Excursions in linear least squares

Daniel Topa

E-mail address: dantopa@gmail.com

2010 *Mathematics Subject Classification.* Primary
Key words and phrases. least squares, L^2, l^2

ABSTRACT.

Contents

Part 1

Rudiments

CHAPTER 1

The least squares problem

This is an unnumbered first-level section head

This is an example of an unnumbered first-level heading.

1.1. Definition

Let us begin straightaway with a definition for the linear least squares problem. The linear system of interest is cast as

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

which has three elements:

(1) A *system matrix* $\mathbf{A} \in \mathbb{C}^{m\times n}$ which encodes measurement locations or times,
(2) A *data vector* $b \in \mathbb{C}^{m}$ which encodes the measurements,
(3) A *solution vector* $x \in \mathbb{C}^{n}$, the desired solution.

If the data vector b is in the range space of the matrix $\mathbf{A}$, that is if

$$b \in \mathcal{R}(\mathbf{A})$$

then there is a direct solution and no need to use the method of least squares. In the general case we must consider a data vector which "sticks into" the null space:

$$b \in \mathcal{R}(\mathbf{A}) \oplus \mathcal{N}(\mathbf{A}^{*}).$$

There is no direct solution to (1.1). We generalize the equation and instead of demanding the equality be satisfied, we instead ask for the nearest solution, depicted in figure (1.1). This is the orthogonal projection onto the range space $\mathcal{R}(\mathbf{A})$.

A more precise definition for the least squares solution x_{LS} is

$$x_{LS} = \min_{x\in\mathbb{C}^{n}} \lVert \mathbf{A}x - b \rVert_2^2 . \tag{1.2}$$

Hence the eponym "least squares" describes minimizing the square of the errors. This solution may be a point or hyperplane.

1.2. The Fundamental Theorem of Linear Algebra

There are many ways to state the Fundamental Theorem of Linear Algebra, the bedrock of linear algebra theory. Start with the canonical $\mathbf{A} \in \mathbb{C}^{m\times n}$

$$\begin{aligned} \mathbb{C}^{m} &= \mathcal{R}(\mathbf{A}) \oplus \mathcal{N}(\mathbf{A}^{*}) \qquad &\text{column space} \\ \mathbb{C}^{n} &= \mathcal{R}(\mathbf{A}^{*}) \oplus \mathcal{N}(\mathbf{A}) \qquad &\text{row space} \end{aligned} \tag{1.3}$$

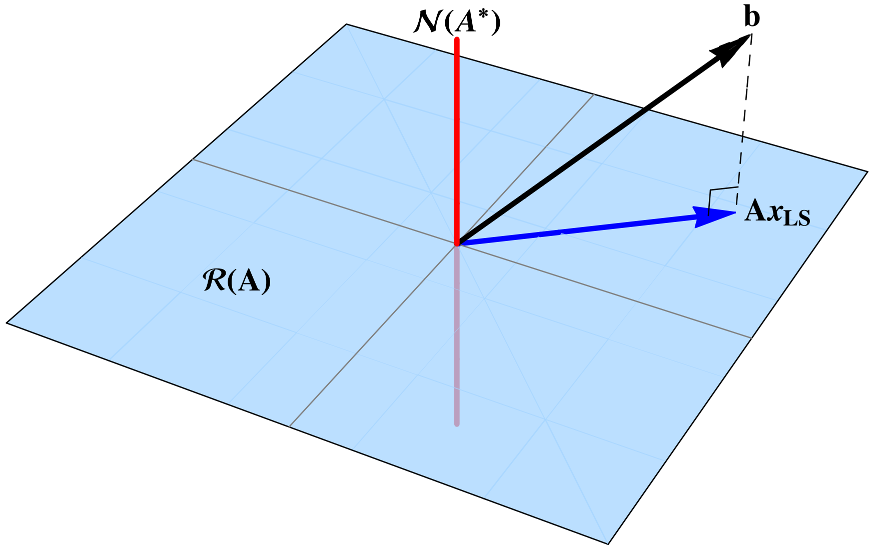


FIGURE 1.1. The least squares solution x_{LS} is an orthogonal projection of the data vector b onto the range space $\mathcal{R}(\mathbf{A})$. The subspace decomposition for the column space represented here is $\mathbb{C}^m = \mathcal{R}(\mathbf{A}) \oplus \mathcal{N}(\mathbf{A}^*)$.

$\mathbb{C}^m$	$=$	$\mathcal{R}(\mathbf{A})$	$\oplus$	$\mathcal{N}(\mathbf{A}^*)$	(column space)
$\mathbb{C}^n$	$=$	$\mathcal{R}(\mathbf{A}^*)$	$\oplus$	$\mathcal{N}(\mathbf{A})$	(row space)

TABLE 1.1. Subspace decomposition and the Fundamental Theorem of Linear Algebra.

$$\begin{aligned} \mathbf{A} &\colon \mathbb{C}^n \mapsto \mathbb{C}^m \\ \mathbf{A}^* &\colon \mathbb{C}^m \mapsto \mathbb{C}^n \end{aligned} \tag{1.4}$$

1.3. General solution to the least squares problem

$$x_{LS} = \mathbf{A}^\dagger b + \left(\mathbb{I}_n - \mathbf{A}^\dagger \mathbf{A}\right) y \tag{1.5}$$

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

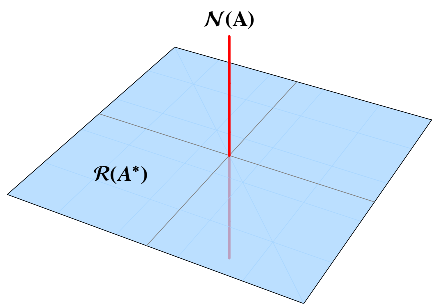
1.4. Singular Value Decomposition

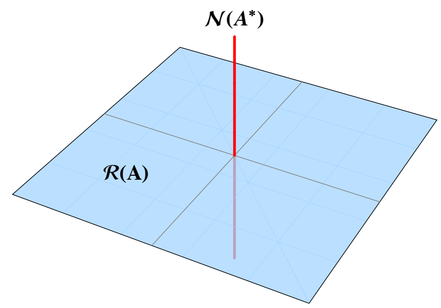
1.4.1. Decomposition.

1.4.2. Motivation: the pseudoinverse solution. Given a matrix $\mathbf{X} \in \mathbb{C}^m$ and a unitary matrix $\mathbf{Q} \in \mathbb{C}^m$ Unitary invariance under the $2-$norm and the Frobenius norm.

$$\|\mathbf{X}\|_2 = \|\mathbf{QX}\|_2 = \|\mathbf{XQ}\|_2 \tag{1.6}$$

Domain		Codomain
	$\mathbf{A} \in \mathbb{C}^{m \times n}$	
	$\mathbf{A}\,x = y$ $\longrightarrow$ $\mathbf{A}^* y = x$	





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

TABLE 1.2. The Fundamental Theorem of Linear Algebra.

rank deficiency	$\mathcal{N}(\mathbf{A}^*)$	$\mathcal{N}(\mathbf{A})$	$\mathbf{U}$	Σ	$\mathbf{V}^*$
none	$\{\mathbf{0}\}$	$\{\mathbf{0}\}$	$\left[\ \mathbf{U}_{\mathcal{R}}\ \right]$	$\left[\ \mathbf{S}\ \right]$	$\left[\ \mathbf{V}^*_{\mathcal{R}}\ \right]$
column	$\{\mathbf{0}\}$	nontrivial	$\left[\ \mathbf{U}_{\mathcal{R}}\ \right]$	$\left[\begin{array}{c\|c} \mathbf{S} & \mathbf{0} \end{array}\right]$	$\left[\begin{array}{c} \mathbf{V}^*_{\mathcal{R}} \\ \hline \mathbf{V}^*_{\mathcal{N}} \end{array}\right]$
row	nontrivial	$\{\mathbf{0}\}$	$\left[\begin{array}{c\|c} \mathbf{U}_{\mathcal{R}} & \mathbf{U}_{\mathcal{N}} \end{array}\right]$	$\left[\begin{array}{c} \mathbf{S} \\ \hline \mathbf{0} \end{array}\right]$	$\left[\ \mathbf{V}^*_{\mathcal{R}}\ \right]$
column & row	nontrivial	nontrivial	$\left[\begin{array}{c\|c} \mathbf{U}_{\mathcal{R}} & \mathbf{U}_{\mathcal{N}} \end{array}\right]$	$\left[\begin{array}{c\|c} \mathbf{S} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right]$	$\left[\begin{array}{c} \mathbf{V}^*_{\mathcal{R}} \\ \hline \mathbf{V}^*_{\mathcal{N}} \end{array}\right]$

TABLE 1.3. Decomposition block forms for $\mathbf{A} = \mathbf{U}\,\Sigma\,\mathbf{V}^*$.

Definition of the least square problem

$$x_{LS} = \left\{ x \in \mathbb{C}^n \colon \ \|r\|_2 \ \text{ is minimized} \right\} \tag{1.7}$$

$$\|r\|_2 = \|\mathbf{A}\,x - b\|_2 = \|\mathbf{U}\,\Sigma\,\mathbf{V}^* x - b\|_2 \tag{1.8}$$

Unitary transformation delivers a simpler problem

$$\|\mathbf{U}^* r\|_2 = \|r\|_2 \tag{1.9}$$

rank deficiency	$\mathcal{N}(\mathbf{A}^*)$	$\mathcal{N}(\mathbf{A})$	$\mathbf{U}$	Σ	$\mathbf{V}^*$
none	$\{\mathbf{0}\}$	$\{\mathbf{0}\}$	$\left[\ \mathbf{U}_{\mathcal{R}}\ \right]$	$\left[\ \mathbf{S}\ \right]$	$\left[\ \mathbf{V}^*_{\mathcal{R}}\ \right]$
column	$\{\mathbf{0}\}$	nontrivial	$\left[\ \mathbf{U}_{\mathcal{R}}\ \right]$	$\left[\ \mathbf{S} \mid \mathbf{0}\ \right]$	$\left[\begin{array}{c}\mathbf{V}^*_{\mathcal{R}}\\ \hline \mathbf{V}^*_{\mathcal{N}}\end{array}\right]$
row	nontrivial	$\{\mathbf{0}\}$	$\left[\ \mathbf{U}_{\mathcal{R}} \mid \mathbf{U}_{\mathcal{N}}\ \right]$	$\left[\begin{array}{c}\mathbf{S}\\ \hline \mathbf{0}\end{array}\right]$	$\left[\ \mathbf{V}^*_{\mathcal{R}}\ \right]$
column & row	nontrivial	nontrivial	$\left[\ \mathbf{U}_{\mathcal{R}} \mid \mathbf{U}_{\mathcal{N}}\ \right]$	$\left[\begin{array}{c\|c}\mathbf{S} & \mathbf{0}\\ \hline \mathbf{0} & \mathbf{0}\end{array}\right]$	$\left[\begin{array}{c}\mathbf{V}^*_{\mathcal{R}}\\ \hline \mathbf{V}^*_{\mathcal{N}}\end{array}\right]$

TABLE 1.4. Decomposition block forms for $\mathbf{A} = \mathbf{U}\,\Sigma\,\mathbf{V}^*$.

$$\lVert r \rVert_2 = \lVert \mathbf{U}^* r \rVert_2 = \lVert \Sigma\,\mathbf{V}^* x - \mathbf{U}^* b \rVert_2 \tag{1.10}$$

