SANDIA REPORT
SAND2023-12630
Printed February 2023

Sandia National Laboratories

Notes on Regression Analysis for Radar Parameter Estimation

Armin W Doerry and Douglas L Bickel

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico
87185 and Livermore,
California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods/

Notes on Regression Analysis for Radar Parameter Estimation

Armin W Doerry and Douglas L Bickel

Abstract

A fundamental task of radar, beyond merely detecting a target, is to estimate some parameters associated with it. For example, this might include range, direction, velocity, etc. In any case, multiple measurements, often noisy, need to be processed to yield a 'best estimate' of the parameter. A common mathematical method for doing so is called "Regression" analysis. The goal is to minimize the expected squared error in the estimate. Even when alternate algorithms are considered, the least squared-error regression analysis is the benchmark against which alternatives are compared.

Acknowledgements

This report is the result in part of a Cooperative Research and Development Agreement (CRADA) between Sandia National Laboratories and General Atomics Aeronautical Systems, Inc. – CRADA No. SC08/01749.

GA-ASI, an affiliate of privately-held General Atomics, is a leading manufacturer of unmanned aircraft systems (UAS), tactical reconnaissance radars, and surveillance systems, including the Predator UAS series and Lynx Multi-Mode radar systems.

The authors thank Brandon Conder for his review of this report and helpful suggestions.

Contents

List of Figures 6
Acronyms and Definitions 7
Foreword 8
Classification 8
Author Contact Information 8
1 Introduction and Background 9
2 Linear Algebra and Least Squares 11
2.1 Weighted Least Squares 12
2.2 Covariance Analysis 13
2.3 Comments 14
3 Dilution of Precision 15
4 Linear Regression Analysis 19
4.1 Simple Linear Regression – Fitting to a Line 19
4.2 Simple Linear Regression – Fitting to a Polynomial 23
4.3 Simple Linear Regression – Fitting to a Sum of Functions 25
4.4 Multi-Linear Regression – Fitting to a Hyperplane 27
4.5 Multi-Linear Regression – Fitting to a Sum of Functions 29
4.6 Comments 31
5 Radar Applications 33
5.1 Phase Variance 33
5.2 Frequency Estimation 36
5.2.1 Two Phase Measurements 36
5.2.2 Multiple Phase Measurements 37
5.2.3 Relationship to Fourier Analysis 40
5.3 Delay Estimation 43
5.4 DOA Estimation 45
5.4.1 Two-Element Interferometer 45
5.4.2 Multi-Element DOA 48
6 Conclusions 57
Appendix A – Orthogonal Projections and Errors 59
Appendix B – Least Squares and Maximum Likelihood 61
Reference 63
Distribution 64

List of Figures

Figure 1. Signal space diagram for signal in noise. 33
Figure 2. Two-element interferometer geometry. 45
Figure 3. Interferometric SAR geometry. 47
Figure 4. Multi-element receiving array. In general, receive element spacing need not be regular, but in practice often is. 48
Figure 5. Interferometric SAR geometry for arbitrary multiple elements. 55
Figure 6. “OK everybody, is this a good enough line?” (Miramar Beach, Florida, USA, November 2022) 58

Acronyms and Definitions

3-D	3-Dimensional
AWGN	Additive White Gaussian Noise
CRADA	Cooperative Research and Development Agreement
DFT	Discrete Fourier Transform
DOA	Direction of Arrival
DOP	Dilution of Precision HDOP Horizontal DOP VDOP Vertical DOP TDOP Time DOP GDOP Geometric DOP PDOP Position DOP
EM	Electromagnetic
GMTI	Ground Moving Target Indicator [radar]
GPS	Global Positioning System
I, Q	In-phase, Quadrature
IFSAR	Interferometric SAR
InSAR	Interferometric SAR
MDV	Minimum Detectable Velocity
OLS	Ordinary Least Squares
RMS	Root Mean Square
SAR	Synthetic Aperture Radar
SNR	Signal to Noise Ratio
TLS	Total Least Squares
UAS	Unmanned/Uncrewed Aerial Systems

Foreword

This report details the results of an academic study. It does not presently exemplify any modes, methodologies, or techniques employed by any operational system known to the authors.

Classification

The specific mathematics and algorithms presented herein do not bear any release restrictions or distribution limitations.

This report formalizes preexisting informal notes and other documentation on the subject matter herein.

This report has been approved as Unclassified – Unlimited Release.

Author Contact Information

Armin Doerry awdoerr@sandia.gov 505-845-8165 www.doerry.us

1 Introduction and Background

Radar signal processing is replete with parameter estimation tasks. Parameters such as frequency, delay, Direction of Arrival (DOA), location, etc., are commonly estimated often from noisy measurements. So, not only do we desire a good estimate of the parameter(s), but then we often wish to know just how good that estimate is, i.e., an evaluation on its accuracy and precision.

A common measure for "goodness" is the expected squared-error, or variance, of the result. The variance in the error is a particularly attractive metric as it relates to the energy or power in the error component of a signal. This is intuitively satisfying to us. However, other measures are also sometimes used, an example being average absolute deviation, or sometimes higher-order statistics. Nevertheless, in this report we will assume the basic error statistic is the variance, or more generally the covariance when multiple outputs are simultaneously calculated. Algorithms that try to minimize the variance are often called "least-squares" algorithms. A closely related metric is Dilution of Precision (DOP).

In radar parameter estimation problems, we often wish to fit noisy data to some signal model, in such a manner to minimize any variance of the resulting fit. This overall task is called "regression." When we wish to fit to a linear sum of functions, then this is termed "linear regression." Some problems are in terms of a single independent variable, and some problems are in terms of more than one independent variable.

We have arranged the contents of this report as follows.

Section 2 reviews basic linear algebra concepts, including covariance calculations.

Section 3 discusses DOP measurements and their relationship to covariance measures.

Section 4 reviews linear regression analysis and calculations.

Section 5 presents a number of example radar applications.

One might reasonably ask "So, what exactly is 'new' in this report?" We answer by suggesting that we are collecting concepts and results from a number of sources into a single document, albeit with enhanced explanations, to serve as a common backdrop for several subsequent reports being written. We thereby expect to forego somewhat extensive appendices in several future reports, and especially duplicating those appendices in multiple reports, by referencing this report in their stead.

“Begin at the beginning,” the King said, very gravely,
“and go on till you come to the end: then stop.”
-- Lewis Carroll, from Alice in Wonderland

2 Linear Algebra and Least Squares

We present here some basic linear algebra concepts. Ample texts and publications exist that cover these concepts in detail.[1,2,3,4]

Consider the vector dot product of column vectors **a** and **x**, which we write as

$$\mathbf{a} \bullet \mathbf{x} = b, \tag{1}$$

where b is the scalar result of the dot product. We may then write the system of multiple such equations for a common **x** vector in matrix form as

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{2}$$

where we denote different **a** vectors and b values with a subscript as

$\mathbf{x}$ = common $M \times 1$ column vector,

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_N^T \end{bmatrix} = N \times M \text{ matrix, and}$$

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix} = N \times 1 \text{ column vector.} \tag{3}$$

Matrix **A** and vector **b** are known. We wish to solve for unknown vector **x**.

If **A** is square and full rank, we can solve

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}. \tag{4}$$

If the system is over-constrained, but still has full rank, we can solve

$$\mathbf{x} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}. \tag{5}$$

This solution and its properties are discussed in Appendix A and Appendix B.

We note that Eq. (5) reduces to Eq. (4) for a square matrix **A**. The entity $\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T$ is sometimes called the pseudo-inverse of **A**, or the Moore–Penrose pseudoinverse matrix of **A**.

Some additional useful matrix identities include for generic matrices **G** and **H**,

$$\begin{aligned}
&\left(\mathbf{GH}\right)^T = \mathbf{H}^T\mathbf{G}^T, \\
&\left(\mathbf{GH}\right)^{-1} = \mathbf{H}^{-1}\mathbf{G}^{-1}, \quad \text{if individual inverses exist,} \\
&\left(\mathbf{G}^T\right)^{-1} = \left(\mathbf{G}^{-1}\right)^T.
\end{aligned} \tag{6}$$

2.1 Weighted Least Squares

A weighted least squares solution to Eq. (2) may be found by solving

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

where the weight matrix **V** is a diagonal weight matrix. We may now solve the equation

$$\left(\mathbf{VA}\right)^T\ \mathbf{V}\ \mathbf{A}\ \mathbf{x} = \left(\mathbf{VA}\right)^T\ \mathbf{V}\ \mathbf{b}\,, \tag{8}$$

noting that $\left(\left(\mathbf{VA}\right)^T\ \mathbf{V}\ \mathbf{A}\ \right)$ is square, and invertible if full rank. The solution to Eq. (7) then becomes

$$\mathbf{x} = \left(\mathbf{A}^T\left[\mathbf{V}^T\mathbf{V}\right]\mathbf{A}\right)^{-1}\mathbf{A}^T\left[\mathbf{V}^T\mathbf{V}\right]\mathbf{b}\,. \tag{9}$$

If we cast

$$\mathbf{W} = \left[\mathbf{V}^T\mathbf{V}\right]\,, \tag{10}$$

then Eq. (9) can be written as

$$\mathbf{x} = \left(\mathbf{A}^T\mathbf{W}\ \mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{W}\ \mathbf{b}\,. \tag{11}$$

This is then the weighted least squares solution to Eq. (2). We observe that because **V** is diagonal, then so is **W**, and **W** is the same as its transpose, that is

$$\mathbf{W}^T = \mathbf{W}. \tag{12}$$

In addition, we note that $\left(\mathbf{A}^T\mathbf{W}\ \mathbf{A}\ \right)$ is square, and invertible if full rank.

Weighting is used to essentially favor more accurate observations at the expense of less accurate observations. Weights are often proportional to the inverse of the variance of respective observations, but this is not required. Weighting in this manner is related to "whitening" in signal processing.[5]

2.2 Covariance Analysis

Covariance provides us a measure of the uncertainty of the result, and how uncertainties are expected to behave relative to other uncertainties.

Using Eq. (9), consider calculating the covariance matrix of **x** in terms of the covariance matrix of **b**. Let

$$\boldsymbol{\Sigma}_{\mathbf{x}} = E\left\{\mathbf{x}\mathbf{x}^T\right\} = E\left\{\left(\left(\mathbf{V}\mathbf{A}\right)^T \mathbf{V}\mathbf{A}\right)^{-1}\left(\mathbf{V}\mathbf{A}\right)^T \mathbf{V}\mathbf{b}\left(\left(\left(\mathbf{V}\mathbf{A}\right)^T \mathbf{V}\mathbf{A}\right)^{-1}\left(\mathbf{V}\mathbf{A}\right)^T \mathbf{V}\mathbf{b}\right)^T\right\}. \tag{13}$$

This can be expanded and simplified to

$$\boldsymbol{\Sigma}_{\mathbf{x}} = \left(\mathbf{A}^T\left[\mathbf{V}^T\mathbf{V}\right]\mathbf{A}\right)^{-1}\mathbf{A}^T\left[\mathbf{V}^T\mathbf{V}\right]\boldsymbol{\Sigma}_{\mathbf{b}}\left[\mathbf{V}^T\mathbf{V}\right]\mathbf{A}\left(\mathbf{A}^T\left[\mathbf{V}^T\mathbf{V}\right]\mathbf{A}\right)^{-1}, \tag{14}$$

where the covariance matrix of **b** is

$$\boldsymbol{\Sigma}_{\mathbf{b}} = E\left\{\mathbf{b}\mathbf{b}^T\right\}. \tag{15}$$

If we assume that $\boldsymbol{\Sigma}_{\mathbf{b}}$ is also diagonal, meaning its elements are uncorrelated with each other, and we set the weighting such that

$$\left[\mathbf{V}^T\mathbf{V}\right] = \boldsymbol{\Sigma}_{\mathbf{b}}{}^{-1}, \tag{16}$$

then Eq. (14) reduces to

$$\boldsymbol{\Sigma}_{\mathbf{x}} = \left(\mathbf{A}^T\left[\mathbf{V}^T\mathbf{V}\right]\mathbf{A}\right)^{-1} = \left(\mathbf{A}^T\boldsymbol{\Sigma}_{\mathbf{b}}{}^{-1}\mathbf{A}\right)^{-1}. \tag{17}$$

In this case the weights in **V** are the reciprocals of the standard deviation. However, what gets applied in the calculation of Eq. (9) is the product $\left[\mathbf{V}^T\mathbf{V}\right]$. This product contains the reciprocals of the variances.

