Matrix Problems

From now on, the coefficient matrix A is allowed to have more rows than columns, i.e.,

$$A \in \mathbb{R}^{m \times n} \qquad \text{with} \qquad m \geq n.$$

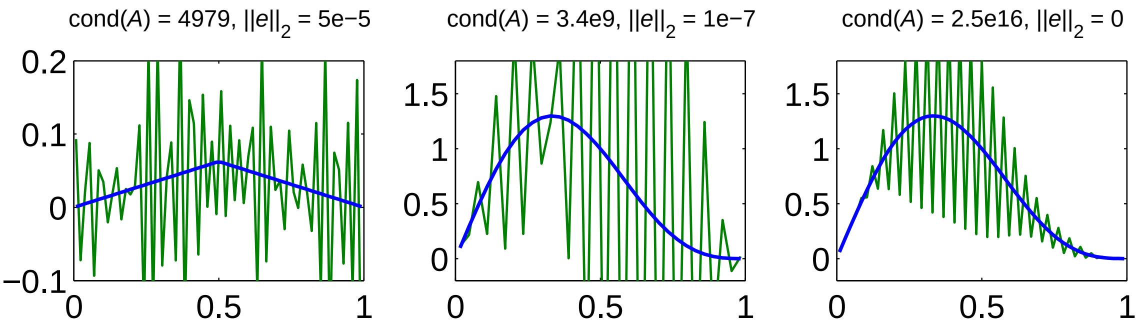
For $m > n$ it is natural to consider the least squares problem $\min_x \|A\,x - b\|_2$.

When we say "naive solution" we either mean the solution $A^{-1}b$ (when $m = n$) or the least squares solution (when $m > n$).

We emphasize the convenient fact that the naive solution has precisely the same SVD expansion in both cases:

$$x^{\text{naive}} = \sum_{i=1}^{n} \frac{u_i^T b}{\sigma_i}\, v_i.$$

Naive Solutions are Useless



Exact solutions (blue smooth lines) together with the naive solutions (jagged green lines) to two test problems.

Left: `deriv2` with $n = 64$.

Middle and right: `gravity` with $n = 32$ and $n = 53$.

Need For Regularization

Discrete ill-posed problems are characterized by having coefficient matrices with a very large condition number.

The naive solution is very sensitive to any perturbation of the right-hand side, representing the errors in the data.

