Contents

List of Figures

List of Tables

The Devil is in the details.

0.1 Life is complex

$\mathbf{A}^*$ or $\mathbf{A}^{\mathrm{T}}$

$$\mathbb{R} \in \mathbb{C}$$

Table 0.1. *Matrix manipulations for* $\boldsymbol{A}^*$ *and* $\boldsymbol{A}^{\mathrm{T}}$.

Application	$\mathbf{A}^*$ $\mathbf{A} \in \mathbb{C}^{m\times n}$ and $\mathbf{A} \in \mathbb{R}^{m\times n}$	$\mathbf{A}^{\mathrm{T}}$ only $\mathbf{A} \in \mathbb{R}^{m\times n}$
Fortran	`conjg(transpose(A))`	`transpose(A)`
Mathematica	$\mathtt{A}^{\mathtt{H}}$	$\mathtt{A}^{\mathtt{T}}$
	`ConjugateTranspose[A]`	`Transpose[A]`
MATLAB	`A'`	`A.'`
	`ctranspose(A)`	`transpose(A)`
Python	`A.T.conj()`	`A.T`

0.2 Readings

There are many excellent books available examining many facets of the least squares problem. Fuller references are in the bibliography.

Carl Freidrich Gauss
Theory of the Combination of Observations Least Subject to Errors

0.3 Least Squares

The titles are ranked by brevity.

Ilse C. F. Ipsen
Numerical Matrix Analysis: Linear System and Least Squares (128 pp)

Charles L. Lawson, and **Richard J. Hanson**
Solving Least Squares Problems (337 pp)

Åke Björk
Numerical Methods for Least Squares Problems (408 pp)

0.4 Linear Algebra and Matrix Analysis

The titles are ranked by brevity.

Alan J. Laub
Matrix Analysis for Scientists and Engineers (157 pp)

Carl D. Meyer
Matrix Analysis and Applied Linear Algebra (718 pp)

0.5 Numerical Linear Algebra

The titles are ranked by brevity.

Alan J. Laub
Computational Matrix Analysis (154 pp)

Lloyd N. Trefethen, and **David Bau, III**
Numerical Linear Algebra (361 pp)

E. Anderson, **Z. Bai**, **C. Bischof**, **S. Blackford**, **J. Demmel**, **J. Dongarra**, **J Du Croz**, **A. Greenbaum**, **S. Hammarling**, **A. McKenney**, and **D. Sorenson**
LAPACK Users' Guide (407 pp)

G. W. Stewart
Matrix Algorithms:
Volume I: Basic Decompositions (460 pp)
Volume II: Eigensystems (474 pp)

0.6 Discussions on Least Squares

Books with chapters dedicated to the topic. Sorted by author.

Gene H. Golub, **Charles F. Van Loan**
Matrix Computations, ch. 5, 6

Nicholas J. Higham
Accuracy and Stability of Numerical Algorithms, ch. 20

Cleve B. Moler
Numerical Computing with MATLAB, ch. 5

David S. Watkins
Numerical Analysis: a mathematical introduction, ch. 5

David S. Watkins
Fundamentals of Matrix Computations, ch. 3

Part I

Rudiments

Chapter 1

Least Squares Problems

1.1 Linear Systems

This story begins with the archetypal matrix–vector equation

$$\mathbf{A}x = b. \tag{1.1}$$

The matrix $\mathbf{A}$ has m rows, n columns, and has rank ρ; the vector b encodes m measurements. The solution vector x represents the n free parameters in the model. In mathematical shorthand,

$$\mathbf{A} \in \mathbb{C}^{m\times n}_{\rho}, \quad b \in \mathbb{C}^{m}, \quad x \in \mathbb{C}^{n} \tag{1.2}$$

with $\mathbb{C}$ representing the field of complex numbers. The matrix $\mathbf{A}$ and the vector b are given, and the task is to find the vector x.

1.1.1 $\|\mathbf{A}x - b\| = 0$

The letters in (1.1) will change, but the operation remains the same: a matrix operates on an $n-$vector and returns an $m-$vector. We can think of the matrix as a map from vectors of dimension n to vectors of dimension m:

$$\mathbf{A} \colon \mathbb{C}^{n} \mapsto \mathbb{C}^{m}.$$

If the vector b can be expressed a combination of the columns of the matrix $\mathbf{A}$ then there is a direct solution:

$$\mathbf{A}x = b \quad \Longrightarrow \quad x_1 a_1 + \cdots + x_n a_n = b$$

and the residual error vanishes:

$$\mathbf{A}x - b = \mathbf{0}$$

where the zero vector $\mathbf{0}$ is a list of m zeros. The total error, the norm of this vector, is 0.

For the problem where the system matrix $\mathbf{A}$ is the identity matrix $\mathbf{I}_2$:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix},$$

the solution is

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix};$$

there is no residual error

$$\mathbf{A}x - b = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

1.1.2 $\|\mathbf{A}x - b\| > 0$

But what happens when the vector b is *not* in the column space of the matrix $\mathbf{A}$? The solution criteria must relax. Instead of seeking zero residual error, seek minimal residual error

$$r = \mathbf{A}x - b.$$

Instead of a perfect solution, ask for the best solution. One such class of solutions are least squares solutions.

1.2 Least Squares Solutions

In both the zonal and modal approximations, the goal is to minimize the residual error

$$\|r\| = \|\mathbf{A}x - b\|.$$

This work explores the minimal solutions under the $2-$norm, the familiar norm of Pythagorus:

$$\left\| \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\|_2 = \sqrt{x_1^2 + x_2^2}.$$

Let's construct a sample problem with $\|\mathbf{A}x - b\| > 0$ by modifying the previous example:

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}. \tag{1.3}$$

When $b_2 \neq 0$ there is no solution with $\|\mathbf{A}x - b\| > 0$. Consider the solutions given by

$$x_* = \begin{bmatrix} x_1 \\ 0 \end{bmatrix}; \tag{1.4}$$

the error is

$$\mathbf{A}x_* - b = -\begin{bmatrix} 0 \\ b_2 \end{bmatrix} \tag{1.5}$$

which has a norm $\|\mathbf{A}x_* - b\| = |b_2|$, plotted in figure 1.1. This is the least possible error for the problem and (1.4) is the best solution. In this light, the transition from an exact solution to an inexact solution is natural.

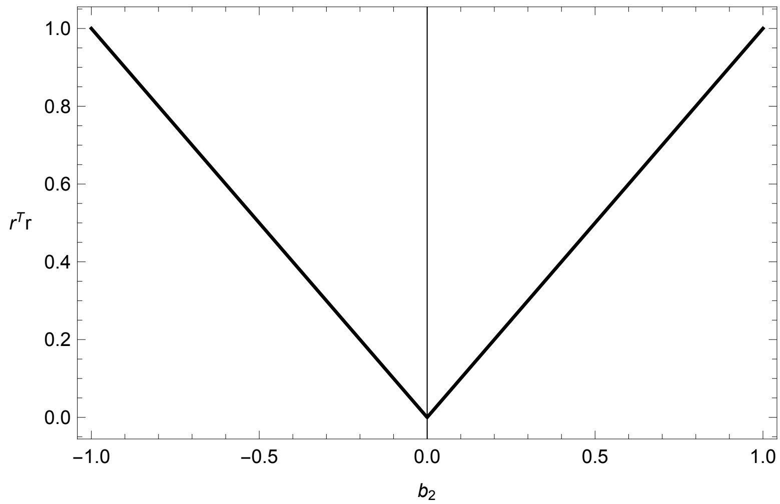


Figure 1.1. *The least squares solution* (1.5) *for* (1.3)

.

Think of the solutions to the linear system

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

as being described by the inequality

$$\|\mathbf{A}x - b\|_2 \geq 0.$$

In some cases the equality is attained.

Least squares solutions are classified by the interpretation of the output. In the first case, *zonal approximation*, the output represents data at a physical zone, a point or a region. In the second case, *modal approximation*, the output represents an amplitude, a contribution for a mode. Basic examples follow.

1.2.1 Zonal Approximation

Consider the vector field F described by the gradient of a scalar field ϕ.

$$F = \nabla\phi$$

Zonal Problem

In practice one measures the vector field and solves the inverse problem. The input and outputs are represented in 1.2. The physical field is $\phi(x)$, $0 \leq x \leq 2$, the approximation is φ_{x_k}, $k = 0, 1, 2$. For a cleaner presentation let $\varphi_{x_k} \to \varphi_k$. The first measurement x_1 represents the potential change between $\phi(0)$ and $\phi(1)$; the second measurement x_2 the change between $\phi(1)$ and $\phi(2)$.

$$\varphi_1 - \varphi_0 \approx \delta_1$$
$$\varphi_2 - \varphi_1 \approx \delta_2$$

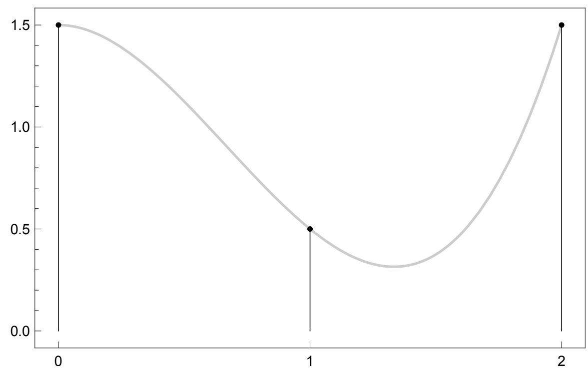


Figure 1.2. *Scalar function $\phi(x)$ (curve) and approximation φ_k (sticks).*

The system matrix

$$\mathbf{A} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \in \mathbb{R}_2^{2\times 3}.$$

There are $m = 2$ measurements, $n = 3$ measurement locations, and the matrix rank is $\rho = 2$. Because the rank is less than the number of columns, $\rho < n$, this problem is *underdetermined* .

The linear system is

$$\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \varphi_0 \\ \varphi_1 \\ \varphi_2 \end{bmatrix} = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}. \tag{1.6}$$

Zonal Solution

The solutions for the linear system in (1.6) which minimize $\|\mathbf{A}x - b\|_2$ are

$$\begin{bmatrix} \varphi_0 \\ \varphi_1 \\ \varphi_2 \end{bmatrix} = \color{blue}{\begin{bmatrix} -2 & -1 \\ 1 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}} + \gamma \color{red}{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}, \qquad \gamma \in \mathbb{C}.$$

The color blue represents range space vectors, red null space vectors. In this way, the fundamental spaces spring to life.

There is a continuum of solutions due to the fact that

$$\mathbf{A}x = \mathbf{A}\left(x + \alpha \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right).$$

An easy demonstration is to write

$$\mathbf{A}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

1.2.2 Modal Approximation

Modal Problem

In the modal approximation, the user first selects a set of basis functions to describe measurements. Popular basis functions include orthogonal polynomials, trigonometric functions, or monomials. For example, a linear regression implies a basis set of two elements: a constant function, and a linear function. This leads to the familiar equation for a straight line:

$$y(x) = a_0 + a_1 x$$

The $n = 2$ parameters represent the intercept (a_0), and the slope (a_1); each of the m measurements represents a straight line:

$$\begin{aligned} a_0 + a_1 x_1 &= y_1 \\ &\vdots \\ a_0 + a_1 x_m &= y_m. \end{aligned}$$

The goal is to simultaneously solve the set of equations.

Modal Solution

The first step is to compose the system

$$\begin{array}{cccc} \mathbf{A} & \alpha & = & y \\ \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} & \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} & = & \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}, \end{array} \tag{1.7}$$

which can be expressed using the column vectors

$$\mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}.$$

The columns of the system matrix $\mathbf{A} = \left[\begin{array}{c|c} \mathbf{1} & x \end{array} \right]$. The solution parameters can be expressed in terms of the column vectors:

$$\left[\begin{array}{c} \alpha_0 \\ \alpha_1 \end{array} \right] = \left(\left(\mathbf{1}^{\mathrm{T}}\mathbf{1}\right)\left(x^{\mathrm{T}}x\right) - \left(\mathbf{1}^{\mathrm{T}}x\right)^2 \right)^{-1} \left[\begin{array}{cc} x^{\mathrm{T}}x & -\mathbf{1}^{\mathrm{T}}x \\ -\mathbf{1}^{\mathrm{T}}x & \mathbf{1}^{\mathrm{T}}\mathbf{1} \end{array} \right] \left[\begin{array}{c} \mathbf{1}^{\mathrm{T}}y \\ x^{\mathrm{T}}y \end{array} \right].$$

1.2.3 Errors

When measurements are not exact, solutions are not exact. A great beauty of the method of least squares in that the quality of the solution can be quantified. An ability to discern answers like 3.0 ± 1.0 from 3.000 ± 0.0010 is invaluable. The machinery needed to compute uncertainties will be developed in following chapters.

1.3 Least Squares Problem

Emboldened by solutions to two basic problems, we turn attention towards formalities. Starting with a the linear system $\mathbf{A}x = b$ where the matrix $\mathbf{A} \in \mathbb{C}^{m\times n}$, the data vector $b \in \mathbb{C}^m$, the least squares solution x_{LS} is defined as the set

$$x_{LS} = \left\{ x \in \mathbb{C}^n \colon \ \lVert \mathbf{A}x - b \rVert_2^2 \ \text{ is minimized} \right\}. \tag{1.8}$$

The least squares solution may be a point or it may be a hyperplane. The general solution is a combination of a particular solution (in blue) and a homogenous solution (in red):

$$\begin{aligned} x_{LS} &= \mathbf{A}^{\dagger}b &+& \left(\mathbf{I}_n - \mathbf{A}^{\dagger}\mathbf{A}\right)y, \qquad y \in \mathbb{C}^n \\ &= x_{\dagger} &+& \qquad x_{\mathcal{N}} \end{aligned} \tag{1.9}$$

where the matrix $\mathbf{A}^{\dagger}$ is the pseudoinverse.

Chapter 2

Least Squares Solutions

Bolstered from producing concrete results, attention now turns to an examination of solution methods through the lens of the Fundamental Theorem.

2.1 Fundamental Theorem of Linear Algebra

Table 2.1. *The Fundamental Theorem of Linear Algebra for* $\boldsymbol{A} \in \mathbb{C}^{m\times n}$

domain:	$\mathbb{C}^n$	$=$	$\mathcal{R}(\mathbf{A}^*)$	$\oplus$	$\mathcal{N}(\mathbf{A})$
codomain:	$\mathbb{C}^m$	$=$	$\mathcal{R}(\mathbf{A})$	$\oplus$	$\mathcal{N}(\mathbf{A}^*)$

Table 2.2. *The Fundamental Theorem of Linear Algebra and Least Squares for* $\boldsymbol{A} \in \mathbb{C}^{m\times n}$

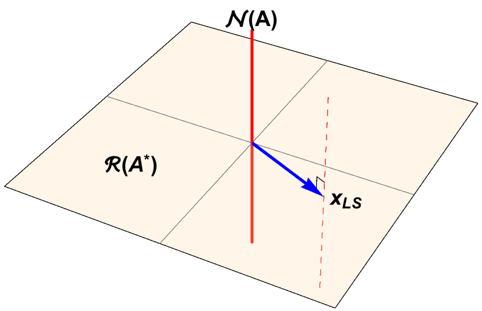
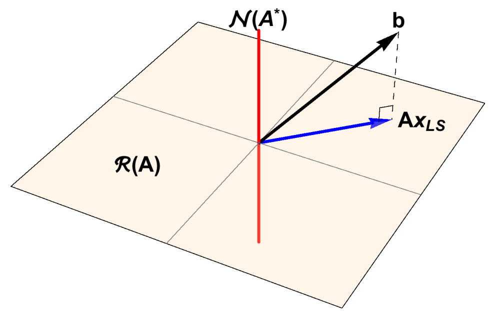
Domain	Mapping	Codomain
	$\mathbf{A}\colon \mathbb{C}^n \quad \mapsto \quad \mathbb{C}^m$ $\mathbb{C}^n \quad \leftarrow\!\shortmid \quad \mathbb{C}^m\colon \mathbf{A}^*$	
$\mathbb{C}^n = \mathcal{R}(\mathbf{A}^*) \oplus \mathcal{N}(\mathbf{A})$		$\mathbb{C}^m = \mathcal{R}(\mathbf{A}) \oplus \mathcal{N}(\mathbf{A}^*)$

Table 2.3. *Dimensions of the fundamental subspaces for* $\boldsymbol{A} \in \mathbb{C}_{\rho}^{m\times n}$.

$$
\begin{array}{rclrcl}
\dim\left(\mathcal{R}(\mathbf{A})\right) & = & \rho & \dim\left(\mathcal{N}(\mathbf{A}^*)\right) & = & m-\rho \\
\dim\left(\mathcal{R}(\mathbf{A}^*)\right) & = & \rho & \dim\left(\mathcal{N}(\mathbf{A})\right) & = & n-\rho
\end{array}
$$

2.2 Singular Value Decomposition - I

The Fundamental Theorem describes the world as a orthogonal decomposition of the domain and codomain. Why not ask for an orthonormal decomposition? This is precisely what we get from the singular value decomposition.

2.2.1 SVD Theorem

Given a matrix $\mathbf{A} \in \mathbb{C}_{\rho}^{m\times n}$, a matrix with complex entries with m rows, n columns, and matrix rank $0 < \rho \le \min(m,n)$, then there exists a decomposition of the form

$$\mathbf{A} = \mathbf{U}\,\Sigma\,\mathbf{V}^*$$

where

1. column vectors of unitary matrix $\mathbf{V} \in \mathbb{C}^{n\times n}$ represent an orthonormal span of the domain,
2. column vectors of unitary matrix $\mathbf{U} \in \mathbb{C}^{m\times m}$ represent an orthonormal span of the codomain,
3. Diagonal entries of $\Sigma \in \mathbb{R}^{m\times n}$ contain the singular values; the ordered, nonzero eigenvalues of the product matrix $\mathbf{A}^*\mathbf{A}$.

   In block form

$$\mathbf{A} = \mathbf{U}\,\Sigma\,\mathbf{V}^* = \left[\begin{array}{c|c} \mathbf{U}_{\mathcal{R}} & \mathbf{U}_{\mathcal{N}} \end{array}\right] \left[\begin{array}{c|c} \mathbf{S} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \left[\begin{array}{c} \mathbf{V}_{\mathcal{R}}^* \\ \hline \mathbf{V}_{\mathcal{N}}^* \end{array}\right]$$

Column vectors span the subspaces:

$$
\begin{aligned}
\mathbf{V} &= \left[\begin{array}{ccc|ccc} v_1 & \dots & v_\rho & v_{\rho+1} & \dots & v_n \end{array}\right], \\
\mathbf{U} &= \left[\begin{array}{ccc|ccc} u_1 & \dots & u_\rho & u_{\rho+1} & \dots & u_m \end{array}\right].
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{U} &\in \mathbb{C}^{m\times m}, \\
\mathbf{V} &\in \mathbb{C}^{n\times n}, \\
\Sigma &\in \mathbb{R}^{m\times n}.
\end{aligned}
$$

Table 2.4. *Orthonormal spans for the invariant subspaces.*

		domain			codomain
$\mathcal{R}(\mathbf{A}^*)$	$=$	$\text{span}\{v_1,\dots,v_\rho\}$	$\mathcal{R}(\mathbf{A})$	$=$	$\text{span}\{u_1,\dots,u_\rho\}$
$\mathcal{N}(\mathbf{A})$	$=$	$\text{span}\{v_{\rho+1},\dots,v_n\}$	$\mathcal{N}(\mathbf{A}^*)$	$=$	$\text{span}\{u_{\rho+1},\dots,u_m\}$

$$\begin{aligned} u_j \cdot u_k &= \delta_{jk}, \\ v_j \cdot v_k &= \delta_{jk}. \end{aligned}$$

Decomposition for (1.6):

$$\begin{array}{ccccc} \mathbf{A} & = & \mathbf{U} & \Sigma & \mathbf{V}^* \\ \left[\begin{array}{rrr} -1 & 1 & 0 \\ 0 & -1 & 1 \end{array}\right] & = & \frac{1}{\sqrt{2}}\left[\begin{array}{rr} -1 & 1 \\ 1 & 1 \end{array}\right] & \left[\begin{array}{cc|c} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{array}\right] & \left[\begin{array}{ccc} \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{array}\right] \end{array}$$

$$\mathbf{S} = \left[\begin{array}{cc} \sqrt{3} & 0 \\ 0 & 1 \end{array}\right]$$

2.2.2 SVD and Least Squares

A direct implication of the singular value decomposition is the homogeneous solution.

Unitary transformation

The definition of the least squares problem in (1.8) shows that the target of minimization is the quantity

$$r^{\mathrm{T}} r = r^2 = \|\mathbf{A}\,x - b\|_2^2 .$$

One minimization strategy invokes a unitary transformation to create a simpler problem:

$$\|\mathbf{A}\,x - b\|_2^2 = \|\mathbf{U}^*\left(\mathbf{A}\,x - b\right)\|_2^2 . \tag{2.1}$$

This remarkable insight opens a door to solution. Rearranging the singular value decomposition

$$\mathbf{U}^*\mathbf{A} = \Sigma\,\mathbf{V}^*,$$

and using the block form in (2.2.1) leads to

$$
\begin{aligned}
\lVert \mathbf{A}x - b \rVert_2^2 = \lVert \Sigma\, \mathbf{V}^* x - \mathbf{U}^* b \rVert_2^2 &= \left\lVert \left[\begin{array}{c|c} \mathbf{S} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \left[\begin{array}{c} \mathbf{V}_{\mathcal{R}}^* \\ \hline \mathbf{V}_{\mathcal{N}}^* \end{array}\right] x - \left[\begin{array}{c} \mathbf{U}_{\mathcal{R}}^* \\ \hline \mathbf{U}_{\mathcal{N}}^* \end{array}\right] b \right\rVert_2^2 \\
&= \left\lVert \left[\begin{array}{c} \mathbf{S}\mathbf{V}_{\mathcal{R}}^* x \\ \hline \mathbf{0} \end{array}\right] - \left[\begin{array}{c} \mathbf{U}_{\mathcal{R}}^* b \\ \hline \mathbf{U}_{\mathcal{N}}^* b \end{array}\right] \right\rVert_2^2 .
\end{aligned}
$$

The range space components are now untangled from the null space components.

