Zonal Pricing in a Deregulated Electricity Market

Mette Bjørndal
Kurt Jørnsten

Institute of Finance and Management Science
The Norwegian School of Economics and Business Administration
Helleveien 30
N-5045 Bergen
Norway

Abstract

In the deregulated Norwegian electricity market a zonal transmission pricing system is used to cope with network capacity problems. In this paper we will illustrate some of the problems that the zonal pricing system, as implemented in Norway, has. With the use of small network examples we illustrate the difficulties involved in defining the zones, the redistribution effects of the surplus that a zonal pricing system has, as well as the conflicting interests concerning zone boundaries that are present among the various market participant. We also show that a zone allocation mechanism based on nodal prices does not necessarily lead to a zone system with maximal social surplus. Finally, we formulate an optimization model that when solved yields the zone system that maximizes social surplus given a pre-specification of the number of zones to be used.

1. Introduction

A zonal approach to managing congestion has been adopted in the Norwegian scheduled power market. The trading process works approximately as follows:

1) Based on the supply and demand schedule bids given by the market participants, the market is cleared while ignoring any grid limitations. This produces a system price p of energy.
2) If the resulting flows induce capacity problems, the nodes of the grid are partitioned into zones.
3) Considering the case with two zones defined, the zone with net supply is defined as the low-price area, whereas the net demand zone is determined the high-price area.
4) Net transmission over the zone-boundary is fixed when curtailed to meet the violated capacity limit.

5) The zonal markets are now cleared separately giving one price for each zone, p_L being the low price and p_H the high price. If the flow resulting from this equilibrium still violates the capacity limit, the process is repeated from step 4). If any new limits are violated the process would be repeated from step 2), possibly generating additional zones.
6) The revenue of the grid-company, (from capacity charges), is equal to the price difference times the transmission across the zone-boundary.

An assumption made in the six steps given above is that a zone boundary should cut the link with the capacity problem. In a large network this still leaves the grid-company, Statnett, with a huge flexibility when defining the zone-boundaries. According to Statnett [11] the Norwegian system can be interpreted as inflicting a positive capacity charge $p - p_L$ in the low price area and a negative charge $p - p_H$ in the high price area (relative to the system price of energy). This means that withdrawals are charged in the high price area and compensated in the low price area. For net injections the opposite is valid.

As pointed out above it is not exactly clear how the number of zones and zone-boundaries are to be determined. Stoft [12], [14] shows that the partition of the network into zones generally is not obvious[1], but states that it should be based on price differences, the reason being that the dead-weight loss resulting from erroneous prices is generally proportional to the square of the pricing error. Walton and Tabors [15] also focus on price differentials and suggest that it might be possible to use statistical methods using the standard deviation of nodal price distributions as a criterion to determine the number of zones and which nodes belong to/do not belong to the different zones.

In this paper we will show the multitude of possible cuts, representing zone-boundaries, that exists even in a small example and study the resulting welfare effects. Different zone allocations will affect both the overall efficiency and the allocation of social surplus. We will also illustrate that the partition of the network into zones based on absolute values of optimal nodal price differences does not necessarily lead to a zone system with maximal social surplus. Gaming is not considered since we assume nodal markets to be competitive when

[1] Networks are generally not zonable.

calculating the market outcomes. The gaming possibilities in a zone price system, as the Norwegian one, will be studied in a forthcoming paper.

2. Computing Zonal Prices

We consider real power in a lossless and linear "DC"-model with all line-reactances equal to 1. When the number of zones K ($K \le n$, where n is the number of nodes in the network) and the allocation of nodes to zones $Z_1, \dots, Z_K$ are determined, the optimal zonal prices can be found by solving the following problem:

(2-1) $$\max \sum_{i=1}^{n} \left(\int_0^{q_i^d} p_i^d(q)dq - \int_0^{q_i^s} p_i^s(q)dq \right)$$

(2-2) $$\text{s.t.} \quad q_i^s - q_i^d = \sum_{j \ne i} q_{ij} \qquad i = 1, \dots, n-1$$

(2-3) $$\sum_{ij \in L_l} q_{ij} = 0 \qquad l = 1, \dots, m-n+1$$

(2-4) $$\sum_{i=1}^{n} (q_i^s - q_i^d) = 0$$

(2-5) $$q_{ij} \le C_{ij} \qquad 1 \le i, j \le n$$

(2-6) $$\begin{cases} p_i^s(q_i^s) = p_{Z_k} \\ p_i^d(q_i^d) = p_{Z_k} \end{cases} \qquad i \in Z_k,\ k = 1, \dots, K.$$

Here $p_i^d(q_i^d)$ is the demand function of node i and q_i^d is the quantity of real power consumed. $p_i^s(q_i^s)$ is the supply function of node i, while q_i^s is the quantity of real power produced. C_{ij} is the capacity of link ij, q_{ij} is the power flow over the link from i to j, and p_{Z_k} is the price in zone Z_k.

The objective function (2-1) expresses the difference between consumer benefit (the area under the demand curve) and the cost of production (the area under the supply curve). Equations (2-2) correspond to Kirchhoff's junction rule, and there are $n-1$ independent equations. Equations (2-3) represent Kirchhoff's loop rules where $L=(L_1,\ldots,L_{m-n+1})$ represents a set of independent loops (Dolan and Aldous [2]), and L_l is the set of directed arcs in a path going through loop l. Equation (2-4) stands for conservation of energy, while inequalities (2-5) are the capacity constraints. Equations (2-6) guarantee that prices are uniform over nodes belonging to the same zone.

Solving (2-1) (or alternatively (2-1)-(2-4) to obtain line flows) gives the unconstrained dispatch and the system price. Problem (2-1)-(2-5) corresponds to the optimal dispatch problem from which optimal nodal prices are found (see Schweppe et al. [9], Hogan [7], Wu et al. [16] or Chao and Peck [1]). Solving (2-1)-(2-6) provides us with the optimal zonal prices. It is obvious that the social surplus of the optimal dispatch is less than or equal to the unconstrained social surplus and greater than or equal to the social surplus of the zonal solution. Moreover, it is obvious that a finer partition of the grid (dividing a zone into two or more "subzones" by allowing more prices) will increase social surplus or leave it unchanged.

In practice we would not solve problem (2-1)-(2-6) to find the zonal solution, because this would be equally complicated as solving the optimal dispatch problem. A practical algorithm based on the described procedure of the Norwegian system could be based on curtailment of the unconstrained dispatch. When capacity limits are violated, the grid is partitioned and trades between zones are curtailed until limits are restored. Zonal markets are then cleared separately and new flows are calculated. If these flows still violate the constraints, the flows are curtailed further and we repeat the process. Following the description of the Norwegian system, defining high price and low price areas, we could alternatively lower the price in the low price area and increase the price in the high price area until balance is restored. We will discuss possible problems pertaining to these procedures in relation to the examples of the next section.

3. Examples

The network considered contains 5 nodes connected by 8 edges like the grid of Figure 3-1. In every node there is both production and consumption, and we assume quadratic cost and benefit functions implying linear supply and demand curves. Demand in node i is given by $p_i = a_i - b_i q_i^d$, where p_i is the price in node i and a_i and b_i are positive constants. Supply is given by $p_i = c_i q_i^s$ where c_i is a positive constant. In the specific example considered, we assume identical demand curves in every node, while the cost functions vary as shown in Table 3-1.

