ECEN 615
Methods of Electric Power Systems Analysis

Lecture 16: Least Squares, Singular Value Decomposition (SVD), and State Estimation

Prof. Tom Overbye
Dept. of Electrical and Computer Engineering
Texas A&M University
overbye@tamu.edu

TEXAS A&M UNIVERSITY

Announcements

- Read Chapter 9
- Homework 5 is due on Tuesday Nov 1, 2022

Forward and Inverse Problems

- In science and engineering analysis we are often dealing with two classes of problems
  - Forward or direct problems; we're using a model with inputs to determine a set of outputs; power flow is an example of a forward problem
  - Inverse problems are the opposite: we're using a set of outputs, perhaps coupled with a model, to determine a set of inputs; power system state estimation is an example of an inverse problem
- Both forward and inverse problems can be linear or nonlinear
- Inverse problems can present many challenges including whether there are enough observations of sufficient quality to obtain an answer

Least Squares

- So far we have considered the solution of $\mathbf{Ax} = \mathbf{b}$ in which $\mathbf{A}$ is a square matrix; as long as $\mathbf{A}$ is nonsingular there is a single solution
  - That is, we have the same number of equations (m) as unknowns (n)
  - This is a forward problem
- Many problems are overdetermined in which there more equations than unknowns (m > n)
  - Overdetermined systems are usually inconsistent, in which no value of $\mathbf{x}$ exactly solves all the equations
- Underdetermined systems have more unknowns than equations (m < n); they never have a unique solution but are usually consistent

Method of Least Squares

- The least squares method is a solution approach for determining an approximate solution for an overdetermined system
- If the system is inconsistent, then not all of the equations can be exactly satisfied
- The difference for each equation between its exact solution and the estimated solution is known as the error
- Least squares seeks to minimize the sum of the squares of the errors
- Weighted least squares allows differ weights for the equations

Least Squares Solution History

- The method of least squares developed from trying to estimate actual values from a number of measurements
- Several persons in the 1700's, starting with Roger Cotes in 1722, presented methods for trying to decrease model errors from using multiple measurements
- Legendre presented a formal description of the method in 1805; evidently Gauss claimed he did it in 1795
- Method is widely used in power systems, with state estimation the best known application, dating from Fred Schweppe's work around 1970
  - Fred also did a lot of work associated with locational marginal prices (LMPs)
  - He was a professor at MIT who died in 1988 at age 54

Initial State Estimation Paper

120 TRANSACTIONS ON POWER APPARATUS AND SYSTEMS, VOL. PAS-89, NO. 1, JANUARY 1970

Power System Static-State Estimation, Part I: Exact Model

FRED C. SCHWEPPE, MEMBER, IEEE, AND J. WILDES, MEMBER, IEEE

Abstract—The static state of an electric power system is defined as the vector of the voltage magnitudes and angles at all network buses. The static-state estimator is a data processing algorithm far converting redundant meter readings and other available information into an estimate of the static-state vector. Discussions center on the general nature of the problem, mathematical modeling, an interative technique for calculating the state estimate, and concepts underlying the detection and identification of modeling errors. Problems of interconnected systems are considered. Results of some initial computer simulation tests are discussed.

INTRODUCTION

The static-state estimator is based on classical mathematical techniques such as estimation, detection, probability, statistics, and filtering. However, an attempt is made to present the ideas without drawing on an extensive background in these theories. A more theoretical presentation would have had some advantages, but the chosen approach hopefully makes the concepts more widely accessible. Most of the specialized mathematical jargon is relegated to background discussions which can be skipped.

This paper is the first of a three-part series on the static-state estimator. In this paper, the general problem, model, and theory of solution are developed. In [2] approximate models and solutions

It was a three part paper with part two focused on an approximate model and part three on implementation

Least Squares and Sparsity

- In many contexts least squares is applied to problems that are not sparse. For example, using a number of measurements to optimally determine a few values
  - Regression analysis is a common example, in which a line or other curve is fit to potentially many points)
  - Each measurement impacts each model value
- In the classic power system application of state estimation the system is sparse, with measurements only directly influencing a few states
  - Power system analysis classes have tended to focus on solution methods aimed at sparse systems; we'll consider both sparse and nonsparse solution methods