Recalling and employing Eq. (10), we also recall that Eq. (9) may be written as Eq. (11), which we repeat here as

$$\mathbf{x} = \left(\mathbf{A}^T\mathbf{W}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{W}\mathbf{b}. \tag{18}$$

Furthermore, if we still assume that $\boldsymbol{\Sigma}_{\mathbf{b}}$ is also diagonal, then Eq. (16) becomes

$$\mathbf{W} = \boldsymbol{\Sigma}_{\mathbf{b}}{}^{-1}, \tag{19}$$

and Eq. (17) can be written as

$$\boldsymbol{\Sigma}_{\mathbf{x}} = \left(\mathbf{A}^T \mathbf{W} \mathbf{A}\right)^{-1} = \left(\mathbf{A}^T \boldsymbol{\Sigma}_{\mathbf{b}}{}^{-1} \mathbf{A}\right)^{-1}. \tag{20}$$

In this case the weights in **W** are the reciprocals of the variances. So, what gets applied in the calculation of Eq. (18) are the reciprocals of the variances. This is not uncommon.

2.3 Comments

Several comments in no particular order.

- We stipulate that the derivation of Eq. (18) and Eq. (20) could have been directly in terms of **W**.
- Regardless of whether we develop the solution in terms of matrix **W** or matrix **V**, we still get the same answers.
- Note: The Strang[1] text derives the Weighted Least Squares derivation above where his **W** is our **V**. Then he defines an **H** matrix that is our **W** matrix. In the end, the result is the same.

3 Dilution of Precision

The concept of Dilution of Precision (DOP) is essentially a figure of merit that is an error sensitivity estimation or calculation. It is common for analysis of precision for the Global Positioning System (GPS),[6] but useful in a number of other applications, too.

We continue with the definitions of the last section. We recall the conventional linear system of equations given in matrix form as

$$\mathbf{A}\mathbf{x} = \mathbf{b}\,. \tag{21}$$

Recall that if **A** is full rank, we can solve for a weighted least-squares solution as

$$\mathbf{x} = \left(\mathbf{A}^T\mathbf{W}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{W}\mathbf{b}\,, \tag{22}$$

where **W** is a diagonal weighting matrix. This is the result previously given in Eq. (11). In the uniform weighting case, $\mathbf{W} = \mathbf{I}$. This formulation works, even if **A** is square. For convenience we will rewrite Eq. (22) as

$$\mathbf{x} = \mathbf{C}\,\mathbf{b}\,, \tag{23}$$

where generally

$$\mathbf{C} = \left(\left(\mathbf{A}^T\mathbf{W}\,\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{W}\right). \tag{24}$$

Now, an error in vector **b** will yield an error in the solution to vector **x**. That is

$$\left(\mathbf{x} + \boldsymbol{\varepsilon}_x\right) = \mathbf{C}\left(\mathbf{b} + \boldsymbol{\varepsilon}_b\right), \tag{25}$$

where the individual error vectors are defined as

$$\begin{aligned} &\boldsymbol{\varepsilon}_b \text{ = column vector of errors in } \mathbf{b}, \text{ and} \\ &\boldsymbol{\varepsilon}_x \text{= column vector of errors in } \mathbf{x}, \end{aligned} \tag{26}$$

which of course means that the errors are related as

$$\boldsymbol{\varepsilon}_x = \mathbf{C}\boldsymbol{\varepsilon}_b\,. \tag{27}$$

Covariances can be calculated as

$$E\left\{\boldsymbol{\varepsilon}_x{\boldsymbol{\varepsilon}_x}^T\right\} = \mathbf{C}E\left\{\boldsymbol{\varepsilon}_b{\boldsymbol{\varepsilon}_b}^T\right\}\mathbf{C}^T\,, \tag{28}$$

where the expected value function is defined as

$E\{z\}$ = expected value of z. (29)

We may write Eq. (28) somewhat more succinctly as

$$\mathbf{\Sigma}_x = \mathbf{C}\,\mathbf{\Sigma}_b\,\mathbf{C}^T, \tag{30}$$

where

$$\mathbf{\Sigma}_x = E\left\{\boldsymbol{\varepsilon}_x \boldsymbol{\varepsilon}_x{}^T\right\} = \text{covariance matrix for } \boldsymbol{\varepsilon}_x, \text{ and}$$
$$\mathbf{\Sigma}_b = E\left\{\boldsymbol{\varepsilon}_b \boldsymbol{\varepsilon}_b{}^T\right\} = \text{covariance matrix for } \boldsymbol{\varepsilon}_b. \tag{31}$$

Now for some special cases.

Case 1.

In this case, we stipulate equal independent variances in the **b** vector, so that

$$\mathbf{\Sigma}_b = \sigma_{ref}^2 \mathbf{I}, \tag{32}$$

where

σ_{ref}^2 = variance of the individual entries in **b**, common to all elements. (33)

In this case, and assuming an unweighted solution where $\mathbf{W} = \mathbf{I}$,

$$\mathbf{\Sigma}_x = \sigma_{ref}^2 \left(\mathbf{C}\,\mathbf{C}^T\right) = \sigma_{ref}^2 \left(\mathbf{A}^T\mathbf{A}\right)^{-1}. \tag{34}$$

Normalizing to σ^2 yields

$$\frac{\mathbf{\Sigma}_x}{\sigma_{ref}^2} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}. \tag{35}$$

The diagonal elements of $\mathbf{\Sigma}_x / \sigma_{ref}^2$ are ratios of variances. The square-root of these ratios are relative standard deviations. So, the square-root of the diagonal elements of $\mathbf{\Sigma}_x / \sigma_{ref}^2$ represent a scaling of standard deviations, that is, a scaling of precision due to errors in **b**. These are the DOP values.

Case 2.

In this case, we stipulate independent unequal variances in the **b** vector, so that

$$\boldsymbol{\Sigma}_b = \sigma_{ref}^2 \mathbf{K}, \tag{36}$$

where

σ_{ref}^2 = reference variance of the individual entries in **b**, and

K = diagonal matrix of weights meant to scale σ_{ref}^2. (37)

Since σ_{ref}^2 is a somewhat arbitrary reference value, elements of **K** may generally be more or less than one. In this case, the covariance matrix for $\boldsymbol{\varepsilon}_x$ becomes

$$\boldsymbol{\Sigma}_x = \sigma_{ref}^2 \left(\mathbf{C}\, \mathbf{K} \mathbf{C}^T \right). \tag{38}$$

Normalizing to σ_{ref}^2 yields

$$\frac{\boldsymbol{\Sigma}_x}{\sigma_{ref}^2} = \mathbf{C}\, \mathbf{K} \mathbf{C}^T. \tag{39}$$

When the variances in $\boldsymbol{\Sigma}_b$ are not equal,† then it is common to use a weighted least squares solution where the weights are the reciprocal of the variance of the observations. This essentially favors more accurate observations at the expense of less accurate observations, providing overall minimum variances in the output covariance matrix $\boldsymbol{\Sigma}_x$. This means

$$\mathbf{W} = \mathbf{K}^{-1}, \tag{40}$$

in which case Eq. (39) reduces to

$$\frac{\boldsymbol{\Sigma}_x}{\sigma_{ref}^2} = \left(\mathbf{A}^T \mathbf{W}\, \mathbf{A} \right)^{-1} = \left(\mathbf{A}^T \mathbf{K}^{-1}\, \mathbf{A} \right)^{-1}. \tag{41}$$

† A diagonal $\boldsymbol{\Sigma}_b$ with unequal elements is termed heteroscedasticity.

The diagonal elements of $\boldsymbol{\Sigma}_x \big/ \sigma_{ref}^2$ are still ratios of variances. So, the square-root of the diagonal elements of $\boldsymbol{\Sigma}_x \big/ \sigma_{ref}^2$ are still the DOP values, with the constraint that they are relative to a reference error in **b**.

Some DOP values, or combinations of DOP values, have been given specific nomenclature.

If elements of vector **x** represent position, with *x* and *y* coordinates representing horizontal positions, and *z* representing a vertical position coordinate, and the diagonal elements of $\boldsymbol{\Sigma}_x$ are

$$\operatorname{diag}\left(\boldsymbol{\Sigma}_x\right) = \begin{bmatrix} \sigma_x^2 & \sigma_y^2 & \sigma_z^2 \end{bmatrix}^T , \tag{42}$$

then

$$\begin{aligned} &\text{HDOP = Horizontal DOP} = \sqrt{\left(\sigma_x^2 + \sigma_y^2\right) \big/ \sigma_{ref}^2} \ , \\ &\text{VDOP = Vertical DOP} = \sqrt{\sigma_z^2 \big/ \sigma_{ref}^2} \ , \\ &\text{PDOP = Position DOP} = \sqrt{\left(\sigma_x^2 + \sigma_y^2 + \sigma_z^2\right) \big/ \sigma_{ref}^2} \end{aligned} \tag{43}$$

Other DOP values are also employed. For example, GPS figure of merits also include Time DOP (TDOP) and Geometric DOP (GDOP).

4 Linear Regression Analysis

Regression analysis is about estimating the relationship between variables. Generally, a model function is assumed for the relationship, and what are estimated are coefficients or other model parameters. This is a parameter estimation problem. Linear regression assumes that some outcome is the weighted linear sum of functions of some independent variables. The functions themselves need not be linear functions of the independent variables, but the weighted sum does need to be a linear sum.

Entire books have been written about regression, with a text by Draper and Smith being one such example.[7] Additional texts that include discussions on regression include those by Johnson and Leone,[8] and Kendall and Stuart.[9]

4.1 Simple Linear Regression – Fitting to a Line

"Simple linear regression" assumes one independent variable. Here now we assume a very simple case, that of a relationship that can be expressed as a line, namely

$$y = y_0 + m\,x\,, \tag{44}$$

where

x = independent variable,
y = dependent variable,
m = slope of the line, and
y_0 = 'y' intercept. (45)

The slope and intercept define the line that relates y to x.

In a typical experiment, if we believe a linear model for the relationship, then we might choose a set of x values and measure corresponding y values, and from these ordered pairs attempt to estimate slope and intercept, noting that our y values will generally exhibit measurement errors, or noise. We will identify specific x and corresponding y values with subscripts, namely

x_i = specific independent variable, and
y_i = specific corresponding measured dependent variable, (46)

both for $i = 1, 2, 3, \ldots, N$.

With this set of ordered pairs, we wish to estimate a "best fit" line to the data, and derive the optimum m and y_0 values. Accordingly, we may set up the system of linear equations in matrix form as

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} \begin{bmatrix} y_0 \\ m \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}. \tag{47}$$

The system of equations is clearly over-constrained, but we want this because of measurement errors in the y variable. However, we also recognize Eq. (47) to be of the form

$$\mathbf{A}\mathbf{x} = \mathbf{b}\,. \tag{48}$$

where in this case the matrix and vector quantities are

$$\mathbf{A} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix},$$

$$\mathbf{b} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \text{ and}$$

$$\mathbf{x} = \begin{bmatrix} y_0 \\ m \end{bmatrix}. \tag{49}$$

It is not unreasonable to expect that the uncertainty in y_i is statistically equal for all measured values. Consequently, using the results of the previous section, we may calculate the uniformly-weighted (a.k.a. unweighted) least-squares fit of a line to the data as

$$\begin{bmatrix} y_0 \\ m \end{bmatrix} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}\,. \tag{50}$$

We note that

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} N & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i^2 \end{bmatrix}. \tag{51}$$

The inverse of Eq. (51) can be calculated to be

$$\left(\mathbf{A}^T\mathbf{A}\right)^{-1} = \frac{1}{\left(N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2\right)}\begin{bmatrix} \sum_{i=1}^{N} x_i^2 & -\sum_{i=1}^{N} x_i \\ -\sum_{i=1}^{N} x_i & N \end{bmatrix}. \tag{52}$$

This allows us to expand Eq. (50) to

$$\begin{bmatrix} y_0 \\ m \end{bmatrix} = \frac{1}{\left(N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2\right)}\begin{bmatrix} \left(\sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} x_i y_i\right) \\ \left(-\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i + N\sum_{i=1}^{N} x_i y_i\right) \end{bmatrix}. \tag{53}$$

Specifically, we identify the components to be

$$y_0 = \frac{\left(\sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} x_i y_i\right)}{\left(N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2\right)}, \text{ and}$$

$$m = \frac{\left(-\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i + N\sum_{i=1}^{N} x_i y_i\right)}{\left(N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2\right)}. \tag{54}$$

Given that the errors in the various y_i measurements are independent, but identically distributed with zero mean, we may define the covariance matrix

$$\mathbf{\Sigma_b} = E\left\{\mathbf{b}\mathbf{b}^T\right\} = \sigma_y^2\mathbf{I}. \tag{55}$$

The covariance matrix for the estimates of line parameters then is calculated as

$$\mathbf{\Sigma_x} = \begin{bmatrix} \sigma_{y_0}^2 & \sigma_{y_0,m} \\ \sigma_{y_0,m} & \sigma_m^2 \end{bmatrix} = \left(\mathbf{A}^T\mathbf{\Sigma_b}^{-1}\mathbf{A}\right)^{-1} = \sigma_y^2\left(\mathbf{A}^T\mathbf{A}\right)^{-1}. \tag{56}$$

This result indicates the "goodness" of the line fit to the data.

