

# State Estimation

## 1.0 Introduction

State estimation for electric transmission grids was first formulated as a weighted least-squares problem by Fred Schweppe and his research group [1] in 1969 (Schweppe also developed spot pricing, the precursor of modern-day locational marginal prices – LMPs – which are a central feature of electricity markets). Figure 0 below shows Dr. Schweppe (the one seated in the chair).



Fig. 0

The basic motivation for state estimation is that we want to perform computer analysis of the

network under the conditions characterized by the current set of measurements.

Specifically, we want to know the values of the bus voltage phasor magnitudes and angles  $|V_k|$ ,  $\theta_k$  for all  $k=1,\dots,N$  buses in the network (we assume  $\theta_1=0$  so we do not need to find that one).

We begin with linear least squares estimation.

## **2.0 Linear least squares estimation**

The material in this section closely follows that in [2, chapter 2].

Consider the circuit given in Fig. 1 below where current injections  $I_1$ ,  $I_2$ , and voltage  $E$  are unknown. Let  $R_1=R_2=R_3=1.0 \ \Omega$ . The measurements are as follows:

- meter  $A_1$ :  $i_{1,2}=1.0$  Ampere
- meter  $A_2$ :  $i_{3,1}=-3.2$  Ampere
- meter  $A_3$ :  $i_{2,3}=0.8$  Ampere
- meter  $V$ :  $e=1.1$  volt

The problem is to determine the state of the circuit, which in this case is nodal voltages  $v_1$ ,  $v_2$ , and the voltage  $e$  across the voltage source.

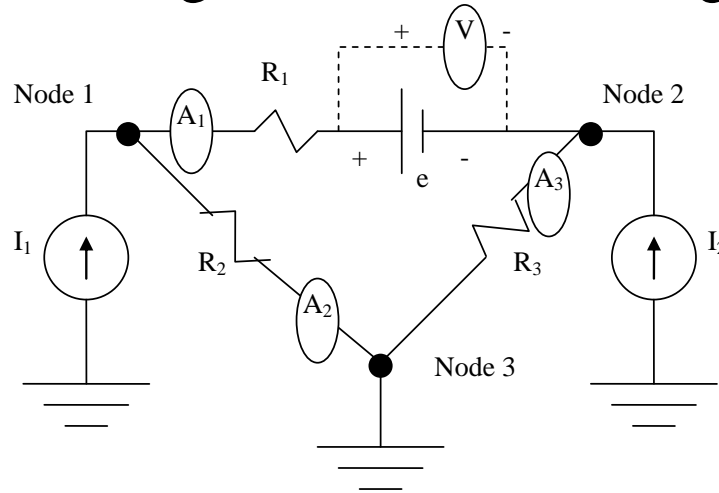


Fig. 1

Let's write each one of the measured currents in terms of the node voltages, and we may also write down our one voltage measurement.

$$i_{1,2}^m = \frac{v_1 - v_2 - e}{1} = v_1 - v_2 - e = 1.0 \quad (1)$$

$$i_{3,1}^m = \frac{0 - v_1}{1} = -v_1 = -3.2 \quad (2)$$

$$i_{2,3}^m = \frac{v_2 - 0}{1} = v_2 = 0.8 \quad (3)$$

$$e = 1.1 \quad (4)$$

Expressing all of the above in matrix form:

$$\begin{bmatrix} 1 & -1 & -1 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ e \end{bmatrix} = \begin{bmatrix} 1.0 \\ -3.2 \\ 0.8 \\ 1.1 \end{bmatrix} \quad (5)$$

Let's denote terms in eq. (5) as  $\underline{A}$ ,  $\underline{x}$ , and  $\underline{b}$ , so:

$$\underline{Ax} = \underline{b} \quad (6a)$$

where

$$\underline{A} = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \underline{x} = \begin{bmatrix} v_1 \\ v_2 \\ e \end{bmatrix}, \quad \underline{b} = \begin{bmatrix} 1.0 \\ -3.2 \\ 0.8 \\ 1.1 \end{bmatrix} \quad (6b)$$

How do we solve eq. (5)?

Observe that the multiplying matrix is not square, i.e., there are 4 rows but only 3 columns. The reason for this is because there are 4 equations but only 3 variables. This means that the system of equations defined by eq. (5) is over-determined. This is a standard feature in state-estimation. Noticing that there is one equation for each measurement, the implication is that we will attempt to always obtain as many measurements as we can.

There is no single solution to eq. (5), but there is a single solution that is normally thought of as “best.” This solution is the one that minimizes the sum of the squared “error” between what should be computed by each equation, which is:

$$\underline{b} = \begin{bmatrix} 1.0 \\ -3.2 \\ 0.8 \\ 1.1 \end{bmatrix} \quad (7)$$

and what is computed by each equation, which is:

$$\underline{Ax} = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ e \end{bmatrix} \quad (8)$$

The difference, or error, is then:

$$\underline{b} - \underline{Ax} = \begin{bmatrix} 1.0 \\ -3.2 \\ 0.8 \\ 1.1 \end{bmatrix} - \begin{bmatrix} 1 & -1 & -1 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ e \end{bmatrix} = \begin{bmatrix} 1.0 - v_1 + v_2 + e \\ -3.2 + v_1 \\ 0.8 - v_2 \\ 1.1 - e \end{bmatrix} \quad (9)$$

The squared error is then

$$\begin{bmatrix} (1.0 - v_1 + v_2 + e)^2 \\ (-3.2 + v_1)^2 \\ (0.8 - v_2)^2 \\ (1.1 - e)^2 \end{bmatrix}$$

and the sum of the squared errors is:

$$(1.0 - v_1 + v_2 + e)^2 + (-3.2 + v_1)^2 + (0.8 - v_2)^2 + (1.1 - e)^2$$

Careful tracking of the previous expression will indicate that it could be written as

$$(\underline{b} - \underline{Ax})^T (\underline{b} - \underline{Ax})$$

Let's multiply the above by  $\frac{1}{2}$  and give it a name:

$$J = \frac{1}{2} (\underline{b} - \underline{Ax})^T (\underline{b} - \underline{Ax}) \quad (10)$$

Our problem is then to choose  $\underline{x}$  so as to minimize  $J$ . (This is an unconstrained minimization problem.) We can do this setting the gradient of  $J$  with respect to  $\underline{x}$  to zero and solving for  $\underline{x}$ .

