Coupled Optimization Models for Planning and Operation of Power Systems on Multiple Scales

Michael C. Ferris

November 30, 2010, Revised February 2011

- Decision processes are predominantly hierarchical. Models to support such decision processes should also be layered or hierachical.
- Coupling collections of (sub)-models with well defined (information sharing) interfaces facilitates:
  - appropriate detail and consistency of sub-model formulation (each of which may be very large scale, of different types (mixed integer, semidefinite, nonlinear, variational, etc) with different properties (linear, convex, discrete, smooth, etc))
  - ability for individual subproblem solution verification and engagement of decision makers
  - ability to treat uncertainty by stochastic and robust optimization at submodel level and with evolving resolution
  - ability to solve submodels to global optimality (by exploiting size, structure and model format specificity)

  (A monster model that mixes several modeling formats loses its ability to exploit the underlying structure and provide guarantees on solution quality)
- Developing interfaces and exploiting hierarchical structure using computationally tractable algorithms will provide overall solution speed, understanding of localized effects, and value for the coupling of the system.

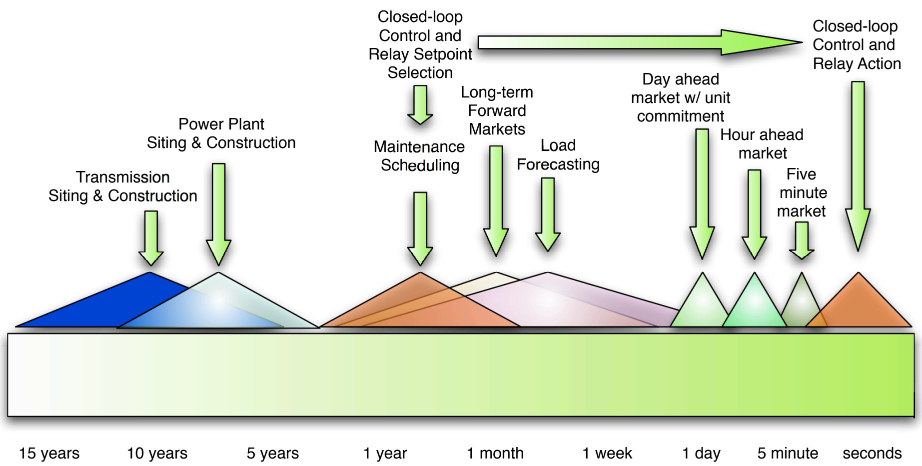

Figure 1: Representative decision-making timescales in electric power systems

1 Problem Hierarchies and Timescales

Technological and economic trends imply significant growth in our nation's reliance on the power grid in the coming decades; well-accepted estimates cite 35% growth in electricity demand over the next 20 years [60]. Planning and operating the Next Generation Electric Grid involves decisions ranging from time scales of perhaps 15 years, for major grid expansion, to time scales of 5-minute markets, and must also account for phenomena at time scales down to fractions of a second. A representation of the decision process over timescales of interest is shown in Figure 1. What makes this setting particularly interesting is that behaviors at very fast time scales (e.g., requirements for grid resilience against cascading failures) potentially impose constraints on longer time scale decisions, such as maintenance scheduling and grid expansion. We argue here against building a single "monster model" that tries to capture all these scales, but propose using a collection of coupled or layered models for both planning and operation, interfacing via information/solution sharing over multiple time scales and layers of decision making. Such approaches have been successful in other application domains [18].

In addition to the multiple time scales in the decision process, the problem is confounded by uncertainties in estimates and structural makeup of the system. For example, plug hybrid electric vehicles are a visible technology that could dramatically alter the patterns, nature, and quantity of U.S. electricity use, and yet the ultimate market penetration of such technology is highly uncertain. Similarly, future grid penetration for non-traditional energy sources such as wind and solar, and for carbon-sequestration-equipped coal plants, also remains highly uncertain. These structural uncertainties present profound challenges to decision methodologies, and to the optimization tools that inform them. While traditional optimization approaches might seek to build a large-scale model that combines all instances together, such approaches are impractical as the size and ranges of the spatial and temporal scales expand, let alone treating the uncertainties that are inherently present in these decision problem settings.

