

IFN Working Paper No. 1208, 2018

Supply Function Equilibrium over a Constrained Transmission Line I: Calculating Equilibria

Keith Ruddell

Supply function equilibrium over a constrained transmission line I: calculating equilibria

Keith Ruddell*
Research Institute of Industrial Economics (IFN),
Stockholm, Sweden

19th April 2018

Abstract

Competition between oligopolist electricity generators is inhibited by transmission constraints. I present a supply function equilibrium (SFE) model of an electricity market with a single lossless, but constrained, transmission line. The market admits equilibria in which generator withhold energy in order to induce congestion, which further increases their local market power.

Under appropriate assumptions on cost and demand functions, I obtain a planar autonomous system of ordinary differential equations for the SFE. Computational methods are developed to solve the system while respecting monotonicity constraints on the supply functions.

Using these methods I can calculate SFE in network markets that range from fully isolated to fully integrated. I also find network markets for which the SFE is not unique.

JEL: C62, C65, D43, L13, L94

Keywords: supply function equilibrium; electricity markets; market power; locational pricing of electricity.

1 Introduction

In this paper are developed methods to solve for supply function equilibrium (SFE) in electricity markets with nodal pricing and transmission constraints. A network SFE models the effects of transmission constraints on the competition among an oligopoly of electricity suppliers.

High-voltage electric transmission makes it efficient to transmit electricity over distances of hundreds of kilometres. This technology makes available generation sources that cannot be built near centers of demand, such as hydroelectric and geothermal power. Transmission lines are large investments, and it is important to understand the value of transmission when making investment decisions. Transmission's value comes from efficiency, reliability and enabling a

*I thank my supervisors Andy Philpott and Anthony Downard for their patience and support, also Pär Holmberg, Shmuel Oren, and Ross Baldick for helpful comments. Financial support from the University of Auckland, the Jan Wallander and Tom Hedelius Foundation, the Torsten Söderberg Foundation (E37/13), the Swedish Energy Agency (40653-1), and the Research Program the Economics of Electricity Markets is gratefully acknowledged.

Email: Keith.Ruddell@ifn.se.

competitive market. Efficiency gains come from the ability to use distant lower-cost generation to meet demand. Reliability gains come from the ability to access remote reserves as well as energy. Transmission lines improve competition by increasing the number of suppliers able to meet demand at a specific location, and by reducing the likelihood of local monopolies.

When deciding on a transmission investment, transmission operators and regulators must evaluate all of these benefits against the cost of the investment. An SFE model will help us understand the competition benefits in particular. We shall see that ownership transfers or other hedging contract arrangements may be sufficient to mitigate the anti-competitive effects of network congestion, without having to build new lines.

SFE in networks

The main contribution of this paper is to develop and apply numerical methods to the computation of SFE in a network setting. It builds on the theoretical work of Wilson (2008) and Holmberg and Philpott (2017a) on equilibrium conditions for SFE in a transmission network. By restricting the marginal cost and demand functions, the first order equilibrium conditions can be transformed into a linear autonomous ordinary differential equation in two dependent variables.

This work goes beyond previous work on supply function and Cournot equilibria in electricity networks in that it has both non-linear supply curves and line congestion as an endogenous variable.

Niu (2005) calculated linear SFE over a transmission network, assuming that generators would correctly predict the congestion state of the network before making their offers. In the SFE model presented here, the presence of transmission congestion is both endogenous (depends on generators' offers) and stochastic (for given offers, the congestion state is known *ex ante* only as a probability).

Borenstein, Bushnell, and Stoft (2000) showed that, even in a two-node one-line network, the Cournot game between generators can fail to have a pure-strategy equilibrium. In certain circumstances only a mixed-strategy equilibrium exists.

Downward, Zakeri, and Philpott (2010) characterize the pure-strategy Cournot equilibrium over a transmission network, and extend the model to account for transmission losses. We shall see that a pure-strategy supply function equilibrium exists in these circumstances, though it may not be unique.

Anderson, Philpott, and Xu (2007) use market distribution functions to analyze the response of a supplier to the opening of a new interconnection with a previously isolated market. They solve the best-response problem in terms of this market distribution function but do not derive a full SFE for the interconnected market. Their model distinguishes itself by taking account of transmission losses.

Computing SFE

The first-order conditions that describe an SFE form a system of differential equations. Since only the simplest models admit analytic solutions, several numerical methods have been employed for SFE.

Anderson and Hu (2008, 2012) approximate solutions by discretizing the demand shock to create a large system of linear constraints. This allows them to solve for the coefficients defining the SFE as a non-linear program.

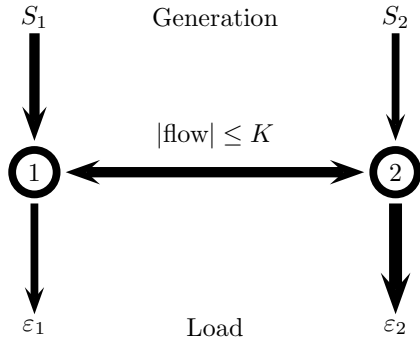


Figure 1: A network with two nodes and one line.

Holmberg (2009) solves the system of ODE resulting from the equilibrium conditions of an SFE in which generators have asymmetric costs. He treats it as a boundary value problem, and uses Runge-Kutta integration to integrate the system which is unstable in both directions. In this paper we embed such Runge-Kutta methods in a scheme which ensures stable integration at all stages of the computation.

Outline

The paper is laid out as follows. Section 2 presents the network, pricing rules, stochastic demand, and the necessary and sufficient conditions for optimal supply functions. In Section 3 we restrict the cost and demand functions in order to arrive at equilibrium conditions that give a linear autonomous system of ordinary differential equations.

We take the approach of first solving these systems in the case where the coefficient matrix is constant — in Section 4 — then treating more general instances — in Section 5. We give sufficient conditions for uniqueness of equilibrium. Section 6 extends our computational methods to systems with singularities, in which uniqueness may fail.

In a follow-on paper, Ruddell (2018), the two-node market is extended to consider generators that own multiple plants, as well as the strategic effects of financial derivatives such as contracts for differences and financial transmission rights.

2 A transmission network

The market uses *locational marginal pricing*. Pricing is uniform, in the sense that every producer receives the local price for all of their production. The usual assumption is made of a lossless transmission network, so that the shadow price on the transmission line obeys a complementarity condition: the prices differ at either end of the line only if the flow constraint is binding. The load, net of any wind and solar generation, is known to the generators only as a random variable whose distribution is common knowledge.

This section follows Holmberg and Philpott (2017a) in modeling the demand uncertainty as a shock at each node. These authors derive optimality conditions for SFE in a general radial network, and the calculation of the market distribution function from a demand shock distribution is due to them. Wilson (2008) takes a slightly different approach. He derives equivalent

equilibrium conditions but the shocks in his model are for the capacities of the lines, plus an additional random shock for the total demand. The present work extends the results of these papers by numerically calculating SFE in which there is some breaking of symmetry among the agents.

An electricity network consists of a set of buses or nodes $\mathcal{N}$ connected by circuits or lines. We will work with loop-free or radial networks, whose underlying graphs are trees. An important property of a tree is that there is a unique path connecting any pair of nodes. Real transmission grids carry alternating current (except on grid interconnectors and undersea cables that use high voltage direct current (HVDC) technology), and experience losses. In using a radial network, we may ignore Kirchhoff's laws for the DC load-flow model, and treat energy transmission as a transshipment problem. Conservation of energy requires that the flow into a node, plus energy injected, is equal to the flow out of the node, plus energy withdrawn. The nodal conservation of energy laws sum up to the global constraint that total supply and demand in the network are equal.

The market clearing mechanism can be formulated as an optimization problem, the *optimal dispatch problem*, which is solved by the system operator (auctioneer). The system operator minimizes total cost over all supply functions offered, to meet demand, subject to a great many constraints. In a real electricity system many of these constraints are either non-convex, binary or generally non-linear, and solving the market clearing problem is a non-trivial computation. Only linear constraints on the line flows are admitted, so the feasible set is a simple polyhedron. Integrating over this polyhedron gives the conditional probabilities of congestion that appear in the first-order conditions.

2.1 Constrained demand distribution

To aid in the calculation, we define a conditional probability distribution that captures the effect of the transmission constraint on distribution of a generator's residual demand.

First, some notation: at every node m the system operator will choose a nodal price p_m . Each power plant (identified by its owner i and node m) submits an offer curve $q_{i,m}(p_m)$ for the amount offered as a function of the local price. The market demand at m is the sum of a price-responsive component $D_m(p_m)$ and a stochastic shock ε_m . For the total offers net of price-responsive demand at node m , write

$$S_m(p_m) := \sum_{j \in m} q_{j,m}(p_m) - D_m(p_m),$$

where $q_{j,m}(p_m)$ is the quantity offered by generator j at node m at the local price and D_m is the nodal demand function representing price responsive but non-strategic demand at node m . For the total net offers of all generators, other than i , at node m write

$$S_{-i,m}(p) = \sum_{j \in m, j \neq i} q_{j,m}(p_m) - D_m(p_m),$$

where $j \in m$ means that agent j is located at node m . The vector S_{-m} consists of the net offers at all nodes other than m .

As shown by Wilson (2008) and Holmberg and Philpott (2017a), in a lossless and loop-free network the market distribution function $\psi_{i,n}(q_{i,n}, p_n; \mathcal{S}_{-(i,n)})$ of a generator i located at node n will depend, besides the generator's own offer $q_{i,n}$ and local price p_n , only on the net offers at the other nodes at price p_n . Let S be a vector of net injections at each node of the network, and define the *constrained demand distribution* $\phi_n(S)$ to be the probability that the quantity

dispatched at node n is less than S_n , given that other nodes have net offers S_{-n} at the same local price.¹ Thus

$$\psi_{i,n}(q_{i,n}, p) = \phi_n(q_{i,n} + S_{-i,n}(p), S_{-n}(p)). \quad (1)$$

So ϕ_n is the joint distribution on the net injection at each node induced by the demand shock distribution and the transmission constraint. We can calculate ϕ_n by integrating over the space of demand shocks. See the Appendix for an explanation of the calculation of ϕ .

Figure 1 shows the two-node network over which we will calculate SFE. The transmission line has a capacity of K in either direction, and we assume there are no transmission losses. The network has three possible *congestion states*. In state ω_1 , the line constraint is not binding but the total demand must equal total supply, $\varepsilon_1 + \varepsilon_2 = S_1 + S_2$. In state ω_2 , the line constraint is binding from node 1 to node 2, so there is a fixed outflow from node 1, $\varepsilon_1 + K = S_1$. In the last state ω_3 , the line constraint is binding the other way, so there is a fixed inflow to node 1, $\varepsilon_1 = S_1 + K$.

The constrained demand distribution can be written

$$\begin{aligned} \phi_1(S_1, S_2) &= \Pr[(\varepsilon_1 + \varepsilon_2 < S_1 + S_2 \text{ and } \varepsilon_1 < S_1 + K) \text{ or } \varepsilon_1 < S_1 - K] \\ \phi_2(S_1, S_2) &= \Pr[(\varepsilon_1 + \varepsilon_2 < S_1 + S_2 \text{ and } \varepsilon_2 < S_2 + K) \text{ or } \varepsilon_2 < S_2 - K]. \end{aligned}$$

The stochastic part of demand has a bivariate normal distribution with mean

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

and variance-covariance matrix

$$\begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

Assume that $\sigma_1 > 0$, $\sigma_2 > 0$ and $\rho \in (-1, 1)$, so that the demand shock density is always positive and smooth, and the constrained demand distribution is continuous. This rules out the existence of multiple equilibria arising from a demand shock with bounded support, as described by Klemperer and Meyer (1989). It also avoids the non-existence of equilibria due to jumps in the market distribution function, as described by Holmberg and Philpott (2017a).

Now consider the market distribution function for an agent at node m . Since $S_m = \sum_{i \in m} q_{i,m}$, we get

$$\frac{\partial \psi_n}{\partial q_{i,m}} = \frac{\partial \phi_n}{\partial S_m} \text{ for all } i \in m.$$

We will henceforth write these partial derivatives with the subscript notation

$$\phi_{n,m} := \frac{\partial \phi_n}{\partial S_m}.$$

From (1) we get generator i 's market distribution function $\psi_{i,m}$ in terms of the constrained demand distribution. Taking partial derivatives and applying the chain rule,

¹In a network with loops, it can arise that there are congested and uncongested paths between a pair of nodes on a loop. The supply curve of a generator at one of these nodes contributes to the residual demand function of a generator in the other, even though their local prices differ. Thus a generator's market distribution function will depend on their competitors' offers *over a range of prices*. In such a network the system of first-order optimality conditions is a system of *integro-differential equations*, which is a rather different problem mathematically to the ordinary differential equations we consider in this work.

$$\begin{aligned}\frac{\partial \psi_{i,m}}{\partial q} &= \phi_{m,m} \text{ and} \\ \frac{\partial \psi_{i,m}}{\partial p} &= \phi_{m,m} \cdot S'_{-i,m}(p) + \sum_{n \neq m} \phi_{m,n} \cdot S'_n.\end{aligned}$$

Of course, in the two-node network there will only be one $n \neq m$.

2.2 Optimal offers

Formally we suppose that each supplier submits to the system operator a piecewise differentiable curve

$$\mathcal{S} = \{(q(t), p(t)) : t \in [0, T]\},$$

whose quantity and price components q and p are non-decreasing with respect to the parameter t

$$\dot{q} = \frac{dq}{dt} \geq 0 \text{ and } \dot{p} = \frac{dp}{dt} \geq 0. \quad (2)$$

These are the *monotonicity constraints*. We will usually have $\dot{p} > 0$, so that the supply function can be reparametrized as a function of price $q(p)$. Each supplier chooses its supply curve $\mathcal{S}$ to maximize an expected profit functional of the form

$$\Pi[\mathcal{S}] = \int_{\mathcal{S}} R(q(t), p(t)) d\psi(q(t), p(t)) = \int_0^T R(q(t), p(t)) (\psi_q \dot{q} + \psi_p \dot{p}) dt, \quad (3)$$

i.e. a line integral along the curve $\mathcal{S}$. We suppose that R , ψ_q and ψ_p are (weakly) differentiable.

Following Anderson and Philpott (2002) we define the field

$$Z(q, p) = \frac{\partial R}{\partial q} \psi_p - \frac{\partial R}{\partial p} \psi_q, \quad (4)$$

which we will call the *first-variation derivative*. It is used to define the following optimality conditions.

Theorem 1 (Necessary conditions). *If $\mathcal{S}^* = \{(q(t), p(t)) : t \in [0, T]\}$ is a local maximum of $\Pi(\mathcal{S})$ with $\dot{q}(t) > 0$ and $\dot{p}(t) > 0$ for all $t \in [0, T]$, then $Z(q(t), p(t)) = 0$ for all $t \in [0, T]$. Moreover, if $Z(q, p)$ is differentiable at $(q, p) = (q(t), p(t))$, then we have $\frac{\partial Z}{\partial p}(q(t), p(t)) \geq 0$, and $\frac{\partial Z}{\partial q}(q(t), p(t)) \leq 0$.*

Theorem 2 (Sufficient conditions). *Let $\mathcal{S}^* = \{(q(t), p(t)) : t \in [0, T]\}$ be a continuous piecewise differentiable curve. If both components of $\mathcal{S}^*$ are non-decreasing in t , then a sufficient condition for $\mathcal{S}^*$ to be a global maximum of $\Pi(\mathcal{S})$ is that $Z = 0$ along $\mathcal{S}^*$, and that at every t , $Z(\tilde{q}, p(t)) \geq 0$ for all $\tilde{q} < q(t)$ and $Z(\tilde{q}, p(t)) \leq 0$ for all $\tilde{q} > q(t)$.*

Under uniform pricing, the profit made by agent i at node m when dispatched quantity $q_{i,m}$ at price p_m is

$$R_{i,m}(q_{i,m}, p_m) = p_m q_{i,m} - C_{i,m}(q_{i,m}).$$

The expected payoff to a generator offering a curve $\mathcal{S} = \{(q(t), p(t)) : t \in [0, T]\}$ is (dropping the agent and node subscripts)

$$\begin{aligned}\Pi^U(\mathcal{S}) &= \int_{\mathcal{S}} R(q, p) d\psi(q, p) \\ &= \int_0^T (q(t)p(t) - C(q(t))) (\psi_q \dot{q} + \psi_p \dot{p}) dt.\end{aligned} \quad (5)$$

The partial derivatives ψ_q and ψ_p give the marginal change in dispatch probability from changing the offer quantity and price. The first-variation derivative is obtained from (4):

$$\begin{aligned} Z^U(q, p) &= \frac{\partial}{\partial q} \left((qp - C(q)) \right) \psi_p - \frac{\partial}{\partial p} \left((qp - C(q)) \right) \psi_q \\ &= (p - C'(q)) \psi_p - q \psi_q. \end{aligned} \quad (6)$$

In terms of the constrained distribution function, this is

$$Z_{i,m} = (p - C'_{i,m}(q_{i,m})) \left(\phi_{m,m} S'_{-i,m} + \sum_{n \neq m} \phi_{m,n} S'_n \right) - \phi_{m,m} q_{i,m} \quad (7)$$

for generator i at node m .

We can obtain some qualitative insights into SFE by examining the optimal offer conditions. In particular, we can deduce the strategic response of a generator to changing congestion patterns.

Suppose that a generator is offering optimally and that some change in market conditions causes the constrained distribution function ϕ to change. To a first approximation, our agent supposes that its competitor's offers remain unchanged. How then does the optimal offer curve move in response to a change in ϕ ? Divide the optimal offer condition through by the local density of demand

$$\frac{Z_{i,m}}{\phi_{m,m}} = (p - C'_{i,m}(q_{i,m})) \left(S'_{-i,m}(p) + \sum_{n \neq m} \frac{\phi_{m,n}}{\phi_{m,m}} S'_n(p) \right) - q_{i,m} = 0.$$

The change in the optimal supply function will come from the change in $\frac{\phi_{m,n}}{\phi_{m,m}}$ near the old supply curve. This ratio gives the conditional probability of uncongested flow between m and n , given offers $S(p)$.² In general, if congestion worsens, $\frac{\phi_{m,n}}{\phi_{m,m}}$ will decrease, so agent m 's residual demand will get less elastic, so $(p - C'_{i,m}(q_{i,m})) \left(S'_{-i,m}(p) + \sum_{n \neq m} \frac{\phi_{m,n}}{\phi_{m,m}} S'_n(p) \right)$ gets smaller, so q must get smaller as well. That is, to a first order of response, agents respond to increased congestion by offering less competitively — reducing their offer quantities or increasing their mark-ups.

Increased congestion causes an increase in generator mark-ups. This is undesirable for a market regulator. To counter the incentive to mark up, the regulator could seek to reduce congestion through direct means (expanding transmission lines), or they can alter the incentive mechanisms in the market. In Ruddell (2018), we shall see how an asset swap (real or virtual) reduces the incentive for generators to mark up in the presence of transmission constraints.

2.3 Binding constraints

To describe the optimality conditions for supply curves with binding constraints, we introduce the adjoint function from optimal control theory; it measures the deviation from $Z = 0$. The *constrained offer optimization problem* is to maximize Π in (3), subject to the monotonicity constraints (2) and the state constraints on output and price.

$$\begin{aligned} q &\in [\underline{q}, \bar{q}], \text{ and} \\ p &\in [\underline{p}, \bar{p}]. \end{aligned} \quad (8)$$

²This interpretation gives a justification for the simplification of the entire New Zealand transmission grid; if within-island transmission constraints have zero probability of binding, and there are no transmission losses with islands, then we can model all injections and withdrawals for each island as if they were at one node.

These constraints arise from the physical production limits of the electricity generators, and the market-imposed bounds on offer prices.

Definition 3 (Adjoint function). Let $\mathcal{S} = \{(q(t), p(t)) : t \in [0, T]\}$ be a supply curve and Π a profit functional of the form (3) with first-variation derivative Z . An *adjoint function* to $\mathcal{S}$ for the constrained offer optimization problem is a function $w : [0, T] \rightarrow \mathbb{R}$ with

$$\dot{w} = \frac{dw}{dt} = Z(q(t), p(t)). \quad (9)$$

Given a candidate curve $\mathcal{S}$, we can calculate an adjoint function w from

$$w(t) = w_0 + \int_0^t Z(q(\tau), p(\tau)) d\tau. \quad (10)$$

So the adjoint function is defined up to a constant of integration w_0 . The following theorem gives a necessary conditions for an optimal constrained supply function in terms of the adjoint.

