Settlement

1.0 Settlement without congestion

Let's consider the base case solution obtained from the notes called LPOPF2. How would the suppliers and the loads be paid?

To answer this question, we repeat here the solution in terms of the one-line diagram and in terms of the table of Lagrange multipliers.

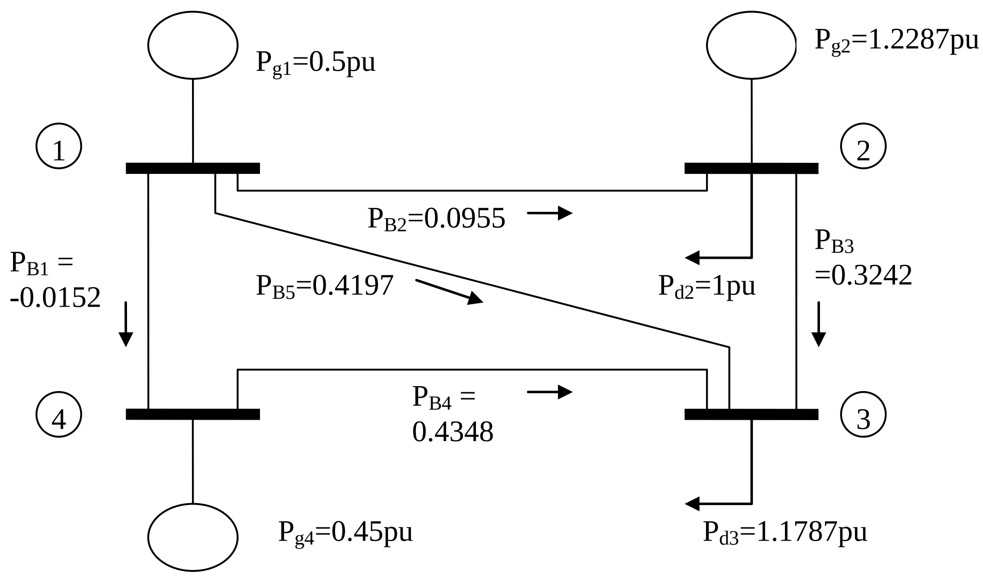


Fig. 1: Result in terms of generation levels and flows for base case

The objective function is Z=2705.8 \$/hr.

Table 2: Lagrange multipliers for P_{d2}=1.0, P_{d3}=1.1787 and infinite transmission capacity ($/per unit-hr)

Equality constraints		Lower bounds		Upper bounds	
Equation	Value*10^3	Variable	value	variable	value
P_{B1}	0.0000	P_{g1}	96.0000	P_{g1}	0.0000
P_{B2}	0.0000	P_{g2}	0	P_{g2}	0.0000
P_{B3}	0.0000	P_{g4}	43.0000	P_{g4}	0.0000
P_{B4}	0.0000	P_{B1}	0	P_{B1}	0.0000
P_{B5}	0.0000	P_{B2}	0	P_{B2}	0.0000
P_1	1.2110	P_{B3}	0	P_{B3}	0.0000
P_2	1.2110	P_{B4}	0	P_{B4}	0.0000
P_3	1.2110	P_{B5}	0	P_{B5}	0.0000
P_4	1.2110	θ_1	0	θ_1	0.0000
		θ_2	0	θ_2	0.0000
		θ_3	0	θ_3	0.0000
		θ_4	0	θ_4	0.0000

The settlement for this case would occur like this:
Amount paid to generators:

$$Payment_{g1} = P_{g1} \times LMP_1 = 50MW \times 12.11\$/MWhr = 605.50\$/hr$$

$$Payment_{g2} = P_{g2} \times LMP_2 = 122.87MW \times 12.11\$/MWhr = 1487.96\$/hr$$

$$Payment_{g4} = P_{g4} \times LMP_4 = 45MW \times 12.11\$/MWhr = 544.95\$/hr$$

The total payments to the generators will be 605.50+1487.96+544.95=2638.41$/hr.

Now what do the loads have to pay?

$$Payment_{d2} = P_{d2} \times LMP_2 = 100MW \times 12.11\$/MWhr = 1211.00\$/hr$$

$$Payment_{d3} = P_{d3} \times LMP_2 = 117.87MW \times 12.11\$/MWhr = 1427.41\$/hr$$

The total payments by the loads will be 1211.00+1427.41=2638.41$/hr, and so we see that the market settles with total payment to the generators equaling total payment to the loads.

Question: Why does this differ from the objective function of 2705.80 $/hr?

Answer: We *optimize* on the offers. We *settle* at the LMPs.

➔The bus *k* LMP is the change in the objective function for increasing the load at bus *k* by a unit. It is determined by the least expensive regulating generator. So we are paying generators at the offer of generator 2.

You can see this clearly by recomputing the total payment if we paid each generator according to the offers they make: In this case, it would be

$$Payment_{g1} = P_{g1} \times s_1 = 50MW \times 13.07\$/MWhr = 653.50\$/hr$$

$$Payment_{g2} = P_{g2} \times s_2 = 122.87MW \times 12.11\$/MWhr = 1487.96\$/hr$$

$$Payment_{g4} = P_{g4} \times s_4 = 45MW \times 12.54\$/MWhr = 564.30\$/hr$$

In this case, if we paid according to the offers, the total payments to the generators will be 653.50+1487.96+564.30=2705.76$/hr, which agrees with the value of the objective function (there is a little round-off error).

So why do we settle at the LMPs rather than the offers? According to the paper I placed on the web page [i, pg. 26],

> "The primary reason for this conclusion is that under the pay-as-bid settlement scheme, market participants would bid substantially higher than their marginal costs (since there is no incentive for participants to bid their operating cost) to try to increase their revenue and, thus, offset and very likely exceed the expected consumer payment reduction. As a result, currently all ISOs in the United States adopt the pay-at-MCP principle."

Note to make a comment regarding what happens when you hit lower limit and constrain-on units as in this example. Otherwise, we should have objective function=sum of payments.

In other words,

- A pay-as-bid settlement scheme incentivizes participants to bid high since the bid is what they will be paid if their bid is accepted. The disincentive to bidding high is that their bid might not be accepted.
- A pay-at MCP settlement scheme provides no incentive to bid high. The disincentive to bid high because their bid might not be accepted remains.

2.0 Settlement with congestion

Now let's consider the case 2 solution obtained from the notes called LPOPF. This is the case where line 3 was congested. How would the suppliers and the loads be paid?

To answer this question, we repeat here the solution in terms of the one-line diagram and in terms of the table of Lagrange multipliers.

