



Munich Personal RePEc Archive

An electricity transmission primer for energy economists

Benjamin, Richard M

Round Table Group

25 June 2012

Online at <https://mpra.ub.uni-muenchen.de/39666/>
MPRA Paper No. 39666, posted 25 Jun 2012 23:41 UTC

An Electricity Transmission Primer for Energy Economists
Parts 1 and 2

By: Richard M. Benjamin, Ph.D.
Round Table Group

Part 1: Basic EE and PTDFs

1. Introduction

This paper provides a technical primer for the energy economist wishing to study the optimal power flow problem, but, who like this paper's author, find it difficult to simply plow through Schweppe et al. (1988). This paper's aim is to provide the reader with sufficient technical background sufficient to calculate power transfer distribution factors (PTDF's, or network shift factors) for a more complicated model than the simple 3-node model common to energy economics studies.¹ The point of departure for this work is basic electrical engineering fundamentals. We start off Section 2 with a simple direct current (DC) circuit analysis, demonstrating Ohm's law, and Kirchhoff's laws. As these are captured by very basic equations, we quickly move on to alternating current (AC) analysis. Here, we spend quite a bit of effort manipulating fairly straightforward (i.e. sinusoidal) AC voltage equations to derive equations for average voltage, current, and power in an AC circuit. From there, we delve into the concept of "phase angles," showing how electrical components such as inductors and capacitors bring voltage and current "out of phase." We then demonstrate both graphically and algebraically the effect of non-zero phase angles on resistance and power relations in an AC circuit. The goal of section 2 is to familiarize the reader with concepts that economists and operations researchers generally take for granted, but are quite foreign to those of us who are self-taught.

Getting this background under our belts, we then move on to DC network analysis. We take this intermediate step to demonstrate circuit concepts to the reader in the simple setting of a DC network. Having introduced the reader to basic circuit analysis, we then introduce the reader to the linearized version of the AC network model in section 3. In this section we demonstrate a methodology for linearizing the real-power AC network equations (the latter of which we derive in Appendix A). After linearizing the AC network equations, we demonstrate how one calculates the PTDFs for the linearized version of the AC network. Sections 2 and 3 of this paper should

¹ See, e.g., Bushnell and Stoft (1997) for a detailed analysis of a three-node model.

thus provide the reader with sufficient technical background to understand basic power equations, as presented in energy economics papers, along with deriving PTFDs.

In Part II, we present the AC power flow model and a powerful method for solving the AC power flow problem, the Newton-Rhapson Method.

2. Electrical Engineering Fundamentals

We begin our power flow analysis with the examination of a DC series circuit.² In a basic series circuit, a source of electromagnetic force (emf) is connected to one or more sources of resistance (e.g. lights), which connect back to the emf source, as shown below, with a battery as the emf source:

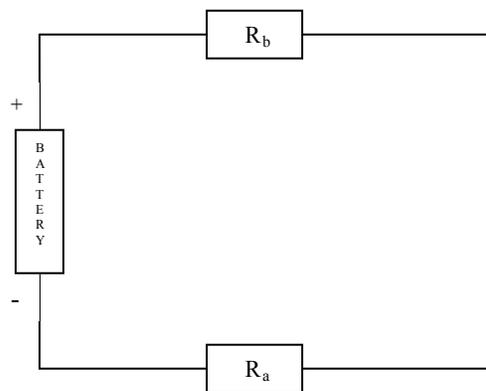


Figure 1.1: Basic DC Circuit

In a DC circuit, current flows in one direction only: from negative to positive.³ Thus, in the circuit above, current would flow in a counterclockwise direction. The mathematics of this circuit are governed by three laws: Ohm's Law, Kirchhoff's Current Law (KCL), and Kirchhoff's Voltage Law (KVL). The three laws are as follows:

Ohm's law: current, resistance and voltage in a circuit are related as follows:

$$V = I \times R \tag{1.1}$$

² Other DC circuits include parallel circuits, where current flow splits at junctures into parallel paths.

³ As opposed to AC, where current alternates direction several times per second.

where

V = voltage, or emf, measured in volts,

I = current, measured in amperes, and

R = resistance, measured in Ohms.

Applied to our analysis, Ohm's law tells us that the voltage "drop" across resistors a and b is equal to the resistor's resistance value time the current flowing through the resistor.

Kirchhoff's Voltage Law (KVL): The directed sum of the electrical potential differences (voltage) around any closed circuit is zero.

Kirchhoff's Voltage law tells us that the sum of the voltage dropped across the two resistors is equal to the emf produced by the battery.

Kirchhoff's Current Law (KCL): At any node in an electrical circuit, the sum of currents flowing into that node is equal to the sum of currents flowing out of that node.

Kirchhoff's current law is a conservation law. For a circuit with only one path, KCL implies that current is constant throughout the circuit.

Now that we know these three laws, given sufficient values for the circuit, we may solve for the remaining unknown values. Let us assume then, that we have a 9-volt battery, that the value of the first resistor, R_a , is 3 ohms (Ω), and that an ammeter yields a measurement of 2 amps, and we must find the remaining values.

The easiest, most organized answer is to fill in these values on a table, as shown below:

Table 1.1: V, I, R Calculations

	R_a	R_b	Total
V	6	3	9
I	2	2	2
R	3	1.5	4.5

(numbers in bold are given)

By KCL, current at all points in the circuit is equal to 2 amps.⁴ Since we are given $R_a = 3\Omega$, Ohm's law allows us to solve for the voltage drop across R_a as 6 volts. That is, running 2 amps through a 3 Ohm resistor produces 6 volts of force. Next, KVL tells us that the 9 volts produced by the battery are dissipated across the two resistors. Since the first resistor dissipated 6 volts of force, the second one must dissipate 3 volts, or $V_b = 3$. Further application of Ohm's law reveals that the second resistor is a $3/2 = 1.5\Omega$ resistor. Ohm's law also applies to the circuit as a whole. So, if a 9-volt battery produces 2 amps of current, it must be the case that total resistance in the circuit is $9/2 = 4.5\Omega$. This reveals another law of series circuits. Resistance, as well as voltage, is additive.⁵

While batteries are a DC voltage source, most electrical devices run on AC. An AC generator, or alternator, produces alternating current as a magnetic field rotates around a set of stationary wire coils. When the magnet starts its rotation, at the reference angle of zero degrees, it is completely out of alignment with the coil. It moves into closer alignment until it reaches an angle of 90 degrees, and moves further out of alignment again. Because the electromagnetic force, or voltage created in this operation varies directly with degree of alignment between the magnet and the coil, we may (for starters) express voltage mathematically as:

$$v(t) = V_{\max} \sin(\omega t) \tag{1.2a}$$

where

V_{\max} = maximum instantaneous voltage, or voltage amplitude,

ω = angular frequency, in radians per second,⁶ and

t = time, in seconds.

⁴ When we look at multiple lines across which current flows (known as parallel circuits), this relation will no longer hold (but KCL still will).

⁵ This rule does not hold for a parallel circuit (although it does hold for parallel components in series). See *ref.*

⁶ For those of us who have long since forgotten, there are 2π radians in a circle, so a radian equals approximately 57.3° . Thus we may also write $v(t) = V_{\max} \sin(2\pi ft)$, where f is frequency, in rotations per second.

Multiplying ω (radians per second) times t , number of seconds gives us the angle, in radians, covered by the magnet in t seconds. Thus, we may write eq. (1) as either $v(t) = V_{\max} \sin(\omega t)$, or

$$v(t) = V_{\max} \sin(\theta) \quad (1.2b)$$

where θ is the familiar symbol for measuring angles. *e.g.*: An alternator with an amplitude of 100 volts, produced by a magnet turning one radian per second, has an instantaneous voltage of $v(1) = 100 \sin(1) = 84.15$ volts at one second.

From here, let us note that voltage is of importance in the study of optimal power flow because it is a determining factor of the power a generator produces. As Ohm's law reminds us, $V = I \times R$. Ohm's law is useful in manipulating the formula for power, as shown below:

$$P = I \times V^7 \quad (1.3)$$

where

P = Power, measured in Watts.

From Ohm's law, we can substitute for either V or I , giving us two alternative expressions for power:

$$P = I^2 R, \text{ and } P = \frac{V^2}{R}.$$

As resistance is a many-headed beast, for the time being we will focus on the derivation of power given by eq. (3).

Intuitively, though, since the most basic resistance source, the resistor, has a fixed resistance, Ohm's law tells us that in a circuit with a fixed resistance source, voltage and current will vary only by a constant of proportionality, the fixed resistance.⁸ Therefore, since the formula for voltage is given by eq. (1), the formula for current will be given by:

$$i(t) = I_{\max} \sin(\omega t) \quad (1.4)$$

⁷ We get the much more familiar expression, $P = I \times E$, by denoting voltage as E rather than V .

⁸ When we get to alternative resistance sources (inductors and capacitors) we will see that this is no longer true.

where

I_{max} = Maximum current, or current amplitude.

Combining equations (1.1) and (1.4), then, we arrive at the formula for AC power:

$$p(t) = V_{max} \sin(\omega t) \cdot I_{max} \sin(\omega t) \quad (1.5)$$

Using trigonometric identities, we may write $\sin^2(x)$ as $\frac{1}{2}(1 - \cos 2x)$, thus we rewrite eq. (5) as:

$$P(t) = \frac{1}{2}V_{max}I_{max}(1 - \cos 2\omega t), \text{ or } \frac{1}{2}V_{max}I_{max} - \frac{1}{2}V_{max}I_{max}(\cos 2\omega t) \quad (1.6)$$

Of course, when examining a generator's power output, we are not concerned with its instantaneous value, which varies continuously, but rather its average value. Looking at the second expression in eq. (1.6), we see that instantaneous power has both a fixed, $\frac{1}{2}V_{max}I_{max}$, and a variable component, $-\frac{1}{2}V_{max}I_{max}(\cos 2\omega t)$. Since, like that of $\sin(x)$, the average value of $\cos(x)$ is zero over one full cycle, the average value of power reduces to the first component

$$P_{avg} = \frac{1}{2}V_{max}I_{max} \quad (1.7)$$

Alternatively, we may calculate the average value of power directly from the average values of voltage and current, which we do below. By inspection, we will not get very far using either eqs. (1.1) or (1.4) directly, because we run into the same phenomenon of zero average value for a sinusoidal signal. This is unimportant, however, as the sign of the voltage or current in question is altogether arbitrary.⁹ Intuitively, one will feel the same shock whether a voltmeter registers -50V or +50V (don't try this at home). We get around the sign problem for voltage and current by squaring the instantaneous values of each and integrating to arrive at their respective averages. Starting with eq. (1.2b), $v(t) = V_{max} \sin(\theta)$, we are reminded that the magnet travels 2π

⁹ In a direct current (DC) circuit, for example, a voltmeter may register a positive or negative value, depending on the reference point. See, *e.g.* Dale, p. 5-29.

radians per complete rotation. Therefore, the average value for the voltage-squared over one complete cycle is

$$\frac{V_{\max}^2}{2\pi} \int_0^{2\pi} \sin^2 \theta d\theta = \frac{V_{\max}^2}{2\pi} \int_0^{2\pi} \frac{1 - \cos 2\theta}{2} d\theta \quad (1.8)$$

Since the integral of $\cos(2x)$ is $\frac{1}{2}\sin(2x)$, the average value of voltage-squared over a complete

cycle is $\frac{V_{\max}^2}{2}$.¹⁰ Taking the square root of this value, we find that the average value for voltage

is

$$V_{avg} = |V| = \frac{V_{\max}}{\sqrt{2}}. \quad (1.9)$$

Going through the same steps, we find that the average value for current is

$$I_{avg} = |I| = \frac{I_{\max}}{\sqrt{2}}. \quad (1.10)$$

Multiplying eqs. (1.9) and (1.10), we again find that P_{avg} is given by equation (1.7).

If electrical systems contained only voltage sources, wires, and resistance sources (e.g., resistors and load), the world would be a happy place for the energy economist, with current and voltage always and everywhere in phase. Unfortunately (for us) though, electrical devices contain capacitors and inductors, casting us out of the Eden of synchronous voltage and current.¹²

Let us begin by demonstrating how inductors bring voltage and current out of phase (creating a phase angle). Inductors (typically an inductor is a conducting wire shaped as a coil) oppose changes in current, as expressed by the formula:

$$V = L \frac{dI}{dt} = L \cdot I_{\max} \cos(\omega t) \quad (1.11)$$

¹⁰ See, e.g. Mittle and Mittal (2006), Section 6.4.1 – 6.4.2.

¹¹ The “| |” notation is unfortunate, since most of us will immediately think, “absolute value,” not “average.” Electrical engineering has no set notation, however.

¹² Inductors and capacitors, of course, do serve useful purposes, such as storing electrical energy and filtering out specific signal frequencies, but this is strictly beside the point.

where:

L = inductance, measured in Henrys.

eq. (1.11) states that the voltage drop across an inductor is equal to the inductor's inductance times the rate of change of current over time. To see how this creates a phase shift between current and voltage, we must refer to the sinusoidal nature of current and voltage.¹³

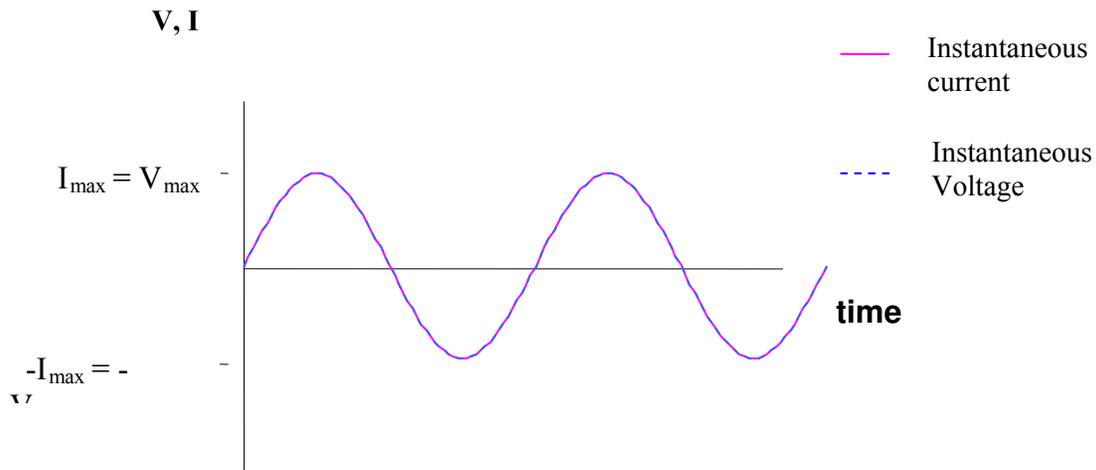


Figure 1.2: Instantaneous AC Current and Voltage

Since $\frac{dI}{dt} = 0$ at a maximum or a minimum, from eq. (1.11) instantaneous voltage, V , is equal to zero when $I = (I_{\max}, -I_{\max})$. Thus, voltage and current are 90° out of phase for an inductor, as shown below:

¹³ AC current and voltage will be the same only when resistance = 1Ω , but nothing is gained in this example by incorporating a different value for resistance.

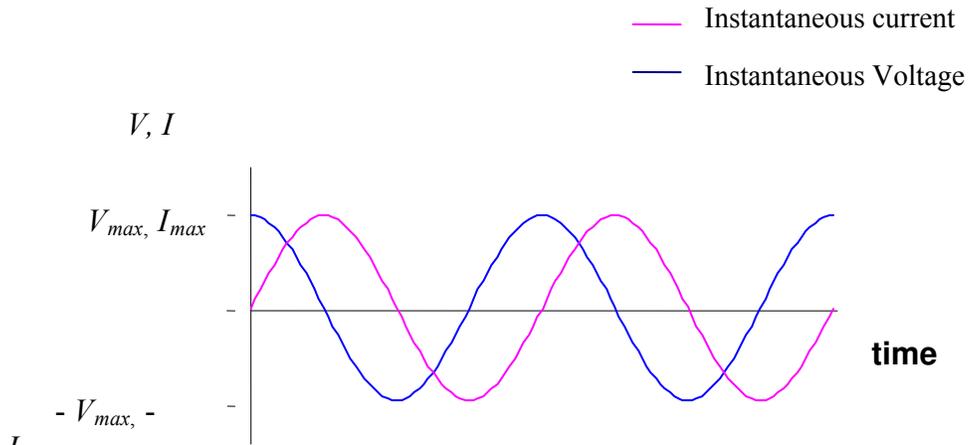


Figure 1.3: Voltage and Current for an Inductor

(Note that we arrive at the same conclusion by noting that instantaneous voltage is at its maximum or minimum synchronously with di/dt . di/dt reaches these values at its points of inflection, which occur at $i(t) = 0$).

Observe that while we may still write $i(t) = I_{max} \sin(\theta)$, it is no longer the case that

$v(t) = V_{max} \sin(\theta)$. Taking advantage of the fact that $\cos(\theta) = \sin(\theta + 90^\circ)$, we may now write

$$i(t) = I_{max} \sin(\theta), \quad (1.12a)$$

$$v(t) = V_{max} \sin(\theta + 90^\circ). \quad (1.12b)$$

In this case we say that the voltage leads the current by 90° , or, equivalently, the current lags the voltage by 90° . More generally, we may write

$$v(t) = V_{max} \sin(\theta + \delta), \quad (1.12c)$$

where δ is commonly known as the *phase angle* between voltage and current.

Like inductors, capacitors also introduce phase angles between current and voltage. A capacitor contains two conductors separated by an insulator. Current flow through a capacitor is directly related to the derivative of voltage with respect to time, as follows:

$$i(t) = C \frac{dV}{dt} \quad (1.13)$$

where

C = capacitance, measured in Farads.

A capacitor's capacitance will depend on the materials composing the insulator. From eq. (1.13), we see that instantaneous current will be zero when voltage is at its maximum and minimum values and will reach its maximum and minimum values at the inflection points for voltage.

Thus, graphically, the relationship between voltage and current in a capacitor is as shown below:

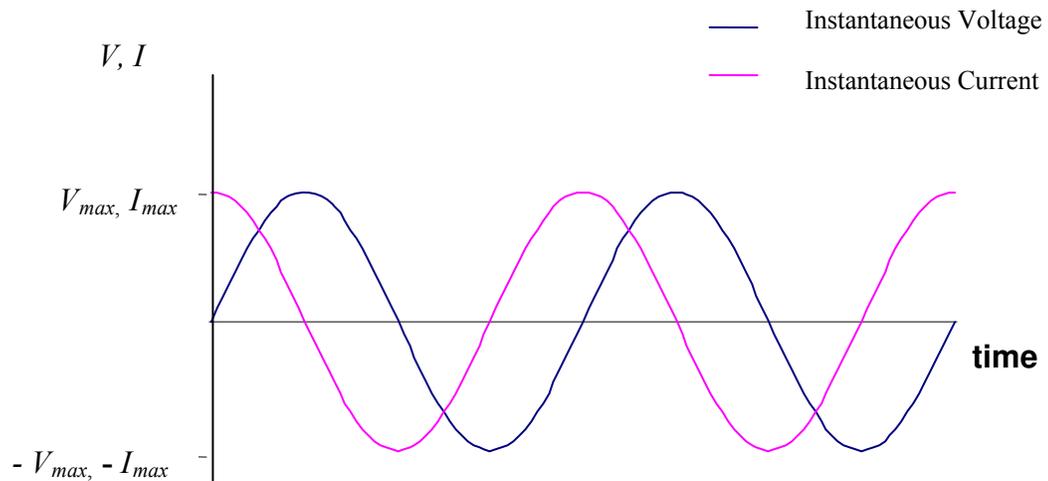


Figure 1.4: Voltage and Current for a Capacitor

In a capacitor, current lags voltage by 90° , and thus we may write

$$v(t) = V_{\max} \sin(\theta - 90^\circ), \text{ or } v(t) = V_{\max} \sin(\theta - \delta). \quad (1.14)$$

Note the difference between a resistor versus an inductor or a capacitor. A resistor simply opposes the flow of current, like sand opposes the flow of water. An inductor or a capacitor, however, does not oppose a constant flow of current. Inductors oppose changes in *current flow*, while capacitors oppose changes in *voltage*, both at angles of $\pm 90^\circ$. The name for

the opposition imposed by inductors and capacitors is reactance, denoted by the symbol X and measured in Ohms.¹⁴ For an inductor, we denote reactance by a vector 90° out of phase with current, causing a 90° phase angle between voltage and current.¹⁵

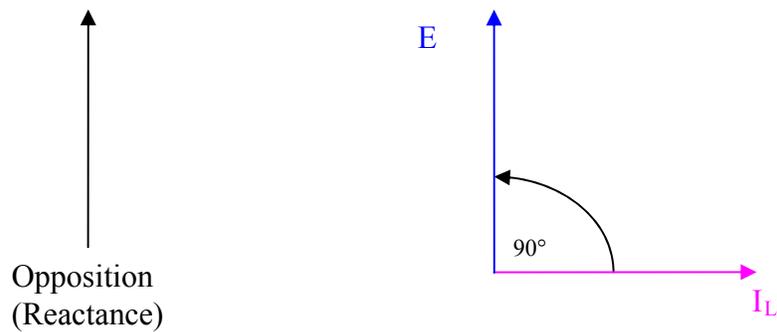


Figure 1.5: Geometric representation of reactance

In general, though, electrical circuits will not simply contain inductors or capacitors or resistors, but some combination of the three. Therefore, real-world electricity systems will generally be characterized by phase angles of other than 0° or 90° . To determine the phase angle for such systems, we add the vectors for reactance and resistance, arriving at a quantity called impedance, denoted by Z and measured in ohms.

Example: Let us examine a circuit with a 4 ohm resistor and a 3 ohm inductor, shown below:

¹⁴ A good distinction between resistance and reactance is found at http://www.allaboutcircuits.com/vol_2/chpt_5/1.html, which describes resistance as essentially friction against the motion of electrons, and reactance as inertia against the motion of electrons.

¹⁵ For a capacitor, reactance forms an angle of -90° to current, so the direction of the reactance vector is reversed.

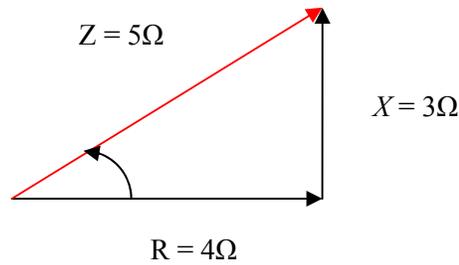


Figure 1.6: Vector representation of impedance

Impedance is simply the sum of the resistance and reactance vectors. A circuit with a 4Ω resistor and a 3Ω inductor will thus have total impedance equal to 5Ω . While we correctly calculated impedance as 5Ω , it is more common to denote impedance by not only its magnitude, but also its direction as well. We can do this by referring to either polar or rectangular notation.

Polar notation uses the vector compass to express the variable of interest's quantity in terms of magnitude and direction (that is, phase angle). In this case, we add the vectors of resistance (4Ω at 0°) and reactance (3Ω at 90°), arriving at

$$Z = \left(\sqrt{4^2 + 3^2} \right) \text{phase angle} = \cos^{-1}(4/5) = 36.87^\circ.$$

Thus, we would express this example's impedance in polar notation as

$$Z = 5 \angle 36.87^\circ. \tag{1.15}$$

This tells us that the combined effect of the resistor and the inductor is to produce an impedance of 5Ω at 36.87° . Notice that no information is lost in switching to polar notation. We may “go the other way” to calculate the resistive and reactive components of impedance as follows:

$$\text{Resistance} = 5 \cdot \cos(36.87^\circ) = 4\Omega \tag{1.16}$$

$$\text{Reactance} = 5 \cdot \sin(36.87^\circ) = 3\Omega \tag{1.17}$$

This yields a second way to express impedance: in terms of resistance plus reactance. We do this using rectangular notation. To switch from polar to rectangular notation graphically, we switch

from the vector compass to representing the horizontal axis as real units, and the vertical axis as imaginary units. In doing so, we place resistance (to real power) on the horizontal axis and reactance on the vertical, as shown below:

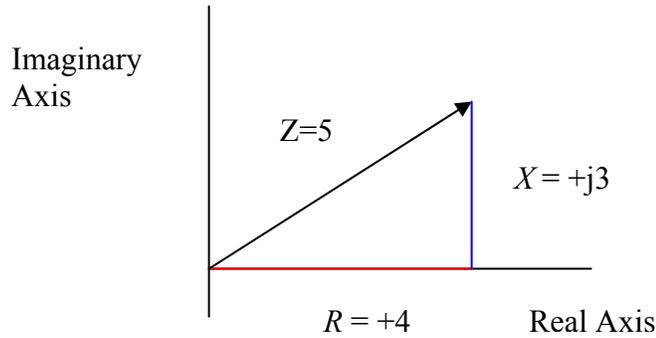


Figure 1.7: Impedance in Rectangular Form

Thus, we write $Z = 4 + j3$ in rectangular form.¹⁶

Having introduced impedance, we may compute power in a mixed resistive/reactive circuit. Let us then draw a circuit with a 4-ohm resistor (R) and a 3-ohm inductor (L) (alternatively, reactor):

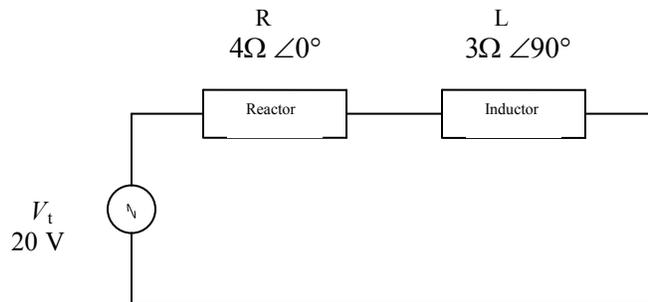


Figure 1.8: AC Circuit with Resistance and Reactance

¹⁶ In electrical engineering, $\sqrt{-1}$ is commonly denoted as j , instead of i . This change in axes accounts for the (initially) confusing switch in the formulas for voltage and current from $v(t) = V_{\max} \sin(\theta)$ to $v(t) = V_{\max} \cos(\theta)$. The real quantity is now the adjacent, not the opposite.