Pseudoinverse solution

Using the Pythagorean theorem to isolate the range and null space components of the total error for the least squares problem

$$
\lVert \mathbf{A}x - b \rVert_2^2 = \underbrace{\lVert \mathbf{S}\mathbf{V}_{\mathcal{R}}^* x - \mathbf{U}_{\mathcal{R}}^* b \rVert_2^2}_{\substack{x \text{ dependence} \\ \text{under control}}} + \underbrace{\lVert \mathbf{U}_{\mathcal{N}}^* b \rVert_2^2}_{\substack{\text{residual} \\ \text{no control}}}
$$

There are now two terms; the first depends upon the solution vector x, the second does not. We only have control over the first term. To minimize the total error we must drive the first term to zero. Then the total error will be given by the residual error term. The error term that is controlled by the solution vector x is this

$$
\mathbf{S}\mathbf{V}_{\mathcal{R}}^* x - \mathbf{U}_{\mathcal{R}}^* b. \tag{2.2}
$$

Choosing the vector x which forces this term to zero leads to the SVD solution for the least squares problem:

$$
x_{\dagger} = \mathbf{V}_{\mathcal{R}} \mathbf{S}^{-1} \mathbf{U}_{\mathcal{R}}^* b.
$$

This is also the pseudoinverse solution

$$
x_{\dagger} = \mathbf{A}^{\dagger} b
$$

where the (thin) pseudoinverse is

$$
\mathbf{A}^{\dagger} = \mathbf{V}_{\mathcal{R}} \mathbf{S}^{-1} \mathbf{U}_{\mathcal{R}}^* .
$$

The error that can be controlled is forced to 0; but this leaves an error which cannot be removed, a residual error defined as

$$
r^2 = \lVert \mathbf{U}_{\mathcal{N}}^* b \rVert_2^2 .
$$

The usually silent null space term can be heard as it pronounces the value of the total error.

To recap, the singular value decomposition leads immediately to the pseudoinverse solution and residual error.

In retrospect

Decompose the data vector in range and null space components:

$$b = b_{\mathcal{R}} + b_{\mathcal{N}}$$

$$\|\mathbf{A}\,x - b\|_2^2 = \|\mathbf{A}\,x - b_{\mathcal{R}} - b_{\mathcal{N}}\|_2^2 = \|b_{\mathcal{N}}\|_2^2$$

Because the vector $b_{\mathcal{R}} \in \mathcal{R}(\mathbf{A})$, there exists a vector x such that $\mathbf{A}\,x = b_{\mathcal{R}}$. Again, the error that cannot be removed is the residual error

$$\|b_{\mathcal{N}}\|_2^2$$

What we shown is that the vector x which minimizes the least squares error in (**??**) is exactly the same vector given by the SVD solution in equation (2.2.2). Using a unitary transform we were able to convert the general least squares problem into a form amenable to solution using the singular value decomposition.

For the overdetermined case as we have here the usually silent null space term can be heard as it pronounces the value of the total error

$$r^2 = \|\mathbf{U}_{\mathcal{N}}^* b\|_2^2 = (\mathbf{U}_{\mathcal{N}}^* b)^* \, (\mathbf{U}_{\mathcal{N}}^* b) = b^* \, (\mathbf{U}_{\mathcal{N}} \mathbf{U}_{\mathcal{N}}^*) \, b \tag{2.3}$$

2.3 Singular Value Decomposition - II

2.3.1 Σ Gymnastics

Success in manipulating the singular value decomposition includes success in manipulating the Σ matrices. Think of the Σ matrix as a sabot matrix for the matrix of singular values $\mathbf{S}$. The breakdown of block diagrams: the matrix Σ $(m \times n)$ and the matrix Σ^{T} $(n \times m)$ have different shapes, but equivalent block diagrams.

$$\begin{aligned}
\Sigma &= \left[\begin{array}{cc} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array} \right], \\
\Sigma^{\mathrm{T}} &= \left[\begin{array}{cc} \mathbf{S}^{\mathrm{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array} \right] = \left[\begin{array}{cc} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array} \right], \\
\Sigma^{\dagger} &= \left[\begin{array}{cc} \mathbf{S}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array} \right].
\end{aligned}$$

$$\Sigma = \left[\begin{array}{c|c} \mathbf{S}_{\rho\times\rho} & \mathbf{0}_{\rho\times n-\rho} \\ \hline \mathbf{0}_{m-\rho\times\rho} & \mathbf{0}_{m-\rho\times n-\rho} \end{array} \right]$$

Notice the interchange of the indices m and n in the transpose:

$$\Sigma^{\mathrm{T}} = \left[\begin{array}{c|c} \mathbf{S}_{\rho\times\rho} & \mathbf{0}_{\rho\times m-\rho} \\ \hline \mathbf{0}_{n-\rho\times\rho} & \mathbf{0}_{n-\rho\times m-\rho} \end{array} \right].$$

The pseudoinverse matrix has the same dimension as the parent matrix:

$$\Sigma = \left[\begin{array}{c|c} \mathbf{S}_{\rho\times\rho} & \mathbf{0}_{\rho\times n-\rho} \\ \hline \mathbf{0}_{m-\rho\times\rho} & \mathbf{0}_{m-\rho\times n-\rho} \end{array}\right]$$

$$\begin{aligned} \Sigma\,\Sigma^{\dagger} &= \mathbb{I}_{m,\rho}, \\ \Sigma^{\dagger}\Sigma &= \mathbb{I}_{n,\rho}. \end{aligned}$$

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \Sigma^{\dagger} = \begin{bmatrix} 1/\sigma_1 & 0 & 0 & 0 \\ 0 & 1/\sigma_2 & 0 & 0 \end{bmatrix}, \quad \Sigma^{\mathrm{T}} = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \end{bmatrix}$$

$$\begin{aligned} \Sigma\Sigma^{\dagger} = \mathbb{I}_{4,2} &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \\ \Sigma^{\dagger}\Sigma = \mathbb{I}_{2,2} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}_2. \end{aligned}$$

$$\begin{aligned} \Sigma\Sigma^{\mathrm{T}} &= \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{S}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \\ \Sigma^{\mathrm{T}}\Sigma &= \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} = \mathbf{S}^2. \end{aligned}$$

$$\begin{aligned} \Sigma\Sigma^{\dagger} &= \mathbb{I}_{2,2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}_2, \\ \Sigma\Sigma^{\mathrm{T}} &= \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} = \mathbf{S}^2, \end{aligned}$$

$$
\begin{aligned}
\Sigma^{\dagger}\Sigma &= \mathbb{I}_{4,2} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \\
\Sigma^{\mathrm{T}}\Sigma &= \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{S}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},
\end{aligned}
$$

2.3.2 Fundamental Projectors

Given a matrix $\mathbf{A} \in \mathbb{C}_{\rho}^{m\times n}$, a matrix with complex entries with m rows, n columns, and matrix rank $0 < \rho \leq \min(m,n)$, then there exists a

Table 2.5. *Fundamental Projectors using the pseudoinverse.*

	Range space			Null space		
domain	$\mathbf{P}_{\mathcal{R}(\mathbf{A}^*)}$	$=$	$\mathbf{A}^{\dagger}\mathbf{A}$	$\mathbf{P}_{\mathcal{N}(\mathbf{A})}$	$=$	$\mathbf{I}_n - \mathbf{A}^{\dagger}\mathbf{A}$
codomain	$\mathbf{P}_{\mathcal{R}(\mathbf{A})}$	$=$	$\mathbf{A}\mathbf{A}^{\dagger}$	$\mathbf{P}_{\mathcal{N}(\mathbf{A}^*)}$	$=$	$\mathbf{I}_m - \mathbf{A}\mathbf{A}^{\dagger}$

Table 2.6. *Fundamental Projectors using domain matrices.*

	Range space			Null space		
domain	$\mathbf{P}_{\mathcal{R}(\mathbf{A}^*)}$	$=$	$\mathbf{V}_{\mathcal{R}}\mathbb{I}_{n,\rho}\mathbf{V}_{\mathcal{R}}^*$	$\mathbf{P}_{\mathcal{N}(\mathbf{A})}$	$=$	$\mathbf{I}_n - \mathbf{V}_{\mathcal{R}}\mathbb{I}_{n,\rho}\mathbf{V}_{\mathcal{R}}^*$
codomain	$\mathbf{P}_{\mathcal{R}(\mathbf{A})}$	$=$	$\mathbf{U}_{\mathcal{R}}\mathbb{I}_{m,\rho}\mathbf{U}_{\mathcal{R}}^*$	$\mathbf{P}_{\mathcal{N}(\mathbf{A}^*)}$	$=$	$\mathbf{I}_m - \mathbf{U}_{\mathcal{R}}\mathbb{I}_{m,\rho}\mathbf{U}_{\mathcal{R}}^*$

2.4 Least Squares Solution - Again

Let's revisit the canonical linear system in (1.1) the general solution in (1.9):

$$
\begin{aligned}
x_{LS} &= \mathbf{A}^{\dagger}b + \left(\mathbf{I}_n - \mathbf{A}^{\dagger}\mathbf{A}\right) y \\
&= \mathbf{A}^{\dagger}b + \mathbf{P}_{\mathcal{R}(\mathbf{A}^*)}y
\end{aligned}
$$

where the arbitrary vector $y \in \mathbb{C}^n$.

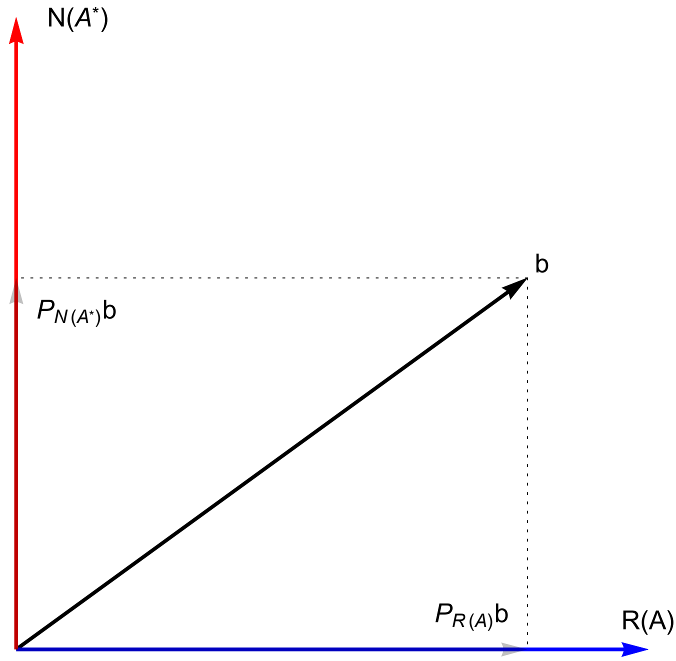


Figure 2.1. *Projections of the data vector.*

The projector onto the range space $\mathcal{R}(\mathbf{A}^*)$

$$\mathbf{A}^\dagger \mathbf{A} = \mathbf{V}\Sigma^\dagger \Sigma \mathbf{V}^* = \mathbf{V}_\mathcal{R} \mathbb{I}_{\rho,m} \mathbf{V}_\mathcal{R}^*$$

Part II

Archetypes

Chapter 3

Modal Example

3.1 Modal Approximation

The following example represents a problem in linear regression. A sequence of m data points (x_k, T_k), $k = 1\colon m$ is are recorded. The goal is to find the best approximation to a straight line. The *trial function* is

$$y(x) = a_0 + a_1 x.$$

The residual errors are the difference between the measurements and predictions:

$$\text{residual error}_k = \text{measurement}_k - \text{prediction}_k.$$

More formally the residual error is

$$r_k = T_k - y(x_k), \quad k = 1\colon m.$$

From this springs the *merit function*, the target of minimization,

$$\begin{aligned} M(a) &= \sum_{k=1}^{m} r_k^2 \\ &= \sum_{k=1}^{m} \left(\text{measurement}_k - \text{prediction}_k\right)^2 \\ &= \sum_{k=1}^{m} \left(T_k - y\left(x_k\right)\right)^2 \\ &= \sum_{k=1}^{m} \left(T_k - a_0 - a_1 x_k\right)^2 \end{aligned} \tag{3.1}$$

The least squares solution a_{LS} is formally defined as

$$a_{LS} = \left\{ a \in \mathbb{C}^2 \colon \; \|y(x_k) - a_0 - a_1 x_k\|_2^2 \; \text{ is minimized} \right\}.$$

The solution satisfies

$$\nabla M(a)|_{a_{LS}} = 0. \tag{3.2}$$

3.2 Bevington Example

To provide a common reference, see the example in Bevington [2, ch 6]. The data is summarized below in table 3.2. The problem involves temperature measurements T_k made at position x_k on a bar in contact with two heat baths (Dirichlet boundary conditions). A conceptualization is shown in figure 3.1. Arrowheads on the bottom show the nine locations where the temperature is measured.

In the ideal linear case, the temperature at the endpoints matches the temperature of the baths, $T(x = 0 \text{ cm}) = 0°\text{C}$ and $T(x = 10 \text{ cm}) = 100°\text{C}$ which describes a line with intercept $a_0 = 0°\text{C}$ and slope $a_1 = 10°\text{C/cm}$. Such an expectation is a crude quality measure, a "sniff test".

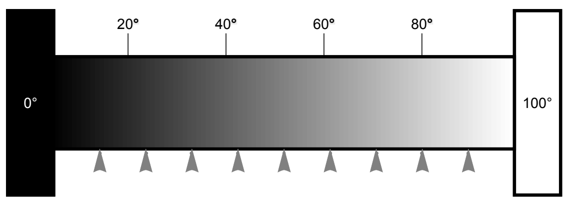


Figure 3.1. *Measuring the temperature of a bar held between two constant-temperature heat baths.*

3.2.1 Problem Statement

Muddled conceptions are wellsprings for muddled execution. Success in dealing with complicated problems in least squares flows from being able to see the problem cleanly; a good practice is to start with a table specifying the problem of interest, such as table 3.1.

The first entry is the *trial function* which defines the functional form to be applied to the data. As the name indicates, this function is an initial guess. Whether or not Nature has chosen this model remains to be seen.

The *merit function* is created by inserting the trial function into (3.1). This function is the target of minimization and can provide a crude check on the solution. Given a candidate solution a_*, compute the value of $M(a_*)$. The least squares solution has the property that $M(a_*)$ has minimum value in the neighborhood of a_*. If the solution is perturbed, one must have $M(a_*) < M(a_* + \delta a)$. When you are developing and refining least squares algorithms, you may see that the computed solution a_* changes. For overdetermined problems, the solution is unique and comparing the values of the merit function helps discriminate solutions. In a later section figure **??** will demonstrate this behavior.

The *measurements* define the quantities to be measured. It seems an obvious step, but more complex models may have ambiguities start here.

Results, or fit parameters, define the quantities to be computed using the least squares algorithm. Together with the trial function, and the measurements, we now have a clear idea of what will be measured and what will be computed.

The *residual error* specifies the difference between measurement and prediction at each point. A simple matter, apparent in the merit function, it is helpful to write it out, particularly for those who may be using the results and not intimate with the derivation.

The *system matrix* describes the measurement apparatus and contains the dependent variables. In this example we have $m = 9$ rows (measurements), $n = 2$ columns (fit parameters) and a matrix rank $\rho = 2$ (full column rank and overdetermined).

The *linear system* shows the application of the trial function to every measurement. It's a good idea to keep this image in mind.

The *ideal solution* is an infrequent visitor which helps provide a rough measure of quality. Caution is required, though. The ideal solution typically represents a concatenation of miracles which Nature may avoid. In this example, the ideal solution assumes magic barriers which absorb no heat, a bar of exact length, thermometers with exact measurements, heat baths at exact temperatures, no interaction with the local environment, etc. The hope is that systematic effects will be negligible and random effects will have 0 mean.

Table 3.1. *Problem statement for linear regression.*

trial function	$T(x) = a_0 + a_1 x$	
residual error	$r_k = T_k - a_0 - a_1 x$	°C
merit function	$M(a) = \sum_{k=1}^{m} \left(T_k - a_0 - a_1 x\right)^2$	
measurements	$x_k,\ k = 1\colon m$	position, cm
	$T_k,\ k = 1\colon m$	temperature, °C
results	$a_0 \pm \epsilon_0$	intercept, °C
	$a_1 \pm \epsilon_1$	slope, °C / cm
system matrix	$\mathbf{A} \in \mathbb{R}_2^{9\times 2}$	
linear system	$\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} T_1 \\ \vdots \\ T_m \end{bmatrix}$	
ideal solution	$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 10 \end{bmatrix}$	

The next phase is to gather and record the data as shown in table 3.2. Discussion of significant digits in the input data is deferred.

Table 3.2. *Raw data and results.*

	Input		Output	
k	$x_k(cm)$	$T_k(°C)$	$T(x_k)\,(°C)$	$r_k(°C)$
1	1	15.6	14.2222	–1.37778
2	2	17.5	23.6306	6.13056
3	3	36.6	33.0389	–3.56111
4	4	43.8	42.4472	–1.35278
5	5	58.2	51.8556	–6.34444
6	6	61.6	61.2639	–0.336111
7	7	64.2	70.6722	6.47222
8	8	70.4	80.0806	9.68056
9	9	98.8	89.4889	–9.31111

3.2.2 Normal Equations via Calculus

In §6.4, Bevington solves the problem by applying calculus to the final form in (3.1), effectively solving (3.2). Introducing the notation

$$\partial_j M(a_0, a_1) = \frac{\partial M(a_0, a_1)}{\partial a_j}$$

the simultaneous equations to solve are

$$\begin{aligned} -2\sum_{k=1}^{m}\left(T_k - a_0 - a_1 x_k\right) \quad &= 0, \\ -2\sum_{k=1}^{m}\left(T_k - a_0 - a_1 x_k\right) x_k &= 0. \end{aligned}$$

Distributing the summation operators creates a more revealing form

$$\begin{aligned} \sum_{k=1}^{m} T_k \quad &= \quad a_0 \sum 1 \quad + \quad a_1 \sum x_k, \\ \sum_{k=1}^{m} T_k x_k \quad &= \quad a_0 \sum x_k \quad + \quad a_1 \sum x_k^2, \end{aligned}$$

where summation from 1 to m is implied. (Therefore $\sum 1 = m$.) The minimization criteria are now recast as the linear system

$$\begin{bmatrix} \sum 1 & \sum x_k \\ \sum x_k & \sum x_k^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum T_k \\ \sum T_k x_k \end{bmatrix}. \tag{3.3}$$

The solution can be written immediately. Defining the determinant

$$\Delta = m \sum x_k^2 - \left(\sum x_k \right)^2,$$

the matrix inverse is

$$\begin{bmatrix} m & \sum x_k \\ \sum x_k & \sum x_k^2 \end{bmatrix}^{-1} = \Delta^{-1} \begin{bmatrix} \sum x_k^2 & -\sum x_k \\ -\sum x_k & m \end{bmatrix}. \tag{3.4}$$

The solution to equation (3.3) is the matrix product

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \Delta^{-1} \begin{bmatrix} \sum x_k^2 & -\sum x_k \\ -\sum x_k & \sum 1 \end{bmatrix} \begin{bmatrix} \sum T_k \\ \sum T_k x_k \end{bmatrix}$$

Compare the final results to Bevington's equations 6–17:

$$\begin{aligned} a_0 &= \Delta^{-1} \left(\sum x_k^2 \sum T_k - \sum x_k \sum T_k x_k \right), \\ a_1 &= \Delta^{-1} \left(m \sum T_k x_k - \sum x_k \sum T_k \right). \end{aligned}$$

Bevington's §6–5 is a succinct explanation of error propagation. In short, measurements are inexact, therefore results will be inexact. The beauty of the method of least squares is that the error in the solution parameters can be expressed in terms of the error in the data. Measurements without uncertainties are incomplete measurements.

The computation chain begins with an estimate of the parent standard deviation which is based upon the total error:

$$s^2 \approx \frac{r^* r}{m - n}.$$

Error contributions for individual parameters are harvested from the diagonal elements of the matrix inverse $(\mathbf{A}^* \mathbf{A})^{-1}$ in (3.4):

$$\begin{aligned} \epsilon_0^2 &= \frac{r^{\mathrm{T}} r}{\Delta (m - n)} \sum x_k^2 \\ \epsilon_1^2 &= \frac{r^{\mathrm{T}} r}{\Delta (m - n)} \sum 1 \end{aligned}$$

3.3 Numerical Results

Results are stated in two forms. The first is an integer form free of numerical errors inherent in binary representations with finite length. This liberates one from trying to determine if errors are in the algorithm or in machine arithmetic. To aid debugging, intermediate results are also provided.

The second form represents the answer which would be provided to a customer: the fit parameters and associated errors quoted with the proper amount of significant digits.

3.3.1 Exact Form

Exact results for the fit parameters are error follow. The product matrix in (3.3) is

$$\mathbf{A}^*\mathbf{A} = \begin{bmatrix} 9 & 45 \\ 45 & 285 \end{bmatrix},$$

with determinant

$$\Delta = \det\left(\mathbf{A}^*\mathbf{A}\right) = 540.$$

The inverse of this matrix, (3.4), is

$$\left(\mathbf{A}^*\mathbf{A}\right)^{-1} = \Delta^{-1} \begin{bmatrix} 285 & -45 \\ -45 & 9 \end{bmatrix}.$$

The solution vector, (3.2.2), is

$$a = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \frac{1}{360} \begin{bmatrix} 1733 \\ 3387 \end{bmatrix}.$$

The residual error vector is $r = \mathbf{A}^*\mathbf{A}a - \mathbf{A}^*T$

$$r = \frac{1}{360} \begin{bmatrix} -496 \\ 2207 \\ -1282 \\ -487 \\ -2284 \\ -121 \\ 2330 \\ 3485 \\ -3352 \end{bmatrix},$$

making the total error

$$r^*r = \frac{1\,139\,969}{3600}.$$

The uncertainties are then

$$\epsilon = \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \end{bmatrix} = \left(360\sqrt{35}\right)^{-1} \begin{bmatrix} 108\,297\,055 \\ 3\,419\,907 \end{bmatrix}.$$

3.3.2 Computed Form

The results in the previous section are a debugging tool. This section deals with formats appropriate for formal reporting. One way to quote numbers with uncertainties is using the $\pm$ (plus – minus) notation:

$$\begin{array}{lclclll} a_0 & = & 4.8 & \pm & 4.9 & \text{intercept} & {}^\circ\text{C}, \\ a_1 & = & 9.41 & \pm & 0.87 & \text{slope} & {}^\circ\text{C / cm.} \end{array}$$

An alternative presentation uses parentheses:

$$\begin{array}{rcll} a_0 & = & 4.8\,(4.9) & \text{intercept} \quad {}^\circ\text{C}, \\ a_1 & = & 9.41\,(0.87) & \text{slope} \quad {}^\circ\text{C / cm}. \end{array}$$

The total error is $r^* r \approx 317$.

The uncertainty determines the number of significant digits. Common practice quotes the first two digits in the uncertainty; the location of these two digits determines the number of digits in the solution. The double precision computations are

$$\begin{array}{ll} \epsilon_1 = 0.8683016476563611 & \text{rounded to } 0.87, \\ a_1 = 9.408333333333333 & \text{rounded to } 9.41. \end{array}$$

At this point the model can be explored and evaluated. If the model is not acceptable, another trial function can be posed. Otherwise, the trial function becomes the solution function and is stated with error:

$$\begin{aligned} T(x) &= a_0 + a_1 x, \\ \epsilon_T^2(x) &= \epsilon_0^2 + x^2 \epsilon_1^2 + a_1^2 \epsilon_x^2, \end{aligned}$$

which allows for interpolation and extrapolation. What happens when the solution is extrapolated to the heat baths? The expected answers are 0°C at 0 cm, and 100°C at 10 cm:

$$\begin{aligned} T\,(0) &= (4.8 \pm 4.9)^\circ\,\text{C}, \\ T\,(10) &= (99. \pm 10.)^\circ\,\text{C}. \end{aligned}$$

One final thought. The method of least squares minimizes the sums of the squares of the residual errors. And in linear regression, the sum of these residuals must be 0. That is,

$$\sum_{k=1}^{m} r_k = 0.$$

This can be an quick method for evaluating solutions and data sets. Given the data and the solution parameters a a quick Python or Mathematica script can compute and sum the residuals. If a data point is omitted, the sum will not be 0. If the parameters are misquoted, the sum will not be 0. If a data point is corrupted, the sum will not be 0. Or if the solutions are for another data set, the sum will not be 0. The 0 test is simple and powerful.