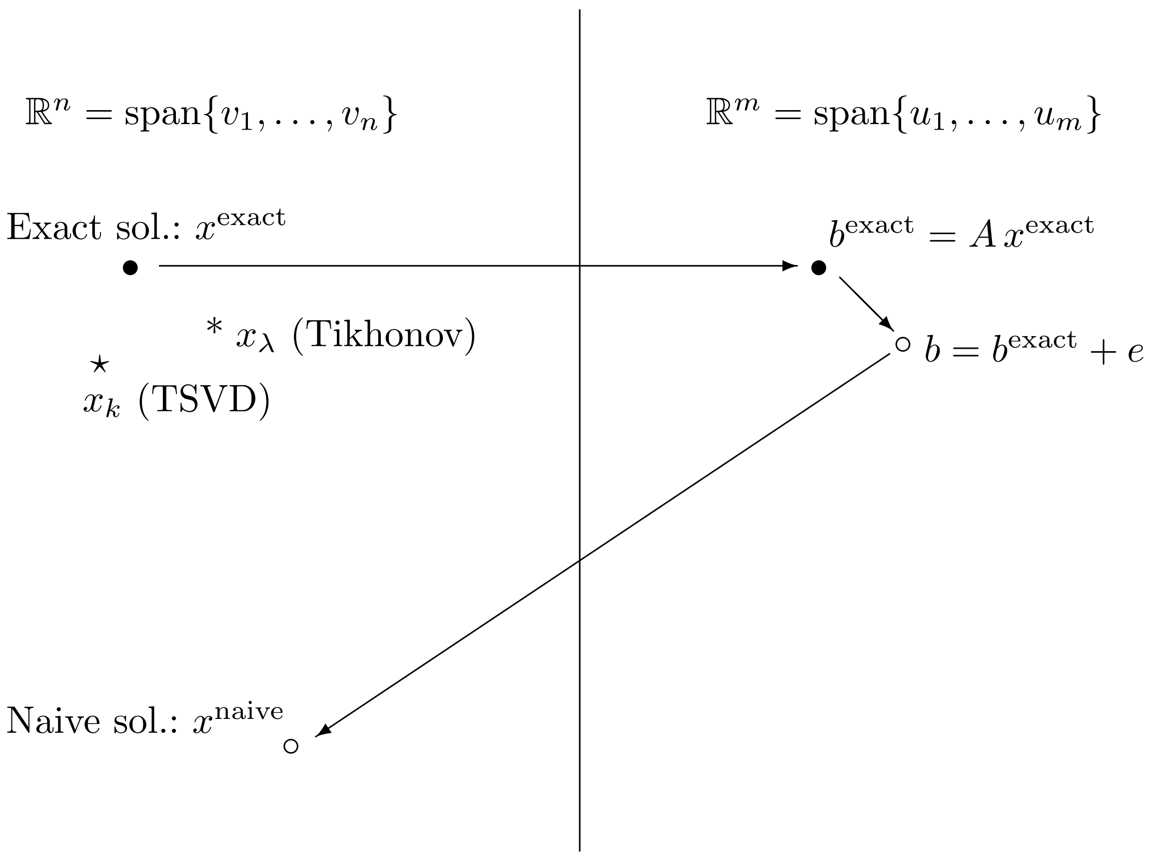
Specifically, assume that the exact and perturbed solutions x^{exact} and x satisfy

$$A\,x^{\text{exact}} = b^{\text{exact}}, \qquad A\,x = b = b^{\text{exact}} + e,$$

where e denotes the perturbation. Then classical perturbation theory leads to the bound

$$\frac{\|x^{\text{exact}} - x\|_2}{\|x^{\text{exact}}\|_2} \leq \text{cond}(A)\,\frac{\|e\|_2}{\|b^{\text{exact}}\|_2}.$$

Since $\text{cond}(A) = \sigma_1/\sigma_n$ is large, this implies that x can be very far from x^{exact}.



Regularization Methods → Spectral Filtering

Almost all the regularization methods treated in this course produce solutions which can be expressed as a filtered SVD expansion of the form

$$x_{\mathrm{reg}} = \sum_{i=1}^{n} \varphi_i \, \frac{u_i^T b}{\sigma_i} \, v_i,$$

where φ_i are the *filter factors* associated with the method.

These methods are called *spectral filtering methods* because the SVD basis can be considered as a spectral basis.

Truncated SVD

A simple way to reduce the influence of the noise is to discard the SVD coefficients corresponding to the smallest singular values.