Now we can separate range and null spaces

$$\lVert \Sigma\,\mathbf{V}^* x - \mathbf{U}^* b \rVert_2^2 = \left\lVert \left[\begin{array}{c}\mathbf{S}\\ \hline \mathbf{0}\end{array}\right] \mathbf{V}^*_{\mathcal{R}} x - \left[\begin{array}{c}\mathbf{U}^*_{\mathcal{R}}\\ \hline \mathbf{U}^*_{\mathcal{N}}\end{array}\right] b \right\rVert_2^2 \tag{1.11}$$

The subspace components correspond to error terms which can and can't be controlled by varying the minimizer x:

$$\left\lVert \left[\begin{array}{c}\mathbf{S}\\ \hline \mathbf{0}\end{array}\right] \mathbf{V}^*_{\mathcal{R}} x - \left[\begin{array}{c}\mathbf{U}^*_{\mathcal{R}}\\ \hline \mathbf{U}^*_{\mathcal{N}}\end{array}\right] b \right\rVert_2^2 = \underbrace{\lVert \mathbf{S}\mathbf{V}^*_{\mathcal{R}} x - \mathbf{U}^*_{\mathcal{R}} b \rVert_2^2}_{\text{controlled}} + \underbrace{\lVert \mathbf{U}^*_{\mathcal{N}} b \rVert_2^2}_{\text{uncontrolled}} \tag{1.12}$$

The piece under control can be driven to zero by making the choice

$$x = \mathbf{V}_{\mathcal{R}} \mathbf{S}^{-1} \mathbf{U}^*_{\mathcal{R}}. \tag{1.13}$$

This is the pseudoinverse solution. The portion that cannot be controlled leaves a residual error

$$r^{\mathrm{T}} r = \lVert \mathbf{U}^*_{\mathcal{N}} b \rVert_2^2 \tag{1.14}$$

This term measures how much of the data vector intrudes into the null space $\mathcal{N}(\mathbf{A}^*)$

1.4.3. Σ gymnastics. 0

1.4.4. Identities.

CHAPTER 2

Least squares archetype

A good starting point is the foundation example of Bevington, Let's get some traction.

$$
\begin{aligned}
\mathbf{A} &= \mathbf{U}\,\Sigma\,\mathbf{V}^{*} \\
&= \left[\ \color{blue}{\mathcal{R}(\mathbf{A})}\ \middle|\ \color{red}{\mathcal{N}(\mathbf{A}^{*})}\ \right] \left[\begin{array}{c|c} \mathbf{S} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \left[\ \color{blue}{\mathcal{R}(\mathbf{A}^{*})}\ \middle|\ \color{red}{\mathcal{N}(\mathbf{A})}\ \right]^{*}
\end{aligned}
$$

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

$$
\mathbf{A} \in \mathbb{C}^{m\times n}_{\rho}
$$

$$
\mathbf{S} = \operatorname{diag}\left(\sigma_1, \sigma_2, \sigma_\rho\right)
$$

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

TABLE 2.1. default

$\mathbf{A}$	$=$	$\mathbf{U}$	Σ	$\mathbf{V}^{*}$
		$\left[\ \color{blue}{\mathcal{R}(\mathbf{A})}\ \middle\|\ \color{red}{\mathcal{N}(\mathbf{A}^{*})}\ \right]$	$\left[\begin{array}{c\|c} \mathbf{S} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right]$	$\left[\ \color{blue}{\mathcal{R}(\mathbf{A}^{*})}\ \middle\|\ \color{red}{\mathcal{N}(\mathbf{A})}\ \right]^{*}$

$\mathcal{SAMPLE}$

2.1. Bevington's example

Given a set of measurements which constant best typifies the data?

2.1.1. Trial function. His *trial function* is a straight line

$$
y(x) = a_0 + a_1 x \tag{2.1}
$$

where a_0 represents the intercept and a_1 the slope. The *residual error* vector describes the difference between the measurements and the approximation:

$$
r_k = \text{measurement}_k - \text{prediction}_k = y_k - y\left(x_k\right) \qquad k = 1:m \tag{2.2}
$$

We will see a delightful representation of this vector in chapter ???. The least squares solution is find the set of parameters a which minimize the sum of the squares of the residual errors:

$$
a_{LS} = \left\{ a \in \mathbb{C}^2 \colon r^{\mathrm{T}} r \text{ is minimized} \right\}. \tag{2.3}
$$

2.1.2. Merit function. This leads to the concept of the *merit function*

$$
\begin{aligned}
M(a) &= \sum_{k=1}^{m} \left(\text{measurement}_k - \text{prediction}_k\right)^2 \\
&= \sum_{k=1}^{m} \left(y_k - y\left(x_k\right)\right)^2 \\
&= \sum_{k=1}^{m} \left(y_k - a_0 - a_1 x_k\right)^2
\end{aligned}
\tag{2.4}
$$

The merit function quantifies the quality of the approximation: smaller values are better. For an exact solution $M(a) = 0$. We will gain valuable insight into the method of least squares by studying plots of the *solution space*.

The merit function can be minimized with basic calculus by finding where the gradient function ∇M vanishes. That is look for a solution where the derivatives in the a_0 direction and the a_1 direction are simultaneously 0.[1] Introducing the shorthand

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

the simultaneous equations to solve are

$$
\begin{aligned}
\partial_0 M &= 0, \\
\partial_1 M &= 0,
\end{aligned}
\tag{2.5}
$$

which become

$$
\begin{aligned}
-2 \sum_k \left(y_k - a_0 - a_1 x_k\right) &= 0, \\
-2 \sum_k \left(y_k - a_0 - a_1 x_k\right) a_1 &= 0.
\end{aligned}
\tag{2.6}
$$

2.1.3. Linear system. This generates a the linear system with $m = 9$ measurements and $n = 2$ free parameters. Introducing a shorthand for summation, for example

$$
\sum x_k = \sum_{k=1}^{m} x_k,
\tag{2.7}
$$

the linear system can be written as[2]

$$
\begin{aligned}
\mathcal{A} a &= b, \\
\left[\begin{array}{cc} \sum 1 & \sum x_k \\ \sum x_k & \sum x_k^2 \end{array}\right] \left[\begin{array}{c} a_0 \\ a_1 \end{array}\right] &= \left[\begin{array}{c} \sum y_k \\ \sum x_k y_k \end{array}\right]
\end{aligned}
\tag{2.8}
$$

The determinant of the matrix $\mathcal{A}$ is

$$
\Delta = \det\left(\mathcal{A}\right) = m \sum x_k^2 - \left(\sum x_k\right)^2.
\tag{2.9}
$$

[1] Will these extrema be maxima or minima? In chapter ??? we will learn they must be minima.

[2] A common stumbling block is the realization that $\sum_{k=1}^{m} 1 = \underbrace{1 + 1 + \cdots + 1}_{m \text{ instances}} = m$.

k	$x_k(cm)$	$y_k(°C)$	$y(x_k)\,(°C)$	$r_k.(°C)$
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

TABLE 2.2. Bevington's raw data

The solution to equation (2.8) is the matrix product

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \Delta^{-1} \begin{bmatrix} \sum x_k^2 & -\sum x_k \\ -\sum x_k & \sum 1 \end{bmatrix} \begin{bmatrix} \sum y_k \\ \sum x_k y_k \end{bmatrix} \tag{2.10}$$

Compare the final results to Bevington's equations 6-19:

$$\begin{aligned} a_0 &= \Delta^{-1}\left(\sum x_k^2 \sum y_k - \sum x_k \sum x_k y_k\right), \\ a_1 &= \Delta^{-1}\left(m \sum x_k y_k - \sum x_k \sum y_k\right). \end{aligned} \tag{2.11}$$

2.1.4. Error propagation. Bevington's greatest contribution may be his masterful explanation of error propagation.

$$s^2 \approx \frac{r^{\mathrm{T}} r}{m-n} \tag{2.12}$$

$$\begin{aligned} \epsilon_0^2 &= \frac{r^{\mathrm{T}} r}{\Delta\,(m-n)} \sum x_k^2 \\ \epsilon_1^2 &= \frac{r^{\mathrm{T}} r}{\Delta\,(m-n)} \sum 1 \end{aligned} \tag{2.13}$$

2.1.5. Data and results. The raw data is presented in table (2.2), and the results in equation (2.14).

$$\begin{aligned} a_0 &= 4.8 \pm 4.9 && \text{(intercept)} \\ a_1 &= 9.41 \pm 0.87 && \text{(slope)} \end{aligned} \tag{2.14}$$

An alternative presentation is

$$\begin{aligned} a_0 &= 4.8\,(4.9) && \text{(intercept)} \\ a_1 &= 9.41\,(0.87) && \text{(slope)} \end{aligned} \tag{2.15}$$

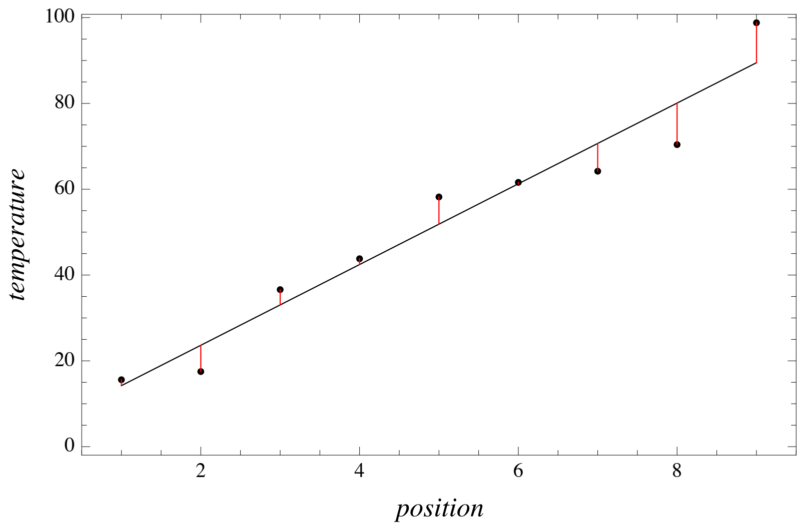


FIGURE 2.1. The solution curve (2.11) plotted against the data in table (**??**). The red lines represent the residual error, the difference between measurement and prediction.

The residual error vector is

$$r = \frac{1}{360}\begin{bmatrix} -496 \\ 2207 \\ -1282 \\ -487 \\ -2284 \\ -121 \\ 2330 \\ 3485 \\ -3352 \end{bmatrix} \tag{2.16}$$

making the total error

$$r^{\mathrm{T}}r = 317. \tag{2.17}$$

2.1.6. Visualization. The trial function is plotted against the data in figure (2.1). One sees the straight line approximation against the data points. The distance between measurement and prediction is the residual error.

Plotting solution against the data is a good first step. Stopping at this first step is a bad practice.The next important step is to examine the residuals in a separate plot. Typically the prediction and the residual error have different magnitudes. In this example, the measurements are contained in the interval $[0, 100]$, yet the errors are contained in $[-10, 10]$.

