Stability of Least Squares

In the previous section, we saw that given a vector $\boldsymbol{y} \in \mathbb{R}^M$ and a $M \times N$ matrix $\boldsymbol{A}$ of rank R, the solution[1] to

$$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{minimize}} \;\; \|\boldsymbol{x}\|_2^2 \quad \text{subject to} \quad \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}, \tag{1}$$

is found by applying the pseudo-inverse of $\boldsymbol{A}$ to $\boldsymbol{y}$:

$$\hat{\boldsymbol{x}}_{\text{ls}} = \boldsymbol{A}^{\dagger}\boldsymbol{y} = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{y},$$

where $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$ is the SVD of $\boldsymbol{A}$.

We will now discuss the stability of this solution, in particular how perturbing the response vector $\boldsymbol{y}$ affects the esimate. The analysis here will be very similar to when we looked at the stability of solving square sym+def systems, and in fact this is one of the main reasons we introduced the SVD.

Suppose we observe

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{e},$$

where $\boldsymbol{e} \in \mathbb{R}^M$ is an unknown perturbation. We apply the pseudo-inverse to $\boldsymbol{y}$ in an attempt to recover $\boldsymbol{x}_0$:

$$\hat{\boldsymbol{x}}_{\text{ls}} = \boldsymbol{A}^{\dagger}\boldsymbol{y} = \boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{A}^{\dagger}\boldsymbol{e}$$

We will separate the least-squares error into two parts

$$\hat{\boldsymbol{x}}_{\text{ls}} - \boldsymbol{x}_0 = \underbrace{\boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{x}_0}_{\text{“null space error”}} + \underbrace{\boldsymbol{A}^{\dagger}\boldsymbol{e}}_{\text{“noise error”}}$$

[1] Recall also that the (unique) solution to (1) is also the minimum norm solution of $\text{minimize}_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2$.

The null space error depends only on the matrix $\boldsymbol{A}$ and the target $\boldsymbol{x}_0$; it is completely independent of the perturbation $\boldsymbol{e}$. The noise error depends only on the perturbation $\boldsymbol{e}$ and the matrix $\boldsymbol{A}$; it is completely independent of the target $\boldsymbol{x}_0$.

We now discuss each of these errors in turn.

Null space error. We can immediately see from the expressions above that even when there is no perturbation ($\boldsymbol{e} = \boldsymbol{0}$), we might not get $\boldsymbol{x}_0$ back exactly, as

$$\begin{aligned}\boldsymbol{A}^\dagger \boldsymbol{A}\boldsymbol{x}_0 &= \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_0 \\ &= \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_0.\end{aligned}$$

If $\operatorname{rank}(\boldsymbol{A}) < N$, then it has a non-trivial null space, and $\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} \neq \mathbf{I}$. This means that any part of $\boldsymbol{x}_0$ that is not in the range of $\boldsymbol{V}$ (i.e. the row space) will not show up in (or influence) $\hat{\boldsymbol{x}}_{\mathrm{ls}}$.

Recall that the right singular vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_R \in \mathbb{R}^N$ are an orthobasis for $\operatorname{Row}(\boldsymbol{A})$. As before, let $\boldsymbol{v}_{R+1}, \ldots, \boldsymbol{v}_N \in \mathbb{R}^N$ be an orthobasis for $\operatorname{Null}(\boldsymbol{A})$ — these would be the columns of the $N \times (N - R)$ matrix $\boldsymbol{V}_0$ in the last set of notes. Together, $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N$ form an orthobasis for all of $\mathbb{R}^N$. Thus we can decompose $\boldsymbol{x}_0$ as

$$\begin{aligned}\boldsymbol{x}_0 &= \sum_{n=1}^{N} \langle \boldsymbol{x}_0, \boldsymbol{v}_n\rangle \boldsymbol{v}_n \\ &= \underbrace{\sum_{n=1}^{R} \langle \boldsymbol{x}_0, \boldsymbol{v}_n\rangle \boldsymbol{v}_n}_{\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_0} + \underbrace{\sum_{n=R+1}^{N} \langle \boldsymbol{x}_0, \boldsymbol{v}_n\rangle \boldsymbol{v}_n}_{\boldsymbol{V}_0\boldsymbol{V}_0^{\mathrm{T}}\boldsymbol{x}_0} \\ &= \boldsymbol{x}_{\mathrm{row}} + \boldsymbol{x}_{\mathrm{null}}.\end{aligned}$$

The vector $\boldsymbol{x}_{\text{row}}$ is the closest point in Row($\boldsymbol{A}$) to $\boldsymbol{x}_0$, and $\boldsymbol{x}_{\text{null}}$ is the closest point in Null($\boldsymbol{A}$) to $\boldsymbol{x}_0$, and of course $\boldsymbol{x}_{\text{null}} \perp \boldsymbol{x}_{\text{row}}$.