Least Squares Problem

- Consider $\mathbf{Ax} = \mathbf{b}$ $\quad \mathbf{A} \in \mathbb{R}^{m \times n},\ \mathbf{x} \in \mathbb{R}^{n},\ \mathbf{b} \in \mathbb{R}^{m}$

  or

$$\begin{bmatrix} (\mathbf{a}^1)^T \\ (\mathbf{a}^2)^T \\ \vdots \\ (\mathbf{a}^m)^T \end{bmatrix} \mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{22} & \dots & a_{2n} \\ & & \vdots & & \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Least Squares Solution

- We write $(\mathbf{a}^i)^T$ for the row i of $\mathbf{A}$ and $\mathbf{a}^i$ is a column vector
- Here, $m \geq n$ and the solution we are seeking is that which minimizes $\|\mathbf{Ax} - \mathbf{b}\|_p$, where p denotes some norm
- Since usually an overdetermined system has no exact solution, the best we can do is determine an $\mathbf{x}$ that minimizes the desired norm.

Choice of *p*

- We discuss the choice of p in terms of a specific example
- Consider the equation $\mathbf{Ax} = \mathbf{b}$ with

$$\mathbf{A} = \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \\ \mathbf{1} \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \qquad \textbf{with } b_1 \geq b_2 \geq b_3 \geq 0$$

(hence three equations and one unknown)

- We consider three possible choices for p:

Choice of p

$(i)\quad p = 1$

$$\|\mathrm{Ax} - \mathrm{b}\|_1 \text{ is minimized by } x^* = b_2$$

$(ii)\quad p = 2$

$$\|\mathrm{Ax} - \mathrm{b}\|_2 \text{ is minimized by } x^* = \frac{b_1 + b_2 + b_3}{3}$$

$(iii)\quad p = \infty$

$$\|\mathrm{Ax} - \mathrm{b}\|_\infty \text{ is minimized by } x^* = \frac{b_1 + b_3}{2}$$

The Least Squares Problem

- In general, $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_p$ is not differentiable for $p = 1$ or $p = \infty$
- The choice of $p = 2$ (Euclidean norm) has become well established given its least-squares fit interpretation
- The problem $\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2$ is tractable for two major reasons
  - First, the function is differentiable

$$\phi(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 = \frac{1}{2}\sum_{i=1}^{m}\left[\left(\mathbf{a}^i\right)^T \mathbf{x} - b_i\right]^2$$

The Least Squares Problem, cont.

– Second, the Euclidean norm is preserved under orthogonal transformations:

$$\left\| \left(\mathbf{Q}^T \mathbf{A} \right) \mathbf{x} - \mathbf{Q}^T \mathbf{b} \right\|_2 = \left\| \mathbf{A}\mathbf{x} - \mathbf{b} \right\|_2$$

with **Q** an arbitrary orthogonal matrix; that is, **Q** satisfies

$$\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I} \qquad \mathbf{Q} \in \mathbb{R}^{n \times n}$$

The Least Squares Problem, cont.

- We introduce next the basic underlying assumption: **A** is full rank, i.e., the columns of **A** constitute a set of linearly independent vectors
- This assumption implies that the rank of **A** is n
because n ≤ m since we are dealing with an overdetermined system
- Fact: The least squares solution $\mathbf{x}^*$ satisfies

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

Proof

- Since by definition the least squares solution $\mathbf{x}^*$ minimizes $\phi(\bullet)$ at the optimum, the derivative of this function zero:

$$\phi(\mathbf{x}) = \frac{1}{2}\|\mathbf{Ax}-\mathbf{b}\|_2^2 = \frac{1}{2}\left(\mathbf{x}^T\mathbf{A}^T\mathbf{Ax} - \mathbf{x}^T\mathbf{A}^T\mathbf{b} - \mathbf{b}^T\mathbf{A}\,\mathbf{x} + \mathbf{b}^T\mathbf{b}\right)$$

$$\mathbf{0} = \left.\frac{\partial\,\phi(\mathbf{x})}{\partial\mathbf{x}}\right|_{\mathbf{x}^*} = \left.\frac{\partial}{\partial\mathbf{x}}\left\{\frac{1}{2}\left(\mathbf{x}^T\mathbf{A}^T\mathbf{Ax} - \mathbf{x}^T\mathbf{A}^T\mathbf{b} - \mathbf{b}^T\mathbf{A}\,\mathbf{x} + \mathbf{b}^T\mathbf{b}\right)\right\}\right|_{\mathbf{x}^*}$$

