2. Solving linear systems, Truncated SVD

Consider the innocent looking linear system,

$$Ax = b,$$

$A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and the problem is to find $x \in \mathbb{R}^n$.

The SVD of A gives us tools to completely analyse the problem. Let

$$A = UDV^T,$$

i.e.,

$$UDV^Tx = b$$

Remembering that $U \in \mathbb{R}^{m \times m}$ is orthogonal, by multiplying by U^T from the left, we get

$$\underbrace{U^TU}_{I} \, D \underbrace{V^Tx}_{\tilde{x}} = \underbrace{U^Tb}_{\tilde{b}}$$

or

$$D\tilde{x} = \tilde{b}$$

Consider different cases:

<u>I $\quad m \le n$.</u> Assume that

$$d_1 \ge d_2 \ge \dots \ge d_r > d_{r+1} = \dots = d_m = 0$$

$$D\tilde{x} = \left[\begin{array}{ccc|c} d_1 & & & \\ & d_2 & & 0 \\ & & \ddots \; d_r & \\ \hline & 0 & & 0 \end{array}\right] \begin{bmatrix} \tilde{x}_1 \\ \vdots \\ \vdots \\ \tilde{x}_n \end{bmatrix} = \begin{bmatrix} \tilde{b}_1 \\ \vdots \\ \tilde{b}_m \end{bmatrix}$$

We see that the solution is

$$\tilde{x}_j = \frac{1}{d_j}\tilde{b}_j \quad , 1 \le j \le r, \qquad \tilde{x}_j \text{ arbitrary, if } j \ge r.$$

Writing

$$V = [\underbrace{V_1}_{r} \mid \underbrace{V_2}_{n-r}] \quad , \quad U = [\underbrace{U_1}_{r} \mid \underbrace{U_2}_{m-r}]$$

the general solution is

$$x = \sum_{j=1}^{r} \frac{1}{d_j}\tilde{b}_j \, \tilde{v}_j + \underbrace{\sum_{j=r+1}^{n} \tilde{x}_j v_j}_{= x_0}$$

$$= V_1 \begin{bmatrix} 1/d_1 & & \\ & \ddots & \\ & & 1/d_r \end{bmatrix} U_1^T b + x_0 \quad , \quad x_0 \in N(A).$$

We have special cases:

$$m = n \quad \text{and} \quad r = m = n.$$

Then the solution is unique, and

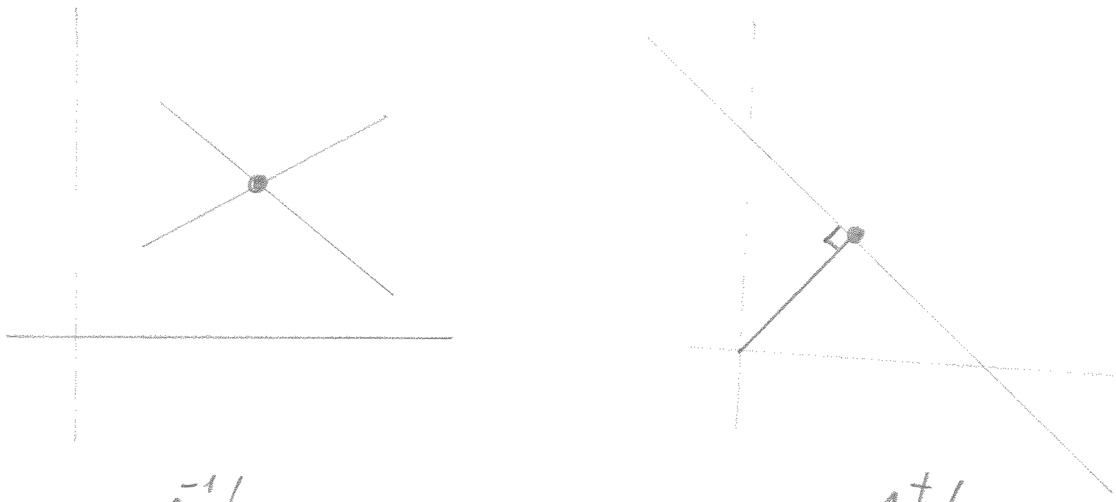
$$x = V \begin{bmatrix} 1/d_1 & & \\ & \ddots & \\ & & 1/d_n \end{bmatrix} U^T b = A^{-1} b.$$

If we choose $x_0 = 0$, we have the Minimum Norm (MN) solution,

$$x = V_1 \begin{bmatrix} 1/d_1 & & \\ & \ddots & \\ & & 1/d_r \end{bmatrix} U_1^T b$$

$$= \underbrace{V \begin{bmatrix} 1/d_1 & & & & \\ & \ddots & & & \\ & & 1/d_r & & \\ & & & 0 & \\ & & & & \ddots \; 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} U^T}_{\equiv A^+} b$$

A^+ = pseudo-inverse of A.