An elementary way to explore the errors in equations (2.13) is to randomly sample the solution space. The following figure shows 250 solution curves from normally distributed random samples, the point being to compare the spread in the data against the spread in the solutions. Given a mean μ and a standard

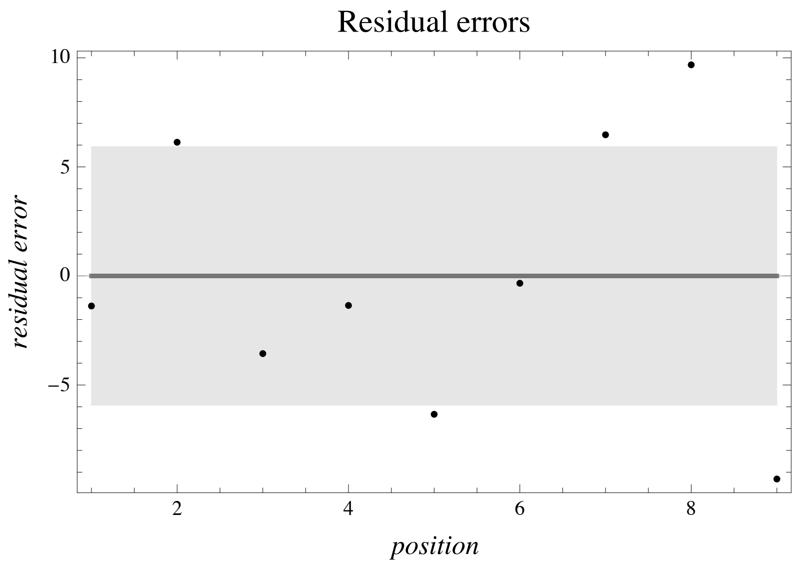


FIGURE 2.2. A closer look at the residual errors. The errors shown in figure (2.1) are plotted on an absolute scale. Ideally these errors should be randomly distributed, but we see signals of correlation suggesting the linear model may be inadequate.

deviation σ, the normal distribution takes the form

$$f(x) = e^{\frac{(x-\mu)^2}{2\sigma^2}}. \tag{2.18}$$

An example is shown in figure (2.5). For the intercept term, $\mu = 4.8$ and $\sigma = 4.0$; for the slope term $\mu = 4.8$ and $\sigma = 4.0$.

We have been sloppy by dividing by a quantity which may be 0.

2.2. Geometry of least squares

Let us repose the problem in a more fundamental way.

$$\begin{aligned} a_0 + a_1 x_1 &= y_1 \\ a_0 + a_1 x_2 &= y_2 \\ a_0 + a_1 x_3 &= y_3 \\ a_0 + a_1 x_4 &= y_4 \\ a_0 + a_1 x_5 &= y_5 \\ a_0 + a_1 x_6 &= y_6 \\ a_0 + a_1 x_7 &= y_7 \\ a_0 + a_1 x_8 &= y_8 \\ a_0 + a_1 x_9 &= y_9 \end{aligned} \tag{2.19}$$

Notice that we can not solve this problem one point at a time. That is given any point (x_k, y_k) we cannot find a solution set a_k - there is not enough information. We would have to consider pairs of points to find a solution set. Yet by looking at

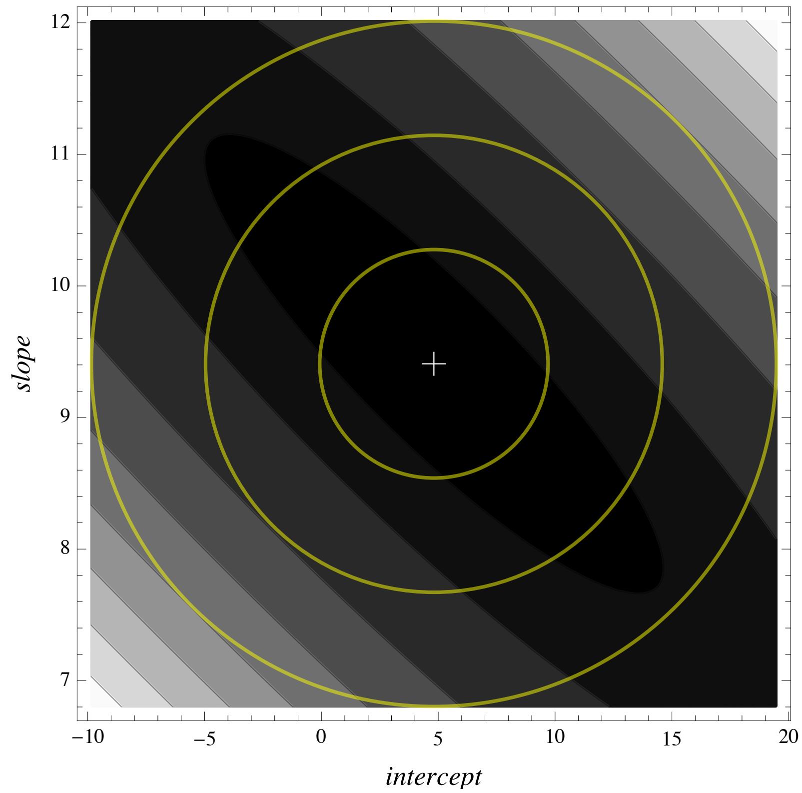


FIGURE 2.3. Looking at the merit function in solution space. The white cross in the center marks the solution found in (2.11). The yellow curves represent the error ellipses with radii of (a_0, a_1), $2\,(a_0, a_1)$, and $3\,(a_0, a_1)$. This provides a visual representation of how stable the solution is against perturbations in the data.

the plot of the data, each pair of points will produce a distinctly different solution.

$$\mathbf{A}a = Y$$

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ 1 & x_5 \\ 1 & x_6 \\ 1 & x_7 \\ 1 & x_8 \\ 1 & x_9 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{bmatrix} \tag{2.20}$$

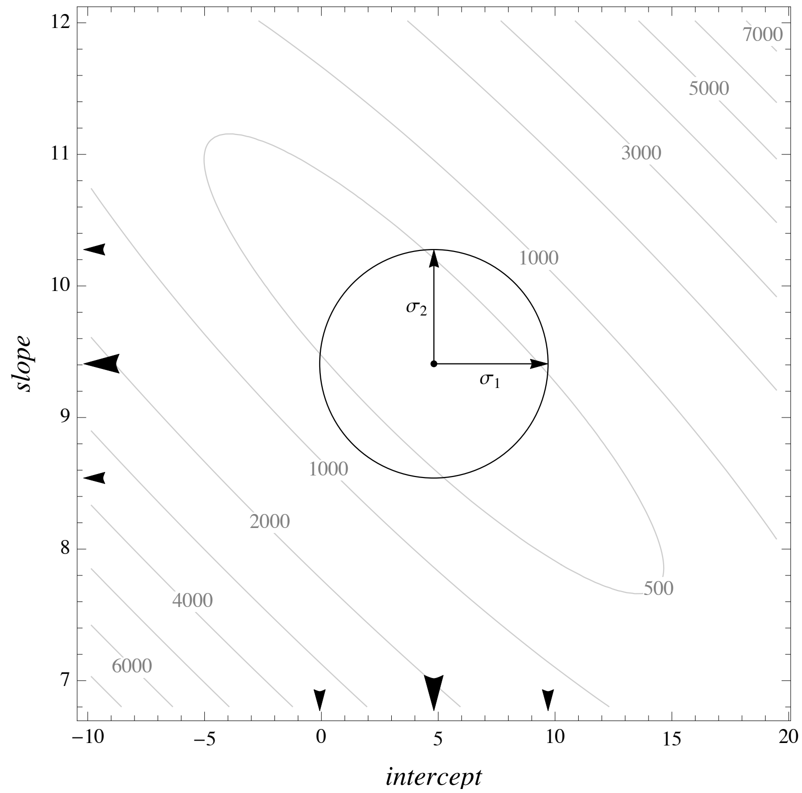


FIGURE 2.4. The solution and the error ellipse. Another look at figure (2.3), this time with the contours labelled. Also, arrowheads are used on the axis to mark the value of the solution and the value of the error in that parameter. The value of the merit function at the minimum is $M(a_0, a_1) \approx 317$.

From this viewpoint the essential vectors are evident:

$$1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{bmatrix} \tag{2.21}$$

The matrix $\mathbf{A} \in \mathbb{C}_{\rho}^{m\times n}$ has $m = 9$ rows, $n = 2$ columns, and has full column rank $\rho = 2$. The parameter m counts the number of measurements, the parameter n counts the number of solutions variables, here slope and intercept.

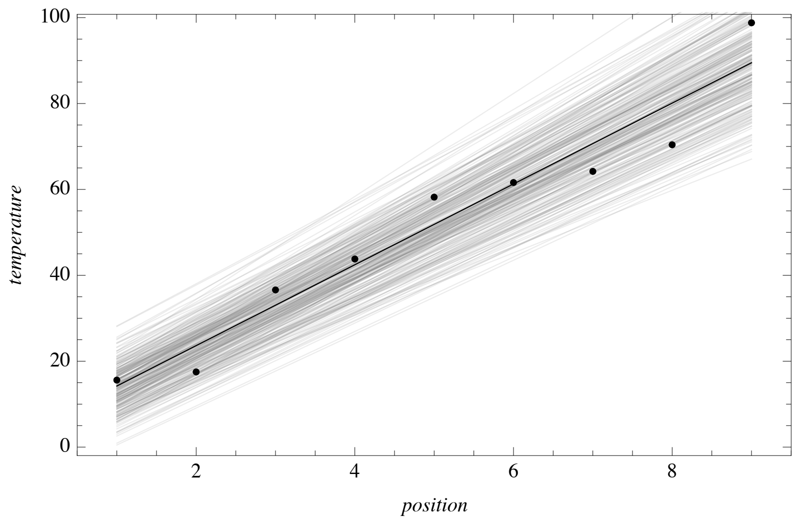


FIGURE 2.5. A look at 250 random solutions, distributed normally. The mean and variance for the distributions are given by the solution vector in (2.14).

2.3. Singular value decomposition

2.3.1. Singular value decomposition. 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

$$\mathbf{A} = \mathbf{U}\,\Sigma\,\mathbf{V}^{\cdot} \tag{2.22}$$

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

$$\mathbf{A} = \left[\begin{array}{c|c} \mathbf{U}_{\mathcal{R}} & \mathbf{U}_{\mathcal{N}} \end{array}\right] \left[\begin{array}{c} \mathbf{S} \\ \hline \mathbf{0} \end{array}\right] \left[\begin{array}{c} \mathbf{V}^{*}_{\mathcal{R}} \end{array}\right] \tag{2.23}$$

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

$$\mathbf{S} = \left[\begin{array}{cc} \sigma_1 & 0 \\ 0 & \sigma 2 \end{array}\right]. \tag{2.24}$$

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).

