

A Nonlinear Complementarity Approach for the National Energy Modeling System

Steven A. Gabriel

Mathematics and Computer Science Division

Argonne National Laboratory

9700 South Cass Avenue

Argonne, Illinois 60439

gabriel@mcs.anl.gov

Andy S. Kydes

Energy Information Administration

U.S. Department of Energy

1000 Independence Avenue, SW

Washington, D.C. 20585

andy.kydes@hq.doe.gov

March 10, 1995

Preprint MCS-P504-0395

Mathematics and Computer Science Division

Argonne National Laboratory

Abstract

The National Energy Modeling System (NEMS) is a large-scale mathematical model that computes equilibrium fuel prices and quantities in the U.S. energy sector. At present, to generate these equilibrium values, NEMS sequentially solves a collection of linear programs and nonlinear equations. The NEMS solution procedure then incorporates the solutions of these linear programs and nonlinear equations in a nonlinear Gauss-Seidel approach.

We describe how the current version of NEMS can be formulated as a particular nonlinear complementarity problem (NCP), thereby possibly avoiding current convergence problems. In addition, we show that the NCP format is equally valid for a more general form of NEMS. We also describe several promising approaches for solving the NCP form of NEMS based on recent Newton type methods for general NCPs. These approaches share the feature of needing to solve their direction-finding subproblems only approximately. Hence, they can effectively exploit the sparsity inherent in the NEMS NCP.

1 Introduction

As a result of the oil embargo of the early 1970s, the U.S. energy community began extensive mathematical modeling to analyze various energy issues and develop a national energy policy. In what follows, we provide a brief overview of three of these energy models that are relevant to this paper. The interested reader can consult Energy Information Administration (1994), and Gabriel (1993) for a more complete history.

One of the prominent energy models developed in the 1970s was the Project Independence Evaluation System (PIES). The goal of this large-scale energy system was to calculate an equilibrium in prices and quantities of fuels in the U.S. energy sector. Equilibrium prices and quantities were values that satisfied the constraints of both the supply and the demand sides of the market. In essence, PIES was a combination of linear programming and econometric demand equations used to determine valid prices and quantities of fuels; see Hogan (1975), and Ahn and Hogan (1982). In his Ph.D. dissertation, Ahn (1979) rigorously analyzed convergence properties of the specialized PIES algorithm and established a connection with the well-known nonlinear Jacobi method for solving a system of nonlinear equations.

Around 1980, the U.S. Congress expressed interest in focusing on near-term energy policy issues. The large PIES model proved to be unsuitable for the resulting year-by-year analyses and so a new energy model, the Intermediate Future Forecasting System (IFFS) was developed (Murphy et al. 1988).

IFFS also sought to compute an equilibrium in fuel prices and quantities, but used a different approach from PIES. IFFS was modular, where each module related to a specific energy activity such as electricity generation or natural gas distribution. Each module computed a trial equilibrium in fuel prices and quantities, holding values outside of the module constant, and then passing these trial equilibrium values on to the next module. This modular structure proved to be more amenable to energy analysis than the PIES approach; see Murphy (1983), Murphy (1993), and Murphy et al. (1988). The solution algorithm used in IFFS was the nonlinear Gauss-Seidel method for solving a system of nonlinear equations related to an energy equilibrium.

About 1990, it was determined that IFFS alone was not sufficient to handle many of the emerging energy issues such as the Clean Air Act Amendments and natural gas deregulation. Consequently, the National Energy Modeling System (NEMS) was initiated. Like IFFS, NEMS is highly modular and uses a nonlinear Gauss-Seidel approach to compute an equilibrium. However, NEMS went well beyond IFFS in several ways. For example, certain market components such as international oil markets, demand side management, and interregional electricity trade and transmission that had been calculated exogenously are now calculated within the NEMS system. Also, more structural detail was added to certain energy sectors.

At present, some convergence problems have been identified with the Gauss-Seidel NEMS algorithm. One of the main problems is that this approach “freezes” the inputs from outside modules while calculating a trial equilibrium within a specific module. This can cause problems for those modules that interact heavily with other ones; for example, the Electricity Market Module interacts heavily with the Natural Gas Transmission and Distribution Module and the Coal Market Module. In addition, there have been convergence problems related to approximating supply and/or demand curves by step functions. These step functions are meant to approximate a nonlinear relationship that cannot be explicitly incorporated as such into the relevant LP modules.

While some of these problems have been fixed, NEMS modelers and other researchers wish to develop a generalized version of NEMS to avoid such obstacles and perhaps improve the speed of convergence of the current approach. One such strategy for achieving these goals is to model the NEMS equilibrium as a nonlinear complementarity problem (NCP). This more general format not only has the potential to reduce some of the convergence problems cited above, but it allows for much more general NEMS modules than at present. This last point is crucial because there is interest in using nonlinear programs or even nonoptimization-based equilibrium models in some of the NEMS modules. An example of a nonoptimization model is the spatial price equilibrium problem; see Harker and Pang (1990) for details. Note that these new modeling directions would be less restrictive than linear programs, potentially allowing for a more realistic model of energy activity. Hence, we see that for various reasons, the NCP format has merit for NEMS.

The organization of the rest of this paper is as follows. In section 2, we give a brief overview of the current NEMS setup; in section 3 we describe the general nonlinear com-

plementarity problem and the specific NCP that arises from NEMS; and in sections 4–6 we describe how several recent iterative Newton type methods for the general NCP can be specialized to efficiently solving the NEMS NCP.

Note that throughout this paper, for vectors $v \in R^n$, we have indicated subvectors by v_y . Here y is a vector of variables so that v_y refers to all components of v relating to these variables. Alternatively, we have also used the index set $\gamma \subseteq \{1, 2, \dots, n\}$ to describe a subvector v_γ of v . Unless otherwise stated, for vectors, superscripts will denote iterates such as y^k , whereas for matrices or scalars subscripts will denote a component, (i.e., y_i). Lastly, unless stated otherwise $\|\cdot\|$ is meant to denote the usual Euclidean norm.

2 A Brief Overview of NEMS

Like its predecessor models, e.g. the Intermediate Future Forecasting System (IFFS), NEMS incorporates a market-based approach to energy analysis. NEMS balances the supply of and demand for energy for each fuel and consuming sector, taking into account the economic competition between energy sources.

NEMS is partitioned into a modular system, which is solved by applying the Gauss-Seidel convergence method with successive over-relaxation. The modules of NEMS represent each of the fuel supply markets, conversion sectors, and end-use consumption sectors, and also include interactive macroeconomic and international modules. The primary flows between these modules are the delivered prices of energy and the quantities consumed by product, region, and sector, but include other information such as economic activity and technology characteristics. The delivered prices of fuel encompasses all the activities necessary to produce, import, and transport fuels to the end user.

At present, NEMS consists of an integrating module as well as the following other modules:

Energy Demand

1. Residential Demand Module
2. Commercial Demand Module
3. Transportation Demand Module
4. Industrial Demand Module

Energy Supply

1. Oil and Gas Supply Module
2. Renewable Fuels Module
3. Natural Gas Transmission and Distribution Module
4. Coal Market Module

Energy Conversion

1. Electricity Market Module
2. Petroleum Market Module

In addition, there are two other modules for modeling economic activities: (1) The Macroeconomic Activity Module and (2) the International Energy Module. At present, the conversion, transmission, and distribution of energy are modeled by using appropriate linear programs (LPs). The presumption is that LPs adequately capture those selected aspects of the energy sector. Also, various prices and quantities are calculated as a function of the output from these LPs; by output we mean optimal decision variables and multipliers. In addition, certain prices and quantities serve as inputs to these linear programs. Lastly, any remaining quantities not calculated from the output of these LPs are generated via nonlinear demand equations. All together, NEMS is a collection of linear programs and nonlinear equations whose simultaneous solution determines equilibrium prices and quantities; we will comment more on the specific nature of these modules in what follows.

It is important to understand that this equilibration process is carried out annually up to the year 2010. The NCP formulation for NEMS to be presented below should be interpreted for an individual year in this series. Hence, for each year, we have a different but related NCP to solve.

In NEMS, we are concerned with calculating fuel prices and quantities in equilibrium between the supply and demand sides of the energy market. We suppose that there are n prices and n quantities of fuels denoted, respectively, by the vectors p and q , where $p = (p_1, \dots, p_n)^T$ and $q = (q_1, \dots, q_n)^T$.

In many instances, we will need to distinguish when a certain variable is being used as an input or an output to a particular NEMS module. Given a vector $y \in R^n$, we will denote a subvector as $y_S = \{y_j : j \in S\}$ where $S \subseteq \{1, \dots, n\}$. When the index set S refers to the variables in the vector y that are used as inputs to the i th NEMS mathematical program (currently an LP), we will use the notation $Iy(i)$; for outputs. The associated set of indices will be designated $Oy(i)$.

When it is appropriate, we will abbreviate this notation for convenience. Hence, $p_{Iy(i)}, q_{Iy(i)}$ are those prices and quantities, respectively, used as inputs to the i th mathematical program, and are thus constants in that module.¹ On the other hand, $p_{Oy(i)}, q_{Oy(i)}$ are, respectively, those prices and quantities calculated as a function of the output of the i th LP. Note that the effect of allowing both prices and quantities to feed back into each LP is one of the more advanced yet computationally complicating features of NEMS.

Each module does not always work at the same level of regional aggregation for the variables involved. For example, one module may work with census divisions, whereas another module may use a completely different regional level. The translation between aggregation schemes is important when we deal with q , the fuel quantities demanded.

Each fuel quantity is calculated in exactly one of four places: in the demand modules, the electricity module, the natural gas module, or the petroleum module. We will partition the vector q into four pieces as follows:

- q_D quantities calculated in the demand modules,
- q_E quantities calculated in the electricity module,
- q_G quantities calculated in the natural gas module,
- q_P quantities calculated in the petroleum module.

For example, the quantities computed from the demand modules are at the level of nine census divisions, whereas for the vectors q_E, q_G and q_P , the level of aggregation is respectively, thirteen North American Electric Reliability Council (NERC) regions, twenty-one supply regions and five Petroleum Administration for Defense Districts (PADDs). Hence, we need to translate between these various aggregation schemes when going between modules.

2.1 Demand Modules

NEMS has four demand modules covering residential, commercial, transportation, and industrial demand for various fuels. These modules involve complex sets of equations relating various economic factors as well as fuel prices to determining fuel quantities to be demanded.

While it is not practical to enumerate each of the defining equations involved in computing demand, from empirical testing it has been determined that these demand functions possess some interesting properties, which we will now explain.

First, let $Q : R_+^m \rightarrow R_+^m$ denote the demand function for all four of the demand modules taken together; here $m = |D|$ is the number of quantities calculated in the demand modules and p_D is the associated subvector of prices.

In general, we will focus only on the equilibrium prices p_D as arguments to this function, since other quantities can be ignored from the point of view of the equilibrium problem. From empirical testing, it was determined that the own price effect on demand dominated the cross price effect. In addition, the price effects were symmetric, and at most six prices (including own price) were involved in determining the demand for a particular fuel; in many cases it was just the own price. If we consider the Jacobian of $-Q$, we see the following structure:

$$\nabla(-Q(p_D)) = \begin{pmatrix} D_{11} & 0 & \dots & 0 \\ 0 & D_{22} & \dots & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \dots & D_{nn} \end{pmatrix}, \quad (1)$$

where D_{ii} is a symmetric matrix of size at most 6×6 . The dominance of the own price over cross prices means that this matrix is strictly diagonally dominant and hence nonsingular. However, we can say even more about $\nabla(-Q)$.

The quantities $\frac{\partial(-Q_i(p_D))}{\partial p_i}$ were observed to be strictly positive so that the diagonals of $\nabla(-Q)$ are strictly positive. If the own price dominates the cross prices sufficiently, then the matrix D_{ii} has positive diagonals with off-diagonals sufficiently small. We first note that the eigenvalues of $\nabla(-Q)$ are just the union of the eigenvalues of each D_{ii} . If there were no cross price effects, the matrix D_{ii} would be a diagonal positive definite matrix with all the eigenvalues positive. Since the eigenvalues are continuous functions of the entries of the matrix, for reasonably small cross price effects, one can say that it is reasonable that the eigenvalues of each D_{ii} would be strictly positive. This, of course, results in the matrix $\nabla(-Q)$ being symmetric positive definite so that $-Q$ is a strictly monotone function.