$$= \left.\frac{\partial}{\partial\mathbf{x}}\left\{\frac{1}{2}\left(\mathbf{x}^T\mathbf{A}^T\mathbf{A}\,\mathbf{x} - 2\,\mathbf{x}^T\mathbf{A}^T\mathbf{b} + \mathbf{b}^T\mathbf{b}\right)\right\}\right|_{\mathbf{x}^*}$$

$$= \mathbf{A}^T\mathbf{A}\,\mathbf{x}^* - \mathbf{A}^T\mathbf{b}$$

Implications

- This underlying assumption implies that
  $\mathbf{A}$ is full rank $\Leftrightarrow \exists\ \mathbf{x} \neq \mathbf{0} \ni \mathbf{A}\mathbf{x} \neq \mathbf{0}$
- Therefore, the fact that $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ is positive definite (*p.d.*) follows from considering any $\mathbf{x} \neq \mathbf{0}$ and evaluating

$$\mathbf{x}^T \mathbf{A}^T \mathbf{A}\,\mathbf{x} = \|\mathbf{A}\,\mathbf{x}\|_2^2 > 0\,,$$

  which is the definition of a *p.d.* matrix
- We use the shorthand $\mathbf{A}^{\mathrm{T}}\mathbf{A} > \mathbf{0}$ for $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ being a symmetric, positive definite matrix

Implications

- The underlying assumption that **A** is full rank and therefore $\mathbf{A}^T\mathbf{A}$ is *p.d.* implies that there exists a unique least squares solution
- Note: we use the inverse in a conceptual, rather than a computational, sense

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

- The below formulation is known as the normal equations, with the solution conceptually straightforward

$$\left(\mathbf{A}^T\mathbf{A}\right)\mathbf{x} = \mathbf{A}^T\mathbf{b}$$

Example: Curve Fitting

- Say we wish to fit five points to a polynomial curve of the form

  $$\mathrm{f}(t, \mathbf{x}) = x_1 + x_2 t + x_3 t^2$$

- This can be written as

Here we are solving for $\mathbf{x}$

$$\mathbf{A}\mathbf{x} = \mathbf{y} = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ 1 & t_3 & t_3^2 \\ 1 & t_4 & t_4^2 \\ 1 & t_5 & t_5^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix}$$

Example: Curve Fitting

- Say the points are **t** =[0,1,2,3,4] and **y** = [0,2,4,5,4]. Then

$$\mathbf{Ax} = \mathbf{y} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 4 \\ 5 \\ 4 \end{bmatrix}$$

$$\mathbf{x}^* = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b} = \begin{bmatrix} 0.886 & 0.257 & -0.086 & -0.143 & 0.086 \\ -0.771 & 0.186 & 0.571 & 0.386 & -0.371 \\ 0.143 & -0.071 & -0.143 & -0.071 & 0.143 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \\ 4 \\ 5 \\ 4 \end{bmatrix}$$

$$\mathbf{x}^* = \begin{bmatrix} -0.2 \\ 3.1 \\ -0.5 \end{bmatrix}$$

Implications

- An important implication of positive definiteness is that we can factor $\mathbf{A}^T\mathbf{A}$ since $\mathbf{A}^T\mathbf{A} > \mathbf{0}$

$$\mathbf{A}^T\mathbf{A} = \mathbf{U}^T\mathbf{D}\,\mathbf{U} = \mathbf{U}^T\mathbf{D}^{1/2}\mathbf{D}^{1/2}\,\mathbf{U} = \mathbf{G}^T\mathbf{G}$$

- The expression $\mathbf{A}^T\mathbf{A} = \mathbf{G}^T\mathbf{G}$ is called the Cholesky factorization of the symmetric positive definite matrix $\mathbf{A}^T\mathbf{A}$

A Least Squares Solution Algorithm

Step 1: Compute the lower triangular part of $\mathbf{A}^T\mathbf{A}$

Step 2: Obtain the Cholesky Factorization $\mathbf{A}^T\mathbf{A} = \mathbf{G}^T\mathbf{G}$