II $m > n$.

$$\begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \\ \hline & & 0 & \end{bmatrix} \begin{bmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_n \end{bmatrix} = \begin{bmatrix} \tilde{b}_1 \\ \tilde{b}_2 \\ \vdots \\ \\ \tilde{b}_m \end{bmatrix}$$

Assume that
We have

$$d_1 \geq d_2 \dots \geq d_k > d_{k+1} = \dots = d_n = 0.$$

$$d_j \tilde{x}_j = \tilde{b}_j \quad , \quad 1 \leq j \leq k$$

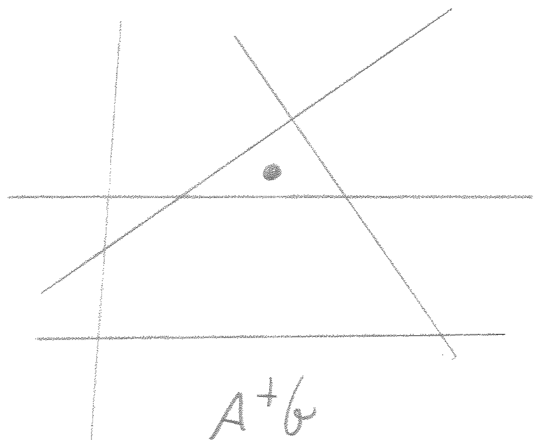
$$0 = \tilde{b}_j \quad , \quad k+1 \leq j \leq m.$$

Whether the latter system is satisfied does not depend on $\tilde{x}$. We can only guarantee the first ones by setting $\tilde{x}_j = \dfrac{\tilde{b}_j}{d_j}$, $1 \leq j \leq k$.

By defining

$$x = \underbrace{V \begin{bmatrix} 1/\tilde{d}_1 & & & & \\ & \ddots & & & \\ & & 1/\tilde{d}_r & & \\ & & & 0 & \\ & & & & \ddots \; 0 \cdots \end{bmatrix} U^T}_{A^+} b$$

we satisfy the equation $Ax = b$ as well as we can, and minimize the norm of the solution. The solution is the Minimum Norm Least Squares (MNLSQ) solution


$A^+ b$

In all cases, the pseudoinverse A^+ of A is

$$A^+ = V D^+ U^T \in \mathbb{R}^{n \times m}$$

$$D^+ = \operatorname{diag}\left(\frac{1}{d_1}, \frac{1}{d_2}, \ldots, \frac{1}{d_r}, 0, \ldots, 0 \right),$$

d_r = smallest non-zero singular value.

The pseudoinverse does not resolve the problem of how to find x from $Ax = b$ if b contains noise. One possible solution is to use the truncated SVD, or TSVD method.

Assume that $x \in \mathbb{R}^n$ is the <u>true</u> solution, i.e.

$$b = Ax + e$$

$b_0 = Ax$ = noiseless signal, e = noise. Let

$$A = UDV^T, \quad D = \mathrm{diag}(d_1, \ldots, d_{\min\{n,m\}})$$

Write first x in the V-basis,

$$x = \sum_{j=1}^{n} \tilde{x}_j v_j, \qquad \tilde{x}_j = (v_j^T x)$$

Let us define a TSVD(k)-<u>estimate</u> of x,

$$\hat{x}^{(k)} = \sum_{j=1}^{k} \frac{1}{d_j} (u_j^T b) v_j$$

$$= \underbrace{[v_1, \ldots, v_k]}_{V_k \in \mathbb{R}^{n \times k}} \begin{bmatrix} 1/d_1 & & & \\ & 1/d_2 & & \\ & & \ddots & \\ & & & 1/d_k \end{bmatrix} \underbrace{[u_1, \ldots, u_k]}_{U_k \in \mathbb{R}^{m \times k}}^T b$$