As one moves between time scales, some of that uncertainty gets resolved, and some new uncertainties become relevant. The decision problems need to capture that uncertainty and allow a decision maker the flexibility to structurally change the system to the new environment. All encompassing models are typically not nimble enough to facilitate adaptation of the decisions as the real process evolves (both structurally and data-wise). Thus, our thesis is not simply about solution speed, but hinges on the added value that arises from modeling and solution in a structured (and better scaled and theoretically richer) setting.

The decision timeline of Figure 1 is intended to highlight the severe challenges the electric power environment presents. As an example of coupling of decisions across time scales, consider decisions related to the siting of major interstate transmission lines. These require economic forecasts, supply and demand forecasts, and an interplay between political and engineering concerns. Typically, relatively few possible choices are available – not only due to engineering or even economic constraints – but arising from public and political concerns that are often hard to justify rationally, but severely limit the possible layouts. Models that demonstrate the benefits and drawbacks of a particular siting decision at an aggregate level are of critical importance for informing discussions and decision makers. The key issue is to facilitate appropriate aggregations (or summarizations of details irrelevant to the decision at hand) that enable a quick, even interactive, and thorough exploration of the actual decision space. It is, and will remain, a significant challenge to identify and manage the interface between a given model and the other models that are connected – from a conceptual, modeling and computational viewpoint.

Note that transmission expansion decisions influence the capital investment decisions made by generating companies, again at a 1-10 year time scale. Much of the same data used for transmission expansion is pertinent to these models, but the decisions are made by independent agents without overall system control, so different types of models (game theoretic for example) more readily capture the decision process here. Specific decision models are typically governed by a overriding principle and can be formulated using the most appropriate modeling tools. The monster model is more likely to be a conglomeration of multiple principles, and becomes unmanageable, intractable and hard to understand the driving issues.

New generation capabilities subsequently affect bids into the power market which are then balanced using economic and reliability objectives on a day-ahead or 5-minute time scale. At this level, models are needed for electric pricing and market control to determine which units are to be deployed and at what price and quantity, accounting for the uncertainties in new forms of energy provision such as wind and solar. Such planning, deployment and commitment of specific resources must be carried out to ensure both operation reserves are sufficient, and to provide robust solutions for these choices that ensure security of the overall system (when confronted with the vast number of uncertainties that can confound the efficient operation of the electric grid). Note that long term planners will not be able to accurately forecast all the structural (and potentially disruptive) changes to the system, and thus wind speed and weather patterns at fine scales are largely irrelevant – efficient sampling and (automated) information aggregation are key to allow informed decision making at widely different time scales.

Finally, power grid dynamics are operating at the millisecond to minutes time scales and involve decisions for settings of protective relays that remove lines and generators from service when operating thresholds are exceeded to guard against cascading failures. At this level, efficient nonlinear optimization must be carried out to match the varying demand for electricity with the ever increasing and uncertain supply of energy, without interruptions or catastrophic cascading failures of the system. While the underlying question may well be "Is there a better choice for the transmission line expansion to reduce the probability of a major blackout?", it is contended here that the additional knowledge gained from understanding the effects of one decision upon another in a structured fashion will facilitate better management and operation of the system when it is built and provide understanding and information to the operators as to the consequences of their decisions.

In addition to this coupling across time scales, one has the challenge of structural differences amongst classes of decision makers and their goals. At the longest time frame, it is often the Independent System Operator (ISO), in collaboration with Regional Transmission Organizations (RTO) and regulatory agencies, that are charged with the transmission design and siting decisions. These decisions are in the hands of regulated monopolies and their regulator. From the next longest time frame through the middle time frame, the decisions are dominated by capital investment and market decisions made by for-profit, competitive generation owners. At the shortest time frames, key decisions fall back into the hands of the Independent System Operator, the entity typically charged with balancing markets at the shortest time scale (e.g., day-ahead to 5-minute ahead), and with making any out-of-market corrections to maintain reliable operation in real time.

Each of these problems involves coordinating a large number of decision making agents in an uncertain environment. The computational needs for such solutions are immense and will require both modeling sophistication and decompostion methodology to exploit problem structure, and a large array of computing devices – whose power is seamlessly provided and available to critical decision makers (not just optimization or computational science experts) – to process resulting subproblems. Subsets of these subproblems may be solved repeatedly when resulting information in their interfaces changes. Model updates can then be coupled to evolving information flow. Quite apart from the efficiency gains achieved, the smaller coupled modelds are more easily verified by their owners, and otherwise hidden deficiencies of a monster model formulation are quickly detected and fixed.