Theorem 4 (Slope-constrained optimal supply curve). *Let $\mathcal{S}^*$ be a supply curve with quantity and price components $q(t)$, $p(t)$. If the curve $\mathcal{S}^*$ is a maximum for the constrained offer optimization problem — with objective (3) and constraints (2) and (8) — then there exists an adjoint function w such that*

$$\dot{q}(t) \cdot w(t) \leq 0 \leq \dot{p}(t) \cdot w(t) \quad (11)$$

for all $t \in [0, T]$.

Proof. See Anderson and Philpott (2002). □

Usually, price is taken to be the parameter of the supply curves (i.e. $t = p$). This implies that $\dot{p} > 0$ everywhere, so we can actually work with simpler complementarity conditions.

Corollary 5. *Let $\mathcal{S}^* = \{(q(t), p(t)) : t \in [0, T]\}$ be a supply curve that satisfies the constraints (2) and (8), with $\dot{p} > 0$. If the curve $\mathcal{S}^*$ is a maximum for the constrained offer optimization problem then there exists an adjoint function w such that*

$$\begin{aligned} \dot{q}(t) \cdot w(t) &= 0 \text{ and} \\ w(t) &\geq 0 \end{aligned} \quad (12)$$

for all $t \in [0, T]$.

Proof. Theorem 4 gives us the existence of a w satisfying (11). From $\dot{p} > 0$ and $\dot{p} \cdot w \geq 0$ in (11), it follows that $w \geq 0$. From $\dot{q} \geq 0$ and $w \geq 0$ we have that $\dot{q} \cdot w \geq 0$. But (11) gives also $\dot{q} \cdot w \leq 0$, so it must be that $\dot{q} \cdot w = 0$. □

From (11) it is clear that if there is a t where both $\dot{q}(t), \dot{p}(t) > 0$, then $w(t)$ must be zero. This determines w_0 in (10). Moreover $w = 0$ at every point where $\dot{q}, \dot{p} > 0$, and so $\dot{w} = Z(q, p) = 0$ over every interval where $\dot{q}, \dot{p} > 0$. Thus, the necessary conditions of Theorem 1 apply locally to the constrained problem, on segments where $\dot{q}, \dot{p} > 0$. The sufficient conditions of Theorem 2 apply locally too. This is easy to see from the proof of that theorem, since the improving deviations, where they exist, are local.

Ironing

Intervals for which $\dot{q} = 0$ we shall call *ironed segments*, and agents who offer supply curves with such segments are said to be *ironing*. When an agent is ironing, its adjoint is no longer bound to zero, so Z can be non-zero over ironed segments. A direct consequence of Theorem 4 is Wilson's (1993) result that over an interval where the $\dot{q} \geq 0$ or $\dot{p} \geq 0$ constraint binds (but the state constraints (8) do not), the first-order conditions $Z = 0$ are satisfied *on average*.³

Horizontal (perfectly elastic) segments of a supply function are intervals where $\dot{p} = 0$. Holmberg (2007) showed that, in equilibrium, horizontal segments may only occur where all other suppliers are at their output capacity bounds. However, they can occur in supply curves which are optimal responses to certain market distribution functions.

3 Tractable equilibrium conditions

In order to focus on the strategic effect of transmission constraints, we would like the other parts of the model to be as simple as possible. It turns out that if the demand and cost functions are simple enough, then an SFE can be described by an autonomous linear differential equation. For our two node network, this gives a planar system of ordinary differential equations, to which we can apply phase plane analysis. The following assumptions will be sufficient.

Assumption 6 (Two-node linearizable market).

1. Demand is perfectly inelastic in price.
2. The demand shock has a bivariate normal distribution.
3. Generators have identical and constant marginal costs of production c .
4. There are $N_1 \geq 2$ of these generators located at node 1 and $N_2 \geq 2$ at node 2.
5. All the generators at node 1 have output capacity $\bar{Q}_1$ and all the generators at node 2 have output capacity $\bar{Q}_2$.

Proposition 7, together with Proposition 11 below, will show that identical generators at the same node will offer identical supply curves in SFE. Because of this symmetry, we get a reduction in order of the Klemperer-Meyer equation $Z = 0$, as we only need to know one curve from each node to know the whole SFE. Each generator at node 1 offers a curve $Q_1(p)$ and each at node 2 offers $Q_2(p)$. The total nodal offers are thus

$$S_1 = N_1 Q_1 \text{ and } S_2 = N_2 Q_2.$$

³Suppose we have an optimal supply curve with $\dot{q} = 0$ over an interval $[t_1, t_2]$, and that this interval is the maximal such interval. Since it is the maximal such interval, we have $\dot{q} > 0$ on $(t_1 - \delta, t_1)$ and $(t_2, t_2 + \delta)$ for some $\delta > 0$. Hence

$$\lim_{t \nearrow t_1} w(t) = 0 \text{ and } \lim_{t \searrow t_2} w(t) = 0,$$

and so since w is continuous, $w(t_1) = 0$ and $w(t_2) = 0$. Using the fundamental theorem of calculus, we see that the average of Z over the interval is zero:

$$\int_{t_1}^{t_2} Z(q(t), p(t)) dt = w(t_2) - w(t_1) = 0.$$

The system of first-order equilibrium conditions under Assumption 6 is derived from (7):

$$\begin{aligned} Z_1 &= (p - c) (\phi_{1,1} (N_1 - 1) Q'_1 + \phi_{1,2} N_2 Q'_2) - \phi_{1,1} Q_1 = 0 \\ Z_2 &= (p - c) (\phi_{2,1} N_1 Q'_1 + \phi_{2,2} (N_2 - 1) Q'_2) - \phi_{2,2} Q_2 = 0. \end{aligned} \quad (13)$$

Note that the partial derivatives $\phi_{m,n}$ of the constrained demand distribution function depend on the net nodal offers $S_1 = N_1 Q_1$ and $S_2 = N_2 Q_2$. Define the matrix

$$\mathbf{A} = \begin{bmatrix} (N_1 - 1) & N_2 \frac{\phi_{1,2}}{\phi_{1,1}} \\ N_1 \frac{\phi_{2,1}}{\phi_{2,2}} & (N_2 - 1) \end{bmatrix}, \quad (14)$$

so that the system (13) can be written

$$(p - c) \mathbf{A} Q' = Q.$$

Further, make a change of variable

$$ae^t = p - c. \quad (15)$$

It follows that $\dot{Q} = \frac{dQ}{dt} = \frac{dQ}{dp} \frac{dp}{dt} = Q' ae^t = Q' (p - c)$, so

$$\mathbf{A} \dot{Q} = Q. \quad (16)$$

Though the matrix $\mathbf{A}$ depends on the offer quantities Q_1 and Q_2 , it does not depend explicitly on either price or the parameter t . Hence the ODE (17) is autonomous, i.e the independent variable does not appear in the equation. We can analyze it as a direction field in *offer-quantity space*, the space of all possible offer quantities $Q = (Q_1, Q_2)$. Autonomy gives a degree of freedom in choosing a scalar multiple on the price variable when we change from t back to p . Another way to express this is that we may rescale the mark-up $(p - c)$ by any positive scalar multiple, i.e. if $r = c + \frac{1}{a} (p - c)$ for some scalar constant a , then

$$a(r - c) \frac{dQ}{dr} = (p - c) \frac{dQ}{dp} = \mathbf{A}^{-1} Q.$$

We define the function $U(Q) = \mathbf{A}^{-1} Q$ so that the slope of the (unconstrained) trajectory $Q(t)$ through offer-quantity space is

$$\dot{Q} = U(Q). \quad (17)$$

In general, the matrix $\mathbf{A}$ is a function of Q . In order to understand the SFE that occur in the presence of constrained transmission, we will first look at a constant-coefficient version of (16), one in which $\mathbf{A}$ does *not* vary with Q .

The elements of $\mathbf{A}$ have the following interpretation: $\mathbf{A}_{m,n}$ is the number of competitors that an agent at node m faces who are at node n , times the probability that there is no transmission congestion between m and n . Hence $\mathbf{A}_{m,n}$ is the expected number of competitors at node n that agents at node m face. This means that the components of $\mathbf{A} Q'$ are the expected slopes of residual demand for agents at the two nodes.

When integrating to find solutions to (16), we want to integrate in a direction for which the solution remains stable. The stability of the solution is determined by the eigenvalues of the matrix $\mathbf{A}$. The Perron-Frobenius theorem (Meyer 2000), tells us that since all the entries of $\mathbf{A}$ are positive, $\mathbf{A}$ has at least one positive eigenvalue and a corresponding eigenvector that lies in the positive orthant. Our $\mathbf{A}$ is 2×2 , so it can have either two positive eigenvalues or one positive and one negative eigenvalue. If there are two positive eigenvalues, then the two nodes

behave more like independent markets. If the eigenvalues are mixed, then the market is more integrated.

The origin is a singular point of the system (16). When solving near the origin, the eigenvalues of $\mathbf{A}$ become particularly important, and force us to integrate in a particular direction to keep the error bounded. If, on the one hand, the eigenvalues of $\mathbf{A}$ at the origin are mixed, then the origin is a saddle point. Integrating towards the origin will always diverge, so it is necessary to integrate moving away from the origin. If, on the other hand, the eigenvalues are both positive, then the origin is a source node; in this case integrating towards the origin is the more stable direction.

As Klemperer and Meyer (1989) showed for SFE without transmission, we now show that when 0 is in the support of the demand shocks, then all SFE with identical generators in two nodes will have all agents in the same node making identical offers near zero output. It is sufficient to show that any pair of generators in the same node have identical offers.

Proposition 7. *Under Assumption 6, if the supply functions of both generators are strictly monotone on some neighbourhood of c , then they are equal.*

3.1 SFE with all offers identical

We proceed to solve for SFE in some special cases. In our network SFE model, if we assume that in equilibrium every agent's optimal offer problem is identical, then the equilibrium can be found by solving a first-order (scalar) ODE. This will be true if there are the same number of agents in each node and symmetrical distributions of demand in each node.

Suppose there are N agents in each node. All $2N$ agents will offer the same supply function $Q(p)$ in equilibrium. Because the offer quantities at a given price are going to be identical between the two nodes. We require $\mathbf{A}$ to be symmetric when $S_1 = S_2$, i.e.

$$\frac{\phi_{1,2}(NQ, NQ)}{\phi_{1,1}(NQ, NQ)} = \frac{\phi_{2,1}(NQ, NQ)}{\phi_{2,2}(NQ, NQ)}. \quad (18)$$

We already have that

$$\phi_{1,2}(NQ, NQ) = \phi_{2,1}(NQ, NQ)$$

for all $Q \in [0, \bar{Q}]$, so for (18) to hold for all Q , we must have

$$\phi_{1,1}(NQ, NQ) = \phi_{2,2}(NQ, NQ). \quad (19)$$

This can be achieved by making the shock distribution a bivariate normal with the same mean and standard deviation in both components.

When $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$, then the Gaussian density function $f(\varepsilon_1, \varepsilon_2)$ is symmetric in its arguments. And the symmetry of $\phi_{m,n}$ follows from this.

The equilibrium condition (13) then becomes

$$(p - c) \left((N - 1) + N \frac{\phi_{1,2}}{\phi_{1,1}} \right) Q' - Q = 0. \quad (20)$$

We interpret the term $\left((N - 1) + N \frac{\phi_{1,2}}{\phi_{1,1}} \right)$ as the expected number of competitors reachable along an uncongested line.⁴ This probability is endogenous and will change with Q , varying

⁴This is $N\tilde{\mu} - 1$ in Holmberg and Philpott's (2017b) 'well-behaved two-node network.'

between $N - 1$ (when the line is almost-certainly congested) and $2N - 1$ (when it is almost certainly uncongested). The general solution to (20) is

$$\log(p - c) = (N - 1) \log(Q) + N \int \frac{\phi_{1,2}(NQ, NQ)}{\phi_{1,1}(NQ, NQ)} \frac{1}{Q} dQ.$$

The last integral might not have an analytic solution, but we can estimate bounds for it as follows. The ratio $\frac{\phi_{1,2}(NQ, NQ)}{\phi_{1,1}(NQ, NQ)}$ is the probability of uncongested flow, so is bounded by 0 and 1.

When the ratio is 0, we have local oligopolies, each with N agents, so the symmetric SFE is

$$Q(p) = k(p - c)^{\frac{1}{N-1}},$$

where k is a constant of integration. The curve is a straight line for $N = 2$ and a hockey-stick shape for $N > 2$. For a given endpoint, the supply function equilibrium is more hooked for larger N . The more hooked curve is more competitive because the price stays closer to marginal cost for longer.

When the ratio is 1, we have all $2N$ agents competing in supply functions, so the symmetric SFE is

$$Q(p) = k(p - c)^{\frac{1}{2N-1}}.$$

Example 8. The demand shock is normal in each node with a mean of 20MW and standard deviation of 50MW. There is no correlation between nodal shocks. There are two generators at each node. Each generator has a production capacity of 120MW. Costs of production are zero and there is a price cap at \$1000. The line has a capacity of 40MW.

By Holmberg (2008), there is a unique SFE in this market where the production capacities are exhausted at the price cap.

Figure 2 shows, on the left, a plot of how the expected number of competitors varies with Q relative to the (nodal) mean of the demand shock distribution. It varies between 1 and a maximum of 2.26. The line is most likely to be uncongested when the offer at each node is equal to the mean of the demand shock there. In some demand shock outcomes (28% of the time), the total demand shock is negative, which will lead to negative prices. You could think of this as the ‘up’ side of a balancing market.

On the right of Figure 2 is the SFE for this market. The supply curve is approximately parabolic near zero output, which is where the probability of uncongested flow is highest, and the expected number of competitors is roughly two. As output grows, so does the probability of congestion. Near the production capacities the generators are expected to face only one competitor, so the SFE curve becomes approximately linear.

3.2 Meeting capacity constraints

Under Assumption 6 each generator in node m has production capacity $\overline{Q}_m$, and write $\overline{Q}$ for the vector

$$\begin{bmatrix} \overline{Q}_1 \\ \overline{Q}_2 \end{bmatrix}.$$

Suppose that the support of demand shock contains all of $[0, \overline{Q}_1] \times [0, \overline{Q}_2]$. Hence Proposition 7 applies near zero output. Suppose that offers prices are required to be below the *price cap* $\overline{p}$ and above the *price floor* $\underline{p}$.

If demand is so high that the market cannot clear even when one of the nodal prices reaches the price cap, we suppose that load is shed. Similarly, if demand is so low that the market would clear below the price floor, then supply is rationed.

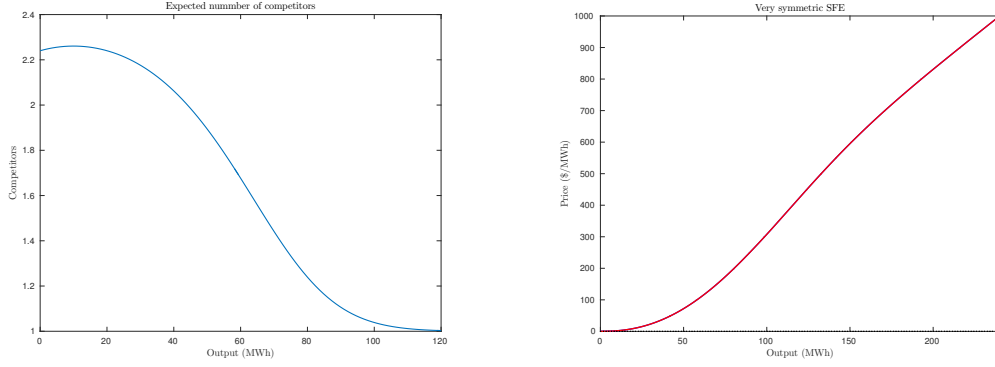


Figure 2: Symmetric SFE. On the left is the expected number of competitors for each firm, which changes with the output. On the right is the symmetric SFE $\{(Q, p) ; p \in [0, \bar{p}]\}$.

The production capacity constraints and price cap give us boundary conditions for the outer (high-price) end of the SFE. When demand is totally inelastic, all generators' production is required in order to meet the highest levels of demand, and a price cap is imposed, Holmberg (2008) showed there is a unique SFE. In this SFE, all generators except perhaps one hit their production capacity exactly at the price cap. His argument for a pool market applies exactly the same to a network market with more than one player at each node.

In our model, since the generators in each node are perfectly symmetric, there are only two possible cases.

1. The generators at one node offer their last unit at the price cap, and those in the other node offer their last unit at a lower price, or
2. all generators offer their last unit at the price cap.

In the first case, generators that offer their last unit at the price cap play SFE only against each other at prices above the point where the other generators offer their last unit; this is as in Holmberg (2008).

3.3 Monotonicity constraints

In certain parts of the offer-quantity space, the slope given by (13) has a negative component. Further, it is not always possible to trace a trajectory from the origin to the production capacity point that maintains positive slope everywhere, even if the capacity constraints bind. We will need to *iron* the solution so that it obeys the monotonicity constraint. As Anderson and Philpott (2002) showed, there are restrictions on the sign that the first-variation derivative of expected profit Z can have at the ends of the interval. Formulas for Z_1 and Z_2 are given in (13). In matrix form,

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = (p - c) \begin{bmatrix} \phi_{1,1}(N_1 - 1) & \phi_{1,2}N_2 \\ \phi_{2,1}N_1 & \phi_{2,2}(N_2 - 1) \end{bmatrix} Q' - \begin{bmatrix} \phi_{1,1} & 0 \\ 0 & \phi_{2,2} \end{bmatrix} Q.$$

Under the change of variables $p - c = e^t$, we get a price-independent expression for Z along a supply curve:

$$\begin{aligned} Z(Q, \dot{Q}) &:= \begin{bmatrix} (N_1 - 1)\phi_{1,1} & N_2\phi_{1,2} \\ N_1\phi_{2,1} & (N_2 - 1)\phi_{2,2} \end{bmatrix} \begin{bmatrix} \dot{Q}_1 \\ \dot{Q}_2 \end{bmatrix} - \begin{bmatrix} \phi_{1,1}Q_1 \\ \phi_{2,2}Q_2 \end{bmatrix} \\ &= \begin{bmatrix} \phi_{1,1} & 0 \\ 0 & \phi_{2,2} \end{bmatrix} (A\dot{Q} - Q). \end{aligned} \quad (21)$$

We will now describe the simultaneous optimization of constrained supply curves. The adjoint for a curve through offer-space is used to define necessary conditions for ironing in SFE.

Trajectories through offer-quantity space

Segments of the supply curve where $\dot{q} = 0$ are said to be *ironed*. On such segments, the optimality conditions are stated in terms of the adjoint function w . When the agents in one node iron, they hold their offer quantities constant over some interval of prices, while the agents in the other node will continue to offer at an increasing slope, following $Z = 0$. In order to find SFE by integration in offer-space, we first define the candidate trajectory as the vector of supply curves, together with an adjoint function for each. As we show below, in Proposition 11, all the agents in a given node offer supply curves that are identical; on a given segment of curve, either none of them iron or they all do.

Definition 9 (Candidate trajectory). A *candidate trajectory* is a (piecewise differentiable) curve $Q : [0, T] \rightarrow \mathbb{R}^2$ together with an adjoint function $w : [0, T] \rightarrow \mathbb{R}^2$ such that

$$\begin{aligned} \dot{w}(t) &= Z(Q(t), \dot{Q}(t)), \\ \dot{Q}_n(t) w_n(t) &= 0 \text{ for } n = 1, 2, \\ \dot{Q}(t) &\geq 0, \text{ and} \\ w(t) &\geq 0 \text{ for all } t \in [0, T]. \end{aligned} \quad (22)$$

We can find a candidate trajectory in t and then transform back to prices. Under the change of variables

$$p - c = ae^t,$$

the parameter t varies from $-\infty$ to T as p goes from c to $\bar{p}$. The zero mark-up at zero output condition becomes a limit; $\lim_{t \rightarrow -\infty} Q(t) = 0$. The outer boundary condition is simply $Q(\bar{p}) = \bar{Q}$. By the correct choice of constant a , we can meet this exactly at the price cap. If $Q(T) = \bar{Q}$, then just pick

$$a = (\bar{p} - c) e^{-T}.$$

Lemma 10. Suppose we have a candidate trajectory $Q(t)$, together with its adjoint $w(t)$. Define $p(t) = c + ae^t$ with $a > 0$. Then the set of supply curves $\mathcal{S}_{i,n} = (Q_n(t), p(t))$ for all $i \in n$, $n = 1, 2$ satisfies the necessary conditions of Theorem 4.

