4

Efficient Short-Term Operation of an Electricity Industry with no Network Constraints

Let us start by putting aside network issues. Let us suppose that we have a set of generators and loads that (for the moment) are all located at the same point on the network. Now imagine that you are an omniscient all-powerful operator, able to control all of the generation and consumption resources in the economy. How should you use this stock of generation and consumption resources in the most efficient way possible? We start by focusing on the question of the efficient use of a set of generation resources. Consideration of this question leads us to the notion of the merit order. In the next section, we shall ask whether it is possible to achieve this efficient outcome using decentralised decision-making and market signals alone.

4.1 The Cost of Generation

Let us focus first on the optimal use of a set of existing generation assets. As we saw in Part II, electricity is produced by physical equipment, which we will refer to as an electricity generator.

We will focus in this section on the short-run economic costs incurred by a generator in producing its output. In general, the costs incurred by a generator will depend on many parameters (such as the price of the input fuel and the cost of capital). We will focus on how the costs incurred by a generator might vary with the level of output of the generator (i.e. the rate at which the generator produces electrical energy). Let us call the rate of production of the generator Q (Q is measured in megawatts). Recall that this output represents a flow of electrical energy per unit of time. A generator that produces at the constant rate Q for a period of time T (measured in hours), produces QT MWh of energy.

As mentioned, we will assume that each generator can control the rate at which it produces electrical energy. While this assumption is true for most generators, such as thermal (coal or gas-fired) or hydro generators, it is not true for some other generators, such as wind- or solar-powered generators, which depend on the availability of resources. These generators are not

The Economics of Electricity Markets, First Edition. Darryl R. Biggar and Mohammad Reza Hesamzadeh.

necessarily able to follow instructions from the system operator as to the rate at which they should produce electrical energy. These generators raise specific issues that will be addressed later.

Let $C(Q)$ (\$/h) be the short-run cost function of the generator – that is, the rate at which the generator is incurring expenditure when producing energy at the rate Q MW. A generator that produces at a constant rate Q MW for a period of time T incurs costs of $C(Q)T$ in dollars (or some other unit of currency).

It is common to distinguish these costs into two components:

- The *fixed costs*, which are independent of the output of the generator. These costs include the costs of leasing the generating facilities (such as the land on which the generator is sited) and/or the costs of financing the purchase of the facilities. These costs also include the costs of any permanent operations and maintenance or management staff, which must be maintained whether or not the generator is in operation. These costs do not enter into the output decision calculus, which we will discuss below.
- The *variable costs* (also known as the 'production costs'), which vary with the output of the generator. These costs include the costs of any fuel consumed, any operating or maintenance costs, which vary with output, and the costs incurred in starting and stopping the generator.

In addition, a generator will typically have some minimum level of output below which it cannot physically operate effectively (without shutting down entirely), and some maximum level of output above which it cannot produce any more output.

Furthermore, there may be other important costs that might need to be taken into account, such as the *startup costs* of the generator. In many thermal generators energy must be consumed to heat the water in the boiler before electrical energy can be produced at all. For the moment we will put these costs to one side.

How might the variable costs of a typical generator vary with its output? At low levels of production (close to the minimum operating level), the average variable costs (that is the variable cost divided by the output of the generator) tend to be relatively high since there are often 'auxiliary' costs that must be incurred whenever the generator is in service and producing nonzero levels of output. The average variable cost then typically declines as the output of the generator increases, but may start to rise again as the output of the generator approaches the maximum operating level.

The cost function of a typical generator is sometimes approximated as a quadratic function of its output. For example, the cost function of a typical generator might be assumed to take the following form:

$$C(Q) = \begin{cases} F, & \text{where } Q = 0 \\ cQ + aQ^2 + b + F, & \underline{Q} \le Q \le \overline{Q} \end{cases}$$

where F is the fixed costs of the generator, a, b and c, are the parameters of the cost function, and $\underline{Q}$ and $\overline{Q}$ are the minimum and maximum operating levels of the generator, respectively.

As we saw in Section 1.3, in the short-run, economists usually focus primarily on the marginal cost function of producers such as generators. The marginal cost function is the

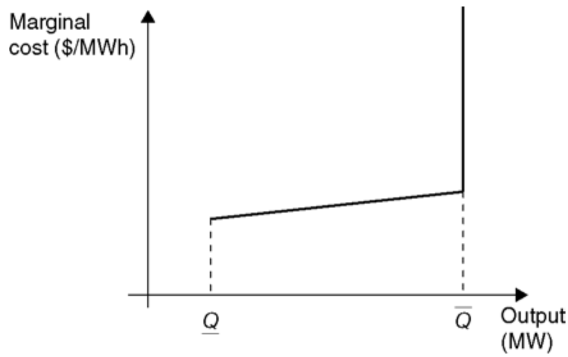

Figure 4.1 Marginal cost curve of a typical generator

derivative of the cost function with respect to the level of output, which we will denote $C'(Q)$. The marginal cost function of the typical generator referred to earlier would be

$$C'(Q) = \begin{cases} \text{undefined}, & \text{where } Q = 0 \\ c + 2aQ, & \underline{Q} \leq Q \leq \overline{Q} \end{cases}$$

The graph of the marginal cost of this hypothetical generator is illustrated in Figure 4.1.

Note that, as the output of each generator reaches its physical limit, it becomes very costly to increase its output further – at this point the marginal cost function turns sharply upward.

A couple of points should be borne in mind. First, the variable cost of a generator (and therefore the marginal cost) reflects the opportunity cost of using the generator's energy source in some other way or at some other time. This opportunity cost may change from moment to moment, shifting the marginal cost function up or down. For example, where there is a spot market for the fuel used in the generator, the price of the fuel on that spot market may change from one time interval to the next.

Even for those generators that have access to fuel under long-term contracts at fixed prices, as long as the generator has the opportunity to sell any fuel it does not use on the fuel spot market, or the opportunity to buy any additional fuel on the fuel spot market, the relevant opportunity cost for the generator is not the long-term or contract price for the fuel, but the short-run fuel spot price. When the fuel spot price is high the generator may find it more profitable to produce less electricity and to sell its input fuel on the spot market. Therefore, even for generators that have long-term fuel price contracts, if those generator have access to a short-term market on which they can buy and sell their input fuel, the marginal cost function will not be static but will depend on the spot price of the fuel.

Some generators, particularly some hydroelectric generators, have a limited stock of energy available to them. In the case of hydroelectric generators, the limited stock of energy is reflected in the limited stock of water in high-elevation lakes. Such generators are known as *energy-limited generators*. For energy-limited generators, the opportunity cost of producing at one instant in time is the foregone profit from not being able to produce at another point in time. Therefore, the marginal cost function of such generators depends on the potential value of the output produced at other times.

For these generators there is a very significant time-interdependence in their costs of production. To model the optimal use of these generators, we should consider the optimal production strategy over a period of time. If the price of electricity is expected to be higher tomorrow, every extra MW produced today implies there will be less energy available to produce electricity tomorrow. As a result, the variable cost of the electricity produced today is the foregone profit from production tomorrow. We will come back to this issue in Section 4.10.

In addition, as noted earlier, we have put to one side the issue of startup costs, which we will return to later in this chapter in Section 4.10.

It is worth noting that the cost function of a typical generator is not necessarily convex. A function is said to be convex if, when we take the weighted average of the function at two different points, the weighted average lies above the function evaluated at the weighted average of the two points. The cost function set out earlier is nonconvex due to (a) the presence of fixed auxiliary costs, which can be avoided when the output reduces to zero, and (b) the presence of a minimum operating level, which is greater than zero.

As we will see later, these nonconvexities complicate the task of finding the optimal combination of output of different generators. Even when the optimal combination of outputs exist, with nonconvexities some generators may not be able to cover their costs. For these reasons, it is common in electricity market analysis to assume these problems away. This is equivalent to assuming that $\underline{Q} = 0$ and $b = 0$ in the earlier cost function. This is precisely what we will do in the next section.

