

Chapter 7

Inverse Problems and Regularization

AT THIS STAGE we have developed all the tools we need to deeply understand when $Ax = b$ has a solution, and, when not, to compute the next best thing, the minimizer of $\|b - Ax\|$. We shall start this chapter by drawing deep insight about these problems from the singular value decomposition, but then make a startling observation: for some important problems from applications, this optimal answer might not be the one that is most physically revealing. We conclude the course by seeing how the truncated singular value decomposition and regularization lead us to superior solutions.

7.1 The pseudoinverse

At this point you must feel like we have already extracted every possible insight from the equation $Ax = b$, but there is just a bit more to learn that will tie the entire theory together. Now is a good time to revisit our flowchart from Figure 4.2, extending it in Figure 7.1 to now cover the case where $Ax = b$ does not have a solution and we must be content with the best approximation that comes from solving the least squares problem

$$\min_{x \in \mathbb{R}^n} \|b - Ax\|.$$

When $\mathcal{N}(A)$ contains nonzero vectors, the solution to $Ax = b$ and $\min_{x \in \mathbb{R}^n} \|b - Ax\|$ will not be unique. From the infinite set of solutions we would like to pick out one particular vector that we can designate as *the best* solution. One natural option is to minimize $\|x\|$,

$$\min_{\substack{x \in \mathbb{R}^n \\ Ax=b}} \|x\|, \quad \min_{\substack{x \in \mathbb{R}^n \\ x \text{ minimizes } \|b - Ax\|}} \|x\|.$$

This criterion often has physical appeal, e.g., minimizing energy. The point of this section is to find this norm-minimizing solution.

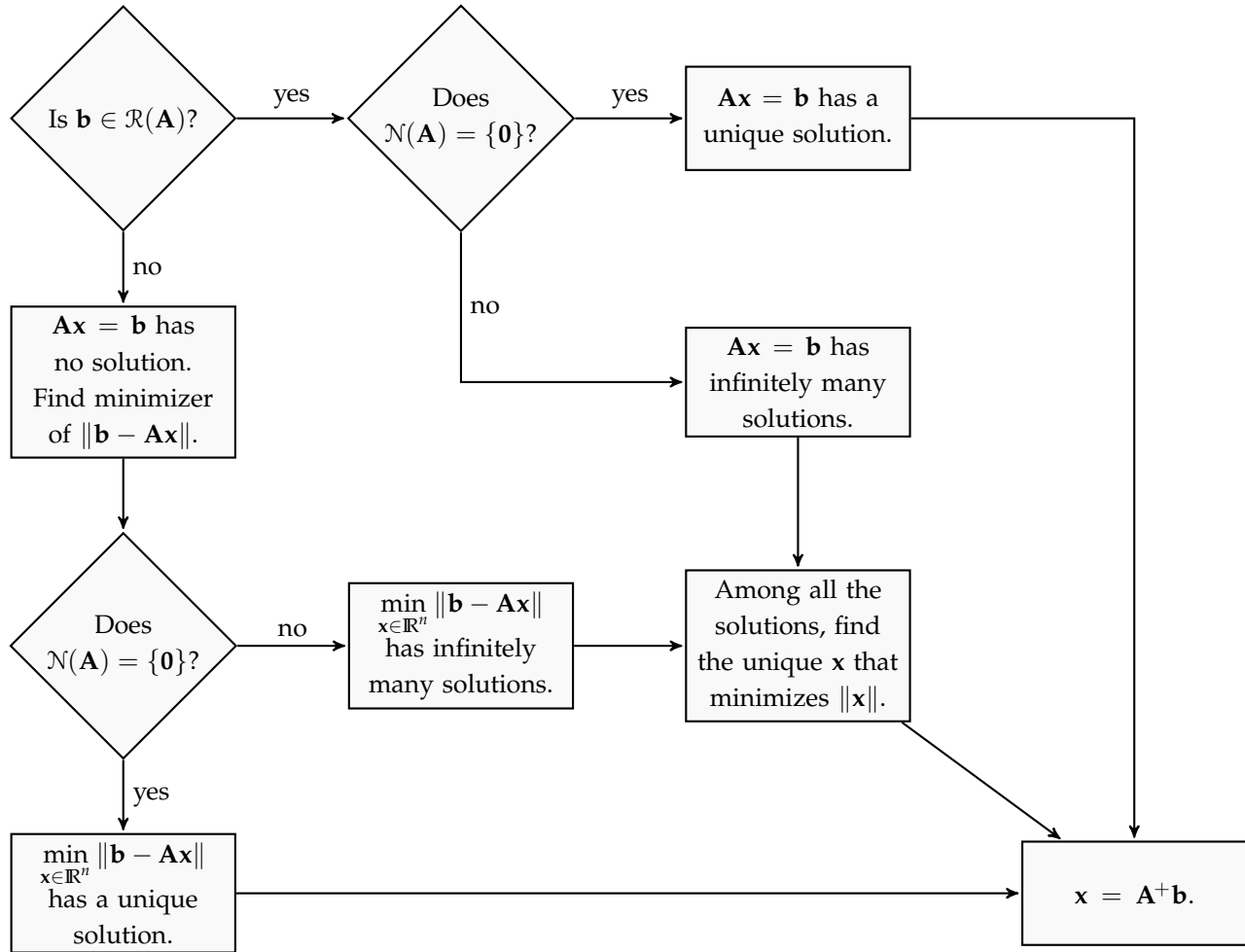


Figure 7.1: Revisiting the $\mathbf{Ax} = \mathbf{b}$ flowchart from Chapter 4, but now adding in the least squares problem. When infinite solutions exist, we pick the unique one of minimal norm. We shall see in the pages ahead that the unique smallest norm solution can always be written as $\mathbf{x} = \mathbf{A}^+\mathbf{b}$.

As we saw in Chapter 5, if $\mathcal{N}(\mathbf{A}) \neq \{\mathbf{0}\}$, the solution to $\mathbf{Ax} = \mathbf{b}$ or $\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - \mathbf{Ax}\|$ will not be unique. By the Fundamental Theorem of Linear Algebra, any such solution can be decomposed in the form

$$\mathbf{x} = \mathbf{x}_R + \mathbf{x}_N,$$

for $\mathbf{x}_R \in \mathcal{R}(\mathbf{A}^T)$ and $\mathbf{x}_N \in \mathcal{N}(\mathbf{A})$. This decomposition will be crucial to finding the minimal norm solution.