Proof. The necessary conditions are

$$\dot{Q}_n(t) \cdot w_n(t) \leq 0 \leq \dot{p}(t) \cdot w_n(t).$$

Since $\dot{p} = ae^t > 0$, we only need to show

$$\begin{aligned}\dot{Q}_n(t) \cdot w_n(t) &= 0 \text{ and} \\ w_n(t) &\geq 0.\end{aligned}$$

But these follow immediately from definition of the adjoint (22). $\square$

To show that a set of curves $\{\mathcal{S}_{i,n}\}_{i,n}$ is an SFE we only have to verify the second-order conditions. In Theorem 16 below, we will show that when the matrix $\mathbf{A}$ is constant over all Q , then the SFE candidate that follows a given trajectory is maximal. However, when $\mathbf{A}$ varies it will generally be necessary to calculate $Z_{i,n}(q_{i,n}, p)$ for all permitted $q_{i,n}$ and p in order to verify that $Z > 0$ to the left of the candidate curve and $Z < 0$ to the right.

Ironed trajectories

On strictly monotone segments of the candidate trajectory, the adjoint w will be zero. Here we study the adjoint when the agents in one node are ironing. Define the *ironing function* $Y_m(Q)$ for node $m = 1, 2$ as the value of $Z_m(Q, \dot{Q})$ if the generators at node m were to make an offer with $\dot{Q}_m = 0$, i.e. if they were to iron over the point Q , with the other node n offering so that $\dot{Q}_n = U_n(Q)$. With all costs identical and constant, we find that Y is, like $\mathbf{A}$, a function of Q alone. To see this for Y_1 , first set Z_2 to zero and solve for $\dot{Q}_2$ when $\dot{Q}_1 = 0$,

$$\begin{aligned}Z_2\left(Q, \begin{bmatrix} 0 \\ \dot{Q}_2 \end{bmatrix}\right) &= (N_2 - 1)\phi_{2,2}\dot{Q}_2 - \phi_{2,2}Q_2 = 0, \text{ so} \\ \dot{Q}_2 &= \frac{Q_2}{(N_2 - 1)}.\end{aligned}$$

Note that the assumption that $N_2 \geq 2$ guarantees this slope is finite. If we had $N_2 = 1$, then when the generators at node 2 ironed, the monopolist at node 2 would face perfectly inelastic residual demand, so their optimal strategy would be to offer all their output at the highest possible price. The presence of a local competitor ensures that all generators face elastic residual demand and so precludes such degeneracy. When we substitute the above value for $\dot{Q}_2$ into (21) to find Z_1 , we get

$$\begin{aligned}Y_1 &= N_2\phi_{1,2}\dot{Q}_2 - \phi_{1,1}Q_1, \\ &= \frac{N_2}{N_2 - 1}\phi_{1,2}Q_2 - \phi_{1,1}Q_1.\end{aligned}$$

By a similar calculation,

$$Y_2 = \frac{N_1}{N_1 - 1}\phi_{2,1}Q_1 - \phi_{2,2}Q_2.$$

Note that Y is a function of Q only, and does not depend on p or $\dot{Q}$. As a matrix equation,

$$\begin{aligned}Y(Q) &= \begin{bmatrix} -\phi_{1,1} & \frac{N_2}{N_2-1}\phi_{1,2} \\ \frac{N_1}{N_1-1}\phi_{2,1} & -\phi_{2,2} \end{bmatrix} Q \\ &= |\mathbf{A}| \begin{bmatrix} \frac{-\phi_{1,1}}{N_2-1} & 0 \\ 0 & \frac{-\phi_{2,2}}{N_1-1} \end{bmatrix} \mathbf{A}^{-1} Q \\ &= |\mathbf{A}| \begin{bmatrix} \frac{-\phi_{1,1}}{N_2-1} & 0 \\ 0 & \frac{-\phi_{2,2}}{N_1-1} \end{bmatrix} U(Q). \end{aligned} \tag{23}$$

Proposition 11 completes the argument that identical generators in a two-node network will offer identical supply curves in equilibrium.

Proposition 11. *Under Assumption (6), at any given point along a candidate trajectory they either all agents iron or none do.*

The proof uses the following properties of Y and U , which later will also serve to control integration along candidate SFE trajectories.

Lemma 12. *The vector function $Y : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ has the following properties:*

1. If $Y(Q) = 0$, then either $Q = 0$ or $|\mathbf{A}| = 0$.
2. For $n \neq m$, if $Q_m = 0$ and $Q_n > 0$, then $Y_m(Q) > 0$ and $Y_n(Q) < 0$.

Lemma 13. *The slope $\dot{Q} = U(Q)$ has both components positive if and only if either $|\mathbf{A}| < 0$ and $Y_1, Y_2 > 0$, or $|\mathbf{A}| > 0$ and $Y_1, Y_2 < 0$.*

The sign restriction on the adjoint function to an optimal slope-constrained supply curve gives a necessary condition for the endpoints of ironed segments in terms of the signs of Z .

Lemma 14 (Corners). *Suppose $\mathcal{S}^* = \{(q(t), p(t)) : t \in [0, T]\}$ is a local maximum of $\Pi(\mathcal{S})$, subject to $\dot{q}, \dot{p} \geq 0$. Suppose that $\dot{p}(t) > 0$ for all $t \in [0, T]$. If $\dot{q} = 0$ on some interval $(\hat{t}, t_1)$ and $\dot{q} > 0$ on an adjacent interval $(t_0, \hat{t})$, i.e. $\hat{t}$ is at the low-priced end of an ironed segment, then there exists an $\epsilon > 0$ such that $Z(q(t), p(t)) \geq 0$ for all $t \in (\hat{t}, \hat{t} + \epsilon)$.*

Similarly, if $\dot{q} = 0$ on some interval $(t_0, \hat{t})$ and $\dot{q} > 0$ on an adjacent interval $(\hat{t}, t_1)$, i.e. $\hat{t}$ is at the high-priced end of an ironed segment, then there exists an $\epsilon > 0$ such that $Z(q(t), p(t)) \leq 0$ for all $t \in (\hat{t} - \epsilon, \hat{t})$.

Proof. For $\hat{t}$ at the low-priced end of an ironed segment, consider the adjoint w . Because $\dot{q} > 0$ on $(t_0, \hat{t})$, it must be that $w = 0$ on this interval. So $w = 0$ at $\hat{t}$. Because $w \geq 0$, it must be that $Z = \dot{w} \geq 0$ in some neighbourhood of $\hat{t}$.

The proof for $\hat{t}$ at the high-priced end of an ironed segment is similar. □

Lemma 14 means that ironed segments can only have their ends in certain parts of offer-quantity space. Suppose $|\mathbf{A}| > 0$. From the optimality conditions discussed in Section 2.3, the inner end of a segment with node m ironing needs

$$U(Q) \geq 0 \text{ and } Y_m \geq 0. \tag{24}$$

Lemma 13 implies that $U(Q) \geq 0$ if and only if $Y_m \leq 0$, but this means that the only points where (24) is possible are those where $Y_m = 0$.

On the other hand, suppose now that $|\mathbf{A}| < 0$. In this case Lemma 13 implies that $Y \geq 0$ wherever $U(Q) \geq 0$, so that ironing may begin from any point on an unconstrained segment.

At the lower priced (inner) end of an ironed segment, where node 1 is ironing, Y_1 must be positive

$$Y_1 = \frac{N_2}{N_2 - 1} \phi_{1,2} Q_2 - \phi_{1,1} Q_1 > 0.$$

Rearranging the expression for Y_1 , we see that $Y_1 > 0$ if and only if

$$Q_2 \frac{\phi_{1,2}}{\phi_{1,1}} > \frac{N_2 - 1}{N_2} Q_1.$$

Note that the right-hand side of this inequality is constant over the ironed interval.

Ironing will often occur at the production capacity of the generators in one of the nodes. Ironing in the interior of an SFE curve can only occur where Y_m passes from positive to negative. For node 1 ironing, this means that $Q_2 \frac{\phi_{1,2}}{\phi_{1,1}}$ must be decreasing; so the non-congestion probability $\frac{\phi_{1,2}}{\phi_{1,1}}$ must be decreasing faster than the offers at node 2 are growing. The generators at a particular node will withhold energy by offering an inelastic (vertical) segment when it increases the probability of congestion in their favor; i.e. when it increases the probability they will be on the high-price end of the line.

4 Constant-coefficient ODE

Suppose that the network congestion probabilities are *exogenously* given, that is, they do not depend on the offers made by generators. Then in (16) we would have a matrix $\mathbf{A}$ that is constant over all of offer-quantity space. Two situations in which this occurs are when the line is either surely congested or surely uncongested. In a pool market, there is no congestion, and the interaction probability $\frac{\phi_{m,n}}{\phi_{n,n}} = 1$, so $|\mathbf{A}| = 1 - N_1 - N_2 < 0$.

Whereas in 3.1, we obtained a scalar ODE, with constant $\mathbf{A}$, (16) becomes a linear system of ODE with constant coefficients that can be solved analytically.

Let there be two generators at each node: $N_1 = N_2 = 2$. Then in the surely congested case we have

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

which has positive determinant and two positive eigenvectors. In the surely uncongested case we have

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix},$$

which has negative determinant and one eigenvalue of each sign. These two cases correspond to different types of market behavior.

Fences

To guide our integration across offer-quantity space, we apply the concepts of fences from planar dynamics (see Hubbard and West 1995). By Lemma 13, strictly monotone segments of a candidate trajectory will be found only in the region where $U(Q) > 0$. The boundaries of this region are the loci $Y_m = 0$ for $m = 1, 2$. The shape of these loci can tell us a lot about the strictly monotone segments.

Definition 15. A *right-hand fence* for the differential equation $\dot{Q} = U(Q)$ is a curve $Q = h(s) = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}(s)$, $s \in [0, 1]$ whose right-hand normal

$$N(s) = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix}(s)$$

satisfies

$$N(s) \cdot U(Q(s)) > 0 \quad \forall s.$$

A solution to $\dot{Q} = U(Q)$ can cross a right-hand fence from left to right, but never the other way. A *left-hand fence* is a curve $Q = h(s) = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} (s)$, $s \in [0, 1]$ whose right-hand normal

$$N(s) = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} (s)$$

satisfies

$$N(s) \cdot U(Q(s)) < 0 \quad \forall s.$$

With some help from the implicit function theorem, it is possible to parametrize the $Y_m = 0$ loci as curves. When these curves are monotone in both components they form fences. These fences are bounds on the solution to $\dot{Q} = U(Q)$ when integrating in the stable direction.

Positive determinant

When the matrix $\mathbf{A}$ has positive determinant, the local markets at either end of the congested line behave more independently of each other, and there are many monotone trajectories of $\dot{Q} = \mathbf{A}^{-1}Q$ through the origin.

The matrix $\mathbf{A}$ will have two positive eigenvalues λ_1 and λ_2 , with eigenvectors v_1 and v_2 . Then $\mathbf{A}^{-1}$ has the same eigenvectors with eigenvalues $\frac{1}{\lambda_1}$ and $\frac{1}{\lambda_2}$. Solutions to $\mathbf{A}\dot{Q} = Q$ will have the form

$$Q = a_1 e^{\frac{t}{\lambda_1}} v_1 + a_2 e^{\frac{t}{\lambda_2}} v_2, \quad (25)$$

with constants of integration a_1 and a_2 . For any a_1 there will be a range of a_2 that gives

$$\lim_{t \rightarrow -\infty} \dot{Q}(t) \geq 0.$$

Figure 3 shows, on the left, a candidate trajectory for positive $|\mathbf{A}|$, together with the $Y_1 = 0$ and $Y_2 = 0$ loci. If they are parametrized away from the origin, then $Y_1 = 0$ forms a left-hand fence and $Y_2 = 0$ a right-hand fence. The trajectory is solved ‘backwards’, in the decreasing t direction, so the fences force the trajectory towards the origin.

Negative determinant

When the matrix $\mathbf{A}$ has negative determinant, the markets at either end of the line behave more like a pool market.

There will be two eigenvalues $\lambda_1 > 0 > \lambda_2$, with eigenvectors v_1 and v_2 . Local solutions to $\mathbf{A}\dot{Q} = Q$ will still have the form (25) but now, since $t \rightarrow -\infty$ as $p \rightarrow c$, the exponent on $e^{\frac{t}{\lambda_2}}$ goes to infinity and this term blows up as price approaches marginal cost, so $a_2 = 0$.

There is only one monotone trajectory through the origin, the one along v_1 . This is the same argument used by Klemperer and Meyer (1989) to show the uniqueness of a symmetric SFE passing through zero output at marginal cost.

Figure 3 shows, on the right, a candidate trajectory for negative $|\mathbf{A}|$, together with the $Y_1 = 0$ and $Y_2 = 0$ loci. As before, if they are parametrized away from the origin, then $Y_1 = 0$ forms a left-hand fence and $Y_2 = 0$ a right-hand fence. However, they now act as a funnel in the opposite direction. Solving in the ‘forwards’, increasing t , direction, the trajectory is forced towards the dominant eigenvector of $\mathbf{A}$.

Algorithm 1 Constant-coefficient algorithm.

- Calculate $|\mathbf{A}|$.
 - If $|\mathbf{A}| > 0$, start from the output capacities $\bar{Q}$ and work in the decreasing t direction. If either entry of $U(\bar{Q})$ is negative, agents in that node iron inwards at their output capacity until $U(Q(\hat{t})) \geq 0$. Integrate $\mathbf{A}\dot{Q} = Q$ to connect that point to the origin by a monotone trajectory.
 - If $|\mathbf{A}| < 0$, start from the origin and work in the increasing t direction. Integrate $\mathbf{A}\dot{Q} = Q$ so that $Q_m(\hat{t}) = \bar{Q}_m$ for one of the nodes m . Generators in that node then iron outwards at their output capacity until $Q_n(\bar{t}) = \bar{Q}_n$ for $n \neq m$.
 - Choose T so that $Q(T) = \bar{Q}$ and then choose a to solve $\bar{p} - c = ae^T$. Then define $p(t) = c + ae^t$.
-

An algorithm to solve for SFE under constant coefficients

These solutions we can now put together with the ironing rules above to give an algorithm for computing SFE trajectories. The *Constant-coefficient Algorithm 1* will find analytic solutions for SFE when the matrix $\mathbf{A}$ is constant. It tells us in what direction to integrate and in which mode to integrate — not-ironing or ironing — to find an SFE.

There are two sorts of integration involved. When the capacity constraints are not binding, we integrate $\mathbf{A}\dot{Q} = Q$, and when they are we integrate the local N -opoly in the node that is not constrained. With constant $\mathbf{A}$, we find that the monotonicity constraints only ever bind when the capacity constraints bind.

To integrate $\mathbf{A}\dot{Q} = Q$ with constant coefficients $\mathbf{A}$, use the general solution (25). When both eigenvalues of $\mathbf{A}$ are positive ($|\mathbf{A}| > 0$), then both eigenvectors are admissible in a non-decreasing solution, and a_1 and a_2 should be chosen so that $Q(\hat{t})$ has the required value. When the smaller eigenvalue $\lambda_2 < 0$ — the $|\mathbf{A}| < 0$ case — then $a_2 = 0$, and a_1 can be chosen so that $Q(\hat{t})$ has the required value.

To integrate for generators at node m ironing at their output capacity, hold $Q_m = \bar{Q}_m$ constant, and for the other node n solve the differential equation

$$\dot{Q}_n = \frac{Q_n}{N_n - 1} \quad \left[\dot{Q}_m = 0 \right] \quad (26)$$

on an interval $[\hat{t}, \bar{t}]$. The solution to (26) will have the form

$$Q_n(t) = ae^{\frac{t}{N_n - 1}}.$$

and we can choose a and $\hat{t}$ to satisfy the boundary conditions. Set $\bar{t} = \log \bar{p}$ and solve $Q_n(\bar{t}) = ae^{\frac{\bar{t}}{N_n - 1}} = \bar{Q}_n$ to find the constant of integration a . The inner boundary conditions will determine the inner endpoint of the ironed interval $\hat{t}$; these depend on the direction. If ironing outwards, we set $U_m(Q(\hat{t})) = 0$ for the ironing node m . If ironing inwards, we instead set $Q_m(\hat{t}) = \bar{Q}_m(\hat{t})$ and $Q(\hat{t}) = \kappa \mathbf{v}_1$ for some $\kappa \in \mathbb{R}$, i.e. $Q(\hat{t})$ lies on the leading eigenvector $\mathbf{v}_1$ and has its m th component at the upper bound Q_m .

Theorem 16. *The Constant Coefficient Algorithm 1 will always find a candidate trajectory that is an SFE. Moreover, this SFE is unique.*

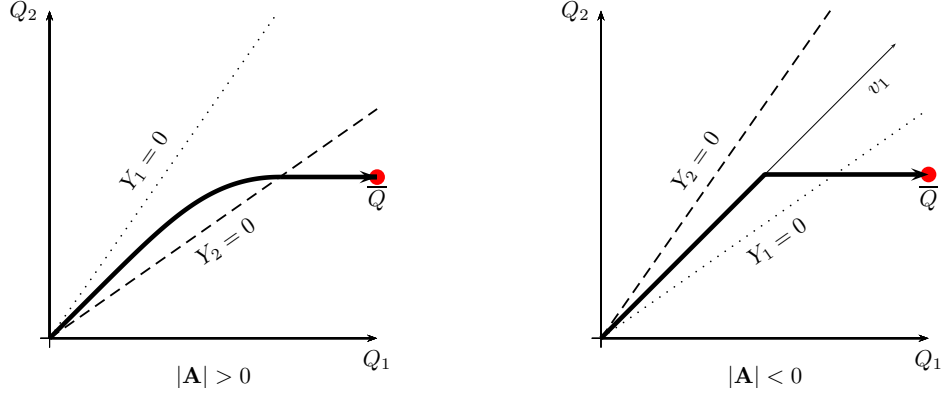


Figure 3: Constant-coefficient algorithm, outward ($|\mathbf{A}| > 0$) and inward ($|\mathbf{A}| < 0$) cases.

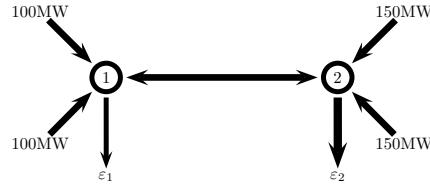


Figure 4: Output capacities for generators in Example 17.

Example 17 (Network SFE with constant integration probabilities). Suppose there are two generators at each end of the transmission line, as shown in Figure 4. The generators at node 1 have 100MWh of generation capacity, while those at node 2 each have 150MWh, i.e.

$$\bar{Q} = \begin{bmatrix} 100 \\ 150 \end{bmatrix}.$$

The price cap is \$1000 and the cost of production is zero. We calculate SFE trajectories for three different integration matrices $\mathbf{A}$.

First, in a completely integrated market we have $\frac{\phi_{m,n}}{\phi_{n,n}} = 1$ for $m \neq n$, so

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}.$$

Taking the determinant, we see $|\mathbf{A}| = -3 < 0$. The positive eigenvalue is $\lambda = 3$, with eigenvector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. The trajectory through the origin is thus

$$Q(t) = a_1 e^{\frac{t}{3}} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

This trajectory hits a production capacity constraint at $\hat{Q} = \begin{bmatrix} 100 \\ 100 \end{bmatrix}$, so agents at node 1 must iron while agents at node 2 play as a duopoly out to their production capacity, so

$$Q_2(t) = a_2 e^t.$$

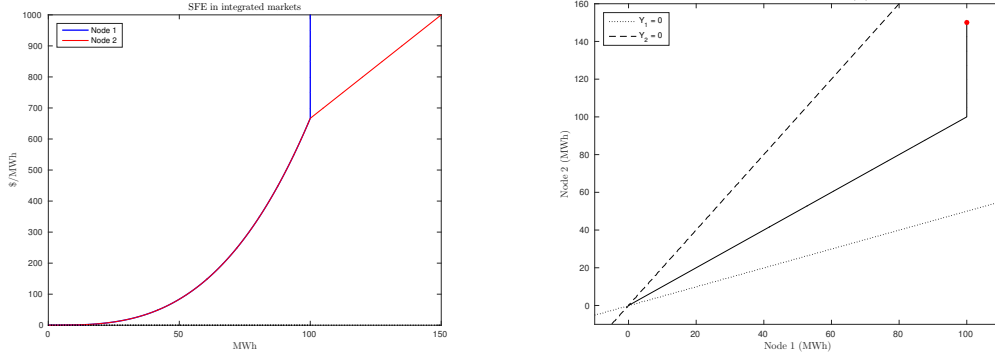


Figure 5: Fully integrated market.