Using Eq. (52). , we may now expand the calculation of the covariance of the estimates to become

$$\boldsymbol{\Sigma}_{\mathbf{x}} = \begin{bmatrix} \sigma_{y_0}^2 & \sigma_{y_0,m} \\ \sigma_{y_0,m} & \sigma_m^2 \end{bmatrix} = \frac{\sigma_y^2}{\left(N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i \right)^2 \right)} \begin{bmatrix} \sum_{i=1}^{N} x_i^2 & -\sum_{i=1}^{N} x_i \\ -\sum_{i=1}^{N} x_i & N \end{bmatrix}. \tag{57}$$

From these, individual variances can be extracted. For example, the variance in the estimate of the slope becomes

$$\sigma_m^2 = \sigma_y^2 \frac{N}{\left(N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i \right)^2 \right)}, \tag{58}$$

and the variance in the estimate of the intercept becomes

$$\sigma_{y_0}^2 = \sigma_y^2 \frac{\sum_{i=1}^{N} x_i^2}{\left(N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i \right)^2 \right)}. \tag{59}$$

Note that these variances decrease as the number of independent samples *N* increases, ***and*** as the spread in x_i increases. This provides important intuition in setting up these types of problems.

This analysis is easily extended to employing a weighted least squares solution.

4.2 Simple Linear Regression – Fitting to a Polynomial

Linear regression analysis does not need to presume a linear relationship to the independent variable. Indeed, it is frequently used for fitting polynomials to data. This is sometimes called "Polynomial Regression."

Consider the model of a polynomial of order *J*, namely

$$y = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 \ldots + \alpha_J x^J , \tag{60}$$

where

x = the independent variable,
y = dependent variable, and
α_j = the j^{th} coefficient of the polynomial. (61)

We wish to estimate the best fit coefficients for a specific set of defined x and corresponding y values, which we again identify with subscripts as

x_i = specific independent variable, and
y_i = specific corresponding measured dependent variable, (62)

both of these for $i = 1, 2, 3, \ldots, N$.

This fits into our least-squares structure

$$\mathbf{A}\mathbf{x} = \mathbf{b} , \tag{63}$$

where in this case the matrix and vector quantities are

$$\mathbf{A} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^J \\ 1 & x_2 & x_2^2 & \cdots & x_2^J \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & x_N^2 & \cdots & x_N^J \end{bmatrix},$$

$$\mathbf{b} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \text{ and}$$

$$\mathbf{x} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_J \end{bmatrix}. \tag{64}$$

We will again assume that the uncertainty in y_i is statistically equal for all measured values. However, we might have other reasons for wanting a "better" match for some x_i than others. Consequently, using the results of the previous section for a weighted least squares solution, we may calculate the weighted least-squares fit of a set of polynomial coefficients to the data as

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_J \end{bmatrix} = \left(\mathbf{A}^T\mathbf{W}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{W}\mathbf{b}\,. \tag{65}$$

Given that the errors in the various y_i measurements are independent, but identically distributed with zero mean, we may again define the covariance matrix

$$\mathbf{\Sigma_b} = E\left\{\mathbf{b}\mathbf{b}^T\right\} = \sigma_y^2\mathbf{I}\,. \tag{66}$$

The covariance matrix for the estimates of the coefficients in Eq. (65) is calculated as

$$\mathbf{\Sigma_x} = \sigma_y^2\left(\mathbf{A}^T\mathbf{W}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{W}^T\mathbf{W}\mathbf{A}\left(\mathbf{A}^T\mathbf{W}\mathbf{A}\right)^{-1}. \tag{67}$$

If uniform weighting, where $\mathbf{W} = \mathbf{I}$, then this reduces to

$$\mathbf{\Sigma_x} = \sigma_y^2\left(\mathbf{A}^T\mathbf{A}\right)^{-1}. \tag{68}$$

We offer some summary comments.

- We can reasonably well fit a polynomial approximation to data using a weighted least-squares regression.
- For a polynomial of order *N*, we need at least *N*+1 independent samples of the function and its inputs. However, more is better to help accommodate outliers and noise in the data.
- Weights can be employed to force the polynomial to be more accurate for certain samples at the expense of other samples.

4.3 Simple Linear Regression – Fitting to a Sum of Functions

Simple linear regression merely requires the linear sum of functions of a single independent variable. So, we may presume a general model for the data as

$$y = \alpha_0 f_0(x) + \alpha_1 f_1(x) + \alpha_2 f_2(x) + \alpha_3 f_3(x) \ldots + \alpha_J f_J(x), \tag{69}$$

where

x = the independent variable,
y = dependent variable,
$f_j(x)$ = the j^{th} function of x, and
α_j = the coefficient of the j^{th} function in the linear sum. (70)

Our task is to estimate the best fit coefficients for a specific set of defined x and corresponding y values, which we again identify with subscripts as

x_i = specific independent variable, and
y_i = specific corresponding measured dependent variable, (71)

both of these for $i = 1, 2, 3, \ldots, N$.

This fits into our least-squares structure

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{72}$$

where in this case the matrix and vector quantities are

$$\mathbf{A} = \begin{bmatrix} f_0(x_1) & f_1(x_1) & f_2(x_1) & \cdots & f_J(x_1) \\ f_0(x_2) & f_1(x_2) & f_2(x_2) & \cdots & f_J(x_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_0(x_N) & f_1(x_N) & f_2(x_N) & \cdots & f_J(x_N) \end{bmatrix},$$

$$\mathbf{b} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \text{ and}$$

$$\mathbf{x} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_J \end{bmatrix}. \tag{73}$$

The weighted least squares solution can be calculated as

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_J \end{bmatrix} = \left(\mathbf{A}^T\mathbf{W}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{W}\mathbf{b}\,. \tag{74}$$

Given that the errors in the various y_i measurements are independent, but identically distributed with zero mean, we may again define the covariance matrix

$$\mathbf{\Sigma_b} = E\left\{\mathbf{b}\mathbf{b}^T\right\} = \sigma_y^2\mathbf{I}\,. \tag{75}$$

The covariance matrix for the estimates of the coefficients in Eq. (74) is calculated as

$$\mathbf{\Sigma_x} = \sigma_y^2\left(\mathbf{A}^T\mathbf{W}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{W}^T\mathbf{W}\mathbf{A}\left(\mathbf{A}^T\mathbf{W}\mathbf{A}\right)^{-1}. \tag{76}$$

If uniform weighting, where $\mathbf{W} = \mathbf{I}$, then this reduces to

$$\mathbf{\Sigma_x} = \sigma_y^2\left(\mathbf{A}^T\mathbf{A}\right)^{-1}. \tag{77}$$

4.4 Multi-Linear Regression – Fitting to a Hyperplane

Multi-linear regression assumes more than one independent variable. A simple case is a relationship that can be expressed as one of a hyperplane, namely

$$y = y_0 + m_1\, x_1 + m_2\, x_2 + m_2\, x_3 + \ldots, \tag{78}$$

where

x_j = the j^{th} independent variable, for $1 \le j \le J$,
y = dependent variable,
m_j = slope associated with x_j, and
y_0 = 'y' intercept. (79)

In two dimensions, a hyperplane equates to a one-dimensional line. In three dimensions, a hyperplane equates to a two-dimensional flat plane.

In a typical experiment, we will choose a set of x_j and measure a corresponding y value, and repeat this N times. We will again define specific x_j and corresponding y values with subscripts, namely

$x_{j,i}$ = the i^{th} value for the j^{th} independent variable, and
y_i = specific corresponding i^{th} measured dependent variable, (80)

all for $i = 1, 2, 3, \ldots, N$.

This fits into our least-squares structure

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{81}$$

where in this case the matrix and vector quantities are

$$\mathbf{A} = \begin{bmatrix} 1 & x_{1,1} & x_{2,1} & \cdots & x_{J,1} \\ 1 & x_{1,2} & x_{2,2} & \cdots & x_{J,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,N} & x_{2,N} & \cdots & x_{J,N} \end{bmatrix},$$

$$\mathbf{b} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \text{ and}$$

$$\mathbf{x} = \begin{bmatrix} y_0 \\ m_1 \\ m_2 \\ \vdots \\ m_J \end{bmatrix}. \tag{82}$$

We will again assume that the uncertainty in y_i is statistically equal for all measured values. Consequently, using the results of the previous section, we may calculate the uniformly-weighted (a.k.a. unweighted) least-squares fit of a hyperplane to the data as

$$\begin{bmatrix} y_0 \\ m_1 \\ m_2 \\ \vdots \\ m_J \end{bmatrix} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}\,. \tag{83}$$

Given that the errors in the various y_i measurements are independent, but identically distributed with zero mean, we may again define the covariance matrix

$$\boldsymbol{\Sigma}_{\mathbf{b}} = E\left\{\mathbf{b}\mathbf{b}^T\right\} = \sigma_y^2\mathbf{I}\,. \tag{84}$$

The covariance matrix for the estimates of hyperplane parameters in Eq. (83) is again calculated as

$$\boldsymbol{\Sigma}_{\mathbf{x}} = \left(\mathbf{A}^T{\boldsymbol{\Sigma}_{\mathbf{b}}}^{-1}\mathbf{A}\right)^{-1} = \sigma_y^2\left(\mathbf{A}^T\mathbf{A}\right)^{-1}. \tag{85}$$

This result indicates the "goodness" of the hyperplane fit to the data.

4.5 Multi-Linear Regression – Fitting to a Sum of Functions

Multi-linear regression merely requires the linear sum of functions of perhaps multiple independent variables. So, we may presume a general model for the data as

$$y = \alpha_0 f_0(x_1,\ldots,x_K) + \alpha_1 f_1(x_1,\ldots,x_K) + \alpha_2 f_2(x_1,\ldots,x_K) + \ldots + \alpha_J f_J(x_1,\ldots,x_K), \quad (86)$$

where

$$\begin{aligned} &x_k \text{ = the independent variables, for } 1 \le k \le K, \\ &y \text{ = dependent variable,} \\ &f_j(x_1,\ldots,x_K) \text{ = the } j^{\text{th}} \text{ function of } x_1,\ldots,x_K, \text{ for } 0 \le j \le J, \text{ and} \\ &\alpha_j \text{ = the coefficient of the } j^{\text{th}} \text{ function in the linear sum.} \end{aligned} \quad (87)$$

In a typical experiment, we will choose a set of x_k and measure a corresponding y value, and repeat this N times. We will again define specific x_k and corresponding y values with subscripts, namely

$$\begin{aligned} &x_{k,i} \text{ = the } i^{\text{th}} \text{ value for the } k^{\text{th}} \text{ independent variable, and} \\ &y_i \text{ = specific corresponding } i^{\text{th}} \text{ measured dependent variable,} \end{aligned} \quad (88)$$

all for $i = 1,2,3,\ldots,N$.

This fits into our least-squares structure

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \quad (89)$$

where in this case the matrix and vector quantities are

$$\mathbf{A} = \begin{bmatrix} f_0(x_{1,1},\ldots,x_{K,1}) & f_1(x_{1,1},\ldots,x_{K,1}) & f_2(x_{1,1},\ldots,x_{K,1}) & \cdots & f_J(x_{1,1},\ldots,x_{K,1}) \\ f_0(x_{1,2},\ldots,x_{K,2}) & f_1(x_{1,2},\ldots,x_{K,2}) & f_2(x_{1,2},\ldots,x_{K,2}) & \cdots & f_J(x_{1,2},\ldots,x_{K,2}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_0(x_{1,N},\ldots,x_{K,N}) & f_1(x_{1,N},\ldots,x_{K,N}) & f_2(x_{1,N},\ldots,x_{K,N}) & \cdots & f_J(x_{1,N},\ldots,x_{K,N}) \end{bmatrix},$$

$$\mathbf{b} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \text{ and}$$

$$\mathbf{x} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_J \end{bmatrix}. \tag{90}$$

The weighted least squares solution can be calculated as

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_J \end{bmatrix} = \left(\mathbf{A}^T\mathbf{W}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{W}\mathbf{b}\,. \tag{91}$$

Given that the errors in the various y_i measurements are independent, but identically distributed with zero mean, we may again define the covariance matrix

$$\mathbf{\Sigma_b} = E\left\{\mathbf{b}\mathbf{b}^T\right\} = \sigma_y^2\mathbf{I}\,. \tag{92}$$

The covariance matrix for the estimates of the coefficients in Eq. (74) is calculated as

$$\mathbf{\Sigma_x} = \sigma_y^2\left(\mathbf{A}^T\mathbf{W}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{W}^T\mathbf{W}\mathbf{A}\left(\mathbf{A}^T\mathbf{W}\mathbf{A}\right)^{-1}. \tag{93}$$

If uniform weighting, where $\mathbf{W} = \mathbf{I}$, then this reduces to

$$\mathbf{\Sigma_x} = \sigma_y^2\left(\mathbf{A}^T\mathbf{A}\right)^{-1}. \tag{94}$$

4.6 Comments

Several comments in no particular order.