To do this, we expand  $J$  as follows:

$$\begin{aligned} J &= \frac{1}{2} (\underline{b} - \underline{Ax})^T (\underline{b} - \underline{Ax}) = \frac{1}{2} (\underline{b}^T - (\underline{Ax})^T) (\underline{b} - \underline{Ax}) \\ &= \frac{1}{2} [\underline{b}^T \underline{b} - (\underline{Ax})^T \underline{b} - \underline{b}^T \underline{Ax} + (\underline{Ax})^T \underline{Ax}] \end{aligned} \quad (11)$$

Using  $(\underline{Ax})^T = \underline{x}^T \underline{A}^T$ , we have:

$$J = \frac{1}{2} \left[ \underline{b}^T \underline{b} - \underline{x}^T \underline{A}^T \underline{b} - \underline{b}^T \underline{A} \underline{x} + \underline{x}^T \underline{A}^T \underline{A} \underline{x} \right] \quad (12a)$$

Consider the second and third terms in (12a). Using a 2x2 to illustrate,

$$\underline{x}^T \underline{A}^T \underline{b} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = b_1 x_1 A_{11} + b_1 x_2 A_{12} + b_2 x_1 A_{21} + b_2 x_2 A_{22}$$

$$\underline{b}^T \underline{A} \underline{x} = \begin{bmatrix} b_1 & b_2 \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = b_1 x_1 A_{11} + b_2 x_1 A_{21} + b_1 x_2 A_{12} + b_2 x_2 A_{22}$$

we observe these terms are equal. Therefore, (12a) becomes

$$J = \frac{1}{2} \left[ \underline{b}^T \underline{b} - 2 \underline{b}^T \underline{A} \underline{x} + \underline{x}^T \underline{A}^T \underline{A} \underline{x} \right] \quad (12b)$$

To remind us all about gradients, we recall that it is given by (13):

$$\nabla_x J = \begin{bmatrix} \frac{\partial J}{\partial x_1} \\ \frac{\partial J}{\partial x_2} \\ \vdots \\ \frac{\partial J}{\partial x_n} \end{bmatrix} \quad (13)$$

Now (12b) is written in compact notation, and it may not be obvious how to differentiate each term in it. To assist with this, I provide the following relations:

	Function	Gradient
#1	$F = \underline{x}^T \underline{b}$	$\nabla_{\underline{x}} F = \underline{b}$
#2	$F = \underline{b}^T \underline{x}$	$\nabla_{\underline{x}} F = \underline{b}$
#3	$F = \underline{x}^T \underline{A} \underline{u}$	$\nabla_{\underline{x}} F = \underline{A} \underline{u}$
#4	$F = \underline{u}^T \underline{A} \underline{x}$	$\nabla_{\underline{x}} F = \underline{A}^T \underline{u}$
#5	$F = \underline{x}^T \underline{A} \underline{x}$	$\nabla_{\underline{x}} F = 2 \underline{A}^T \underline{x}$

The above gradient relations apply to (12b) as illustrated in (12c).

$$J = \frac{1}{2} \left[ \underline{b}^T \underline{b} - \overbrace{2 \underline{b}^T \underline{A} \underline{x}}^{\#4} + \overbrace{\underline{x}^T \underline{A}^T \underline{A} \underline{x}}^{\#5} \right] \quad (12c)$$

Using the appropriate relations in the above table (#4 to the second term and #5 to the third term), the gradient of (12c) can be expressed as:

$$\nabla_{\underline{x}} J = \frac{1}{2} \left[ -2 \underline{A}^T \underline{b} + 2 (\underline{A}^T \underline{A})^T \underline{x} \right] \quad (14)$$

Again, using  $(\underline{A} \underline{x})^T = \underline{x}^T \underline{A}^T$ , the second term is

$$2 (\underline{A}^T \underline{A})^T \underline{x} = 2 \underline{A}^T \underline{A} \underline{x}$$

so (14) becomes

$$\nabla_{\underline{x}} J = \frac{1}{2} \left[ -2 \underline{A}^T \underline{b} + 2 \underline{A}^T \underline{A} \underline{x} \right] = -\underline{A}^T \underline{b} + \underline{A}^T \underline{A} \underline{x} \quad (15)$$

The minimum of J is obtained when

$$\frac{\partial J}{\partial \underline{x}} = -\underline{A}^T \underline{b} + \underline{A}^T \underline{A} \underline{x} = \underline{0} \quad (16)$$



And this implies that

$$\underline{A}^T \underline{b} = \underline{A}^T \underline{A} \underline{x} \quad (17)$$

Note:

- Equation (17) is referred to in statistics as the *normal equations*.
- We could have obtained (17) by just multiplying  $\underline{A}\underline{x}=\underline{b}$  through by  $\underline{A}^T$ .
- $\underline{A}^T \underline{A}$  multiplies an  $m \times n$  by an  $n \times m$  to get an  $m \times m$  matrix  $\rightarrow$  Square!
- $(\underline{A}^T \underline{A})^T = \underline{A}^T \underline{A}$ , so the transpose of  $\underline{A}^T \underline{A}$  is itself. This may only occur if  $\underline{A}^T \underline{A}$  is symmetric, implying that  $\underline{A}^T \underline{A}$  is symmetric.
- Reference [3, p. 157] shows that if  $A$  has linearly independent columns, then  $\underline{A}^T \underline{A}$  is invertible.