When we change variables to $p = c + e^t$, we get the piecewise SFE

$$Q(p) = \begin{cases} a_1 (p - c)^{\frac{1}{3}} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} & p < \hat{p} \\ \begin{bmatrix} 100 \\ a_2 (p - c) \end{bmatrix} & p > \hat{p}. \end{cases}$$

We now choose the scale constants a_1, a_2 so that the supply functions are continuous and reach the output capacity point $\bar{Q}$ at the price cap $\bar{p}$. It is simple to calculate

$$\begin{aligned} \hat{p} &= 0.0015 \\ a_1 &= 11.4471 \text{ and} \\ a_2 &= 0.15. \end{aligned}$$

The SFE and trajectory are plotted in Figure 5.

Second, in a fully separated market we have $\frac{\phi_{m,n}}{\phi_{n,n}} = 0$ for $m \neq n$, so

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The determinant of $\mathbf{A}$ is 1, which is positive. At the production capacity point

$$U(\bar{Q}) = \mathbf{A}^{-1} \bar{Q} = \begin{bmatrix} 100 \\ 150 \end{bmatrix} > 0,$$

so there is no need of ironing. To integrate inwards, we can solve two independent ODEs to obtain

$$Q(p) = a_1 (p - c) \begin{bmatrix} 1 \\ 0 \end{bmatrix} + a_2 (p - c) \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

To meet the production capacities at the price cap $Q(\bar{p}) = \bar{Q}$, we set

$$\begin{aligned} a_1 &= 0.1 \text{ and} \\ a_2 &= 0.15. \end{aligned}$$

The SFE and trajectory are plotted in Figure 6.

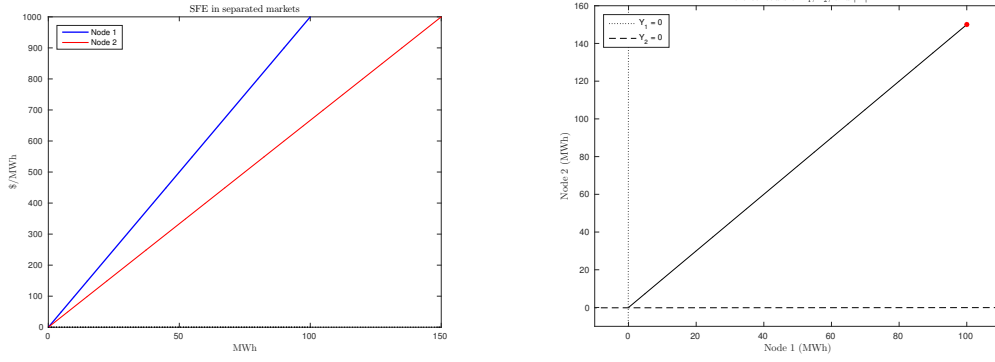


Figure 6: Fully separated markets.

Third and finally, consider a partially integrated market. In general a partially integrated market will *not* have constant $\mathbf{A}$, so this final example is rather artificial, however it demonstrates how Algorithm 1 will generalize to variable $\mathbf{A}$. Let $\frac{\phi_{m,n}}{\phi_{n,n}} = 0.45$ for $m \neq n$, so

$$\mathbf{A} = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}.$$

The determinant $|\mathbf{A}| = 0.09 > 0$. The slope at the production capacity point is given by

$$U(\bar{Q}) = \mathbf{A}^{-1}\bar{Q} = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 100 \\ 150 \end{bmatrix} = \begin{bmatrix} -184.2 \\ 315.8 \end{bmatrix},$$

which has a negative component, so we must iron at node 1's production capacity. The ironing stops when $Y_1 = 0.9Q_2 - Q_1 = 0$. Solving for Q_2 gives $\begin{bmatrix} 100 \\ 111.1 \end{bmatrix}$ as the point where U becomes non-negative. From this point we can solve $\dot{Q} = U(Q)$ to the origin by the eigenvector method. The eigenvectors of $\mathbf{A}$ are

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ and } \begin{bmatrix} -1 \\ 1 \end{bmatrix},$$

with eigenvalues 1.9 and 0.1 respectively. Thus the general solution to $\dot{Q} = U(Q) = \mathbf{A}^{-1}Q$ is

$$Q = a_1 e^{\frac{t}{1.9}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + a_2 e^{\frac{t}{0.1}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

for constants of integration a_1 and a_2 . For the solution to pass through $\begin{bmatrix} 100 \\ 111.1 \end{bmatrix}$ at $\hat{t}$, we must have

$$a_1 e^{\frac{\hat{t}}{1.9}} = 105.55 \text{ and } a_2 e^{\frac{\hat{t}}{0.1}} = 5.55.$$

When node 1 is ironing, node 2 generators offer a curve

$$Q_2 = a_3 e^t$$

and the boundary condition $Q_2(\log 1000) = 150$ gives $a_3 = 0.15$. We can then find $\hat{t}$ by solving $100 = 0.15e^{\hat{t}}$, for $\hat{t} = \log 666.66$. Knowing $\hat{t}$, we can solve for a_1 and a_2 . Putting it all together,

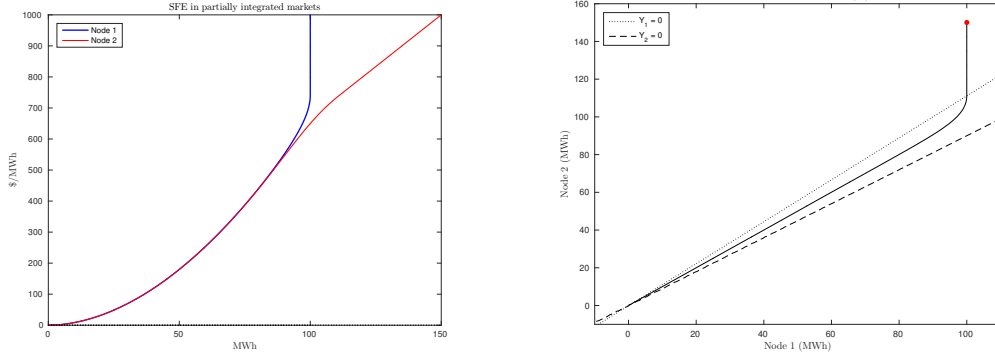


Figure 7: Partially integrated markets

the piecewise SFE is

$$Q(p) = \begin{cases} a_1 (p - c)^{\frac{1}{1.9}} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} + a_2 (p - c)^{10} \cdot \begin{bmatrix} -1 \\ 1 \end{bmatrix} & p < \hat{p} \\ \begin{bmatrix} 100 \\ a_3 (p - c) \end{bmatrix} & p > \hat{p} \end{cases}$$

with

$$\begin{aligned} a_1 &= 3.4452 \\ a_2 &= 3.2036 \cdot 10^{-28} \\ a_3 &= 0.15 \text{ and} \\ \hat{p} &= 666.66. \end{aligned}$$

The SFE and trajectory are plotted in Figure 7. Because the a_2 component is so small, the gap between the supply curves is imperceptible at prices below \$500/MW.

5 Varying coefficients: ironing within SFE

The analysis of the ODE (16) with constant coefficients will inform the solution methods for the general problem, in which the matrix $\mathbf{A}$ varies with the offer quantities Q . We shall extend Algorithm 1 to find candidate trajectories first in markets where $|\mathbf{A}|$ has the same sign everywhere, then in markets where $\mathbf{A}$ becomes singular along the candidate trajectory. Here is an example of the difficulties we can expect.

Example 18 (Bad behavior). Suppose there are two players in each node, and independent normal demand shocks in the two nodes with means (120, 120) and standard deviations (60, 50). The two nodes are connected by a line with a capacity of 45MW. There are two generators at each node.

Figure 8 shows, on the left, the zero-contours of Y_m and $|\mathbf{A}|$. The determinant $|\mathbf{A}|$ is negative inside an elliptical region centered on the mean of the demand shock distribution. This shows that the transmission constraint is least likely to bind when the total offers at each node are exactly equal to the mean of the demand shock distribution, since in a sense this is when the demand for transmission is the least.

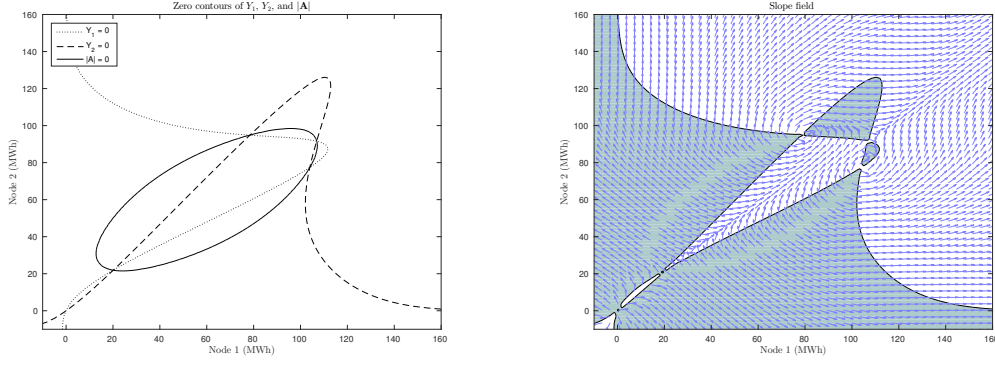


Figure 8: The signs of $|A|$ and Y_m for the two node model, and the slope field $U(Q)$, with bad behavior.

On the right of Figure 8 is the slope field $U(Q) = \mathbf{A}^{-1}Q$ with the regions where $U(Q)$ has a negative component shaded. This is a poorly-behaved instance of our two-node model. We have a very asymmetric demand shock distribution, and a line whose transmission capacity is not quite large enough to integrate the two nodal markets in all demand outcomes. It is clear that there are points that can only be reached by a candidate trajectory with an ironed segment in its interior.

Under constant $\mathbf{A}$, the rays $Q = \mathbf{A} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $Q = \mathbf{A} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ are the boundaries between positive and negative Y_m . These rays are fences so will be crossed at most once by a non-decreasing trajectory from the origin, even counting ironed segments. With varying coefficients, the boundary $Y_m = 0$ can have other shapes, even something like that shown in Figure 8.

With constant $\mathbf{A}$, the differential equations can be solved analytically. In what follows, we will turn to numerical methods, in particular the `ode23t` initial-value solver in MATLAB Gladwell et al. (2003). It is a Runge-Kutta method with trapezoid rule, which works like the Euler method, but is better suited to stiff ODEs. It takes in a function $f(t, Q)$ and an initial point $Q(t_0) = Q_0$ and keeps extending an approximate solution to $\dot{Q} = f(t, Q)$ until some stopping condition is met. Because of the existence of singular points (in particular the origin) in our ODE, we are concerned with numerical stability. Depending on the sign of $|\mathbf{A}|$, numerical solutions can diverge very rapidly if we integrate in the wrong direction.

Not all instances of the two-node model are as complicated as the one in Figure 8. If the region where $U(Q) \geq 0$ is sufficiently cone-like, then Algorithm 1 can be applied, using numerical integration instead of taking eigenvalues of $\mathbf{A}$ to follow $\dot{Q} = U(Q)$. If $|\mathbf{A}|$ keeps the same sign across a neighbourhood of the valid trajectory, then the algorithm needs only a little modification.

Our improved algorithm for finding the SFE will work when the coefficient matrix $\mathbf{A}$ is allowed to vary, provided that the determinant of $\mathbf{A}$ does not change sign along the candidate curve. In order for it to work, we need to make some technical assumptions on the behavior of the problem, in addition to Assumption 6.

Assumption 19. [Conditions for ironing] We make the following three assumptions about the ironing functions Y_m .

1. For $m = 1, 2$, there are only a finite number of points where

$$Y_m = 0, \text{ and } \frac{\partial}{\partial Q_m} Y_m = 0.$$

2. For any two points $Q, \tilde{Q}$ with $Q_1 = \tilde{Q}_1$ and $Q_2 < \tilde{Q}_2$, if $Y_1(\tilde{Q}_1, \tilde{Q}_2) < 0$ and $Y_1(\tilde{Q}_1, Q_2) > 0$ then $Y_2(\tilde{Q}_1, \tilde{Q}_2) < 0$. Similarly, for any two points $Q, \tilde{Q}$ with $Q_1 < \tilde{Q}_1$ and $Q_2 = \tilde{Q}_2$, if $Y_2(\tilde{Q}_1, \tilde{Q}_2) < 0$ and $Y_2(Q_1, \tilde{Q}_2) > 0$, then $Y_1(\tilde{Q}_1, \tilde{Q}_2) < 0$.
3. Regions where $Y_m > 0$ are simply connected (connected and without holes).

These assumptions are sufficient to ensure that the solution algorithm terminates after finitely many iterations and that the resulting trajectory is unique.

The first part of Assumption 19 means that the curves defined by $Y_m = 0$ for $m = 1, 2$ change between upward and downward sloping only a finite number of times. The functions Y_m are continuous functions of two variables, so generically the loci $Y_m = 0$ will be curves. At any point where $Y_m = 0$ we can see whether the direction $\dot{Q} = U(Q)$ points into the $Y_m > 0$ region or out of it. We assume there are only a finite number of points where the direction changes from in to out. At each such change, a curve along $Y_m = 0$ changes between a left-hand fence and a right-hand fence. This part of the assumption is true for any reasonable choice of network and shock distribution (i.e. the shock density is continuous with a finite number of local maxima).

The second part of Assumption 19 means that there is no point at which we can turn a hard corner from one node ironing to the other node ironing, without a strictly increasing segment in between. Note that the ironing functions in Figure 8 violate condition 2 of Assumption 19. Take, for instance $Q = (90, 80)$ and $\tilde{Q} = (90, 110)$. We shall see in Example 29 that this configuration leads to multiple equilibrium trajectories, since it allows one node to iron and then the other in adjoining segments of the trajectory.

The third part is more for convenience, it cuts down the number of different cases that must be considered in the proofs of uniqueness. For both the second and third parts it is possible to construct counterexamples with a bivariate normal shock distribution.

As in the constant-coefficient problem above, the direction of integration depends on the determinant of $\mathbf{A}$. We will define algorithms for computing candidate trajectories in the positive and negative determinant cases, and then use them to calculate some example SFEs.

5.1 Varying coefficients and positive determinant

The key idea of the Inward Algorithm 2 is to get closer to the origin with every round of the main loop. After each iteration, the number of changes in handedness of the fences $Y_m = 0$ between the end of the partial trajectory and the origin decreases. Hence the algorithm will eventually terminate. After some initial ironing, we begin and end each round of the main loop at a point $\hat{t}$ with

$$w(\hat{t}) = 0 \text{ and } U(Q(\hat{t})) \geq 0.$$

Upon exit from the loop we will have $t \rightarrow -\infty$, so $p \rightarrow c$.

The algorithm drives a Runge-Kutta integrator in the stable direction of the slope field. We integrate either the unconstrained differential equation $\dot{Q} = U(Q)$, or a specially modified

Algorithm 2 Inward Algorithm

Initialization: Initial trajectory is defined at the single point $Q(T) = \bar{Q}$. The adjoint w will be initialized after the first ironed segment.

Initial ironing: Calculate $Y(\bar{Q})$; either A) $Y_m > 0$ for $m = 1$ or 2 or B) $Y \geq 0$.

- A) Node m irons inwards until $Y_m = 0$. Set $\hat{t}$ to be the parameter at the inner end of this ironed segment. Set $w(\hat{t}) = 0$, and $w_m(t) = \int_{\hat{t}}^t Y_m dt$ and $w_n(t) = 0$ for $t \in [\hat{t}, \bar{t}]$, $n \neq m$.
- B) There is no ironing. Set $w(\bar{t}) = 0$ and $\hat{t} = \bar{t}$.

Main loop:

Integrate inwards (in the direction of decreasing t) from $Q(\hat{t})$ along $\dot{Q} = U(Q)$ until either: C) $Q(t) \rightarrow 0$ or D) $Y_m = 0$ for $m = 1$ or 2 .

- C) We are DONE, exit the loop
- D) Apply the Inward Interval Search Algorithm 3 with node m ironing to obtain a trajectory and its adjoint on an interval (t_*, t^*) . Update $\hat{t}$ to t_* . We now have $w(\hat{t}) = 0$ and $U(Q(\hat{t})) \geq 0$, and can go round the loop again.
-

Algorithm 3 Inward Interval Search

Until tolerance satisfied:

Choose an initial t^* in the interval $(\hat{t}, \bar{t}]$.

Node m irons inwards from a point $Q(t^*)$ that is on the partial trajectory with $w(t^*) = 0$.

Stop integration when A) $Q_n \rightarrow 0$ for $n \neq m$ or B) $w_m = 0$.

- A) Decrease t^* .
- B) Increase t^* .

Repeat until the inner endpoint of the ironed segment, at parameter t_* , has $w_m(Q(t_*)) = 0$ and $Y_m(Q(t_*)) = 0$, to within desired tolerance.

Output: An ironed trajectory on the interval (t_*, t^*) , with $w_m(t_*) = 0$ and $Y_m(Q(t_*)) = 0$.

ordinary differential equation if the agents in one node are ironing. Both types of differential equation will be integrated in the decreasing t direction.

When node m is ironing, we integrate the differential equation system

$$\dot{Q}_m = 0, \quad \dot{Q}_n = U_n(Q), \quad \dot{w}_m = Y_m(Q), \quad \dot{w}_n = 0.$$

This has four scalar functions but is essentially a two-dimensional system as the Q_m and w_n terms remain constant. It derives from the optimality conditions of Lemma 10.

The interval search Algorithm 3 is implemented as a bisection method search. It is illustrated in Figure 9. Above is a plot of trajectory space. The thin black line is the zero contour of Y_1 , with coloured shading showing it is positive on the left and negative on the right. The initial segment of trajectory starts from the production capacity point (red dot) and continues down until it meets the $Y_1 = 0$ contour. Four trajectory segments with node 1 agents ironing run down from this to meet the lower bound $Q_2 \geq 0$. Below are plots of the adjoint values as a function of price $w_1(p)$, for these four ironed curves.

The goal of the interval search is to find an ironed segment of trajectory where the top end is

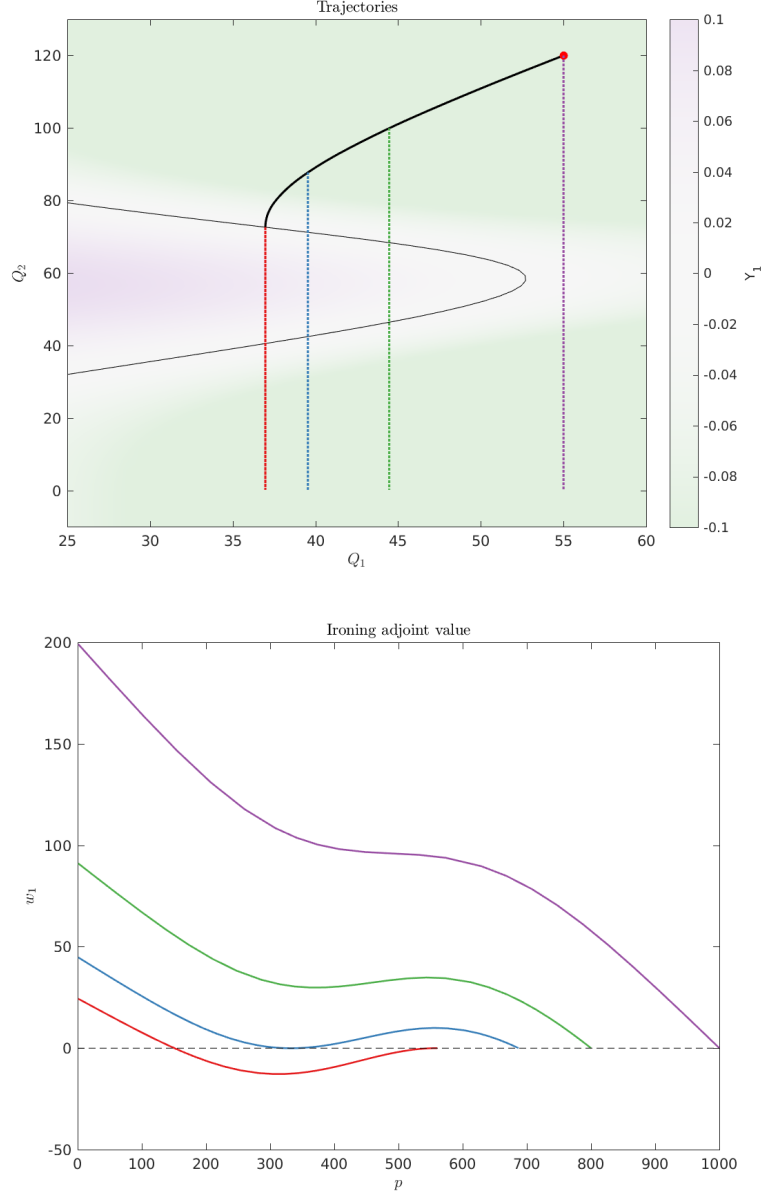


Figure 9: Typical interval search problem in case D) of the Inward Algorithm. In the top plot, the initial trajectory is the thick solid curve, found by integrating down from the production capacity point (red dot). Four ironed trajectory segments come down from this trajectory segment (coloured lines). In the bottom plot are the adjoint values w_1 for the four ironed trajectories. The blue curve is the solution we seek, with its gentle *frôlement* of zero.

on the existing partial trajectory, and the bottom end has $w_m = 0$ and $Y_m = 0$ simultaneously. This means that the adjoint function reaches zero with a slope of zero, as does the blue curve in Figure 9. If node 1 starts ironing from the point $\bar{t}$, the purple segment, then at no point on the ironed segment is $w_1 = 0$, so Q_2 will reach zero. This is case A), and we must then decrease the starting parameter for the ironing. If on the other hand w_1 reaches zero, it must be that $Y_1 \geq 0$, and we will want to increase the ironing start point so that Y_1 is closer to zero at the end of the ironed segment.

