6

Optimization

The basic objective of any optimization method is to find the values of the system state variables and/or parameters that minimize some cost function of the system. The types of cost functions are system dependent and can vary widely from application to application and are not necessarily strictly measured in terms of dollars. Examples of engineering optimizations can range from minimizing

- the error between a set of measured and calculated data,
- active power losses,
- the weight of a set of components that comprise the system,
- particulate output (emissions),
- system energy, or
- the distance between actual and desired operating points,

to name a few possibilities. The basic formulation of any optimization can be represented as minimizing a defined cost function subject to any physical or operational constraints of the system:

$$\begin{aligned} \text{minimize} \;\; & f(x,u) \;\; x \in R^n \\ & \qquad\quad u \in R^m \end{aligned} \tag{6.1}$$

subject to

$$g(x,u) = 0 \;\; \text{equality constraints} \tag{6.2}$$

$$h(x,u) = 0 \;\; \text{inequality constraints} \tag{6.3}$$

where x is the vector of system states and u is the vector of system parameters. The basic approach is to find the vector of system parameters that, when substituted into the system model, will result in the state vector x that minimizes the cost function $f(x,u)$.

6.1 Least Squares State Estimation

In many physical systems, the system operating condition cannot be determined directly by an analytical solution of known equations using a given set of known, dependable quantities. More frequently, the system operating condition is determined by the measurement of system states at different points throughout the system. In many systems, more measurements are made than are necessary to uniquely determine the operating point. This redundancy is often purposely designed into the system to counteract the effect of inaccurate or missing data due to instrument failure. Conversely, not all of the states may be available for measurement. High temperatures, moving parts, or inhospitable conditions may make it difficult, dangerous, or expensive to measure certain system states. In this case, the missing states must be estimated from the rest of the measured information of the system. This process is often known as *state estimation* and is the process of estimating unknown states from measured quantities. State estimation gives the "best estimate" of the state of the system in spite of uncertain, redundant, and/or conflicting measurements. A good state estimation will smooth out small random errors in measurements, detect and identify large measurement errors, and compensate for missing data. This process strives to minimize the error between the (unknown) true operating state of the system and the measured states.

The set of measured quantities can be denoted by the vector z, which may include measurements of system states (such as voltage and current) or quantities that are functions of system states (such as power flows). Thus

$$z^{true} = Ax \tag{6.4}$$

where x is the set of system states and A is usually not square. The error vector is the difference between the measured quantities z and the true quantities:

$$e = z - z^{true} = z - Ax \tag{6.5}$$

Typically, the minimum of the square of the error is desired to negate any effects of sign differences between the measured and true values. Thus a state estimator endeavors to find the minimum of the squared error, or a *least squares minimization*:

$$\text{minimize } \|e\|^2 = e^T \cdot e = \sum_{i=1}^{m} \left[z_i - \sum_{j=1}^{m} a_{ij} x_j \right]^2 \tag{6.6}$$

The squared error function can be denoted by $U(x)$ and is given by

$$U(x) = e^T \cdot e = (z - Ax)^T (z - Ax) \tag{6.7}$$

$$= \left(z^T - x^T A^T\right)(z - Ax) \tag{6.8}$$

$$= z^T z - z^T Ax - x^T A^T z + x^T A^T Ax \tag{6.9}$$

Note that the product $z^T Ax$ is a scalar and so it can be equivalently written as

$$z^T Ax = \left(z^T Ax\right)^T = x^T A^T z$$

Therefore, the squared error function is given by

$$U(x) = z^T z - 2x^T A^T z + x^T A^T Ax \tag{6.10}$$

The minimum of the squared error function can be found by an unconstrained optimization where the derivative of the function with respect to the states x is set to zero:

$$\frac{\partial U(x)}{\partial x} = 0 = -2A^T z + 2A^T Ax \tag{6.11}$$

Thus

$$A^T Ax = A^T z \tag{6.12}$$

Thus, if $b = A^T z$ and $\hat{A} = A^T A$, then

$$\hat{A}x = b \tag{6.13}$$

which can be solved by LU factorization. This state vector x is the best estimate (in the squared error) to the system operating condition from which the measurements z were taken. The measurement error is given by

$$e = z^{meas} - Ax \tag{6.14}$$

Example 6.1

A set of measurements for the circuit shown in Figure 6.1 is given by

Ammeter 1	z_1	4.27 A
Ammeter 2	z_2	–1.71 A
Voltmeter 1	z_3	3.47 V
Voltmeter 2	z_4	2.50 V

where $R_1 = R_3 = R_5 = 1.5\Omega$ and $R_2 = R_4 = 1.0\Omega$. Find the node voltages V_1 and V_2.

Solution 6.1 The Kirchoff voltage and current law equations for this system can be written as

$$\begin{aligned} -V_1 + R_1 z_1 + z_3 &= 0 \\ -V_2 - R_5 z_2 + z_4 &= 0 \\ z_3/R_2 - z_1 + (z_3 - z_4)/R_3 &= 0 \\ z_4/R_4 + z_2 + (z_4 - z_3)/R_3 &= 0 \end{aligned}$$

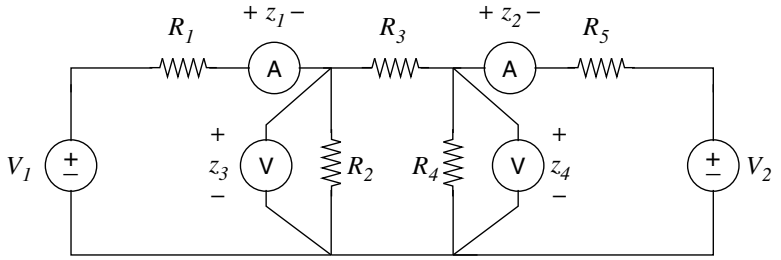

FIGURE 6.1
Circuit for Example 6.1

These equations can be rewritten in matrix form as

$$\begin{bmatrix} R_1 & 0 & 1 & 0 \\ 0 & -R_5 & 0 & 1 \\ 1 & 0 & -\left(\frac{1}{R_2}+\frac{1}{R_3}\right) & \frac{1}{R_3} \\ 0 & 1 & -\frac{1}{R_3} & \frac{1}{R_3}+\frac{1}{R_4} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \tag{6.15}$$

To find the relationship between the measurements z and x, this equation must be reformulated as $z = Ax$. Note that this equation can be solved easily by LU factorization by considering each column of A individually. Thus

$$\begin{bmatrix} R_1 & 0 & 1 & 0 \\ 0 & -R_5 & 0 & 1 \\ 1 & 0 & -\left(\frac{1}{R_2}+\frac{1}{R_3}\right) & \frac{1}{R_3} \\ 0 & 1 & -\frac{1}{R_3} & \frac{1}{R_3}+\frac{1}{R_4} \end{bmatrix} [A(:,1)] = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{6.16}$$

Similarly,

$$\begin{bmatrix} R_1 & 0 & 1 & 0 \\ 0 & -R_5 & 0 & 1 \\ 1 & 0 & -\left(\frac{1}{R_2}+\frac{1}{R_3}\right) & \frac{1}{R_3} \\ 0 & 1 & -\frac{1}{R_3} & \frac{1}{R_3}+\frac{1}{R_4} \end{bmatrix} [A(:,2)] = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \tag{6.17}$$

yielding

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} = \begin{bmatrix} 0.4593 & -0.0593 \\ 0.0593 & -0.4593 \\ 0.3111 & 0.0889 \\ 0.0889 & 0.3111 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \tag{6.18}$$

Thus

$$b = A^T z = \begin{bmatrix} 0.4593 & 0.0593 & 0.3111 & 0.0889 \\ -0.0593 & -0.4593 & 0.0889 & 0.3111 \end{bmatrix} \begin{bmatrix} 4.27 \\ -1.71 \\ 3.47 \\ 2.50 \end{bmatrix}$$

$$= \begin{bmatrix} 3.1615 \\ 1.6185 \end{bmatrix}$$

and

$$\hat{A} = A^T A = \begin{bmatrix} 0.3191 & 0.0009 \\ 0.0009 & 0.3191 \end{bmatrix}$$

leading to

$$\begin{bmatrix} 0.3191 & 0.0009 \\ 0.0009 & 0.3191 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} 3.1615 \\ 1.6185 \end{bmatrix} \tag{6.19}$$

Solving this equation yields

$$V_1 = 9.8929$$
$$V_2 = 5.0446$$

The error between the measured values and the estimated values of this system is given by

$$\begin{aligned} e &= z - Ax \qquad (6.20) \\ &= \begin{bmatrix} 4.27 \\ -1.71 \\ 3.47 \\ 2.50 \end{bmatrix} - \begin{bmatrix} 0.4593 & -0.0593 \\ 0.0593 & -0.4593 \\ 0.3111 & 0.0889 \\ 0.0889 & 0.3111 \end{bmatrix} \begin{bmatrix} 9.8929 \\ 5.0446 \end{bmatrix} \\ &= \begin{bmatrix} 0.0255 \\ 0.0205 \\ -0.0562 \\ -0.0512 \end{bmatrix} \end{aligned}$$

■

6.1.1 Weighted Least Squares Estimation

If all measurements are treated equally in the least squares solution, then the less accurate measurements will affect the estimation as significantly as the more accurate measurements. As a result, the final estimation may contain large errors due to the influence of inaccurate measurements. By introducing a weighting matrix to emphasize the more accurate measurements more heavily than the less accurate measurements, the estimation procedure can then force the results to coincide more closely with the measurements of greater accuracy. This leads to the weighted least squares estimation

$$\text{minimize } \|e\|^2 = e^T \cdot e = \sum_{i=1}^{m} w_i \left[z_i - \sum_{j=1}^{m} a_{ij} x_j \right]^2 \tag{6.21}$$

where w_i is a weighting factor reflecting the level of confidence in the measurement z_i.

Example 6.2

Suppose that the ammeters are known to have been more recently calibrated than the voltmeters; thus the level of confidence in the current measurements is greater than the voltage measurements. Using the following weighting matrix, find the node voltages V_1 and V_2.

$$W = \begin{bmatrix} 100 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 \\ 0 & 0 & 50 & 0 \\ 0 & 0 & 0 & 50 \end{bmatrix}$$

Solution 6.2 By introducing the weighting matrix, the new minimum is given by

$$A^T W A x = A^T W z \tag{6.22}$$

The matrix $A^T W A$ is also known as the *gain* matrix. Using the same procedure as before, the weighted node voltage values are given by

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} 9.9153 \\ 5.0263 \end{bmatrix} \tag{6.23}$$

and the error vector is given by

$$e = \begin{bmatrix} 0.0141 \\ 0.0108 \\ -0.0616 \\ 0.0549 \end{bmatrix} \tag{6.24}$$

■

Note that the added confidence in the current measurements has decreased the estimation error in the current, but the voltage measurement error is approximately the same.

Example 6.2 illustrates the impact of confidence weighting on the accuracy of the estimation. All instruments add some degree of error to the measured values, but the problem is how to quantify this error and account for it during the estimation process. In general, it can be assumed that the introduced errors have normal (Gaussian) distribution with zero mean and that each measurement is independent of all other measurements. This means that each measurement error is as likely to be greater than the true value as it is to be less than the true value. A zero mean Gaussian distribution has several attributes. The standard deviation of a zero mean Gaussian distribution is denoted by σ. This means that 68% of all measurements will fall within $\pm\sigma$ of the expected value, which is zero in a zero mean distribution. Further, 95% of all measurements will fall within $\pm 2\sigma$, and 99% of all measurements will fall within $\pm 3\sigma$. The variance of the measurement distribution is given by σ^2.

This implies that, if the variance of the measurements is relatively small, then the majority of measurements are close to the mean. One interpretation of this is that accurate measurements lead to small variance in the distribution.

This relationship between accuracy and variance leads to a straightforward approach from which to develop a weighting matrix for the estimation. Consider the squared error matrix given by

$$e \cdot e^T = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_m \end{bmatrix} \begin{bmatrix} e_1 & e_2 & e_3 & \dots & e_m \end{bmatrix} \tag{6.25}$$

$$= \begin{bmatrix} e_1^2 & e_1e_2 & e_1e_3 & \dots & e_1e_m \\ e_2e_1 & e_2^2 & e_2e_3 & \dots & e_2e_m \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ e_me_1 & e_me_2 & e_me_3 & \dots & e_m^2 \end{bmatrix} \tag{6.26}$$

where each e_i is the error in the ith measurement. The expected, or mean, value of each error product is given by $E[\cdot]$. The expected value of each of the diagonal terms is the variance of the ith error distribution σ_i^2. The expected value of each of the off-diagonal terms, or covariance, is zero because each measurement is assumed to be independent of every other measurement. Therefore, the expected value of the squared error matrix (also known as the covariance matrix) is

$$E\left[e \cdot e^T\right] = \begin{bmatrix} E\left[e_1^2\right] & E[e_1e_2] & \dots & E[e_1e_m] \\ E[e_2e_1] & E\left[e_2^2\right] & \dots & E[e_2e_m] \\ \vdots & \vdots & \vdots & \vdots \\ E[e_me_1] & E[e_me_2] & \dots & E\left[e_m^2\right] \end{bmatrix} \tag{6.27}$$

$$= \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_m^2 \end{bmatrix} \tag{6.28}$$

$$= R \tag{6.29}$$

With measurements taken from a particular meter, the smaller the variance of the measurements (i.e., the more consistent they are), the greater the level of confidence in that set of measurements. A set of measurements that has a high level of confidence should have a higher weighting than a set of measurements that has a larger variance (and therefore less confidence). Therefore, a plausible weighting matrix that reflects the level of confidence in each measurement set is the inverse of the covariance matrix $W = R^{-1}$. Thus measurements that come from instruments with good consistency (small variance) will carry

greater weight than measurements that come from less accurate instruments (high variance). Thus one possible weighting matrix is given by

$$W = R^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2^2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sigma_m^2} \end{bmatrix} \tag{6.30}$$