Solving eq. (17) for  $\underline{x}$  results in

$$\underline{x} = (\underline{A}^T \underline{A})^{-1} \underline{A}^T \underline{b} \quad (18)$$

Define the gain matrix  $\underline{G}$  as

$$\underline{G} = \underline{A}^T \underline{A} \quad (19)$$

Also define the *pseudo-inverse* of  $\underline{A}$  as

$$\underline{A}^I = (\underline{A}^T \underline{A})^{-1} \underline{A}^T = \underline{G}^{-1} \underline{A}^T \quad (20)$$

Now we can find the answer to our problem as.

$$\underline{x} = (\underline{A}^T \underline{A})^{-1} \underline{A}^T \underline{b} = \underline{G}^{-1} \underline{A}^T \underline{b} = \underline{A}^I \underline{b} \quad (21)$$

First, the gain matrix is given as

$$\underline{G} = \underline{A}^T \underline{A} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{bmatrix}$$

The inverse of the gain matrix is then found from Matlab as

$$\underline{G}^{-1} = \frac{1}{4} \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix}$$

The pseudo-inverse is then

$$\underline{A}^I = \underline{G}^{-1} \underline{A}^T = \frac{1}{4} \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & -3 & 1 & 1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix}$$

Then we can obtain the least squares estimate of the 3 states from the 4 measurements as

$$\underline{x} = \underline{A}^I \underline{b} = \frac{1}{4} \begin{bmatrix} 1 & -3 & 1 & 1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix} \begin{bmatrix} 1.0 \\ -3.2 \\ 0.8 \\ 1.1 \end{bmatrix} = \begin{bmatrix} 3.125 \\ 0.875 \\ 1.175 \end{bmatrix}$$

It is also interesting at this point to look at the difference between the measurements we actually had, which is  $\underline{b}$ , and the values corresponding to those measurements that we would compute using the state vector  $x$ , which is  $\underline{Ax}$ .

This difference is referred to as the residual,  $\underline{r}$ , and given by eq. (9) as

$$\begin{aligned} \underline{r} = \underline{b} - \underline{Ax} &= \begin{bmatrix} 1.0 \\ -3.2 \\ 0.8 \\ 1.1 \end{bmatrix} - \begin{bmatrix} 1 & -1 & -1 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ e \end{bmatrix} \\ &= \begin{bmatrix} 1.0 \\ -3.2 \\ 0.8 \\ 1.1 \end{bmatrix} - \begin{bmatrix} 1 & -1 & -1 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3.125 \\ 0.875 \\ 1.175 \end{bmatrix} = \begin{bmatrix} 1.0 \\ -3.2 \\ 0.8 \\ 1.1 \end{bmatrix} - \begin{bmatrix} 1.075 \\ -3.125 \\ 0.875 \\ 1.175 \end{bmatrix} = \begin{bmatrix} -0.075 \\ -0.075 \\ -0.075 \\ -0.075 \end{bmatrix} \end{aligned}$$

### 3.0 A motivating system and some basics

Consider Fig. 2 where we obtain measurements on the indicated quantities.

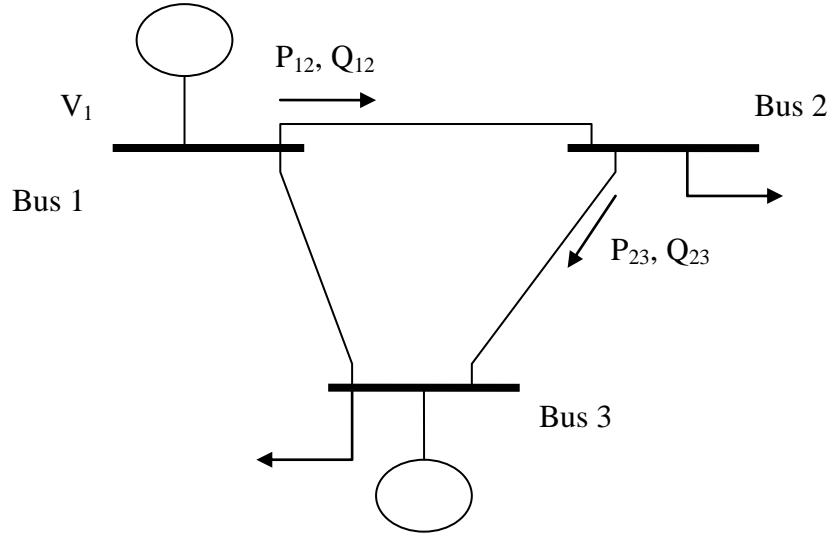


Fig. 2

We denote the measured quantities as follows:

$$\begin{aligned}
 z_1 &= V_1 \\
 z_2 &= P_{12} \\
 z_3 &= Q_{12} \\
 z_4 &= P_{23} \\
 z_5 &= Q_{23}
 \end{aligned} \tag{22}$$

Then we can write

$$z_i = \hat{z}_i + \eta_i \tag{23}$$

where

- $z_i$  is the measured value
- $\hat{z}_i$  is the true value (unknown)

- $\eta_i$  is the error (unknown)

Not knowing  $\hat{z}_i$  and  $\eta_i$  is a problem. However, we may obtain statistical information from calibration curves (error as a function of measurement) of measuring instruments. It is usually assumed that  $\eta_i$  is a *random variable* with a normal (Gaussian) distribution having zero mean, as illustrated in Fig. 3.