Proposition 20 (Inward Algorithm finds a unique trajectory). *Suppose that $|\mathbf{A}| > 0$ within the offer region $[0, \bar{Q}_1] \times [0, \bar{Q}_2]$, and that the regularity assumptions 19 are satisfied. Then the Inward Algorithm will find a candidate trajectory from the origin to $\bar{Q}$. Moreover, if every interval search subproblem has a unique solution, then this trajectory is the unique candidate trajectory from the origin to $\bar{Q}$.*

5.2 Varying Coefficients and negative determinant

The key idea of the Outward Algorithm 4 is that on each round of the main loop the trajectory is extended out through one swerve of the fence loci. Integration cannot begin from $p = c$, $Q = 0$, as our ODE is singular there. At this point, however, $|\mathbf{A}| < 0$ implies $\mathbf{A}$ has one positive and one negative eigenvalue. We use the eigenvector associated with the positive eigenvalue as an initial direction to integrate along. At the end of each round, either we have reached the production capacity point $Q(\hat{t}) = \bar{Q}$, or we have a partial trajectory on $[0, \hat{t}]$ with

$$w(\hat{t}) = 0 \text{ and } U(Q(\hat{t})) \geq 0$$

The ODEs that are integrated in the Outward Algorithm are exactly the same as in the Inward Algorithm above. The difference is in the direction of integration; the Outward Algorithm always integrates in the increasing t direction.

The goal of the Outward Interval Search Algorithm 5 is to find an segment with agents at node m ironing on an interval (t_*, t^*) where $Q(t_*)$ lies on the existing partial trajectory, with $w(t_*) = 0$, while at the other end we have either $Y_m(Q(t^*)) = 0$, or $Q_n(t^*) = \bar{Q}_n$ for $m \neq n$. In Figure 10, we obtain this second case; node $m = 1$ is ironing and the blue trajectory reaches $Q_2(t) = \bar{Q}_2$ and $w_1(t) = 0$ at the same parameter value t . The algorithm is essentially the same as the Inward Interval Search Algorithm 3. The differences are that the integration goes in the opposite direction and that instead of $Q_n \rightarrow 0$, we stop when Q_n reaches an upper bound. There are two ways to get into the interval search now: either agents at node m reach $w_m = 0$ while ironing at their output capacity, or the slope component m reaches zero when the trajectory crosses the locus $Y_m = 0$.

Proposition 21 (Outward Algorithm finds a unique trajectory). *Suppose that the regularity assumptions 19 are satisfied within the rectangle $[0, \bar{Q}_1] \times [0, \bar{Q}_2]$. If the Outward Algorithm maintains a negative determinant $|\mathbf{A}(Q(t))| < 0$ at every point on its partial trajectory, then it will find a candidate trajectory from the origin to $\bar{Q}$. Moreover, if every interval search subproblem has a unique solution, then this trajectory is the unique candidate trajectory from the origin to $\bar{Q}$.*

Theorems 20 and 21 take us most of the way to showing existence and uniqueness of SFE for specific market models. We just have to check, in the specific instance we are modeling, that the line searches do not admit multiple solutions (there will usually only be one interval search to check), and that the second-order conditions for SFE are met.

Algorithm 4 Outward Algorithm

Initialization: Initial trajectory is defined on a single point. Let $\mathbf{v}_1 > 0$ be the leading eigenvector of $\mathbf{A}$ at 0. Choose a $\delta > 0$ so that $\delta \mathbf{v}_1$ is small enough to be within some tolerance bound of the origin. Set $t = 0$, $Q(0) = \delta \mathbf{v}_1$, and $w(0) = 0$. Begin integrating from this point in the direction of increasing t , following $\dot{Q} = U(Q)$.

Main loop:

Integrate outwards, following $\dot{Q} = U(Q)$, until either: A) $Q_m(\hat{t}) = \bar{Q}_m$ for some m or B) $Y_m(Q(\hat{t})) = 0$ for some m .

A) Node m irons outwards until either: C) $Q_n = \bar{Q}_n$ for $n \neq m$ and $w_m > 0$ or D) $w_m = 0$.

C) We are DONE, exit the loop.

D) Apply the Outward Interval Search Algorithm 5 with node m ironing to obtain a trajectory and its adjoint on an interval (t_*, t^*) . Update $\hat{t}$ to t^* . We have either E) $w_m(\hat{t}) = 0$ and $Y_m(Q(\hat{t})) = 0$ or F) $w_m(\hat{t}) = 0$ and $Q_n(\hat{t}) = \bar{Q}_n(\hat{t})$.

E) Since we have $w(\hat{t}) = 0$, $U_m(Q(\hat{t})) = 0$, and $U_n(Q(\hat{t})) > 0$, we can begin a new round of the main loop.

F) Node n irons outwards until $Q_m = \bar{Q}_m$. We are DONE, exit the loop.

B) Apply the Outward Interval Search Algorithm 5 with node m ironing to obtain a trajectory and its adjoint on an interval (t_*, t^*) . Update $\hat{t}$ to t^* . We have either E) $w_m(\hat{t}) = 0$ and $Y_m(Q(\hat{t})) = 0$ or F) $w_m(\hat{t}) = 0$ and $Q_n(\hat{t}) = \bar{Q}_n(\hat{t})$.

E) Since we have $w(\hat{t}) = 0$, $U_m(Q(\hat{t})) = 0$, and $U_n(Q(\hat{t})) > 0$, we can begin a new round of the main loop.

F) Node n irons outwards until $Q_m = \bar{Q}_m$. We are DONE, exit the loop.

End of main loop.

Algorithm 5 Outward Interval Search

Until tolerance satisfied:

Choose an initial t_* in the interval $(0, \hat{t})$.

Node m irons outwards from a point $Q(t_*)$ that is on the partial trajectory with $w(t^*) = 0$.

Stop integrating when either A) $Q_n = \bar{Q}_n$ for $n \neq m$ or B) $w_m = 0$

A) Decrease t_* .

B) Increase t_* .

Repeat until the outer endpoint of the ironed segment, at parameter t^* , has $w_m(t^*) = 0$ and $Y_m(Q(t^*)) = 0$ to within desired tolerance.

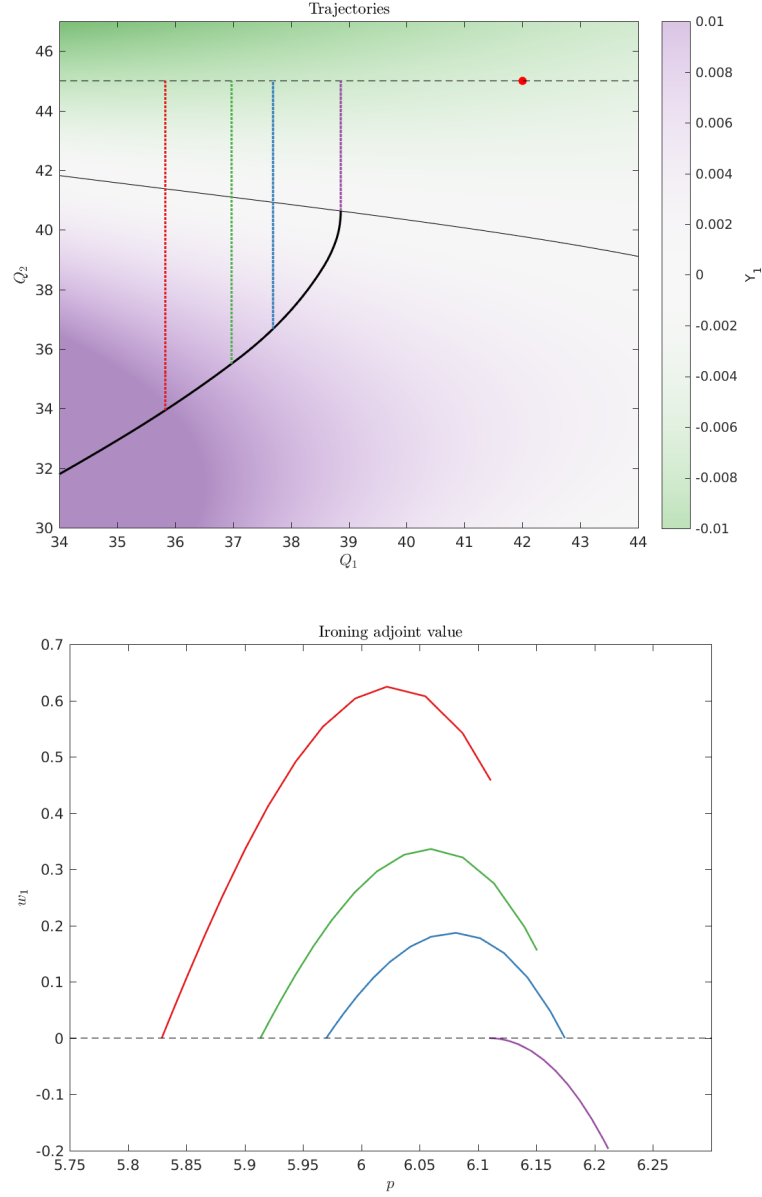


Figure 10: Typical interval search problem in case B) of the Outward Algorithm. In the top plot, the partial trajectory is the solid thick curve entering from the bottom right, found by integrating outwards from the origin. Four ironed trajectories shoot up to the node 2 production capacity at $\bar{Q}_2 = 45$, from this partial trajectory. In the bottom plot are shown the adjoint functions w_1 for these four trajectory segments. The blue curve is the solution we seek; it reaches $Q_2 = \bar{Q}_2$ at the same point that w_1 hits 0.

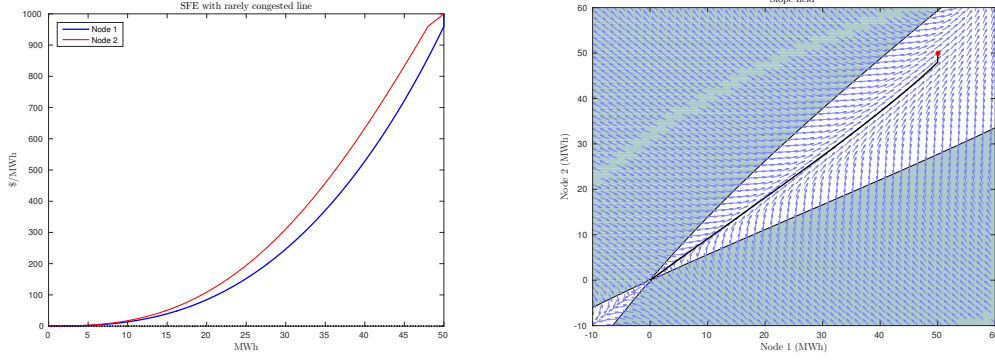


Figure 11: Fully integrated market

Examples

We now present two examples of SFE in the two-node network with two generators at each node. The first example is one where the line is rarely congested, so $|\mathbf{A}|$ is always negative along the candidate trajectory.

Example 22 (SFE over a rarely-congested line). The demand shock distribution is an independent bivariate normal with mean $(100, 50)$, and standard deviations $(60, 40)$. The line has a transmission capacity of 60MW. There are two generators in each node and each generator has a production capacity of 50MW and zero cost of production.

Figure 11 shows the SFE on the left and the trajectory through the slope field on the right. This is similar to the four-player pool market. The node 2 generators offer slightly more competitively than the node 1 generators, so they exhaust their output capacity at a price of about \$950 (slightly below the price cap). The node 2 generators iron over the highest prices while the node 1 generators compete as a duopoly.

In a second example we see that when the line is congested often enough for $|\mathbf{A}|$ to be always positive, there can be strategic withholding by generators in the downstream node.

Example 23 (SFE over an oft-congested line, with strategic withholding). Consider a two-node network with an independent normal demand shock distribution with mean $(100, 100)$, standard deviations $(60, 40)$, and zero correlation. The line has a transmission capacity of only 25MW. There are two generators in each node, and each generator has a production capacity of 120MW and zero cost.

Figure 12 shows the SFE on the left and the trajectory through the slope field on the right. Notice that the supply curve of the node 1 generators is very inelastic at around 50MW output. This is because withholding energy from the market increases the probability of constraining the line so node 1's price exceeds that of node 2. The advantage of greater market power through congestion is almost enough to overwhelm the gain from selling more energy.

We can check the second-order conditions by plotting $Z_{i,m}(q_{i,m}, p)$ for one agent at each node. The field Z is evaluated taking every other agent's offer to be either Q_1 or Q_2 depending on which node they are in. In Figure 13 we plot heat maps of the Z values for an agent in node 1 (left) and in node 2 (right). The SFE curves are plotted in black. It is easy to see that the sufficient conditions of Theorem 2 are met; the black SFE curves lie on the $Z = 0$ contour and $Z > 0$ everywhere to the left of the SFE and $Z < 0$ everywhere to the right.

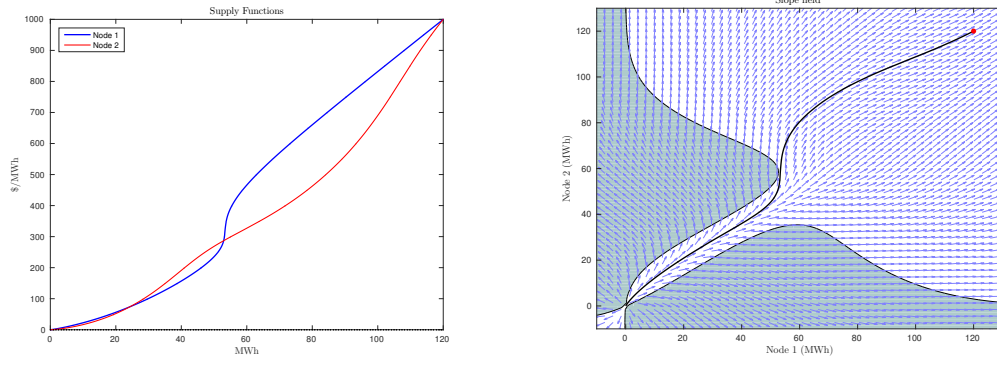


Figure 12: Highly congested market

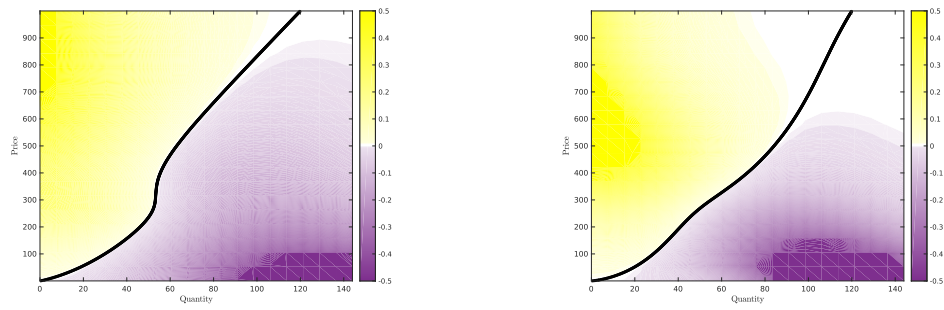


Figure 13: Z values around the SFE

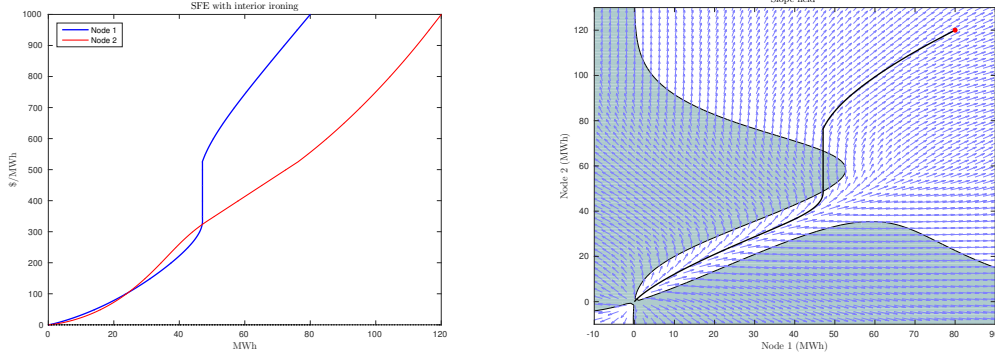


Figure 14: Highly congested market

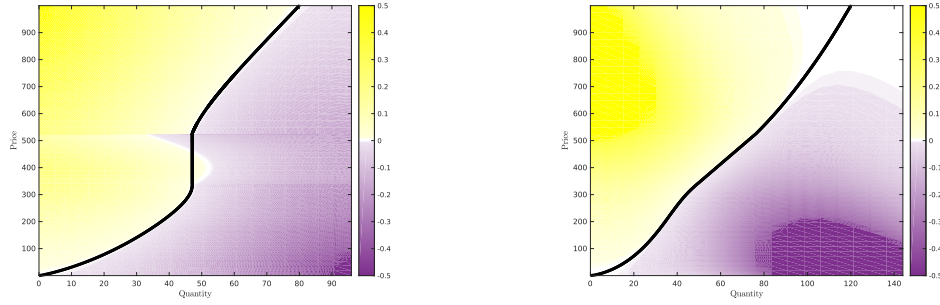


Figure 15: Z values around the SFE

Example 24. Figure 14 shows the SFE and trajectories for the network and demand of Example 23 when the node 1 generators have their production capacity reduced to 80MW each. They now find an equilibrium where they iron at output values below capacity. This has the effect of pushing the outer end of the SFE into a regime where congestion into node 1 is much more likely, so the node 1 generators exploit this by withholding.

Figure 15 shows the verification of the second-order conditions for this SFE with an ironed segment.⁵ On the left is plotted Z for a generator at node 1. On the unironed segments of the supply curve, we have $Z > 0$ to the left and $Z < 0$ to the right as before, with $Z = 0$ along the curve. On the ironed segment, $Z > 0$ on the inner half of the curve and $Z < 0$ on the outer half. This is consistent with Lemma 14. The level of Q_1 where ironing occurs is chosen so that the average of Z over this segment is zero. On the right is plotted Z for a generator at node 2. This generator has a strictly monotone supply curve in the SFE with $Z = 0$ along the entire length with $Z > 0$ to the left and $Z < 0$ to the right.

6 Singularities in SFE

We will now relax the assumption that $\mathbf{A}$ is everywhere non-singular, to allow the sign of $|\mathbf{A}|$ to change once along the trajectory of an SFE. The techniques used in the Inward and Outward Algorithms to form partial equilibrium trajectories can be adapted to this situation.

⁵For the remaining examples we omit the plots for the second-order conditions because they are satisfied in all of them.

The SFE trajectory will transition from a region with positive determinant to one with negative determinant. There will be no proof of existence and uniqueness of solutions in this section, but the example SFE can still be verified by plotting the first-variation derivative for agents at each node and using this to check the first- and second-order equilibrium conditions.

Assume that there exists a candidate trajectory between the origin and the production capacity point which crosses the line $|\mathbf{A}| = 0$ exactly once. There are two cases to consider, either:

1. At the origin $|\mathbf{A}(0)| > 0$, and at the production capacity point $|\mathbf{A}(\bar{Q})| < 0$, or
2. At the origin $|\mathbf{A}(0)| < 0$, and at the production capacity point $|\mathbf{A}(\bar{Q})| > 0$.

We will analyze first one, then the other.

Origin positive

If $|\mathbf{A}| > 0$ at the origin, then there are ironed and non-ironed ways to cross the $|\mathbf{A}| = 0$ line.