3.4 Visualization

3.4.1 Seeing the Solution

Numbers tell a story and plots bring a story to life. Besides a powerful and immediate impact, the plots are often a defense against a wide host of problems.

The first plot is the solution curve plotted against the measurements as shown in figure 3.4.2. The residual errors are shown as red segments which are the actual components of the residual error vector in $\mathbb{C}^9$.

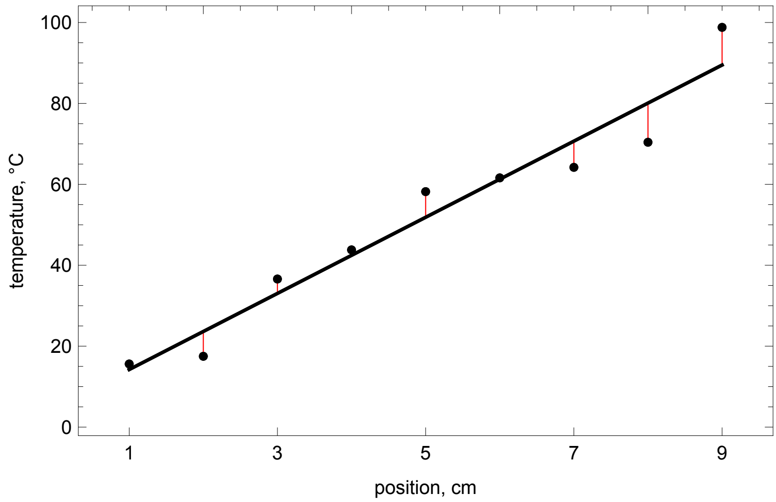


Figure 3.2. *Solution plotted against data with residual errors shown in red.*

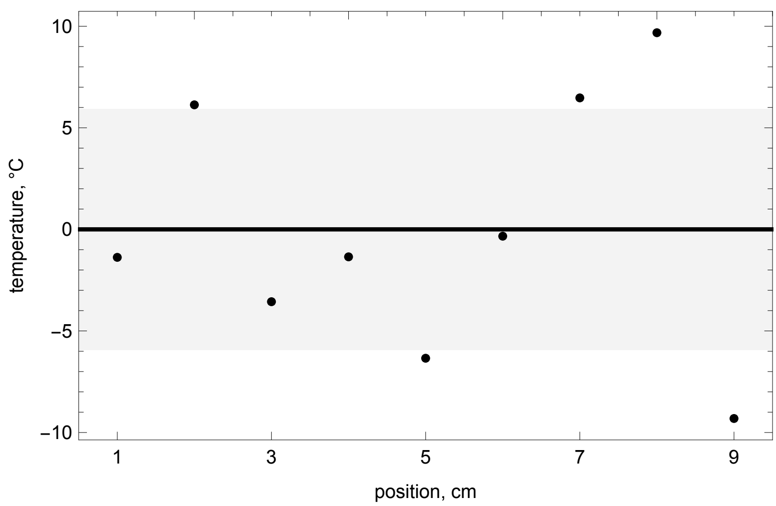


Figure 3.3. *Scatter plot of residual errors.*

3.4.2 Digger Deeper

3.4.3 Seeing the Uncertainty

How should one interpret the uncertainties in slope and intercept? First understand the concept of distribution of errors.

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

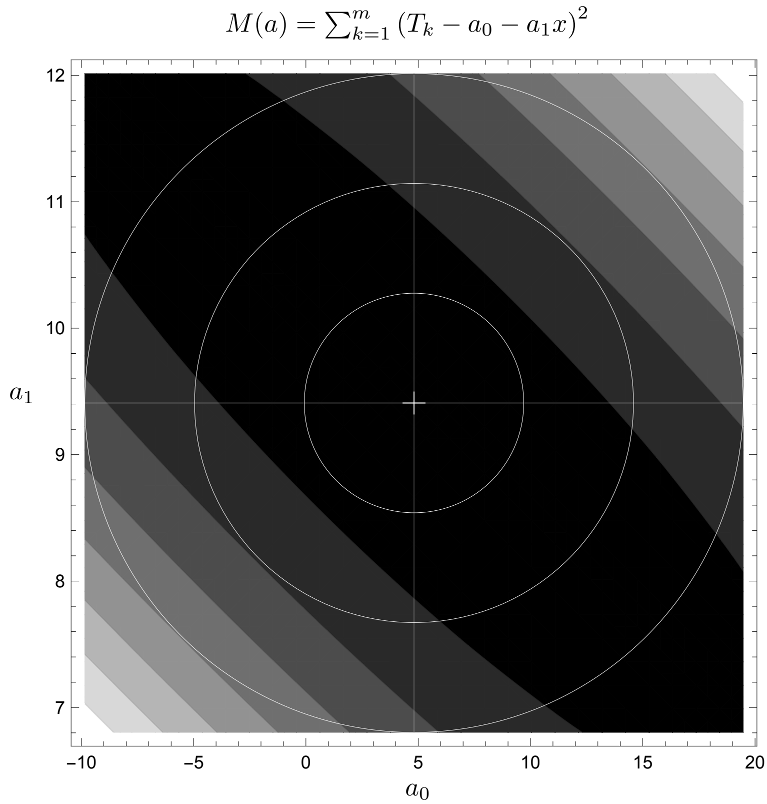


Figure 3.4. *Contour plot of merit function showing the solution (white cross) and three concentric error bands.*

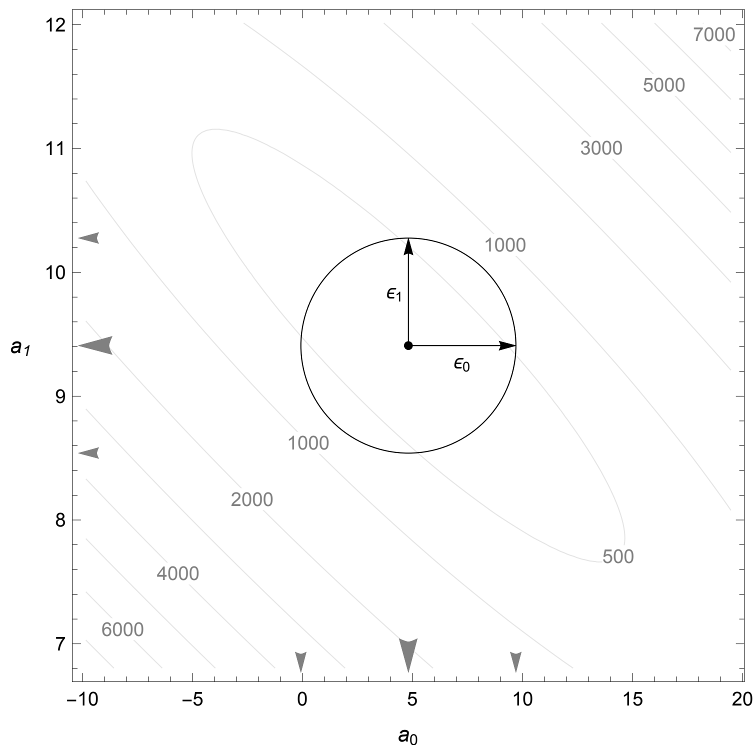


Figure 3.5. *Another look at the merit function in* (3.1) *showing the uncertainty parameters as elliptic radii.*

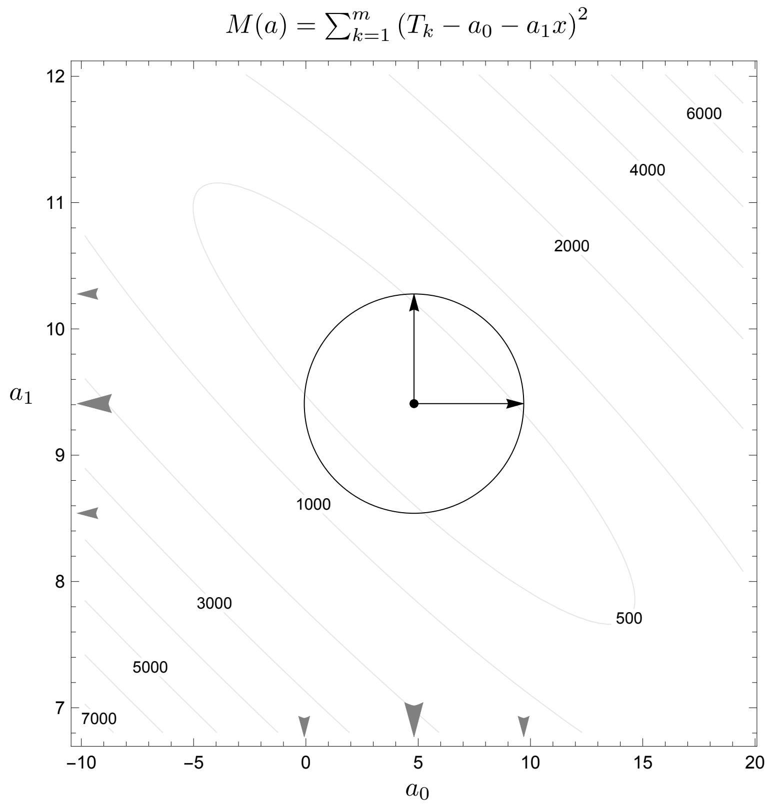


Figure 3.6. *Contour plot of merit function showing the solution (white cross) and three concentric error bands.*

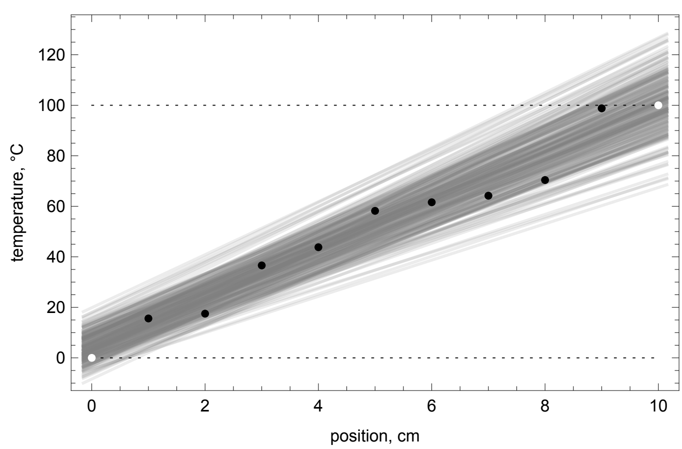


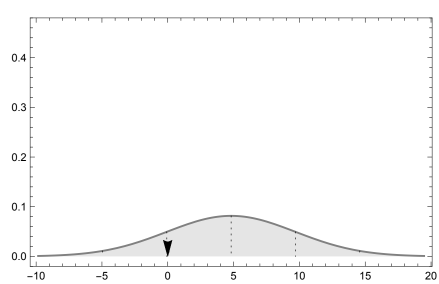
Figure 3.7. *Whisker plot showing 250 randomly sampled solutions.*

Table 3.3. *Results for linear regression.*

fit parameters	a_0	intercept, °C
	a_1	slope, °C / cm
solution function	$T(x) = a_0 + a_1 x$	°C
solution error	$\epsilon_T^2(x) = \epsilon_0^2 + x^2\epsilon_1^2 + a_1^2\epsilon_x^2$	°C
computed solution	$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 4.8 \\ 9.41 \end{bmatrix} \pm \begin{bmatrix} 4.9 \\ 0.87 \end{bmatrix}$	
ideal solution	$\begin{bmatrix} \tilde{a}_0 \\ \tilde{a}_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 10 \end{bmatrix}$	
r^*r	316.6	
curvature matrix $(\mathbf{A}^*\mathbf{A})^{-1}$	$\frac{1}{180}\begin{bmatrix} 95 & -15 \\ -15 & 3 \end{bmatrix}$	
problem statement	table 3.1	
input data	table 3.2	
plots	figure 3.2	data and solution
	figure 3.4.2	residual errors
	figure **??**	merit function

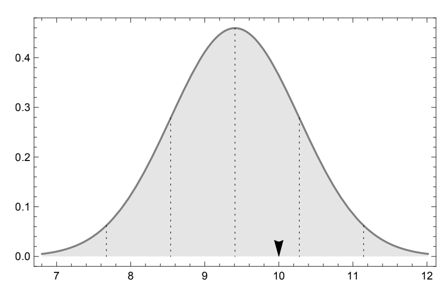
Table 3.4. *The solution parameters expressed as normal distributions.*

intercept



$$f(x; a_0, \epsilon_0) = \frac{1}{\epsilon_0\sqrt{2\pi}} e^{-\frac{(x-a_0)^2}{2\epsilon_0^2}}$$

slope



$$f(x; a_1, \epsilon_1) = \frac{1}{\epsilon_1\sqrt{2\pi}} e^{-\frac{(x-a_1)^2}{2\epsilon_1^2}}$$

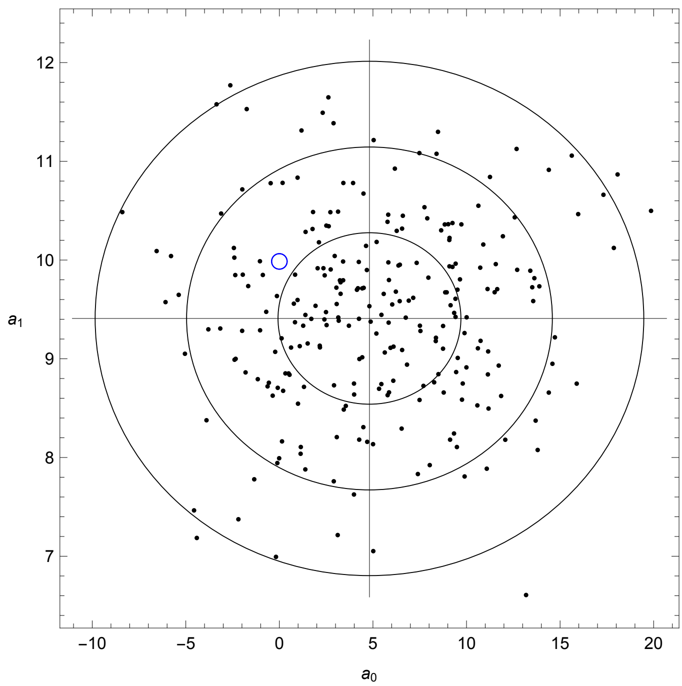


Figure 3.8. *Scatter plot showing sampling of solutions.*

Table 3.5. *Comparing samples to ideal normal distribution.*

ring	count	area	density	(actual) cumulative	(ideal) limit
1	88	1	64.83%	64.83%	68.27%
2	115	3	28.24%	93.07%	95.45%
3	41	5	6.41%	99.16%	99.73%
4	6	5	0.88%	100.0%	99.99%

Chapter 4

Modal Example Continued

Other solution methods

4.1 Normal Equations - Again

$$\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad x = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{bmatrix}, \quad T = \frac{1}{10} \begin{bmatrix} 156 \\ 175 \\ 366 \\ 438 \\ 582 \\ 616 \\ 642 \\ 704 \\ 988 \end{bmatrix}$$

$$\mathbf{A} = \left[\begin{array}{c|c} \mathbf{1} & x \end{array}\right]$$

The linear system looks like this

$$\begin{array}{cccc} \mathbf{A} & a & = & T \\ \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \\ 1 & 7 \\ 1 & 8 \\ 1 & 9 \end{bmatrix} & \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} & = & \frac{1}{10} \begin{bmatrix} 156 \\ 175 \\ 366 \\ 438 \\ 582 \\ 616 \\ 642 \\ 704 \\ 988 \end{bmatrix}. \end{array} \tag{4.1}$$

$$
\begin{aligned}
\mathbf{1}^{\mathrm{T}}\mathbf{1} &= m = 9\\
\mathbf{1}^{\mathrm{T}}x &= x^{\mathrm{T}}\mathbf{1} = 45\\
x^{\mathrm{T}}x &= 285\\
\mathbf{1}^{\mathrm{T}}T &= \frac{4667}{10}\\
x^{\mathrm{T}}T &= 2898
\end{aligned}
$$

$$
\mathbf{A}^{*}\mathbf{A} = \left[\begin{array}{cc} \mathbf{1}^{\mathrm{T}}\mathbf{1} & \mathbf{1}^{\mathrm{T}}x \\ x^{\mathrm{T}}\mathbf{1} & x^{\mathrm{T}}x \end{array}\right] = \left[\begin{array}{cc} 9 & 45 \\ 45 & 285 \end{array}\right]
$$

$$
\mathbf{A}^{*}T = \left[\begin{array}{c} \mathbf{1}^{\mathrm{T}}T \\ x^{\mathrm{T}}T \end{array}\right] = \frac{1}{10}\left[\begin{array}{c} 4667 \\ 28\,980 \end{array}\right]
$$

(4.1) becomes

$$
\left[\begin{array}{cc} \mathbf{1}^{\mathrm{T}}\mathbf{1} & \mathbf{1}^{\mathrm{T}}x \\ x^{\mathrm{T}}\mathbf{1} & x^{\mathrm{T}}x \end{array}\right]\left[\begin{array}{c} a_0 \\ a_1 \end{array}\right] = \left[\begin{array}{c} \mathbf{1}^{\mathrm{T}}T \\ x^{\mathrm{T}}T \end{array}\right] \tag{4.2}
$$

4.2 Singular Value Decomposition

4.2.1 Computing the SVD

Solution steps

1. Compute $\lambda\left(\mathbf{A}^{*}\mathbf{A}\right)$.
2. Educated guess at domain matrix $\mathbf{V}$.
3. Compute codomain matrix $\mathbf{U}$.

The least squares problem delivers a singular value decomposition (SVD) without the muss and fuss of solving an eigensystem. The SVD is given by the matrix product

$$
\mathbf{A} = \mathbf{U}\,\Sigma\,\mathbf{V}^{*}.
$$

For the full column rank problem we have we can expand in the following block decomposition

$$
\mathbf{A} = \left[\begin{array}{c|c} \mathbf{U}_{\mathcal{R}} & \mathbf{U}_{\mathcal{N}} \end{array}\right]\left[\begin{array}{c} \mathbf{S} \\ \hline \mathbf{0} \end{array}\right]\left[\begin{array}{c} \mathbf{V}_{\mathcal{R}}^{*} \end{array}\right]
$$

The $\mathbf{S}$ matrix contains the singular values

$$
\mathbf{S} = \left[\begin{array}{cc} \sigma_1 & 0 \\ 0 & \sigma_2 \end{array}\right].
$$

The column vectors of the matrix $\mathbf{V}_{\mathcal{R}}$ represent an orthonormal basis for the row space (domain). The column vectors of the matrix $\mathbf{U}_{\mathcal{R}}$ represent two of the nine vectors in an orthonormal basis for the column space (codomain).

Singular values

The singular value spectrum of the matrix $\mathbf{A}$ is the square root of the (non-zero) eigenvalues of the product matrix $\mathbf{A}^*\mathbf{A}$

$$\sigma\left(\mathbf{A}\right) = \sqrt{\lambda\left(\mathbf{A}^*\mathbf{A}\right)}.$$

The eigenvalues of the product matrix are the roots of the characteristic polynomial $p(\lambda)$ for said matrix.

$$p(\lambda) = \lambda^2 - \lambda\,\mathrm{tr}\left(\mathbf{A}^*\mathbf{A}\right) + \det\left(\mathbf{A}^*\mathbf{A}\right)$$

We are well familiar with the determinant by now; the trace is $\mathrm{tr}\left(\mathbf{A}^*\mathbf{A}\right) = \mathbf{1}^{\mathrm{T}}\mathbf{1} + x^{\mathrm{T}}x$. The singular values are then

$$\sigma = \sqrt{\frac{1}{2}\left(\mathbf{1}^{\mathrm{T}}\mathbf{1} + x^{\mathrm{T}}x \pm \sqrt{4\left(\mathbf{1}^{\mathrm{T}}x\right)^2 - \left(\mathbf{1}^{\mathrm{T}}\mathbf{1} - x^{\mathrm{T}}x\right)^2}\right)}.$$

(The astute reader will notice that the discriminant does not seem to have the familiar form of $b^2 - 4ac$. The earnest reader will discover why this is so.) The singular value spectrum for these data is

$$\sigma = \sqrt{3\left(49 \pm \sqrt{2341}\right)} \approx \left(17.0924, 1.35954\right).$$

We now have the sigma matrix and the matrix of singular values $\mathbf{S}$:

$$\Sigma = \left[\begin{array}{c}\mathbf{S}\\ \hline \mathbf{0}\end{array}\right] = \left[\begin{array}{cc}\sigma_1 & 0\\ 0 & \sigma_2\\ \hline 0 & 0\\ 0 & 0\\ 0 & 0\\ 0 & 0\\ 0 & 0\\ 0 & 0\\ 0 & 0\end{array}\right] \approx \left[\begin{array}{cc}17.0924 & 0\\ 0 & 1.35954\\ \hline 0 & 0\\ 0 & 0\\ 0 & 0\\ 0 & 0\\ 0 & 0\\ 0 & 0\\ 0 & 0\end{array}\right].$$

Domain matrix

We can skip the eigenvector problem. To find the domain matrix we exploit the singular value decomposition of the product matrix

$$\Sigma^{\mathrm{T}}\Sigma = \left[\begin{array}{cc}\mathbf{S} & \mathbf{0}\end{array}\right]\left[\begin{array}{c}\mathbf{S}\\ \mathbf{0}\end{array}\right] = \mathbf{S}^2$$

$$\mathbf{V}_{\mathcal{R}}\mathbf{S}^2\mathbf{V}_{\mathcal{R}}^* = \mathbf{A}^*\mathbf{A}. \tag{4.3}$$

