Trefethen and Bau (top of p.133) writes: “We could proceed by calculating Jacobians algebraically.” That could be another method. Maybe do this too. And then it would be nice to read Wedin’s paper. And then we need to create examples that stress these bounds and prove that the bounds are sharp.”)

The norm, $\| \cdot \|$, used here is the 2-norm.

We assume that A is full rank and we also assume that $A + \Delta A$ is full rank.

This second assumption is in general obtained by assuming that

$$\|\Delta A\| \leq \sigma_{\text{min}}(A)$$

or equivalently

$$\frac{\|\Delta A\|}{\|A\|} \cdot \kappa \leq 1.$$

This assumption is often written as A is “numerically full rank”. That is κ is not larger than the inverse of our numerical precision.

Quantities:

$$\kappa = \frac{\|A\|}{\sigma_{\text{min}}(A)}$$

$$\eta = \frac{\|A\| \; \|x\|}{\|Ax\|}$$

$$\cos\theta = \frac{\|Ax\|}{\|b\|} \qquad \tan\theta = \frac{\|b - Ax\|}{\|Ax\|}$$

Note that

$$1 \leq \eta \leq \kappa$$

$$\eta = \kappa \text{ when } x = v_{\text{min}}$$

$$\eta = 1 \text{ when } x = v_{\text{max}}$$

Things to know:

$$\|(A^T A)^{-1}\| = \frac{1}{\sigma_{\text{min}}(A)^2}$$

$$(A^T A)^{-1} A^T = A^+$$

$$\|(A^T A)^{-1} A^T\| = \frac{1}{\sigma_{\text{min}}(A)}$$

$$A(A^T A)^{-1} A^T = P_{\text{Range}(A)}$$

$$\|A(A^T A)^{-1} A^T\| = 1$$

$$I - A(A^TA)^{-1}A^T = P_{\text{Range}(A)^\perp}$$

$$\|I - A(A^TA)^{-1}A^T\| = 1$$

We start with linear system of equations first. We solve $Ax = b$. Now we perturb A and b and we get

$$(A + \Delta A)(x + \delta x) \quad = \quad (b + \delta b).$$

Now developing, removing the leading term $Ax = b$ (0th order), and the second order term (negligible), leads to

$$\Delta A \cdot x + A\delta x \quad = \quad \delta b,$$

where we should have used an $\approx$ or a $\mathcal{O}(\|\Delta A\| \cdot \|\delta x\|)$, but omitted it.

Solving for δx, we get

$$\delta x \quad = \quad -A^{-1}\Delta A \cdot x + A^{-1}\delta b,$$

And so

$$\frac{\|\delta x\|}{\|x\|} \quad \leq \quad \frac{\|\Delta A\|}{\|A\|}\left(\|A^{-1}\|\|A\|\right) + \frac{\|\delta b\|}{\|b\|}\left(\|A^{-1}\|\|A\|\right)\left(\frac{\|Ax\|}{\|A\|\|x\|}\right)$$

Here we used the fact that $b = Ax$. Note that we won't be able to use $b = Ax$ in the linear least squares case. (This is where θ will come into play.)

And so

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

The point of doing this is that we see that, for the linear least squares case, we shall find

$$\frac{\|\delta x\|}{\|x\|} \quad \leq \quad \frac{\|\Delta A\|}{\|A\|}\left(\kappa + \frac{\kappa^2 \tan\theta}{\eta}\right) + \frac{\|\delta b\|}{\|b\|}\left(\frac{\kappa}{\eta\cos(\theta)}\right).$$

We see that the formulae are consistent when we take $\theta = 0$ and so $\cos\theta = 1$ and $\tan\theta = 0$.

Now let us do the linear least squares case.

We solve $\min_x \|b - Ax\|$, the solution is given by

$$A^TAx = A^Tb.$$

Now we perturb A and b and we get

$$(A+\Delta A)^T(A+\Delta A)(x+\delta x) \quad = \quad (A+\Delta A)^T(b+\delta b),$$

Now developing, removing the leading term $A^TAx = A^Tb$ (0th order), and the second order terms (negligible), leads to

$$\Delta A^TAx + A^T\Delta Ax + A^TA\delta x \quad = \quad \Delta A^Tb + A^T\delta b,$$

where we should have used an $\approx$ or a $\mathcal{O}(\cdot)$, but omitted it.