The upshot is that if one were to attempt to consolidate the demand modules, using a strictly monotone function for $-Q$ would be a reasonable place to start. This conclusion is relevant because in the NCP algorithms to be presented, we need to calculate $\nabla(-Q)$.

2.2 Supply Modules

2.2.1 Oil and Gas Supply Module, Renewables Supply Module

The Oil and Gas Supply Module's purpose is to produce a supply function for oil and gas that is used in other modules. That is, having last year's fuel prices and production quantities, this module produces appropriate supply curves. In general, log-linear functions are used to approximate the supply relationship. Specifically, for a particular fuel i , the following model is used:

$$p_i = \hat{p}_i \left(\frac{q_i}{\hat{q}_i} \right)^\alpha,$$

where q_i , p_i are the quantity and price for fuel i in the current year, \hat{q}_i, \hat{p}_i are last year's reference values, and α is the own price elasticity. If natural logs of both sides are taken, one ends up with the form

$$\log\left(\frac{p_i}{\hat{p}_i}\right) = \alpha \log\left(\frac{q_i}{\hat{q}_i}\right),$$

hence the name "log-linear." These relationships are then approximated in a step-function manner and incorporated in the various linear programs used in other modules. The supply module for the renewable fuels also operates in this manner, namely, using a log-linear function for supply, then approximating it by a step function for use in other modules. Consequently, the effect of both the Oil and Gas and the Renewables Modules is made in the objective functions of the various linear programming formulations where the costs of the fuels in question are used.

2.2.2 The Natural Gas Transmission and Distribution Module

The purpose of the Natural Gas Transmission and Distribution Module (NGTDM) is to model the network of pipelines and storage facilities that link suppliers (including importers) and consumers of natural gas. At present, a linear programming formulation is used.²

The following linear program will be used to model the activities of NGTDM as well as other relevant modules, the only differences being the dimension of the constraint matrices, the number of variables, and the specific form of the objective function. We have

$$\text{minimize}_x \{ \theta(x, p) : Ax \geq \tilde{q}, Bx \geq 0, x \geq 0 \} \quad (2)$$

where the first set of constraints is associated with demand quantities q and the second set of constraints is nondemand related. In light of our earlier comment concerning q , we see that \tilde{q} must be at the level of the NGTDM demand regions to be compatible with the other NGTDM values. Hence, we see that $\tilde{q} = Nq$ where the matrix N converts fuel quantities to the NGTDM level of regionality. The other LPs will have a similar translation whose particular form will, of course, depend on the level of aggregation.³ Note that the variables x represent the decision variables for this LP. We will let u and v be the multiplier vectors for these two sets of constraints, respectively. The objective function $\theta(x, p)$ is the sum of supply costs, pipeline tariffs for local distribution companies (LDCs) using the network, storage charges and distribution charges initiated by the LDCs. In this way, the objective

function takes on the following form:

$$\theta(x, p) = \sum_{j \in I_p^c} x_j c_j + \sum_{j \in I_p} x_j p_j, \quad (3)$$

that is, costs independent of the prices of other fuels plus costs using these prices. Note that the index set I_p is understood to be for the NGTDM module; only when it is unclear from the context will the module index i be added as in $I_p(i)$ for the i th submodule.

We see that prices (or supply costs) enter into the objective function and demands enter as right-hand side constraint values; this is one of two possibilities for the other LP-based modules. The other is that just the quantities are used as right-hand sides without any supply costs in the objective function.

In the current version of NEMS, the natural gas prices computed in this module are average prices from the firm markets.⁴ In particular, having the vector of demand multipliers u , we see that the fuel prices computed in this module p^{ngtdm} are calculated as follows:

$$p^{ngtdm} = D(p)Cu,$$

where $D(p)$ is a diagonal matrix whose diagonals are positive and C is a matrix representing the average pricing process (as applied to multipliers). The effect of the diagonal matrix is to scale up or down the average prices based on relative prices of certain fuels. Note that no equilibrium quantities are calculated in this module.

2.2.3 The Coal Module

The Coal Market Module (CMM) represents the mining, transportation, and pricing of coal subject to end-use demand for coal differentiated by physical characteristics such as heat, sulfur, and ash content. The CMM also determines U.S. coal exports as a part of the worldwide market for coal trade.

A linear programming formulation is used to model the activities in the coal market. The objective function does not include prices of competing fuels, as was the case in NGTDM. Consequently, the form of the objective function is

$$\theta(x, p) = \sum_j x_j c_j.$$

Using the LP notation from NGTDM, we see that the coal prices are just the demand multipliers,

$$p^{coal} = Iu,$$

where, of course, the vector u is now specific to the coal LP (and similarly for the other LPs to follow). Additionally, we note that the comment about aggregating demand quantities is also valid here. Lastly, we note that no equilibrium quantities are output from this module.

2.3 The Conversion Modules

2.3.1 The Electricity Market Module

The Electricity Market Module (EMM) is concerned with the generation, transmission, and pricing of electricity subject to delivered prices for various other fuels. At present,

a linear programming formulation as well as an optimization heuristic is used.⁵ The objective function for the LP is of the form (3), since the prices of the various fuels used in the generation of electricity (coal, natural gas, etc.) need to be taken into account when generating electricity. In the EMM, the dual values are used for market penetration and a separate pricing module is used to allocate costs so that total costs are recovered as currently determined in rate case proceedings. Consequently, the price of electricity is calculated as

$$p^{elec} = f(u^{elec}),$$

where $f(\cdot)$ is a function representing the cost recovery calculations. The output quantities (namely, how much of the various fuels is used to generate electricity) are calculated from the optimal solution as

$$q = Rx,$$

where the matrix R reflects the appropriate aggregation levels discussed above.

2.3.2 The Petroleum Market Module

The Petroleum Market Module (PMM) models the refining activities of the energy sector for which a linear programming formulation is used. Since the prices for natural gas, coal, oil, and electricity are needed in the refining process, the objective function for PMM takes on the form of (3). The delivered prices for petroleum products are determined from the multiplier vector u via the affine transformation

$$p^{petro} = Su + b,$$

where S is a matrix representing the effects of regional sharing and b is a vector of tariffs. The output quantities are a function of the vector x via

$$q = Tx,$$

where T takes into account the aggregation from five PADDs to nine census regions.

Note that for a particular year, the International Energy Module has fixed supply curves and thus need not be considered in the computation of equilibrium p and q . Also, the effects of the Macroeconomic Activity Module have been accounted for in the discussion of the demand modules.

As will be shown, the collection of math programs and nonlinear equations that comprise NEMS can be alternatively viewed as an instance of a nonlinear complementarity problem (NCP). Before commenting on the specific form of the NEMS NCP, in the next section, we first introduce the general form of the NCP.

3 The Nonlinear Complementarity Problem and NEMS

3.1 Statement of the Nonlinear Complementarity Problem

In this section we describe the general form of the nonlinear complementarity problem of which NEMS is a special case. Having a function $F : R_+^n \rightarrow R^n$, the nonlinear complementarity problem $NCP(F)$ is to find an $x \in R^n$ such that

$$\begin{array}{llll} x_i \geq 0 & F_i(x) \geq 0 & F_i(x)x_i = 0 & \forall i \in T_1 \\ x_i \text{ free} & F_i(x) = 0 & F_i(x)x_i = 0 & \forall i \in T_2, \end{array} \quad (4)$$

where $T_1 \cup T_2$ is a partition of the indices $\{1, 2, \dots, n\}$. When T_1 is empty, this formulation reduces to solving a set of nonlinear equations. When both T_1 and T_2 are nonempty, we have what is called the *mixed* NCP; the term *mixed* refers to the fact that there is a mixture of inequalities and equations as well as the complementarity conditions $F_i(x)x_i = 0$, $i = 1, \dots, n$. And when T_2 is empty, we have the *pure* NCP, which is the conventional form of the problem. Throughout this paper, we will assume that $\text{NCP}(F)$ refers to the pure NCP formulation. However, for many results, the distinction between mixed and pure NCP is not necessary.

The NCP is a very general format for modeling various equilibrium problems in a variety of application areas. In particular, every nonlinear program is an instance of an NCP via the Karush-Kuhn-Tucker (KKT) optimality conditions. In addition, the NCP format includes as special cases, problems in game theory, network equilibrium modeling, traffic systems, and mechanical engineering; see Harker and Pang (1990). The NCP format is particularly attractive for NEMS because it offers such a wide range of useful generalizations to the current setup.

3.2 The NEMS Equilibrium Problem as a Nonlinear Complementarity Problem

3.2.1 Conversion, Transmission and Distribution of Energy

We will model the conversion, transmission and distribution of energy by $m = 4$ separate nonlinear programs (NLPs). These NLPs correspond, for example to the conversion of fuels into electricity in the Electricity Market module of NEMS, and the distribution of coal to meet demands.

The use of nonlinear programs (as opposed to linear ones) is a worthwhile generalization of what is currently employed in NEMS. There are several attractive reasons for analyzing a more general setting. First, as was noted in the introduction, there have been convergence problems with the current setup. In part, these difficulties are due to discontinuities of the solution mapping from the linear programs being used. In some cases, linear programs were used to approximate nonlinear programs. The hope is that by directly using NLPs, these and other convergence problems will be mitigated. In addition, the linear programming formulation represents a tractable simplification of activity in the energy sector, based in part on the relatively easy access to existing LP software. With the current favorable state of software for NLPs, the previous justification for use of LPs based on reasons of software availability may no longer hold.

We will ultimately be formulating the NEMS equilibrium problem as a nonlinear complementarity problem. This NCP will be formed by considering the Karush-Kuhn-Tucker optimality conditions of the nonlinear programs cited above, as well as various nonlinear equations related to NEMS. To this end, we will need to be sure that solving the KKT conditions will in fact lead us to a solution to the associated NLP. For this reason, we will make the conventional assumption that the KKT conditions are sufficient for optimality; note that this does not depart from the current NEMS format of using linear programs.

However, some comments concerning KKT conditions for NLPs are in order. For linear programs, these conditions are both necessary and sufficient. Hence, the set of optimal solutions is completely characterized by the set of KKT points. For nonlinear programs,

the KKT conditions are necessary only when certain constraint qualifications hold. In some sense, this is a small price to pay for including more realistic nonlinearities.

The sufficiency of these KKT conditions is guaranteed if the objective function and inequality constraint functions are convex (for less than or equal to constraints) and if the equality constraints are affine.⁶ We will assume throughout this paper that the constraint and objective functions are indeed of this form.

We begin by considering the i th nonlinear program in NEMS. It will take as inputs certain prices and quantities as well as other values that we can exclude from the equilibrium calculations. For notational convenience, we will denote the input prices and quantities demanded as p^i and q^i , respectively, where of course we mean that $p^i = p_{Ip(i)}$ and $q^i = q_{Iq(i)}$. The solution will be a vector denoted as x^i ; x^i is the same as $x_{Ox(i)}$. In addition, there will be multiplier vectors u^i and v^i associated, respectively, with the demand and nondemand constraints of this i th mathematical program, formally defined in an analogous way to p, q , and x . More specifically, we assume that the form of the i th NLP is

$$\text{minimize}_{x^i} \{ \theta^i(x^i, p^i) : -g^i(x^i) \geq N^i q^i, \quad -h^i(x^i) \geq 0, \quad x^i \geq 0 \},$$

where $x^i \in R^{var_i}$ and $g^i : R^{var_i} \rightarrow R^{ng_i}$ and $h^i : R^{var_i} \rightarrow R^{nh_i}$; var_i, ng_i and nh_i represent, respectively, the number of variables, g constraints, and h constraints. We will let $n_i = var_i + ng_i + nh_i$ denote the total number of variables as well as multipliers involved in this i th NLP. The objective function θ^i and the constraint functions g^i, h^i are assumed to be twice-continuously differentiable and convex, so that the KKT conditions are sufficient for solving this i th NLP. The assumption that $x^i \geq 0$ is made without loss of realism. Lastly, the matrix N^i converts the fuel quantities to appropriate regional levels analogously to the matrix N used in the LP formulations presented above.