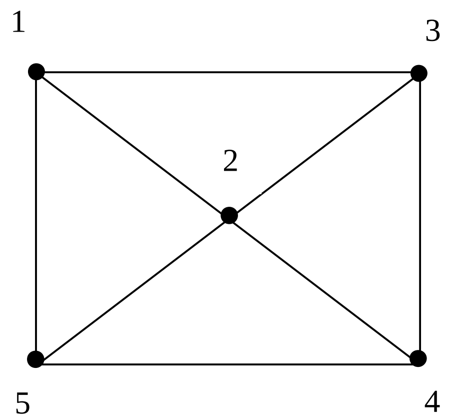


Figure 3-1 Grid Topology

Table 3-1 Supply and Demand Data

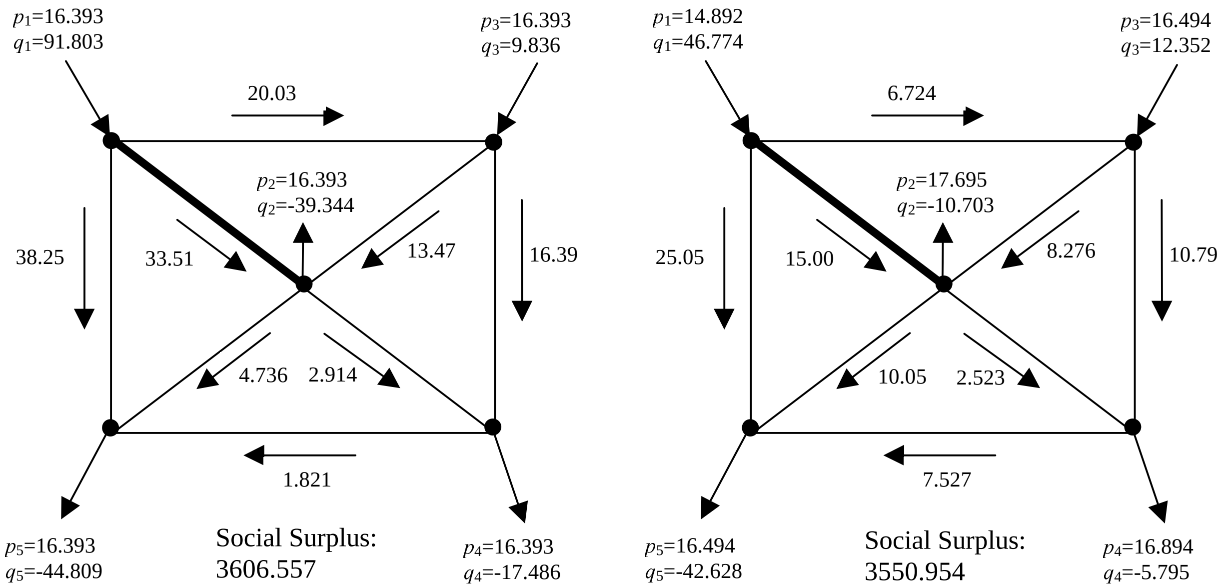
NODE	CONSUMPTION		PRODUCTION
	a_i	b_i	c_i
1	20	0.05	0.1
2	20	0.05	0.5
3	20	0.05	0.2
4	20	0.05	0.3
5	20	0.05	0.6

In the unconstrained dispatch we get a uniform nodal price of 16.393 (the system price of energy). Net injections, $q_i = q_i^s - q_i^d$, and line flows are shown in Figure 3-2 part A. Line 1-2 is assumed to have a capacity of 15 units and is overloaded in the unconstrained dispatch. Taking into account the flow limit, we get the optimal dispatch shown in Figure 3-2 part B.

In the following we will examine zonal pricing. Even if we are restricted to use only two zones in the example, several allocations are possible. In practical zonal implementations[2], the nodes at the endpoints of the congested line would typically be allocated to different zones. However, as is shown later, this is not necessarily optimal when there are more than one congested link. When restricting the attention to the case where the endpoints of the congested

[2] Consider for instance the WEPEX (Western Power Exchange) proposal discussed by Stoft [14].

link are allocated to different zones, there would be 8 different zone allocations in the example. They are all exhibited in Figure 3-3.



Part A:
Unconstrained Dispatch

Part B:
(Constrained) Optimal Dispatch

Figure 3-2 Optimal Dispatch

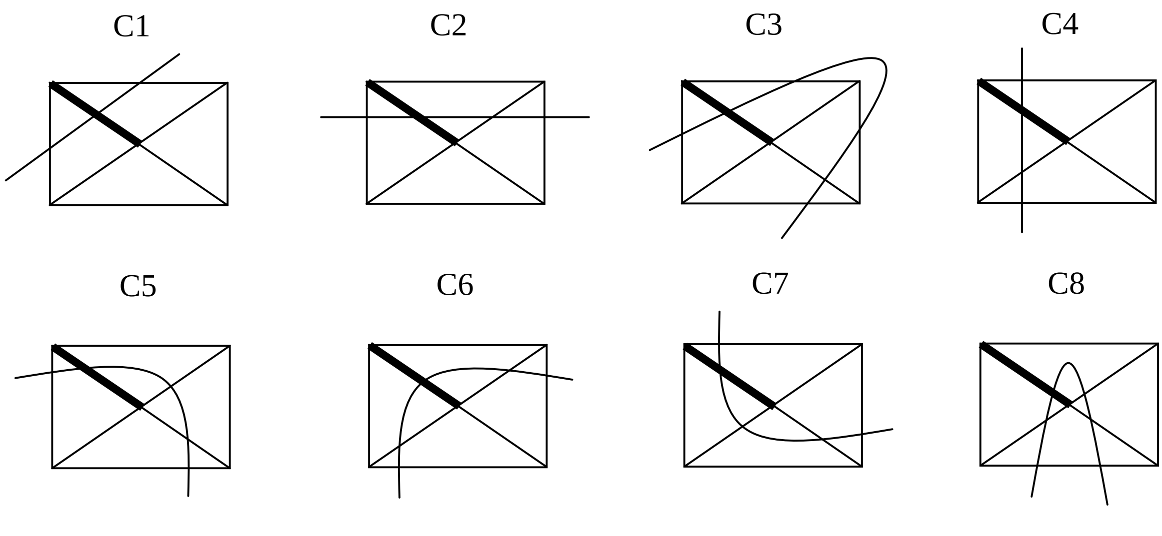


Figure 3-3 Zonal Allocations

Generally, if we consider a single congested line in an n- node network, and if we assume that the endpoints of the congested link are to be allocated to different zones, the number of allocations to two zones is equal to[3]

$$\sum_{i=0}^{n-2} \binom{n-2}{i}.$$

It may be questioned whether all these cuts are meaningful, if not, this is an "at most" number. For instance, as regards cut C3, the zone containing nodes 1 and 4 is not connected, so it can be argued that the network has in practice 3 zones and should be treated accordingly. In the given example, introducing 3 different prices would increase total social surplus from 3439.552 to 3536.556. The grid revenue, equal to the merchandizing surplus (see Wu et al. [16]) would increase from –86.111 to 88.762.

In Table 3-2 and Table 3-3 we show the results for the different zone allocations C1-C8. In addition, we show the results for the constrained (OD) and unconstrained (UD) optimal dispatch. In the first part of Table 3-2 total social surplus and grid revenue are exhibited. The next parts show the results for the different nodes, i.e. prices *p*, quantities (generation *q(s)*, consumption *q(d)*, and net injection *q*) and surpluses (*S(s)* for the producers and *S(d)* for the consumers with a total of *S* to the region as a whole). Table 3-3 shows how individual line flows vary with different zone allocations (a negative entry in row $i - j$ implies that power flows from node j to node i). The highest and lowest zonal surpluses are in boldface types, and so are also the maximal and minimal (zonal) line-flows.

[3] Allowing nodes 1 and 2 to be in the same zone would add 7 more possibilities in the example. In an *n*-node network the total number of different allocations to two zones is equal to $\sum_{i=0}^{n-2} \binom{n-1}{i}$.