6.1.2 Bad Data Detection

Frequently, a set of measurements will contain one or more data points from faulty or poorly calibrated instruments. Telemetered measurements are subject to noise or error in metering and communication. These "bad" data points typically fall outside of the standard deviation of the measurements and may affect the reliability of the state estimation process. In severe cases, the bad data may actually lead to grossly inaccurate results. Bad data may cause the accuracy of the estimate to deteriorate because of the "smearing" effect as the bad data will pull, or smear, the estimated values away from the true values. Therefore, it is desirable to develop a measure of the "goodness" of the data upon which the estimation is based. If the data lead to a good estimate of the states, then the error between the measured and calculated values will be small in some sense. If the error is large, then the data contain at least one bad data point. One error that is useful to consider is the estimated measurement error $\hat{e}$. This error is the difference between the actual measurements z and the estimated measurements $\hat{z}$. Recall the error vector from Equation (6.14) where $e = z - Ax$; then the estimated measurement error becomes

$$\begin{aligned} \hat{e} &= z - \hat{z} && (6.31)\\ &= z - A\hat{x} && (6.32)\\ &= z - A\left(A^T W A\right)^{-1} A^T W z && (6.33)\\ &= \left(I - A\left(A^T W A\right)^{-1} A^T W\right) z && (6.34)\\ &= \left(I - A\left(A^T W A\right)^{-1} A^T W\right)(e + Ax) && (6.35)\\ &= \left(I - A\left(A^T W A\right)^{-1} A^T W\right) e + A\left(I - \left(A^T W A\right)^{-1} A^T W A\right) x && (6.36)\\ &= \left(I - A\left(A^T W A\right)^{-1} A^T W\right) e && (6.37) \end{aligned}$$

Thus the variance of $\hat{e}$ can be calculated from

$$\begin{aligned} \hat{e}\hat{e}^T &= (z - \hat{z})(z - \hat{z})^T && (6.38)\\ &= \left[I - A\left(A^T W A\right)^{-1} A^T W\right] e e^T \left[I - W A\left(A^T W A\right)^{-1} A^T\right] && (6.39) \end{aligned}$$

The expected, or mean, value of $\hat{e}\hat{e}^T$ is given by

$$E\left[\hat{e}\hat{e}^T\right] = \left[I - A\left(A^T W A\right)^{-1} A^T W\right] E\left[ee^T\right] \left[I - WA\left(A^T W A\right)^{-1} A^T\right] \tag{6.40}$$

Recall that $E\left[ee^T\right]$ is just the covariance matrix $R = W^{-1}$, which is a diagonal matrix. Thus

$$E\left[\hat{e}\hat{e}^T\right] = \left[I - A\left(A^T W A\right)^{-1} A^T W\right] \left[I - A\left(A^T W A\right)^{-1} A^T W\right] R \tag{6.41}$$

The matrix

$$\left[I - A\left(A^T W A\right)^{-1} A^T W\right]$$

has the unusual property that it is an *idempotent* matrix. An idempotent matrix M has the property that $M^2 = M$; thus, no matter how many times M is multiplied by itself, it will still return the product M. Therefore,

$$\begin{aligned} E\left[\hat{e}\hat{e}^T\right] &= \left[I - A\left(A^T W A\right)^{-1} A^T W\right] \left[I - A\left(A^T W A\right)^{-1} A^T W\right] R && (6.42)\\ &= \left[I - A\left(A^T W A\right)^{-1} A^T W\right] R && (6.43)\\ &= R - A\left(A^T W A\right)^{-1} A^T && (6.44)\\ &= R' && (6.45) \end{aligned}$$

To determine whether the estimated values differ significantly from the measured values, a useful statistical measure is the χ^2 (chi-squared) test of inequality. This measure is based on the χ^2 probability distribution, which differs in shape depending on its degrees of freedom k, which is the difference between the number of measurements and the number of states. By comparing the weighted sum of errors with the χ^2 value for a particular degree of freedom and significance level, it can be determined whether the errors exceed the bounds of what would be expected by chance alone. A significance level indicates the level of probability that the measurements are erroneous. A significance level of 0.05 indicates there is a 5% likelihood that bad data exist, or conversely, a 95% level of confidence in the goodness of the data. For example, for $k = 2$ and a significance level $\alpha = 0.05$, if the weighted sum of errors does not exceed a χ^2 of 5.99, then the set of measurements can be assured of being good with 95% confidence; otherwise the data must be rejected as containing at least one bad data point. Although the χ^2 test is effective in signifying the presence of bad data, it cannot identify locations. The identification of bad data locations continues to be an open research topic.

χ^2 Values				
	α			
k	0.10	0.05	0.01	0.001
1	2.71	3.84	6.64	10.83
2	4.61	5.99	9.21	13.82
3	6.25	7.82	11.35	16.27
4	7.78	9.49	13.23	18.47
5	9.24	11.07	15.09	20.52
6	10.65	12.59	16.81	22.46
7	12.02	14.07	18.48	24.32
8	13.36	15.51	20.09	26.13
9	14.68	16.92	21.67	27.88
10	15.99	18.31	23.21	29.59
11	17.28	19.68	24.73	31.26
12	18.55	21.03	26.22	32.91
13	19.81	22.36	27.69	34.53
14	21.06	23.69	29.14	36.12
15	22.31	25.00	30.68	37.70
16	23.54	26.30	32.00	39.25
17	24.77	27.59	33.41	40.79
18	25.99	28.87	34.81	42.31
19	27.20	30.14	36.19	43.82
20	28.41	31.41	37.67	45.32
21	29.62	32.67	38.93	46.80
22	30.81	33.92	40.29	48.27
23	32.00	35.17	41.64	49.73
24	33.20	36.42	42.98	51.18
25	34.38	37.65	44.31	52.62
26	35.56	38.89	45.64	54.05
27	36.74	40.11	46.96	55.48
28	37.92	41.34	48.28	56.89
29	39.09	42.56	49.59	58.30
30	40.26	43.77	50.89	59.70

A test procedure to test for the existence of bad data is given by

Test Procedure for Bad Data

1. Use z to estimate x.
2. Calculate the error
$$e = z - Ax$$
3. Evaluate the weighted sum of squares
$$f = \sum_{i=1}^{m} \frac{1}{\sigma_i}^2 e_i^2$$
4. For $k = m - n$ and a specified probability α, if $f < \chi^2_{k,\alpha}$, then the data are good; otherwise at least one bad data point exists.

Example 6.3
Using the chi-square test of inequality with $\alpha = 0.01$, check for the presence of bad data in the measurements of Example 6.1.

Solution 6.3 The number of states in Example 6.1 is 2 and the number of measurements is 4; therefore $k = 4 - 2 = 2$. The weighted sum of squares is given by

$$\begin{aligned} f &= \sum_{i=1}^{m=4} \frac{1}{\sigma_i} e_i^2 \\ &= 100(0.0141)^2 + 100(0.0108)^2 + 50(-0.0616)^2 + 50(0.0549) \\ &= 0.3720 \end{aligned}$$

From the table of chi-squared values, the chi-square value for this example is 9.21. The weighted least squares error is less than the chi-square value; this indicates that the estimated values are good to a confidence level of 99%. ■

6.1.3 Nonlinear Least Squares State Estimation

As in the linear least squares estimation, the nonlinear least squares estimation attempts to minimize the square of the errors between a known set of measurements and a set of weighted nonlinear functions:

$$\text{minimize } f = \|e\|^2 = e^T \cdot e = \sum_{i=1}^{m} \frac{1}{\sigma^2} [z_i - h_i(x)]^2 \tag{6.46}$$

where $x \in R^n$ is the vector of unknowns to be estimated, $z \in R^m$ is the vector of measurements, σ_i^2 is the variance of the ith measurement, and $h(x)$ is the

function vector relating x to z, where the measurement vector z can be a set of geographically distributed measurements, such as voltages and power flows.

In state estimation, the unknowns in the nonlinear equations are the state variables of the system. The state values that minimize the error are found by setting the derivatives of the error function to zero:

$$F(x) = H_x^T R^{-1} [z - h(x)] = 0 \tag{6.47}$$

where

$$H_x = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \frac{\partial h_2}{\partial x_1} & \frac{\partial h_2}{\partial x_2} & \cdots & \frac{\partial h_2}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial h_m}{\partial x_1} & \frac{\partial h_m}{\partial x_2} & \cdots & \frac{\partial h_m}{\partial x_n} \end{bmatrix} \tag{6.48}$$

and R is the matrix of measurement variances. Note that Equation (6.47) is a set of nonlinear equations that must be solved using the Newton–Raphson or another iterative numerical solver. In this case, the Jacobian of $F(x)$ is

$$J_F(x) = H_x^T(x) R^{-1} \frac{\partial}{\partial x} [z - h(x)] \tag{6.49}$$

$$= -H_x^T(x) R^{-1} H_x(x) \tag{6.50}$$

and the Newton–Raphson iteration becomes

$$\left[H_x^T \left(x^k\right) R^{-1} H_x \left(x^k\right)\right] \left[x^{k-1} - x^k\right] = H_x^T \left(x^k\right) R^{-1} \left[z - h(x^k)\right] \tag{6.51}$$

which is solved repeatedly using LU factorization. At convergence, x^{k+1} is equal to the set of states that minimize the error function f of Equation (6.46). The test procedure for bad data is the same as that for the linear state estimation.

6.2 Linear Programming

Linear programming is one of the most successful forms of optimization. Linear programming can be used when a problem can be expressed by a linear objective (or cost) function to be maximized (or minimized) subject to linear equality or inequality constraints. A general linear programming problem can be formulated as

$$\text{minimize} \quad f(x) = c^T x \tag{6.52}$$

$$\text{subject to} \quad Ax \leq b \tag{6.53}$$

$$x \geq 0 \tag{6.54}$$

Note that almost any linear optimization problem can be put into this form via one of the following tranformations:

1. Maximizing $c^T x$ is the same as minimizing $-c^T x$,
2. Any constraint of the form $a^T x \geq \beta$ is equivalent to $-a^T x \leq -\beta$,
3. Any constraint of the form $a^T x = \beta$ is equivalent to $a^T x \leq \beta$ and $-a^T x \leq -\beta$,
4. If a problem does not require x_i to be nonnegative, then x_i can be replaced by the difference of two variables $x_i = u_i - v_i$ where u_i and v_i are nonnegative.

In any linear programming problem described by (A, b, c), there exists another equivalent, or *dual* problem $(-A^T, -c, -b)$. If a linear programming problem and its dual both have feasible points (i.e., any point that satisfies $Ax \leq b, x \geq 0$ or $-A^T y \leq -c, y \geq 0$ for the dual problem), then both problems have solutions and their values are the negatives of each other.

6.2.1 Simplex Method

One of the most common methods of solving linear programming problems is the well-known *simplex* method. The simplex method is an iterative method that moves the x vector from one feasible basic vector to another in such a way that $f(x)$ always decreases. It gives the exact result after a number of steps, which is usually much less than $\binom{n}{n-m}$, generally taking $2m$ to $3m$ iterations at most (where m is the number of equality constraints) [9]. However, its worst-case complexity is exponential, as can be demonstrated with carefully constructed examples.

The simplex method is often accomplished by representing the problem in *tableau* form, which is then modified in successive steps according to given rules. Every step of the simplex method begins with a tableau. The top row contains the coefficients that pertain to the objective function $f(x)$. The current value of $f(x)$ is displayed in the top right corner of the tableau. The next m rows in the tableau represent the equality constraints. The last row of the tableau contains the current x vector. The rows in the tableau pertaining to the equality constraints can be transformed by elementary row operations without altering the solution.

The initial state of a simplex tableau has the form

c^T	0	0
A	I	b
$\underbrace{0}_{x}$	b^T	

The rules of the tableau that must be satisfied are

1. The x vector must satisfy the equality constraints $Ax = b$.
2. The x vector must satisfy the inequality $x \geq 0$.
3. There are n components of x (designated *nonbasic variables*) that are zero. The remaining m components are usually nonzero and are designated as *basic variables*.
4. In the matrix that defines the constraints, each basic variable occurs in only one row.
5. The objective function $f(x)$ must be expressed only in terms of nonbasic variables.
6. An artificial variable may be added to one or more constraints to obtain a starting solution.

The Simplex Algorithm is summarized:

Simplex Algorithm

1. If all coefficients in $f(x)$ (i.e., top row of the tableau) are greater than or equal to zero, then the current x vector is the solution.
2. Select the nonbasic variable whose coefficient in $f(x)$ is the largest negative entry. This variable becomes the new basic variable x_j.
3. Divide each b_i by the coefficient of the new basic variable in that row, a_{ij}. The value assigned to the new basic variable is the least of these ratios (i.e., $x_j = b_k/a_{kj}$).
4. Using pivot element a_{kj}, create zeros in column j of A with Gaussian elimination. Return to 1.

The series of inequalities in Equation (6.53), when taken together, form intersecting hyperplanes. The feasible region is the interior of this n-dimensional polytope, and the minimum $f(x)$ must occur on an edge or vertex of this polytope. The simplex method is an organized search of the vertices by moving along the steepest edge of the polytope until $x^\star$ is obtained, as shown in Figure 6.2.

Example 6.4

Minimize

$$f(x): \quad -6x_1 - 14x_2$$

subject to the following constraints:

$$\begin{aligned} 2x_1 + x_2 &\leq 12 \\ 2x_1 + 3x_2 &\leq 15 \\ x_1 + 7x_2 &\leq 21 \\ x_1 \geq 0,\ x_2 &\geq 0 \end{aligned}$$

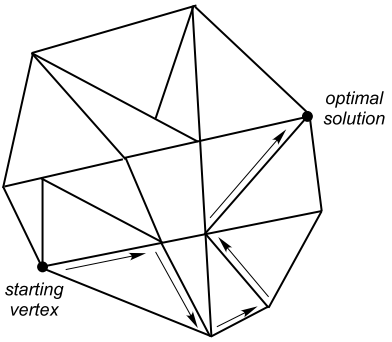

FIGURE 6.2
Example of a simplex method search

Solution 6.4 Introduce slack variables $x_3,\ x_4$, and x_5 such that the problem becomes

Minimize

$$f(x): \ -6x_1 - 14x_2 + 0x_3 + 0x_4 + 0x_5$$

subject to the following constraints:

$$\begin{aligned} 2x_1 + \ x_2 + x_3 \qquad\quad &= 12 \\ 2x_1 + 3x_2 \qquad + x_4 \quad &= 15 \\ x_1 + 7x_2 \qquad\quad + x_5 &= 21 \end{aligned}$$

$$x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0,\ x_4 \geq 0,\ x_5 \geq 0$$

Form the tableau:

-6	-14	0	0	0	0
2	1	1	0	0	12
2	3	0	1	0	15
1	7	0	0	1	21
0	0	12	15	21	

The starting vector is $x = [0\ 0\ 12\ 15\ 21]^T$. The current value of $f(x)$ is displayed in the top right corner. In the next step, the function $f(x)$ is examined to determine which variable will cause the greatest decrease. Since -14 is more negative than -6, a unit increase in x_2 will decrease $f(x)$ faster than a unit increase in x_1. Therefore, hold x_1 constant at zero and allow x_2 to increase as much as possible (i.e., traverse one edge of the polytope to the next vertex). To determine the new value of x_2 consider the constraints

$$0 \leq x_3 = 12 - x_2$$

$$0 \le x_4 = 15 - 3x_2$$
$$0 \le x_5 = 21 - 7x_2$$

From these constraints, the possible values of x_2 are $x_2 \le 12$, $x_2 \le 5$, $x_2 \le 3$. The most stringent is $x_2 \le 3$; therefore, x_2 can be increased to 3, and

$$x_3 = 12 - x_2 = 9$$
$$x_4 = 15 - 3x_2 = 6$$
$$x_5 = 21 - 7x_2 = 0$$

yielding the new vector $x = [0\ 3\ 9\ 6\ 0]^T$ and $f(x) = -42$.

