsilon_b |b| + \epsilon_A |A||x|\right) \leq (\epsilon_b + \epsilon_A)|A^{-1}||A||x|.$$

Taking any vector norm such that $\| \, |x| \, \| = \|x\|$, we have

$$\|\delta x\| \leq (\epsilon + \epsilon') \| \, |A^{-1}| \, |A| \, \|.$$

The quantity $\kappa_{\text{rel}}(A) = \| \, |A^{-1}| \, |A| \, \|$ is the componentwise relative condition number (also known as the Skeel condition number).

1.4 Residual-based bounds

The *residual* for an approximate solution $\hat{x}$ to the equation $Ax = b$ is

$$r = A\hat{x} - b.$$

We can express much simpler error bounds in terms of the residual, using the relation

$$\hat{x} - x = A^{-1} r;$$

taking norms immediately gives

$$\|\hat{x} - x\| \leq \|A^{-1}\| \|r\|$$

and for any vector norm such that $\| \, |x| \, \| = \|x\|$, we have

$$\|\hat{x} - x\| \leq \| \, |A^{-1}| |r| \, \|.$$

Note that we can re-cast a residual error as a backward error on A via the relation

$$\left(A - \frac{r\hat{x}^T}{\|\hat{x}\|^2}\right)\hat{x} = b.$$

1.5 Shape of error

So far, we have only really discussed the *magnitude* of errors in a linear solve, but it is worth taking a moment to consider the *shape* of the errors as well. In particular, suppose that we want to solve $Ax = b$, and we have the singular value decomposition

$$A = U\Sigma V^T.$$

If $\sigma_n(A) \ll \sigma_1(A)$, then $\kappa_2 = \sigma_1/\sigma_n \gg q$, and we expect a large error. But is this the end of the story? Suppose that A satisfies

$$1 \geq \sigma_1 \geq \ldots \geq \sigma_k \geq C_1 \quad > \quad C_2 \geq \sigma_{k+1} \geq \ldots \geq \sigma_n > 0.$$

where $C_1 \gg C_2$. Let $r = A\hat{x} - b$, so that $Ae = r$ where $e = \hat{x} - x$. Then

$$e = A^{-1}r = V\Sigma^{-1}U^Tr = V\Sigma^{-1}\tilde{r} = \sum_{j=1}^{n} \frac{\tilde{r}_j}{\sigma_j} v_j.$$

where $\|\tilde{r}\| = \|U^Tr\| = \|r\|$. Split this as

$$e = e_1 + e_2$$

where we have a controlled piece

$$\|e_1\| = \left\| \sum_{j=1}^{k} \frac{\tilde{r}_j}{\sigma_j} v_j \right\| \leq \frac{\|r\|}{C_1}$$

and a piece that may be large,

$$e_2 = \sum_{j=k+1}^{n} \frac{\tilde{r}_j}{\sigma_j} v_j.$$

Hence, backward stability implies that the error consists of a small part and a part that lies in the "nearly-singular subspace" for the matrix.

2 Iterative refinement

If we have a solver for $\hat{A} = A + E$ with E small, then we can use *iterative refinement* to "clean up" the solution. The matrix $\hat{A}$ could come from finite precision Gaussian elimination of A, for example, or from some factorization of a nearby "easier" matrix. To get the refinement iteration, we take the equation

$$Ax = \hat{A}x - Ex = b, \tag{1}$$

and think of x as the fixed point for an iteration

$$\hat{A}x_{k+1} - Ex_k = b. \tag{2}$$

Note that this is the same as

$$\hat{A}x_{k+1} - (\hat{A} - A)x_k = b,$$

or

$$x_{k+1} = x_k + \hat{A}^{-1}(b - Ax_k).$$

If we subtract (1) from (2), we see

$$\hat{A}(x_{k+1} - x) - E(x_k - x) = 0,$$

or

$$x_{k+1} - x = \hat{A}^{-1}E(x_k - x).$$

Taking norms, we have

$$\|x_{k+1} - x\| \leq \|\hat{A}^{-1}E\|\|x_k - x\|.$$

Thus, if $\|\hat{A}^{-1}E\| < 1$, we are guaranteed that $x_k \to x$ as $k \to \infty$. In fact, this holds even if the backward error varies from step to step, as long as it satisfies some uniform bound that is less than one. At least, this is what happens in exact arithmetic.

In practice, the residual is usually computed with only finite precision, and so we would stop making progress at some point — usually at the point where we have a truly backward stable solution. In general, iterative refinement is mainly used when either the residual can be computed with extra precision or when the original solver suffers from relatively large backward error.