Table 3-2 Prices, Quantities, and Surpluses

	UD	OD	C1	C2	C3	C4	C5	C6	C7	C8
Total	3606.557	3550.954	**3537.568**	3503.220	3439.552	3506.243	3424.065	3498.269	**3422.521**	3470.630
Grid	0.000	90.092	**88.741**	59.194	-86.111	-78.663	-61.359	-59.881	**-262.753**	-119.092

NODE 1	UD	OD	C1	C2	C3	C4	C5	C6	C7	C8
p	16.393	14.892	14.531	14.957	14.516	14.897	15.102	15.369	15.018	15.693
q(s)	163.934	148.925	145.311	149.569	145.160	148.972	151.020	153.689	150.184	156.933
S(s)	1343.725	1108.927	1055.764	1118.544	**1053.567**	1109.638	1140.359	1181.015	1127.767	**1231.395**
q(d)	72.131	102.151	109.378	100.862	109.681	102.055	97.959	92.622	99.631	86.134
S(d)	130.073	260.869	299.089	254.329	**300.746**	260.382	239.899	214.471	248.160	**185.479**
q	91.803	46.774	35.933	48.707	35.479	46.917	53.061	61.067	50.553	70.798
S	1473.797	1369.797	1354.853	1372.873	**1354.313**	1370.020	1380.259	1395.486	1375.927	**1416.873**

NODE 2	UD	OD	C1	C2	C3	C4	C5	C6	C7	C8
p	16.393	17.695	17.001	17.573	17.852	17.493	18.710	18.126	18.588	19.576
q(s)	32.787	35.391	34.001	35.145	35.703	34.985	37.420	36.252	37.175	39.152
S(s)	268.745	313.124	**289.025**	308.800	318.682	305.989	350.067	328.557	345.499	**383.220**
q(d)	72.131	46.094	59.985	48.546	42.967	50.149	25.799	37.477	28.248	8.480
S(d)	130.073	53.116	**89.956**	58.918	46.153	62.873	16.639	35.114	19.949	**1.798**
q	-39.344	-10.703	-25.984	-13.401	-7.263	-15.164	11.622	-1.225	8.927	30.672
S	398.818	366.240	378.981	367.718	364.835	368.862	366.706	**363.670**	365.448	**385.018**

NODE 3	UD	OD	C1	C2	C3	C4	C5	C6	C7	C8
p	16.393	16.494	17.001	14.957	17.852	17.493	15.102	15.369	18.588	15.693
q(s)	81.967	82.470	85.004	74.784	89.258	87.463	75.510	76.844	92.938	78.466
S(s)	671.862	680.138	722.562	**559.272**	796.706	764.974	570.180	590.507	**863.748**	615.697
q(d)	72.131	70.118	59.985	100.862	42.967	50.149	97.959	92.622	28.248	86.134
S(d)	130.073	122.914	89.956	**254.329**	46.153	62.873	239.899	214.471	**19.949**	185.479
q	9.836	12.352	25.018	-26.078	46.292	37.314	-22.449	-15.778	64.690	-7.668
S	801.935	803.052	812.518	813.601	842.859	827.846	810.079	804.979	**883.696**	**801.176**

NODE 4	UD	OD	C1	C2	C3	C4	C5	C6	C7	C8
p	16.393	16.894	17.001	17.573	14.516	17.493	15.102	18.126	15.018	15.693
q(s)	56.645	56.315	56.669	58.576	48.387	58.309	50.340	60.420	50.061	52.311
S(s)	447.908	475.707	481.708	514.666	**351.189**	509.982	380.120	**547.594**	375.922	410.465
q(d)	72.131	62.110	59.985	48.546	109.681	50.149	97.959	37.477	99.631	86.134
S(d)	130.073	96.441	89.956	58.918	**300.746**	62.873	239.899	**35.114**	248.160	185.479
q	-17.486	-5.795	-3.316	10.030	-61.294	8.160	-47.619	22.943	-49.570	-33.824
S	577.981	572.148	**571.664**	573.584	**651.935**	572.855	620.019	582.708	624.082	595.944

NODE 5	UD	OD	C1	C2	C3	C4	C5	C6	C7	C8
p	16.393	16.494	17.001	17.573	17.852	14.897	18.710	15.369	15.018	15.693
q(s)	27.322	27.490	28.335	29.288	29.753	24.829	31.183	25.615	25.031	26.155
S(s)	223.954	226.713	240.854	257.333	265.569	**184.940**	**291.722**	196.836	187.961	205.232
q(d)	72.131	70.118	59.985	48.546	42.967	102.055	25.799	92.622	99.631	56.134
S(d)	130.073	122.914	89.956	58.918	46.153	**260.382**	**16.639**	214.471	248.160	185.479
q	-44.809	-42.628	-31.651	-19.258	-13.214	-77.227	5.385	-67.007	-74.601	-59.979
S	354.027	349.626	330.810	316.251	311.722	**445.322**	**308.361**	411.307	436.121	390.711

Table 3-3 Line Flows under Different Zone Allocations

FLOW	UD	OD	C1	C2	C3	C4	C5	C6	C7	C8
1-2	33.515	15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000
1-3	20.036	6.724	1.022	17.990	0.322	-3.132	**23.670**	14.495	**-5.439**	19.181
1-5	38.251	25.050	19.911	15.717	20.157	35.049	**14.392**	31.572	**40.992**	36.618
2-3	-13.479	-8.276	-13.978	2.990	-14.678	-18.132	**8.670**	-0.505	**-20.439**	4.181
2-4	2.914	2.523	-1.917	-2.108	17.258	-2.081	18.560	**-2.292**	18.374	**19.874**
2-5	4.736	10.050	4.911	0.717	5.157	20.049	**-0.608**	16.572	**25.992**	21.618
3-4	16.393	10.799	12.061	**-5.098**	31.936	16.051	9.890	-1.787	**38.813**	15.693
4-5	1.821	7.527	6.828	2.824	-12.101	**22.129**	**-19.168**	18.864	7.617	1.744

Variations in Total Social Surplus

As can be seen from Table 3-2, the zonal allocations show considerable variations when it comes to total social surplus. C1 is best with total surplus of 3537.568, only 13.386 below optimal dispatch (a difference of 0.376%). The poorest allocation is C7 with a surplus of 3422.521, which is 115.047 below C1 or 3.616% below optimal dispatch.

Allocation of Surplus to Individual Agents

For individual agents the outcome is heavily influenced by the allocation to zones. In the example, the greatest difference is experienced by the grid-company, which would prefer C1 with a merchandizing surplus of 88.741, which is 351.494 greater than the –262.753 of C7. For the individual producers and consumers the difference in surplus between the best and worst allocation can be several hundreds, for instance the surplus of producer 3 in C7 is 304.476 greater than the surplus attained in C2. Likewise consumers 3, 4 and 5 preferring C2, C3 and C4 respectively, will be better off by more than 200 compared to their least favorable allocation. It is also evident from the tables that between producers and consumers there is a conflict of interest, the allocation preferred by the producer is the allocation *least* favored by the consumer, and the contrary.

Based on these observations it is questionable that the grid-company shall have this power to effect the distribution of surplus among the participants in the market. Also since the selection of zone boundaries affect the surplus allocation to the grid-company, there might be a conflict of interest between the grid-company and the market participants. One way to handle this conflict of interest could be to specify a rule for selecting zone boundaries, for instance a regulation specifying that the grid-company shall select zone boundaries in order to maximize total social surplus. This would in the example mean C1. However, such a regulation is dependent on well-behaved players, i.e. suppliers and consumers must truthfully reveal to the system operator their cost and demand schedules. As pointed to by Wu et al. [16] there are strong incentives for players not to behave in this way.

If the zonal pricing system is to be based on pre-specified zones, which is not the case in the current Norwegian system, one could base the zone allocation on some form of bargaining mechanism based on results from a typical load flow situation. This approach is something that we are currently investigating.

Line Flows

As displayed in Table 3-3, line-flows vary greatly from one zone definition to another. In some cases lines may be heavily loaded while other allocations leave the links practically unused. In addition, the direction of the line-flows depend on which cut is considered. This may have the effect that lines that are not congested in optimal dispatch may be congested in the zonal solution, i.e. additional limitations may be introduced.