4.2 Simple Stylised Representation of a Generator

For our purposes it is useful, at least at the outset, to assume a particularly simple stylised shape for the marginal cost function of generators. In particular we will assume that there is no minimum operating level, no fixed auxiliary costs, and constant marginal costs of operation up to the maximum operating level.

In mathematical notation, we will assume that the minimum operating level is zero, and the parameters a and b in the cost function mentioned earlier are also zero. The maximum operating level will be said to be the generator's *capacity* and will be denoted K. In other words, we will assume $\underline{Q} = 0$ and $\overline{Q} = K$. The cost function of the generator is then simply: $C(Q) = cQ + F$ for $0 \leq \overline{Q} \leq K$.

Under these assumptions the marginal cost curve for the generator is flat (horizontal) up to the generator's maximum operating level at which point the marginal cost curve becomes vertical. Since there is no ambiguity, we can refer to the generator's marginal cost in the flat part of the curve as the generator's *variable cost*. This marginal cost is illustrated in Figure 4.2.

Even in those cases where the marginal cost function of a generator is not flat over a wide range of output, it is often the case that the marginal cost function can be approximated by a 'step function' – that is, a function that is flat over a range of output. Mathematically this is equivalent to breaking a single generator up into smaller units each of which has a constant marginal cost, allowing us to use the simple stylised representation of a generator as set out in Figure 4.2. In principle, this approximation can be made arbitrarily accurate.

Figure 4.3 shows how the cost function of a generator might be approximated using three hypothetical generators each with a constant marginal cost.

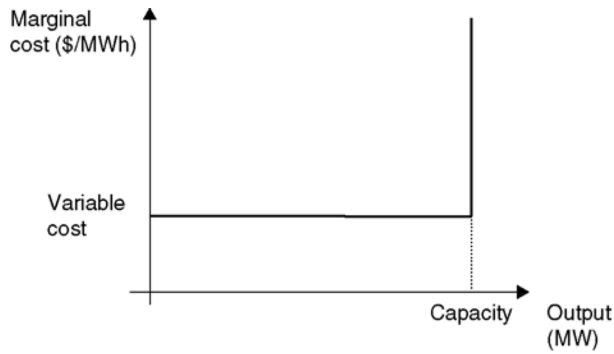

Figure 4.2 Stylised representation of the marginal cost of a generator

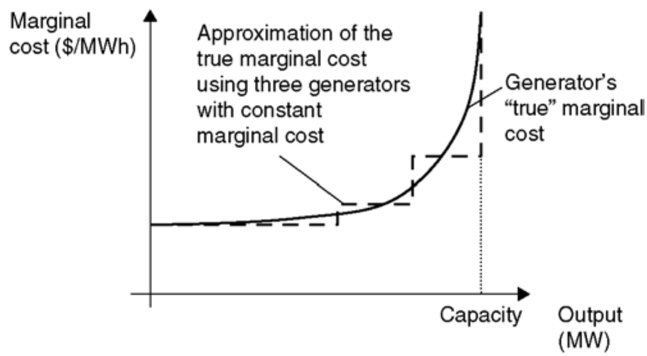

Figure 4.3 Approximation of an arbitrary marginal cost curve using generators with constant marginal cost

In practice, many liberalised electricity markets require generators to submit an offer curve (that is, a statement of how much they are willing to produce at a given price) that is a step function of just this kind.

4.3 Optimal Dispatch of Generation with Inelastic Demand

Let us consider first a power system with no physical limits except one – the constraint that total power produced must equal total power consumed. We will ignore the transmission and distribution networks entirely – in effect, we will assume that all production and consumption of electricity takes place at a single location.

In addition, in this section, we will ignore any controllable consumption assets. We will consider these further in Section 4.4. Finally, to keep things simple, let us put to one side, at least temporarily, the nonconvexities mentioned earlier. Specifically, we will assume that the cost function of every generator $C_i(Q_i)$ is convex. This implies, amongst other things, that there is no minimum operating level. We will also ignore startup costs.

Let us consider the task of an omniscient system operator who has a portfolio of N generators over which it has perfect information and total control. The system operator is able to direct each generator as to how much it should produce. If generator i is directed to produce energy at the rate Q_i (MW) it incurs cost at the rate $C_i(Q_i)$ ($/h). The total demand for electricity is Q (MW).

The set of all the orders given by the system operator to the generators is known as the *dispatch* (also often spelled 'despatch'). The dispatch is a vector $Q = (Q_1, Q_2, \ldots, Q_N)$, which specifies for each generator the total rate at which it is to produce energy.

The task of the system operator is to find the dispatch that minimises the total cost of generation subject to the constraint that the total amount of generation must equal the total load. Let us assume for the moment that the total demand is less than the total capacity of the generators to produce so we do not have to worry about rationing.

4.3.1 Optimal Least Cost Dispatch of Generation Resources

The task of the system operator is to find a dispatch that minimises the total overall cost of meeting demand. In other words, the task of the system operator can be written as follows:

$$\min \sum_{i=1}^{N} C_i(Q_i)$$

$$\text{subject to } \sum_{i=1}^{N} Q_i = Q \leftrightarrow \lambda$$

$$\text{and } \forall i,\ Q_i \leq K_i \leftrightarrow \mu_i \text{ and } Q_i \geq 0 \leftrightarrow \nu_i$$

The KKT conditions for this problem are as follows:

$$\forall i,\ \ C_i'(Q_i) = \lambda - \mu_i + \nu_i$$

In addition,

$$\forall i,\ \mu_i \geq 0 \text{ and } \mu_i(Q_i - K_i) = 0$$

$$\forall i, \nu_i \geq 0 \text{ and } \nu_i Q_i = 0$$

From these expressions we can see that for each generator there are three possibilities: Either the generator is dispatched to an intermediate level (between its maximum and minimum output, so that $\mu_i = 0$ and $\nu_i = 0$) in which case the generator is dispatched to a point where its marginal cost is equal to some common value:

For each i for which $0 < Q_i < \overline{Q}_i$, $C_i'(Q_i) = \lambda$ for some constant λ.

Alternatively, the generator is dispatched to its maximum output $Q_i = K_i$, in which case its marginal cost is smaller than the common value $C_i'(Q_i) \leq \lambda$, or the generator is dispatched to its minimum output $Q_i = 0$ in which case the marginal cost is greater than the common value: $C_i'(Q_i) \geq \lambda$.

The common marginal cost λ is known as the *system marginal cost* or SMC. Note that the SMC is equal to the marginal cost of producing an additional unit to meet an additional unit of demand.

We can define the total cost of the optimal dispatch as $C(Q) = \sum_{i=1}^{N} C_i(Q_i)$. If the cost function of each generator is convex, the total cost is convex in the total load.

> *Result*: Given a set of controllable generators with convex cost functions and upper and lower bounds on production, the least cost dispatch (ignoring network constraints) has the following characteristics:
>
> a. For every generator that is despatched to a rate of production that lies between its minimum and maximum operating level, the marginal cost of each generator is the same; this common value is known as the 'system marginal cost' or SMC;
> b. All generators that are despatched to a target of zero have a marginal cost that is above SMC, and all generators that are despatched to their maximum operating level have a marginal cost that is below the SMC.
> c. The sum of the output of all generators is equal to the total demand.
>
> In addition, it is straightforward to demonstrate that (i) the total cost of generation is convex in the total system load Q, (ii) the SMC is monotonically increasing in the system load Q, and (iii) the output of each generator is monotonically increasing in both the SMC and load.

4.3.2 Least Cost Dispatch for Generators with Constant Variable Cost

The results just mentioned hold whatever the shape of the marginal cost function of each generator. Let us focus now on the special case where the marginal cost curve of each generator takes the simple stylised form of Figure 4.2.