Step 3: Compute $\mathbf{A}^T\mathbf{b} = \hat{\mathbf{b}}$

Step 4: Solve for $\mathbf{y}$ using forward substitution in

$$\mathbf{G}^T\mathbf{y} = \hat{\mathbf{b}}$$

and for $\mathbf{x}$ using backward substitution in

$$\mathbf{G}\,\mathbf{x} = \mathbf{y}$$

Note, our standard **LU** factorization approach would work; we can just solve it twice as fast by taking advantage of it being a symmetric matrix

Practical Considerations

- The two key problems that arise in practice with the triangularization procedure are:
  - First, while **A** maybe sparse, $\mathbf{A}^T\mathbf{A}$ is much less sparse and consequently requires more computing resources for the solution
    - In particular, with $\mathbf{A}^T\mathbf{A}$ second neighbors are now connected! Large networks are still sparse, just not as sparse
  - Second, $\mathbf{A}^T\mathbf{A}$ may actually be numerically less well-conditioned than **A**

Loss of Sparsity Example

- Assume the **B** matrix for a network is

$$\mathbf{B} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

- Then $\mathbf{B}^T\mathbf{B}$ is $\mathbf{B}^T\mathbf{B} = \begin{bmatrix} 2 & -3 & 1 & 0 \\ -3 & 6 & -4 & 1 \\ 1 & -4 & 6 & -3 \\ 0 & 1 & -3 & 2 \end{bmatrix}$

- Second neighbors are now connected!

Singular Value Decomposition

- Traditionally power system analysis has mostly been focused on the sparse matrices associated with the electric grid; there was not much signal analysis
- This is rapidly changing as the power industry get more signals and need to extract information from them, with PMUs one example
- This data is often presented in the form of a matrix, for example with the rows being sample points
- A key technique for extracting information from matrices is known as the singular value decomposition

Matrix Singular Value Decomposition (SVD)

- The SVD is a factorization of a matrix that generalizes the eigendecomposition to any m by n matrix to produce

$$\mathbf{Y} = \mathbf{U\Sigma V}^{\mathbf{T}}$$

The original concept is more than 100 years old, but has founds lots of recent applications

where **S** is a diagonal matrix of the singular values, and **U** and **V** are orthogonal matrices

- The singular values are non-negative real numbers that can be used to indicate the major components of a matrix (the gist is they provide a way to decrease the rank of a matrix
- A key application is image compression