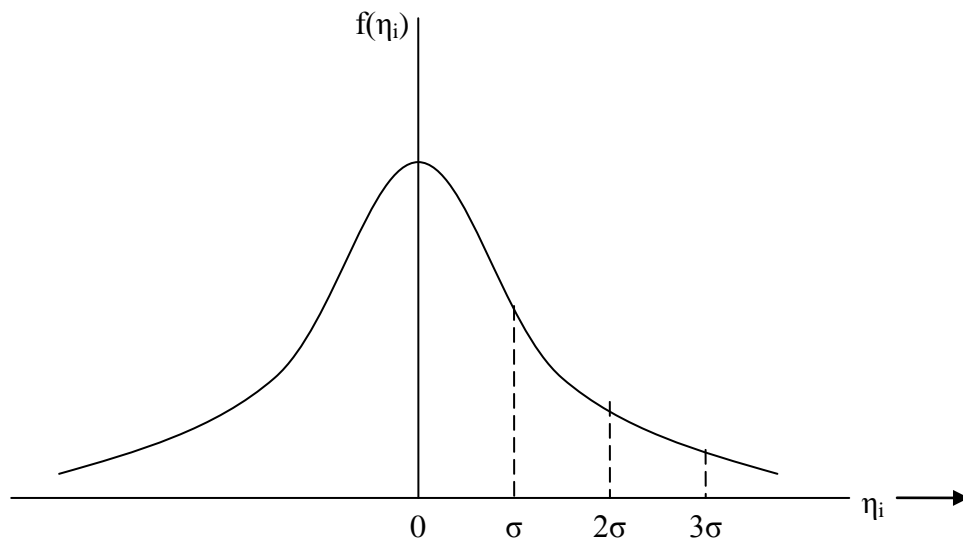


Fig. 3

This makes sense because a particular measuring instrument, if it is reasonably calibrated, may read a little high (positive error)

at times or a little low (negative error) at times, so that the average error is zero. Calibration curves enable determination of the variance  $\sigma_i$ .

Recall the expectation operator, which we denote as  $E(\bullet)$  (i.e., the expected value of  $\bullet$ ). It is defined as:

$$E(x) = \int_{-\infty}^{\infty} xf(x)dx \quad (24)$$

which gives the mean value of a variable  $x$  described by the probability distribution function  $f(x)$ .

We also define variance as:

$$\sigma_x^2 = \text{var}(x) = \int_{-\infty}^{\infty} (x - E(x))^2 f(x)dx \quad (25)$$

We relate variance to mean beginning with (25).

$$\begin{aligned}
\sigma_x^2 = \text{var}(x) &= \int_{-\infty}^{\infty} (x - E(x))^2 f(x) dx = \\
&= \int_{-\infty}^{\infty} (x^2 - 2xE(x) + (E(x))^2) f(x) dx \\
&= \int_{-\infty}^{\infty} x^2 f(x) dx - 2E(x) \int_{-\infty}^{\infty} xf(x) dx + (E(x))^2 \int_{-\infty}^{\infty} f(x) dx \\
&= E(x^2) - 2(E(x))^2 + (E(x))^2 \\
&= E(x^2) - (E(x))^2
\end{aligned} \tag{26}$$

From eq. (26) (which is true for any random variable  $x$ ), we see that if the mean is 0 ( $E(x)=0$ ), then the last term in eq. (26) is 0 and

$$\sigma_x^2 = E(x^2) \tag{27}$$

In regards to the calibration error, characterized by the random variable  $\eta_i$ , we have then:

- $E(\eta_i) = 0$  (zero mean)
- $E(\eta_i^2) = \sigma_i^2$  (variance)

Note that the larger the variance, the less accurate is the measuring device.

Since we have multiple measuring instruments, we also need to understand how the statistics of

one random variable relate to the statistics of another.

The covariance measure is effective in doing this, and is defined as

$$\sigma_{xy}^2 = \text{cov}(x, y) = \int_{-\infty}^{\infty} (x - E(x))(y - E(y))f(x, y)dx \quad (28)$$

Note that variance is a special case of covariance when  $x=y$ , i.e.,  $\text{var}(x)=\text{cov}(x,x)$ .

The covariance  $\text{cov}(x,y)$  is a measure of how two variables change together.

- If  $\text{cov}(x,y)>0$ , then  $x$  tends to increase when  $y$  increases.
- If  $\text{cov}(x,y)<0$ , then  $x$  tends to decrease when  $y$  increases.
- If  $x$  and  $y$  are independent, then  $\text{cov}(x,y)=0$  (but note that  $\text{cov}(x,y)=0$  does not necessarily imply independence) because covariance reflects linear dependence. Two variables can be nonlinearly dependent (and therefore not independent) but have a covariance of 0.



It can be shown [4] from eq. (26) that

$$\sigma_{xy}^2 = \text{cov}(x, y) = E(xy) - E(x)E(y) \quad (29)$$

• If  $x$  and  $y$  are independent, then  $E(xy) = E(x)E(y)$ .

Therefore, for two independent random variables, the covariance is:

$$\sigma_{xy}^2 = \text{cov}(x, y) = 0$$

Now, back to our state estimation problem...

A basic assumption:

The errors  $\eta_i$  and  $\eta_j$  for any two measuring instruments  $i$  and  $j$  are independent. This means that

$$\text{cov}(\eta_i, \eta_j) = \begin{cases} 0, & i \neq j \\ \sigma_i^2, & i = j \end{cases} \quad (30)$$

Thus, we can define a covariance matrix  $\underline{\mathbf{R}}$ , where the element in position  $(i, j)$  is  $\text{cov}(\eta_i, \eta_j)$ . Given (28), the matrix will appear as:

$$\underline{R} = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_m^2 \end{bmatrix} \quad (31)$$

where it is assumed that we have  $m$  measuring instruments.

We will use eq. (31) in our development.

## 4.0 Problem for AC State Estimator

We will define the state vector as for an N-bus network as:

$$\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_{2N-1} \end{bmatrix} = \begin{bmatrix} \theta_2 \\ \vdots \\ \theta_N \\ \hline V_1 \\ \vdots \\ V_N \end{bmatrix} \quad (32)$$

We will have  $n=2N-1$  states.

For each parameter for which we have a measurement, we want to write an equation in terms of the states. In other words, if we have a

measurement  $z_i$ , then the true value of that measurement will be

$$\hat{z}_i = h_i(\underline{x}) \quad (33)$$

For a voltage measurement, the function  $h_i$  is very simple:

$$\hat{z}_i = V_k \quad (34)$$

where measurement  $i$  occurs at bus  $k$ .