Specifically, let us assume that when there are K_i units of capacity of generation of type i, and when that type of generation is producing at the rate Q_i, costs are incurred at the rate

$$C_i(Q_i) = c_i Q_i \text{ for } 0 \leq Q_i \leq K_i$$

The social optimisation problem is as follows:

$$\min \sum_i c_i Q_i$$

$$\text{subject to } \sum_i Q_i = Q \leftrightarrow \lambda$$

$$\text{and } \forall i, Q_i \leq K_i \leftrightarrow \mu_i \text{ and } Q_i \geq 0 \leftrightarrow \nu_i$$

The KKT conditions for this problem include the following condition:

$$\forall i, c_i = \lambda - \mu_i + \nu_i$$

From the KKT conditions we find that the least cost dispatch has the following properties: Each generator can be ranked in order, from the generator with the lowest variable cost to the generator with the highest variable cost. This is known as the *merit order*. Intuitively, the system operator can then work its way up the merit order dispatching each generator up to its maximum capacity in order until all demand is satisfied.

> *Result*: In the least cost dispatch of a set of controllable generators with a constant variable cost and ignoring network constraints, each generator is dispatched according to the merit order. Generators are dispatched in order from the lowest-variable-cost to the highest, working up the merit order until all demand is satisfied.
> More formally, for each generator:
>
> a. if the variable cost of the generator is below the SMC, the generator is dispatched for its full available capacity;
> b. if the variable cost of the generator is above the SMC, the generator is not dispatched at all;
> c. if the variable cost of the generator is equal to the SMC, the output of the generator is indeterminate, but the total output of all generators is equal to the total load.

Intuitively, the optimal dispatch can be found as follows. We can sum the marginal cost curves for each generator horizontally to find the *industry supply curve*. When the marginal cost curves take the simple stylised form of Figure 4.2 the industry supply curve is a simple step function. We can then find the intersection between this industry supply curve and the (vertical) demand curve. This intersection yields the SMC.

The optimal output for each generator is where the marginal cost of each generator is equal to the SMC. The output of each generator is either at its maximum capacity, if the generator's variable cost is below the SMC, or zero, if the generator's variable cost is above the SMC, or at some intermediate level when the generator's variable cost is equal to the SMC (Figure 4.4).

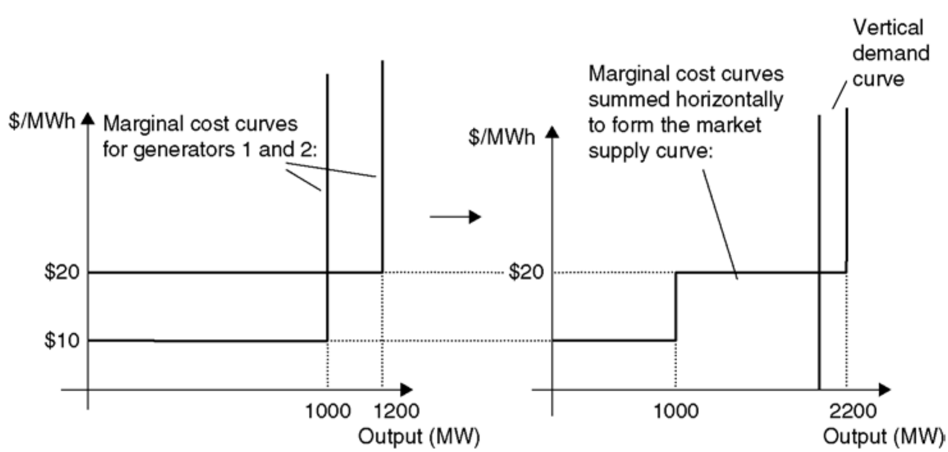

Figure 4.4 Intersection of market supply and market demand gives market-clearing price

Table 4.1 Cost data for a simple three-generator electricity supply industry

Generator	Variable Cost ($/MWh)	Capacity (MW)
A	$10	1000
B	$25	500
C	$100	200

As we will see later, in a liberalised electricity market, the SMC will correspond to the wholesale spot price for electricity. However, nothing we have said so far relies on the existence of a wholesale spot market. The same principles would apply for a vertically integrated electricity industry as for a liberalised market.

We can say that the SMC is always equal to the marginal cost of every generator that is dispatched for a positive amount. Strictly speaking, however, the marginal cost of some generators is indeterminate at their maximum capacity (indicated here by the vertical marginal cost curve at this point). However, mathematically we can define a left-hand marginal cost as the marginal cost saving of a small reduction in output and a right-hand marginal cost as the marginal cost addition from a small increase in output. It is perhaps more correct to say that the SMC lies between the left-hand marginal cost and the right-hand marginal cost for all generators. However, if we allow the marginal cost to take a range of values at this point (all values between the left-hand marginal cost and the right-hand marginal cost) we can make the assertion that the SMC is always equal to the marginal cost of every generator that is dispatched for a positive amount.

The SMC is sometimes said to be equal to the variable cost of the last generator to be dispatched – which is also known as the *marginal generator* or sometimes as the *price setter* (this last term is something of a misnomer since, of course, it is the output of all generators combined, together with the level of demand, which sets the level of the price).

4.3.3 Example

For example, suppose we have an electricity industry with three generators with constant variable cost and capacities as shown in Table 4.1.

Suppose that load varies between 500 and 1700 MW. Given this information about supply and demand conditions we can work out the dispatch of each generator for each level of demand, as shown in Table 4.2. For example, when the load is 1100 MW, generator A is dispatched up to its maximum capacity (1000 MW) and generator B is dispatched for the remainder (100 MW). Since generator B is the marginal generator, the SMC (the wholesale spot price) is $25/MWh.

Table 4.2 The optimal dispatch and SMC for different levels of demand in a simple three-generator industry

Load (MW)	Dispatch (MW)			SMC (Price) ($/MWh)
	Gen A	Gen B	Gen C	
$500 \leq Q < 1000$	Q	0	0	$10
1000	1000	0	0	$\$10 \leq P \leq \$25$
$1000 \leq Q < 1500$	1000	$Q-1000$	0	$25
1500	1000	500	0	$\$25 \leq P \leq \$100$
$1500 \leq Q < 1700$	1000	500	$Q-1500$	$100
1700	1000	500	200	$P \geq \$100$

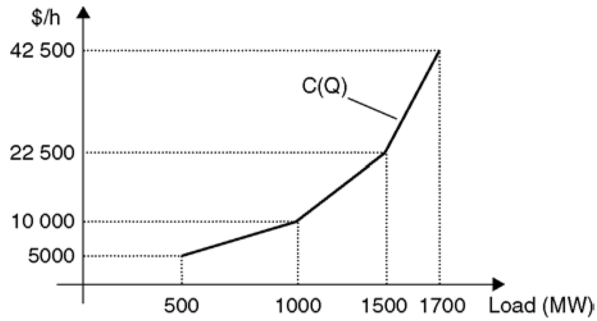

Figure 4.5 Cost function for the simple three-generator case

We can also express the total cost of the optimal dispatch in this example as follows:

$$C(Q) = \begin{cases} 10Q, & 500 \le Q \le 1000 \\ 25Q - 15{,}000, & 1000 \le Q \le 1500 \\ 100Q - 127{,}500, & 1500 \le Q \le 1700 \end{cases}$$

As we noted before, this function is convex in the total demand, Q, as illustrated in Figure 4.5.

4.4 Optimal Dispatch of Both Generation and Load Assets

Section 4.3 focused on the optimal use of a set of generation assets to serve a given quantity of load. However, this analysis is incomplete. Both production and consumption assets can, in principle, respond to wholesale market conditions. Overall efficiency requires that we consider efficiency in the use of both production and consumption assets.

As in Part I, we will assume that each electricity customer has a utility function from the consumption of electricity $U_i(Q_i)$. Each customer is assumed to be a price-taker on the electricity market. As we know from Part I, if each consumer faces a simple linear price for electricity, we can derive a downward sloping demand curve for each customer.