- The models of Eq. (44), Eq. (60), Eq. (69), and Eq. (78), are all merely special cases of Eq. (86).
- Choosing the underlying functions $f_j(x_1,\ldots,x_K)$ for which coefficients are calculated generally requires some insight or intelligent guess as to a reasonably good model for the underlying behavior.

How Many Samples?

So, given that we need to find *J* coefficients, how many *N* independent measurements should we collect? We readily identify some basic truths.

1. We minimally need $N \geq J$ independent samples.
2. More *N* is better, all else equal. (Note the qualifier.)

After this, it gets a little more complicated, because other things matter, too.

Ultimately, we want enough of the right kind of data to achieve an acceptable covariance matrix $\boldsymbol{\Sigma}_{\mathbf{x}}$ for our coefficient estimation. We emphasize that it is not merely selecting the right number *N*, but also the nature of the data set.

For example, in the relatively simple example of determining the slope of a line fitted to some data, recall that the overall covariance of the linear regression result was given by Eq. (56), and specifically the variance of the slope was given by Eq. (58). Now consider two cases

> Case 1. Consider the case where $\sigma_y^2 = 1$ and *N*=100, but samples of x_i are uniformly distributed over the interval [0,1]. The variance of the estimate of the slope is approximately 0.1176.
>
> Case 2. Now consider the case where $\sigma_y^2 = 1$ and *N*=2, which is the minimum number of samples needed, but samples of x_i are at the endpoints of the wider interval [0,100]. The variance of the estimate of the slope is now approximately 0.0002. This indicates a more accurate result than Case 1, in spite of the substantially fewer number of samples.

The point of this exercise is to illustrate that sometimes 'better' data is more useful than 'more' data. The ultimate measure of goodness is the variance of the parameter we wish to estimate.

Without elaboration, we further state that when multiple parameters are being estimated, by selecting the right data set, we can often trade the goodness of one parameter's estimate for the goodness of another parameter's estimate.

"The purpose of computing is insight, not numbers."
-- Richard Hamming

5 Radar Applications

Frequently in radar applications, given a signal in Additive White Gaussian Noise (AWGN), the question arises "how well can we estimate phase, frequency, time-delay, or some other parameter?" Texts that address this question include one by McDonough and Whalen.[10]

We now address several related problems. However, now that we move into radar applications, we are compelled to reuse some variable names from earlier in this report to represent some of the new parameters with which we will deal. We apologize for any confusion, and hope that context will help the reader to sort them out.

Nevertheless, the following examples are somewhat simple ones, mainly meant to illustrate the breadth and utility of regression analysis in radar parameter estimation.

5.1 Phase Variance

Consider a sinusoidal signal in AWGN. The signal space diagram is in Figure 1.

We define some signals as follows.

$$\begin{aligned} s &= Ae^{j\theta} = \text{"truth" signal, and} \\ \eta &= \text{complex zero-mean AWGN.} \end{aligned} \tag{95}$$

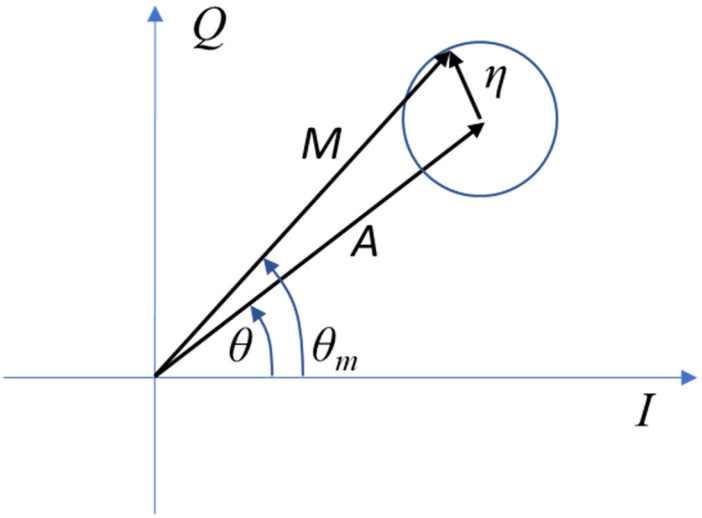

Figure 1. Signal space diagram for signal in noise.

We stipulate that η has uncorrelated In-Phase (I) and Quadrature (Q) components, with

$$\sigma_\eta^2 = \text{variance of orthogonal components of } \eta. \tag{96}$$

Our measured signal is identified as follows

$$s_m = A_m e^{j\theta_m} = Ae^{j\theta} + \eta \ = \text{“measured” signal with noise,} \tag{97}$$

where

$$\begin{aligned} A_m &= \left|s_m\right| \text{ = magnitude of measured signal with noise,} \\ \theta_m &= \arg\left(s_m\right) \text{ = phase of measured signal with noise.} \end{aligned} \tag{98}$$

The idea is to estimate the true signal's θ based on measurements of s_m. We will assume that $A_m \approx A$ is a reasonable assumption. This is effectively assuming reasonably good Signal-to-Noise Ratio (SNR).

We identify the SNR of a single measurement s_m as

$$SNR_1 = \frac{A^2}{2\sigma_\eta^2} \ . \tag{99}$$

Due to circular symmetry, for good SNR, the variance of the phase measurement of s_m is

$$\sigma_{\theta_m}^2 = E\left\{\left|\theta_m - \theta\right|^2\right\} = \frac{\sigma_\eta^2}{A^2} = \frac{1}{2\ SNR_1} \ . \tag{100}$$

This is the expected squared error of the phase measurement with respect to the true signal phase.

<u>Multiple Signal Measurements</u>

Let us make *N* measurements of signal+noise, each with an independent noise sample, and create the weighted linear sum

$$z = \sum_{i=1}^{N} a_i \left(Ae^{j\theta} + \eta_i \right) \ . \tag{101}$$

The new SNR of the linear sum of signals is calculated as

$$SNR_\Sigma = \left(\frac{A^2}{2\sigma_\eta^2} \right) \frac{\left(\sum_{i=1}^{N} a_i \right)^2}{\sum_{i=1}^{N} a_i^2} = SNR_1 \frac{\left(\sum_{i=1}^{N} a_i \right)^2}{\sum_{i=1}^{N} a_i^2} \ . \tag{102}$$

Thus, SNR has been improved by $\left(\sum_{i=1}^{N} a_i\right)^2 \Big/ \sum_{i=1}^{N} a_i^2$. For uniform weighting, this equates to an improvement by a factor *N*.

Recall from Eq. (100) that a phase measurement from a single signal+noise measurement exhibits a variance of $\sigma_{\theta_m}^2 = \left(2\ SNR_1\right)^{-1}$. The phase of the linear weighted sum of multiple signal+noise measurements then has variance

$$\sigma_{\theta_\Sigma}^2 = \frac{1}{2\ SNR_\Sigma} = \frac{1}{2\ SNR_1 \left(\left(\sum_{i=1}^{N} a_i\right)^2 \Big/ \sum_{i=1}^{N} a_i^2\right)} \ . \tag{103}$$

For uniform weighting, this becomes

$$\sigma_{\theta_\Sigma}^2 = \frac{1}{2\ SNR_\Sigma} = \frac{1}{N\left(2\ SNR_1\right)} \ . \tag{104}$$

<u>Multiple Phase Measurements</u>

A weighted linear sum of phase measurements themselves can be described as

$$\theta_\Sigma = \sum_{i=1}^{N} a_i \theta_{m,i} \ , \tag{105}$$

where

$$\theta_{m,i} \ = \text{the phase measurement of the } i^{\text{th}} \text{ measurement.} \tag{106}$$

We will assume no ambiguities in phase, that is, all phases are unwrapped and phase aliasing has been mitigated. Since the phase measurement errors are presumed to be independent, but identically distributed,

$$\sigma_{\theta_\Sigma}^2 = \sigma_{\theta_m}^2 \sum_{i=1}^{N} a_i^2 = \frac{\left(\sum_{i=1}^{N} a_i^2\right)}{2\ SNR_1} \ . \tag{107}$$

Note that for a phase difference, where $a_1 = 1$ and $a_2 = -1$, we calculate $\sigma_{\theta_\Sigma}^2 = \left(SNR_1\right)^{-1}$.

If we set $a_i = 1/N$, for a uniformly weighted average, then $\sigma_{\theta_\Sigma}^2 = \frac{1}{N\left(2SNR_1\right)}$, which is identical to Eq. (104).

5.2 Frequency Estimation

Frequency is defined as the time-derivative of phase. Here we examine the accuracy of a frequency estimation based on phase measurements.

5.2.1 Two Phase Measurements

For a constant-frequency sinusoid, we are given that

$$\omega = \frac{d\theta}{dt} = \left.\frac{\theta_2 - \theta_1}{T}\right|_{T\to 0}, \tag{108}$$

where

$$\begin{aligned} &\theta_1 \text{ = phase at the beginning of the observation interval,} \\ &\theta_2 \text{ = phase at the end of the observation interval,} \\ &T \text{ = observation time interval.} \end{aligned} \tag{109}$$

Of course, if the frequency is indeed constant, then any non-zero T will do as well, aliasing issues notwithstanding. This is the unwrapped phase assumption mentioned earlier.

Accordingly, we identify the "measured" frequency based on two phase measurements as

$$\omega_m = \frac{\theta_{m,2} - \theta_{m,1}}{T} \text{ = "measured" radian frequency,} \tag{110}$$

Consequently, we may calculate the variance of the frequency measurement as

$$\sigma_{\omega_m}^2 = \frac{1}{T^2} 2\sigma_{\theta_m}^2 = \frac{1}{T^2 SNR_1}. \tag{111}$$

We might further note that nominal frequency resolution may be defined as

$$\rho_{\omega_m} = \frac{2\pi}{T}. \tag{112}$$

The variance of the frequency measurement can then be equated to

$$\sigma_{\omega_m}^2 = \frac{\rho_{\omega_m}^2}{(2\pi)^2 SNR_1}. \tag{113}$$

Nevertheless, SNR here is for a single phase measurement, and this equation is for a single phase difference. Did we mention that aliasing issues are notwithstanding?

Comments

Eq. (111) will often be written in textbooks as the equation

$$\sigma_{\omega_m}^2 = \frac{1}{T^2 2E_r/N_0}, \tag{114}$$

where

$$\begin{aligned} &E_r \text{ = Energy in the received signal, and} \\ &N_0/2 \text{ = two-sided power spectral density of the white noise.} \end{aligned} \tag{115}$$

An example is the text by McDonough & Whalen.[10]

We note that a signal of duration τ will have a two-sided nominal bandwidth

$$\beta = 1/\tau, \tag{116}$$

Consequently, SNR is equated to

$$SNR = \frac{E_r/\tau}{(N_0/2)\beta}, \tag{117}$$

where the numerator is the signal power in a single measurement, and the denominator is the noise power in that measurement. This may be simplified to

$$SNR = \frac{E_r}{(N_0/2)} = 2E_r/N_0, \tag{118}$$

allowing the variance of the frequency measurement to be

$$\sigma_{\omega_m}^2 = \frac{1}{T^2 SNR_1}. \tag{119}$$

This, of course, is identical to Eq. (111).

5.2.2 Multiple Phase Measurements

While the preceding development is for two phase measurements, we now examine the case of N equally spaced phase measurements. This will require a linear regression analysis.

