Linear Least Squares

Klaus Mosegaard

Niels Bohr Institute
Copenhagen University

2009

Outline

What is a least squares problem?

Given an equation

$$f(\mathbf{x}) = \mathbf{b} \tag{1}$$

where the vector $\mathbf{b}$ and the function f are known, and the vector $\mathbf{x}$ is unknown.

Define the *misfit*:

$$E(\mathbf{x}) = \|f(\mathbf{x}) - \mathbf{b}\|^2 \tag{2}$$

The Least-Squares solution to (1) is then

$$\boxed{\hat{\mathbf{x}} = \text{Argmin } E(\mathbf{x})} \tag{3}$$

The linear least squares problem

If the relation between $\mathbf{x}$ and $\mathbf{b}$ is *linear* :

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{4}$$

the *Linear least squares problem* is to minimize

$$E(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2. \tag{5}$$

This can be done analytically, and a solution vector $\hat{\mathbf{x}}$ satisfies:

$$\forall j : \quad \frac{\partial E}{\partial \hat{x}_j} = 0 \tag{6}$$

Existence and Uniqueness

The overdetermined (overconstrained) problem

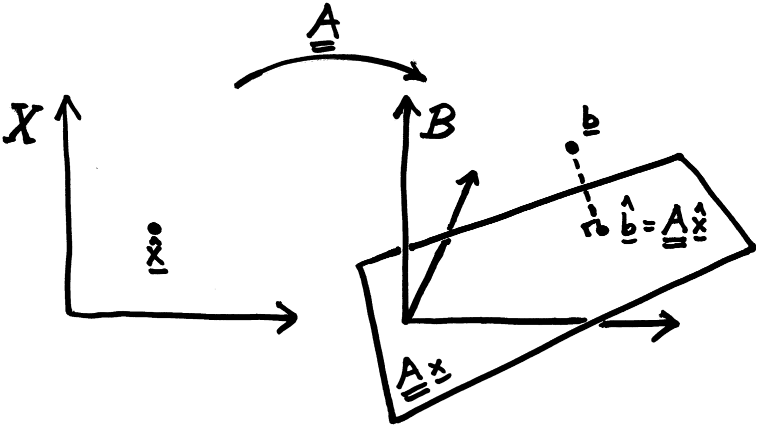


Figure: The overdetermined problem is characterized by a unique, but (usually) inexact solution.

A solution to the overdetermined problem

If the linear problem

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{7}$$

is overdetermined, minimizing the misfit

$$E(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2. \tag{8}$$

through

$$\forall j : \quad \frac{\partial E}{\partial \hat{x}_j} = 0 \tag{9}$$

leads to the following formula for the least squares estimate:

$$\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} \tag{10}$$

The underdetermined problem

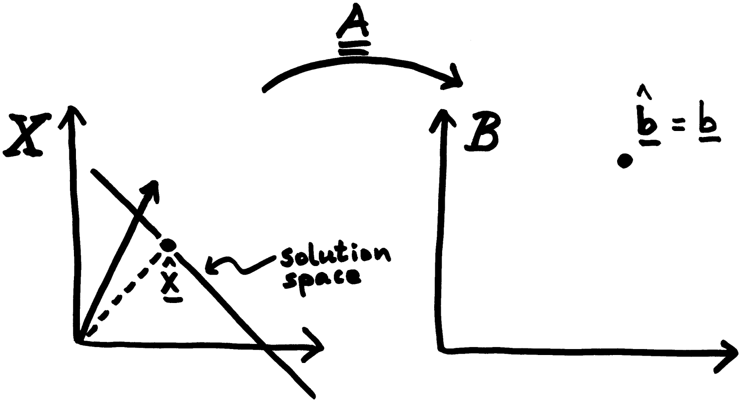


Figure: The underdetermined problem is characterized by infinitely many exact solutions.