In Section 4.3, the task of the market operator was to choose a set of production rates (one for each generator), known as the dispatch, which minimises the overall cost of meeting demand.

Now the market operator must choose a combination of a rate of production for each generator and a rate of consumption for each consumer, which maximises the total economic welfare. In other words, the dispatch now has two components. We can write the optimal dispatch as (Q^S, Q^B) where, as before, $Q^S = Q_1^S, Q_2^S, \ldots$ is a vector that specifies for each generator the total rate at which that generator is to produce energy. In addition, $Q^B = Q_1^B, Q_2^B, \ldots$ specifies the rate at which each electricity consumer is to consume electrical energy.

For simplicity, let us temporarily set aside the generator production bounds (this is without loss of generality since these bounds are embodied in the generator cost function in any case).

The problem of finding the efficient use of a set of production and consumption assets was considered in Section 1.4. Here we apply that analysis to the case of the electricity market.

Given a set of generation and consumption resources, the task of the system operator is to find a dispatch that minimises the total economic surplus. In other words, the task of the system operator can be written as follows:

$$\max \sum_i U_i(Q_i^B) - \sum_i C_i(Q_i^S)$$

Subject to

$$\sum_i Q_i^B = \sum_i Q_i^S \leftrightarrow \lambda$$

This problem has the KKT conditions

$$\forall i,\ U_i'(Q_i^B) = \lambda \text{ and } \forall i, C_i'(Q_i^S) = \lambda \text{ and } \sum_i Q_i^B = \sum_i Q_i^S$$

In other words, this problem has the following solution: the point at which the demand and supply curves intersect yields a price. All generators produce at a rate where their marginal cost is equal to this common price or SMC. In exactly the same way, loads are dispatched to the point where their marginal valuation is equal to the SMC.

Result: The welfare-maximising dispatch of a set of controllable electricity production and consumption assets has the following characteristic: There is a common system-wide marginal cost. Each generator produces at a rate where the marginal cost is equal to the common system-wide marginal cost. Each consumer consumes at a rate where the marginal value of consumption (the point on the demand curve) is equal to the common system-wide marginal cost.

For example, it might be that there are three different types of consumers in the market. Consumers of type A are prepared to consume at the rate of up to 200 MW of electricity and value that electricity at \$200/MWh. Consumers of type B are prepared to consume at a rate of up to 300 MW of electricity and value that electricity at \$100/MWh. Consumers of type C are prepared to consume up to 50 MW of electricity and value that electricity at \$20/MWh. The resulting market demand curve (the sum of the individual demand curves) is illustrated in Figure 4.6.

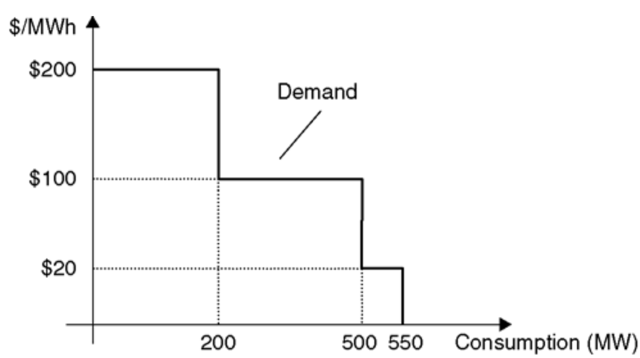

Figure 4.6 Market demand curve for a simple market

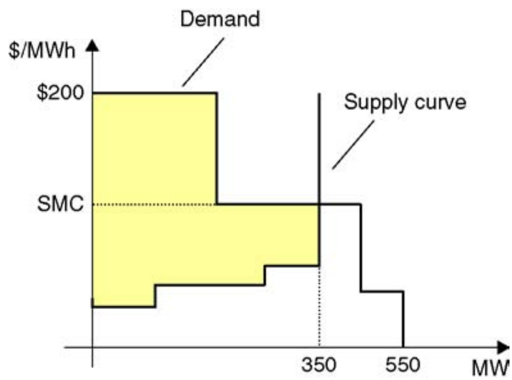

Figure 4.7 Illustration of optimal dispatch in a simple electricity market

Given a market supply curve such as the one in Figure 4.7, we can determine the optimal dispatch. From Figure 4.7 we see that the SMC is equal to \$100/MWh. The generators are dispatched according to this SMC as before. In addition, customers of type A consume at the rate of 200 MW. Customers of type B consume at some intermediate rate, while customers of type C choose not to consume at all. The total economic surplus is the shaded area in Figure 4.7.

4.5 Symmetry in the Treatment of Generation and Load

4.5.1 Symmetry Between Buyer-Owned Generators and Stand-Alone Generators

In the previous analysis we have assumed that consumers (loads) have some controllable consumption for electricity given by the utility function, without enquiring further into how that controllable consumption comes about.

This controllable consumption could arise from an interruptible production process or from the deferment of consumption in cases where there is substantial thermal inertia (i.e. when the electricity is used for heating or cooling). However, in addition, an important source of controllability in the level of load arises when the buyer has an on-site controllable generator. In this case, even if the underlying load of the buyer is fairly stable, the net load as it appears to the market (that is, the underlying load less the on-site production) can appear controllable.

As we noted in Part II, one of the key potential future developments in electric power systems is the increasing penetration of small-scale generation including roof-top solar PV, small wind generation and local storage. A key question therefore, is whether the earlier analysis should treat this generation differently.

Does the analysis distinguish in any significant way between stand-alone generation and generation integrated with a buyer? The simple answer is no. We can treat the utility function of an integrated load-generator as equal to the underlying utility from the total electricity consumption less the cost of electricity produced on site. Specifically, if the underlying utility function is denoted $V(Q)$ and the local cost of production is $C(Q)$, then the utility reported to the central market operator is as follows:

$$U(Q) = V\left(Q + Q^S\right) - C\left(Q^S\right)$$

If the buyer operates the local generator in such a way as to maximise his/her utility then he/she will choose the rate of production for which the marginal value of consumption is equal to the marginal cost of production:

$$V'\left(Q+Q^S\right)=C'\left(Q^S\right)$$

Moreover, this common marginal value and marginal cost is equal to the common marginal value and marginal cost for the other generators and loads in the market. This is the condition for the overall efficient outcome above.

In other words, there is no difference in the treatment of integrated generators and stand-alone generators. We can combine a generator with a load and the combined entity can be treated the same way in the market as a separate load and a stand-alone generator.

4.5.2 Symmetry Between Total Surplus Maximisation and Generation Cost Minimisation

We are focussing on the question of the efficient use of a given set of production and consumption assets. We have viewed this as the problem of finding the dispatch that maximises total surplus, but we can also view this problem as a minimisation problem very similar to the problem of minimising the total cost of generation as we saw in Section 4.3.

Let us suppose that, at the price of zero, the demand for electricity by customer *i* is finite, and takes the value $Q_i^{\mathrm{B\ max}}$. This implies that the utility function $U_i(Q_i^{\mathrm{B}})$ is bounded above. Let the maximum of this function be $U_i^{\max}$. Let us define a new variable $Q_i^{\mathrm{Sdr}}=Q_i^{\mathrm{B\ max}}-Q_i^{\mathrm{B}}$ and define a new function $V_i(Q_i^{\mathrm{Sdr}})=U_i^{\max}-U_i(Q_i^{\mathrm{B\ max}}-Q_i^{\mathrm{Sdr}})$. This function $V_i(\cdot)$ is positive and upward sloping (and has a positive second derivative), just like a generator cost function.

In fact we can think *any* demand-side responsiveness to the wholesale price as being exactly equivalent to a generator. We can imagine that the demand for each customer is represented as a single, fixed, demand of Q_i^{Bmax}, which is valued at the amount $U_i^{\max}$. Any responsiveness of the customer to the wholesale market is reflected in the form of a generator with output Q_i^{Sdr} and cost function $V_i\left(Q_i^{\mathrm{Sdr}}\right)$. The task of the market operator can then be viewed as simply minimising the cost of generation.