Define truncated SVD (TSVD) solution as

$$x_k = \sum_{i=1}^{k} \frac{u_i^T b}{\sigma_i} \, v_i, \qquad k < n.$$

Reg. Tools: `tsvd`. Can show that if $\mathrm{Cov}(b) = \eta^2 I$ then

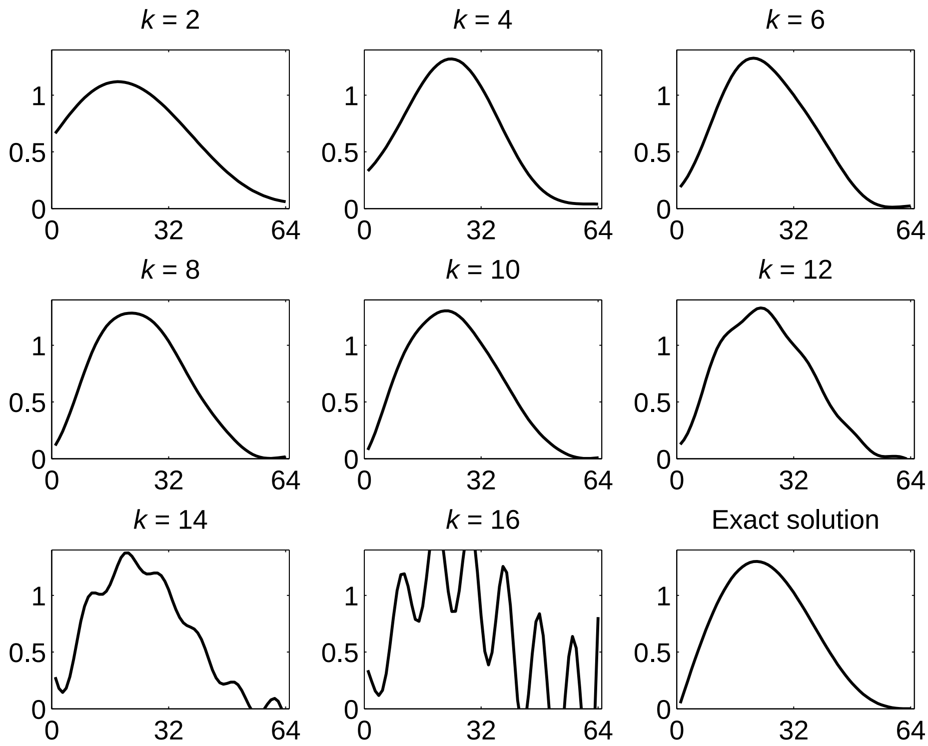
$$\mathrm{Cov}(x_k) = \eta^2 \sum_{i=1}^{k} \frac{1}{\sigma_i^2} \, v_i \, v_i^T$$

and thus we can expect that

$$\|x_k\|_2 \ll \|x^{\mathrm{naive}}\|_2 \qquad \text{and} \qquad \|\mathrm{Cov}(x_k)\|_2 \ll \|\mathrm{Cov}(x^{\mathrm{naive}})\|_2 .$$

The prize we pay for smaller covariance is *bias:* $\mathcal{E}(x_k) \neq \mathcal{E}(x^{\mathrm{naive}})$.

Truncated SVD Solutions



The Truncation Parameter

Note: the truncation parameter k in

$$x_k = \sum_{i=1}^{k} \frac{u_i^T b}{\sigma_i} \, v_i$$

is dictated by the coefficients $u_i^T b$, not the singular values!

Basically we should choose k as the index i where $|u_i^T b|$ start to "level off" due to the noise.

Selective SVD

Consider a problem in which, say, every second SVD component is zero ($v_2^T x^{\text{exact}} = v_4^T x^{\text{exact}} = v_6^T x^{\text{exact}} = \ldots = 0$). There is no need to include these SVD components.

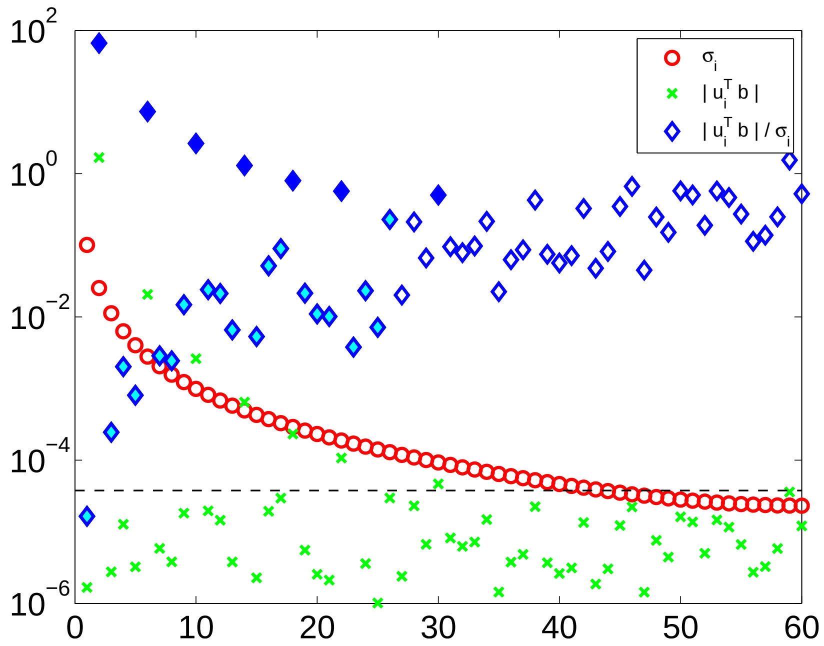
A variant of the TSVD method called *selective SVD* (SSVD) includes, or selects, only those SVD components which make significant contributions to the regularized solution:

$$x_\tau \equiv \sum_{|u_i^T b| > \tau} \frac{u_i^T b}{\sigma_i} v_i .$$

Thus, the filter factors for the SSVD method are

$$\varphi_i^{[\tau]} = \begin{cases} 1, & |u_i^T b| \geq \tau \\ 0, & \text{otherwise.} \end{cases}$$

SSVD Example



Only the filled diamonds contribute to the SSVD solution.

Regularization – A General Approach

Regularization = stabilization: how to deal with (and filter) solution components corresponding to the small singular values.

Most approaches involve the residual norm

$$\rho(f) = \left\| \int_0^1 K(s,t)\, f(t)\, dt - g(s) \right\|_2 ,$$

and a *smoothing norm* $\omega(f)$ that measure the "size" of the solution f. Example of a common choices:

$$\omega(f)^2 = \int_0^1 |f(t)|^2\, dt \qquad \text{or} \qquad \omega(f)^2 = \int_0^1 |f^{(p)}(t)|^2\, dt$$

1. Minimize $\rho(f)$ s.t. $\omega(f) \leq \delta$.
2. Minimize $\omega(f)$ s.t. $\rho(f) \leq \alpha$.
3. Tikhonov: $\min_f \left\{ \rho(f)^2 + \lambda^2\, \omega(f)^2 \right\}$.

Discrete Tikhonov Regularization

Replace the continuous problem with a linear algebra problem.

Minimization of the residual ρ is replaced by

$$\min_x \|A\, x - b\|_2 , \qquad A \in \mathbb{R}^{m \times n} ,$$

where A and b are obtained by discretization of the integral eq.

Must also discretize the smoothing norm

$$\Omega(x) \approx \omega(f).$$

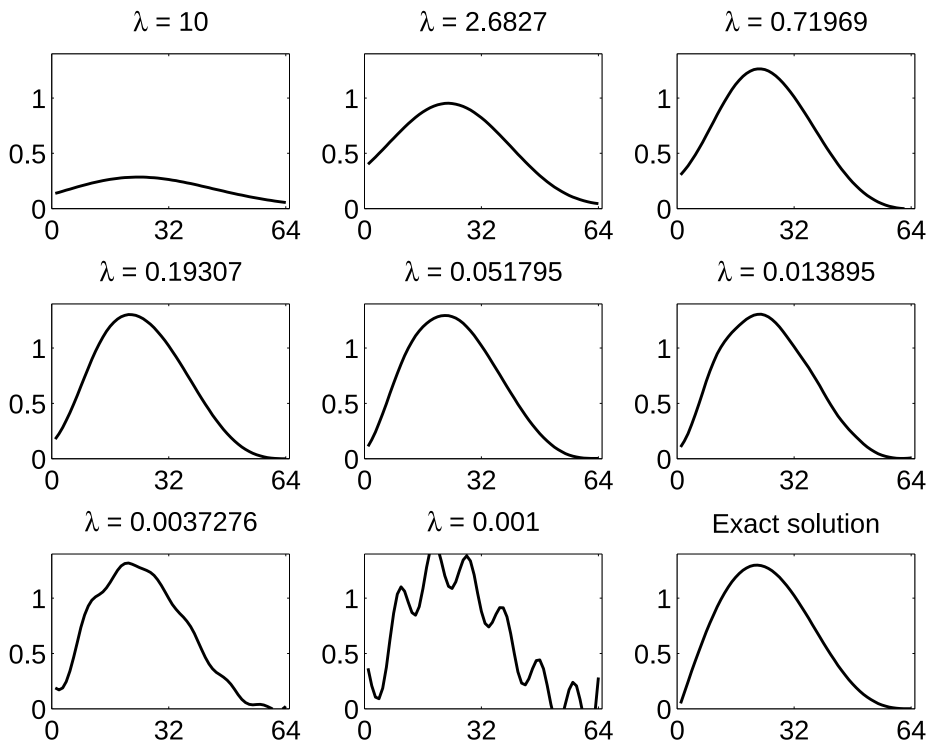
We focus on a common choice: $\Omega(x) = \|x\|_2$.