Now suppose two vectors \mathbf{x} and $\mathbf{y} \in \mathbb{R}^n$ satisfy

$$\mathbf{Ax} = \mathbf{b}, \quad \mathbf{Ay} = \mathbf{b}.$$

What can be said of their difference, $\mathbf{x} - \mathbf{y}$? Notice that

$$\mathbf{A}(\mathbf{x} - \mathbf{y}) = \mathbf{Ax} - \mathbf{Ay} = \mathbf{b} - \mathbf{b} = \mathbf{0},$$

so the difference $\mathbf{x} - \mathbf{y}$ must be contained in the null space of \mathbf{A} :

$$\mathbf{x} - \mathbf{y} \in \mathcal{N}(\mathbf{A}).$$

Recall that $\mathbf{x} \in \mathbb{R}^n = \mathcal{R}(\mathbf{A}^T) \oplus \mathcal{N}(\mathbf{A})$, and $\mathcal{R}(\mathbf{A}^T) \perp \mathcal{N}(\mathbf{A})$.

From this we conclude that all solutions to $\mathbf{Ax} = \mathbf{b}$ must have the same component in $\mathcal{R}(\mathbf{A}^T)$, and so can be written in the form

$$\mathbf{x} = \mathbf{x}_+ + \mathbf{x}_N,$$

where $\mathbf{x}_+ \in \mathcal{R}(\mathbf{A}^T)$ is a unique vector, and \mathbf{x}_N is any vector in $\mathcal{N}(\mathbf{A})$.

To see that the same argument applies to the least squares problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - \mathbf{Ax}\|,$$

suppose \mathbf{x} and \mathbf{y} $\in \mathbb{R}^n$ both solve this minimization, with

$$\|\mathbf{b} - \mathbf{Ax}\| = \|\mathbf{b} - \mathbf{Ay}\|.$$

Our discussion in Chapter 5 showed how to approach the least squares problem: decompose $\mathbf{b} = \mathbf{b}_R + \mathbf{b}_N$, where $\mathbf{b}_R \in \mathcal{R}(\mathbf{A})$ and $\mathbf{b}_N \in \mathcal{N}(\mathbf{A}^T)$, and then pick $\mathbf{x} \in \mathbb{R}^n$ so that $\mathbf{Ax} = \mathbf{b}_R$. Since \mathbf{y} is also a solution to the least squares problem, the same must apply: $\mathbf{Ay} = \mathbf{b}_R$. The difference of the solutions thus satisfies

$$\mathbf{A}(\mathbf{x} - \mathbf{y}) = \mathbf{b}_R - \mathbf{b}_R = \mathbf{0},$$

and again we have $\mathbf{x} - \mathbf{y} \in \mathcal{N}(\mathbf{A})$. Just as above, we conclude that \mathbf{x} and \mathbf{y} have the same component in $\mathcal{R}(\mathbf{A}^T)$; the only difference between these vectors lies in $\mathcal{N}(\mathbf{A})$. Hence, all solutions of the minimization problem also have the form

$$\mathbf{x} = \mathbf{x}_+ + \mathbf{x}_N,$$

where $\mathbf{x}_+ \in \mathcal{R}(\mathbf{A}^T)$ is a unique vector, and \mathbf{x}_N is any vector in $\mathcal{N}(\mathbf{A})$. Notice that $\mathbf{x}_+ \perp \mathbf{x}_N$ follows from $\mathcal{R}(\mathbf{A}^T) \perp \mathcal{N}(\mathbf{A})$, so the Pythagorean Theorem gives

$$\|\mathbf{x}\|^2 = \|\mathbf{x}_+\|^2 + \|\mathbf{x}_N\|^2.$$

To make $\|\mathbf{x}\|$ as small as possible, there is but one choice: set $\mathbf{x}_N = \mathbf{0}$.

Any solution to $\mathbf{Ax} = \mathbf{b}$ or $\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - \mathbf{Ax}\|$ has the form

$$\mathbf{x} = \mathbf{x}_+ + \mathbf{x}_N,$$

where $\mathbf{x}_+ \in \mathcal{R}(\mathbf{A})$ and $\mathbf{x}_N \in \mathcal{N}(\mathbf{A}^T)$. The vector \mathbf{x}_+ is unique.

A small example helps to reveal what is happening here.

Example 3 Consider the equation

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

The first equation gives $x_1 + x_2 = 2$, while the second is vacuous, $0 + 0 = 0$. Since $\mathbf{b} \in \mathcal{R}(\mathbf{A})$, there exist solutions to $\mathbf{Ax} = \mathbf{b}$, and since $\mathcal{N}(\mathbf{A})$ contains nonzero vectors, there are infinitely many of them. Each solution takes the general form

$$\mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \gamma \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

where

$$\mathbf{x}_+ = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \in \mathcal{R}(\mathbf{A}^T) = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

and

$$\gamma \begin{bmatrix} 1 \\ -1 \end{bmatrix} \in \mathcal{N}(\mathbf{A}) = \text{span}\left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}.$$

Since the norm of the solution satisfies

$$\|\mathbf{x}\|^2 = \left\| \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\|^2 + |\gamma|^2 \left\| \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\|^2,$$

the solution with smallest norm requires $\gamma = 0$, yielding $\mathbf{x} = \mathbf{x}_+$.

You might visualize how the picture accompanying the last example would generalize to three dimensions. If $\dim(\mathcal{N}(\mathbf{A})) = 2$, then the black arrow of solutions would become a plane, and there would still be a single point on that plane that was closest to $\mathbf{0}$. This closest point is $\mathbf{x}_+ \in \mathcal{R}(\mathbf{A}^T)$.