> *Result*: In the economic analysis of power systems there is no need to draw a distinction between controllable generation, controllable loads, or sites that include both generation and consumption. Economically, we can treat the market as consisting of all controllable generation or all controllable loads. Similarly, we can allow colocated generation and load assets to be treated collectively (as either a generator or load) or separately as a generator separate from a load. (Later we will see that this result depends on the assumption that generators and loads face the same market price).

4.6 The Benefit Function

We can express the mathematical problem of finding the optimal dispatch of generation and consumption resources in a slightly different way. This alternative way is mathematically identical to the problem set out in Section 4.5. However, this alternative formulation proves more convenient when we introduce network constraints in Chapter 5.

This alternative formulation works as follows: Consider a partition of the set of producers and consumers.[1] The jth partition will be labelled N_j. Let us define the net injection of the producers and consumers in the jth partition as the total rate of production of all the producers in this partition less the total rate of consumption of all the consumers:

$$Z_j = \sum_{i \in N_j} Q_i^{S} - \sum_{i \in N_j} Q_i^{B}$$

Now let us define the benefit function for each partition as follows: The benefit function for the jth partition for the net injection Z_j is the maximum level of economic surplus that can be achieved for the producers and consumers in the jth partition while holding constant the net injection Z_j. In other words, the benefit function is defined as follows:

$$B_j(Z_j) = \max_{Q_i^{B}, Q_i^{S}} \sum_{i \in N_j} U_i(Q_i^{B}) - \sum_{i \in N_j} C_i(Q_i^{S})$$

$$\text{subject to } Z_j = \sum_{i \in N_j} Q_i^{S} - \sum_{i \in N_j} Q_i^{B} \ \leftrightarrow \ \alpha_j$$

As before, at the optimum, the marginal valuation of each customer in the jth partition and the marginal cost of each supplier in the partition are the same and equal to α_j. Furthermore, an increase in the net injection for this partition by a small amount increases the benefit function by the amount α_j:

$$B_j'(Z_j) = \alpha_j$$

Importantly, the optimal dispatch task can be rewritten more simply in terms of the benefit function. The optimal dispatch task is now:

$$W = \max_{Z_j} \sum_j B_j(Z_j)$$

$$\text{subject to } \sum_j Z_j = 0$$

4.7 Nonconvexities in Production: Minimum Operating Levels

In the previous sections we assumed, for convenience, that the cost function of each generator was convex. However, in practice, many generators have a minimum level of output below which they cannot physically operate. This introduces a new set of constraints into the optimal dispatch problem. How do these new constraints affect the optimal dispatch?

It turns out that the presence of minimum operating levels affects the optimal dispatch in four ways:

- It may not simply be feasible to meet certain levels of demand;
- As load increases, some generator's output may need to be *reduced* (rather than increased) in the optimal dispatch;

[1] A partition of set A is a set of mutually dis-joint subsets with a union equal to the original set A.

Table 4.3 Key cost data for a simple illustration of the impact of nonconvexities

Generator	Variable Cost ($/MWh)	Minimum Operating Level (MW)	Maximum Operating Level (Capacity, MW)
A	$10	250	500
B	$20	260	500

- The marginal or price-setting generator may not be the 'last' generator in the merit order to be dispatched;
- The total cost of the optimal dispatch may not be convex in the load.

These results can be illustrated using the following simple example. Let us suppose an electricity industry has just two generating units, each with a capacity of 500 MW. Unit A has a constant variable cost of $10/MWh, and a minimum operating level of 250 MW. Unit B has a constant variable cost of $20/MWh and a minimum operating level of 260 MW, as summarised in Table 4.3.

Clearly, when demand is below 250 MW, there is no feasible dispatch – that is, there is no dispatch that satisfies the operating constraints of the generators.

If demand is larger than 250 MW but less than 500 MW, the load is met by unit A. However, what about if load increases to 501 MW? This load exceeds the capacity of unit A to supply alone, so we need to increase the output of unit B. However, the minimum output of unit B is 260 MW. If we dispatch unit B to 260 MW, we would have to reduce the output of unit A to 241 MW, which is below its minimum operating level. We see that for loads between 500 and 510 MW there is, again, no feasible dispatch.

For loads larger than 510 MW and up to 760 MW, we can dispatch unit B for its minimum (260 MW), and then dispatch unit A for the remainder. Note that unit A is, in this case, the marginal generator, even though unit B is also dispatched. For loads larger than 760 MW, up to 1000 MW, unit A is dispatched for 500 MW, and unit B is dispatched for the remainder.

Figure 4.8 illustrates the shape of the total cost of optimal dispatc as a function of system load. As we can see, this function is no longer convex.

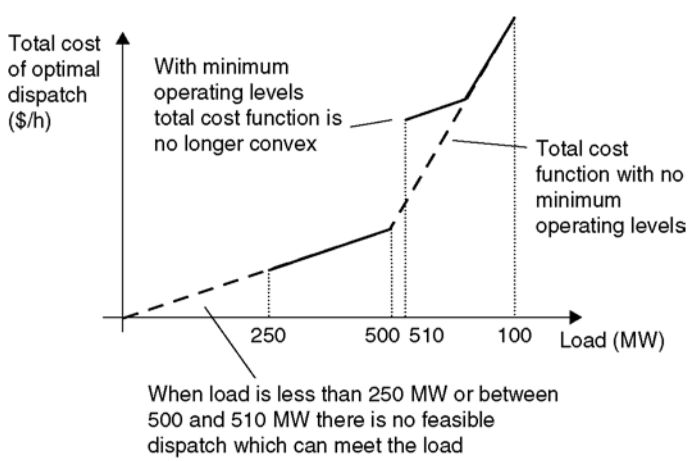

Figure 4.8 Cost of dispatch in the presence of nonconvexities

4.8 Efficient Dispatch of Energy-Limited Resources

Some generators have only a limited stock of their fuel or key energy input on hand, and cannot easily obtain more at short notice. These generators are known as *energy limited*. As we will see, it will typically not make sense to put an energy-limited generator in the merit order at the variable cost of obtaining its key input. Instead, as we will see, we have to take into account the *opportunity cost* of not being able to produce at other times.

The classic example of an energy-limited generator is a hydro-generator with a limited stock of water held back in a dam. Once the dam is full, any additional in-flows must be spilled down the river. Since this water cannot be stored and used at any other time, the opportunity cost of using any further inflows of water is zero – the hydro-generator may as well use this water for generating electricity provided the spot price for electricity exceeds the variable cost of converting that water into electricity (which is typically close to zero for a hydro-generator).

Let us focus on the case where there are no more inflows forecast over the period in consideration so that generating electricity at the current time involves consuming water that could otherwise be used to generate electricity at some other time. What is the most efficient way to use this resource of stored water?

If the water in the dam is used for generating electricity at a particular time, it cannot be used for generating electricity at another time. The opportunity cost of using the water for generation at a particular point in time is the value of that water in generation at some other time.

Let us suppose, for simplicity, that the variable cost of the energy-limited generator is close to zero. This is a reasonable approximation for hydro-generators.

It turns out that the most efficient use of the energy-limited resource is to use the resource to generate at those times when the SMC is the highest over the period in question. In other words, if a hydro-generator captures enough water to produce, say, 1000 MWh of output each week, the hydro-generator should generate those 1000 MWh precisely at those times when the SMC is the highest.

Note that this implicitly requires a degree of *future price forecasting and intertemporal optimisation*. In deciding whether or not to generate 1 MWh from a hydro resource, an efficient market operator must look at the generation that will be displaced by that 1 MWh of production today and the generation that will be displaced by that 1 MWh of production at some point in the future. It makes no sense to use an energy-limited resource to displace 1 MWh of low-cost generation today if that same resource could be used to displace 1 MWh of very high cost generation at a time in the near future.