The null space error is then

$$\boldsymbol{x}_0 - \boldsymbol{A}^\dagger \boldsymbol{A}\boldsymbol{x}_0 = \sum_{n=R+1}^{N} \langle \boldsymbol{x}_0, \boldsymbol{v}_n\rangle \boldsymbol{v}_n = \boldsymbol{x}_{\text{null}},$$

and has size

$$\|\boldsymbol{x}_0 - \boldsymbol{A}^\dagger \boldsymbol{A}\boldsymbol{x}_0\|_2^2 = \|\hat{\boldsymbol{x}}_{\text{null}}\|_2^2 = \sum_{n=R+1}^{N} |\langle \boldsymbol{x}_0, \boldsymbol{v}_n\rangle|^2.$$

Again, this error is completely determined by (the row space of) $\boldsymbol{A}$ and $\boldsymbol{x}_0$, and does not depend on $\boldsymbol{e}$ at all.

Noise error. The noise error is

$$\begin{aligned} \boldsymbol{A}^\dagger \boldsymbol{e} &= \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{e} \\ &= \sum_{r=1}^{R} \frac{1}{\sigma_r}\langle \boldsymbol{e}, \boldsymbol{u}_r\rangle \boldsymbol{v}_r, \end{aligned}$$

and has size

$$\|\boldsymbol{A}^\dagger \boldsymbol{e}\|_2^2 = \sum_{r=1}^{R} \frac{1}{\sigma_r^2}|\langle \boldsymbol{e}, \boldsymbol{u}_r\rangle|^2.$$

There is some good news here: this error only depends on the part of $\boldsymbol{e}$ that is in Col($\boldsymbol{A}$); anything in the left null space Null($\boldsymbol{A}^{\mathrm{T}}$) gets completely filtered out by the pseudo inverse. Recall that the left singular vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_R \in \mathbb{R}^M$ are an orthobasis for Col($\boldsymbol{A}$). As before, let $\boldsymbol{u}_{R+1}, \ldots, \boldsymbol{u}_M \in \mathbb{R}^M$ be an orthobasis for Null($\boldsymbol{A}^{\mathrm{T}}$) —

these would be the columns of the $M \times (M-R)$ matrix $\boldsymbol{U}_0$ in the last set of notes. Together, $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_M$ form an orthobasis for all of $\mathbb{R}^M$. We can thus decompose the perturbation $\boldsymbol{e}$ as

$$
\begin{aligned}
\boldsymbol{e} &= \sum_{m=1}^{M} \langle \boldsymbol{e}, \boldsymbol{u}_m \rangle \boldsymbol{u}_m \\
&= \underbrace{\sum_{m=1}^{R} \langle \boldsymbol{e}, \boldsymbol{u}_m \rangle \boldsymbol{u}_m}_{\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{e}} + \underbrace{\sum_{m=R+1}^{M} \langle \boldsymbol{e}, \boldsymbol{u}_m \rangle \boldsymbol{u}_m}_{\boldsymbol{U}_0\boldsymbol{U}_0^{\mathrm{T}}\boldsymbol{e}} \\
&= \boldsymbol{e}_{\mathrm{col}} + \boldsymbol{e}_{\mathrm{ln}}.
\end{aligned}
$$

The vector $\boldsymbol{e}_{\mathrm{col}}$ is the closest point in $\mathrm{Col}(\boldsymbol{A})$ to $\boldsymbol{x}_0$, and $\boldsymbol{e}_{\mathrm{ln}}$ is the closest point in $\mathrm{Null}(\boldsymbol{A}^{\mathrm{T}})$ to $\boldsymbol{x}_0$, and of course $\boldsymbol{e}_{\mathrm{col}} \perp \boldsymbol{e}_{\mathrm{ln}}$. Note in particular that

$$
\|\boldsymbol{e}_{\mathrm{col}}\|_2^2 = \|\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{e}\|_2^2 = \|\boldsymbol{U}^{\mathrm{T}}\boldsymbol{e}\|_2^2 = \sum_{r=1}^{R} |\langle \boldsymbol{e}, \boldsymbol{u}_r \rangle|^2 \leq \|\boldsymbol{e}\|_2^2.
$$

Using analysis that is almost identical to that applied to solving symmetric systems of equations, we have

$$
\frac{1}{\sigma_1^2}\|\boldsymbol{e}_{\mathrm{col}}\|_2^2 \;\leq\; \|\boldsymbol{A}^{\dagger}\boldsymbol{e}\|_2^2 \;\leq\; \frac{1}{\sigma_R^2}\|\boldsymbol{e}_{\mathrm{col}}\|_2^2.
$$

The upper bound (worst case) is achieved when $\boldsymbol{e}_{\mathrm{col}}$ is aligned with $\boldsymbol{u}_R$. The best case is actually when $\boldsymbol{e}$ is completely in the left null space $\mathrm{Null}(\boldsymbol{A}^{\mathrm{T}})$, and so $\|\boldsymbol{A}^{\dagger}\boldsymbol{e}\|_2^2 = 0$.