If we are to cross without ironing, we need to find a *critical point*, one where $|\mathbf{A}|$, Y_1 , and Y_2 are all zero. We can integrate away from this critical point; in towards the origin and out towards the production capacities. Each of these subproblems reduces to the case where $\mathbf{A}$ is everywhere non-singular.

No interior ironing: There is a region where $|\mathbf{A}| < 0$ and $Y_1, Y_2 > 0$. Then there must be a critical point where $|\mathbf{A}|$, Y_1 , and Y_2 are all zero. In the phase portrait classification, this point is a *saddle point*. Starting at the critical point, the solution is found by integrating outwards until one of the production capacities is reached, and integrate inwards towards the origin.

Ironing across the boundary: If there is no region where $|\mathbf{A}| < 0$ and $Y_1, Y_2 > 0$, then at the production capacities, Y_1 or Y_2 must be negative, say Y_2 . In this case the other node, node 1, should iron at its production cap, integrating down from the price cap. Eventually Y_2 will reach zero, at which point we can integrate along $Z = 0$ to the origin.

Example 25. Two examples, again with two generators at each node.

Figure 16 shows the SFE and trajectory for the case where there is a repelling node in the interior of the offer region. It is located at (17.46, 19, 75).

The SFE is solved piecewise in three pieces. One from the interior node to the origin, another from the interior node to the capacity constraint of node 2, and the third with node 2 ironing at their production capacity up to node 1's production capacity.

Figure 17 shows the SFE and trajectory for the case where there is no intersection between $|\mathbf{A}| < 0$ and $Y_2 > 0$. Node 1 irons from their production capacity down until $Y_1 = 0$. Then there is a monotone trajectory of $\mathbf{A}\dot{Q} = Q$ down to the origin. Note that node 1's SFE curve has continuous derivative at the point where they reach their production capacity; this is because the generators at node 1 begin ironing on the locus $Y_1 = 0$.

Origin negative

If $|\mathbf{A}(0)| < 0$, then a candidate trajectory is found as follows. We integrate inwards from the production capacity point using the Inward Algorithm and outwards from the origin using the Outward Algorithm. These two segments of trajectory will each reach the line $|\mathbf{A}| = 0$. Several cases are possible, but we will consider three:

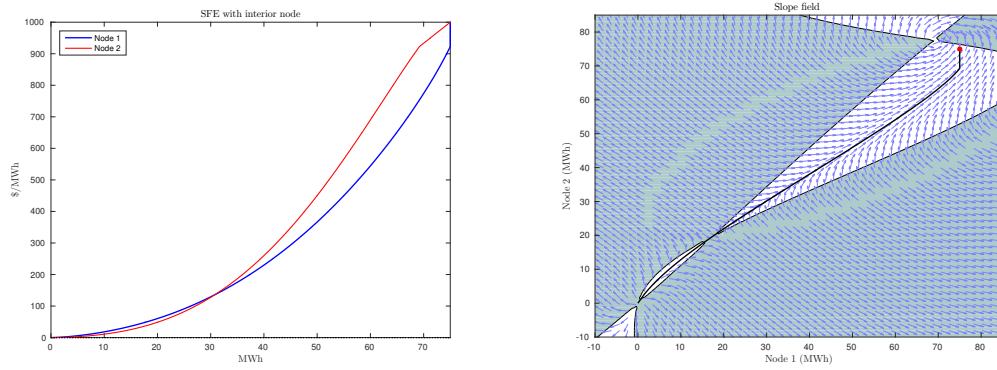


Figure 16: Interior singular point

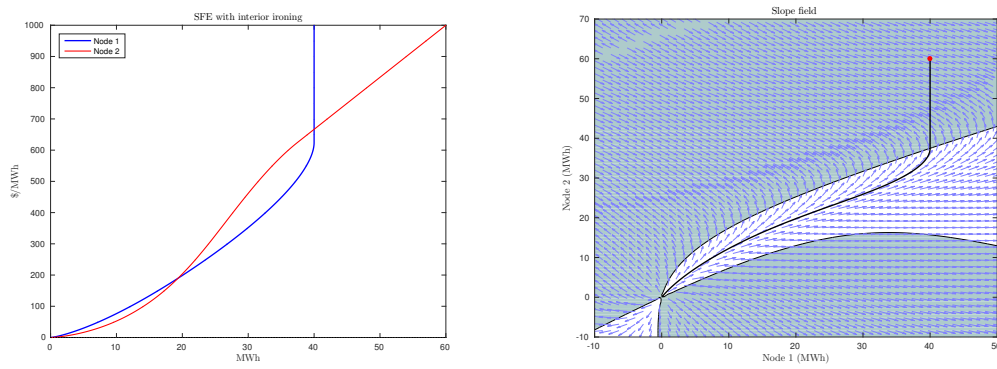


Figure 17: Ironing across boundary

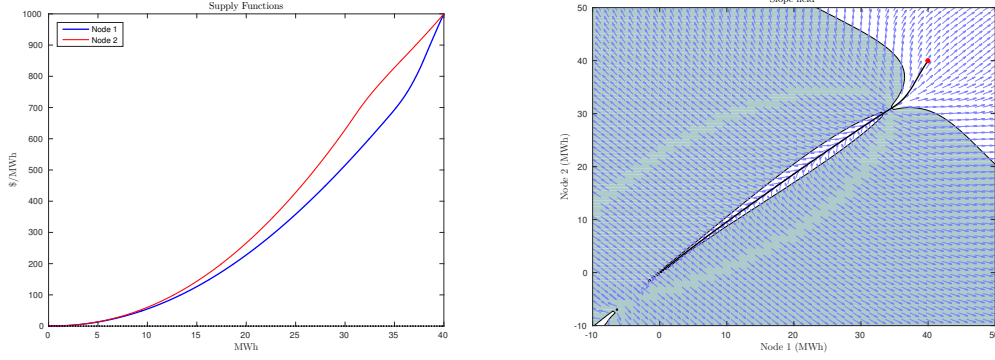


Figure 18: $|\mathbf{A}(0)| < 0$: SFE and trajectory on slope field where inward and outward parts meet without ironing.

1. Both inward and outward parts meet $|\mathbf{A}| = 0$ on a non-ironing segment.
2. Both inward and outward parts reach $|\mathbf{A}| = 0$ while ironing, with the same node ironing.
3. Both inward and outward parts reach $|\mathbf{A}| = 0$ while ironing, with different nodes ironing.

In the first case, the two parts will generally meet $|\mathbf{A}| = 0$ at the same point, so it is a simple matter of piecing them together into a candidate trajectory. In the second case, an interval search procedure is applied to iron between the two parts of the trajectory. In the third case we find that the regularity Assumption 19.2 is violated, whereby we may obtain a candidate trajectory with two ironed segments one after the other. We also lose uniqueness of the candidate trajectory in this last case, as we shall see in Example 29.

To solve all three cases we integrate inwards from the production capacity point and outwards from the origin. We calculate two partial trajectories $Q(t)$ and $\tilde{Q}(u)$, in parameters t and u defined on intervals

$$t \in (-\infty, t_1] \quad u \in [u_0, u_1].$$

We will have $Q(t) \rightarrow 0$ as $t \rightarrow -\infty$, $\tilde{Q}(u_1) = \bar{Q}$, and $Q(t_1) = \tilde{Q}(u_0)$. To make a single continuous trajectory, we just have to rescale u . Set $u(t) = u_0 - t_1 + t$ and $T = t_1 + u_1 - u_0$, and define $Q(t) = \tilde{Q}(u(t))$ on $[t_1, T]$. Then

$$u(t_1) = u_0 - t_1 + t_1 = u_0,$$

so Q is extended to be continuous on $(-\infty, T]$ and connect the origin to $\bar{Q}$.

Example 26 (Monotone trajectory across singularity). The demand shock distribution is an independent bivariate normal with mean $(20, 30)$ and standard deviations $(50, 42)$. The line has a transmission capacity of 32MW. There are two generators in each node and each generator has a production capacity of 40MW.

Figure 18 shows an SFE and trajectory where the solution to $\dot{Q} = U(Q)$ crosses the line $|\mathbf{A}| = 0$ at the point $(33.94, 30.54)$ with a positive slope on both sides. The candidate trajectory is just the concatenation of these two segments.

Example 27 (Ironing across singularity). The demand shock distribution is an independent bivariate normal with mean $(80, 80)$, standard deviations $(60, 40)$. The line has a transmission capacity of 60MW. There are two generators in each node and each generator in node 1 has a production capacity of 50MW while those in node 2 have 150MW each.

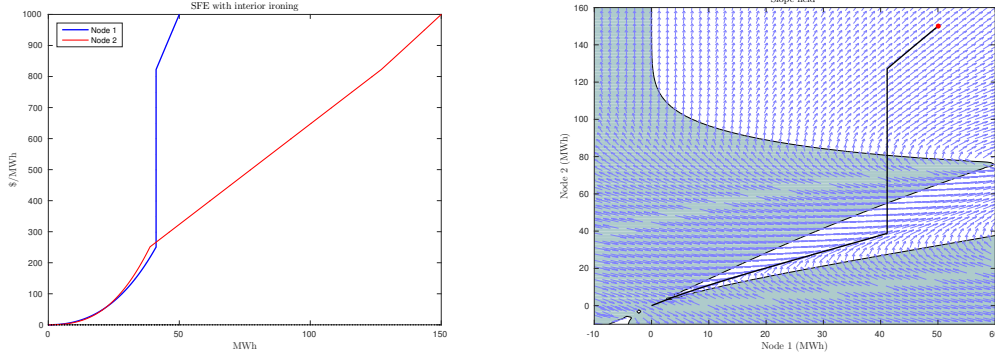


Figure 19: $|\mathbf{A}(0)| < 0$: SFE and trajectory on slope field where node 1 irons across $|\mathbf{A}| = 0$ boundary

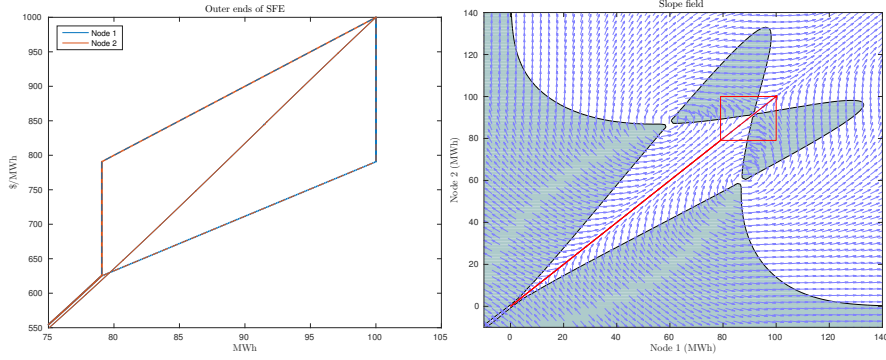


Figure 20: The outer ends of the multiple SFE, and their trajectories through the slope field $U(Q)$.

Figure 19 shows an SFE and trajectory where the node 1 generators must iron over the singularity $|\mathbf{A}| = 0$. The quantity at which they should iron is found by an interval search similar to that in the Inward and Outward Algorithms. The difference is that here the ironed section should join the partial trajectories determined by the origin and the production capacity point and have the mean of Y_1 be zero over the whole segment.

6.1 Trajectories are not always unique

The following two examples are counterexamples to the conjecture that all two-node networks with ‘well-behaved’ demand shock distributions give rise to unique equilibrium trajectories. The non-uniqueness is different in kind to that identified by Klemperer and Meyer (1989), in that even with fixed boundary conditions on the equilibrium curves there can be multiple SFE.

Example 28. Non-unique trajectories in the two-node normal model

We calculate three equilibrium trajectories for a two node network in a market that violates condition 2 of Assumption 19. To see this, take $Q = (80, 60)$ and $\tilde{Q} = (80, 100)$. We have $Y_1(Q) > 0$ and $Y_2(Q) < 0$, but $Y_2(\tilde{Q}) > 0$.

The demand shocks are independently normally distributed with mean 80MW and standard deviation 50MW. The line has a transmission capacity of 50MW. There are two generators in

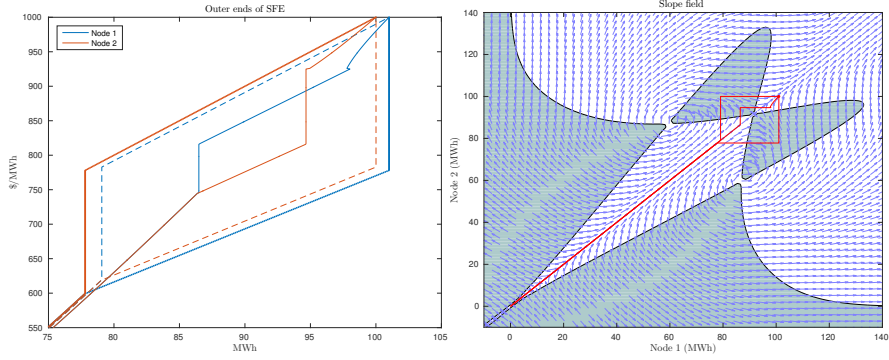


Figure 21: SFE and slopes for same market, but with node 1 generators each having an extra 1MW capacity.

each node and each of these has a production capacity of 100MW.

In Figure 20 are plotted the outer ends of the SFE and their trajectories through the slope field. The three SFE have different ways of reaching the production capacity point from the unique trajectory through the origin:

1. The perfectly symmetric monotone trajectory passing clean through the critical point at approximately (91,91) (thin lines).
2. Node 1 iron until node 2 hit their cap, then node 2 iron at the cap (dashed lines).
3. Node 2 iron until node 1 hit their cap, then node 1 iron at the cap (thick lines).

Example 29. To show that the non-uniqueness of the candidate trajectories is not just due to the symmetry of the market between the two nodes, we give each of the node 1 generators an extra 1MW of production capacity. We find there are also three equilibrium trajectories to the new production capacity point of (101,100) MW.

In Figure 21 are plotted the outer ends of the SFE and their trajectories through the slope field. The three SFE have different ways of reaching the production capacity point from the unique trajectory through the origin:

1. Node 1 iron, then node 2 iron, to end on a strictly increasing segment (thin lines).
2. Node 1 iron until node 2 hit their cap, then node 2 iron at the cap (dashed lines).
3. Node 2 iron until node 1 hit their cap, then node 1 iron at the cap (thick lines).

This second example suggests that the set of parameters for which such multiple trajectories exist is non-null.

6.2 Correlated shocks

The correlation between demand shocks in a two node network can have a significant influence on how the transmission constraints affect competition. All the examples so far have used independent bivariate normal distributions on the demand shocks, but it is no more difficult to compute SFE when the shocks are correlated.

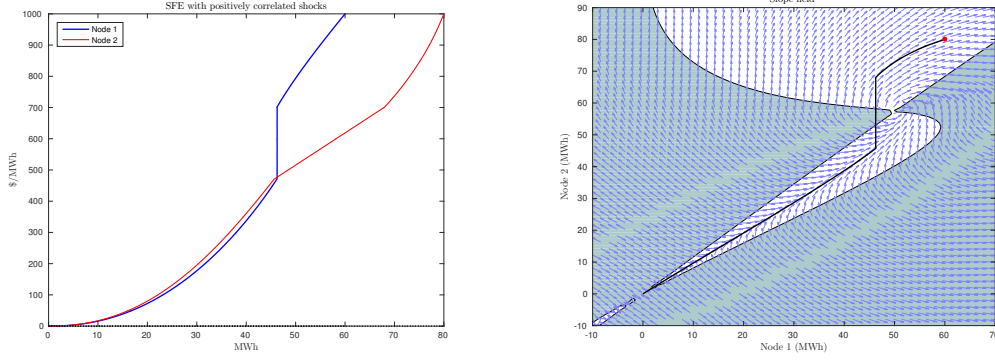


Figure 22: An SFE for positively correlated nodal demand shocks, and its trajectory through the slope field $U(Q)$.

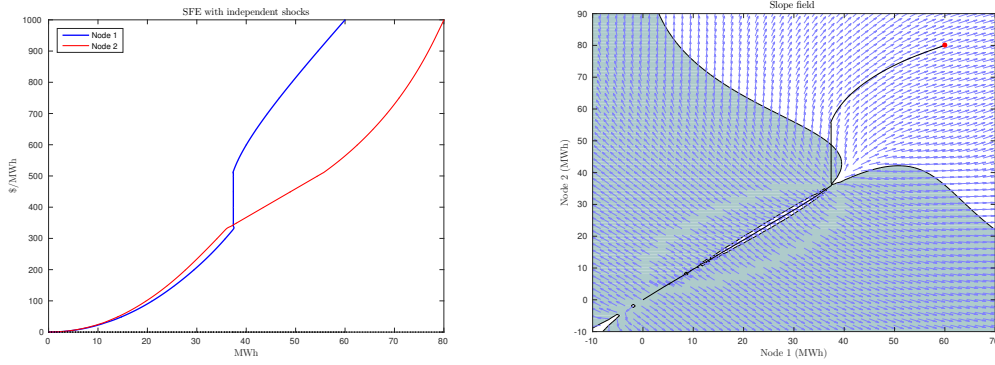


Figure 23: An SFE for independent nodal demand shocks, and its trajectory through the slope field $U(Q)$.

Example 30. The demand shock distribution is a bivariate normal with mean $(40, 40)$ and standard deviations $(60, 50)$. The line has a transmission capacity of 36MW. There are two generators in each node; the node 1 generators each have a production capacity of 60MW, while those in node 2 have 80MW.

The correlation of the demand shock distribution takes three values: 0.3, 0, and -0.3 . SFE for these three values are plotted in Figures 22-24. The most important effect of correlation is to shift the $|\mathbf{A}| = 0$ boundary, where offer strategy switches from very competitive to more monopolistic. When shocks are positively correlated, the SFE stays in the competitive regime up to higher output levels than when the correlation is negative.

6.3 Multiple singularities

It can happen that a candidate trajectory will require two or more crossings of $|\mathbf{A}| = 0$. If this happens, we can find the interior singular points, and use those to subdivide the problem of finding a trajectory into subproblems where there is negative $|\mathbf{A}|$ on the inner part and positive $|\mathbf{A}|$ on the outer part.

In the examples we have studied, the region where $|\mathbf{A}| < 0$ is convex. The worst cases that we see are like Figure 8, where the $Y_1 = 0$ and $Y_2 = 0$ loci cross four times in the positive

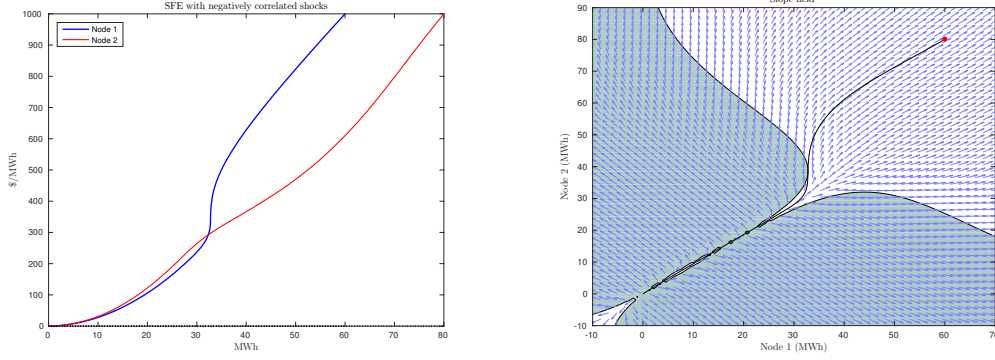


Figure 24: An SFE for negatively correlated nodal demand shocks, and its trajectory through the slope field $U(Q)$.

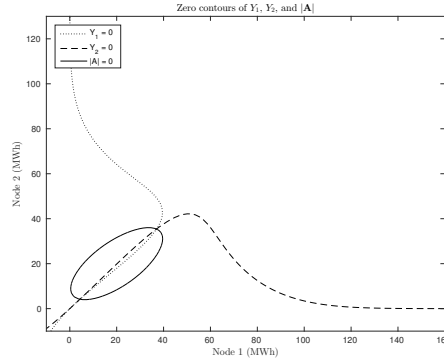


Figure 25: The outer ends of the multiple SFE, and their trajectories through the slope field $U(Q)$.

quadrant, to give three singular points plus the origin. Although our numerical techniques can be extended to handle such cases, there will be issues with non-uniqueness of the solution, as we see in example 28.