For MW and MVAR flows, the function  $h_i$  is given by the expressions for power flow across a line from bus  $p$  to bus  $q$ . These are given by

$$P_{pq} = V_p^2 g_{pq} - V_p V_q g_{pq} \cos(\theta_p - \theta_q) - V_p V_q b_{pq} \sin(\theta_p - \theta_q) \quad (35)$$

$$Q_{pq} = -V_p^2 (b_{pq} + b_p) - V_p V_q g_{pq} \sin(\theta_p - \theta_q) + V_p V_q b_{pq} \cos(\theta_p - \theta_q) \quad (36)$$

where the line has

- series admittance of  $g_{pq} + jb_{pq}$ ;  $g_{pq} > 0$ ,  $b_{pq} < 0$  for inductive line.
- Shunt susceptance at bus  $p$  of  $b_p$  (which includes any reactive shunt at the bus plus half of the line charging). If capacitive, then  $b_p > 0$ .

Now define some vectors:

$$\text{Measured values: } \underline{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix} \quad (37)$$

$$\text{True values: } \underline{\hat{z}} = \begin{bmatrix} \hat{z}_1 \\ \vdots \\ \hat{z}_m \end{bmatrix} \quad (38)$$

$$\text{Errors } \underline{\eta} = \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_m \end{bmatrix} \quad (39)$$

Generalizing eq. (23), we have:

$$\underline{z} = \underline{\hat{z}} + \underline{\eta} \quad (40)$$

From (33), we also have for a vector of functions expressing the measurement values in terms of the states:

$$\underline{\hat{z}} = \underline{h}(\underline{x}) \quad (41)$$

Substituting eq. (41) into (40), we have:

$$\underline{z} = \underline{h}(\underline{x}) + \underline{\eta} \quad (42)$$

Now consider what we have here. The number of unknowns is  $n=2N-1$  (the states in  $\underline{x}$  which

are the angle and voltage variables), and then we have some number of measurements  $m$ .

Let's assume that  $m > n$ , i.e., that we have more measurements than states.

One thing we could do is

- set  $\underline{\eta} = \underline{0} \rightarrow \underline{z} = \underline{h}(\underline{x})$
- choose  $m = n$  equations (each corresponding to a measurement)
- Solve for  $\underline{x}$  (it would need to be non-linear solver but once done, solution is unique).

However, the tough question would be: Which measurements to choose to keep? Which are the best?

Since we do not know which measurements are the best, we instead make sure that we have more measurements than states, i.e., we will solve the problem for  $m > n = 2N - 1$ .

So our strategy is as we saw in our earlier example, to choose  $\underline{x}$  so as to minimize the sum

of the squared errors between the measured values and the actual values.

From eq. (40), we have the error is

$$\underline{\eta} = \underline{z} - \underline{h}(\underline{x}) \quad (43)$$

Similar to eq. (10), we can then express the sum of squared errors as

$$J' = \frac{1}{2} \sum_{i=1}^m \eta_i^2 = \frac{1}{2} \underline{\eta}^T \underline{\eta} = \frac{1}{2} (\underline{z} - \underline{h}(\underline{x}))^T (\underline{z} - \underline{h}(\underline{x})) \quad (44)$$

We denote the above as  $J'$  because we will find it convenient to modify a bit.

By minimizing  $J'$ , we are effectively choosing  $\underline{x}$  that best “fits” the measurements. Remember, however, that some measurement devices are better than others (which is a different statement than some measurements are better than others).

It is reasonable, then, to place more weight on the better measuring devices.

A good choice for this weight is  $\frac{1}{\sigma_i^2}$  since

• Good device  $\rightarrow$  small  $\sigma_i^2$ , large  $\frac{1}{\sigma_i^2}$

• Bad device  $\rightarrow$  large  $\sigma_i^2$ , small  $\frac{1}{\sigma_i^2}$

Therefore, we will modify eq. (44) to be:

$$J = \frac{1}{2} \sum_{i=1}^m \frac{\eta_i^2}{\sigma_i^2} \quad (45)$$

And so  $\sigma_i^2$  will increase the error terms for accurate measurements, making the optimization (and the solution) more dependent on those measurements.

Recall the covariance matrix given by eq. (31), repeated here for convenience:

$$\underline{R} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_m^2 \end{bmatrix} \quad (31)$$

Because  $R$  is diagonal, its inverse is easy to find:

$$\underline{R}^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2^2} & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & 0 & \frac{1}{\sigma_m^2} \end{bmatrix} \quad (46)$$

We can therefore express eq. (45) in as:

$$J = \frac{1}{2} \sum_{i=1}^m \frac{\eta_i^2}{\sigma_i^2} = \frac{1}{2} \underline{\eta}^T \underline{R}^{-1} \underline{\eta} = \frac{1}{2} (\underline{z} - \underline{h}(\underline{x}))^T \underline{R}^{-1} (\underline{z} - \underline{h}(\underline{x})) = \frac{1}{2} \sum_{i=1}^m \frac{(z_i - h_i(\underline{x}))^2}{\sigma_i^2} \quad (47)$$

The problem then becomes to find  $\underline{x}$  that minimizes J. Note, however, that  $\underline{h}$  is nonlinear, and so our solution will necessarily be iterative.