Aside: SVD Image Compression Example

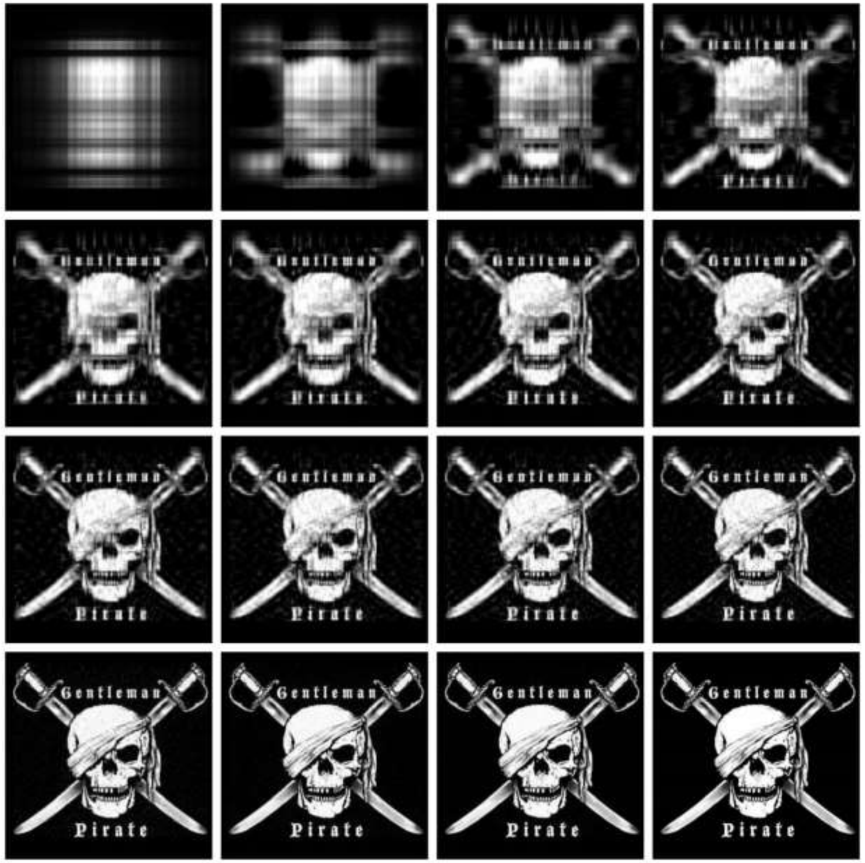

Figure 3.1: Image size 250x236 – modes used {{1,2,4,6},{8,10,12,14},{16,18,20,25},{50,75,100,original image}}

Images can be represented with matrices. When an SVD is applied and only the largest singular values are retained the image is compressed.

Image Source: www.math.utah.edu/~goller/F15_M2270/BradyMathews_SVDImage.pdf

An SVD Application, the Pseudoinverse of a Matrix

- The pseudoinverse of a matrix generalizes concept of a matrix inverse to an m by n matrix, m >= n
  - Specifically this is a Moore-Penrose Matrix Inverse
- Notation for the pseudoinverse of **A** is $\mathbf{A}^+$
- Satisfies $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$
- If **A** is a square matrix, then $\mathbf{A}^+ = \mathbf{A}^{-1}$
- Quite useful for solving the least squares problem since the least squares solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = \mathbf{A}^+\mathbf{b}$
- Can be calculated using an SVD

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

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

Pseudoinverse Least Squares Matrix Example

- Assume we wish to fit a line (mx + b = y) to three data points: (1,1), (2,4), (6,4)
- Two unknowns, m and b; hence $\mathbf{x} = [m \;\; b]^T$
- Setup in form of $\mathbf{Ax} = \mathbf{b}$

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 6 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 4 \end{bmatrix} \quad \text{so} \quad \mathbf{A} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 6 & 1 \end{bmatrix}$$

Aside: Pseudoinverse Least Squares Matrix Example

- Doing an economy SVD

$$\mathbf{A} = \mathbf{U\Sigma V}^T = \begin{bmatrix} -0.182 & -0.765 \\ -0.331 & -0.543 \\ -0.926 & 0.345 \end{bmatrix} \begin{bmatrix} 6.559 & 0 \\ 0 & 0.988 \end{bmatrix} \begin{bmatrix} -0.976 & -0.219 \\ 0.219 & -0.976 \end{bmatrix}$$

In an economy SVD the Σ matrix has dimensions of m by m if m < n or n by n if n < m

- Computing the pseudoinverse

$$\mathbf{A}^+ = \mathbf{V}\,\mathbf{\Sigma}^+\mathbf{U}^T = \begin{bmatrix} -0.976 & 0.219 \\ -0.219 & -0.976 \end{bmatrix} \begin{bmatrix} 0.152 & 0 \\ 0 & 1.012 \end{bmatrix} \begin{bmatrix} -0.182 & -0.331 & -0.926 \\ -0.765 & -0.543 & 0.345 \end{bmatrix}$$

$$\mathbf{A}^+ = \mathbf{V}\,\mathbf{\Sigma}^+\mathbf{U}^T = \begin{bmatrix} -0.143 & -0.071 & 0.214 \\ 0.762 & 0.548 & -0.310 \end{bmatrix}$$

Least Squares Matrix Pseudoinverse Example, cont.

- Computing $\mathbf{x} = [\mathrm{m\ b}]^T$ gives

$$\mathbf{A}^{+}\mathbf{b} = \begin{bmatrix} -0.143 & -0.071 & 0.214 \\ 0.762 & 0.548 & -0.310 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \\ 4 \end{bmatrix} = \begin{bmatrix} 0.429 \\ 1.71 \end{bmatrix}$$

- With the pseudoinverse approach we immediately see the sensitivity of the elements of **x** to the elements of **b**
  - New values of m and b can be readily calculated if **y** changes
- Computationally the SVD is order mn^2+n^3 (with $n < m$)

Often n is much less than m, so the result tends to scale linearly with m

SVD and Principal Component Analysis (PCA)