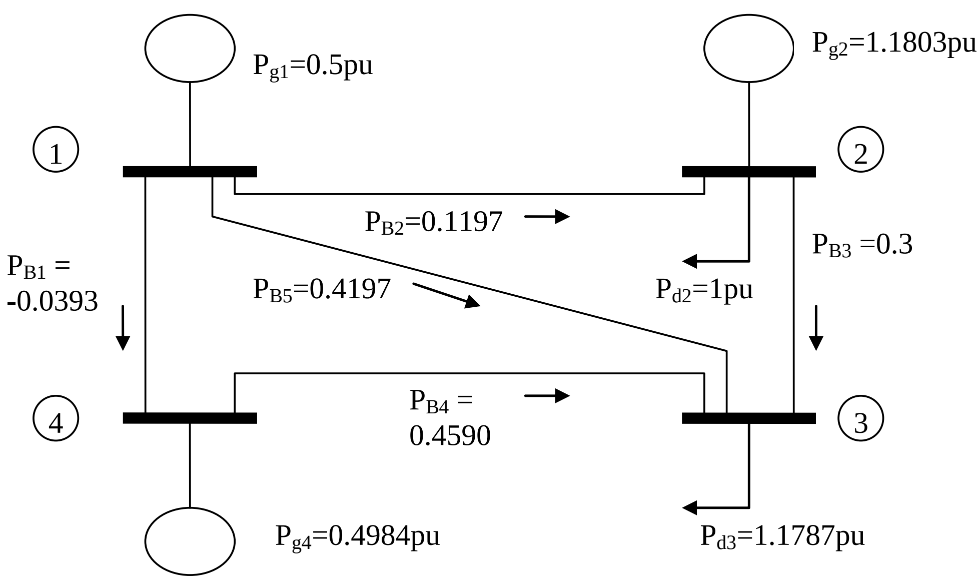


Fig. 6: Cases 2 flows

Table 3: Lagrange multipliers for P_{d2}=1.0, P_{d3}=1.1787 and infinite transmission capacity except for and 0.3 capacity constraint on P_{B3} ($/per unit-hr)

Equality constraints		Lower bounds		Upper bounds	
Equation	Value*10^3	Variable	value	variable	value
P_{B1}	0.0000	P_{g1}	63.7500	P_{g1}	0.0000
P_{B2}	0.0000	P_{g2}	0.0000	P_{g2}	0.0000
P_{B3}	0.0860	P_{g4}	0.0000	P_{g4}	0.0000
P_{B4}	0.0000	P_{B1}	0.0000	P_{B1}	0.0000
P_{B5}	0.0000	P_{B2}	0.0000	P_{B2}	0.0000
P_1	1.2432	P_{B3}	0.0000	P_{B3}	86.0000
P_2	1.2110	P_{B4}	0.0000	P_{B4}	0.0000
P_3	1.2647	P_{B5}	0.0000	P_{B5}	0.0000
P_4	1.2540	θ_1	0.0000	θ_1	0.0000
		θ_2	0.0000	θ_2	0.0000
		θ_3	0.0000	θ_3	0.0000
		θ_4	0.0000	θ_4	0.0000

The settlement for this case would occur like this:
Amount paid to generators:

$$Payment_{g1} = P_{g1} \times LMP_1 = 50MW \times 12.432\$/MWhr = 621.60\$/hr$$

$$Payment_{g2} = P_{g2} \times LMP_2 = 118.03MW \times 12.11\$/MWhr = 1429.34\$/hr$$

$$Payment_{g4} = P_{g4} \times LMP_4 = 49.84MW \times 12.54\$/MWhr = 624.99\$/hr$$

The total payments to the generators will be 621.60+1429.34+624.99=2675.93$/hr.

Now what do the loads have to pay?

$$Payment_{d2} = P_{d2} \times LMP_2 = 100MW \times 12.11\$/MWhr = 1211.00\$/hr$$

$$Payment_{d3} = P_{d3} \times LMP_2 = 117.87MW \times 12.647\$/MWhr = 1490.70\$/hr$$

The total payments by the loads will be 1211.00+1490.70=2701.70$/hr.

Notice: The amount paid by the loads exceeds that paid to the generators by 2701.7-2675.93=25.77$/hr.

Why is this?

This is due to the congestion charges, denoted as CC and given by:

$$CC = \sum_{j=1}^{M} \mu_j P_{bj}$$

In our example, since we have only one congested line, this is:

$$CC = \sum_{j=1}^{M} \mu_j P_{bj} = 86 * 0.3 = 25.8$$

Note that the units of μ_j are $/per-unit hr and the units of P_{bj} are per unit, and so the units of CC are $/hr. Congestion charges are allocated by the market operator to holders of financial transmission rights (FTRs). We will discuss more about FTRs when we talk about (a) risk management and (b) planning.

Comparison to the difference between payment to the generators and payment by the loads, 25.77$/hr, indicates that this accounts for it (within roundoff error).

Thus we are led to conclude that

$$CC = \sum_{j=1}^{M} \mu_j P_{bj} = \sum_{k \in load} LMP_j * P_{dk} - \sum_{k \in gen} LMP_k * P_{gk}$$

Let's try to prove this.

We show in the Appendix A (see eq. 26) that the LMPs at each bus are given by:

$$k \in load: \quad LMP_k = \lambda \quad \text{Energy component}$$
$$+\lambda \frac{\partial P_{loss}}{\partial P_{dk}} \quad \text{Loss component}$$
$$+\sum_{j=1}^{M} \mu_j t_{jk} \quad \text{Congestion component}$$

If we ignore the loss component, then LMP's are given by:

$$LMP_k = \lambda + \sum_{j=1}^{M} \mu_j t_{jk}$$

where t_{jk} are called *shift factors* and give the change in flow on circuit *j* to a change in real power injection at bus *k*, under a specified slack distribution, according to

$$t_{jk} = \frac{\Delta F_j}{\Delta P_k} \tag{1}$$

If the network is linear over its entire operating range, then (1) applies even when

$$\Delta F_j = F_j - 0, \quad \Delta P_k = P_k - 0 \tag{2}$$

so that

$$t_{jk} = \frac{F_j}{P_k} \tag{3}$$

or

$$F_j = t_{jk} P_k \tag{4}$$

In matrix form, (4) becomes:

$$\underline{F} = \underline{T}\,\underline{P} \tag{5}$$

We show in Appendix B how to compute shift factors.