Also as in the symmetric case, we can look at the average error when the perturbation $\boldsymbol{e}$ is random. Say that each entry $e[m]$ of $\boldsymbol{e}$ is a random variable independent of all the other entries, and distributed

$$e[m] \ \sim \ \text{Normal}(0, \nu^2).$$

Then, as we have argued before, the expected size of $\boldsymbol{e}$ is

$$\mathrm{E}[\|\boldsymbol{e}\|_2^2] = M\nu^2,$$

and the average noise error is

$$\begin{aligned}
\mathrm{E}\left[\|\boldsymbol{A}^\dagger\boldsymbol{e}\|_2^2\right] &= \mathrm{E}\left[\boldsymbol{e}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{e}\right] \\
&= \mathrm{E}\left[\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}^{-2}\boldsymbol{\beta}\right], \quad \text{where } \ \boldsymbol{\beta} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{e} \\
&= \sum_{r=1}^{R} \frac{1}{\sigma_r^2}\,\mathrm{E}\left[\beta_r^2\right] \\
&= \nu^2 \cdot \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \cdots + \frac{1}{\sigma_R^2}\right)
\end{aligned}$$

where the last step uses the fact (see the bottom of page 16) that

$$\mathrm{E}[\beta_r^2] = \mathrm{E}[|\langle \boldsymbol{e}, \boldsymbol{u}_r\rangle|^2] = \nu^2\|\boldsymbol{u}_r\|_2^2 = \nu^2.$$

Again, if σ_R is tiny, $1/\sigma_R^2$ will dominate the sum above, and the average reconstruction error will be quite large.

Noting that

$$\mathrm{E}[\|\boldsymbol{e}_{\mathrm{col}}\|_2^2] = \sum_{r=1}^{R}\mathrm{E}[|\langle \boldsymbol{e}, \boldsymbol{u}_r\rangle|^2] = R\nu^2,$$

we can write the noise error as

$$
\begin{aligned}
\mathrm{E}\left[\|\boldsymbol{A}^{\dagger}\boldsymbol{e}\|_2^2\right] &= \frac{\nu^2}{R}\left(\sum_{r=1}^{R}\frac{1}{\sigma_r^2}\right)\mathrm{E}[\|\boldsymbol{e}_{\mathrm{col}}\|_2^2] \\
&= \frac{\nu^2}{M}\left(\sum_{r=1}^{R}\frac{1}{\sigma_r^2}\right)\mathrm{E}[\|\boldsymbol{e}\|_2^2].
\end{aligned}
$$

The size of the noise error only depends on the portion of $\boldsymbol{e}$ that is in the column space, and on average, this will be only a fraction of the total power in the perturbation, as $\mathrm{E}[\|\boldsymbol{e}_{\mathrm{col}}\|_2^2] = \frac{R}{M}\,\mathrm{E}[\|\boldsymbol{e}\|_2^2]$.

Combined error. Note that the null space error is in the span of the $\boldsymbol{v}_{R+1},\ldots,\boldsymbol{v}_N$ (i.e. $\mathrm{Null}(\boldsymbol{A})$) while the noise error is in the span of $\boldsymbol{v}_1,\ldots,\boldsymbol{v}_N$ (i.e. $\mathrm{Row}(\boldsymbol{A})$). This means that the errors are orthogonal to one another and so

$$
\begin{aligned}
\|\hat{\boldsymbol{x}}_{\mathrm{ls}} - \boldsymbol{x}_0\|_2^2 &= \|\text{null space error}\|_2^2 + \|\text{noise error}\|_2^2 \\
&= \|\boldsymbol{V}_0\boldsymbol{V}_0^{\mathrm{T}}\boldsymbol{x}_0\|_2^2 + \|\boldsymbol{A}^{\dagger}\boldsymbol{e}\|_2^2 \\
&= \sum_{r=R+1}^{N}|\langle\boldsymbol{x}_0,\boldsymbol{v}_r\rangle|^2 + \sum_{r=1}^{R}\frac{1}{\sigma_r^2}|\langle\boldsymbol{e},\boldsymbol{u}_r\rangle|^2
\end{aligned}
$$

We close this section by re-iterating two points:

- The null space error if fixed, it depends only on $\boldsymbol{x}_0$ and $\mathrm{Row}(\boldsymbol{A})$ and not the perturbation.
- If any of the σ_r are small, then the noise error can be very large; we can see above that the component of $\boldsymbol{e}$ in direction $\boldsymbol{u}_r$ gets amplified by a factor of $\frac{1}{\sigma_r^2}$.

Stable Reconstruction with the Truncated SVD

We have seen that if $\boldsymbol{A}$ has very small singular values and we apply the pseudo-inverse in the presence of noise, the results can be disastrous. But it doesn't have to be this way. There are several ways to stabilize the pseudo-inverse. We start be discussing the simplest one, where we simply "cut out" the part of the reconstruction which is causing the problems.