The new basic (nonzero) variables are x_2, x_3, and x_4; therefore, $f(x)$ must be expressed in terms of x_1 and x_5. By substitution,

$$x_2 = \frac{(21 - x_5 - x_1)}{7}$$

and

$$f(x) = -6x_1 - 14x_2 = -6x_1 - 14\frac{(21 - x_5 - x_1)}{7} = -4x_1 + 2x_5 - 42$$

To satisfy the rule that a basic variable must occur in only one row, Gaussian elimination is used to eliminate x_2 from every row but one using the pivot as defined in Step 3 of the algorithm.

The new tableau after Gaussian elimination is

-4	0	0	0	2	-42
$\frac{13}{7}$	0	1	0	$-\frac{1}{7}$	9
$\frac{11}{7}$	0	0	1	$-\frac{3}{7}$	6
$\frac{1}{7}$	1	0	0	$\frac{1}{7}$	3
0	3	9	6	0	

The method is again repeated. Any increase in x_5 will increase $f(x)$, so x_1 is chosen to become a basic variable. Therefore, x_5 is held at zero and x_1 is allowed to increase as much as possible. The new constraints are

$$0 \le x_3 = 9 - \frac{13}{7}x_1$$
$$0 \le x_4 = 6 - \frac{11}{7}x_1$$
$$0 \le 7x_2 = 21 - x_1$$

or $x_1 \le \frac{63}{13}, x_1 \le \frac{42}{11}$, and $x_1 \le 21$. The most stringent constraint is $x_1 \le \frac{42}{11}$, therefore, x_1 is set to $\frac{42}{11}$, and the new values of x_2, x_3, and x_4 are computed. The new vector is $x = \left[\frac{42}{11}\ \frac{27}{11}\ \frac{21}{11}\ 0\ 0\right]^T$ and $f(x)$ is rewritten in terms of x_4

and x_5:

$$\begin{aligned} x_1 &= \frac{7}{11}(6 - x_4) \\ f(x) &= -4x_1 + 2x_5 - 42 \\ &= \frac{28}{11}x_4 + 2x_5 - \frac{630}{11} \end{aligned}$$

Since all coefficients in $f(x)$ are positive, the simplex method terminates because any increase in x_4 or x_5 will increase $f(x)$. This signifies that the current x vector is the solution and the final value of $f(x)$ is $-\frac{630}{11}$. ■

6.2.2 Interior Point Method

A different type of method for linear programming problems is interior point methods (also known as Karmarkar's methods), whose complexity is polynomial for both average and worst case. The simplex method has the potential to have a worst case scenario of exponential complexity that can occur in the situation in which the solution visits *every vertex* in the feasible region before reaching the optimum. For this reason, interior point methods have received considerable attention over the past few decades. Interior point methods construct a sequence of strictly feasible points (i.e., lying in the interior of the polytope but never on its boundary) that converges to the solution.

The interior point method constructs a series of feasible solutions $x^0, x^1, \ldots$ that must satisfy $Ax_i = b$. Since x^0 satisfies $Ax^0 = b$ and the next point must satisfy $Ax^1 = b$, the difference in solutions must satisfy $A\Delta x = 0$. In other words, each step must lie in the nullspace of A, which is parallel to the feasible set. Projecting $-c$ onto that nullspace gives the direction of most rapid change. However, if the iterate x^k is close to the boundary (as with $\hat{x}^k$ in Figure 6.3), very little improvement will occur. If, however, the current iterate is near the center (as with $\bar{x}^k$), there could be significant improvement. One of the key aspects of the interior point method is that a transformation is applied such that the current feasible point is moved (through the transformation) to the center of the interior. The new direction is then computed and the interior point is transformed back to the original space.

This direction of change is the *projected gradient direction*, or p^k, and the feasible points are updated through

$$x^{k+1} = x^k + \alpha p^k \tag{6.55}$$

where $\alpha > 0$ is the step length. Since the feasible points must lie in the null space of A, each p^k must be orthogonal to the rows of A. The projection matrix P

$$P = I - A^T(AA^T)^{-1}A \tag{6.56}$$

will transform any vector v into $Pv = p$, and p will be in the null space of A because AP is the zero matrix.

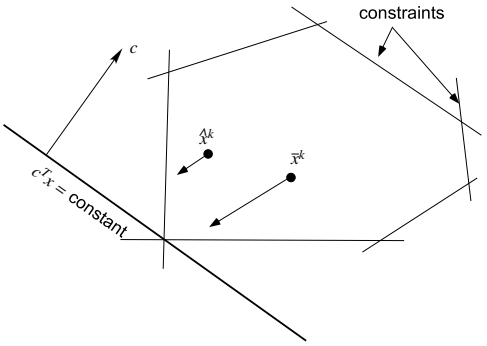

FIGURE 6.3
Comparison of two different points in the interior

Since projecting $-c$ onto the nullspace gives the direction of most rapid change, then to maintain feasibility a new iterate must satisfy $p^k = -Pc$. To remain in the interior of the space, the step length α is chosen at each step to ensure feasibility of the nonnegativity constraints. To ensure that the updates remain in the interior of the feasible space, the step length is chosen to be less than the full distance to the boundary, usually $0.5 \leq \alpha \leq 0.98$.

The last aspect is the transformation is required to center the iterate in the feasible space. This is accomplished by a scaling, such that the iterate is equidistant from all constraint boundaries in the transformed feasible space. Therefore, after rescaling, $x^k = e$, where $e = [1\ 1\ \ldots\ 1]^T$. Let $D = \text{diag}(x^k)$ be the diagonal matrix with each component of the current iterate x^k on the diagonals. This is accomplished by letting $x = D\hat{x}$ so that $\hat{x}^k = e$. The new problem to be solved then becomes

$$\text{minimize} \quad \hat{c}^T \hat{x} = z \tag{6.57}$$

$$\text{subject to} \quad \hat{A}\hat{x} \leq b \tag{6.58}$$

$$\hat{x} \geq 0 \tag{6.59}$$

where $\hat{c} = Dc$ and $\hat{A} = AD$. After scaling, the projection matrix becomes

$$\hat{P} = I - \hat{A}^T(\hat{A}\hat{A}^T)^{-1}\hat{A} \tag{6.60}$$

Hence, at each iteration k, the iterate x^k is rescaled to $\hat{x}^k = e$ and the update is given by

$$\hat{x}^{k+1} = e - \alpha\hat{P}\hat{c} \tag{6.61}$$

and then the updated iterate is transformed back to the original space:

$$x^{k+1} = D\hat{x}^{k+1} \tag{6.62}$$

This process repeats until $\|x^{k+1} - x^k\| < \varepsilon$. This process is often referred to as the *primal affine* interior point method [10]. The steps of the method are summarized:

Primal Affine Interior Point Method

1. Let $k = 0$.
2. Let $D = \text{diag}(x^k)$.
3. Compute $\hat{A} = AD,\ \hat{c} = Dc$.
4. Compute $\hat{P}$ from Equation (6.60).
5. Set $p^k = \hat{P}\hat{c}$.
6. Set $\theta = -\min_j p_j^k$. The factor θ is used to determine the maximum step length that can be taken before exiting the feasible region.
7. Compute
$$\hat{x}^{k+1} = e + \frac{\alpha}{\theta} p^k$$
8. Compute $x^{k+1} = D\hat{x}^{k+1}$.
9. If $\|x^{k+1} - x^k\| < \varepsilon$, then done. Else set $k = k + 1$. Go to step 2.

Example 6.5
Repeat Example 6.4 using the primal affine interior point method.

Solution 6.5 The problem is restated (with slack variables included) for convenience:

Minimize
$$f(x): \quad -6x_1 - 14x_2 + 0x_3 + 0x_4 + 0x_5 = z$$
subject to the following constraints:

$$\begin{aligned} 2x_1 + \ x_2 + x_3 \qquad\qquad &= 12 \\ 2x_1 + 3x_2 \qquad + x_4 \qquad &= 15 \\ x_1 + 7x_2 \qquad\qquad + x_5 &= 21 \end{aligned}$$

$$x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0,\ x_4 \geq 0,\ x_5 \geq 0$$

A feasible initial starting solution is

$$x^o = [1 \ \ 1 \ \ 9 \ \ 10 \ \ 13]$$

with $z^0 = c^T x^0 = -20$. The first scaling matrix is

$$D = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 9 & & \\ & & & 10 & \\ & & & & 13 \end{bmatrix}$$

The rescaled matrix $\hat{A}$ and objective function vector $\hat{c}$ are computed as

$$\hat{A} = AD = \begin{bmatrix} 2 & 1 & 1 & 0 & 0 \\ 2 & 3 & 0 & 1 & 0 \\ 1 & 7 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 9 & & \\ & & & 10 & \\ & & & & 13 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 9 & 0 & 0 \\ 2 & 3 & 0 & 10 & 0 \\ 1 & 7 & 0 & 0 & 13 \end{bmatrix}$$

$$\hat{c} = Dc = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 9 & & \\ & & & 10 & \\ & & & & 13 \end{bmatrix} \begin{bmatrix} -6 \\ -14 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -6 \\ -14 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

The projection matrix $\hat{P}$ is

$$\hat{P} = I - \hat{A}^T(\hat{A}\hat{A}^T)^{-1}\hat{A}$$

$$= \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix} - \begin{bmatrix} 2 & 2 & 1 \\ 1 & 3 & 7 \\ 9 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 13 \end{bmatrix} \left[\begin{bmatrix} 2 & 1 & 9 & 0 & 0 \\ 2 & 3 & 0 & 10 & 0 \\ 1 & 7 & 0 & 0 & 13 \end{bmatrix} \begin{bmatrix} 2 & 2 & 1 \\ 1 & 3 & 7 \\ 9 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 13 \end{bmatrix} \right]^{-1} \begin{bmatrix} 2 & 1 & 9 & 0 & 0 \\ 2 & 3 & 0 & 10 & 0 \\ 1 & 7 & 0 & 0 & 13 \end{bmatrix}$$

$$= \begin{bmatrix} 0.9226 & -0.0836 & -0.1957 & -0.1595 & -0.0260 \\ -0.0836 & 0.7258 & -0.0621 & -0.2010 & -0.3844 \\ -0.1957 & -0.0621 & 0.0504 & 0.0578 & 0.0485 \\ -0.1595 & -0.2010 & 0.0578 & 0.0922 & 0.1205 \\ -0.0260 & -0.3844 & 0.0485 & 0.1205 & 0.2090 \end{bmatrix}$$

The projected gradient is

$$p^0 = -\hat{P}\hat{c} = - \begin{bmatrix} 0.9226 & -0.0836 & -0.1957 & -0.1595 & -0.0260 \\ -0.0836 & 0.7258 & -0.0621 & -0.2010 & -0.3844 \\ -0.1957 & -0.0621 & 0.0504 & 0.0578 & 0.0485 \\ -0.1595 & -0.2010 & 0.0578 & 0.0922 & 0.1205 \\ -0.0260 & -0.3844 & 0.0485 & 0.1205 & 0.2090 \end{bmatrix} \begin{bmatrix} -6 \\ -14 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 4.3657 \\ 9.6600 \\ -2.0435 \\ -3.7711 \\ -5.5373 \end{bmatrix}$$

Calculate $\theta = -\min_j p_j^0 = 5.5373$. Rescale the current iterate to $\hat{x}^0 = D^{-1}x^0 = e$ and move to $\hat{x}^1$ in the transformed space with $\alpha = 0.9$:

$$\hat{x}^1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \alpha p^0 = \begin{bmatrix} 1.7096 \\ 2.5701 \\ 0.6679 \\ 0.3871 \\ 0.1000 \end{bmatrix}$$

Transforming this point back to the original space:

$$x^1 = D\hat{x}^1 = \begin{bmatrix} 1.7096 \\ 2.5701 \\ 6.0108 \\ 3.8707 \\ 1.3000 \end{bmatrix}$$

and the updated cost function is $c^T x^1 = -46.2383$.

Performing one more iteration (and omitting the detailed text) yields

$$\hat{A} = \begin{bmatrix} 3.4191 & 2.5701 & 6.0108 & 0 & 0 \\ 3.4191 & 7.7102 & 0 & 3.8707 & 0 \\ 1.7096 & 17.9904 & 0 & 0 & 1.3000 \end{bmatrix}$$

$$\hat{c} = \begin{bmatrix} -10.2574 \\ -35.9809 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\hat{P} = \begin{bmatrix} 0.5688 & -0.0584 & -0.2986 & -0.3861 & 0.0606 \\ -0.0584 & 0.0111 & 0.0285 & 0.0295 & -0.0766 \\ -0.2986 & 0.0285 & 0.1577 & 0.2070 & -0.0017 \\ -0.3861 & 0.0295 & 0.2070 & 0.2822 & 0.0991 \\ 0.0606 & -0.0766 & -0.0017 & 0.0991 & 0.9803 \end{bmatrix}$$

$$p^1 = \begin{bmatrix} 3.7321 \\ -0.2004 \\ -2.0373 \\ -2.8975 \\ -2.1345 \end{bmatrix}$$

$$\hat{x}^1 = \begin{bmatrix} 2.1592 \\ 0.9378 \\ 0.3672 \\ 0.1000 \\ 0.3370 \end{bmatrix}$$

$$x^1 = \begin{bmatrix} 3.6914 \\ 2.4101 \\ 2.2072 \\ 0.3871 \\ 0.4381 \end{bmatrix}$$

and the updated cost function is $c^T x^1 = -55.8892$.