By the singular value theorem the matrix is unitary and will be a rotation matrix, a reflection matrix or a convolution. We begin by trying a rotation matrix

$$\mathbf{V}_{\mathcal{R}} = \left[\begin{array}{cc} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{array} \right] \tag{4.4}$$

which is colored blue because the column vectors belong to $\mathcal{R}(\mathbf{A}^*)$. The objective is to find the angle θ. The immediate result of equations (**??**), (4.3), and (4.4) is

$$\left[\begin{array}{cc} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{array} \right] \mathbf{S}^2 \left[\begin{array}{cc} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{array} \right]^{\mathrm{T}} = \mathbf{A}^*\mathbf{A},$$

$$\left[\begin{array}{cc} \sigma_1^2\cos^2\theta + \sigma_2^2\sin^2\theta & \left(\sigma_1^2 - \sigma_2^2\right)\cos\theta\sin\theta \\ \left(\sigma_1^2 - \sigma_2^2\right)\cos\theta\sin\theta & \sigma_2^2\cos^2\theta + \sigma_1^2\sin^2\theta \end{array} \right] = \left[\begin{array}{cc} \mathbf{1}^{\mathrm{T}}\mathbf{1} & \mathbf{1}^{\mathrm{T}}x \\ x^{\mathrm{T}}\mathbf{1} & x^{\mathrm{T}}x \end{array} \right].$$

which presents multiple solution paths for the angle θ. For example

$$\cos\theta = \sqrt{\frac{\sigma_2^2 - \mathbf{1}^{\mathrm{T}}\mathbf{1}}{\sigma_2^2 - \sigma_1^2}} = \sqrt{\frac{x^{\mathrm{T}}x - \sigma_1^2}{\sigma_2^2 - \sigma_1^2}}. \tag{4.5}$$

This implies

$$\sin\theta =$$

The domain matrix is now

$$\mathbf{V} = \left[\begin{array}{c} \mathbf{V}_{\mathcal{R}} \end{array} \right] = \left(\sigma_2^2 - \sigma_1^2\right)^{-\frac{1}{2}} \left[\begin{array}{cc} \sqrt{\sigma_2^2 - \mathbf{1}^{\mathrm{T}}\mathbf{1}} & -\sqrt{\mathbf{1}^{\mathrm{T}}\mathbf{1} - \sigma_1^2} \\ \sqrt{\mathbf{1}^{\mathrm{T}}\mathbf{1} - \sigma_1^2} & \sqrt{\sigma_2^2 - \mathbf{1}^{\mathrm{T}}\mathbf{1}} \end{array} \right].$$

Using the data set at hand

$$\left[\begin{array}{c} \mathbf{V}_{\mathcal{R}} \end{array} \right] = \left[\begin{array}{cc} \sqrt{\frac{1}{2} - \frac{23}{\sqrt{2341}}} & -\sqrt{\frac{1}{2} + \frac{23}{\sqrt{2341}}} \\ \sqrt{\frac{1}{2} + \frac{23}{\sqrt{2341}}} & \sqrt{\frac{1}{2} - \frac{23}{\sqrt{2341}}} \end{array} \right] \approx \left[\begin{array}{cc} 0.156956 & -0.987606 \\ 0.987606 & 0.156956 \end{array} \right]$$

Codomain matrix

The final component is of course the codomain matrix. Knowing the decomposition for the adjoint matrix $\mathbf{A}^*$ and that the linear system is overdetermined we can write

$$\mathbf{U}_{\mathcal{R}}^* = \mathbf{S}^{-1}\mathbf{V}_{\mathcal{R}}^*\mathbf{A}^*.$$

The kth column vector of this matrix has the compact form

$$[\mathbf{U}_{\mathcal{R}}^*]_k = \left(\sigma_1^2 - \sigma_2^2\right)^{-\frac{1}{2}} \left[\begin{array}{c} \sigma_1^{-2}\left(\sqrt{\sigma_2^2 - \mathbf{1}^{\mathrm{T}}\mathbf{1}} - x\sqrt{\mathbf{1}^{\mathrm{T}}\mathbf{1} - \sigma_1^2}\right) \\ \sigma_2^{-2}\left(\sqrt{\sigma_2^2 - \mathbf{1}^{\mathrm{T}}\mathbf{1}} + x\sqrt{\mathbf{1}^{\mathrm{T}}\mathbf{1} - \sigma_1^2}\right) \end{array} \right].$$

Using the following shorthand,

$$f(x,y)=\sqrt{x+y/\sqrt{2341}},$$

the range space component of the codomain matrix can be written as

$$\mathbf{U}_{\mathcal{R}}=\left(6\sqrt{10}\right)^{-1}\left[\begin{array}{ll} f(68,-3212) & -f(68,3212)\\ f(47,-2003) & -f(47,2003)\\ f(32,-968) & -f(32,968)\\ f(23,-107) & -f(23,107)\\ f(50,580) & -f(50,-580)\\ f(23,1093) & -f(23,-1093)\\ f(32,1432) & f(32,-1432)\\ f(47,1597) & f(47,-1597)\\ f(68,1588) & f(68,-1588)\end{array}\right]$$

If one wishes to complete the codomain matrix, use the Gram-Schmidt orthonormalization process on the matrix

$$\mathbf{U}=\left[\begin{array}{ccccccccc}[\mathbf{U}_{\mathcal{R}}]_1 & [\mathbf{U}_{\mathcal{R}}]_2 & e_{1,9} & e_{2,9} & e_{3,9} & e_{4,9} & e_{5,9} & e_{6,9} & e_{7,9}\end{array}\right]$$

starting with column three.

Completing the codomain matrix with an orthonormal span of the null space is optional and can be done by feeding the range space components and a complementary set of unit vectors into a Gram-Schmidt algorithm.

Error terms

$$\epsilon_k^2=\frac{(\mathbf{A}\alpha-T)^*(\mathbf{A}\alpha-T)}{m-n}\left[(\mathbf{A}^*\mathbf{A})^{-1}\right]_{kk}$$

The error terms can be computed after this step. Given that the product matrix decomposition in (4.3) is a singular value decomposition we can trivially write the inverse matrix as

$$(\mathbf{A}^*\mathbf{A})^{-1}=\mathbf{V}_{\mathcal{R}}\mathbf{S}^{-2}\mathbf{V}_{\mathcal{R}}^*=\frac{1}{180}\left[\begin{array}{rr}95 & -15\\ -15 & 3\end{array}\right]$$

The error terms in (**??**) become

$$\begin{aligned}\epsilon &=\sqrt{\frac{r^*r}{\mathbf{1}^{\mathrm{T}}\mathbf{1}-n}}\sqrt{\frac{1}{\sigma_1\sigma_2}}\sqrt{\left[\begin{array}{c}\frac{\cos^2\theta}{\sigma_2^2}+\frac{\sin^2\theta}{\sigma_1^2}\\ \frac{\cos^2\theta}{\sigma_1^2}+\frac{\sin^2\theta}{\sigma_2^2}\end{array}\right]},\\ &=\sqrt{\frac{r^*r}{m-n}}\sqrt{\left[\begin{array}{c}\sigma_1^2-\sqrt{(\sigma_2^2-\mathbf{1}^{\mathrm{T}}\mathbf{1})(\sigma_2^2-\sigma_1^2)}\\ \sigma_2^2+\sqrt{(\sigma_2^2-\mathbf{1}^{\mathrm{T}}\mathbf{1})(\sigma_2^2-\sigma_1^2)}\end{array}\right]}.\end{aligned}$$

$$\mathbf{U} = \begin{bmatrix} 0.0670 & -0.611 & 0.656 & -0.0132 & -0.0780 & -0.137 & -0.193 & -0.240 & -0.270 \\ 0.125 & -0.496 & -0.749 & -0.0927 & -0.113 & -0.147 & -0.184 & -0.215 & -0.233 \\ 0.183 & -0.380 & 0 & 0 & 0 & 0 & 0 & 0 & 0.907 \\ 0.240 & -0.265 & 0 & 0 & 0 & 0 & 0 & 0.920 & -0.159 \\ 0.298 & -0.149 & 0 & 0 & 0 & 0 & 0.924 & -0.142 & -0.123 \\ 0.356 & -0.0337 & 0 & 0 & 0 & 0.910 & -0.150 & -0.117 & -0.0858 \\ 0.414 & 0.0817 & 0 & 0 & 0.867 & -0.198 & -0.141 & -0.0930 & -0.0490 \\ 0.471 & 0.197 & 0 & 0.755 & -0.321 & -0.209 & -0.132 & -0.0685 & -0.0123 \\ 0.529 & 0.313 & 0.0937 & -0.649 & -0.356 & -0.219 & -0.124 & -0.0441 & 0.0245 \end{bmatrix}$$

Visualization

With the singular value decomposition in hand, the domain space plots become more concrete and we do so below beginning in figure (4.2). The black vector represents the measurements

$$T = \mathbf{A}\,a - \mathbf{R}$$

$$y \in \mathbb{C}^9, \quad \mathbf{A}\,a \in \mathcal{R}(\mathbf{A}) \subseteq \mathbb{C}^2, \quad r \in \mathcal{N}(\mathbf{A}^*) \subseteq \mathbb{C}^7$$

$$T = \frac{1}{10}\begin{bmatrix} 156 \\ 175 \\ 366 \\ 438 \\ 582 \\ 616 \\ 642 \\ 704 \\ 988 \end{bmatrix}$$

The closest point to the data vector in the range $\mathcal{R}(\mathbf{A})$ is

$$\mathbf{A}\,a = \frac{1}{360}\begin{bmatrix} 5120 \\ 8507 \\ 11\,894 \\ 15\,281 \\ 18\,668 \\ 22\,055 \\ 25\,442 \\ 28\,829 \\ 32\,216 \end{bmatrix} = \alpha_1[\mathbf{U}_{\mathcal{R}}]_1 + \alpha_2[\mathbf{U}_{\mathcal{R}}]_2 \ \in \mathcal{R}(\mathbf{A})$$

where the coordinates are

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \left(30\left(\sqrt{2341}-31\right)\sqrt{4682}+58\sqrt{2341}\right)^{-1} \times \begin{bmatrix} 4\,104\,889 + 75\,341\sqrt{2341} \\ 3\sqrt{15}\left(753\,593 - 15\,933\sqrt{2341}\right) \end{bmatrix}$$
$$\approx \begin{bmatrix} 171.733 \\ -4.45594 \end{bmatrix}$$
$$T_{\mathcal{R}} \approx 171.733 u_1 - 4.45594 u_2$$

Another way to think of the geometry is to pose the merit function in terms of the

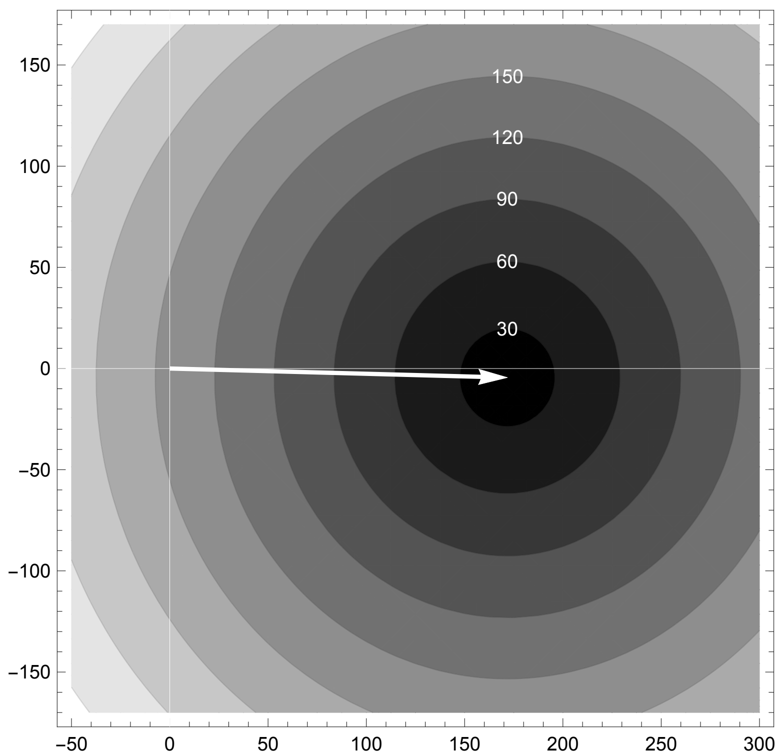


Figure 4.1. *The solution vector (white arrow) is the mixture of u_1 and u_2 which eliminates the error.*

range space vectors

$$M(\alpha) = \|T - \alpha_1 u_1 - \alpha_2 u_2\|_2^2 .$$

The minimizer is given by (4.2.1) and displayed in figure 4.1.

$$M(171.733, -4.45594) \approx 17.7949.$$

The residual error vector lies entirely in the null space $\mathcal{N}(\mathbf{A}^{*})$

$$
\begin{aligned}
r = -\frac{1}{360}&\begin{bmatrix} 496 \\ -2207 \\ 1282 \\ 487 \\ 2284 \\ 121 \\ -2330 \\ -3485 \\ 3352 \end{bmatrix} \\
&= \alpha_3[\mathbf{U}_{\mathcal{N}}]_1 + \alpha_4[\mathbf{U}_{\mathcal{N}}]_2 + \alpha_5[\mathbf{U}_{\mathcal{N}}]_3 + \alpha_6[\mathbf{U}_{\mathcal{N}}]_4 + \alpha_7[\mathbf{U}_{\mathcal{N}}]_5 + \alpha_8[\mathbf{U}_{\mathcal{N}}]_6 + \alpha_9[\mathbf{U}_{\mathcal{N}}]_7 \\
&\in \mathcal{N}(\mathbf{A}^{*})
\end{aligned}
$$

The coordinates are now

$$
\begin{bmatrix} \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7 \\ \alpha_8 \\ \alpha_9 \end{bmatrix} = \left(60\sqrt{24\,747\,709}\right)^{-1} \begin{bmatrix} 680\sqrt{7\,815\,066} \\ -1933\sqrt{3\,907\,533} \\ -3621\sqrt{186\,073} \\ 6679\sqrt{10\,434} \\ 13\,406\sqrt{29\,526} \\ 2196\sqrt{85\,386} \\ 641\sqrt{3\,344\,285} \end{bmatrix}
$$

$$
T = \mathbf{U}\alpha = \mathbf{A}\,a - r
$$

The result is 8.090170000000000000

$$
\begin{array}{ccccc}
T & = & T_{\mathcal{R}} & + & T_{\mathcal{N}} \\
\frac{1}{360}\begin{bmatrix} 5616 \\ 6300 \\ 13\,176 \\ 15\,768 \\ 20\,952 \\ 22\,176 \\ 23\,112 \\ 25\,344 \\ 35\,568 \end{bmatrix} & = & \frac{1}{360}\begin{bmatrix} 5120 \\ 8507 \\ 11\,894 \\ 15\,281 \\ 18\,668 \\ 22\,055 \\ 25\,442 \\ 28\,829 \\ 32\,216 \end{bmatrix} & + & \frac{1}{360}\begin{bmatrix} 496 \\ -2207 \\ 1282 \\ 487 \\ 2284 \\ 121 \\ -2330 \\ -3485 \\ 3352 \end{bmatrix}
\end{array}
$$

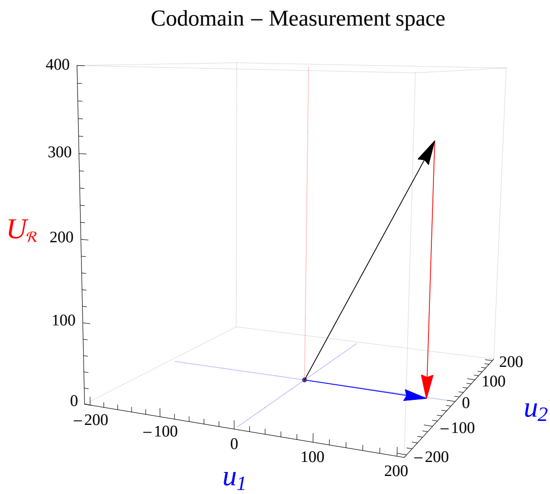


Figure 4.2. *Measurement space $\mathcal{R}(\boldsymbol{A})$ for the Bevington example.*

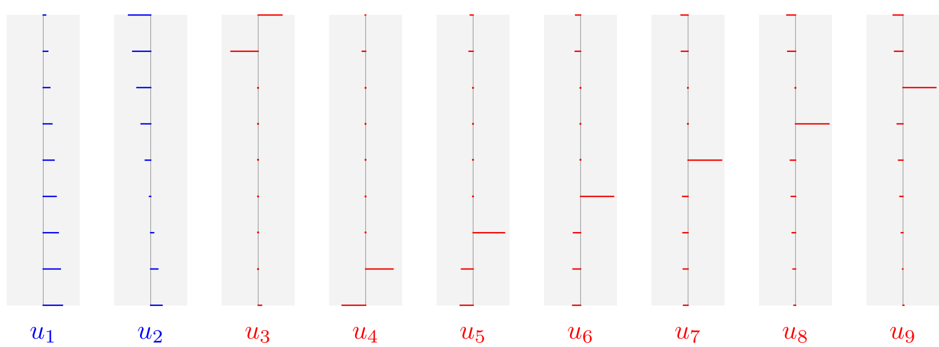


Table 4.1. *The column vectors of* $\mathbf{U}$. *The gray box represents the maximum length an element may have:* $[-1, 1]$.

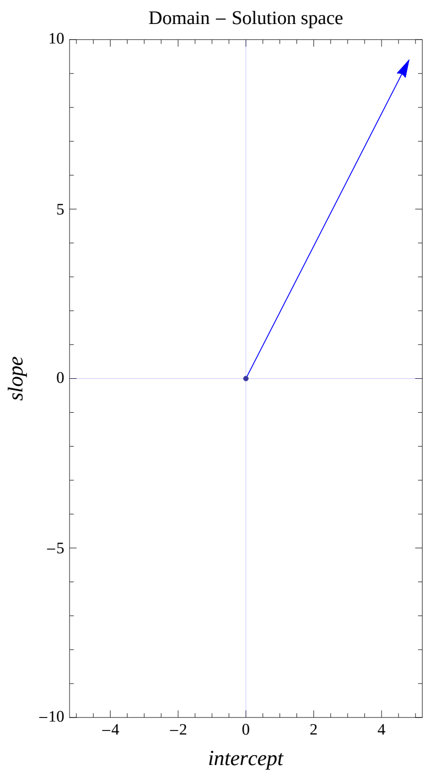


Figure 4.3. *Minimization occurs in the codomain.*

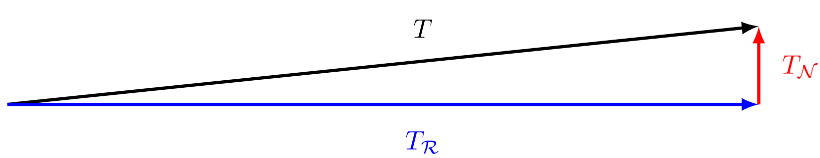


Figure 4.4. *Data vector resolved into range and null space components.*

Figure 4.5. *Decomposing* $\|r = T_{\mathcal{N}}\|_2^2$ *into residual error terms* r_k^2 *of table 4.3.*

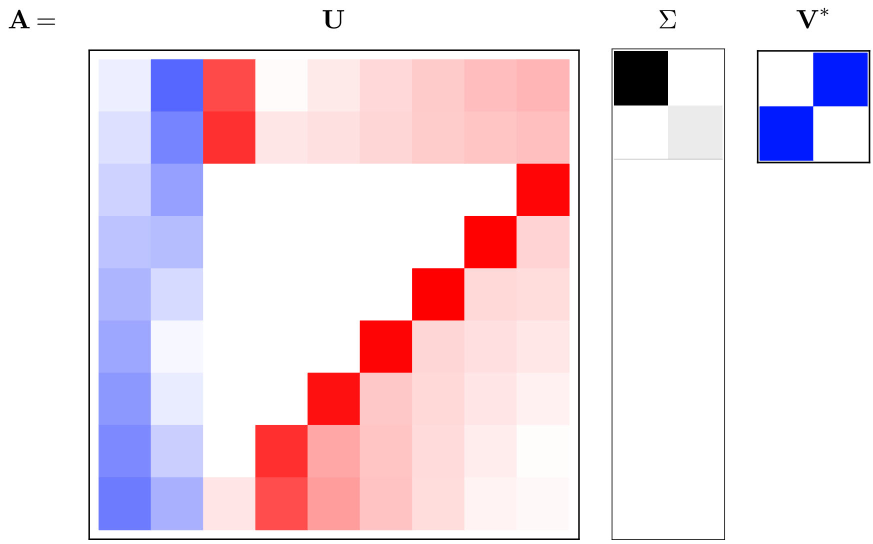


Table 4.2. *Singular value decomposition for the system matrix* $\boldsymbol{A}$ *in* (4.1) *showing range and null space components.*

Table 4.3. *A summary of the residual errors and their contributions to* $\|r\|_2$.

k	r_k	r_k^2
1	-1.4	1.9
2	6.1	37.6
3	-3.6	12.7
4	-1.4	1.8
5	-6.3	40.3
6	-0.3	0.1
7	6.5	41.9
8	9.7	93.7
9	-9.3	86.7