As before, we are given noisy indirect observations of a vector $\boldsymbol{x}_0$ through a $M \times N$ matrix $\boldsymbol{A}$:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{e}. \tag{2}$$

The matrix $\boldsymbol{A}$ has SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$, and pseudo-inverse $\boldsymbol{A}^{\dagger} = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathrm{T}}$. We can rewrite $\boldsymbol{A}$ and its pseudo-inverse as a sum of rank-1 matrices:

$$\boldsymbol{A} = \sum_{r=1}^{R} \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^{\mathrm{T}}, \quad \boldsymbol{A}^{\dagger} = \sum_{r=1}^{R} \frac{1}{\sigma_r} \boldsymbol{v}_r \boldsymbol{u}_r^{\mathrm{T}},$$

where R is the rank of $\boldsymbol{A}$, the σ_r are the singular values, and $\boldsymbol{u}_r \in \mathbb{R}^M$ and $\boldsymbol{v}_r \in \mathbb{R}^N$ are columns of $\boldsymbol{U}$ and $\boldsymbol{V}$, respectively. Given $\boldsymbol{y}$ as above, we can write the least-squares estimate of $\boldsymbol{x}$ from the noisy measurements as

$$\hat{\boldsymbol{x}}_{\mathrm{ls}} = \boldsymbol{A}^{\dagger}\boldsymbol{y} = \sum_{r=1}^{R} \frac{1}{\sigma_r} \langle \boldsymbol{y}, \boldsymbol{u}_r \rangle \boldsymbol{v}_r. \tag{3}$$

As we can see (and have seen before) if any one of the σ_r are very small, the least-squares reconstruction can be a disaster.

A simple way to avoid this is to simply truncate the sum (3), leaving out the terms where σ_r is too small ($1/\sigma_r$ is too big). Exactly how many terms to keep depends a great deal on the application, as there are competing interests. On the one hand, we want to ensure that each of the σ_r we include has an inverse of reasonable size, on the other, we want the reconstruction to be accurate (i.e. does not deviate from the noiseless least-squares solution by too much).

We form an approximation $\boldsymbol{A}_{\mathrm{t}}$ to $\boldsymbol{A}$ by taking[2]

$$\boldsymbol{A}_{\mathrm{t}} = \sum_{r=1}^{R'} \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^{\mathrm{T}},$$

for some $R' < R$. Again, our final answer will depend on which R' we use, and choosing R' is often times something of an art. It is clear that the approximation $\boldsymbol{A}_{\mathrm{t}}$ has rank R'. Note that the pseudo-inverse of $\boldsymbol{A}_{\mathrm{t}}$ is also a truncated sum

$$\boldsymbol{A}_{\mathrm{t}}^{\dagger} = \sum_{r=1}^{R'} \frac{1}{\sigma_r} \boldsymbol{v}_r \boldsymbol{u}_r^{\mathrm{T}}.$$

Given noisy data $\boldsymbol{y}$ as in (2), we reconstruct $\boldsymbol{x}$ by applying the truncated pseudo-inverse to $\boldsymbol{y}$:

$$\hat{\boldsymbol{x}}_{\mathrm{t}} = \boldsymbol{A}_{\mathrm{t}}^{\dagger} \boldsymbol{y} = \sum_{r=1}^{R'} \frac{1}{\sigma_r} \langle \boldsymbol{y}, \boldsymbol{u}_r \rangle \boldsymbol{v}_r.$$

We can again analyze the reconstruction using the same framework

[2]The subscript t stands for "truncated".

as we used for least squares. We have

$$\begin{aligned}
\hat{\boldsymbol{x}}_{\mathrm{t}} - \boldsymbol{x}_0 &= \boldsymbol{A}_{\mathrm{t}}^{\dagger}\boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{x}_0 + \boldsymbol{A}_{\mathrm{t}}^{\dagger}\boldsymbol{e} \\
&= \sum_{r=1}^{R'} \langle \boldsymbol{x}_0, \boldsymbol{v}_r \rangle \boldsymbol{v}_r - \sum_{r=1}^{N} \langle \boldsymbol{x}_0, \boldsymbol{v}_r \rangle \boldsymbol{v}_r + \sum_{r=1}^{R'} \frac{1}{\sigma_r} \langle \boldsymbol{e}, \boldsymbol{u}_r \rangle \boldsymbol{v}_r \\
&= -\underbrace{\sum_{r=R+1}^{N} \langle \boldsymbol{x}_0, \boldsymbol{v}_r \rangle \boldsymbol{v}_r}_{\text{null space error}} - \underbrace{\sum_{r=R'+1}^{R} \langle \boldsymbol{x}_0, \boldsymbol{v}_r \rangle \boldsymbol{v}_r}_{\text{truncation error}} + \underbrace{\sum_{r=1}^{R'} \frac{1}{\sigma_r} \langle \boldsymbol{e}, \boldsymbol{u}_r \rangle \boldsymbol{v}_r}_{\text{noise error}}
\end{aligned}$$

Note that each of these three errors is in the span of a different subset of the orthonormal basis vectors $\{\boldsymbol{v}_n\}$, this means that they are all orthogonal to one another and

$$\|\hat{\boldsymbol{x}}_{\mathrm{t}} - \boldsymbol{x}_0\|_2^2 = \underbrace{\sum_{r=R+1}^{N} |\langle \boldsymbol{x}_0, \boldsymbol{v}_r \rangle|^2}_{\|\text{null space error}\|_2^2 = \|\boldsymbol{x}_{\text{null}}\|_2^2} + \underbrace{\sum_{r=R'+1}^{R} |\langle \boldsymbol{x}_0, \boldsymbol{v}_r \rangle|^2}_{\|\text{truncation error}\|_2^2} + \underbrace{\sum_{r=1}^{R'} \frac{1}{\sigma_r^2} |\langle \boldsymbol{e}, \boldsymbol{u}_r \rangle|^2}_{\|\text{noise error}\|_2^2}$$