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

$$\sigma\left(\mathbf{A}\right) = \sqrt{\lambda\left(\mathbf{A}^{*}\mathbf{A}\right)}. \tag{2.25}$$

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

$$p(\lambda) = \lambda^2 - \lambda\,\mathrm{tr}\left(\mathbf{A}^{*}\mathbf{A}\right) + \det\left(\mathbf{A}^{*}\mathbf{A}\right) \tag{2.26}$$

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

$$\sigma = \sqrt{\frac{1}{2}\left(\mathbf{1}^{\mathrm{T}}\mathbf{1} + x^{\mathrm{T}}x \pm \sqrt{4\left(\mathbf{1}^{\mathrm{T}}x\right)^2 - \left(\mathbf{1}^{\mathrm{T}}\mathbf{1} - x^{\mathrm{T}}x\right)^2}\right)}. \tag{2.27}$$

(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

$$\sigma = \sqrt{3\left(49 \pm \sqrt{2341}\right)} \approx (17.0924, 1.35954). \tag{2.28}$$

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

$$\Sigma = \left[\begin{array}{c} \mathbf{S} \\ \hline \mathbf{0} \end{array}\right] = \left[\begin{array}{cc} \sigma_1 & 0 \\ 0 & \sigma_2 \\ \hline 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{array}\right] \approx \left[\begin{array}{cc} 17.0924 & 0 \\ 0 & 1.35954 \\ \hline 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{array}\right]. \tag{2.29}$$

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

$$\Sigma^{\mathrm{T}}\Sigma = \left[\begin{array}{cc} \mathbf{S} & \mathbf{0} \end{array}\right] \left[\begin{array}{c} \mathbf{S} \\ \mathbf{0} \end{array}\right] = \mathbf{S}^2 \tag{2.30}$$

$$\mathbf{V}_{\mathcal{R}}\mathbf{S}^2\mathbf{V}_{\mathcal{R}}^* = \mathbf{A}^*\mathbf{A}. \tag{2.31}$$

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

$$\mathbf{V}_{\mathcal{R}} = \left[\begin{array}{cc} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{array}\right] \tag{2.32}$$

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 (2.53), (2.31), and (2.32) is

$$\begin{gathered} \left[\begin{array}{cc} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{array}\right] \mathbf{S}^2 \left[\begin{array}{cc} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{array}\right]^{\mathrm{T}} = \mathbf{A}^*\mathbf{A}, \\ \left[\begin{array}{cc} \sigma_1^2\cos^2\theta + \sigma_2^2\sin^2\theta & \left(\sigma_1^2 - \sigma_2^2\right)\cos\theta\sin\theta \\ \left(\sigma_1^2 - \sigma_2^2\right)\cos\theta\sin\theta & \sigma_2^2\cos^2\theta + \sigma_1^2\sin^2\theta \end{array}\right] = \left[\begin{array}{cc} \mathbf{1}^{\mathrm{T}}\mathbf{1} & \mathbf{1}^{\mathrm{T}}x \\ x^{\mathrm{T}}\mathbf{1} & x^{\mathrm{T}}x \end{array}\right]. \end{gathered} \tag{2.33}$$

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

$$\cos\theta = \sqrt{\frac{\sigma_2^2 - \mathbf{1}^{\mathrm{T}}\mathbf{1}}{\sigma_2^2 - \sigma_1^2}} = \sqrt{\frac{x^{\mathrm{T}}x - \sigma_1^2}{\sigma_2^2 - \sigma_1^2}}. \tag{2.34}$$

This implies

$$\sin\theta = \tag{2.35}$$

The domain matrix is now

$$\mathbf{V} = \left[\ \mathbf{V}_{\mathcal{R}}\ \right] = \left(\sigma_2^2 - \sigma_1^2\right)^{-\frac{1}{2}} \left[\begin{array}{cc} \sqrt{\sigma_2^2 - \mathbf{1}^{\mathrm{T}}\mathbf{1}} & -\sqrt{\mathbf{1}^{\mathrm{T}}\mathbf{1} - \sigma_1^2} \\ \sqrt{\mathbf{1}^{\mathrm{T}}\mathbf{1} - \sigma_1^2} & \sqrt{\sigma_2^2 - \mathbf{1}^{\mathrm{T}}\mathbf{1}} \end{array} \right]. \tag{2.36}$$

Using the data set at hand

$$\left[\ \mathbf{V}_{\mathcal{R}}\ \right] = \left[\begin{array}{cc} \sqrt{\frac{1}{2} - \frac{23}{\sqrt{2341}}} & -\sqrt{\frac{1}{2} + \frac{23}{\sqrt{2341}}} \\ \sqrt{\frac{1}{2} + \frac{23}{\sqrt{2341}}} & \sqrt{\frac{1}{2} - \frac{23}{\sqrt{2341}}} \end{array} \right] \approx \left[\begin{array}{cc} 0.156956 & -0.987606 \\ 0.987606 & 0.156956 \end{array} \right] \tag{2.37}$$

2.3.1.3. *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

$$\mathbf{U}_{\mathcal{R}}^* = \mathbf{S}^{-1}\mathbf{V}_{\mathcal{R}}^*\mathbf{A}^*. \tag{2.38}$$

The kth column vector of this matrix has the compact form

$$\left[\mathbf{U}_{\mathcal{R}}^*\right]_k = \left(\sigma_1^2 - \sigma_2^2\right)^{-\frac{1}{2}} \left[\begin{array}{c} \sigma_1^{-2}\left(\sqrt{\sigma_2^2 - \mathbf{1}^{\mathrm{T}}\mathbf{1}} - x\sqrt{\mathbf{1}^{\mathrm{T}}\mathbf{1} - \sigma_1^2}\right) \\ \sigma_2^{-2}\left(\sqrt{\sigma_2^2 - \mathbf{1}^{\mathrm{T}}\mathbf{1}} + x\sqrt{\mathbf{1}^{\mathrm{T}}\mathbf{1} - \sigma_1^2}\right) \end{array} \right]. \tag{2.39}$$

$$\mathbf{U}_{\mathcal{R}} = \left(6\sqrt{10}\right)^{-1} \left[\begin{array}{cc} \sqrt{68 - 3212/\sqrt{2341}} & -\sqrt{68 + 3212/\sqrt{2341}} \\ \sqrt{47 - 2003/\sqrt{2341}} & -\sqrt{47 + 2003/\sqrt{2341}} \\ \sqrt{32 - 968/\sqrt{2341}} & -\sqrt{32 + 968/\sqrt{2341}} \\ \sqrt{23 - 107/\sqrt{2341}} & -\sqrt{23 + 107/\sqrt{2341}} \\ \sqrt{50 + 580/\sqrt{2341}} & -\sqrt{50 - 580/\sqrt{2341}} \\ \sqrt{23 + 1093/\sqrt{2341}} & -\sqrt{23 - 1093/\sqrt{2341}} \\ \sqrt{32 + 1432/\sqrt{2341}} & \sqrt{32 - 1432/\sqrt{2341}} \\ \sqrt{47 + 1597/\sqrt{2341}} & \sqrt{47 - 1597/\sqrt{2341}} \\ \sqrt{68 + 1588/\sqrt{2341}} & \sqrt{68 - 1588/\sqrt{2341}} \end{array} \right] \tag{2.40}$$

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

$$\mathbf{U} = \left[\ [\mathbf{U}_{\mathcal{R}}]_1 \quad [\mathbf{U}_{\mathcal{R}}]_2 \quad e_{1,9} \quad e_{2,9} \quad e_{3,9} \quad e_{4,9} \quad e_{5,9} \quad e_{6,9} \quad e_{7,9}\ \right] \tag{2.41}$$

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.

2.3.1.4. *Error terms.*

$$\epsilon_k^2 = \frac{\left(\mathbf{A}\alpha - y\right)^{\mathrm{T}}\left(\mathbf{A}\alpha - y\right)}{m - n}\left[\left(\mathbf{A}^*\mathbf{A}\right)^{-1}\right]_{kk} \tag{2.42}$$

The error terms can be computed after this step. Given that the product matrix decomposition in (2.31) is a singular value decomposition we can trivially write the inverse matrix as

$$(\mathbf{A}^*\mathbf{A})^{-1} = \mathbf{V}_{\mathcal{R}}\mathbf{S}^{-2}\mathbf{V}_{\mathcal{R}}^* = \frac{1}{180}\left[\begin{array}{rr} 95 & -15 \\ -15 & 3 \end{array}\right] \tag{2.43}$$

The error terms in (2.58) become

$$\begin{aligned} \epsilon &= \sqrt{\frac{r^{\mathrm{T}}r}{\mathbf{1}^{\mathrm{T}}\mathbf{1}-n}}\sqrt{\frac{1}{\sigma_1\sigma_2}}\sqrt{\left[\begin{array}{c} \frac{\cos^2\theta}{\sigma_2^2}+\frac{\sin^2\theta}{\sigma_1^2} \\ \frac{\cos^2\theta}{\sigma_1^2}+\frac{\sin^2\theta}{\sigma_2^2} \end{array}\right]}, \\ &= \sqrt{\frac{r^{\mathrm{T}}r}{m-n}}\sqrt{\left[\begin{array}{c} \sigma_1^2-\sqrt{(\sigma_2^2-\mathbf{1}^{\mathrm{T}}\mathbf{1})(\sigma_2^2-\sigma_1^2)} \\ \sigma_2^2+\sqrt{(\sigma_2^2-\mathbf{1}^{\mathrm{T}}\mathbf{1})(\sigma_2^2-\sigma_1^2)} \end{array}\right]}. \end{aligned} \tag{2.44}$$

(2.45)

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

2.3.1.5. *Visualization.* With the singular value decomposition in hand, we can the domain space plots more concrete and we do so below beginning in figure (2.6). The black vector represents the measurements

$$\mathbf{Y} = \mathbf{A}\,a - \mathbf{R} \tag{2.46}$$

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

$$y = \frac{1}{100}\left[\begin{array}{c} 156 \\ 175 \\ 366 \\ 438 \\ 582 \\ 616 \\ 642 \\ 704 \\ 988 \end{array}\right] \tag{2.47}$$

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

$$\mathbf{A}\,a = \frac{1}{360}\left[\begin{array}{r} 5120 \\ 8507 \\ 11\,894 \\ 15\,281 \\ 18\,668 \\ 22\,055 \\ 25\,442 \\ 28\,829 \\ 32\,216 \end{array}\right] = \alpha_1[\mathbf{U}_{\mathcal{R}}]_1 + \alpha_2[\mathbf{U}_{\mathcal{R}}]_2 \ \in \mathcal{R}(\mathbf{A}) \tag{2.48}$$

where the coordinates are

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

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

(2.49)
$$-r = \frac{1}{360}\begin{bmatrix} 496 \\ -2207 \\ 1282 \\ 487 \\ 2284 \\ 121 \\ -2330 \\ -3485 \\ 3352 \end{bmatrix}$$
$$= \alpha_3[\mathbf{U}_{\mathcal{N}}]_1 + \alpha_4[\mathbf{U}_{\mathcal{N}}]_2 + \alpha_5[\mathbf{U}_{\mathcal{N}}]_3 + \alpha_6[\mathbf{U}_{\mathcal{N}}]_4 + \alpha_7[\mathbf{U}_{\mathcal{N}}]_5 + \alpha_8[\mathbf{U}_{\mathcal{N}}]_6 + \alpha_9[\mathbf{U}_{\mathcal{N}}]_7 \ \in \mathcal{N}(\mathbf{A}^*)$$

The coordinates are now

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

$$y = \mathbf{U}\alpha = \mathbf{A}\,a - r \tag{2.50}$$

$$M(\alpha) = \|y - \alpha_1 u_1 - \alpha_2 u_2\|_2 \tag{2.51}$$

The minimizer is given by (2.3.1.5).

$$M(171.733, -4.45594) \approx 17.7949. \tag{2.52}$$

2.4. Normal equations

$$\mathbf{A}^*\mathbf{A} = \begin{bmatrix} \mathbf{1}^{\mathrm{T}}\mathbf{1} & \mathbf{1}^{\mathrm{T}}x \\ x^{\mathrm{T}}\mathbf{1} & x^{\mathrm{T}}x \end{bmatrix} = \begin{bmatrix} m & \sum_{k=1}^m x_k \\ \sum_{k=1}^m x_k & \sum_{k=1}^m x_k^2 \end{bmatrix}. \tag{2.53}$$

This is enough to complete the solution (compare to (2.10)):

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \left(\left(\mathbf{1}^{\mathrm{T}}\mathbf{1}\right)\left(x^{\mathrm{T}}x\right) - \left(\mathbf{1}^{\mathrm{T}}x\right)^2\right)^{-1}\begin{bmatrix} x^{\mathrm{T}}x & -\mathbf{1}^{\mathrm{T}}x \\ -\mathbf{1}^{\mathrm{T}}x & \mathbf{1}^{\mathrm{T}}\mathbf{1} \end{bmatrix}\begin{bmatrix} \mathbf{1}^{\mathrm{T}}y \\ x^{\mathrm{T}}y \end{bmatrix}. \tag{2.54}$$