Then the congestion charge is: (Note there is a minus sign error on the below somewhere, need to find it).

$$CC = \sum_{k=1}^{N} LMP_k * P_{gk} - \sum_{k=1}^{N} LMP_k * P_{dk}$$

$$= \sum_{k=1}^{N} \left(\lambda + \sum_{j=1}^{M} \mu_j t_{jk} \right) * P_{gk} - \sum_{k=1}^{N} \left(\lambda + \sum_{j=1}^{M} \mu_j t_{jk} \right) * P_{dk}$$

$$= \sum_{k=1}^{N} \left[\left(\lambda + \sum_{j=1}^{M} \mu_j t_{jk} \right) * P_{gk} - \left(\lambda + \sum_{j=1}^{M} \mu_j t_{jk} \right) * P_{dk} \right]$$

$$= \sum_{k=1}^{N} \left[\left(\lambda + \sum_{j=1}^{M} \mu_j t_{jk} \right) \left(P_{gk} - P_{dk} \right) \right]$$

$$= \sum_{k=1}^{N} \left[\lambda \left(P_{gk} - P_{dk} \right) + \sum_{j=1}^{M} \mu_j t_{jk} \left(P_{gk} - P_{dk} \right) \right]$$

$$= \sum_{k=1}^{N} \lambda \left(P_{gk} - P_{dk} \right) + \sum_{k=1}^{N} \sum_{j=1}^{M} \mu_j t_{jk} \left(P_{gk} - P_{dk} \right)$$

$$= \lambda \sum_{k=1}^{N} \left(P_{gk} - P_{dk} \right) + \sum_{k=1}^{N} \sum_{j=1}^{M} \mu_j t_{jk} \left(P_{gk} - P_{dk} \right)$$

$$= 0 + \sum_{k=1}^{N} \sum_{j=1}^{M} \mu_j t_{jk} \left(P_{gk} - P_{dk} \right)$$

Now interchange the summation to obtain:

$$CC = \sum_{j=1}^{M} \sum_{k=1}^{N} \mu_j t_{jk} \left(P_{gk} - P_{dk} \right)$$

$$= \sum_{j=1}^{M} \mu_j \sum_{k=1}^{N} t_{jk} \left(P_{gk} - P_{dk} \right)$$

$$= \sum_{j=1}^{M} \mu_j \sum_{k=1}^{N} t_{jk} P_k = \sum_{j=1}^{M} \mu_j F_j$$

Appendix A (LMPs)

1.0 Objective function

We make two simplifications on the objective function:

- We assume that demand bids are not price sensitive, i.e., that demand is fixed, independent of the price. Therefore, the problem of maximizing the social surplus becomes one of maximizing the producer's surplus, which is the same as minimizing the producer's cost. So our objective function will be to minimize producers's cost.
- We assume that producers cost is expressed linearly as a function of their generation. You can think about this in two different ways.
    - The producers represent their cost functions using a piecewise linear approximation, or
    - The producers are simply making offers of certain quantities, at fixed prices.

In either case, we may represent the objective function as

$$G(\underline{P}_g) = \sum_{k=1}^{N} s_k P_{gk} \tag{1}$$

where s_k are the \$/MWhr offers being made on an amount of generation of P_{gk} over 1 hour, and there are N generators making such offers. If bus k is a pure load bus, then P_{gk}=s_k=0.

2.0 Power balance

If our analysis is based on linearized network representation, then it is implied that resistance has been assumed zero. Therefore, losses should be zero, and the power balance equation would be

$$\sum_{k=1}^{N} P_{gk} - P_{dk} = \sum_{k=1}^{N} P_k = 0 \tag{2}$$

That left-hand summation of (2), when placed equal to 0, says that the sum of generation is exactly equal to the sum of demand. The right-hand summation of (2), when placed equal to 0, says the sum of injections is exactly equal to zero.

However, we will, for the moment, give a more general relation that accounts for losses, i.e.,

$$\sum_{k=1}^{N} P_{gk} - P_{dk} = P_{loss} \qquad (3)$$

3.0 Line flow constraints

Regarding network representation, we assume that we have a constraint sensitivity matrix $\underline{T}$ with elements t_{jk} that give the change in flow on circuit j to a change in real power injection at bus k, under a specified slack distribution, according to

$$t_{jk} = \frac{\Delta F_j}{\Delta P_k} \qquad (4)$$

If the network is linear over its entire operating range, then (4) applies even when

$$\Delta F_j = F_j - 0, \quad \Delta P_k = P_k - 0 \qquad (5)$$

so that

$$t_{jk} = \frac{F_j}{P_k} \qquad (6)$$

or

$$F_j = t_{jk} P_k \qquad (7)$$

In matrix form, (7) becomes:

$$\underline{F} = \underline{T}\,\underline{P} \qquad (8)$$

It is important to note that the sensitivity factors of (6) are computed under a so-called "slack-bus" assumption which indicates how the change ΔP_k is assumed to be compensated. It could be compensated from another certain bus, or from several other buses, or from all other buses, and the elements of T will change depending on which of these is assumed. It is generally considered best to employ a so-called *distributed slack bus assumption* here where the compensation is assumed to come from all other generator buses.

Given that we know the “normal” flow constraints on every circuit, then

$$-\underline{F}_{\max} \leq \underline{F} \leq \underline{F}_{\max} \tag{9}$$

Substitution of (8) into (9) results in

$$-\underline{F}_{\max} \leq \underline{T}\,\underline{P} \leq \underline{F}_{\max} \tag{10}$$

We assume at this point that high flows in our network are unidirectional, i.e., we need not be concerned with high flows in both directions. This does not prevent bidirectional flows, it merely enables us to be concerned with reaching the upper bound in only one direction. Therefore, we may ignore the lower bound in (10) so that our circuit flow constraint is

$$\underline{T}\,\underline{P} \leq \underline{F}_{\max} \tag{11}$$

In scalar form, (11) is

$$\sum_{k=1}^{N} t_{jk} P_k \leq F_{j\max},\ j = 1,\ldots,M \tag{12}$$

and replacing injection with difference between generation and load, we obtain:

$$\sum_{k=1}^{N} t_{jk} (P_{gk} - P_{dk}) \leq F_{j\max},\ j = 1,\ldots,M \tag{13}$$

4.0 Optimization problem for LOPF-1

A linearized optimal power flow (OPF) problem minimizes (1) subject to (3) and (12), that is,

$$\min \quad G(\underline{P}) = \sum_{k=1}^{N} s_k P_{gk}$$

$$s.t.$$

$$\sum_{k=1}^{N} P_{gk} - P_{dk} = P_{loss}$$

$$\sum_{k=1}^{N} t_{jk}(P_{gk} - P_{dk}) \le F_{j\max}, \; j = 1,\ldots,M \tag{14}$$

We will call (14) LOPF-1.

The Lagrangian function for LOPF-1 is

$$L(\underline{P}_g, \lambda, \underline{\mu}) = \sum_{k=1}^{N} s_k P_{gk} - \lambda\left[\sum_{k=1}^{N} P_{gk} - P_{dk} - P_{loss}\right] - \sum_{j=1}^{M} \mu_j \left[\sum_{k=1}^{N} t_{jk}(P_{gk} - P_{dk}) - F_{j\max}\right] \tag{15}$$

The first order conditions for finding the optimum to LOPF-1 include:

$$k \in gen: \qquad \frac{\partial L}{\partial P_{gk}} = s_k - \lambda\left(1 - \frac{\partial P_{loss}}{\partial P_{gk}}\right) - \sum_{j=1}^{M} \mu_j t_{jk} = 0 \tag{16}$$

But we are more interested in the load buses. Consider

$$k \in load: \qquad \frac{\partial L}{\partial P_{dk}} = \lambda\left(1 + \frac{\partial P_{loss}}{\partial P_{dk}}\right) + \sum_{j=1}^{M} \mu_j t_{jk} \tag{17}$$

Note carefully that P_{dk} is not a decision variable, and therefore we do not set it equal to 0.