What happens to these errors as we change R'? First, as before, the null space error is fixed an is not going away. In fact, as we chose R' smaller and smaller, we are adding terms to the truncation error — conceptually, by truncating $\boldsymbol{A}$ we are effectively increasing the dimension of its null space, and more of $\boldsymbol{x}_0$ gets cut out when $\boldsymbol{A}_{\mathrm{t}}^{\dagger}$ is applied. But making R' smaller also effectively decreasing the dimension of the column space, meaning that the size of the noise error goes down.

In fact, our worst case upper bound for the noise error is now

$$\sum_{r=1}^{R'} \frac{1}{\sigma_r} |\langle \boldsymbol{e}, \boldsymbol{u}_r \rangle|^2 \leq \frac{1}{\sigma_{R'}^2} \|\boldsymbol{e}\|_2^2,$$

which gets smaller and smaller as R' decreases (remember that we have ordered the singular values in decreasing order, so $\sigma_1 \geq \sigma_2 \geq \cdot \geq \sigma_R$, so in particular $\sigma_{R'} \geq \sigma_R$) and is always less than $1/\sigma_R^2$. If the entries of $\boldsymbol{e}$ are random, zero mean, and uncorrelated with variance ν^2, the expected noise error is

$$\begin{aligned}
\mathrm{E}[\|\text{noise error}\|_2^2] &= \sum_{r=1}^{R'} \frac{1}{\sigma_r^2} \ \mathrm{E}[|\langle \boldsymbol{e}, \boldsymbol{u}_r \rangle|^2] \\
&= \nu^2 \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \cdots + \frac{1}{\sigma_{R'}^2} \right) \\
&= \frac{1}{M} \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \cdots + \frac{1}{\sigma_{R'}^2} \right) \cdot \mathrm{E}[\|\boldsymbol{e}\|_2^2].
\end{aligned}$$

As R' gets smaller, we are removing terms from the sum above, and the noise error is decreasing. Not only that, but the terms that we are removing are the ones that are most problematic, since again the $\{\sigma_r\}$ are indexed in decreasing order.

Of course, choosing the truly optimal R' would require knowledge of both $\boldsymbol{x}_0$ and $\boldsymbol{e}$, information that is in general not available. You can, however, get a reasonable idea of its effect on the noise error by looking at the singular value of $\boldsymbol{A}$ ... in general, singular values that are comparatively very small should be "cut out".

Stable Reconstruction using Tikhonov Regularization (Ridge Regression)

Tikhonov[3] regularization is another way to stabilize the least-squares recovery. It has the nice features that: 1) it can be interpreted using optimization, and 2) it can be computed without direct knowledge of the SVD of $\boldsymbol{A}$. When $\boldsymbol{A}, \boldsymbol{y}$ come from a regression problem, then this amounts to **ridge regresssion**, something that we are already familiar with.

Recall that we motivated the pseudo-inverse by showing that $\hat{\boldsymbol{x}}_{ls} = \boldsymbol{A}^{\dagger}\boldsymbol{y}$ is a solution to

$$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{minimize}} \;\; \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2. \tag{4}$$

When $\boldsymbol{A}$ has full column rank, $\hat{\boldsymbol{x}}_{ls}$ is the unique solution, otherwise it is the solution with smallest energy. When $\boldsymbol{A}$ has full column rank but has singular values which are very small, huge variations in $\boldsymbol{x}$ (in directions of the singular vectors $\boldsymbol{v}_k$ corresponding to the tiny σ_k) can have very little effect on the residual $\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2$. As such, the solution to (4) can have wildly inaccurate components in the presence of even mild noise.

One way to counteract this problem is to modify (4) with a **regularization** term that penalizes the size of the solution $\|\boldsymbol{x}\|_2^2$ as well as the residual error $\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2$:

$$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{minimize}} \;\; \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \delta\|\boldsymbol{x}\|_2^2. \tag{5}$$

The parameter $\delta > 0$ gives us a trade-off between accuracy and regularization; we want to choose δ small enough so that the residual

[3] Andrey Tikhonov (1906-1993) was a 20th century Russian mathematician.

for the solution of (5) is close to that of (4), and large enough so that the problem is well-conditioned.