The symbol, \surd , represents an AC voltage source. As is standard, we will assign a phase angle of 0° to the voltage source, with the phase angle and magnitude of impedance given by eq. (1.15). Given total voltage and impedance, we apply Ohm's law, as applied to AC circuits, $V = IZ$, to calculate current.¹⁷

$$I = \frac{V}{Z} = \frac{20 \text{ V} \angle 0^\circ}{5 \Omega \angle 36.87^\circ} = 4 \text{ A} \angle -36.87^\circ \quad (1.18)$$

From here, calculating the remaining voltages is straightforward (answers in the table below).

Table 1.2: V, I, Z Calculations

	R	L	Total
V	$16 \angle -36.87^\circ$	$12 \angle +53.13^\circ$	$20 \angle 0^\circ$
I	$4 \angle -36.87^\circ$	$4 \angle -36.87^\circ$	$4 \angle -36.87^\circ$
Z	$4 \angle 0^\circ$	$3 \angle 90^\circ$	$5 \angle 36.87^\circ$

Finally, we move on to AC power calculation. Real power is the rate at which energy is expended. According to Grainer and Stevenson (1994), reactive, or imaginary power, “expresses the flow of energy alternately toward the load and away from the load.”¹⁸ Intuitively, we can divide a power generator's output into power capable of doing work (real power), and power not capable of doing work (reactive power). Reactive power is the result of current moving out of phase with voltage. The greater the phase angle between voltage and current, the less efficient is power output in terms of capability to do work, and the greater is reactive power. Apparent, or complex, power is the (geometric) sum of real and reactive power. Thus, we derive the “power triangle” in the same manner as we derived the relationship between resistance, reactance, and impedance:

¹⁷ Since there is only one pathway for current, the amount (amperage) of current is constant across the entire circuit. The angle, -36.87 degrees, is the amount by which current lags voltage (*i.e.*, the phase angle).

¹⁸ p. 8. The exact nature of reactive power is not well understood.

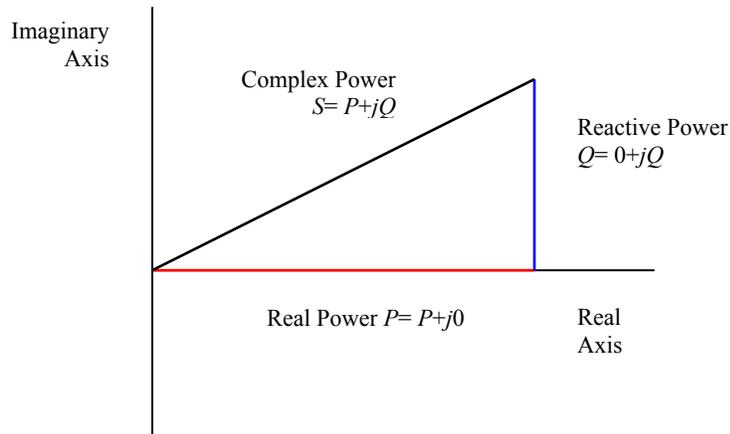


Figure 1.9: Graph of Complex Power

Calculating complex power in rectangular notation is straightforward: Simply take real power plus reactive power equals complex power, or complex power, $S = P + jQ$. Calculating complex power in polar form is a little more complicated. Taking the phasors for voltage and current as $V = |V| \angle \alpha$ and $I = |I| \angle \beta$,¹⁹ the calculation for complex power is

$$VI^* = |V||I| \angle \alpha - \beta \quad (1.19)$$

where * denotes the complex conjugate (as will be demonstrated shortly). Note that the angle, $(\alpha - \beta)$ is once again the phase angle between voltage and current, as may be verified by designating voltage as the reference phasor.

To see why we take the complex conjugate of current (*i.e.*, we switch the sign on the angle for current), let us turn to eqs. (1A.5) and (1A.6) for reactive and complex power:

$$Q = |I|^2 \times X \quad (1.20)$$

$$S = |I|^2 \times Z \quad (1.21)$$

Dividing (1.20) by (1.21), we see that

¹⁹ Note that in this example we have not chosen either quantity as reference.

$$\frac{Q}{S} = \frac{X}{Z} \quad (1.22)$$

That is, the cosines of the phase angles for impedance and for the power triangle are equal, and thus the phase angles themselves are equal. Since the phase angle for impedance is $(\alpha - \beta)$, the phase angle for the power triangle is this quantity as well. But we have to switch the sign on the current phasor to maintain this result.

Euler's identity: $e^{j\theta} = \cos(\theta) + j\sin(\theta)$ offers further insight into both rectangular and polar representation of electrical quantities, such as impedance and power. Since resistance is equal to $|Z|\cos(\theta)$, and reactance is equal to $|Z|\sin(\theta)$, if we denote the horizontal axis as the real number line, and the vertical axis as the imaginary number line, we may express impedance as

$$Z = |Z|\cos(\theta) + j|Z|\sin(\theta) = |Z|e^{j\theta} . \quad (1.23)$$

and complex power as

$$S = |I|^2 |Z| e^{j\theta} \quad (1.24)$$

3. DC Network Calculations

a. Nodal Voltages and Current Flows

Now that we have demonstrated the basics of voltage, current, resistance (and admittance), and power, we may examine demonstrate the calculation of these values in circuit analysis. We start with a DC circuit, then move on to its AC counterpart. First we specify the network itself. We will examine an elementary circuit, called a ladder circuit, as shown below.²⁰

²⁰ Taken from Baldick (2006), p. 163 *et. seq.*

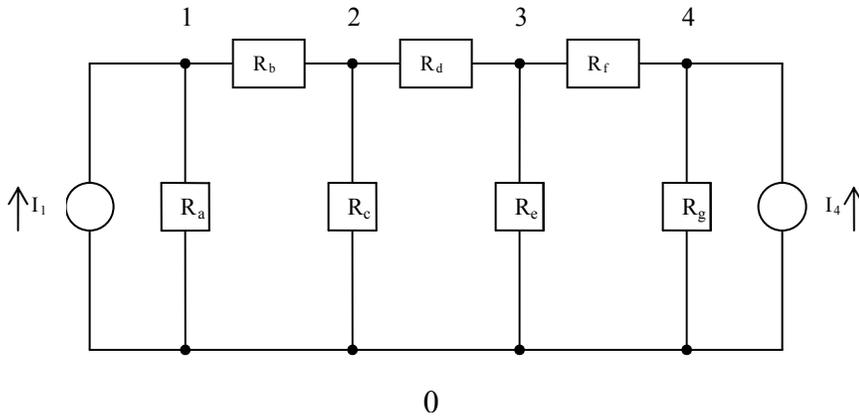


Figure 1.10: A five-node ladder network

A network is composed of transmission lines, or branches. The junctions formed when two or more transmission lines (more generally, circuit elements) are connected with each other are called nodes (e.g., in the figure above, branch (1,2) connects node 1 with node 2). In the above example, then, there are seven transmission lines [(1,2), (2,3), (3,4), (0,1), (0,2), (0,3), and (0,4)] connecting five nodes [0, 1, 2, 3, 4, 5].

Denote:

N = Number of nodes in a transmission system

k = Number of lines in a transmission system

R_{ij} = Resistance of the transmission line connecting nodes i and j ; alternatively

R_l = Resistance of transmission line l .

$Y_l = 1/R_l =$ Admittance of transmission line l ($= Y_{ij}$).

I_j = Current source, located at node j .

I_{ij} = Current flowing over line ij .

V_j = Node j voltage

V_{ij} = voltage differential between nodes i and j (also known as the “voltage drop” across line ij)

Note that all of the above network’s $N = 5$ nodes are interconnected, directly or indirectly, by transmission lines and are subject to Kirchoff’s laws.²¹ Since the (directional) voltage drops across the system must sum to zero, satisfaction of KVL takes away one degree of freedom in the system. Thus we need write voltage equations for only $N-1$ nodes to fully identify an N – node

²¹ This usually holds for small networks, but it is not generally true.

system. KVL also allows us to choose one node as the datum node (or ground) node, and set voltage at this node equal to zero. The researcher is free to choose any node as datum. To minimize computational cost, one generally chooses as datum the node with the most transmission lines connected to it. It is customary to denote the datum node as node 0.

There are two current sources in this circuit, I_1 and I_4 . The current sources are generators located at nodes 1 and 4. These generators are connected to the transmission grid by lines (0,1) and (0,4), or, alternatively, lines a and g . These lines are generally denoted “limited interconnection facilities.”²² In the electrical engineering literature, such lines are known as “shunt elements.”²³ The ladder network above might represent, say, a transmission line stretching across New York State, where node 1 represents Buffalo, node 2 represents Rochester, node 3 represents Syracuse, and Node 4 represents Albany. Node 0 then, actually represents four physically distinct locations, where power plants at or around these four cities are connected to the transmission grid by limited transmission interconnection facilities. Again, in standard electrical engineering parlance, node 0 is the ground “node.”²⁴

Note that one may specify the resistance associated with a particular transmission lines by naming the nodes the line connects, or simply assigning the line its own, alternative subscript.²⁵ Resistance in transmission lines is the opposition to current as current “bumps into” the material composing the transmission lines. Good conductors, or materials that offer relatively little resistance, are ideal candidates for transmission lines. Metals with free electrons, like copper and aluminum, make good conductors, as they provide little resistance.

²² Generally, limited transmission facilities’ sole purpose is to interconnect power plants to the grid. Such facilities are not required to provide open access to customers wishing to transmit power over these facilities. However, the distinction between a limited interconnection facility and a transmission line which must provide open access (that is, a facility which must submit an Open Access Transmission Tariff to the FERC) can become blurry when the “limited interconnection facility” is, say, 40+ miles long. This was the case with the Sagebrush line (see, e.g. the case filings before the Federal Energy Regulatory Commission in docket numbers ER09-666-000 and its progeny (e.g. ER09-666-001). Two relevant orders in these cases are found in 127 FERC ¶ 61,243 (2009) and 130 FERC ¶ 61,093 (2010).

²³ See, e.g., Glover *et al.* (2012), section 4.11, and Grainger and Stevenson (1994), chapter 6.

²⁴ While the electrical engineer may think it is ok to use any node as ground, it makes a lot more sense to use the “ground” and the ground node.

²⁵ We take the latter tack to save space in Figure 9.

Before we analyze the equations corresponding to Kirchhoff's Laws, let us introduce the concept of admittance. Admittance measures the ease with which electrons flow through a circuit's elements, and is the inverse of impedance. Since there are no reactive components in our example, though, impedance and resistance are equivalent. We thus loosen our terminology and let admittance denote the inverse of resistance.²⁶ Labeling admittance by Y , we thus have $Y = \frac{1}{R}$.

This allows us to write Ohm's Law as

$$I = \frac{1}{R}V = Y * V. \quad (1.25)$$

Ohm's Law allows us to write current flow along a particular line as a product of the voltage drop between the two nodes the line connects and the admittance of that line. Ohm's Law thus yields the flow of current across the $k = 7$ transmission lines in the example. Thus, we write the current flow along line (1,0), I_{10} , as

$$I_{10} = V_{10} \cdot Y_{10} = (V_1 - V_0)Y_{10} = V_1 \cdot Y_{10} \quad (1.26a)$$

(since $V_0 = 0$)

We write current flow across the other 6 lines analogously:

$$I_{12} = (V_1 - V_2)Y_{12} \quad (1.26b)$$

$$I_{20} = V_2 \cdot Y_{20} \quad (1.26c)$$

$$I_{23} = (V_2 - V_3)Y_{23} \quad (1.26d)$$

$$I_{30} = V_3 \cdot Y_{30} \quad (1.26e)$$

$$I_{34} = (V_3 - V_4)Y_{34} \quad (1.26f)$$

$$I_{40} = V_4 \cdot Y_{40} \quad (1.26g)$$

KCL expresses the conservation of current at the $N = 4$ nodes. That is, KCL states that the sum of the currents entering a node equals the sum of currents exiting that node. We will use

²⁶ The inverse of resistance is actually conductance.

the convention that only current sources (generators) produce current entering a node, while all transmission lines carry current away from the node.

Let us begin with node 1. Current I_1 enters node 1, while transmission lines (1,0) and (1,2) carry current “away” from node 1. Thus KCL implies that $I_1 = \text{current (1,0)} + \text{current (1,2)}$, or $I_1 = I_{10} + I_{12}$. Note that I_{ij} denotes flow of current from node i to node j (when the actual current flow is in the opposite direction, from node j to node i , we write $I_{ij} < 0$. Otherwise $I_{ij} \geq 0$). The important point is that we specify the assumed direction of current flow by the ordering of the subscripts.

Because we do not incorporate load in this example,²⁷ node 1 current flow will be nonnegative, or $I_1 \geq 0$ ($I_1 = 0$ if the current source is not currently operational, as when a power plant is shut down for maintenance). I_{10} and I_{12} may be either positive or negative. Substituting eqs. (1.26a) and (1.26b) for I_{10} and I_{12} , respectively, yields:

$$I_1 = V_1 \cdot Y_{10} + (V_1 - V_2)Y_{12} \quad (1.27a)$$

Collecting the voltage terms yields:

$$I_1 = (Y_{01} + Y_{12})V_1 - Y_{12}V_2 \quad (1.27b)$$

Moving to node 2, notice that there is no current source here, simply three transmission lines which, by convention, carry current away from node 2. Our three current equations for lines (2,1), (2,0), and (2,3) are as follows:

$$Y_{21}(V_2 - V_1) = I_{21} \quad (1.28)$$

$$Y_2 V_2 = I_{20} \quad (1.29)$$

$$Y_{23}(V_2 - V_3) = I_{23} \quad (1.30)$$

²⁷ If there were a load located at node 1, then current flow from node 1 would be negative (*i.e.*, net current would flow to node 1, rather than away from it) whenever node 1 load were greater than node 1 generation.

Let us remark at this point that $Y_{ij} = Y_{ji}$, *i.e.*, admittance is not a directional value.²⁸ KCL,

applied to node 2 yields $I_{21} + I_{20} + I_{23} = 0$,²⁹ or

$$-Y_{21}V_1 + (Y_{21} + Y_{20} + Y_{23})V_2 - Y_{23}V_3 = 0 \quad (1.31)$$

The alert reader will notice that node 3 is similar to node 2, in that there is no current source and three branches emanating there. Thus KCL yields:

$$-Y_{32}V_2 + (Y_{32} + Y_{30} + Y_{34})V_3 - Y_{43}V_4 = 0 \quad (1.32)$$

Finally, our treatment of node 4 is symmetric to node 1, yielding:

$$-Y_{43}V_3 + (Y_{43} + Y_{40})V_4 = I_4 \quad (1.33)$$

Note that eqs. (1.27b), (1.31) – (1.33) form a system of four equations in four unknowns, which we express in matrix form below (after switching to single subscript notation for line admittances, for expositional ease):

$$\begin{bmatrix} Y_a + Y_b & -Y_b & 0 & 0 \\ Y_b & Y_b + Y_c + Y_d & -Y_d & 0 \\ 0 & -Y_d & Y_d + Y_e + Y_f & -Y_f \\ 0 & 0 & -Y_f & Y_f + Y_g \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \end{bmatrix} = \begin{bmatrix} I_1 \\ 0 \\ 0 \\ I_4 \end{bmatrix} \quad (1.34)$$

The LHS matrix is known as the admittance matrix. The (square) admittance matrix is usually denoted as $A \in R^{N \times N}$, with generic element A_{ij} . The vectors $V \in R^N$ and $I \in R^N$ are the vectors of unknown voltages and known current injections at each of the network's nodes, respectively.

Given values for admittances and current injections, we may solve for nodal voltages and current flows over each line. As the simplest example, consider resistances of one unit for each resistor in the circuit and current injections of one unit at both nodes 1 and 4. In this case,

$$Y_i = \frac{1}{R_i} = 1 \quad \forall i = a, b, \dots, g$$

²⁸ This is a likely explanation for the single-subscript notation for resistance and admittance common in *EE* texts.

²⁹ Note that this equation does not imply that no current flows through node 2. It states that current traveling toward node two, which takes on a negative sign, equals the amount of current flowing out of node 2 (which has a positive sign), and thus the principle of conservation of current.

The system to be solved is then given in matrix form below:

$$\begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 3 & -1 & 0 \\ 0 & -1 & 3 & 0 \\ 0 & 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad (1.36)$$

Yielding $V^{*T} = [2/3 \ 1/3 \ 1/3 \ 2/3]^{30}$. (1.37)

We may now solve for current flows across all lines in the network, using Eqs. (1.26a) – (1.26g).

$$I_{10} = Y_a V_1 = 2/3 \quad (1.38a)$$

$$I_{12} = Y_b (V_1 - V_3) = 1/3 \quad (1.38b)$$

$$I_{20} = Y_c V_2 = 1/3 \quad (1.38c)$$

$$I_{23} = Y_d (V_2 - V_3) = 0 \quad (1.38d)$$

$$I_{30} = Y_e V_3 = 1/3 \quad (1.38e)$$

$$I_{34} = Y_f (V_3 - V_4) = -1/3 \quad (1.38f)$$

$$I_{40} = Y_g V_4 = 2/3 \quad (1.38g)$$

(all currents in amps).

The negative sign on I_{34} means that current is flowing in the opposite direction than we assumed (*i.e.*, current is flowing from node 4 to node 3, not from node 3 to node 4). One may easily check that KCL is indeed satisfied at the four nodes (remembering that $I_{ji} = -I_{ij}$, $j \neq i$).

b. A Quick DC Sensitivity

Sensitivity analysis studies how the value of a problem's solution, or some function of that solution, changes as we tweak the value(s) of either the solution (vector) or some other model parameter. As a quick introduction to sensitivity analysis, we examine the sensitivity of the

³⁰ Where we denote the transpose of a vector/matrix as T .

vector of voltages in the ladder circuit problem when we change the current injected at a single node. We do this analysis to demonstrate a basic method for calculating sensitivities.

To calculate a sensitivity of a solution to a change in an input into an equation, we start with a square system of equations $Ax = b$ (for our example, $A \in \mathfrak{R}^{4 \times 4}$, $x \in \mathfrak{R}^{4 \times 1}$, and $b \in \mathfrak{R}^{4 \times 1}$).³¹ Since we are examining the sensitivity of the solution of this system of equations, we assume such a solution exists, and denote it as x^* , or the “base case” solution. We will denote the amount by which we wish to change a variable as χ . At the base case, then, none of the problem inputs have changed at all. Therefore, the base case corresponds to $\chi = 0$. A non-zero value for χ , then, will refer to a “change case.” In the change case, we will have changed one or more inputs to the original problem. For example, we may change an admittance from Y_i to $Y_i + \chi$, or change a current source from I_i to $I_i + \chi$. Let us denote the change-case equation, then, as

$$A(\chi)x^*(\chi) = b(\chi). \quad (1.39)$$

That is to say, the coefficient matrix and right hand side vector are now dependent on, or functions of, χ . With the change in any of the inputs to the problem, the solution, x^* , will change, depending on χ as well. We will calculate a sensitivity for x , then, as the partial derivative of the original solution vector (*i.e.*, the base case) to a change in a specific variable, which we denote as χ_j . First, since the base case corresponds to no change in any variable (*i.e.*, $\chi = \mathbf{0}$), we will denote the base case equation as

$$A(\mathbf{0})x^* = b(\mathbf{0}). \quad (1.40)$$

To calculate the sensitivity of the original solution to a change in a variable, we simply solve for the derivative of the base-case solution with respect to the variable in question. To solve for this derivative, we start by totally differentiating eq. (1.39).

³¹ We assume the reader is familiar with matrix algebra, so we do not go into details such as conditions for matrix invertibility.

$$\frac{\partial A(\chi)}{\partial \chi_j} x^*(\chi) d\chi + A(\chi) \frac{\partial x^*(\chi)}{\partial \chi_j} d\chi = \frac{\partial b(\chi)}{\partial \chi_j} d\chi.$$

Since the $d\chi$ terms all cancel, we solve the remaining equation for $\frac{\partial x^*(\chi)}{\partial \chi}$, obtaining

$$\frac{\partial x^*}{\partial \chi_j}(\chi) = [A(\chi)]^{-1} \left[\frac{\partial b}{\partial \chi_j}(\chi) - \frac{\partial A}{\partial \chi_j}(\chi) x^* \right] \quad (1.41)$$

Finally, note that we are evaluating this partial derivative at the base-case solution, so we must set $\chi = \mathbf{0}$ in Eq. (41), obtaining

$$\frac{\partial x^*}{\partial \chi_j}(\mathbf{0}) = [A(\mathbf{0})]^{-1} \left[\frac{\partial b}{\partial \chi_j}(\mathbf{0}) - \frac{\partial A}{\partial \chi_j}(\mathbf{0}) x^* \right] \quad (1.42)$$

Let us return to the example shown in eqs. (1.26) and (1.34) the ladder network with current injections I_1 and I_4 equal to one amp each, and admittances $Y_a - Y_g$ equal to one unit each. We wish to calculate the sensitivities of the voltages obtained, $\mathbf{x}^{*T} = [2/3 \ 1/3 \ 1/3 \ 2/3]$, to a change in the node 1 current source from I_1 to $I_1 + \chi$. We show this change in eq. (1.43), below:

$$\begin{bmatrix} Y_a + Y_b & -Y_b & 0 & 0 \\ Y_b & Y_b + Y_c + Y_d & -Y_d & 0 \\ 0 & -Y_d & Y_d + Y_e + Y_f & -Y_f \\ 0 & 0 & -Y_f & Y_f + Y_g \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} I_1 + \chi \\ 0 \\ 0 \\ I_4 \end{bmatrix} \quad (1.43)$$

we see that $\frac{\partial A}{\partial \chi}(\cdot) = 0$ and $\frac{\partial b}{\partial \chi}(\cdot)^T = [1 \ 0 \ 0 \ 0]$. Finally, inverting the matrix, $A(\mathbf{0})$ (as found

on the rhs of eq. (1.36)), and multiplying, we obtain

$$\frac{\partial x^*}{\partial \chi}(\mathbf{0}) = [A(\mathbf{0})]^{-1} \left[\frac{\partial b}{\partial \chi}(\mathbf{0}) \right] = \begin{bmatrix} 0.6191 \\ 0.2381 \\ 0.0952 \\ 0.0476 \end{bmatrix} \quad (1.44)$$

4. The DC Approximation to the AC Network: Calculating PTDFs

We started with DC network calculations to familiarize the reader with concepts we will use in our analysis of the AC Network. Electrical engineers, intrepid folks that they are, will actually set up and solve AC power flow problems as systems on non-linear equations.³² Energy economists generally study linearized versions of the AC power equations, though. The obvious advantage of linearizing is that one need not run iterative algorithms to find the critical points for systems of non-linear equations.

This section focuses on only one element of linearized AC analysis, the calculation of PTDFs. Doing so, we will present the reader important concepts such as the Jacobian of power balance equations (as derived in Appendix A), noting the similarity between this matrix and the admittance matrix of the linearized AC system (derived in Appendix B). We believe that the advantage of this approach is that it is incremental. We introduce the reader to these important concepts in the context of the derivation of a single concept, rather than throwing the reader the several derivations necessary to solve the optimal power flow problem all at once.

a. Linearizing AC Power Equations

We derive the equations for power balance and power flow across transmission lines in Appendix A (which the more advanced reader is free to skip). Appendix equations (1A.23) and (1A.24), shown below:

$$P_\ell = \sum_{k=1}^n u_\ell u_k [G_{\ell k} \cos(\theta_\ell - \theta_k) + B_{\ell k} \sin(\theta_\ell - \theta_k)] \quad (1.45)$$

$$Q_\ell = u_\ell u_k [G_{\ell k} \sin(\theta_\ell - \theta_k) - B_{\ell k} \cos(\theta_\ell - \theta_k)] \quad (1.46)$$

³² Well, perhaps they are not that intrepid, since available computer packages solve the problems for them!

are called power flow equality constraints,³³ and must be satisfied at each bus ℓ for the power system to maintain a constant frequency.³⁴

Likewise, we write equations (1A.27) and (1A.28) for power flow across transmission lines as

$$\tilde{P}_{\ell k} = (u_\ell)^2 G_{\ell\ell} + u_\ell u_k [G_{\ell k} \cos(\theta_\ell - \theta_k) + B_{\ell k} \sin(\theta_\ell - \theta_k)] \quad (1.47)$$

$$\tilde{Q}_{\ell k} = -(u_\ell)^2 B_{\ell\ell} + u_\ell u_k [G_{\ell k} \sin(\theta_\ell - \theta_k) - B_{\ell k} \cos(\theta_\ell - \theta_k)]. \quad (1.48)$$

The linearized AC model is generally known as the DC approximation to the AC network. Thus, in the linearized AC model, terms such as reactive power and impedance are simplified (assumed) away. As discussed in Appendix B, we make two simplifying assumptions to transform the complex admittance matrix to its DC approximation: we set its real terms, $G_{\ell k}$, equal to zero. Further, we ignore the shunt elements (*i.e.*, the terms corresponding to node 0, or the ground). This reduces eqs. (1.45) and (1.47) to

$$P_\ell = \sum_{k=2}^n u_\ell u_k [B_{\ell k} \sin(\theta_\ell - \theta_k)]^{35} \quad (1.49)$$

$$\tilde{P}_{\ell k} = u_\ell u_k [B_{\ell k} \sin(\theta_\ell - \theta_k)] \quad (1.50)$$

We complete the process of linearizing these equations by assuming that our base-case solution (the point of reference for our derivatives) involves zero net power flow at nodes $2 - n$,³⁶ and all voltage magnitudes equal to one per unit, so that $u^{(0)} = \mathbf{1}$. This reduces eqs. (1.49) and (1.50) to:

³³ The power flow equality constraints express the relationship that complex power generated at bus ℓ is equal to voltage times (the complex conjugate of) current.