How DO WE FIND the minimum norm solution $\mathbf{x}_+ \in \mathcal{R}(\mathbf{A}^T)$? Suppose \mathbf{A} has rank r and singular value decomposition

$$\mathbf{A} = \sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^T.$$

Recall from the last chapter's SVD interpretation of the Fundamental Theorem that

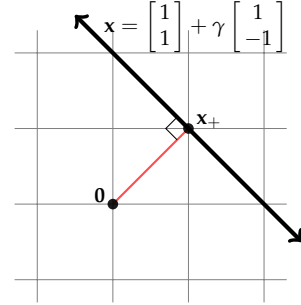
$$\mathcal{R}(\mathbf{A}^T) = \text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_r\}.$$

Thus we can always write $\mathbf{x}_+ \in \mathcal{R}(\mathbf{A}^T)$ as

$$\mathbf{x}_+ = c_1 \mathbf{v}_1 + \dots + c_r \mathbf{v}_r.$$

To find \mathbf{x}_+ , we need only find the coefficients c_1, \dots, c_r . If \mathbf{x}_+ is a solution to $\mathbf{Ax} = \mathbf{b}$, then we want find the coefficients that satisfy $\mathbf{Ax}_+ = \mathbf{b}$. Substitute the SVD for \mathbf{A} to find

$$\begin{aligned} \mathbf{b} = \mathbf{Ax}_+ &= \left(\sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^T \right) \left(\sum_{k=1}^r c_k \mathbf{v}_k \right) \\ &= \sum_{j=1}^r \sum_{k=1}^r \sigma_j c_k \mathbf{u}_j \mathbf{v}_j^T \mathbf{v}_k \end{aligned}$$



The black arrow denotes the set of all solutions \mathbf{x} to $\mathbf{Ax} = \mathbf{b}$. Among these vectors, $\mathbf{x}_+ \in \mathcal{R}(\mathbf{A}^T)$ is the unique solution that has smallest norm, i.e., is closest to $\mathbf{0}$.

Using the orthonormality of the right singular vectors,

$$\mathbf{v}_j^T \mathbf{v}_k = \begin{cases} 1, & j = k; \\ 0, & j \neq k, \end{cases}$$

we can collapse the double sum down to a single sum, finding

$$\mathbf{b} = \sum_{j=1}^r \sigma_j c_j \mathbf{u}_j.$$

Now to find the value of c_k , premultiply this last equation by \mathbf{u}_k^T and use the orthonormality of the left singular vectors,

$$\mathbf{u}_k^T \mathbf{u}_j = \begin{cases} 1, & k = j; \\ 0, & k \neq j, \end{cases}$$

to find

$$\mathbf{u}_k^T \mathbf{b} = \sum_{j=1}^r \sigma_j c_j \mathbf{u}_k^T \mathbf{u}_j = \sigma_k c_k.$$

Solve this last expression for c_k to obtain

$$c_k = \frac{\mathbf{u}_k^T \mathbf{b}}{\sigma_k},$$

from which we can build the norm-minimizing solution

$$\mathbf{x}_+ = \sum_{k=1}^r \left(\frac{\mathbf{u}_k^T \mathbf{b}}{\sigma_k} \right) \mathbf{v}_k.$$

It is helpful to think about \mathbf{x}_+ in the rearranged form

$$\mathbf{x}_+ = \left(\sum_{k=1}^r \frac{1}{\sigma_k} \mathbf{v}_k \mathbf{u}_k^T \right) \mathbf{b}. \quad (7.1)$$

We have thus found the minimum norm solution \mathbf{x}_+ for all cases where it is possible to satisfy $\mathbf{Ax} = \mathbf{b}$. What about the case where $\mathbf{b} \notin \mathcal{R}(\mathbf{A})$, where we can only solve the least squares problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - \mathbf{Ax}\|?$$

Recall that all solutions to this problem satisfy $\mathbf{Ax} = \mathbf{b}_R$, where $\mathbf{b} = \mathbf{b}_R + \mathbf{b}_N$ is a decomposition of \mathbf{b} into its components in $\mathcal{R}(\mathbf{A})$ and $\mathcal{N}(\mathbf{A}^T)$. Since the left singular vectors $\mathbf{u}_1, \dots, \mathbf{u}_r, \mathbf{u}_{r+1}, \dots, \mathbf{u}_m$ form an orthonormal basis for \mathbb{R}^m , we can write

$$\mathbf{b} = \left(\sum_{j=1}^m \mathbf{u}_j \mathbf{u}_j^T \right) \mathbf{b} = \sum_{j=1}^r (\mathbf{u}_j^T \mathbf{b}) \mathbf{u}_j + \sum_{j=r+1}^m (\mathbf{u}_j^T \mathbf{b}) \mathbf{u}_j,$$

$$\text{Recall } \mathbf{I} = \mathbf{U}\mathbf{U}^T = \sum_{j=1}^m \mathbf{u}_j \mathbf{u}_j^T.$$

where

$$\mathbf{b}_R = \sum_{j=1}^r (\mathbf{u}_j^T \mathbf{b}) \mathbf{u}_j \in \mathcal{R}(\mathbf{A}), \quad \mathbf{b}_N = \sum_{j=r+1}^m (\mathbf{u}_j^T \mathbf{b}) \mathbf{u}_j \in \mathcal{N}(\mathbf{A}^T),$$

since

$$\mathcal{R}(\mathbf{A}) = \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_r\}, \quad \mathcal{N}(\mathbf{A}^T) = \text{span}\{\mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}.$$

We seek the minimum norm solution $\mathbf{x} \in \mathcal{R}(\mathbf{A}^T)$ that solves $\mathbf{Ax} = \mathbf{b}_R$. Apply the same ideas that led to (7.1) for the $\mathbf{Ax} = \mathbf{b}$ case to get

$$\begin{aligned} \mathbf{x}_+ &= \left(\sum_{k=1}^r \frac{1}{\sigma_k} \mathbf{v}_k \mathbf{u}_k^T \right) \mathbf{b}_R = \left(\sum_{k=1}^r \frac{1}{\sigma_k} \mathbf{v}_k \mathbf{u}_k^T \right) \left(\sum_{j=1}^r (\mathbf{u}_j^T \mathbf{b}) \mathbf{u}_j \right) \\ &= \sum_{k=1}^r \sum_{j=1}^r \frac{\mathbf{u}_j^T \mathbf{b}}{\sigma_k} \mathbf{v}_k \mathbf{u}_k^T \mathbf{u}_j \\ &= \sum_{k=1}^r \frac{\mathbf{u}_k^T \mathbf{b}}{\sigma_k} \mathbf{v}_k, \end{aligned} \tag{7.2}$$

where we have used the orthonormality of the left singular vectors for the last step.