Consider for instance the case where there is a flow limit of 15 on line 2-5. This constraint does not bind, neither in the unconstrained solution nor in (the constrained) optimal dispatch. Choosing a zone definition corresponding to C4 however (or C6, C7 or C8) activates the constraint. Holding on to zone definition C4, it is not possible to find *two* zonal prices that clear the zonal markets and induce a feasible flow in the given example. Adding a third zone by separating nodes 1 and 5 solves the problem, and the partition of the network with new zonal prices is shown in Figure 3-4 part A. Due to the new constraint requiring 3 zones, social surplus has increased, also compared to C1.

Figure 3-4 part B illustrates that the degree of improvement depends partly on the system operator being allowed to make an efficient redispatch. As is also discussed by Stoft [12], restricting the system operator to redispatch only until congestion is relieved (implying $q_{25} = 15$) might reduce social surplus. Moreover, line-flows that are left at its limit may constitute a security threat.

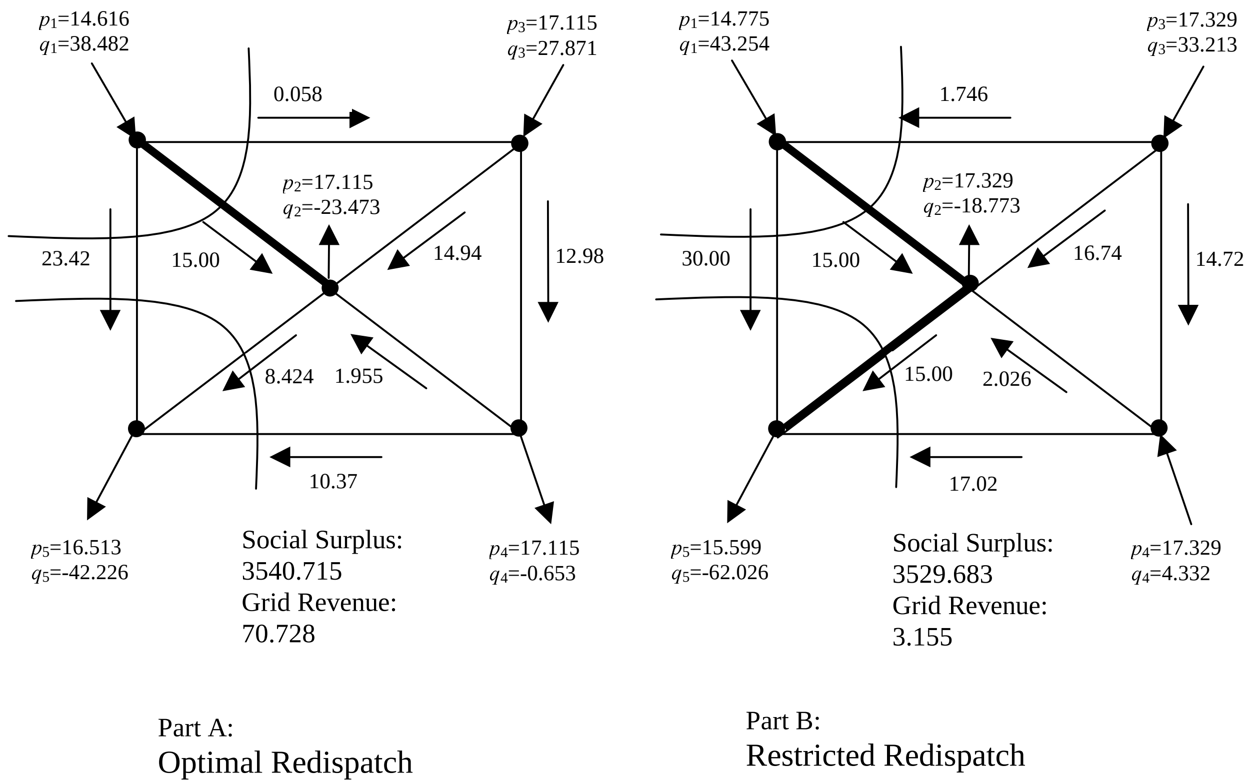


Figure 3-4 Secondary Constraints

Merchandizing Surplus

In Table 3-2 there are several examples of negative merchandizing surplus. This is closely related to line-flows varying as a consequence of choosing different zone allocations. Consider for instance C7, letting area I consist of nodes 1, 4 and 5, while nodes 2 and 3 belong to area II. In unconstrained dispatch, area I is a surplus area with a combined net injection of 29.508 which is exported to area II. In C7 however, flow over the zone-boundary from area I to area II has been reduced to –73.618, i.e. there is net flow from the high price area to the low price area, with the result that the revenue from the grid is equal to $-73.618 \cdot (18.588 - 15.018) \approx -262.816$.

Changing the parameters of the example, for instance by reducing the capacity of line 1-2 to 1 unit, gives a negative merchandizing surplus even for the best cut (which is still C1). Even if so in Table 3-2, the best cut does not necessarily give the maximal merchandizing surplus.

This is illustrated in Figure 3-5 where we assume two congested lines, 1-2 with a capacity of 15 units and 4-5 with a capacity of 5 units. The figure displays optimal nodal prices as well as zonal prices in the case of three (upper part) and four (lower part) zones. Only zone allocations corresponding to maximal social surplus and maximal grid revenue are exhibited.

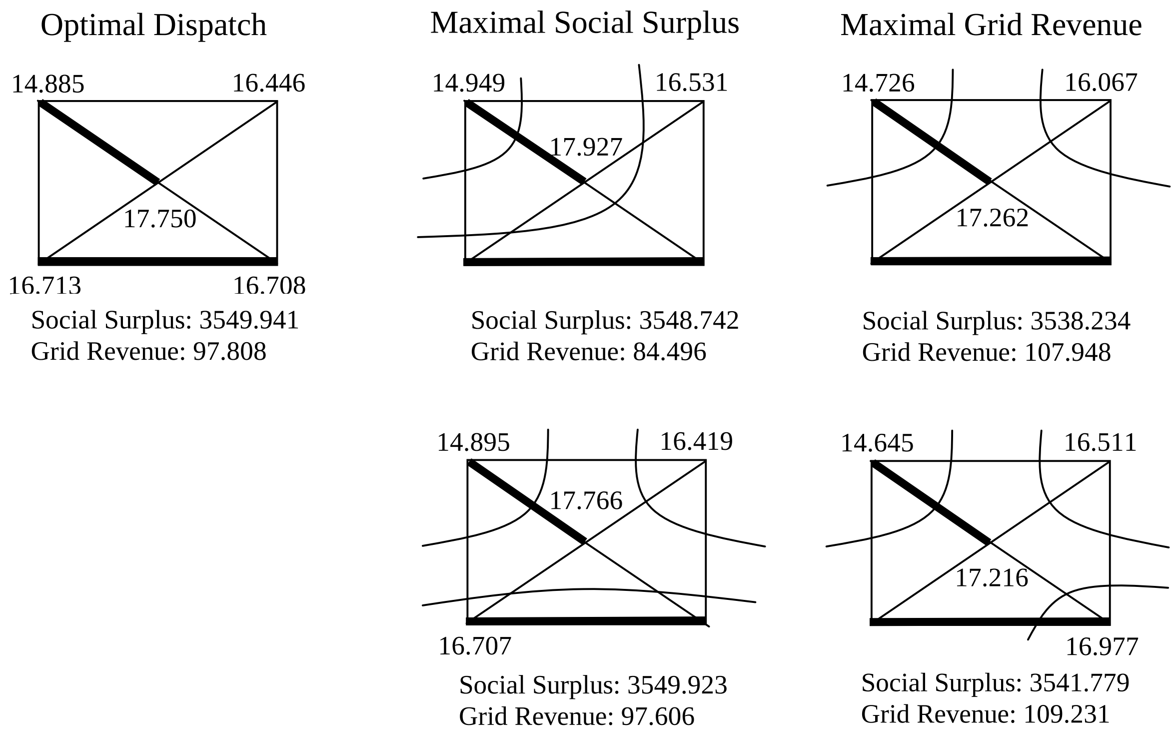


Figure 3-5 Two Congested Lines With Three and Four Zones

Practical Implementations

The results of Table 3-2 and Table 3-3 are based on optimizing the zonal prices. A question that may be raised is whether the practical procedures outlined at the end of section 2 will converge to these prices. We can think of two possible problems.