We have already seen that the (unique) solution to (5) is

$$\hat{\boldsymbol{x}}_{\text{tik}} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \delta\mathbf{I})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}.$$

(We show why this is true in the Technical Details section below.) This can be re-written in terms of the SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$ as

$$\begin{aligned}\hat{\boldsymbol{x}}_{\text{tik}} &= \boldsymbol{V}(\boldsymbol{\Sigma}^2 + \delta\mathbf{I})^{-1}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{y} \\ &= \sum_{r=1}^{R} \frac{\sigma_r}{\sigma_r^2 + \delta}\langle \boldsymbol{y}, \boldsymbol{u}_r\rangle \boldsymbol{v}_r.\end{aligned}$$

We can get a better feel for what Tikhonov regularization is doing by comparing it directly to the pseudo-inverse. The least-squares reconstruction $\hat{\boldsymbol{x}}_{\text{ls}}$ can be written as

$$\begin{aligned}\hat{\boldsymbol{x}}_{\text{ls}} &= \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{y} \\ &= \sum_{r=1}^{R} \frac{1}{\sigma_r}\langle \boldsymbol{y}, \boldsymbol{u}_r\rangle \boldsymbol{v}_r,\end{aligned}$$

while the Tikhonov reconstruction $\hat{\boldsymbol{x}}_{\text{tik}}$ derived above is

$$\hat{\boldsymbol{x}}_{\text{tik}} = \sum_{r=1}^{R} \frac{\sigma_r}{\sigma_r^2 + \delta}\langle \boldsymbol{y}, \boldsymbol{u}_r\rangle \boldsymbol{v}_r. \tag{6}$$

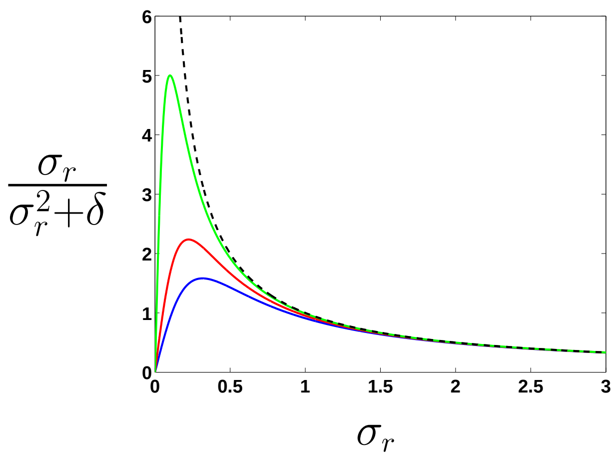
Notice that when σ_r is much larger than δ,

$$\frac{\sigma_r}{\sigma_r^2 + \delta} \approx \frac{1}{\sigma_r}, \quad \sigma_r \gg \delta,$$

but when σ_r is small

$$\frac{\sigma_r}{\sigma_r^2+\delta} \approx 0, \quad \sigma_r \ll \delta.$$

Thus the Tikhonov reconstruction modifies the important parts (components where the σ_r are large) of the pseudo-inverse very little, while ensuring that the unimportant parts (components where the σ_r are small) affect the solution only by a very small amount. This **damping** of the singular values, is illustrated below.



Above, we see the damped multipliers $\sigma_r/(\sigma_r^2+\delta)$ versus σ_r for $\delta = 0.1$ (blue), $\delta = 0.05$ (red), and $\delta = 0.01$ (green). The black dotted line is $1/\sigma_r$, the least-squares multiplier. Notice that for large σ_r ($\sigma_r > 2\sqrt{\delta}$, say), the damping has almost no effect.

This damping makes the Tikhonov reconstruction exceptionally stable; large multipliers never appear in the reconstruction (6). In fact it is easy to check that

$$\frac{\sigma_r}{\sigma_r^2+\delta} \leq \frac{1}{2\sqrt{\delta}}$$

no matter the value of σ_r.

Technical Details: Tikhonov Regularization and the SVD

In this section, we will quickly argue why

$$(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \delta\mathbf{I})^{-1}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{V}(\boldsymbol{\Sigma}^2 + \delta\mathbf{I})^{-1}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{T}},$$

where $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$ is the SVD of $\boldsymbol{A}$.

When $\boldsymbol{A}$ has full column rank, $R = N$, then $\boldsymbol{V}$ is a square orthonormal matrix (so $\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \mathbf{I}$) and we have

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \delta\mathbf{I} = \boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^{\mathrm{T}} + \delta\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{V}(\boldsymbol{\Sigma}^2 + \delta\mathbf{I})\boldsymbol{V}^{\mathrm{T}},$$

and so $(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \delta\mathbf{I})^{-1} = \boldsymbol{V}(\boldsymbol{\Sigma}^2 + \delta\mathbf{I})^{-1}\boldsymbol{V}^{\mathrm{T}}$, and

$$\begin{aligned}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \delta\mathbf{I})^{-1}\boldsymbol{A}^{\mathrm{T}} &= \boldsymbol{V}(\boldsymbol{\Sigma}^2 + \delta\mathbf{I})^{-1}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{T}} \\ &= \boldsymbol{V}(\boldsymbol{\Sigma}^2 + \delta\mathbf{I})^{-1}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{T}}.\end{aligned}$$

When $R < N$, we have to be a little more careful. Let $\boldsymbol{V}_{\text{full}}$ be as in the SVD notes (the column concatenation of the orthobasis for the row space from the SVD and an orthobasis for the null space), and let $\boldsymbol{\Sigma}_N$ be the $N \times N$ diagonal matrix

$$\boldsymbol{\Sigma}_N = \begin{bmatrix} \sigma_1 & & & & & & \\ & \sigma_2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_R & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}.$$

