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

Least Squares

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,

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

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.

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.

1.1.1. Zonal Approximation.

$$F = \nabla\phi$$

$$\begin{aligned} \varphi_1 - \varphi_0 &= x_1 \\ \varphi_2 - \varphi_1 &= x_2 \end{aligned}$$

$$\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \varphi_0 \\ \varphi_1 \\ \varphi_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

There are $m = 2$ measurements, $n = 3$ fit parameters, and the matrix rank is $\rho = 2$.

1.1.2. Modal Approximation.

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

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

$$\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

1.2. Least Squares Solution

Going back to (1.1), there is a solution when the data vector b can be expressed as a combination of the column vectors of $\mathbf{A}$. In this

1.2.1. Zonal Solution.

The solutions for the linear system in (1.1.1)

$$\left[\begin{array}{c}\varphi_0\\ \varphi_1\\ \varphi_2\end{array}\right]=\left[\begin{array}{rr}-2&-1\\ 1&-1\\ 1&2\end{array}\right]\left[\begin{array}{c}y_1\\ y_2\end{array}\right]+\alpha\left[\begin{array}{c}1\\ 1\\ 1\end{array}\right],\qquad \alpha\in\mathbb{C}.$$

$$\mathbf{A}=\left[\begin{array}{rrr}-1&1&0\\ 0&-1&1\end{array}\right]$$

$$\mathbf{A}x=\mathbf{A}\left(x+\alpha\left[\begin{array}{c}1\\ 1\\ 1\end{array}\right]\right)$$

because

$$\mathbf{A}\left[\begin{array}{c}1\\ 1\\ 1\end{array}\right]=\left[\begin{array}{rrr}-1&1&0\\ 0&-1&1\end{array}\right]\left[\begin{array}{c}1\\ 1\\ 1\end{array}\right]=\left[\begin{array}{c}0\\ 0\end{array}\right]$$

1.2.2. Modal Solution.

First define the vectors

$$\mathbf{1}=\left[\begin{array}{c}1\\ \vdots\\ 1\end{array}\right],\quad x=\left[\begin{array}{c}x_1\\ \vdots\\ x_m\end{array}\right],\quad y=\left[\begin{array}{c}y_1\\ \vdots\\ y_m\end{array}\right]$$

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

1.3. Least Squares Problem

$$x_{LS}=\left\{x\in\mathbb{C}^n\colon\ \|\mathbf{A}x-b\|_2^2\ \text{is minimized}\right\}\tag{1.2}$$

$$x_{LS}=\mathbf{A}^{\dagger}b+\left(\mathbb{I}_n-\mathbf{A}^{\dagger}\mathbf{A}\right)y,\qquad y\in\mathbb{C}^n$$

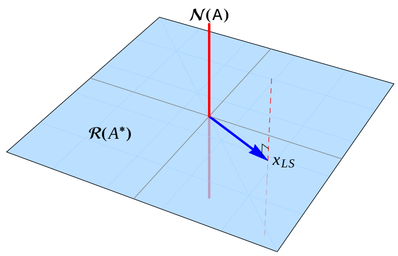
1.4. Fundamental Theorem of Linear Algebra

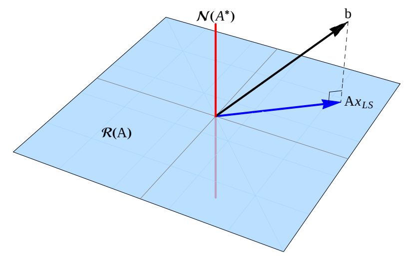
TABLE 1.1. The Fundamental Theorem of Linear Algebra

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 1.2. The Fundamental Theorem of Linear Algebra in pictures

Domain	Codomain
$\mathbb{C}^n$	$\mathbb{C}^m$





$$\begin{array}{rcl} \mathbf{A}\colon \mathbb{C}^n & \mapsto & \mathbb{C}^m \\ \mathbb{C}^n & \leftarrow\!\shortmid & \mathbb{C}^m : \mathbf{A}^* \end{array}$$

Part 2

Zonal fits

CHAPTER 2

Gradient Problems

The archetypal problem is

$$\mathbf{F} = \nabla\phi$$

where the vector $\mathbf{F} \in \mathbb{C}^n$ is measured and the goal is to compute the scalar function $\phi(x)\colon \mathbb{R}^n \mapsto \mathbb{C}$ in the Sobolev space

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

where the integrals in L^2 are in the Lebesgue sense and the derivative in the weak sense.

2.1. Rank Defect

Start the discussion in the context of a full rectangular grid. The derivative action of the gradient introduces a rank defect stemming from fundamental invariance:

$$D_x\phi(x) = D_x\left(\phi_x + const\right)$$

This goes back to the saying "you never solve for potential, you solve for potential differences." For example, if the steps in you house are 20 cm high, then going up three stair takes you 60 cm higher. But we cannot say what at what height we starting; we only know the change of height.

In n dimensions this will manifest is a solution with n free variables

$$\varphi = \begin{bmatrix} \varphi_1 \\ \vdots \\ \varphi_m \end{bmatrix} = \phi_{general} + \phi_{homogeneous} = \xi + \alpha_1\eta_1 + \cdots + \alpha_n\eta_n$$

where $\alpha \in \mathbb{C}^n$, and $\xi, \eta \in \mathbb{C}^n$. The general solution is in the range space

$$\phi_{general} = \xi \in \mathcal{R}(\mathbf{A}^*)$$

and the homogenous solution vectors are in the null space

$$\eta_k \in \mathcal{N}(\mathbf{A}), \qquad k = 1\colon n$$

and are mutually orthogonal: $\eta_j \cdot \eta_k = 0$ for $j = k$.

2.2. Average Gradient

Of particular interest is the case where a device measures the average of a gradient.