Let's consider (17). What is this? To answer this question, we need to learn a theorem.

5.0 Envelope theorem

Consider the following optimization problem.

$$\max_{x} f(x,\theta)$$
$$s.t. \quad g(x,\theta) \le 0 \tag{18}$$

where x is the decision variable and θ is some parameter that is influential in the problem, but it is not a decision variable, i.e., we may not select its value. We desire to find how the optimal value of f changes with respect to θ.

Let's give a name to the optimal value of f. Let's call it V; it is a function of θ. That is,

$$V(\theta) = f(x^*(\theta),\theta) \tag{19}$$

Then what we are trying to find is

$$\frac{\partial V(\theta)}{\partial \theta} \tag{20}$$

Note that V will change both because θ affects f and because it also affects the optimal choice of x.

The Lagrangian function is

$$L(x,\theta,\lambda) = f(x,\theta) - \lambda g(x,\theta) \tag{21}$$

<u>Envelope theorem</u>: The total rate of change in the optimal value of the objective function due to a small change in the parameter θ is the rate of change in the Lagrangian L evaluated at the optimal value of x. That is,

$$\frac{\partial V(\theta)}{\partial \theta} = \left.\frac{\partial L(x(\theta),\theta,\lambda)}{\partial \theta}\right|_{x=x^*} \tag{22}$$

The proof requires several lines of calculus that we omit here.

6.0 Locational marginal price

Need to review following paper to consider updating this part.

T. Organogianni and G. Gross, "A General Formulation for LMP Evaluation," IEEE Trans. On Power Systems, Vol 22, No 3, Aug 2007.

Armed with the envelope theorem, we may now identify the meaning to (17), which is repeated here for convenience:

$$k \in load: \qquad \frac{\partial L}{\partial P_{dk}} = \lambda(1+\frac{\partial P_{loss}}{\partial P_{dk}}) + \sum_{j=1}^{M} \mu_j t_{jk} \tag{17}$$

Equation (17) gives the change in the optimal value of the objective function due to a small change in the parameter P_{dk}.

In other words, if we solve the optimization problem with $P_{dk}=P_{dk0}$, obtaining $G^*(P_{dk0})$, and then resolve the optimization problem with $P_{dk}=P_{dk0}+1$, obtaining $G^*(P_{dk0}+1)$, then

$$G^*(P_{dk0}+1) - G^*(P_{dk0}) = \frac{\partial L}{\partial P_{dk}} \tag{23}$$

We call $\frac{\partial L}{\partial P_{dk}}$ the LMP for bus k, that is,

$$k \in load: \qquad LMP_k = \lambda(1+\frac{\partial P_{loss}}{\partial P_{dk}}) + \sum_{j=1}^{M} \mu_j t_{jk} \tag{24}$$

Written slightly different, it is

$$k \in load: \qquad LMP_k = \lambda + \lambda\frac{\partial P_{loss}}{\partial P_{dk}} + \sum_{j=1}^{M} \mu_j t_{jk} \tag{25}$$

And (25) show us a very useful way to think about LMPs. They consist of three components:

$$\begin{aligned} k \in load: \qquad LMP_k &= \lambda && \text{Energy component} \\ &+ \lambda\frac{\partial P_{loss}}{\partial P_{dk}} && \text{Loss component} \\ &+ \sum_{j=1}^{M} \mu_j t_{jk} && \text{Congestion component} \end{aligned} \tag{26}$$

We discuss each one of these terms in what follows.

7.0 Energy component

We are considering the components of the LMP at a particular bus *k*. The first component is the energy component, represented by λ.

To gain better understanding of exactly what this is, we will neglect losses in our original formulation (14), resulting in

$$\begin{aligned} &\min \quad G(\underline{P}) = \sum_{k=1}^{N} s_k P_{gk} \\ &s.t. \\ &\sum_{k=1}^{N} P_{gk} - P_{dk} = 0 \\ &\sum_{k=1}^{N} t_{jk}(P_{gk} - P_{dk}) \le F_{j\max},\ j = 1,\ldots,M \end{aligned} \tag{27}$$

Rewriting the equality constraint in (27) so that the function of decision variables is on the left-hand-side and constants on the right-hand-side, we have

$$\begin{aligned} &\min \quad G(\underline{P}) = \sum_{k=1}^{N} s_k P_{gk} \\ &s.t. \\ &\sum_{k=1}^{N} P_{gk} = \sum_{k=1}^{N} P_{dk} = P_{D,tot} \\ &\sum_{k=1}^{N} t_{jk}(P_{gk} - P_{dk}) \le F_{j\max},\ j = 1,\ldots,M \end{aligned} \tag{28}$$

Now write the Lagrangian function:

$$L(\underline{P}_g, \lambda, \underline{\mu}) = \sum_{k=1}^{N} s_k P_{gk} - \lambda \left[\sum_{k=1}^{N} P_{gk} - \sum_{k=1}^{N} P_{dk} \right] - \sum_{j=1}^{M} \mu_j \left[\sum_{k=1}^{N} t_{jk} (P_{gk} - P_{dk}) - F_{j\max} \right] \tag{29}$$

or

$$L(\underline{P}_g, \lambda, \underline{\mu}) = \sum_{k=1}^{N} s_k P_{gk} - \lambda \left[\sum_{k=1}^{N} P_{gk} - P_{D,tot} \right] - \sum_{j=1}^{M} \mu_j \left[\sum_{k=1}^{N} t_{jk} (P_{gk} - P_{dk}) - F_{j\max} \right] \tag{30}$$

Notice that λ is the Lagrange multiplier (or dual variable) on the power balance equality constraint. This immediately gives us an interpretation of λ.

> ➔The energy component λ of the LMP is the increase in the objective function (in this case, cost per hour) if demand $P_{D,tot}$ increases by 1 unit.

Without losses, the LMP expression becomes (from (29)):

$$k \in load: \qquad LMP_k = \frac{\partial L}{\partial P_{dk}} = \lambda + \sum_{j=1}^{M} \mu_j t_{jk} \tag{31}$$

The summation is the congestion component. If there is no congestion, then

$$k \in load: \qquad LMP_k = \lambda \tag{32}$$

Equation (32) makes the interesting point that, if we ignore losses, and if there is no congestion, then the LMP will equal to λ, *and this will be true for every load bus in the network*.

One last comment here. It is worthwhile to identify what determines λ. We may gain insight to this via the first order condition (16) which, without losses, becomes

$$k \in gen: \qquad \frac{\partial L}{\partial P_{gk}} = s_k - \lambda - \sum_{j=1}^{M} \mu_j t_{jk} = 0 \tag{33}$$

Solving for λ, we obtain:

$$k \in gen: \quad \lambda = s_k - \sum_{j=1}^{M} \mu_j t_{jk} \tag{34}$$

Under the condition of no congestion, then

$$k \in gen: \quad \lambda = s_k \tag{35}$$

What does this mean?...