³⁴ We derive these equations in Appendix A by dividing the term for the amount of complex power injected into a generic node, ℓ , as the sum of its real (P_ℓ) and reactive ($Q_\ell \sqrt{-1}$) components.

³⁵ Notice that we sum over only nodes 2 through n in Eq. (49), because we are examining only nodes $1 - n$, and system balance implies that we have only $n - 1$ degrees of freedom.

³⁶ And thus, no net flow from the reference node either, due to the first law of thermodynamics and Kirchoff's laws.

$$P_\ell = \sum_{k=2}^n B_{\ell k} \sin(\theta_\ell - \theta_k) \quad (1.49a)$$

$$\tilde{p}_{\ell k} = B_{\ell k} \sin(\theta_\ell - \theta_k) \quad (1.50a)$$

At this point we deviate from the standard methodology by linearizing these equations about the base-case solution for power. As per Baldick (2006, p. 343), we define a flat start as the state of the system corresponding to phase angles equal to zero at all nodes and all voltage magnitudes equal to one (which, in fact, is the base-case solution to the problem when we ignore the circuit's shunt elements.³⁷ As we will see shortly, the model is now linear at the base-case solution, because $\left. \frac{\partial \sin(\theta)}{\partial \theta} \right|_{\theta=0} = 1$.

b. Calculating PTDFs

Now we are ready to solve for a system's PTDFs. flow with respect to phase angles, and the Jacobian of line flow with respect to phase angles (note that the latter is not a reduced matrix, because we wish to solve for incremental flows for all lines in the system).

The term PTDF, or “shift factor,” means the change in the flow of power across a particular transmission line (ℓk), induced by an incremental increase in power output at a given node. The alert reader will note that we can calculate N matrices of these shift factors, corresponding to the change in power flow across all lines in the system as the result of an increase in power generation at any of the system's N nodes. Mathematically, we will calculate a real-power shift factor as $\frac{\partial \tilde{p}_{\ell k}}{\partial P_j}$. We denote this matrix as $L_j = \frac{\partial \tilde{\mathbf{P}}_{\ell k}}{\partial \mathbf{P}_{(m)j}}$. That is, the matrix of

incremental line flows across all of a system's lines due to an injection of power at node j and

³⁷ The traditional interpretation of DC power flow emphasizes small angle approximation to sine and cosine (i.e. $\cos(\theta_i - \theta_k) \approx 1$, $\sin(\theta_i - \theta_j) \approx (\theta_i - \theta_j)$), and the solution of DC power flow being the same as the solution of an analogous DC circuit with current sources specified by the power injections and voltages specified by the angles (see Schweppe *et al.*, Appendix D). Our presentation differs from the standard. However, once the reader digests our presentation, s/he should be able solve the traditional method as well.

withdrawal of that power at node m , for each of the system's other nodes.³⁸ We cannot take the desired partial derivative directly, because there is no explicit term for P_ℓ in eq. (1.50a).

Therefore, we incorporate the chain rule to solve for the desired term based on a variable common to both eqs. (1.49a) and (1.50a), finding that:

$$\frac{\partial \tilde{p}_{\ell k}}{\partial P_j} = \frac{\partial \tilde{p}_{\ell k}}{\partial \theta_i} \frac{\partial \theta_i}{\partial P_j}. \quad (1.51)$$

We find the second term, of course, by inverting the Jacobian corresponding to $\frac{\partial P_j}{\partial \theta_i}$.

Differentiating these equations is now fairly straightforward. First, we shorthand our equation (1.49a) to express net power flow from node ℓ as a function of phase angles and voltage injections and the vector of net power flows as:

$$P_\ell = p_\ell \left(\begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{u} \end{bmatrix} \right) \quad (1.52a)$$

$$P = p \left(\begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{u} \end{bmatrix} \right), \quad (1.52b)$$

respectively. As alluded to earlier, there are $n-1$ degrees of freedom in the power balance equations, because power balance indicates that one cannot specify the power balance equation at the n^{th} node independently of the power balance in the rest of the system. Therefore, we subtract the row and column of Jacobian for power balance. It is customary to denote this node as the “reference” node.³⁹ It is also customary to denote the reference node as node 0. The only drawback of this approach is that we have already denoted the ground node as node 0, and the ground node and the reference node are *not* the same thing. However, since we have assumed

³⁸ One might denote the fact that the matrix expresses power withdrawal for each of the other system nodes (other than j . Power injected and withdrawn as the same node has $\text{PTDF}_{(j)j} = \mathbf{0}$, trivially) as $-j$, but this notation might lead the reader to believe that power is being withdrawn at all other nodes *simultaneously*, which is not the case.

³⁹ Alternatively, we may denote the reference node as the “base” node. But we have already used the term “base-case,” and the base node and the base case are two different concepts, so we will stick with reference node from this point on.

away the shunt elements, we are not examining the ground node any longer. Thus, there should be no further confusion regarding denoting the reference node as node 0.

We will demonstrate the calculation of two different Jacobians here, for expositional clarity. We denote the power-balance Jacobian for all of the systems nodes (the “full” Jacobian) as $J_{p\theta}$, and the Jacobian for all the system’s nodes, excepting the reference node (the “reduced” Jacobian) as $\hat{J}_{p\theta}$, respectively. The term $p\theta$ indicates that we are taking the derivatives of real power flows with respect to phase angles. When solving for PTDFs, we will use the reduced Jacobian because the power-balance equations have $n - 1$ degrees of freedom. That is, we cannot specify the power-balance equation for the reference node independently of the power balance equations at the system’s $n - 1$ other nodes, because power balance at the reference node is determined residually by Kirchoff’s voltage law (and the first law of thermodynamics). Therefore, we will delete both the row and column corresponding to the reference node when evaluating the reduced Jacobian. Remembering that we are evaluating the linearized power balance equations, we write the reduced Jacobian for our calculation of network shift factors as:

$$\hat{J}_{p\theta}^{(0)} = \hat{J}_{p\theta} \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix} \right) = \frac{\partial \hat{p}}{\partial \theta} \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix} \right).^{40} \quad (1.53)$$

(that is, we evaluate this Jacobian at a flat start).

Now let us examine the Jacobian for the line flow equations: $K^{(0)} \in \Re^{k \times (n-1)}$. This matrix has k rows, one for each of the equations for the system’s k transmission lines (again, ignoring shunt elements). This matrix shows the PTDFs corresponding to power injected at a

⁴⁰ Note that upon examination of Eqs. (45, 46), there are actually four Jacobians of interest in AC power derivation: $\hat{J}_{p\theta}(x)$, $\hat{J}_{pu}(x)$, $\hat{J}_{q\theta}(x)$, $\hat{J}_{qu}(x)$. Where we write “ x ” instead of “(0),” because the methodologies used to solve the AC problem involve iterations of a non-linear solution method. The first iteration will be at the base-case, using $J_{(c)}^{(0)}$, but, depending on the method used, the Jacobian will be revised at either every step of the process, at regular intervals in the process (or not at all). In the first two cases, the Jacobian changes from $J_{(c)}^{(0)}$ to $J_{(c)}^{(v)}(x_v)$, where “ v ” denotes the iteration number (or index) of the process.

specific node. Because the system behaves differently when power is withdrawn at a node ℓ versus a different node, k , there are $n - 1$ shift factors corresponding to each line of the transmission system, one for each of the nodes at which we may withdraw power when power is injected at a specific node (all PTDFs for power injected *and* withdrawn at the same node are identically zero). We denote the Jacobian for the line flow equations, evaluated at the base-case of a flat start, as

$$K^{(0)} = K \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix} \right) = \frac{\partial \tilde{p}_{\ell k}}{\partial \theta} \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix} \right).^{41} \quad (1.54)$$

For expositional simplicity, let us now solve eqs. (1.53) and (1.54) and the resulting shift factors for a sample network before introducing the general notation corresponding to eqs. (1.53) and (1.54).

Consider the 5-node model shown below. The value shown between each node pair is that line's impedance. When calculating a shift factor matrix, we must choose one node to be the reference node (*i.e.*, we will have to decide at which node we withdraw power when calculating a shift-factor matrix. Let us choose node 1 as the reference node, so that we will replace the notation "1" with "0" whenever we refer to the reference node.

⁴¹ In the electrical engineering literature, the Jacobian of Eq. 54 is often denoted $K_{(\ell k)}^{(0)}$. Note that we do not denote this Jacobian as $J_{(\ell k)}^{(0)}$, because this notation denotes the ℓk^{th} term of the Jacobian for power balance.

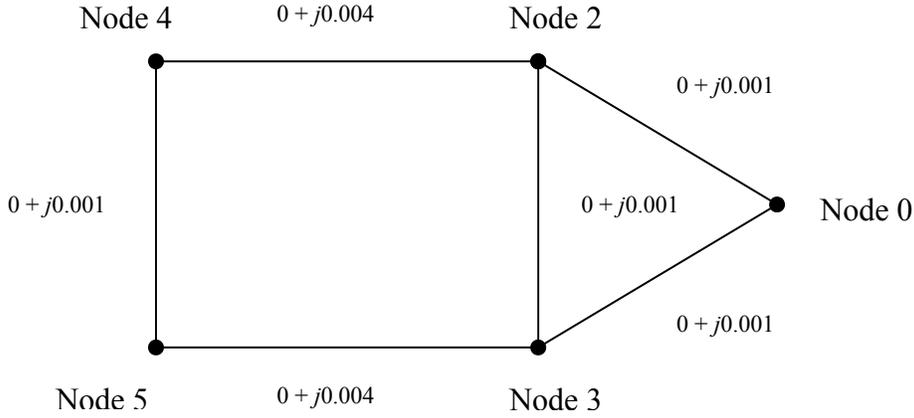


Figure A

Note that each line ℓk has $R_{\ell k} = 0$. That is, the real part of impedance (resistance) is zero for all lines. This corresponds to the assumption that the real component of conductance, the inverse of impedance, is zero as well). The complex component of impedance (reactance) is given by $X_{\ell k} \sqrt{-1}$. We invert each line's impedance to derive its admittance, denoted $Y_{\ell k}$:

$$\begin{aligned}
 Y_{02} = Y_{03} = Y_{23} = Y_{45} &= \frac{1}{(0 + j0.001)} = j1000 \\
 Y_{24} = Y_{35} &= \frac{1}{(0 + j0.004)} = j250
 \end{aligned} \tag{1.54}$$

Note also that $Y_{\ell k} = B_{\ell k} \sqrt{-1}$.

For any node ℓ , we write $\sin(\theta_\ell - \theta_j)$ for all nodes j connected to node ℓ by a transmission line. We may thus express Eqs. (49a) for each of the five nodes in our example as:

$$P_0 = 1,000 \sin(\theta_0 - \theta_2) + 1,000 \sin(\theta_0 - \theta_3) \tag{1.55a}$$

$$P_2 = 1,000 \sin(\theta_2 - \theta_0) + 1,000 \sin(\theta_2 - \theta_3) + 250 \sin(\theta_2 - \theta_4) \tag{1.55b}$$

$$P_3 = 1,000 \sin(\theta_3 - \theta_0) + 1,000 \sin(\theta_3 - \theta_2) + 250 \sin(\theta_3 - \theta_5) \tag{1.55c}$$

$$P_4 = 250 \sin(\theta_4 - \theta_2) + 1,000 \sin(\theta_4 - \theta_5) \tag{1.55d}$$

$$P_5 = 250 \sin(\theta_5 - \theta_3) + 1,000 \sin(\theta_5 - \theta_4) \tag{1.55e}$$

Deriving the Jacobian for this example is straightforward, nonetheless, we show explicitly the terms corresponding to the first row of this matrix:

$$J_{11}^{(0)} = \frac{\partial P_0}{\partial \theta_0} = 1,000 \cos(\theta_0 - \theta_2) + 1,000 \cos(\theta_0 - \theta_3) = 2,000$$

$$J_{12}^{(0)} = \frac{\partial P_0}{\partial \theta_2} = 1,000 \cos(\theta_0 - \theta_2)(-1) = -1,000$$

$$J_{13}^{(0)} = \frac{\partial P_0}{\partial \theta_3} = 1,000 \cos(\theta_0 - \theta_3)(-1) = -1,000$$

Since neither θ_4 nor θ_5 are arguments of P_1 , their corresponding terms are obviously zero.

Taking the remaining partial derivatives yields the matrix:

$$J^{(0)} = \begin{bmatrix} 2,000 & -1,000 & -1,000 & 0 & 0 \\ -1,000 & 2,250 & -1,000 & -250 & 0 \\ -1,000 & -1,000 & 2,250 & 0 & -250 \\ 0 & -250 & 0 & 1,250 & -1,000 \\ 0 & 0 & -250 & -1,000 & 1,250 \end{bmatrix} \quad (1.56)$$

A couple of characteristics of this matrix stand out immediately. First, notice the diagonal terms are equal to the sum of the off-diagonal terms times minus one. Recognizing that the off-diagonal term, $J_{\ell k}^{(0)} = -B_{\ell k}$, this allows us to shorthand the terms of the Jacobian as:

$$\frac{\partial \tilde{p}_\ell}{\partial \theta_k} \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) = \begin{cases} \sum_{j \in \mathfrak{Z}(\ell)} B_{\ell j} & \text{if } k = \ell, \\ -B_{\ell k} & \text{if } k \in \mathfrak{Z}(\ell), \\ 0, & \text{otherwise.} \end{cases} \quad (1.57)$$

Second, notice that the Jacobian is symmetric (this fact makes calculation of the full AC model less computationally expensive).

As mentioned above, we omit the power flow equality constraint and partial derivatives corresponding to the base node in order to derive the reduced Jacobian. Again, we do this because there are $n - 1$ degrees of freedom in the power flow equality constraints, so including the base node would give us a redundant equation. Thus, the “reduced” system of equations is:

$$P_2 = 1,000 \sin(\theta_2 - \theta_0) + 1,000 \sin(\theta_2 - \theta_3) + 250 \sin(\theta_2 - \theta_4) \quad (1.58b)$$

$$P_3 = 1,000 \sin(\theta_3 - \theta_0) + 1,000 \sin(\theta_3 - \theta_2) + 250 \sin(\theta_3 - \theta_5) \quad (1.58c)$$

$$P_4 = 250 \sin(\theta_4 - \theta_2) + 1,000 \sin(\theta_4 - \theta_5) \quad (1.58d)$$

$$P_5 = 250 \sin(\theta_5 - \theta_3) + 1,000 \sin(\theta_5 - \theta_4), \quad (1.58e)$$

and we take the derivatives of these equations with respect to $\theta_2, \theta_3, \theta_4,$ and $\theta_5,$ obtaining.⁴²

$$\hat{J}^{(0)} = \begin{bmatrix} 2,250 & -1,000 & -250 & 0 \\ -1,000 & 2,250 & 0 & -250 \\ -250 & 0 & 1,250 & -1,000 \\ 0 & -250 & -1,000 & 1,250 \end{bmatrix}, \quad (1.59)$$

with inverse

$$[\hat{J}^{(0)}]^{-1} = \begin{bmatrix} 0.000655 & 0.000345 & 0.000517 & 0.000483 \\ 0.000345 & 0.000655 & 0.000483 & 0.000517 \\ 0.000517 & 0.000483 & 0.002724 & 0.002276 \\ 0.000483 & 0.000517 & 0.002276 & 0.002724 \end{bmatrix} \quad (1.60)$$

To calculate the Jacobian of line flows with respect to phase angles, remember that the Jacobian of line flows wrt phase angles is of dimension $k \times (n-1)$. We will take derivatives on flows across the $k = 6$ transmission lines with respect to the $(n-1) = 4$ nodes (other than the reference node) at which power is injected.

Before taking these derivatives, however, we must determine the (reference) direction of power flow across all k lines. That is, we know that the equation for power flow across a line is $\tilde{P}_{\ell k} = B_{\ell k} \sin(\theta_{\ell} - \theta_k)$, where power this equation indicates that power flows from node ℓ to node k . We just have to figure out whether, for any given node pair, if power flows from one node to the other, or vice-versa.

⁴² Note that we do not remove the terms in power flow equality constraints $P_2 - P_5$ corresponding to the base node, because these terms capture the physical characteristics of the network (*i.e.* the transmission lines between the node in question and the base node).

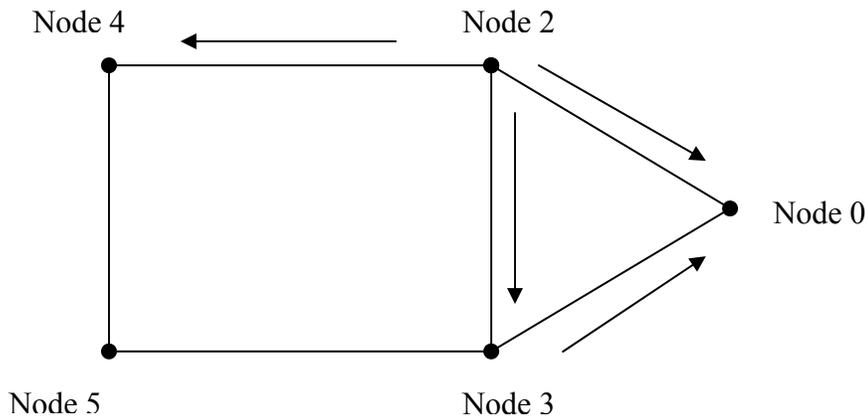
There is no one “right” answer to this question, though, because the direction of power flow in the network will change whenever one changes the node at which power flows into the network (*e.g.*, the direction of flow across line 45 will be different if one injects power at node 2, versus injecting it at node 3).

Fortunately, this observation leads us to the correct method for assigning power flow. To assign a direction for power flow at each of a system’s nodes, simply designate both the node at which power is withdrawn and the node at which power is injected. From here, assign power flows such that no physical laws (first law of thermodynamics, Kirchhoff’s laws) are violated.

So, let us arbitrarily designate node 2 as the “injection” node. Once one specifies the injection node, one may follow two shortcuts to power flow direction:

- (1) Power flows to the reference node along the lines *directly* connected to it.
- (2) Power flows away from the node of injection along the lines directly connected to it.⁴³

Following these two shortcuts, we may assign directional power flows as follows:



The first observation tells us that power flows from nodes 2 and 3 to node 0. Load at node 0 draws power along lines (2,0) and (3,0). The second observation tells us that power flows away from node 2, to nodes 0, 3, and 4.

⁴³ These rules of thumb change with multiple points of injection and withdrawal, but that observation is not relevant to our present analysis.

Now all that remains is to determine directional flows along the lines connected to node 5. Kirchhoff's current law tells us that the sum of currents entering a node equals the sum of currents exiting the node. Since no current (or the power corresponding to it) is being withdrawn at node 4, all of the power entering node 4 will flow to node 5, indicating the direction of flow on this line. The same reasoning applies at node 5, so power flows from node 5 to node 3.⁴⁴

Using the directions indicated above, we write the 6 equations for current flow as:

$$\tilde{p}_{20} = 1,000 \sin(\theta_2 - \theta_0) \quad (1.61a)$$

$$\tilde{p}_{30} = 1,000 \sin(\theta_3 - \theta_0) \quad (1.61b)$$

$$\tilde{p}_{23} = 1,000 \sin(\theta_2 - \theta_3) \quad (1.61c)$$

$$\tilde{p}_{24} = 250 \sin(\theta_2 - \theta_4) \quad (1.61d)$$

$$\tilde{p}_{53} = 250 \sin(\theta_5 - \theta_3) \quad (1.61e)$$

$$\tilde{p}_{45} = 1,000 \sin(\theta_4 - \theta_5) \quad (1.61f)$$

To calculate these derivatives, however, one has to first specify the direction of power flow. Doing so, one obtains

$$K \begin{pmatrix} \mathbf{0} \\ \mathbf{1} \end{pmatrix} = \begin{bmatrix} 1,000 & 0 & 0 & 0 \\ 0 & 1,000 & 0 & 0 \\ 1,000 & -1,000 & 0 & 0 \\ 250 & 0 & -250 & 0 \\ 0 & -250 & 0 & 250 \\ 0 & 0 & 1,000 & -1,000 \end{bmatrix} \quad (1.62)$$

We shorthand this matrix as:

$$K_{(ij)\ell} = \begin{cases} B_{ij}, & \text{if } \ell = i \\ -B_{ij}, & \text{if } \ell = j \end{cases} \quad (1.63)$$

⁴⁴ Kirchhoff's current law also tells us that the current flowing from node 3 to node 0 is equal to the sum of currents flowing from nodes 2 and 5 to node 3.

Multiplying the two matrices gives us the shift factor matrix below:

$$K \begin{pmatrix} \mathbf{0} \\ \mathbf{1} \end{pmatrix} = \begin{matrix} & \theta_2 & \theta_3 & \theta_4 & \theta_5 \\ \text{line 20} & \left[\begin{array}{cccc} 0.6552 & 0.3448 & 0.5172 & 0.4828 \\ 0.3448 & 0.6552 & 0.4828 & 0.5127 \\ 0.3103 & -0.3103 & 0.0345 & -0.0345 \\ 0.0345 & -0.0345 & -0.5517 & -0.4483 \\ 0.0345 & -0.0345 & 0.4483 & 0.5517 \\ 0.0345 & -0.0345 & 0.4483 & -0.4483 \end{array} \right] \end{matrix} \quad (1.64)$$

To interpret this matrix, observe that column values indicate the source of the power, that is, the node at which power is generated (as well as the phase angle corresponding to this node).⁴⁵

The row value corresponds to the line over which this power flows. Therefore, the columns of the matrix yield the PTDFs for every line in the network when power is injected at the node corresponding to that column and withdrawn at the reference node. Reading down the first column, then, 65.52% of the power generated at node 2 flows directly to node 1. The remaining 34.48% flows in a more circuitous fashion along the network's other lines. 31.03% flows through node 3 to node 1, while 3.45% flows around the horn, from node 2 to node 4 to node 5 to node 3 and on to node 1. As mentioned above, the amount exiting node 3, 34.48% of the increment of power generated, is equal to the sum of the flows entering that node: 31.03%, plus 3.45%

5. Conclusion

This paper has taken the reader through the steps to calculate PTDFs for the DC approximation to the AC power system. It first introduced the reader to electrical engineering fundamentals, in an effort to familiarize the reader with the various AC power flow equations,

⁴⁵ Even though we assume that no power is generated at nodes 1-3, we may, in principle, calculate shift factors for any node in a transmission system.

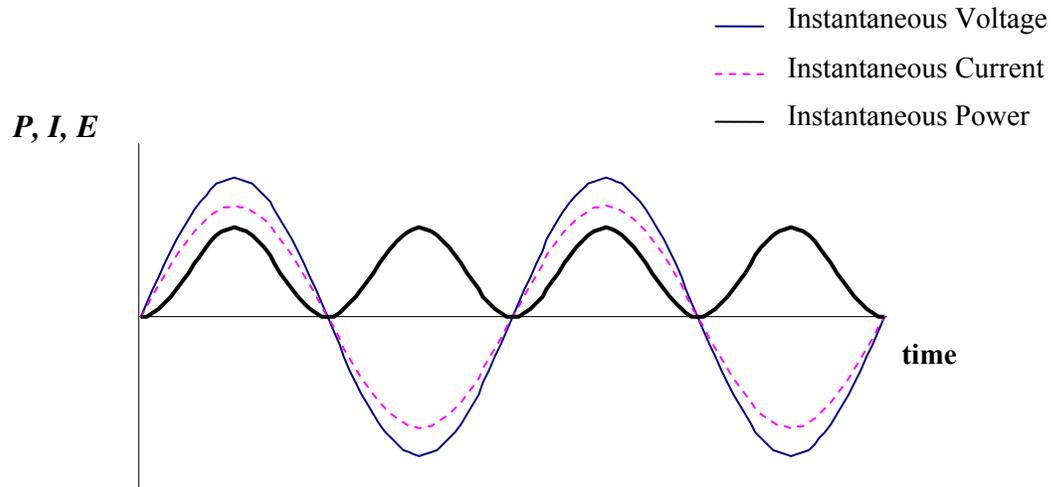
particular with the equations for real power. It then introduced the reader to circuit analysis, presenting a basic DC circuit. Finally, it presented the reader with two tasks of AC analysis: linearizing the AC circuit equations and calculating the network shift factors associated with the DC approximation to the AC circuit. The paper's mathematical demonstrations are meant to allow the reader to solve for PTDFs for circuits beyond the simple 3-node model and to make the reader more comfortable with AC power equations.