The determinant of the product matrix

$$\det\left(\mathbf{A}^*\mathbf{A}\right) = \Delta = \left(\mathbf{1}^{\mathrm{T}}\mathbf{1}\right)\left(x^{\mathrm{T}}x\right) - \left(\mathbf{1}^{\mathrm{T}}x\right)^2 \tag{2.55}$$

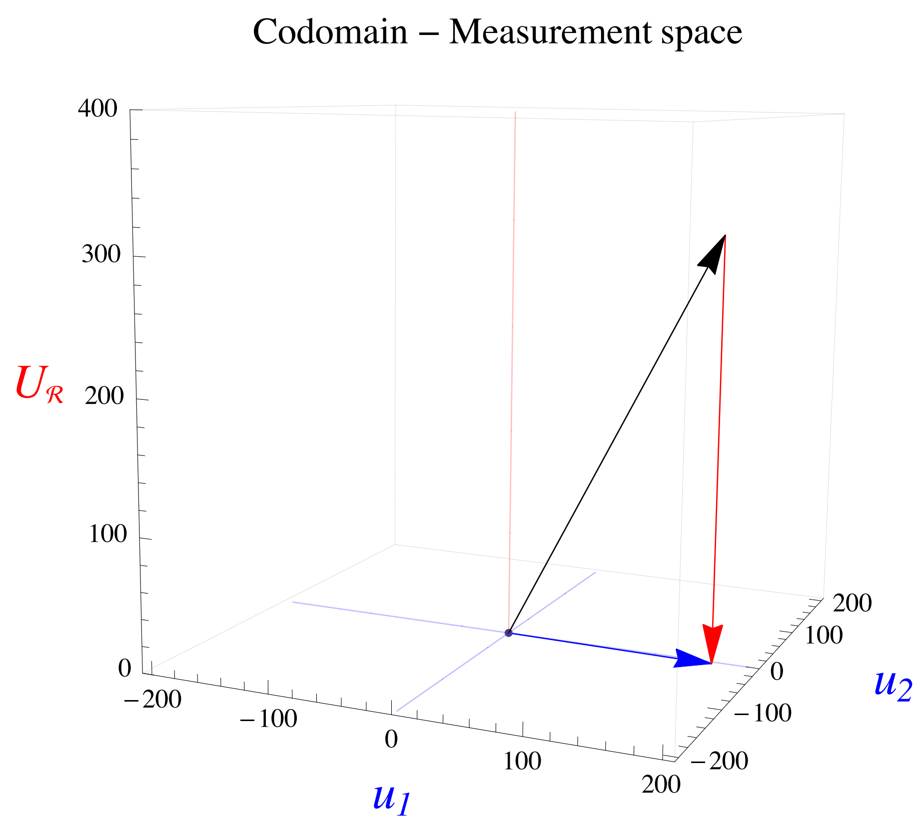


FIGURE 2.6. Measurement space $\mathcal{R}(\mathbf{A})$ for Bevington example. The black vector is y, the measured data. The range space $\mathbf{U}_{\mathcal{R}}$ is spanned by the vectors $[\mathbf{U}_{\mathcal{R}}]_1$ and $[\mathbf{U}_{\mathcal{R}}]_2$; the least squares solution $\mathbf{A}\,a$ lies within this plane and is represented by the blue line. The difference between the measurement y and the approximation $\mathbf{A}\,a$ is the residual error vector, shown in red. The black and red lines are actually hyperplanes of dimension 7.

When is the determinant 0? When the x vector is proportional to the 1 vector

$$x = \alpha 1, \tag{2.56}$$

that is, when the measurement locations are all identical: $x_1 = x_2 = \cdots = x_m = \alpha$. (In the case the rank of the matrix $\mathbf{A}$ would be $\rho = 1$.) Put in a form comparable to (2.11)

$$\begin{aligned} a_0 &= \Delta^{-1}\left(\left(x^{\mathrm{T}}x\right)\left(\mathbf{1}^{\mathrm{T}}y\right) - \left(\mathbf{1}^{\mathrm{T}}x\right)\left(x^{\mathrm{T}}y\right)\right), \\ a_1 &= \Delta^{-1}\left(\left(\mathbf{1}^{\mathrm{T}}\mathbf{1}\right)\left(x^{\mathrm{T}}y\right) - \left(\mathbf{1}^{\mathrm{T}}x\right)\left(\mathbf{1}^{\mathrm{T}}y\right)\right). \end{aligned} \tag{2.57}$$

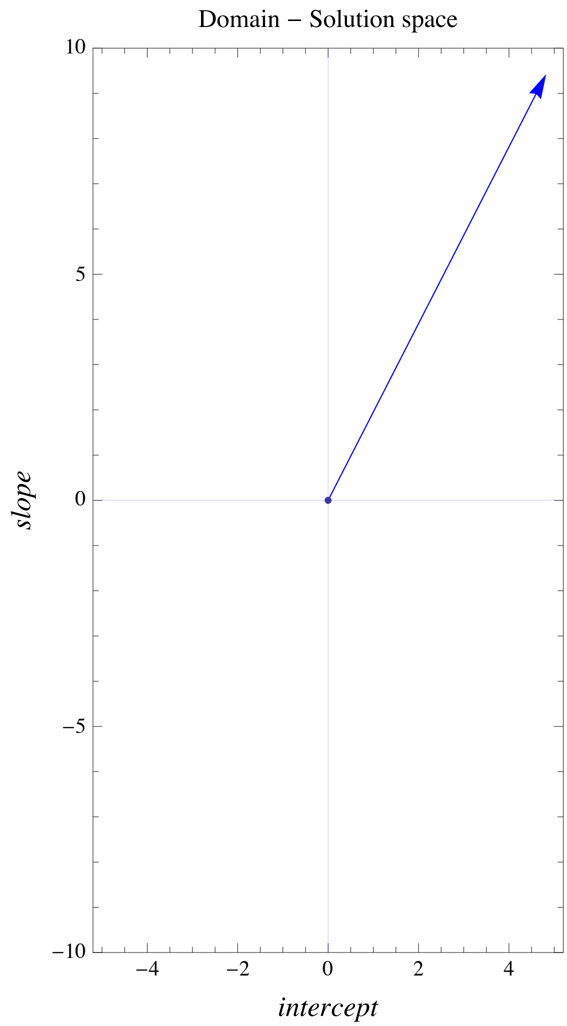


FIGURE 2.7. Solution space $\mathcal{R}(\mathbf{A}^*)$ for the Bevington example. The minimization occurs in codomain and is the projection of the measurement vector onto the range space $\mathcal{R}(\mathbf{A})$. The solution is in the domain $\mathcal{R}(\mathbf{A}^*)$ and this is the point which maps into to codomain.

$$\begin{aligned} r &= \mathbf{A}\alpha - y, \\ \epsilon^2 &= \frac{r^{\mathrm{T}}r}{\mathbf{1}^{\mathrm{T}}\mathbf{1} - n}\, diag\left(\left(\mathbf{A}^*\mathbf{A}\right)^{-1}\right). \end{aligned} \tag{2.58}$$

The final result would be quoted as having intercept $\alpha_0 \pm \epsilon_0$ and slope $\alpha_1 \pm \epsilon_1$. The corresponding errors are ((2.13))

$$\begin{bmatrix} \epsilon_0 \\ \epsilon_1 \end{bmatrix} = \sqrt{\frac{r^{\mathrm{T}}r}{\left(\mathbf{1}^{\mathrm{T}}\mathbf{1} - n\right)\Delta}} \sqrt{\begin{bmatrix} x^{\mathrm{T}}x \\ \mathbf{1}^{\mathrm{T}}\mathbf{1} \end{bmatrix}} \tag{2.59}$$

$$\mathbf{A} = \begin{bmatrix} 1 & x \end{bmatrix} \tag{2.60}$$

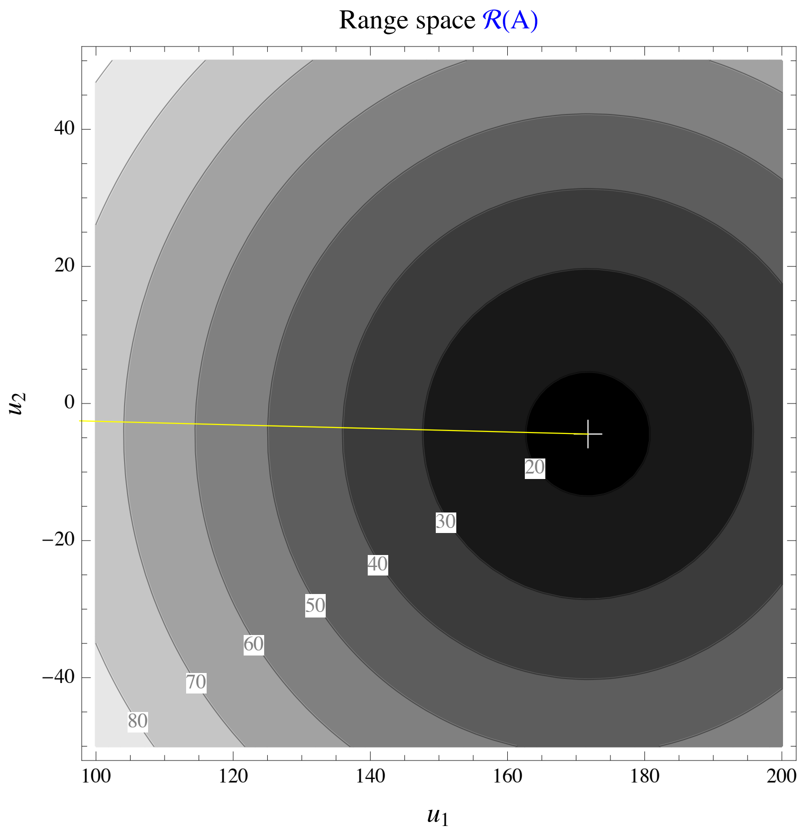


FIGURE 2.8. Minimization occurs in the codomain. The minimization occurs in codomain and is the projection of the measurement vector onto the range space $\mathcal{R}(\mathbf{A})$. The yellow line is the shadow of the data vector in the range space.

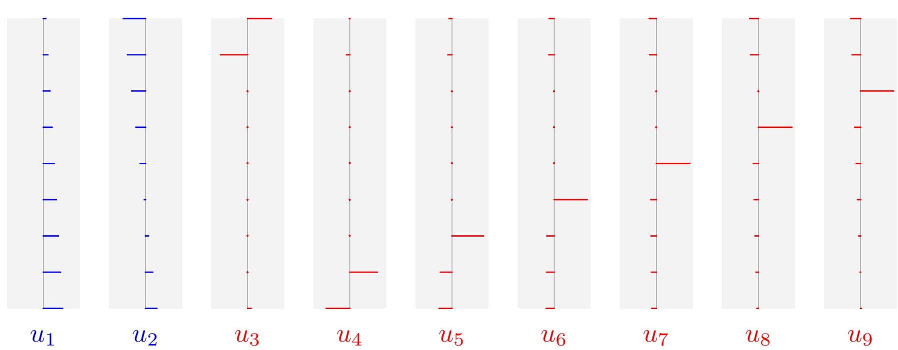


TABLE 2.3. The column vectors of $\mathbf{U}$. The gray box represents the maximum length an element may have: $[-1, 1]$.