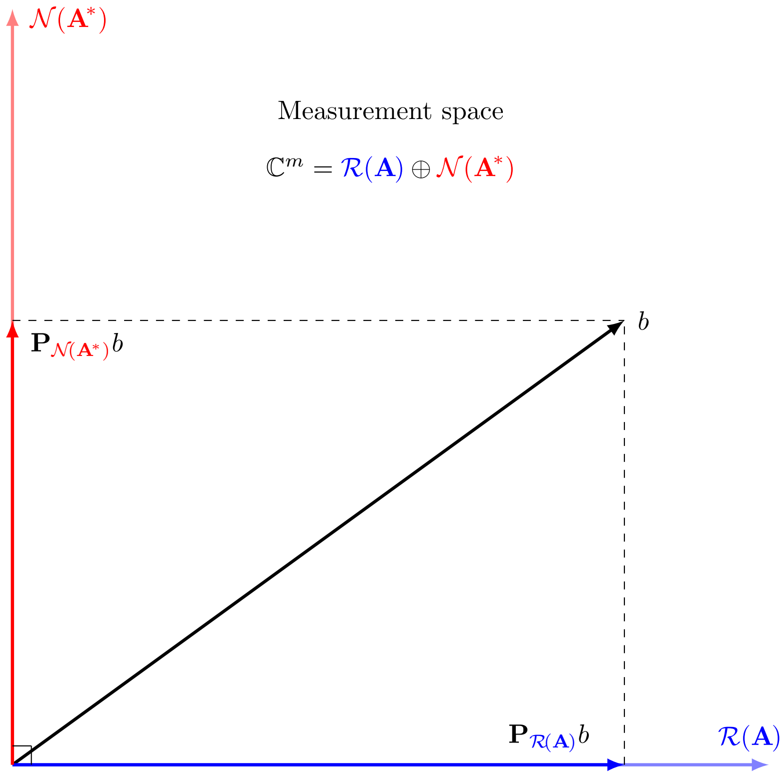


Figure 4.6. *Projection basics.*

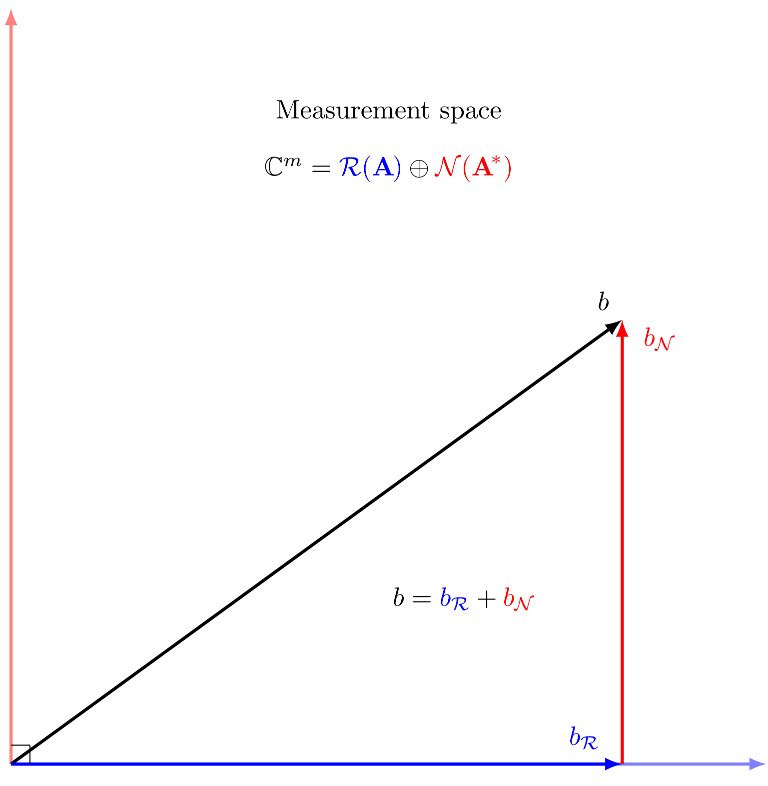
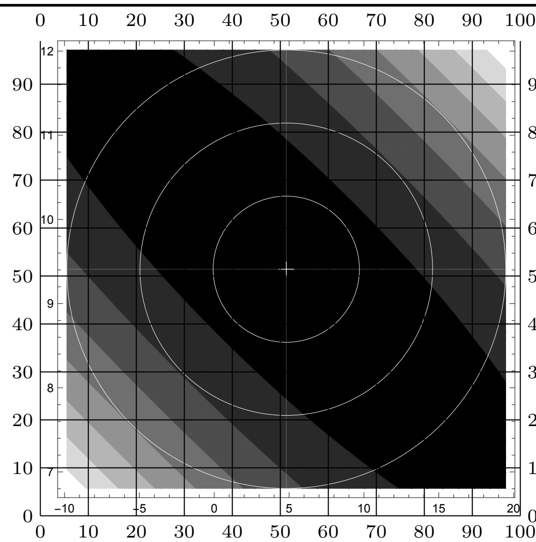
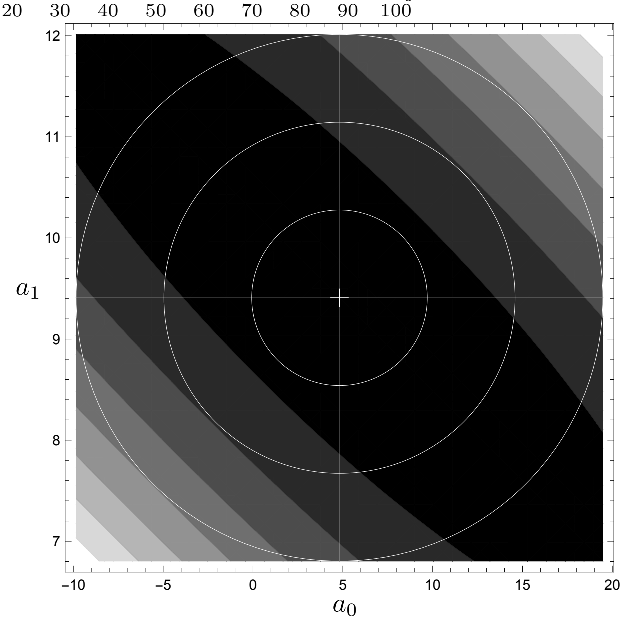


Figure 4.7. *Decomposing the data vector.*





4.3 QR Decomposition

4.3.1 Problem Statement

Chapter 5

Zonal Example

5.1 Problem

5.1.1 Zonal Subsection

Part III

Applications: Finding Patterns

Chapter 6
Lines

6.1 Face-centered cubic lattice

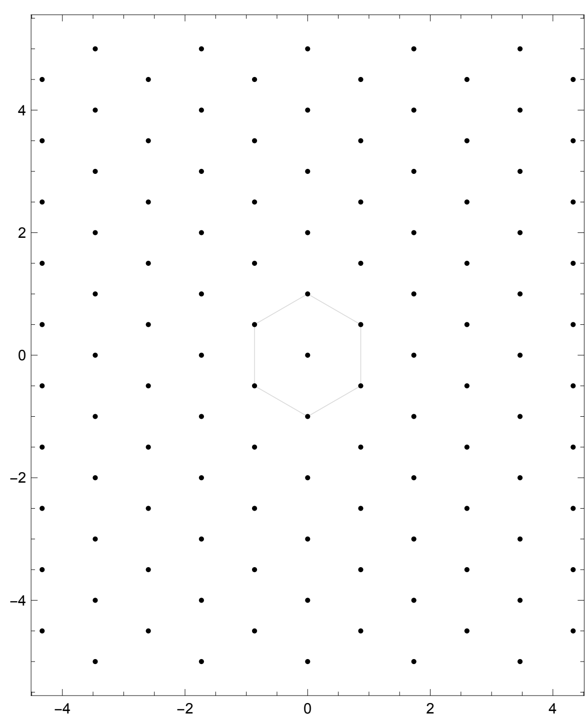


Figure 6.1. *A slice of a face-centered cubic lattice showing a single crystal.*

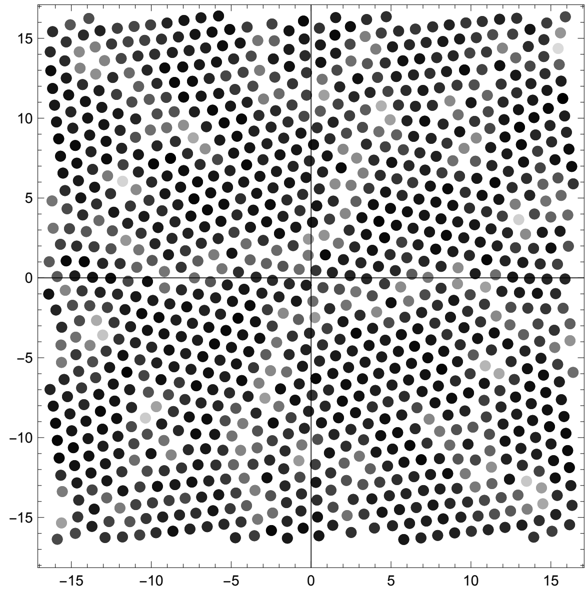


Figure 6.2. *Simulation output showing atomic shades shaded by potential energy.*

6.2 Model

$$y_{(\mu)}(x) = \mu\alpha_* + \alpha_0 + \alpha_1 x, \qquad \mu = 0, 1, 2, \ldots, M-1. \tag{6.1}$$

$$\begin{aligned}
&\begin{cases} 0 \cdot \alpha_* + \alpha_0 + \alpha_1 x_{1_1} = y_{1_1} \\ \qquad\qquad\qquad\qquad \vdots & \quad \textit{row } 1 \\ 0 \cdot \alpha_* + \alpha_0 + \alpha_1 x_{m_1} = y_{m_1} \end{cases} \\
&\begin{cases} 1 \cdot \alpha_* + \alpha_0 + \alpha_1 x_{1_2} = y_{1_2} \\ \qquad\qquad\qquad\qquad \vdots & \quad \textit{row } 2 \\ 1 \cdot \alpha_* + \alpha_0 + \alpha_1 x_{m_2} = y_{m_2} \end{cases} \\
&\begin{cases} 2 \cdot \alpha_* + \alpha_0 + \alpha_1 x_{1_3} = y_{1_3} \\ \qquad\qquad\qquad\qquad \vdots & \quad \textit{row } 3 \\ 2 \cdot \alpha_* + \alpha_0 + \alpha_1 x_{m_3} = y_{m_3} \end{cases}
\end{aligned} \tag{6.2}$$

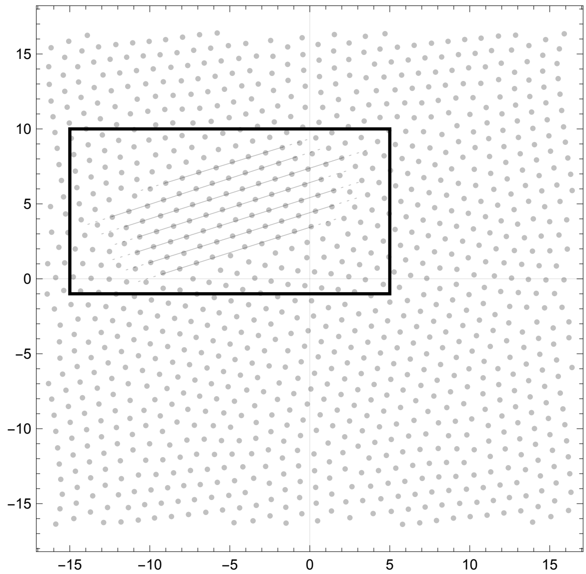


Figure 6.3. *Full data set showing inset.*

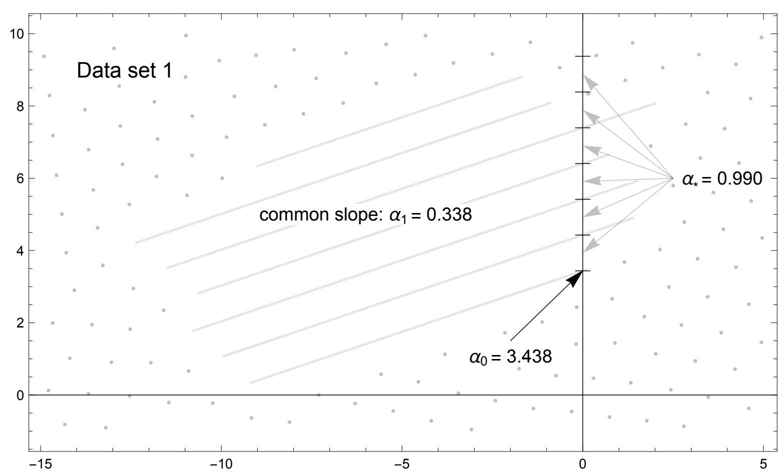
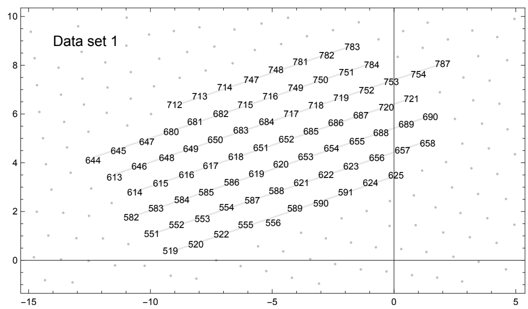
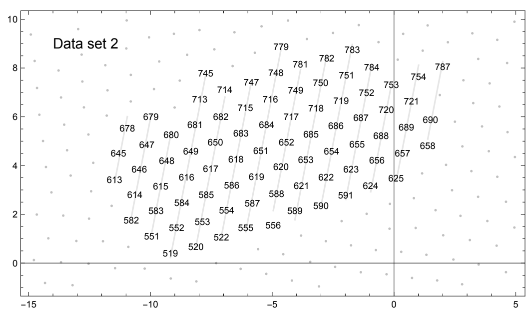


Figure 6.4. *Sample data set showing fit parameters.*

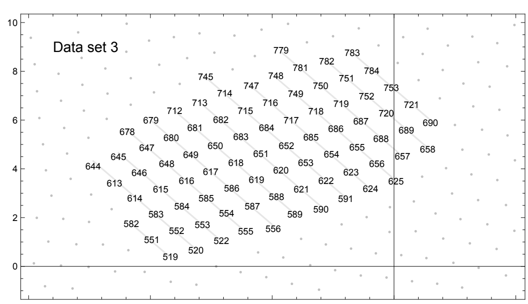
Table 6.1. *Data sets and basic results*



$$\begin{aligned} \alpha_* &= 0.9899 \pm 0.0032 \\ \alpha_0 &= 3.438 \pm 0.013 \\ \alpha_1 &= 0.3376 \pm 0.0017 \\ \sqrt{\langle r^2 \rangle} &= 0.052 \end{aligned}$$



$$\begin{aligned} \alpha_* &= 4.974 \pm 0.052 \\ \alpha_0 &= -2.075 \pm 0.093 \\ \alpha_1 &= 5.168 \pm 0.052 \\ \sqrt{\langle r^2 \rangle} &= 0.18 \end{aligned}$$



$$\begin{aligned} \alpha_* &= 1.2322 \pm 0.0039 \\ \alpha_0 &= -7.505 \pm 0.043 \\ \alpha_1 &= -0.8576 \pm 0.0038 \\ \sqrt{\langle r^2 \rangle} &= 0.054 \end{aligned}$$

$$
\begin{array}{ccccc}
\mathbf{A} & \alpha & = & y & \\
\left[\begin{array}{ccc}
0 & 1 & x_{1_1} \\
\vdots & \vdots & \vdots \\
0 & 1 & x_{\mu_1} \\ \hline
1 & 1 & x_{1_2} \\
\vdots & \vdots & \vdots \\
1 & 1 & x_{\mu_2} \\ \hline
2 & 1 & x_{1_3} \\
\vdots & \vdots & \vdots \\
2 & 1 & x_{\mu_3} \\ \hline
\vdots & \vdots & \vdots \\ \hline
M-1 & 1 & x_{1_M} \\
\vdots & \vdots & \vdots \\
M-1 & 1 & x_{m_M}
\end{array}\right]
&
\left[\begin{array}{c}
\alpha_* \\ \alpha_0 \\ \alpha_1
\end{array}\right]
& = &
\left[\begin{array}{c}
y_{1_1} \\
\vdots \\
y_{\mu_1} \\ \hline
y_{1_2} \\
\vdots \\
y_{\mu_2} \\ \hline
y_{1_3} \\
\vdots \\
y_{\mu_3} \\ \hline
\vdots \\ \hline
y_{1_M} \\
\vdots \\
y_{\mu_M}
\end{array}\right]
\end{array}
\tag{6.3}
$$

6.3 Solution

Once again the normal equations offer the easy path to solution as in (**??**). The first step is to compute the inverse of the product matrix. Recall that the dot product is a commutative operator; therefore only six of the nine matrix entries are unique:

$$
\mathbf{A}^{\mathrm{T}}\mathbf{A} = \begin{bmatrix} \mathrm{J}\cdot\mathrm{J} & \mathrm{J}\cdot\mathbf{1} & \mathrm{J}\cdot x \\ \mathbf{1}\cdot\mathrm{J} & \mathbf{1}\cdot\mathbf{1} & \mathbf{1}\cdot x \\ x\cdot\mathrm{J} & x\cdot\mathbf{1} & x\cdot x \end{bmatrix} = \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}.
$$

For clarity, the unique elements are specified:

$$
\begin{array}{ccc}
a = \mathrm{J}\cdot\mathrm{J} & b = \mathrm{J}\cdot\mathbf{1} & c = \mathrm{J}\cdot x \\
 & d = \mathbf{1}\cdot\mathrm{J} & e = \mathbf{1}\cdot x \\
 & & f = x\cdot x
\end{array}
$$

In advance of the computing the inverse, first compute the determinant

$$
\det\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right) = \Delta = 2bce + adf - ae^2 - c^2d - fb^2.
$$

Using (**??**) the inverse is

$$
\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1} = \Delta^{-1} \begin{bmatrix} df - e^2 & ce - bf & be - cd \\ \cdot & af - c^2 & bc - ae \\ \cdot & \cdot & ad - b^2 \end{bmatrix}.
$$

The right-hand side in (6.3) is

$$\mathbf{A}^{\mathrm{T}} y = \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} \mathrm{J} \cdot y \\ \mathbf{1} \cdot y \\ x \cdot y \end{bmatrix}.$$

The least squares solution is provided as

$$\begin{bmatrix} \alpha_0 \\ \alpha_* \\ \alpha_1 \end{bmatrix} = \left(\mathbf{A}^{\mathrm{T}} \mathbf{A} \right)^{-1} \mathbf{A}^{\mathrm{T}} y$$

which distills down to

$$\alpha = \Delta^{-1} \begin{bmatrix} \beta_1 \left(df - e^2 \right) & + \beta_2 \left(ce - bf \right) & + \beta_3 \left(be - cd \right) \\ \beta_1 \left(ce - bf \right) & + \beta_2 \left(af - c^2 \right) & + \beta_3 \left(bc - ae \right) \\ \beta_1 \left(be - cd \right) & + \beta_2 \left(bc - ae \right) & + \beta_3 \left(ad - b^2 \right) \end{bmatrix}.$$

The errors associate with the fit parameters are

$$\begin{bmatrix} \sigma_* \\ \sigma_0 \\ \sigma_1 \end{bmatrix} = \sqrt{\frac{r^{\mathrm{T}}}{(m-n)\,\Delta}} \sqrt{\begin{bmatrix} df - e^2 \\ af - c^2 \\ ad - b^2 \end{bmatrix}}.$$

The solutions are expressed in terms of dot products readily available in Fortran.

6.4 Problem Statement

6.5 Data

6.6 Results

6.6.1 Least Squares Results

6.6.2 Apex Angles

6.6.3 Qualitative Results

Table 6.2. *Problem statement for grain identification by rows (coupled linear regression).*

trial function	$y_{(\mu)}(x) = \mu\alpha_* + \alpha_0 + \alpha_1 x$	
merit function	$M(p) = \sum_{k=1}^{n} (y_k - \mu\alpha_* + \alpha_0 + \alpha_1 x_k)^2$	
number of zones	$m = 5$	
number of overlaps	$n = 4$	
rank defect	$m - n = 1$	
measurements	$\lambda = \{11, 13, 13, 13, 12\}$	
measurements	(x_k, y_k), $k = 1\colon 1024$	
system matrix	$\mathbf{A} = \begin{bmatrix} \mathbf{1} & x \end{bmatrix}$	
data vector	y	
linear system	(6.3)	
results	α_*	gap
	α_0	$y-$axis intercept
	α_1	slope
residual error	$r = \mathbf{A}\alpha - y$	

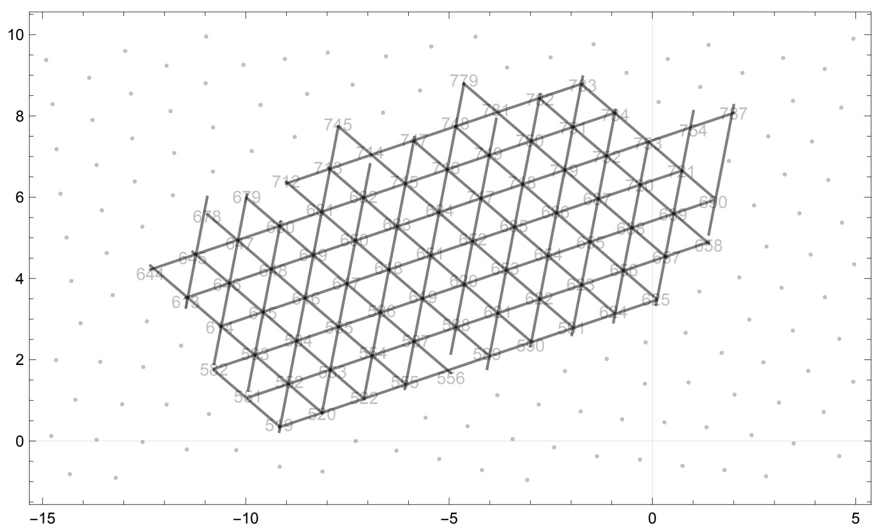


Figure 6.5. *Solutions for three data sets.*

Table 6.3. *Point membership in data sets shown in figure 6.1.*

set	row	1	2	3	4	5	6	7	8	9	10	11	12
1	1	519	520	522	555	556	589	590	591	624	625		
1	2	551	552	553	554	587	588	621	622	623	656	657	658
1	3	582	583	584	585	586	619	620	653	654	655	688	689
1	4	614	615	616	617	618	651	652	685	686	687	720	721
1	5	613	646	648	649	650	683	684	717	718	719	752	753
1	6	644	645	647	680	681	682	682	715	716	749	750	751
1	7	712	713	714	747	748	781	782	783				
2	1	658	690	787									
2	2	625	657	689	721	754							
2	3	624	656	688	720	753							
2	4	591	623	655	687	752	784						
2	5	590	622	654	686	719	751	783					
2	6	589	621	653	685	718	750	782					
2	7	556	588	620	652	717	749	781					
2	8	555	587	619	651	684	716	748	779				
2	9	522	554	586	618	683	715	747					
2	10	520	553	585	617	650	682	714					
2	11	519	552	584	616	649	681	713	745				
2	12	551	583	615	648	680							
2	13	582	614	646	647	679							
2	14	613	645	678									
3	1	582	551	519									
3	2	644	613	614	583	552	520						
3	3	645	646	615	584	553	522						
3	4	647	648	616	585	554	555	678					
3	5	679	680	649	617	586	587	556					
3	6	712	681	650	618	619	588	589					
3	7	713	682	683	651	620	621	590					
3	8	745	714	715	684	652	653	622	591				
3	9	747	716	717	685	654	623	624					
3	10	748	749	718	686	655	656	625					
3	11	779	781	750	719	687	688	657					
3	12	782	751	752	720	689	658						
3	13	783	784	753	721	690							

Table 6.4. *Excerpted data set.*

k	x_k	y_k	ϕ_k
1	−15.879001	−16.365496	−2.597531
2	−14.749446	−15.995488	−2.613017
3	−13.905339	16.242941	−2.557543
⋮			
1 022	13.927362	−16.235010	−2.780323
1 023	14.741765	15.957687	−2.687929
1 024	15.905518	16.346979	−2.599001

Table 6.5. *Least squares results for three axes.*

axis		gap			intercept			slope		$\sqrt{\langle r^2 \rangle}$
1	0.9899	±	0.0032	3.438	±	0.013	0.3376	±	0.0017	0.052
2	4.974	±	0.052	−2.075	±	0.093	5.168	±	0.052	0.18
3	1.2322	±	0.0039	−7.505	±	0.043	−0.8576	±	0.0038	0.054

Table 6.6. *Intermediate results: angles for the axes.*

axis	θ	±	σ_θ				
1	0.3256	±	0.0015	=	(18.655	±	0.086)°
2	1.380	±	0.018	=	(79.0	±	1.0)°
3	−0.7089	±	0.0025	=	(−40.62	±	0.14)°

Table 6.7. *Final results: apex angle measurements*

	θ	±	σ_θ				
α	1.040	±	0.018	=	(59.6	±	1.0)°
β	1.0345	±	0.0029	=	(59.27	±	0.17)°
γ	1.041	±	0.018	=	(59.7	±	1.0)°
total	3.116	±	0.026	=	(178.5	±	1.7)°