References

- Baldick R (2006) Applied Optimization: Formulation and Algorithms for Engineering Systems, Cambridge University Press, New York
- Bushnell J, Stoft S (1997) Improving Private Incentives for Electric Grid Investment. Resource and Energy Economics 19(1-2):85-108.
- Dale C (1995) Basic Electricity & DC Circuits, Prompt Publications, Indianapolis, IN.
- Glover J, Sarma T, Overbye T (2012) Power system Analysis and Design, 5th Edition. Cengage Learning, Stamford, CT.
- Grainger J, Stevenson W (1994) Power System Analysis. McGraw-Hill, Inc., New York, NY.
- Mittle V, Mittal A (2006) Basic Electrical Engineering, 2nd Edition, Tata McGraw-Hill Education Private Limited. New Delhi, India.
- Schwepe F, Caramanis M, Tabors R, Bohn R (1988) Spot Pricing of Electricity. Kluwer Academic Publishing, Norwell, MA.
- Wood A, Wollenberg B (2003) Power Generation, Operation, and Control. Wiley, New York, NY.

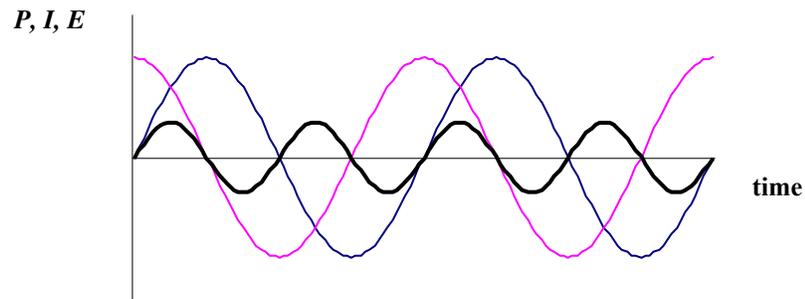
Appendix 1.A: A Further Discussion of Power

When current and voltage are in phase, an electrical system is producing its maximum (real) power output, for a given amount of generation. A graphical analysis is illustrative. Consider first the graph of power, current, and voltage for a purely resistive system (using $R \neq 1\Omega$, for purposes of illustration).



When current and voltage are in phase, power is strictly nonnegative because voltage is never positive when current is negative, and vice-versa. As voltage and current move out of phase, though, this synchronicity vanishes. Consider the inductor example from Figure 2, with instantaneous power added to instantaneous current and voltage.

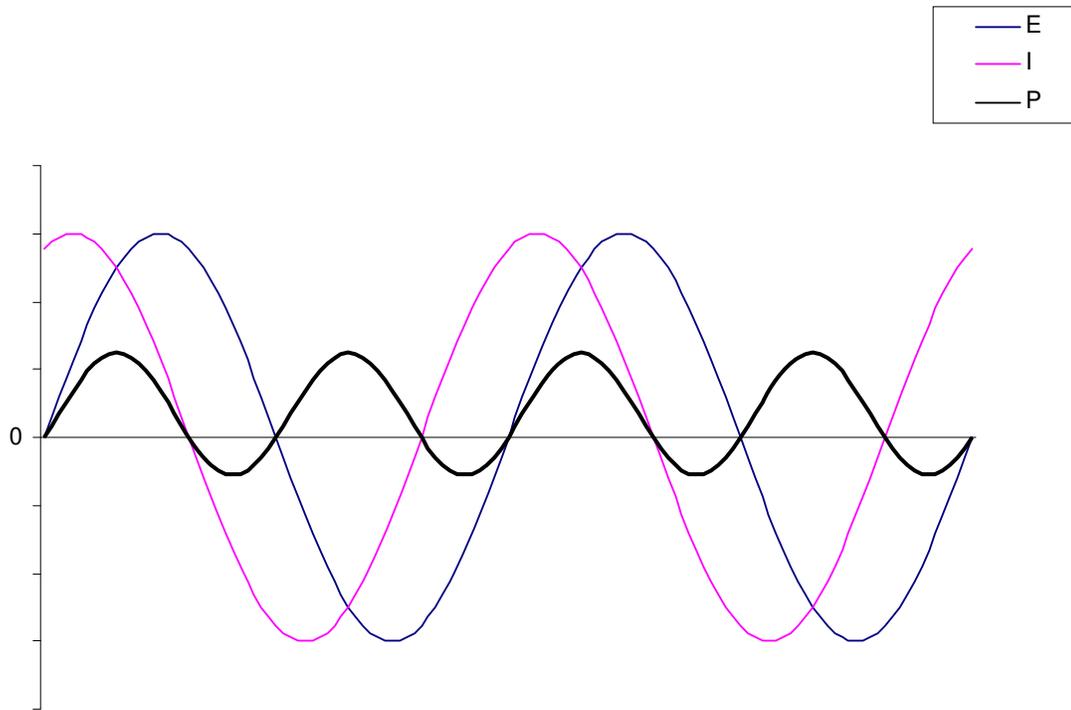
— Instantaneous Voltage
— Instantaneous Current
— Instantaneous Power



Since voltage and current are 90° out of phase for a purely inductive current, half of the time voltage and current will be of the same sign, while they will be of opposite sign the other half of the time. This causes power in this circuit to alternate in cycles of positive and negative values. Positive power means that load in the system is receiving (absorbing) power from the generation source(s), when negative power means that power is actually being returned to the generation source. This means that reactive components (inductors and capacitors) dissipate zero power, as they equally absorb power from, and return power to, the rest of the circuit. ⁴⁶

Adding resistance into the picture, of course, means that the circuit's phase angle will decrease, moving voltage and current closer into phase and allowing for positive power output, as shown graphically below:

⁴⁶ http://www.allaboutcircuits.com/vol_2/chpt_11/1.html



In a circuit characterized by both resistance and reactance, voltage and current are out of phase by less than $|90^\circ|$, and power cycles between positive and negative values, but net power output is positive.

When calculating power in circuits with reactive (and resistive) elements, we apply Ohm's law to reactive circuits. Once again denoting reactance by X , resistance by R , impedance by Z , current by I , and voltage by E , we have the following three expressions for Ohm's law:

$$E = I \times R \quad (\text{circuit w/resistive elements only}) \quad (1A.1)$$

$$E = I \times X \quad (\text{circuit w/reactive elements only}) \quad (1A.2)$$

$$E = I \times Z \quad (\text{circuit w/resistive and reactive elements}) \quad (1A.3)$$

Note that the basic relationship, voltage equals current times resistance, is unchanged. The point is simply that the nature of resistance varies with the different resistance sources. Since the basic relationship for power is unchanged as well, we see that power for resistive, reactive, and resistive plus reactive circuits is, respectively:

$$P = |I|^2 \times R \quad (1A.4)$$

$$Q = |I|^2 \times X \quad (1A.5)$$

$$S = |I|^2 \times Z \quad (1A.6)$$

where

P = real power, measured in Watts

Q = reactive power, measured in volt-amperes reactive (VAR), and

S = apparent power, measured in volt-amperes.

Just as we expressed the relationship between resistance, reactance, and impedance with a right triangle, we express the relationships between different power elements with the *power triangle*.

Consider a circuit containing a 60 volt power source, a capacitor rated at 30 farads, and a 20 Ω resistor. In order to calculate P , Q , and S for this circuit, we must first calculate the current flowing through it. Impedance is simply:

$$Z = \sqrt{R^2 + X^2} = 36.06\Omega .$$

Now that we have E and Z , we calculate current as

$$I = \frac{E}{Z} = \frac{60}{36.06} = 1.66 \text{ amps.}$$

Real power is thus

$$P = I^2 R = 55.1 \text{ watts,}$$

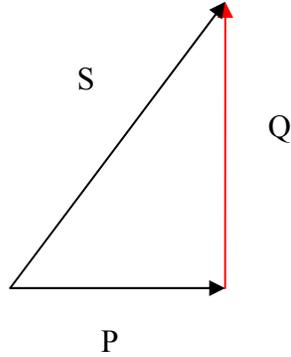
reactive power is

$$Q = I^2 X = 82.7 \text{ VAR, and}$$

complex power is

$$S = I^2 Z = 99.4 \text{ VA.}$$

The power triangle in this case is shown graphically as:



It is straightforward to verify that $P/S = R/Z$, so that one may derive the phase angle either through either power or impedance manipulations. The latter quantity is commonly known as the *power factor*. The ability to reduce a system's phase angle is known as *reactive power supply*. Only generators with this ability may compete in the reactive power market.

Just as we calculated instantaneous power for purely resistive circuits in eq. (1A.4), we may also calculate instantaneous power for a purely inductive circuit, a purely capacitive circuit, and a combined resistive-reactive circuit. For a purely inductive load, current lags voltage by 90° .⁴⁷ Note that we make a notational switch here. While it is common to write voltage as E in Ohm's law, it is customary to write instantaneous voltage as $v(t)$, rather than $e(t)$. Doing so, we write

$$i(t) = I_{\max} \cos(\theta - 90^\circ), \quad (1A.7)$$

$$v(t) = V_{\max} \cos(\theta). \quad (1A.8)$$

Instantaneous power is then

$$\begin{aligned} p(t) &= V_{\max} I_{\max} \cos(\theta) \cos(\theta - 90^\circ) \\ &= V_{\max} I_{\max} \cos(\theta) \sin(\theta) \\ &= |V||I| \sin(2\theta). \end{aligned} \quad (1A.9)$$

⁴⁷ Electrical engineering texts switch between expressing voltage and current in terms of $\sin(\theta)$ and $\cos(\theta)$, because one is free to use either voltage or current as the reference vector.

Integrating over $[0, 2\pi]$, we find that instantaneous power for an inductor circuit does indeed have an average value of zero, as argued above. One may also show that average power for a purely capacitive circuit is zero, as found by integrating

$$-|V||I|\sin(2\theta) \quad (1A.10)$$

Finally, for mixed resistive and reactive circuits, we may write

$$\begin{aligned} p(t) &= V_{\max} I_{\max} \cos(\theta) \cos(\theta - \delta) \\ &= \frac{V_{\max} I_{\max}}{2} [\cos^2(\theta) \cos(\delta) - \sin^2(\theta) \cos(\delta) + \cos(\delta) + 2 \sin(\theta) \cos(\theta) \sin(\delta)]. \end{aligned}$$

The first three terms inside the parentheses reduce to $\cos(\delta)(1 + \cos 2\theta)$. The fourth reduces to $\sin(\delta)\sin(2\theta)$. Thus, we may express instantaneous power in a single-phase AC circuit as

$$s(t) = \frac{V_{\max} I_{\max}}{2} \cos(\delta)(1 + \cos 2\theta) + \frac{V_{\max} I_{\max}}{2} \sin(\delta)\sin(2\theta). \quad (1A.11)$$

Alternatively, since $|V| = \frac{V_{\max}}{\sqrt{2}}$ and $|I| = \frac{I_{\max}}{\sqrt{2}}$ we have

$$s(t) = |V||I| \cos(\delta)(1 + \cos 2\theta) + |V||I| \sin(\delta)\sin(2\theta). \quad (1A.12)$$

The first term is once again real power, while the second term is reactive power. Note that when the phase angle is equal to zero, reactive power is equal to zero, as expected. These equations give us an alternative method for calculating average real, reactive, and complex power in AC circuits. We calculate average real power as

$$\frac{1}{2\pi} \int_0^{2\pi} |V||I| \cos(\delta)(1 + \cos 2\theta) d\theta = |V||I| \cos(\delta). \quad (1A.13)$$

By inspection, eq. (1A.13) is at a maximum when $\delta = 0$. Notice that since $k \int_0^{2\pi} \sin(2\theta) d\theta = 0$ the average value of reactive power from eq. (1A.12) is always zero.

We will also find it useful to express power in terms of the system's admittance matrix.

To do this, we note that complex power injected into the network at bus ℓ is given by

$$S_\ell = V_\ell I_\ell^* \quad (1A.14)$$

Post-multiplying the admittance matrix, given by Eq. (36), by the voltage vector V once again yields $AV = I$, with individual elements

$$I_\ell = A_{\ell\ell}V_\ell + \sum_{k \in \mathfrak{I}(\ell)} A_{\ell k}V_k \quad (1A.15)$$

Substituting Eq. (1A.15) into (1A.14) yields

$$\begin{aligned} S_\ell &= V_\ell \left[A_{\ell\ell}V_\ell + \sum_{k \in \mathfrak{I}(\ell)} A_{\ell k}V_k \right]^* \\ &= |V_\ell|^2 A_{\ell\ell}^* + \sum_{k \in \mathfrak{I}(\ell)} A_{\ell k}^* V_\ell V_k \end{aligned} \quad (1A.16)$$

We may break down complex power into real and reactive power, starting by dividing the admittance terms into their real and imaginary components. Let us start with the expression for admittance as the inverse of impedance. That is, given $Z_{\ell k} = R_{\ell k} + jX_{\ell k}$,

$$\begin{aligned} Y_{\ell k} &= \frac{1}{Z_{\ell k}} = \frac{1}{R_{\ell k} + jX_{\ell k}} = \frac{1}{R_{\ell k} + jX_{\ell k}} \times \frac{R_{\ell k} - jX_{\ell k}}{R_{\ell k} - jX_{\ell k}} \\ &= \frac{R_{\ell k} - jX_{\ell k}}{(R_{\ell k})^2 + (X_{\ell k})^2} \end{aligned} \quad (1A.17)$$

Using Eq. (1A.17), we define the real and imaginary parts of resistance, $G_{\ell k}$ and $B_{\ell k}$, respectively, as follows:

$$G_{\ell k} = \frac{R_{\ell k}}{(R_{\ell k})^2 + (X_{\ell k})^2} \quad (1A.18)$$

$$B_{\ell k} = \frac{X_{\ell k}}{(R_{\ell k})^2 + (X_{\ell k})^2}. \quad (1A.19)$$

Therefore,

$$Y_{\ell k} = G_{\ell k} - jB_{\ell k}. \quad (1A.20)$$

When we get to the equations for complex power, employing the notation V_{\max} gets cumbersome. So, at this point, we replace V_{\max} with u .⁴⁸ We employ Euler's identity to write complex voltage at node ℓ as:

$$V_\ell = u_\ell e^{j\theta_\ell}, \quad (1A.21)$$

and complex power as

$$\begin{aligned} P_\ell + jQ_\ell &= \sum_{k \in \mathfrak{S}(\ell) \cup \ell} u_\ell u_k (G_{\ell k} - jB_{\ell k}) e^{j(\theta_\ell - \theta_k)} = \\ & \sum_{k \in \mathfrak{S}(\ell) \cup \ell} u_\ell u_k [G_{\ell k} \cos(\theta_\ell - \theta_k) + B_{\ell k} \sin(\theta_\ell - \theta_k)] \\ & + ju_\ell u_k [G_{\ell k} \sin(\theta_\ell - \theta_k) - B_{\ell k} \cos(\theta_\ell - \theta_k)]; \end{aligned} \quad (1A.22)$$

where

$$P_\ell = \sum_{k \in \mathfrak{S}(\ell) \cup \ell} u_\ell u_k [G_{\ell k} \cos(\theta_\ell - \theta_k) + B_{\ell k} \sin(\theta_\ell - \theta_k)] \quad (1A.23)$$

$$Q_\ell = \sum_{k \in \mathfrak{S}(\ell) \cup \ell} u_\ell u_k [G_{\ell k} \sin(\theta_\ell - \theta_k) - B_{\ell k} \cos(\theta_\ell - \theta_k)]. \quad (1A.24)$$

Eqs. (1A.23) and (1A.24) are known as power flow equality constraints, and must be satisfied at each bus ℓ of the power system. Note that the lhs of this equation represents net power flow out of node ℓ . If (real, reactive) power generation at this node is greater than (real, reactive) power consumption, then $(P_\ell, Q_\ell) > 0$, respectively. The rhs of the equation represents the fact that this net power flow travels across the transmission network, from the given node, ℓ , to the nodes with which it is connected $k \in \mathfrak{S}(\ell)$. Summing up all of these positive (indicating net power flow from node ℓ to node k) and negative (vice-versa) power flows, we arrive at the net (real, reactive) power flow out of node ℓ . Finally, the first law of thermodynamics implies that net power flow out of node ℓ is equal to the sum of the power flows along each line connected to that node, so we write the power balance constraints as:

⁴⁸ We use alternate notations because electrical engineering texts do, and we wish to expose the reader to these stylistic differences.

$$p_\ell(x) = \sum_{k \in \mathfrak{S}(\ell) \cup \ell} u_\ell u_k [G_{\ell k} \cos(\theta_\ell - \theta_k) + B_{\ell k} \sin(\theta_\ell - \theta_k)] - P_\ell = 0 \quad (1A.25)$$

$$q_\ell(x) = \sum_{k \in \mathfrak{S}(\ell) \cup \ell} u_\ell u_k [G_{\ell k} \sin(\theta_\ell - \theta_k) - B_{\ell k} \cos(\theta_\ell - \theta_k)] - Q_\ell = 0 \quad (1A.26)$$

We also derive the formulas for power flow across a given line ℓk , connecting buses ℓ and k , in the direction of ℓ to k . Now, if we had a simple circuit with no shunt elements, the flow of power across line ℓk would be equal to $u_\ell u_k [G_{\ell k} \cos(\theta_\ell - \theta_k) + B_{\ell k} \sin(\theta_\ell - \theta_k)]$.⁴⁹ This equation simply refers to the drop in voltage (and thus power) moving across nodes in a series circuit. Once we take explicit account of the elements connecting generation to the transmission system (the shunt elements), we have to take into the account the power sourcing at these elements. We derive this element of the power flow equation by setting $k = \ell$ in equations (1A.23) and (1A.24). Doing so yields the quantities

$u_\ell u_\ell [G_{\ell \ell} \cos(\theta_\ell - \theta_\ell) + B_{\ell \ell} \sin(\theta_\ell - \theta_\ell)] = (u_\ell)^2 G_{\ell \ell}$ and, likewise, $-(u_\ell)^2 B_{\ell \ell}$. Therefore, denoting real power flow across line ℓk in the direction of k as $\tilde{p}_{\ell k}$ (and doing so analogously for reactive power flow), we have:

$$\tilde{p}_{\ell k} = (u_\ell)^2 G_{\ell \ell} + u_\ell u_k [G_{\ell k} \cos(\theta_\ell - \theta_k) + B_{\ell k} \sin(\theta_\ell - \theta_k)] \quad (1A.27)$$

$$\tilde{q}_{\ell k} = -(u_\ell)^2 B_{\ell \ell} + u_\ell u_k [G_{\ell k} \sin(\theta_\ell - \theta_k) - B_{\ell k} \cos(\theta_\ell - \theta_k)]. \quad (1A.28)$$

⁴⁹ This is parallel to how one captures a simple DC circuit. Having started with DC circuit analysis myself, this made the transition to AC analysis confusing. The point being that if you want to examine the DC approximation to an AC circuit, the obvious step of learning DC first is, in fact, the exact thing you do NOT want to do.

Appendix 1B: Admittance Matrix for the DC Approximation of the AC Circuit

For convenience, we return to the admittance matrix for the ladder circuit from Eq. (34). For notational simplicity, however, we now denote the admittances with double subscript notation $Y_{\ell k}$, to denote the admittance of the line connecting nodes i and j (e.g., Y_a now becomes $Y_{0,1}$, because line a connects nodes 0 and 1). We may thus express the admittance matrix from Eq. (34) as

$$\begin{bmatrix} Y_{02} + Y_{12} & -Y_{12} & 0 & 0 \\ Y_{12} & Y_{12} + Y_{02} + Y_{23} & -Y_{23} & 0 \\ 0 & -Y_{23} & Y_{23} + Y_{30} + Y_{34} & -Y_{34} \\ 0 & 0 & -Y_{34} & Y_{34} + Y_{40} \end{bmatrix} \quad (1B.1)$$

From Eq. (1A.20), we have $Y_{\ell k} = G_{\ell k} - jB_{\ell k}$, which we could plug in for each value in the matrix above. However, at this point we simplify the matrix by setting the real component of admittance equal to zero. Electrical engineers justify this assumption by noting that $|G_{\ell k}| \ll |B_{\ell k}|$.⁵⁰ Further, we neglect the shunt elements in the line models. The electrical engineering justification for this is that shunt components sometimes have values such that their effect on the circuit is negligible.⁵¹ This reduces the admittance matrix to:

$$\begin{bmatrix} -jB_{12} & jB_{12} & 0 & 0 \\ jB_{12} & -j(B_{12} + B_{23}) & jB_{23} & 0 \\ 0 & jB_{23} & -j(B_{23} + B_{34}) & jB_{34} \\ 0 & 0 & jB_{34} & -jB_{34} \end{bmatrix} \quad (1B.2)$$

⁵⁰ See, e.g. Baldick (2006), pp. 281-283.

⁵¹ See, e.g., Baldick (2006), section 6.2.

PART 2: The Newton-Rhapson Method and AC OPF Formulation

Having tackled the basic concepts of AC power flow, and linearization of AC power flow equations and related shift factor calculations, we are ready to move on to the optimal power flow problem itself.

At this point, even though most energy economists will not delve into solutions of systems of nonlinear equations (rather, the energy economist will normally just skip directly to the DC approximation of the AC optimal power flow problem), we start with the nonlinear AC system to present the reader with additional background material for the presentation of the DC approximation. Thus, we will delve into the AC optimal power flow problem on our way to the solution method for the DC approximation. Some readers may wish to skip ahead to the DC approximation of the AC system at this point.

2.1 Solving Systems of Nonlinear Equations: The Newton-Rhapson Method

AC power flow is captured by a system of nonlinear equations. Thus, instead of moving straight to the solution of the optimal power flow problem (OPF), we first introduce a solution method for systems of nonlinear equations.

Except in relatively simple cases, analytical solutions to systems of nonlinear equations will not exist. Therefore, the best we can hope for is to approximate the solution to this system of equations with a high degree of accuracy (to be specified by the researcher). One uses an iterative algorithm to find this approximate solution. In an iterative algorithm, the researcher makes an initial “guess” as to a problem’s solution and tries to improve upon this guess with successive iterates of the algorithm. The Newton-Rhapson method is quite popular in electrical engineering texts, so we will demonstrate this mechanism.

The general form for the iterates created by an iterative algorithm is:

$$x^{v+1} = x^v + \alpha^v \Delta x^v, \quad v = 0, 1, 2, \dots \quad (2.1)$$

Where:

- x^0 is the initial guess of the solution⁵²
- v is the iteration counter
- x^v is the value of the iterate at the v -th iteration
- α^v is the step size, usually $0 < \alpha \leq 1$ ⁵³
- Δx^v is the step direction, and
- The product $\alpha^v \Delta x^v$ is the update to x^v to obtain x^{v+1} .

Now we consider a function $g(\cdot): \mathfrak{R}^n \rightarrow \mathfrak{R}^n$ and suppose we wish to solve the system of simultaneous nonlinear equations:⁵⁴

$$g(\mathbf{x}) = \mathbf{0}. \quad (2.2)$$

As indicated above, we start with an initial guess to the solution, \mathbf{x}^0 . We may think of the initial guess as our predicted answer, based on our understanding of the problem. Except in the rare circumstance that our first guess is exactly correct, we will have $g(\mathbf{x}) \neq \mathbf{0}$ and we will wish to find an updated guess $\mathbf{x}^1 = \mathbf{x}^0 + \Delta \mathbf{x}^0$ such that:

$$g(\mathbf{x}^1) = g(\mathbf{x}^0 + \Delta \mathbf{x}^0) = \mathbf{0}. \quad (2.3)$$

A first-order Taylor approximation of $g(\cdot)$ about \mathbf{x}^0 yields:

$$g(\mathbf{x}^1) = g(\mathbf{x}^0 + \Delta \mathbf{x}^0) \approx g(\mathbf{x}^0) + \frac{\partial g}{\partial \mathbf{x}} \mathbf{x}^0 \Delta \mathbf{x}^0. \quad (2.4)$$

For a square system of equations we may write $g(\mathbf{x}^0 + \Delta \mathbf{x}^0) \approx g(\mathbf{x}^0) + J(\mathbf{x}^0) \Delta \mathbf{x}^0$.

Therefore, we have that

$$g(\mathbf{x}^1) = g(\mathbf{x}^0 + \Delta \mathbf{x}^0) \approx g(\mathbf{x}^0) + J(\mathbf{x}^0) \Delta \mathbf{x}^0 \quad (2.5)$$

Since we seek an \mathbf{x}^1 such that $g(\mathbf{x}^1) = \mathbf{0}$, and the rhs of eq. (2.5) approximates this value, we will choose a value for $\Delta \mathbf{x}^0$ such that

⁵² We will use k^v to denote the value of k at the v^{th} iteration, rather than k raised to the v^{th} power, which we will denote as $(k)^v$.

