CHAPTER 2

PROBLEM FORMATION

2.1 Development of the notions of state estimation

Statistical estimation is a procedure that calculates the value of one or more unknown parameters in a system by using measurements from the system. Since the measurements are inexact, then the values of the unknown parameters are also inexact. Hence, a "best estimate" of the unknown parametric values given the available measurements needs to be formulated. The unknown parameters in a power system are the so called state variables.

The development of the notions of state estimation is dependent on the statistical criterion that is selected. Of the many criteria that have been examined and used in various applications, there are three most commonly encountered criteria .

The maximum likelihood criterion

The objective is to maximize the probability that the estimates of the state variables are the true values of the state variables.

The weighted least square criterion

The objective is to minimize the sum of the squares of the weighted deviations of the estimated measurements from the actual measurements.

The minimum variance criterion

The objective is to minimize the expected value of the sum of the squares of the deviations of the estimated components of the state variables vector from the corresponding components of the true state variables vector.

Each of these approaches results in identical estimators when the measurements errors are normally distributed and unbiased meter error distributions are assumed. The maximum likelihood approach is utilized since the method introduces the measurement error weighting matrix in a straightforward manner (Wood and Wollenberg, 1996).

In this approach, the probability that the estimates of the state variables are the true values of the state variables depends on the random error in the measurement device as well as the unknown parameters to be estimated. Therefore, a reasonable procedure would be one that simply chooses the estimate as the value that maximizes the probability (Wood and Wollenberg, 1996).

2.1.1 The principle of maximum likelihood estimation

The maximum likelihood estimator (Wood and Wollenberg, 1996) assumes that the probability density function (PDF) of the random errors in the measurements is known. The development of the state formula using the maximum likelihood criterion is by assuming that the measurement errors are normally distributed. Hence, the result will be a “weighted least squares” estimation formula.

The measurements are assumed to be in error, where the value obtained from the measurement device is close to the true value of the parameter being measured but differs by an unknown error. Mathematically, it can be modeled as

$$z^{meas} = z^{true} + h \, , \tag{2.1}$$

where z^{meas} is the measurement received from a measurement device, z^{true} is the true value of the quantity being measured, and h is the random measurement error.

The random number h serves to model the uncertainty in the measurements. The probability distribution of h is chosen to be a normal distribution with zero mean. The normal distribution is commonly used for modeling measurement error since it is the distribution that is followed when many factors contribute to the overall error.

The probability density function of h is

$$\mathrm{PDF}(h) = \frac{1}{s\sqrt{2p}} \exp\left(\frac{-h^2}{2s^2}\right), \quad (2.2)$$

where s is the standard deviation and s^2 is the variance of the random number. $\mathrm{PDF}(h)$ describes the behavior of h, and together with the standard deviation, s, provides a way to model the seriousness of the random measurement error. If s is large, then the measurement is relatively inaccurate, whereas a small value of s implies a small error spread.

A simple DC (Direct current) circuit example is used to illustrate the principle of maximum likelihood estimation of state estimation.

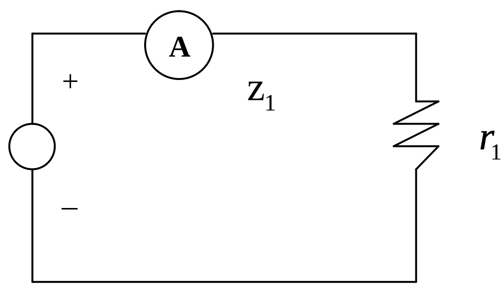

Figure 2.1: Simple DC circuit with current measurement

In this example, the objective is to estimate the value of the DC voltage source, x^{true}. The ammeter used is with an error having a known standard deviation. The ammeter gives a reading z_1^{meas} which is the sum of the true current flowing in the circuit, z_1^{true} and the error present in the ammeter, h_1. It can be written as

$$z_1^{meas} = z_1^{true} + h_1. \quad (2.3)$$

As mentioned above, the mean value of h_1 is zero, and hence the mean value of z_1^{meas} is equal to the mean value of z_1^{true}. Thus, the probability density function for z_1^{meas} can be written as

$$\mathrm{PDF}(z_1^{meas}) = \frac{1}{s_1\sqrt{2p}} \exp\left[\frac{-(z_1^{meas} - z_1^{true})^2}{2s_1^2}\right], \quad (2.4)$$

where s_1 is the standard deviation of the random error h_1. Assuming that the value of the resistance, r_1 in the circuit is known, then the probability density function for z_1^{meas} becomes

$$\mathrm{PDF}(z_1^{meas}) = \frac{1}{s_1\sqrt{2p}}\exp\left[\frac{-\left(z_1^{meas} - \frac{1}{r_1}x\right)^2}{2s_1^2}\right]. \tag{2.5}$$