We will use N phase measurements, where we equate the phase measurement to the model as

$$\theta_{m,i} = \theta_0 + \omega_m t_i, \text{ for } 1 \le i \le N. \tag{120}$$

Eq. (120) is just the equation of a line, as was explored in section 4.1, making linear regression appropriate. For equally spaced measurements in time, we set

$$t_i = \Delta_t i \,, \tag{121}$$

where

$$\Delta_t \text{ = sample spacing in time.} \tag{122}$$

Note that the set of time delays is $\Delta_t i$ for $1 \le i \le N$. We calculate the overall interval to be

$$T = \left(N-1\right)\Delta_t \,. \tag{123}$$

Accordingly, we identify the matrix and vector quantities as

$$\mathbf{A} = \begin{bmatrix} 1 & \Delta_t \\ 1 & 2\Delta_t \\ \vdots & \vdots \\ 1 & N\Delta_t \end{bmatrix},$$

$$\mathbf{b} = \begin{bmatrix} \theta_{m,1} \\ \theta_{m,2} \\ \vdots \\ \theta_{m,N} \end{bmatrix}, \text{ and}$$

$$\mathbf{x} = \begin{bmatrix} \theta_0 \\ \omega_m \end{bmatrix}. \tag{124}$$

For uniform weighting, the best fit parameters are calculated as

$$\begin{bmatrix} \theta_0 \\ \omega_m \end{bmatrix} = \left(\mathbf{A}^T \mathbf{A}\right)^{-1} \mathbf{A}^T \mathbf{b} \,. \tag{125}$$

Recall from Eq. (100) that the variance in the phase measurements is

$$\sigma_{\theta_m}^2 = \frac{1}{2\ SNR_1} \,. \tag{126}$$

From section 4.1, the variance in the frequency measurement is then

$$\sigma_{\omega_m}^2 = \sigma_{\theta_m}^2 \frac{N}{\left(N\sum_{i=1}^{N}(\Delta_t i)^2 - \left(\sum_{i=1}^{N}(\Delta_t i)\right)^2 \right)}. \tag{127}$$

The series in the denominator may be evaluated so that Eq. (127) can be expanded and simplified to

$$\sigma_{\omega_m}^2 = \sigma_{\theta_m}^2 \frac{12}{\Delta_t^2 N\left(N^2-1\right)}. \tag{128}$$

This is essentially the same result reported by Rife and Boorstyn.[11]

Incorporating Eq. (123) into Eq. (128) yields

$$\sigma_{\omega_m}^2 = \frac{12\left(N-1\right)}{T^2 N\left(N+1\right)}\sigma_{\theta_m}^2. \tag{129}$$

Expanding with Eq. (126) yields

$$\sigma_{\omega_m}^2 = \frac{6\left(N-1\right)}{T^2 N\left(N+1\right)}\left(\frac{1}{SNR_1}\right). \tag{130}$$

For $N = 2$, this calculates to Eq. (111), which is comforting. However, as *N* increases, this variance does in fact decrease. Note that for large *N*, this approaches

$$\sigma_{\omega_m}^2 \to \frac{6}{T^2 N}\left(\frac{1}{SNR_1}\right). \tag{131}$$

So, the longer the observation time interval, the lesser is the variance in our frequency measurement. The same is true for higher SNR per sample, or more independent measurement samples.

We stipulate that these results are for equally spaced sample times of the phase. If sample times are spaced unequally, then we would need to back up to a more general linear regression calculation.

5.2.3 Relationship to Fourier Analysis

The customary way to determine frequency is to employ a Fourier analysis of the data, typically calculating a Fourier Transform of the data, for sampled data being a Discrete-Time Fourier Transform (DFT). Here we now examine the relationship of the previous analysis to the results of a DFT.

Recall that samples of the noise-free truth signal are described by the constant modulus function

$$s_i = Ae^{j\theta_i} \text{ = "truth" signal,} \tag{132}$$

where we now allow θ_i to explicitly vary with sample number, for samples $1 \le i \le N$.

Recall that our measured signals were defined as

$$s_{m,i} = A_{m,i}e^{j\theta_{m,i}} = s_i + \eta_i \text{ = "measured" signal containing noise,} \tag{133}$$

where

$$\eta_i \text{ = complex noise values at sample times.} \tag{134}$$

A matched filter will attempt to select the correct reference signal function to maximize the magnitude of the response

$$S_r = \sum_{i=1}^{N} s_{m,i}s_{r,i}^{*}, \tag{135}$$

where the "*" denotes complex conjugate, and

$$s_{r,i} = e^{j\theta_{r,i}} \text{ = the reference signal that we wish to match.} \tag{136}$$

Consequently, we may expand Eq. (135) to

$$S_r = \sum_{i=1}^{N} s_{m,i}e^{-j\theta_{r,i}}. \tag{137}$$

If we define the reference phase function as a time-varying phase, such that

$$\theta_{r,i} = \omega_r t_i \text{ = reference phase for } i^{\text{th}} \text{ sample,} \tag{138}$$

where

ω_r = constant reference frequency, and
t_i = sample times, (139)

then Eq. (137) becomes

$$S_r = \sum_{i=1}^{N} s_{m,i} e^{-j\omega_r t_i} \ . \tag{140}$$

This is, in fact, the DFT of the measured input signal with sample values $s_{m,i}$, albeit it is usually written with constant increments in time, with $t_i = \Delta_t i$, and evaluated with ω_r chosen at constant increments in frequency.

We now stipulate that the true signal is a constant-modulus constant-frequency complex sinusoid, with phase function as a time-varying phase, such that

$\theta_i = \theta_0 + \omega_0 t_i$ = true signal phase for i^{th} sample, (141)

where

θ_0 = true signal reference phase, and
ω_0 = true signal constant frequency. (142)

In addition, we model the phase of the measured signal in Eq. (133) as

$\theta_{m,i} = \theta_m + \omega_m t_i + \theta_{\varepsilon,i}$ = phase of measured signal for i^{th} sample, (143)

where

θ_m = best-fit reference phase to the phase measurements,
ω_m = best-fit linear phase increment (frequency) to the phase measurements, and
$\theta_{\varepsilon,i}$ = residual error in the measured phase for each sample. (144)

The nature of $\theta_{\varepsilon,i}$ is that it is devoid of constant and linear terms with respect to t_i, and thus contains only higher-order terms which we will assume to be small, in fact negligible.

We now define the deviation of the measured phase from the true phase as

$$\theta_{m,i} - \theta_i \approx (\theta_m - \theta_0) + (\omega_m - \omega_0) t_i = \theta_\Delta + \omega_\Delta t_i \ , \tag{145}$$

where

θ_Δ = error in best-fit reference phase, and
ω_Δ = error in best-fit frequency. (146)

These errors by construct have zero mean, otherwise they wouldn't be "best-fit," i.e., an unbiased estimator.

For good SNR, we will again assume $A_{m,i} \approx A$. Combining all these results and assumptions allows us to write Eq. (140) as

$$S_r = A\sum_{i=1}^{N} e^{j\left((\theta_0+\theta_\Delta)+(\omega_0+\omega_\Delta)t_i\right)}e^{-j\omega_r t_i} = Ae^{j(\theta_0+\theta_\Delta)}\sum_{i=1}^{N} e^{j(\omega_0+\omega_\Delta-\omega_r)t_i}\,. \tag{147}$$

Neglecting any aliasing, this summation is maximum when $\omega_r = \omega_0 + \omega_\Delta$. Note that if $t_i = \Delta_t i$, then this equates to

$$\begin{aligned} S_r &= Ae^{j(\theta_0+\theta_\Delta)}\sum_{i=1}^{N} e^{j(\omega_0+\omega_\Delta-\omega_r)\Delta_t i} \\ &= ANe^{j(\theta_0+\theta_\Delta)}e^{j\frac{(\omega_0+\omega_\Delta-\omega_r)\Delta_t(N+1)}{2}}\left(\frac{\sin\left(\frac{(\omega_0+\omega_\Delta-\omega_r)\Delta_t N}{2}\right)}{N\sin\left(\frac{(\omega_0+\omega_\Delta-\omega_r)\Delta_t}{2}\right)}\right). \end{aligned} \tag{148}$$

So, we have a peak value at $\omega_r = \omega_0 + \omega_\Delta$. Statistically, the expected peak is located at $\omega_r = \omega_0$, but ω_Δ will manifest as an error. The variance of this error is its expected squared error with respect to its mean, which is the same as the variance in ω_m itself. So

$$\sigma_{\omega_\Delta}^2 = \sigma_{\omega_m}^2\,, \tag{149}$$

where $\sigma_{\omega_m}^2$ is exactly that value calculated in Eq. (128).

So, what is the point of all this?

In summary, whether we calculate the DFT of samples of the signal itself and locate its peak, or we extract the phase of the samples and fit a line to the phase samples, we get the same answer for the estimated frequency, with the same variance of the frequency estimate.

5.3 Delay Estimation

Time is the dual of frequency, so the analysis of the preceding section can be applied here. It is essentially the same development, but with parameters' roles reversed.

We will use N phase measurements, where we now equate the phase measurement to the model with the expression

$$\theta_{m,i} = \theta_0 + \omega_i t_m, \text{ for } 1 \le i \le N. \tag{150}$$

The intent now is to use the measured phase as a function of frequency to estimate time delay.

Eq. (150) is again just the equation of a line, as was explored in section 4.1. For equally spaced measurements in frequency, we set

$$\omega_i = \Delta_\omega i, \tag{151}$$

where

$$\Delta_\omega = \text{sample spacing in frequency.} \tag{152}$$

Note that the set of frequencies is $\Delta_\omega i$ for $1 \le i \le N$. We calculate the overall frequency bandwidth to be

$$\Omega = (N-1)\Delta_\omega. \tag{153}$$

Accordingly, we identify the matrix and vector quantities as

$$\mathbf{A} = \begin{bmatrix} 1 & \Delta_\omega \\ 1 & 2\Delta_\omega \\ \vdots & \vdots \\ 1 & N\Delta_\omega \end{bmatrix},$$

$$\mathbf{b} = \begin{bmatrix} \theta_{m,1} \\ \theta_{m,2} \\ \vdots \\ \theta_{m,N} \end{bmatrix}, \text{ and}$$

$$\mathbf{x} = \begin{bmatrix} \theta_0 \\ t_m \end{bmatrix}. \tag{154}$$

For uniform weighting, the best fit parameters are calculated as

$$\begin{bmatrix} \theta_0 \\ t_m \end{bmatrix} = \left(\mathbf{A}^T \mathbf{A} \right)^{-1} \mathbf{A}^T \mathbf{b} \,. \tag{155}$$

Recall from Eq. (100) that the variance in the phase measurements is

$$\sigma_{\theta_m}^2 = \frac{1}{2\, SNR_1} \,. \tag{156}$$

From section 4.1, the variance in the time delay measurement is then

$$\sigma_{t_m}^2 = \sigma_{\theta_m}^2 \frac{N}{\left(N \sum_{i=1}^{N} \left(\Delta_\omega i \right)^2 - \left(\sum_{i=1}^{N} \left(\Delta_\omega i \right) \right)^2 \right)} \,. \tag{157}$$

The series in the denominator may be evaluated so that Eq. (157) can be expanded and simplified to the expression

$$\sigma_{t_m}^2 = \sigma_{\theta_m}^2 \frac{12}{\Delta_t^2 N \left(N^2 - 1 \right)} \,. \tag{158}$$

Incorporating Eq. (153) into Eq. (158) yields

$$\sigma_{t_m}^2 = \frac{12 \left(N - 1 \right)}{\Omega^2 N \left(N + 1 \right)} \sigma_{\theta_m}^2 \,. \tag{159}$$

Expanding with Eq. (156) yields

$$\sigma_{t_m}^2 = \frac{6 \left(N - 1 \right)}{\Omega^2 N \left(N + 1 \right)} \left(\frac{1}{SNR_1} \right) \,. \tag{160}$$

As N increases, this variance does in fact decrease. Note that for large N, this approaches

$$\sigma_{t_m}^2 \to \frac{6}{\Omega^2 N} \left(\frac{1}{SNR_1} \right) \,. \tag{161}$$

So, the wider the observation bandwidth, the lesser is the variance in our time delay measurement. The same is true for higher SNR per sample, or more independent measurement samples.

We stipulate that these results are for equally spaced sample frequencies of the phase. If sample frequencies are spaced unequally, then we would need to back up to a more general linear regression calculation.

5.4 DOA Estimation

Measuring the Direction of Arrival of a transmitted signal is often done by analyzing the phase of received signals across a dispersed set of receiving antenna elements. This is a rich area of analysis and can quickly explode into many different specialty scenarios and situations.

We note that the previous sections dealt with phase as a function of time or temporal frequency. However, we now generally deal with phase as a function of position, or distance. Strictly speaking, phase as a function of distance is termed "wavenumber," although "spatial frequency" is also often heard.

In this major section, we will exemplify DOA estimation with some rather simple studies, but nevertheless quite relevant to even state-of-the-art radar modes. We will leave more sophisticated DOA estimation for future reports.

5.4.1 Two-Element Interferometer

Consider the two-dimensional construct where linear Electro-Magnetic (EM) signal wavefronts are travelling in free space, across a region where we have two receiving antenna elements. This is illustrated in Figure 2. The DOA is the direction opposite the direction of propagation.