## 5.0 Solution for AC State Estimator

So the problem can be stated as follows:

mimimize



$$J = \frac{1}{2} \sum_{i=1}^m \frac{\eta_i^2}{\sigma_i^2} = \frac{1}{2} \underline{\eta}^T \underline{R}^{-1} \underline{\eta} = \frac{1}{2} (\underline{z} - \underline{h}(\underline{x}))^T \underline{R}^{-1} (\underline{z} - \underline{h}(\underline{x})) = \frac{1}{2} \sum_{i=1}^m \frac{(z_i - h_i(\underline{x}))^2}{\sigma_i^2} \quad (48)$$

We can apply first order conditions, which means that all first derivatives of the objective function with respect to decision variables must be zero, i.e., that  $\nabla_{\underline{x}} J = \underline{0}$ . That is,

$$\nabla_{\underline{x}} J = \frac{\partial J}{\partial \underline{x}} = \begin{bmatrix} \frac{\partial J}{\partial x_1} \\ \vdots \\ \frac{\partial J}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \quad (49)$$

For a single element in  $\nabla_{\underline{x}} J$ , we have:

$$J = \frac{1}{2} \sum_{i=1}^m \frac{(z_i - h_i(\underline{x}))^2}{\sigma_i^2} \quad (50)$$

$$\frac{\partial J}{\partial x_1} = \frac{1}{2} \sum_{i=1}^m \frac{-2(z_i - h_i(\underline{x}))}{\sigma_i^2} \frac{\partial h_i(\underline{x})}{\partial x_1} = \sum_{i=1}^m \frac{-(z_i - h_i(\underline{x}))}{\sigma_i^2} \frac{\partial h_i(\underline{x})}{\partial x_1} \quad (51)$$

This can be written in matrix form as

$$\frac{\partial J}{\partial x_1} = - \begin{bmatrix} \frac{\partial h_1(\underline{x})}{\partial x_1} & \frac{\partial h_2(\underline{x})}{\partial x_1} & \dots & \frac{\partial h_m(\underline{x})}{\partial x_1} \end{bmatrix} \underline{R}^{-1} \begin{bmatrix} z_1 - h_1(\underline{x}) \\ z_2 - h_2(\underline{x}) \\ \vdots \\ z_m - h_m(\underline{x}) \end{bmatrix} \quad (52)$$

And we then see how to write the vector of derivatives, according to:

$$\frac{\partial J}{\partial \underline{x}} = - \begin{bmatrix} \frac{\partial h_1(\underline{x})}{\partial x_1} & \frac{\partial h_2(\underline{x})}{\partial x_1} & \dots & \frac{\partial h_m(\underline{x})}{\partial x_1} \\ \frac{\partial h_1(\underline{x})}{\partial x_2} & \frac{\partial h_2(\underline{x})}{\partial x_2} & \dots & \frac{\partial h_m(\underline{x})}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_1(\underline{x})}{\partial x_n} & \frac{\partial h_2(\underline{x})}{\partial x_n} & \dots & \frac{\partial h_m(\underline{x})}{\partial x_n} \end{bmatrix} \underline{R}^{-1} \begin{bmatrix} z_1 - h_1(\underline{x}) \\ z_2 - h_2(\underline{x}) \\ \vdots \\ z_m - h_m(\underline{x}) \end{bmatrix} \quad (53)$$

We recognize the matrix of partial derivatives in (53) as a sort of Jacobian matrix but

- (a) it is  $n \times m$ , i.e., it is not square and
- (b) unlike standard Jacobian, here the rows vary with variable  $(x_1, x_2, \dots)$ , not function  $(h_1, h_2, \dots)$ .

Let's define a matrix  $\underline{H}$  that does not have the second (b) attribute, i.e.,

$$\underline{H} = \begin{bmatrix} \frac{\partial h_1(\underline{x})}{\partial x_1} & \frac{\partial h_1(\underline{x})}{\partial x_2} & \dots & \frac{\partial h_1(\underline{x})}{\partial x_n} \\ \frac{\partial h_2(\underline{x})}{\partial x_1} & \frac{\partial h_2(\underline{x})}{\partial x_2} & \dots & \frac{\partial h_2(\underline{x})}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_m(\underline{x})}{\partial x_1} & \frac{\partial h_m(\underline{x})}{\partial x_2} & \dots & \frac{\partial h_m(\underline{x})}{\partial x_n} \end{bmatrix} \quad (54)$$

Note that  $\underline{H}$  is  $m \times n$ , and it is the negative transpose of the first matrix in eq. (53).

Then we see that the optimality condition can be written as:

$$\nabla_{\underline{x}} J = \frac{\partial J}{\partial \underline{x}} = \begin{bmatrix} \frac{\partial J}{\partial x_1} \\ \vdots \\ \frac{\partial J}{\partial x_n} \end{bmatrix} = -\underline{H}^T(x)\underline{R}^{-1}[(\underline{z} - \underline{h}(x))] = \underline{0} \quad (55)$$

The solution to eq. (55) will yield the estimated state vector  $\underline{x}$  which minimizes the square error.

Because there are  $n$  elements in the partial-J vector on the left, we observe that eq. (55) gives  $n$  equations. Since there are  $n$  variables in  $\underline{x}$ , it is possible to solve eq. (55) explicitly for  $\underline{x}$ .

Now we need to determine a solution procedure for eq. (55). To do so, let's define the left-hand-side of eq. (55) as  $\underline{G}(\underline{x})$ , i.e.,

$$\underline{G}(x) = -\underline{H}^T(x)\underline{R}^{-1}[(\underline{z} - \underline{h}(x))] = \underline{0} \quad (56)$$

Perform a Taylor series expansion of  $\underline{G}(\underline{x})$  around a certain solution  $\underline{x}_0$ .

$$\underline{G}(\underline{x}_0 + \Delta\underline{x}) = \underline{G}(\underline{x}_0) + \nabla_{\underline{x}} \underline{G}(\underline{x}) \Big|_{\underline{x}_0} \Delta\underline{x} + h.o.t. = \underline{0} \quad (57)$$

Note that eq. (57) indicates that if  $\underline{x}_0 + \Delta\underline{x}$  is to be a solution, then the right-hand-side of eq. (57) must be zero.

Recall that in a Taylor series expansion, the higher order terms contain products of  $\Delta\underline{x}$ , and so if  $\Delta\underline{x}$  is relatively small, terms containing products of  $\Delta\underline{x}$  will be very small, and in fact, negligible. So we will neglect the h.o.t. in eq. (57). This results in:

$$\underline{G}(\underline{x}_0 + \Delta\underline{x}) = \underline{G}(\underline{x}_0) + \nabla_{\underline{x}} \underline{G}(\underline{x}) \Big|_{\underline{x}_0} \Delta\underline{x} = \underline{0} \quad (58)$$

Since eq. (58) is nonlinear, we must resort to an iterative algorithm to solve it. We will use a Newton-type algorithm

Let's assume that we can make a pretty good guess at the solution to eq. (58), i.e., that the difference between our guess and the real solution is relative small.