- The previous image compression example demonstrates PCA, which reduces dimensionality
  - Extracting the principal components
- The principal components are associated with the largest singular values
  - This helps to extract the key features of the data and removes redundancy
- PCA is a statistical method for reducing the dimensionality of a dataset
  - One example of PCA is facial recognition; another is market research
- PCA is starting to be more widely used in power system analysis, particularly when doing signal analysis
  - In electrical engineering a signal is defined as any time-varying quantity, which hopefully contains some information

Numerical Conditioning

- To understand the point on numerical ill-conditioning, we need to introduce terminology
- We define the norm of a matrix $\mathbf{B} \in \mathbb{R}^{m\times n}$ to be

$$\|\mathbf{B}\| = \max_{\mathbf{x} \neq \mathbf{0}} \left\{ \frac{\|\mathbf{B}\,\mathbf{x}\|}{\|\mathbf{x}\|} \right\}$$

$$= \textit{maximum stretching of the matrix } \mathbf{B}$$

- This is the maximum singular value of **B**

Numerical Conditioning Example

- Say we have the matrix

$$\mathbf{B} = \begin{bmatrix} 10 & 0 \\ 0 & 0.1 \end{bmatrix}$$

- What value of $\mathbf{x}$ with a norm of 1 that maximizes $\|\mathbf{Bx}\|$?
- What value of $\mathbf{x}$ with a norm of 1 that minimizes $\|\mathbf{Bx}\|$?

$$\|\mathbf{B}\| = \max_{\mathbf{x} \neq \mathbf{0}} \left\{ \frac{\|\mathbf{B\,x}\|}{\|\mathbf{x}\|} \right\}$$

$$= \textit{maximum stretching of the matrix } \mathbf{B}$$

Numerical Conditioning

$$= \max_i \left\{ \sqrt{\lambda}_i \,,\ \lambda_i \textit{ is an eigenvalue of } \underline{B}^T \underline{B} \right\},$$

Keep in mind the eigenvalues of a p.d. matrix are positive

i.e., λ_i is a root of the polynomial

$$p(\lambda) = det\left[\mathbf{B}^T\mathbf{B} - \lambda\mathbf{I}\right]$$

- In other words, the ⍰$_2$ norm of **B** is the square root of the largest eigenvalue of $\mathbf{B}^T\mathbf{B}$

Numerical Conditioning

- The conditioning number of a matrix **B** is defined as

$$\kappa(\mathbf{B}) = \|\mathbf{B}\| \, \|\mathbf{B}^{-1}\| = \frac{|\sigma_{max}(\mathbf{B})|}{|\sigma_{min}(\mathbf{B})|} \left. \vphantom{\frac{a}{b}} \right\} \begin{array}{l} \textit{the max / min stretching} \\ \textit{ratio of the matrix } \mathbf{B} \end{array}$$

- A well–conditioned matrix has a small value of $\kappa(\mathbf{B})$ close to 1; the larger the value of $\kappa(\mathbf{B})$, the more pronounced the ill-conditioning

Power System State Estimation (SE)

- The need is because in power system operations there is a desire to do “what if” studies based upon the actual “state” of the electric grid
  - An example is an online power flow or contingency analysis
- Overall goal of SE is to come up with a power flow model for the present "state" of the power system based on the actual system measurements
- SE assumes the topology and parameters of the transmission network are mostly known
- Measurements from SCADA and increasingly PMUs
- Overview is given in ECEN 615; much more details are provided in 614
  - Prof Ali Abur has done a lot of work in state estimation; he was at TAMU from 1985 to 2005, and is now at Northeastern University

Power System State Estimation

- Problem can be formulated in a nonlinear, weighted least squares form as

$$\min J(\mathbf{x}) = \sum_{i=1}^{m} \frac{\left[z_i - f_i(\mathbf{x}) \right]^2}{\sigma_i^2}$$

where J($\mathbf{x}$) is the scalar cost function, **x** are the state variables (primarily bus voltage magnitudes and angles), z_i are the m measurements, **f**($\mathbf{x}$) relates the states to the measurements and σ_i is the assumed standard deviation for each measurement

Assumed Error

- Hence the goal is to decrease the error between the measurements and the assumed model states **x**
- The ⍰$_i$ term weighs the various measurements, recognizing that they can have vastly different assumed errors

$$\min J(\mathbf{x}) = \sum_{i=1}^{m} \frac{\left[z_i - f_i(\mathbf{x}) \right]^2}{\sigma_i^2}$$