From the definition of maximum likelihood, one wishes to find an estimate of x, x^{est} that maximizes the probability that the observed measurement z_1^{meas} would occur. By using $\mathrm{PDF}(z_1^{meas})$, the probability of z_1^{meas} can be written as

$$\mathrm{prob}(z_1^{meas}) = \int_{z_1^{meas}}^{z_1^{meas}+dz_1^{meas}} \mathrm{PDF}(z_1^{meas})\mathrm{d}z_1^{meas} \qquad \text{as } \mathrm{d}z_1^{meas} \to 0$$

$$= \mathrm{PDF}(z_1^{meas})\mathrm{d}z_1^{meas}. \tag{2.6}$$

The maximum likelihood procedure then requires maximization of the value of the $\mathrm{prob}(z_1^{meas})$, which is a function of x:

$$\max_x \mathrm{prob}(z_1^{meas}) = \max_x \mathrm{PDF}(z_1^{meas})\mathrm{d}z_1^{meas}. \tag{2.7}$$

One convenient transformation that can be used at this point is to maximize the natural logarithm of $\mathrm{PDF}(z_1^{meas})$ since this will also maximize $\mathrm{PDF}(z_1^{meas})$. Hence, the transformation can be written as

$$\mathrm{Ln}\left(\mathrm{PDF}(z_1^{meas})\right) = -\mathrm{Ln}(s_1\sqrt{2p}) - \frac{\left(z_1^{meas} - \frac{1}{r_1}x\right)^2}{2s_1^2}. \tag{2.8}$$

Then one wishes to find the maximum of the natural logarithm of $\mathrm{PDF}(z_1^{meas})$,

$$\max_x \mathrm{Ln}\left(\mathrm{prob}(z_1^{meas})\right) = \max_x\left(-\mathrm{Ln}\left(s_1\sqrt{2p}\right) - \frac{\left(z_1^{meas} - \frac{1}{r_1}x\right)^2}{2s_1^2}\right). \tag{2.9}$$

Since the first term in the right hand side of eq. 2.9, $\mathrm{Ln}\left(s_1\sqrt{2p}\right)$ is a constant, then the right hand side function can be maximized by minimizing the second term, $\left(-\frac{\left(z_1^{meas}-\frac{1}{r_1}x\right)^2}{2s_1^2}\right)$ since it has a negative coefficient. This means

$$\max_x\left(-\mathrm{Ln}\left(s_1\sqrt{2p}\right)-\frac{\left(z_1^{meas}-\frac{1}{r_1}x\right)^2}{2s_1^2}\right)=\min_x\left(\frac{\left(z_1^{meas}-\frac{1}{r_1}x\right)^2}{2s_1^2}\right). \tag{2.10}$$

The minimum of the function in eq. 2.10 can be obtained by minimizing the value of x. And the minimum value of x is found by simply taking the first derivative and setting the result to zero:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\left(z_1^{meas}-\frac{1}{r_1}x\right)^2}{2s_1^2}\right)=\frac{-\left(z_1^{meas}-\frac{1}{r_1}x\right)}{r_1 s_1^2}=0. \tag{2.11}$$

Hence, the maximum likelihood estimate of the voltage is simply the product of the measured current and the known resistance, that is, $x^{est}=r_1 z_1^{meas}$.

However, the situation changes if a second measurement circuit (Figure 2.2) is added as the best estimate is not so obvious.

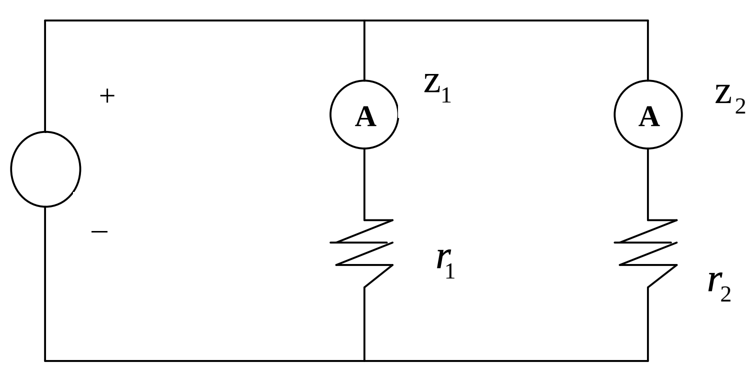

Figure 2.2: DC circuit with two current measurements

Assuming that the values of both resistances, r_1 and r_2 are known, then each meter reading can be modeled as the sum of the true value and a random error as below:

$$z_1^{meas} = z_1^{true} + h_1 ,$$
$$z_2^{meas} = z_2^{true} + h_2 , \tag{2.12}$$

where the errors, h_1 and h_2 will be represented as independent zero mean, normally distributed random variables with probability density functions:

$$\mathrm{PDF}(h_1) = \frac{1}{s_1\sqrt{2p}}\exp(\frac{-h_1^{\ 2}}{2s_1^{\ 2}}),$$
$$\mathrm{PDF}(h_2) = \frac{1}{s_2\sqrt{2p}}\exp(\frac{-h_2^{\ 2}}{2s_2^{\ 2}}). \tag{2.13}$$

As before, the probability density functions of z_1^{meas} and z_2^{meas} can be written as:

$$\mathrm{PDF}(z_1^{meas}) = \frac{1}{s_1\sqrt{2p}}\exp\left[\frac{-(z_1^{meas} - \frac{1}{r_1}x)^2}{2s_1^{\ 2}}\right],$$
$$\mathrm{PDF}(z_2^{meas}) = \frac{1}{s_2\sqrt{2p}}\exp\left[\frac{-(z_2^{meas} - \frac{1}{r_2}x)^2}{2s_2^{\ 2}}\right]. \tag{2.14}$$

The maximum likelihood function is the probability of obtaining the measurements z_1^{meas} and z_2^{meas}. Under the assumption that the random errors h_1 and h_2 are independent random variables, the probability of obtaining z_1^{meas} and z_2^{meas} is simply the probability of obtaining z_1^{meas} times the probability of obtaining z_2^{meas}.

$$\mathrm{prob}\left(z_1^{meas} \cap z_2^{meas}\right) = \mathrm{prob}\left(z_1^{meas}\right) \times \mathrm{prob}\left(z_2^{meas}\right)$$
$$= \mathrm{PDF}\left(z_1^{meas}\right)\mathrm{PDF}\left(z_2^{meas}\right)\mathrm{dz}_1^{meas}\mathrm{dz}_2^{meas}$$

$$= \frac{1}{s_1\sqrt{2p}} \exp\left[\frac{-(z_1^{meas} - \frac{1}{r_1}x)^2}{2s_1^2}\right]$$

$$\times \frac{1}{s_2\sqrt{2p}} \exp\left[\frac{-(z_2^{meas} - \frac{1}{r_2}x)^2}{2s_2^2}\right] \mathrm{d}z_1^{meas}\mathrm{d}z_2^{meas}. \tag{2.15}$$

Again, to maximize the function, the natural logarithm of $\mathrm{PDF}\left(z_1^{meas}\right)\mathrm{PDF}\left(z_2^{meas}\right)$ is taken:

$$\mathrm{Ln}\left(\mathrm{PDF}\left(z_1^{meas}\right)\mathrm{PDF}\left(z_2^{meas}\right)\right) = -\mathrm{Ln}(s_1\sqrt{2p}) - \frac{\left(z_1^{meas} - \frac{1}{r_1}x\right)^2}{2s_1^2}$$

$$-\mathrm{Ln}(s_2\sqrt{2p}) - \frac{\left(z_2^{meas} - \frac{1}{r_2}x\right)^2}{2s_2^2}, \tag{2.16}$$

and

$$\max \mathrm{prob}\left(z_1^{meas} \cap z_2^{meas}\right)$$

$$= \max_x \left[-\mathrm{Ln}(s_1\sqrt{2p}) - \frac{\left(z_1^{meas} - \frac{1}{r_1}x\right)^2}{2s_1^2} - \mathrm{Ln}(s_2\sqrt{2p}) - \frac{\left(z_2^{meas} - \frac{1}{r_2}x\right)^2}{2s_2^2}\right]$$

$$= \min_x \left[\frac{\left(z_1^{meas} - \frac{1}{r_1}x\right)^2}{2s_1^2} + \frac{\left(z_2^{meas} - \frac{1}{r_2}x\right)^2}{2s_2^2}\right]. \tag{2.17}$$

As before, the minimum sought is found by simply taking the first derivative and setting the result to zero:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\left(z_1^{meas}-\frac{1}{r_1}x\right)^2}{2s_1^2}+\frac{\left(z_2^{meas}-\frac{1}{r_2}x\right)^2}{2s_2^2}\right)=0\,,$$

or
$$-\frac{\left(z_1^{meas}-\frac{1}{r_1}x\right)}{r_1 s_1^2}-\frac{\left(z_2^{meas}-\frac{1}{r_2}x\right)}{r_2 s_2^2}=0\,. \tag{2.18}$$