Let us modify the optimal dispatch problem set out earlier, to incorporate the possibility of energy-limited resources. Since this is an intertemporal optimisation problem we now have to keep track of the time dimension of each variable. As before, let us suppose we have N generating units, operating over T periods. Suppose that the duration of period t is represented by σ_t (with units of time). The total demand in period t is Q_t (MW). The output of generating unit i in period t is Q_{it} (MW). Let us assume that the physical characteristics of each generating unit remains the same over time – that is, the marginal cost of generating unit i is $C_i(Q_{it})$ (\$/MWh) over the range $0 \leq Q_{it} \leq K_i$. In addition, the Nth generator is assumed to face an energy constraint, that its total output over the T periods does not exceed an energy limit E_N.

The task of the system operator can now be written as follows:

$$\text{Min} \sum_t \sigma_t \sum_i C_i(Q_{it})$$

$$\text{Subject to } \forall t, \sum_{i=1}^{N} Q_{it} = Q_t \ \leftrightarrow \ \sigma_t \lambda_t$$

$$\text{And } \forall i, t, Q_{it} \leq K_{it} \leftrightarrow \sigma_t \mu_{it} \text{ and } Q_{it} \geq 0 \leftrightarrow \sigma_t \nu_{it}$$

$$\text{And } \sum_t \sigma_t Q_{Nt} \leq E_N \leftrightarrow \gamma$$

The KKT conditions are very similar to the problem solved earlier (see Section 4.3) except for the *N*th generator. The KKT condition for the *N*th generator is as follows:

$$C'_N(Q_{Nt}) = \gamma + \lambda_t - \mu_{it} + \nu_{it}$$

In other words, if the energy constraint on the energy-limited generator is not binding, the optimal dispatch is exactly as before (recall that every generator is dispatched according to merit order; all generators that are dispatched for an intermediate level of output have the same marginal cost). However, if the energy constraint is binding, the optimal dispatch is the same as before except that the energy-constrained generator is dispatched as though it has a marginal cost that is above its true or underlying marginal cost (by the amount γ). In effect, the energy-constrained generator is dispatched according to its opportunity cost, not its true marginal cost.

Result: When a generator is limited in the amount of energy it can produce in a given period of time, the optimal dispatch for that generator should reflect not the generator's true marginal cost of production, but a higher value reflecting the opportunity cost of not being able to produce at other times.

4.8.1 Example

An example will make this clearer. Let us suppose we have an energy market with three conventional thermal generators, each with a capacity of 500 MW. These generators have a variable cost of, say, \$10/MWh, \$50/MWh and \$100/MWh, respectively. In addition there is a single energy-limited generator. Let us suppose each day is divided into six periods of equal length. The demand over these six periods is, say, 400, 800, 1200, 900, 700 and 200 MW.

What is the optimal way to dispatch these generators? The solution to this problem is illustrated in Figure 4.9. As before, we construct a merit order, from the lowest cost to the highest cost. The lowest-cost generator is dispatched first, and then the next-to-lowest and so on. However, when should the energy-limited generator be dispatched?

If the energy-limited generator has 500 MWh of energy available, the efficient dispatch is for the energy-limited generator to only produce at the peak period (period 3). At this time it

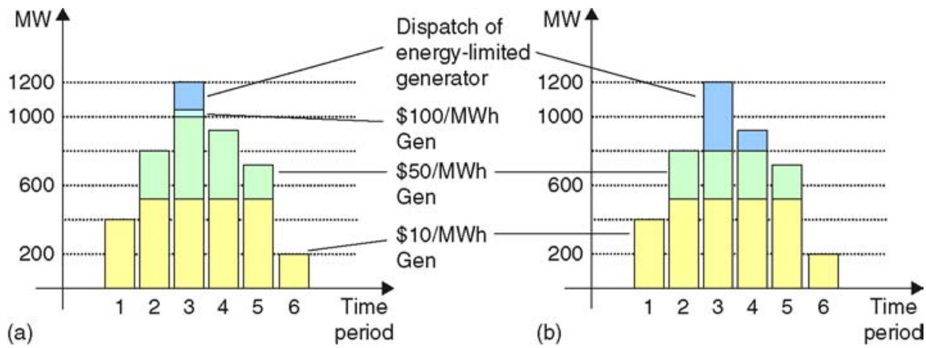

Figure 4.9 (a) Energy-limited generator has 500 MWh of energy available. (b) Energy-limited generator has 2000 MWh of energy available

displaces the highest cost generator. It can produce at the rate of 125 MW over this 4 h period, producing 500 MWh. This is illustrated in Figure 4.9a.

If the energy-limited generator has 2000 MWh of energy available, an efficient dispatch is for the energy-limited generator to produce at both the peak period (period 3) and the next-highest peak (period 4). It produces at the rate of 400 MW in period 3 and 100 MW in period 4, for a total of 2000 MWh of energy produced. In this case the energy-limited generator completely displaces the highest-cost generator in the peak period and partly displaces the $50/MWh generator in the next highest period. This is illustrated in Figure 4.9b.

4.9 Efficient Dispatch in the Presence of Ramp-Rate Constraints

Generators vary in how quickly they can increase or decrease their output. Some generators, such as large coal-fired generators, can only increase their output by a few MW every minute. On the other hand, some other generators, such as hydro-generators, are able to respond much more quickly to changes in the supply–demand balance. In practice, these *ramp rate constraints* must also be taken into account in the optimal dispatch.

In the presence of ramp rate constraints, it may be necessary to dispatch generators out of merit-order for a period of time. This can lead to large swings in the wholesale spot price.

Let us consider again the optimal dispatch task we have seen above, but this time let us explore the impact of ramp-rate constraints. Let us suppose that we have a power system with a number of generators. Let us suppose that the output of each generator is initially in some steady state Q_{i0}, and the total demand is Q_0. The power system then evolves over time $t = 1, \ldots, T$, following the demand Q_t. During this evolution ramp rate constraints may be binding. Let us suppose that at time $t = T$ the ramp rate constraints are no longer binding. The optimal dispatch task is now formulated as follows:

$$\min \sum_{i,t} c_i Q_{it}$$

$$\text{subject to } \forall t, \sum_{i} Q_{it} = Q_t \leftrightarrow \lambda_t$$

and

$$\forall i, t, Q_{it} \leq K_{it} \leftrightarrow \mu_{it} \text{ and } Q_{it} \geq 0 \leftrightarrow \nu_{it}$$

and

$$\forall i, t,\ Q_{it} - Q_{it-1} \leq R_i \leftrightarrow \tau_{it}$$

Here, R_i is the (upward) ramp rate constraint on generator i.

The KKT conditions for this problem include the following:

$$\forall i, t, c_i = \lambda_t - \mu_{it} + \nu_{it} - \tau_{it} + \tau_{it+1}$$

Paradoxically, when ramp-up constraints will be binding in the future, the wholesale spot price (i.e. the SMC) can drop to very low levels. Intuitively, the reason is that when ramp-up constraints will be binding in the future, adding more production today reduces the duration of the ramping period, reducing costs in subsequent periods as long as those ramp rate constraints are binding.

4.9.1 *Example*

To illustrate this let us suppose we have an electricity industry with 1000 MW of generation with a marginal cost of \$10/MWh, with a low ramp rate of, say, 100 MW every 5 min. In addition, let us suppose we have 500 MW of generation with a marginal cost of \$20/MWh with a ramp rate of, say, 75 MW every 5 min, and 500 MW of generation with a marginal cost of \$100/MWh and a ramp rate of say 500 MW every 5 min. (This fast-ramping generation could be energy-limited hydro-generation in which case, as we know from the discussion in Section 4.8, this marginal cost is the 'opportunity cost' of generation rather than the true marginal cost). For simplicity, let us ignore other startup costs or minimum operating levels. The generation assets are summarised in Table 4.4.