A solution to the underdetermined problem

If the linear problem

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{11}$$

is underdetermined, minimizing the misfit

$$E(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2. \tag{12}$$

through

$$\forall j : \quad \frac{\partial E}{\partial \hat{x}_j} = 0 \tag{13}$$

leads to the following formula for the least squares estimate:

$$\hat{\mathbf{x}} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b} \tag{14}$$

The mixed-determined problem

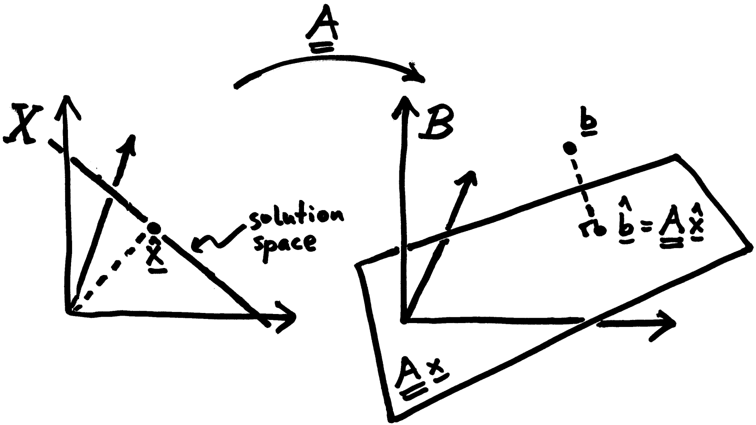


Figure: The mixed-determined problem is characterized by infinitely many (usually) inexact solutions.

An approximate solution to the mixed-determined problem

If the linear problem

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{15}$$

is mixed-determined, minimizing the modified misfit

$$E(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \epsilon^2 \|\mathbf{x}\|^2. \tag{16}$$

for suitable small ϵ leads to the following approximate formula for the least squares estimate:

$$\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A} + \epsilon^2 \mathbf{I})^{-1} \mathbf{A}^T \mathbf{b} \tag{17}$$

This method is called **Tikhonov Regularization**.

Example: The inverse geomagnetic problem

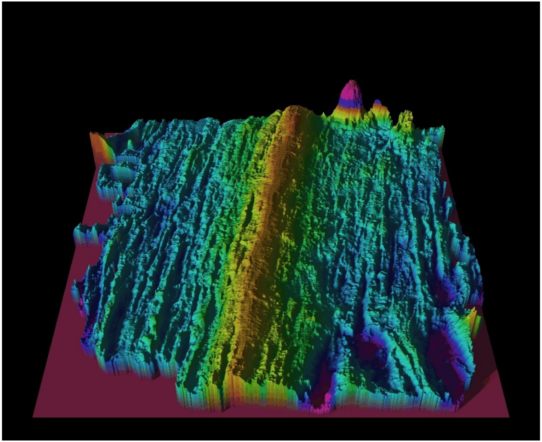


Figure: Magnetization of the ocean floor.

Model of the ocean bottom

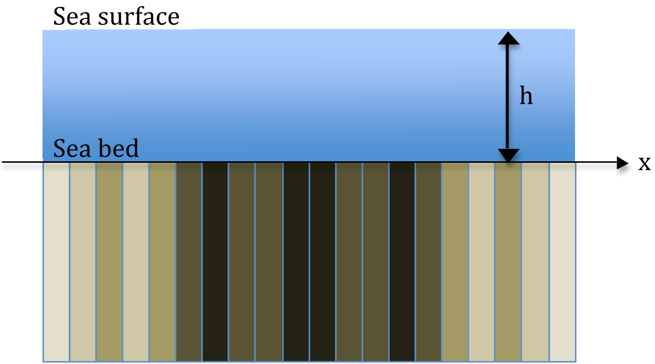


Figure: Model of the ocean bottom. The magnetization below the sea bottom is represented by a series of vertical, thin plates of constant magnetization.

Magnetic data

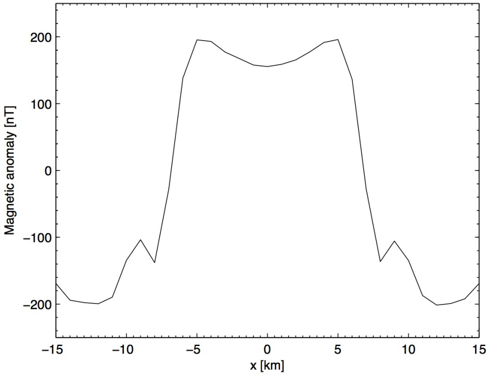


Figure: Observed vertical magnetic field profile perpendicular to the ocean ridge.