The resulting discrete Tikhonov problem is thus

$$\boxed{\min_x \left\{ \|A\, x - b\|_2^2 + \lambda^2\, \|x\|_2^2 \right\}.}$$

Regularization Tools: `tikhonov`.

Tikhonov Solutions



Other Smoothing Norms → Chapter 8

Another common choice:

$$\Omega(x) = \|L\,x\|_2,$$

where L approximates a derivative operator.

Examples of the 1. and 2. derivative operator on a regular mesh

$$L_1 = \begin{pmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{pmatrix} \in R^{(n-1)\times n}$$

$$L_2 = \begin{pmatrix} 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \end{pmatrix} \in R^{(n-2)\times n}.$$

Regularization Tools: `get_l`.

Efficient Implementation

The original formulation

$$\min \left\{ \|A\,x - b\|_2^2 + \lambda^2 \, \|x\|_2^2 \right\} .$$

Two alternative formulations

$$(A^T A + \lambda^2 I)\, x = A^T b$$

$$\min \left\| \begin{pmatrix} A \\ \lambda\, I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|_2$$

The first shows that we have a linear problem. The second shows how to solve it stably:

- treat it as a least squares problem,
- utilize any sparsity or structure.

SVD and Tikhonov Regularization

We can write the discrete Tikhonov solution x_λ in terms of the SVD of A as

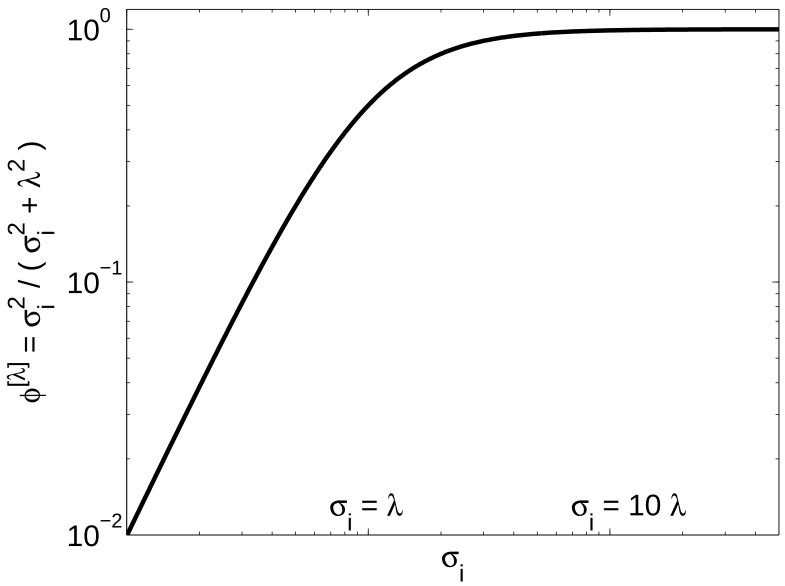
$$x_\lambda = \sum_{i=1}^{n} \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \, \frac{u_i^T b}{\sigma_i} \, v_i = \sum_{i=1}^{n} \phi_i^{[\lambda]} \, \frac{u_i^T b}{\sigma_i} .$$

The *filter factors* are given by

$$\phi_i^{[\lambda]} = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \; ,$$

and their purpose is to dampen the components in the solution corresponding to small σ_i.