⁵³ We will ignore the issue of step size and simply let $\alpha^{(v)} = 1, \forall v$, but *see* Baldick (2006 section 7.4.2.)

⁵⁴ This analysis is taken from Baldick (2006, section 7.1).

$$g(\mathbf{x}^0) + J(\mathbf{x}^0)\Delta\mathbf{x}^0 = \mathbf{0}. \quad (2.6)$$

Solving eq. (2.6) for $\Delta\mathbf{x}^0$ yields $\Delta\mathbf{x}^0 = J(\mathbf{x}^0)^{-1}g(\mathbf{x}^0)$.

Our stage zero update is thus will then be:

$$\Delta\mathbf{x}^0 = -J(\mathbf{x}^0)^{-1}g(\mathbf{x}^0) \quad (2.7)$$

$$\mathbf{x}^1 = \mathbf{x}^0 + \Delta\mathbf{x}^0. \quad (2.8)$$

Eqs. (2.7) and (2.8) constitute the stage zero **Newton-Rhapson update** and $\Delta\mathbf{x}^0$ is the stage zero **Newton-Rhapson step direction**. More generally, at any stage v we have:

$$\Delta\mathbf{x}^v = -J(\mathbf{x}^v)^{-1}g(\mathbf{x}^v) \quad (2.9)$$

$$\mathbf{x}^{v+1} = \mathbf{x}^v + \Delta\mathbf{x}^v \quad (2.10)$$

Example 2.1: Newton-Rhapson Method

Consider the system of non-linear simultaneous equations $g(\mathbf{x}) = \mathbf{0}$ where $g(\cdot): \mathfrak{R}^2 \rightarrow \mathfrak{R}^2$ is defined by:

$$\forall \mathbf{x} \in \mathfrak{R}^2, \quad g(\mathbf{x}) = \begin{bmatrix} \frac{1}{2}(x_1)^2 + x_2 - 4 \\ x_1 + \frac{1}{2}(x_2)^2 - 7 \end{bmatrix} \quad (2.11)$$

We expect that the pair of quadratic equations will have multiple roots. We will focus on the positive root pair. Since x_2 is of larger power in the greater expression, our initial guess will have x_2 greater than x_1 . We prefer to solve one of the equations exactly for the initial guess, while

“eyeballing” the second one.⁵⁵ Let us start, then, with an initial guess of $\mathbf{x}^0 = \begin{bmatrix} 1 \\ 3.5 \end{bmatrix}$. This guess

⁵⁵ The same rule-of-thumb for the root with x_1 positive and x_2 negative should yield an initial guess of approximately $(\sqrt{13}, -2.5)$.

yields $g(\mathbf{x}^0) = \begin{bmatrix} 0 \\ 0.125 \end{bmatrix}$. To calculate the Newton-Rhapson update, $\Delta\mathbf{x}^0$, we must find, then

invert, the Jacobian $J(\mathbf{x}^0)$. By inspection,

$$J(\mathbf{x}) = \begin{bmatrix} x_1 & 1 \\ 1 & x_2 \end{bmatrix}, \quad (2.12)$$

and inverting this matrix, we find:

$$J(\mathbf{x})^{-1} = \begin{bmatrix} \frac{x_2}{x_1x_2 - 1} & \frac{-1}{x_1x_2 - 1} \\ \frac{-1}{x_1x_2 - 1} & \frac{x_1}{x_1x_2 - 1} \end{bmatrix} \quad (2.13)$$

Therefore,

$$-J(\mathbf{x}^0)^{-1}g(\mathbf{x}^0) = -\begin{bmatrix} 1.4 & -0.4 \\ -0.4 & 0.4 \end{bmatrix} \begin{bmatrix} 0 \\ 0.125 \end{bmatrix}, \quad (2.14)$$

or

$$\Delta\mathbf{x}^0 = \begin{bmatrix} 0.05 \\ -0.05 \end{bmatrix}. \quad (2.15)$$

Therefore, we have that

$$\mathbf{x}^1 = \mathbf{x}^0 + \Delta\mathbf{x}^0 = \begin{bmatrix} 1.05 \\ 3.45 \end{bmatrix}, \quad (2.16)$$

Next, we must decide if we wish to stop here, or perform another iteration. There are three basic rules for determining when to stop iterating: (1) the step size, $\Delta\mathbf{x}_v$, is sufficiently small; (2) the error term, $g(\mathbf{x}^v)$, is sufficiently small, where “small” is determined by a metric of the researcher’s choosing; or (3) after a fixed number of iterations. Since we are simply demonstrating the method, we will stop after two iterations.

$$\text{Given the update, } \mathbf{x}^1 = \begin{bmatrix} 1.05 \\ 3.45 \end{bmatrix}, \text{ we compute } g(\mathbf{x}^1) = \begin{bmatrix} 0.00125 \\ 0.00125 \end{bmatrix}.$$

To perform our second iteration we plug the values \mathbf{x}^1 into eq. (2.13) and iterate on eq.

2.14, finding:

$$-J(\mathbf{x}^1)^{-1}g(\mathbf{x}^1) = -\begin{bmatrix} 1.31554 & -.38132 \\ -.38132 & 0.40038 \end{bmatrix} \begin{bmatrix} 0.00125 \\ 0.00125 \end{bmatrix},$$

or

$$\Delta\mathbf{x}^1 = \begin{bmatrix} -0.00117 \\ -0.00002 \end{bmatrix}$$

Therefore, $\mathbf{x}^2 = \begin{bmatrix} 1.04998 \\ 3.44883 \end{bmatrix}$, and $g(\mathbf{x}^2) = \begin{bmatrix} 0.00006 \\ -0.00281 \end{bmatrix}$.

Notice that we have “overshot” the value of x_2 , in that $g(x_2^2)$ is greater than $g(x_2^1)$. This is where the issue of step size sets in. We generally want to decrease absolute values of the errors for all elements of \mathbf{x} from one iteration to the next. Therefore, one would generally reiterate to obtain

$$\Delta\mathbf{x}^2 = \alpha^1 \Delta\mathbf{x}^1, \quad \alpha^1 < 1.$$

It is common to iterate using $\alpha_i^v = \left(\frac{1}{2}\right)^i$, $i = 0, 1, 2, \dots$,

where α_i^v is the i^{th} iterate of α at stage v of the Newton-Rhapson update, and to stop iterating on α when the error term has fallen by some desired amount.

Having demonstrated (if briefly) the Newton-Rhapson method, the next step in solving an AC optimal power flow problem is to set up the AC power flow problem using this method. To do this, we return to the equations for real and reactive power flow:

$$P_\ell = \sum_{k \in \mathfrak{S}(\ell) \cup \ell} u_\ell u_k [G_{\ell k} \cos(\theta_\ell - \theta_k) + B_{\ell k} \sin(\theta_\ell - \theta_k)] \quad (2.18)$$

$$Q_\ell = \sum_{k \in \mathfrak{S}(\ell) \cup \ell} u_\ell u_k [G_{\ell k} \sin(\theta_\ell - \theta_k) - B_{\ell k} \cos(\theta_\ell - \theta_k)]. \quad (2.19)$$

Note that we may separate out the terms in the above summations corresponding to the bus self-admittances as follows:

$$P_\ell = u_\ell^2 G_{\ell\ell} + \sum_{k \in \mathfrak{N}(\ell)} u_\ell u_k [G_{\ell k} \cos(\theta_\ell - \theta_k) + B_{\ell k} \sin(\theta_\ell - \theta_k)] \quad (2.20)$$

$$Q_\ell = -u_\ell^2 B_{\ell\ell} + \sum_{k \in \mathfrak{N}(\ell)} u_\ell u_k [G_{\ell k} \sin(\theta_\ell - \theta_k) - B_{\ell k} \cos(\theta_\ell - \theta_k)] \quad (2.21)$$

Example 2.2: 2-node power flow problem.

Let us examine the power flow equations in a very simple two-node network, as illustrated below:

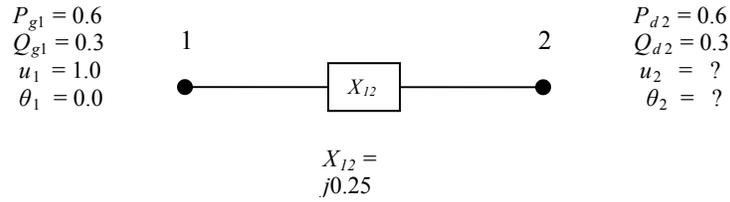


Figure 1: The Two-node Network

As discussed in Part I, we will solve this power flow problem by designating a node (node 1) as the reference node. By convention, we set the phase angle at the reference node equal to 0.0 (radians). We will also set the reference voltage equal to one per unit (1 p.u.). In this example, the generator at node 1 produces 0.6 units of real power and 0.3 units of reactive power, consumed by load at node 2 (thus, we assume line (1,2) is lossless). As above, we also assume that line (1,2) exerts no resistance, but only impedance, equal to $j0.25$, where again, $j = \sqrt{-1}$. We also ignore the impedance of any ground elements. Our problem is now to solve for two unknowns, the voltage angle and magnitude at node 2.⁵⁶

Applying the power flow equations for real and reactive power at node 2 to this example, we have:

$$-0.6 = (u_2)^2 \cdot 0 + 1 \cdot u_2 [4 \sin \theta_2] \quad (2.22)$$

⁵⁶ To extend this methodology to a larger circuit, we simply solve for all unknown voltage magnitudes and angles at every node except reference node, using analogous equations for real and reactive power.

$$-0.3 = 4(u_2)^2 - u_2 \cdot 4[\cos \theta_2]^{57} \quad (2.23)$$

That is, we have a system of two nonlinear equations in two unknowns, which we will (approximately) solve using the Newton-Rhapson method. Carrying all terms over to the rhs, we have

$$u_2[4 \sin \theta_2] + 0.6 = 0 \quad (2.22a)$$

$$4(u_2)^2 - u_2 \cdot 4[\cos \theta_2] + 0.3 = 0 \quad (2.23a)$$

To use the Newton-Rhapson method, we make initial guesses for the two unknowns as

$$\mathbf{x}^0 = \begin{bmatrix} \theta_2^0 \\ u_2^0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

This guess yields initial errors,

$$\mathbf{g}(\mathbf{x}^0) = \begin{bmatrix} 0.6 \\ 0.3 \end{bmatrix}$$

Next, we partially differentiate the equations wrt the unknowns, yielding:

$$\mathbf{J}(\mathbf{x}) = \begin{bmatrix} \frac{\partial p_2}{\partial \theta_2} & \frac{\partial p_2}{\partial u_2} \\ \frac{\partial q_2}{\partial \theta_2} & \frac{\partial q_2}{\partial u_2} \end{bmatrix} = \begin{bmatrix} 4u_2 \cos \theta_2 & 4 \sin \theta_2 \\ 4u_2 \sin \theta_2 & 8u_2 - 4 \cos \theta_2 \end{bmatrix} \quad (2.24)$$

Our stage zero update, from eqs. (2.7, 2.8) is the solution to:

$$\begin{bmatrix} 4 \cos(0) & 4 \sin(0) \\ 4 \sin(0) & 8 - 4 \cos(0) \end{bmatrix} \begin{bmatrix} \Delta \theta_2^0 \\ \Delta u_2^0 \end{bmatrix} = \begin{bmatrix} -0.6 \\ -0.3 \end{bmatrix} \quad (2.25)$$

Inverting and multiplying, we obtain

$$\Delta \mathbf{x}^0 = \begin{bmatrix} -0.15 \\ -0.075 \end{bmatrix} \quad (2.26)$$

$$\theta_2^1 = \theta_2^0 + \Delta \theta_2^0 = 0.0 - 0.15 = -0.15 \text{ rad} \quad (2.27a)$$

⁵⁷ Note that the sign switch on the first term of eq. (2.23) is due to the fact that, ignoring shunt elements, $B_{22} = -B_{21}$. Also note that since power is being withdrawn at node 2, the signs on real and reactive power injection are negative.

$$u_2^1 = u_2^0 + \Delta u_2^0 = 1.0 - 0.075 = 0.925 \text{ p.u.} \quad (2.27b)$$

The new errors are

$$g(\mathbf{x}^1) = \begin{bmatrix} g(x_1^1) \\ g(x_2^1) \end{bmatrix} = \begin{bmatrix} 4(0.925)\sin(-0.15) + 0.6 \\ -4(0.925)^2 + 4(0.925)\cos(-0.15) + 0.3 \end{bmatrix} = \begin{bmatrix} 0.047079 \\ 0.064047 \end{bmatrix}$$

Stopping Criterion:

We will choose to stop when each of the errors are less than 3×10^{-3} . The first-iteration mismatches do not satisfy this criteria, so we re-run the mechanism. Our stage one update is then:

$$\begin{bmatrix} 3.658453 & -0.597753 \\ -0.552921 & 3.444916 \end{bmatrix}^{-1} \begin{bmatrix} -0.047079 \\ -0.064047 \end{bmatrix} = \begin{bmatrix} -0.016335 \\ -0.021214 \end{bmatrix}$$

Our new corrected values are

$$\theta_2^2 = -0.15 - 0.016335 = -0.16635 \text{ rad}$$

$$u_2^2 = 0.925 - 0.021214 = 0.903786 \text{ p.u.}$$

Our new errors are

$$(g(x_1^2), g(x_2^2)) = (0.001391, -0.002077).$$

Since these errors satisfy the stopping criterion, the procedure ends here.

Electrical engineers use such circuit-flow studies to predict the behavior of a power system under such contingencies as lightning strikes that cause short-circuits to occur between transmission lines. Power-flow studies allow system operators to operate the transmission system so as to keep line flows within operating limits even if the most important element in the system fails (known as the “N-1 Criterion”). We have introduced the power flow problem simply to aid our understanding of the optimal power flow problem, though, and will not consider such issues as contingency analysis.

Our next step is to use the Newton-Raphson method to solve nonlinear optimization problems, as the AC optimal power flow problem falls in this category. First, we introduce a

generic nonlinear optimization problem. After understanding the basic problem, we will then add constraints onto the problem and finally address the optimal power flow problem itself. Since the optimal power flow problem involves cost minimization, we will illustrate the Newton-Rhapson method for a minimization problem. We will work with a strictly convex function, so as to use the Newton-Rhapson method to find a unique global minimum/minimizer.

Example 2.3 Minimization of a convex, nonlinear function:

Find the minimum of the function:

$$\forall \mathbf{x} \in \mathfrak{R}^2, \quad f(\mathbf{x}) = 0.01(x_1 - 1)^4 + 0.01(x_2 - 3)^4 + (x_1 - 2)^2 + (x_2 - 1)^2, \quad (2.28)$$

using the Newton-Rhapson method.

The FONC for a minimum (and FOOSC, since the problem is strictly convex) are:

$$\begin{aligned} \frac{\partial f}{\partial x_1} &= 0.04(x_1 - 1)^3 + 2(x_1 - 2) = 0 \\ \frac{\partial f}{\partial x_2} &= 0.04(x_2 - 3)^3 + 2(x_2 - 1) = 0 \end{aligned} \quad (2.29)$$

As the alert reader has already guessed, since this is a minimization problem, the FOCs constitute the set of equations which we wish to set equal to zero. From the cubic term in each equation, this is a system of nonlinear equations (a square system, with two variables in two unknowns), which we may solve using the Newton-Rhapson method. More formally,

$$\mathbf{g}(\mathbf{x}) = \nabla f(\mathbf{x}), \quad (2.30)$$

where $\nabla(\cdot)$ is the gradient function. Applying eq. (2.4):

$$\mathbf{g}(\mathbf{x}^1) = \mathbf{g}(\mathbf{x}^0 + \Delta \mathbf{x}^0) \approx \mathbf{g}(\mathbf{x}^0) + \frac{\partial \mathbf{g}}{\partial \mathbf{x}} \mathbf{x}^0 \Delta \mathbf{x}^0$$

To eq. (2.30) yields

$$\nabla f(\mathbf{x}^1) = \nabla f(\mathbf{x}^0 + \Delta \mathbf{x}^0) \approx \nabla f(\mathbf{x}^0) + \nabla^2 f(\mathbf{x}^0) \Delta \mathbf{x}^0 \quad (2.31)$$

Setting the last term in eq. (2.31) equal to zero and solving for $\Delta \mathbf{x}$ and generalizing to generic step v , we find that the Newton-Rhapson update is:

$$\Delta \mathbf{x}^v = -\left[\nabla^2 f(\mathbf{x}^v)\right]^{-1}\left[\nabla f(\mathbf{x}^v)\right]. \quad (2.32)$$

Therefore, in setting up the Newton-Rhapson update, we must take the Jacobian of the system of first-derivatives. Doing such for this example, we obtain:

$$\nabla f^2(\mathbf{x}^0) = \begin{bmatrix} 0.12(x_1 - 1)^2 + 2 & 0 \\ 0 & 0.12(x_2 - 3)^2 + 2 \end{bmatrix}$$

A sensible initial guess is $(x_1, x_2) = (2, 1)$, as the contribution of the second rhs terms is much larger than that of the first in the error equations (2.29). Using this guess, our initial update is:

$$\begin{bmatrix} \Delta x_1^0 \\ \Delta x_2^0 \end{bmatrix} = -\begin{bmatrix} 2.12 & 0 \\ 0 & 2.96 \end{bmatrix}^{-1} \begin{bmatrix} 0.04 & 0 \\ 0 & -0.32 \end{bmatrix} = \begin{bmatrix} -0.0189 & 0 \\ 0 & 0.1081 \end{bmatrix},$$

and the first iterate is

$$(x_1^1, x_2^1) = (1.981132, 1.108108)$$

The error is then $(0.000042, -0.054647)$. Performing another iterate, we have:

$$\begin{bmatrix} \Delta x_1^1 \\ \Delta x_2^1 \end{bmatrix} = -\begin{bmatrix} 2.1155 & 0 \\ 0 & 2.4295 \end{bmatrix}^{-1} \begin{bmatrix} 0.000042 & 0 \\ 0 & -0.054647 \end{bmatrix} = \begin{bmatrix} -0.000020 & 0 \\ 0 & 0.022493 \end{bmatrix}$$

and the second iterate is $(x_1^2, x_2^2) = (1.981112, 1.1306)$,

with associated error: $(2 * 10^{-8}, 0.000114)$, which we will deem close enough for our purposes.

Since the OPF problem is one of *constrained* minimization of a nonlinear function, we must determine how to incorporate constraints into the Newton-Rhapson method. Though demonstrating this problem based on examination of the null space of the coefficient matrix, A , is

quite intuitive,⁵⁸ we will stick with the Lagrange multiplier approach since energy economists are, in general, more familiar with it.⁵⁹

Example 2.4: Minimization of a convex objective with linear constraints

Let us apply the Newton-Rhapson method to minimize the objective function $f(\mathbf{x}): \mathfrak{R}^2 \rightarrow \mathfrak{R}$, defined as:

$$\forall \mathbf{x} \in \mathfrak{R}^2, \quad f(\mathbf{x}) = 0.01(x_1 - 1)^4 + 0.01(x_2 - 3)^4 + (x_1 - 1)^2 + (x_2 - 3)^2 - 1.8(x_1 - 1)(x_2 - 3)$$

$$s.t. \quad x_1 - x_2 = 8$$

We may set up the Lagrangian as:

$$L = 0.01(x_1 - 1)^4 + 0.01(x_2 - 3)^4 + (x_1 - 1)^2 + (x_2 - 3)^2 - 1.8(x_1 - 1)(x_2 - 3) + \lambda(x_1 - x_2 - 8) \quad (2.33)$$

The first-order necessary conditions (and sufficient, since the problem is strictly convex) for a minimum are:

$$\begin{aligned} \partial L / \partial x_1 &= 0.04(x_1 - 1)^3 + 2(x_1 - 1) - 1.8(x_2 - 3) + \lambda = 0. \\ \partial L / \partial x_2 &= 0.04(x_2 - 3)^3 + 2(x_2 - 3) - 1.8(x_1 - 1) - \lambda = 0. \\ \partial L / \partial \lambda &= x_1 - x_2 - 8 = 0 \end{aligned} \quad (2.34)$$

From eqs. (2.32) and (2.10), we calculate the Newton-Rhapson update as

$$\Delta \mathbf{x}^v = -\left[\nabla^2 L(\mathbf{x}^v) \right]^{-1} \left[\nabla L(\mathbf{x}^v) \right]$$

$$\mathbf{x}^{v+1} = \mathbf{x}^v + \Delta \mathbf{x}^v.$$

and once again the values that the Jacobian and error term take on are determined by the initial guesses for x_1 , x_2 , and λ , and the sequence of iterates generated therefrom. Just eyeing the problem, it appears that the optimal values for x_1 and x_2 will be 6 and -2 , respectively, so we will choose $(x_1, x_2) = (5.5, -2.5)$, since this guess should be close to the final answer. Without

⁵⁸ See Baldick (2006, section 5.8.1 and Ch 13) for an excellent demonstration of this approach, which involves traversing the null space of A in search of the minimizer(s).

⁵⁹ The following explanation is based on Baldick (2006, Section 13.3.2.3).

any intuition on the value the Lagrange multiplier will take, we set $\lambda^0 = 0$.⁶⁰ We may now calculate the components of the Newton-Rhapson update as:

$$\nabla L(\mathbf{x}^0) = \begin{bmatrix} 0.04(5.5-1)^3 + 2(5.5-1) - 1.8(-2.5-3) + 0 \\ 0.04(-2.5-3)^3 + 2(-2.5-3) - 1.8(5.5-1) - 0 \\ -5.5 - 2.5 + 8 \end{bmatrix} = \begin{bmatrix} 22.545 \\ -25.755 \\ 0 \end{bmatrix}$$

and

$$\begin{bmatrix} \Delta x_1^0 \\ \Delta x_2^0 \\ \Delta \lambda \end{bmatrix} = - \begin{bmatrix} 4.43 & -1.8 & 1 \\ -1.8 & 5.63 & -1 \\ 1 & -1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 22.545 \\ -25.755 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.497 \\ 0.497 \\ -23.852 \end{bmatrix}, \quad (2.35)$$

yielding $\begin{bmatrix} x_1^1 \\ x_2^1 \\ \lambda_1 \end{bmatrix} = \begin{bmatrix} 5.997 \\ -2.003 \\ -23.852 \end{bmatrix}$.

It might seem obvious, but the addition of the Lagrangian multiplier adds a third component to the system of FONC, yielding a 3×3 Jacobian, as opposed to a 2×2 matrix.

As a double-check on our intuition as to the problem's solution, we calculate the errors

$$\Delta L(\mathbf{x}^1) = \begin{bmatrix} 0.138 \\ -0.158 \\ 0 \end{bmatrix}.$$

These errors indicate that the answer to this problem was pretty obvious.

Since the OPF problem has nonlinear constraints, as opposed to simply linear ones, we move from the Newton-Rhapson method for solving nonlinear problems with linear constraints to solving nonlinear problems with nonlinear constraints. Doing so is analytically straightforward, though the presence of nonlinear constraints is problematic because the constraint set is no longer

⁶⁰ Well, actually we know that since this is a constrained minimization problem the value of the Lagrange multiplier (on the constraint) will be non-positive, because relaxing the constraint will allow us to do no worse than the constrained minimum. However, since one is not quite sure about how close we are to the unconstrained minimizers, it is difficult to predict the value of λ_0 , so setting $\lambda^0 = 0$ is as good a guess as any.

convex. Once a nonlinear constraint set is introduced, instead of simply searching for the minimizer of the objective function by moving around the hyperplane in n -dimensional space that corresponds to the affine constraint, we must adjust our search by tracing out the constraint function as best we can while scoping out a descent direction for the objective function.⁶¹

Formally, we address a problem of the form:

$$\min_{\mathbf{x} \in S} f(\mathbf{x}) \quad s.t. \quad S = \{ \mathbf{x} \in \mathfrak{R}^n \mid g(\mathbf{x}) = \mathbf{0} \}. \quad (2.36)$$

Where $f(\cdot): \mathfrak{R}^n \rightarrow \mathfrak{R}$ and $g(\cdot): \mathfrak{R}^n \rightarrow \mathfrak{R}^m$. Alternatively, we write this problem as:

$$\min_{\mathbf{x} \in \mathfrak{R}^n} \{ f(\mathbf{x}) \mid g(\mathbf{x}) = \mathbf{0} \}. \quad (2.37)$$

So, we have a problem that is non-affine, both in the n variables and the m constraints.

For a general physical problem, we will have a choice of functions that approximate the feasible set (electric energy being an exception, since it itself operates according to well-defined mathematical functions). In this case that we do have such choice, however, we will wish to choose a function $g(\cdot)$ that satisfies the following definition:

Definition 1: Let $g(\cdot): \mathfrak{R}^n \rightarrow \mathfrak{R}^m$. We say that \mathbf{x}^* is a regular point of the equality constraints $g(\mathbf{x}) = \mathbf{0}$ if:

- (i) $g(\mathbf{x}^*) = \mathbf{0}$,
- (ii) $g(\cdot)$ is partially differentiable with continuous partial derivatives at \mathbf{x}^* , and
- (iii) the m rows of the Jacobian $J(\mathbf{x}^*)$ of $g(\cdot)$ evaluated at \mathbf{x}^* are linearly independent.

