

Contents

List of Figures

List of Tables

Part I

Rudiments

Chapter 1

Least Squares Problems

1.1 Linear Systems

This story begins with the archetypal matrix–vector equation

$$\mathbf{A}x = b. \quad (1.1)$$

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,

$$\mathbf{A} \in \mathbb{C}^{m \times n}, \quad b \in \mathbb{C}^m, \quad x \in \mathbb{C}^n \quad (1.2)$$

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

1.1.1 Exact Solution

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

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

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

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

and the residual error vanishes:

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

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

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

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

The solution is

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

There is no residual error

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

1.1.2 No Exact Solution

But what happens when the vector b is *not* in the column space of the matrix \mathbf{A} ? We must relax our solution criteria and instead of insisting upon zero residual error, we ask for minimal residual error. Instead of a perfect solution, we ask for the best solution. One such class of solutions are least squares solutions.

To demonstrate this point, the previous example is modified to be:

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

When $b_2 \neq 0$ there is no exact solution. Consider the solutions given by

$$x_* = \begin{bmatrix} x_1 \\ 0 \end{bmatrix}. \quad (1.3)$$

The error is

$$\mathbf{A}x - b = - \begin{bmatrix} 0 \\ b_2 \end{bmatrix}$$

which has a norm of b_2 . This is the least possible error for the problem and we can safely call (??) the best solution. In this light, the transition from an exact solution to an inexact solution is natural.

There are two fundamental types of least squares solutions of interest and they are classified by the interpretation of the output. In the first case, zonal approximation, the output represents data at a physical zone, a point or a region. In the second case, the output represents an amplitude, a contribution for a mode. Basic examples follow.

1.1.3 Zonal Approximation

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

$$F = \nabla\phi$$

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

$$\begin{aligned}\varphi_1 - \varphi_0 &\approx \delta_1 \\ \varphi_2 - \varphi_1 &\approx \delta_2\end{aligned}$$

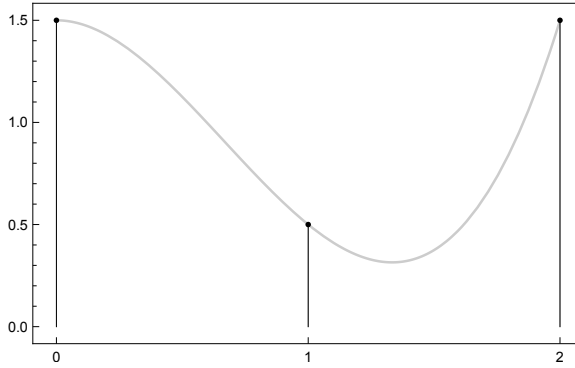


Figure 1.1. *Scalar function $\phi(x)$ (curve) and approximation φ_k (sticks).*

The system matrix

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

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

The linear system is

$$\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \varphi_0 \\ \varphi_1 \\ \varphi_2 \end{bmatrix} = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}. \quad (1.4)$$

1.1.4 Modal Approximation

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

$$y(x) = a_0 + a_1x$$

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

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

Taken together, the measurements compose the system

$$\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \quad (1.5)$$

1.2 Least Squares Solutions

In both the zonal and modal approximations, there is exact solution and we to minimize the residual error

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

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

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

1.2.1 Zonal Solution

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

$$\begin{bmatrix} \varphi_0 \\ \varphi_1 \\ \varphi_2 \end{bmatrix} = \begin{bmatrix} -2 & -1 \\ 1 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} + \alpha \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \alpha \in \mathbb{C}.$$

There is a continuum of solutions due to the fact that

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

An easy demonstration is to write

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

1.2.2 Modal Solution

The linear system is (??) can be case using the column vectors

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

The columns of the system matrix $\mathbf{A} = \begin{bmatrix} \mathbf{1} & x \end{bmatrix}$. The solution parameters can be expressed in terms of the column vectors:

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

1.3 Least Squares Problem

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

$$x_{LS} = \left\{ x \in \mathbb{C}^n : \|\mathbf{A}x - b\|_2^2 \text{ is minimized} \right\}. \quad (1.6)$$

The least squares solution may be a point or it may be a hyperplane. A general solution is

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

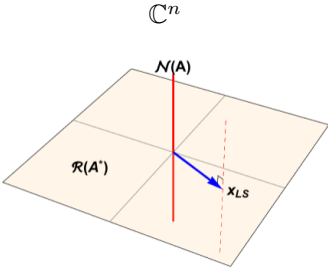
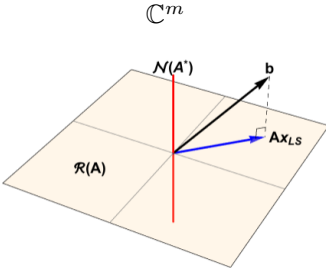
where the matrix \mathbf{A}^\dagger is the pseudoinverse.