- Measurement error is assumed Gaussian (whether it is or not is another question); outliers (bad measurements) are often removed

State Estimation for Linear Functions

- First we'll consider the linear problem. That is where

$$\mathbf{z}^{meas} - \mathbf{f}(\mathbf{x}) = \mathbf{z}^{meas} - \mathbf{H}\mathbf{x}$$

- Let **R** be defined as the diagonal matrix of the variances (square of the standard deviations) for each of the measurements

$$\mathbf{R} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_m^2 \end{bmatrix}$$

State Estimation for Linear Functions

- We then differentiate J(**x**) w.r.t. **x** to determine the value of **x** that minimizes this function

$$J(\mathbf{x}) = \left[\mathbf{z}^{meas} - \mathbf{H}\mathbf{x}\right]^T \mathbf{R}^{-1} \left[\mathbf{z}^{meas} - \mathbf{H}\mathbf{x}\right]$$

$$\nabla J(\mathbf{x}) = -2\mathbf{H}^T \mathbf{R}^{-1} \mathbf{z}^{meas} + 2\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}\mathbf{x}$$

At the minimum we have $\nabla J(\mathbf{x}) = \mathbf{0}$. So solving for $\mathbf{x}$ gives

$$\mathbf{x} = \left[\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}\right]^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{z}^{meas}$$

Simple DC System Example

- Say we have a two bus power system that we are solving using the dc approximation. Say the line's per unit reactance is j0.1. Say we have power measurements at both ends of the line. For simplicity assume **R**=**I**. We would then like to estimate the bus angles. Then

$$z_1 = 2.2,\ f_1(\mathbf{x}) = \frac{\theta_1 - \theta_2}{0.1}, \quad z_2 = -2.0,\ f_2(\mathbf{x}) = \frac{\theta_2 - \theta_1}{0.1}$$

$$\mathbf{x} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}, \mathbf{H} = \begin{bmatrix} 10 & -10 \\ -10 & 10 \end{bmatrix}, \mathbf{H}^T\mathbf{H} = \begin{bmatrix} 200 & -200 \\ -200 & 200 \end{bmatrix}$$

We have a problem since $\mathbf{H}^T\mathbf{H}$ is singular.
This is because of lack of an angle reference.

Simple DC System Example, cont.

- Say we directly measure θ_1 (with a PMU) to be zero; set this as the third measurement. Then

$$z_1 = 2.2,\ f(\mathbf{x}) = \frac{\theta_1 - \theta_2}{0.1}, \quad z_2 = -2.0,\ f_2(\mathbf{x}) = \frac{\theta_2 - \theta_1}{0.1}, \quad z_3 = 0,\ f_3(\mathbf{x}) = 0$$

$$\mathbf{x} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} 2.2 \\ -2 \\ 0 \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} 10 & -10 \\ -10 & 10 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{H}^T\mathbf{H} = \begin{bmatrix} 201 & -200 \\ -200 & 200 \end{bmatrix}$$

$$\mathbf{x} = \left[\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right]^{-1} \mathbf{H}^T\mathbf{R}^{-1}\mathbf{z}^{meas}$$

$$\mathbf{x} = \begin{bmatrix} 201 & -200 \\ -200 & 200 \end{bmatrix}^{-1} \begin{bmatrix} 10 & -10 & 1 \\ -10 & 10 & 0 \end{bmatrix} \begin{bmatrix} 2.2 \\ -2 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -0.21 \end{bmatrix}$$

Note that the angles are in radians

Nonlinear Formulation

- A regular ac power system is nonlinear, so we need to use an iterative solution approach. This is similar to the Newton power flow. Here assume m measurements and n state variables (usually bus voltage magnitudes and angles) Then the Jacobian is the **H** matrix

$$\mathbf{H}(\mathbf{x}) = \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f_1}{x_1} & \cdots & \frac{\partial f_1}{x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{x_1} & \cdots & \frac{\partial f_m}{x_n} \end{bmatrix}$$