Solving for δx, we get

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

And so

$$\frac{\|\delta x\|}{\|x\|} \quad \leq \quad (\frac{\|\Delta A\|}{\|A\|})\left(\frac{\|A\|^2}{\sigma_{\min}(A)^2}\frac{\|b-Ax\|}{\|Ax\|}\frac{\|Ax\|}{\|A\|\|x\|}+\frac{\|A\|}{\sigma_{\min}(A)}\right)+\frac{\|\delta b\|}{\|b\|}(\frac{\|A\|}{\sigma_{\min}(A)})\frac{\|b\|}{\|Ax\|}\frac{\|Ax\|}{\|A\|\|x\|}$$

And so

$$\frac{\|\delta x\|}{\|x\|} \quad \leq \quad (\frac{\|\Delta A\|}{\|A\|})\left(\kappa+\frac{\kappa^2\tan\theta}{\eta}\right)+\frac{\|\delta b\|}{\|b\|}\left(\kappa\frac{1}{\cos\theta}\frac{1}{\eta}\right)$$

We see that we find Trefethen and Bau terms, Theorem 18.1, page 131, for the second column.

To find Higham's expression, (Theorem 20.1, Equation (20.1), page 382,) we set $\frac{\|\Delta A\|}{\|A\|} = \varepsilon$ and $\frac{\|\delta b\|}{\|b\|} = \varepsilon$ so that

$$\frac{\|\delta x\|}{\|x\|} \quad \leq \quad \kappa\varepsilon\left(1+\frac{\kappa\tan\theta}{\eta}+\frac{1}{\cos\theta}\frac{1}{\eta}\right)$$

and then we use that

$$\frac{\|r\|}{\|A\|\|x\|} = \frac{\|r\|}{\|Ax\|}\frac{\|Ax\|}{\|A\|\|x\|} = \tan\theta\frac{1}{\eta}$$

$$\frac{\|b\|}{\|A\|\|x\|} = \frac{\|b\|}{\|Ax\|}\frac{\|Ax\|}{\|A\|\|x\|} = \frac{1}{\cos\theta}\frac{1}{\eta}$$

to get

$$\frac{\|\delta x\|}{\|x\|} \quad \leq \quad \kappa\varepsilon\left(1+\kappa\frac{\|r\|}{\|A\|\|x\|}+\frac{\|b\|}{\|A\|\|x\|}\right).$$

Higham's expression is

$$\frac{\|\delta x\|}{\|x\|} \quad \leq \quad \frac{\kappa\varepsilon}{1-\kappa\varepsilon}\left(2+(\kappa+1)\frac{\|r\|}{\|A\|\|x\|}\right).$$

This looks about right but this is not the same thing. Not sure where the issues are.

We are now interested in perturbation in y which is Ax. Now we perturb A and b and we get

$$(A+\Delta A)^T(A+\Delta A)(x+\delta x) \quad = \quad (A+\Delta A)^T(b+\delta b),$$

Do we look at $A(x+\delta x)$ or $(A+\Delta A)(x+\delta x)$????? I think we look at $(A+\Delta A)(x+\delta x)$ but it does not really matter much at the end, it's just a "2" or a "1" in front of the $\frac{1}{\eta}$.