This continues until $\|x^{k+1} - x^k\| < \varepsilon$, at which time the solution is

$$x^\star = \begin{bmatrix} 3.8182 \\ 2.4545 \\ 1.9091 \\ 0.0000 \\ 0.0000 \end{bmatrix}$$

and the updated cost function is $c^T x^1 = -57.2727$. Both the resulting solution and cost function are the same as the simplex method. ■

6.3 Nonlinear Programming

Continuous nonlinear optimization problems are typically of the following form:

$$\text{minimize} \quad f(x) \quad x \in \mathbb{R}^n \tag{6.63}$$

$$\text{subject to} \quad c_i(x) = 0, \quad i \in \xi \tag{6.64}$$

$$h_i(x) \geq 0, \quad i \in \Xi \tag{6.65}$$

where $[c(x) \;\; h(x)]$ is an m-vector of nonlinear constraint functions such that ξ and Ξ are nonintersecting index sets. The function $f(x)$ is sometimes referred to as a "cost" function. It is assumed throughout that f, c, and h are twice-continuously differentiable. Any point x satisfying the constraints of Equations (6.64) and (6.65) is called a *feasible* point, and the set of all such points is the *feasible region*. These types of problems are known generically as *nonlinear programming problems*, or NLP.

Often in optimization problems, it is convenient to refer to the "Karush–Kuhn–Tucker" (or KKT) conditions. The first-order KKT conditions for the inequality-constrained problem hold at the point $x^\star$, if there exists an m-vector $\lambda^\star$ called a Lagrange-multiplier vector, such that [4]

$$c(x^\star) \geq 0 \;\; \text{(feasibility condition)} \tag{6.66}$$

$$g(x^\star) = J(x^\star)^T \lambda^\star \;\; \text{(stationarity condition)} \tag{6.67}$$

$$\lambda^\star \geq 0 \;\; \text{(nonnegativity of the multipliers)} \tag{6.68}$$

$$c(x^\star) \cdot \lambda^\star = 0 \;\; \text{(complementarity)} \tag{6.69}$$

The stationarity condition (6.67) can be written as

$$\nabla_x L\left(x^\star, \lambda^\star\right) = 0, \quad \text{where } L\left(x, \lambda\right) \triangleq f(x) - \lambda^T c(x) \tag{6.70}$$

where λ is often generically known as a *Lagrangian multiplier* and Equation (6.70) as the Lagrangian equation. The Karush–Kuhn–Tucker conditions are necessary for a solution in nonlinear programming to be optimal.

6.3.1 Quadratic Programming

A special subset of nonlinear problems is *quadratic problems* that are characterized by the following formulation:

$$\text{minimize} \quad f(x) = \frac{1}{2} x^T Q x + c^T x \quad x \in \mathbb{R}^n \tag{6.71}$$

$$\text{subject to} \quad Ax \leq b \tag{6.72}$$

$$x \geq 0 \tag{6.73}$$

If Q is a positive semidefinite matrix, then $f(x)$ is a convex function. If Q is zero, then the problem becomes a linear program. The Lagrangian function for the quadratic program is given by

$$L(x, \lambda) = c^T x + \frac{1}{2} x^T Q x + \lambda \left(Ax - b\right) \tag{6.74}$$

where λ is an m dimensional row vector. The KKT conditions for a local minima are

$$c^T + x^T Q + \lambda A \geq 0 \tag{6.75}$$

$$Ax - b \leq 0 \tag{6.76}$$

$$x^T \left(c + Qx + A^T \lambda\right) = 0 \tag{6.77}$$

$$\lambda(Ax - b) = 0 \tag{6.78}$$

$$x \geq 0 \tag{6.79}$$

$$\lambda \geq 0 \tag{6.80}$$

To put these equations in a more manageable form, the nonnegative slack variable $y \in \mathbb{R}^n$ is introduced to the inequalities in Equation (6.75) and the slack variable $v \in \mathbb{R}^m$ in Equation (6.76) to obtain

$$c + Qx + A^T \lambda^T - y = 0 \tag{6.81}$$

$$Ax - b + v = 0 \tag{6.82}$$

and the KKT equations are

$$Qx + A^T \lambda^T - y = -c^T \tag{6.83}$$

$$Ax + v = b \tag{6.84}$$

$$x \geq 0 \tag{6.85}$$
$$y \geq 0 \tag{6.86}$$
$$v \geq 0 \tag{6.87}$$
$$\lambda \geq 0 \tag{6.88}$$
$$y^T x = 0 \tag{6.89}$$
$$\lambda v = 0 \tag{6.90}$$

Now any linear programming method can be used to solve this set of equations by treating the complementary slackness conditions (Equations (6.89 and 6.90)) implicitly with a restricted basis entry rule. The goal is to find the solution to the linear program problem with the additional requirement that the complementary slackness conditions be satisfied at each iteration. The objective function is satisfied by adding an artificial variable to each equation and minimizing the sum of the artificial variables.

Example 6.6
Minimize

$$f(x): \quad -10x_1 - 8x_2 + x_1^2 + 2x_2^2$$

subject to the following constraints:

$$x_1 + x_2 \leq 10$$
$$x_2 \leq 5$$
$$x_1 \geq 0,\ x_2 \geq 0$$

Solution 6.6 This can be written in matrix form as
Minimize

$$f(x): \quad [-10 \quad -8] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \frac{1}{2} [x_1 \quad x_2] \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

subject to the following constraints:

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leq \begin{bmatrix} 10 \\ 5 \end{bmatrix}$$

The linear program problem becomes (where the artificial variables are given by $a_1 - a_4$)

Minimize $a_1 + a_2 + a_3 + a_4$
subject to

$$\begin{array}{rrrrrrrrrrrrr}
2x_1 & & +\lambda_1 & & -y_1 & & & & +a_1 & & & & = 10 \\
 & 4x_2 & +\lambda_1 & +\lambda_2 & & -y_2 & & & & +a_2 & & & = 8 \\
x_1 & +x_2 & & & & & +v_1 & & & & +a_3 & & = 10 \\
 & x_2 & & & & & & +v_2 & & & & +a_4 & = 5
\end{array}$$

This can now be solved using either the simplex or interior point method. The solution is

$$x = \begin{bmatrix} 5.0000 \\ 2.0000 \end{bmatrix}$$
$$\lambda = \begin{bmatrix} 0.6882 \\ 0.7439 \end{bmatrix}$$
$$y = \begin{bmatrix} 0.6736 \\ 1.3354 \end{bmatrix}$$
$$v = \begin{bmatrix} 3.0315 \\ 3.0242 \end{bmatrix}$$
$$a = [0 \;\; 0 \;\; 0 \;\; 0]$$

with the cost function $f(x) = -33$. ■

6.3.2 Steepest Descent Algorithm

For engineering applications, general nonlinear programming problems are generally solved by two classes of approaches:

1. gradient methods such as steepest descent, or
2. iterative programming techniques such as successive quadratic programming

In an unconstrained system, the usual approach to minimizing the function $f(x)$ is to set the function derivatives to zero and then solve for the system states from the set of resulting equations. In the majority of applications, however, the system states evaluated at the unconstrained minimum will not satisfy the constraint equations. Thus an alternate approach is required to find the constrained minimum. One approach is to introduce an additional set of parameters λ, frequently known as *Lagrange multipliers*, to impose the constraints on the cost function. The augmented cost function then becomes

$$\text{minimize} \;\; f(x) - \lambda c(x) \tag{6.91}$$

The augmented function in Equation (6.91) can then be minimized by solving for the set of states that result from setting the derivatives of the augmented function to zero. Note that the derivative of Equation (6.91) with respect to λ effectively enforces the equality constraint of Equation (6.64).

Example 6.7
Minimize

$$C: \;\; \frac{1}{2}\left(x^2 + y^2\right) \tag{6.92}$$

subject to the following constraint:

$$2x - y = 5$$

Solution 6.7 Note that the function to be minimized is the equation for a circle. The unconstrained minimum of this function is the point at the origin with $x = 0$ and $y = 0$ which defines a circle with a radius of zero length. However, the circle must also intersect the line defined by the constraint equation; thus the constrained circle must have a nonzero radius. The augmented cost function becomes

$$C^* : \quad \frac{1}{2}\left(x^2 + y^2\right) - \lambda\left(2x - y - 5\right) \tag{6.93}$$

where λ represents the Lagrange multiplier. Setting the derivatives of the augmented cost function to zero yields the following set of equations:

$$0 = \frac{\partial C^*}{\partial x} = x - 2\lambda$$
$$0 = \frac{\partial C^*}{\partial y} = y + \lambda$$
$$0 = \frac{\partial C^*}{\partial \lambda} = 2x - y - 5$$

Solving this set of equations yields $[x \;\; y \;\; \lambda]^T = [2 \;\; -1 \;\; 1]^T$. The cost function of Equation (6.92) evaluated at the minimum of the augmented cost function is

$$C : \quad \frac{1}{2}\left((2)^2 + (-1)^2\right) = \frac{5}{2}$$

■

Both the cost function f and the constraint equations c may be a function of an external input u, which may be varied to minimize the function $f(x)$. In this case, Equation (6.91) may be more generally written as

$$\text{minimize} \;\; f(x, u) - \lambda c(x, u) \tag{6.94}$$

If there is more than one equality constraint, then λ becomes a vector of multipliers and the augmented cost function becomes

$$C^* : \quad f(x) - [\lambda]^T c(x) \tag{6.95}$$

where the derivatives of C^* become

$$\left[\frac{\partial C^*}{\partial \lambda}\right] = 0 = c(x) \tag{6.96}$$

$$\left[\frac{\partial C^*}{\partial x}\right] = 0 = \left[\frac{\partial f}{\partial x}\right] - \left[\frac{\partial c}{\partial x}\right]^T [\lambda] \tag{6.97}$$

$$\left[\frac{\partial C^*}{\partial u}\right] = 0 = \left[\frac{\partial f}{\partial u}\right] - \left[\frac{\partial c}{\partial u}\right]^T [\lambda] \tag{6.98}$$

Note that, for any *feasible* solution to the equality constraint, Equation (6.96) is satisfied, but the feasible solution may not be the optimal solution which minimizes the cost function. In this case, $[\lambda]$ can be obtained from Equation (6.97) and then only

$$\left[\frac{\partial C^*}{\partial u}\right] \neq 0$$

This vector can be used as a gradient vector $[\nabla C]$ which is orthogonal to the contour of constant values of the cost function C. Thus

$$[\lambda] = \left[\left[\frac{\partial c}{\partial x}\right]^T\right]^{-1} \left[\frac{\partial f}{\partial x}\right] \tag{6.99}$$

which gives

$$\nabla C = \left[\frac{\partial C^*}{\partial u}\right] = \left[\frac{\partial f}{\partial u}\right] - \left[\frac{\partial c}{\partial u}\right]^T [\lambda] \tag{6.100}$$

$$= \left[\frac{\partial f}{\partial u}\right] - \left[\frac{\partial c}{\partial u}\right]^T \left[\left[\frac{\partial c}{\partial x}\right]^T\right]^{-1} \left[\frac{\partial f}{\partial x}\right] \tag{6.101}$$

This relationship provides the foundation of the optimization method known as the *steepest descent* algorithm.

Steepest Descent Algorithm

1. Let $k = 0$. Guess an initial vector $u^k = u^0$.
2. Solve the (possibly nonlinear) system of Equations (6.96) for a feasible solution x.
3. Calculate C^{k+1} and ∇C^{k+1} from Equation (6.101). If $\|C^{k+1} - C^k\|$ is less than some predefined tolerance, stop.
4. Calculate the new vector $u^{k+1} = u^k - \gamma \nabla C^{k+1}$, where γ is a positive number which is the user-defined "step size" of the algorithm.
5. $k = k + 1$. Go to step 2.

In the steepest descent method, the u vector update direction is determined at each step of the algorithm by choosing the direction of the greatest change of the augmented cost function C^*. For example, consider a person skiing from the top of a mountain to the bottom, as illustrated in Figure 6.4. The skier will travel in a straight path for a certain distance. At that point, he may no longer be pointed directly down the mountain. Therefore, he will adjust his direction so that his skis point in the direction of steepest descent. The direction of steepest descent is perpendicular to the tangent of the curve of

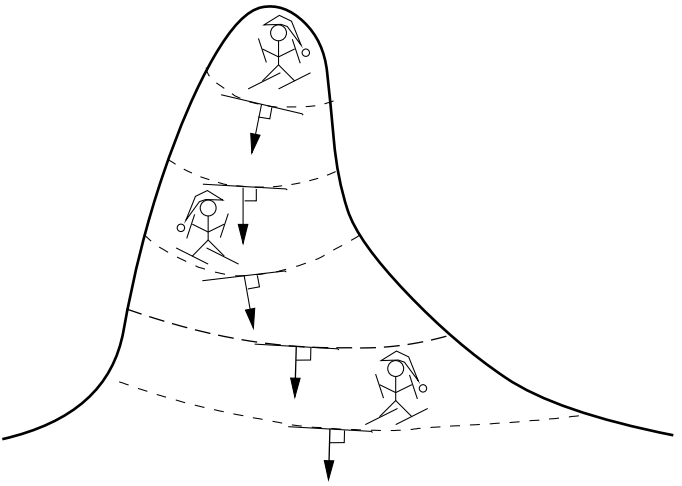

FIGURE 6.4
Example of steepest descent

constant altitude (or cost). The distance the skier travels between adjustments is analogous to the step size γ of the algorithm. For small γ, the skier will frequently alter direction; thus his descent will be slow. For large γ, however, he may overshoot the foot of the mountain and will start ascending again. Thus the critical part of the steepest descent algorithm is the choice of γ. If γ is chosen small, then convergence to a minimum value is more likely, but may require many iterations, whereas a large value of γ may result in oscillations about the minimum.