Then we have

$$\begin{aligned}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}+\delta\mathbf{I} &= \boldsymbol{V}_{\text{full}}\boldsymbol{\Sigma}_N^2\boldsymbol{V}_{\text{full}}^{\mathrm{T}}+\delta\boldsymbol{V}_{\text{full}}\boldsymbol{V}_{\text{full}}^{\mathrm{T}} \\ &= \boldsymbol{V}_{\text{full}}(\boldsymbol{\Sigma}_N^2+\delta\mathbf{I})\boldsymbol{V}_{\text{full}}^{\mathrm{T}},\end{aligned}$$

and

$$(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}+\delta\mathbf{I})^{-1}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{V}_{\text{full}}(\boldsymbol{\Sigma}_N^2+\delta\mathbf{I})^{-1}\boldsymbol{V}_{\text{full}}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{T}}.$$

The matrix $\boldsymbol{V}_{\text{full}}^{\mathrm{T}}\boldsymbol{V}$ is $N\times R$, and has structure

$$\boldsymbol{V}_{\text{full}}^{\mathrm{T}}\boldsymbol{V} = \begin{bmatrix}\mathbf{I}_{R\times R} \\ \mathbf{0}_{(N-R)\times R}\end{bmatrix}.$$

So multiplying this matrix on the right of $\boldsymbol{V}_{\text{full}}(\boldsymbol{\Sigma}_N^2+\delta\mathbf{I})^{-1}$ has the effect of simply removing the last $N-R$ columns, and

$$\boldsymbol{V}_{\text{full}}(\boldsymbol{\Sigma}_N^2+\delta\mathbf{I})^{-1}\begin{bmatrix}\mathbf{I}_{R\times R} \\ \mathbf{0}_{(N-R)\times R}\end{bmatrix} = \boldsymbol{V}(\boldsymbol{\Sigma}^2+\delta\mathbf{I})^{-1}.$$

(Note that the diagonal matrices $\boldsymbol{\Sigma}_N,\mathbf{I}$ on the left above are both $N\times N$, while $\boldsymbol{\Sigma},\mathbf{I}$ on the right are $R\times R$.) Thus

$$\begin{aligned}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}+\delta\mathbf{I})^{-1}\boldsymbol{A}^{\mathrm{T}} &= \boldsymbol{V}(\boldsymbol{\Sigma}^2+\delta\mathbf{I})^{-1}\boldsymbol{\Sigma}\boldsymbol{U} \\ &= \sum_{r=1}^{R}\frac{\sigma_r}{\sigma_r^2+\delta}\boldsymbol{v}_r\boldsymbol{u}_r^{\mathrm{T}}.\end{aligned}$$

Technical Details: Tikhonov Error Analysis

Given noisy observations $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{e}$, how well will Tikhonov regularization work? The answer to this questions depends on multiple factors including the choice of δ, the nature of the perturbation $\boldsymbol{e}$, and how well $\boldsymbol{x}_0$ can be approximated using a linear combination of the singular vectors $\boldsymbol{v}_r$ corresponding to the large (relative to δ) singular values. Since a closed-form expression for the solution to (5) exists, we can quantify these trade-offs precisely.

The Tikhonov regularized solution is given by

$$\begin{aligned}
\hat{\boldsymbol{x}}_{\text{tik}} &= \sum_{r=1}^{R} \frac{\sigma_r}{\sigma_r^2 + \delta} \langle \boldsymbol{y}, \boldsymbol{u}_r \rangle \boldsymbol{v}_r \\
&= \sum_{r=1}^{R} \frac{\sigma_r}{\sigma_r^2 + \delta} \langle \boldsymbol{A}\boldsymbol{x}_0, \boldsymbol{u}_r \rangle \boldsymbol{v}_r \; + \; \sum_{r=1}^{R} \frac{\sigma_r}{\sigma_r^2 + \delta} \langle \boldsymbol{e}, \boldsymbol{u}_r \rangle \boldsymbol{v}_r \\
&= \sum_{r=1}^{R} \frac{\sigma_r^2}{\sigma_r^2 + \delta} \langle \boldsymbol{x}_0, \boldsymbol{v}_r \rangle \boldsymbol{v}_r \; + \; \sum_{r=1}^{R} \frac{\sigma_r}{\sigma_r^2 + \delta} \langle \boldsymbol{e}, \boldsymbol{u}_r \rangle \boldsymbol{v}_r .
\end{aligned}$$