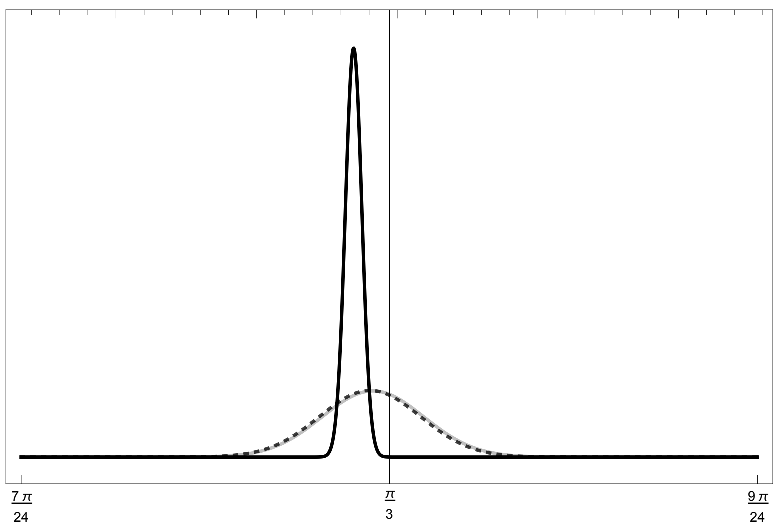


Figure 6.6. *Apex angles displayed in table 6.7.*

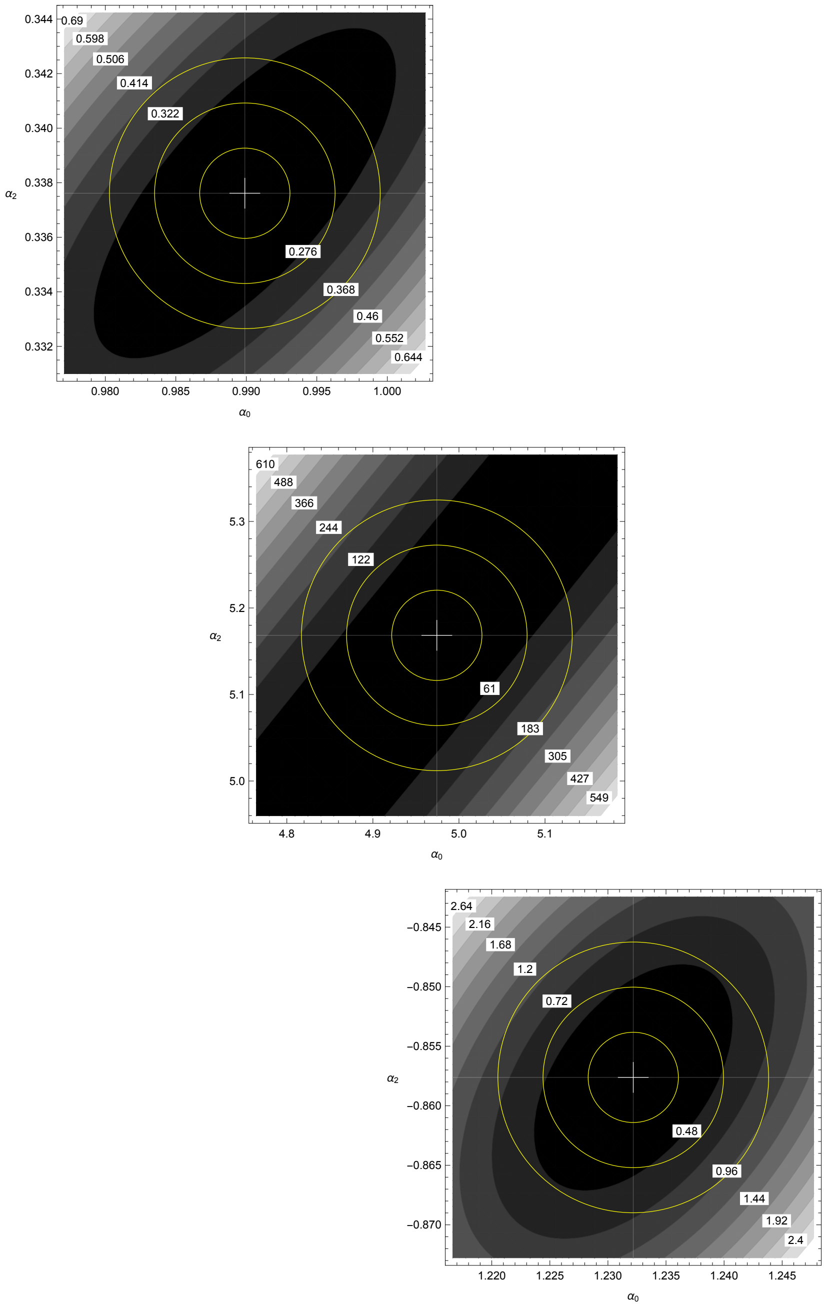


Figure 6.7. *Merit functions for the three data sets.*

Chapter 7

Crystals

In the previous model, the rows of atoms were treated independently. In this section the basic unit is not a row, it is instead a crystal. Mathematically, the process will imitate Nature: a seed crystal is picked, and other crystals will be identified from that.

Part IV

Applications: Stitching

Chapter 8

Stitching Local Maps

8.1 What is stitching?

Stitching is the process of combining local maps to create a global map.

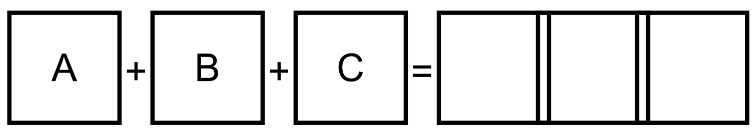


Figure 8.1. *Stitching local maps together to form a global map.*

1. ϕ
2. $\nabla\phi$
3. ϕ and $\nabla\phi$

8.2 Stitch ϕ

8.2.1 Genesis

$$\phi(x) = \exp\left(-\frac{x}{5}\right)\sin\left(\pi x\right)$$

8.2.2 Data

The central idea is simple; the mathematical expression is a tedious exercise in index gymnastics.

$$\zeta = 3$$

Table 8.1. *The input data in continuous and discrete form.*

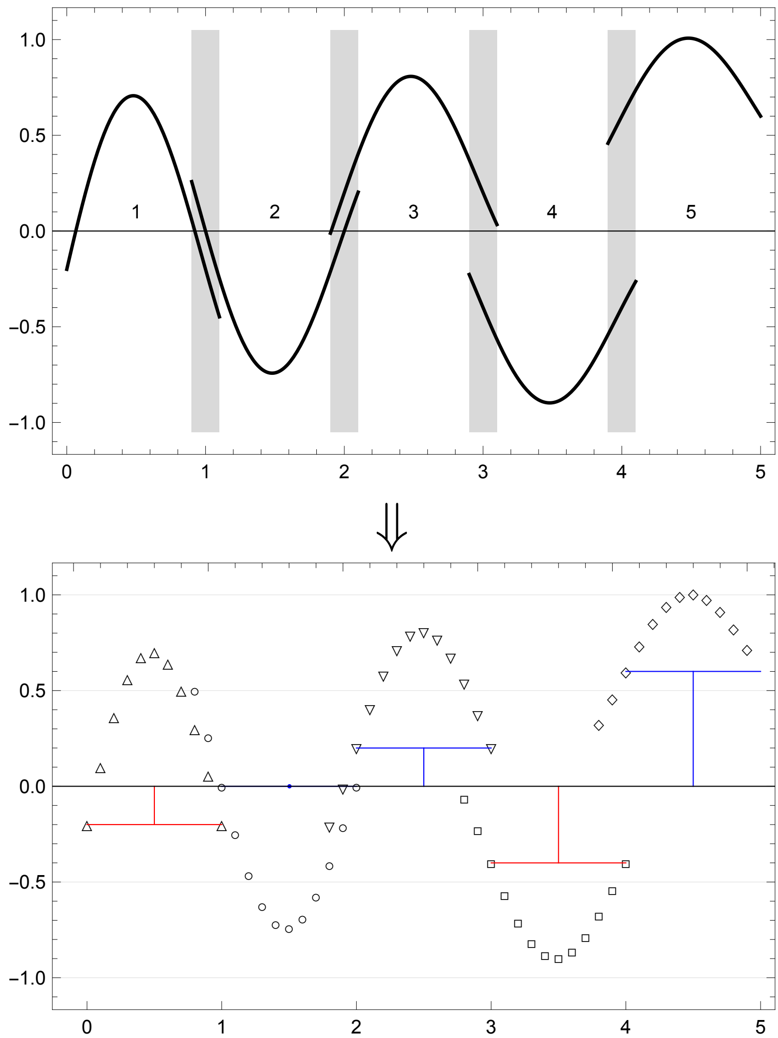


Table 8.2. *Sample showing an overlap of $\zeta = 3$ between the first two zones.*

$\phi_{1,1}$	$\phi_{2,1}$	$\cdots$	$\phi_{\lambda_1-2,1}$	$\phi_{\lambda_1-1,1}$	$\phi_{\lambda_1,1}$				
			$\phi_{1,2}$	$\phi_{2,2}$	$\phi_{3,2}$	$\cdots$	$\phi_{\lambda_1-2,2}$	$\phi_{\lambda_2-1,2}$	$\phi_{\lambda_2,2}$

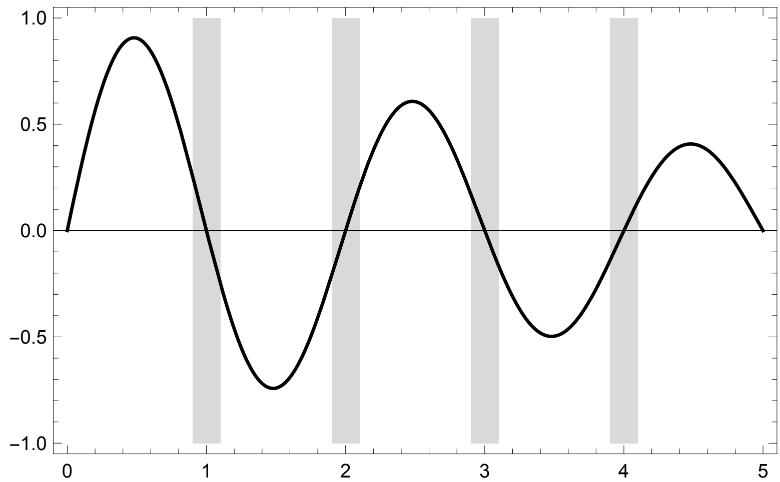


Figure 8.2. *The ideal potential function showing five measurement zones and four overlap bands.*

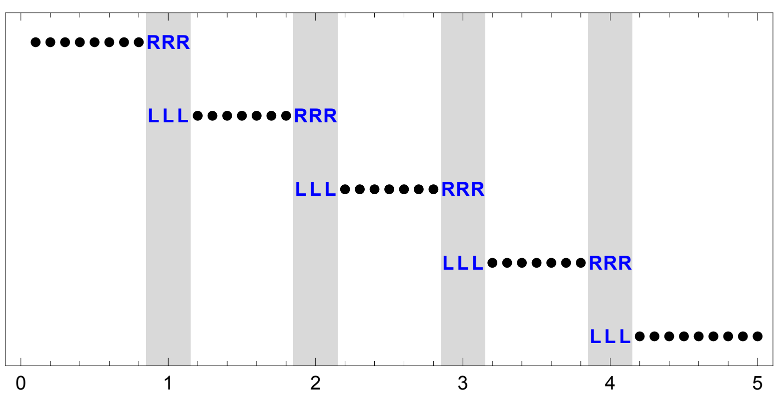


Figure 8.3. *Waterfall diagram showing discretization within measurement zones with left and right zone overlaps.*

$$\Delta_{12} = \zeta^{-1}\left(\underbrace{(\phi_{\lambda_1-2,1}+\phi_{\lambda_1-1,1}+\phi_{\lambda_1,1})}_{\text{zone 1}} - \underbrace{(\phi_{1,2}+\phi_{2,2}+\phi_{3,2})}_{\text{zone 2}}\right)$$

$$\Delta_{12} = \zeta^{-1}\left(\underbrace{(\phi_{\lambda_1-2,1}-\phi_{1,2})}_{\text{pair 1}} + \underbrace{(\phi_{\lambda_1-1,1}-\phi_{2,2})}_{\text{pair 2}} + \underbrace{(\phi_{\lambda_1,1}-\phi_{3,2})}_{\text{pair 3}}\right)$$

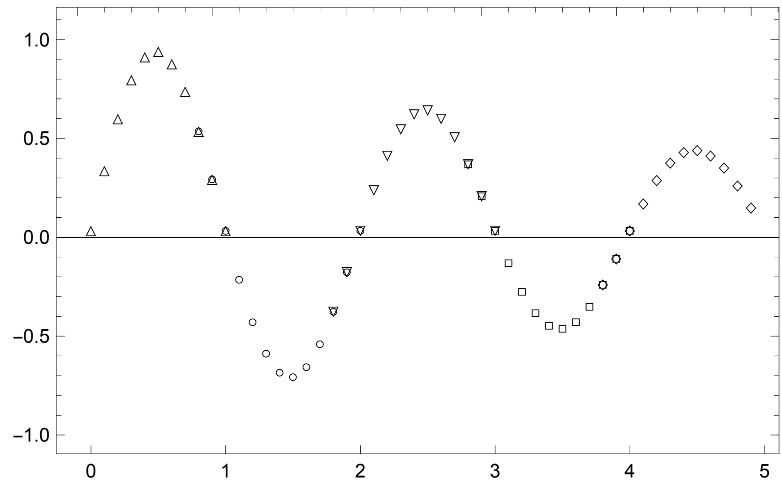


Figure 8.4. *Stitching unifies the data.*

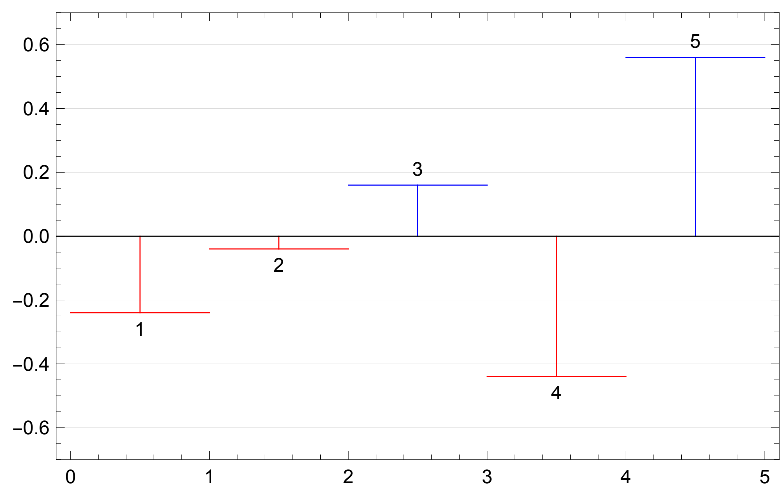


Figure 8.5. *A set of piston adjustments which restores continuity across the domain.*

Mean value of the differences.

$$\Delta_{j,j+1} = \zeta^{-1} \sum_{k=1}^{\zeta} p_{j,\lambda_j-\zeta+k} - p_{j+1,k}$$

k	y_1	y_2	y_3	y_4	y_5
1	−0.2	0.500 878	−0.210 084	−0.064 251 7	0.325 113
2	0.102 898	0.258 113	−0.011 324 8	−0.226 982	0.458 345
3	0.364 738	0.0	0.2	−0.4	0.6
4	0.561 904	−0.247 992	0.403 039	−0.566 234	0.736 101
5	0.677 936	−0.462 368	0.578 555	−0.709 935	0.853 753
6	0.704 837	−0.623 794	0.710 719	−0.818 142	0.942 345
7	0.643 511	−0.718 793	0.788 498	−0.881 821	0.994 482
8	0.503 326	−0.740 818	0.806 531	−0.896 585	1.006 57
9	0.300 878	−0.690 609	0.765 423	−0.862 929	0.979 014
10	0.058 112 7	−0.575 834	0.671 453	−0.785 993	0.916 025
11	−0.2	−0.410 084	0.535 748	−0.674 887	0.825 059
12		−0.211 325	0.373 018	−0.541 655	0.715 978
13		0.0	0.2	−0.4	

Table 8.3. *Measurements displaying the connection between overlap bands in figure 8.3.*

First overlap region.

$$\Delta_{12} = \zeta^{-1}\left(\underbrace{(\phi_{9,1}+\phi_{10,1}+\phi_{11,1})}_{\text{last 3 elements of zone 1}} - \underbrace{(\phi_{1,2}+\phi_{2,2}+\phi_{3,2})}_{\text{first 3 elements of zone 2}}\right)$$

8.2.3 Data and results

Table 8.4. *Computation of the zone shift values.*

$$\begin{aligned}
\Delta_{12} &= \tfrac{1}{3}\left((\phi_{9,1}+\phi_{10,1}+\phi_{11,1})-(\phi_{1,2}+\phi_{2,2}+\phi_{3,2})\right)\\
\Delta_{23} &= \tfrac{1}{3}\left((\phi_{11,2}+\phi_{12,2}+\phi_{13,2})-(\phi_{1,3}+\phi_{2,3}+\phi_{3,3})\right)\\
\Delta_{34} &= \tfrac{1}{3}\left((\phi_{11,3}+\phi_{12,3}+\phi_{13,3})-(\phi_{1,4}+\phi_{2,4}+\phi_{3,4})\right)\\
\Delta_{45} &= \tfrac{1}{3}\left((\phi_{11,4}+\phi_{12,4}+\phi_{13,4})-(\phi_{1,5}+\phi_{2,5}+\phi_{3,5})\right)
\end{aligned}$$

8.2.4 Linear System

$$\mathbf{A}p = \Delta$$

Table 8.5. *Computation of the zone shift values.*

$$
\begin{aligned}
\Delta_{12} &= \tfrac{1}{3}\left((0.300878 + 0.0581127 - 0.2) - (0.500878 + 0.258113 + 0.)\right) \\
\Delta_{23} &= \tfrac{1}{3}\left((-0.410084 - 0.211325 + 0.) - (-0.210084 - 0.0113248 + 0.2)\right) \\
\Delta_{34} &= \tfrac{1}{3}\left((0.535748 + 0.373018 + 0.2) - (-0.0642517 - 0.226982 - 0.4)\right) \\
\Delta_{45} &= \tfrac{1}{3}\left((-0.674887 - 0.541655 - 0.4) - (0.325113 + 0.458345 + 0.6)\right)
\end{aligned}
$$

Table 8.6. *Input data*

	Shift	Value
1	Δ_{12}	–0.2
2	Δ_{23}	0.0
3	Δ_{34}	0.6
4	Δ_{45}	–1.

$$
\begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}
\begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{bmatrix}
=
\begin{bmatrix} \Delta_{12} \\ \Delta_{23} \\ \Delta_{34} \\ \Delta_{45} \end{bmatrix}
$$

$$
p_{LS} = \frac{1}{5}
\begin{bmatrix} 4 & 3 & 2 & 1 \\ -1 & 3 & 2 & 1 \\ -1 & -2 & 2 & 1 \\ -1 & -2 & -3 & 1 \\ -1 & -2 & -3 & -4 \end{bmatrix}
\begin{bmatrix} \Delta_{12} \\ \Delta_{23} \\ \Delta_{34} \\ \Delta_{45} \end{bmatrix}
+ \alpha
\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}
$$

$$
\mathbf{A}^{\dagger} b = \frac{1}{25}
\begin{bmatrix} -6 \\ -1 \\ 4 \\ -11 \\ 14 \end{bmatrix}
$$

These are the actual plot values used in figure 8.5.

$$
\Phi_{corrected} = \Phi_{measured} - \mathbf{A}^{\dagger} b
$$

Table 8.7. *Problem statement for linear regression.*

trial function	$p_k - p_{k+1} = \Delta_{k,k+1},\ k = 1\colon n$
merit function	$M(p) = \sum_{k=1}^{n} \left(\Delta_{k,k+1} - p_k + p_{k+1}\right)^2$
number of zones	$m = 5$
number of overlaps	$n = 4$
rank defect	$m - n = 1$
measurements per zone	$\lambda = \{11, 13, 13, 13, 12\}$
measurements	$\phi_{k,j},\ k = 1\colon m,\ j = 1\colon \lambda_m$
input data	$\Delta_{k,k+1},\ k = 1\colon n$
results	$p_k,\ k = 1\colon m$
residual error	$r = \mathbf{A}^{\dagger} b - \Delta$
linear system	$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{bmatrix} = \begin{bmatrix} \Delta_{12} \\ \Delta_{23} \\ \Delta_{34} \\ \Delta_{45} \end{bmatrix}$
gauge condition	$\sum_{k=1}^{m} p_k = 0$

8.2.5 Least Squares Arbitration

There is a fundamental ambiguity arising from gradient measurements stemming from the basic fact that

$$\frac{d}{dx}\phi(x) = \frac{d}{dx}\left(\phi(x) + c\right).$$

We can recover the function shape, but not the offset. In other words, there is a translation invariance. This lone constant is the poster child for the rank one deficiency in the linear system of (8.2.4). Realizing this, the one dimensional problem could be solved without resort to least squares.

The system can be solved, for example, by moving from left to right and manually forcing the data to match. If the the overlap difference between zone 1 and zone 2 is Δ_{12}, add Δ_{12} to every value in zone 2. Now zones 1 and 2 are stitched together. Compute Δ_{23}, add this value to every point in zone 3. Zones 1, 2, and 3 are now stitched together. Continue as needed.

The least squares problem is obviated. How did this happen? The process of least squares is an exercise error arbitration which takes a peanut butter approach by trying to distribute the error evenly. In one dimension, there is no need for arbitration as there is no conflict in measurements.

In two dimensions, the problem changes. Consider the typical cell with a neighbor to the right and a neighbor above. The right–left overlap adjustment conflicts with the up–down overlap adjustment. The least squares process takes

Table 8.8. *Results for stitching with piston.*

fit parameters	$p_k,\ k = 1\colon m$	pistons
computed solution	$p = \frac{1}{25}\begin{bmatrix} -6 \\ -1 \\ 4 \\ -11 \\ 14 \end{bmatrix}$	$\mathbf{A}^{\dagger}b$
data vector	$\frac{1}{5}\begin{bmatrix} -1 \\ 0 \\ 1 \\ -2 \\ 3 \end{bmatrix} = \frac{1}{25}\begin{bmatrix} -6 \\ -1 \\ 4 \\ -11 \\ 14 \end{bmatrix} - \frac{1}{5}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$	$p_{\mathcal{R}} + p_{\mathcal{N}}$
residual error	$r \cdot r = 0$	
problem statement	table 8.7	
measurements	table 8.3	
input data	table 8.6	
plots	figure 8.2	raw data (bottom)
	figure 8.4	corrected data

all off the overlap conflicts and provides a set of adjustments which minimizes the global error. To close, note that the least squares solution was used even though it is not necessary until dimension 2 or higher.