Denote this guess as  $\underline{x}^{(k)}$ . Because it is not the solution,  $\underline{G}(\underline{x}^{(k)}) \neq \underline{0}$ .

So we want a better guess. Denote the better guess as  $\underline{x}^{(k+1)}$ . The difference between the old guess  $\underline{x}^{(k)}$  and the new guess  $\underline{x}^{(k+1)}$  is  $\Delta \underline{x}^{(k+1)}$ , i.e.,

$$\underline{x}^{(k+1)} = \underline{x}^{(k)} + \Delta \underline{x} \quad (59)$$

Or we can write

$$\Delta \underline{x} = \underline{x}^{(k+1)} - \underline{x}^{(k)} \quad (60)$$

Evaluating  $\underline{G}$  at the better guess, we have

$$\underline{G}(\underline{x}^{(k+1)}) = \underline{G}(\underline{x}^{(k)} + \Delta \underline{x}) = \underline{G}(\underline{x}^{(k)}) + \nabla_{\underline{x}} \underline{G}(\underline{x}) \Big|_{\underline{x}^{(k)}} \Delta \underline{x} \quad (61)$$

We desire that  $\underline{G}(\underline{x}^{(k+1)}) = \underline{0}$ . Under this desired condition, eq. (61) becomes:

$$\underline{G}(\underline{x}^{(k+1)}) = \underline{G}(\underline{x}^{(k)} + \Delta \underline{x}) = \underline{G}(\underline{x}^{(k)}) + \nabla_{\underline{x}} \underline{G}(\underline{x}) \Big|_{\underline{x}^{(k)}} \Delta \underline{x} = \underline{0} \quad (62)$$

Solving for  $\nabla_{\underline{x}} \underline{G}(\underline{x}) \Big|_{\underline{x}^{(k)}}$ , we have:

$$\nabla_{\underline{x}} \underline{G}(\underline{x}) \Big|_{\underline{x}^{(k)}} \Delta \underline{x} = -\underline{G}(\underline{x}^{(k)}) \quad (63)$$

In considering eq. (63), we already understand the right-hand-side, this is just the negative of eq. (56), evaluated at  $\underline{x}^{(k)}$ , i.e.,

$$-\underline{G}(\underline{x}^{(k)}) = \underline{H}^T(\underline{x}^{(k)}) \underline{R}^{-1} \left[ \left( \underline{z} - \underline{h}(\underline{x}^{(k)}) \right) \right] \quad (64)$$

There are  $n$  functional expressions in eq. (64).

But what is  $\nabla_{\underline{x}} \underline{G}(\underline{x})|_{\underline{x}^{(k)}}$ ? This is the derivatives, with respect to each of the  $n$  state variables, of each of the  $n$  functional expressions in eq. (56):

$$\underline{G}(\underline{x}) = -\underline{H}^T(\underline{x})\underline{R}^{-1}[(\underline{z} - \underline{h}(\underline{x}))] = \underline{0} \quad (56)$$

Since there are  $n$  functional expressions and  $n$  derivatives to take for each one, we can see that  $\underline{G}(\underline{x})$  will be  $n \times n$ , a square matrix.

Remembering eq. (55)

$$\nabla_{\underline{x}} J = \frac{\partial J}{\partial \underline{x}} = \begin{bmatrix} \frac{\partial J}{\partial x_1} \\ \vdots \\ \frac{\partial J}{\partial x_n} \end{bmatrix} = -\underline{H}^T(\underline{x})\underline{R}^{-1}[(\underline{z} - \underline{h}(\underline{x}))] = \underline{0} \quad (55)$$

reminds us that  $\underline{G}(\underline{x})$  are also derivatives with respect to each of the  $n$  state variables.

Therefore,  $\nabla_{\underline{x}} \underline{G}(\underline{x})|_{\underline{x}^{(k)}}$  are second derivatives of  $J$  with respect to the state variables.

Can we obtain a form for these second derivatives. Let's start from eq. (56):

$$\underline{G}(\underline{x}) = -\underline{H}^T(\underline{x})\underline{R}^{-1}[(\underline{z} - \underline{h}(\underline{x}))] \quad (56)$$

So what we want is:

$$\nabla_{\underline{x}} \underline{G}(\underline{x}) = \frac{\partial \underline{G}(\underline{x})}{\partial \underline{x}} = \frac{\partial}{\partial \underline{x}} \left\{ -\underline{H}^T(\underline{x}) \underline{R}^{-1} [(\underline{z} - \underline{h}(\underline{x}))] \right\} \quad (65)$$

The differentiation of what is inside the brackets of eq. (65) is formidable. We will make it easier for ourselves by assuming that  $\underline{H}(\underline{x})$  is a constant matrix. This implies  $\underline{H}(\underline{x} + \Delta \underline{x}) \approx \underline{H}(\underline{x})$ , i.e., the derivatives of the power flow equations do not change.

Remember, from eq. (54),

$$\underline{H} = \frac{\partial \underline{h}(\underline{x})}{\partial \underline{x}} = \begin{bmatrix} \frac{\partial h_1(\underline{x})}{\partial x_1} & \frac{\partial h_1(\underline{x})}{\partial x_2} & \dots & \frac{\partial h_1(\underline{x})}{\partial x_n} \\ \frac{\partial h_2(\underline{x})}{\partial x_1} & \frac{\partial h_2(\underline{x})}{\partial x_2} & \dots & \frac{\partial h_2(\underline{x})}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial h_m(\underline{x})}{\partial x_1} & \frac{\partial h_m(\underline{x})}{\partial x_2} & \dots & \frac{\partial h_m(\underline{x})}{\partial x_n} \end{bmatrix} \quad (54)$$

where the functions  $h_i$  are the expressions for the measurements (voltages, real and reactive power flows) in terms of the states (angles and voltage magnitudes).