Example 31 (Multiple singularities). The demand shock distribution is an independent bivariate normal with mean $(40, 40)$ and standard deviations $(60, 50)$. The line has a transmission capacity of 36MW. There are two generators in each node; the node 1 generators each have a production capacity of 120MW, while those in node 2 have 150MW. The results are plotted in Figure 25. There are critical points at $(4.31, 4.17)$ and $(36.99, 35.42)$, and the candidate trajectory goes smoothly through both of them.

7 Conclusion

In this paper, we apply the optimality conditions for a supply function equilibrium in a lossless and loop-free (radial) transmission network derived by Wilson (2008) and adapted to the market distribution function approach by Holmberg and Philpott (2017a). We consider a very simple transmission network, with two nodes connected by a single constrained line. Under Assumption 6, we obtain equilibrium conditions that transform into a planar linear autonomous system of ordinary differential equations. Solutions to this system are not always monotone,

so we define another system of ODEs which the equilibrium trajectories must follow when the monotonicity constraints bind.

The constant-coefficient version of this ODE reveals some important qualitative features of the general system. It also allows us to compute the limit cases of a line with zero or infinite capacity. We then extend these computational techniques to markets where the probability of congestion varies with the quantities offered. Under some strict sufficient conditions, we show that the extended algorithms will compute a unique candidate trajectory. We may check the second-order conditions to ensure that such a trajectory corresponds to an SFE. When the probability of congestion becomes sufficiently high, we find SFE in which all players at some node will offer a vertical segment in the interior of their supply curves. This can be viewed as a strategic withholding of energy from the market in order to bring about a more favourable congestion state in high-price outcomes.

An important contribution of the present work is to show non-uniqueness of SFE in a network setting. Uniqueness of equilibrium is a desirable property in market models. In Example 29 there are multiple trajectories at high prices in markets where the system of differential equations has singularities. In these cases, boundary conditions are not sufficient to select a unique equilibrium.

References

- E. Anderson and X. Hu. Finding supply function equilibria with asymmetric firms. *Operations Research*, 56(3):697–711, May–June 2008.
- E. Anderson and X. Hu. Asymmetric supply function equilibria with forward contracts. *Journal of Optimization Theory and Applications*, 152(1):198–224, 2012.
- E. Anderson and A. Philpott. Optimal offer construction in electricity markets. *Mathematics of Operations Research*, 27(1):82–100, 2002.
- E. Anderson, A. Philpott, and H. Xu. Modelling the effects of interconnection between electricity markets subject to uncertainty. *Mathematical Methods of Operations Research*, 65(1):1–26, 2007.
- S. Borenstein, J. Bushnell, and S. Stoft. The competitive effects of transmission capacity in a deregulated electricity industry. *RAND Journal of Economics*, 31(2):294–325, 2000.
- A. Downward, G. Zakeri, and A. Philpott. On Cournot equilibria in electricity transmission networks. *Operations Research*, 58(4, part 2):1194–1209, 2010.
- I. Gladwell, L. Shampine, and S. Thompson. *Solving ODEs with MATLAB*. Cambridge University Press, New York, 2003.
- P. Holmberg. Supply function equilibrium with asymmetric capacities and constant marginal costs. *The Energy Journal*, 28(2):pp. 55–82, 2007.
- P. Holmberg. Unique supply function equilibrium with capacity constraints. *Energy Economics*, 30(1): 148, 2008.
- P. Holmberg. Numerical calculation of an asymmetric supply function equilibrium with capacity constraints. *European Journal of Operational Research*, 199(1):285, 2009.
- P. Holmberg and A. Philpott. On supply function equilibria in radial transmission networks. Working paper, Electric Power Optimization Centre, www.epoc.org.nz, 2017a.
- P. Holmberg and A. Philpott. Supply function equilibria in networks with transport constraints. Working paper no. 945, Research Institute for Industrial Economics, www.ifn.se, 2017b.
- J. H. Hubbard and B. H. West. *Differential equations : a dynamical systems approach*. Springer, New York, 1995.
- P. D. Klemperer and M. A. Meyer. Supply function equilibria in oligopoly under uncertainty. *Econometrica*, 57(6):1243–1277, November 1989.

- C. D. Meyer, editor. *Matrix Analysis and Applied Linear Algebra*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2000.
- H. Niu. *Models For Electricity Market Efficiency and Bidding Strategy Analysis*. PhD thesis, University of Texas, Austin, 2005.
- K. Ruddell. Supply function equilibrium over a constrained transmission line ii: multiple plants and nodal price derivatives. Working Paper Series XXXX, IFN – Research Institute for Industrial Economics, Stockholm, April 2018.
- A. Seierstad and K. Sydsaeter. *Optimal control theory with economic applications*. North-Holland, New York, 1987.
- R. Wilson. *Nonlinear Pricing*. Oxford University Press, 1993.
- R. Wilson. Supply function equilibrium in a constrained transmission system. *Operations Research*, 56(2):369–382, 2008.

Appendix

Proof of Theorem 1. See Anderson and Philpott (2002) for a proof using Green’s Theorem. The condition $Z = 0$ can also be derived from the Euler-Lagrange condition from the calculus of variations.

Given a functional

$$\Pi = \int_0^T \Lambda(q, p, \dot{q}, \dot{p}) dt,$$

the *Euler-Lagrange condition* is

$$\begin{aligned} \frac{\partial \Lambda}{\partial q} - \frac{d}{dt} \frac{\partial \Lambda}{\partial \dot{q}} &= 0 \text{ and} \\ \frac{\partial \Lambda}{\partial p} - \frac{d}{dt} \frac{\partial \Lambda}{\partial \dot{p}} &= 0, \end{aligned}$$

where $\frac{d}{dt}$ is the total derivative.⁶ This condition must hold for any stationary curve, in particular for a local maximum. To apply the Euler-Lagrange condition to (3), note that the integrand has the linear (in $\dot{q}$ and $\dot{p}$) form

$$\Lambda(q, p, \dot{q}, \dot{p}) = u(q, p) \dot{q} + v(q, p) \dot{p}, \quad (27)$$

with

$$\begin{aligned} u(q, p) &= R(q, p) \psi_q \quad \text{and} \\ v(q, p) &= R(q, p) \psi_p \end{aligned} \quad (28)$$

and take the appropriate derivatives:

$$\begin{aligned} \frac{\partial \Lambda}{\partial q} &= \frac{\partial u}{\partial q} \dot{q} + \frac{\partial v}{\partial q} \dot{p}, \\ \frac{\partial \Lambda}{\partial \dot{q}} &= u(q(t), p(t)), \text{ and} \\ \frac{d}{dt} \frac{\partial \Lambda}{\partial \dot{q}} &= \frac{\partial u}{\partial q} \dot{q} + \frac{\partial u}{\partial p} \dot{p}. \end{aligned}$$

⁶see, for instance, Seierstad and Sydsaeter 1987

So

$$\frac{\partial \Lambda}{\partial q} - \frac{d}{dt} \frac{\partial \Lambda}{\partial \dot{q}} = \dot{p} \left(\frac{\partial v}{\partial q} - \frac{\partial u}{\partial p} \right). \quad (29)$$

By similar working

$$\frac{\partial \Lambda}{\partial p} - \frac{d}{dt} \frac{\partial \Lambda}{\partial \dot{p}} = \dot{q} \left(\frac{\partial u}{\partial p} - \frac{\partial v}{\partial q} \right). \quad (30)$$

For both (29) and (30) to vanish, either $\dot{q}$ and $\dot{p}$ both vanish (which we rule out by assumption) or

$$Z = \frac{\partial v}{\partial q} - \frac{\partial u}{\partial p} = 0.$$

Substituting in the values for u and v from (28), we obtain

$$\begin{aligned} Z &= \frac{\partial v}{\partial q} - \frac{\partial u}{\partial p} \\ &= \frac{\partial R}{\partial q} \psi_p + R \cdot \psi_{pq} - \frac{\partial R}{\partial p} \psi_q - R \cdot \psi_{qp} \\ &= \frac{\partial R}{\partial q} \psi_p - \frac{\partial R}{\partial p} \psi_q. \end{aligned} \quad \square$$

Proof of Theorem 2. By assumption, the field Z is continuous. By Green's theorem, the integral of

$$\Lambda = u(q, p) \dot{q} + v(q, p) \dot{p}$$

around any simple closed curve $(q(t), p(t))$ in the anticlockwise direction is equal to the integral of Z over the area enclosed by the curve. If $Z = 0$ along $\mathcal{S}^*$, then there can be an improving deviation to the left only if $Z < 0$ somewhere in that region. Similarly, there can be an improving deviation to the right only if $Z > 0$ somewhere in that region. Thus the conditions guarantee that there are no improving deviations, therefore $\mathcal{S}^*$ is maximal. $\square$

Calculating constrained distribution function

Each congestion state ω has different constraints on the demand shock and supply. The demand shocks satisfying the constraints define the sets $\Gamma_{n,\omega}(S)$, of ε for which the market clears in congestion state ω and the total offers at all nodes at the price $p = p_n$ is the vector S . For node $n = 1$ in the two-node network, we have

$$\begin{aligned} \Gamma_{1,\omega_1}(S) &= \{\varepsilon : \varepsilon_1 + \varepsilon_2 = S_1 + S_2 \wedge -K \leq S_1 - \varepsilon_1 \leq K\} \\ \Gamma_{1,\omega_2}(S) &= \{\varepsilon : \varepsilon_1 = S_1 - K \wedge \varepsilon_2 - S_2 \geq K\} \\ \Gamma_{1,\omega_3}(S) &= \{\varepsilon : \varepsilon_1 = S_1 + K \wedge \varepsilon_2 - S_2 \leq -K\}. \end{aligned}$$

For node $n = 2$ we have

$$\begin{aligned} \Gamma_{2,\omega_1}(S) &= \{\varepsilon : \varepsilon_1 + \varepsilon_2 = S_1 + S_2 \wedge -K \leq S_1 - \varepsilon_1 \leq K\} \\ \Gamma_{2,\omega_2}(S) &= \{\varepsilon : \varepsilon_2 = S_2 + K \wedge \varepsilon_1 - S_1 \leq -K\} \\ \Gamma_{2,\omega_3}(S) &= \{\varepsilon : \varepsilon_2 = S_2 - K \wedge \varepsilon_1 - S_1 \geq K\}. \end{aligned}$$

The sets $\Gamma_{n,\omega}(S)$ are traced out by dashed lines in Figure 26: blue for node 1 and red for node 2. The constraints on ε_2 in states ω_2 and ω_3 ensure consistency with S_2 being offered at the same price as S_1 ; the price must be higher at the upstream end of the transmission constraint.

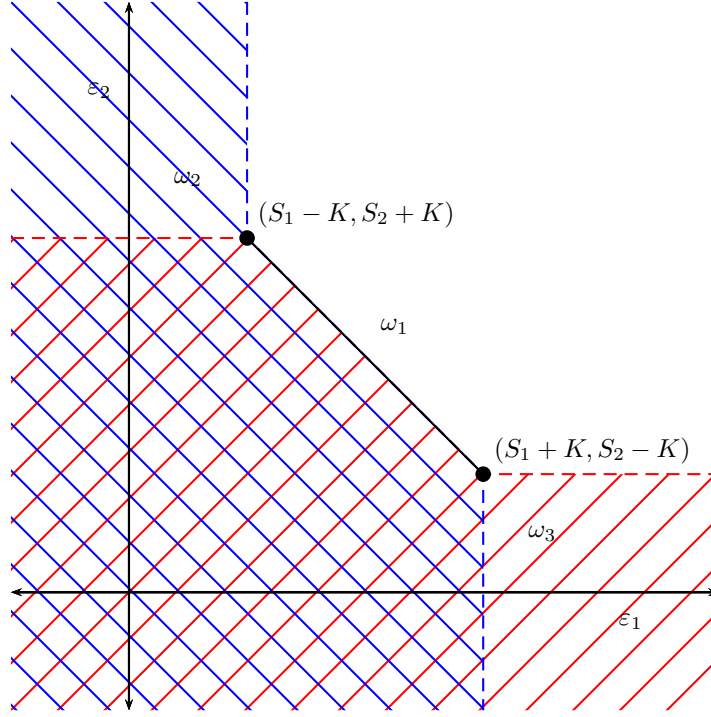


Figure 26: Calculating constrained demand distribution in two-node network

A demand shock ε can be satisfied at n if and only if it is Pareto dominated by an $\tilde{\varepsilon} \in \cup_{\omega} \Gamma_{n,\omega}$, i.e. $\varepsilon_m \leq \tilde{\varepsilon}_m$ for all m .

The constrained demand distributions are

$$\begin{aligned}\phi_1(S_1, S_2) &= \Pr \left[\varepsilon \leq \bigcup_{\omega} \Gamma_{1,\omega}(S_1, S_2) \right] \\ \phi_2(S_1, S_2) &= \Pr \left[\varepsilon \leq \bigcup_{\omega} \Gamma_{2,\omega}(S_1, S_2) \right].\end{aligned}$$

Writing this out explicitly as conditions on ε_1 and ε_2 gives

$$\begin{aligned}\phi_1(S_1, S_2) &= \Pr [(\varepsilon_1 + \varepsilon_2 < S_1 + S_2 \text{ and } \varepsilon_1 < S_1 + K) \text{ or } \varepsilon_1 < S_1 - K] \\ \phi_2(S_1, S_2) &= \Pr [(\varepsilon_1 + \varepsilon_2 < S_1 + S_2 \text{ and } \varepsilon_2 < S_2 + K) \text{ or } \varepsilon_2 < S_2 - K].\end{aligned}$$

The areas in ε -space dominated by the Γ_n are the hatched regions in Figure 26. The probability ϕ_1 is the probability mass that falls on the blue area to the left of the blue dashed line, and ϕ_2 is the probability mass that falls in the red hatched area below the dotted red line.

In the two-node network we calculate the partial derivatives of ϕ_1, ϕ_2 as follows: the boundaries of the event regions in shock space, Γ_1 and Γ_2 , move with S_1 and S_2 , so the partial derivatives of ϕ_1, ϕ_2 are the line integrals of the shock density $f(\varepsilon_1, \varepsilon_2)$ along Γ_1 and Γ_2 . Define $P_n(S, \omega)$ as the density of the shock density on the set $\Gamma_{n,\omega}$:

$$P_n(S, \omega) = \int_{\Gamma_{n,\omega}} df(\varepsilon). \quad (31)$$

The calculation for ϕ_1 is as follows. The three relevant line integrals are

$$\begin{aligned} P_1(S, \omega_1) &= \int_{\Gamma_1, \omega_1} f = \int_{-K}^K f(S_1 - t, S_2 + t) dt \\ P_1(S, \omega_2) &= \int_{\Gamma_1, \omega_2} f = \int_{S_2+K}^{\infty} f(S_1 - K, t) dt \\ P_1(S, \omega_3) &= \int_{\Gamma_1, \omega_3} f = \int_{-\infty}^{S_2-K} f(S_1 + K, t) dt. \end{aligned} \quad (32)$$

In each case $P_1(S, \omega)$ is the joint density function of residual demand S_1 and congestion state ω , given the other node's offer S_2 . The joint densities for node 2 are

$$\begin{aligned} P_2(S, \omega_1) &= \int_{\Gamma_2, \omega_1} f = \int_{-K}^K f(S_1 - t, S_2 + t) dt \\ P_2(S, \omega_2) &= \int_{\Gamma_2, \omega_2} f = \int_{-\infty}^{S_1-K} f(t, S_2 + K) dt \\ P_2(S, \omega_3) &= \int_{\Gamma_2, \omega_3} f = \int_{S_1+K}^{\infty} f(t, S_2 - K) dt. \end{aligned} \quad (33)$$

From these we can calculate the partial derivatives of ϕ_1 . Under a small change in S_1 , both the dashed and solid lines move to the right. So

$$\frac{\partial \phi_1}{\partial S_1} = P_1(S, \omega_1) + P_1(S, \omega_2) + P_1(S, \omega_3).$$

For the cross-derivative (off-diagonal), $\frac{\partial \phi_1}{\partial S_2}$, only the sloping (solid) part of the boundary moves with changing S_2 . It shifts upwards when S_2 increases. So

$$\frac{\partial \phi_1}{\partial S_2} = P_1(S, \omega_1).$$

The partial derivatives of ϕ_2 are similarly

$$\begin{aligned} \frac{\partial \phi_2}{\partial S_2} &= P_2(S, \omega_1) + P_2(S, \omega_2) + P_2(S, \omega_3) \text{ and} \\ \frac{\partial \phi_2}{\partial S_1} &= P_2(S, \omega_1). \end{aligned}$$

With a bivariate normal shock distribution, the line integrals in (32) and (33) are simple to calculate. Each will be the product of a marginal density, which is constant over the whole line, and a conditional probability, which is the integral of the density of a uni-variate normal over the interval. There are three different angles of line, so three different sorts of marginal and conditional distribution:

- For integrands of the form $f(\varepsilon_1, t)$, the marginal distribution has mean μ_1 and standard deviation σ_1 and the conditional has mean $\mu_1 + \frac{\sigma_1}{\sigma_2} \rho(t - \mu_2)$ and standard deviation $\sigma_1 \sqrt{1 - \rho^2}$.
- For integrands of the form $f(t, \varepsilon_2)$, the marginal distribution has mean μ_2 and standard deviation σ_2 and the conditional has mean $\mu_2 + \frac{\sigma_2}{\sigma_1} \rho(t - \mu_1)$ and standard deviation $\sigma_2 \sqrt{1 - \rho^2}$.

- For integrands of the form $f(S_1 - t, S_2 + t)$ the marginal distribution has mean $\mu_1 - \mu_2$ and standard deviation $\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$ and the conditional has mean $\mu_1 + \mu_2$ and standard deviation $\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2$

In the numerical examples we use MATLAB's built-in `normpdf` and `normcdf` functions to compute the line integrals.

Note that since the demand shock density f is everywhere non-negative, all of the $P_m(S, \omega)$ are positive. The partial derivatives $\phi_{m,n}$ are sums over the $P_m(S, \omega)$, so they are all positive. Observe furthermore that the diagonal terms $\phi_{m,m}$ are greater than the off-diagonal terms $\phi_{m,n}$, $m \neq n$. Also, the terms $P_1(S, \omega_1)$ and $P_2(S, \omega_1)$ are equal, so $\phi_{1,2} = \phi_{2,1}$. All this implies that the Jacobian matrix $\frac{\partial \phi}{\partial S}$ has all positive entries and is positive-definite.

Example 32 (Plotting ϕ_m). Suppose the demand shock has a bivariate normal distribution with means $(\mu_1, \mu_2) = (180, 150)$, standard deviations $(\sigma_1, \sigma_2) = (60, 40)$, and correlation coefficient $\rho = 0$. The density of this distribution is plotted in Figure 27. Suppose the transmission line has a capacity of $K = 60$ MW. Then we can use (32) and (33) to calculate the values of $P_m(S, \omega)$ for $m = 1, 2$ and $\omega = \omega_1, \omega_2, \omega_3$, which are shown in Figure 28. Then we can sum up the $P_m(S, \omega)$ by congestion states to obtain the partial derivatives $\phi_{m,n}$ of the constrained demand distribution, shown in Figure 29. Finally, we divide $\phi_{m,n}$ for $m \neq n$ by $\phi_{m,m}$ to obtain the conditional probability of uncongested flow $\hat{P}_m$. The regions of least congested flow occur in different shock realizations depending on where you are located.

It is clear in Figure 28 that the density functions $P_m(S, \omega)$ are the convolution of the normal density function $f(\varepsilon)$, plotted in Figure 27 with the sets $\Gamma_{n,\omega}$ (more precisely, with their indicator functions), plotted in Figure 26.

Proof of Lemma 12. 1: Suppose $Y(Q) = 0$. Then

$$\begin{aligned} \frac{N_2}{N_2 - 1} \phi_{1,2} Q_2 - \phi_{1,1} Q_1 &= 0 \text{ and} \\ \frac{N_1}{N_1 - 1} \phi_{2,1} Q_1 - \phi_{2,2} Q_2 &= 0. \end{aligned}$$

This implies that if one of Q_1, Q_2 is zero, the other is too. Suppose both are non-zero. Then

$$\begin{aligned} \frac{N_2}{N_2 - 1} \frac{\phi_{1,2}}{\phi_{1,1}} Q_2 &= Q_1 = \frac{\phi_{2,2}}{\phi_{2,1}} \frac{N_1 - 1}{N_1} Q_2 \\ \frac{N_2}{N_2 - 1} \frac{\phi_{1,2}}{\phi_{1,1}} &= \frac{\phi_{2,2}}{\phi_{2,1}} \frac{N_1 - 1}{N_1} \\ (N_1 - 1)(N_2 - 1) - N_1 N_2 \frac{\phi_{1,2}}{\phi_{1,1}} \frac{\phi_{2,1}}{\phi_{2,2}} &= 0, \end{aligned}$$

but the left-hand side of the last line is just the determinant $|\mathbf{A}|$.