Consider the domain

$$\Gamma = \left\{x \in \mathbb{R}\colon 0 < x < \lambda_1 m\right\}, \qquad \lambda_1 \in \mathbb{R}, m \in \mathbb{N}$$

with the equipartition

$$\omega_k = \left\{x \in \mathbb{R}\colon \lambda_1(k-1) < x < \lambda_1 k\right\}, k = 1\colon m.$$

There are neither underlaps:

$$\Gamma = \bigcup_{k=1}^{m} \omega_k,$$

nor overlaps in this covering

$$\omega_j \cap \omega_k = \emptyset, j \neq k.$$

Part 3

Modal fits

Part 4

Nonlinear problems

Blindly applying linear tools to nonlinear problems presents many paths to perdition. Hope, no matter how fervent, cannot remedy mathematical maladies.

We stress the definition of the least squares problem in (**??**) as the first indication that something is amiss. We stress visualization methods to help reveal the status of a calculation.

(1) finding reasonable approximations which nudge the problem into linearity;
(2) iterating the solution to a linear problem to improve a nonlinear problem;
(3) separating a problem into linear and nonlinear components.

CHAPTER 3

Best Circle

3.1. 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.

$$O = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \tag{3.1}$$

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

$$(x - x_0)^2 - (y - y_0)^2 = \rho^2 \tag{3.2}$$

This implies a trial function

$$\chi^2(O, \rho) = \sum_{j=1}^{m} \left(\rho^2 - (x_j - x_0)^2 + (y_j - y_0)^2\right)^2 \tag{3.3}$$

In equation (3.2) 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.

3.2. Linear formulation

We start with the simple vector equation

$$p_j = r_k + O \tag{3.4}$$

from which we conclude

$$p_j^2 = r_j^2 + O^2 + 2r_j \cdot O \tag{3.5}$$

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

$$\Delta_{jk} = p_j - p_k = r_j - r_k \tag{3.6}$$

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

$$\tau = \frac{1}{2}m(m - 1) \tag{3.7}$$

For example, when $m = 4$

$$p_j^2 - p_k^2 = r_j^2 - r_k^2 + 2(r_j - r_k) \cdot O \tag{3.8}$$

$$r_j^2 = \rho^2 \qquad j = 1\colon m \tag{3.9}$$

Table 3.1. 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		$\Delta_{34} = p_3 - p_4$

$$r_j^2 - r_k^2 = 0 \qquad j, k = 1 \colon m \tag{3.10}$$

The final trial function is this using equation (3.6)

$$p_j^2 - p_k^2 = 2\left(p_j - p_k\right) \cdot O \tag{3.11}$$

The trial function is then

$$p_j^2 - p_k^2 = 2\left(p_j - p_k\right) \cdot O \tag{3.12}$$

and the merit function

$$\chi^2\left(O\right) = \sum_{j=1}^{m-1} \sum_{k=1}^{m} \left(p_j^2 - p_k^2 - 2\left(p_j - p_k\right) \cdot O\right)^2 \tag{3.13}$$

Label the pairs

$$\xi = \left\{ \left[\begin{array}{c} 1 \\ 2 \end{array} \right], \left[\begin{array}{c} 1 \\ 3 \end{array} \right], \dots, \left[\begin{array}{c} m-1 \\ m \end{array} \right] \right\} \tag{3.14}$$

$$\chi^2\left(O\right) = \sum_{\mu=1}^{\tau} \left(2\Delta_\xi \cdot O - p_{\xi_1}^2 + p_{\xi_2}^2\right)^2 \tag{3.15}$$

Linear system

$$\begin{aligned} p_1^2 - p_2^2 &= 2\left(p_1 - p_2\right) \cdot O \\ p_1^2 - p_3^2 &= 2\left(p_1 - p_3\right) \cdot O \\ &\vdots \\ p_{m-1}^2 - p_m^2 &= 2\left(p_{m-1} - p_m\right) \cdot O \end{aligned} \tag{3.16}$$

solve for the origin O. The problem statement

$$\Delta O = b \tag{3.17}$$

In d dimensions the matrix dimensions are

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

and the matrices are defined as

$$\Delta = 2 \left[\begin{array}{c} p_1 - p_2 \\ \vdots \\ p_{m-1} - p_m \end{array} \right], \quad O = \left[\begin{array}{c} x_1 \\ \vdots \\ x_d \end{array} \right], \quad b = \left[\begin{array}{c} p_1^2 - p_2^2 \\ \vdots \\ p_{m-1}^2 - p_m^2 \end{array} \right] \tag{3.18}$$

CHAPTER 4

Focal Length

This is a delightful example from the wild for it represents not only interesting mathematics but also represents the sociology of applied mathematics. In simplest form, a colleague walks in with a set of measurements and asks for linear analysis. You, the analyst, has no input on the design of the measurement apparatus or the measurement scheme. The apparatus may have been assembled for another task and the measurements represent a sanity check on the design. The apparatus may have been assembled in great haste or after great deliberation. The apparatus may represent significant financial and temporal investment, or it may be trivial. The device may be available for further measurement refinements, or it may be at a distant facility, or it may have been harvested for parts or deconstructed to free bench space.

Applied mathematics contains elements of applied sociology. The customer may have significant emotional investment in the device, the data, and the linear model. The customer may be phobic to higher order fits. The spectrum runs from someone who wants a number for a report to someone who wants insight. "Here is my spreadsheet." The hope being that the magic of least squares will salvage the experimental effort or just a subconscious association between computer results and legitimacy.

Critical elements are beyond our control. A good approach is to provide not just an answer, but also motivation for our colleagues to involve data analysts in device and experiment design.

Here we sit witnessing the collision of practicality and quality.

Often we are reliving a cautionary tale in designing the device you want to build instead of the device you need to make a measurement.

Is the best fit a good fit? A beauty of the method of least squares comes from the qualitative evaluation of the fit parameters. 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. We may think of the task as asking the question "how well does a linear model describe the data?"

In the case at hand, we had purchased a high quality lens with an expected focal length of $f = 1$ meter and the measurements were a quick test of a concept to measure focal length with a wavefront sensor.