To understand what this means, it is important to understand that P_{gk} for which we differentiate to obtain (35) must be "regulating," i.e., it cannot be at its limit. We could have exposed this idea more clearly by including constraints on P_{gk} in the optimization problem formulation, in which case we would have obtained corresponding terms in the objective that would have vanished for regulating units and would have contributed for non-regulating units.

Now consider how an electricity market works. Each generation owner offers in their s_k with a corresponding range. The algorithm selects the lowest offer, and takes the full range of that offer, and then selects the next lowest offer, and then the next, and so on until the demand is met. Figure 1 illustrates.

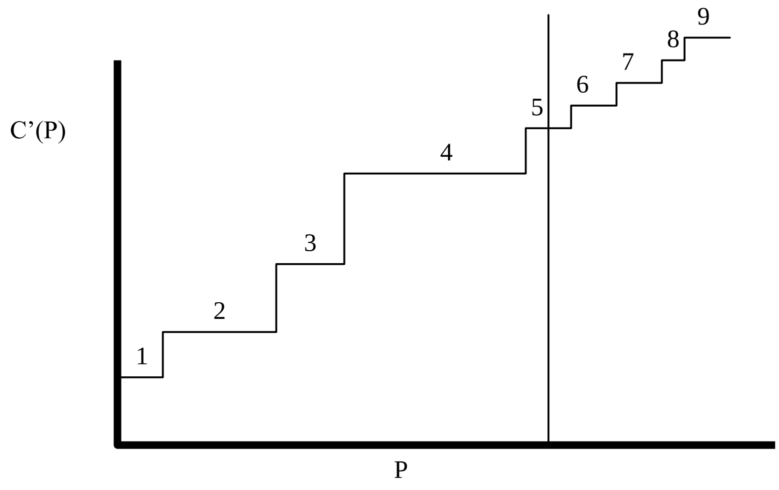


Fig. 1

The only unit that is selected, and is regulating, is unit 5. This is the unit for which $\lambda=s_k$. It is the unit that will pick up the extra demand when the demand is increased by 1 unit. We say that unit 5 is "on the margin."

8.0 Loss component

Consider the expression for LMP again, from (25)

$$k \in load: \qquad LMP_k = \lambda + \lambda \frac{\partial P_{loss}}{\partial P_{dk}} + \sum_{j=1}^{M} \mu_j t_{jk} \tag{25}$$

Assuming no congestion, we have

$$k \in load: \qquad LMP_k = \lambda + \lambda \frac{\partial P_{loss}}{\partial P_{dk}} \tag{36}$$

When we increase the demand at bus *k* by one unit, the losses will increase due to more current flowing through the network. Therefore the term $\frac{\partial P_{loss}}{\partial P_{dk}}$ will be positive. This results in each bus seeing a higher LMP than that set by the energy component λ.

For a particular bus *k*, the increase in LMP_k beyond λ will depend on how an increase in that buses demand P_{dk} would be compensated. The way it would *really* be compensated is that the marginal unit would increase its generation. This would require $\frac{\partial P_{loss}}{\partial P_{dk}}$ to be recomputed each time the marginal unit changes which is very frequent. What is really done is that $\frac{\partial P_{loss}}{\partial P_{dk}}$ is computed for each bus under an assumed compensation strategy. For example, reference [ii][1] shows how to compute $\frac{\partial P_{loss}}{\partial P_{dk}}$ relative to a distributed slack bus reference. We will not cover this but will simply assume the availability of $\frac{\partial P_{loss}}{\partial P_{dk}}$.

[1] I have placed this reference on the web page. It is an excellent paper on LMP calculation.

9.0 Congestion component

Finally, we reconsider the expression for LMP once again, from (25)

$$k \in load: \qquad LMP_k = \lambda + \lambda \frac{\partial P_{loss}}{\partial P_{dk}} + \sum_{j=1}^{M} \mu_j t_{jk} \tag{25}$$

At this point, our interest is the last term. Let's ignore the losses, resulting in

$$k \in load: \qquad LMP_k = \lambda + \sum_{j=1}^{M} \mu_j t_{jk} \tag{37}$$

The summation in (37) will contain zero terms for all circuits j for which flow is not at the rating, i.e., the only non-zero terms in the summation will be for circuits that are at their rating, i.e., that are *congested*. Let's consider that there is only one such circuit in the network, circuit 5. Then

$$k \in load: \qquad LMP_k = \lambda + \mu_5 t_{5k} \tag{38}$$

The Lagrange multiplier (dual variable) μ_5 is on the flow constraint for circuit 5, and it will always be nonnegative. On the other hand, t_{5k}, the generation shift factor, representing the change in flow on circuit 5 for an increase in injection at bus k, may be positive or negative. Thus we see that congestion, although usually increasing LMPs for most buses, can also decrease LMPs under certain conditions.

We will study the effects of congestion on LMPs in some depth in the next set of notes.

Appendix B (Sensitivities)

1.0 Introduction

Operation of the Eastern Interconnection has become heavily reliant on using the so-called Interchange Distribution Calculator (IDC). This is an internet-accessed system that interfaces with OASIS and allows market participants and network operators to efficiently, but approximately, determine the change in MW flow on a *flowgate* given a set of changes in MW bus injections.

A flowgate is a circuit or set of circuits that interconnect different regions of a network that can be limiting under some condition.

The IDC does not represent buses but rather represents control areas, and there are 97 of them in the eastern interconnection. Therefore the flowgates often represent interconnections between these control areas; however, a flowgate may also be internal to a single control area as well.

For our purposes, a control area is a bus, and the flowgates are interconnections between the buses.

One of the most important uses of the IDC is in the coordination of Transmission Loading Relief (TLR) actions. TLR procedures are in place to guide operators in mitigating flows that exceed operational security limits. TLR *levels*, summarized in Table 1 [1] have been defined that correspond to different types of actions that may be taken for which curtailments must be made. When a TLR level 5 is declared, all ongoing transactions including those with firm transmission service are subject to curtailment.

What we desire to obtain, then, is an expression for computing the change in flow on a branch in a network for a given change in MW bus injection.