Traditional optimization approaches are no longer effective in solving the practical, large scale, complex problems that require robust answers in such domains. While the study of linear programming, convex optimization, mixed integer and stochastic programming have in themselves led to significant advances in our abilities to solve large scale instances of these problems, typical application problems such as those outlined above require a sophisticated coupling of a number of these approaches with specific domain knowledge and expertize to generate solutions in a timely manner that are robust to uncertainties in an operating environment and in the data that feeds the model. Rather than attempting to model all these features together, we propose a methodology that utilizes layering and information sharing interfaces between collections of models, that allows decisions to be made using appropriately scaled problems, each of which approximates external features by aggregate variables and constraints. It could be argued that by using a collection of coupled models, we are leaving some optimization possibilities "on the table". Clearly, poorly defined interfaces will have this issue. The challenge for modelers and algorithms is to define these interfaces correctly, manage them automatically, understand the hierarchy of decision makers and match this to the model, thereby facilitating solution of the overall system by processing of (modified) problems at each level.

In short, there is clearly a need for optimization tools that effectively inform and integrate decisions across widely separated time scales, by different agents who have differing individual objectives, in the presence of uncertainty.

2 Motivating Problems

The purpose of the electric power industry is to generate and transport electric energy to consumers [56]. At time frames beyond those of electromechanical transients (i.e. beyond perhaps, 10's of seconds), the core of almost all power system representations is a set of equilibrium equations known as the power flow model. This set of nonlinear equations relates bus (nodal) voltages to the flow of active and reactive power through the network and to power injections into the network. With specified load (consumer) active and reactive powers, generator (supplier) active power injections and voltage magnitude, the power flow equations may be solved to determine network power flows, load bus voltages, and generator reactive powers. Current research is still ongoing to reliably solve these equations; approaches involve Newton based methods and techniques from semidefinite programming.

At the next level of sophistication, an Optimal Power Flow (OPF) can be used to determine least cost generation dispatch, subject to physical grid constraints such as power flow equations, power line flow limits, generator active and reactive power limits, and bus voltage limits. Typically the problem is solved by the ISO, and is characterized by a number of different methods that generator firm's can make bids to supply electricity. For simplicity here, we assume that bids are characterized by a decision variable α, resulting in the following optimization problem:

OPF(α):	$\min_q$	energy dispatch cost (q, α)
	s.t.	conservation of power flow at nodes
		Kirchoff's voltage law, and simple bound constraints

Note that since α are (given) price bids, this problem is a parametric optimization for dispatch quantities q. We assume this problem has a unique solution for each α for ease of exposition.

Each generator firm i has to determine its bid α_i. Assuming no generator has market power (perhaps an unreasonable assumption), the problem faced by firm i is

$$\begin{array}{lll} \text{Bid}(\bar{\alpha}_{-i}): & \max_{\alpha_i,q,p} & \text{firm } i\text{'s profit } (\alpha_i, q, p) \\ & \text{s.t.} & 0 \le \alpha_i \le \hat{\alpha}_i \\ & & q \text{ solves OPF}(\alpha_i, \bar{\alpha}_{-i}) \end{array}$$

where the objective function involves the multiplier p determined from the OPF problem. This multiplier is not exposed to the decision maker. To overcome this issue, we can replace the lower level optimization problem by its first order (KKT) conditions and thus expose the multipliers directly to the upper level optimization problem:

$$\begin{array}{lll} \text{Bid}(\bar{\alpha}_{-i}): & \max_{\alpha_i,q,p} & \text{firm } i\text{'s profit } (\alpha_i, q, p) \\ & \text{s.t.} & 0 \le \alpha_i \le \hat{\alpha}_i \\ & & q, p \text{ solves KKT(OPF}(\alpha_i, \bar{\alpha}_{-i})) \end{array}$$

This process takes a bilevel program and converts it to a mathematical program with complementarity constraints (MPCC) since the KKT conditions form what are called complementarity constraints. We outline in the sequel methods to write down and solve problems of this form, but note that they are computationally difficult, and theoretically the MPCC is hard due to the lack of a constraint qualification. It may even be the case that this transformation is incorrect: the KKT may not be necessary and sufficient for global optimality of the lower level problem.