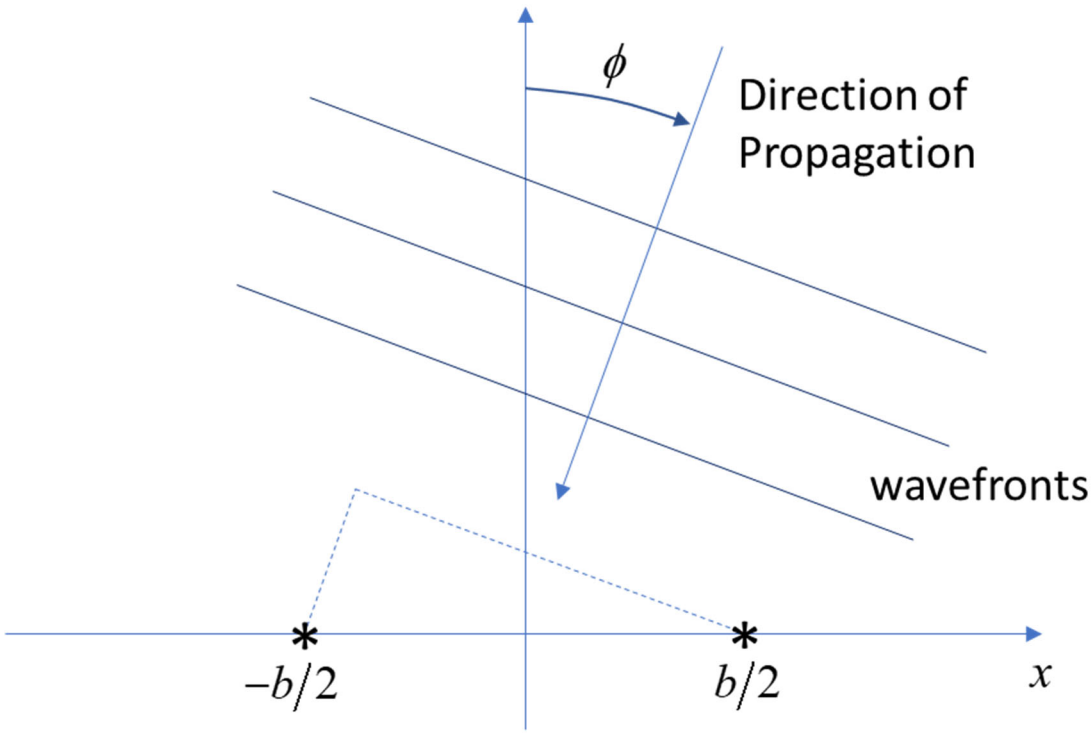

Figure 2. Two-element interferometer geometry.

In this geometry, we define

$\theta_{m,1}$ = measured phase at the location $x = -b/2$,
$\theta_{m,2}$ = measured phase at the location $x = b/2$, and
b = separation of the antenna phase centers, often called the "baseline." (162)

For a constant-frequency sinusoidal signal, we identify from the geometry that

$$\theta_{m,2} - \theta_{m,1} = -k\, b \sin\phi\,, \tag{163}$$

where

$k = \dfrac{2\pi}{\lambda}$ = wavenumber in the direction of propagation, where
λ = wavelength of the propagating signal. (164)

The DOA can then be calculated as

$$\sin\phi = -\frac{\left(\theta_{m,2} - \theta_{m,1}\right)}{k\, b}. \tag{165}$$

Using Eq. (100), we calculate the variance of the DOA as

$$\sigma_\phi^2 = \frac{1}{\left(k\, b\cos\phi\right)^2} 2\sigma_{\theta_m}^2 = \frac{1}{\left(k\, b\cos\phi\right)^2 SNR_1}. \tag{166}$$

Note that $b\cos\phi$ is merely the projection of the baseline separation in the direction of the signal source, i.e., the DOA.

Classic Interferometric SAR

The two-element interferometer is routinely used in Interferometric Synthetic Aperture Radar (IFSAR, or InSAR) to estimate land surface topography.[12]

We will examine a simple model for height measurement, illustrated in Figure 3. For simplicity, we will assume that the baseline normal is pointed to a point on the reference plane, which is not the same as the actual land surface level.

We now identify the following two points

P = point at range R on the reference plane, and
P' = point at range R on actual land surface. (167)

We define the spatial relationships between points P and P' as

d = 3-D distance between points P and P', and
h = vertical height of point P' above point P. (168)

For small d with respect to range R,

$$d \approx \frac{h}{\cos\psi} \approx R\sin\phi \tag{169}$$

Combining this with Eq. (165) yields

$$h \approx R\cos\psi\sin\phi \approx -\frac{R\cos\psi}{k\,b}\left(\theta_{m,2}-\theta_{m,1}\right) \tag{170}$$

Note that for small ϕ, we may assume $\cos\phi \approx 1$.

The variance in the height estimate for our geometry is then

$$\sigma_h^2 \approx \left(\frac{R\cos\psi}{k\,b}\right)^2 \frac{1}{SNR_1}. \tag{171}$$

This is the familiar expression given in multiple sources. Some factor-of-two improvements might be had by slightly altering the operating modality.[13]

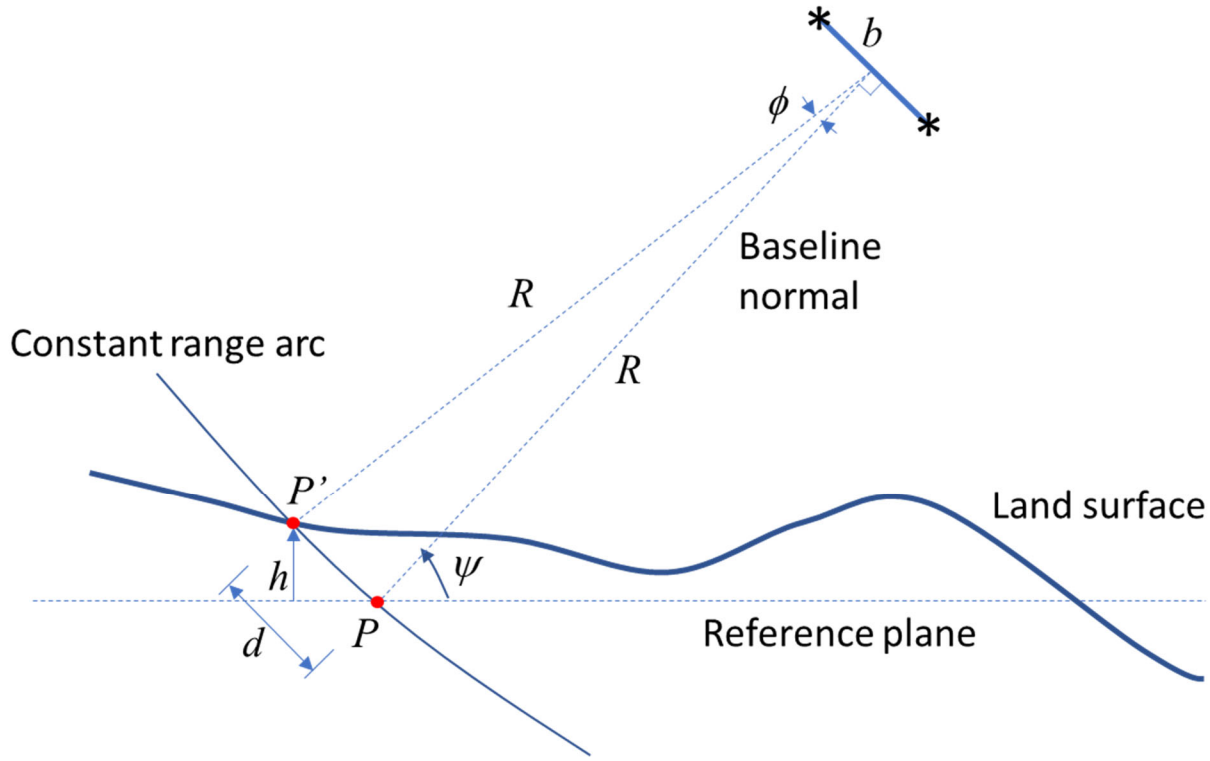

Figure 3. Interferometric SAR geometry.

5.4.2 Multi-Element DOA

Consider the two-dimensional construct where linear Electro-Magnetic (EM) signal wavefronts are travelling in free space, across a region where we now have multiple receiving antenna elements. This is illustrated in Figure 4. The DOA is still the direction opposite the direction of propagation.

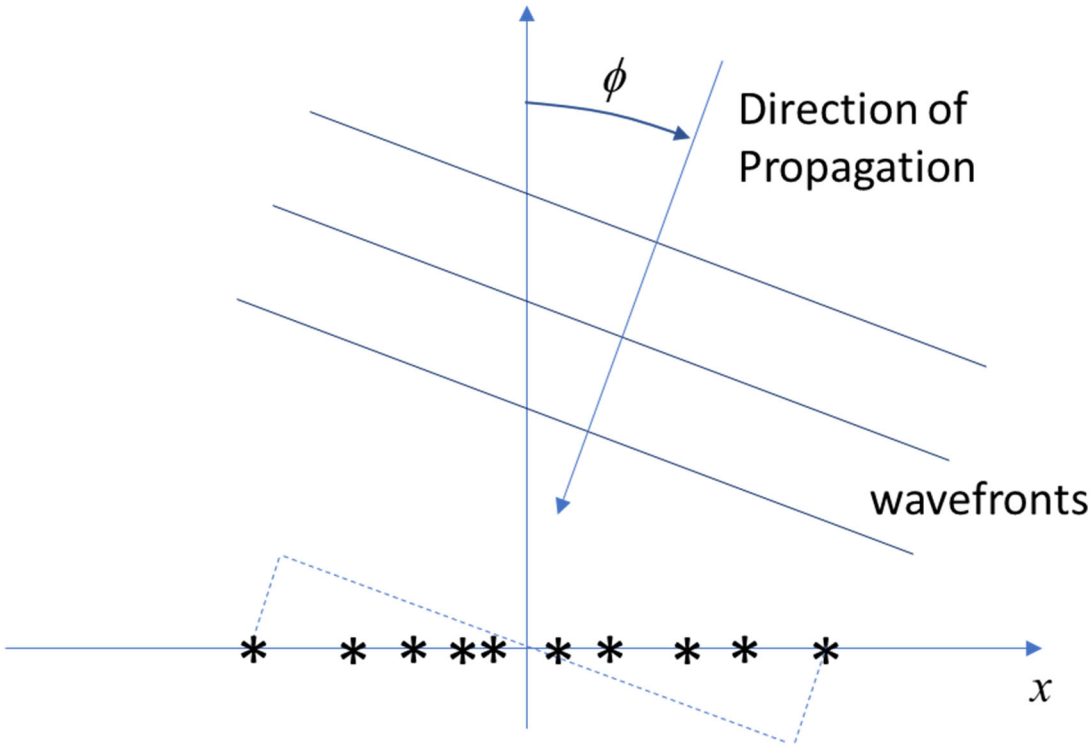

Figure 4. Multi-element receiving array. In general, receive element spacing need not be regular, but in practice often is.

In this geometry, the phase of the signal received at the i^{th} antenna element is modelled as

$$\theta_{m,i} = \theta_0 - x_i \, k \sin\phi_m \text{, for } 1 \le i \le N \tag{172}$$

where our task is to estimate the best fit direction of arrival ϕ_m, or equivalently the best fit function $\sin\phi_m$. The antenna element locations are specifically their phase centers.

This model makes linear regression appropriate. Although Figure 4 shows irregular receive element spacing, for our convenience we will subsequently assume equally spaced antenna elements along the x axis, and set

$$x_i = \Delta_x i \text{,} \tag{173}$$

where

Δ_x = sample spacing along the x axis. (174)

Note that the set of spatial locations is $\Delta_x i$ for $1 \le i \le N$, that is assumed to be limited to a finite overall span. We then calculate the overall span of antenna element phase centers to be

$$b = (N-1)\Delta_x . \tag{175}$$

The separation of the ends of this array in fact define the maximum baseline available to us.

Accordingly, we identify the matrix and vector quantities as

$$\mathbf{A} = \begin{bmatrix} 1 & -k\Delta_x \\ 1 & -2k\Delta_x \\ \vdots & \vdots \\ 1 & -Nk\Delta_x \end{bmatrix},$$

$$\mathbf{b} = \begin{bmatrix} \theta_{m,1} \\ \theta_{m,2} \\ \vdots \\ \theta_{m,N} \end{bmatrix}, \text{ and}$$

$$\mathbf{x} = \begin{bmatrix} \theta_0 \\ \sin\phi_m \end{bmatrix}. \tag{176}$$

For uniform weighting, the best fit parameters are calculated as

$$\begin{bmatrix} \theta_0 \\ \sin\phi_m \end{bmatrix} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b} . \tag{177}$$

Specifically, we identify

$$\sin\phi_m = \left(\frac{6}{k\Delta_x}\right)\frac{\left((N+1)\sum_{i=1}^{N}\theta_{m,i} - 2\sum_{i=1}^{N} i\theta_{m,i}\right)}{N\left(N^2-1\right)} \tag{178}$$

Note that for $N = 2$, where $\Delta_x = b$, this equates to Eq. (165).