Table 1: Summary of TLR Levels [1]

TLR Level	Reliability Coordinator Action	Comments	
1	**Notify** Reliability Coordinators of potential Operating Security Limit violations		
2	**Hold** Interchange Transactions at current levels to prevent Operating Security Limit violations	Of those transactions at or above the Curtailment Threshold, only those under existing Transmission Service reservations will be allowed to continue, and only to the level existing at the time of the hold. Transactions using Firm Point-to-Point Transmission Service are not held. See Section B.1.	System Secure
3a	**Reallocation** Transactions using Non-firm Point-to-Point Transmission Service are curtailed to allow Transactions using higher priority Point-to-Point Transmission Service	Curtailment follows Transmission Service priorities. Higher priority transactions are enabled to start by the Reallocation process. See Section B.3.	
3b	**Curtail** Transactions using Non-firm Point-to-Point Transmission Service to mitigate Operating Security Limit Violation	Curtailment follows Transmission Service priorities. There are special considerations for handling Transactions using Firm Point-to-Point Transmission Service. See Section B.4.	Security Limit Violation
4	**Reconfigure** transmission system to allow Transactions using Firm Point-to-Point Transmission Service to continue	There may or may not be an Operating Security Limit violation. There are special considerations for handling Transactions using Firm Point-to-Point Transmission Service. See Section B.5.	
5a	**Reallocation** Transactions using Firm Point-to-Point Transmission Service are curtailed (pro rata) to allow new Transactions using Firm Point-to-Point Transmission Service to begin (pro rata).	Attempts to accommodate all Transactions using Firm Point-to-Point Transmission Service, though at a reduced ("pro rata") level. Pro forma tariff also requires curtailment / Reallocation on pro rata basis with Network Integration Transmission Service and Native Load. See Section B.6.	System Secure
5b	**Curtail** Transactions using Firm Point-to-Point Transmission Service to mitigate Operating Security Limit Violation	Pro forma tariff requires curtailment on pro rata basis with Network Integration Transmission Service and Native Load. See Section B.7.	Security Limit Violation
6	**Emergency Action**	Could include demand-side management, re-dispatch, voltage reductions, interruptible and firm load shedding. See Section B.8.	
0	**TLR Concluded**	**Restore** transactions. See Section B.9.	System Secure

TLR Lev	"Risk" Criteria		Transaction criteria	RELIABILITY COORD Action	Comments
	IMMINENCE	State			
1	Forsee possible condition resulting in violation	Secure		Notify	
2	Expected to approach, is approaching, SOL			Hold	Not > 30 minutes before going to higher levels so xactions may be made based on priority.
3a	Expected to approach is approaching, SOL		Some non-firm ptp at or above curtailment thres – holds, higher priority ptp reservation approved	Reallocate	Curtailments made at top of hour.
3b	Existing or imminent SOL violation or will occur on element removal	Insecure or about to be	Some non-firm ptp at or above their curtailment thresholds.	Hold and Curtail	Hold on onfarm; Curtailments made immediately.
4	Existing or imminent SOL violation	Insecure or about to be		Hold and Reconfigure	Hold on onfarm.
5a	At SOL, no further reconfig possible	Secure	All non-firm ptp at or above curtailment thresholds curtailed; xaction request for previously arranged firm xmission service.	Reallocate	Curtailments made at top of upcoming hour.
5b	Existing or imminent SOL violation or one will occur on element removal, no further reconfig possible	Insecure or about to be	All non-firm ptp at or above curtailment thresholds curtailed.	Curtail	Curtailments made immediately.
6	Existing SOL violation or one will occur upon element removal	Insecure or about to be		Emergency Action	Could include redispatch, reconfiguration, voltage reductions, interruptible and firm load shedding.

2.0 Calculation of Generation Shift Factors

The desired quantity is referred to as the generation shift factor and will be denoted by $t_{b,k}$. It gives the fraction of a change in injection at bus k that appears on branch b. The Power Transfer Distribution Factor (PTDF) is a generalization of the generation shift factor.

This calculation of generation shift factors is relatively straightforward based on what we have done using the DC power flow model.

Recall the DC power flow equations and the corresponding matrix relation for flows across branches.

$$\underline{P} = \underline{B}'\,\underline{\theta} \tag{1}$$

$$\underline{P}_B = (\underline{D} \times \underline{A}) \times \underline{\theta} \tag{2}$$

Inverting eq (1) yields:

$$\underline{\theta} = \left[\underline{B}'\right]^{-1} \underline{P} \tag{3}$$

Substitution of (3) into (2) yields:

$$\underline{P}_B = (\underline{D} \times \underline{A})\left[\underline{B}'\right]^{-1} \underline{P} \tag{4}$$

As we have previously defined in the notes on DC PowerFlow:

- $\underline{P}_B$ is the vector of branch flows. It has dimension of M x 1. Branches are ordered arbitrarily, but whatever order is chosen must also be used in $\underline{D}$ and $\underline{A}$.
- $\underline{D}$ is an M x M matrix having non-diagonal elements of zeros; the diagonal element in position row k, column k contains the negative of the susceptance of the k^{th} branch.
- $\underline{A}$ is the M x (N-1) *node-arc incidence matrix*.
- $\underline{B}$’ is the DC power flow matrix of dimension (N-1)x(N-1), where N is the number of buses in the network, obtained from the Y-bus as follows:

1. Replace diagonal element $\underline{B}'_{kk}$ with the sum of the non-diagonal elements in row k. Alternatively, subtract b_k (the shunt term) from B_{kk}, and multiply by -1.
2. Multiply all off-diagonals by -1.
3. Remove row 1 and column 1.

- $\underline{P}$ is the vector of nodal injections for buses 2, …, N

The calculation of eq. (4) provides the flows on all lines given the injections at all buses.

Bus this is not what we want. What we want is the fraction change in flow on all lines given a change in injection at one bus.

In other words, given a change in injection vector $\Delta\underline{P}$:

$$\Delta\underline{P} = \begin{bmatrix} P_2 \\ P_3 \\ \vdots \\ P_k \\ \vdots \\ P_N \end{bmatrix} - \begin{bmatrix} P_2^0 \\ P_3^0 \\ \vdots \\ P_k^0 \\ \vdots \\ P_N^0 \end{bmatrix} = \begin{bmatrix} \Delta P_2 \\ \Delta P_3 \\ \vdots \\ \Delta P_k \\ \vdots \\ \Delta P_N \end{bmatrix} = \underline{P} - \underline{P}^0 \tag{5}$$

Therefore,

$$\begin{aligned} \Delta\underline{P}_B &= \underline{P}_B - \underline{P}_B^0 \\ &= (\underline{D}\times\underline{A})\left[\underline{B}'\right]^{-1}\underline{P} - (\underline{D}\times\underline{A})\left[\underline{B}'\right]^{-1}\underline{P}^0 \\ &= (\underline{D}\times\underline{A})\left[\underline{B}'\right]^{-1}\left(\underline{P} - \underline{P}^0\right) \\ &= (\underline{D}\times\underline{A})\left[\underline{B}'\right]^{-1}\Delta\underline{P} \end{aligned} \tag{6}$$

Now let the $\Delta\underline{P}$ vector be all zeros except for the element corresponding to the k[th] bus, and assign this bus an injection change of 1.

$$\Delta\underline{P} = \begin{bmatrix} \Delta P_2 \\ \Delta P_3 \\ \vdots \\ \Delta P_k \\ \vdots \\ \Delta P_N \end{bmatrix} = \begin{bmatrix} \Delta P_2 \\ \Delta P_3 \\ \vdots \\ \Delta P_k \\ \vdots \\ \Delta P_N \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \tag{7}$$

Then

$$\Delta\underline{P}_B = \begin{bmatrix} \Delta P_{B1} \\ \Delta P_{B2} \\ \vdots \\ \Delta P_{Bb} \\ \vdots \\ \Delta P_{BM} \end{bmatrix} = \begin{bmatrix} t_{1,k} \\ t_{2,k} \\ \vdots \\ t_{b,k} \\ \vdots \\ t_{M,k} \end{bmatrix} = (\underline{D} \times \underline{A})[\underline{B}']^{-1} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \tag{8}$$

Question: Does the above equation imply that the injection is changed at only one bus? Explain.