For $g(\cdot)$ to have any regular points, it must be the case that $m \leq n$, since otherwise the m rows of $J(\mathbf{x}^*)$ cannot be linearly independent. Also, if $g(\cdot)$ has a regular point, then we can find an

invertible $m \times m$ submatrix: $\frac{\partial g}{\partial \omega}(\mathbf{x}^*)$. This last point is important, of course, because the Newton-

Rhapson method uses the inverse of the relevant Jacobian.

⁶¹ The following discussion is based on Baldick (2006, pp. 529-533). The temptation here is to just say “use the Newton-Rhapson method,” and move on. But for those not quite comfortable with this explanation at this point, we provide a sketch of the details.

Next, we define the tangent plane to the feasible set as follows:

Definition 2: Let $g(\cdot): \mathfrak{R}^n \rightarrow \mathfrak{R}^m$ be partially differentiable and $\mathbf{x}^* \in \mathfrak{R}^n$. Let $J(\cdot): \mathfrak{R}^n \rightarrow \mathfrak{R}^{m \times n}$ be the Jacobian of $g(\cdot)$. Suppose that \mathbf{x}^* is a regular point of the constraints $g(\mathbf{x}) = \mathbf{0}$. Then the tangent plane to the set $S = \{\mathbf{x} \in \mathfrak{R}^n \mid g(\mathbf{x}) = \mathbf{0}\}$ at the point \mathbf{x}^* is the set $T = \{\mathbf{x} \in \mathfrak{R}^n \mid J(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) = \mathbf{0}\}$.

The tangent plane at \mathbf{x}^* is the set of points such that the first-order Taylor approximation of $g(\mathbf{x})$ about \mathbf{x}^* is $\mathbf{0}$. The alert reader will recognize that the set T corresponds to an affine constraint set with Jacobian $J(\mathbf{x})$ passing through the point \mathbf{x}^* . A little math (Baldick (2006, Ch. 13) demonstrates that the set T is the null space of the affine system of constraints $A\mathbf{x} = \mathbf{b}$.

Therefore, if we had an affine system of constraints, we could just scan the null space of A in search of the minimizer of $f(\cdot)$. Since this is not the case, however, the best we can hope to do is move along the tangent plane T of S , evaluated at the regular point \mathbf{x}^* . As we move away from \mathbf{x}^* along T , we will necessarily move out of the feasible set S . Because the tangent plane is generally a better approximation of the constraint set for small values of $(\mathbf{x} - \mathbf{x}^*) = \Delta\mathbf{x}^*$, we will wish to keep our step sizes small to avoid straying far from the feasible set S . Using this update, we arrive at a new (hopefully) regular point \mathbf{x}^{**} , and perform another update.⁶²

Having performed our first update, however, we will necessarily have moved away from the feasible set. This creates a certain tension in our analysis. At the next update we will, at the same time, wish to: (1) move back towards the constraint set and, (2) reduce the value of the objective function.

⁶² At this point, Baldick notes that finite precision in our calculations may not allow us to exactly calculate a step direction that allows us to stay in T . However, this is true for an affine constraint set as well. His point may be, though, that every time we take a step away from the previous regular point, we will move to a new tangent plane, T^v , *i.e.*, the tangent plane corresponding to the v^{th} iteration of the Newton-Raphson update. Having to calculate a new Jacobian at every step certainly decreases the precision of the process.

Focusing on the first point, we will have chosen our initial value, \mathbf{x}^0 , such that $g(\mathbf{x}^0) = \mathbf{0}$. As we have just argued, though, taking a linear approximation to the constraint set in making our initial iterate implies that $g(\mathbf{x}^1) \neq \mathbf{0}$. Therefore, we will try to make the next update, $\Delta \mathbf{x}^1$, bring us back to the constraint (*i.e.*, we will try to choose $\Delta \mathbf{x}^1$ such that $g(\mathbf{x}^2) = \mathbf{0}$). As we have already discussed, one may use one of several methodologies to construct the update $\Delta \mathbf{x}^1$, and we will choose the Newton-Rhapson update. At \mathbf{x}^1 , then, the Newton-Rhapson update is simply $\Delta \mathbf{x}^1 = -J(\mathbf{x}^1)^{-1} g(\mathbf{x}^1)$. This update is consistent with the definition of the tangent plane, above, except that now we are no longer on the constraint set (*i.e.* $g(\mathbf{x}^1) \neq \mathbf{0}$).

Moving on to the second point, we once again will set up the FONC to solve for the minimizer of the objective function. However, since the constraint set is no longer convex, the first-order conditions are no longer sufficient. Formally, the SOSC for a local minimizer are given by the following theorem.

Theorem 1 *Suppose that $f(\cdot): \mathfrak{R}^n \rightarrow \mathfrak{R}$ and $g(\cdot): \mathfrak{R}^n \rightarrow \mathfrak{R}^m$ are twice partially differentiable with continuous second partial derivatives. Let $J(\cdot): \mathfrak{R}^n \rightarrow \mathfrak{R}^{m \times n}$ be the Jacobian of $g(\cdot)$. Consider problem (2.37) and points $\mathbf{x}^* \in \mathfrak{R}^n$ and $\boldsymbol{\lambda}^* \in \mathfrak{R}^m$. Suppose that*

$$\Delta f(\mathbf{x}^*) + J(\mathbf{x}^*)^T \boldsymbol{\lambda}^* = 0 \tag{2.38a}$$

$$g(\mathbf{x}^*) = 0 \tag{2.38b}$$

$\nabla^2 f(\mathbf{x}^*) + \sum_{\ell=1}^m \lambda_\ell^* \nabla^2 g_\ell(\mathbf{x}^*)$ is positive definite on the null space:

$$N = \{ \Delta \mathbf{x} \in \mathfrak{R}^n \mid J(\mathbf{x}^*) \Delta \mathbf{x} = \mathbf{0} \}.$$

Then \mathbf{x}^* is a strict local minimizer of Problem (2.37).

Proof See D.G. Luenberger (1984, Section 10.5)

This condition, requiring that the Hessian of the Lagrangian be positive definite, is to be expected, but it is convenient to have it on hand, anyway. Additionally, it gives us a compact representation of the first-order conditions to problem (2.37). That is, the FONC for minimization of the Lagrangian:

$$L = f(\mathbf{x}) + \lambda(g(\mathbf{x}) - \mathbf{0}), \quad (2.39)$$

are given by eq. (2.38a) and (2.38b).⁶³

Applying the Newton-Rhapson method to this system of equations yields:

$$\begin{bmatrix} \nabla_{xx}^2 L(\mathbf{x}^v, \boldsymbol{\lambda}^v) & J(\mathbf{x}^v)^T \\ J(\mathbf{x}^v) & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}^v \\ \Delta \boldsymbol{\lambda}^v \end{bmatrix} = - \begin{bmatrix} \nabla f(\mathbf{x}^v) + J(\mathbf{x}^v)^T \boldsymbol{\lambda}^v \\ g(\mathbf{x}^v) \end{bmatrix}. \quad (2.40)$$

At this point the reader should be able to go back to eqs. (2.34) and (2.35) deduce that the difference between the Newton-Rhapson update for the nonaffine vs. affine constraint set is the replacement of terms of the matrix $A^{(T)}$ by terms of the matrix $J(\mathbf{x}^v)^{(T)}$, and $g(\mathbf{x}^v)$ for $\mathbf{0}$ for the second-row on the rhs, since we no longer can guarantee that we will stay with the feasible set, S . Alternatively, we may say that the only difference between the two solution methods is that when we have a nonlinear constraint set, we must use the Newton-Rhapson approximation of the corresponding linear constraint set. Since we are already familiar with this method, adding a nonlinear constraint set imposes no new analytical difficulties.

Once solve for $(\Delta \mathbf{x}^v, \Delta \boldsymbol{\lambda}^v)$ using the system of Newton-Rhapson equations, we write the Newton-Rhapson update as:

$$\begin{bmatrix} \mathbf{x}^{v+1} \\ \boldsymbol{\lambda}^{v+1} \end{bmatrix} = \begin{bmatrix} \mathbf{x}^v \\ \boldsymbol{\lambda}^v \end{bmatrix} + \alpha^v \begin{bmatrix} \Delta \mathbf{x}^v \\ \Delta \boldsymbol{\lambda}^v \end{bmatrix} \quad (2.41)$$

⁶³ At this point, one may verify that the corresponding FO(S)C for the linear counterpart to this problem are: $\nabla f(\mathbf{x}^*) + A^T \boldsymbol{\lambda}^* = \mathbf{0}$, and $A \mathbf{x}^* - \mathbf{b} = \mathbf{0}$, where A is defined in the second equation as the coefficient matrix for the linear constraint set.

We explicitly include the step size, α^y , because it reminds us of the tradeoff we face between satisfaction of the constraints and improvement in the objective. We may feel it necessary to set $\alpha^y < 1$ should the full Newton-Rhaphson update lead us to stray away from the constraint $g(\mathbf{x}) = \mathbf{0}$.

The Newton-Rhaphson method handles affine- and non-affine equality constraints easily because the method is meant to solve problems of the form $g(\mathbf{x}) = \mathbf{0}$, whether $g(\mathbf{x})$ is affine or non-affine. Incorporating inequality constraints is more problematic, however, for precisely the same reason. We will examine two types of inequality constraints: (1) non-negativity constraints; and (2) feasible operating ranges (*i.e.* feasible production levels).

We address the first problem non-negativity constrained minimization, subject to affine equality constraints. In this setting, if the objective function is convex, then we have a convex minimization problem. We introduce the FONC for this problem and the familiar Kuhn-Tucker conditions in the following theorem:

Theorem 2 *Let $f(\cdot): \mathfrak{R}^n \rightarrow \mathfrak{R}$ be partially differentiable with continuous partial derivatives, $A \in \mathfrak{R}^{m \times n}$, and $\mathbf{b} \in \mathfrak{R}^m$. Consider the problem,*

$$\min_{\mathbf{x} \in \mathfrak{R}^n} \{f(\mathbf{x}) \mid A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}, \quad (2.42)$$

and a point $\mathbf{x}^* \in \mathfrak{R}^n$. If \mathbf{x}^* is a local minimizer of (&) then:

$$\begin{aligned} \exists \lambda^* \in \mathfrak{R}^m, \exists \mu^* \in \mathfrak{R}^n \text{ such that: } & \nabla f(\mathbf{x}^*) + A^T \lambda^* - \mu^* = \mathbf{0}; \\ & M^* \mathbf{x}^* = \mathbf{0}; \\ & A\mathbf{x}^* = \mathbf{0}; \\ & \mathbf{x}^* \geq \mathbf{0}; \text{ and} \\ & \mu^* \geq \mathbf{0}, \end{aligned} \quad (2.43)$$

where $M^* = \text{diag}\{\mu_\ell^*\} \in \mathfrak{R}^{n \times n}$ is a diagonal matrix with diagonal entries equal to μ_ℓ^* , $\ell = 1, \dots, n$. The vectors λ^* and μ^* satisfying the conditions (2.42) are called the vectors of Lagrange multipliers for the constraints $A\mathbf{x} = \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$, respectively. The conditions $M^* \mathbf{x}^* = \mathbf{0}$ are the familiar complementary slackness conditions.

Proof Nash and Sofer (1996, Section 14.4).

To see why the addition of non-negativity constraints is problematic for the Newton-Rhapson method, consider that the complementary slackness conditions for minimization, $M^* \mathbf{x}^* = \mathbf{0}$, require that for each choice variable x_ℓ , either $u_\ell = 0$, or $x_\ell = 0$ (or both). To apply the Newton-Rhapson method to this condition, we linearize the equation $M\mathbf{x} = \mathbf{0}$ about the values of $\boldsymbol{\mu}$ and \mathbf{x} at iteration v , and use the linearized equations to construct an update. Doing so, we would obtain:

$$\mu_\ell^{v+1} \cdot x_\ell^{v+1} = (\mu_\ell^v + \Delta\mu_\ell^v)(x_\ell^v + \Delta x_\ell^v) \approx \mu_\ell^v \cdot x_\ell^v + \Delta\mu_\ell^v \cdot x_\ell^v + \mu_\ell^v \cdot \Delta x_\ell^v,$$

with the goal of obtaining the complementary slackness conditions corresponding to the minimizer, \mathbf{x}^* . Unfortunately, though, it is entirely possible that we arrive at an iterate for which $x_\ell^v = 0$ *before* we have found the minimizer \mathbf{x}^* . If this happens, then, by the Newton-Rhapson method, we would have:

$$\mu_\ell^{v+1} \cdot x_\ell^{v+1} \approx \mu_\ell^v \cdot x_\ell^v + \Delta\mu_\ell^v \cdot x_\ell^v + \mu_\ell^v \cdot \Delta x_\ell^v = \mu_\ell^v \cdot \Delta x_\ell^v,$$

(since $x_\ell^v = 0$). But the Newton-Rhapson method implies that

$$J(\cdot)\Delta(\cdot) = -g(\cdot),$$

and thus $-\mu_\ell^v \cdot x_\ell^v = 0 \Rightarrow \Delta(x_\ell^v \cdot \mu_\ell^v) = 0 \Rightarrow \Delta x_\ell^v = 0$. Therefore, once we get to the point where $x_\ell^v = 0$ (or $\mu_\ell^v = 0$, for that matter), we will never move from that point. Successful application of the Newton-Rhapson method in this case requires that we avoid the outcomes $x_\ell^v = 0$ and $\mu_\ell^v = 0$.

Fortunately there is a simple (and rather ingenious) method for doing just that, called the interior point algorithm.⁶⁴ Conceptually, the interior point algorithm erects a “barrier” that

⁶⁴ Alternatively, one might use the active set method (*see, e.g.* Baldick (2006, section 16.3))

prevents violation of all of the equality constraints so that $\mathbf{x}^v \in \mathfrak{R}_{++}^n$, $\boldsymbol{\mu}^v \in \mathfrak{R}_{++}^n \quad \forall v$. To avoid the case where an entry of \mathbf{x} or $\boldsymbol{\mu}$ is on the boundary of the positive orthant, the **barrier function** should increase sufficiently rapidly as the objective function approaches the boundary of the feasible region.

To illustrate the interior point algorithm and its associated barrier function, let us consider the following problem:

$$\min_{\mathbf{x} \in \mathfrak{R}^n} f(\mathbf{x}) \quad s.t. \quad \mathbf{x} \geq \mathbf{0}. \quad (2.44)$$

We add a barrier function to the objective function to form the **barrier objective**:

$$\phi(\mathbf{x}) = f(\mathbf{x}) + f_b(\mathbf{x}) \quad (2.45)$$

A suitable barrier function will be differentiable on the interior of the constraint set, but will be unbounded as we approach the boundary of this set. Two such barrier functions are the reciprocal function and the negative of the logarithm function, known as the **logarithmic barrier function**, which we will study further.

Let us define the logarithmic barrier function $f_b(\cdot): \mathfrak{R}_{++}^n \rightarrow \mathfrak{R}$ for the constraints $\mathbf{x} \geq \mathbf{0}$ by:

$$\forall \mathbf{x} \in \mathfrak{R}_{++}^n, f_b(\mathbf{x}) = -\ln(\mathbf{x}). \quad (2.46)$$

At this point, the typical response for an economist is probably a lot of head scratching, because adding a function to the objective that becomes infinite just as the objective is approaching its minimum value makes no sense at all. Fortunately (and this is the point where the approach becomes exceedingly clever), this is not where the methodology terminates. At this juncture, we multiply the barrier function by the **barrier parameter**, $t > 0$, obtaining the revised barrier function:

$$t \cdot f_b(\mathbf{x}) = -t \cdot \ln(\mathbf{x}), \quad (2.47)$$

and the revised barrier objective:

$$\phi(\mathbf{x}) = f(\mathbf{x}) + t \cdot f_b(\mathbf{x})$$

Thus, instead of solving (2.44), we will solve the **barrier problem**:

$$\min_{\mathbf{x} \in \mathfrak{R}^n} \{\phi(\mathbf{x}) \mid A\mathbf{x} = \mathbf{b}, \mathbf{x} > 0\}. \quad (2.48)$$

That is, we minimize $\phi(\mathbf{x})$ over values of $\mathbf{x} \in \mathfrak{R}^n$ that satisfy $A\mathbf{x} = \mathbf{b}$ and which are also in the interior of $\mathbf{x} \geq 0$.

Notice that, for a fixed value of t , the swift increase of the barrier as the boundary nears means that the minimizer of problem (2.47), $\mathbf{x}^*(t) \in \mathfrak{R}_{++}^n$. Thus, Problem (2.48) will not have a solution unless $\exists \mathbf{x} \in \mathfrak{R}_{++}^n \mid A\mathbf{x} = \mathbf{b}$.⁶⁵ Notice also that because we wish to avoid iterates that fall outside of \mathfrak{R}_{++}^n , we must carefully monitor the step size to prevent this from happening.⁶⁶

Since the barrier function distorts the objective function, actually preventing us from finding the optimal value of the decision vector \mathbf{x} , the key to the interior point algorithm is that its initial iterate starts at some positive value for t , but then constructs a path of iterates for successive values of the barrier objective as $t \rightarrow 0$.

As mentioned above, for a fixed value of t , we minimize the barrier objective, obtaining the minimizer $\mathbf{x}^*(t)$. We then run another iterate of the Newton-Rhapson mechanism for a smaller value of the barrier parameter. That is, $t^{y+1} < t^y$, y is the iteration counter for the iterates of t (as opposed to v , the iteration counter for the Newton-Rhapson iterates). At the first run of the interior point algorithm, we are free to set the initial value, $\mathbf{x}^0(t^0)$. However, for every successive iteration, $y > 0$, we set the initial value of the vector of choice variables equal to the optimal value for the previous iteration. That is, $\mathbf{x}^0(t^{y+1}) = \mathbf{x}^*(t^y)$.

⁶⁵ This is known as the Slater condition (*see, e.g.* Nash and Sofer (1996, p. 485)).

⁶⁶ Even though we penalize values of the choice variables that fall near the boundary of the feasible set, this alone does not guarantee that we will not step outside of the feasible set. If the Newton-Rhapson step direction, $\Delta \mathbf{x}$, generated by the Newton-Rhapson method is too large, we might still find ourselves on (or outside) the value of the choice set. We stop this from happening by appropriate choice of α^v .

As the alert reader will already have surmised, the limit of this sequence of iterates is the minimizer, \mathbf{x}^* , of the objective function, $f(\mathbf{x})$. The trajectory of the sequence of minimizers of problem (2.48) is known as the **central path** (Nash and Sofer (1996, Section 17.4)). If the sequences of minimizers and Lagrange multipliers associated with Problem (2.48) converge, then the limits of these sequences satisfy the FONC for Problem (2.41).⁶⁷

Provided that we have successfully implemented the barrier function, so that the central path falls strictly within \mathfrak{R}_{++}^n , we may (partially) ignore the inequality constraints and the domain of the barrier function. That is, computationally, we may examine the problem:

$$\min_{\mathbf{x} \in \mathfrak{R}^n} \{\phi(\mathbf{x}) \mid A\mathbf{x} = \mathbf{b}\}. \quad (2.49)$$

Thus, we have effectively reduced the problem of non-negatively constrained minimization, subject to equality constraints, to equality-constrained minimization alone. So long as f and f_b are convex, we may write the FOSC for minimization of (2.49) as:

$$\nabla \phi(\mathbf{x}) - A^T \boldsymbol{\lambda} = \mathbf{0}, \quad (2.50a)$$

$$\mathbf{b} - A\mathbf{x} = \mathbf{0}. \quad (2.50b)$$

These are a set of nonlinear simultaneous equations which we will once again solve using the Newton-Rhapson method.

To use the Newton-Rhapson method to solve (2.49), let us first write the barrier objective function in terms of its separate components:

$$\phi(\mathbf{x}) = f(\mathbf{x}) + t \cdot f_b(\mathbf{x}). \quad (2.51)$$

⁶⁷ As Baldick (2006, problem 16.20) demonstrates, it is difficult to solve problem (2.47) directly for a very small value of t , because a badly chosen initial guess leads to a poor update of the algorithm. Baldick (2006, p. 638) also notes that, in problems with more than one variable, if the initial guess that is far from the minimizer of Problem (2.48) for the current value of the barrier parameter t , then the coefficient matrix to determine the step-direction may be ill-conditioned. Thus, we are well-advised to start with a fairly large value of t and work toward the minimizer of problem (2.42) using the central path generated from problem (2.48), instead of “cheating” by starting with a small value for t .

⁶⁸ Here, we deviate from Baldick (2006), in that we write the Lagrangian multiplier term as: $\lambda(\mathbf{b} - \mathbf{x}A)$, as opposed to $\lambda(A\mathbf{x} - \mathbf{b})$. The rest of the demonstration follows accordingly.

Therefore, we may write the first term in Eq. (2.50a) as:

$$\nabla \phi(\mathbf{x}) = \nabla f(\mathbf{x}) + t \cdot \nabla f_b(\mathbf{x}). \quad (2.52)$$

Writing Eq. (2.52) in terms of the logarithmic barrier function yields:

$$\begin{aligned} \nabla \phi(\mathbf{x}) &= \nabla f(\mathbf{x}) + t \cdot \nabla(-\ln \mathbf{x}) \\ &= \nabla f(\mathbf{x}) + t \begin{bmatrix} -\frac{1}{x_1} \\ \vdots \\ -\frac{1}{x_n} \end{bmatrix} \\ &= \nabla f(\mathbf{x}) - t[X]^{-1}\mathbf{1}, \end{aligned} \quad (2.53)$$

where $X = \text{diag}\{x_\ell\} \in \mathfrak{R}^{n \times n}$ is a diagonal matrix with diagonal entries equal to x_ℓ , $\ell = 1, \dots, n$.

The Newton-Rhapson update for this problem is expressed by the familiar equations:

$$\begin{bmatrix} \nabla^2 \phi(\mathbf{x}^v) & -A^T \\ -A & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}^v \\ \Delta \boldsymbol{\lambda}^v \end{bmatrix} = \begin{bmatrix} -\nabla \phi(\mathbf{x}^v) + A^T \boldsymbol{\lambda}^v \\ A\mathbf{x}^v - \mathbf{b} \end{bmatrix}. \quad (2.54)$$

Decomposing the barrier function, this system becomes:

$$\begin{bmatrix} \nabla^2 f(\mathbf{x}^v) + t[X^v]^{-2} & -A^T \\ -A & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}^v \\ \Delta \boldsymbol{\lambda}^v \end{bmatrix} = \begin{bmatrix} -\nabla f(\mathbf{x}^v) + t[X^v]^{-1}\mathbf{1} + A^T \boldsymbol{\lambda}^v \\ A\mathbf{x}^v - \mathbf{b} \end{bmatrix} \quad (2.55)$$

Updates of this family are known as the **primal interior point algorithm**.

As we move along here, implementation of the Newton-Rhapson method becomes increasingly complicated. The primal interior point algorithm does not simply involve finding a single vector of minimizers, but rather successive rounds of them. This simply emphasizes the indispensability of computer programming in successfully addressing any but the simplest of problems using this algorithm. Nevertheless, it is instructive to work out an example using this method.

Example 2.5: Primal interior point algorithm

Consider the following problem:

$$\min_{\mathbf{x} \in \mathbb{R}^2} \{(x_1)^2 - (x_2)^2 \mid x_1 + x_2 = 1, x_1 \geq 0, x_2 \geq 0\}.$$

To implement the primal interior point algorithm, with a negative logarithmic penalty function, we set up the augmented Lagrangian as follows:

$$L = (x_1)^2 - (x_2)^2 - t \ln x_1 - t \ln x_2 + \lambda(1 - x_1 - x_2) \quad (2.56)$$

To demonstrate the algorithm effectively, we want to avoid too large of a step size. We will set

$$t=1. \text{ Having done so (and cheating a little bit), setting } \mathbf{x}^0 = \begin{bmatrix} 0.2875 \\ 0.7125 \\ -2.825 \end{bmatrix} \text{ will avoid a large step.}$$

To solve for the step zero Newton-Rhapson update, we must solve for the FOC for a minimum and the Hessian of the Jacobian, which we show below:

$$\begin{aligned} \frac{\partial L}{\partial x_1} &= 2x_1 - \frac{t}{x_1} - \lambda = 0 \\ \frac{\partial L}{\partial x_2} &= -2x_2 - \frac{t}{x_2} - \lambda = 0 \\ \frac{\partial L}{\partial \lambda} &= 1 - x_1 - x_2 = 0 \end{aligned} \quad (2.57)$$

$$H = \begin{bmatrix} 2 + \frac{t}{(x_1)^2} & 0 & -1 \\ 0 & -2 + \frac{t}{(x_2)^2} & -1 \\ -1 & -1 & 0 \end{bmatrix} \quad (2.58)$$

Notice that eq. (2.58) and the first rhs term in eq. (2.55) are, in fact, equivalent, since

$$\nabla^2 f(\mathbf{x}^v) + t[X^v]^{-2} = \begin{bmatrix} 2 + \frac{t}{(x_1)^2} & 0 \\ 0 & -2 + \frac{t}{(x_2)^2} \end{bmatrix}, \text{ and}$$

$$A = [-1 \ -1].$$

Notice also that the lhs of eq. (2.55) is simply the system of FOCs (with the exception that the zeros are explicit in eq. (2.57), whereas they are implicit in the lhs terms of eq. (2.55) (because we want $g(x^v) = \mathbf{0}$, so the lhs of eq. (2.55) is $-\Delta g(x^v) = -g(x^v) - \mathbf{0}$, but the zero vector is implicit.