Let us suppose that demand is initially at 500 MW and in a steady state. This load can be supplied entirely by the low-cost generator. The price is \$10/MWh. However, let us suppose that it is known that at a particular point in time (say $t = 3$) demand will increase over 5 min to (just under) 1000 MW (later, in Section 12.2, we will deal with the case where demand varies in an uncertain manner).

At time $t = 3$, as demand increases, the lowest-cost generation cannot increase from 500 to 1000 MW in a single 5 min period. It can only increase from 500 to 600 MW. We can also increase the output of the next-lowest-cost generation from 0 to 100 MW in the 5 min period, increasing total output to 700 MW. We must use the highest-cost generator to provide the remaining 300 MW of output.

Table 4.4 Key cost data for a simple illustration of the impact of ramp rate limits

Generator	Variable Cost (\$/MWh)	Maximum Ramp Rate (MW/5 min)	Capacity (MW)
A	\$10	100	1000
B	\$20	75	500
C	\$100	500	500

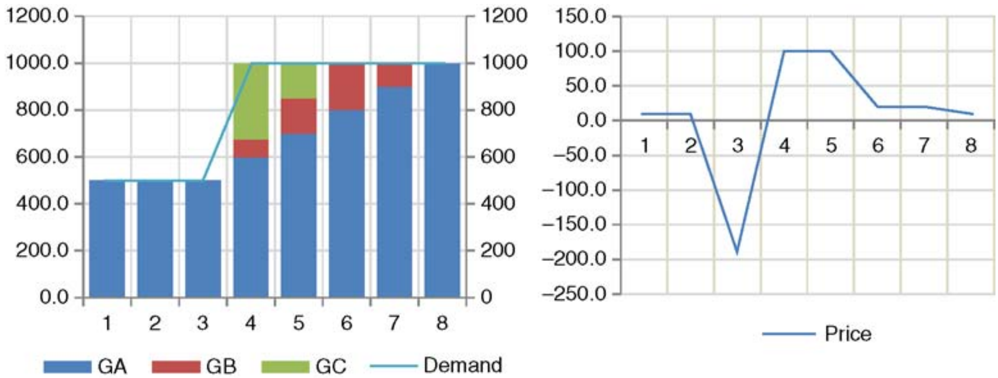

Figure 4.10 Efficient pricing and dispatch outcomes in the presence of ramp rate constraints

Importantly, the spot price at time $t = 3$ (before the increase in demand occurs) is \$−190/MWh. The reason is that any increase in production by generator A at this point increases the cost of dispatch by \$10/MWh in that period, but reduces the dispatch cost over the next several intervals – by \$10/MWh in periods 6 and 7 and by \$90/MWh in periods 4 and 5.

In period 4 any additional output must be met by the \$100/MWh generation, so the price in this period is \$100/MWh. The price remains at this level until the ramp rate constraints on generator B are no longer binding, at which point the price drops to \$20/MWh. These results are illustrated in Figure 4.10.

The key result here is the counter-intuitive variation in the price. The price drops to a low level (here the price is negative) in the period *before* the ramp rate constraints start to bind.

In principle, a similar outcome can arise in the event of a rapid decline in demand. Again, certain generators, especially large thermal generators, may not be able to efficiently reduce their output rapidly. In this case the price may temporarily increase to significantly above its normal level.

Even more importantly, it turns out that if a period of binding ramp rate constraints is anticipated in the future, the optimal dispatch may require that this episode be anticipated. Specifically, it may be efficient to ramp up a low-ramp-rate plant in advance in order to reduce the cost of transition to the new steady-state equilibrium.

Actions that are taken by the power system to adjust to a new steady-state equilibrium after an event occurs are known as *corrective actions*. Actions taken in advance of an event occurring in order to reduce the cost of corrective actions *ex post* are known as *preventive actions*. These are discussed in more detail in Part V. Here we can observe that the presence of ramp rate constraints limits the ability of the power system to adjust to changing demand *ex post*, thereby raising the cost of corrective actions. It may be efficient for the power system to take actions in advance (preventive actions) to reduce the costs of those corrective actions *ex post*. In this case, the preventive action consists of ramping up the low-ramp-rate plant in advance of the increase in demand.

To see this consider the earlier example again, but this time let us suppose that demand starts at 1200 MW. In period 4 this demand ramps up to (just under) 1500 MW. If no preventive actions were taken, following the increase in the demand, generator B would ramp up slowly to 500 MW. However, this requires substantial use of the \$100/MWh generator. This is not the

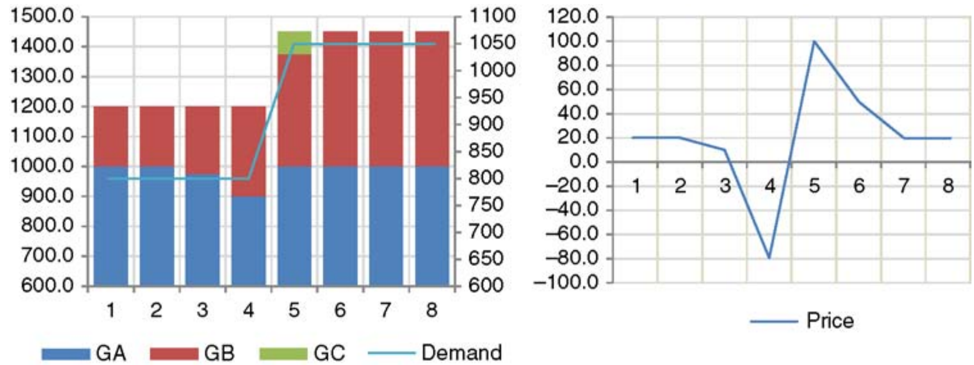

Figure 4.11 Efficient dispatch in the presence of ramp rate constraints may require taking preventive actions

least cost dispatch. Instead, it is efficient to take preventive actions – to ramp up generator B in advance of the increase in demand.

As shown in Figure 4.11, the price first drops in anticipation of the ramp-rate-binding episode and then increases during the episode. Intuitively, the reason is that, in this example, an increase in demand before the episode reduces the cost of adjustment. This occurs because the burden of the adjustment falls on the generator with the lowest ramp rate. In this example, if the overall demand were lower (as in the previous case) or higher, the burden of adjusting to the new steady state would fall on generators with a higher ramp rate, and the adjustment cost would be lower.

Result: When the supply, demand or network conditions change rapidly, the change of production or consumption targets may exceed the ramp-rate limits of some assets. In this case some assets may need to be dispatched out of merit order. This raises the cost of adjustment to the new steady state, known as corrective actions.

Where the change in supply/demand conditions is anticipated it will often be efficient to take actions in advance of the change in conditions, known as preventive actions, to reduce the cost of corrective actions. This may involve dispatch out-of-merit-order before the change in conditions occurs.

During the period of preventive actions the price may move in a counter-intuitive manner. Specifically the price will be lower than otherwise expected when an increase in the load in that period reduces the cost of taking preventive actions in subsequent periods.

4.10 Startup Costs and the Unit-Commitment Decision

Many generators must incur material costs before they are able to produce any output at all. For example, the generator might need to boil water to create steam. The cost of the energy required to get the boiler up to normal operating temperature must be incurred whether or not the generator operates for 1 h or 1 year. These costs are known as *startup costs*.

Up until now we have assumed that we have a stock of generating assets that are ready and willing to produce at a moment's notice. In the presence of startup costs, however, there is another decision to make – whether or not to bring a generating unit into a position where it is ready to produce output. This decision is known as the *unit commitment decision*.