Having identified the best fit $\sin\phi_m$, we can calculate the specific angle

$$\phi_m = \operatorname{asin}\left(\sin\phi_m\right). \tag{179}$$

Recall from Eq. (100) that the variance in the phase measurements is

$$\sigma_{\theta_m}^2 = \frac{1}{2\,SNR_1}\ . \tag{180}$$

From section 4.1, the variance in the $\sin\phi_m$ calculation is then

$$\sigma_{\sin\phi_m}^2 = \sigma_{\theta_m}^2 \frac{N}{\left(N\sum_{i=1}^{N}\left(k\Delta_x i\right)^2 - \left(\sum_{i=1}^{N}\left(k\Delta_x i\right)\right)^2 \right)}. \tag{181}$$

The series in the denominator may be evaluated so that Eq. (181) can be expanded and then simplified to

$$\sigma_{\sin\phi_m}^2 = \sigma_{\theta_m}^2 \frac{12}{k^2\Delta_x^2 N\left(N^2-1\right)}. \tag{182}$$

Incorporating Eq. (175) into Eq. (182) yields

$$\sigma_{\sin\phi_m}^2 = \frac{12\left(N-1\right)}{\left(kb\right)^2 N\left(N+1\right)}\sigma_{\theta_m}^2\ . \tag{183}$$

Expanding with Eq. (180) yields

$$\sigma_{\sin\phi_m}^2 = \frac{6\left(N-1\right)}{\left(kb\right)^2 N\left(N+1\right)}\left(\frac{1}{SNR_1}\right). \tag{184}$$

This is essentially the same result as Eq. (130) in section 5.2.2. If we want to calculate the variance of the DOA angle itself, then we observe that a small change in angle yields a small change in the sine of the angle as

$$\frac{d\left(\sin\phi_m\right)}{d\phi_m} = \cos\phi_m\ . \tag{185}$$

Consequently, we identify the relationship of variances as

$$\sigma_{\phi_m}^2 = \frac{1}{\cos^2\phi_m}\sigma_{\sin\phi_m}^2\ , \tag{186}$$

which is then expanded to

$$\sigma_{\phi_m}^2 = \frac{6\left(N-1\right)}{\left(kb\cos\phi_m\right)^2 N\left(N+1\right)}\left(\frac{1}{SNR_1}\right). \tag{187}$$

Note that for $N=2$ this equates to Eq. (166), a nice check.

Antenna Constraints

The development above, culminating in Eq. (187), suggests that if we keep adding independent antenna elements, then the DOA variance keeps decreasing at an asymptotic rate of $1/N$. This is a little misleading. In fact, for a finite overall length of the antenna, as we keep subdividing the antenna into smaller and smaller non-overlapping subapertures, then the antenna effective area will also diminish at a $1/N$ rate, which means the SNR of the received signal from an individual antenna element will also decrease at a $1/N$ rate.

So, we define the overall aperture of the summation of all equal and nonoverlapping subapertures to be

$$L_\Sigma = N\Delta_x \text{ = overall aperture length.} \tag{188}$$

Furthermore, the baseline span is related to the overall aperture length by

$$b = \frac{N-1}{N} L_\Sigma \ . \tag{189}$$

This also suggests that the SNR of the received signal from an individual antenna element that is of length $\Delta_x = L_\Sigma / N$, will behave as

$$SNR_1 = \frac{1}{N} SNR_\Sigma , \tag{190}$$

where

$$SNR_\Sigma \text{ = the SNR if the signals from all subapertures were coherently added.} \tag{191}$$

Putting these results into Eq. (187) yields

$$\sigma^2_{\phi_m} = \frac{6}{\left(kL_\Sigma \cos\phi_m\right)^2}\left[\frac{N^2}{N^2-1}\right]\left(\frac{1}{SNR_\Sigma}\right). \tag{192}$$

If we assume that SNR_Σ remains constant, then observe that behavior as a function of N is limited to the quantity in the square brackets, and in going from $N = 2$ to $N \to \infty$, decreases from values 4/3 to 1. Consequently, adding more antenna phase centers within some finite overall aperture length at the expense of a corresponding lesser SNR for any one phase center, will decrease the variance of the DOA estimate somewhat, but arguably not by much.

Otherwise, a larger overall antenna aperture, and or larger wavenumber (i.e., shorter wavelength) will improve the DOA measurement variance with significantly greater impact.

We observe that this theme of subdividing a fixed aperture into ever more subapertures not significantly enhancing some performance aspects is not limited to DOA, with Minimum Detectable Velocity (MDV) in Ground Moving Target Indicator (GMTI) radar being another example.[14]

RMS Baseline

Consider the concept of a Root-Mean-Square (RMS) baseline. This is essentially twice the second-moment measure of antenna element positions, i.e., the span between one standard deviation on either side of center mean. For convenience we will assume an even number of uniformly spaced antenna elements centered at zero. The RMS baseline is then twice the standard deviation of the elements, or

$$b_{RMS} = 2\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \ . \tag{193}$$

This evaluates to

$$b_{RMS} = \Delta_x \sqrt{\frac{N^2 - 1}{3}} \ . \tag{194}$$

Using Eq. (188), we may write this as

$$b_{RMS} = L_\Sigma \sqrt{\frac{N^2 - 1}{3N^2}} \ . \tag{195}$$

This lets us write Eq. (192) as

$$\sigma_{\phi_m}^2 = \frac{2}{\left(kb_{RMS}\cos\phi_m\right)^2}\left(\frac{1}{SNR_\Sigma}\right). \tag{196}$$

Note that for $N = 2$, this equates to Eq. (166), where we recall from Eq. (190) that for this specific case $SNR_\Sigma = 2SNR_1$.

The overall message here is that for uniformly spaced antenna elements, dividing the overall aperture into more subapertures will increase the RMS baseline somewhat, and give some improvement in DOA estimation, albeit with diminishing returns.

Phase Ambiguities

It is imperative that phase measurements $\theta_{m,i}$ be unwrapped prior to DOA calculations. Since DOA is essentially a beamforming or beam-steering problem, the proper parameter from which to calculate DOA is relative time delay of the received signals across the antenna elements. Our construct is using phase shift as a surrogate for time delay, which is adequate for narrow-band signals. Nevertheless, we require

$$\theta_{m,i} - \theta_0 = \frac{2\pi c}{\lambda}\left(t_{m,i} - t_0\right), \tag{197}$$

where

c = velocity of propagation,
$t_{m,i}$ = the "measured" time corresponding to the phase measurement $\theta_{m,i}$, and
t_0 = the reference time corresponding to the reference phase θ_0. (198)

Note that for our calculations, the phase difference $\left(\theta_{m,i} - \theta_0\right)$ can *not* be modulo-2π, i.e., it needs to be unwrapped.

Simple Beamforming

As with frequency estimation that was addressed in 5.2.3, we can use Fourier analysis to estimate DOA. Using complex weighted sums of signals at the antenna elements is termed "beamforming." Building on the previous analysis, we identify the uniformly spaced elemental signals as

$$s_{m,i} = A_{m,i} e^{j\theta_{m,i}} = A_{m,i} e^{j\left(\theta_0 - k\sin\phi_m \Delta_x i\right)}, \tag{199}$$

and our reference signal as

$$s_{r,i} = e^{j\Delta_\theta i} \text{ = the reference signal that we wish to match.} \tag{200}$$

The matched filter response then becomes

$$S_r = \sum_{i=1}^{N} s_{m,i} e^{-j\Delta_\theta i}. \tag{201}$$

For good SNR, we will again assume $A_{m,i} \approx A$. Combining all these results and assumptions allows us to write

$$S_r = A\sum_{i=1}^{N} e^{j\left(\theta_0 - k\sin\phi_m \Delta_x i\right)} e^{-j\Delta_\theta i} = A e^{j\theta_0} \sum_{i=1}^{N} e^{-j\left(k\sin\phi_m \Delta_x + \Delta_\theta\right) i}. \tag{202}$$

This evaluates to

$$S_r = AN e^{j\theta_0} e^{-j\frac{\left(k\sin\phi_m \Delta_x + \Delta_\theta\right)\left(N+1\right)}{2}} \left(\frac{\sin\left(\frac{\left(k\sin\phi_m \Delta_x + \Delta_\theta\right) N}{2} \right)}{N\sin\left(\frac{\left(k\sin\phi_m \Delta_x + \Delta_\theta\right)}{2} \right)} \right). \tag{203}$$

Statistically, based on the discussion on Fourier Analysis in section 5.2.3, this will have a peak magnitude when $\Delta_\theta = -k\Delta_x \sin\phi_m$, from which we can calculate

$$\phi_m = \text{asin}\left(-\frac{\Delta_\theta}{k\Delta_x} \right)\Bigg|_{\Delta_\theta = peak}. \tag{204}$$

The variance in this peak location is given by Eq. (192). The trick is in identifying the exact location of the peak, which is outside the scope of this report. We do note that identifying something even close to this peak DOA can aid the aforementioned phase unwrapping need.

We also note that the width of the mainlobe, nominally its resolution, is nominally defined when

$$\frac{\left(k \sin \phi_m \, \Delta_x\right) N}{2} = \pi \, . \tag{205}$$

This yields the resolution of $\sin \phi_m$ to be

$$\rho_{\sin \phi_m} = \frac{2\pi}{kN\Delta_x} = \frac{2\pi}{k \, L_\Sigma} \, . \tag{206}$$

Interferometric SAR Reexamined

We now revisit the Interferometric SAR discussion in section 5.4.1, where we now have an array of antenna elements, as illustrated in Figure 5. Here we now have an arbitrary number N of antenna elements.

As in the previous development, it remains true that,

$$\frac{h}{\cos \psi} \approx R \sin \phi \tag{207}$$

Using Eq. (178), we identify

$$h \approx \left(\frac{6R \cos \psi}{k \Delta_x} \right) \frac{\left((N+1) \sum_{i=1}^{N} \theta_{m,i} - 2 \sum_{i=1}^{N} i \theta_{m,i} \right)}{N\left(N^2 - 1\right)} \tag{208}$$

The variance in the height estimate for our geometry is then

$$\sigma_h^2 \approx \left(R \cos \psi\right)^2 \sigma_{\sin \phi}^2 \, . \tag{209}$$

Using Eq. (184), Eq. (189), and Eq. (190), this is expanded to

$$\sigma_h^2 \approx \left(\frac{R \cos \psi}{k \, L_\Sigma} \right)^2 \left(\frac{6N^2}{N^2 - 1} \right) \left(\frac{1}{SNR_\Sigma} \right) . \tag{210}$$

For $N = 2$, this equates to Eq. (171).

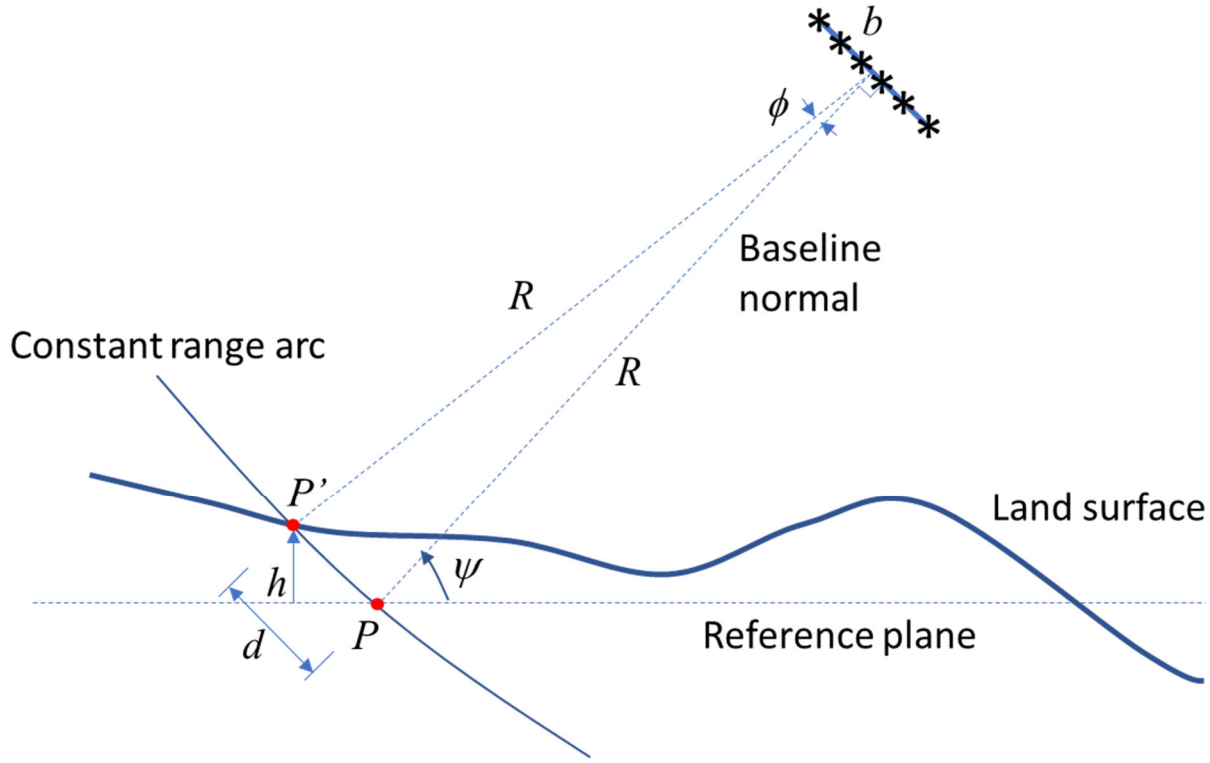

Figure 5. Interferometric SAR geometry for arbitrary multiple elements.