So  $\underline{H}$  is really a power flow Jacobian matrix. It is well known that the power flow Jacobian is relatively insensitive to relatively small variations in state.

With the above assumption, differentiating the right-hand-side of eq. (63) becomes not-so-bad:

$$\begin{aligned}\nabla_{\underline{x}} \underline{G}(\underline{x}) &= \frac{\partial \underline{G}(\underline{x})}{\partial \underline{x}} = \frac{\partial}{\partial \underline{x}} \left\{ -\underline{H}^T(\underline{x}) \underline{R}^{-1} [(\underline{z} - \underline{h}(\underline{x}))] \right\} \\ &= -\underline{H}^T(\underline{x}) \underline{R}^{-1} \left( -\frac{\partial \underline{h}(\underline{x})}{\partial \underline{x}} \right) = \underline{H}^T(\underline{x}) \underline{R}^{-1} \frac{\partial \underline{h}(\underline{x})}{\partial \underline{x}}\end{aligned}\quad (66)$$

But we recognize from eq. (54) the term  $\frac{\partial \underline{h}(\underline{x})}{\partial \underline{x}}$  in eq. (66) as  $\underline{H}$ . Therefore, eq. (66) becomes:

$$\nabla_{\underline{x}} \underline{G}(\underline{x}) = \underline{H}^T(\underline{x}) \underline{R}^{-1} \underline{H}(\underline{x})\quad (67)$$

Making this substitution into eq. (61) results in:

$$\nabla_{\underline{x}} \underline{G}(\underline{x}) \Big|_{\underline{x}^{(k)}} \Delta \underline{x} = -\underline{G}(\underline{x}^{(k)})\quad (63)$$

$$\underline{H}^T(\underline{x}) \underline{R}^{-1} \underline{H}(\underline{x}) \Big|_{\underline{x}^{(k)}} \Delta \underline{x} = -\underline{G}(\underline{x}^{(k)})\quad (68)$$

Finally, replacing the right-hand-side of eq. (68) with eq. (54) evaluated at  $\underline{x}^{(k)}$  yields:

$$\underline{H}^T(\underline{x}) \underline{R}^{-1} \underline{H}(\underline{x}) \Big|_{\underline{x}^{(k)}} \Delta \underline{x} = \underline{H}^T(\underline{x}) \underline{R}^{-1} [(\underline{z} - \underline{h}(\underline{x}))] \Big|_{\underline{x}^{(k)}}\quad (69)$$

Equation (69) provides a way to solve for  $\Delta \underline{x}$ .

## 6.0 Solution Algorithm

Given:

- measurements  $\underline{z} \quad [z_1, \dots, z_m]$
- standard deviations  $\sigma_1, \dots, \sigma_m$



• the network

Compute : state estimate  $\underline{x}$   $[x_1, \dots, x_n]$

(this is all voltage magnitudes and all voltage angles except for swing bus angle)

1. Form measurement expressions  $\underline{h}(\underline{x})$

2. Form derivative expressions  $\underline{H} = \frac{\partial \underline{h}(\underline{x})}{\partial \underline{x}}$

3. Form  $\underline{R}$

4. Let  $k=0$ . Guess solution  $\underline{x}^{(0)}$ .

5. Compute  $\underline{H}(\underline{x}^{(k)})$ ,  $\underline{h}(\underline{x}^{(k)})$

6. Compute  $\underline{A} = \underline{H}^T(\underline{x})\underline{R}^{-1}\underline{H}(\underline{x})\big|_{\underline{x}^{(k)}}$ ,  $\underline{b} = \underline{H}^T(\underline{x})\underline{R}^{-1}[(\underline{z} - \underline{h}(\underline{x}))]\big|_{\underline{x}^{(k)}}$

7. Solve  $\underline{A}\Delta\underline{x} = \underline{b}$  for  $\Delta\underline{x}$ .

8. Compute  $\underline{x}^{(k+1)} = \underline{x}^{(k)} + \Delta\underline{x}$

9. If  $\max_i \{|\Delta x_i|\} > \varepsilon$  then

$k=k+1$

Go to 5

Else Stop

Homework #10:

For the lossless network shown below, the following data is given:

$$z_1 = V_1 = 1.02, \sigma_1 = 0.1$$

$$z_2 = V_2 = 1.0, \sigma_2 = 0.1$$

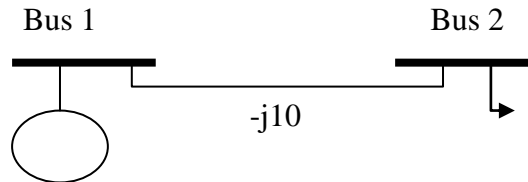
$$z_3 = P_{12} = 2.0, \sigma_3 = 0.05$$

$$z_4 = Q_{12} = 0.2, \sigma_4 = 0.05$$

Let

$$\underline{x}^{(0)} = \begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \\ x_3^{(0)} \end{bmatrix} = \begin{bmatrix} \theta_2^{(0)} \\ V_1^{(0)} \\ V_2^{(0)} \end{bmatrix} = \begin{bmatrix} 0 \\ 1.02 \\ 1.0 \end{bmatrix}$$

and perform one iteration of the solution procedure to find  $\underline{x}^{(1)}$ .



## References:

- 
- [1] F. Schweppe, J. Wildes, and D. Rom, "Power system static state estimation: Parts I, II, and III," Power Industry Computer Conference (PICA), Denver, Colorado June, 1969.
  - [2] A. Monticelli, "State Estimation in Electric Power Systems: A Generalized Approach," Kluwer, Boston, 1999.
  - [3] G. Strang, "Linear algebra and its applications," third edition, Harcourt Brace, 1988.
  - [4] A. Papoulis, "Probability, Random Variables, and Stochastic Processes," McGraw-Hill, New York, 1984.