Tikhonov Filter Factors

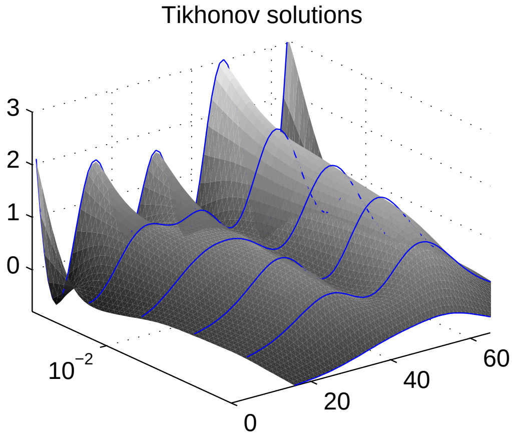


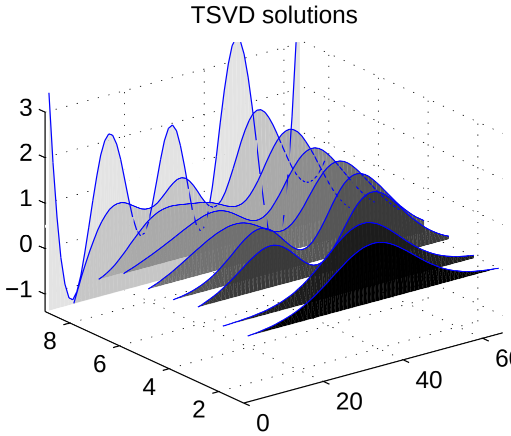
$$\phi_i^{[\lambda]} = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \approx \begin{cases} 1\ , & \sigma_i \gg \lambda \\ \sigma_i^2/\lambda^2\ , & \sigma_i \ll \lambda. \end{cases}$$

TSVD and Tikhonov Regularization

TSVD and Tikhonov solutions are both filtered SVD expansions.

The regularization parameter is either k or λ.





For each k, there exists a λ such that $x_\lambda \approx x_k$.

Wiener Filtering

In certain applications, e.g., in image deblurring, the SVD basis vectors u_i and v_i can be replaced by the discrete Fourier vectors (that underly the discrete Fourier transform).

In these applications, Tikhonov regularization is known as Wiener filtering. It is typically derived in a stochastic setting.

Here, λ^{-2} is the signal-to-noise power, i.e., the power of the exact solution divided by the power of the noise in the right-hand side.

Available in MATLAB's Image Processing Toolbox as `deconvwnr`.

Other Spectral Filtering Methods

A few spectral filtering methods not mentioned in the book.

- Damped SVD:

$$\varphi_i^{[\lambda]} = \frac{\sigma_i}{\sigma_i + \lambda}, \qquad \lambda \geq 0.$$

- Exponential filtering:

$$\varphi_i^{[\beta]} = 1 - \exp(-\beta\, \sigma_i^2), \qquad \beta \geq 0.$$

Regularization Tools: `fil_fac` computers filter factors for DSVD, TSVD, Tikhonov, and TTLS (not covered here).

TSVD Perturbation Bound

Theorem.

Let $b = b^{\mathrm{exact}} + e$ and let x_k and x_k^{exact} denote the TSVD solutions computed with the *same* k.

Then

$$\frac{\|x_k^{\mathrm{exact}} - x_k\|_2}{\|x_k\|_2} \leq \frac{\sigma_1}{\sigma_k} \frac{\|e\|_2}{\|A\,x_k\|_2}.$$

We see that the condition number for the TSVD solution is

$$\kappa_k = \frac{\sigma_1}{\sigma_k}$$

and it can be much smaller than $\mathrm{cond}(A) = \sigma_1/\sigma_n$.