Thus, the estimate of x is:

$$x^{est}=\frac{\left(\frac{z_1^{meas}}{r_1 s_1^2}+\frac{z_2^{meas}}{r_2 s_2^2}\right)}{\left(\frac{1}{r_1^2 s_1^2}+\frac{1}{r_2^2 s_2^2}\right)}. \tag{2.19}$$

If one of the ammeters is of superior quality, its variance will be much smaller than that of the other meter. It is obvious that the estimation problem does not need to be expressed as a maximum of the product of probability density functions. Instead, a direct way of writing what is needed can be observed from eqs. 2.10 and 2.17. In these equations, the maximum likelihood estimate of the unknown parameter is always expressed as the value of the parameter that gives the minimum of the sum of the squares of the difference between each measured value and the true value being measured with each squared difference divided or “weighted” by the variance of the meter error. Thus, the estimation of a single parameter, x using N_m measurements, can be expressed as:

$$\min_x \mathrm{J}(x)=\sum_{i=1}^{N_m}\frac{\left[z_i^{meas}-\mathrm{f}_i(x)\right]^2}{s_i^2}\quad, \tag{2.20}$$

where

f_i = function that is used to calculate the value being measured by the i_{th} measurement;

s_i^2 = variance for the i_{th} measurement;

$J(x)$ = measurement residual;

N_m = number of measurements;

z_i^{meas} = i_{th} measured quantity.

Eq. 2.20 can also be expressed in per unit or in physical units such as MW for active power, MVAR for reactive power, or kV for voltage.

If one tries to estimate N_s unknown parameters (N_s >1) using N_m measurements, then the estimation of the unknown parameters can be expressed as:

$$\min_{\{x_1, x_2, x_3 ...., x_{N_s}\}} J(x_1, x_2, x_3 ...., x_{N_s}) = \sum_{i=1}^{N_m} \frac{\left[z_i^{meas} - f_i(x_1, x_2, x_3 ...., x_{N_s})\right]^2}{s_i^2} . \qquad (2.21)$$

The estimation calculation shown in eqs. 2.20 and 2.21 are known as "weighted least squares" estimator which is equivalent to a maximum likelihood estimator if the measurement errors are modeled as normally distributed random numbers.

2.2 Matrix formulation:

If the functions $f_i(x_1, x_2, x_3 ...., x_{N_s})$ are linear functions, then eq. 2.21 has a closed form solution. The functions $f_i(x_1, x_2, x_3 ...., x_{N_s})$ can be written as:

$$f_i(x_1, x_2, x_3 ...., x_{N_s}) = f_i(\mathbf{x}) = h_{i1}\, x_1 + h_{i2}\, x_2 + h_{i3}\, x_3 + ... + h_{iN_s}\, x_{N_s} , \qquad (2.22)$$

where h_{ij} (j =1, 2, …, N_s) are arbitrary constants.

Placing all the f_i functions in a vector gives us

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_{N_m}(\mathbf{x}) \end{bmatrix} = \mathbf{H}\,\mathbf{x} , \qquad (2.23)$$

where

$\mathbf{H}$ = N_m by N_s matrix containing the coefficients of the linear functions $f_i(\mathbf{x})$;

N_m = number of measurements;

N_s = number of unknown parameter being estimated.

By placing the measurements in a vector:

$$\mathbf{z}^{meas} = \begin{bmatrix} z_1^{meas} \\ z_2^{meas} \\ \vdots \\ z_{N_m}^{meas} \end{bmatrix}, \tag{2.24}$$

eq. 2.21 can be written in a compact form:

$$\min_{\mathbf{x}} \mathrm{J}(\mathbf{x}) = \left[\mathbf{z}^{meas} - \mathbf{f}(\mathbf{x})\right]^T \left[\mathbf{R}^{-1}\right]\left[\mathbf{z}^{meas} - \mathbf{f}(\mathbf{x})\right], \tag{2.25}$$

where

$$\mathbf{R} = \begin{bmatrix} s_1^2 & & & \\ & s_2^2 & & \\ & & \ddots & \\ & & & s_{N_m}^2 \end{bmatrix}.$$