Form the normal equations

$$\mathbf{A}a = y \quad \longrightarrow \quad \mathbf{A}^* \mathbf{A}\, a = \mathbf{A}^* y \tag{2.61}$$

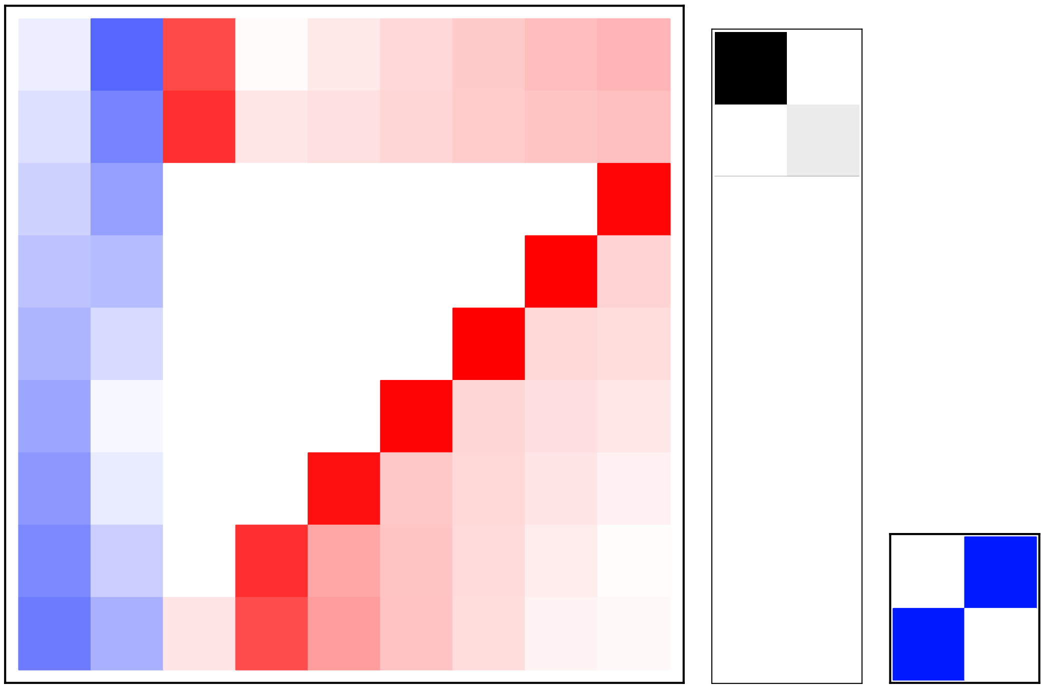


TABLE 2.4. The decomposition for the system matrix $\mathbf{A}$

The existence of a solution (when $y \neq 0$) is guaranteed. How do we know this? Consider the form recast so

$$\mathbf{A}^{*}\left(\mathbf{A}a\right) = \mathbf{A}^{*}y \tag{2.62}$$

Read this as "a matrix times a vector equals the same matrix times another vector".

$$\mathbf{A}^{*}v_{1} = v_{2} \tag{2.63}$$

This system has a solution whenever the vector v_2 can be expressed as a combination of the columns of the matrix $\mathbf{A}^{*}$.

$$a = \left(\mathbf{A}^{*}\mathbf{A}\right)^{-1}\mathbf{A}^{*}y \tag{2.64}$$

$$\mathcal{A} = \mathbf{A}^{*}\mathbf{A} \tag{2.65}$$

2.5. Full rank factorization

$$\mathbf{A} = \mathbf{B}\mathbf{C} \tag{2.66}$$

$$\mathbf{E_A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{2.67}$$

$$\mathbf{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{2.68}$$

2.6. Summary

à la Bevington Repeat equation (2.20).

$$\mathbf{A}a = Y$$

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ 1 & x_5 \\ 1 & x_6 \\ 1 & x_7 \\ 1 & x_8 \\ 1 & x_9 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{bmatrix} \tag{2.20}$$

2.6.1. SVD Decomposition. Every matrix has a singular value decomposition of the form

$$\mathbf{A} = \mathbf{U}\,\Sigma\,\mathbf{V}^* \tag{2.69}$$

In block form the overdetermined linear system looks like

$$\mathbf{A} = \left[\begin{array}{c|c} \mathbf{U}_{\mathcal{R}} & \mathbf{U}_{\mathcal{N}} \end{array} \right] \left[\begin{array}{c} \mathbf{S} \\ \hline \mathbf{0} \end{array} \right] \left[\begin{array}{c} \mathbf{V}^*_{\mathcal{R}} \end{array} \right] \tag{2.70}$$

Solution via pseudoinverse

$$\mathbf{A}^\dagger = \mathbf{V}\,\Sigma^{(\dagger)}\,\mathbf{U}^* \tag{2.71}$$

which has the block form

$$\mathbf{A}^\dagger = \left[\begin{array}{c} \mathbf{V}_{\mathcal{R}} \end{array} \right] \left[\begin{array}{c|c} \mathbf{S}^{-1} & \mathbf{0} \end{array} \right] \left[\begin{array}{c} \mathbf{U}^*_{\mathcal{R}} \\ \hline \mathbf{U}^*_{\mathcal{N}} \end{array} \right] \tag{2.72}$$

The pseudoinverse of the sigma matrix

$$\begin{aligned} \Sigma^{(\dagger)} = \left[\begin{array}{c|c} \mathbf{S}^{-1} & \mathbf{0} \end{array} \right] &= \left[\begin{array}{cc|ccccccc} \sigma_1^{-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2^{-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right] \\ &= \left[\begin{array}{cc|ccccccc} 0.0585054 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.735542 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right]. \end{aligned} \tag{2.73}$$

Computation

$$\mathbf{A}^{\dagger} = \frac{1}{180}\begin{bmatrix} 80 & 65 & 50 & 35 & 20 & 5 & -10 & -25 & -40 \\ -12 & -9 & -6 & -3 & 0 & 3 & 6 & 9 & 12 \end{bmatrix} \tag{2.74}$$

$$a = \mathbf{A}^{\dagger}\,\mathbf{Y} = \frac{1}{360}\begin{bmatrix} 1733 \\ 3387 \end{bmatrix} \approx \begin{bmatrix} 4.81389 \\ 9.40833 \end{bmatrix} \tag{2.75}$$

2.6.2. QR Decomposition.

$$\mathbf{A} = \mathbf{Q}\mathbf{R} \tag{2.76}$$

$$a = \mathbf{R}^{-1}\mathbf{Q}^{*}\mathbf{Y} \tag{2.77}$$

$$\mathbf{Q} = \begin{bmatrix} \frac{1}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} & \frac{1}{2\sqrt{15}}\begin{bmatrix} -4 \\ -3 \\ -2 \\ -1 \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \end{bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} 3 & 15 \\ 0 & 2\sqrt{15} \end{bmatrix} \tag{2.78}$$

$$\mathbf{R}^{-1} = \begin{bmatrix} \frac{1}{3} & \frac{5}{2\sqrt{15}} \\ 0 & \frac{1}{2\sqrt{15}} \end{bmatrix} \tag{2.79}$$

2.6.3. Normal equations.

$$\mathbf{A}^{*}\mathbf{A}\,a = \mathbf{A}^{*}\mathbf{Y} \tag{2.80}$$

$$a = \left(\mathbf{A}^{*}\mathbf{A}\right)^{-1}\mathbf{A}^{*}\mathbf{Y} \tag{2.81}$$

$$\left(\mathbf{A}^{*}\mathbf{A}\right)^{-1} = \frac{1}{180}\begin{bmatrix} 95 & -15 \\ -15 & 3 \end{bmatrix}, \qquad \mathbf{A}^{*}\mathbf{Y} = \frac{1}{10}\begin{bmatrix} 4667 \\ 28\,980 \end{bmatrix} \tag{2.82}$$

CHAPTER 3

Polynomial fits

3.1. Constant value

Given a set of measurements which constant best typifies the data?

3.2. Linear fit

3.3. The Weierstrass approximation theorem

Theorem 1 (Weierstrass approximation theorem). Polynomials are dense in the space of continuous functions with respect to the uniform norm.

Part 2

Scalar fields

CHAPTER 4

Strays

Topics of uncertain location

This is an example of an unnumbered first-level heading.

4.1. Finding the best circle

This is an example of an unnumbered first-level heading.

4.1.1. Nonlinear formulation. This is an example of a numbered first-level heading.

A circle is characterized by two parameters: an origin and a radius. The origin is a vector quantity, the radius a scalar.

$$O = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \tag{4.1}$$

Given a set of measurements p_j, $j = 1\colon m$.

$$(x - x_0)^2 + (y - y_0)^2 = \rho^2 \tag{4.2}$$

This implies a trial function

$$\chi^2(O, \rho) = \sum_{j=1}^{m} \left(\rho^2 - (x_j - x_0)^2 + (y_j - y_0)^2\right)^2 \tag{4.3}$$

In equation (4.2) the fit parameters for the origin appear in a nonlinear fashion, making this a nonlinear problem. There are many ways to solve such a problem. However, our focus is on linear problems.

4.1.2. Linear formulation. We start with the simple vector equation

$$p_j = r_k + O \tag{4.4}$$

from which we conclude

$$p_j^2 = r_j^2 + O^2 + 2r_j \cdot O \tag{4.5}$$

The trick is make one parameter disappear. To do so examine differences between the measurements

$$\Delta_{jk} = p_j - p_k = r_j - r_k \tag{4.6}$$

The data is no longer a list of m measurements of p; instead it is a list of τ differences where

$$\tau = \frac{1}{2}m(m-1) \tag{4.7}$$

For example, when $m = 4$

TABLE 4.1. The new data set compared to the old. The measured values p are converted to a set of differences Δ_{jk}.

	measurements	inputs
1	p_1	$\Delta_{12} = p_1 - p_2$
2	p_2	$\Delta_{13} = p_1 - p_3$
3	p_3	$\Delta_{14} = p_1 - p_4$
4	p_4	$\Delta_{23} = p_2 - p_3$
5		$\Delta_{24} = p_2 - p_4$
6		$\Delta_{34} = p_3 - p_4$

$$p_j^2 - p_k^2 = r_j^2 - r_k^2 + 2\left(r_j - r_k\right) \cdot O \tag{4.8}$$

$$r_j^2 = \rho^2 \qquad j = 1 \colon m \tag{4.9}$$

$$r_j^2 - r_k^2 = 0 \qquad j, k = 1 \colon m \tag{4.10}$$

The final trial function is this using equation (4.6)

$$p_j^2 - p_k^2 = 2\left(p_j - p_k\right) \cdot O \tag{4.11}$$

The trial function is then

$$p_j^2 - p_k^2 = 2\left(p_j - p_k\right) \cdot O \tag{4.12}$$

and the merit function

$$\chi^2\left(O\right) = \sum_{j=1}^{m-1} \sum_{k=1}^{m} \left(p_j^2 - p_k^2 - 2\left(p_j - p_k\right) \cdot O\right)^2 \tag{4.13}$$

Label the pairs

$$\xi = \left\{ \left[\begin{array}{c} 1 \\ 2 \end{array} \right], \left[\begin{array}{c} 1 \\ 3 \end{array} \right], \dots, \left[\begin{array}{c} m-1 \\ m \end{array} \right] \right\} \tag{4.14}$$

$$\chi^2\left(O\right) = \sum_{\mu=1}^{\tau} \left(2\Delta_{\xi} \cdot O - p_{\xi_1}^2 + p_{\xi_2}^2\right)^2 \tag{4.15}$$