If we were to generalize just the nonfuel costs portion of the objective function used at present, we would end up with

$$\theta^i(x^i, p^i) = \hat{\theta}^i(\hat{x}^i) + \sum_{j \in Ip(i)} p_j \tilde{x}_j^i,$$

where $\hat{\theta}^i$ is a convex function of \hat{x}^i , the nonfuel quantity variables, and \tilde{x}^i are the fuel quantity variables with $x^i = \begin{pmatrix} \hat{x}^i \\ \tilde{x}^i \end{pmatrix}$. While this formulation would generalize the current setup, the approximation to the fuel costs, namely, $\sum_{j \in Ip(i)} p_j \tilde{x}_j^i$, would still assume constant supply prices p_j . This formulation can be improved upon by instead using

$$\sum_{j \in Ip(i)} p_j(\tilde{x}^i) \tilde{x}_j^i,$$

where $p_j(\tilde{x}^i)$ is a better approximation to the supply price function, based on the fuel quantities \tilde{x}^i used in the i th NLP. Note that $j \in Ip(i)$ in this sense refers to those fuels j that are used in the i th NLP.

We require that the overall objective function be convex in x^i . One easy generalization to $\sum_{j \in Ip(i)} p_j \tilde{x}_j^i$ which would satisfy these convexity conditions would be to take

$$p_j(\tilde{x}^i) = p_j(\tilde{x}_j^i).$$

With this choice, we have the Hessian of $(\sum_j p_j(\tilde{x}^i)\tilde{x}_j^i)$ equal to a diagonal matrix with the k th diagonal being

$$p_k''(\tilde{x}_k^i)\tilde{x}_k^i + 2p_k'(\tilde{x}_k^i).$$

If $p_j(\tilde{x}_j^i)$ is convex and strictly increasing for all nonnegative arguments, then the desired convexity condition is satisfied. Indeed, in this case, the Hessian of $p_j(\tilde{x}_j^i)$ would be positive definite. By defining $\theta(x^i) = \hat{\theta}(\hat{x}^i) + \sum_{j \in I_p(i)} p_j(\tilde{x}_j^i)\tilde{x}_j^i$, with the proper convexity conditions holding, we can hope to avoid some of the present convergence problems by modeling the supply price function more accurately and inserting its approximation into the objective function of the relevant NLPS. This is the generalization to the objective function that we will use in what follows.

If we define the Lagrangian function $\mathcal{L}^i(x^i, u^i, v^i) = \theta^i(x^i) + (u^i)^T(g^i(x^i) + N^i q^i) + (v^i)^T(h^i(x^i))$, then we get the following NCP in n_i variables; i.e., the vectors x^i, u^i, v^i (note that $\nabla_{x^i} \mathcal{L}(x^i, u^i, v^i) = \nabla_{x^i} \theta^i(x^i) + (u^i)^T \nabla g^i(x^i)^T + (v^i)^T \nabla h^i(x^i)^T$).

Find a solution vector x^i , and multiplier vectors u^i and v^i to satisfy the conditions

$$\nabla_{x^i} \mathcal{L}^i(x^i, u^i, v^i) \geq 0 \quad x \geq 0 \quad (\nabla_{x^i} \mathcal{L}(x^i, u^i, v^i))^T x^i = 0$$

$$-g^i(x^i) - N^i q^i \geq 0 \quad u^i \geq 0 \quad (u^i)^T (-g^i(x^i) - N^i q^i) = 0$$

$$-h^i(x^i) \geq 0 \quad v^i \geq 0 \quad (v^i)^T h^i(x^i) = 0$$

so that the NCP function corresponding to this i th nonlinear program is just

$$NLP_i(x^i, u^i, v^i, q^i) = \begin{pmatrix} \nabla_{x^i} \mathcal{L}^i(x^i, u^i, v^i) \\ -g^i(x^i) - N^i q^i \\ -h^i(x^i) \end{pmatrix},$$

where $NLP_i : R^{n_i} \rightarrow R^{n_i}$. It is important to note that for each NLP, the vector q^i is fixed and thus not considered a variable in the NCP shown above. However, it is a variable in a larger NCP to be presented below.

3.2.2 Linking Equations

In an earlier section, we described how the fuel prices and some of the fuel quantities were calculated from the optimal solution or multiplier values from the associated linear programs. In this section, we will generalize the specific functions that “linked” the solutions and multipliers with the computed fuel prices and quantities.

We will assume that a certain subset of the prices and quantities will be calculated from the output from the nonlinear programming modules. In particular, for the fuel prices we will assume that the i th NLP will give rise to those prices indexed by the set $Op(i)$. In addition, the i th NLP will generate those quantities indexed by the set $Oq(i)$. The difference between the prices and quantities is that some of the quantities will be calculated via demand

equations (to be explained below). Hence, since we are considering m nonlinear programs, we have $\cup_{i=1}^m Op(i)$ contained in but not equal to $\{1, \dots, n\}$ but $\cup_{i=1}^m Op(i) = \{1, \dots, n\}$. Of course, every price or quantity will be calculated in just one of the ways mentioned above.

Fuel Prices

In the current NEMS setup, in some cases the equilibrium prices are just the multipliers (of affine transformations thereof) associated with demand constraints from a particular LP. In the more general case that we are considering, however, the prices will be allowed to be functions of the multipliers.

In particular, we assume that for each price variable p_j where $j \in Op(i)$, we have the linking equation

$$L_j^i(u^i) = p_j$$

or

$$L^i(u^i) = p_{Op(i)}$$

with $L^i : R^{ng_i} \rightarrow R^{|Op(i)|}$.⁷ Also, let $L : R^{\sum_{i=1}^m ng_i} \rightarrow R^n$ with

$$L = \begin{pmatrix} L^1(u^1) \\ L^2(u^2) \\ \vdots \\ L^m(u^m) \end{pmatrix}.$$

With the exception of the NGTDM prices, the prices from NLP i were a function just of the dual variables u^i . However, as shown above, we have assumed that $p_j = L_j^i(u^i)$ for all modules considered. This assumption was made simply for notational consideration. We could just as easily have defined $p^{NGTDM} = L^{NGTDM}(u)$ instead of $p^{NGTDM} = L^{NGTDM}(u^{NGTDM})$ and made the appropriate changes in ∇F to be shown below.

Fuel Quantities

In an analogous manner, we can define the linking constraints for the fuel quantities as

$$\hat{L}^i(x) = q_{Oq(i)},$$

where $\hat{L}^i : R^{\sum_{i=1}^m var_i} \rightarrow R^{|Oq(i)|}$ and

$$\hat{L} = \begin{pmatrix} \hat{L}^1(x^1) \\ \hat{L}^2(x^2) \\ \vdots \\ \hat{L}^m(x^m) \end{pmatrix}$$

for $\hat{L} : R^{\sum_{i=1}^m var_i} \rightarrow R^{\sum_{i=1}^m |Oq(i)|}$.⁸

3.3 Demand Equations

As was stated above, some of the fuel quantities will be calculated via demand equations. We will write down general nonlinear demand functions that are meant to incorporate what is currently being used in NEMS.

We will assume that the j th quantity q_j is calculated from prices (and other variables not relevant to our equilibrium analysis) via a demand equation of the following form:

$$Q_j(p_D) = q_j.$$

We will collect all those relevant quantity indices j into the set $D = (\bigcup_{i=1}^m Oq(i))^c$ so that we get

$$Q(p_D) = q_D,$$

where $Q : R^{|D|} \rightarrow R^{|D|}$. Without loss of realism, we will assume that L, \hat{L} , and Q are sufficiently smooth functions.

3.4 The NEMS NCP

Putting together the conversion, linking and demand sides of NEMS; we see that the NEMS equilibrium problem can be expressed as solving a pure NCP of size $N = \sum_{i=1}^m n_i + 2n$ whose function F is given as follows:

$$F(x, u, v, p, q) = \left\{ \begin{array}{ll} NLP_1(x^1, u^1, v^1, q^1) & \\ NLP_2(x^2, u^2, v^2, q^2) & \\ \vdots & \text{From nonlinear programs} \\ NLP_m(x^m, u^m, v^m, q^m) & \\ \text{-----} & \\ -L^1(u^1) + p_{Op(1)} & \\ -L^2(u^2) + p_{Op(2)} & \\ \vdots & \\ -L^m(u^m) + p_{Op(m)} & \\ \text{-----} & \\ -\hat{L}^1(x^1) + q_{Oq(1)} & \text{Linking constraints} \\ -\hat{L}^2(x^2) + q_{Oq(2)} & \\ \vdots & \\ -\hat{L}^m(x^m) + q_{Oq(m)} & \\ \text{-----} & \\ -Q(p_D) + q_D & \text{Demand equations} \end{array} \right\}. \quad (5)$$

A few remarks about this NCP are in order.

1. The pure NCP form of this problem implies that the linking and demand constraints are actually inequalities rather than equations as is needed. However, with the reasonable assumption that at a solution, prices and quantities are strictly positive, as opposed to being just nonnegative, the complementarity conditions force these constraints to be equations as desired. Also, note that if equality constraints appear in the nonlinear programs, we will have a mixed NCP rather than a pure one.
2. The division of the function F into components corresponding to nonlinear programs, linking constraints and demand equations is meant to parallel the current configuration in NEMS in which there is a separate module for each activity associated with one of these three components. It is of considerable interest to NEMS modelers and others to view the NCP in this fashion, rather than just substituting the linking and demand constraints for p and q into the NLP sections of F . The main reason is that in this separated form, we will more easily be able to develop NEMS NCP methods that minimally alter the current solution algorithm.
3. To allow for as general a setting as possible, we will take $q_{Iq(i)} \subseteq q_{Oq(i)}^c$.
4. We will need to compute Jacobians for the linking and demand functions. This task may involve computing approximate derivatives via finite differences or analytic derivatives as applied to approximations to the current (or proposed) linking and demand functions.

In what follows, we will group the variables together as

$$w^T = ((x^1)^T, (u^1)^T, (v^1)^T, \dots, (x^m)^T, (u^m)^T, (v^m)^T, p^T, q_D^T, q_{\bar{D}}^T).$$

The Jacobian of the NCP function F shown in (5) takes on the form

$$\begin{matrix} \sum_i n_i & n & |\bar{D}| & |D| \\ \sum_i n_i & \begin{pmatrix} J_{11} & 0 & J_{13} & J_{14} \\ J_{21} & I & 0 & 0 \\ J_{31} & 0 & I & 0 \\ 0 & J_{42} & 0 & I \end{pmatrix} \end{matrix}, \quad (6)$$

where the matrices J_{ij} are defined as follows:

$$J_{11} = \begin{pmatrix} D_{11} & 0 \dots & 0 \\ 0 & D_{22} \dots & 0 \\ 0 & \dots & D_{mm} \end{pmatrix}, \quad (7)$$

$$D_{ii} = \begin{matrix} & x^i & u^i & v^i \\ \begin{matrix} x^i \\ u^i \\ v^i \end{matrix} & \begin{pmatrix} \nabla_{x^i x^i} \mathcal{L} & \nabla g^i(x^i)^T & \nabla h^i(x^i)^T \\ -\nabla g^i(x^i) & 0 & 0 \\ -\nabla h^i(x^i) & 0 & 0 \end{pmatrix} \end{matrix}. \quad (8)$$

After possibly permuting the columns for the prices and quantities, the rows from the i th NLP for J_{12} , J_{13} , and J_{14} are thus

$$J_{13} = \begin{matrix} & q_{I_q(i) \cap \bar{D}} & q_{other} \\ \begin{matrix} \hat{x}^i \\ \tilde{x}^i \\ u^i \\ v^i \end{matrix} & \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ -R^i & 0 \\ 0 & 0 \end{pmatrix} \end{matrix} \quad (9)$$

and

$$J_{14} = \begin{matrix} & q_{I_q(i) \cap D} & q_{other} \\ \begin{matrix} \hat{x}^i \\ \tilde{x}^i \\ u^i \\ v^i \end{matrix} & \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ -S^i & 0 \\ 0 & 0 \end{pmatrix} \end{matrix}, \quad (10)$$

where $[R^i S^i]$ are the (possibly) permuted columns of N^i . Also, we have

$$J_{21} = \begin{matrix} & x^1 & u^1 & v^1 & x^2 & u^2 & v^2 \dots & x^m & u^m & v^m \\ \begin{matrix} p_{O_p(1)} \\ p_{O_p(2)} \\ \vdots \\ p_{O_p(m)} \end{matrix} & \begin{pmatrix} 0 & -\nabla L^1(u^1) & 0 & 0 & 0 & 0 \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\nabla L^2(u^2) & 0 \dots & 0 & 0 & 0 \\ \vdots & & & & & & & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 \dots & 0 & -\nabla L^m(u^m) & 0 \end{pmatrix} \end{matrix}, \quad (11)$$

$$J_{31} = \begin{matrix} & x^1 & u^1 & v^1 & x^2 & u^2 & v^2 \dots & x^m & u^m & v^m \\ \begin{matrix} q_{O_q(1)} \\ q_{O_q(2)} \\ \vdots \\ q_{O_q(m)} \end{matrix} & \begin{pmatrix} -\nabla \hat{L}^1(x^1) & 0 & 0 & 0 & 0 & 0 \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & -\nabla \hat{L}^2(x^2) & 0 & 0 \dots & 0 & 0 & 0 \\ \vdots & & & & & & & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 \dots & -\nabla \hat{L}^m(x^m) & 0 & 0 \end{pmatrix} \end{matrix}, \quad (12)$$

and

$$J_{42} = q_D \begin{pmatrix} p_D & p_{\bar{D}} \\ -\nabla Q(p_D) & 0 \end{pmatrix}. \quad (13)$$

As was shown above, the NEMS problem can be viewed as a large NCP with a good deal of sparsity in ∇F . To effectively solve this NEMS NCP, any proposed methods should take advantage of this sparsity by breaking the overall problem into smaller ones or by performing sparse matrix-vector calculations. The hope is to build methods that use the existing modules as much as possible but use the information in a way that is consistent with the more general NCP approach.