4.1. First analysis

Thin lens equation

$$x\,(R+e) = -f^2 \tag{4.1}$$

TABLE 4.1. Full data set and results for focal length measurement.

k	x_k	R_k	$x_k R_k$	$\phi + e x_k$	residual
1	0.0600	15.782463	0.946948	0.878471	0.0684766
2	0.0500	18.891135	0.944557	0.886614	0.0579429
3	0.0400	23.960752	0.95843	0.894756	0.0636736
4	0.0300	32.353135	0.970594	0.902899	0.0676949
5	0.0200	50.59144	1.0118288	0.911042	0.100787
6	0.0100	104.688717	1.0468872	0.919184	0.127703
7	0.0000	−1839.049364	0.0000	0.927327	−0.927327
8	−0.0100	−103.184318	1.0318432	0.93547	0.0963735
9	−0.0200	−50.736612	1.0147322	0.943612	0.0711199
10	−0.0300	−33.758937	1.0127681	0.951755	0.0610131
11	−0.0400	−25.711537	1.0284615	0.959898	0.0685638
12	−0.0500	−20.803821	1.0401911	0.96804	0.0721507
13	−0.0600	−17.466853	1.0480112	0.976183	0.0718282

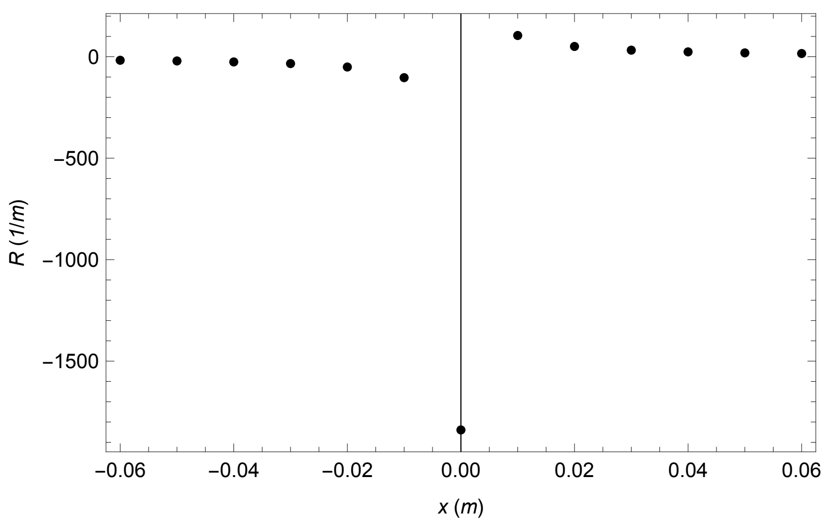


FIGURE 4.1. The complete data set.

Trial function is not of the form $y(x) = a_0 + a_1 x$ but is instead an implicit function $xy = a_0 + a_1 x$:

$$xR = \phi + ex \tag{4.2}$$

$$xR - \phi - ex = 0 \tag{4.3}$$

Physical fact

$$M(x, r) = xR - \phi - ex \tag{4.4}$$

$$|\phi| = f^2 \tag{4.5}$$

TABLE 4.2. Results: focal length f

measurements	$x_k,\ k = 1 : m$	micrometer position
	$R_k,\ k = 1 : m$	radius of curvature
result	$f \pm \epsilon_\phi$	
fit parameters	$\phi \pm \epsilon_\phi$	
	$e \pm \epsilon_e$	
trial function	$\phi + ex = -xR$	
merit function	$M(\phi, e) = \sum_{k=1}^{m} (\phi + ex_k + x_k R_k)^2$	
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.

TABLE 4.4. Results: focal length f

fit parameters	$\phi \pm \epsilon_\phi$	=	-0.927	$\pm$	0.081
	$e \pm \epsilon_e$	=	0.8	$\pm$	2.2
result	$f \pm \epsilon_f$	=	0.963	$\pm$	0.062
$r^{\mathrm{T}} r$	0.08511				
c	$\frac{1}{91} \begin{bmatrix} 7 & 0 \\ 0 & 5000 \end{bmatrix}$				
plots	data vs fit: figure 4.2				
	residuals: figure 4.3				
	merit function in $\mathcal{R}(\mathbf{A}^*)$: figure 4.4				

Linear system

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

$$\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} \begin{bmatrix} \phi \\ e \end{bmatrix} = - \begin{bmatrix} x_1 R_1 \\ \vdots \\ x_m R_m \end{bmatrix} \tag{4.6}$$

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

$$f_{measured} = 0.963 \pm 0.062 \text{ m} \tag{4.7}$$

$$\begin{aligned} f &= \sqrt{-\phi} \\ \sigma_f &= \frac{\sigma_\phi}{\sqrt{2}\,|\phi|} \end{aligned} \tag{4.8}$$

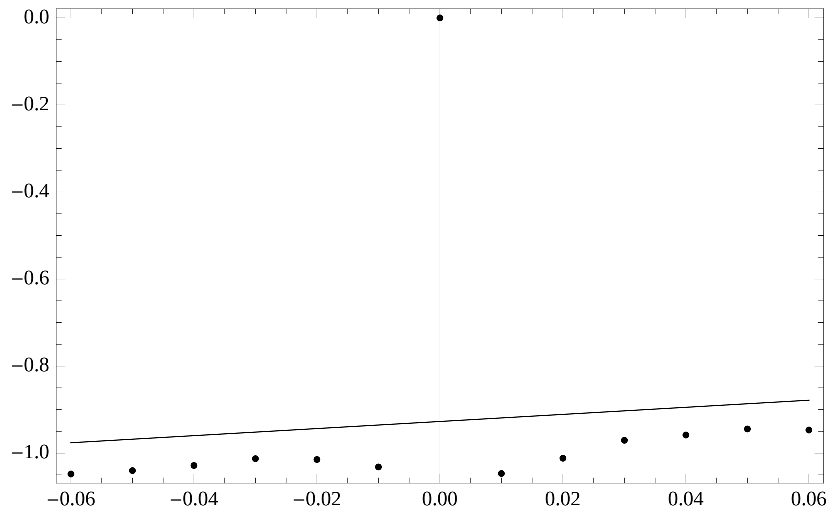


FIGURE 4.2. The solution curve (**??**) plotted against the data in table 4.1.