Example 6.8
Minimize

$$C: \quad x_1^2 + 2x_2^2 + u^2 = f(x_1, x_2, u) \tag{6.102}$$

subject to the following constraints:

$$0 = x_1^2 - 3x_2 + u - 3 \tag{6.103}$$

$$0 = x_1 + x_2 - 4u + 2 \tag{6.104}$$

Solution 6.8 To find ∇C of Equation (6.101), the following partial derivatives are required:

$$\left[\frac{\partial f}{\partial u}\right] = 2u$$

$$\left[\frac{\partial f}{\partial x}\right] = \begin{bmatrix} 2x_1 \\ 4x_2 \end{bmatrix}$$
$$\left[\frac{\partial c}{\partial u}\right]^T = \begin{bmatrix} 1 & -4 \end{bmatrix}$$
$$\left[\frac{\partial c}{\partial x}\right] = \begin{bmatrix} 2x_1 & -3 \\ 1 & 1 \end{bmatrix}$$

yielding

$$\begin{aligned} \nabla C &= \left[\frac{\partial f}{\partial u}\right] - \left[\frac{\partial c}{\partial u}\right]^T \left[\left[\frac{\partial c}{\partial x}\right]^T\right]^{-1} \left[\frac{\partial f}{\partial x}\right] \\ &= 2u - \begin{bmatrix} 1 & -4 \end{bmatrix} \left[\begin{bmatrix} 2x_1 & -3 \\ 1 & 1 \end{bmatrix}^T\right]^{-1} \begin{bmatrix} 2x_1 \\ 4x_2 \end{bmatrix} \end{aligned}$$

Iteration 1

Let $u = 1, \gamma = 0.05$, and choose a stopping criterion of $\epsilon = 0.0001$. Solving for x_1 and x_2 yields two values for each with a corresponding cost function:

$$\begin{aligned} x_1 &= 1.7016 \quad x_2 = 0.2984 \quad f = 4.0734 \\ x_1 &= -4.7016 \; x_2 = 6.7016 \quad f = 23.2828 \end{aligned}$$

The first set of values leads to the minimum cost function, so they are selected as the operating solution. Substituting $x_1 = 1.7016$ and $x_2 = 0.2984$ into the gradient function yields $\nabla C = 10.5705$ and the new value of u becomes

$$\begin{aligned} u^{(2)} &= u^{(1)} - \gamma \nabla C \\ &= 1 - (0.05)(10.5705) \\ &= 0.4715 \end{aligned}$$

Iteration 2

With $u = 0.4715$, solving for x_1 and x_2 again yields two values for each with a corresponding cost function:

$$\begin{aligned} x_1 &= 0.6062 \quad x_2 = -0.7203 \; f = 1.6276 \\ x_1 &= -3.6062 \; x_2 = 3.4921 \quad f = 14.2650 \end{aligned}$$

The first set of values again leads to the minimum cost function, so they are selected as the operating solution. The difference in cost functions is

$$\left|C^{(1)} - C^{(2)}\right| = |4.0734 - 1.6276| = 2.4458$$

which is greater than the stopping criterion. Substituting these values into the gradient function yields $\nabla C = 0.1077$ and the new value of u becomes

$$\begin{aligned} u^{(3)} &= u^{(2)} - \gamma \nabla C \\ &= 0.4715 - (0.05)(0.1077) \\ &= 0.4661 \end{aligned}$$

Iteration 3

With $u = 0.4661$, solving for x_1 and x_2 again yields two values for each with a corresponding cost function:

$$\begin{aligned} x_1 &= 0.5921 \quad x_2 = -0.7278 \; f = 1.6271 \\ x_1 &= -3.5921 \; x_2 = 3.4565 \quad f = 14.1799 \end{aligned}$$

The first set of values again leads to the minimum cost function, so they are selected as the operating solution. The difference in cost functions is

$$\left| C^{(2)} - C^{(3)} \right| = |1.6276 - 1.6271| = 0.0005$$

which is greater than the stopping criterion. Substituting these values into the gradient function yields $\nabla C = 0.0541$ and the new value of u becomes

$$\begin{aligned} u^{(4)} &= u^{(3)} - \gamma \nabla C \\ &= 0.4661 - (0.05)(0.0541) \\ &= 0.4634 \end{aligned}$$

This process is continued until the stopping criterion is reached. The final values are $x_1 = 0.5774, x_2 = -0.7354$, and $u = 0.4605$ which yield the minimum cost function $f = 1.6270$. ■

6.3.3 Sequential Quadratic Programming Algorithm

Gradient descent techniques work well for small nonlinear systems, but become inefficient as the dimension of the search space grows. The nonlinear sequential quadratic programming (SQP) optimization method is computationally efficient and has been shown to exhibit super linear convergence for convex search spaces [5]. The SQP is an iterative procedure which models the nonlinear optimization problem for a given iterate x^k by a quadratic programming subproblem, whose solution is used to construct a new iterate x^{k+1}. Thus it sequentially solves the QP algorithm to produce the solution to the original nonlinear problem in much the same way that the Newton–Raphson sequentially solves a series of linear problems to find the solution to a nonlinear problem.

The SQP method also attempts to solve the system

$$\text{minimize} \quad f(x) \quad x \in \mathbb{R}^n \tag{6.105}$$

$$\text{subject to} \quad c_i(x) = 0, \quad i \in \xi \tag{6.106}$$

$$h_i(x) \geq 0, \quad i \in \Xi \tag{6.107}$$

As before, the usual approach to solving this optimization problem is to use Lagrangian multipliers and minimize the hybrid system:

$$L(x, \lambda) = f(x) + \lambda^T c(x) + \pi^T h(x) \tag{6.108}$$

The KKT conditions are

$$\begin{aligned} \nabla f(x) + C^T\lambda + H^T\pi &= 0 \\ c(x) &= 0 \\ h(x) + s &= 0 \\ \pi^T s &= 0 \\ \pi, s &\geq 0 \end{aligned}$$

where λ is a vector of Lagrange multipliers for the equality constraints, π is a vector of Lagrange multipliers for the inequality constraints, s is a vector of slack variables, and

$$C = \frac{\partial c(x)}{\partial x} \tag{6.109}$$

$$H = \frac{\partial h(x)}{\partial x} \tag{6.110}$$

This nonlinear system can be solved for x, λ, π, and s using the Newton–Raphson method. Consider the case of only x and λ. Using the Newton–Raphson method to solve for $y = [x \;\; \lambda]^T$ with

$$F(y) = 0 = \begin{bmatrix} \nabla f(x) + C^T\lambda \\ c(x) \end{bmatrix}$$

and the Newton–Raphson update becomes

$$y^{k+1} = y^k - [\nabla F_k]^{-1} F(y^k)$$

or in original variables

$$\begin{bmatrix} x^{k+1} \\ \lambda^{k+1} \end{bmatrix} = \begin{bmatrix} x^k \\ \lambda^k \end{bmatrix} - \begin{bmatrix} \nabla^2 L & \nabla c^T \\ \nabla c^T & 0 \end{bmatrix}_k^{-1} \begin{bmatrix} \nabla L \\ c \end{bmatrix}_k \tag{6.111}$$

Example 6.9
Repeat Example 6.8 using the SQP method.

Solution 6.9 The problem is repeated here:

Minimize

$$x_1^2 + 2x_2^2 + u^2 = f(x_1, x_2, u) \tag{6.112}$$

subject to the following constraints:

$$0 = x_1^2 - 3x_2 + u - 3 \tag{6.113}$$

$$0 = x_1 + x_2 - 4u + 2 \tag{6.114}$$

Applying the KKT conditions yields the following nonlinear system of equations:

$$\begin{aligned} 0 &= 2x_1 + 2\lambda_1 x_1 + \lambda_2 \\ 0 &= 4x_2 - 3\lambda_1 + \lambda_2 \\ 0 &= 2u + \lambda_1 - 4\lambda_2 \\ 0 &= x_1^2 - 3x_2 + u - 3 \\ 0 &= x_1 + x_2 - 4u + 2 \end{aligned}$$

The Newton–Raphson iteration becomes

$$\begin{bmatrix} x_1 \\ x_2 \\ u \\ \lambda_1 \\ \lambda_2 \end{bmatrix}^{k+1} = \begin{bmatrix} x_1 \\ x_2 \\ u \\ \lambda_1 \\ \lambda_2 \end{bmatrix}^{k} - \begin{bmatrix} 2+2\lambda_1^k & 0 & 0 & 2x_1^k & 1 \\ 0 & 4 & 0 & -3 & 1 \\ 0 & 0 & 2 & 1 & -4 \\ 2x_1^k & -3 & 1 & 0 & 0 \\ 1 & 1 & -4 & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 2x_1^k + 2\lambda_1^k x_1^k + \lambda_2^k \\ 4x_2^k - 3\lambda_1^k + \lambda_2^k \\ 2u^k + \lambda_1^k - 4\lambda_2^k \\ (x_1^k)^2 - 3x_2^k + u^k - 3 \\ x_1^k + x_2^k - 4u^k + 2 \end{bmatrix} \tag{6.115}$$

Starting with an initial guess $\begin{bmatrix} x_1 & x_2 & u & \lambda_1 & \lambda_2 \end{bmatrix}^T = [1 \ 1 \ 1 \ 1 \ 1]^T$ yields the following updates:

k	x_1	x_2	u	λ_1	λ_2	$f(x)$
0	1.0000	1.0000	1.0000	1.0000	1.0000	4.0000
1	0.5774	–0.7354	0.4605	–0.9859	–0.0162	1.6270
2	0.8462	–0.5804	0.5664	–0.7413	0.0979	1.7106
3	0.6508	–0.7098	0.4853	–0.9442	0.0066	1.6666
4	0.5838	–0.7337	0.4625	–0.9831	–0.0145	1.6314
5	0.5774	–0.7354	0.4605	–0.9859	–0.0162	1.6270

which is the same solution as previously. ■

6.4 Power System Applications

6.4.1 Optimal Power Flow

Many power system applications, such as the power flow, offer only a snapshot of the system operation. Frequently, the system planner or operator is interested in the effect that making adjustments to the system parameters will have on the power flow through lines or system losses. Rather than making the adjustments in a random fashion, the system planner will attempt to optimize the adjustments according to some objective function. This objective function can be chosen to minimize generating costs, reservoir water levels, or system losses, among others. The optimal power flow problem is to formulate the power flow problem to find system voltages and generated powers within the framework of the objective function. In this application, the inputs to the power flow are systematically adjusted to maximize (or minimize) a scalar function of the power flow state variables. The two most common objective functions are minimization of generating costs and minimization of active power losses.

The time frame of optimal power flow is on the order of minutes to one hour; therefore, it is assumed that the optimization occurs using only those units that are currently on-line. The problem of determining whether or not to engage a unit, at what time, and for how long is part of the *unit commitment* problem and is not covered here. The minimization of active transmission losses saves both generating costs and creates a higher generating reserve margin.

Usually, generator cost curves (the curves that relate generated power to the cost of such generation) are given as piecewise linear incremental cost curves. This has its origin in the simplification of concave cost functions with the valve points as cost curve breakpoints [20]. Piecewise linear incremental cost curves correspond to piecewise quadratic cost curves by integrating the incremental cost curves. This type of objective function lends itself easily to the *economic dispatch*, or λ-dispatch problem where only generating units are considered in the optimization. In this process, system losses and constraints on voltages and line powers are neglected. This economic dispatch method is illustrated in the following example.

Example 6.10

Three generators with the following cost functions serve a load of 952 MW. Assuming a lossless system, calculate the optimal generation scheduling.

$$\begin{aligned} C_1 &: \ P_1 + 0.0625P_1^2 \quad \$/\text{hr} \\ C_2 &: \ P_2 + 0.0125P_2^2 \quad \$/\text{hr} \\ C_3 &: \ P_3 + 0.0250P_3^2 \quad \$/\text{hr} \end{aligned}$$

Solution 6.10 The first step in determining the optimal scheduling of the generators is to construct the problem in the general form. Thus the optimization statement is

$$\begin{aligned} \text{Minimize C:} \quad & P_1 + 0.0625P_1^2 + P_2 + 0.0125P_2^2 + P_3 + 0.0250P_3^2 \\ \text{Subject to:} \quad & P_1 + P_2 + P_3 - 952 = 0 \end{aligned}$$

From this statement, the constrained cost function becomes

$$C^*: \quad P_1 + 0.0625P_1^2 + P_2 + 0.0125P_2^2 + P_3 + 0.0250P_3^2 - \lambda\,(P_1 + P_2 + P_3 - 952) \tag{6.116}$$

Setting the derivatives of C^* to zero yields the following set of linear equations:

$$\begin{bmatrix} 0.125 & 0 & 0 & -1 \\ 0 & 0.025 & 0 & -1 \\ 0 & 0 & 0.050 & -1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \\ P_3 \\ \lambda \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ -1 \\ 952 \end{bmatrix} \tag{6.117}$$

Solving Equation (6.117) yields

$$\begin{aligned} P_1 &= 112 \text{ MW} \\ P_2 &= 560 \text{ MW} \\ P_3 &= 280 \text{ MW} \\ \lambda &= \$15/\text{MWhr} \end{aligned}$$

for a constrained cost of \$7,616/hr. ■

This is the generation scheduling that minimizes the hourly cost of production. The value of λ is the *incremental* or *break-even* cost of production. This gives a company a price cut-off for buying or selling generation: if they can purchase generation for less than λ, then their overall costs will decrease. Likewise, if generation can be sold for greater than λ, their overall costs will decrease. Also note that at the optimal scheduling

$$\lambda = 1 + 0.125P_1 = 1 + 0.025P_2 = 1 + 0.050P_3 \tag{6.118}$$

Since λ is the incremental cost for the system, this point is also called the point of "equal incremental cost," and the generation schedule is said to satisfy the "equal incremental cost criterion." Any deviation in generation from the equal increment cost scheduling will result in an increase in the production cost C.

Example 6.11

If a buyer is willing to pay \$16/MWhr for generation, how much excess generation should be produced and sold, and what is the profit for this transaction?