Relation between model parameters and data

If we assume that the magnetization of the ocean bottom depends only on the x-coordinate, the magnetic field d_i measured in x_i can be expressed as

$$d_i = \int_{-\infty}^{\infty} g_i(x) m(x) dx, \tag{18}$$

where $m(x)$ is the magnetization, and

$$g_i(x) = -\frac{\mu_0}{2\pi} \frac{(x_i - x)^2 - h^2}{\left[(x_i - x)^2 + h^2\right]^2} \tag{19}$$

is the magnetic field at x_i generated by an infinitesimally thin vertical "plate" of magnetized material, located at x.

Thin-plate fields

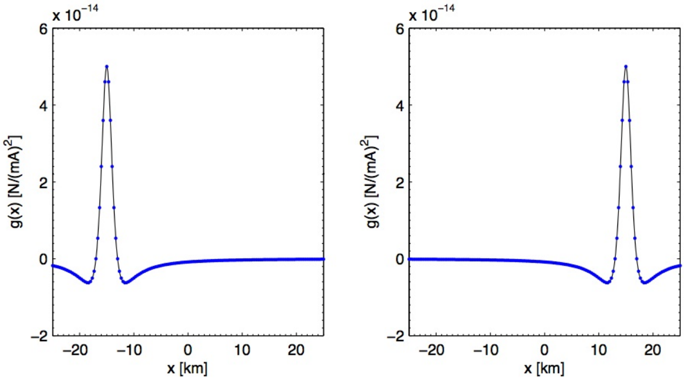


Figure: Magnetic fields from thin, vertical plates of magnetized material below the sea bottom at x = -15 km and x = 15 km

Model discretization 1

Consider a finite set of x-values: $x_1, x_2, \ldots, x_M$. Let us represent $m(x)$ by the vector:

$$\mathbf{m} = (m(x_1), m(x_2), \ldots, m(x_M)) \tag{20}$$

This leads to a discretized expression:

$$g_i(x_j) = -\frac{\mu_0}{2\pi} \frac{(x_i - x_j)^2 - h^2}{\left[(x_i - x_j)^2 + h^2\right]^2} \tag{21}$$

Model discretization 2

We can now discretize the problem:

$$\begin{aligned} d_i &= \int_{-\infty}^{\infty} g_i(x) m(x) dx \\ &\approx \sum_{k=1}^{M} g_i(x_k) m_k \Delta x \end{aligned} \tag{22}$$

Putting $G_{ij} = g_i(x_j)$, we have

$$\mathbf{d} = \mathbf{G}\mathbf{m} \tag{23}$$

which is a matrix equation relating data $\mathbf{d}$ to model parameters $\mathbf{m}$.

A least-squares solution based on Tikhonov Regularization

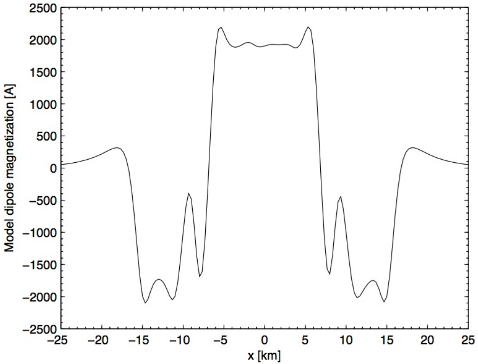


Figure: Estimated (symmetric) magnetization $\hat{\mathbf{m}}$ of the ocean bottom. The regularization parameter ϵ is chosen such that the N data are barely fitted within their uncertainty: $\|\mathbf{d}_{obs} - \mathbf{A}\hat{\mathbf{m}}\|^2 \approx N\sigma^2$

Data residuals

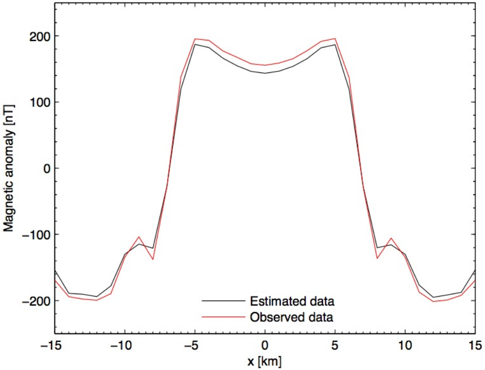


Figure: Re-computed data $\mathbf{A}\hat{\mathbf{m}}$ compared to observed data $\mathbf{d}_{obs}$.