The Newton-Rhapson step-zero step direction is thus derived from the solution of:

$$\begin{bmatrix} 2 + \frac{t}{(x_1^0)^2} & 0 & -1 \\ 0 & -2 + \frac{t}{(x_2^0)^2} & -1 \\ -1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \Delta x_1^0 \\ \Delta x_2^0 \\ \Delta \lambda^0 \end{bmatrix} = - \begin{bmatrix} 2x_1^0 - \frac{t}{x_1^0} - \lambda^0 \\ -2x_2^0 - \frac{t}{x_2^0} - \lambda^0 \\ 1 - x_1^0 - x_2^0 \end{bmatrix} \quad (2.59)$$

Substituting in the initial values for x_1 , x_2 , and λ , we find

$$\begin{bmatrix} \Delta x_1^0 \\ \Delta x_2^0 \\ \Delta \lambda^0 \end{bmatrix} = \begin{bmatrix} 0.005314 \\ -0.005314 \\ -0.003349 \end{bmatrix}, \quad (2.60)$$

and therefore,

$$\begin{bmatrix} x_1^1 \\ x_2^1 \\ \lambda^1 \end{bmatrix} = \begin{bmatrix} 0.2928 \\ 0.7072 \\ -2.8280 \end{bmatrix},$$

with associated errors:

$$g(\mathbf{x}^1) = \begin{bmatrix} -0.0017 \\ -0.00043 \\ 0.0 \end{bmatrix}$$

For our purposes, these errors are sufficiently small that we may move on to the next step in the algorithm. Now we iterate on the barrier parameter, t . Doing so, we decrease t from one to 0.9

⁶⁹ In solving this system, instead of deriving a formal inverse, $H^{-1}(x_1^0, x_2^0, t)$, it is much quicker to simply plug the values of these variables into $H(\cdot)$ and invert numerically.

(and to 0.8 in the next run, and so on). As mentioned above, for our initial guesses for x_1, x_2 , and λ , we use the optimal values, \mathbf{x}^* , we found for $t = 1$. The augmented Lagrangian is again

$$L = (x_1)^2 - (x_2)^2 - t \ln x_1 - t \ln x_2 + \lambda(1 - x_1 - x_2),$$

so that the Newton-Rhapson update is of exactly the same form (as one would expect), but this

time with the reduced value of t . So, plugging the value $\begin{bmatrix} x_1^1 \\ x_1^2 \\ \lambda^1 \end{bmatrix} = \begin{bmatrix} 0.2928 \\ 0.7072 \\ -2.8280 \end{bmatrix}$, $t = 0.9$ into (2.59).

we find:
$$\begin{bmatrix} \Delta x_1^0 \\ \Delta x_2^0 \\ \Delta \lambda \end{bmatrix} = \begin{bmatrix} -0.0162 \\ 0.0162 \\ 0.1377 \end{bmatrix},$$

$$\begin{bmatrix} x_1^1 \\ x_1^2 \\ \lambda^1 \end{bmatrix} = \begin{bmatrix} 0.2766 \\ 0.7234 \\ -2.6903 \end{bmatrix}, \tag{2.62}$$

$$g(\mathbf{x}^1) = \begin{bmatrix} -0.0103 \\ -0.00062 \\ 0.0 \end{bmatrix}.$$

Assuming a tolerance of 0.005 on the error term, we would next run another iterate with $t = 0.9$ before ratcheting t down to 0.8 (assuming, of course, we reach the desired tolerance in the next iteration).

As we would expect, after we decreased the barrier parameter, the values for x_1 and x_2 moved closer to the true minimizers $(x_1^*, x_2^*) = (0, 1)$. The central path, then, would converge to this value as $t \rightarrow 0$.

Having introduced the primal interior point algorithm, we are ready to move on to a variant of this algorithm that includes explicit consideration of non-negativity constraints, called the **primal-dual interior point algorithm**.

If the primal interior point algorithm is clever, the primal-dual interior point algorithm is truly inspired. The algorithm's focus is on the handling of the complementary slackness conditions for non-negatively constrained minimization:

$$\mu_\ell x_\ell = 0, \mu_\ell \geq 0, x_\ell \geq 0; \quad \forall \ell = 1, \dots, n,$$

where n is the number of non-negatively constrained variables.⁷⁰

As argued above, one cannot incorporate the complementary slackness conditions directly into

the Newton-Raphson algorithm, because the ordered pairs $\begin{bmatrix} x_\ell \\ \mu_\ell \end{bmatrix}$ that satisfy these conditions lie

either on the x axis, the μ axis, or at the origin, and linearization of the complementary slackness conditions implies that once a variable hits the axis, it stays there.

Notice, though, if we replace the conventional complementary slackness condition with a revised condition:

$$\mu_\ell x_\ell = t, \quad \forall \ell = 1, \dots, n, \tag{2.63}$$

we have the following:

1. The complementary slack condition no longer brings us to either the x or μ axis;
2. Linearization of (2.63), together with an explicit requirement to avoid the x and μ axes, yields a useful update that can approximate the kink in the complementary slackness conditions;
3. Solving for $\mu_\ell = \frac{t}{x_\ell}$ will come in handy in the FOCs; and
4. As $t \rightarrow 0$, the modified complementary slackness conditions approach the complementary slackness conditions for minimization of the objective function. That is, as t is reduced, the hyperbolic-shaped sets corresponding to the modified complementary slackness conditions come closer to the set of points satisfying the complementary slackness conditions and the non-negativity constraints.

Point 3 suggests use the logarithmic barrier function with barrier parameter t : $t(-\ln x)$.

We may set up the barrier problem for non-negatively constrained minimization with linear constraints as:

⁷⁰ The following discussion is based on Baldick (2006, Section 16.4.3.3), with the one difference mentioned in fn. 17.

$$\min_{\mathbf{x} \in \mathcal{R}^n} L = f(\mathbf{x}) + t(-\ln \mathbf{x}) + \lambda(\mathbf{b} - A\mathbf{x}). \quad (2.64)$$

The FONC for a minimum are:

$$\frac{\partial L}{\partial \mathbf{x}} = \nabla f(\mathbf{x}) - t[X]^{-1} - A^T \lambda = \mathbf{0}, \quad (2.65a)$$

$$\frac{\partial L}{\partial \lambda} = \mathbf{b} - A\mathbf{x} = \mathbf{0}. \quad (2.65b)$$

which, from (2.64), we may re-write as

$$\nabla f(\mathbf{x}) - A^T \lambda - \boldsymbol{\mu} = \mathbf{0}.^{71} \quad (2.65c)$$

Now we are ready to use the Newton-Rhapson method to find a step direction to solve the complementary slackness condition and the two first-order conditions, eq. (2.65a), (2.65a):

$$t\mathbf{1} - X\boldsymbol{\mu} = \mathbf{0}, \quad (2.66)$$

$$\nabla f(\mathbf{x}) - A^T \lambda - \boldsymbol{\mu} = \mathbf{0},$$

$$A\mathbf{x} = \mathbf{b}.$$

By inspection, the Newton-Rhapson step direction is:

$$\begin{bmatrix} X^v & -\mathbf{M}^v & \mathbf{0} \\ -\mathbf{I} & \nabla^2 f(\mathbf{x}^v) & -A^T \\ \mathbf{0} & -A & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta \boldsymbol{\mu}^v \\ \Delta \mathbf{x}^v \\ \Delta \lambda^v \end{bmatrix} = \begin{bmatrix} \mathbf{X}^v \boldsymbol{\mu}^v - t\mathbf{1} \\ -\nabla f(\mathbf{x}^v) + A^T \lambda^v + \boldsymbol{\mu}^v \\ A\mathbf{x}^v - \mathbf{b} \end{bmatrix}, \quad (2.67)$$

where $\mathbf{M}^v = \text{diag}\{\mu_\ell^v\}$ and $\mathbf{X}^v = \text{diag}\{x_\ell^v\}$.

The Newton-Rhapson update is⁷²

⁷¹ Notice that we have implicitly expanded the expression $\mu_\ell x_\ell = t$, across all variables, \mathbf{x} , as

$\boldsymbol{\mu}[X] = t\mathbf{1}$, which implies that $\boldsymbol{\mu} = [X]^{-1}t$, with X as defined above. The true brilliance of this method is that differentiating the augmented Lagrangian yields the same Lagrange multiplier one derives by differentiating the ordinary Lagrangian with respect to the non-negativity constraint.

⁷² Because the full Newton-Rhapson update may take us outside the bounds of the feasible region, we will modify the update according to the step size, α^v , as necessary. For further discussion of the step-size for the primal-dual problem, see Baldick (2006, pp. 643-645).

$$\begin{bmatrix} \boldsymbol{\mu}^{v+1} \\ \mathbf{x}^{v+1} \\ \boldsymbol{\lambda}^{v+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}^v \\ \mathbf{x}^v \\ \boldsymbol{\lambda}^v \end{bmatrix} + \begin{bmatrix} \Delta \boldsymbol{\mu}^v \\ \Delta \mathbf{x}^v \\ \Delta \boldsymbol{\lambda}^v \end{bmatrix}.^{73} \quad (2.67)$$

As before, we will find the optimal vector of choice variables, \mathbf{x}^* as the limit of the central path at t approaches zero, provided that the problem is well-behaved.

Example 2.6: Primal-dual interior point algorithm

Consider again the problem from example 2.5:

$$\min_{\mathbf{x} \in \mathfrak{R}^2} \left\{ (x_1)^2 - (x_2)^2 \mid x_1 + x_2 = 1, x_1 \geq 0, x_2 \geq 0 \right\}. \quad (2.68)$$

We write the augmented Lagrangian for problem (2.65) corresponding to the primal-dual interior point algorithm as:

$$\min_{\mathbf{x} \in \mathfrak{R}^n} L = (x_1)^2 - (x_2)^2 + t(-\ln x_1) + t(-\ln x_2) + \lambda(1 - x_1 - x_2) \quad (2.69)$$

The FOSC for a minimum are:

$$\frac{\partial L}{\partial x_1} = 2x_1 - \lambda - \mu_1 = 0 \quad (2.70a)$$

$$\frac{\partial L}{\partial x_2} = -2x_2 - \lambda - \mu_2 = 0 \quad (2.70b)$$

$$\frac{\partial L}{\partial \lambda} = 1 - x_1 - x_2 = 0, \quad (2.70c)$$

with associated complementary slackness conditions

$$\begin{aligned} t - x_1 \mu_1 &= 0 \\ t - x_2 \mu_2 &= 0 \end{aligned} \quad (2.70d)$$

The Newton-Rhapson update is given by the solution of the following system of equations:⁷⁴

⁷³ Iterating until we approach a minimizer, $\mathbf{x}^{t_0^*}$, corresponding to a fixed value of $t = t_0$ would be very computationally expensive, as we continue this process for smaller and smaller values of t . Therefore, instead of doing so, a more conventional method is to reduce t slightly at every iteration of the process. Reducing t excessively yields a poor update direction because the step-size to maintain non-negativity of the iterates will be very small (See Baldick (2006, Exercise 16.22)).

$$\begin{bmatrix} x_1 & 0 & \mu_1 & 0 & 0 \\ 0 & x_2 & 0 & \mu_2 & 0 \\ -1 & 0 & 2 & 0 & -1 \\ 0 & -1 & 0 & -2 & -1 \\ 0 & 0 & -1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \Delta\mu_1 \\ \Delta\mu_2 \\ \Delta x_1 \\ \Delta x_2 \\ \Delta\lambda \end{bmatrix} = - \begin{bmatrix} t - x_1\mu_1 \\ t - x_2\mu_2 \\ 2x_1 - \lambda - \mu \\ -2x_2 - \lambda - \mu \\ 1 - x_1 - x_2 \end{bmatrix}. \quad (2.68)$$

As you will notice, even for a simple problem, the Newton-Rhapson method is already getting out of hand (which is why for any problem worth solving, one writes code and lets the computer do the rest of the work). Nevertheless, we will demonstrate the first iteration of this process.

Using the same barrier parameter, $t=1$, we will use the revised values for x_1 , x_2 , and λ found in eq. (2.62) for the first run of the primal interior point algorithm. With x_1 and x_2

determining the values for μ_1 and μ_2 , we may write the initial vector, $\mathbf{x}^0 = \begin{bmatrix} 3.6153 \\ 1.3824 \\ 0.2766 \\ 0.7234 \\ -2.6903 \end{bmatrix}$

Inserting these values into (2.68), the stage zero update is:

$$\begin{bmatrix} \mu_1^1 \\ \mu_2^1 \\ x_1^1 \\ x_2^1 \\ \lambda^1 \end{bmatrix} = \begin{bmatrix} 3.6153 \\ 1.3824 \\ 0.2766 \\ 0.7234 \\ -2.6903 \end{bmatrix} + \begin{bmatrix} -0.20325 \\ 0.02972 \\ 0.01555 \\ -0.01555 \\ -0.13748 \end{bmatrix} = \begin{bmatrix} 3.4121 \\ 1.4121 \\ 0.2922 \\ 0.7079 \\ -2.8278 \end{bmatrix}.^{75}$$

As one might expect, the method for solving inequality constrained optimization problems is quite similar to that for solving non-negatively constrained optimization problems. Again, we will wish to keep the Newton-Rhapson iterates bounded away from the constraint by use of a barrier parameter. The difference is that we will include the familiar method of slack variables (along with complementary slackness conditions) in solving the problem.

Formally, we will consider the following problem:

⁷⁴ It does seem strange that we lose the t/x^2 term in the Hessian, but if you look at Baldick (2006, p. 643) the expected derivative term reappears.

⁷⁵ At this point, the temptation is to just pack our bags, say, “you get the idea,” and leave the rest to the operations analysts. But, we will continue on, anyway.

$$\min_{\mathbf{x} \in \mathfrak{R}^n} \{f(\mathbf{x}) \mid A\mathbf{x} = \mathbf{b}, C\mathbf{x} \leq \mathbf{d}\}, \quad (2.69)$$

where $A \in \mathfrak{R}^{m \times n}$, $\mathbf{b} \in \mathfrak{R}^m$, $C \in \mathfrak{R}^{r \times n}$, and $\mathbf{d} \in \mathfrak{R}^r$ are constants. The feasible set defined by the linear equality and inequality constraints is convex (see Baldick (2006, Exercise 2.36)). Thus, if $f(\cdot)$ is convex on the feasible set, the problem is convex. Again, we introduce the FONC for this problem in the following theorem:

Theorem 3 *Let $f : \mathfrak{R}^n \rightarrow \mathfrak{R}$ be partially differentiable with continuous partial derivatives, $A \in \mathfrak{R}^{m \times n}$, $\mathbf{b} \in \mathfrak{R}^m$, $C \in \mathfrak{R}^{r \times n}$, and $\mathbf{d} \in \mathfrak{R}^r$. Consider the problem,*

$$\min_{\mathbf{x} \in \mathfrak{R}^n} \{f(\mathbf{x}) \mid A\mathbf{x} = \mathbf{b}, C\mathbf{x} \leq \mathbf{d}\},$$

and a point $\mathbf{x}^* \in \mathfrak{R}^n$. If \mathbf{x}^* is a local minimizer of (2.69) then:

$$\begin{aligned} \exists \boldsymbol{\lambda}^* \in \mathfrak{R}^m, \exists \boldsymbol{\mu}^* \in \mathfrak{R}^n \text{ such that : } \nabla f(\mathbf{x}^*) + A^T \boldsymbol{\lambda}^* - C^T \boldsymbol{\mu}^* &= \mathbf{0}; \\ M^*(C\mathbf{x}^* - \mathbf{d}) &= \mathbf{0}; \\ A\mathbf{x}^* &= \mathbf{b}; \\ C\mathbf{x}^* &\leq \mathbf{d}; \text{ and} \\ \boldsymbol{\mu}^* &\geq \mathbf{0}, \end{aligned} \quad (2.70)$$

where $M^* = \text{diag}\{\mu_\ell^*\} \in \mathfrak{R}^{r \times r}$. The vectors \mathbf{x}^* and $\boldsymbol{\mu}^*$ satisfying the conditions (2.70) are called the vectors of Lagrange multipliers for the constraints $A\mathbf{x} = \mathbf{b}$ and $C\mathbf{x} \leq \mathbf{d}$, respectively. The conditions $M^*(C\mathbf{x}^* - \mathbf{d}) = \mathbf{0}$ are the complementary slackness conditions.

Proof Nash and Sofer (1996, Section 14.4). Again, the FOCs are sufficient provided that the objective is convex on the feasible set.

Introducing the slack variable, $\mathbf{w} \in \mathfrak{R}^r$, we transform the inequality constraint $C\mathbf{x} \leq \mathbf{d}$ into the equivalent equality constraint: $C\mathbf{x} + \mathbf{w} = \mathbf{d}$. Upon making this substitution, (2.69) becomes:

$$\min_{\mathbf{x} \in \mathfrak{R}^n, \mathbf{w} \in \mathfrak{R}^r} \{f(\mathbf{x}) \mid A\mathbf{x} = \mathbf{b}, C\mathbf{x} + \mathbf{w} = \mathbf{d}, \mathbf{w} \geq \mathbf{0}\}, \quad (2.71)$$

transforming the inequality constrained problem into a non-negatively constrained problem. Of course, this formulation suggests an analogous method for solving the inequality constrained

problem as in solving the non-negatively constrained optimization problem: introduce a barrier function into the objective function and then minimize the barrier objective function.

We form the barrier objective function $\phi(\cdot): \mathfrak{R}^n \times \mathfrak{R}_{++}^r \rightarrow \mathfrak{R}$ as:

$$\forall \mathbf{x} \in \mathfrak{R}^n, \forall \mathbf{w} \in \mathfrak{R}_{++}^r, \phi(\mathbf{x}, \mathbf{w}) = f(\mathbf{x}) + t \cdot f_b(\mathbf{w}).$$

(Note that the barrier function does not apply to x here, unless x is non-negatively constrained as well).

Once again, we must assume that the Slater condition holds in order to solve this problem using an interior point algorithm. We will continue to specify a logarithmic barrier function, so that:

$$\forall \mathbf{w} \in \mathfrak{R}_{++}^r, f_b(\mathbf{w}) = \sum_{\ell=1}^r \ln(w_\ell),$$

$$\nabla f_b(\mathbf{w}) = -[W]^{-1} \mathbf{1},$$

where $W = \text{diag}\{w_\ell\} \in \mathfrak{R}^{r \times r}$ is a diagonal matrix with diagonal entries equal to $w_\ell \quad \ell = 1, \dots, r$.

Setting up the Lagrangian for this general problem, we have:

$$L = f(\mathbf{x}) + t(-\ln \mathbf{w}) + \lambda(\mathbf{b} - A\mathbf{x}) + \mu(\mathbf{d} - C\mathbf{x} - \mathbf{w}). \quad (2.72)$$

Since this problem is quite similar to eq. (2.69), we will not explicitly work out an example,⁷⁶ but rather simply set up the general FONC for problem (2.72), and the Newton-Rhapson step direction, as follows:

FONC:

$$\begin{aligned} \nabla f(\mathbf{x}) - A^T \lambda - C^T \mu &= \mathbf{0} \\ A\mathbf{x} &= \mathbf{0} \\ C\mathbf{x} + \mathbf{w} &= \mathbf{0} \\ t\nabla f_b(\mathbf{w}) + \mu &= \mathbf{0} \end{aligned} \quad (2.73)$$

The Newton-Rhapson step direction is given as the solution to:

⁷⁶ But see Baldick (2006, section 17.3.1.3).

$$\begin{bmatrix} M^v & \mathbf{0} & \mathbf{0} & W^v \\ \mathbf{0} & \nabla^2 f(\mathbf{x}^v) & A^T & C^T \\ \mathbf{0} & A & \mathbf{0} & \mathbf{0} \\ \mathbf{I} & C & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w}^v \\ \Delta \mathbf{x}^v \\ \Delta \boldsymbol{\lambda}^v \\ \Delta \boldsymbol{\mu}^v \end{bmatrix} = - \begin{bmatrix} W^v \boldsymbol{\mu}^v + t \mathbf{1} \\ \nabla f(\mathbf{x}^v) - A^T \boldsymbol{\lambda}^v - C^T \boldsymbol{\mu}^v \\ \mathbf{b} - A \mathbf{x}^v \\ \mathbf{d} - C \mathbf{x}^v - \mathbf{w}^v \end{bmatrix}. \quad (2.74)$$

Again, decreasing the barrier parameter yields the central path, and the limit of the central path, as the barrier parameter approaches zero yields the solution to the inequality constrained problem.

One special case of inequality constrained minimization is when $C = I_r$, the identity matrix of order r . This is the case when there exist non-zero upper and/or lower limits on a choice variable. In the case of a power flow problem, these constraints take the form of (1) limits on flows of energy across power lines, and; (2) minimum and maximum run constraints on generators, both with respect to voltage and power.

Fortunately, we may approach this problem using the tools already in place. Let \bar{x} denote the maximum value the variable x may take. We may transform the inequality constraint $x \leq \bar{x}$ into an equality constraint using the slack variable approach, as follows:

$x + w = \bar{x}$, $w \geq 0$. From there, the method proceeds in the same manner as the general solution method for inequality constrained minimization (*i.e.*, set up problem (2.72), derive FONC (2.73), and set up the Newton-Rhapon step direction (2.74), *etc.*).

Since we have introduced a second variation of the inequality-constrained minimization problem we will present an example for the interested reader.⁷⁷

Example 2.7: Inequality-constrained minimization

Consider the following problem:

$$\min_{\mathbf{x} \in \mathfrak{R}^2} \left\{ (x_1 - 1)^2 + (x_2 - 3)^2 \right\} \text{ s.t. } x_1 - x_2 = 0, \quad x_2 \leq 1.5.$$

Perform one iteration of the primal-dual interior point algorithm, using as the initial guess:

⁷⁷ The reader who is comfortable with the primal-dual interior point algorithm should feel free to move on to the optimal power flow problem.

$$\mu^0 = 1, x_1^0 = 1, x_2^0 = 1, w^0 = 0.5, \lambda^0 = 0, t^0 = 0.5.$$

First, we set up the Lagrangian

$$L = (x_1 - 1)^2 + (x_2 - 3)^2 + t(-\ln w) + \lambda(x_1 - x_2) + \mu(x_2 + w - 1.5)$$

The FONC are:

$$\frac{\partial L}{\partial w} = \frac{-t}{w} + \mu = 0$$

$$\frac{\partial L}{\partial x_1} = 2(x_1 - 1) + \lambda = 0$$

$$\frac{\partial L}{\partial x_2} = 2(x_2 - 3) - \lambda + \mu = 0$$

$$\frac{\partial L}{\partial \lambda} = x_1 - x_2 = 0$$

$$\frac{\partial L}{\partial \mu} = x_2 + w - 1.5 = 0$$

The stage zero Newton-Rhapson update to the barrier problem is thus given by the solution to the following system of equations:

$$\begin{bmatrix} \mu^0 & 0 & 0 & 0 & w^0 \\ 0 & 2 & 0 & 1 & 0 \\ 0 & 0 & 2 & -1 & 1 \\ 0 & 1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta w^0 \\ \Delta x_1^0 \\ \Delta x_2^0 \\ \Delta \lambda^0 \\ \Delta \mu^0 \end{bmatrix} = \begin{bmatrix} -w^0 \mu^0 + t \\ -2(x_1^0 - 1) - \lambda^0 \\ -2(x_2^0 - 3) + \lambda^0 - \mu^0 \\ -x_1^0 + x_2^0 \\ -x_2^0 - w^0 + 1.5 \end{bmatrix},$$

and

$$\begin{bmatrix} w^1 = 0.5 - 0.5 = 0 \\ x_1^1 = 1 + 0.5 = 1.5 \\ x_2^1 = 1 + 0.5 = 1.5 \\ \lambda^1 = 0 - 1 = -1 \\ \mu^1 = 1 + 1 = 2 \end{bmatrix}^{78}$$

⁷⁸ Well, it just goes to show that, while the primal-dual algorithm is very clever, it's not foolproof (that is, sometimes it does take us all the way to the boundary). However, in this example it is clear that we have reached the constrained minimizers x_1^* and x_2^* in one iteration (the objective function falls with increasing values of the two variables until it reaches the unconstrained minimum values of 2.0 each).

At this point, we are (finally) ready to attack AC OPF. The one additional complication that we face in addressing this problem is incorporating nonlinear constraints into the primal-dual algorithm. That is to say, we replace the linear inequality constraints $\{A\mathbf{x} = \mathbf{b}, C\mathbf{x} \leq \mathbf{d}\}$ with the nonlinear constraint sets $\{g(\mathbf{x}) = \mathbf{0}, h(\mathbf{x}) \leq \mathbf{0}\}$ to consider the nonlinear inequality-constrained minimization problem:

$$\min_{\mathbf{x} \in \mathfrak{R}^n} \{f(\mathbf{x}) \mid g(\mathbf{x}) = \mathbf{0}, h(\mathbf{x}) \leq \mathbf{0}\}. \quad (2.75)$$

As we previously remarked, however, once we introduce nonlinear constraint sets, we no longer have a convex minimization problem. We deal with this problem again by choosing functional forms that yield regular points of the constraints $g(\mathbf{x}) = \mathbf{0}$ and $h(\mathbf{x}) \leq \mathbf{0}$. At a regular point of the inequality constraints, linearization of the equality constraints and of the binding inequality constraints will once again yield a useful approximation to the feasible set or its boundary, at least in the vicinity of the regular point.

