Sensitivity of Linear Least Squares

Assume that $A \in \mathbb{R}^{m \times n}$ has full column rank. We know that the problem

$$\min_x \|Ax - b\|_2 \tag{1}$$

has a unique solution, say x_{LS}, which satisfies the (nonsingular) normal equations

$$A^T Ax = A^T b. \tag{2}$$

Here we will exploit the equivalence of (1) and (2) to investigate the sensitivity of (1). Note that how we *compute* x_{LS} is not at issue; the normal equations give us an explicit functional relationship from which we can *analyze the sensitivity* of x_{LS}.

As with $Ax = b$ we consider a perturbed problem

$$\min_x \|(A + tE)x(t) - (b + te)\|_2,$$

where t is a real parameter, and E and e are fixed. The normal equations here are

$$(A + tE)^T (A + tE)x(t) = (A + tE)^T (b + te). \tag{3}$$

If t is small enough, then $A + tE$ has full column rank, $(A + tE)^T (A + tE)$ is s.p.d., $x(t)$ is continuously differentiable and $x(0) = x_{LS}$. From Taylor's theorem

$$\|x_{LS} - x(t)\| \;=\; |t| \; \|\dot{x}(0)\| + \mathrm{O}(t^2), \tag{4}$$

so we differentiate (3) wrt t:

$$(A+tE)^T(A+tE)\dot{x}(t) + [(A+tE)^T E + E^T(A+tE)]x(t) = (A+tE)^T e + E^T(b+te)$$

and solve for $\dot{x}(t)|_{t=0}$:

$$\dot{x}(0) = (A^T A)^{-1}[A^T(e - Ex_{LS}) + E^T(b - Ax_{LS})]. \tag{5}$$

Taking $\|\cdot\| \equiv \|\cdot\|_2$ throughout gives submultiplicativity and provides the useful identities $\kappa(A) = \|A\|\|(A^T A)^{-1}A^T\|$ and $\kappa^2(A) = \|A\|^2\|(A^T A)^{-1}\|$. Now with $r \equiv b - Ax_{LS}$, (4) and (5) give

$$\begin{aligned} |t|\tfrac{\|\dot{x}(0)\|}{\|x_{LS}\|} \;&\leq\; |t|\left[\tfrac{\|(A^T A)^{-1}A^T(e - Ex_{LS})\|}{\|x_{LS}\|} + \tfrac{\|(A^T A)^{-1}E^T r\|}{\|x_{LS}\|}\right] + \mathrm{O}(t^2) \\ &\leq\; \kappa(A)\left[\tfrac{\|te\|}{\|A\|\|x_{LS}\|} + \tfrac{\|tE\|}{\|A\|}\right] + \tfrac{\|A\|^2\|t(A^T A)^{-1}E^T r\|}{\|A\|^2\|x_{LS}\|} + \mathrm{O}(t^2). \end{aligned}$$

If θ is the angle between b and $\mathrm{ColSp}(A)$), then $s \equiv \sin\theta = \|r\|/\|b\|$. With $c \equiv \cos\theta$ $\|Ax_{LS}\| = c\,\|b\| \leq \|A\|\|x_{LS}\|$, and for small t we have the (first order) bound

$$\frac{\|x_{LS} - x(t)\|}{\|x_{LS}\|} \;\lessapprox\; \tfrac{\|tE\|}{\|A\|}\ (\ \kappa(A) + \tan(\theta)\kappa^2(A)\ ) + \tfrac{\|te\|}{\|b\|}\ (\ \sec(\theta)\kappa(A)\ ).$$

Thus, a relative condition number for (1) is arguably

$$\kappa_{LS}(A, b) \;\equiv\; \kappa(A) \;+\; \tan(\theta)\;\kappa^2(A).$$

This generalizes the $Ax = b$ condition number (where $r = 0$), but becomes quadratic in $\kappa(A)$ as b points away from $\mathrm{ColSp}(A)$.