Plug in b:

$$\hat{x}^{(k)} = \sum_{j=1}^{k} \frac{1}{d_j} (u_j^T A x) v_j + \sum_{j=1}^{k} \frac{1}{d_j} (u_j^T e) v_j$$

But

$$u_j^T A x = u_j^T \sum_{\ell=1}^{k} u_\ell d_\ell (v_\ell^T x) = d_j \tilde{x}_j ,$$

so we have

$$\hat{x}^{(k)} = \sum_{j=1}^{k} \tilde{x}_j v_j + \sum_{j=1}^{k} \frac{1}{d_j} (u_j^T e) v_j$$

How far is $\hat{x}^{(k)}$ from x? We have

$$x - \hat{x}^{(k)} = \underbrace{\sum_{j=k+1}^{n} \tilde{x}_j v_j}_{\in \operatorname{sp}[v_{k+1}, \ldots, v_n]} + \underbrace{\sum_{j=1}^{k} \frac{1}{d_j} (u_j^T e) v_j}_{\in \operatorname{sp}[v_1, \ldots, v_k]}$$

and

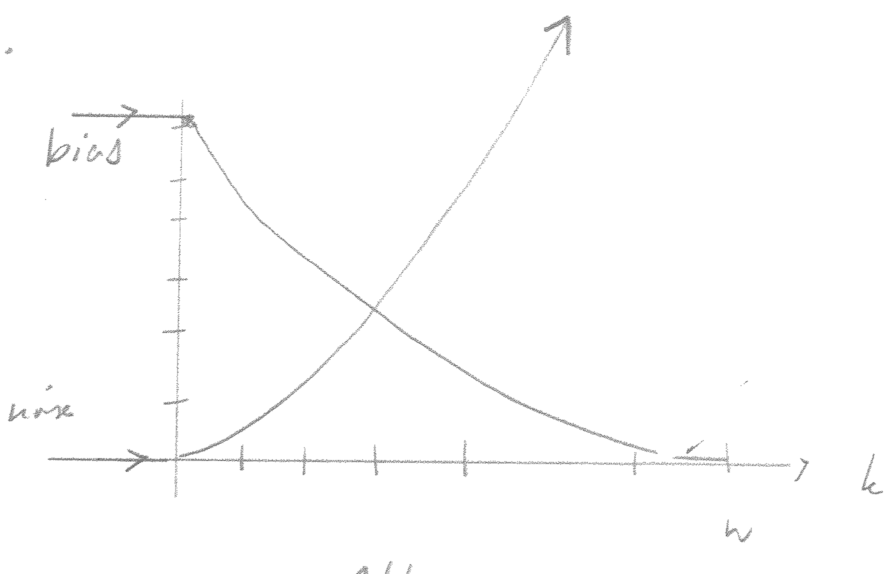
$$\| x - \hat{x}^{(k)} \|^2 = \Big\| \sum_{j=k+1}^{n} \tilde{x}_j v_j \Big\|^2 + \Big\| \sum_{j=1}^{k} \frac{1}{d_j} (u_j^T e) v_j \Big\|^2$$

$$= \sum_{j=k+1}^{n} \tilde{x}_j^2 + \sum_{j=1}^{k} \frac{1}{d_j^2} (u_j^T e)^2$$

The first term is *independent* of the error e, and it is often called the *bias* of the estimate $\hat{x}^{(k)}$. It holds that

$$\sum_{j=k+1}^{n} \tilde{x}_j^{2} \longrightarrow 0, \quad \text{as } k \to n$$

The second term is a reconstruction noise, and it *increases* as $k \to n$. If the last singular values are large, it may get out of control.



The estimate $\hat{x}^{(k)}$ is called the *TSVD estimate* of x. The art is to choose judiciously k so that

(a) The bias is *no more* too large

(b) The reconstruction noise is *not yet* unbearably large.