Solution 6.11 From Example 6.10, the derivatives of the augmented cost function yield the following relationships between generation and λ:

$$\begin{aligned} P_1 &= 8\,(\lambda - 1) \\ P_2 &= 40\,(\lambda - 1) \\ P_3 &= 20\,(\lambda - 1) \end{aligned}$$

from which the equality constraint yields

$$8\,(\lambda - 1) + 40\,(\lambda - 1) + 20\,(\lambda - 1) - 952 = 0 \tag{6.119}$$

To determine the excess amount, the equality Equation (6.119) will be augmented and then evaluated at $\lambda = \$16$/MWhr:

$$8\,(16 - 1) + 40\,(16 - 1) + 20\,(16 - 1) - (952 + \text{excess}) = 0 \tag{6.120}$$

Solving Equation (6.120) yields a required excess of 68 MW, and $P_1 = 120$ MW, $P_2 = 600$ MW, and $P_3 = 300$ MW. The total cost of generation becomes

$$C: \quad P_1 + 0.0625P_1^2 + P_2 + 0.0125P_2^2 + P_3 + 0.0250P_3^2 = \$8,670/hr \tag{6.121}$$

but the amount recovered by the sale of generation is the amount of excess times the incremental cost λ,

$$68 \text{ MW} \times \$16/\text{MWhr} \ = \$1,088/\text{hr}$$

Therefore, the total cost is \$8,670 – 1,088 = \$7,580/hr. This amount is \$34/hr less than the original cost of \$7,616/hr; thus \$34/hr is the profit achieved from the sale of the excess generation at \$16/MWhr. ■

Figure 6.5 shows an incremental cost table for a medium size utility. The incremental cost of generation is listed vertically along the left-hand side of the table. The various generating units are listed across the top from least expensive to most expensive (left to right). Nuclear units are among the least expensive units to operate and the nuclear unit *Washington* at the far left can produce up to 1222 MW at an incremental cost of 7.00 \$/MWhr. This incremental cost is half of the next least expensive unit, *Adams* at 14 \$/MWhr, which is a coal unit. As the available units become increasingly more expensive to operate, the incremental cost also increases.

Example 6.12
What is the incremental cost for the utility to produce 12,500 MW?

Solution 6.12 To find the incremental cost that corresponds to 12,500 MW from the incremental cost table in Figure 6.5, the maximum generation available from each unit is summed until it equals 12,500 MW. This amount is

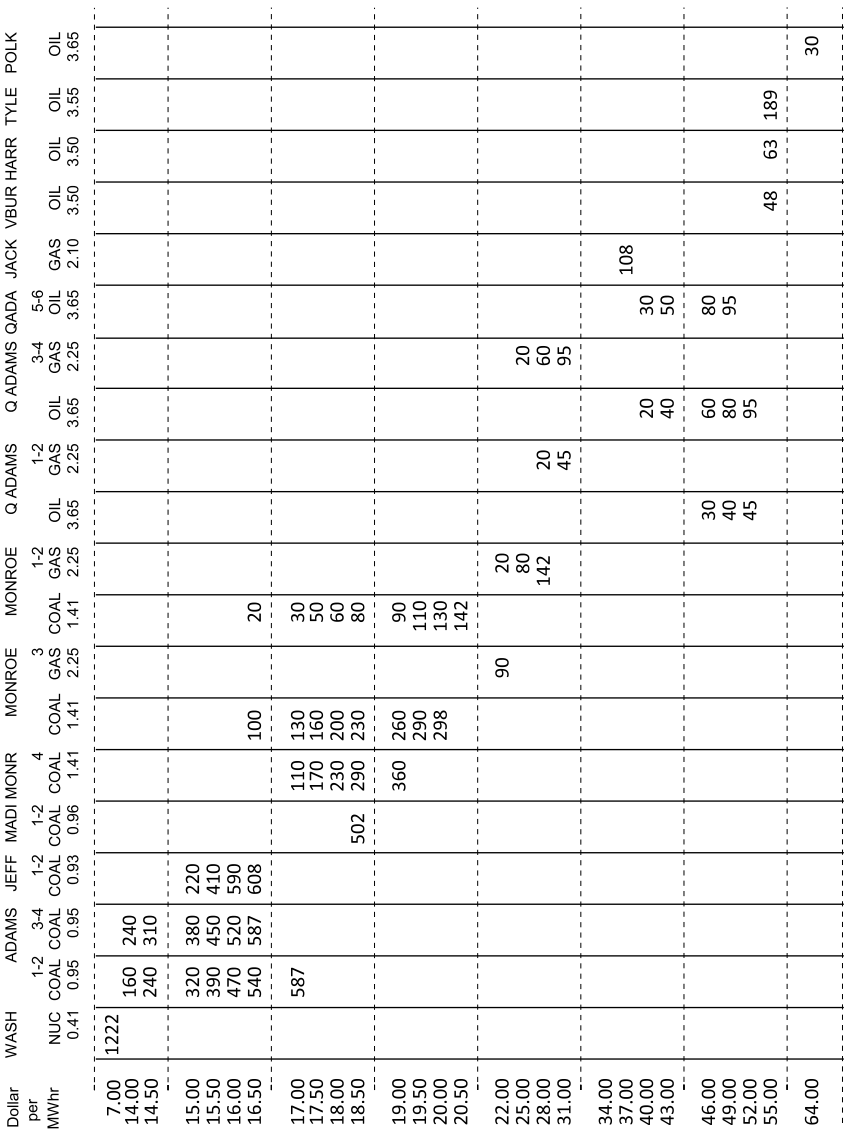

FIGURE 6.5
Incremental cost table

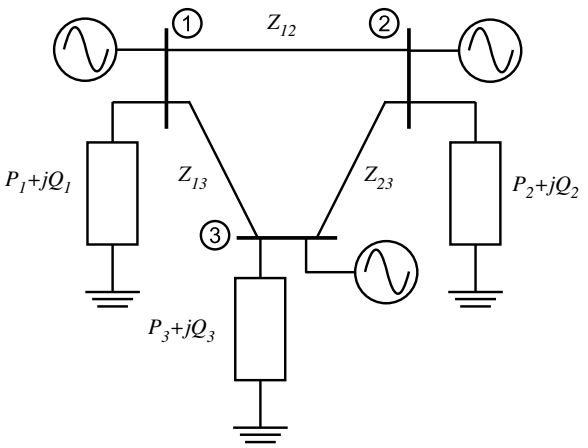

FIGURE 6.6
Figure for Example 6.13

reached by the gas unit of *Monroe 1-2*. This corresponds to an incremental cost of 28.00$/MW hr. This is the breakeven point for 12,500 MW.

If power can be purchased for less than 28.00 $/MW hr, the utility should purchase generation. ■

The primary drawback with the equal incremental cost scheduling is that it neglects all losses in the system. The only enforced equality constraint is that the sum of the generation must equal the total load demand. In reality, however, the sum of the generation must equal the load demand plus any system losses. In the consideration of system losses, the equality constraints must include the set of power flow equations, and the optimization process must to be extended to the steepest descent, or similar, approach [11].

Example 6.13

Consider the three machine system shown in Figure 6.6. This system has the same parameters as the three-bus system of Example 3.11 except that bus 3 has been converted to a generator bus with a voltage magnitude of 1.0 pu. The cost functions of the generators are the same as in Example 6.10. The per unit loads are

bus	P_{load}	Q_{load}
1	0.312	0.1
2	0.320	0.1
3	0.320	0.1

Using $P_{gen} = [0.112 \quad 0.560 \quad 0.280]'$ as a starting point (from Example 6.10), find the optimal scheduling of this system considering losses. Let $\gamma = 1$.

Solution 6.13 Following the steepest descent procedure detailed in Section 6.3.2, the first step is to develop an expression for the gradient ∇C, where

$$\nabla C = \left[\frac{\partial f}{\partial u}\right] - \left[\frac{\partial g}{\partial u}\right]^T \left[\left[\frac{\partial g}{\partial x}\right]^T\right]^{-1} \left[\frac{\partial f}{\partial x}\right] \tag{6.122}$$

where f is the sum of the generator costs

$$f: \quad C_1 + C_2 + C_3 = P_1 + 0.0625P_1^2 + P_2 + 0.0125P_2^2 + P_3 + 0.0250P_3^2$$

g is the set of power flow equations

$$g_1: \quad 0 = P_2 - P_{L2} - V_2 \sum_{i=1}^{3} V_i Y_{2i} \cos\left(\theta_2 - \theta_i - \phi_{2i}\right)$$

$$g_2: \quad 0 = P_3 - P_{L3} - V_3 \sum_{i=1}^{3} V_i Y_{3i} \cos\left(\theta_3 - \theta_i - \phi_{3i}\right)$$

where P_{Li} denotes the active power load at bus i, the set of inputs u is the set of independent generation settings

$$u = \begin{bmatrix} P_2 \\ P_3 \end{bmatrix}$$

and x is the set of unknown states

$$x = \begin{bmatrix} \theta_2 \\ \theta_3 \end{bmatrix}$$

The generator setting P_1 is not an input because it is the slack bus generation and cannot be independently set. From these designations, the various partial derivatives required for ∇C can be derived:

$$\left[\frac{\partial g}{\partial u}\right] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{6.123}$$

$$\left[\frac{\partial g}{\partial x}\right] = \begin{bmatrix} \frac{\partial g_1}{\partial \theta_2} & \frac{\partial g_1}{\partial \theta_3} \\ \frac{\partial g_2}{\partial \theta_2} & \frac{\partial g_2}{\partial \theta_3} \end{bmatrix} \tag{6.124}$$

where

$$\frac{\partial g_1}{\partial \theta_2} = V_2 \left(V_1 Y_{12} \sin\left(\theta_2 - \theta_1 - \phi_{21}\right) + V_3 Y_{13} \sin\left(\theta_2 - \theta_3 - \phi_{23}\right)\right) \tag{6.125}$$

$$\frac{\partial g_1}{\partial \theta_3} = -V_2 V_3 Y_{32} \sin\left(\theta_2 - \theta_3 - \phi_{23}\right) \tag{6.126}$$

$$\frac{\partial g_2}{\partial \theta_2} = -V_3 V_2 Y_{23} \sin\left(\theta_3 - \theta_2 - \phi_{32}\right) \tag{6.127}$$

$$\frac{\partial g_2}{\partial \theta_3} = V_3 \left(V_1 Y_{13} \sin\left(\theta_3 - \theta_1 - \phi_{31}\right) + V_2 Y_{23} \sin\left(\theta_3 - \theta_2 - \phi_{32}\right)\right) \tag{6.128}$$

and

$$\left[\frac{\partial f}{\partial u}\right] = \begin{bmatrix} 1 + 0.025P_2 \\ 1 + 0.050P_3 \end{bmatrix} \tag{6.129}$$

Finding the partial derivative $\left[\frac{\partial f}{\partial x}\right]$ is slightly more difficult since the cost function is not written as a direct function of x. Recall, however, that P_1 is not an input, but is actually a quantity that depends on x, i.e.,

$$\begin{aligned} P_1 = V_1 &\left(V_1 Y_{11} \cos\left(\theta_1 - \theta_1 - \phi_{11}\right)\right. \\ &+ V_2 Y_{12} \cos\left(\theta_1 - \theta_2 - \phi_{12}\right) + \left. V_3 Y_{13} \cos\left(\theta_1 - \theta_3 - \phi_{13}\right)\right) \end{aligned} \tag{6.130}$$

Thus, using the chain rule,

$$\left[\frac{\partial f}{\partial x}\right] = \left[\frac{\partial f}{\partial P_1}\right]\left[\frac{\partial P_1}{\partial x}\right] \tag{6.131}$$

$$= \left(1 + 0.125P_1\right)\begin{bmatrix} V_1 V_2 Y_{12} \sin\left(\theta_1 - \theta_2 - \phi_{12}\right) \\ V_1 V_3 Y_{13} \sin\left(\theta_1 - \theta_3 - \phi_{13}\right) \end{bmatrix} \tag{6.132}$$

The initial values of P_2 and P_3 are obtained from the equal incremental cost rule. Using $P_2 = 0.56$ pu and $P_3 = 0.28$ pu as inputs into the power flow yields the following states:

i	θ_i (rad)	V_i (pu)	$P_{gen,i}$ (pu)
1	0	1.02	0.1131
2	0.0279	1.00	0.5600
3	0.0128	1.00	0.2800

Substituting these values into the partial derivatives yields

$$\left[\frac{\partial g}{\partial u}\right] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{6.133}$$

$$\left[\frac{\partial g}{\partial x}\right] = \begin{bmatrix} -13.3048 & 9.9148 \\ 9.8849 & -19.9960 \end{bmatrix} \tag{6.134}$$

$$\left[\frac{\partial f}{\partial u}\right] = \begin{bmatrix} 1.0140 \\ 1.0140 \end{bmatrix} \tag{6.135}$$

$$\left[\frac{\partial f}{\partial x}\right] = 1.0141\begin{bmatrix} -3.3773 \\ -10.0853 \end{bmatrix} = \begin{bmatrix} -3.4251 \\ -10.2278 \end{bmatrix} \tag{6.136}$$

which yields

$$\nabla C = \begin{bmatrix} 0.0048 \\ 0.0021 \end{bmatrix} \tag{6.137}$$

Thus the new values for the input generation are

$$\begin{bmatrix} P_2 \\ P_3 \end{bmatrix} = \begin{bmatrix} 0.560 \\ 0.280 \end{bmatrix} - \gamma \begin{bmatrix} 0.0048 \\ 0.0021 \end{bmatrix} = \begin{bmatrix} 0.5552 \\ 0.2779 \end{bmatrix} \tag{6.138}$$

Thus $P_2 = 555.2$ and $P_3 = 277.9$ MW.

Already the gradient ∇C is very small, indicating that the generation values from the equal incremental cost process were relatively close to the optimal values, even considering losses.

Proceeding until convergence is achieved yields the final generation values for all of the generators:

$$\begin{bmatrix} P_1 \\ P_2 \\ P_3 \end{bmatrix} = \begin{bmatrix} 136.6 \\ 517.9 \\ 298.4 \end{bmatrix} \quad \text{MW}$$

which yields a cost of \$7,698/MW hr. Note that this amount is greater than the calculated cost for the equal incremental cost function. This increase is due to the extra generation required to satisfy the losses in the system. ■

Often the steepest descent method may indicate either states or inputs lie outside of their physical constraints. For example, the algorithm may result in a power generation value that exceeds the physical maximum output of the generating unit. Similarly, the resulting bus voltages may lie outside of the desired range (usually ± 10% of unity). These are violations of the *inequality constraints* of the problem. In these cases, the steepest descent algorithm must be modified to reflect these physical limitations. There are several approaches to account for limitations and these approaches depend on whether or not the limitation is on the input (independent) or on the state (dependent).

6.4.1.1 Limitations on Independent Variables

If the application of the steepest descent algorithm results in an updated value of input that exceeds the specified limit, then the most straightforward method of handling this violation is simply to set the input state equal to its limit and continue with the algorithm except with one less degree of freedom.