Superresolution

Let us examine the ratio of the noise in the peak location to the width of the mainlobe, that is

$$\frac{\sigma_{\sin\phi_m}}{\rho_{\sin\phi_m}} = \left[\frac{1}{2\pi}\sqrt{\frac{6N^2}{\left(N^2-1\right)}}\right]\frac{1}{\sqrt{SNR_\Sigma}} \approx \frac{1}{2\sqrt{SNR_\Sigma}}. \tag{211}$$

This means that with good SNR, the DOA can be estimated to something much more accurate than the nominal width of the formed beam. Of course, this assumes some a priori knowledge that we have but a single DOA for all received signals. In any case, in some communities this might be termed "superresolution." As such, Interferometric SAR topographic mapping might be considered super-resolved 3-D SAR.[15]

“If you can't explain it to a six-year-old, you don't understand it yourself.”
-- Albert Einstein

6 Conclusions

We offer and repeat some key points.

- Data are noisy. Get over it.
- Fitting the data to a model is called regression. Fitting the data to a linear sum of functions is called "linear regression." This can be done with respect to one independent parameter, or multiple independent parameters.
- Matrix algebra is an excellent tool for regression analysis.
- The goodness of the data fit to a model is embodied in the covariance matrix of the coefficients being estimated.
- A related measure of goodness is the Dilution of Precision (DOP). This is just the square root of the diagonal elements of the covariance matrix, albeit relative to a reference data standard deviation.
- There are many applications for regression analysis of radar data. These include, but are not limited to, frequency estimation, delay estimation, range estimation, height estimation, more general position/location estimation, Doppler/velocity estimation, Direction of Arrival (DOA) estimation, target geometry dependencies, motion measurement, antenna beam pattern description, autofocus, and more.

A Note About Kalman Filtering

Herein this report we have discussed estimating a least-squares solution for a finite data set, with the entire data set available for the necessary calculations. When the data set arrives sequentially over a lengthy interval or duration, we may wish to employ a sequential formulation of the calculation that can provide an updated estimate of the least-squares solution as new data becomes available. This leads to a construct called the "Kalman Filter." It is particularly useful for dynamic system models, where signal or noise parameters might change over the data collection period. The Kalman Filter and its many variations are discussed in many texts, with one by McDonough and Whalen being but one example.[10]

Figure 6. "OK everybody, is this a good enough line?" (Miramar Beach, Florida, USA, November 2022)

Appendix A – Orthogonal Projections and Errors

Here we discuss the derivation of the Least-Squares solution, and some of its properties. This follows the development given by Haykin.[16]

The least-squares solution chooses a projection of the data onto a subspace such that the distance from the observed data is minimized. This means that the error vector is orthogonal to the data subspace.

Recall that we want the best solution to the equation

$$\mathbf{A}\mathbf{x} = \mathbf{b}\,, \tag{A1}$$

where **A** defines the data space, and **b** is the observation vector. Our task is to find the coefficient vector **x** that minimizes any errors that prevent equality in Eq. (A1). We define the error vector as

$$\boldsymbol{\varepsilon} = \mathbf{A}\mathbf{x} - \mathbf{b}\,. \tag{A2}$$

We then identify the squared error as

$$\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = \left(\mathbf{A}\mathbf{x}-\mathbf{b}\right)^T\left(\mathbf{A}\mathbf{x}-\mathbf{b}\right) = \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} - \mathbf{b}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{A}^T\mathbf{b} + \mathbf{b}^T\mathbf{b}\,. \tag{A3}$$

We find the minimum squared error solution by taking the derivative with respect to **x** and setting it to zero, thereby yielding

$$2\mathbf{A}^T\mathbf{A}\hat{\mathbf{x}} - 2\mathbf{A}^T\mathbf{b} = \mathbf{0}\,, \tag{A4}$$

which can be manipulated to

$$\mathbf{A}^T\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^T\mathbf{b}\,. \tag{A5}$$

This resulting system of equations is often referred to as "normal" equations. It may be further manipulated to the well-known least-squares solution

$$\hat{\mathbf{x}} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}\,. \tag{A6}$$

We may further rearrange Eq. (A4) to

$$\mathbf{A}^T\left(\mathbf{A}\hat{\mathbf{x}}-\mathbf{b}\right) = \mathbf{A}^T\boldsymbol{\varepsilon} = \mathbf{0}\,. \tag{A7}$$

This makes it obvious that the least-squares processing chooses a solution $\hat{\mathbf{x}}$ such that the error vector $\boldsymbol{\varepsilon}$ is orthogonal to the subspace of the data **A**. This is an important and well-known property of least-squares calculations. Another way of saying this is that the solution $\hat{\mathbf{x}}$ is one

that projects the observations **b** onto the subspace of **A**. Plugging this into the original equation yields

$$\hat{\mathbf{b}} = \mathbf{A}\hat{\mathbf{x}} = \mathbf{A}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}\,. \tag{A8}$$

The matrix $\mathbf{A}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T$ is the projection operator. The vector $\hat{\mathbf{b}}$ is the projection of **b** onto the subspace of **A**.

As a final comment, we are considering here what is commonly referred to as "Ordinary Least Squares" (OLS), where it is assumed that the errors are confined to the observation vector **b**.

If there are errors in the data in **A**, then we need to consider adjusting our error vector $\boldsymbol{\varepsilon}$ accordingly. This is referred to as "Total Least Squares" (TLS).[17,18]

Appendix B – Least Squares and Maximum Likelihood

Because the least-squares method is quite powerful, versatile, and simple to implement, it is valuable in multiple applications, as we have observed in the text of this document. There is another reason that it is commonly used which is discussed in this Appendix. In general, one of the drawbacks of least-squares is that it is not always optimal. However, there is one important case where it does have optimal properties. When the error has a normal distribution then the least-squares estimate can be shown to be the maximum likelihood estimator which carries with it the asymptotic optimal properties of the maximum likelihood estimator.

We now show that the least-squares solution leads to the maximum likelihood estimator in the normal case. Recall the equation

$$\mathbf{A}\mathbf{x} = \mathbf{b}\,. \tag{B1}$$

We will assume that the errors in **b** are zero-mean and normally distributed, which we will incorporate into the conditional probability with notation

$$P\left(\mathbf{b}\middle|\mathbf{A}\right) \sim N\left(\mathbf{A}\mathbf{x}, \sigma^2\mathbf{I}\right), \tag{B2}$$

where

$$\sigma^2\mathbf{I} \text{ = covariance matrix of the normal distribution.} \tag{B3}$$

We want to find the coefficient matrix **x** that yields the maximum likelihood estimate of observations **b** with respect to data matrix **A**.

The conditional probability can be explicitly written as

$$P\left(\mathbf{b}\middle|\mathbf{A}\right) = \frac{e^{-\frac{1}{2}(\mathbf{b}-\mathbf{A}\mathbf{x})^T\left(\sigma^2\mathbf{I}\right)^{-1}(\mathbf{b}-\mathbf{A}\mathbf{x})}}{\sqrt{(2\pi)^k \det\left(\sigma^2\mathbf{I}\right)}} = \frac{e^{-\frac{1}{2\sigma^2}(\mathbf{b}-\mathbf{A}\mathbf{x})^T(\mathbf{b}-\mathbf{A}\mathbf{x})}}{\sqrt{(2\pi)^k\,\sigma^{2k}}}, \tag{B4}$$

where k is the length of vector **b**.

Conventionally, we will deal with the logarithm of this equation. Consequently, the log-likelihood under the multivariate normal assumption is

$$\ell\left(\mathbf{x}\right) = -\frac{1}{2}\ln\left(\left(2\pi\right)^k\,\sigma^{2k}\right) - \frac{1}{2\sigma^2}\left(\mathbf{b}-\mathbf{A}\mathbf{x}\right)^T\left(\mathbf{b}-\mathbf{A}\mathbf{x}\right). \tag{B5}$$

We wish to find the solution for **x** to maximize this. We do so by setting the derivative to zero, namely

$$\frac{d}{d\mathbf{x}}\ell(\mathbf{x}) = -\frac{1}{2\sigma^2}\frac{d}{d\mathbf{x}}\left((\mathbf{b}-\mathbf{A}\mathbf{x})^T(\mathbf{b}-\mathbf{A}\mathbf{x})\right) = 0\,. \tag{B6}$$

This may be expanded to yield

$$-\mathbf{A}^T\mathbf{b} + \mathbf{A}^T\mathbf{A}\mathbf{x} = 0\,. \tag{B7}$$

which of course can be rearranged to the normal system of equations

$$\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}\,, \tag{B8}$$

for which the solution becomes

$$\hat{\mathbf{x}} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}\,. \tag{B9}$$

This is the same answer as derived in Appendix A. Therefore, the maximum likelihood estimator is the least squares solution in this case.

Maximum likelihood estimators are more comprehensively discussed in texts by among others McDonough & Whalen,[10] and by Van Trees.[19]

Reference

[1] Gilbert Strang, *Linear Algebra and its Applications*, ISBN 0-12-673660-X, Academic Press, Inc., 1976, 1980.

[2] Philippe G. Ciarlet, *Introduction to Numerical Linear Algebra and Optimisation*, ISBN 0-521-33984-7, Cambridge University Press, 1989, 1991.

[3] Gene H. Golub, Charles F. Van Loan, *Matrix Computations*, second edition, ISBN 0-8018-3739-1, The Johns Hopkins University Press, 1989, 1991.

[4] Daniel T. Finkbeiner II, *Introduction to Matrices and Linear Transformations*, third edition, ISBN 0-7167-0084-0, W. H. Freeman and Company, 1960, 1966, 1978.

[5] Steven Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, ISBN-13: 978-0133457117, Prentice Hall Signal Processing Series, 1993.

[6] Elliott D. Kaplan (editor), *Understanding GPS Principles and Applications*, ISBN 0-89006-793-7, Arthech House, Inc., 1996.

[7] Norman R. Draper, Harry Smith, *Applied Regression Analysis (Wiley Series in Probability and Statistics) Third Edition*, ISBN-10: 0471170828, Wiley-Interscience, April 23, 1998.

[8] Norman L Johnson, Fred C. Leone, *Statistics and Experimental Design in Engineering and the Physical Sciences, Volume I, second edition,* ISBN 0-471-01756-6, John Wiley & Sons, Inc., 1977.

[9] Maurice G. Kendall, Alan Stuart, *The Advanced Theory of Statisticsm, Vol 2, Inference and Relationship*, Hafner Publishing Company, New York, 1961.

[10] Robert N. McDonough, Anthony D. Whalen, *Detection of Signals in Noise, second edition*, ISBN 0-12-744852-7, Academic Press, 1995.

[11] David C. Rife, Robert R. Boorstyn, "Single-Tone Parameter Estimation from Discrete-Time Observations," *IEEE Transactions on Information Theory*, Vol. IT-20, No. 5, September 1974.

[12] D. L. Bickel, W. H. Hensley, *Design, Theory, and Applications of lnterferometric Synthetic Aperture Radar for Topographic Mapping*, Sandia National Laboratories Report SAND96-1092, Unlimited Release, May 1996.

[13] A. W. Doerry, D. L. Bickel, "A comparison of interferometric SAR antenna options," SPIE 2013 Defense, Security & Sensing Symposium, Radar Sensor Technology XVII, Vol. 8714, Baltimore MD, 29 April – 3 May 2013.

[14] D. L. Bickel, A. W. Doerry, "On minimum detectable velocity," SPIE 2019 Defense & Commercial Sensing Symposium, Radar Sensor Technology XXIII, Vol. 11003, Baltimore, MD, 14-18 April 2019.

[15] Armin Doerry, Doug Bickel, "A 3-D SAR approach to IFSAR processing," SPIE 2000 International Symposium on Aerospace/Defense Sensing, Simulation, and Controls, Algorithms for Synthetic Aperture Radar Imagery VII, Vol. 4053, pp 78-88, Orlando FL, 24 April 2000.

[16] Simon Haykin, *Adaptive Filter Theory (5th Edition)*, ISBN-13: 978-0132671453, Pearson, 2013.

[17] Gene H. Golub, Charles F. Van Loan, "An analysis of the total least squares problem," *SIAM journal on numerical analysis*, Vol. 17, No. 6, pp. 883-893, 6 December 1980.

[18] Ivan Markovsky, Sabine Van Huffel, "Overview of total least-squares methods," Elsevier, *Signal Processing*, Vol. 87, No. 10, pp. 2283-2302, 1 October 2007.

[19] Harry L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*, ISBN-13: 978-0471093909, Wiley-Interscience, 1st edition, 4 April 2002.

Distribution

Unlimited Release

Email—Internal

all members	0534x	
Technical Library	01911	sanddocs@sandia.gov

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.