In what follows, we analyze several Newton-type NCP approaches which are particularly well-suited to the NEMS NCP. We have decided to focus just on Newton type methods for the NCP based on the robustness and fast convergence rates associated with these approaches and the generally favorable performance; for example, see Gabriel and Pang (1992), Pang and Gabriel (1993), Chen and Harker (1993), Gabriel and Pang (1994), Ralph (1994), Dirkse and Ferris (1994), Chen and Mangasarian (1994), and Dirkse and Ferris (1995). Each of these methods relies on a certain reformulation of the NCP into an equivalent but

computationally more useful problem. The key to implementing each of these methods for NEMS is to show how the direction-finding subproblems can be tailored to take advantage of the specific NEMS structure. For this reason, we concentrate our analysis mostly on the subproblems of these methods and omit other details of these approaches.

We consider only those NCP approaches that require inexact solutions of computationally manageable subproblems. This is significant because given the large-scale nature of NEMS, exact solution of the associated subproblems could be computationally prohibitive. This approach rules out methods that, for example, require the exact solution of linear complementarity or quadratic programming subproblems. (An exception is made for methods that have been successfully tested in practice and that have subproblems that can exploit the sparsity of NEMS). Also, we focus on methods that are applicable to general NCPs (for example, not valid just on monotone NCPs). This is relevant given the NEMS NCP (to be shown below), which is not necessarily monotone. Additionally, we also rule out methods that involve pivoting of a large linear system because this may cause excessive fill-in and make the method inappropriate for such a large-scale model as NEMS.

4 The NE/SQP Method

NE/SQP (for nonsmooth equations/sequential quadratic programming) is a recent method for solving general nonlinear complementarity problems. It has been shown to be globally convergent and fast (Q-quadratic rate), as well as robust in the sense that the direction-finding subproblems are always solvable.

The basis for this method is to solve $\text{NCP}(F)$ by first transforming it into the equivalent problem of finding the zero of a certain set of nonsmooth equations. Specifically, let the function $H : R_+^n \rightarrow R^n$ be defined by

$$H(x)_i = \min(x_i, F_i(x)) \quad i = 1, \dots, n. \quad (14)$$

It is not hard to see that a zero of this function H corresponds exactly to a solution to $\text{NCP}(F)$. Unfortunately, because of the presence of the min operator, this function is not differentiable (in the sense of Fréchet), so that standard algorithms such as Newton's method cannot directly be applied. However, the function H is directionally differentiable with the directional derivative $H'(x, d)$ in the direction d given by

$$H'_i(x, d) = \begin{cases} d_i & \text{if } i \in I_x(x) = \{i : x_i < F_i(x)\} \\ \nabla F_i(x)^T d & \text{if } i \in I_F(x) = \{i : x_i > F_i(x)\} \\ \min(d_i, \nabla F_i(x)^T d) & \text{if } i \in I_e(x) = \{i : x_i = F_i(x)\}, \end{cases} \quad (15)$$

for $i = 1, \dots, n$.

Closely related to H is the *norm function* $\theta : R_+^n \rightarrow R_+^n$ defined by

$$\theta(x) = \frac{1}{2} \|H(x)\|^2, \quad (16)$$

where $\|\cdot\|$ is the Euclidean norm. As a result, we see that $\text{NCP}(F)$ can be recast as the nonsmooth, nonconvex optimization problem

$$\begin{aligned} & \text{minimize}_x \quad \theta(x) \\ & \text{such that} \quad x \geq 0. \end{aligned} \quad (17)$$

Note that the nonnegativity constraints are actually embodied in the definition of θ but are used because their inclusion facilitates the relevant convergence analysis. In addition, for certain applications, the relevant functions are not necessarily defined even for negative values. Hence, these constraints are needed.

A solution to (17) for which θ equals zero corresponds exactly to a solution to $\text{NCP}(F)$. Starting at some initial estimate $x^0 \geq 0$, a natural scheme would then be to iteratively decrease the value of θ at each step, with the objective of driving it down to zero to obtain a solution; this is the essential idea of how NE/SQP works. Note that in general θ is only piecewise differentiable (since we assume that F is continuously differentiable) and in general θ is not even convex; to see this, take $F(x) = e^{-x}$ for $x \in \mathbb{R}$. However, an important characteristic of this function is that it is directionally differentiable with the directional derivative $\theta'(x, d) = H(x)^T H'(x, d)$.

The basic scheme with NE/SQP is thus as follows: having an estimate x^k of the solution, a new iterate x^{k+1} is generated according to the rule

$$x^{k+1} = x^k + \tau_k d^k,$$

where d^k is a suitable search direction and τ_k is the associated step length needed for global convergence of the method. The calculation of the search direction entails the solution of a certain convex quadratic program (QP) which we will now explain.

Let $\phi : \mathbb{R}_+^n \times \mathbb{R}^n \rightarrow \mathbb{R}_+^n$ be defined as

$$\phi(x, d) = \frac{1}{2} \|H(x) + M(x)d\|^2, \quad (18)$$

where $M(x)$ is the $n \times n$ matrix that (after possible reordering of rows and columns) is defined as

$$M(x) = \begin{pmatrix} I_{\alpha\alpha} & 0 \\ \nabla_{\alpha} F_{\beta} & \nabla_{\beta} F_{\beta} \end{pmatrix} \quad (19)$$

for index sets $\alpha = \{i : x_i \leq F_i(x)\}$ and $\beta = \{i : F_i(x) < x_i\}$ and $I_{\alpha\alpha}$ the identity matrix of order α .

With the iterate x^k , the associated direction-finding convex quadratic subproblem can thus be stated as ⁹

$$\begin{aligned} & \text{minimize}_d \quad \phi(x^k, d) \\ & \text{subject to} \quad x^k + d \geq 0. \end{aligned} \quad (20)$$

We note that the direction $d = 0$ is always feasible, since each iterate x^k is maintained nonnegative; see (17). As a result, the feasible region is a nonempty polyhedron, which taken together with the fact that the objective function is a quadratic bounded below by zero means that this QP will always have a solution (see Frank and Wolfe 1956). This result validates the robustness of NE/SQP. In fact, each subproblem need be only approximately solved to maintain the relevant convergent properties associated with exact subproblem solutions. The resulting inexact NE/SQP method has been developed and successfully tested by using a matrix splitting approach on the equivalent linear complementarity problem (LCP) form of the subproblem.

For this QP, the KKT optimality conditions are both necessary and sufficient so one can alternatively solve the equivalent linear complementarity problem. Having the vector x , this LCP is to find an $s = x + d$ such that

$$(c - Ax) + As \geq 0 \quad s \geq 0 \quad ((c - Ax) + As)^T s = 0,$$

where $A = M(x)^T M(x)$, and $c^T = H(x)^T M(x)$. We will refer to this LCP as $LCP(q, A)$ where $q = c - Ax$.

Note that the matrix $M = M(x)$ has a certain sparsity that is lost on A . It is reasonable to ask whether we can solve this LCP using the matrix M rather than A . Such an approach has obvious advantages for the NEMS NCP, where there is considerable additional sparsity in M derived from the special structure of the NEMS NCP. The following examples show that, in general, the answer is no.

Example 4.1 Let $M = \begin{pmatrix} 1 & 0 \\ a & b \end{pmatrix}$ and $q = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$, where the scalars $a, b > 0$. Then the unique solution to $LCP(q, M)$ is $s = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. However, if we take this solution in $LCP(q, M^T M)$, we get $q + M^T M s = \begin{pmatrix} a^2 \\ 1 + ab \end{pmatrix}$, which violates the complementarity condition.

This example clearly shows that using $LCP(q, M)$ to solve the NE/SQP subproblem $LCP(q, A)$ will generally not work. In a similar vein, one can ask whether, if M is invertible, solving $LCP(\tilde{q}, M)$ will provide a solution to the NE/SQP subproblem $LCP(q, M^T M)$ where $q = (M^T)\tilde{q}$. The answer here is also in the negative.

Example 4.2 Let the matrix M be given as in Example (4.1) with $\tilde{q} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$. Suppose we used the unique solution $s = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ in $LCP(q, M^T M)$. This would give $q + M^T M s = \begin{pmatrix} a + a^2 \\ b + ab \end{pmatrix}$, which violates the complementarity condition. So again, we see that solving an LCP with just the matrix M is not helpful for the NE/SQP subproblem.

Having generated a search direction d^k from either the QP or LCP form of the subproblem, we next determine a suitable step length. Such a parameter is used to guarantee sufficient decrease in the norm function θ and thus global convergence of NE/SQP. The well-known Armijo backtracking strategy is used to compute the step length τ_k . Specifically, having x^k and d^k , and a scalar $\rho \in (0, 1)$, we let m_k be the smallest nonnegative integer m such that

$$\theta(x^k + \rho^m d^k) - \theta(x^k) \leq -\sigma \rho^m z(x^k, d^k), \quad (21)$$

where $z(x, d) = \frac{1}{2} \|M(x)d\|^2$, and then let $\tau_k = \rho^{m_k}$ be the chosen step length.

From Lemma 2 (b), and Proposition 2 (b), (c) of Pang and Gabriel (1993), we see that as long as $\phi(x^k, d^k) < \phi(x^k, 0)$ d^k is a descent direction for θ at x^k and $z(x^k, d^k)$ is strictly positive thus forcing descent in θ in a finite number of trials. In fact, any d^k that satisfies

the following will suffice:

$$\left\{ \begin{array}{ll} \text{(a) descent in the norm function } \theta & \phi(x^k, d^k) < \phi(x^k, 0) = \theta(x^k) \\ \text{(b) subproblem feasibility} & x^k + d^k \geq 0 \\ \text{(c) inexact rule met} & \|\min(s^k, y^k)\| \leq \varepsilon_k \end{array} \right\}, \quad (22)$$

where the LCP variables are $s^k = x^k + d^k$ and $y^k = q + M s^k$, and $\{\varepsilon_k\} \downarrow 0$. The importance of (c) is that d^k must be an approximate solution to the LCP subproblem. This is an easy requirement to see because (s^*, y^*) solves the LCP if and only if $\|\min(s^*, y^*)\| = 0$. For other values of s and y , this *residue function* $\|\min(\cdot, \cdot)\|$ is nonnegative. Also, the scalars ε_k control the level of inexactness that must go to zero in the limit.

Clearly, in general, solving each subproblem exactly will satisfy the above conditions on d^k . However, potentially great computational gains can be made by only approximately solving the subproblem at each outer iteration. In short, the strategy with the inexact NE/SQP approach is to apply a sequence of “inner” iterations corresponding to solving the subproblem inexactly. Then, the direction that is obtained is used in an “outer” NE/SQP iteration in conjunction with the Armijo test described above. Thus, one can avoid costly calculations associated with solving each subproblem exactly. In the rest of this section, we will analyze several LCP/QP algorithms for which there is a reasonable chance or a definite certainty that the inexact NE/SQP conditions (22) (a)–(c) can be satisfied for the proposed algorithms.