Measurement Example

- Assume we measure the real and reactive power flowing into one end of a transmission line; then the z_i-$f_i(\mathbf{x})$ functions for these two are

$$P_{ij}^{meas} - \left[-V_i^2 G_{ij} + V_i V_j \left(G_{ij} \cos\left(\theta_i - \theta_j\right) + B_{ij} \sin\left(\theta_i - \theta_j\right) \right) \right]$$

$$Q_{ij}^{meas} - \left[V_i^2 \left(B_{ij} + \frac{B_{cap}}{2} \right) + V_i V_j \left(G_{ij} \sin\left(\theta_i - \theta_j\right) - B_{ij} \cos\left(\theta_i - \theta_j\right) \right) \right]$$

  - Two measurements for four unknowns
- Other measurements, such as the flow at the other end, and voltage magnitudes, add redundancy

SE Iterative Solution Algorithm

- We then make an initial guess of **x**, $\mathbf{x}^{(0)}$ and iterate, calculating $\Delta\mathbf{x}$ each iteration

$$\Delta\mathbf{x} = \left[\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right]^{-1}\mathbf{H}^T\mathbf{R}^{-1}\begin{bmatrix} z_1 - f_1(\mathbf{x}) \\ \vdots \\ z_m - f_m(\mathbf{x}) \end{bmatrix}$$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \Delta\mathbf{x}$$

This is exactly the least squares form developed earlier with $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ an n by n matrix. This could be solved with Gaussian elimination, but this isn't preferred because the problem is often ill-conditioned

Keep in mind that **H** is no longer constant, but varies as **x** changes, and is often ill-conditioned

Nonlinear SE Solution Algorithm, Book Figure 9.11

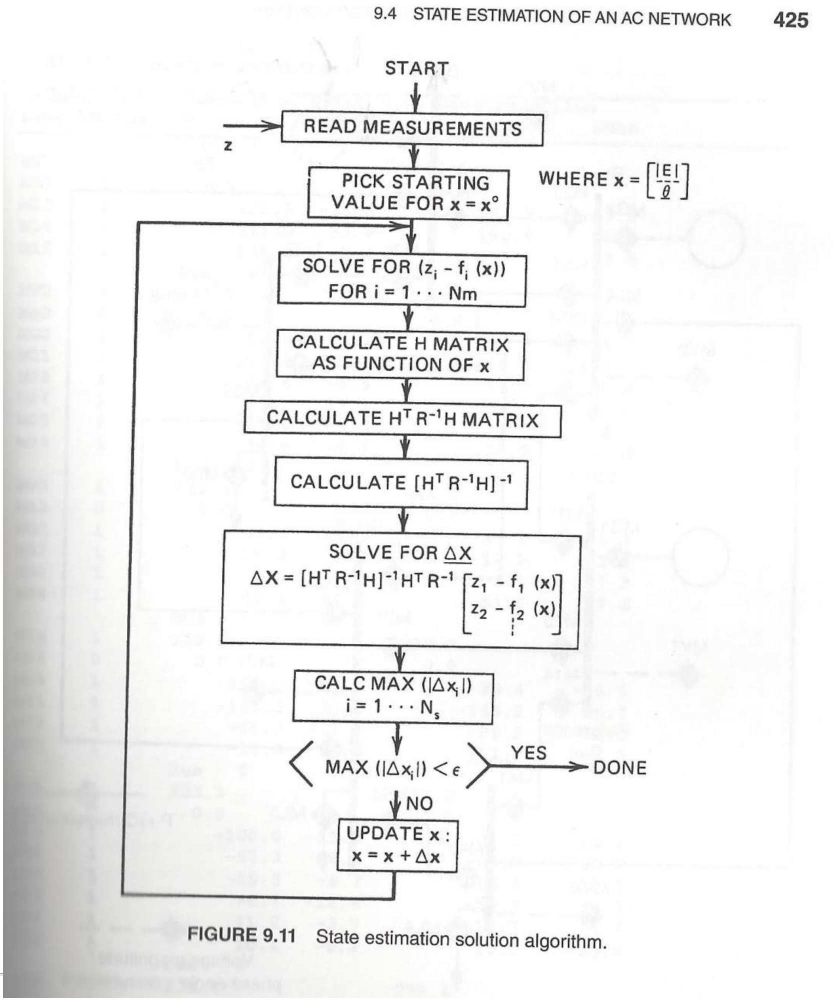

FIGURE 9.11 State estimation solution algorithm.