$\mathbf{R}$ is the covariance matrix of the measurement errors. In order to obtain the general expression for the minimum in eq. 2.25, the expression is expanded and $\mathbf{f}(\mathbf{x})$ is substituted with $\mathbf{H}\,\mathbf{x}$ (from eq. 2.23):

$$\min_{\mathbf{x}} \mathrm{J}(\mathbf{x}) = \left\{\left(\mathbf{z}^{meas}\right)^T \mathbf{R}^{-1}\mathbf{z}^{meas} - \mathbf{x}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{z}^{meas} - \left(\mathbf{z}^{meas}\right)^T\mathbf{R}^{-1}\mathbf{H}\,\mathbf{x} + \mathbf{x}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\,\mathbf{x}\right\} \tag{2.26}$$

Since the second and the third term in eq. 2.26 are identical, then

$$\min_{\mathbf{x}} \mathrm{J}(\mathbf{x}) = \left\{\left(\mathbf{z}^{meas}\right)^T \mathbf{R}^{-1}\mathbf{z}^{meas} - 2\left(\mathbf{z}^{meas}\right)^T \mathbf{R}^{-1}\mathbf{H}\,\mathbf{x} + \mathbf{x}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\,\mathbf{x}\right\}. \tag{2.27}$$

To find the value of $\mathbf{x}$ that minimizes $\mathrm{J}(\mathbf{x})$, the first derivative of $\mathrm{J}(\mathbf{x})$ with respect to each x_i (i = 1, 2,… N_s) is taken and set to zero. That is

$$\frac{\partial\,\mathrm{J}(\mathbf{x})}{\partial x_i} = 0 \text{ for } i = 1, 2, \ldots\ N_s\,. \tag{2.28}$$

Placing these derivatives into a vector gives us what is called the gradient of $\mathrm{J}(\mathbf{x})$, $\nabla_x \mathrm{J}(\mathbf{x})$, which can be written as:

$$\nabla_x \mathrm{J}(\mathbf{x}) = \begin{bmatrix} \frac{\partial \mathrm{J}(\mathbf{x})}{\partial x_1} \\ \frac{\partial \mathrm{J}(\mathbf{x})}{\partial x_2} \\ \vdots \end{bmatrix}. \tag{2.29}$$

Then the goal of forcing each derivative to zero can be written as:

$$\nabla_x \mathrm{J}(\mathbf{x}) = \mathbf{0}. \tag{2.30}$$

The gradient of $\mathrm{J}(\mathbf{x})$ is

$$\nabla_x \mathrm{J}(\mathbf{x}) = -2\mathbf{H}^T \mathbf{R}^{-1} \mathbf{z}^{meas} + 2\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}\,\mathbf{x}. \tag{2.31}$$

Then $\nabla_x \mathrm{J}(\mathbf{x}) = \mathbf{0}$ gives the maximum likelihood estimate of $\mathbf{x}$:

$$\mathbf{x}^{est} = \left[\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}\right]^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{z}^{meas}. \tag{2.32}$$

Eq. 2.32 holds when $N_m > N_s$; that is, when the number of measurements being made is greater than number of parameters being estimated.

2.3 State Estimation of an AC network

An AC (Alternating current) network is an electrical network in which the energy is transmitted by alternating voltage. Alternating current is an electric current whose direction reverses cyclically, as opposed to DC, whose direction remains constant. Used generically, AC refers to the form in which electricity is delivered to businesses and residences.
(http://en.wikipedia.org/wiki/Alternating_current)

2.3.1 Development of the AC network estimation formula

Section 2.1 has shown how the maximum likelihood estimation scheme led to a weighted least squares calculation for measurements from a linear system, where the sum of the measurement residuals is minimized:

$$\min_{\mathbf{x}} \mathrm{J}(\mathbf{x}) = \sum_{i=1}^{N_m} \frac{\left[z_i^{meas} - \mathrm{f}_i(\mathbf{x})\right]^2}{s_i^2} . \tag{2.33}$$

In the case of a linear system, similar to the case in the DC network, the $\mathrm{f}_i(\mathbf{x})$ functions are themselves linear and the minimum of $\mathrm{J}(\mathbf{x})$ can be found directly. However, in an AC network, the measured quantities are active and reactive power on each end of each transmission line and at each load and generator, amperes, transformer tap position, and voltage magnitude. The equation for power entering a bus *i* is:

$$P_i + jQ_i = \sum_{k=1}^{N} \left\{ \begin{array}{l} |E_i||E_k|[G_{ik}\cos(q_i - q_k) + B_{ik}\sin(q_i - q_k)] \\ + j\left[|E_i||E_k|[G_{ik}\sin(q_i - q_k) + B_{ik}\cos(q_i - q_k)]\right] \end{array} \right\} , \tag{2.34}$$

where

P_i = active power;

Q_i = reactive power;

q_i, q_k = the phase angles at buses *i* and *k* respectively;

$|E_i|, |E_k|$ = the bus voltage magnitudes, respectively;

$G_{ik} + jB_{ik} = Y_{ik}$ is the *ik* term in the Y matrix of the power system.