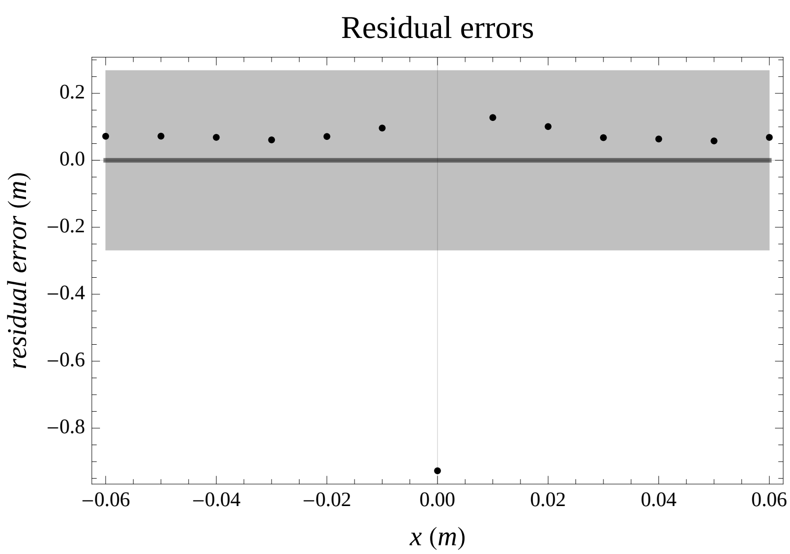


FIGURE 4.3. A closer look at the residual errors plotted on an absolute scale.

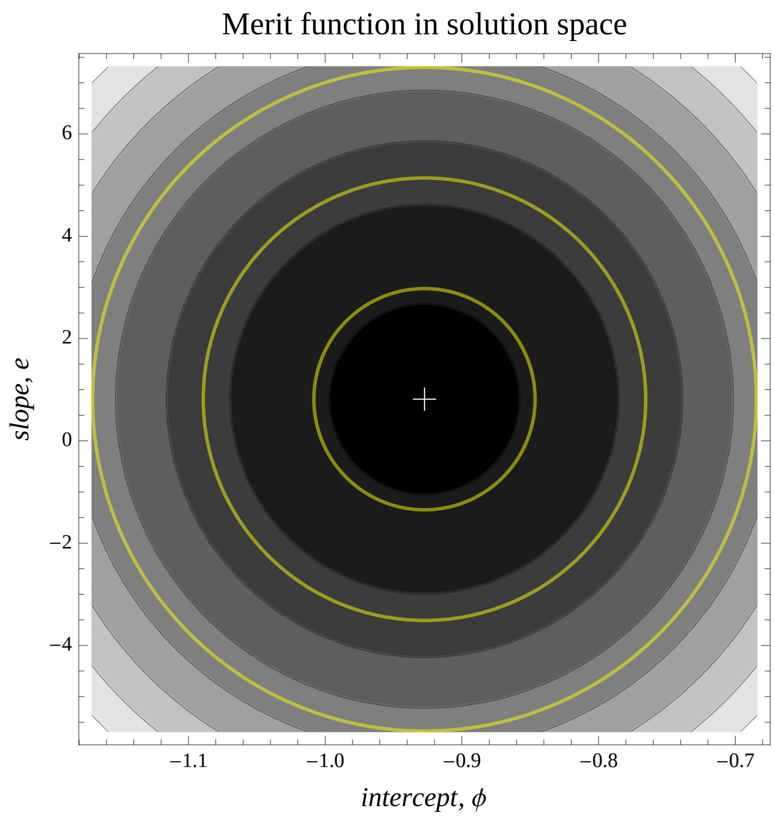


FIGURE 4.4. The merit function in solution space.

TABLE 4.5. Truncated data set and results for focal length measurement

k	x_k	R_k	$x_k R_k$	$\phi + e x_k$	residual
1	0.0600	15.782463	0.946948	0.955748	−0.00880062
2	0.0500	18.891135	0.944557	0.963891	−0.0193343
3	0.0400	23.960752	0.95843	0.972034	−0.0136036
4	0.0300	32.353135	0.970594	0.980176	−0.00958232
5	0.0200	50.59144	1.0118288	0.988319	0.0235098
6	0.0100	104.688717	1.0468872	0.996462	0.0504255
7	−0.0100	−103.184318	1.0318432	1.012747	0.0190962
8	−0.0200	−50.736612	1.0147322	1.0208896	−0.00615739
9	−0.0300	−33.758937	1.0127681	1.0290323	−0.0162642
10	−0.0400	−25.711537	1.0284615	1.0371749	−0.00871344
11	−0.0500	−20.803821	1.0401911	1.0453176	−0.00512651
12	−0.0600	−17.466853	1.0480112	1.0534602	−0.00544907

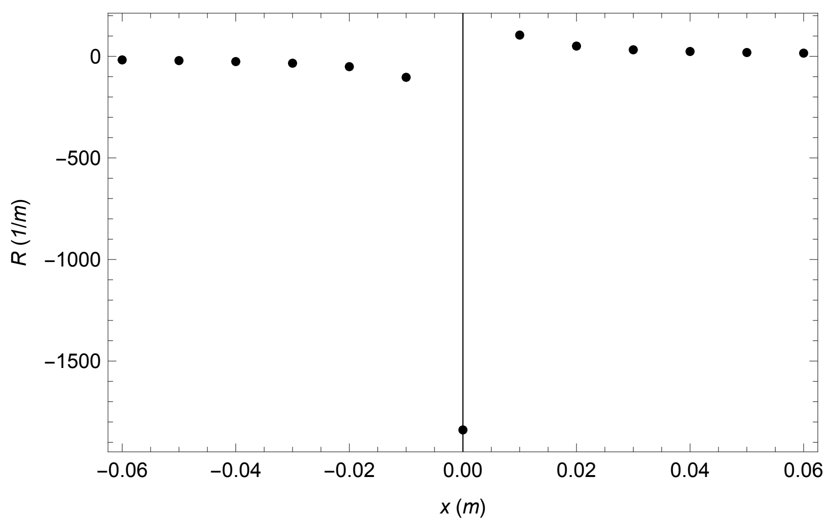


FIGURE 4.5. The cull data set.

4.2. Second Analysis

$$f_{measured} = 1.0023 \pm 0.0042 \text{ m} \tag{4.9}$$

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.