Adjusting Prices

An algorithm relying on price adjustments may run into problems because prices are not to be changed in the expected direction. In the example of Figure 3-6 (with input data in Table 3-4), line 1-2 has a capacity of 20 and is overloaded in unconstrained dispatch. Assuming nodes 1, 3 and 4 to be in one zone and node 2 to be in the second, node 2 is a surplus area and therefore defined to be the “low-price area”. Choosing the optimal zonal prices corresponding to this partition however requires node 2 to have the highest price. This implies that a procedure of adjusting prices in the supposed direction from the system price will not converge. Since the surplus area does not necessarily have the lowest price, the interpretation of Statnett on positive and negative charges is not general. However, if the validity of this interpretation is used as a criterion for choosing zone allocations, it will guarantee a positive revenue from the grid.

Table 3-4 Input Data for 4-Node Example

NODE	CONSUMPTION		PRODUCTION
	a_i	b_i	c_i
1	20	0.05	0.1
2	20	0.05	0.2
3	20	0.05	0.4
4	20	0.05	0.5

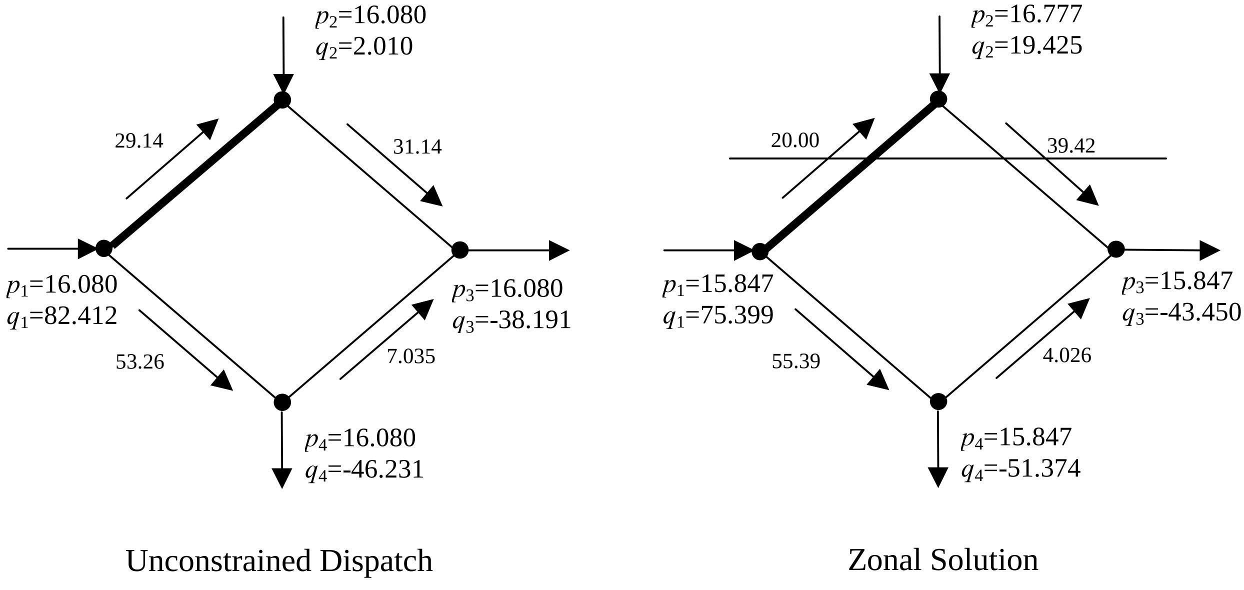


Figure 3-6 "Low Price" Becomes High Price

Curtailing Flow

A procedure based on curtailing flow over the zone-boundary may also run into problems. If for instance a zone contains both the congested link and the nodes adjacent to it, the procedure must curtail *intra*-zonal flows to make progress. We have already stated that practical implementations typically place the endpoints of a congested line in different zones. However, as can be seen from the example of Figure 3-5, this may not be optimal. Both in the three and four zone optimal solutions, nodes 4 and 5 are in the same zone and the congested link 4-5 is intra-zonal. This example shows that restricting attention to zones where zone boundaries should cut congested links, as is done in the current Norwegian system, might lead to a non-optimal solution.

4. The Correct Basis for Zone Allocations

As already mentioned in the beginning of the chapter, both Stoft [12], [14] and Walton and Tabors [15] focus on nodal price differences when evaluating zonal proposals. As the examples of Stoft clearly illustrate, if two nodes have different prices in optimal dispatch, they should in principle belong to different zones. It is however also stated that if a zonal approach renders significant simplification it is no doubt worth some loss of efficiency. The question is then how to allocate nodes to zones such that the loss of social surplus is minimal.

The statistical methods used by Walton and Tabors aim at identifying zones that should be split or combined from means and variances of optimal nodal prices within zones. More specifically, it is reported that a difference-of-the-means test is applied to examine the probability that two zonal samples are, in fact, part of a single sample. Moreover, within each zone outliers are identified (having values further than two standard deviations from the mean), i.e. Walton and Tabors are comparing nodal prices with the average nodal prices in the zones.

Returning to the example at the beginning of section 3, assuming line 1-2 is congested, we have varied the line capacity and changed supply and demand data such that line 1-2 is still congested[4]. In this case it seems like the best zonal division, C1, is quite robust to changes. Also, C1 corresponds to allocating nodes based on absolute price differences, i.e. placing nodes 1 and 2 in different zones and then allocating node i to zone 1 if $|p_i - p_1| < |p_i - p_2|$ and to zone 2 otherwise.

If line 1-2 is the only congested line, it follows from nodal price theory that p_1 would be the lowest price and p_2 the highest, and that p_i can be found as a weighted average of p_1 and p_2 (Stoft [14] or Wu et al. [16]). In the "DC" approximation the exact weights are constants depending on network characteristics only (though p_1 and p_2 depend on the exact capacity and cost and benefit data). Introducing the dual price μ_{ij} of capacity on line ij, prices can be related by applying load factors (Chao and Peck [1]). Since $\mu_{12} > 0$ and $\mu_{ij} = 0 \; \forall ij \neq 12$,

$$p_j = p_i + \mu_{12}\beta_{12}^{ij}$$

where β_{12}^{ij} is the load factor of line 1-2 of a trade from i to j.

In the example

$$|p_i - p_1| = p_i - p_1 = p_1 + \mu_{12}\beta_{12}^{1i} - p_1 = \mu_{12}\beta_{12}^{1i}$$

and

$$|p_i - p_2| = p_2 - p_i = p_2 - (p_2 + \mu_{12}\beta_{12}^{2i}) = -\mu_{12}\beta_{12}^{2i}$$

i.e.[5] $|p_i - p_1| < |p_i - p_2|$ if $\beta_{12}^{1i} < \beta_{12}^{i2}$. Choosing node 2 as the reference point, the condition can be written $\beta_{12}^{1} < 2\beta_{12}^{i}$, meaning that whether the price of node i is closest to p_1 or p_2 depends only on network characteristics and can be decided before any bids are received.

---

[4] Without being in any way exhaustive.

[5] Note that $\beta_{kl}^{ij} = -\beta_{kl}^{ji}$, and with node m being the reference node, $\beta_{kl}^{ij} = \beta_{kl}^{im} - \beta_{kl}^{jm} = \beta_{kl}^{i} - \beta_{kl}^{j}$.

Since $\beta_{12} = (7/15, 0, 1/5, 2/15, 1/5)$ it is easily seen that according to a rule based on (absolute) price differences, nodes 3, 4 and 5 would be allocated to zone 2.