This is clearly not a linear function of the voltage magnitude and phase angle at each bus. Therefore, the $\mathrm{f}_i(x)$ functions will be nonlinear functions, except for a voltage magnitude measurement where $\mathrm{f}_i(x)$ is simply unity times the particular x_i that corresponds to the voltage magnitude being measured. Thus, on a transmission line from bus *i* to bus *j*, the measurement residual $\mathrm{J}(\mathbf{x})$ can be written as:

$$\mathrm{J}(\mathbf{x}) = \frac{\left\{MW_{ij}^{meas} - \left[-|E_i|^2 G_{ij} + |E_i||E_j|[G_{ij}\cos(q_i - q_j) + B_{ij}\sin(q_i - q_j)]\right]\right\}^2}{s_{MW_{ij}}^2} \tag{2.35}$$

for active power flow measurement from bus i to j, MW, or

$$J(\mathbf{x}) = \frac{\left\{ MWAR_{ij}^{meas} - \left[|E_i|^2 [B_{ij} + B_{capij}] + |E_i||E_j| \left[G_{ij} \sin(\theta_i - \theta_j) - B_{ij} \cos(\theta_i - \theta_j) \right] \right] \right\}^2}{\sigma^2_{MWAR_{ij}}} \tag{2.36}$$

(B_{capij} is the total line charging susceptance) for reactive power flow measurement from bus i to j, MWAR, or

$$J(\mathbf{x}) = \frac{\left\{ |E_i|^{meas} - |E_i| \right\}^2}{\sigma^2_{|E_i|}} \tag{2.37}$$

for voltage magnitude measurement, kV.

Since there is no linear relationship between the state variables and the power flows in the network, then an iterative technique is needed to minimize the measurement residual, $J(\mathbf{x})$. A commonly used technique for power system state estimation is to calculate the gradient of $J(\mathbf{x})$ and then force it to zero by using Newton's method. Now, the gradient of $J(\mathbf{x})$ is:

$$\nabla_x \mathbf{J}(\mathbf{x}) = \begin{bmatrix} \frac{\partial J(\mathbf{x})}{\partial x_1} \\ \frac{\partial J(\mathbf{x})}{\partial x_2} \\ \frac{\partial J(\mathbf{x})}{\partial x_3} \\ \vdots \end{bmatrix}$$

$$= -2 \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \frac{\partial f_2(\mathbf{x})}{\partial x_1} & \frac{\partial f_3(\mathbf{x})}{\partial x_1} & \cdots \\ \frac{\partial f_1(\mathbf{x})}{\partial x_2} & \frac{\partial f_2(\mathbf{x})}{\partial x_2} & \frac{\partial f_3(\mathbf{x})}{\partial x_2} & \cdots \\ \frac{\partial f_1(\mathbf{x})}{\partial x_3} & \frac{\partial f_2(\mathbf{x})}{\partial x_3} & \frac{\partial f_3(\mathbf{x})}{\partial x_3} & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1^2} & & & \\ & \frac{1}{\sigma_2^2} & & \\ & & \frac{1}{\sigma_3^2} & \\ & & & \ddots \end{bmatrix} \begin{bmatrix} z_1 - f_1(\mathbf{x}) \\ z_2 - f_2(\mathbf{x}) \\ z_3 - f_3(\mathbf{x}) \\ \vdots \end{bmatrix}. \tag{2.38}$$

Then, the $f_i(\mathbf{x})$ functions are written in the vector form $\mathbf{f}(\mathbf{x})$ and the jacobian of $\mathbf{f}(\mathbf{x})$ is calculated as:

$$\mathbf{H} = \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \mathrm{f}_1(\mathbf{x})}{\partial x_1} & \frac{\partial \mathrm{f}_1(\mathbf{x})}{\partial x_2} & \frac{\partial \mathrm{f}_1(\mathbf{x})}{\partial x_3} & \cdots \\ \frac{\partial \mathrm{f}_2(\mathbf{x})}{\partial x_1} & \frac{\partial \mathrm{f}_2(\mathbf{x})}{\partial x_2} & \frac{\partial \mathrm{f}_2(\mathbf{x})}{\partial x_3} & \cdots \\ \frac{\partial \mathrm{f}_3(\mathbf{x})}{\partial x_1} & \frac{\partial \mathrm{f}_3(\mathbf{x})}{\partial x_2} & \frac{\partial \mathrm{f}_3(\mathbf{x})}{\partial x_3} & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}. \tag{2.39}$$