We state the FONC for minimization of the nonlinear inequality constrained problem as follows:

Theorem 4 *Suppose that the functions $f(\cdot): \mathfrak{R}^n \rightarrow \mathfrak{R}$, $g(\cdot): \mathfrak{R}^n \rightarrow \mathfrak{R}^m$, and $h(\cdot): \mathfrak{R}^n \rightarrow \mathfrak{R}^r$ are partially differentiable with continuous partial derivatives. Let $J(\cdot): \mathfrak{R}^n \rightarrow \mathfrak{R}^{m \times n}$ and $K(\cdot): \mathfrak{R}^n \rightarrow \mathfrak{R}^{r \times n}$ be the Jacobians of g and h , respectively. Consider Problem (2.75):*

$$\min_{\mathbf{x} \in \mathfrak{R}^n} \{f(\mathbf{x}) \mid g(\mathbf{x}) = \mathbf{0}, h(\mathbf{x}) \leq \mathbf{0}\}.$$

Suppose that $\mathbf{x}^ \in \mathfrak{R}^n$ is a regular point of the constraints $g(\mathbf{x}) = \mathbf{0}$ and $h(\mathbf{x}) \leq \mathbf{0}$. If \mathbf{x}^* is a local minimizer of Problem (2.75), then:*

$$\begin{aligned} \exists \boldsymbol{\lambda}^* \in \mathfrak{R}^m, \exists \boldsymbol{\mu}^* \in \mathfrak{R}^r \text{ such that: } & \nabla f(\mathbf{x}^*) + J(\mathbf{x}^*)^T \boldsymbol{\lambda}^* + K(\mathbf{x}^*)^T \boldsymbol{\mu}^* = \mathbf{0}; \\ & M^* h(\mathbf{x}^*) = \mathbf{0}; \\ & g(\mathbf{x}^*) = \mathbf{0}; \\ & h(\mathbf{x}^*) \leq \mathbf{0}; \text{ and} \\ & \boldsymbol{\mu}^* \geq \mathbf{0}. \end{aligned}$$

We again transform the inequality-constrained problem into an equality-constrained problem using slack variables, as follows:

$$\min_{\mathbf{x} \in \mathfrak{R}^n, \mathbf{w} \in \mathfrak{R}^r} \{f(\mathbf{x}) \mid g(\mathbf{x}) = \mathbf{0}, h(\mathbf{x}) + \mathbf{w} = \mathbf{0}, \mathbf{w} \geq \mathbf{0}\}. \quad (2.76)$$

Due to our previously mentioned desire of avoiding the border of the feasible set, we once again add a barrier function $f_b(\mathbf{w}): \mathfrak{R}_{++}^r \rightarrow \mathfrak{R}$ and a barrier parameter $t \in \mathfrak{R}_{++}$ to transform problem (2.76) into the barrier problem:

$$\min_{\mathbf{x} \in \mathfrak{R}^n, \mathbf{w} \in \mathfrak{R}^r} \{\phi(\mathbf{x}, \mathbf{w}) \mid g(\mathbf{x}) = \mathbf{0}, h(\mathbf{x}) + \mathbf{w} = \mathbf{0}, \mathbf{w} > \mathbf{0}\}. \quad (2.77)$$

In review, we minimize $\phi(\mathbf{x}, \mathbf{w})$ over values of $\mathbf{x} \in \mathfrak{R}^n$ and $\mathbf{w} \in \mathfrak{R}^r$ that satisfy $g(\mathbf{x}) = \mathbf{0}$ and $h(\mathbf{x}) + \mathbf{w} = \mathbf{0}$ and are in the interior of $\mathbf{w} \geq \mathbf{0}$. In order to address this problem, we again must assume that the Slater condition holds. As before, we partially ignore the inequality constraints and the domain of the barrier function and solve the following nonlinear equality-constrained problem:

$$\min_{\mathbf{x} \in \mathfrak{R}^n, \mathbf{w} \in \mathfrak{R}^r} \{\phi(\mathbf{x}, \mathbf{w}) \mid g(\mathbf{x}) = \mathbf{0}, h(\mathbf{x}) + \mathbf{w} = \mathbf{0}\}, \quad (2.78)$$

which has first-order necessary conditions:

$$\nabla f(\mathbf{x}^*) + J(\mathbf{x}^*)^T \boldsymbol{\lambda}^* + K(\mathbf{x}^*)^T \boldsymbol{\mu}^* = \mathbf{0}, \quad (2.79a)$$

$$g(\mathbf{x}^*) = \mathbf{0}, \quad (2.79b)$$

$$h(\mathbf{x}^*) + \mathbf{w} = \mathbf{0}, \quad (2.79c)$$

$$t \nabla f_b(\mathbf{w}) + \boldsymbol{\mu} = \mathbf{0}, \quad (2.79d)$$

where $J(\cdot)$ and $K(\cdot)$ are the Jacobians of $g(\cdot)$ and $h(\cdot)$, respectively, and $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are the Lagrange multipliers corresponding to the constraints $g(\mathbf{x}) = \mathbf{0}$ and $h(\mathbf{x}) + \mathbf{w} = \mathbf{0}$, respectively.

2.2 AC Optimal Power Flow

Conceptually, the AC OPF problem is straightforward. We wish to minimize the cost of producing a given amount of electricity⁷⁹ subject to the power flow equality constraints, generator

⁷⁹ Where “given amount” is understood to encompass location-specific electricity demand. “Given amount” may also be interpreted as “optimal amount,” if we assume price-responsive demand.

operating constraints (*i.e.*, minimum and maximum power output),⁸⁰ voltage magnitude constraints,⁸¹ and transmission line capacity constraints. We express this general problem as:

$$\min_{\mathbf{x} \in \mathcal{R}_+^n} \{f(\mathbf{x}) \mid g(\mathbf{x}) = \mathbf{0}, \underline{\mathbf{x}} \leq \mathbf{x} \leq \bar{\mathbf{x}}, \underline{\mathbf{h}} \leq h(\mathbf{x}) \leq \bar{\mathbf{h}}\}. \quad (2.80)$$

Where:

$$f(\mathbf{x}): \mathcal{R}_+^n \rightarrow \mathcal{R}_+$$

The objective, $f(\mathbf{x})$, thus represents the cost of generating (real) power.⁸² Note that $f(\cdot)$ is separable, since the decisions of one generation owner typically do not affect the costs of any other generation owner. The equality constraint system, $g(\mathbf{x}) = \mathbf{0}$ applies to Kirchhoff's current law, applied to each node in the system. The inequality constraint system, $\underline{\mathbf{x}} \leq \mathbf{x} \leq \bar{\mathbf{x}}$, applies to generation constraints. Generators have lower and upper operating bounds on real power output and the range of voltages (generally measured in per unit quantities) at which they may be operated. Finally, the inequality constraint system $\underline{\mathbf{h}} \leq h(\mathbf{x}) \leq \bar{\mathbf{h}}$, applies to transmission line constraints. The function, $h(\mathbf{x})$, represents the fact that flows of power on different transmission lines are functions of the power transfer distribution factors for generators at each node of the system.

With respect to the objective function, Stoft (2002) notes that simplified diagrams of generation supply curves typically assume a constant marginal cost up to the point of maximum generation. This produces jump discontinuities in the market supply curve as we move from one generator's marginal cost to the next.⁸³ Stoft argues that one may "smooth" these point discontinuities by assuming that supply curves have extremely large, but finite, slopes when we

⁸⁰ We will not examine minimum output constraints ("minimum run constraints"), other than to assume that a generator's output cannot be negative.

⁸¹ Which we will not examine.

⁸² Since we are examining the AC power problem, we will bring reactive power back into the discussion. However, the total cost of power generation generally depends only on the entries of \mathbf{x} corresponding to real power generation. See Baldick (2006, p. 595). In some formulations, $f(\cdot)$ will also depend on reactive power generations.

⁸³ See Stoft (2002, Chapter 1-6).

move from one generator to another. This will allow us to assume a twice-continuously differentiable market supply function. Next, let us examine the constraint system $g(\mathbf{x}) = \mathbf{0}$.

Kirchhoff's current law implies that the net flow of power away from a given node, P_ℓ , is equal to the net sum of the flows on all lines connected to that node, as given by the power flow equations, (2.20) and (2.21), which we repeat below:

$$P_\ell = u_\ell^2 G_{\ell\ell} + \sum_{k \in \mathfrak{N}(\ell)} u_\ell u_k [G_{\ell k} \cos(\theta_\ell - \theta_k) + B_{\ell k} \sin(\theta_\ell - \theta_k)],$$

$$Q_\ell = -u_\ell^2 B_{\ell\ell} + \sum_{k \in \mathfrak{N}(\ell)} u_\ell u_k [G_{\ell k} \sin(\theta_\ell - \theta_k) - B_{\ell k} \cos(\theta_\ell - \theta_k)].$$

Since we are working with AC power, we do not linearize these equations, but take them into account as non-linear constraints that must be satisfied in our analysis.

Next, we move on to the equations for real and reactive power flow on line (ℓ, k) , which we repeat here as:

$$\tilde{p}_{\ell k} = (u_\ell)^2 G_{\ell\ell} + u_\ell u_k [G_{\ell k} \cos(\theta_\ell - \theta_k) + B_{\ell k} \sin(\theta_\ell - \theta_k)] \quad (2.81)$$

$$\tilde{q}_{\ell k} = -(u_\ell)^2 B_{\ell\ell} + u_\ell u_k [G_{\ell k} \sin(\theta_\ell - \theta_k) - B_{\ell k} \cos(\theta_\ell - \theta_k)]. \quad (2.82)$$

In the usual case that there is a real power flow limit, $\pm \bar{p}_{\ell k}$ on the line joining two nodes, ℓ and k , we express this constraint as $-\bar{p}_{\ell k} \leq \tilde{p}_{\ell k} \leq \bar{p}_{\ell k}$.⁸⁴ As mentioned above, this is the constraint system $\mathbf{h} \leq h(\mathbf{x}) \leq \bar{\mathbf{h}}$.

At this point, we note that our objective function is typically convex, but because eqs. (2.20) and (2.21) are nonlinear in x , we will generally not have a convex choice set. As Baldick (2006, pp. 599-601) demonstrates, however, it can be shown that the feasible set defined by the constraints:

$$p_\ell(\mathbf{x}) \leq \mathbf{0} \quad (2.83)$$

⁸⁴ Some transmission lines have different flow limits for opposite directional flows, a fact that we will ignore in our analysis.

$$q_\ell(\mathbf{x}) \leq \mathbf{0} \tag{2.84}$$

is convex under the assumption that all voltage magnitudes are constant, provided that

$$|\theta_\ell - \theta_k| \leq 0.1 \text{ radian.}^{85}$$

As one might readily guess, this problem gets complicated really quickly as one examines larger test networks. Therefore, in practically every AC OPF paper you might read, the authors simply present the classical AC OPF formulation, write the Lagrangian of the problem, present first-order conditions, and solve the problem using various computer algorithms.

To keep the problem manageable, we will use a three-node model. Additionally, we will add as few constraints as possible while preserving the flavor of the “power flow” problem.

Example 2.8: 2-node optimal power flow problem.

Let us examine the three-node network shown below:

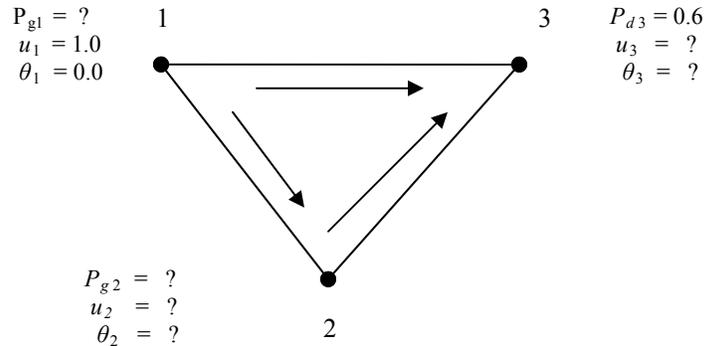


Figure 2: The Three-node Network

Even with just a three-node network, if we include reactive power in the calculations, the problem gets unwieldy, so we focus on real power alone. The transmission line characteristics are given in

Table 2.1 below:

⁸⁵ As Baldick (2006) notes, eqs. (2.83) and (2.84) are equivalent to assuming that power can be “thrown away,” and can be justified by the observation that as long as a generator is not at its minimum operating range, equality can be reestablished by lowering the output of the generator in question. Since we are presenting the AC problem only to make the linearized AC problem easier to understand, we will not go into the proof of this proposition.

Table 2.1 Transmission Line Characteristics

Line	Impedance	Admittance	Real Power Flow Limit
1-2	$0 + j0.001$	$-j1,000$	$-0.3 \leq p_{12} \leq 0.3$
1-3	$0 + j0.002$	$-j500$	$-0.3 \leq p_{13} \leq 0.3$
2-3	$0 + j0.001$	$-j1000$	$-0.4 \leq p_{23} \leq 0.4$

Again, just to keep the problem manageable, we will assume away voltage constraints at nodes 2 and 3, as well as generator operating constraints.

Assume that there are two generators, located at nodes 1 and 2, and let the cost of real power production for the two generators be given by:

$$f_1(P_1) = P_1 + 0.2(P_1)^2 \quad (2.85)$$

$$f_2(P_2) = 1.1P_2 + 0.1(P_2)^2, \quad (2.86)$$

respectively, where P_i , $i = 1, 2$ corresponds to real power generation of the two generators.

Set up the augmented Lagrangian for the OPF problem, using a logarithmic barrier function with barrier parameter $t = 1.0$, and perform one run of the Newton-Rhapson algorithm using initial guesses $P_1^0 = 0.3$, $P_2^0 = 0.3$, $u_2^0 = 1.0$ $\theta_2^0 = 0.0$ $u_3^0 = 1.0$ $\theta_3^0 = 0.0$

Returning to eq. (2.80), we first focus on the constraints, $g(\mathbf{x}) = \mathbf{0}$, and $\underline{\mathbf{h}} \leq h(\mathbf{x}) \leq \bar{\mathbf{h}}$.⁸⁶

The first constraint corresponds to Kirchhoff's current law. This law must hold at all three buses of the network, but since we have only two degrees of freedom, we will express Kirchhoff's law at nodes 2 and 3 only, with node 1 current flow determined residually. Ignoring shunt elements,⁸⁷ Kirchhoff's law applied to nodes 2 and 3 yields:

$$P_2 = u_2 u_1 [G_{21} \cos(\theta_2 - \theta_1) + B_{21} \sin(\theta_2 - \theta_1)] + u_2 u_3 [G_{23} \cos(\theta_2 - \theta_3) + B_{23} \sin(\theta_2 - \theta_3)] + (u_2)^2 G_{22}$$

$$P_3 = u_3 u_1 [G_{31} \cos(\theta_3 - \theta_1) + B_{31} \sin(\theta_3 - \theta_1)] + u_3 u_2 [G_{32} \cos(\theta_3 - \theta_2) + B_{32} \sin(\theta_3 - \theta_2)] + (u_3)^2 G_{33},$$

respectively. Substituting the assumed values for u_1 , G_{ij} , B_{ij} , and θ_1 yields

⁸⁶ Note that we ignore the constraint $\underline{x} \leq x \leq \bar{x}$, which corresponds to the generator's operating range.

⁸⁷ But see Appendix 2 for a discussion of shunt admittances.

$$P_2 = u_2[1000 \cdot \sin \theta_2] + u_2 u_3[1000 \cdot \sin(\theta_2 - \theta_3)] \quad (2.87)$$

$$P_3 = u_3[500 \cdot \sin \theta_3] + u_2 u_3[1000 \cdot \sin(\theta_2 - \theta_3)] \quad (2.88)$$

OK. So far, so good. Moving on to the transmission line constraints, notice that we have three lines, with upper and lower bounds on real power flow, for six constraints. Taking the general formula (eq. 2.81) for real power flow across line (ℓ, k) :

$$\tilde{p}_{\ell k} = (u_\ell)^2 G_{\ell\ell} + u_\ell u_k [G_{\ell k} \cos(\theta_\ell - \theta_k) + B_{\ell k} \sin(\theta_\ell - \theta_k)],$$

and plugging in the parameters corresponding to the three lines yields:

$$\tilde{p}_{12} = u_2[1000 \sin(-\theta_2)] \quad (2.89a)$$

$$\tilde{p}_{13} = u_3[500 \sin(-\theta_3)] \quad (2.89b)$$

$$\tilde{p}_{23} = u_2 u_3[1000 \sin(\theta_2 - \theta_3)]. \quad (2.89c)$$

Since power is being withdrawn at node 3, power will flow along lines (1,3) and (2,3) towards node 3, as reflected in eqs. (2.89b and c), respectively. Because the impedance of line (1,3) is greater than that of line (2,3), our initial guess is that the net power flow on line (1,2) will be from node 1 towards node 2, as reflected in eq. (2.89a). We may thus re-write the line constraints shown in Table 1 as:

$$\begin{aligned} -0.3 &\leq u_2[1000 \sin(-\theta_2)]; & u_2[1000 \sin(-\theta_2)] &\leq 0.3 \\ -0.3 &\leq u_3[500 \sin(-\theta_3)]; & u_3[500 \sin(-\theta_3)] &\leq 0.3 \\ -0.4 &\leq u_2 u_3[1000 \sin(\theta_2 - \theta_3)]; & u_2 u_3[1000 \sin(\theta_2 - \theta_3)] &\leq 0.4 \end{aligned}$$

At this point, we may express the problem as:

$$\min_{P_1, P_2} : P_1 + 0.2(P_1)^2 + 1.1P_2 + 0.1(P_2)^2$$

s.t.

$$\begin{aligned} u_2[1,000 \sin \theta_2] + u_2 u_3[1,000 \sin(\theta_2 - \theta_3)] - P_2 &= 0 \\ u_3[1,000 \sin \theta_3] + u_2 u_3[1,000 \sin(\theta_2 - \theta_3)] + 0.6 &= 0 \\ 0.6 - P_1 - P_2 &= 0 \\ -0.3 &\leq u_2[1,000 \sin(-\theta_2)] \\ u_2[1,000 \sin(-\theta_2)] &\leq 0.3 \\ -0.3 &\leq u_3[500 \sin(-\theta_3)] \\ u_3[500 \sin(-\theta_3)] &\leq 0.3 \\ -0.4 &\leq u_2 u_3[1,000 \sin(\theta_2 - \theta_3)] \\ u_2 u_3[1,000 \sin(\theta_2 - \theta_3)] &\leq 0.4 \end{aligned}$$

As we have seen in the case of inequality constraints, each of these six equations will have an associated slack variable, turning the inequality constraint into an equality constraint. Also, the since we want the slack variable to be bounded away from zero, we will attach a barrier function to the slack variable. This motivates the augmented Lagrangian:

$$\begin{aligned} L = & P_1 + 0.2(P_1)^2 + 1.1P_2 + 0.1(P_2)^2 + t(-\ln P_1) + t(-\ln P_2) + t(-\ln w_1) + t(-\ln w_2) + t(-\ln w_3) \\ & + t(-\ln w_4) + t(-\ln w_5) + t(-\ln w_6) + \lambda_1 \{u_2[1,000 \sin \theta_2] + u_2 u_3[1,000 \sin(\theta_2 - \theta_3)] - P_2\} \\ & + \lambda_2 \{u_3[1,000 \sin \theta_3] + u_2 u_3[1,000 \sin(\theta_2 - \theta_3)] + 0.6\} + \lambda_3 \{0.6 - P_1 - P_2\} \\ & + \lambda_4 \{u_2[1,000 \sin(-\theta_2)] - w_1 + 0.3\} + \lambda_5 \{u_2[1,000 \sin(-\theta_2)] + w_2 - 0.3\} \\ & + \lambda_6 \{u_3[500 \sin(-\theta_3)] - w_3 + 0.3\} + \lambda_7 \{u_3[500 \sin(-\theta_3)] + w_4 - 0.3\} \\ & + \lambda_8 \{u_2 u_3[1,000 \sin(\theta_2 - \theta_3)] - w_5 + 0.4\} + \lambda_9 \{u_2 u_3[1,000 \sin(\theta_2 - \theta_3)] + w_6 - 0.4\}. \end{aligned}$$

The rest of the problem is left as an exercise to the intrepid reader.⁸⁸

In any event, we now see that the AC OPF formulation is quite involved, and its solution is best left to the computer. It is our hope and expectation, though, that the principles contained herein will be of great help in our understanding of the DC approximation to the AC OPF problem.

⁸⁸ Had I really known what I was getting into, I may have just skipped this section, anyway, because this whole exercise is getting unruly.

References

- Baldick R (2006) Applied Optimization: Formulation and Algorithms for Engineering Systems, Cambridge University Press, New York
- Luenberger D (1984) Linear and Nonlinear Programming, 2nd Edition. Addison-Wesley Publishing Company, Reading, MA.
- Nash S, Stofer A (1996) Linear and Nonlinear Programming. McGraw-Hill, Inc., New York, NY.
- Stoft S (2002) Power System Economics. Wiley-Interscience, Piscataway, NJ.

Appendix: Shunt Elements

While it is usual to ignore shunt elements in DC analysis, the same is not true of AC, so we examine this topic here. As mentioned previously, shunt elements ground overhead lines. We will now explore two properties of shunt elements: conductance and capacitance. As per Grainger and Stevenson (1994), conductance exists between conductors or between conductors and the ground. Conductance accounts for current leakage at insulators of overhead lines and the insulation of cables. Since leakage at insulators of overhead lines is trivial, the conductance between conductors of an overhead line is usually neglected. Another reason electric engineers

generally ignore conductance is that leakage at insulators, the principal source of conductance, changes considerably with atmospheric conditions and with the conducting properties of dirt that collects on the insulators. Another effect, corona, which causes line leakages, also changes appreciably with the weather. Since conductance also has a minimal effect on shunt lines, electrical engineers will generally ignore its effect on shunt admittance.

Capacitance of a transmission line results from the potential difference between conductors. Capacitance causes conductors to be charged in the same manner as the plates of a capacitor when there is a potential difference between them. Capacitance between parallel conductors is a constant depending on the size and spacing of the conductors. For lines less than roughly 50 miles (80 km) long, the effect of capacitance is small, and is often neglected.

Without going into the technical details,⁸⁹ the capacitance of a transmission line is an increasing function of the (positive) charge of q (Coulombs/meter) and inversely related to the voltage drop along the line (v). In the most basic example, we consider a transmission line composed of two parallel wires. In this case, capacitance is defined as the charge on the conductors (wires) per unit of potential difference between them. That is:

$$C = \frac{q}{v} \text{ F / m ,} \tag{2A.1}$$

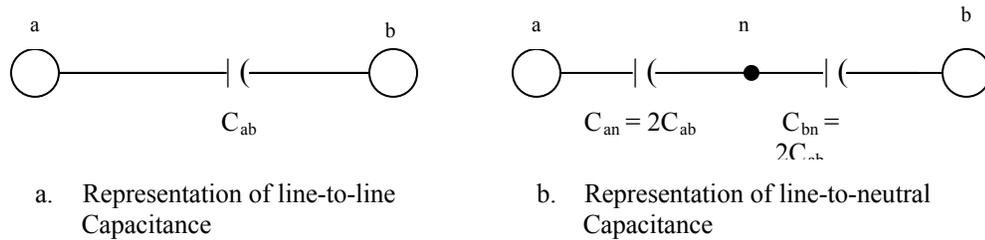
that is, Farads per meter. For a two-wire line, we calculate the voltage drop v_{ab} between the two wires by computing the voltage drop due to the charge q_a on conductor a and the voltage drop due to the charge q_b on conductor b . The voltage drop from one conductor a to conductor b due to the charges on both conductors is the sum of the voltage drops caused by each charge alone.

This last point is central to our (simplified) analysis. For a two-wire line, we have that $q_a = -q_b$.⁹⁰ The other determinant of voltage induced by a given charge is the radius of the wire. Assuming that the two wires have the same radius, it is therefore the case that $|v_{ab}| = |2v_a| = |2v_b|$.

⁸⁹ See, e.g. Grainger and Stevenson (1994, Ch. 5)

⁹⁰ Ask an EE friend why this holds.

Since the ground has voltage = 0 V, by definition, we have the voltage difference between each conductor (of a two-wire shunt element) as $v_s = v_a = v_b = \frac{1}{2}v_{ab}$. Referring to Eq. (2A.1), we then have that capacitance to ground, or capacitance to neutral, is equal to twice the capacitance between the two conductors. That is, $C_n = 2C_{ab}$. We represent these concepts diagrammatically as follows:



Looking ahead to the optimal power flow problem, we will encounter both equality constraints (in terms of Kirchhoff's laws) and inequality constraints (generation and transmission line constraints). Fortunately, moving from the linear non-negativity problem with equality constraints to the linear non-negativity problem with equality and inequality constraints is straightforward.