Error propagation for overdetermined problems

If the linear problem

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{24}$$

is (purely) overdetermined, the *pseudoinverse* of $\mathbf{A}$ is defined as

$$\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T, \tag{25}$$

and the Least Squares solution is $\hat{\mathbf{x}} = \mathbf{A}^+\mathbf{b}$.

A small perturbation $\Delta\mathbf{b}$ of $\mathbf{b}$ will now give rise to a perturbation of the solution:

$$\Delta\hat{\mathbf{x}} = \mathbf{A}^+\Delta\mathbf{b}, \tag{26}$$

that is,

$$\|\Delta\hat{\mathbf{x}}\| \leq \|\mathbf{A}^+\|\|\Delta\mathbf{b}\|. \tag{27}$$

Error propagation for overdetermined problems

Let us compute the relative perturbation (error) of $\hat{\mathbf{x}}$:

$$\begin{aligned}
\frac{\|\Delta\hat{\mathbf{x}}\|}{\|\hat{\mathbf{x}}\|} &\leq \|\mathbf{A}^{+}\|\frac{\|\Delta\mathbf{b}\|}{\|\hat{\mathbf{x}}\|} \\
&= \mathrm{cond}(\mathbf{A})\frac{\|\mathbf{b}\|\cdot\|\boldsymbol{\Delta}\mathbf{b}\|}{\|\mathbf{A}\|\cdot\|\hat{\mathbf{x}}\|\cdot\|\mathbf{b}\|} \\
&\leq \mathrm{cond}(\mathbf{A})\frac{\|\mathbf{b}\|\cdot\|\boldsymbol{\Delta}\mathbf{b}\|}{\|\mathbf{A}\hat{\mathbf{x}}\|\cdot\|\mathbf{b}\|} \\
&= \mathrm{cond}(\mathbf{A})\frac{\mathbf{1}}{\cos(\theta)}\frac{\|\boldsymbol{\Delta}\mathbf{b}\|}{\|\mathbf{b}\|}
\end{aligned} \tag{28}$$

where $\mathrm{cond}(\mathbf{A}) = \|\mathbf{A}\|\|\mathbf{A}^{+}\|$ is $\mathbf{A}$'s condition number, and θ is the angle between $\mathbf{b}$ and $\mathbf{A}\hat{\mathbf{x}}$.

Solving overdetermined problems: $\mathbf{QR}$-Factorization

QR factorization

- reduces a real $n \times m$ matrix $\mathbf{A}$ with $n \geq m$ and full rank to a simple form.
- improves numerical stability by minimizing errors caused by machine roundoffs.

A suitably chosen orthogonal matrix $\mathbf{Q}$ will triangularize $\mathbf{A}$:

$$\mathbf{A} = \mathbf{Q} \begin{pmatrix} \mathbf{R} \\ \mathbf{O} \end{pmatrix} \tag{29}$$

with the $n \times n$ right triangular matrix $\mathbf{R}$.

Solving overdetermined problems: $\mathbf{QR}$-Factorization

The equation

$$\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} \tag{30}$$

now becomes

$$\begin{aligned} \mathbf{x} &= (\mathbf{R}^T\mathbf{Q}^T\mathbf{Q}\mathbf{R})^{-1}\mathbf{R}^T\mathbf{Q}^T\mathbf{b} \\ &= (\mathbf{R}^T\mathbf{R})^{-1}\mathbf{R}^T\mathbf{Q}^T\mathbf{b} \\ &= \mathbf{R}^{-1}\mathbf{Q}^T\mathbf{b} \end{aligned} \tag{31}$$

or,

$$\mathbf{R}\mathbf{x} = \mathbf{Q}^T\mathbf{b} \tag{32}$$

QR-Factorization using the Gram-Schmidt process

Let $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \ldots \mathbf{a}_M)$ and