Example 6.14

Repeat Example 6.13 except that the generators must satisfy the following limitations:

$$\begin{aligned} 80 &\leq P_1 \leq 1200 \ \text{MW} \\ 450 &\leq P_2 \leq 750 \ \text{MW} \\ 150 &\leq P_3 \leq 250 \ \text{MW} \end{aligned}$$

Solution 6.14 From the solution of Example 6.13, the output of generator 3 exceeds the maximum limit of 0.25 pu. Therefore, after the first iteration in the previous example, P_3 is set to 0.25 pu. The new partial derivatives

become

$$\left[\frac{\partial g}{\partial u}\right] = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{6.139}$$

$$\left[\frac{\partial g}{\partial x}\right] = \text{same} \tag{6.140}$$

$$\left[\frac{\partial f}{\partial u}\right] = [1 + 0.025P_2] \tag{6.141}$$

$$\left[\frac{\partial f}{\partial x}\right] = \text{same} \tag{6.142}$$

From the constrained steepest descent, the new values of generation become

$$\begin{bmatrix} P_1 \\ P_2 \\ P_3 \end{bmatrix} = \begin{bmatrix} 147 \\ 556 \\ 250 \end{bmatrix} \quad \text{MW}$$

with a cost of \$7,728/MWhr, which is higher than the unconstrained cost of generation of \$7,698/MWhr. As more constraints are added to the system, the system is moved away from the optimal operating point, increasing the cost of generation. ■

6.4.1.2 Limitations on Dependent Variables

In many cases, the physical limitations of the system are imposed upon states that are dependent variables in the system description. In this case, the inequality equations are functions of x and must be added to the cost function. Examples of limitations on dependent variables include maximum line flows or bus voltage levels. In these cases, the value of the states cannot be independently set, but must be enforced indirectly. One method of enforcing an inequality constraint is to introduce a *penalty function* into the cost function. A penalty function is a function that is small when the state is far away from its limit, but becomes increasingly larger the closer the state is to its limit. Typical penalty functions include

$$p(h) = e^{kh} \quad k > 0 \tag{6.143}$$

$$p(h) = x^{2n}e^{kh} \quad n, k > 0 \tag{6.144}$$

$$p(h) = ax^{2n}e^{kh} + be^{kh} \quad n, k, a, b > 0 \tag{6.145}$$

and the cost function becomes

$$C^* : \quad C(u, x) + \lambda^T g(u, x) + p\left(h(u, x) - h^{max}\right) \tag{6.146}$$

This cost equation is then minimized in the usual fashion by setting the appropriate derivatives to zero. This method has the advantage of simplicity of implementation, but also has several disadvantages. The first disadvantage

is that the choice of penalty function is often a heuristic choice and can vary by application. A second disadvantage is that this method cannot enforce *hard* limitations on states, i.e., the cost function becomes large if the maximum is exceeded, but the state is allowed to exceed its maximum. In many applications this is not a serious disadvantage. If the power flow on a transmission line slightly exceeds its maximum, it is reasonable to assume that the power system will continue to operate, at least for a finite length of time. If, however, the physical limit is the height above ground for an airplane, then even a slightly negative altitude will have dire consequences. Thus penalty functions to enforce limits must be used with caution and is not applicable for all systems.

Example 6.15
Repeat Example 6.13, except use penalty functions to limit the power flow across line 2-3 to 0.1 per unit.

Solution 6.15 The power flow across line 2-3 in Example 6.13 is given by

$$\begin{aligned} P_{23} &= V_2 V_3 Y_{23} \cos(\theta_2 - \theta_3 - \phi_{23}) - V_2^2 Y_{23} \cos\phi_{23} \\ &= 0.1211 \text{ per unit} \end{aligned} \tag{6.147}$$

If P_{23} exceeds 0.1 per unit, then the penalty function

$$p(h) = (P_{23} - 0.1)^2 \tag{6.148}$$

will be appended to the cost function. The partial derivatives remain the same, with the exception of $\left[\frac{\partial f}{\partial x}\right]$, which becomes:

$$\left[\frac{\partial f}{\partial x}\right] = \left[\frac{\partial f}{\partial P_1}\right]\left[\frac{\partial P_1}{\partial x}\right] + \left[\frac{\partial f}{\partial P_{23}}\right]\left[\frac{\partial P_{23}}{\partial x}\right] \tag{6.149}$$

$$\begin{aligned} &= (1 + 0.125P_1)\begin{bmatrix} V_1 V_2 Y_{12} \sin(\theta_1 - \theta_2 - \phi_{1,2}) \\ V_1 V_3 Y_{13} \sin(\theta_1 - \theta_3 - \phi_{1,3}) \end{bmatrix} \\ &\quad +2(P_{23} - 0.1)\begin{bmatrix} -V_2 V_3 Y_{23} \sin(\theta_2 - \theta_3 - \phi_{23}) \\ V_2 V_3 Y_{23} \sin(\theta_2 - \theta_3 - \phi_{23}) \end{bmatrix} \end{aligned} \tag{6.150}$$

Proceeding with the steepest gradient algorithm iterations yields the final constrained optimal generation scheduling:

$$\begin{bmatrix} P_1 \\ P_2 \\ P_3 \end{bmatrix} = \begin{bmatrix} 141.3 \\ 491.9 \\ 319.7 \end{bmatrix} \quad \text{MW}$$

and $P_{23} = 101.2$ MW. The cost for this constrained scheduling is \$7,781/MWhr, which is slightly greater than the nonconstrained cost. Also note that P_{23} is slightly greater than 100 MW since this was considered a soft limit. ■

In the case where hard limits must be imposed, an alternate approach to enforcing the inequality constraints must be employed. In this approach, the

inequality constraints are added as additional equality constraints with the inequality set equal to the limit (upper or lower) that is violated. This in essence introduces an additional set of Lagrangian multipliers. This is often referred to as the dual variable approach, because each inequality has the potential of resulting in two equalities: one for the upper limit and one for the lower limit. However, the upper and lower limits cannot be simultaneously violated; thus, out of the possible set of additional Lagrangian multipliers, only one of the two will be included at any given operating point and thus the dual limits are mutually exclusive.

Example 6.16
Repeat Example 6.15 using the dual variable approach.

Solution 6.16 Introducing the additional equation

$$P_{23} = V_2 V_3 Y_{23} \cos(\theta_2 - \theta_3 - \phi_{23}) - V_2^2 Y_{23} \cos\phi_{23} = 0.100 \text{ per unit} \quad (6.151)$$

to the equality constraints adds an additional equation to the set of $g(x)$. Therefore an additional unknown must be added to the state vector x to yield a solvable set of Equations (three equations in three unknowns). Either P_{G2} or P_{G3} can be chosen as the additional unknown. In this example, P_{G3} will be chosen. The new system Jacobian becomes

$$\left[\frac{\partial g}{\partial x}\right] = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \frac{\partial g_1}{\partial x_3} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \frac{\partial g_2}{\partial x_3} \\ \frac{\partial g_3}{\partial x_1} & \frac{\partial g_3}{\partial x_2} & \frac{\partial g_3}{\partial x_3} \end{bmatrix} \quad (6.152)$$

where

$$\frac{\partial g_1}{\partial x_1} = V_2 \left(V_1 Y_{12} \sin(\theta_2 - \theta_1 - \phi_{21}) + V_3 Y_{13} \sin(\theta_2 - \theta_3 - \phi_{23})\right)$$
$$\frac{\partial g_1}{\partial x_2} = -V_2 V_3 Y_{32} \sin(\theta_2 - \theta_3 - \phi_{23})$$
$$\frac{\partial g_1}{\partial x_3} = 0$$
$$\frac{\partial g_2}{\partial x_1} = -V_3 V_2 Y_{23} \sin(\theta_3 - \theta_2 - \phi_{32})$$
$$\frac{\partial g_2}{\partial x_2} = V_3 V_1 Y_{13} \sin(\theta_3 - \theta_1 - \phi_{31}) + V_2 Y_{23} \sin(\theta_3 - \theta_2 - \phi_{32})$$
$$\frac{\partial g_2}{\partial x_3} = 1$$
$$\frac{\partial g_3}{\partial x_1} = -V_2 V_3 Y_{23} \sin(\theta_2 - \theta_3 - \phi_{23})$$
$$\frac{\partial g_3}{\partial x_2} = V_2 V_3 Y_{23} \sin(\theta_2 - \theta_3 - \phi_{23})$$

$$\frac{\partial g_3}{\partial x_3} = 0$$

and

$$\left[\frac{\partial g}{\partial u}\right] = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}^T, \qquad \left[\frac{\partial f}{\partial u}\right] = [1 + 0.025 P_{G2}]$$

Similar to Example 6.13, the chain rule is used to obtain $\left[\frac{\partial f}{\partial x}\right]$:

$$\left[\frac{\partial f}{\partial x}\right] = \left[\frac{\partial C}{\partial P_{G1}}\right]\left[\frac{\partial P_{G1}}{\partial x}\right] + \left[\frac{\partial C}{\partial P_{G3}}\right]\left[\frac{\partial P_{G3}}{\partial x}\right] \tag{6.153}$$

$$= (1 + 0.125 P_{G1}) \begin{bmatrix} V_1 V_2 Y_{12} \sin(\theta_1 - \theta_2 - \phi_{12}) \\ V_1 V_3 Y_{13} \sin(\theta_1 - \theta_3 - \phi_{13}) \\ 0 \end{bmatrix} + (1 + 0.050 P_{G3}) \times$$

$$\begin{bmatrix} V_3 V_2 Y_{32} \sin(\theta_3 - \theta_2 - \phi_{32}) \\ -V_3 \left(V_1 Y_{13} \sin(\theta_3 - \theta_1 - \phi_{31}) + V_2 Y_{23} \sin(\theta_3 - \theta_2 - \phi_{32})\right) \\ 0 \end{bmatrix} \tag{6.154}$$

Substituting these partial derivatives into the expression for ∇C of Equation (6.122) yields

$$\begin{bmatrix} P_1 \\ P_2 \\ P_3 \end{bmatrix} = \begin{bmatrix} 141.3 \\ 490.3 \\ 321.2 \end{bmatrix} \quad \text{MW}$$

and $P_{23} = 100.0$ MW. ■

6.4.2 State Estimation

In power system state estimation, the estimated variables are the voltage magnitudes and the voltage phase angles at the system buses. The inputs to the state estimator are the active and reactive powers of the system, measured either at the injection sites or on the transmission lines. The state estimator is designed to give the best estimates of the voltages and phase angles, minimizing the effects of the measurement errors. Another consideration for the state estimator is to determine if a sufficient number of measurements are available to fully estimate the power system. This is the notion of system observability.

A set of specified measurements of a power system is said to be *observable* if the entire state vector of bus voltage magnitude and phase angles can be estimated from the set of available measurements. An unobservable system is one in which the set of measurements does not span the entire state space. The power system is observable if the matrix H_x in Equation (6.47) has rank n (full rank), where the number of measurements m is greater than or equal to the number of system states n. A *redundant* measurement is one whose addition to the measurement does not increase the rank of the matrix H_x.

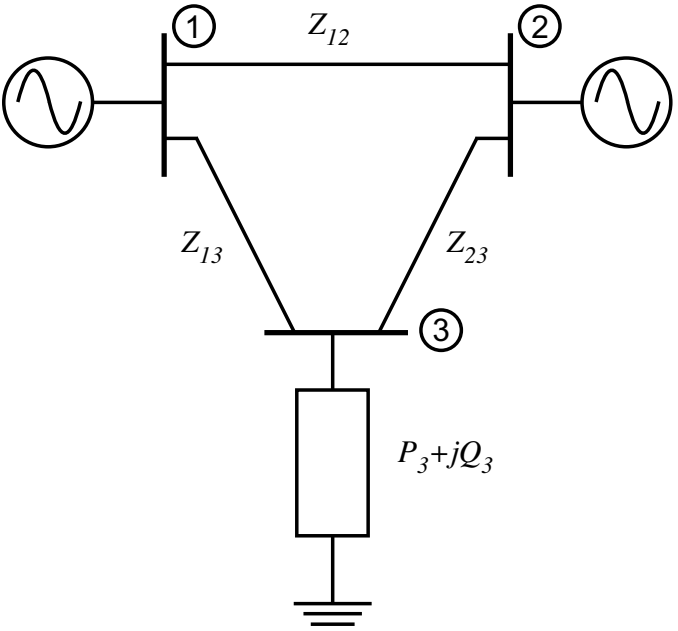

FIGURE 6.7
Example power system

The observability of a power system can be determined by examining the measurement set and the structure of the power system. A *tree* is a set of measurements (either bus or line) that spans the entire set of power system buses. In other words, by graphically connecting the buses and lines that contribute to the set of measurements, the entire set of system buses can be connected by a single connected graph. A power system can be made observable by adding measurements at those lines which will connect disjoint trees.

Example 6.17
The SCADA system for the power network shown in Figure 6.7 reports the following measurements and variances:

z_i	state	measurement	variance (σ^2)
1	V_3	0.975	0.010
2	P_{13}	0.668	0.050
3	Q_{21}	–0.082	0.075
4	P_3	–1.181	0.050
5	Q_2	–0.086	0.075

This system has the same parameters as the three-bus system of Example 3.11. Note that P_3 and Q_3 are *injected* powers and are therefore negative. Estimate the power system states and use the chi-square test of inequality with $\alpha = 0.01$ to check for the presence of bad data in the measurements.