The transpose of $\mathbf{H}$ is

$$\mathbf{H}^T = \begin{bmatrix} \frac{\partial \mathrm{f}_1(\mathbf{x})}{\partial x_1} & \frac{\partial \mathrm{f}_2(\mathbf{x})}{\partial x_1} & \frac{\partial \mathrm{f}_3(\mathbf{x})}{\partial x_1} & \cdots \\ \frac{\partial \mathrm{f}_1(\mathbf{x})}{\partial x_2} & \frac{\partial \mathrm{f}_2(\mathbf{x})}{\partial x_2} & \frac{\partial \mathrm{f}_3(\mathbf{x})}{\partial x_2} & \cdots \\ \frac{\partial \mathrm{f}_1(\mathbf{x})}{\partial x_3} & \frac{\partial \mathrm{f}_2(\mathbf{x})}{\partial x_3} & \frac{\partial \mathrm{f}_3(\mathbf{x})}{\partial x_3} & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}. \tag{2.40}$$

Furthermore, the weighting matrix is

$$\mathbf{R}^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & & & \\ & \frac{1}{\sigma_2^2} & & \\ & & \frac{1}{\sigma_3^2} & \\ & & & \ddots \end{bmatrix}. \tag{2.41}$$

It follows that

$$\nabla_x \mathbf{J}(\mathbf{x}) = -2\mathbf{H}^T \mathbf{R}^{-1} \begin{bmatrix} z_1 - \mathrm{f}_1(\mathbf{x}) \\ z_2 - \mathrm{f}_2(\mathbf{x}) \\ z_3 - \mathrm{f}_3(\mathbf{x}) \\ \vdots \end{bmatrix}. \tag{2.42}$$

To make $\nabla_x \mathbf{J}(\mathbf{x})$ equal zero, Newton's method (see Appendix A) is applied, and yields

$$\Delta \mathbf{x} = \left[\frac{\partial \nabla_x \mathbf{J}(\mathbf{x})}{\partial \mathbf{x}} \right]^{-1} \left[-\nabla_x \mathbf{J}(\mathbf{x}) \right], \tag{2.43}$$

where $\Delta \mathbf{x}$ are small changes in $\mathbf{x}$. The jacobian of $\nabla_x \mathbf{J}(\mathbf{x})$ is calculated by treating $\mathbf{H}$ as a constant matrix:

$$\left[\frac{\partial \nabla_x \mathbf{J}(\mathbf{x})}{\partial \mathbf{x}}\right] = \frac{\partial}{\partial \mathbf{x}}\left\{-2\mathbf{H}^T\mathbf{R}^{-1}\begin{bmatrix} z_1 - \mathrm{f}_1(\mathbf{x}) \\ z_2 - \mathrm{f}_2(\mathbf{x}) \\ z_3 - \mathrm{f}_3(\mathbf{x}) \\ \vdots \end{bmatrix}\right\}$$

$$= -2\mathbf{H}^T\mathbf{R}^{-1}[-\mathbf{H}]$$

$$= 2\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\,. \tag{2.44}$$

Finally, the state estimator equation for an AC network is

$$\Delta\mathbf{x} = \frac{1}{2}\left[\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right]^{-1}\left\{2\mathbf{H}^T\mathbf{R}^{-1}\begin{bmatrix} z_1 - \mathrm{f}_1(\mathbf{x}) \\ z_2 - \mathrm{f}_2(\mathbf{x}) \\ z_3 - \mathrm{f}_3(\mathbf{x}) \\ \vdots \end{bmatrix}\right\}$$

or $$\Delta\mathbf{x} = \left[\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right]^{-1}\mathbf{H}^T\mathbf{R}^{-1}\begin{bmatrix} z_1 - \mathrm{f}_1(\mathbf{x}) \\ z_2 - \mathrm{f}_2(\mathbf{x}) \\ z_3 - \mathrm{f}_3(\mathbf{x}) \\ \vdots \end{bmatrix}$$

or $$\left[\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right][\Delta\mathbf{x}] = \mathbf{H}^T\mathbf{R}^{-1}\begin{bmatrix} z_1 - \mathrm{f}_1(\mathbf{x}) \\ z_2 - \mathrm{f}_2(\mathbf{x}) \\ z_3 - \mathrm{f}_3(\mathbf{x}) \\ \vdots \end{bmatrix}. \tag{2.45}$$

To solve the AC state estimation problem, eq. 2.45 is applied iteratively as shown in Figure 2.3. At each iteration, the state estimator equation (eq. 2.45) is solved by using the proposed orthogonal decomposition method using Householder transformation. Details on how it is solved are presented in section 3.3.2. The sum of the measurement residuals $\mathrm{J}(\mathbf{x})$ will be calculated at the beginning of each iteration. Its value represents a measure of the overall fit of the estimated values to the measurement values.