Ideal solution: Choose $k = k_{opt}$

$$k_{opt} = \min_k \| x - \hat{x}(k) \|^2$$

Here, however, we have a problem, since the optimization requires the knowledge of the true x, so we have to resort to approximate solutions. One such solution is the

<u>Morozov Discrepancy Principle</u> (MDP):

Assume that we have an estimate of the noise norm:

$$\| e \| \lesssim \varepsilon,$$

i.e., we assume that the true x satisfies

$$\| Ax - b \| \lesssim \varepsilon,$$

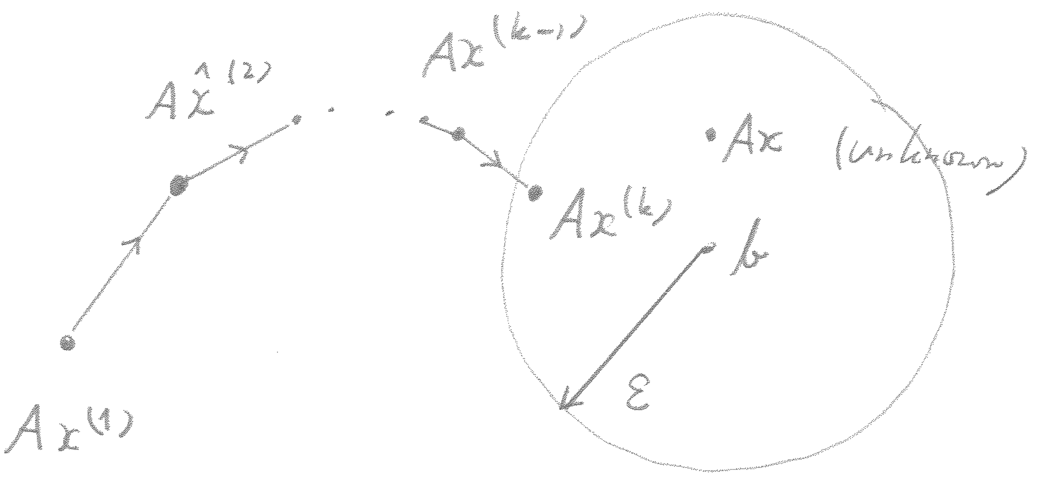
and that is all we know a priori. Then, we may say that any estimate $\hat{x}$ of x that satisfies $\| A\hat{x} - b \| \lesssim \varepsilon$ is, in view of our prior information, feasible.

On the other hand, increasing k in TSVD(k)-estimate seems to fit eventually to noise, so we should use as low k as possible.

Morozov Discrepancy Principle (MDP): Choose k so that

$$\|A\hat{x}^{(k)} - b\| \le \varepsilon, \quad \|A\hat{x}^{(k-1)} - b\| > \varepsilon,$$

i.e., k is the smallest integer for which the discrepancy $A\hat{x}^{(k)} - b$ is smaller in norm than the data.



Is this choice always possible?

Yes, if $b \in R(A)$:

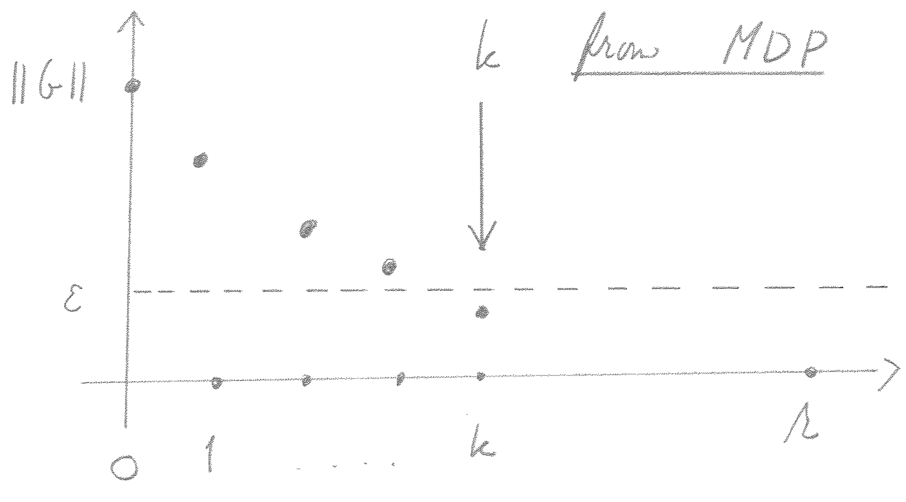
$$Ax = \sum_{j=1}^{r} u_j d_j (v_j^T x), \quad R(A) = \mathrm{sp}\{u_1, \dots, u_r\}$$

$$b \in R(A) \Rightarrow b = \sum_{j=1}^{r} \tilde{b}_j u_j$$

$$\hat{x}^{(r)} = \sum_{j=1}^{r} \frac{1}{d_j} \tilde{b}_j v_j$$

$$\|A\hat{x}^{(k)} - b\|^2 = \sum_{j=k+1}^{r} \tilde{b}_j^2 \xrightarrow{k \to r} 0$$

$\|A\hat{x}^{(k)} - b\| = \delta_k$



The is another, heuristic method to choose k, called the L-curve method.

