Least squares in noise

The real advantage to thinking about least squares through the lens of the SVD is that it tells us exactly what to expect when our measurements are corrupted with *noise*. Specifically, suppose we observe

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

where $\boldsymbol{e} \in \mathbb{R}^M$ is an (unknown) perturbation of our measurements. Suppose that we then form the least squares estimate:

$$\widehat{\boldsymbol{x}}_{\text{ls}} = \boldsymbol{A}^\dagger \boldsymbol{y} = \boldsymbol{A}^\dagger \boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}^\dagger \boldsymbol{e}. \tag{1}$$

We would like to understand the effect of the noise vector $\boldsymbol{e}$ on our estimate of $\boldsymbol{x}$. One way to understand this is to compare $\widehat{\boldsymbol{x}}_{\text{ls}}$ to what we would have obtained if we had been able to use least squares on the noise-free observations $\boldsymbol{y}_{\text{clean}} = \boldsymbol{A}\boldsymbol{x}$. The noise-free reconstruction is

$$\widehat{\boldsymbol{x}}_{\text{clean}} = \boldsymbol{A}^\dagger \boldsymbol{y} = \boldsymbol{A}^\dagger \boldsymbol{A}\boldsymbol{x}.$$

Combining this with (1) we see that the additional error due to noise can be measured as

$$\|\widehat{\boldsymbol{x}}_{\text{ls}} - \widehat{\boldsymbol{x}}_{\text{clean}}\|_2^2 = \|\boldsymbol{A}^\dagger \boldsymbol{e}\|_2^2 = \|\boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{e}\|_2^2. \tag{2}$$

Now suppose for a moment that the error has unit norm, $\|\boldsymbol{e}\|_2 = 1$. Just how bad can the error be? The worst case for (2) in this case is given by

$$\underset{\boldsymbol{e}\in\mathbb{R}^M}{\text{maximize}} \ \|\boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{e}\|_2^2 \quad \text{subject to} \quad \|\boldsymbol{e}\|_2 = 1. \tag{3}$$

Note that $\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{e}$ is a vector in $\mathbb{R}^R$, and that for any vector $\boldsymbol{z} \in \mathbb{R}^R$, we have

$$\|\boldsymbol{V}\boldsymbol{z}\|_2^2 = \boldsymbol{z}^{\mathrm{T}}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{z} = \boldsymbol{z}^{\mathrm{T}}\boldsymbol{z} = \|\boldsymbol{z}\|_2^2,$$

since the columns of $\boldsymbol{V}$ are orthonormal. Thus, we can simplify (3) to

$$\underset{\boldsymbol{e}\in\mathbb{R}^M}{\text{maximize}} \ \ \|\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{e}\|_2^2 \quad \text{subject to} \ \ \|\boldsymbol{e}\|_2 = 1. \tag{4}$$

We can simplify a bit further by noticing that, since the columns of $\boldsymbol{U}$ are orthonormal, $\|\boldsymbol{U}^{\mathrm{T}}\boldsymbol{e}\|_2^2 \leq \|\boldsymbol{e}\|_2^2$. To see why this is the case, recall from our discussion of alternative forms of the SVD that when $R < M$, we can form the matrix

$$\widetilde{\boldsymbol{U}} = \left[\ \boldsymbol{U} \mid \boldsymbol{U}_0\ \right],$$

where $\widetilde{\boldsymbol{U}}$ is an $M \times M$ orthonormal matrix. It follows immediately that

$$\|\widetilde{\boldsymbol{U}}^{\mathrm{T}}\boldsymbol{e}\|_2^2 = \boldsymbol{e}^{\mathrm{T}}\widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{U}}^{\mathrm{T}}\boldsymbol{e} = \boldsymbol{e}^{\mathrm{T}}\boldsymbol{e} = \|\boldsymbol{e}\|_2^2,$$

since $\widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{U}}^{\mathrm{T}} = \mathbf{I}$. But it should also be clear that since $\widetilde{\boldsymbol{U}}^{\mathrm{T}}\boldsymbol{e}$ is just the vector of inner products between the columns in $\widetilde{\boldsymbol{U}}$ and $\boldsymbol{e}$, and $\widetilde{\boldsymbol{U}}$ is just the concatenation of $\boldsymbol{U}$ and $\boldsymbol{U}_0$,

$$\|\widetilde{\boldsymbol{U}}^{\mathrm{T}}\boldsymbol{e}\|_2^2 = \|\boldsymbol{U}^{\mathrm{T}}\boldsymbol{e}\|_2^2 + \|\boldsymbol{U}_0^{\mathrm{T}}\boldsymbol{e}\|_2^2,$$

and hence we must have $\|\boldsymbol{U}^{\mathrm{T}}\boldsymbol{e}\|_2^2 \leq \|\boldsymbol{e}\|_2^2$.