2: Since all the $\phi_{m,n}$, $\frac{N_1}{N_1 - 1}$, and $\frac{N_2}{N_2 - 1}$ are positive, this follows immediately from the definition of Y_m . $\square$

Proof of Lemma 13. Write out $U(Q)$ as

$$\begin{aligned} U(Q) &= \frac{1}{|\mathbf{A}|} \begin{bmatrix} (N_2 - 1) & -N_2 \frac{\phi_{1,2}}{\phi_{1,1}} \\ -N_1 \frac{\phi_{2,1}}{\phi_{2,2}} & (N_1 - 1) \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} \\ |\mathbf{A}| \cdot U(Q) &= \begin{bmatrix} -\frac{N_2 - 1}{\phi_{1,1}} Y_1(Q) \\ -\frac{N_1 - 1}{\phi_{2,2}} Y_2(Q) \end{bmatrix}. \end{aligned}$$

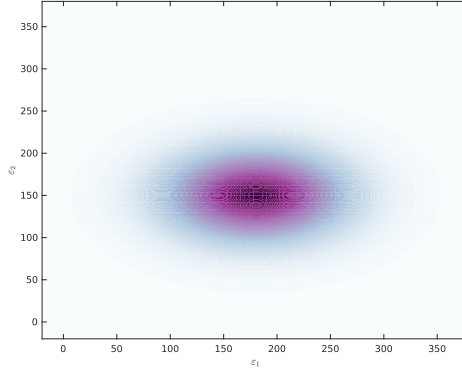


Figure 27: The demand shock density.

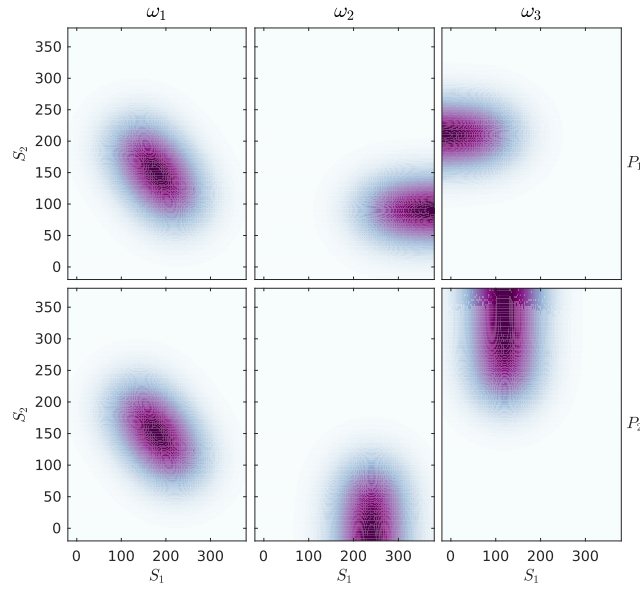


Figure 28: The $P_m(S, \omega)$ from the shock density and a line capacity of $K = 60\text{MW}$. Regions of concentrated density in congestion states ω_2 and ω_3 continue off the edges of these plots out to infinity.

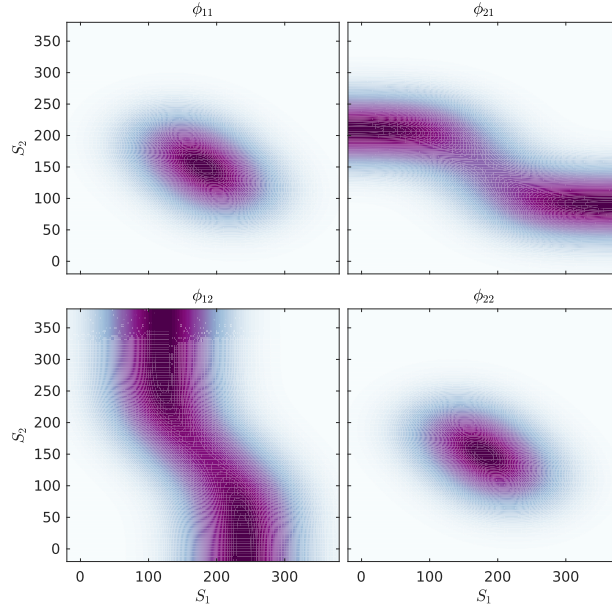


Figure 29: The $\phi_{m,n}$ matrix, found by summing up the $P_m(S, \omega)$ shock density.

Thus it is clear that $U > 0$ if and only if both Y_1, Y_2 have the opposite sign to $|\mathbf{A}|$. $\square$

Proof of Proposition 7. This proposition is similar to Proposition 3 of Klemperer and Meyer (1989), but because we have $C'' = 0$ we must alter the argument to allow for $q'_{i,m}(c) = \infty$.

Consider two generators, labeled 1 and 2, located at node 1. The node 2 generators offer supply functions $q_{i,2}(p)$ whose total is $\sum_i q_{i,2}(p) = S_2(p)$. The other node 1 generators offer supply functions whose total is $\hat{S}_1(p)$; i.e.

$$S_1(p) = q_{1,1}(p) + q_{2,1}(p) + \hat{S}_1(p). \quad (34)$$

The first-order optimality conditions for the two agents are

$$\begin{aligned} Z_{1,1} &= (p - c) \left(\phi_{1,1} (q'_{2,1} + \hat{S}'_1) + \phi_{1,2} S'_2 \right) - \phi_{1,1} q_{1,1} = 0 \\ Z_{2,1} &= (p - c) \left(\phi_{1,1} (q'_{1,1} + \hat{S}'_1) + \phi_{1,2} S'_2 \right) - \phi_{1,1} q_{2,1} = 0. \end{aligned}$$

The slopes of the two generators' supply functions are therefore

$$\begin{aligned} q'_{2,1} &= \frac{q_{1,1}}{p - c} - \frac{\phi_{1,2}}{\phi_{1,1}} S'_2 - \hat{S}'_1 \\ q'_{1,1} &= \frac{q_{2,1}}{p - c} - \frac{\phi_{1,2}}{\phi_{1,1}} S'_2 - \hat{S}'_1. \end{aligned} \quad (35)$$

We treat the $S_{-\{1,2\}}(p)$ and $S_2(p)$ as fixed functions, so this is a second-order system of differential equations in $q_{1,1}$ and $q_{2,2}$. The common residual demand faced by the two agents is

$$\frac{\phi_{1,2}}{\phi_{1,1}} S'_2 - \hat{S}'_1.$$

Consider $G = q_{2,1} - q_{1,1}$. We can use (35) to write an ODE for $G(p)$, in which this common residual demand cancels out;

$$\begin{aligned} G' &= q'_{2,1} - q'_{1,1} \\ &= \frac{q_{1,1} - q_{2,1}}{p - c} \\ &= -\frac{G}{p - c}. \end{aligned}$$

Solutions to this ODE are

$$G = \frac{a}{p - c},$$

but these diverge at c unless $a = 0$, in which case $G \equiv 0$. Therefore the only solutions to (35) that have $\lim_{p \rightarrow c} q_{1,1}(p) = \lim_{p \rightarrow c} q_{2,1}(p) = 0$ must have $q_{1,1} \equiv q_{2,1}$. $\square$

Proof of Proposition 11. Suppose that $k < N_1$ agents in node 1 begin to iron at $\hat{t}$: i.e. their supply curves are strictly monotone (and equal to all the other node 1 supply curves) on $(-\infty, \hat{t})$, and are constant on $(\hat{t}, t^*)$. Over the interval $(\hat{t}, t^*)$, we have $N_1 - k$ agents offering the continuation $Q_1(t)$, and k offering $r(t) = Q_1(\hat{t})$. The $N_1 - k$ agents who are not ironing will each offer a curve Q_1 which satisfies

$$Z_{1,1} = \left(\phi_{1,1}(N_1 - k - 1) \dot{Q}_1 + \phi_{1,2} \dot{S}_2 \right) - \phi_{1,1} Q_1 = 0.$$

Meanwhile, the k agents that iron will each have adjoint w_r that satisfies

$$\begin{aligned} \dot{w}_r &= \left(\phi_{1,1}(N_1 - k) \dot{Q}_1 + \phi_{1,2} \dot{S}_2 \right) - \phi_{1,1} r \\ w_r(\hat{t}) &= 0. \end{aligned}$$

Now $\dot{Q}_1 \geq 0$, so $Q_1 > r$ on $(\hat{t}, t^*)$, and $N_1 - k - 1 < N_1 - k$. Hence $\dot{w}_r > Z_{1,1} = 0$ over $(\hat{t}, t^*)$.

Even if the Q_1 agents start ironing, we still have $Z_{r,1} > Z_{1,1} = Y_1 \geq 0$, so $w_r > w_1 \geq 0$ on all of $(\hat{t}, t^*)$. In particular $w_r(t^*) > 0$, so by Corollary 5, the ironed segment is only admissible if the node 1 agents are at their (common) production capacity. But if this is the case then all node 1 agents must necessarily iron. $\square$

Proof of Theorem 16. The first part of the proof is by cases on the sign of $|\mathbf{A}|$. In each case we show that the constant-coefficient algorithm will compute a unique candidate trajectory. The second part shows that this candidate trajectory gives a unique SFE.

Case $|\mathbf{A}| < 0$

[Existence] If $\bar{Q}$ lies on the ray through the positive eigenvector of $\mathbf{A}$, then the algorithm will connect it to the origin by a straight line trajectory that has $w = 0$ and $Q' \geq 0$ everywhere. This is a candidate trajectory.

If $\bar{Q}$ lies off that ray, then a second segment is added. At the inner end of this segment we have $w = 0$. For node m which is ironing we have $Y_m > 0$ and Q_n increasing, $n \neq m$. Since Y_m is the product of the a positive function $\phi_{m,m}$ and the function

$$\frac{N_n}{N_n - 1} \frac{\phi_{m,n}}{\phi_{m,m}} Q_n - Q_m$$

which is increasing in Q_n , we will have $Y_m > 0$ along the whole ironed segment. Hence $w_m \geq 0$ along the whole ironed segment too. Since $Z_n = 0$ when m is ironing, we have $w_n = 0$ along the whole ironed segment. Thus S and w form a candidate trajectory.

[Uniqueness] By a similar argument to how Klemperer and Meyer (1989) show that the slope of all pool SFE through the origin is the same, it is clear that there is only one monotone and un-ironed trajectory through the origin.

A candidate trajectory can only start or stop ironing where $w = 0$. If we start ironing at any other point than where one of the production capacity constraints begins to bind, then w will never reach 0 to allow the trajectory to turn a corner to iron in a different direction, or to integrate along $\dot{Q} = U(Q)$. Suppose that on a candidate trajectory, agents at some node m start ironing from some point $\hat{Q}$ on the positive eigenvector ray, where $\hat{Q} < \bar{Q}$. When node m stops ironing, we will have $\dot{Q}_m > 0$. The complementarity condition on the candidate trajectory requires that at this corner point, $w_m = 0$. However, as argued in the existence part of this proof, if node m starts ironing then w_m is non-negative and strictly increasing. It cannot reach zero, and so there cannot be a second corner. Thus the two-part trajectory consisting of the positive eigenvector ray and the ironed segment from where the first production capacity constraint binds is the unique candidate trajectory.

Case $|A| > 0$

[Existence] The equation $\dot{Q} = A^{-1}Q$ is a linear ODE, so solutions exist. If there is only one segment of the trajectory the algorithm will find this solution. The loci $Y_1 = 0$ and $Y_2 = 0$ are fences that form a cone-shaped anti-funnel with its point at the origin. So from any point Q with $U(Q) \geq 0$, there is a curve connecting Q to the origin along the direction field $\dot{Q} = U(Q)$. Moreover this curve is monotone.

If there are two segments, then $w = 0$, $Q' > 0$ is clearly true on the inner part. On the ironed part, the ironing node m has $w_m(t) \geq 0$, since $w = 0$ at the inner end of the ironed segment and $Y_m \geq 0$ along the ironed segment. The non-ironing node n has $w_n = 0$ everywhere. Hence the Q and w form a candidate trajectory.

[Uniqueness] The direction field $\dot{Q} = U(Q)$ is Lipschitz-continuous, so the non-ironed segment is the unique trajectory connecting its end points (by the Picard-Lindelöf Theorem).

To rule out multiple ironed segments, look to the corner conditions. By Lemma 14, the only points where node m can begin to iron are where $Y_m = 0$. Moreover, as $Y_m > 0$ at every point further out along the ironed trajectory, $w > 0$ no matter how far we iron. In other words, once agents start ironing, they will never be able to stop and turn a corner. So there is at most one ironed segment in a candidate trajectory. The algorithm finds the only way to iron from a production capacity point $\bar{Q}$ with $U_m(\bar{Q}) < 0$ to the feasible cone $U \geq 0$. Therefore the trajectory produced by the algorithm is unique.

[Trajectory is an SFE] We now show that the supply function for each agent is a maximum of their expected profit. An agent in node 1 maximizes Π^U . Their first-variation derivative is

$$Z_{i,1} = \phi_{1,1} \left((p - c) \left((N_1 - 1) Q'_1 + N_2 \frac{\phi_{1,2}}{\phi_{1,1}} Q'_2 \right) - q_{i,1} \right).$$

This is the product of $\phi_{1,1}$, which is everywhere positive, and

$$\hat{Z}_{i,1} = (p - c) \left((N_1 - 1) Q'_1 + N_2 \frac{\phi_{1,2}}{\phi_{1,1}} Q'_2 \right) - q_{i,1},$$

which has $\frac{\partial \hat{Z}}{\partial q_{i,1}} = -1 < 0$ everywhere (by assumption, $\frac{\partial}{\partial S_1} \frac{\phi_{1,2}}{\phi_{1,1}} = 0$). Therefore the curve $(Q_1(t), p(t))$ and the first-variation derivative $Z_{i,1}$ satisfy the conditions of Theorem 2, and the curve is a maximum for Π^U . A similar argument holds for any agent at node 2. Therefore the set of supply curves is an SFE. As $A(\bar{Q}) \neq \mathbf{0}$, there must be some probability of lost load, and so by Holmberg's (2008) argument, the SFE is unique. $\square$

Proof of Proposition 20. The proof of existence is by showing that the Inward Algorithm 2 will reach the origin from any given production capacity point $\bar{Q}$. The uniqueness proof is by partial trajectories; we show that the extension in each round of the main loop is unique.

Existence. On initial ironing it cannot be that both Y_1 and Y_2 are positive, as then we would be in the $|\mathbf{A}| < 0$ case, by Lemma 12. Whenever node m is ironing inwards, the ironed segment meets $Y_m = 0$ before Q_n reaches zero because of Lemma 12. After the initial ironing, we will have $Y(Q(\hat{t})) \leq 0$, by assumption 19. Thus, by Lemma 13, we find that $U(Q(\hat{t})) \geq 0$.

The same lemma also ensures that the integration along $\dot{Q} = U(Q)$ in the main loop will either reach the origin (case C) or cross $Y_m = 0$ (case D). For case C there is nothing to show. For case D, we must show that a solution to the interval search problem exists. Ironing starts from a point t^* with $Y \leq 0$, $U \geq 0$, and $w = 0$ and ends at t_* with either $Q_n(t_*) = 0$ or $w(t_*) = 0$. Consider the inner endpoint t_* as function of the outer endpoint t^* . As the initial ironing point t^* approaches the end of the partial trajectory $\hat{t}$, $t_*(t^*)$ approaches $\hat{t}$ too, since $Y_m < 0$ immediately above $\hat{t}$ and $Y_m > 0$ immediately below $\hat{t}$ (by the continuity of Y_m and part 1 of Assumption 19). As the initial ironing point t^* moves outwards, the inner endpoint $t_*(t^*)$ decreases until eventually it meets $Y_m = 0$.

Uniqueness. Suppose that the line searches in cases D have unique solutions. Then every time the algorithm stops integrating along its current segment, is because it cannot continue without violating one of the optimality conditions (22). Furthermore, in every place it does start or stop ironing, it is the unique point from which the ironed segment will reach either the production capacity point $\bar{Q}$ or a point with $U(Q) > 0$ and $w = 0$. Therefore the candidate trajectory that it finds is unique.

It remains only to verify that the trajectory conditions (22) are met by every segment of the trajectory. On the un-ironed segments we set $\dot{w} = Z(Q, U(Q)) = 0$, so $w = 0$ everywhere there is no ironing. On the ironed segments, if node m is ironing then $\dot{w}_n = Z_n(Q, \dot{Q}) = 0$, so $w_n = 0$ along the entire segment. Since $\dot{Q}_n = (N_n - 1)Q_n > 0$, the adjoint conditions (22) are satisfied in the n components. For node m , we have $\dot{w}_m = Y_m(Q) = Z_m(Q, \dot{Q})$. Since we always stop ironing at or before w_m reaches zero, we end up with $w_m \geq 0$ along the entire segment. Of course $\dot{Q}_m = 0$, so (22) is satisfied in the m component. $\square$

Proof of Theorem 21. The proof of existence is by showing that the Outward Algorithm 4 can reach any given production capacity point $\bar{Q}$. The uniqueness proof is by partial trajectories; we show that the extension in each round of the main loop is unique.

By the Perron-Frobenius theorem, any matrix with all positive entries has an eigenvector with all positive values. Thus the eigenvector $\mathbf{v}_1$ that initializes the algorithm exists.

We need to show that each case terminates in one of the specified conditions.

- A) We start ironing with $Q_m = \bar{Q}_m$, $w = 0$, and $Y > 0$. The quantity offered by node n , Q_n is increasing, so will eventually surpass $\bar{Q}_n$, unless it reaches $w_m = 0$ first.
- B) Ironing starts from a point t_* with $Y \geq 0$, $U \geq 0$, and $w = 0$, and ends at t^* with either $w_m(t^*) = 0$ or $Q_n(t^*) = \bar{Q}_n$. Consider the outer endpoint t^* as a function of the inner endpoint t_* . As the initial ironing point t_* approaches $\hat{t}$, this $t^*(t_*)$ approaches $\hat{t}$ too, since $Y_m > 0$ immediately below $\hat{t}$ and $Y_m < 0$ immediately above $\hat{t}$ (by the continuity of Y_m and part 1 of Assumption 19). As the initial ironing point t_* moves inwards, the outer endpoint $t^*(t_*)$ increases until eventually it either meets $Y_m(Q(t^*)) = 0$ as it curves back up (where it is an outward fence) or meets $Q_n = \bar{Q}_n$.

- C) Nothing to show.
- D) Very similar to case B, except here ironing out from $\hat{t}$ gives an outer endpoint $t^* (\hat{t})$ with $w_m (t^*) = 0$ and $Y_m (Q (t^*)) < 0$.
- E) Nothing to show.
- F) Part 2 of Assumption 19 gives $w_n > 0$ along this entire ironed segment. Node n ironing means that Q_m increases, eventually it will reach $\bar{Q}_m$.

Existence. Assumption 19 implies that the algorithm will eventually terminate. It is clear that after each round, Q increases. After each round there are two changes in sign of one of the Y_m . Thus after each round of the main loop, the rectangle $[0, Q_1 (\hat{t})] \times [0, Q_2 (\hat{t})]$ contains at least one additional swerve of the $Y_m = 0$.

It remains only to verify that the trajectory conditions (22) are met by every segment of the trajectory. On the un-ironed segments we set $\dot{w} = Z (Q, U (Q)) = 0$, so $w = 0$ everywhere there is no ironing. On the ironed segments, if node m is ironing then $\dot{w}_n = Z_n (Q, \dot{Q}) = 0$, so $w_n = 0$ along the entire segment. Since $\dot{Q}_n = (N_n - 1) Q_n > 0$, (22) is satisfied in the n components. For node m , we have $\dot{w}_m = Y_m (Q) = Z_m (Q, \dot{Q})$. Since we always stop ironing at or before w_m reaches zero, we end up with $w_m \geq 0$ along the entire segment. Of course $\dot{Q}_m = 0$, so (22) is satisfied in the m components.

Uniqueness. Suppose that the line searches in Cases B and D have unique solutions. Then every time the algorithm stops integrating along its current segment, it is because it cannot continue without violating one of the optimality conditions (22). Furthermore, in every place it does start or stop ironing, it is the unique point from which the ironed segment will reach either the production capacity point $\bar{Q}$ or a point with $U (Q) > 0$ and $w = 0$. Therefore the candidate trajectory that it finds is unique.

The trajectory conditions (22) are met by every segment of the trajectory, as for the Inward Algorithm. $\square$