Whether or not it is socially efficient to incur startup costs depends, amongst other things, on the length of time that the output of that generator will be required by the market. If a generator is only expected to be operating for an hour or less, it will typically not be worthwhile incurring substantial startup costs. On the other hand, if the generator is expected to be running continuously for months, the startup costs are largely irrelevant. Therefore, we cannot look at the effect of startup costs on the optimal dispatch in a single dispatch interval. Instead, we need to look at the effects of startup costs over a period of time, as in the earlier discussion on energy-limited generators.

In general, whether or not it will be worthwhile to incur the startup costs of a particular generating unit involves a comparison of (a) the magnitude of the startup costs of that generating unit and (b) the additional costs incurred by the market in the absence of that generating unit. Suppose a generator is expected to generate, say, 1000 MWh of electricity over the course of a day. Suppose that the startup cost of this generator is, say, \$50 000. If this generator is not started, then some other more expensive generator must be called on to supply this 1000 MWh of electricity. This will tend to increase the total cost of generating electricity. In addition there may be a need to incur startup costs from some other generators. Whether or not to incur startup costs depends on a comparison of these startup costs and the additional costs incurred when this generator is not started.

Where startup costs are material they can be incorporated into the optimal dispatch task – but, as with ramp-rates and energy limits, we need to consider an intertemporal dispatch task.

Up until this point we have modelled the dispatch task as a set of linear equations. These can be solved easily using standard algorithms. The introduction of startup costs requires that we introduce binary variables to the constrained optimisation. A binary variable can only take one of two values: zero or one.

As before, let us suppose that we have a power system with a number of generators. The output of generator i at time t is given by Q_{it}. At each time the state of the generator is given by the binary variable S_{it}. $S_{it} = 1$ means that the generator has started and is in an operational state. $S_{it} = 0$ implies the generator is in the state of being shut down. We can model startup costs as the costs of transitioning between states. For example, the transition from shut-down to started can be modelled as a binary variable: $T_{it}^{\text{off}\to\text{on}} = S_{it} - S_{it-1}$. The transition from started to shut-down can be modelled as a binary variable: $T_{it}^{\text{on}\to\text{off}} = S_{it-1} - S_{it}$. Let us suppose that the startup cost is $C_i^{\text{off}\to\text{on}}$ and the shut-down cost is $C_i^{\text{on}\to\text{off}}$.

The optimal dispatch task is now formulated as follows:

$$\min \sum_{i,t} c_i Q_{it} + \sum_{i,t} T_{it}^{\text{off}\to\text{on}} C_i^{\text{off}\to\text{on}} + \sum_{i,t} T_{it}^{\text{on}\to\text{off}} C_i^{\text{on}\to\text{off}}$$

$$\text{subject to } \forall t, \sum\nolimits_i Q_{it} = Q_t \leftrightarrow \lambda_t$$

and

$$\forall i, t, \ Q_{it} \leq K_{it} \leftrightarrow \mu_{it} \text{ and } Q_{it} \geq 0 \leftrightarrow \nu_{it}$$

Table 4.5 Key cost data for the illustration of the impact of startup costs on optimal dispatch

Generator	Variable Cost ($/MWh)	Capacity (MW)	Startup Cost ($)
A	10	1000	
B	20	500	20 000
C	50	500	

and

$$\forall i, t, T_{it}^{\text{off}\to\text{on}} - T_{it}^{\text{on}\to\text{off}} = S_{it} - S_{it-1} \text{ and } T_{it}^{\text{off}\to\text{on}} + T_{it}^{\text{on}\to\text{off}} \leq 1$$

and

$S_{it}, T_{it}^{\text{off}\to\text{on}}$ and $T_{it}^{\text{on}\to\text{off}}$ are binary variables.

To illustrate this let us consider a simple network with three generators as set out in Table 4.5.

Initially, generators A and C are operating; B is shutdown. Let us suppose that there are 8 periods. Demand is 800 MW in the first two periods, 1000 MW in the next two periods, 1300 MW in the next two periods and 750 MW in the last two periods. With this demand pattern, the efficient dispatch outcome involves generator B remaining off. Generator A produces the first 1000 MW and generator C produces the rest.

Now suppose that demand is 300 MW higher in each period (1100 MW in the first two periods, then 1300 MW, then 1600 MW and then 1050 MW). With this demand pattern it is efficient to incur the startup costs for generator B at the outset. The generators are then dispatched in their merit-order.

Result: Where there are material costs associated with the transition from shut-down to operational or vice versa, these can be taken into account in an intertemporal optimal dispatch task. The decision to incur startup costs depends on the contribution of the generator to reducing the cost of dispatch once it is operational.

4.11 Summary

A core task in achieving an overall efficient electricity industry is achieving efficient use of a given set of generation and consumption assets. This requires information about the cost function of generators and the utility function of consumers. In the modelling of electricity markets it is common to assume that generators have a marginal cost function that is constant up to some fixed capacity.

The optimal least-cost dispatch of a set of generators (i.e. the outcome that minimises the cost of generating sufficient electricity to meet demand) is where each generator produces up to the point where its marginal cost is equal to the common industry-wide marginal cost, which is also sometimes known as the system marginal cost or SMC. In the case of constant marginal cost, the efficient outcome involves the construction of a merit order of generators. The efficient outcome is often also referred to as optimal dispatch.

Optimal dispatch of load (consumption) resources occurs in a similar manner. The basic task is to choose the rate of production for each generator and the rate of consumption of each customer in such a way as to maximise total economic welfare or surplus. The optimal dispatch has the characteristic that each customer consumes at a rate where his/her marginal valuation is equal to a common system marginal cost, and each generator produces at a rate where his/her marginal cost is equal to the same common system marginal cost. We can treat customer responsiveness to price as equivalent to a hypothetical generator, converting the task of surplus maximisation into a task of cost minimisation.

Nonconvexities in production, such as minimum load levels, or startup costs significantly complicate the task of finding the optimal dispatch. There may be no feasible solution to the optimal dispatch problem.

Where a generator is limited in how much energy it can produce, the task of finding the least-cost dispatch or the welfare-maximising dispatch is an intertemporal problem involving choosing the rate of production or consumption over time. It will typically make economic sense to hold back the output of the energy-limited generator at times of low system marginal cost in order to increase the output of the generator at times of high system marginal cost.

The optimal dispatch task also requires an intertemporal optimisation where generators are limited in the rate of change of their output or where generators must incur startup costs to transition from shutdown to operational. Where generators are subject to ramp-rate constraints it may be necessary to dispatch generators out of merit order in order to balance supply and demand. It may make sense to adjust the power system *ex ante* (before the ramp rate constraints are binding) in order to reduce the cost of adjustment *ex post*. This illustrates the principle that it may make sense to take preventive actions to reduce the cost of taking corrective actions *ex post*. When ramp rate constraints are binding, the system marginal cost (which is also the price in a liberalised market) may move in a counter-intuitive manner.

Questions

4.1 What does it mean for the cost function of a generator to be nonconvex? What characteristics of a generator might give rise to a nonconvex cost function?

4.2 Under what conditions is there a monotonic-increasing relationship between the SMC (or market price) and the quantity of electricity supplied (i.e. under what conditions does higher demand lead to prices that are equal or higher)?

4.3 True or false: In a general least cost dispatch with upward-sloping generator marginal cost functions, every generator always produces to the point where its marginal cost is equal to the SMC?

4.4 Suppose that all customers have an inelastic demand for electricity up to a marginal value V (at which point demand for electricity drops to zero). Show that the problem of total surplus maximisation is equivalent to a problem of generator cost minimisation by including a hypothetical generator with a marginal cost equal to V.

4.5 An energy-limited generator has a very low marginal cost of \$1/MWh. Should this generator be classified as a ‘baseload’ generator in the merit order?

4.6 True or false: In the presence of ramp rate constraints, price spikes can occur at times of off-peak demand? Explain why or why not.

4.7 In the presence of startup costs, is it still correct to say that generators should be dispatched according to the merit order? Why or why not?

Further Reading

For more on the unit-commitment issue, see Padhy (2004).