With this fact in hand, we can consider a slight "relaxation" of (4) to

$$\underset{\boldsymbol{\beta}\in\mathbb{R}^R}{\text{maximize}} \ \ \|\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}\|_2^2 \quad \text{subject to} \ \ \|\boldsymbol{\beta}\|_2 = 1.$$

We are not explicitly enforcing the fact that $\boldsymbol{\beta}$ should be a linear combination of the columns of $\boldsymbol{U}$ here, which is why I am calling this a relaxation of the original problem. However, this problem has a simple solution that you will verify in the homework. Specifically the worst case $\boldsymbol{\beta}$ will have a 1 in the entry corresponding to the largest entry in $\boldsymbol{\Sigma}^{-1}$, and will be zero everywhere else. Thus

$$\max_{\boldsymbol{\beta}\in\mathbb{R}^R:\|\boldsymbol{\beta}\|_2=1} \|\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}\|_2^2 \;=\; \max_{r=1,\ldots,R} \sigma_r^{-2} \;=\; \frac{1}{\sigma_R^2}.$$

(Recall that by convention, we order the singular values so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_R$.)

Note that, even though we ignored the constraint that $\boldsymbol{\beta}$ should be a linear combination of the columns of $\boldsymbol{U}$ above, it is easy to find an $\boldsymbol{e}$ such that $\boldsymbol{\beta} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{e}$ gives us a 1 in the entry corresponding to σ_R. In particular, $\boldsymbol{e} = \boldsymbol{u}_R$ gives us precisely this $\boldsymbol{\beta}$, and so even though we ignored this constraint, the result would not have changed had we included this requirement.

Returning to the reconstruction error (2), we now see that

$$\|\widehat{\boldsymbol{x}}_{\mathrm{ls}} - \widehat{\boldsymbol{x}}_{\mathrm{clean}}\|_2^2 \;=\; \|\boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{e}\|_2^2 \;\leq\; \frac{1}{\sigma_R^2}\|\boldsymbol{e}\|_2^2.$$

While this is a worst-case bound, notice that if σ_R is small, the worst case reconstruction error can be **very bad**. Later in this course we will discuss some strategies to mitigate this fact.

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 $\boldsymbol{x}$ through a $M \times N$ matrix $\boldsymbol{A}$:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}. \tag{5}$$

The matrix $\boldsymbol{A}$ has SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$, and pseudo-inverse $\boldsymbol{A}^{\dagger} = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathrm{T}}$.

At this point it is useful to recall that we could write the matrix $\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}$ as a sum of outer products of the columns of $\boldsymbol{U}$ and $\boldsymbol{V}$:

$$\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}} = \sum_{r=1}^{R} \boldsymbol{u}_r \boldsymbol{v}_r^{\mathrm{T}},$$

where R is the rank of $\boldsymbol{A}$, 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. See the addendum at the end of these notes on matrix multiplication if this way of writing a matrix product seems unfamiliar.

Since $\boldsymbol{\Sigma}$ is diagonal, we can think of $\boldsymbol{U}\boldsymbol{\Sigma}$ as just rescaling the columns of $\boldsymbol{U}$, so that we can also rewrite $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$ as a sum:

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

where the σ_r are the singular values. Similarly, we can write the pseudo-inverse as

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

Given $\boldsymbol{y}$ as above, we can write the least-squares estimate of $\boldsymbol{x}$ from the noisy measurements as

$$\widehat{\boldsymbol{x}}_{\mathrm{ls}} = \boldsymbol{A}^{\dagger}\boldsymbol{y} = \sum_{r=1}^{R} \frac{1}{\sigma_r}\langle \boldsymbol{y}, \boldsymbol{u}_r\rangle \boldsymbol{v}_r. \tag{6}$$

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 (6), 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., not to deviate from the noiseless least squares solution by too much).