Tikhonov Perturbation Bound

Theorem.

Let $b = b^{\mathrm{exact}} + e$ and let $x_\lambda^{\mathrm{exact}}$ and x_λ denote the solutions to

$$\min\{\|A\,x - b^{\mathrm{exact}}\|_2^2 + \lambda^2\|x\|_2^2\} \quad \text{and} \quad \min\{\|A\,x - b\|_2^2 + \lambda^2\|x\|_2^2\}$$

computed with the *same* λ.

Then

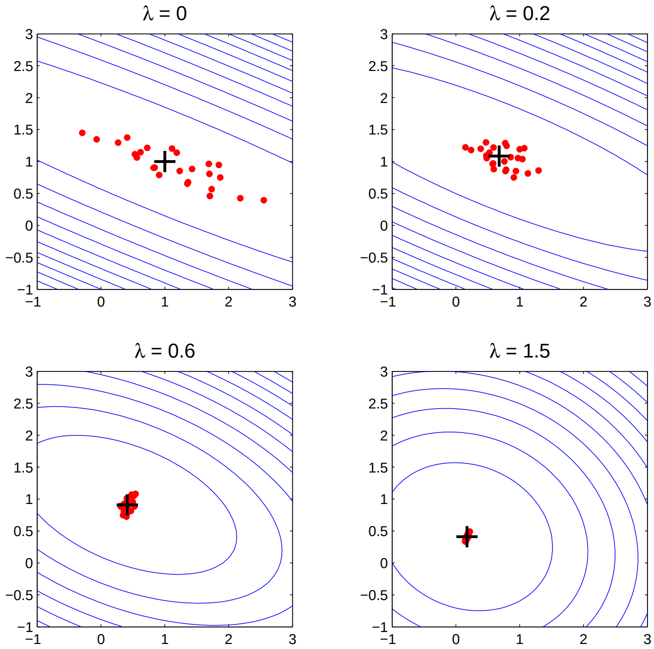
$$\frac{\|x_\lambda^{\mathrm{exact}} - x_\lambda\|_2}{\|x_\lambda\|_2} \leq \frac{\|A\|_2}{\lambda} \frac{\|e\|_2}{\|A\,x_\lambda\|_2}$$

and hence the condition number for the Tikhonov solution is

$$\kappa_\lambda = \frac{\|A\|_2}{\lambda} = \frac{\sigma_1}{\lambda}.$$

Again it can be much smaller than $\mathrm{cond}(A) = \sigma_1/\sigma_n$.

Illustration of Sensitivity



Red dots: x_λ for 25 random perturbations of b.

Black crosses: unperturbed x_λ – note the bias.

Monotonic Behavior of the Norms

The TSVD solution and residual norms vary monotonically with k

$$\|x_k\|_2^2 = \sum_{i=1}^{k} \left(\frac{u_i^T b}{\sigma_i} \right) \leq \|x_{k+1}\|_2^2 \quad \text{(we assume } m = n\text{)},$$

$$\|A\, x_k - b\|_2^2 = \sum_{i=k+1}^{n} (u_i^T b)^2 \geq \|A\, x_{k+1} - b\|_2^2.$$