TABLE 4.6. Improved results: focal length f

fit parameters	$\phi \pm \epsilon_\phi$	$=$	-1.0046	$\pm$	0.0062
	$e \pm \epsilon_e$	$=$	0.81	$\pm$	0.16
result	$f \pm \epsilon_f$	$=$	1.0023	$\pm$	0.0044
$r^{\mathrm{T}} r$	0.004623				
c	$\frac{1}{1092}\begin{bmatrix} 91 & 0 \\ 0 & 60\,000 \end{bmatrix}$				
plots	data vs fit: Fig. 4.6				
	residuals: Fig. 4.7				
	merit function in $\mathcal{R}(\mathbf{A}^*)$: Fig. 4.8				

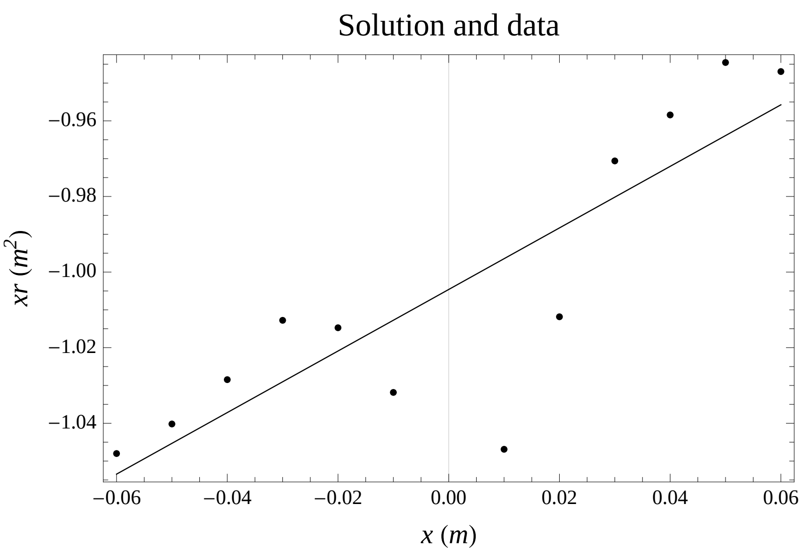


FIGURE 4.6. The solution curve (**??**) plotted against the data in table **??**.

TABLE 4.7. Focal length computation for both data sets.

full set	0.963	$\pm$	0.062
exclude origin	1.0023	$\pm$	0.0042

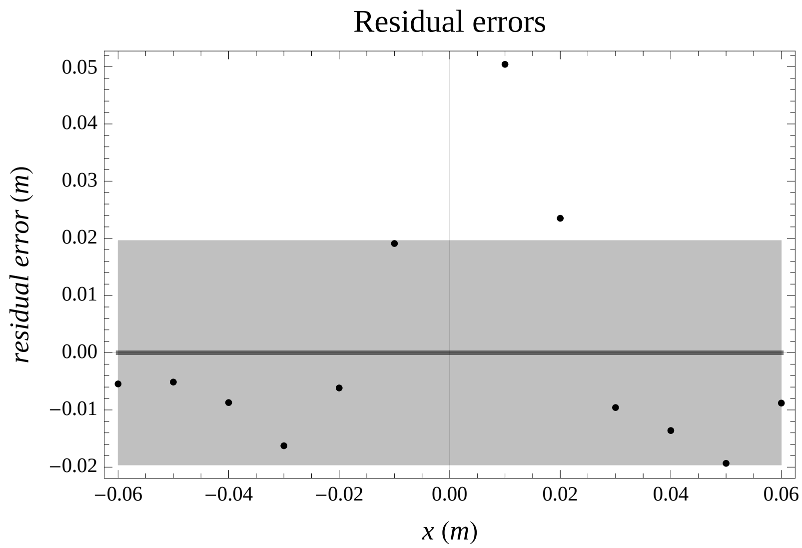


FIGURE 4.7. A closer look at the residual errors plotted on an absolute scale.

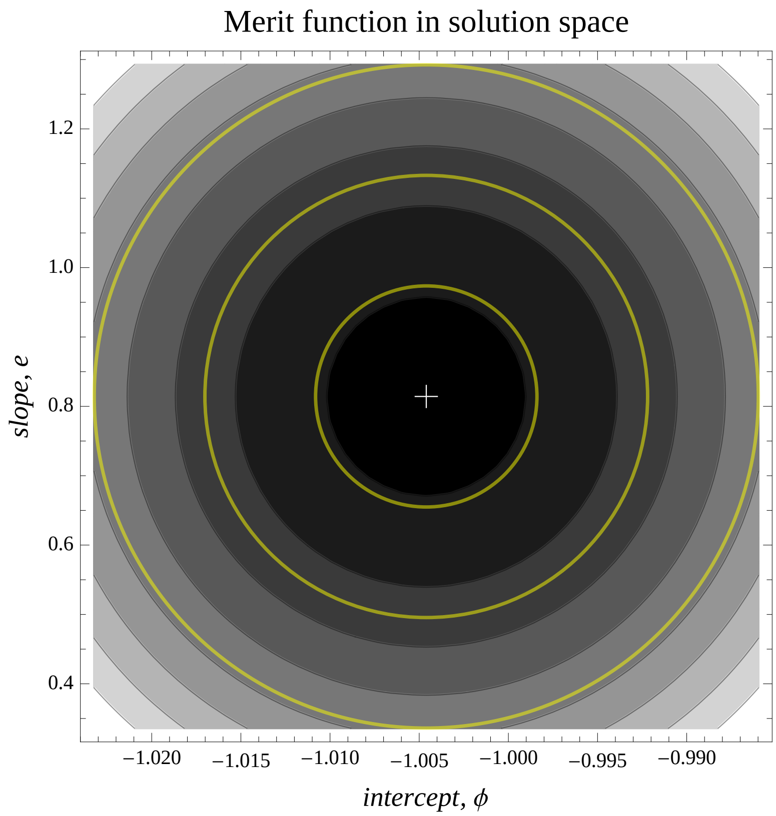


FIGURE 4.8. The merit function in solution space: the white cross in the center marks the solution found in (**??**), the yellow curves represent the error ellipses with radii of (a_0, a_1), $2\,(a_0, a_1)$, and $3\,(a_0, a_1)$.