We form an approximation $\boldsymbol{A}'$ to $\boldsymbol{A}$ by taking

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

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 $\boldsymbol{A}'$ has rank R'. Note that the pseudo-inverse of $\boldsymbol{A}'$ is also a truncated sum

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

Given noisy data $\boldsymbol{y}$ as in (5), we reconstruct $\boldsymbol{x}$ by applying the truncated pseudo-inverse to $\boldsymbol{y}$:

$$\widehat{\boldsymbol{x}}_{\text{trunc}} = \boldsymbol{A}'^{\dagger}\boldsymbol{y} = \sum_{r=1}^{R'} \frac{1}{\sigma_r} \langle \boldsymbol{y}, \boldsymbol{u}_r \rangle \boldsymbol{v}_r.$$

How good is this reconstruction? To answer this question, we will again compare it to the least squares reconstruction corresponding to "noiseless" measurements $\widehat{\boldsymbol{x}}_{\text{clean}} = \boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{x}$. The difference between these two is the reconstruction error (relative to $\widehat{\boldsymbol{x}}_{\text{clean}}$) as

$$\begin{aligned}
\widehat{\boldsymbol{x}}_{\text{trunc}} - \widehat{\boldsymbol{x}}_{\text{clean}} &= \boldsymbol{A}'^{\dagger}\boldsymbol{y} - \boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{x} \\
&= \boldsymbol{A}'^{\dagger}\boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}'^{\dagger}\boldsymbol{e} - \boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{x} \\
&= (\boldsymbol{A}'^{\dagger} - \boldsymbol{A}^{\dagger})\boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}'^{\dagger}\boldsymbol{e}.
\end{aligned}$$

Proceeding further, we can write the matrix $\boldsymbol{A}'^{\dagger} - \boldsymbol{A}^{\dagger}$ as

$$\boldsymbol{A}'^{\dagger} - \boldsymbol{A}^{\dagger} = \sum_{r=R'+1}^{R} -\frac{1}{\sigma_r} \boldsymbol{v}_r \boldsymbol{u}_r^{\mathrm{T}},$$

and so the first term in the reconstruction error can be written as

$$\begin{aligned}
(\boldsymbol{A}'^{\dagger} - \boldsymbol{A}^{\dagger})\boldsymbol{A}\boldsymbol{x} &= \sum_{r=R'+1}^{R} -\frac{1}{\sigma_r} \langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{u}_r \rangle \boldsymbol{v}_r \\
&= \sum_{r=R'+1}^{R} -\frac{1}{\sigma_r} \left\langle \sum_{j=1}^{R} \sigma_j \langle \boldsymbol{x}, \boldsymbol{v}_j \rangle \boldsymbol{u}_j, \boldsymbol{u}_r \right\rangle \boldsymbol{v}_r \\
&= \sum_{r=R'+1}^{R} -\frac{1}{\sigma_r} \sum_{j=1}^{R} \sigma_j \langle \boldsymbol{x}, \boldsymbol{v}_j \rangle \langle \boldsymbol{u}_j, \boldsymbol{u}_r \rangle \boldsymbol{v}_r \\
&= \sum_{r=R'+1}^{R} -\langle \boldsymbol{x}, \boldsymbol{v}_r \rangle \boldsymbol{v}_r \quad \text{(since } \langle \boldsymbol{u}_r, \boldsymbol{u}_j \rangle = 0 \text{ unless } j = r\text{)}.
\end{aligned}$$

The second term in the reconstruction error can also be expanded against the $\boldsymbol{v}_r$:

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

Combining these expressions, the reconstruction error can be written

$$\begin{aligned} \widehat{\boldsymbol{x}}_{\text{trunc}} - \widehat{\boldsymbol{x}}_{\text{clean}} &= \underbrace{\sum_{r=1}^{R'} \frac{1}{\sigma_r} \langle \boldsymbol{e}, \boldsymbol{u}_r \rangle \boldsymbol{v}_r} \quad + \quad \underbrace{\sum_{r=R'+1}^{R} -\langle \boldsymbol{x}, \boldsymbol{v}_r \rangle \boldsymbol{v}_r} \\ &= \quad \text{Noise error} \quad + \quad \text{Approximation error.} \end{aligned}$$

Since the $\boldsymbol{v}_r$ are mutually orthogonal, and the two sums run over disjoint index sets, the noise error and the approximation error will be orthogonal. Also

$$\begin{aligned} \|\widehat{\boldsymbol{x}}_{\text{trunc}} - \widehat{\boldsymbol{x}}_{\text{clean}}\|_2^2 &= \|\text{Noise error}\|_2^2 + \|\text{Approximation error}\|_2^2 \\ &= \sum_{r=1}^{R'} \frac{1}{\sigma_r^2} |\langle \boldsymbol{e}, \boldsymbol{u}_r \rangle|^2 + \sum_{r=R'+1}^{R} |\langle \boldsymbol{x}, \boldsymbol{v}_r \rangle|^2. \end{aligned}$$

The reconstruction error, then, is signal dependent and will depend on how much of the vector $\boldsymbol{x}$ is concentrated in the subspace spanned by $\boldsymbol{v}_{R'+1}, \ldots, \boldsymbol{v}_R$. We will lose everything in this subspace. On the other hand, if it contains a significant part of $\boldsymbol{x}$, then there is not much least squares can do for you.