Solution 6.17 The first step in the estimation process is to identify and enumerate the unknown states. In this example, the unknowns are $[x_1\ x_2\ x_3]^T = [\theta_2\ \theta_3\ V_3]^T$. After the states are identified, the next step in the estimation process is to identify the appropriate functions $h(x)$ that correspond to each of the measurements. The nonlinear function that is being driven to zero to minimize the weighted error is

$$F(x) = H_x^T R^{-1} [z - h(x)] = 0 \tag{6.155}$$

where the set of $z - h(x)$ is given by

$$\begin{aligned}
z_1 - h_1(x) &= V_3 - x_3 \\
z_2 - h_2(x) &= P_{13} - \left(V_1 x_3 Y_{13} \cos(-x_2 - \phi_{13}) - V_1^2 Y_{13} \cos\phi_{13}\right) \\
z_3 - h_3(x) &= Q_{21} - \left(V_2 V_1 Y_{21} \sin(x_1 - \phi_{21}) + V_2^2 Y_{21} \sin\phi_{21}\right) \\
z_4 - h_4(x) &= P_3 - (x_3 V_1 Y_{31} \cos(x_2 - \phi_{31}) + x_3 V_2 Y_{32} \cos(x_2 - x_1 - \phi_{32}) \\
&\quad + x_3^2 Y_{33} \cos\phi_{33}) \\
z_5 - h_5(x) &= Q_2 - \left(V_2 V_1 Y_{21} \sin(x_1 - \phi_{21}) - V_2^2 Y_{22} \sin\phi_{22}\right. \\
&\quad \left. + V_2 x_3 Y_{23} \sin(x_1 - x_2 - \phi_{23})\right)
\end{aligned}$$

and the matrix of partial derivatives for the set of functions (6.155) is $H_x =:$

$$\begin{bmatrix}
0 & 0 & 1 \\
0 & V_1 x_3 Y_{13} \sin(-x_2 - \phi_{13}) & V_1 Y_{13} \cos(-x_2 - \phi_{13}) \\
V_1 V_2 Y_{21} \cos(x_1 - \phi_{21}) & 0 & 0 \\
x_3 V_2 Y_{32} \sin(x_2 - x_1 - \phi_{32}) & -x_3 V_1 Y_{31} \sin(x_2 - \phi_{31}) & V_1 Y_{31} \cos(x_2 - \phi_{31}) \\
 & -x_3 V_2 Y_{32} \sin(x_2 - x_1 - \phi_{32}) & +V_2 Y_{32} \cos(x_2 - x_1 - \phi_{32}) \\
 & & +2 x_3 Y_{33} \cos\phi_{33} \\
V_1 V_2 Y_{21} \cos(x_1 - \phi_{21}) & -V_2 x_3 Y_{23} \cos(x_1 - x_2 - \phi_{23}) & V_2 Y_{23} \sin(x_1 - x_2 - \phi_{23}) \\
+V_2 x_3 Y_{23} \cos(x_1 - x_2 - \phi_{23}) & &
\end{bmatrix} \tag{6.156}$$

This matrix has rank 3; therefore, this set of measurements spans the observable space of the power system.

The covariance matrix of the measurements is

$$R^{-1} = \begin{bmatrix}
\frac{1}{0.010} & & & & \\
 & \frac{1}{0.050} & & & \\
 & & \frac{1}{0.075} & & \\
 & & & \frac{1}{0.050} & \\
 & & & & \frac{1}{0.075}
\end{bmatrix} \tag{6.157}$$

The Newton–Raphson iteration to solve for the set of states x that minimize the weighted errors is

$$\left[H_x^T\left(x^k\right) R^{-1} H_x\left(x^k\right)\right]\left[x^{k-1} - x^k\right] = H_x^T\left(x^k\right) R^{-1}\left[z - h(x^k)\right] \tag{6.158}$$

Iteration 1

The initial condition for the state estimation solution is the same flat start as for the power flow equations; namely, all angles are set to zero and all unknown voltage magnitudes are set to unity. The measurement functions $h(x)$ evaluated at the initial conditions are

$$h(x^0) = \begin{bmatrix} 1.0000 \\ 0.0202 \\ -0.0664 \\ -0.0198 \\ -0.1914 \end{bmatrix}$$

The matrix of partials evaluated at the initial condition yields

$$H_x^0 = \begin{bmatrix} 0 & 0 & 1.0000 \\ 0 & -10.0990 & -1.0099 \\ -0.2257 & 0 & 0 \\ -9.9010 & 20.0000 & 1.9604 \\ -1.2158 & 0.9901 & -9.9010 \end{bmatrix}$$

The nonlinear functions (6.155) are

$$F(x^0) = \begin{bmatrix} 228.2791 \\ -593.9313 \\ -75.0229 \end{bmatrix}$$

The updated states are

$$\begin{bmatrix} \theta_2^1 \\ \theta_3^1 \\ V_3^1 \end{bmatrix} = \begin{bmatrix} -0.0119 \\ -0.0624 \\ 0.9839 \end{bmatrix}$$

where θ_2 and θ_3 are in radians. The error at iteration 1 is

$$\varepsilon^1 = 593.9313$$

Iteration 2

The updated values are used to recalculate the Newton–Raphson iterations:

$$h(x^1) = \begin{bmatrix} 0.9839 \\ 0.6582 \\ -0.0634 \\ -1.1593 \\ -0.0658 \end{bmatrix}$$

The matrix of partials is

$$H_x^1 = \begin{bmatrix} 0 & 0 & 1.0000 \\ 0 & -9.9791 & -0.3780 \\ -0.2659 & 0 & 0 \\ -9.6800 & 19.5351 & 0.7700 \\ -0.7470 & 0.4811 & -9.9384 \end{bmatrix}$$

The nonlinear function evaluated at the updated values yields

$$F(x^1) = \begin{bmatrix} 4.4629 \\ -10.5560 \\ 1.3749 \end{bmatrix}$$

The updated states are

$$\begin{bmatrix} \theta_2^2 \\ \theta_3^2 \\ V_3^2 \end{bmatrix} = \begin{bmatrix} -0.0113 \\ -0.0633 \\ 0.9851 \end{bmatrix}$$

The error at iteration 2 is

$$\varepsilon^2 = 10.5560$$

The iterations are obviously converging. At convergence, the states that minimize the weighted measurement errors are

$$x = \begin{bmatrix} -0.0113 \\ -0.0633 \\ 0.9851 \end{bmatrix}$$

To check for the presence of bad data, the weighted sum of squares of the measurement errors is compared to the chi-square distribution for $k = 2$ and $\alpha = 0.01$. The weighted sum of squares is

$$\begin{aligned} f &= \sum_{i=1}^{5} \frac{1}{\sigma_i^2} \left(z_i - h_i(x)\right)^2 \\ &= \frac{(-0.0101)^2}{0.010} + \frac{(0.0012)^2}{0.050} + \frac{(-0.0184)^2}{0.075} + \frac{(0.0007)^2}{0.050} + \frac{(-0.0076)^2}{0.075} \\ &= 0.0155 \end{aligned}$$

This value is less than $\chi_{2,0.01} = 9.21$; therefore, the data set is good and does not contain any spurious measurements. ■

6.5 Problems

1. The fuel costs for a three-unit plant are given by

$$\begin{aligned} F_1 &: 173.61 + 8.670P_1 + 0.00230P_1^2 \quad \$/\text{MWhr} \\ F_2 &: 180.68 + 9.039P_2 + 0.00238P_2^2 \quad \$/\text{MWhr} \\ F_3 &: 182.62 + 9.190P_3 + 0.00235P_3^2 \quad \$/\text{MWhr} \end{aligned}$$

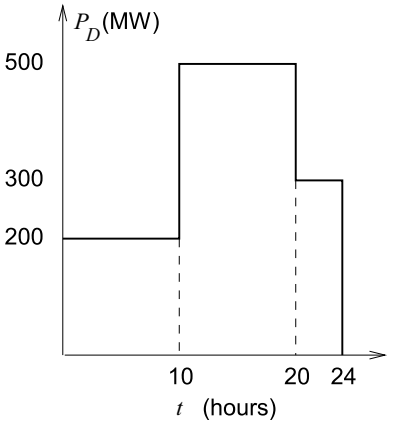

FIGURE 6.8
Load curve for Problem 1

The daily load curve for the plant is given in Figure 6.8. Obtain and sketch the optimal power generated by each unit and the plant's incremental cost of power delivered (λ).

2. Use the method of least squares to find the "best fit" coefficients c_0 and c_1 in the function

$$f(x) = c_0 + c_1 x$$

for the following measured data:

x	$f(x)$
1	–2.1
3	–0.9
4	–0.6
6	0.6
7	0.9

3. Use the method of least squares to find the "best fit" coefficients a_0, a_1, and a_2 in the function

$$f(t) = a_0 + a_1 \sin\frac{2\pi t}{12} + a_2 \cos\frac{2\pi t}{12}$$

for the following measured data:

t	$f(t)$
0	1.0
2	1.6
4	1.4
6	0.6
8	0.2
10	0.8

This function describes the movement of the tide with a 12-hour period.

4. Minimize $-7x_1 - 3x_2 + x_3$ subject to the constraints

$$\begin{aligned} x_1 + x_2 + x_3 &\leq 15 \\ 2x_1 - 3x_2 + x_3 &\leq 10 \\ x_1 - 5x_2 - x_3 &\leq 0 \\ x_1, x_2, x_3 &\geq 0 \end{aligned}$$

   (a) using the simplex method

   (b) using the primal affine method with $\alpha = 0.9$.

   Both methods should be augmented with slack variables x_4, x_5, x_6 and initiated with the feasible vector $x^0 = [1\ 1\ 1\ 12\ 10\ 5]^T$.

5. In bulk transmission systems operation, it is useful to determine the total simultaneous interchange capacity (SIC) available via interties. This SIC may be limited by the transmission system or the available generation. The SIC may be found using the simplex method.

   Missouri State Transmission (MST) has interchanges with three other utilities. The interchanges have the following generation limits:

$$\begin{aligned} &\text{Interchange A: } x_A \leq 1800\text{MW} \\ &\text{Interchange B: } x_B \leq 1500\text{MW} \\ &\text{Interchange C: } x_C \leq 2000\text{MW} \end{aligned}$$

   Each interchange consists of several intertie lines which may limit the SIC. They can be described by their distribution factors:

$$\begin{bmatrix} 2.50 & 1.00 & 0.50 \\ 1.00 & 1.50 & 1.00 \\ 2.00 & 1.25 & 1.00 \end{bmatrix} \begin{bmatrix} x_A \\ x_B \\ x_C \end{bmatrix} = \Delta P$$

   where the maximum incremental load ΔP is 5000 MW.

The simplex method to solve this SIC problem is to maximize $c(x) = x_A + x_B + x_C$, subject to

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2.50 & 1.00 & 0.50 \\ 1.00 & 1.50 & 1.00 \\ 2.00 & 1.25 & 1.00 \end{bmatrix} \begin{bmatrix} x_A \\ x_B \\ x_C \end{bmatrix} \leq \begin{bmatrix} 1800 \\ 1500 \\ 2000 \\ 5000 \\ 5000 \\ 5000 \end{bmatrix}$$

Find the optimal SIC. Is the system generation or transmission limited? Support your answer.

6. Using the steepest descent method, minimize $x_1^2 + x_2^2$ subject to the constraint
$$x_1^2 + 2x_1x_2 + 3x_2^2 - 1 = 0$$
Start with x_1=0.5 and $x_2 = 0.5$.

7. Repeat Problem 6 using the SQP method. Start with $\lambda = 0$.

8. Find the minimum of
$$C: \quad x_1^2 + x_2^2 + u_1x_1 + u_2x_2 + 1$$
(a) subject to
$$\begin{aligned} x_1\cos(x_2) + x_2^2 - u_1\cos(x_1) &= 1 \\ x_1 - x_2 + 3u_2 &= -3 \end{aligned}$$
(b) subject to
$$\begin{aligned} x_1\cos(x_2) + x_2^2 - u_1\cos(x_1) &= 1 \\ x_1 - x_2 + 3u_2 &= -3 \\ u_2 &\geq -0.8 \end{aligned}$$
(c) subject to
$$\begin{aligned} x_1\cos(x_2) + x_2^2 - u_1\cos(x_1) &= 1 \\ x_1 - x_2 + 3u_2 &= -3 \\ x_2 &\leq 0.30 \end{aligned}$$
using the penalty function $f(x_2) = ae^{b(x_2-c)}$, where a and b are positive constants and c is the function offset.

Use an initial guess of $x^0 = [0\ 0]'$ and $u^0 = [0\ 0]'$ and $\gamma = 0.05$. You might want to experiment with other values of γ as well. Your stopping criterion should be $\|\nabla C\| \leq 0.01$.

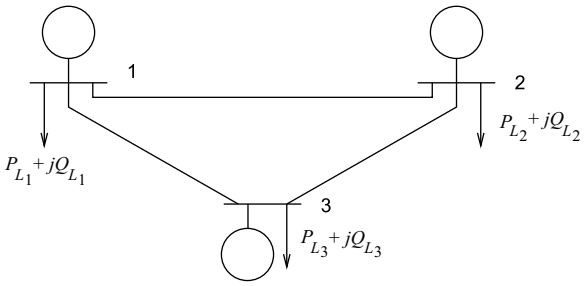

FIGURE 6.9
Three-bus system

9. Consider the system shown in Figure 6.9. The bus and line data are given below:

Line	R	X	B
1-2	0.01	0.1	0.050
1-3	0.05	0.1	0.025
2-3	0.05	0.1	0.025
Bus	$\|V\|$	P_L	Q_L
1	1.00	0.35	0.10
2	1.02	0.40	0.25
3	1.02	0.25	0.10

The fuel costs for the generators are

$$F_1 : P_{G_1} + 1.5P_{G_1}^2$$
$$F_2 : 2P_{G_2} + P_{G_2}^2$$
$$F_3 : 2.5P_{G_3} + 0.5P_{G_3}^2$$

(a) Using the equal incremental cost criterion, find the optimal scheduling for the units (remember that this method neglects system losses).

(b) Using your answer for part (a) as the initial control vector, use the steepest descent method to find the optimal scheduling for this system, which considers system losses.

(c) Now assume the following limits are imposed:

$$F_1 : P_{G_1} + 1.5P_{G_1}^2 \quad 0 \le P_{G_1} \le 0.6$$
$$F_2 : 2P_{G_2} + P_{G_2}^2 \quad 0 \le P_{G_2} \le 0.4$$
$$F_3 : 2.5P_{G_3} + 0.5P_{G_3}^2 \quad 0 \le P_{G_3} \le 0.1$$

Repeat part (b).

(d) Interpret your results relating the generator settings to the cost functions.

10. For the system shown in Figure 6.9, the following measurements were obtained:

V_2	1.04
V_3	0.98
P_{G1}	0.58
P_{G2}	0.30
P_{G3}	0.14
P_{12}	0.12
P_{32}	–0.04
P_{13}	0.10

where $\sigma_V^2 = (0.01)^2, \sigma_{P_G}^2 = (0.015)^2$, and $\sigma_{P_{ij}}^2 = (0.02)^2$.

Estimate the system states, the error, and test for bad data using the chi-square test with $\alpha = 0.01$.