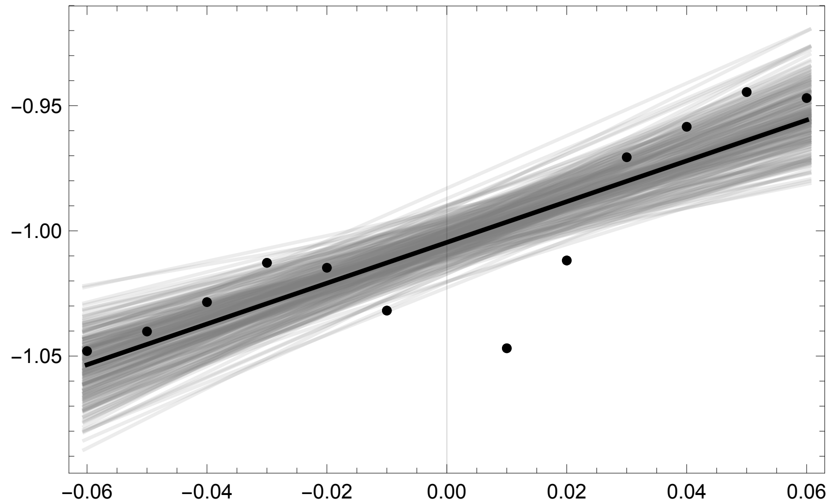


FIGURE 4.9. Whisker plot.

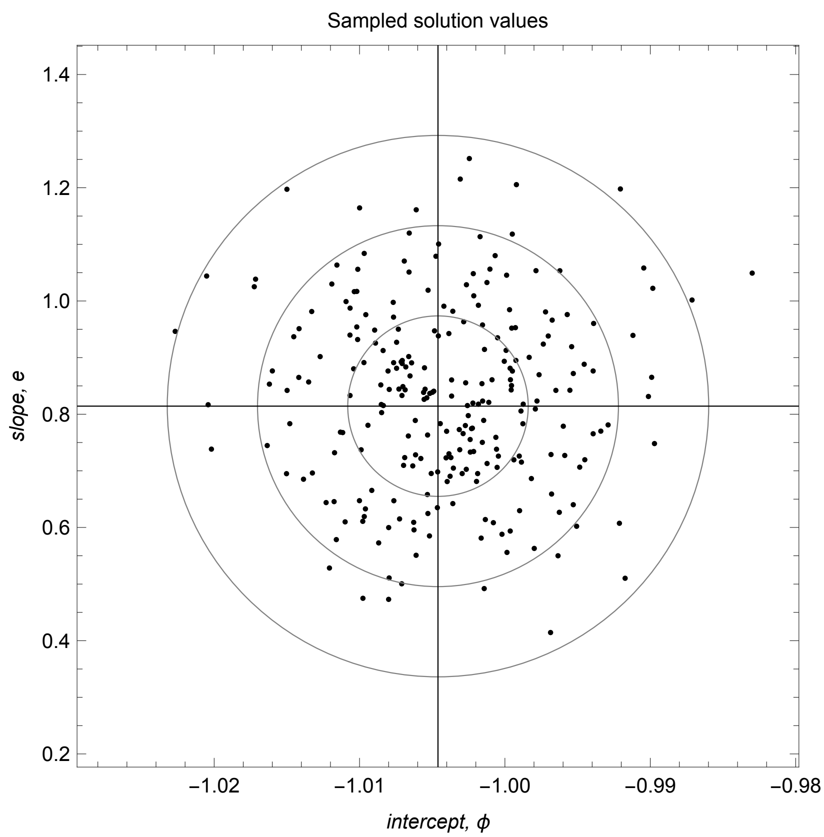


FIGURE 4.10. Sampled solution values.

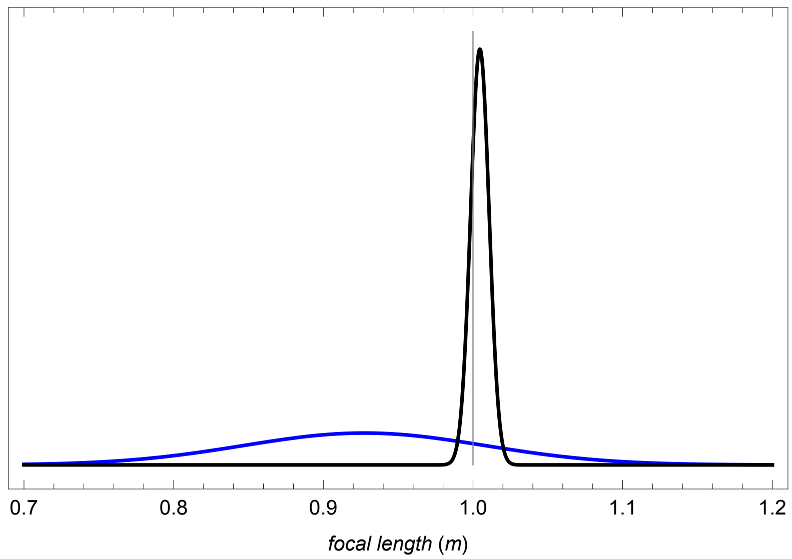


FIGURE 4.11. Accuracy and precision before and after removing data.