Like the least-squares estimate, $\hat{\boldsymbol{x}}_{\text{tik}}$ lies in the row space of $\boldsymbol{A}$. This means that there will we will have the same null space error. But unlike least-squares, this estimate does not match $\boldsymbol{x}_{\text{row}}$ when there is no perturbation, $\boldsymbol{e} = \boldsymbol{0}$. This leads to an additional source of error that does not depend on $\boldsymbol{e}$; it depends only on the regularization parameter δ relative to the singular values of $\boldsymbol{A}$. We call this deviation in the row space the **regularization error**:

$$\begin{aligned}
\text{regularization error} &= \boldsymbol{x}_{\text{row}} - \sum_{r=1}^{R} \frac{\sigma_r^2}{\sigma_r^2+\delta}\langle \boldsymbol{x}_0, \boldsymbol{v}_r\rangle \boldsymbol{v}_r \\
&= \sum_{r=1}^{R} \left(1 - \frac{\sigma_r^2}{\sigma_r^2+\delta}\right)\langle \boldsymbol{x}_0, \boldsymbol{v}_r\rangle \boldsymbol{v}_r \\
&= \sum_{r=1}^{R} \frac{\delta}{\sigma_r^2+\delta}\langle \boldsymbol{x}_0, \boldsymbol{v}_r\rangle \boldsymbol{v}_r,
\end{aligned}$$

with

$$\|\text{regularization error}\|_2^2 \;=\; \sum_{r=1}^{R} \frac{\delta^2}{(\sigma_r^2+\delta)^2}|\langle \boldsymbol{x}_0, \boldsymbol{v}_r\rangle|^2.$$

Note that for the components much smaller than δ,

$$\sigma_r^2 \ll \delta \quad \Rightarrow \quad \frac{\delta^2}{(\sigma_r^2+\delta)^2} \approx 1,$$

so this portion of the approximation error will be about the same as if we had simply truncated these components.

For large components,

$$\sigma_r^2 \gg \delta \quad \Rightarrow \quad \frac{\delta^2}{(\sigma_r^2+\delta)^2} \approx \frac{\delta^2}{\sigma_r^2},$$

the corresponding portion of the approximation error will be very small.

For the noise error, we have

$$\begin{aligned}\|\text{Noise error}\|_2^2 &= \sum_{r=1}^{R}\left(\frac{\sigma_r}{\sigma_r^2+\delta}\right)^2 |\langle \boldsymbol{e}, \boldsymbol{u}_r\rangle|^2 \\ &\leq \frac{1}{4\delta}\sum_{r=1}^{R}|\langle \boldsymbol{e}, \boldsymbol{u}_r\rangle|^2 \\ &\leq \frac{1}{4\delta}\|\boldsymbol{e}\|_2^2.\end{aligned}$$

The worst-case error is more or less determined by the choice of δ. The regularization makes the effective condition number of $\boldsymbol{A}$ about $1/(2\sqrt{\delta})$; no matter how small the smallest singular value is, the noise energy will not increase by more than a factor of $1/(4\delta)$ during the reconstruction process.

Less pessimistic is the average case error. Suppose that the entries of the error vector $\boldsymbol{e}$ are iid Gaussian random variables

$$e[m] \sim \text{Normal}(0, \nu^2).$$

Then the expected value of the noise error reconstruction will be

$$\begin{aligned}\mathrm{E}\left[\|\text{Noise error}\|_2^2\right] &= \sum_{r=1}^{R}\left(\frac{\sigma_r}{\sigma_r^2+\delta}\right)^2 \mathrm{E}\left[|\langle \boldsymbol{e}, \boldsymbol{u}_r\rangle|^2\right] \\ &= \nu^2\cdot\sum_{r=1}^{R}\left(\frac{\sigma_r}{\sigma_r^2+\delta}\right)^2 \\ &= \frac{1}{M}\cdot\left(\sum_{r=1}^{R}\frac{\sigma_r^2}{(\sigma_r^2+\delta)^2}\right)\cdot \mathrm{E}\left[\|\boldsymbol{e}\|_2^2\right]. \qquad (7)\end{aligned}$$

Note that

$$\frac{\sigma_r^2}{(\sigma_r^2+\delta)^2} \;\leq\; \min\left(\frac{1}{\sigma_r^2}, \frac{1}{4\delta}\right),$$

so we can think of the error in (7) as an average of the $\frac{1}{\sigma_r^2}$, with the large values simply replaced by $1/(4\delta)$.

Finally, notice that the noise Error and the regularization error above are not orthogonal to each other; they are both acting in the row space. The null space error, however, still remains orthogonal to both of these. Thus we can write the Tikhonov estimation error as

$$\begin{aligned}\|\hat{\boldsymbol{x}}_{\text{tik}} - \boldsymbol{x}_0\|_2^2 &= \|\boldsymbol{x}_{\text{null}}\|_2^2 + \|\boldsymbol{x}_{\text{row}} - \hat{\boldsymbol{x}}_{\text{tik}}\|_2^2 \\ &= \sum_{r=R+1}^{N} |\langle \boldsymbol{x}_0, \boldsymbol{v}_n\rangle|^2 + \sum_{r=1}^{R} |\epsilon_r|^2,\end{aligned}$$

where

$$\epsilon_r = \frac{1}{\sigma_r^2 + \delta}\left(\delta\langle \boldsymbol{x}_0, \boldsymbol{v}_r\rangle + \sigma_r\langle \boldsymbol{e}, \boldsymbol{u}_r\rangle\right).$$