$$\begin{aligned}
\mathbf{u}_1 &= \mathbf{a}_1 \\
\mathbf{u}_2 &= \mathbf{a}_2 - \mathrm{proj}_{\mathbf{e}_1}(\mathbf{a}_2) \\
\mathbf{u}_3 &= \mathbf{a}_3 - \mathrm{proj}_{\mathbf{e}_1}(\mathbf{a}_3) - \mathrm{proj}_{\mathbf{e}_2}(\mathbf{a}_3)
\end{aligned} \tag{33}$$

where

$$\begin{aligned}
\mathbf{e}_1 &= \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|} \\
\mathbf{e}_2 &= \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|} \\
\mathbf{e}_3 &= \frac{\mathbf{u}_3}{\|\mathbf{u}_3\|} \\
&\vdots
\end{aligned} \tag{34}$$

QR-Factorization using the Gram-Schmidt process

Now the factorization

$$\mathbf{A} = \mathbf{QR} = (\mathbf{Q}_1 \;\; \mathbf{Q}_2) \begin{pmatrix} \mathbf{R}_1 \\ \mathbf{O} \end{pmatrix} = \mathbf{Q}_1 \mathbf{R}_1 \tag{35}$$

is accomplished by

$$\mathbf{Q}_1 = (\mathbf{e}_1, \ldots \mathbf{e}_m) \tag{36}$$

and

$$\mathbf{R_1} = \begin{pmatrix} \langle \mathbf{e}_1, \mathbf{a}_1 \rangle & \langle \mathbf{e}_1, \mathbf{a}_2 \rangle & \langle \mathbf{e}_1, \mathbf{a}_3 \rangle & \ldots \\ \langle \mathbf{e}_2, \mathbf{a}_1 \rangle & \langle \mathbf{e}_2, \mathbf{a}_2 \rangle & \langle \mathbf{e}_2, \mathbf{a}_3 \rangle & \ldots \\ \langle \mathbf{e}_3, \mathbf{a}_1 \rangle & \langle \mathbf{e}_3, \mathbf{a}_2 \rangle & \langle \mathbf{e}_3, \mathbf{a}_3 \rangle & \ldots \\ \vdots & \vdots & \vdots & \end{pmatrix} \tag{37}$$

Singular Value Decomposition (SVD)

The mixed-determined problem (again)

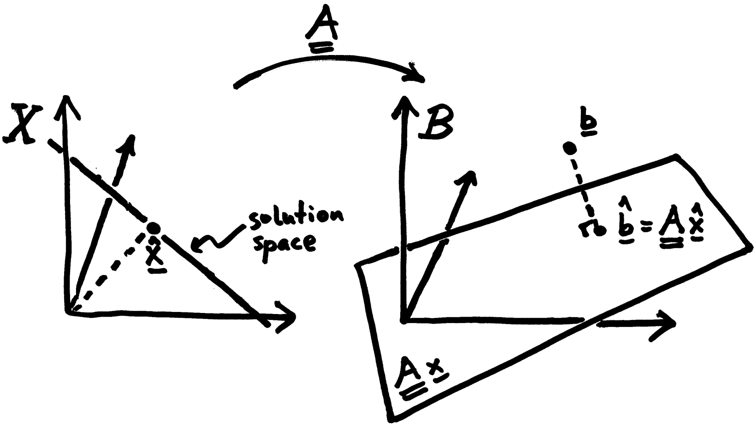
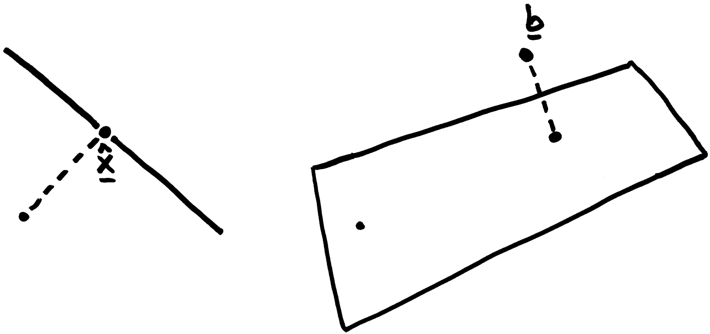
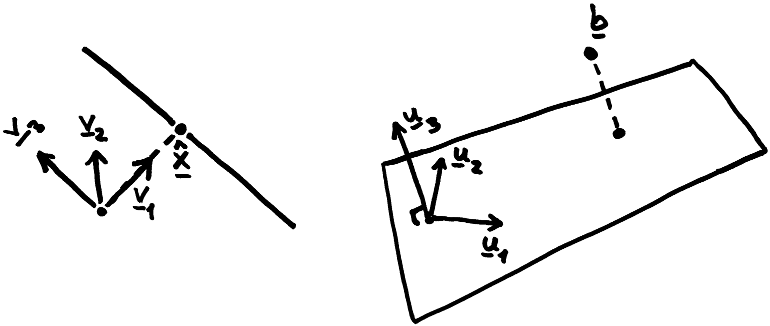