This last expression was built from \mathbf{b}_R , but we would really prefer to work directly with \mathbf{b} itself. Toward that end, notice that by the orthogonality of the left singular vectors,

$$\begin{aligned} \left(\sum_{k=1}^r \frac{1}{\sigma_k} \mathbf{v}_k \mathbf{u}_k^T \right) \mathbf{b} &= \left(\sum_{k=1}^r \frac{1}{\sigma_k} \mathbf{v}_k \mathbf{u}_k^T \right) \left(\sum_{j=1}^m (\mathbf{u}_j^T \mathbf{b}) \mathbf{u}_j \right) \\ &= \sum_{k=1}^r \sum_{j=1}^m \frac{\mathbf{u}_j^T \mathbf{b}}{\sigma_k} \mathbf{v}_k \mathbf{u}_k^T \mathbf{u}_j \\ &= \sum_{k=1}^r \frac{\mathbf{u}_k^T \mathbf{b}}{\sigma_k} \mathbf{v}_k, \end{aligned}$$

which is exactly the same as the expression (7.2) for \mathbf{x}_+ . Hence, we can directly compute

$$\mathbf{x}_+ = \left(\sum_{k=1}^r \frac{1}{\sigma_k} \mathbf{v}_k \mathbf{u}_k^T \right) \mathbf{b}, \tag{7.3}$$

which does not require us to first find \mathbf{b}_R .

THESE TEDIOUS CALCULATIONS have placed us on the verge of a major epiphany. Did you notice that the formula for the minimum norm solution to $\mathbf{Ax} = \mathbf{b}$ given in (7.1) agrees perfectly with the formula for the minimum norm solution to $\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - \mathbf{Ax}\|$ given in (7.3)? In both cases, one distinguished matrix,

$$\sum_{k=1}^r \frac{1}{\sigma_k} \mathbf{v}_k \mathbf{u}_k^T \in \mathbb{R}^{n \times m} \tag{7.4}$$

plays the same role as \mathbf{A}^{-1} in the $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ solution to $\mathbf{Ax} = \mathbf{b}$, expect this new matrix (7.4) works for any matrix: \mathbf{A} can have any rank, and any dimension. We give this distinguished matrix (7.4) a special name: the *pseudoinverse*.

Definition 24 Let $\mathbf{A} = \sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^T \in \mathbb{R}^{m \times n}$ be a rank- r matrix, written in its singular value decomposition. Then the pseudoinverse of \mathbf{A} is

$$\mathbf{A}^+ = \sum_{j=1}^r \frac{1}{\sigma_j} \mathbf{v}_j \mathbf{u}_j^T \in \mathbb{R}^{n \times m}.$$

Example 4 Let us revisit Example 3, where

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

has rank $r = 1$ and full singular value decomposition

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}.$$

The pseudoinverse is given by

$$\mathbf{A}^+ = \frac{1}{\sigma_1} \mathbf{v}_1 \mathbf{u}_1^T = \frac{1}{\sqrt{2}} \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} [1 \quad 0] = \begin{bmatrix} 1/2 & 0 \\ 1/2 & 0 \end{bmatrix}, \quad (7.5)$$

and so we compute

$$\mathbf{x}_+ = \mathbf{A}^+ \mathbf{b} = \begin{bmatrix} 1/2 & 0 \\ 1/2 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

just as we found in Example 4.

Let us summarize our major finding of this section.

Consider the linear system $\mathbf{Ax} = \mathbf{b}$ or the least squares problem $\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - \mathbf{Ax}\|$. Of all the (possibly infinite) solutions \mathbf{x} , the unique solution that minimizes $\|\mathbf{x}\|$ is given by

$$\mathbf{x}_+ = \mathbf{A}^+ \mathbf{b},$$

where \mathbf{A}^+ is the pseudoinverse,

$$\mathbf{A}^+ = \sum_{j=1}^r \frac{1}{\sigma_j} \mathbf{v}_j \mathbf{u}_j^T.$$

7.2 Ill-posed problems

The pseudoinverse gives us the ultimate solution to linear systems and least squares problems. But just before we begin to rest on our laurels, we should look at one last example.

Example 5 Suppose we make a small modification to Examples 3 and 4, giving

$$\begin{bmatrix} 1 & 1 \\ 0 & 0.01 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

This small change to the (2,2) entry makes \mathbf{A} an invertible matrix, and so

$$\mathbf{A}^+ = \mathbf{A}^{-1} = \begin{bmatrix} 1 & -100 \\ 0 & 100 \end{bmatrix},$$

quite different from the earlier pseudoinverse in (7.5). The unique solution to the new linear system is thus

$$\mathbf{x}_+ = \mathbf{A}^{-1}\mathbf{b} = \begin{bmatrix} 1 & -100 \\ 0 & 100 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

Normally we would just solve this problem and move on, but the previous examples should make us pause. The small change to one entry of \mathbf{A} has moved \mathbf{x}_+ from the old solution $[1, 1]^T$ to the new solution $[2, 0]^T$, a major change! Moreover, the new solution has much bigger norm, for

$$\left\| \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\| = \sqrt{2}, \quad \left\| \begin{bmatrix} 2 \\ 0 \end{bmatrix} \right\| = 2.$$

There was nothing special about the value 0.01 in the (2,2) position: any nonzero value has the same effect. For any $\varepsilon \neq 0$, the problem

$$\begin{bmatrix} 1 & 1 \\ 0 & \varepsilon \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

has the unique solution

$$\mathbf{x}_+ = \mathbf{A}^{-1}\mathbf{b} = \begin{bmatrix} 1 & -1/\varepsilon \\ 0 & 1/\varepsilon \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

So, an infinitesimal change to \mathbf{A} causes a significant change to the solution. The situation gets even worse if we also make a small change to \mathbf{b} as well. For example, for any $\varepsilon \neq 0$,