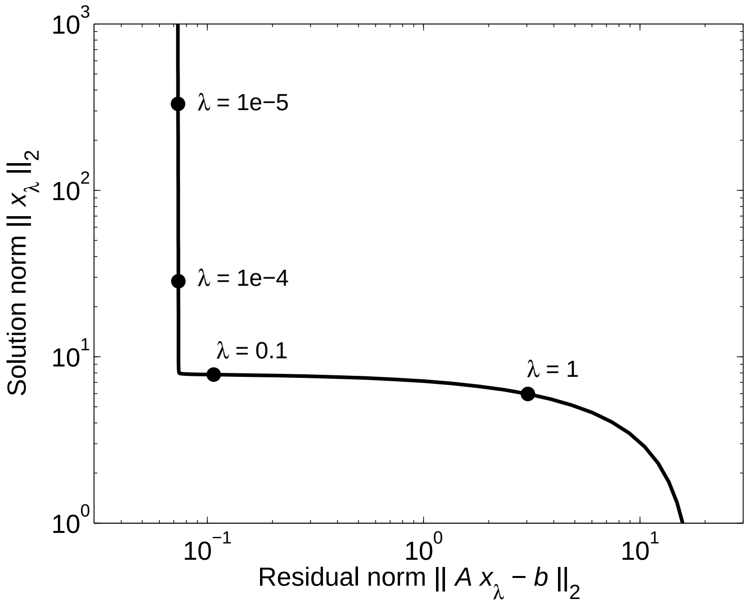
Similarly, the Tikhonov solution and residual norms vary monotonically with λ:

$$\|x_\lambda\|_2^2 = \sum_{i=1}^{n} \left(\phi_i^{[\lambda]} \frac{u_i^T b}{\sigma_i} \right)^2,$$

$$\|A\, x_\lambda - b\|_2^2 = \sum_{i=1}^{n} \left(1 - \phi_i^{[\lambda]} \right) u_i^T b \Big)^2.$$

The L-Curve for Tikhonov Regularization

Plot of $\|x_\lambda\|_2$ versus $\|A\,x_\lambda - b\|_2$ in *log-log scale.*



Properties of the L-Curve

The norm $\|x_\lambda\|_2$ is a monotonically decreasing convex function of the norm $\|A\,x_\lambda - b\|_2$.

Define the "inconsistency"

$$\delta_0^2 = \sum_{i=n+1}^{m} (u_i^T b)^2 \qquad (= 0 \text{ when } m = n.)$$

Then

$$\delta_0 \le \|A\,x_\lambda - b\|_2 \le \|b\|_2$$
$$0 \le \|x_\lambda\|_2 \le \|x^{\text{naive}}\|_2 \ .$$

Any point (δ, η) on the L-curve is a solution to the following two inequality-constrained least squares problems:

$$\delta = \min_x \|A\,x - b\|_2 \qquad \text{subject to} \qquad \|x\|_2 \le \eta$$
$$\eta = \min_x \|x\|_2 \qquad \text{subject to} \qquad \|A\,x - b\|_2 \le \delta \ .$$

More Properties

When λ is large, then x_λ is dominated by SVD coefficients whose main contribution is from the exact right-hand side b^{exact} – the solution is *over-smoothed*.

A careful analysis shows that for large values of λ we have that

$$\|x_\lambda\|_2 \approx \|x^{\text{exact}}\|_2 \text{ (a constant)}, \qquad \|A\,x_\lambda - b\|_2 \text{ increases with } \lambda.$$

For small values of λ the Tikhonov solution is dominated by the perturbation errors coming from the inverted noise – the solution is *under-smoothed*, and we have that

$$\|x_\lambda\|_2 \text{ increases with } \lambda^{-1} \qquad \text{and} \qquad \|A\,x_\lambda - b\|_2 \approx \|e\|_2 \text{ (a constant)}.$$

Thus the L-curve has two distinctly different parts: a part that is approximately horizontal, and a part that is approximately vertical.

Log-Log Scale Separates Over- and Under-Smoothing

The features become more pronounced (and easier to inspect) when the L-curve is plotted in double-logarithmic scale:

$$(\log \|A\,x_\lambda - b\|_2\,,\ \log \|x_\lambda\|_2\,)$$

The "corner" that separates these horizontal and vertical parts is located roughly at the point

$$(\log \|e\|_2\,,\ \log \|x^{\text{exact}}\|_2\,)\ .$$

Towards the right, for $\lambda \to \infty$, the L-curve starts to bend down as the increasing amount of regularization forces the solution norm towards zero.