To use NE/SQP effectively for NEMS, we need methods that can relatively easily generate directions satisfying (22) (a)–(c) while exploiting the structure and sparsity in the NEMS NCP. In general, any method for convex QPs with simple bound constraints or any monotone LCP approach is potentially useful.

Our aim is to present a representative but not totally exhaustive list of methods, each able to reasonably guarantee (a)–(c) as the algorithm proceeds. In addition, the proposed methods should at worst, solve only sparse linear systems. In fact, many of the intermediate calculations in the methods we analyze involve just the sparse matrix M times a vector rather than the matrix $A = M^T M$; this strategy is crucial for solving the NEMS problem.

We will analyze several candidate approaches that work on either the QP or the LCP form of the subproblem. The common feature to all these methods is that they maintain at least one of the inexact conditions (a)–(c) and work toward satisfying the remaining ones.

We highlight several projected matrix-splitting methods (Jacobi, SOR) that work on the LCP form of the subproblem, maintaining nonnegativity of the iterates and producing descent in the QP objective. Hence, having the k th NE/SQP iterate x^k , if we start the QP algorithm at $s^0 = x^k$, then conditions (a) and (b) will be satisfied. The remaining condition (c) is generally satisfied in the limit as a result of some feature of the method. An advantage of these methods is that they decompose the problem into pieces that can roughly match the current set of NEMS modules with closed form solutions for generating iterates. The resulting calculations involve solving much smaller LCPs corresponding to this decomposition, with relatively small sparse linear systems to be solved as the most complicated step. These reduced LCPs can be solved either in parallel (projected Jacobi) or sequentially (projected SOR). Since the current NEMS solution strategy is based on a nonlinear SOR approach, we feel that relative to other proposed methods, these splitting

approaches would most easily be incorporated into the existing framework. This is quite important given the large amount of development time already invested. A drawback of these approaches is the sometimes slow convergence rates in practice.

We also analyze a class of infeasible interior-point methods for the LCP subproblem. The distinguishing feature of these methods is that they maintain nonnegativity of the iterates, condition (b), and simultaneously work toward satisfying (a) and (c). This is accomplished by reducing the complementarity gap $s^T y / n^{10}$ at each stage and reducing the infeasibility, namely, $\|y - q - As\|$. Once a feasible y is found, these methods maintain feasibility and thus concentrate on reducing the complementarity gap. The condition (a) is likely to be satisfied at some intermediate iteration if the iterates are converging to a solution because we initiate the method at $s^0 = x^k$. The advantage of interior-point methods is their speed and relatively low number of iterations for large problems. We end up needing to solve a large sparse linear system which can be effectively handled with the NEMS structure.

The last set of methods we consider are based on the QP form of the subproblem. These active set approaches solve a sequence of smaller equality-constrained QPs relating to a selection of specific variables not at their lower or upper bound. The important feature of these strategies is that they maintain nonnegativity and decrease the objective function at each stage, thereby validating (b) and (a) assuming that we start at $d^k = 0$. The condition (c) is satisfied as the algorithm proceeds, since these methods work toward optimality of the QP. An advantage of these approaches is that they are likely to produce inexact directions with only minimal computational requirements such as matrix-vector products.

4.1 The Projected Block Jacobi Approach for the LCP Subproblem

We consider a modification of the projected block Jacobi approach which can avoid explicitly forming the matrix $A = M^T M$ and is ideal for parallel computing.

The parallelization aspect comes into play if we break up the matrix LCP A into b^2 blocks with block A_{ij} of size $b_i \times b_j$; the trick is how to form these blocks.

As is typically done with splitting methods, we will first split A into the sum of two matrices B and C . Strictly speaking, the projected block Jacobi method would just take B to be a block diagonal matrix with the i th block $B_{ii} = A_{ii}$. In our version of the projected block Jacobi, $B_{ii} = A_{ii} + I$ where the latter identity matrix is of order b_i .¹¹ Also, we will let $q^\nu = q + C s^\nu$ where ν is the inner iteration counter. Then, $\text{LCP}(q^\nu, A)$ can be solved by simultaneously solving for $\text{LCP}(q_i^\nu, B_{ii})$, $i = 1, \dots, b$.

It is worthwhile to analyze how the blocks of the matrix A should be organized. Our overriding concern is to use the existing NEMS routines as much as possible, with the exception that NLPs will replace LPs.

Let us examine the various matrix and vector computations. We see that an equivalent form of the subproblem has A of the following form:

$$A = \begin{pmatrix} I + \nabla_\alpha F_\beta(w)^T \nabla_\alpha F_\beta(w) & \nabla_\alpha F_\beta(w)^T \nabla_\beta F_\beta(w) \\ \nabla_\beta F_\beta(w)^T \nabla_\alpha F_\beta(w) & \nabla_\beta F_\beta(w)^T \nabla_\beta F_\beta(w) \end{pmatrix}. \quad (23)$$

As opposed to a general NCP, the index sets α and β for the NEMS problem are somewhat predetermined. Since the linking and demand relationships are equations, without loss of generality we can place the indices for these components of the NEMS NCP function

into the set $I_F(w)$. The remaining component indices for the KKT conditions for the m nonlinear programs will be broken down into two parts. Let us denote by β_N all those indices for the KKT conditions that are also in $I_F(w)$ and by α the remaining KKT indices. The set β will thus be partitioned into β_N (“N” for NLP) and β_O (“O” for other) where $\beta_O = \beta \setminus \beta_N$. Then, we have

$$\nabla_{\alpha} F_{\beta}(w) = \begin{pmatrix} (J_{11})_{\beta_N \alpha} \\ 0 \\ (J_{31})_{\bullet \alpha} \\ 0 \end{pmatrix}, \nabla_{\beta} F_{\beta}(w) = \begin{pmatrix} (J_{11})_{\beta_N \beta_N} & 0 & (J_{13})_{\beta_N \bullet} & (J_{14})_{\beta_N \bullet} \\ (J_{21})_{\bullet \beta_N} & I & 0 & 0 \\ (J_{31})_{\bullet \beta_N} & 0 & I & 0 \\ 0 & J_{42} & 0 & I \end{pmatrix} \quad (24)$$

where $M_{\bullet S}$ denotes the columns from a matrix M indexed by the set S ; $A_{S\bullet}$ is defined analogously for rows.

Now, let us examine the special structure of A . In light of (24), we see that the diagonal blocks of A have the following form (we assume that the indices have been ordered accordingly)

(a) rows and columns $1, \dots, |\alpha|$:

$$\hat{A}_{11} = I + (J_{11})_{\beta_N \alpha}^T (J_{11})_{\beta_N \alpha} + (J_{21})_{\bullet \alpha}^T (J_{21})_{\bullet \alpha} + (J_{31})_{\bullet \alpha}^T (J_{31})_{\bullet \alpha};$$

(b) rows and columns $|\alpha| + 1, \dots, |\alpha| + |\beta_N|$:

$$\hat{A}_{22} = (J_{11})_{\beta_N \beta_N}^T (J_{11})_{\beta_N \beta_N} + (J_{21})_{\bullet \beta_N}^T (J_{21})_{\bullet \beta_N} + (J_{31})_{\bullet \beta_N}^T (J_{31})_{\bullet \beta_N};$$

(c) rows and columns $|\alpha| + |\beta_N| + 1, \dots, |\alpha| + |\beta_N| + n$:

$$\hat{A}_{33} = J_{42}^T J_{42} + I;$$

(d) rows and columns $|\alpha| + |\beta_N| + n + 1, \dots, |\alpha| + |\beta_N| + n + |\bar{D}|$:

$$\hat{A}_{44} = (J_{13})_{\beta_N \bullet}^T (J_{13})_{\beta_N \bullet} + I;$$

(e) rows and columns $|\alpha| + |\beta_N| + n + |\bar{D}| + 1, \dots, N$:

$$\hat{A}_{55} = (J_{14})_{\beta_N \bullet}^T (J_{14})_{\beta_N \bullet} + I.$$

The natural scheme for splitting the subproblem matrix A will be to take either $B_{ii} = \hat{A}_{ii} + I$ for $i = 1, \dots, b = 5$ or B_{ii} equal to a principal submatrix of $\hat{A}_{ii} + I$ with $b > 5$.

It is important to never actually compute the matrix $A = M^T M$, since this computation would destroy the sparsity of M , which is critical to solving this large-scale problem efficiently. The following is our proposed way to use a projected Jacobi-type splitting algorithm with line search for NEMS; see Cottle, Pang, and Stone (1992) for details on this approach.

In the projected block Jacobi method with line search, many of the steps will involve just matrix-vector products using the sparse M . In addition, there is a line search step for

which the same comment is valid. The possible exception to this occurs when the following intermediate direction is computed for the i th block of variables. We have

$$s_i^{\nu+1/2} = \max [0, s_i^\nu - v_i], \quad (25)$$

where the vector v_i is calculated according to

$$(A_{ii} + I)v_i = (q + M^T M s^\nu)_i, \quad (26)$$

and $A_{ii} = E_{ii}^T E_{ii}$ for

$$E_{ii} = \begin{pmatrix} M_{1i} \\ M_{2i} \\ \vdots \\ M_{b_i} \end{pmatrix}. \quad (27)$$

We wish to avoid actually forming $(A_{ii} + I)$, which would ruin the sparsity of the problem. The next lemma shows how we can solve this system without losing sparsity.

Lemma 4.3 *Let $W, T, U, V \in R^{n \times n}$, $b \in R^n$ with T, U, V positive diagonal matrices such that $T = UV$. Then, the following statements hold:*

- (i) *the matrix $W^T W + T$ is symmetric positive definite,*
- (ii) *with $R = W^T W + T$, the unique solution x^* to*

$$Rx = b \quad (28)$$

can be obtained by taking $x = V^{-1}z^2$ where $z^T = [(z^1)^T (z^2)^T]$ solves the $2n \times 2n$ system

$$Nz = \begin{pmatrix} b \\ 0 \end{pmatrix} \quad (29)$$

where

$$N = \begin{pmatrix} W^T & U \\ I & -WV^{-1} \end{pmatrix}.$$

Proof

Condition (i) follows because R is the sum of a symmetric positive semidefinite matrix and a symmetric positive definite matrix. If we let $P \in R^{2n \times 2n}$ be defined by

$$P = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix},$$

then by the Schur formula for determinants,

$$\begin{aligned} \det(N) \neq 0 &\Leftrightarrow \det(NP) \neq 0 \\ &\Leftrightarrow \det(I + WV^{-1}U^{-1}W^T) \neq 0 \\ &\Leftrightarrow \det(I + WT^{-1}W^T) \neq 0 \\ &\Leftrightarrow \det(I + WT^{-1/2}(WT^{-1/2})^T) \neq 0. \end{aligned}$$

Since $I + WT^{-1/2}(WT^{-1/2})^T$ is symmetric positive definite, we see that the system (29) will have a unique solution for all b . Now, letting $V\hat{x} = z^2$ and substituting into the bottom row of (29) we have $z^1 = W\hat{x}$. Plugging this into the first row of this system gives $W^TW\hat{x} + UV\hat{x} = R\hat{x} = b$. By (i), we see that $\hat{x} = x^*$ as desired. \square

The advantage of solving the larger system (29) versus (28) is that the product W^TW need not be explicitly formed. This result has direct importance for our calculations. Specifically, we can solve (26) by letting

$$W = E_{ii}, U = V = I$$

and applying the method of the above lemma; this can be done for each block in parallel as well. Note that the matrix N in the system (29) is sparse and does not change for the calculations of a given subproblem. Hence, one can use a sparse factorization routine; see Duff, Erismann, and Reid (1986) for a discussion of sparse methods.

Now we state a result concerning the application of this method to the NEMS subproblem.

Theorem 4.4 *Suppose that $\{s^\nu\}$ are iterates generated by the modified projected Jacobi approach with $s^0 = w^k$ and either*

- (i) $\nabla_\beta F_\beta(w)$ is nonsingular or
- (ii) $\exists z \in R^{|\beta|}$ such that if $\nabla_\beta F_\beta(w)z = 0, 0 \leq z \neq 0$, then $q_\beta^T z > 0$ where $q_\beta = \{q_i : i \in \beta\}$.