Linear system

$$\begin{aligned} p_1^2 - p_2^2 &= 2\left(p_1 - p_2\right) \cdot O \\ p_1^2 - p_3^2 &= 2\left(p_1 - p_3\right) \cdot O \\ &\vdots \\ p_{m-1}^2 - p_m^2 &= 2\left(p_{m-1} - p_m\right) \cdot O \end{aligned} \tag{4.16}$$

solve for the origin O. The problem statement

$$\Delta O = b \tag{4.17}$$

In d dimensions the matrix dimensions are

$$\Delta \in \mathbb{R}_d^{\tau \times d}, \quad O \in \mathbb{R}^{d \times 1}, \quad b \in \mathbb{R}^{\tau \times 1}$$

k	$.x_k$	$.r_k$	$.x_k r_k$	$.g\left(x_k, r_k\right)$	.residual
1	0.0600	15.78246300	0.94694800	0.87847100	0.068476600
2	0.0500	18.89113500	0.94455700	0.88661400	0.057942900
3	0.0400	23.96075200	0.9584300	0.89475600	0.063673600
4	0.0300	32.35313500	0.97059400	0.90289900	0.067694900
5	0.0200	50.5914400	1.011828800	0.91104200	0.10078700
6	0.0100	104.68871700	1.046887200	0.91918400	0.12770300
7	0.0000	−1839.04936400	0.0000	0.92732700	−0.92732700
8	−0.0100	−103.18431800	1.031843200	0.9354700	0.096373500
9	−0.0200	−50.73661200	1.014732200	0.94361200	0.071119900
10	−0.0300	−33.75893700	1.012768100	0.95175500	0.061013100
11	−0.0400	−25.71153700	1.028461500	0.95989800	0.068563800
12	−0.0500	−20.80382100	1.040191100	0.9680400	0.072150700
13	−0.0600	−17.46685300	1.048011200	0.97618300	0.071828200

TABLE 4.2. Raw data for focal length measurement

and the matrices are defined as

$$\Delta = 2\begin{bmatrix} p_1 - p_2 \\ \vdots \\ p_{m-1} - p_m \end{bmatrix}, \quad O = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}, \quad b = \begin{bmatrix} p_1^2 - p_2^2 \\ \vdots \\ p_{m-1}^2 - p_m^2 \end{bmatrix} \tag{4.18}$$

4.2. Dubious applications

Let's explore the application of linear methods to nonlinear problems. Laboratory constrains mathematics. You inherit a spreadsheet and are asked to do basic analysis.

4.2.1. First analysis. Thin lens equation

$$x\left(r + e\right) = -f^2 \tag{4.19}$$

Trial function

$$\phi + ex = -xr \tag{4.20}$$

Physical fact

$$|\phi| = f^2 \tag{4.21}$$

Linear system

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

$$\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} \begin{bmatrix} \phi \\ e \end{bmatrix} = -\begin{bmatrix} x_1 R_1 \\ \vdots \\ x_m R_m \end{bmatrix} \tag{4.22}$$

The expected focal length is $f = 1$ m.

$$f_{measured} = 0.963 \pm 0.062 \text{ m} \tag{4.23}$$

measurements	$x_k,\ k = 1 : m$ $R_k,\ k = 1 : m$
fit parameters	$\phi \pm \epsilon_\phi$ $e \pm \epsilon_e$
trial function	$\phi + ex = -xR$
merit function	$\sum_{k=1}^{m} \left(\phi + ex_k + x_k R_k\right)^2$
linear system	$\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} \begin{bmatrix} \phi \\ e \end{bmatrix} = - \begin{bmatrix} x_1 R_1 \\ \vdots \\ x_m R_m \end{bmatrix}$

TABLE 4.3. Problem statement: determine focal length f

fit parameters	$\phi \pm \epsilon_\phi = -0.927 \pm 0.081$ $e \pm \epsilon_e = 0.8 \pm 2.2$
$r^{\mathrm{T}} r$	0.08511
c	$\frac{1}{91} \begin{bmatrix} 7 & 0 \\ 0 & 5000 \end{bmatrix}$
plots	data vs fit (**??**) residuals (**??**) merit function in $\mathcal{R}(\mathbf{A}^*)$ (4.3)

TABLE 4.4. Results: focal length f

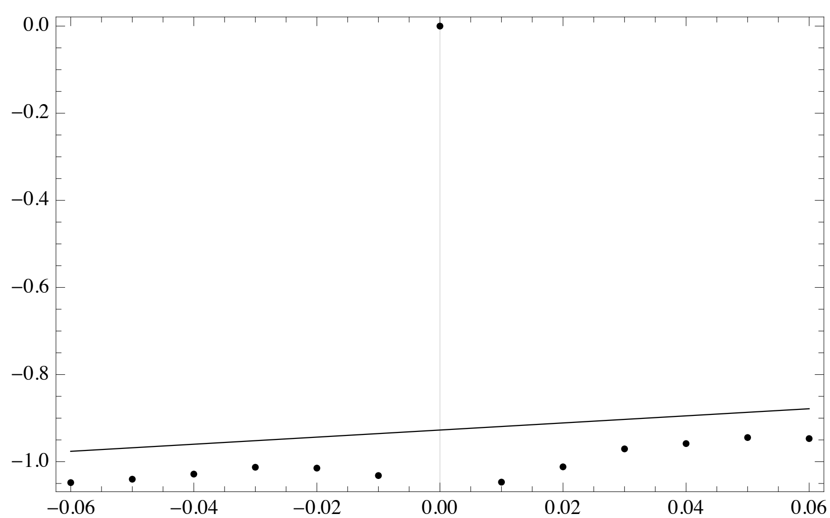


FIGURE 4.1. The solution curve (2.11) plotted against the data in table (**??**). The red lines represent the residual error, the difference between measurement and prediction.

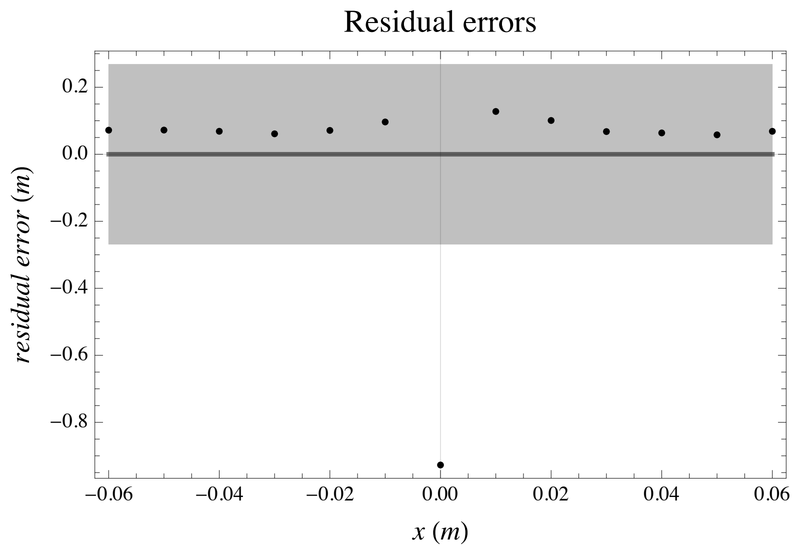


FIGURE 4.2. A closer look at the residual errors. The errors shown in figure (2.1) are plotted on an absolute scale. Ideally these errors should be randomly distributed, but we see signals of correlation suggesting the linear model may be inadequate.

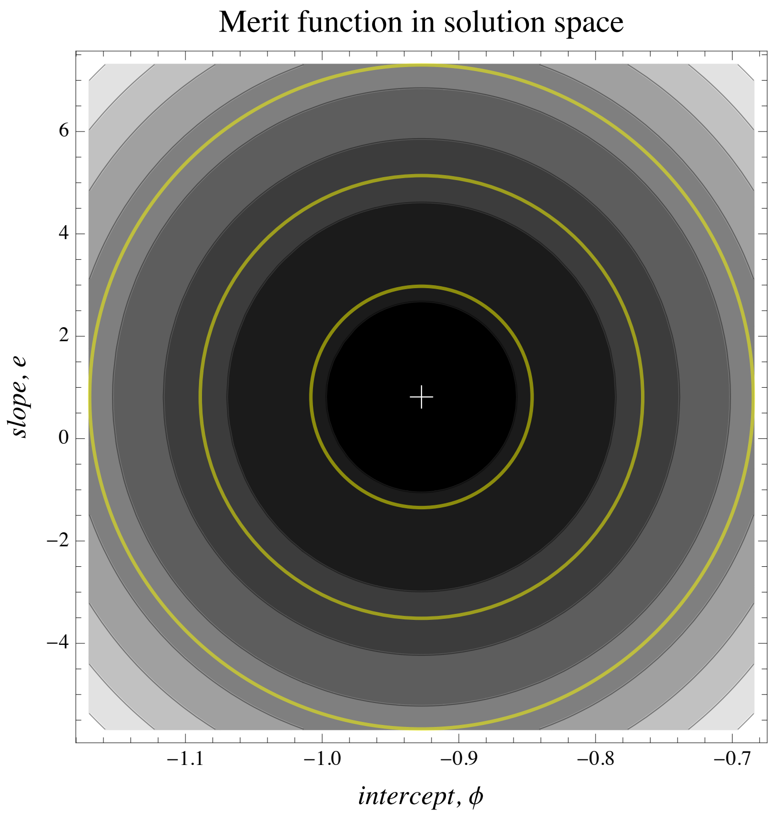


FIGURE 4.3. Looking at the merit function in solution space. The white cross in the center marks the solution found in (2.11). The yellow curves represent the error ellipses with radii of (a_0, a_1), $2\,(a_0, a_1)$, and $3\,(a_0, a_1)$. This provides a visual representation of how stable the solution is against perturbations in the data.

k	$.x_k$	$.r_k$	$.x_k r_k$	$.g\,(x_k, r_k)$	.residual
1	0.0600	15.78246300	0.94694800	−0.95574800	−0.0088006200
2	0.0500	18.89113500	0.94455700	−0.96389100	−0.019334300
3	0.0400	23.96075200	0.9584300	−0.97203400	−0.013603600
4	0.0300	32.35313500	0.97059400	−0.98017600	−0.0095823200
5	0.0200	50.5914400	1.011828800	−0.98831900	0.023509800
6	0.0100	104.68871700	1.046887200	−0.99646200	0.050425500
7	−0.0100	−1839.04936400	18.39049400	−1.01274700	0.019096200
8	−0.0200	−103.18431800	2.063686400	−1.020889600	−0.0061573900
9	−0.0300	−50.73661200	1.52209800	−1.029032300	−0.016264200
10	−0.0400	−33.75893700	1.35035700	−1.037174900	−0.0087134400
11	−0.0500	−25.71153700	1.28557700	−1.045317600	−0.0051265100
12	−0.0600	−20.80382100	1.24822900	−1.053460200	−0.0054490700

TABLE 4.5. Input data for focal length measurement

fit parameters	$\phi \pm \epsilon_\phi = -1.0046 \pm 0.0062$
	$e \pm \epsilon_e = 0.81 \pm 0.16$
$r^{\mathrm{T}} r$	0.004623
c	$\frac{1}{1092}\begin{bmatrix} 91 & 0 \\ 0 & 60\,000 \end{bmatrix}$
plots	data vs fit (**??**)
	residuals (**??**)
	merit function in $\mathcal{R}(\mathbf{A}^*)$ (4.6)

TABLE 4.6. Improved results: focal length f

4.2.2. Second analysis.

$$f_{measured} = 1.0023 \pm 0.0042 \text{ m} \tag{4.24}$$

Precision improves by an order of magnitude when the point at the origin is excluded. The exclusion criteria is based on the statistics of the data set, not difficulty in the measurement.