Definition: The generation shift factor $t_{b,k}$ is defined as

$$t_{b,k} = \left.\frac{\Delta P_{Bb}}{\Delta P_k}\right|_{\substack{\text{Reallocation}\\ \text{Policy}}}$$

Example 1:

Consider the example that we started in the “PowerFlow” notes. Compute the generation shift factors for all branches corresponding to an increase in bus 2 injection and a decrease in bus 3 injection.

$$\begin{bmatrix} t_{1,23} \\ t_{2,23} \\ t_{3,23} \\ t_{4,23} \\ t_{5,23} \end{bmatrix} = \begin{bmatrix} 10 & 0 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0 \\ 0 & 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 0 & 10 \end{bmatrix} \begin{bmatrix} 0 & 0 & -1 \\ -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & -1 & 1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} 20 & -10 & 0 \\ -10 & 30 & -10 \\ 0 & -10 & 20 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & -10 \\ -10 & 0 & 0 \\ 10 & -10 & 0 \\ 0 & -10 & 10 \\ 0 & -10 & 0 \end{bmatrix} \begin{bmatrix} 0.0625 & 0.025 & 0.0125 \\ 0.025 & 0.05 & 0.025 \\ 0.0125 & 0.025 & 0.0625 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & -10 \\ -10 & 0 & 0 \\ 10 & -10 & 0 \\ 0 & -10 & 10 \\ 0 & -10 & 0 \end{bmatrix} \begin{bmatrix} 0.0375 \\ -0.025 \\ -0.0125 \end{bmatrix} = \begin{bmatrix} 0.125 \\ -0.375 \\ -0.2125 \\ 0.125 \\ 0.25 \end{bmatrix}$$

Note that the above generation shift factors are for a "double shift."

You can think of it like this. A generation shift factor for branch b, bus k would be $t_{b,k}$ and another generation shift factor for branch b, bus j would be $t_{b2,j}$. If we have an injection increase at bus k of ΔP_k and an injection increase at bus j of ΔP_j, then

$$\Delta P_b = t_{b,k}\Delta P_k + t_{b,j}\Delta P_j \qquad (9)$$

Increase P_k, Decrease P_1

Decrease P_j, Increase P_1

Therefore, if ΔP_k=-ΔP_j, then

$$\Delta P_b = \left(t_{b,k} - t_{b,j}\right)\Delta P_k \tag{10}$$

3.0 Generation Shift Factors with Distributed Slack

Equation (8) shows how to compute the generation shift factors for the case when a single specified slack bus corresponds to bus 1.

Example 1 above shows how to compute the generation shift factors for the case when a single specified slack bus corresponds to some other bus in the network (not the bus corresponding to the reference by way of omission from its corresponding row and column in the B' matrix).

What we are interested in here is computation of generation shift factors for the case when we would like to distribute the slack, or the compensation, throughout the network. The key criterion to guide this is that the elements in the nodal injection vector should correspond to the percentage of desired compensation for each bus.

This criterion is illustrated below:

$$\Delta \underline{P}_B = \begin{bmatrix} \Delta P_{B1} \\ \Delta P_{B2} \\ \vdots \\ \Delta P_{Bb} \\ \vdots \\ \Delta P_{BM} \end{bmatrix} = \begin{bmatrix} t_{1,k} \\ t_{2,k} \\ \vdots \\ t_{b,k} \\ \vdots \\ t_{M,k} \end{bmatrix} = (\underline{D} \times \underline{A})\left[\underline{B}'\right]^{-1} \begin{bmatrix} c_2 \\ c_3 \\ \vdots \\ 1 \\ \vdots \\ c_N \end{bmatrix} \tag{11}$$

where

$$c_1 = -\sum_{i=2}^{N} c_i = -\left(1 + \sum_{\substack{i=2 \\ i \neq k}}^{N} c_i\right) \tag{12}$$

is the allocation desired for the reference bus.

One way to distribute the slack is to distribute equally to all buses. In this case,

$$c_i = \frac{-1}{N-1} \tag{13}$$

where we use *N-1* in the denominator because one bus, bus *k*, is the bus for which the computation is being made (and therefore $c_k=1$). If we use (13), then we can substitute into (12) to obtain:

$$c_1 = -\sum_{i=2}^{N} c_i = -\left(1 + \sum_{\substack{i=2 \\ i \neq k}}^{N} c_i\right) = -\left(1 + \sum_{\substack{i=2 \\ i \neq k}}^{N} \frac{-1}{N-1}\right)$$

$$= -\left(1 + (N-2)\frac{-1}{N-1}\right) = -\left(\frac{N-1}{N-1} - \frac{N-2}{N-1}\right) = \frac{-1}{N-1}$$

<u>Example 2</u>:

Using the system from Example 1 above, compute generation shift factors for all branches corresponding to an increase in bus 2 injection, when the slack is equally distributed to all buses.

$$\begin{bmatrix} t_{1,2all} \\ t_{2,2all} \\ t_{3,2all} \\ t_{4,2all} \\ t_{5,2all} \end{bmatrix} = \begin{bmatrix} 10 & 0 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0 \\ 0 & 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 0 & 10 \end{bmatrix} \begin{bmatrix} 0 & 0 & -1 \\ -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & -1 & 1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} 20 & -10 & 0 \\ -10 & 30 & -10 \\ 0 & -10 & 20 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ -0.333 \\ -0.333 \end{bmatrix}$$

$$=\begin{bmatrix} 0 & 0 & -10 \\ -10 & 0 & 0 \\ 10 & -10 & 0 \\ 0 & -10 & 10 \\ 0 & -10 & 0 \end{bmatrix}\begin{bmatrix} 0.0625 & 0.025 & 0.0125 \\ 0.025 & 0.05 & 0.025 \\ 0.0125 & 0.025 & 0.0625 \end{bmatrix}\begin{bmatrix} 1 \\ -0.333 \\ -0.333 \end{bmatrix}$$

$$=\begin{bmatrix} 0 & 0 & -10 \\ -10 & 0 & 0 \\ 10 & -10 & 0 \\ 0 & -10 & 10 \\ 0 & -10 & 0 \end{bmatrix}\begin{bmatrix} 0.05 \\ 0 \\ -0.0166 \end{bmatrix}=\begin{bmatrix} 0.1664 \\ -0.5001 \\ 0.4999 \\ -0.1666 \\ -0.0002 \end{bmatrix}$$

It is of interest to compare the answer from the example where the slack was distributed entirely to bus 3 and the example where the slack was distributed to all buses.

$$\begin{bmatrix} t_{1,23} \\ t_{2,23} \\ t_{3,23} \\ t_{4,23} \\ t_{5,23} \end{bmatrix}=\begin{bmatrix} 0.125 \\ -0.375 \\ -0.2125 \\ 0.125 \\ 0.25 \end{bmatrix} \qquad \begin{bmatrix} t_{1,2all} \\ t_{2,2all} \\ t_{3,2all} \\ t_{4,2all} \\ t_{5,2all} \end{bmatrix}=\begin{bmatrix} 0.1664 \\ -0.5001 \\ 0.4999 \\ -0.1666 \\ -0.0002 \end{bmatrix}$$

Clearly the assumption on slack distribution is important!