Then, the following statements hold:

1. *the method will solve the subproblem in a finite number of steps, or*
2. *the inexact NE/SQP conditions (22) (a)–(c) can be met for each subproblem.*

Proof

First note that the matrix A is symmetric and B is positive definite. Hence, by Lemma 5.5.1 in Cottle, Pang, and Stone (1992), we see that either s^ν solves the subproblem or, for $g(s) = q^Ts + \frac{1}{2}s^TAs$, we have that $g(s^\nu) < g(s^{\nu-1}) < \dots < g(s^0)$ which is equivalent to $\phi(x^k, d^\nu) < \dots < \phi(x^k, 0)$. Since $s^0 = w^k$ and $s^\nu + w^k = d^k$, we see that the condition (a) is met; (b) is also met because the iterates are maintained nonnegative. Lastly, by (i) or (ii), and Lemma 5.3.4 and Theorem 5.5.3 in Cottle, Pang, and Stone (1992), $\{s^\nu\} \in R_+^n$ is bounded and every accumulation point solves the subproblem so that the remaining inexact NE/SQP conditions are also met subsequentially. \square

Remarks:

1. When neither (i) nor (ii) hold, but the iterates $\{s^\nu\}$ are still bounded, by Theorem 5.5.3 in the cited reference, every accumulation point of this sequence is a solution to the NE/SQP subproblem so that (22) (a)–(c) are satisfied subsequentially.
2. Even when the iterates contain no convergent subsequence, we still have $g(s^\nu) < g(s^{\nu-1})$ thus, since $s^0 = w^k$, $d^\nu = s^\nu - x^k$ will be a descent direction for the merit function θ at w^k .

3. In the splitting approach described above, each block could refer to a current NEMS module (or pieces thereof). The overall effect would be to make use of the current information being passed between NEMS modules in a different and (one hopes) more effective manner.

4.2 The Projected Block Gauss-Seidel and SOR Approaches for the LCP Subproblem

One theoretical weakness of the Jacobi approach described above is that the boundedness of the iterates $\{s^\nu\}$ cannot be guaranteed for the NE/SQP subproblem. The projected block Gauss-Seidel approach overcomes this deficiency but is not as parallelizable as the former method. In particular, one must solve the associated inner subproblems sequentially rather than at the same time. Convergence of the iterates is then assured if the splitting of A is *regular* a term we now define.

Definition 4.5 *Let A be an $n \times n$ matrix with $A = B + C$. Then the splitting (B, C) is called weakly regular if $B - C$ is positive semidefinite and regular if $B - C$ is positive definite.*

The projected block Gauss-Seidel approach begins by partitioning the set of variables $\{1, \dots, n\}$ into b distinct sets I_1, \dots, I_b and for each $j = 1, \dots, b$, choosing a regular splitting $(B_{I_j I_j}, C_{I_j I_j})$. Then, having an iterate s^ν , we take $s_{I_j}^{\nu+1}$ as a solution to the LCP($q^j, B_{I_j I_j}$), where

$$q^j = q_{I_j} + C_{I_j I_j} s_{I_j}^\nu + A_{I_j \tilde{I}_j} s_{\tilde{I}_j}^\nu \quad (30)$$

for $\tilde{I}_j = (I_j)^c$. Note that q^j can be computed by making use of the sparse structure of A and C . The key for using this approach with the NEMS NCP is to select a regular splitting $(B_{I_j I_j}, C_{I_j I_j})$ and solve the associated LCP efficiently.

Lemma 4.6 *Let $A \in R^{n \times n}$ be symmetric positive semidefinite with $A = B + C$. Let L, U, D be respectively, the strictly lower, strictly upper, and diagonal parts of A . Then for $B = L + D + I$, (B, C) is a regular splitting.*

Proof

We see that $B - C = L - U + D + 2I$. We note that $B - C$ is positive definite if and only if $(B - C) + (B - C)^T$ is symmetric positive definite. But since A is symmetric, $(L - U)^T = U - L$, so that

$$(B - C) + (B - C)^T = 2D + 4I,$$

which is positive definite given that A , hence D is symmetric positive semidefinite. \square

Since every principal submatrix of a positive definite matrix is also positive definite, this result shows that $(B_{I_j I_j}, C_{I_j I_j})$ will also be regular for all $j = 1, \dots, b$ given the choice for B described above. It remains to show how we can efficiently implement this approach for NEMS taking into account the sparsity of the NCP formulation.

With the projected block Gauss-Seidel approach, we will be solving LCPs with the matrix $B_{I_j I_j} = (L + D + I)_{I_j I_j}$. For notational simplicity, let $B_{I_j I_j} = \tilde{B}$ with \tilde{B} of size b ,

and let the constant vector be denoted simply as q . With the iterate s^ν , for $\text{LCP}(q, \tilde{B})$, the projected Gauss-Seidel recursively computes $s_i^{\nu+1}$ ¹² by solving

$$s_i^{\nu+1} = \max \left[0, s_i^\nu - \frac{1}{\tilde{b}_{ii}} \left(q_i + \sum_{j < i} \tilde{b}_{ij} s_j^{\nu+1} + \sum_{j \geq i} \tilde{b}_{ij} s_j^\nu \right) \right], i = 1, \dots, m. \quad (31)$$

The only potentially challenging step is to compute

$$\left(\sum_{j < i} \tilde{b}_{ij} s_j^{\nu+1} + \sum_{j \geq i} \tilde{b}_{ij} s_j^\nu \right)$$

without explicitly forming the matrix \tilde{B} which would ruin the sparsity in the problem. This can be easily handled as follows. Let v^i be the following intermediate vector¹³

$$(v^i)^T = (s_1^{\nu+1}, \dots, s_{i-1}^{\nu+1}, s_i^\nu, \dots, s_m^\nu),$$

and notice that

$$v_j^{i+1} - v_j^i = \begin{cases} 0 & i \neq j \\ s_i^{\nu+1} - s_i^\nu & i = j. \end{cases}$$

Since \tilde{B} is a principal submatrix of $L + D + I$, we have¹⁴

$$\tilde{b}_{ij} = \begin{cases} M_{\bullet i}^T M_{\bullet j} & i > j \\ M_{\bullet i}^T M_{\bullet j} + 1 & i = j \\ 0 & i < j. \end{cases}$$

Hence, we have the following for row i :

$$\begin{aligned} \sum_{j < i} \tilde{b}_{ij} s_j^{\nu+1} + \sum_{j \geq i} \tilde{b}_{ij} s_j^\nu &= \sum_{j=1}^i \tilde{b}_{ij} v_j^i \\ &= M_{\bullet i}^T (\sum_{j=1}^i M_{\bullet j} v_j^i) + v_i^i. \end{aligned}$$

If we denote the quantity $(\sum_{j=1}^i M_{\bullet j} v_j^i)$ by σ_i , for row $i+1$ we have just a simple adjustment as follows:

$$\begin{aligned} \sum_{j < i+1} \tilde{b}_{i+1,j} s_j^{\nu+1} + \sum_{j \geq i+1} \tilde{b}_{i+1,j} s_j^\nu &= \sum_{j=1}^{i+1} \tilde{b}_{i+1,j} v_j^{i+1} \\ &= M_{\bullet i+1}^T (\sum_{j=1}^{i+1} M_{\bullet j} v_j^{i+1}) + v_{i+1}^{i+1} \\ &= M_{\bullet i+1}^T (\sigma_i + M_{\bullet i+1} (s_i^{\nu+1} - s_i^\nu)) + v_{i+1}^{i+1}. \end{aligned}$$

In this way, we never need to form the matrix \tilde{B} explicitly and can exploit the nonzero structure of M . We have the following result concerning the application of the projected block Gauss-Seidel method.

Theorem 4.7 *Suppose that $\{s^\nu\}$ are iterates generated by the projected block Gauss-Seidel method with $s^0 = w^k$. Then, the inexact NE/SQP conditions (22) (a)–(c) can be met for each subproblem.*

Proof

From Theorem 3 of Luo and Tseng (1991), we know that the nonnegative iterates of this algorithm will converge to a solution of the NE/SQP problem. This means that since $s^0 = w^k$, all three conditions (a)–(c) can be satisfied. \square

Remarks: We can modify this algorithm to include overrelaxation and underrelaxation parameters as well. Depending on the particular choice of relaxation, Theorem 4.7 either remains valid or at worst we know that

$$\min\{\|s^\nu - s^*\| : s^* \text{ is a subproblem solution}\} \rightarrow 0.$$

In this case, we cannot assert that the iterates converge to a solution, but we can see that every accumulation point (if any exist) will be a subproblem solution. Hence, subsequentially, the inexact conditions can be met.

4.3 An Infeasible-Interior-Point Approach for the LCP Subproblem

Finding a solution s and complementary vector y to $\text{LCP}(q, A)$ can be recast as finding the zero of the following system of constrained nonlinear equations: find $s, y \in R_+^n$ to satisfy

$$G(s, y) = \begin{pmatrix} As + q - y \\ SYe \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad (32)$$

where $S = \text{diag}(s)$, $Y = \text{diag}(y)$, and $e^T = [1 \dots 1]$. Recent infeasible- interior-point approaches for monotone LCPs (like the NE/SQP subproblem) use this formulation and apply a variation of Newton's method with centering; for example, see the recent work of Wright (1995), and Zhang (1994).

Specifically, having the iterates $s^\nu, y^\nu \in R_{++}^n$, with $\mu_\nu = (s^\nu)^T y^\nu / n$, we generate the Newton search direction for

$$G(s, y) = \begin{pmatrix} As + q - y \\ SYe \end{pmatrix} = \begin{pmatrix} 0 \\ \tilde{\sigma}, \mu_\nu e \end{pmatrix} \quad (33)$$

where $\tilde{\sigma} \in [0, 1)$ is a parameter whose value can vary by iteration. This system is solved (at least once but perhaps several times with different right-hand sides) to obtain an appropriate search direction; in addition, a line search step is added.

In terms of applicability to NEMS, we are primarily concerned with how this linear system can be solved efficiently, ensuring that the inexact NE/SQP conditions can be met. These methods typically maintain nonnegativity (positivity) of s^ν, y^ν and strive to decrease the complementarity gap μ_ν and the infeasibility, in other words, $\|y^\nu - (As^\nu + q)\|$.

The resulting search directions $\Delta s^\nu, \Delta y^\nu$ for s^ν and y^ν , respectively, are computed as solutions to the following system of linear equations

$$\begin{pmatrix} A & -I \\ Y^\nu & S^\nu \end{pmatrix} \begin{pmatrix} \Delta s \\ \Delta y \end{pmatrix} = \begin{pmatrix} r^\nu \\ -S^\nu Y^\nu e + \tilde{\sigma} \mu_\nu e \end{pmatrix}, \quad (34)$$

where $S^\nu = \text{diag}(s^\nu)$, $Y^\nu = \text{diag}(y^\nu)$, and $r^\nu = y^\nu - As^\nu - q$ is the residual vector at iteration ν .

It is not hard to see that (34) can be reduced to solving for Δs^ν in

$$(A + (S^\nu)^{-1}Y^\nu)\Delta s = r^\nu - y^\nu + \tilde{\sigma}\mu_\nu(s^\nu)^{-1}, \quad (35)$$

where $(s^\nu)_i^{-1} = 1/s_i^\nu$ $i = 1, \dots, n$ and for Δy^ν by

$$\Delta y^\nu = A\Delta s^\nu - r^\nu. \quad (36)$$

The only challenging computation is thus to solve for Δs .

Of course, since for NEMS the matrix A is of the form $M^T M$, we wish to avoid explicitly forming this product of matrices. We appeal to the result from Lemma 4.3. Specifically, we see that the system (35) is of the required form if we let

$$W = M, U = (S^\nu)^{-1}, V = Y^\nu. \quad (37)$$

Hence, we need to solve a linear system whose matrix is of the form

$$\begin{pmatrix} M^T & (S^\nu)^{-1} \\ I & -M(Y^\nu)^{-1} \end{pmatrix}. \quad (38)$$

This matrix is clearly nonsingular, and it is easy to see that the sparsity pattern for this matrix is unchanging throughout the LCP subproblem calculations. Moreover, many of the actual values do not even change. This feature has obvious advantages if we employ a sparse factorization of this matrix. However, the disadvantage with solving this entire system together is the size.

We can solve this system by an iterative method such as QMR; see Barrett et al. (1994), Freund and Nachtigal (1991), and Freund and Nachtigal (1994). This approach is ideal for large systems of the form $Bx = b$, where B is sparse and nonsingular. The most challenging step in this approach is forming the product Bv where v is a particular intermediate vector. This can be facilitated given the structure of the matrix in (38) and the fact that the current NEMS modules (or generalized versions of them) can generate the necessary parts of this matrix.