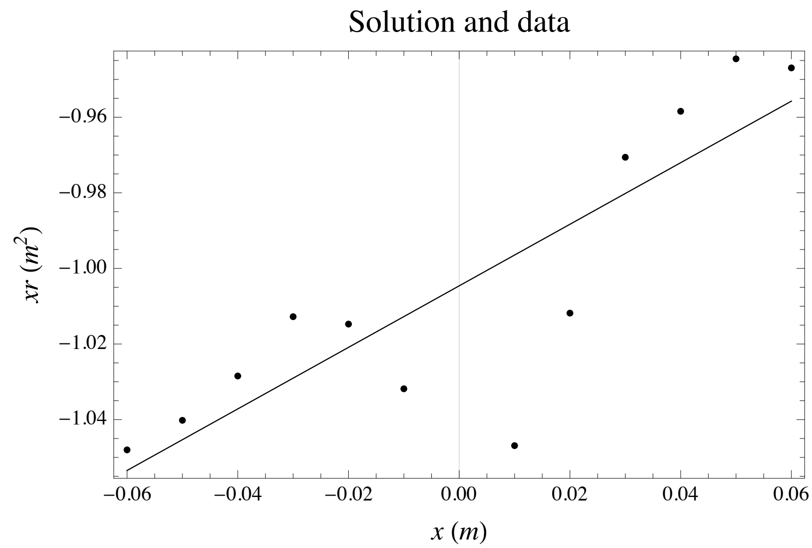


FIGURE 4.4. The solution curve (2.11) plotted against the data in table (**??**). The red lines represent the residual error, the difference between measurement and prediction.

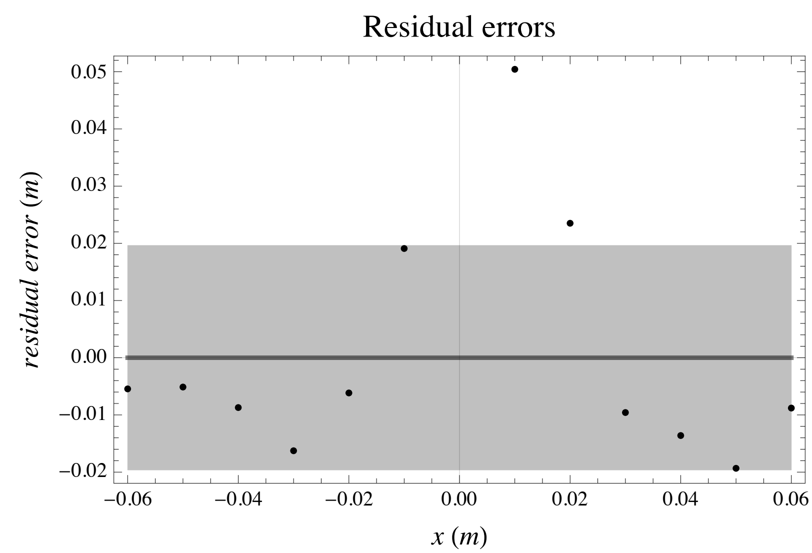


FIGURE 4.5. A closer look at the residual errors. The errors shown in figure (2.1) are plotted on an absolute scale. Ideally these errors should be randomly distributed, but we see signals of correlation suggesting the linear model may be inadequate.

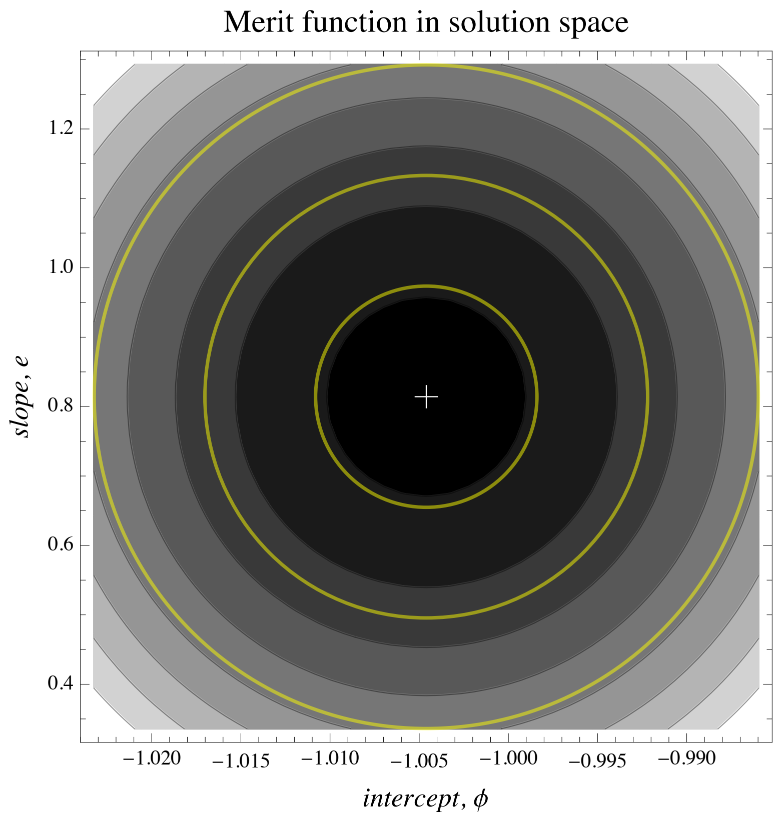


FIGURE 4.6. Looking at the merit function in solution space. The white cross in the center marks the solution found in (2.11). The yellow curves represent the error ellipses with radii of (a_0, a_1), $2\,(a_0, a_1)$, and $3\,(a_0, a_1)$. This provides a visual representation of how stable the solution is against perturbations in the data.

Part 3

Vector fields

Part 4

Tensor fields

Part 5

Zonal fits

Part 6

Stitching

APPENDIX A

Least squares with exemplar matrices

A.1. Linear systems

The essential concepts of least squares and the fundamental subspaces spring to life using exemplar matrices. The canonical linear system is

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

The matrix $\mathbf{A}$ has m rows and n columns of complex numbers. 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.

The residual error from the best fit is

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

Ignore the trivial cases where $b = 0$.

Exemplar matrices have obvious singular value decompositions.

$$\left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right], \quad \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \\ \hline 0 & 0 \end{array}\right], \quad \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right], \quad \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

$$\left[\begin{array}{c} \mathbf{I}_2 \end{array}\right], \quad \left[\begin{array}{c} \mathbf{I}_2 \\ \mathbf{0} \end{array}\right], \quad \left[\begin{array}{cc} \mathbf{I}_2 & \mathbf{0} \end{array}\right], \quad \left[\begin{array}{cc} \mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & 0 \end{array}\right]$$

A.2. Full rank: $\rho = m = n$

Start with an ideal linear system

$$\begin{aligned} \mathbf{A}x &= b \\ \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right] \left[\begin{array}{c} x_1 \\ x_2 \end{array}\right] &= \left[\begin{array}{c} b_1 \\ b_2 \end{array}\right]. \end{aligned} \tag{A.2}$$

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

TABLE A.1. Subspace decomposition for the $\mathbf{A}$ matrix in (A.2). Both null spaces are trivial and we have existence and uniqueness.

A.2.1. Subspace decomposition. Because the matrix $\mathbf{A}$ has full column rank the null space $\mathcal{N}(\mathbf{A}^*)$ is trivial. Because the matrix $\mathbf{A}^*$ has full row rank the null space $\mathcal{N}(\mathbf{A})$ is trivial.

A.2.2. Existence and uniqueness. We have unconditional existence and uniqueness without regard to the data vector. The exact solution is

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \tag{A.3}$$

which is also the least squares solution

$$x_{LS} = x = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \tag{A.4}$$

with $r^{\mathrm{T}} r = 0$ residual error. More formally, the linear system has a unique solution for any value of $b_1, b_2 \in \mathbb{C}$.

A.3. Full column rank: $\rho = n < m$

Foreshadowing the resolution of the range and null spaces, we show a partitioning

$$\begin{gathered} \mathbf{A}x = b \\ \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \\ \hline 0 & 0 \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \left[\begin{array}{c} b_1 \\ b_2 \\ \hline b_3 \end{array}\right]. \end{gathered} \tag{A.5}$$

A.3.1. Subspace decomposition. Fundamental Theorem of Linear Algebra

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

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

$$b = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ 0 \end{bmatrix}}_{\in \mathcal{R}(\mathbf{A})} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ b_3 \end{bmatrix}}_{\in \mathcal{N}(\mathbf{A}^*)} \tag{A.6}$$

A.3.2. Existence and uniqueness. When the data vector component $b_3 = 0$,

$$b = \begin{bmatrix} b_1 \\ b_2 \\ 0 \end{bmatrix} \in \mathcal{R}(\mathbf{A}) \tag{A.7}$$

the linear system is consistent and we have a unique solution

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \tag{A.8}$$

statement	subspace condition	data conditions
existence and uniqueness	$b \in \mathcal{R}(\mathbf{A})$	$(b_1 \ne 0 \text{ or } b_2 \ne 0)$ and $b_3 = 0$
existence	$b \in \mathcal{R}(\mathbf{A}) \oplus \mathcal{N}(\mathbf{A}^*)$	$(b_1 \ne 0 \text{ or } b_2 \ne 0)$ and $b_3 \ne 0$
no existence	$b \in \mathcal{N}(\mathbf{A}^*)$	$b_1 = b_2 = 0$, $b_3 \in \mathbb{C}^2$

TABLE A.2. Existence and uniqueness for the full column rank linear system in equation (A.5).

which is also the least squares solution

$$x_{LS} = x = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \tag{A.9}$$

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

$$x \in \mathcal{R}(\mathbf{A}^*). \tag{A.10}$$

A.3.3. No existence: When the data vector inhabits the null space we do not even have a least squares solution.

A.3.4. Existence, no uniqueness:

A.4. Full row rank: $\rho = m < n$

Foreshadowing the resolution of the range and null spaces, we show a partitioning

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

$$\left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}. \tag{A.11}$$

A.4.1. Subspace decomposition. Fundamental Theorem of Linear Algebra

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

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

$$x = \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix}}_{\in \mathcal{R}(\mathbf{A}^*)} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ x_3 \end{bmatrix}}_{\in \mathcal{N}(\mathbf{A})} \tag{A.12}$$

statement	subspace condition	data conditions
existence and uniqueness	$b \in \mathcal{R}(\mathbf{A})$	$(b_1 \neq 0$ or $b_2 \neq 0)$ and $b_3 = 0$
existence	$b \in \mathcal{R}(\mathbf{A}) \oplus \mathcal{N}(\mathbf{A}^*)$	$(b_1 \neq 0$ or $b_2 \neq 0)$ and $b_3 \neq 0$
no existence	$b \in \mathcal{N}(\mathbf{A}^*)$	$b_1 = b_2 = 0$, $b_3 \in \mathbb{C}^2$

TABLE A.3. Existence and uniqueness for the full column rank linear system in equation (A.5).

A.4.2. Existence and uniqueness. When the data vector component $b_3 = 0$,

$$b = \begin{bmatrix} b_1 \\ b_2 \\ 0 \end{bmatrix} \in \mathcal{R}(\mathbf{A}) \tag{A.13}$$

the linear system is consistent and we have a unique solution

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \tag{A.14}$$

which is also the least squares solution

$$x_{LS} = x = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \tag{A.15}$$

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

$$x \in \mathcal{R}(\mathbf{A}^*). \tag{A.16}$$

A.4.3. No existence. When the data vector inhabits the null space we do not even have a least squares solution.

A.4.4. Existence, no uniqueness.