1.4 Fundamental Theorem of Linear Algebra

Table 1.1. *The Fundamental Theorem of Linear Algebra for $\mathbf{A} \in \mathbb{C}^{m \times n}$*

$$\begin{array}{llll} \text{domain:} & \mathbb{C}^n & = & \mathcal{R}(\mathbf{A}^*) \oplus \mathcal{N}(\mathbf{A}) \\ \text{codomain:} & \mathbb{C}^m & = & \mathcal{R}(\mathbf{A}) \oplus \mathcal{N}(\mathbf{A}^*) \end{array}$$

Table 1.2. *The Fundamental Theorem of Linear Algebra for $\mathbf{A} \in \mathbb{C}^{m \times n}$*

Domain		Codomain
<div> <div> \mathbb{C}^n  </div> </div>	<div> $\mathbf{A}: \mathbb{C}^n \mapsto \mathbb{C}^m$ $\mathbb{C}^n \leftarrow \mathbb{C}^m: \mathbf{A}^*$ </div>	<div> <div> \mathbb{C}^m  </div> </div>

Chapter 2

Least Squares Solutions

2.1 Fundamental Theorem of Linear Algebra

[The Fundamental Theorem of Linear Algebra]The Fundamental Theorem of Linear Algebra for $\mathbf{A} \in \mathbb{C}^{m \times n}$

domain:

\mathbb{C}^n

=

$\mathcal{R}(\mathbf{A}^*)$

\oplus

$\mathcal{N}(\mathbf{A})$

codomain:

\mathbb{C}^m

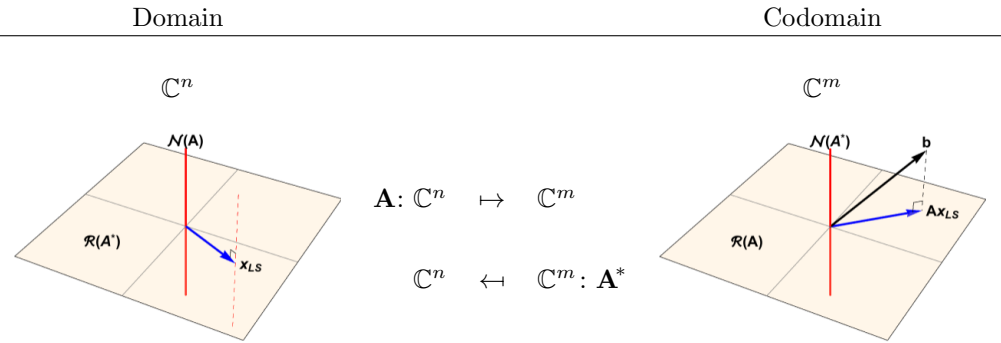
=

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

\oplus

$\mathcal{N}(\mathbf{A}^*)$

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



Part II

Appendices

Appendix A

Notation

A brief listing of notation.

Table A.1. *Matrices*

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

Table A.2. *Vectors*

a_k	k th column vector of matrix \mathbf{A}
$a_{[k]}$	k th row vector of matrix \mathbf{A}

Table A.3. *Vector spaces*

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

Table A.4. *Fields*

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

Table A.5. *Constants*

m	number of rows in a matrix
n	number of columns in a matrix
η_C	rank deficiency of the <i>column</i> space
η_R	rank deficiency of the <i>row</i> space
ρ	rank of a matrix

Table A.6. *Symbols*

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

Table A.7. *Abbreviations*

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

Appendix B

Lexicon

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

row space	column space
domain	codomain
preimage	image

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

Table B.2. *Matrix shapes.*

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

Table B.3. *Rank conditions.*

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

Part III

Backmatter

Bibliography

- [1] Richard Bellman, *Introduction to matrix analysis*, SIAM, Society for Industrial and Applied Mathematics; 2nd 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*, 3rd 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*, 3rd 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>