Since s^ν is maintained nonnegative, the inexact condition (b) is satisfied. Also, as described in Wright (1995), and Zhang (1994), once a feasible vector y^ν is found, y^λ is feasible for all $\lambda > \nu$. Since infeasibility and complementarity are being driven to zero, this means that since $\mu_\nu \downarrow 0$, the condition (c) will be met for each subproblem iteration. As long as the iterates converge (this was shown in Wright 1995), then for $s^0 = w^k$, it is likely that the condition (a) will also be met for each LCP iteration as well. Hence, we see that interior-point methods of the kind described are likely to solve the inexact conditions.

4.4 An Active Set Approach for the QP Subproblem

Suppose that we are trying to solve the convex quadratic program

$$\min_d \{q(d) | d \in \Omega\} \quad (39)$$

where $q : R^n \rightarrow R$ is a convex quadratic function, and $\Omega = \{d \in R^n | l \leq d \leq u\}$ for $l \leq u$, $l, u \in R^n$, (this, of course, includes the NE/SQP QP where $l = -x^k$ and $u = +\infty$). We will let n be the size of the subproblem.

If we knew the active set of indices at a solution d^* , that is, $\mathcal{A}(d^*) = \{i : d_i = l_i \text{ or } d_i = u_i\}$, then (39) could be solved by the following unconstrained QP in just the free variables $d_i, i \in \mathcal{A}(d^*)^c$,

$$\min_w \{q_\nu(w) : w \in R^{m_k}\} \quad (40)$$

where m_k is the number of free variables and $q_\nu(w) = \frac{1}{2}w^T A_\nu w + r_\nu^T w$ for A_ν, r_ν the reduced Hessian and gradient of $q(\cdot)$, respectively.

Active set approaches solve a sequence of reduced QPs of the form (40) and either find a solution to the overall problem or modify the active set of indices. For our purposes, we will analyze the recent active set approach described in Moré and Toraldo (1991). This method is particularly suitable for NEMS because the conjugate gradient approach is used to generate directions and the most computationally challenging aspect of this method is to compute matrix vector products; these sorts of calculation can effectively make use of the sparse structure for NEMS. In addition, as will be shown, under reasonable hypotheses, the inexact NE/SQP conditions can be met.¹⁵

In the cited reference, the conjugate gradient method is used to compute directions w^0, w^1, \dots until a w^j is generated that satisfies sufficient decrease for $q_\nu(\cdot)$, namely,

$$q_\nu(w^{j-1}) - q_\nu(w^j) \leq \eta_1 \max\{q_\nu(w^{l-1}) - q_\nu(w^l) : 1 \leq l < j\} \quad (41)$$

for $\eta_1 > 0$. The approximate solution to the reduced QP is then given as $\hat{d}^\nu = Z_\nu w^{j_\nu}$, where j_ν is the first index j satisfying (41) and $Z_\nu \in R^{n \times m_\nu}$ has as its j th column the i_j th column of the $n \times n$ identity matrix, where i_j refers to the j th free variable.

In order to be able to pick up more than one constraint at a time,¹⁶ a projected search is used to define the step length α_ν which is used to compute the next iterate as follows:

$$d^{\nu+1} = P(d^\nu + \alpha_\nu \hat{d}^\nu),$$

where P is the projection onto Ω .¹⁷ The key point is that the projected search selects an $\alpha_\nu > 0$ so that $q(d^{\nu+1}) < q(d^\nu)$, which is useful for the inexact condition (a). In the QP algorithm under consideration, the conjugate gradient method is used to explore a face of the feasible region that has been chosen by the projected gradient algorithm. Based on this active set strategy, we have the following result for the inexact NE/SQP method.

Theorem 4.8 *Suppose that the active set method described above is used on the QP form of the NE/SQP subproblem with $d^0 = w^k$. Then, for $\beta = I_F(w^k)$, if $\nabla \beta F_\beta(w)$ is nonsingular, the inexact conditions (22) (a)–(c) can be met.*

Proof

Under the nonsingularity assumption, we see that the NE/SQP subproblem has a strictly convex objective function. By construction, the iterates $\{d^\nu\}$ satisfy the nonnegativity conditions (b), (a) is guaranteed by choice of $d^0 = w^k$ and the calculation of the step length since $q(d^{\nu+1}) < q(d^\nu)$. The condition (c) then follows from Theorem 5.1 in Moré and Toraldo (1991), since the sequence $\{d^\nu\}$ terminates finitely or converges to an exact solution. \square

Remark: When the nonsingularity condition cited above is not met, the active set approach is still useful because the condition (a) is satisfied. Hence we have a descent direction for θ at w^k . In practice, we can use an upper bound on the number of inner QP iterations, which may actually preclude the condition (c) from always being satisfied. However, this has not caused problems in previous work; see Gabriel and Pang (1992).

5 A Bound-Constrained Nonlinear Least Squares Method

Another recent Newton type approach for the general NCP is the method of Moré (1994). In this approach, the NCP is formulated as the following equivalent constrained system of nonlinear equations. Find $x, y \in R^n$ such that

$$h(x, y) = 0, x \geq 0, y \geq 0 \quad (42)$$

where $h : R^{2n} \rightarrow R^{2n}$ is defined by

$$h(x, y) = \begin{pmatrix} F(x) - y \\ Yx \end{pmatrix} \quad (43)$$

and Y is the diagonal matrix $\text{diag}(y_1, \dots, y_{2n})$.

Based on this formulation, one can solve $\text{NCP}(F)$ by the following bound-constrained nonlinear least squares problem:

$$\min_{x, y} \left\{ \frac{1}{2} \|h(x, y)\|_2^2 : x \geq 0, y \geq 0 \right\}. \quad (44)$$

Under reasonable conditions global and superlinear or quadratic convergence to a solution can be obtained.

As advocated in this work, the trust region method of Burke, Moré, and Toraldo (1990) is particularly suitable. With this formulation, this method can be applied to the general minimization problem

$$\min \{f_0(z) : z \in \Omega\}$$

for Ω a closed convex set. In the NCP case, we have $f_0(z) = \frac{1}{2} \|h(z)\|^2$, and $\Omega = R_+^{2n}$. The proposed trust region method is appropriate because its convergence rate covers the case of degenerate minimization. Nondegeneracy here means that at a solution z^* , $[\nabla f_0(z^*)]_i \neq 0$ for $i = 1, \dots, 2n$. The NCP problem is degenerate because at a solution we have $h(z^*) = 0$ but

$$\nabla f_0(z^*) = \nabla h(z^*)^T h(z^*) = 0.$$

In addition, the proposed trust region method is valid when projected searches (as advocated in Moré 1994) are used, this is not always the case for other methods that cover degenerate minimization.

The idea of the trust region approach is to create a local model of the decrease in the objective function f_0 and then, depending on how well this local model performs, update accordingly the trust region radius and the local model itself. More specifically, at iteration k , having the iterate $z^k = (x^k, y^k)$, we try to predict the decrease $f_0(z^k + d) - f_0(z^k)$ for a step d that satisfies the trust region bound $\|d\| \leq \Delta_k$. We use the local model

$$\psi(d) = \frac{1}{2} (\|h(z^k) + \nabla h(z^k)d\|^2 - \|h(z^k)\|^2).$$

Given a step s^k with $z^k + s^k \in R_+^{2n}$ and $\psi(s^k) < 0$, we update z^{k+1} and Δ_{k+1} , depending on how well the local model predicted the desired reduction in f_0 .

The Cauchy step \tilde{s}^k generated by the gradient projection method is used as a benchmark for accepting a candidate step s^k . It is desired to have the candidate step perform at least

as well as \tilde{s}^k and, one hopes, better in predicting the decrease in the objective function. The Cauchy step is defined as a solution to the problem

$$\min_d \{\psi_k(d) : z^k + d \in \Omega, \|d\| \leq \Delta_k\}.$$

The Cauchy step is of the form $s^k(\alpha_k)$ where

$$s^k(\alpha) = [z^k - \alpha \nabla f_0(z^k)]_+ - z^k,$$

where $[y]_+ = \max(0, y)$ componentwise and α is a suitably chosen step length.

For a candidate direction s^k we require that the following be satisfied

$$\psi_k(s^k) \leq \mu_0 \psi_k(\tilde{s}^k), \quad \text{with } \|s^k\| \leq \mu_1 \Delta_k, \quad \text{and } z^k + s^k \in R_+^{2n}.$$

This is the required improvement over the Cauchy step for a candidate direction. Clearly, taking the Cauchy step will work, but for faster convergence we need to employ a different direction from \tilde{s}^k .

One useful strategy is to base the computation of the search direction s^k on the following problem

$$\min_z \{q_k(z) : z_i = 0, i \in \mathcal{A}(z^{k,1})\}, \quad (45)$$

where $z^{k,1} = z^k + \tilde{s}^k$, $\mathcal{A}(z^{k,1})$ is the active set of indices and $q_k : R^{2n} \rightarrow R_+$ is the following convex function of z

$$q_k(z) = \frac{1}{2} \|h(z^k) + \nabla h(z^k)(z - z^k)\|^2.$$

It is not hard to see that for a fixed set of active indices; this is just a linear least squares problem.

With the iterate z^k , we can generate a suitable search direction s^k , $\mu_0 \in (0, \frac{1}{2})$, $\mu_1 > 0$, as follows. Let $z^{k,0} = z^k$ and compute l minor iterates $z^{k,1}, z^{k,2}, \dots, z^{k,l}$ to satisfy the conditions

$$\begin{aligned} z^{k,1} &= z^k + \tilde{s}^k \\ z^{k,j} &\in R_+^{2n}, \|z^{k,j} - z^k\| \leq \mu_1 \Delta_k \\ q_k(z^{k,j+1}) &\leq q_k(z^{k,j}) + \mu_0 \nabla q_k(z^{k,j})^T (z^{k,j+1} - z^{k,j}). \end{aligned} \quad (46)$$

The last condition ensures that the quadratic function is sufficiently decreased. We form the step as $s^k = z^{k,l} - z^k$. Notice that there is some flexibility as to how we obtain the minor iterates: various strategies suitable for the NEMS NCP can be employed.

To gain superlinear convergence, we need to impose stronger conditions on the direction s^k . In particular, this can be accomplished if we demand that the minor iterate $z^{k,l}$ be an approximate minimizer of q_k on the active set of indices. More specifically, let P_k be the projection operator into the subspace $\{z \in R^{2n} : z_i = 0, i \in \mathcal{A}(z^{k,l})\}$. Then we also require the following condition to be met for $\xi_k \in [0, 1]$:

$$\|\nabla_{d=0} q_k(z^{k,l} + P_k d)\| \leq \xi_k \|h(z^k)\|.$$

Clearly for $\xi_k = 0$ we are demanding that $z_{k,l}$ be an exact minimizer with respect to the active set of indices.

These inexact conditions give rise to a host of possible methods to compute the subproblem. The important idea is to use the sparse structure of the NEMS NCP in performing

the various calculations. One such method is the active set approach by Moré and Toraldo (1991) discussed in the preceding section. This method is particularly effective for NEMS because the most complicated computational step involves just a sparse matrix times a vector.

6 A Smoothing Function Approach

A recent method by Chen and Mangasarian (1994) has potential application for the NEMS NCP. The starting point is that the (pure) NCP(F) can be recast as solving the following unconstrained system of nonlinear equations. Find $x \in R^n$ such that

$$x - [x - F(x)]_+ = 0 \quad (47)$$

where $(y) = \max(0, y)$ componentwise. Since this system is nonsmooth, Newton's method cannot be used directly. However, if we approximate the function $[\cdot]_+$ by a smooth function $\hat{p}(\cdot, \beta)$, where the parameter $\beta > 0$ controls the approximation (i.e., $\beta = 0$ means an exact match), we could just apply Newton's method to

$$R(x) = x - \hat{p}(x - F(x), \beta) = 0 \quad (48)$$

and solve for small enough values of β ; this is the chosen strategy discussed in this work. An example of the function \hat{p} is the neural networks function given as follows: $\hat{p}_i : R \times R_+ \rightarrow R_+$ defined by

$$\hat{p}_i(x_i, \beta) = x_i + \beta \log(1 + e^{-x_i/\beta})$$

and $\hat{p}(x, \beta) = (\hat{p}_i(x_i, \beta), i = 1, \dots, n)$.