Figure: The mixed-determined problem is characterized by infinitely many (usually) inexact solutions.

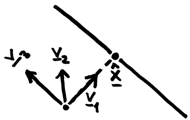
A coordinate free picture

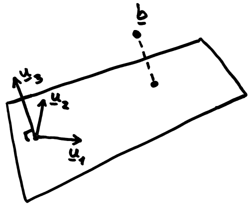


Rotated coordinate systems in X and B spaces



Rotated coordinate systems in X and B spaces





Orthogonal matrix of coordinate vectors in X:

$$\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) \tag{38}$$

Orthogonal matrix of coordinate vectors in B:

$$\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) \tag{39}$$

Singular value decomposition

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$$
$$= \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} \left\{ \begin{array}{ccc} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{array} \right\} \left\{ \begin{array}{c} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \mathbf{v}_3^T \end{array} \right\} \quad (40)$$

where

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0. \quad (41)$$

The transformed problem

If we put

$$\mathbf{x}' = \mathbf{V}^T\mathbf{x} \tag{42}$$

and

$$\mathbf{b}' = \mathbf{U}^T\mathbf{b} \tag{43}$$

we obtain

$$\begin{aligned} \mathbf{A}\mathbf{x} &= \mathbf{b} \\ \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T\mathbf{x} &= \mathbf{b} \\ \boldsymbol{\Sigma}\mathbf{V}^T\mathbf{x} &= \mathbf{U}^T\mathbf{b} \\ \boldsymbol{\Sigma}\mathbf{x}' &= \mathbf{b}' \end{aligned} \tag{44}$$

Solution to the the transformed problem

The solution is now trivial. Assume that $\lambda_1 \geq \lambda_2 > \lambda_3 = 0$. Then

$$\begin{aligned} \lambda_1 {x'}_1 = {b'}_1 \quad &\Rightarrow \quad {x'}_1 = \frac{{b'}_1}{\lambda_1} \\ \lambda_2 {x'}_2 = {b'}_2 \quad &\Rightarrow \quad {x'}_2 = \frac{{b'}_2}{\lambda_2} \\ \lambda_3 {x'}_3 = {b'}_3 \quad &\Rightarrow \quad {x'}_3 \text{ can be chosen arbitrarily} \end{aligned} \tag{45}$$

This shows that small singular values amplify noise:

If λ_i is small, a noisy ${b'}_i$ results in a very noisy ${x'}_i$!

and that zero singular values result in underdetermination:

If $\lambda_i = 0$, ${x'}_i$ is unconstrained !

Returning to the untransformed problem

Once we have found $\mathbf{x}'$, we can find $\mathbf{x}$ through

$$\mathbf{x} = \mathbf{V}\mathbf{x}' \tag{46}$$

If we have chosen the unconstrained components of $\mathbf{x}'$ to be 0, we arrive at the least squares solution:

$$\begin{aligned} \hat{\mathbf{x}} &= \mathbf{V}_p \boldsymbol{\Sigma}_p^{-1} \mathbf{U}_p^T \mathbf{b} \\ &= \{\mathbf{v}_1, \mathbf{v}_2\} \left\{ \begin{array}{cc} \lambda_1 & 0 \\ 0 & \lambda_2 \end{array} \right\}^{-1} \left\{ \begin{array}{c} \mathbf{u}_1^T \\ \mathbf{u}_2^T \end{array} \right\} \end{aligned} \tag{47}$$

Note that well-determined, ill-determined and undetermined components of $\mathbf{x}'$ mix in the expression for $\mathbf{x}$!