We observe that the sequence

$$\delta_k = \|A\hat{x}^{(k)} - b\|$$

is monotonically decreasing.

$$\delta_0 = \|b\| \geq \delta_1 \geq \dots \geq \delta_r = 0$$

On the other hand, consider the norm of $\hat{x}^{(k)}$,

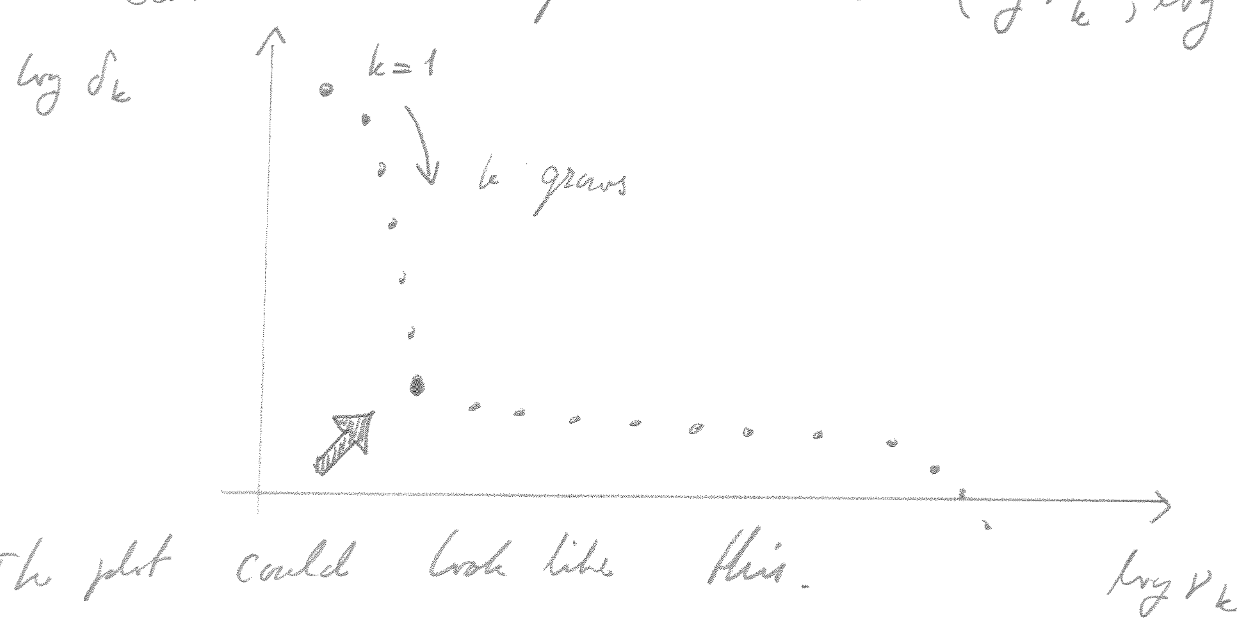
$$\nu_k = \| \hat{x}^{(k)} \| = \Big\| \sum_{j=1}^{k} \frac{\tilde{b}_j}{d_j} v_j \Big\|^2 ,$$

which is an increasing sequence. Moreover, remember that

$$\| \hat{x}^{(k)} \|^2 = \sum_{j=1}^{k} \left(\tilde{x}_j + \frac{u_j^T e}{d_j} \right)^2 ,$$

so we see that when $d_j \to 0$, the noise term starts to dominate.

Consider a plot $k \mapsto (\log \nu_k, \log \delta_k)$:



The plot could look like this.

What happens at the "kink" of the L-curve (if it is at all like an L!):

(1) The discrepancy does not go down any more, but

(2) The noise $\frac{u_j^T b}{d_j}$ takes over.

<u>Conclusion:</u> When you have reached the kink, <u>stop</u>!