One last tidbit. Figure 8.7 shows the piston values that were input to distort the values. Least squares chooses a distinct set of corrections. Why was this set selected? A tantalizing clue is given by the null space vector in (8.2.4). Notice this vector is perpendicular to every column vector in $\mathbf{A}^{\dagger}$ which implies that the sum of each column vector must be 0. Therefore, the gauge condition is that the solution vector will have sum 0:

$$p_1 + p_2 + p_3 + p_4 + p_5 = 0.$$

We may now eliminate a variable; choose the last one:

$$p_5 = -p_1 - p_2 - p_3 - p_4$$

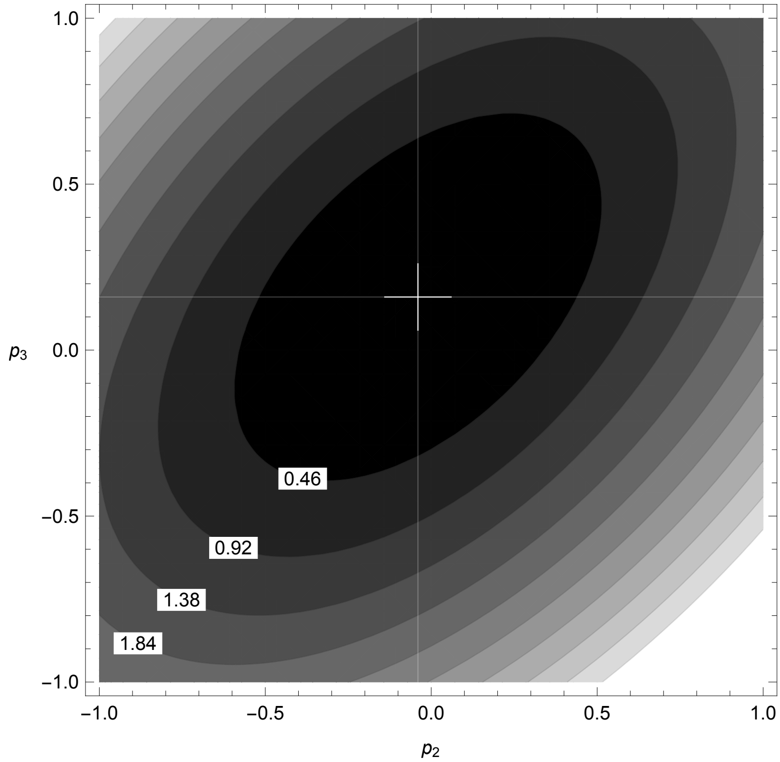


Figure 8.6. *Looking at the merit function on the $p_2 - p_3$ axis.*

Instead of (8.2.4), there is now

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 1 & 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix} = \begin{bmatrix} \Delta_{12} \\ \Delta_{23} \\ \Delta_{34} \\ \Delta_{45} \end{bmatrix}.$$

The solution is the same:

$$\hat{p}_{gauge} = \begin{bmatrix} 4 & 3 & 2 & 1 \\ -1 & 3 & 2 & 1 \\ -1 & -2 & 2 & 1 \\ -1 & -2 & -3 & 1 \end{bmatrix} \begin{bmatrix} \Delta_{12} \\ \Delta_{23} \\ \Delta_{34} \\ \Delta_{45} \end{bmatrix}.$$

The 0 sum, or equivalently 0 mean, condition is a gauge condition which restores the column rank of the problem.

The piston values used to create the data set are decomposed into range and

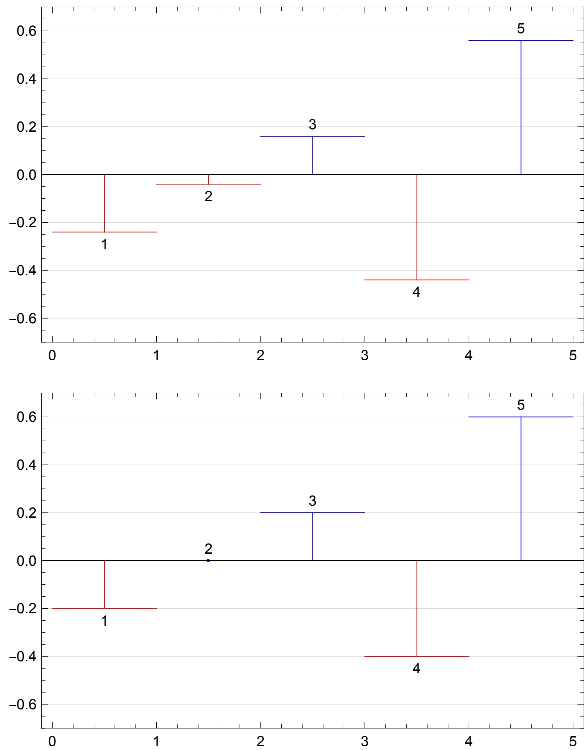


Figure 8.7. *On top, pistons output from the solution; on bottom, pistons input to create the data.*

null space terms.

$$\frac{1}{5}\begin{bmatrix}-1\\0\\1\\-2\\3\end{bmatrix}=\frac{1}{25}\begin{bmatrix}-6\\-1\\4\\-11\\14\end{bmatrix}-\frac{1}{5}\begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix}$$

8.3 Stitch $\nabla\phi$

The next challenge is to stitch data together using the gradient $\nabla\phi$ rather than the function value ϕ. The outputs now will be a set of piston adjustments called tilts which restore continuity of the gradient.

The problem arises in the field of wavefront sensing. Modern devices make exquisite measurements of tilts. The process of wavefront reconstruction takes these tilts and reconstructs the wavefront. Measure $\nabla\phi(x)$ and compute $\nabla\phi(x)$.

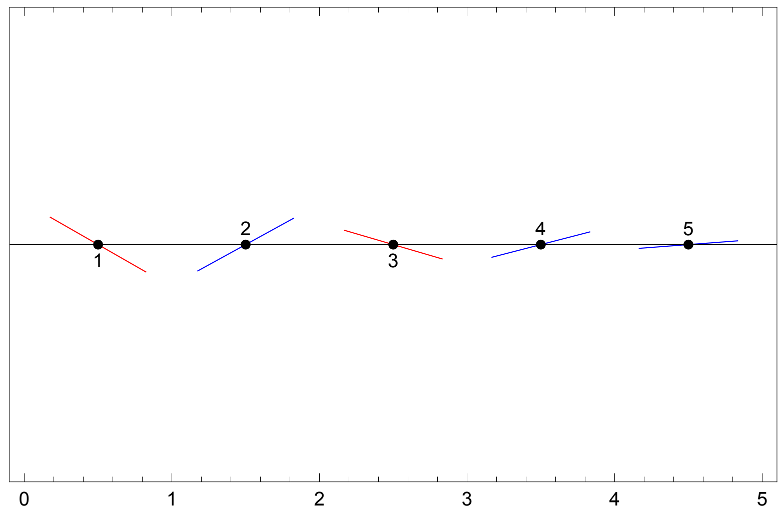


Figure 8.8. *A set of tilt adjustments which restores continuity of the gradient across the domain.*

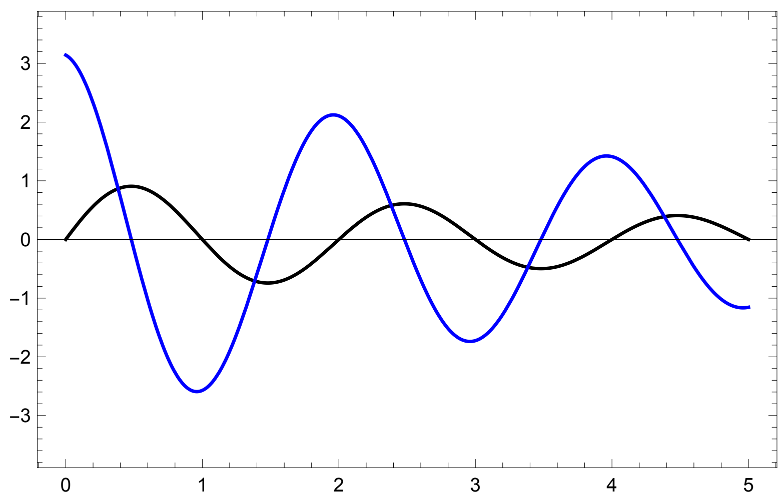


Figure 8.9. *A function (black) and its gradient (blue).*

Gradient of (8.3)

$$\nabla\phi(x) = \frac{1}{5}\exp\left(-\frac{x}{5}\right)(5\pi\cos(\pi z) - \sin(\pi z))$$

$$\tau = \frac{1}{100} \begin{bmatrix} -90 \\ 85 \\ -40 \\ 35 \\ 10 \end{bmatrix}$$

A scaled version of these values is plotted in figure 8.8.

Part V

Applications: Inverting the Gradient

Chapter 9

Gradient I

$$F = \nabla \phi$$

$$W^{1,2}(\Omega) = \left\{ \phi \in L^2(\Omega) : \partial_x^1 \phi \in L^2(\Omega) \right\}$$

9.1 One Dimension

$$\Omega = \bigcup_k \omega_k$$

Interval

$$\omega = \{x \in \mathbb{R} \colon a < x < b\}$$

Average gradient

$$\langle \nabla \phi(x) \rangle_\omega = \phi(b) - \phi(a)$$

$$\begin{bmatrix} -1 & 1 & 0 & \dots & & \\ 0 & -1 & 1 & & & \\ \vdots & & \ddots & \ddots & & \vdots \\ & & & -1 & 1 & 0 \\ & & & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \varphi_0 \\ \varphi_1 \\ \vdots \\ \varphi_{m-1} \\ \varphi_m \end{bmatrix} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}$$

Part VI

Applications: Nonlinear Problems

Chapter 10

Linearization

10.1 Linear Transformation

A linear transformation T satisfies the requirement

$$T(x + \alpha y) = T(x) + \alpha T(y).$$

An immediate consequence is

$$T(\underbrace{x - x}_{0}) = \underbrace{T(x) - T(x)}_{0} = 0,$$

therefore, because $x - x = 0$, we must have $T(0) = 0$.

Is the transformation $T(x) = a_0 + a_1 x$ a linear transformation? No, because $T(0) = b$. This is, however, an example of an affine transformation.

10.2 Mythology

An enduring misadventure in least squares is to hope that an exponential function like

$$y(x) = a_0 e^{a_1 x}$$

can be linearized with a logarithmic transformation:

$$\tilde{y}(x) = \ln (y(x)) = \ln a_0 + a_1 x.$$

Is the logarithm a linear transformation? Of course not:

$$\ln (x + \alpha y) \neq \ln x + \alpha \ln y.$$

Chapter 11

Population Growth

In this section we take a nonlinear model for population growth and separate the linear and nonlinear terms.

11.1 Model

$$y(\tau) = a_1 + a_2\tau + a_3 e^{d\tau} \tag{11.1}$$

$$\mathbf{A}\left(d+\gamma\right) \neq \mathbf{A}\left(d\right) + \mathbf{A}\left(\gamma\right)$$

$$\begin{array}{cccc}
\mathbf{A}(d) & a & = & y \\
\left[\begin{array}{ccc}
1 & \tau_1 & e^{d\tau_1} \\
1 & \tau_2 & e^{d\tau_2} \\
1 & \tau_3 & e^{d\tau_3} \\
1 & \tau_4 & e^{d\tau_4} \\
1 & \tau_5 & e^{d\tau_5} \\
1 & \tau_6 & e^{d\tau_6} \\
1 & \tau_7 & e^{d\tau_7} \\
1 & \tau_8 & e^{d\tau_8}
\end{array}\right] &
\left[\begin{array}{c} a_0 \\ a_1 \\ a_3 \end{array}\right] & = &
\left[\begin{array}{c} y_0 \\ y_1 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{array}\right]
\end{array}$$

$$\min_{\substack{a\in\mathbb{R}^3 \\ d\in\mathbb{R}}} \left\| \mathbf{A}(d) \left[\begin{array}{c} a_1 \\ a_2 \\ a_3 \end{array}\right] - y \right\|_2^2 \tag{11.2}$$

Table 11.1. *Problem statement for population model with linear and exponential growth.*

trial function	$y(\tau) = a_0 + a_1\tau + a_2 e^{d\tau}$	$a \in \mathbb{R}^3$
		$d \in \mathbb{R}$
merit function	$M(a,d) = \sum\limits_{k=1}^{m} \left(y_k - a_1 + a_2\tau + a_3 e^{d\tau_k} \right)^2$	
# measurements	$m = 8$	
# parameters	$n = 4$	
rank defect	$\rho = n$	overdetermined
input data	$(\tau_k, y_k),\ k = 1\colon 8$	table 11.2
results	a_0	constant
	a_1	linear
	a_2	exponential
	d	power term
residual error	$r = \mathbf{A}^{\dagger} b - \Delta$	
linear system	$\mathbf{A}(d) \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = y$	

11.2 Problem Statement

11.3 Data

11.4 Example

$\text{year} = 1900 + 10(\tau - 1)$

11.5 Polynomials

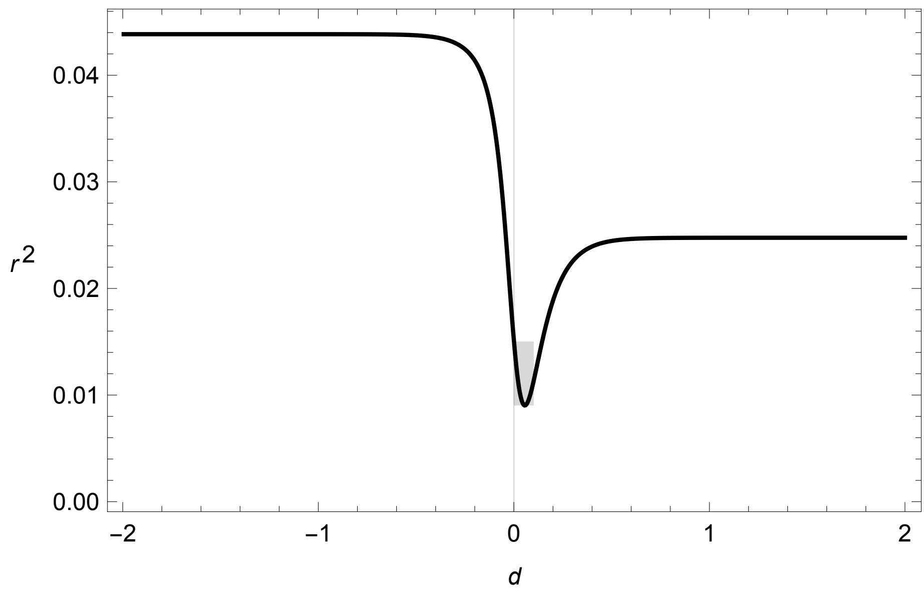
There is the model we choose and the model which nature chooses. Are they the same?

Table 11.2. *Data v. prediction.*

year	census	fit	r	rel. error
1900	76.00	77.51	1.51	2.0%
1910	91.97	90.98	−0.99	−1.1%
1920	105.71	104.87	−0.84	−0.8%
1930	122.78	119.48	−3.29	−2.7%
1940	131.67	135.36	3.69	2.8%
1950	150.70	153.46	2.76	1.8%
1960	179.32	175.45	−3.87	−2.2%
1970	203.24	204.26	1.029	0.5%

Table 11.3. *Results: census*

fit parameters	$c = \begin{bmatrix} 0.010 \\ 0.0170 \\ 0.0096 \end{bmatrix} \pm \begin{bmatrix} 0.031 \\ 0.0014 \\ 0.0020 \end{bmatrix}$
	$d = 0.056136 \pm ?.?$
r^*r	0.009025
a	$\begin{bmatrix} 0.5397 & -0.0188 & 0.0165 \\ -0.0188 & 0.0011 & -0.0014 \\ 0.0165 & -0.0014 & 0.0022 \end{bmatrix}$
plots	data vs fit (**??**)
	residuals (**??**)
	merit function in $\mathcal{R}(\mathbf{A}^*)$ (**??**)



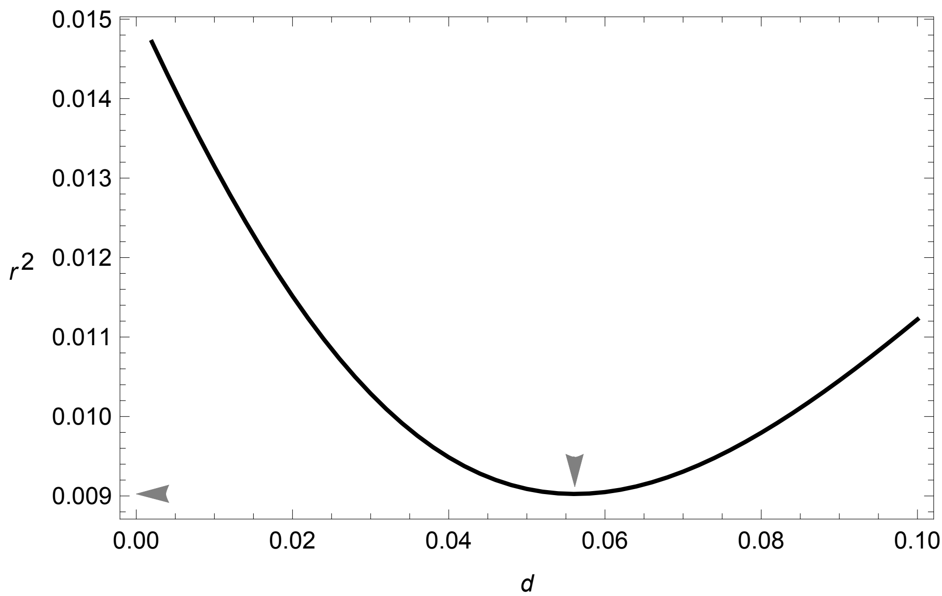


Figure 11.1. *The shaded region in this plot is shown below.*

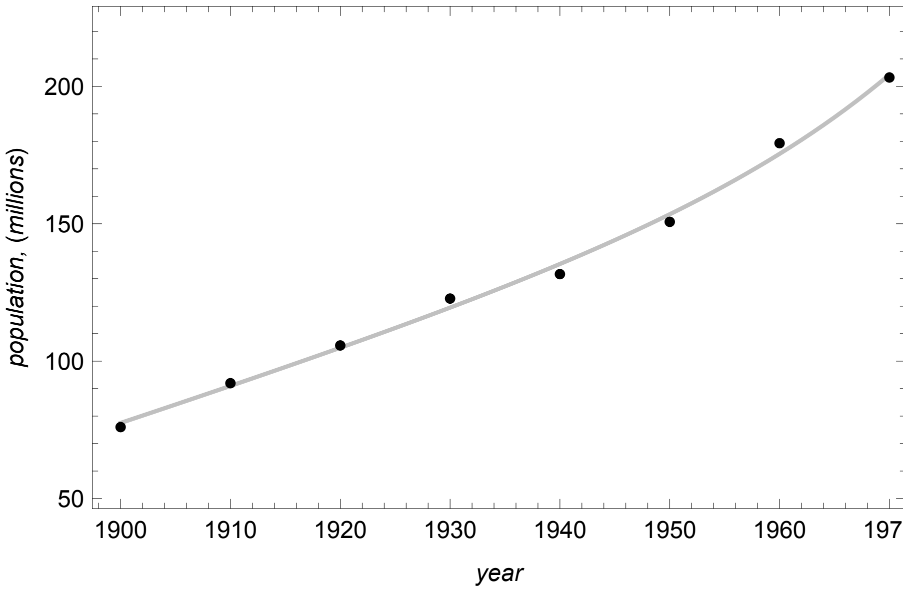


Figure 11.2. *Solution plotted against data.*

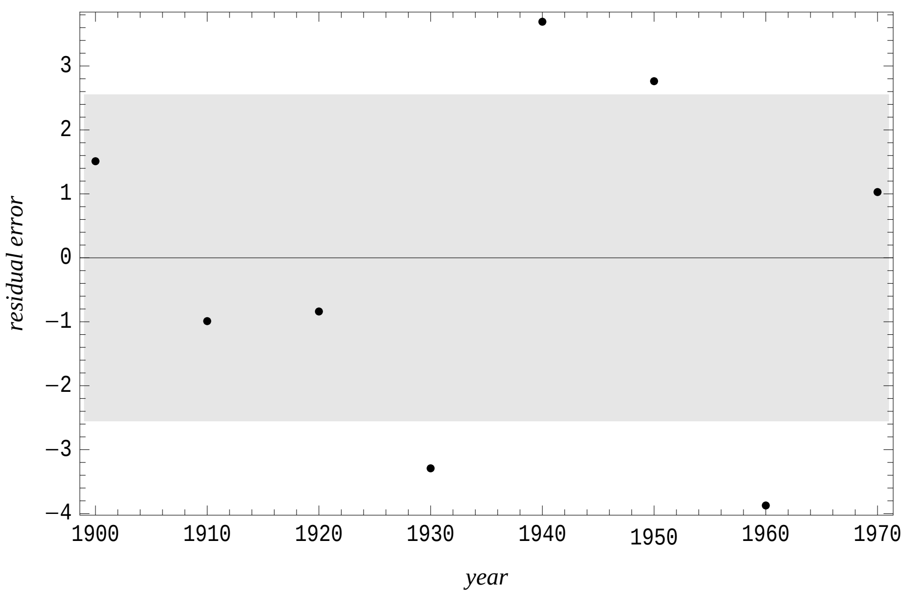


Figure 11.3. *Residual errors.*

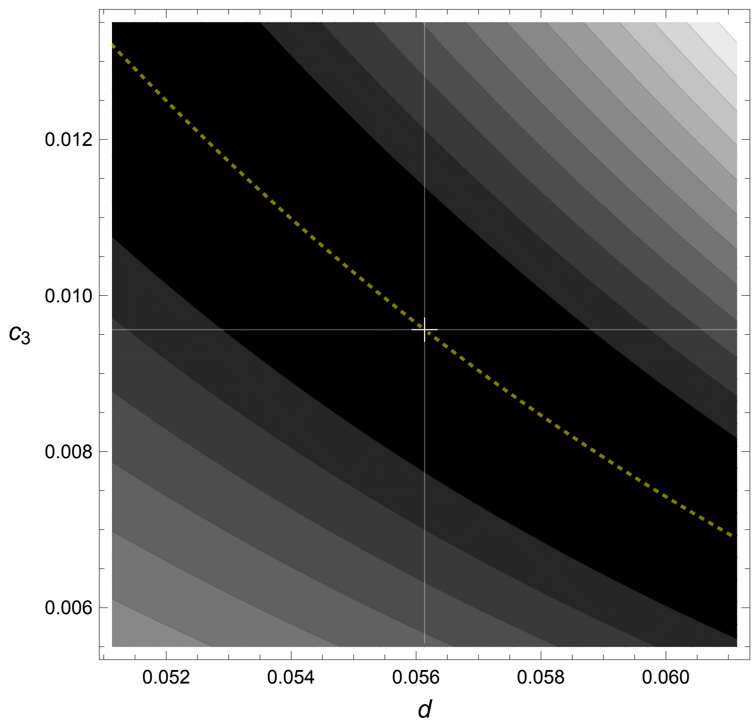


Figure 11.4. *The merit function with a_1 and a_2 fixed at best values showing least squares solution (center) and null cline (dashed, yellow).*

Part VII

Appendices

Appendix A

Least squares with exemplar matrices

A broad brush paints the primal elements in a portrait of the linear algebra pertinent to the practice of least squares.