At the first order, since $(y+\delta y)=(A+\Delta A)(x+\delta x)$, we get

$$\delta y \quad = \quad A\delta x + \Delta A x \tag{1}$$

Inserting our expression for δx found above, that we repeat below for conveniency,

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

we get

$$\delta y \quad = \quad A(A^TA)^{-1}\Delta A^T(b-Ax)+(I-A(A^TA)^{-1}A^T)\Delta A x+A(A^TA)^{-1}A^T\delta b$$

And so taking norms

$$\frac{\|\delta y\|}{\|y\|} \quad = \quad \frac{\|A\|}{\sigma_{\text{min}}(A)}\frac{\|\Delta A\|}{\|A\|}\frac{\|b-Ax\|}{\|Ax\|}+\frac{\|\Delta A\|}{\|A\|}\frac{\|A\|\|x\|}{\|Ax\|}+\frac{\|\delta b\|}{\|b\|}\frac{\|b\|}{\|Ax\|}$$

$$\frac{\|\delta y\|}{\|y\|} \quad = \quad \frac{\|\Delta A\|}{\|A\|}\left(\kappa\tan\theta+\frac{1}{\eta}\right)+\frac{\|\delta b\|}{\|b\|}\left(\frac{1}{\cos\theta}\right)$$

Looking at Trefethen and Bau terms, Theorem 18.1, page 131, for the first column, we read:

$$\frac{\|\delta y\|}{\|y\|} \quad = \quad \frac{\|\Delta A\|}{\|A\|}\frac{\kappa}{\cos\theta}+\frac{\|\delta b\|}{\|b\|}\left(\frac{1}{\cos\theta}\right)$$

This looks about right but this is not exactly the same thing for $\frac{\|\Delta A\|}{\|A\|}$. Not sure where the issues are.

Let us work on r since this is given in Higham. (Higham gives x and r in Theorem 20.1.)

At the first order, since $(r + \delta r) = (b + \delta b) - (A + \Delta A)(x + \delta x)$, we get

$$\delta r \quad = \quad \delta b - A \delta x - \Delta A x$$

We note that we can also get δr from δy with

$$\delta r \quad = \quad \delta b - \delta y$$

Inserting our expression for δx found above, that we repeat below for conveniency,

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

we get

$$\delta r \quad = \quad \delta b - A(A^T A)^{-1} \Delta A^T (r) + A(A^T A)^{-1} A^T \Delta A x - A(A^T A)^{-1} A^T \delta b - \Delta A x$$

And rearranging to get our projections, we get

$$\delta r \quad = \quad (I - A(A^T A)^{-1} A^T) \delta b - A(A^T A)^{-1} \Delta A^T (r) - (I - A(A^T A)^{-1} A^T) \Delta A x$$

And so taking norms, (be careful $\|\delta r\|$ is divided by $\|b\|$ for some reasons,)

$$\frac{\|\delta r\|}{\|b\|} \quad \leq \quad \frac{\|\delta b\|}{\|b\|} + \frac{\|A\|}{\sigma_{\text{min}}(A)} \frac{\|\Delta A\|}{\|A\|} \frac{\|r\|}{\|b\|} + \frac{\|\Delta A\|}{\|A\|} \frac{\|A\| \|x\|}{\|Ax\|} \frac{\|Ax\|}{\|b\|}$$

And so taking norms

$$\frac{\|\delta r\|}{\|r\|} \quad \leq \quad \frac{\|\Delta A\|}{\|A\|} (\kappa \sin\theta + \eta \cos\theta) + \frac{\|\delta b\|}{\|b\|}$$

To find Higham's expression, (Theorem 20.1, Equation (20.2), page 382,) we set $\frac{\|\Delta A\|}{\|A\|} = \varepsilon$ and $\frac{\|\delta b\|}{\|b\|} = \varepsilon$ and then we use that $\eta \leq \kappa$ and $\sin\theta \leq 1$ and $\cos\theta \leq 1$

$$\frac{\|\delta r\|}{\|b\|} \quad \leq \quad (1 + 2\kappa)\, \varepsilon \quad \checkmark$$

Note that since we need to bound something like $\sin\theta + \cos\theta$. We could use $\sqrt{2}$ instead of 2 and so we could get the stronger bound

$$\frac{\|\delta r\|}{\|b\|} \quad \leq \quad \left(1 + \sqrt{2}\kappa\right) \varepsilon$$

However the work of Wedin is not neglecting second order approximation so we cannot claim improving Wedin's work.

Look at the matrix from Higham.