The interpretation is that since price differentials are based on electrical distances and influences (through distribution factors), it is natural that node 2 has a stronger influence on nodes 3, 4 and 5 than node 1 has. Similarly, if only line 4-5 is congested (in direction 4 to 5, i.e. $\mu_{45} > 0$) we find that $|p_i - p_4| < |p_i - p_5|$ if $\beta_{45}^4 + \beta_{45}^5 < 2\beta_{45}^i$ (node 2 still being the reference node). Since $\beta_{45} = (-1/15, 0, 1/15, 4/15, -4/15)$ and $\beta_{45}^4 + \beta_{45}^5 = 0$, node 1 is allocated to 5, node 3 is allocated to 4, and node 2 can be allocated to either, which is also expected since node 2 is equally "far" from both nodes 4 and 5.

Consider now both lines 1-2 and 4-5 to be congested, with $\mu_{12} > 0$ and $\mu_{45} > 0$. Now the general Chao-Peck-expression for relating prices is

$$p_j = p_i + \mu_{12}\beta_{12}^{ij} + \mu_{45}\beta_{45}^{ij}.$$

Examining the relationship between prices of nodes 1 and 2 now give $p_2 - p_1 = \mu_{12}\beta_{12}^{12} + \mu_{45}\beta_{45}^{12}$, and since $\beta_{12}^{12} > 0$ while $\beta_{45}^{12} < 0$, the size and also the sign of $p_2 - p_1$ depend on load factors *and* the size of μ_{12} and μ_{45}, and therefore on the specific input data, i.e. line capacities and cost and benefit data. By considering the other pairs of nodes, some qualitative statements can be made, and they are given in Table 4-1. For instance, the entry of row p_3 and column p_5 is <, implying that $p_3 < p_5$. A question mark indicates that the relationship cannot be decided without knowledge of shadow prices. The given relationships could possibly be used to assess zone definitions, at least in a heuristic sense.

Table 4-1 Price Relationships

	p_2	p_3	p_4	p_5
P_1	?	?	?	$<$
P_2		$>$	$>$	?
P_3			?	$<$
P_4				?

Allocating nodes to zones based on optimal nodal prices requires clustering techniques. An overview over cluster analysis is given by Hansen and Jaumard [5], in which the steps of a clustering study are discussed. In particular, different criteria for evaluating homogeneity and separation are reviewed, together with algorithms for different clustering types. In our setting, the criterion to base zone allocations on, is of special interest.

In the example of Figure 3-1 it seems useful to allocate nodes to zones based on price differentials. In general however, optimal nodal price differentials are in themselves not indicative of the best zone allocation. As exhibited in the example of Figure 4-1 the best zone allocation (Z1 or Z2) varies with the capacity of line 4-5 (all the other parameters are fixed)[6]. When the capacity is equal to 4.2, Z1 is the best partition, allocating node 5 to node 2. Reducing capacity by 0.1 to 4.1, Z2 is best, allocating node 5 to nodes 3 and 4. A capacity of 4.14858 makes Z1 and Z2 equally good when it comes to total social surplus, although, as exhibited in and Figure 4-1 and Table 4-2, the allocation of surplus to individual agents vary considerably.

This switch of best zone allocation occurs even if price differentials are almost identical in the two cases. Both with capacity equal to 4.1 and 4.2, p_5 is closer to p_2 than p_3 and p_4 are, and more so when capacity is 4.1 than when capacity is 4.2. However, in both cases p_5 is closer to p_3 and p_4 (or their average) than to p_2.

[6] Note that the Z2 solutions are identical under the exhibited capacities of line 4-5, the reason being that this constraint is not binding in Z2.

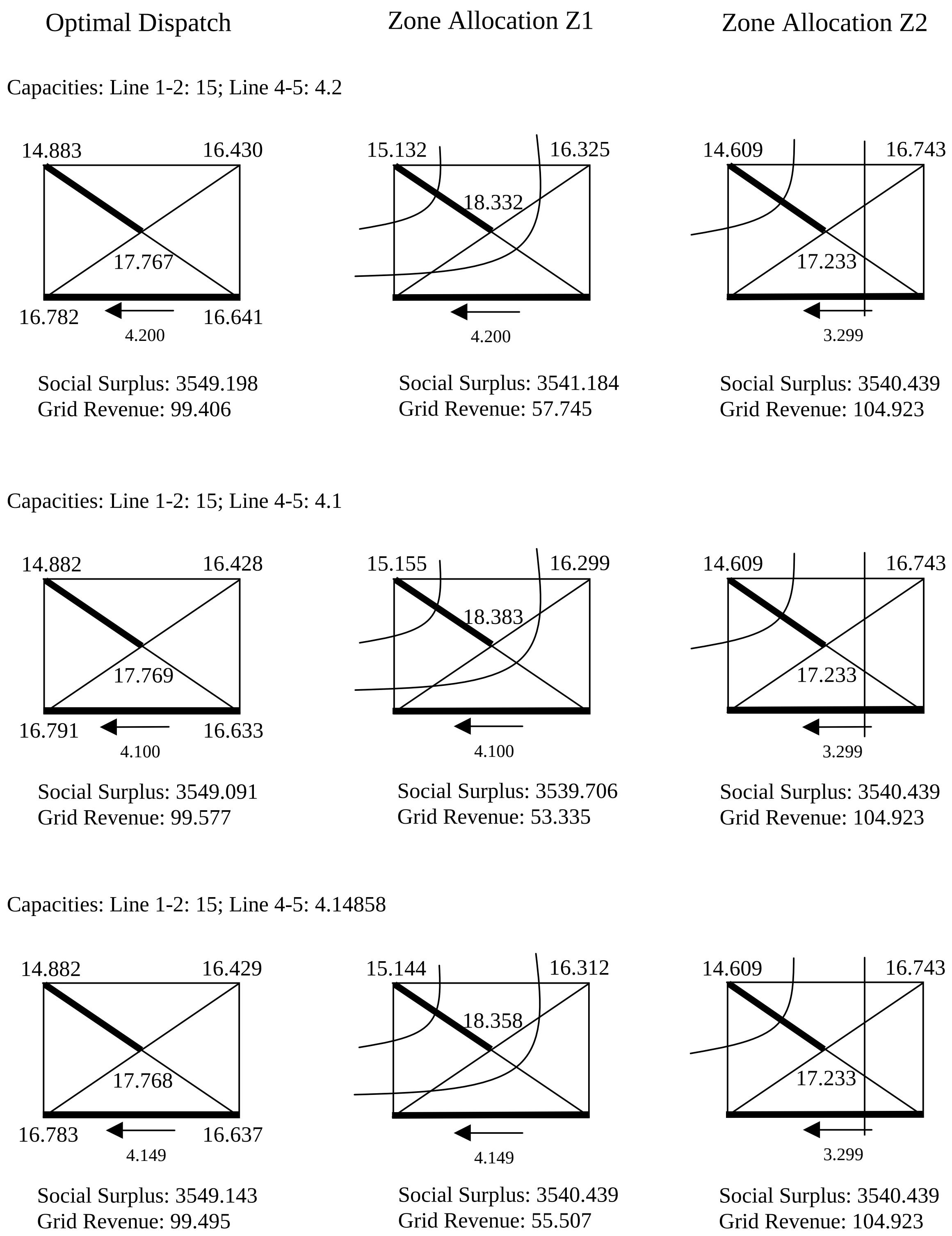


Figure 4-1 Capacity and Zone Allocation

Table 4-2 Quantities and Surpluses

Node	q(s)			q(d)			q		
	OD	Z1	Z2	OD	Z1	Z2	OD	Z1	Z2
1	148.823	151.436	146.086	102.353	97.128	107.828	46.470	54.308	38.258
2	35.536	36.716	34.466	44.642	32.841	55.345	-9.106	3.875	-20.879
3	82.146	81.558	83.715	71.417	73.766	65.140	10.728	7.792	18.575
4	55.457	54.372	55.810	67.257	73.766	65.140	-11.800	-19.394	-9.330
5	27.977	27.186	28.721	64.270	73.766	55.345	-36.293	-46.580	-26.623

Node	S(s)			S(d)			S		
	OD	Z1	Z2	OD	Z1	Z2	OD	Z1	Z2
1	1107.421	1146.642	1067.057	261.904	235.847	290.671	1369.325	1382.489	1357.728
2	315.698	337.015	296.968	49.823	26.963	76.575	365.521	363.978	373.544
3	674.791	665.177	700.819	127.511	136.037	106.082	802.302	801.214	806.900
4	461.324	443.451	467.212	113.088	136.037	106.082	574.412	579.489	573.294
5	234.822	221.726	247.474	103.266	136.037	76.575	338.088	357.763	324.049

The rationale for using price-differences when evaluating zone allocations can be illustrated by comparing the unconstrained and constrained dispatch in a two-node example with consumption in node i and production in node j, the nodes being connected by a line of limited capacity. For simplicity we still assume linear supply and demand functions.