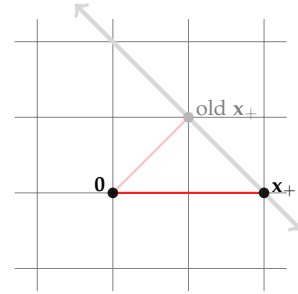
$$\begin{bmatrix} 1 & 1 \\ 0 & \varepsilon \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ \sqrt{\varepsilon} \end{bmatrix}$$

has the unique solution

$$\mathbf{x}_+ = \mathbf{A}^{-1}\mathbf{b} = \begin{bmatrix} 1 & -1/\varepsilon \\ 0 & 1/\varepsilon \end{bmatrix} \begin{bmatrix} 2 \\ \sqrt{\varepsilon} \end{bmatrix} = \begin{bmatrix} 2 - 1/\sqrt{\varepsilon} \\ 1/\sqrt{\varepsilon} \end{bmatrix}.$$

When ε is a small nonzero number, we have made a very minor change to \mathbf{A} and \mathbf{b} that results in a solution \mathbf{x}_+ with enormous entries. For example, when $\varepsilon = 0.01$ as above, this equation with perturbed \mathbf{b} gives

$$\mathbf{x}_+ = \begin{bmatrix} -8 \\ 10 \end{bmatrix}.$$



The solution \mathbf{x}_+ to the new, slightly perturbed problem is quite different from the original solution found in Example 3 (and larger in norm).

Reduce the size of the perturbation even more, to $\varepsilon = 0.0001 = 10^{-4}$. Now

$$\mathbf{x}_+ = \begin{bmatrix} -98 \\ 100 \end{bmatrix}.$$

This is a strange thing! As ε gets smaller, \mathbf{A} and \mathbf{b} get closer and closer to their unperturbed counterparts in Example 3, yet the solution \mathbf{x}_+ gets farther and farther from the unperturbed solution $[1, 1]^T$.

Here is a crucial observation: in all these cases, the new solution \mathbf{x}_+ is actually in that larger space of all solutions for the unperturbed problem from Example 3,

$$\begin{bmatrix} 2 - 1/\sqrt{\varepsilon} \\ 1/\sqrt{\varepsilon} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + (1 + 1/\sqrt{\varepsilon}) \begin{bmatrix} 1 \\ -1 \end{bmatrix} \in \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \gamma \begin{bmatrix} 1 \\ -1 \end{bmatrix} : \gamma \in \mathbb{C} \right\},$$

even though the new solution \mathbf{x}_+ is far from the old solution of minimal norm. This is a particularly clean outcome due to the structure of our example, but the general pattern is seen in more sophisticated settings.

7.2.1 Application: deblurring

7.3 Truncated SVD solutions

Look back to Example 5. Small perturbations ε to \mathbf{A} and \mathbf{b} caused the solution \mathbf{x}_+ to change by large amounts. To resolve this difficulty, your natural reaction might be: “Can’t we just ignore the ε entries in \mathbf{A} and \mathbf{b} ?” This shows good instinct, but we would like a more systematic way to identify and neglect the “small” entries in a system. Thankfully the singular value decomposition provides just the right tool for the job. The basic idea is simple: instead of using all the terms in the SVD to compute the pseudoinverse,

$$\mathbf{A}^+ = \sum_{j=1}^r \frac{1}{\sigma_j} \mathbf{v}_j \mathbf{u}_j^T,$$

simply leave out the terms with small σ_j (and hence large $1/\sigma_j$):

$$\mathbf{A}_k^+ = \sum_{j=1}^k \frac{1}{\sigma_j} \mathbf{v}_j \mathbf{u}_j^T$$

All we did here was change the upper limit on the sum from r to k .

for some $k < r$. Let us explain why this might be a good idea.

Recall that in Section 8.11, we computed optimal rank- k approximations to the rank- r matrix

$$\mathbf{A} = \sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^T$$

by truncating the singular value decomposition to the first k terms in the sum,

$$\mathbf{A}_k = \sum_{j=1}^k \sigma_j \mathbf{u}_j \mathbf{v}_j^T, \quad k \leq r.$$

As we saw in Chapter 8, the accuracy of this approximation is controlled by the first neglected singular value:

$$\|\mathbf{A} - \mathbf{A}_k\| = \left\| \sum_{j=k+1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^T \right\| = \sigma_{k+1}.$$

SUPPOSE WE WISH TO SOLVE the least squares problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|, \quad (7.6)$$

where \mathbf{A} has some very small singular values that make the pseudoinverse solution derived in Section 7.1,

$$\mathbf{x}_+ = \mathbf{A}^+ \mathbf{b} = \sum_{j=1}^r \frac{1}{\sigma_j} \mathbf{v}_j \mathbf{u}_j^T \mathbf{b},$$

very large in norm. In fact, using the orthonormality of the right singular vectors and the Pythagorean Theorem, we can compute

$$\|\mathbf{x}_+\|^2 = \sum_{j=1}^r \frac{1}{\sigma_j^2} |\mathbf{u}_j^T \mathbf{b}|^2. \quad (7.7)$$

Now if we replace \mathbf{A} in the least squares problem (7.6) with the truncated SVD \mathbf{A}_k ,

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - \mathbf{A}_k \mathbf{x}\|, \quad (7.8)$$

the solution changes to

$$\mathbf{x}_k = \mathbf{A}_k^+ \mathbf{b} = \sum_{j=1}^k \frac{1}{\sigma_j} \mathbf{v}_j \mathbf{u}_j^T \mathbf{b}$$

having smaller norm, since

$$\|\mathbf{x}_k\|^2 = \sum_{j=1}^k \frac{1}{\sigma_j^2} |\mathbf{u}_j^T \mathbf{b}|^2. \quad (7.9)$$

By picking k so that σ_k is not too small, we can prevent $\|\mathbf{x}_k\|$ from being offensively large. But how well does \mathbf{x}_k satisfy the original least squares problem we really want to solve? With the help of the SVD, we can readily check:

$$\begin{aligned} \|\mathbf{b} - \mathbf{A}\mathbf{x}_k\| &= \left\| \mathbf{b} - \left(\sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^T \right) \left(\sum_{\ell=1}^k \frac{1}{\sigma_\ell} \mathbf{v}_\ell \mathbf{u}_\ell^T \mathbf{b} \right) \right\| \\ &= \left\| \mathbf{b} - \sum_{j=1}^r \sum_{\ell=1}^k \frac{\sigma_j}{\sigma_\ell} \mathbf{u}_j \mathbf{v}_j^T \mathbf{v}_\ell \mathbf{u}_\ell^T \mathbf{b} \right\| \\ &= \left\| \mathbf{b} - \sum_{j=1}^k \mathbf{u}_j \mathbf{u}_j^T \mathbf{b} \right\| \\ &= \left\| \sum_{j=k+1}^m \mathbf{u}_j \mathbf{u}_j^T \mathbf{b} \right\|. \end{aligned}$$

Recall that $\mathbf{b} = \sum_{j=1}^n \mathbf{u}_j \mathbf{u}_j^T \mathbf{b}$.