CHAPTER 5

Population Growth

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

5.1. Model

$$y(t) = c_1 + c_2\,(t - 1900) + c_3 e^{d(t-1900)} \tag{5.1}$$

$$\min_{c \in \mathbb{R}^3} \left\| \mathbf{A}(d) \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} - y \right\|_2^2 \tag{5.2}$$

5.2. Example

5.3. Comparison

Table 5.1. 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 5.2. Results: census

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

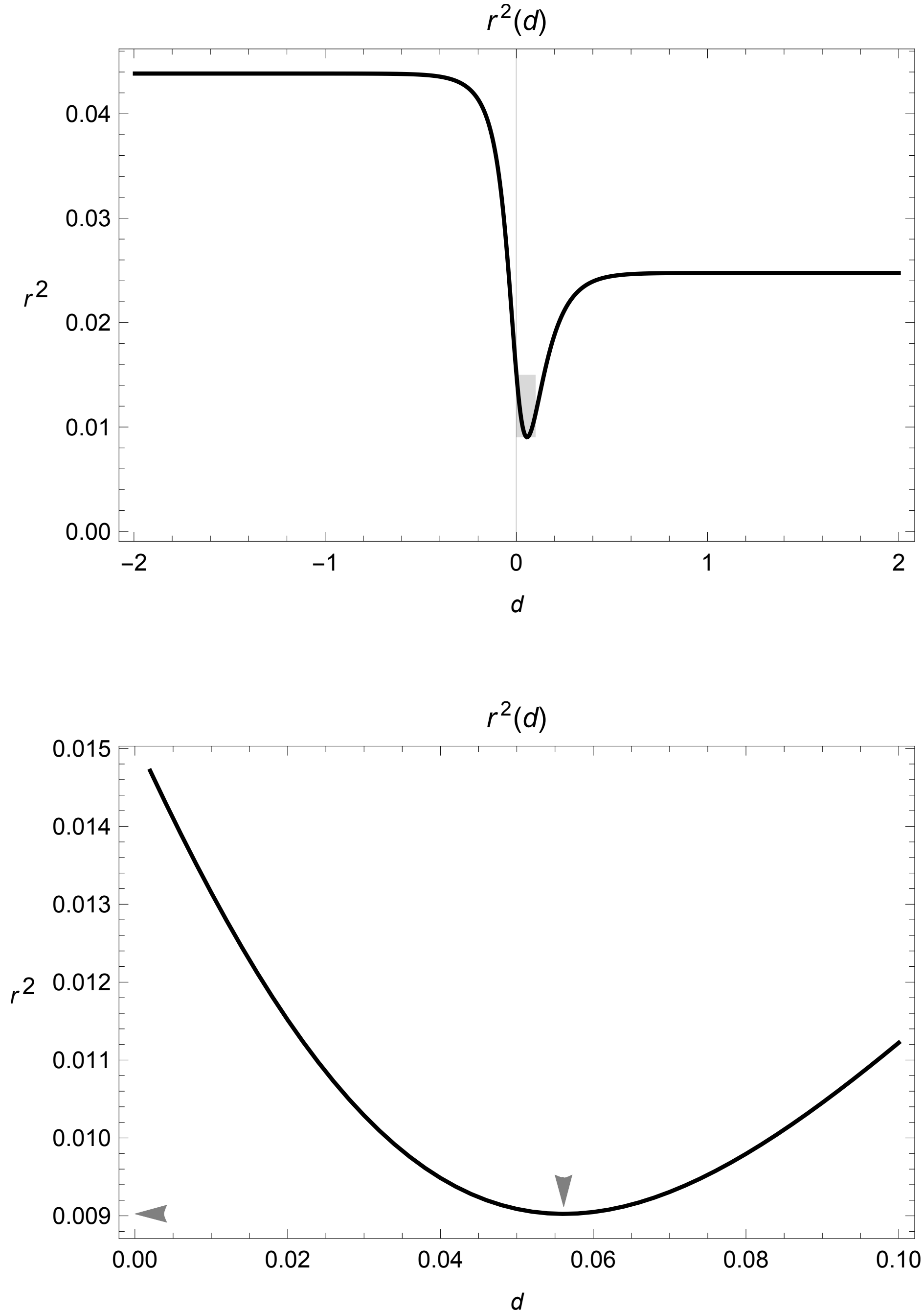


FIGURE 5.1. The shaded region in this plot is shown below.

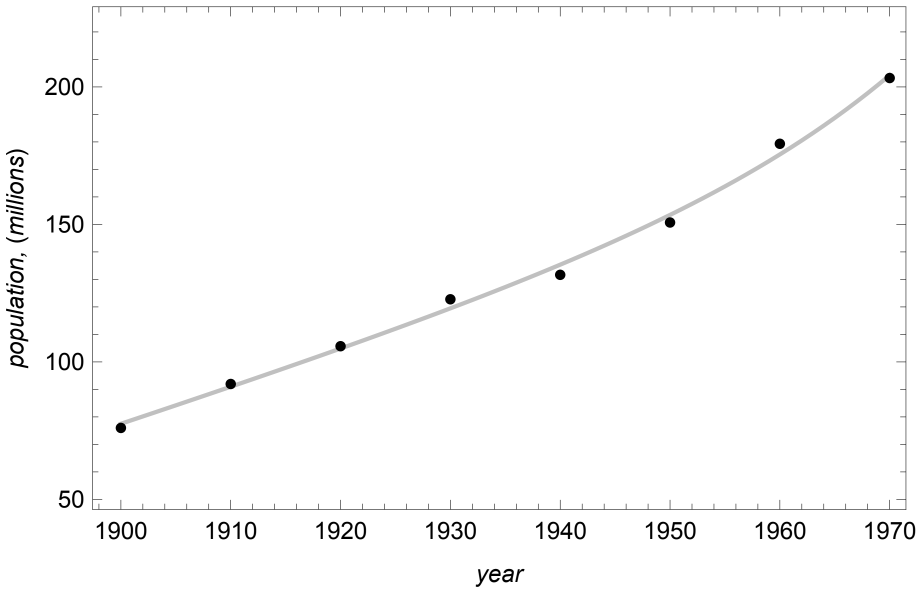


FIGURE 5.2. Solution plotted against data.

