Section 2.3: Perturbing $A$.

Perturbing the coefficients matrix also leads to a measurable perturbation in the solution, even in exact arithmetic.

**Theorem 1.** If $A$ is non-singular, $b \neq 0$, and $x$ and $\hat{x} = x + \delta x$ be the solutions to $Ax = b$, respectively, $(A + \delta x)\hat{x} = b$ for some $\delta A$ a “noise” matrix, then

$$\frac{||\delta x||}{||\hat{x}||} \leq \kappa(A) \frac{||\delta A||}{||A||},$$

for any induced norm $|| \cdot ||$.

**Proof.** Subtracting the two equations and solving for $\delta x$ we get

$$\delta x = -A^{-1}\delta A\hat{x}.$$ 

and after taking norms, $||\delta x|| \leq ||A^{-1}|| \cdot ||\delta A|| \cdot ||\hat{x}||$. Switching $||\hat{x}||$ to the other side and writing

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

we obtain the desired conclusion. \qed

Finally, with a little work, one can put the result above together with the motivating inequality we showed about how $\kappa(A)$ relates the noise in the right-hand-side to the noise in the solution, to get

**Theorem 2.** (2.3.8) If $A$ is non-singular, $x$ and $\hat{x}$ satisfy $Ax = b$ and $\hat{A}\hat{x} = \hat{b}$, with $\hat{A} = A + \delta A$, $\hat{b} = b + \delta b$, and $\hat{x} = x + \delta x$, then

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

This inequality is a powerful way to link the condition number of the matrix as well as the relative sizes of the perturbations to the relative error in the solution. Remember, though, that all this is under an assumption of exact arithmetic. *Things can only get worse under finite-precision (that is, realistic) arithmetic.*

Section 4.1: The Singular Value Decomposition (SVD). And with this, we can move on to a deeper understanding of the way in which norms and condition numbers relate to matrices and affect linear systems: via the SVD.

We will start with a simple theorem, whose proof we will return to in a couple of weeks.
**Theorem 3.** Let $A$ be an $m \times n$ matrix, non-zero, with rank $r > 0$. Then there exist orthogonal matrices $U$, which is $m \times m$, and $V$, which is $n \times n$, as well as positive numbers $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ such that $A = U\Sigma V^T$, with $\Sigma$ an $m \times n$ “diagonal” matrix with form

$$
\Sigma = \begin{bmatrix}
\sigma_1 \\
\sigma_2 \\
\ddots \\
\sigma_r \\
0 \\
\ddots 
\end{bmatrix}.
$$

The above is called the **full Singular Value Decomposition (SVD)** of $A$.

Due to the fact that we can always change the signs of the columns of $U$ and $V$ simultaneously, the SVD is never unique.

One other way to see interpret the full SVD is to say that for any $m \times n$ matrix of rank $r > 0$, there exists an orthonormal basis of $\mathbb{R}^m$, call it $\{u_1, \ldots, u_m\}$ and an orthonormal basis of $\mathbb{R}^n$, call it $\{v_1, \ldots, v_n\}$, and $r$ positive numbers $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ such that

$$
Av_i = \begin{cases}
\sigma_i u_i, & 1 \leq i \leq r \\
0, & (r + 1) \leq i \leq n,
\end{cases} \quad A^T u_i = \begin{cases}
\sigma_i v_i, & 1 \leq i \leq r \\
0, & (r + 1) \leq i \leq m,
\end{cases}
$$

**Remark 1.** Think about this at home: $\{v_{r+1}, \ldots, v_n\}$ is an orthonormal basis for the nullspace of $A$; $\{u_{r+1}, \ldots, u_m\}$ is an orthonormal basis for the nullspace of $A^T$; $\{v_1, \ldots, v_r\}$ is an orthonormal basis for the range of $A^T$, and finally $\{u_1, \ldots, u_r\}$ is an orthonormal basis for the range of $A$.

Visually, the SVD illustrates the following phenomenon. If we think of the unit sphere in the Euclidean norm (a.k.a. the 2-norm) in $\mathbb{R}^n$, and pick an orthogonal system of coordinates for it, given by the semiaxes $\{v_1, \ldots, v_n\}$, then $A$ maps the unit sphere in $\mathbb{R}^n$ into a convex body in $\mathbb{R}^m$ with orthogonal coordinate system $\{u_1, \ldots, u_m\}$ by stretching the $i$th semiaxis by a factor $\sigma_i$ if $i \leq r$, and flattening all others to 0. For an illustration when $m = n = 2$, see picture below.

[Diagram showing the mapping of the unit sphere $S$ to a convex body $AS$ under the action of $A$.]

**The spectral norm and the SVD.** We have seen that the induced $\infty$-norm and the induced 1-norm for matrices can be expressed in relatively simple ways in terms of the matrix entries. This
will not be true for the induced 2-norm (a.k.a. the spectral norm, $||A|| = \max_{x \neq 0} \frac{||Ax||_2}{||x||_2}$). But the spectral norm of a matrix, and as a consequence, its condition number, is easily expressible in terms of the singular values. Moreover, this notion extends to the induced 2-norm for rectangular matrices (although in that case we cannot call it the spectral norm, as there will be no spectrum).

**Theorem 4.** For any matrix $A \in \mathbb{R}^{m \times n}$, define $||A||_2 = \max_{x \neq 0} \frac{||Ax||_2}{||x||_2}$. Then $||A||_2 = \sigma_1$, where $\sigma_1$ is the largest singular value of $A$.

**Remark 2.** Note that in the definition of the induced 2-norm above, $Ax \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$, but in both cases the 2-norms make sense and so does their ratio.

You may read the rigorous proof for this theorem in the textbook; we will offer some insight into what happens.

As the norms are multiplicative with respect to constants,

$$||A||_2 = \max_{x \neq 0} \frac{||Ax||_2}{||x||_2} = \max_{x \neq 0} \left| \frac{x}{||x||_2} \right|_2 = \max_{||y||_2=1} ||Ay||_2.$$

So, in other words, the induced 2-norm represents the maximum stretch that a unit vector in $\mathbb{R}^n$ undergoes when mapped to into a vector in $\mathbb{R}^m$ via $A$. As the picture above shows, that will happen precisely when we map $v_1$ into $\sigma_1 u_1$, via multiplication by $A$. So $||A||_2 = \sigma_1$. 