Compare this to the error from the least squares problem when the pseudoinverse solution is used:

$$\|\mathbf{b} - \mathbf{Ax}_+\| = \left\| \sum_{j=r+1}^m \mathbf{u}_j \mathbf{u}_j^T \mathbf{b} \right\|.$$

The residual $\|\mathbf{b} - \mathbf{Ax}_k\|$ is larger, but perhaps not by all that much:

$$\|\mathbf{b} - \mathbf{Ax}_k\|^2 - \|\mathbf{b} - \mathbf{Ax}_+\|^2 = \sum_{j=k+1}^r |\mathbf{u}_j^T \mathbf{b}|^2.$$

In summary, using the truncated SVD can greatly reduce $\|\mathbf{x}_k\|$ in (7.7) by omitting the $\mathbf{u}_j^T \mathbf{b} / \sigma_j$ terms for small σ_j , but the increase in the norm of the residual, $\|\mathbf{b} - \mathbf{Ax}_k\|$ is comparatively modest, only adding terms like $\mathbf{u}_j^T \mathbf{b}$.

Compare the pseudoinverse solution

$$\mathbf{x}_+ = \sum_{j=1}^r \frac{1}{\sigma_j} \mathbf{v}_j \mathbf{u}_j^T \mathbf{b} \quad (7.10)$$

to the truncated SVD approximation

$$\mathbf{x}_k = \sum_{j=1}^k \frac{1}{\sigma_j} \mathbf{v}_j \mathbf{u}_j^T \mathbf{b} \quad (7.11)$$

for the least squares problem $\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - \mathbf{Ax}\|$:

- The norm of the truncated SVD *solution* is smaller:

$$\|\mathbf{x}_k\|^2 = \|\mathbf{x}_+\|^2 - \sum_{j=k+1}^r \frac{1}{\sigma_j^2} |\mathbf{u}_j^T \mathbf{b}|^2.$$

- The norm of the truncated SVD *residual* is larger:

$$\|\mathbf{b} - \mathbf{Ax}_k\|^2 = \|\mathbf{b} - \mathbf{Ax}_+\|^2 + \sum_{j=k+1}^r |\mathbf{u}_j^T \mathbf{b}|^2.$$

In applications where $\sigma_{k+1}, \dots, \sigma_r$ are small, the reduction in the norm of the solution can yield much more physically realistic answer by removing the artificial but overwhelming effects of noisy data, while the modest increase in the residual is not such a big concern.

7.4 Regularization

The truncated SVD has great appeal: one must appreciate the simplicity and potency of this approach. However, it requires us to first

compute the singular value decomposition of \mathbf{A} . When \mathbf{A} is a large matrix (often “sparse,” meaning most entries are zero), one cannot afford to compute the SVD of \mathbf{A} . Another approach is equally intuitive but computationally more appealing: *regularization*.

The boxed paragraph at the end of the last section highlights the two conflicting tensions that arise when solving ill-posed problems: we seek to make the least-squares residual $\|\mathbf{b} - \mathbf{Ax}\|$ as small as possible, while also controlling the size of the solution $\|\mathbf{x}\|$. The fundamental problem is that the \mathbf{x} that minimizes $\|\mathbf{b} - \mathbf{Ax}\|$ often gives large $\|\mathbf{x}\|$. We are willing to accept a *suboptimal* \mathbf{x} that gives a *slightly* larger $\|\mathbf{b} - \mathbf{Ax}\|$ but a *much smaller* $\|\mathbf{x}\|$.

Why not combine these two goals into one optimization problem? Replace the usual least squares problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - \mathbf{Ax}\|$$

with the *penalized* problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - \mathbf{Ax}\|^2 + \lambda^2 \|\mathbf{x}\|^2 \quad (7.12)$$

for some choice of the *regularization parameter* λ . Two fundamental questions arise:

- How should one choose λ ?
- How does one find the optimal \mathbf{x} in (7.12)?

We shall tackle these questions in reverse order.

7.4.1 Solving regularized least squares problems

Suppose $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $m \geq n$, and consider the alternative least squares problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \left\| \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{A} \\ \lambda \mathbf{I} \end{bmatrix} \mathbf{x} \right\| \quad (7.13)$$

involving the $(m+n) \times n$ matrix

$$\mathbf{A}_\lambda = \begin{bmatrix} \mathbf{A} \\ \lambda \mathbf{I} \end{bmatrix}.$$

Equation (7.13) has a neat property. Square the norm of the residual:

$$\begin{aligned} \left\| \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{A} \\ \lambda \mathbf{I} \end{bmatrix} \mathbf{x} \right\|^2 &= \left(\begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{A} \\ \lambda \mathbf{I} \end{bmatrix} \mathbf{x} \right)^T \left(\begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{A} \\ \lambda \mathbf{I} \end{bmatrix} \mathbf{x} \right) \\ &= [\mathbf{b}^T \quad \mathbf{0}] \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} - \mathbf{x}^T [\mathbf{A}^T \quad \lambda \mathbf{I}] \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} - [\mathbf{b}^T \quad \mathbf{0}] \begin{bmatrix} \mathbf{A} \\ \lambda \mathbf{I} \end{bmatrix} \mathbf{x} + \mathbf{x}^T [\mathbf{A}^T \quad \lambda \mathbf{I}] \begin{bmatrix} \mathbf{A} \\ \lambda \mathbf{I} \end{bmatrix} \mathbf{x} \\ &= \mathbf{b}^T \mathbf{b} - \mathbf{x}^T \mathbf{A}^T \mathbf{b} - \mathbf{b}^T \mathbf{Ax} + \mathbf{x}^T (\mathbf{A}^T \mathbf{A} + \lambda^2 \mathbf{I}) \mathbf{x} \end{aligned}$$

$$\begin{aligned}
&= (\mathbf{b} - \mathbf{Ax})^T (\mathbf{b} - \mathbf{Ax}) + \lambda^2 \mathbf{x}^T \mathbf{x} \\
&= \|\mathbf{b} - \mathbf{Ax}\|^2 + \lambda^2 \|\mathbf{x}\|^2.
\end{aligned}$$