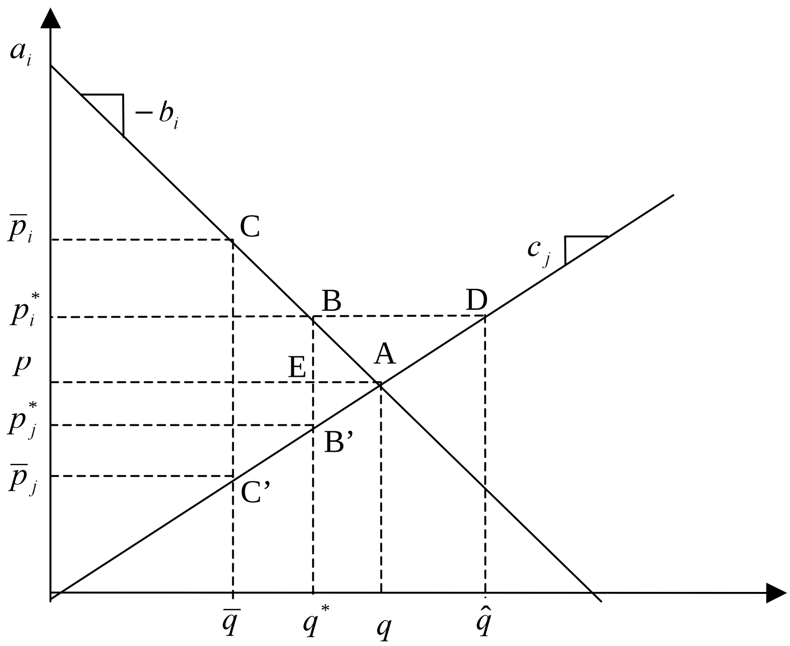


Figure 4-2 Dead-Weight Loss

In Figure 4-2 unconstrained dispatch is given by point A. Because of the limited capacity of the line connecting nodes i and j, this cannot be attained, and optimal dispatch is given by B and B', p_i^* and p_j^* being the optimal prices inducing a production of q^* equal to the capacity of the line. The reduction of social surplus resulting from the capacity limit is equal to the area of triangle ABB'. Points C and C' correspond to deviating from the optimal prices. In the single line case considered, the resulting equilibrium does not fully utilize the capacity of line *ij*. In a larger and more general network involving loop flow, the congested line could still be fully utilized even if prices are not optimal.

The reduction of social surplus, or dead-weight loss, resulting from the price error can be expressed as a function of prices. Comparing with unconstrained dispatch, the reduction corresponds to triangle ACC' and is equal to

$$\begin{aligned}
\tfrac{1}{2}(\overline{p}_i - \overline{p}_j)(q - \overline{q}) &= \tfrac{1}{2}(\overline{p}_i - p)(q - \overline{q}) + \tfrac{1}{2}(p - \overline{p}_j)(q - \overline{q}) \\
&= \tfrac{1}{2}(\overline{p}_i - p)\left(\frac{a_i - p}{b_i} - \frac{a_i - \overline{p}_i}{b_i}\right) + \tfrac{1}{2}(p - \overline{p}_j)\left(\frac{p}{c_j} - \frac{\overline{p}_j}{c_j}\right) \\
&= \tfrac{1}{2b_i}(\overline{p}_i - p)^2 + \tfrac{1}{2c_j}(p - \overline{p}_j)^2,
\end{aligned}$$

showing that the reduction of social surplus in a node due to the capacity limit *and* price error is proportional to the square of the difference between the (unconstrained) system price p and the prevailing price.

Alternatively we could consider the reduction in social surplus from choosing non-optimal prices only. This is equal to the area of trapezium BCC'B', i.e.

$$\begin{aligned}
&\tfrac{1}{2}(\overline{p}_i - \overline{p}_j + p_i^* - p_j^*)(q^* - \overline{q}) \\
&= \tfrac{1}{2}(\overline{p}_i - p_i^*)(q^* - \overline{q}) + \tfrac{1}{2}(p_j^* - \overline{p}_j)(q^* - \overline{q}) + (p_i^* - p_j^*)(q^* - \overline{q}) \\
&= \tfrac{1}{2b_i}(\overline{p}_i - p_i^*)^2 + \tfrac{1}{2c_j}(p_j^* - \overline{p}_j)^2 + \tfrac{1}{2b_i}(p_i^* - p_j^*)(\overline{p}_i - p_i^*) + \tfrac{1}{2c_j}(p_i^* - p_j^*)(p_j^* - \overline{p}_j) \\
&= \tfrac{1}{2b_i}(\overline{p}_i - p_i^*)(\overline{p}_i - p_j^*) + \tfrac{1}{2c_j}(p_j^* - \overline{p}_j)(p_i^* - \overline{p}_j)\,.
\end{aligned}$$

This is also a function of price-differences, involving optimal nodal prices and the prevailing prices $\overline{p}_i$ and $\overline{p}_j$.

The example also illustrates why a uniform market price is not possible without rationing producers or consumers. Raising the price from p in the example reduces consumption, and at point B the total quantity demanded can be handled by the capacitated line. However, at this point suppliers prefer quantity $\hat{q}$, which is greater than q^*, and production must be curtailed or rationed by some mechanism. Using price p_j^* in node j is of course one alternative, the price constituting the rationing mechanism, but this implies that prices are no longer uniform.

Another alternative is *counter-purchases*, buying off some production by compensating producers with the difference between p_i^* and their cost of production. The cheapest way to implement this would be to compensate the costlier producers. In the given example this would imply a cost equal to the area of triangle BB'D. In general, consumers could also take part in this process, in which case optimal dispatch is attainable by transferring ABE to the least valued demand and AB'E to the most expensive suppliers. Moreover, in a network involving loop flow we should take into account the effect of individual agents on the congestion considered. For producers generating counter-flows and consumers relieving congestion this could imply being compensated for *increasing* output. In general, finding the least cost redispatch involves solving an optimization problem of the same possible complexity as the optimal dispatch problem (Fang and David [3], [4] and Singh et al. [10]).

In principle, this arrangement corresponds to the Swedish system of managing congestion, where the cost of counter-purchases is recovered through the fixed network charges[7]. Also the (real time) regulation power markets of both Norway and Sweden manage congestion by redispatching based on incremental and decremental bids. The exact curtailment procedure determines the allocation of social surplus to individual agents. In the discussion above we assumed competitive markets, however, as is illustrated by Stoft [13], a counter-purchase arrangement is vulnerable to gaming.

[7] Note that attaining optimal dispatch by counter-purchases involves a *cost* on the hands of the grid-company, whereas under optimal nodal pricing a positive revenue (merchandizing surplus) is collected (Wu et al. [16]).

Generally in a meshed network, possibly containing both production and consumption in each node, some agents may loose while others are better off due to price errors. However, the dead-weight loss can also in this case be expressed as a function of prices. If p_i^* is the optimal nodal price of node i and $\overline{p}_i$ is the zonal price resulting from a given zone allocation, the difference between surplus in optimal dispatch and in the zonal solution is equal to

$$\sum_i (\tfrac{1}{2b_i} + \tfrac{1}{2c_i})(\overline{p}_i + p_i^*)(\overline{p}_i - p_i^*),$$

assuming linear demand and supply functions. Each part of the expression can be positive or negative depending on the sign of $(\overline{p}_i - p_i^*)$, which again depends on the exact zone allocations that determine $\overline{p}_i$. It is far from obvious how to construct zones from this expression.