There are other ways to distribute the slack. For example, we may distribute the slack equally to all generation buses. Or we may distribute the slack equally to all load buses. Or we may distribute the slack to all generation buses in proportion to the MVA rating of the generation that is located there (this approach conforms best to reality, as we will see when we study AGC).

4.0 Generation Shift Factor Matrix

Given a specified slack distribution, we may compute a matrix of generation shift factors according to

$$
\underline{T}=\underbrace{\begin{bmatrix} t_{1,1} & t_{1,2} & \cdots & t_{1,k} & \cdots & t_{1,N} \\ t_{2,1} & t_{2,2} & \cdots & t_{2,k} & \cdots & t_{2,N} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ t_{b,1} & t_{b,2} & \cdots & t_{b,k} & \cdots & t_{b,N} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ t_{M,1} & t_{M,2} & \cdots & t_{M,k} & \cdots & t_{M,N} \end{bmatrix}}_{M\times N}
$$

$$
=(\underbrace{\underline{D}}_{M\times M} \times \underbrace{\underline{A}}_{M\times(N-1)})\ \underbrace{[\underline{B}']^{-1}}_{(N-1)\times(N-1)} \underbrace{\begin{bmatrix} c & 1 & c & c & c & c & c \\ c & c & 1 & c & c & c & c \\ c & c & c & 1 & c & c & c \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ c & c & c & c & c & \vdots & c \\ c & c & c & c & c & \cdots & 1 \end{bmatrix}}_{(N-1)\times N}
$$

The above assumes that we desire generation shift factors for every branch (a row of $\underline{T}$) and every bus (a column of $\underline{T}$). Note that the first column of $\underline{T}$ is for a shift at the bus 1, which is the one assumed to be deleted from the $\underline{B}'$ matrix.

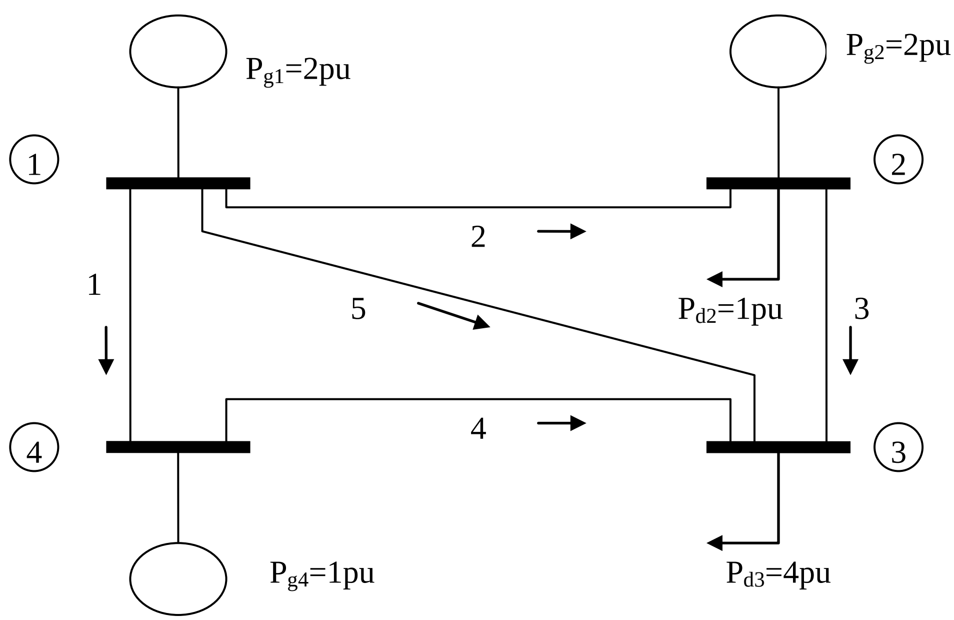
However, we need not include every branch. There may be some branches that we know from experience will never overload, or there may be policy that requires a particularly application to only monitor certain branches. The latter is the case for NERC's IDC described at the beginning of this document.

Example 3: Let's compute the T-matrix for Example 2. We assume a distributed slack bus, where, c=-1/3. Therefore

$$\underline{T} = \begin{bmatrix} 10 & 0 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0 \\ 0 & 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 0 & 10 \end{bmatrix} \begin{bmatrix} 0 & 0 & -1 \\ -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & -1 & 1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} 20 & -10 & 0 \\ -10 & 30 & -10 \\ 0 & -10 & 20 \end{bmatrix}^{-1} \begin{bmatrix} -0.333 & 1 & -0.333 & -0.333 \\ -0.333 & -0.333 & 1 & -0.333 \\ -0.333 & -0.333 & -0.333 & 1 \end{bmatrix}$$

$$\underline{T} = \begin{bmatrix} 0.3333 & 0.1667 & 0 & -0.5 \\ 0.3333 & -0.5 & 0 & 0.1667 \\ 0 & 0.5 & -0.3333 & -0.1667 \\ 0 & -0.1667 & -0.3333 & 0.5 \\ 0.3333 & 0 & -0.3333 & 0 \end{bmatrix}$$

The one-line diagram is shown below to facilitate understanding of the relation between increased injection at bus k (the columns) and how branch flows are affected (the rows).



Remember: each column is the set of shift factors for a unit increase in injection (generation) at a certain bus. Column 1 is when the injection at bus 1 is increased (there is no "1" in that column because that is the one corresponding to the bus that was deleted in the B' matrix). Column 2 is when the injection at bus 2 is increased, and so on.

References:
[1] North American Electric Reliability Council (NERC) Operating Manual, Appendix 9C1, May, 2004, available at www.nerc.com.

---

[i] J. Yan, G. Stern, P. Luh, and F. Zhao, "Payment versus bid cost," IEEE Power and Energy Magazine, March/April 2008.
[ii] Eugene Litvinov, Tongxin Zheng, Gary Rosenwald, and Payman Shamsollahi, "Marginal Loss Modeling in LMP Calculation," IEEE Transactions On Power Systems, Vol. 19, No. 2, May 2004.