Thus, remarkably, we can solve the penalized problem (7.12) by solving the standard least squares problem (7.13):

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - \mathbf{Ax}\|^2 + \lambda^2 \|\mathbf{x}\|^2 = \min_{\mathbf{x} \in \mathbb{R}^n} \left\| \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{A} \\ \lambda \mathbf{I} \end{bmatrix} \mathbf{x} \right\|.$$

If one of these problems has a unique solution, the other does as well. We shall compute that solution \mathbf{x}_λ using the conventional least squares problem (7.13).

Take a few moments to think about \mathbf{A}_λ . To understand if (7.13) has a unique solution, we need to understand the rank and null space of \mathbf{A}_λ . What is $\text{rank}(\mathbf{A}_\lambda)$? You can determine the rank in several ways.

Consider the $m = 3, n = 2$ case:

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{bmatrix}, \quad \mathbf{A}_\lambda = \begin{bmatrix} \mathbf{A} \\ \lambda \mathbf{I} \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \\ \lambda & 0 \\ 0 & \lambda \end{bmatrix}.$$

Regardless of \mathbf{A} , the $n = 2$ columns of \mathbf{A}_λ must be linearly independent due to the bottom 2×2 block $\lambda \mathbf{I}$, for all $\lambda \neq 0$. Thus the column space $\mathcal{R}(\mathbf{A})$ must have dimension $n = 2$, so $\text{rank}(\mathbf{A}) = 2$. Since

$$\begin{aligned}
\dim(\mathcal{N}(\mathbf{A}_\lambda)) &= n - \dim(\mathcal{R}(\mathbf{A}_\lambda)) \\
&= n - \text{rank}(\mathbf{A}_\lambda),
\end{aligned}$$

we conclude that $\dim(\mathcal{N}(\mathbf{A}_\lambda)) = 0$, and so $\mathcal{N}(\mathbf{A}_\lambda) = \{\mathbf{0}\}$. The same holds for general $\mathbf{A} \in \mathbb{R}^{m \times n}$. Since $\dim(\mathcal{N}(\mathbf{A}_\lambda)) = \{\mathbf{0}\}$, as we see in the flowchart in Figure 7.1, the least squares problem (7.13) will have a unique solution, and thus too our problem of interest (7.12). Let us summarize where we now stand.

This is true even if $\mathbf{A} = \mathbf{0}$!

You might also study the rank of \mathbf{A}_λ by computing its singular values. Recall that the singular values are square roots of the eigenvalues of $\mathbf{A}_\lambda^T \mathbf{A}_\lambda$. Since

$$\mathbf{A}_\lambda^T \mathbf{A}_\lambda = \mathbf{A}^T \mathbf{A} + \lambda^2 \mathbf{I},$$

you can see that the eigenvalues of $\mathbf{A}_\lambda^T \mathbf{A}_\lambda$ are just the eigenvalues of $\mathbf{A}^T \mathbf{A}$, plus λ^2 : for if

$$\mathbf{A}^T \mathbf{A} \mathbf{v}_j = \sigma_j^2 \mathbf{v}_j,$$

then

$$\mathbf{A}_\lambda^T \mathbf{A}_\lambda \mathbf{v}_j = (\mathbf{A}^T \mathbf{A} + \lambda^2 \mathbf{I}) \mathbf{v}_j = (\sigma_j^2 + \lambda^2) \mathbf{v}_j.$$

Thus we see that for any nonzero λ ,

j th singular value of $\mathbf{A}_\lambda = \sqrt{\sigma_j^2 + \lambda^2} > 0$,

even if $\sigma_j = 0$. Since \mathbf{A}_λ has n nonzero singular values, its rank must be n . The right singular vectors \mathbf{v}_j of \mathbf{A}_λ are also the right singular vectors of \mathbf{A} .

For any $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\lambda \neq 0$, the regularized matrix \mathbf{A}_λ satisfies

$$\text{rank}(\mathbf{A}_\lambda) = n$$

and hence $\mathcal{N}(\mathbf{A}_\lambda) = \{\mathbf{0}\}$. Thus the least squares problem (7.13) has the unique solution

$$\begin{aligned} \mathbf{x}_\lambda &= (\mathbf{A}_\lambda^T \mathbf{A}_\lambda)^{-1} \mathbf{A}_\lambda^T \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \\ &= (\mathbf{A}^T \mathbf{A} + \lambda^2 \mathbf{I})^{-1} \mathbf{A}^T \mathbf{b}, \end{aligned}$$

which is also the unique solution to the regularized problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \left\| \begin{bmatrix} \mathbf{A} \\ \lambda \mathbf{I} \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \right\|.$$

In this last box we have the formula for the solution

$$\mathbf{x}_\lambda = (\mathbf{A}^T \mathbf{A} + \lambda^2 \mathbf{I})^{-1} \mathbf{A}^T \mathbf{b}.$$

In terms of the SVD of \mathbf{A} (setting $\sigma_j = 0$ for $j > r$), we have

$$\mathbf{A}^T \mathbf{A} + \lambda^2 \mathbf{I} = \sum_{j=1}^n \sigma_j^2 \mathbf{v}_j \mathbf{v}_j^T + \lambda^2 \sum_{j=1}^n \mathbf{v}_j \mathbf{v}_j^T = \sum_{j=1}^n (\sigma_j^2 + \lambda^2) \mathbf{v}_j \mathbf{v}_j^T,$$

which can be readily inverted:

$$(\mathbf{A}^T \mathbf{A} + \lambda^2 \mathbf{I})^{-1} = \sum_{j=1}^n \frac{1}{\sigma_j^2 + \lambda^2} \mathbf{v}_j \mathbf{v}_j^T.$$