The worst-case noise error occurs when $\boldsymbol{e}$ is aligned with $\boldsymbol{u}_{R'}$:

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

Stable Reconstruction using Tikhonov Regularization

Tikhonov[1] 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 $\boldsymbol{A}$.

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

$$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{minimize}} \ \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2. \tag{7}$$

When $\boldsymbol{A}$ has full column rank, $\widehat{\boldsymbol{x}}_{\text{ls}}$ 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 $\boldsymbol{v}_r$ corresponding to the tiny σ_r) can have very little effect on the residual $\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2$. As such, the solution to (7) can have wildly inaccurate components in the presence of even mild noise.

One way to counteract this problem is to modify (7) with a **regularization** term that penalizes the size of the solution $\|\boldsymbol{x}\|_2^2$ as well as the residual error $\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2$:

$$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{minimize}} \ \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \delta\|\boldsymbol{x}\|_2^2. \tag{8}$$

The parameter $\delta > 0$ gives us a trade-off between accuracy and regularization; we want to choose δ small enough so that the residual for the solution of (8) is close to that of (7), and large enough so that the problem is well-conditioned (i.e., stable in the presence of noise).

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

Just as with (7), which is solved by applying the pseudo-inverse to $\boldsymbol{y}$, we can write the solution to (8) in closed form.

We have two ways of writing this solution. In terms of the SVD of $\boldsymbol{A}$, the minimizer of (8) is

$$\begin{aligned}\widehat{\boldsymbol{x}}_{\text{tik}} &= \boldsymbol{V}\widehat{\boldsymbol{\alpha}}_{\text{tik}} \\ &= \boldsymbol{V}(\boldsymbol{\Sigma}^2 + \delta\mathbf{I})^{-1}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{y}. \end{aligned} \tag{9}$$

Alternatively, we can write the solution in terms of $\boldsymbol{A}$ as

$$\widehat{\boldsymbol{x}}_{\text{tik}} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \delta\mathbf{I})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}. \tag{10}$$

Note that even if $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is not invertible $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}+\delta\mathbf{I}$ always is (if $\delta > 0$). Thus, the expression (10) holds for all M, N, and R. It is also the case that

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

You will show that all of these expressions are equivalent on the next homework.

Tikhonov regularization is in some sense very similar to the truncated SVD, but with one significant advantage: because of the second way of writing the solution, we do not need to explicitly calculate the SVD to solve (8). The importance of not needing to explicitly compute the SVD is significant when we are solving large problems. When $\boldsymbol{A}$ is large ($M, N > 10^5$, say) it may be expensive or even impossible to construct the SVD and compute with it explicitly. However, if it has special structure (if it is sparse, for example), then it may take many fewer than MN operations to compute a matrix vector product $\boldsymbol{A}\boldsymbol{x}$.

In these situations, a **matrix free** iterative algorithm can be used to perform the inverse required in (10). A prominent example of such an algorithm is **gradient descent** and its many variants, which we will see very soon.

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

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

The Tikhonov reconstruction $\widehat{\boldsymbol{x}}_{\mathrm{tik}}$ derived above is

$$\widehat{\boldsymbol{x}}_{\mathrm{tik}} = \sum_{r=1}^{R} \frac{\sigma_r}{\sigma_r^2 + \delta} \langle \boldsymbol{y}, \boldsymbol{u}_r \rangle \boldsymbol{v}_r. \tag{11}$$

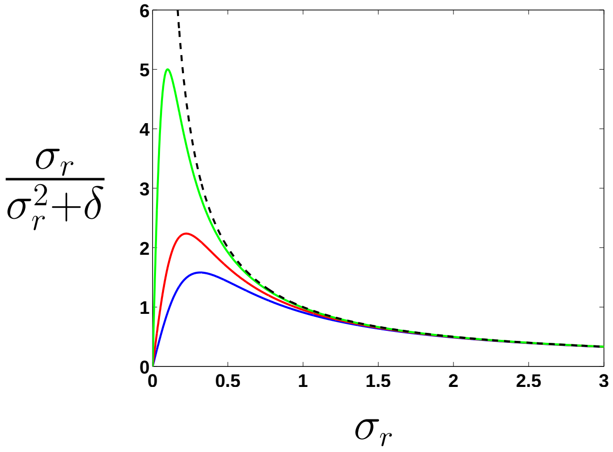
Notice that when σ_r is much larger than δ,

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

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 (11). 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.

You can perform a very similar kind of noise analysis for Tikhonov reconstruction as we just did for the truncated SVD, but we will leave you to do this at home.