We additionally define the scalar function $f(x) = \frac{1}{2} \|R(x)\|^2$ and note that solving (48) is equivalent to minimizing $f(x)$ and achieving an optimal value of zero. For a fixed value of $\beta > 0$, $\nabla R(x)$ has the following form

$$\nabla R(x) = I - D(x, \beta)(I - \nabla F(x)) \quad (49)$$

where $D(x, \beta)$ is a diagonal matrix with the i th diagonal equal to $\frac{\partial p_i(x_i - F_i(x), \beta)}{\partial x_i} \in (0, 1)$. Thus, if $\nabla F(x)$ is sparse, then so is $\nabla R(x)$. This fact is particularly useful for the NEMS NCP in light of the sparsity of $\nabla F(x)$, as demonstrated above. We note that when F is monotone so that $\nabla F(x)$ is positive semidefinite, $\nabla R(x)$ is nonsingular so that (49) is nonsingular.

With an iterate x^k , the method consists of solving the linear system $\nabla R(x)d = -R(x)$ for a search direction d^k . Then, an Armijo-type search is performed for the step length λ_k to ensure that sufficient descent is made for the merit function f , in other words,

$$f(x_{k+1}) - f(x_k) \leq -\sigma \lambda_k |d_k^T \nabla f(x_k)|.$$

The smoothing parameter is then updated. This approach was successfully tested on a wide range of problems and thus appears promising for use with the NEMS NCP.

The specialization to NEMS comes down to solving the system $\nabla R(x)d = -R(x)$ efficiently. We note that when the smoothing parameter β changes, only the matrix $D(x, \beta)$ is

affected. When x is updated, $\nabla F(x)$ and $D(x, \beta)$ are affected. But the important point is that the position of the nonzeros is unchanging for each iteration. Hence, one can potentially make use of the sparse factorization of $\nabla R(x)$ from one iteration to the next.

One disadvantage with this approach is that a linear system of the size of the entire NEMS system needs to be solved for each iteration of this method. A more promising approach would be to use a method such as QMR as described in an earlier section. QMR simply uses sparse matrix-vector-type products to solve linear systems where the matrix is asymmetric and nonsingular as in (49).

7 Conclusions

In this work, we have described how the National Energy Modeling System (NEMS) can be viewed as an instance of a nonlinear complementarity problem (NCP). This perspective leads to a more general and perhaps more realistic and successful modeling format than is currently being used. We have described the details of several iterative NCP approaches specialized to the NEMS NCP, that can effectively exploit the structure of this large-scale problem.

Notes

¹ While $p_{I(i)}, q_{I(i)}$ would be notationally simpler than $p_{Ip(i)}, q_{Iq(i)}$, the former would ignore the fact that in general, $Ip(i) \neq Iq(i)$.

² In actuality, there are several linear programs in this module. However, for our purposes, we can consider consolidating them into one or just focus on the one that is relevant for calculating equilibrium prices and quantities. This strategy will be adopted for other modules as well, that is, just one (possibly consolidated) linear program for each relevant module.

³ The quantities from the petroleum module are aggregated at the level of five PADDS. However, before being used in NGTDM, they are run through the demand modules, which convert them to the level of nine census regions.

⁴ There are firm and interruptible customers in the natural gas market. The former type reserve a certain amount of gas and consequently pay a premium for this. The latter type are not guaranteed the gas and thus pay less.

⁵ The heuristic concerns finding a suitable pollution penalty parameter and estimating inter-regional electricity trade and could be incorporated into a consolidated linear programming formulation.

⁶ For notational simplification only, we will assume that no equality constraints are present. The inclusion of such constraints does not change the arguments to be presented, but without them things are somewhat simplified in that only pure rather than mixed NCPs need

be considered.

⁷ $L_j^i(\cdot)$ denotes the j th component of the i th linking function L^i ; a similar definition holds for \hat{L}^i .

⁸ For completeness, we have assumed that every NLP is involved with producing prices or quantities from its solution and multiplier vectors. In actuality, not every NLP may do so but this is not restrictive.

⁹ The actual QP subproblem does not require those components i where $x_i = f_i(x) = 0$. Hence, after eliminating these components, we end up with a QP of the form as shown.

¹⁰ Note: for $u, v \in R_+^n$, there exists a $b > 0$ such that $b(u^T v) \geq \|\min(u, v)\|^2 \geq 0$ so that reducing the complementarity gap forces condition (c) to be satisfied. To see this, let $S = \{i : v_i < u_i\}$. Then we have

$$0 \leq \|\min(u, v)\|_2^2 = \sum_{i \in S} v_i^2 + \sum_{i \in S^c} u_i^2 \leq u^T v.$$

Hence, by the equivalence of vectors norms, we have the desired result.

¹¹ This is done to ensure the nonsingularity of B_{ii} given that A_{ii} is symmetric positive semidefinite for all i .

¹² The subscript i refers to the i th block of variables as was shown in (25).

¹³ We define $v_j^0 = s_j^\nu$ for $j = 1, \dots, b$.

¹⁴ In actuality, the i, j element may not use the i th and j th columns of M since \tilde{B} is a principal submatrix of $B + D + I$, but there is no loss of generality in using this form.

¹⁵ Of course, other active set approaches involving mostly just matrix vector products are also of potential interest. However, for concreteness, we have decided to focus on the particular active set approach in Moré and Toraldo (1991).

¹⁶ Some active set methods allow only one index to be picked up per iteration. This constraint could slow down the subproblem calculations for a large-scale problem such as NEMS.

¹⁷ $P(d) = \text{mid}(l, d, u)$, where mid is the componentwise median operator.

The research for Dr. Gabriel was conducted while a member of the Mathematical Sciences and Geography and Environmental Engineering Departments at The Johns Hopkins University in Baltimore, Maryland, and the Mathematics and Computer Science Division at

Argonne National Laboratory in Argonne, Illinois. All correspondence should be addressed to Dr. Gabriel.

This paper was based on research supported by (1) the U.S. Department of Energy, contract #DE-AC01-92EI21946 with Decision Analysis Corporation of Virginia, and (2) the Office of Scientific Computing, U.S. Department of Energy, under Contract W-31-109-Eng-38.

Disclaimer

This product was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe on privately owned rights. Reference herein to any specific commercial product, process or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Acknowledgments

We thank Gail Pieper of the Mathematics and Computer Science Division of Argonne National Laboratory for her useful editorial suggestions. Also, we thank the following people for various insightful discussions concerning this work: the NEMS modeling team at the U.S. Department of Energy, Professor Fred Murphy of Temple University, Professors Alan Goldman and Jong-Shi Pang of Johns Hopkins University, and Drs. Joe Czyzyk, Stephen Wright, and Jorge Moré of Argonne National Laboratory; in particular, Dr. Moré brought to our attention several useful works.

REFERENCES

- Ahn, B. H. 1979. *Computation of Market Equilibria for Policy Analysis*, Garland Publishing, Inc., New York.
- Ahn, B. H., and W. W. Hogan. 1982. On Convergence of the PIES Algorithm for Computing Equilibria. *Opns. Res.* **30**, 281–300.
- Barrett, R. , M. Berry, T. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. van der Vorst. 1994. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, Society for Industrial and Applied Mathematics, Philadelphia.
- Burke, J. V., J. J. Moré, and G. Toraldo. 1990. Convergence Properties of Trust Region Methods for Linear and Convex Constraints. *Math. Prog.*, **47**, 305–336.

- Chen, B. T., and P. T. Harker. 1993.** A Non-Interior Point Continuation Method for Linear Complementarity Problems. *SIAM J. Matrix Analysis and Applications*. **20**, 1168–1190.
- Chen, C. H., and O. L. Mangasarian. 1994.** A Class of Smoothing Functions for Non-linear and Mixed Complementarity Problems. *Mathematical Programming Technical Report 94-11*, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin.
- Cottle, R. W., J. S. Pang, and R. E. Stone. 1992.** *The Linear Complementarity Problem*, Academic Press, Inc., San Diego.
- Dirkse, S. P. and M. C. Ferris. 1994.** A Pathsearch Damped Newton Method for Computing General Equilibria. *Mathematical Programming Technical Report # 94-03*, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin.
- Dirkse, S. P. and M. C. Ferris. 1995.** The PATH Solver: A Non-Monotone Stabilization Scheme for Mixed Complementarity Problems. *Optimization Methods & Software*, forthcoming.
- Duff, I. S., A. M. Erisman, and J. K. Reid. 1986.** *Direct Methods for Sparse Matrices*, Oxford University Press, Oxford.
- Energy Information Administration. 1994.** *The National Energy Modeling System: An Overview* Office of Integrated Analysis and Forecasting, U.S. Department of Energy, Washington, D.C.
- Frank, M. and P. Wolfe. 1956.** An Algorithm for Quadratic Programming. *Nav. Res. Log. Quart.* **3**, 95–110.
- Freund, R. W., and N. M. Nachtigal. 1991.** QMR: A Quasi-Minimal Residual Method for Non-Hermitian Linear Systems. *Numer. Math.* **60**, 315–339.
- Freund, R. W., and N. M. Nachtigal. 1994.** An Implementation of the QMR Method Based on Coupled Two-Term Recurrences. *SIAM J. Sci. Comp.* **15**, 313–337.
- Gabriel, S. A. 1993.** Literature Survey for Solving Convergence Issues in NEMS Final Report Subtask 1. Technical report prepared for *Energy Information Administration, U.S. Department of Energy* in conjunction with *Decision Analysis Corporation of Virginia* Contract # DE-AC01-92EI21946, Vienna, Virginia.
- Gabriel, S. A., and J. S. Pang. 1992.** An Inexact NE/SQP Method for Solving the Nonlinear Complementarity Problem. *Comp. Opt. and Appl.* **1**, 67–91.
- Gabriel, S. A., and J. S. Pang. 1994.** A Trust Region Method for Constrained Nonsmooth Equations, in *Large-Scale Optimization: State of the Art*, edited by W. W. Hager, D. W. Hearn. and P. M. Pardalos, Kluwer Academic Publishers, 159–186.
- Harker, P. T., and J. S. Pang. 1990.** Finite-Dimensional Variational Inequality and Non-linear Complementarity Problems: A Survey of Theory, Algorithms and Applications. *Math. Prog.* **48**, 161–220.

- Hogan, W. W. 1975.** Energy Policy Models for Project Independence. *Comp. & Oper. Res.* **2**, 251–271.
- Luo, Z. H., and P. Tseng. 1991.** On the Convergence of a Matrix Splitting Algorithm for the Symmetric Monotone Linear Complementarity Problem. *SIAM J. Control and Optim.* **29**, 1037–1060.
- Moré, J. J. 1994.** Global Methods for Nonlinear Complementarity Problems. Mathematics and Computer Science Division, preprint MCS-P429-0494, Argonne National Laboratory, Argonne, Illinois.
- Moré, J. J., and G. Toraldo. 1991.** On the Solution of Large Quadratic Programming Problems with Bound Constraints, *Siam J. Opt.* **1**, 93–113.
- Murphy, F. H. 1983.** An Overview of the Intermediate Future Forecasting System. in *Ener. Mod. and Sim.*, edited by A. S. Kydes et al., North Holland Publishing Company, 66–73.
- Murphy, F. H. 1993.** Making Large-Scale Models Manageable: Modeling from an Operations Management Perspective. *Oper. Res.* **41**, 241–252.
- Murphy, F. H., J. J. Conti, S. H. Shaw, and R. Sanders. 1988.** Modeling and Forecasting Energy Markets with the Intermediate Future Forecasting System. *Oper. Res.* **36**, 406–420.
- Ortega, J. M., and W. C. Rheinboldt. 1970.** *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, Inc., New York.
- Pang, J. S., and S. A. Gabriel. 1993.** NE/SQP: A Robust Algorithm for the Nonlinear Complementarity Problem. *Math. Prog.* **60**, 295–337.
- Ralph, D. 1994.** Global Convergence of Damped Newton’s Method for Nonsmooth Equations, Via the Path Search. *Math. of O.R.* **19**, 352–389.
- Wright, S. J. 1995.** A Path-Following Interior-Point Algorithm for Linear and Quadratic Programs. Mathematics and Computer Science Division, preprint MCS-P401-1293, Argonne National Laboratory, Argonne, Illinois.
- Zhang, Y. 1994.** On the Convergence of a Class of Infeasible Interior-Point Methods for the Horizontal Linear Complementarity Problem. *SIOPT* **4**, 208–227.