The best allocation of nodes to a given number of zones K in the presence of a capacity constraint on line kl (in direction k to l) can be formulated as a non-linear mixed integer program.

(4-1) $$\max \sum_i \sum_j \left[\tfrac{1}{2}(a_i + p_j)q_{ij}^d - \tfrac{1}{2} p_j q_{ij}^s\right]$$

(4-2) s.t. $$p_i = c_i \cdot \sum_j q_{ij}^s \qquad i = 1,\dots,n$$

(4-3) $$p_i = a_i - b_i \cdot \sum_j q_{ij}^d \qquad i = 1,\dots,n$$

(4-4) $$q_{ij}^s \le M_1 \delta_{ij} \qquad i = 1,\dots,n; j = 1,\dots,K$$

(4-5) $$q_{ij}^d \le M_2 \delta_{ij} \qquad i = 1,\dots,n; j = 1,\dots,K$$

(4-6) $$\sum_j \delta_{ij} = 1 \qquad i = 1,\dots,n$$

(4-7)	$p_j \leq (1-\delta_{ij})M_3 + p_i$	$i=1,\ldots,n; j=1,\ldots,K$
(4-8)	$p_j \geq p_i - (1-\delta_{ij})M_4$	$i=1,\ldots,n; j=1,\ldots,K$
(4-9)	$\sum_i \sum_j (q_{ij}^s - q_{ij}^d) = 0$	
(4-10)	$\sum_i \beta_{kl}^i \cdot \sum_j (q_{ij}^s - q_{ij}^d) \leq C_{kl}$	
(4-11)	$q_{ij}^s, q_{ij}^d \geq 0$	$i=1,\ldots,n; j=1,\ldots,K$
(4-12)	$\delta_{ij} = 0/1$	$i=1,\ldots,n; j=1,\ldots,K$

Here, δ_{ij} is a binary variable, which is equal to 1 if node i belongs to zone j and 0 otherwise. It can be interpreted as an indicator of whether node i is allocated to zonal market j. Production in node i when allocated to zone j is q_{ij}^s and total production in node i is $\sum_j q_{ij}^s$. Consumption q_{ij}^d is similar, and M_1-M_4 are arbitrarily large positive constants ("big Ms").

Assuming linear cost and demand functions the objective function (4-1) expresses the difference between consumers' willingness to pay and the cost of production. Constraints (4-2) and (4-3) define the price in node i, (4-6) allocates each node to exactly one zone, and (4-4) and (4-5) guarantee that only q_{ij}^s and q_{ij}^d corresponding to $\delta_{ij} = 1$ can be strictly positive. (4-7) and (4-8) set $p_j = p_i$ if $\delta_{ij} = 1$, otherwise they put no restriction on the relationship between p_i and p_j. Constraint (4-9) balances total supply and demand. (4-10) is the capacity constraint, where the left hand side, multiplying load factors and net injections, is equal to the flow over link kl in direction from *k* to *l*. Finally, (4-11) and (4-12) specify q_{ij}^s / q_{ij}^d and δ_{ij} as non-negative and binary variables, respectively.

Due to the non-linear and discrete nature of the problem, it is difficult to solve. However the non-linearity occurs in the objective function only. Hansen et al. [6] have studied zonal pricing in relation to facility location and developed solution methods for this related problem.

We are currently investigating if the optimization problem stated above can be solved using an equivalent solution approach.

5. Conclusions and Future Research

From the analyses of this paper, it is evident that

- Zonal pricing is second-best
- Zone allocations affect the surplus of individual agents, thus possibly emphasizing conflicts of interest
- Merchandizing surplus may be negative even if the zone allocation is optimal
- The best partition may not have the maximal merchandizing surplus
- The best zone allocation may or may not separate the endpoints of congested lines
- Optimal nodal price differences may or may not be indicative of the best partition

It has also been demonstrated that zonal pricing and zone allocation is difficult if it is to be optimal. This raises the question whether a zonal approach to managing congestion is really a useful simplification of nodal pricing. It may be so if the main point of managing congestion is to obtain feasibility, or if it can be established that the disadvantages of not finding the optimum is outweighed by the perceived simplicity of having only a few prices.

This paper has also identified a number of interesting topics for future research, including developing solution methods for the non-linear mixed integer program given by (4-1)-(4-12). Moreover, there may be a need for further investigation of whether it is useful to base zone allocations on optimal nodal prices, in which case the specific clustering criteria must be identified. In this context we should look for possibilities of making judgements on the error resulting from using optimal nodal prices as the basis for allocating nodes to zones.

In Norway there has recently started a discussion on changing the current flexible zonal pricing system into a system with a few a priori determined zones. The findings in this paper indicate that it is very important to make a thorough investigation on the number of zones needed in a fixed zone system, if the fixed zones shall be the same in all load situations or different according to some pre-specified criteria. Given that a fixed zone system is to be used, there is also a need to investigate the redistribution effects the zone system has on the various market participants and take this into account when defining the fixed zones. Hence, even though a fixed zone system is simpler to handle and may make it easier to develop a market for transmission capacity reservation trading, it is far from obvious that a fixed zone system would be efficient.

6. References

[1] Chao, Hung-Po and Stephen Peck (1996), "A Market Mechanism For Electric Power Transmission", Journal of Regulatory Economics, vol.10, pp.25-59.

[2] Dolan, Alan and Joan Aldous (1993) "Networks and Algorithms: An Introductory Approach", John Wiley & Sons.

[3] Fang, R.S and A. K. David (1999) "Optimal Dispatch Under Transmission Contracts ", IEEE Transactions on Power Systems, vol.14, pp.732-737.

[4] Fang, R.S and A. K. David (1999) "Transmission Congestion Management in an Electricity Market", IEEE Transactions on Power Systems, vol.14, pp.877-883.

[5] Hansen, Pierre and Brigitte Jaumard (1997), "Cluster Analysis and Mathematical Programming", Mathematical Programming, vol.79, pp.191-215.

[6] Hansen, Pierre, Dominique Peeters and Jacques-François Thisse (1997) "Facility Location Under Zone Pricing", Journal of Regional Science, vol.37, pp.1-22.

[7] Hogan, William W. (1992), "Contract Networks for Electric Power Transmission", Journal of Regulatory Economics, vol.4, pp.211-242.

[8] Jörnsten, Kurt, Linda Rud and Balbir Singh (1997), "Samspillet mellom nettariffering og energimarkeder: En illustrasjon", SNF Working Paper 55/97.

[9] Schweppe, F.C., M.C. Caramanis, R.D. Tabors, and R.E. Bohn, "Spot Pricing of Electricity", Kluwer Academic Publishers, 1988.

[10] Singh, Harry, Shangyou Hao and Alex Papalexopoulos (1998) "Transmission Congestion Management in Competitive Electricity Markets", IEEE Transactions on Power Systems, vol.13, pp.672-680.

[11] Statnett (1997), "Tariffer for Sentralnettet 1997".

[12] Stoft, Steven (1996), “Analysis of the California WEPEX Applications to FERC”, PWP-042A, University of California Energy Institute.

[13] Stoft, Steven (1998), “Using Game Theory to Study Market Power in Simple Networks”, Federal Energy Regulatory Commission.

[14] Stoft, Steven (1997), “Zones: Simple or Complex?”, The Electricity Journal, Jan/Feb, pp. 24-31.

[15] Walton, Steven and Richard D. Tabors (1996), “Zonal Transmission Pricing: Methodology and Preliminary Results from the WSCC”, The Electricity Journal, vol.9, pp.34-41.

[16] Wu, Felix, Pravin Varaiya, Pablo Spiller and Shmuel Oren (1996) “Folk Theorems on Transmission Access: Proofs and Counterexamples”, Journal of Regulatory Economics vol. 10, pp. 5-23.