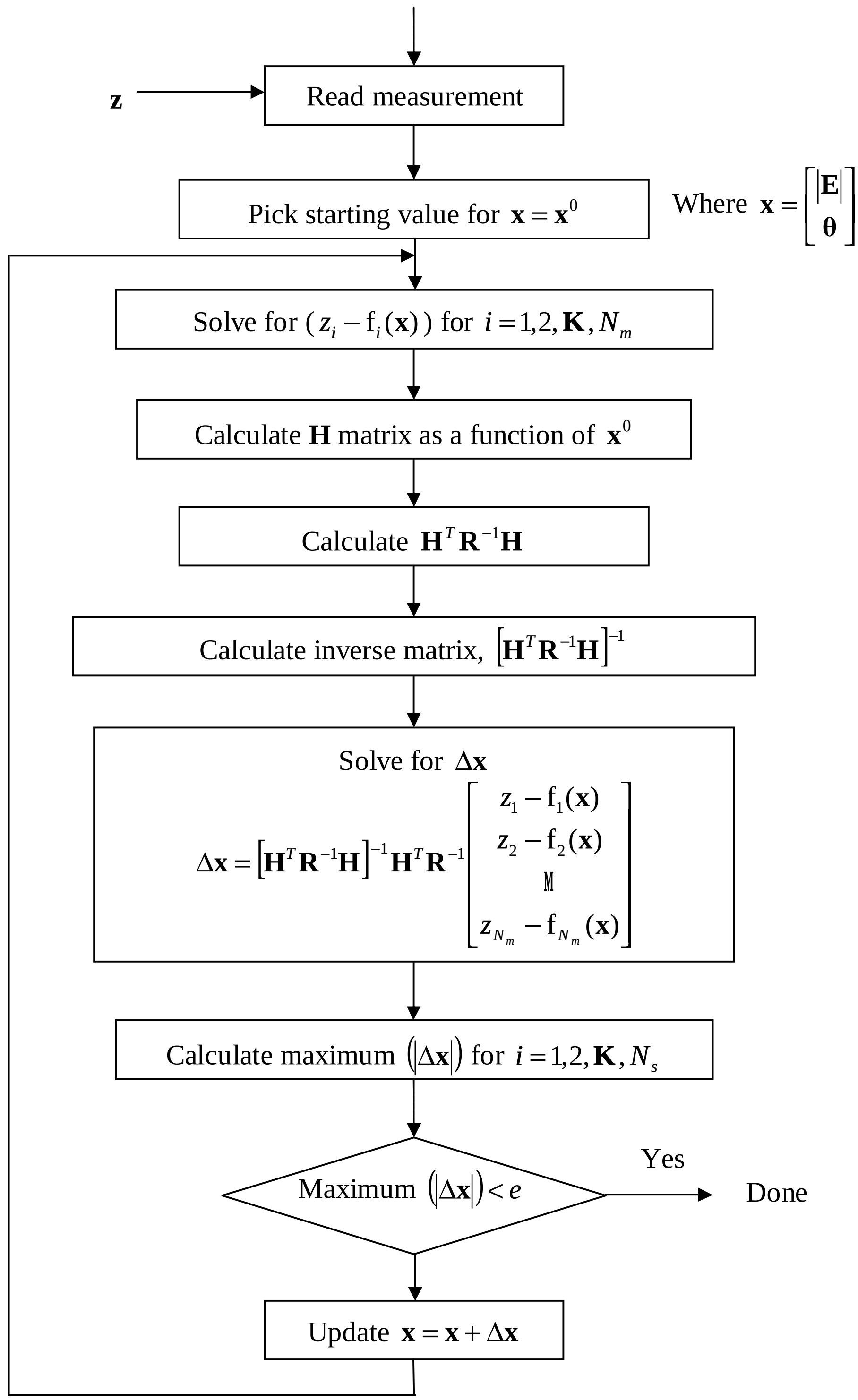

Figure 2.3: General state estimation solution algorithm