A.1 Linear systems

Begin with the canonical linear system described by the matrix–vector equation

$$\mathbf{A}x = b. \tag{A.1}$$

The matrix $\mathbf{A}$ has m rows and n columns of complex numbers. (Recall the real number line $\mathbb{R}$ is part of the complex plane $\mathbb{C}$.) The matrix rank is $\rho \le \min(m, n)$. In shorthand, the three components are

1. $\mathbf{A} \in \mathbb{C}^{m\times n}_{\rho}$: the system matrix, an input;
2. $b \in \mathbb{C}^{m}$: the data vector, an input;
3. $x \in \mathbb{C}^{n}$: the solution vector, the output.

Given the matrix $\mathbf{A}$ and the data vector b, find the vector x which provides the best solution, in the least squares sense, to (A.1). This best solution minimizes the residual error given by

$$r = \mathbf{A}x - b. \tag{A.2}$$

In all instances, ignore the trivial cases where $b = 0$ which corresponds to the data vector lying within the null space $\mathcal{N}(\mathbf{A}^{*})$.

The general solution for (A.1) has the form

$$x_{LS} = x_{\dagger} + x_{\mathcal{N}}, \tag{A.3}$$

a set of $n-$vectors where the blue component inhabits $\mathcal{R}(\mathbf{A}^{*})$ and the red $\mathcal{N}(\mathbf{A})$. While it is true that

$$\mathbf{A}x_{\dagger} = \mathbf{A}\left(x_{\dagger} + x_{\mathcal{N}}\right),$$

the *solutions* $x_\dagger$ and $x_\dagger + x_\mathcal{N}$ are equivalent, it is also true that

$$\|x_\dagger\|_2 \geq \|x_\dagger + x_\mathcal{N}\|_2 \,, \tag{A.4}$$

the *norms* are different. The *solution of minimum norm* is $x_\dagger$. Hence a subtlety: equation (A.3) describes all leasts squares solutions (which have a common residual error vector). Amongst these solutions, there is one of minimum norm. As seen in (A.4), this is the pseudoinverse solution $x_\dagger$. While x_{LS} represents, in general, a set of solutions, $x_\dagger$ represents a special solution, a point in $\mathcal{R}(\mathbf{A}^*)$, the solution of least error norm.

Exemplar matrices have immediate singular value decompositions providing an $x-$ray image of the fundamental subspaces. The decompositions connect to the foundational concepts of solutions: existence and uniqueness. The exemplar set takes an identity matrix which is then extended to study null spaces.

Table A.1. *Exemplar matrices and their block forms.*

exemplar	block form
$\left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right]$	$\left[\begin{array}{c} \mathbf{I}_2 \end{array}\right]$
$\left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \\ \hline 0 & 0 \end{array}\right]$	$\left[\begin{array}{c} \mathbf{I}_2 \\ \hline \mathbf{0} \end{array}\right]$
$\left[\begin{array}{cc\|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right]$	$\left[\begin{array}{c\|c} \mathbf{I}_2 & \mathbf{0} \end{array}\right]$
$\left[\begin{array}{cc\|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 0 \end{array}\right]$	$\left[\begin{array}{c\|c} \mathbf{I}_2 & \mathbf{0} \\ \hline \mathbf{0} & 0 \end{array}\right]$

A.2 Exemplars

Essential concepts of least squares and the fundamental subspaces spring to life using exemplar matrices. Exemplar systems can be solved by inspection which invites introspection into the invariant subspaces.

A.2.1 Full rank: $\rho = m = n$

The simplest linear system is

$$\begin{array}{ccc} \mathbf{A} & x & = & b \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} & = & \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \end{array} \tag{A.5}$$

which has least squares solution

$$x_{LS} = x_{\dagger} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix},$$

which is an exact solution

$$r^{\mathrm{T}} r = 0.$$

Table A.2. *Subspace decomposition for the* $\boldsymbol{A}$ *matrix in equation* (A.5).

$$\begin{array}{rcccccc}
\text{domain:} & \mathbb{C}^2 & = & \mathcal{R}(\mathbf{A}^*) & \oplus & \mathcal{N}(\mathbf{A}) \\
& & = & \text{sp}\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\} & \oplus & \text{sp}\left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\} \\
\text{codomain:} & \mathbb{C}^2 & = & \mathcal{R}(\mathbf{A}) & \oplus & \mathcal{N}(\mathbf{A}^*) \\
& & = & \text{sp}\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\} & \oplus & \text{sp}\left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}
\end{array}$$

Table A.3. *Rank and invariant subspaces in equation* (A.5).

space	rank	range space			null space		
domain	$\rho = n = 2$	$\mathcal{R}(\mathbf{A}^*)$	$=$	$\text{sp}\ \{e_k^n\}_{k=1,n}$	$\mathcal{N}(\mathbf{A})$	$=$	$\{\mathbf{0}\}$
codomain	$\rho = m = 2$	$\mathcal{R}(\mathbf{A})$	$=$	$\text{sp}\ \{e_k^m\}_{k=1,m}$	$\mathcal{N}(\mathbf{A}^*)$	$=$	$\{\mathbf{0}\}$

Table A.4. *Existence and uniqueness for the full column rank linear system in equation* (A.5).

statement	subspace condition	data conditions
existence and uniqueness	$b \in \mathcal{R}(\mathbf{A})$	$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \neq \mathbf{0}$
no existence	$b \in \mathcal{N}(\mathbf{A}^*)$	$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \mathbf{0}$

A.2.2 Full column rank: $\rho = n < m$

Adding a row of zeros to the identity matrix induces a null space:

$$\begin{array}{ccccc} \mathbf{A} & x & = & b \\ \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \\ \hline 0 & 0 \end{array}\right] & \left[\begin{array}{c} x_1 \\ x_2 \end{array}\right] & = & \left[\begin{array}{c} b_1 \\ b_2 \\ \hline b_3 \end{array}\right]. \end{array} \tag{A.6}$$

The least squares solution is

$$x_{LS} = x_{\dagger} = \left[\begin{array}{c} b_1 \\ b_2 \end{array}\right]$$

which has error

$$r^{\mathrm{T}} r = |b_3| \, .$$

Table A.5. *Subspace decomposition for the* $\boldsymbol{A}$ *matrix in* (A.6).

$$\begin{array}{rcccccc} \text{domain:} & \mathbb{C}^2 & = & \mathcal{R}(\mathbf{A}^*) & \oplus & \mathcal{N}(\mathbf{A}) \\ & & = & \text{sp}\left\{ \left[\begin{array}{c} 1 \\ 0 \end{array}\right], \left[\begin{array}{c} 0 \\ 1 \end{array}\right] \right\} & \oplus & \text{sp}\left\{ \left[\begin{array}{c} 0 \\ 0 \end{array}\right] \right\} \\ \text{codomain:} & \mathbb{C}^3 & = & \mathcal{R}(\mathbf{A}) & \oplus & \mathcal{N}(\mathbf{A}^*) \\ & & = & \text{sp}\left\{ \left[\begin{array}{c} 1 \\ 0 \\ 0 \end{array}\right], \left[\begin{array}{c} 0 \\ 1 \\ 0 \end{array}\right] \right\} & \oplus & \text{sp}\left\{ \left[\begin{array}{c} 0 \\ 0 \\ 1 \end{array}\right] \right\} \end{array}$$

Table A.6. *Rank and invariant subspaces in equation* (A.5).

space	rank	range space			null space		
domain	$\rho = n = 2$	$\mathcal{R}(\mathbf{A}^*)$	$=$	$\text{sp}\,\{e_k^n\}_{k=1,n}$	$\mathcal{N}(\mathbf{A})$	$=$	$\{\mathbf{0}\}$
codomain	$\rho < m = 3$	$\mathcal{R}(\mathbf{A})$	$=$	$\text{sp}\,\{e_k^m\}_{k=1,\rho}$	$\mathcal{N}(\mathbf{A}^*)$	$=$	$\text{sp}\,\{e_k^m\}_{k=\rho+1,m}$

Conditions for existence and uniqueness are clear once the data vector is de-

Table A.7. *Existence and uniqueness for the full column rank linear system in equation* (A.6).

statement	subspace condition	data conditions
existence and uniqueness	$b \in \mathcal{R}(\mathbf{A})$	$(b_1 \neq 0 \text{ or } b_2 \neq 0)$ and $b_3 = 0$
existence	$b \in \mathcal{R}(\mathbf{A}) \oplus \mathcal{N}(\mathbf{A}^*)$	$(b_1 \neq 0 \text{ or } b_2 \neq 0)$ and $b_3 \neq 0$
no existence	$b \in \mathcal{N}(\mathbf{A}^*)$	$b_1 = b_2 = 0$, $b_3 \in \mathbb{C}$

composed:

$$b = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ 0 \end{bmatrix}}_{\in\mathcal{R}(\mathbf{A})} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ b_3 \end{bmatrix}}_{\in\mathcal{N}(\mathbf{A}^*)} \tag{A.7}$$

A.2.3 Full row rank: $\rho = m < n$

Adding a column of zeros to the identity matrix induces a different null space:

$$\begin{array}{ccccc} \mathbf{A} & & x & = & b \\ \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right] & & \left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array}\right] & = & \left[\begin{array}{c} b_1 \\ b_2 \end{array}\right]. \end{array} \tag{A.8}$$

The least squares solution is

$$x_{LS} = x_{\dagger} = \left[\begin{array}{c} b_1 \\ b_2 \\ 0 \end{array}\right] + \alpha \left[\begin{array}{c} 0 \\ 0 \\ 1 \end{array}\right], \qquad \alpha \in \mathbb{C}.$$

The residual error is

$$r^{\mathrm{T}} r = 0.$$

Table A.8. *Subspace decomposition for the* $\boldsymbol{A}$ *matrix in* (A.8).

$$\begin{array}{rcccccc} \text{domain:} & \mathbb{C}^3 & = & \mathcal{R}(\mathbf{A}^*) & \oplus & \mathcal{N}(\mathbf{A}) \\ & & = & \operatorname{sp}\left\{ \left[\begin{array}{c} 1 \\ 0 \\ 0 \end{array}\right], \left[\begin{array}{c} 0 \\ 1 \\ 0 \end{array}\right] \right\} & \oplus & \operatorname{sp}\left\{ \left[\begin{array}{c} 0 \\ 0 \\ 1 \end{array}\right] \right\} \\ \text{codomain:} & \mathbb{C}^2 & = & \mathcal{R}(\mathbf{A}) & \oplus & \mathcal{N}(\mathbf{A}^*) \\ & & = & \operatorname{sp}\left\{ \left[\begin{array}{c} 1 \\ 0 \end{array}\right], \left[\begin{array}{c} 0 \\ 1 \end{array}\right] \right\} & \oplus & \operatorname{sp}\left\{ \left[\begin{array}{c} 0 \\ 0 \end{array}\right] \right\} \end{array}$$

Thanks to the gentle behavior of the exemplar matrix, the range and null space components for the solution vector are apparent:

$$x = \underbrace{\left[\begin{array}{c} x_1 \\ x_2 \\ 0 \end{array}\right]}_{\in \mathcal{R}(\mathbf{A}^*)} + \underbrace{\left[\begin{array}{c} 0 \\ 0 \\ x_3 \end{array}\right]}_{\in \mathcal{N}(\mathbf{A})} \tag{A.9}$$

Existence and uniqueness: When the data vector component $b_3 = 0$,

$$b = \left[\begin{array}{c} b_1 \\ b_2 \\ 0 \end{array}\right] \in \mathcal{R}(\mathbf{A}) \tag{A.10}$$

Table A.9. *Existence and uniqueness for the full column rank linear system in equation* (A.8).

statement	subspace condition	data conditions
existence	$b \in \mathcal{R}(\mathbf{A})$	$b \neq \mathbf{0}$
no existence	$b \in \mathcal{N}(\mathbf{A}^*)$	$be\mathbf{0}$
uniqueness	no uniqueness because $\mathcal{R}(\mathbf{A})$ is non trivial	

the linear system is consistent and we have a unique solution

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \tag{A.11}$$

which is also the least squares solution

$$x_{LS} = x = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \tag{A.12}$$

with $r^{\mathrm{T}} r = 0$ residual error. Notice that the solution vector is in the complementary range space, the range space of $\mathbf{A}^*$:

$$x \in \mathcal{R}(\mathbf{A}^*). \tag{A.13}$$

A.2.4 Row and column rank deficit: $\rho < m, \rho < n$

Partitioning

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

$$
\left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 0 \end{array}\right] \left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array}\right] = \left[\begin{array}{c} b_1 \\ b_2 \\ b_3 \end{array}\right]. \tag{A.14}
$$

The least squares solution is

$$
x_{LS} = x_{\dagger} + x_{\mathcal{N}} = \left[\begin{array}{c} b_1 \\ b_2 \end{array}\right] + \alpha \left[\begin{array}{c} 0 \\ 0 \\ 1 \end{array}\right]
$$

which has error

$$
r^{\mathrm{T}} r = |b_3|\,.
$$

Singular Value Decomposition

$$
\left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 0 \end{array}\right] = \mathbf{U}\,\Sigma\,\mathbf{V}^{*} = \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right] \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \\ \hline 0 & 0 \end{array}\right] \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right] \tag{A.15}
$$

Subspace decomposition:

Table A.10. *Subspace decomposition for the* $\boldsymbol{A}$ *matrix in* (A.14).

$$
\begin{array}{rclcccc}
\text{domain:} & \mathbb{C}^3 & = & & \mathcal{R}(\mathbf{A}^*) & \oplus & \mathcal{N}(\mathbf{A}) \\
 & & = & \mathrm{sp} & \left\{ \left[\begin{array}{c} 1 \\ 0 \\ 0 \end{array}\right], \left[\begin{array}{c} 0 \\ 1 \\ 0 \end{array}\right] \right\} & \oplus & \mathrm{sp} \left\{ \left[\begin{array}{c} 0 \\ 0 \\ 1 \end{array}\right] \right\} \\
\text{codomain:} & \mathbb{C}^3 & = & & \mathcal{R}(\mathbf{A}) & \oplus & \mathcal{N}(\mathbf{A}^*) \\
 & & = & \mathrm{sp} & \left\{ \left[\begin{array}{c} 1 \\ 0 \\ 0 \end{array}\right], \left[\begin{array}{c} 0 \\ 1 \\ 0 \end{array}\right] \right\} & \oplus & \mathrm{sp} \left\{ \left[\begin{array}{c} 0 \\ 0 \\ 1 \end{array}\right] \right\}
\end{array}
$$

Thanks to the gentle behavior of the exemplar matrix, the range and null space components for the solution vector are apparent:

$$x = \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix}}_{\in \mathcal{R}(\mathbf{A}^*)} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ x_3 \end{bmatrix}}_{\in \mathcal{N}(\mathbf{A})} \tag{A.16}$$

Existence and uniqueness: When the data vector component $b_3 = 0$,

$$b = \begin{bmatrix} b_1 \\ b_2 \\ 0 \end{bmatrix} \in \mathcal{R}(\mathbf{A}) \tag{A.17}$$

the linear system is consistent and we have a unique solution

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \tag{A.18}$$

which is also the least squares solution

$$x_{LS} = x = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \tag{A.19}$$

with residual error $r^{\mathrm{T}} r = 0$. Notice that the solution vector is in the complementary range space, the range space of $\mathbf{A}^*$:

$$x \in \mathcal{R}(\mathbf{A}^*). \tag{A.20}$$

No existence When the data vector inhabits the null space

$$b \in \mathcal{N}(\mathbf{A}),$$

there is no least squares solution.

Existence, no uniqueness:

Table A.11. *Existence and uniqueness for the full column rank linear system in equation* (A.6).

statement	subspace condition	data conditions
existence and uniqueness	$b \in \mathcal{R}(\mathbf{A})$	$(b_1 \neq 0 \text{ or } b_2 \neq 0)$ and $b_3 = 0$
existence	$b \in \mathcal{R}(\mathbf{A}) \oplus \mathcal{N}(\mathbf{A}^*)$	$(b_1 \neq 0 \text{ or } b_2 \neq 0)$ and $b_3 \neq 0$
no existence	$b \in \mathcal{N}(\mathbf{A}^*)$	$b_1 = b_2 = 0,\ b_3 \in \mathbb{C}$

Appendix B

Error Propagation

B.1 Arithmetic Cases

$$\begin{aligned} y &= a_1 x_1 \pm a_2 x_2 \\ \epsilon_y^2 &= a_1^2 \epsilon_1^2 + a_2^2 \epsilon_2^2 \end{aligned} \tag{B.1}$$

$$\begin{aligned} y &= a x_1 x_2 \\ \epsilon_y^2 &= a^2 \left(x_1^2 \epsilon_2^2 + x_2^2 \epsilon_1^2 \right) \end{aligned} \tag{B.2}$$

$$\begin{aligned} y &= a \frac{x_1}{x_2} \\ \epsilon_y^2 &= a^2 \left(\frac{\epsilon_1^2}{x_2^2} + \frac{\epsilon_2^2}{x_1^2} \right) \end{aligned} \tag{B.3}$$

B.2 Powers and Exponential Cases

$$\begin{aligned} y &= a x^{\pm b} \\ \epsilon_y &= a b x^{\pm b - 1} \epsilon_x \end{aligned} \tag{B.4}$$

B.3 Example I: Polynomials

$$\begin{aligned} y(x) &= a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots \\ \epsilon_y^2 &= a_1^2 \epsilon_1^2 + a_1^2 \epsilon_1^2 \end{aligned}$$

$$\begin{aligned} y(x) &= a_0 + \sum_{k=1}^{d} a_k x^k \\ \epsilon_y &= 1 \end{aligned}$$

B.4 Example II: Quadratic Formula

$$y(x) = a_0 + a_1 x + a_2 x^2$$

Appendix C

Notation

A brief listing of notation.

Table C.1. *Matrices*

$\mathbf{A}^{\dagger}$	pseudoinverse of matrix $\mathbf{A}$
$\mathbf{A}^{*}$	Hermitian conjugate of matrix $\mathbf{A}$
$\mathbf{A}^{\mathrm{T}}$	transpose of matrix $\mathbf{A}$
$\mathbf{A}^{-\mathrm{L}}$	left inverse of matrix $\mathbf{A}$: $\mathbf{A}^{-\mathrm{L}}\mathbf{A} = \mathbf{I}_n, \quad \mathbf{A} \in \mathbb{C}_n^{m\times n}$
$\mathbf{A}^{-\mathrm{R}}$	right inverse of matrix $\mathbf{A}$: $\mathbf{A}\mathbf{A}^{-\mathrm{R}} = \mathbf{I}_m, \quad \mathbf{A} \in \mathbb{C}_m^{m\times n}$
$\mathbf{I}_k$	identity matrix of dimension $k \times k$
$\mathbb{I}_{j,k}$	stencil matrix, $j \le k$
$\mathbf{T}$	an upper triangular matrix

Table C.2. *Vectors*

a_k	kth column vector of matrix $\mathbf{A}$
$a_{[k]}$	kth row vector of matrix $\mathbf{A}$
e_k^j	unit vector of length j with 1 in the kth position
x_{LS}	least squares solution defined in (1.9)
$x_{\dagger}$	pseudoinverse solution defined in (2.2.2)

Table C.3. *Vector spaces*

$\mathcal{R}(\cdot)$	range space
$\mathcal{N}(\cdot)$	null space

Table C.4. *Fields*

$\mathbb{C}$	field of complex numbers
$\mathbb{R}$	field of real numbers
$\mathbb{Z}$	field of integers
$\mathbb{Z}^+$	field of positive integers
$\mathbb{N}$	field of natural numbers 0, 1, 2, ...

Table C.5. *Constants*

m	number of rows in a matrix
n	number of columns in a matrix
η_C	rank deficiency of the *column* space
η_R	rank deficiency of the *row* space
ρ	rank of a matrix

Table C.6. *Symbols*

$\oplus$	direct sum
$\otimes$	outer product
$\cdot$	dot product
$\Rightarrow\Leftarrow$	contradiction

Table C.7. *Abbreviations*

tr	matrix trace: sum of diagonal elements
set	matrix determinant
sp	span

Appendix D

Lexicon

Table D.1. *Row and column spaces.*

row space	column space
domain	codomain
preimage	image

1. pseudoinverse
2. Moore-Penrose pseudoinverse
3. generalized matrix inverse

Table D.2. *Matrix shapes.*

$m = n$	square	equal number of rows and columns
$m \geq n$	tall	more rows than columns
$n \geq m$	wide	more columns than rows

Table D.3. *Rank conditions.*

$\rho = m = n$	full rank	square	
$\rho = n \leq m$	full column rank	tall	overdetermined
$\rho = m \leq n$	full row rank	wide	underdetermined

Part VIII

Backmatter

Bibliography

[1] Richard Bellman, *Introduction to matrix analysis*, SIAM, Society for Industrial and Applied Mathematics; 2^{nd} edition (1997).

[2] Philip R. Bevington, *Data Reduction and Error Analysis in the Physical Sciences*, McGraw-Hill (1969).

[3] Raymond H. Chan, and Chen Greif, and Diane P. O'Leary, *Milestones in matrix computation: Selected works of Gene H. Golub, with commentaries*, Oxford University Press (2007).

[4] James W. Demmel, *Applied numerical linear algebra*, SIAM, Society for Industrial and Applied Mathematics (1997).

[5] Gene H. Golub, and Charles Van Loan, *Matrix Computations*, 3^{rd} Edition. Johns Hopkins University Press (1996).

[6] Nicholas J. Higham, *Functions of Matrices: Theory and Computation*, SIAM, Society for Industrial and Applied Mathematics (2008).

[7] Roger A. Horn, and Charles R. Johnson, *Matrix analysis*, Cambridge University Press (1990).

[8] Roger A. Horn, and Charles R. Johnson, *Topics in Matrix analysis*, 3^{rd} Edition. Cambridge University Press (1991).

[9] Idris C. Mercer *Finding nonobvious nilpotent matrices*, (2005) `http://www.idmercer.com/nilpotent.pdf`

[10] Alan J. Laub, *Matrix analysis for scientists and engineers*, SIAM, Society for Industrial and Applied Mathematics (2005).

[11] Carl D. Meyer, *Matrix analysis and applied linear algebra*, SIAM, Society for Industrial and Applied Mathematics (2000).

[12] Gilbert Strang, *Linear Algebra and Its Applications*, SIAM, Society for Industrial and Applied Mathematics (2005).

[13] Lloyd N. Trefethen, and David Bau, *Numerical linear algebra*, SIAM, Society for Industrial and Applied Mathematics (2000).

[14] Eric W. Weisstein, "Characteristic Polynomial", from MathWorld–A Wolfram Web Resource.
http://mathworld.wolfram.com/CharacteristicPolynomial.html