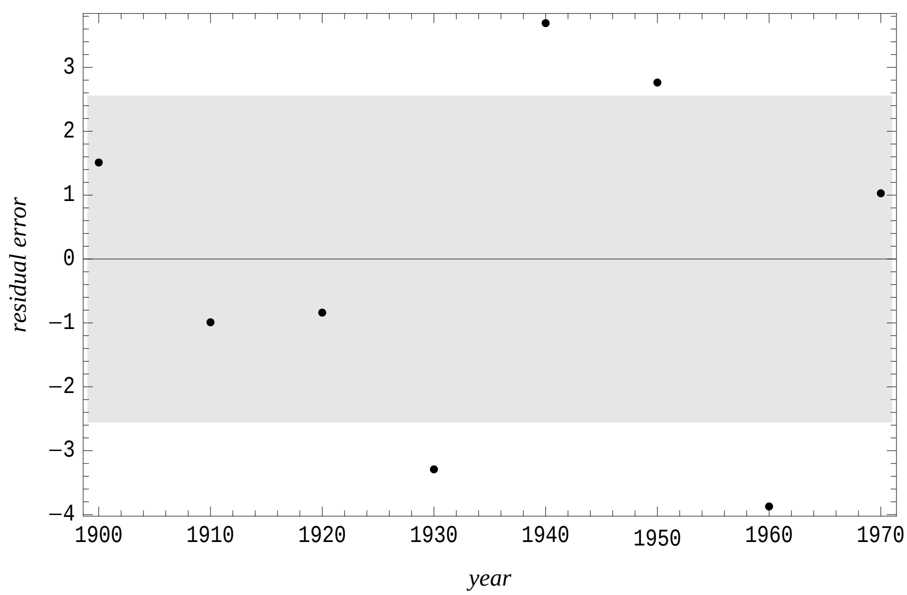


FIGURE 5.3. Residual errors.

FIGURE 5.4. The merit function with c_1 and c_2 fixed at best values showing least squares solution (center) and null cline (dashed, yellow).

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.

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

A.2.1. Full rank: $\rho = m = n$. Start with an ideal linear system

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

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}. \tag{A.2}$$

Subspace decomposition:

TABLE A.2. Subspace decomposition for the $\mathbf{A}$ matrix in (A.2).

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

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.

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.2.2. Full column rank: $\rho = n < m$. Foreshadowing the resolution of the range and null spaces, we show a partitioning

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

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

Subspace decomposition:

TABLE A.3. Subspace decomposition for the $\mathbf{A}$ matrix in (A.5).

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

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	$b \in \mathcal{R}(\mathbf{A})$	$(b_1 \neq 0$ 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	$b \in \mathcal{N}(\mathbf{A}^*)$	$b_1 = b_2 = 0$, $b_3 \in \mathbb{C}$

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

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

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

$$b = \left[\begin{array}{c} b_1 \\ b_2 \\ 0 \end{array}\right] \in \mathcal{R}(\mathbf{A}) \tag{A.7}$$

the linear system is consistent and we have a unique solution

$$x = \left[\begin{array}{c} x_1 \\ x_2 \end{array}\right] = \left[\begin{array}{c} b_1 \\ b_2 \end{array}\right] \tag{A.8}$$

which is also the least squares solution

$$x_{LS} = x = \left[\begin{array}{c} b_1 \\ b_2 \end{array}\right] \tag{A.9}$$

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

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

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

Existence, no uniqueness:

A.2.3. Full row rank: $\rho = m < n$. Foreshadowing the resolution of the range and null spaces, we show a partitioning

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

Subspace decomposition:

TABLE A.5. Subspace decomposition for the $\mathbf{A}$ matrix in (A.11).

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

TABLE A.6. Existence and uniqueness for the full column rank linear system in equation (A.5).

statement	subspace condition	data conditions
existence and uniqueness	$b \in \mathcal{R}(\mathbf{A})$	$(b_1 \neq 0$ 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	$b \in \mathcal{N}(\mathbf{A}^*)$	$b_1 = b_2 = 0$, $b_3 \in \mathbb{C}$

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

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

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

$$
b = \left[\begin{array}{c} b_1 \\ b_2 \\ 0 \end{array} \right] \in \mathcal{R}(\mathbf{A}) \tag{A.13}
$$

the linear system is consistent and we have a unique solution

$$
x = \left[\begin{array}{c} x_1 \\ x_2 \end{array} \right] = \left[\begin{array}{c} b_1 \\ b_2 \end{array} \right] \tag{A.14}
$$

which is also the least squares solution

$$
x_{LS} = x = \left[\begin{array}{c} b_1 \\ b_2 \end{array} \right] \tag{A.15}
$$

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

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

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

Existence, no uniqueness:

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

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

$$\left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 0 \end{array}\right] \left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array}\right] = \left[\begin{array}{c} b_1 \\ b_2 \\ b_3 \end{array}\right]. \tag{A.17}$$

Singular Value Decomposition

$$\left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 0 \end{array}\right] = \mathbf{U}\,\Sigma\,\mathbf{V}^* = \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right] \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \\ \hline 0 & 0 \end{array}\right] \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right] \tag{A.18}$$

Subspace decomposition:

TABLE A.7. Subspace decomposition for the **A** matrix in (A.17).

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

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

$$x = \underbrace{\left[\begin{array}{c} x_1 \\ x_2 \\ 0 \end{array}\right]}_{\in\mathcal{R}(\mathbf{A}^*)} + \underbrace{\left[\begin{array}{c} 0 \\ 0 \\ x_3 \end{array}\right]}_{\in\mathcal{N}(\mathbf{A})} \tag{A.19}$$

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

$$b = \left[\begin{array}{c} b_1 \\ b_2 \\ 0 \end{array}\right] \in \mathcal{R}(\mathbf{A}) \tag{A.20}$$

the linear system is consistent and we have a unique solution

$$x = \left[\begin{array}{c} x_1 \\ x_2 \end{array}\right] = \left[\begin{array}{c} b_1 \\ b_2 \end{array}\right] \tag{A.21}$$

which is also the least squares solution

$$x_{LS} = x = \left[\begin{array}{c} b_1 \\ b_2 \end{array}\right] \tag{A.22}$$

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

$$x \in \mathcal{R}(\mathbf{A}^*). \tag{A.23}$$

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

Existence, no uniqueness:

TABLE A.8. Existence and uniqueness for the full column rank linear system in equation (A.5).

statement	subspace condition	data conditions
existence and uniqueness	$b \in \mathcal{R}(\mathbf{A})$	$(b_1 \neq 0 \text{ or } b_2 \neq 0)$ 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	$b \in \mathcal{N}(\mathbf{A}^*)$	$b_1 = b_2 = 0$, $b_3 \in \mathbb{C}$

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`