From this we can compute

$$\begin{aligned} \mathbf{x}_\lambda &= (\mathbf{A}^T \mathbf{A} + \lambda^2 \mathbf{I})^{-1} \mathbf{A}^T \mathbf{b} \\ &= \left(\sum_{j=1}^n \frac{1}{\sigma_j^2 + \lambda^2} \mathbf{v}_j \mathbf{v}_j^T \right) \left(\sum_{\ell=1}^r \sigma_\ell \mathbf{v}_\ell \mathbf{u}_\ell^T \right) \mathbf{b} \\ &= \sum_{j=1}^r \frac{\sigma_j}{\sigma_j^2 + \lambda^2} \mathbf{v}_j \mathbf{u}_j^T \mathbf{b}, \end{aligned}$$

using orthogonality of the singular vectors, as usual. We summarize:

The unique solution to the regularized least squares problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|^2 + \lambda^2 \|\mathbf{x}\|^2$$

is given by

$$\mathbf{x}_\lambda = \sum_{j=1}^r \frac{\sigma_j}{\sigma_j^2 + \lambda^2} \mathbf{v}_j \mathbf{u}_j^T \mathbf{b}. \quad (7.14)$$

Contrast three formulas:

$$\text{pseudoinverse solution} \quad \mathbf{x}_+ = \sum_{j=1}^r \frac{1}{\sigma_j} \mathbf{v}_j \mathbf{u}_j^T \mathbf{b}$$

$$\text{truncated SVD solution} \quad \mathbf{x}_k = \sum_{j=1}^k \frac{1}{\sigma_j} \mathbf{v}_j \mathbf{u}_j^T \mathbf{b}$$

$$\text{regularized solution} \quad \mathbf{x}_\lambda = \sum_{j=1}^r \frac{\sigma_j}{\sigma_j^2 + \lambda^2} \mathbf{v}_j \mathbf{u}_j^T \mathbf{b}$$

All three of these “solutions” involve the terms $\mathbf{v}_j \mathbf{u}_j^T \mathbf{b}$, but different weight is put on them by the different formulas. When the regularization term is small (or zero), the $\lambda^2 \|\mathbf{x}\|^2$ term does not much influence the least squares problem, and \mathbf{x}_λ will be quite close to the pseudoinverse solution \mathbf{x}_+ . Large λ values place much greater influence on the $\lambda^2 \|\mathbf{x}\|^2$ term, and in the limit as $\lambda \rightarrow \infty$ we see that $\mathbf{x}_\lambda \rightarrow \mathbf{0}$.

7.4.2 Selecting the regularization parameter

How then should one select the regularization parameter λ to yield the best results? One seeks to strike a perfect balance between keeping $\|\mathbf{x}_\lambda\|$ at a moderate size while making $\|\mathbf{b} - \mathbf{A}\mathbf{x}_\lambda\|$ as small as possible. One way to select λ is to create a plot with $\log \|\mathbf{b} - \mathbf{A}\mathbf{x}_\lambda\|$ on the horizontal axis, and $\log \|\mathbf{x}_\lambda\|$ on the vertical axis, sampled

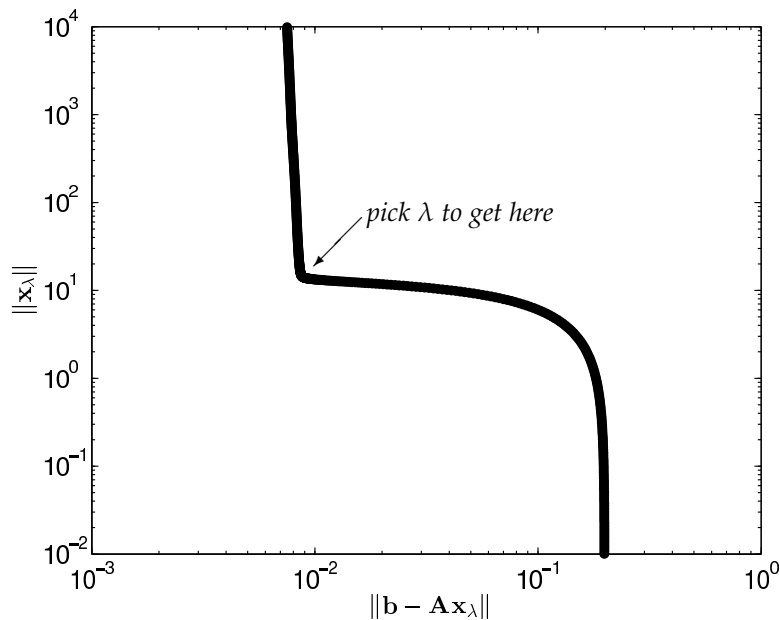


Figure 7.2: An illustration of the relationship between $\|\mathbf{b} - \mathbf{A}\mathbf{x}_\lambda\|$ and $\|\mathbf{x}_\lambda\|$ as λ changes over orders of magnitude. (In this particular plot, from a UPC barcode recovery problem, λ varies from roughly 10^{-8} (top left) to 10^0 (bottom right).) To make $\|\mathbf{x}_\lambda\|$ as small as possible without increasing $\|\mathbf{b} - \mathbf{A}\mathbf{x}_\lambda\|$ too much, choose a λ value that corresponds to a $(\|\mathbf{b} - \mathbf{A}\mathbf{x}_\lambda\|, \|\mathbf{x}_\lambda\|)$ pair at the corner of the “L.”

over a wide range of λ values (varying over orders of magnitude). Often this plot shows a distinct bend, as seen in one example in Figure 7.2. For many applications, the best choice for λ will yield values of $\|\mathbf{b} - \mathbf{Ax}_\lambda\|$ and $\|\mathbf{x}_\lambda\|$ that land at the sharp bend in this “L curve.”

For many more details of regularization problem, ranging from applications to algorithms, we recommend the excellent introductory book by Hansen.¹

1