This problem is a single firm's problem. Adventurous modelers require further conditions, in that the firms collectively should have no incentive to change their bids in equilibrium: $(\bar{\alpha}_1, \bar{\alpha}_2, \ldots, \bar{\alpha}_m)$ is an equilibrium if

$$\bar{\alpha}_i \text{ solves Bid}(\bar{\alpha}_{-i}), \quad \forall i.$$

This is an example of a (Nonlinear) Nash Equilibrium where each player solves an MPCC. It is known that such a Nash Equilibrium is PPAD-complete [17, 19]. While complexity results of this nature lead to an appreciation of the extreme difficulty of the underlying problem, it is clear that such problems must be solved (repeatedly) for effective operation of the power system.

Unfortunately, in practice, a Security Constrained Optimal Power Flow (SCOPF) adds the additional constraint that the solution for powers and voltages must remain within limits for a user-specified set of contingencies (scenarios) [57, 70]. To some extent, this is a simplification made to the problem to gain tractability. Even so, such problems are currently beyond the state of the art for solution methodologies. We outline an extended mathematical programming (EMP) framework that allows such problems to be written down. We firmly believe that the underlying structure in these models will be necessary to exploit for any realistic solution method to be successful.

The constraints in the OPF and SCOPF problems make them more difficult to solve [80], and some programs use simplified models to quickly "solve" these equations. A common simple model is the so-called DC power flow, which is a simple linearized form of the true power flow equations. The industry uses this form extensively. However, "proxy constraints" are often added to the formulations to recapture effects that linearization (or approximations) lose. This is fraught with danger and possibilities for exploitation (of the difference in the approximate model and the nonlinear physics) when such constraints are used for pricing in markets for reactive power, for example. Nonlinear models are necessary and can and should be reliably solved at appropriate levels of detail.

Another example that motivates our work is the notion of transmission line switching [31, 41]. In this setting, an optimization problem of the form:

$$
\begin{array}{lll}
\min_{g,f,\theta} & c^T g & \text{generation cost} \\
\text{s.t.} & g - d = Af, f = BA^T\theta & A \text{ is node-arc incidence} \\
 & \bar{\theta}_L \leq \theta \leq \bar{\theta}_U & \text{bus angle constraints} \\
 & \bar{g}_L \leq g \leq \bar{g}_U & \text{generator capacities} \\
 & \bar{f}_L \leq f \leq \bar{f}_U & \text{transmission capacities}
\end{array}
$$

can be solved to determine the flows and generation with a simplified DC power flow model. Transmission switching is a design philosophy that allows a subset of the lines to be opened to improve the dispatch cost. The additional discrete choice is whether the line i is open or closed and can be modeled using the following disjunction:

$$
\begin{array}{lll}
\min_{g,f,\theta} & c^T g & \\
\text{s.t.} & g - d = Af & \\
 & \bar{\theta}_L \leq \theta \leq \bar{\theta}_U & \\
 & \bar{g}_L \leq g \leq \bar{g}_U & \\
\text{either} & f_i = (BA^T\theta)_i, \bar{f}_{L,i} \leq f_i \leq \bar{f}_{U,i} & \text{if } i \text{ closed} \\
\text{or} & f_i = 0 & \text{if } i \text{ open}
\end{array}
$$

This disjunction is not a typical constraint for an optimization problem, but can be directly modeled in EMP. The framework allows automatic problem reformulations - in the above case, this can generate mixed integer programming problems, or indeed nonlinear mixed integer programs when the linearized DC model is replaced by the full AC model.

The final example concerns the transmission line expansion outlined in the introduction [48]. We suggest considering a hierarchical approach to this problem formulated as follows. If we let x represent the transmission line expansion decision and assume the RTO can postulate a (discrete) distribution of future demand scenarios (at the decade level scale), then the RTO problem is:

$$\min_{x \in X} \sum_{\omega} \pi_{\omega} \sum_{i \in N} d_i^{\omega} p_i^{\omega}(x)$$

where ω runs over the scenarios, π are the probabilities, and d_i^{ω} is the resulting demand in such a scenario at a given node in the network. The constraints $x \in X$ reflect budgetary and other constraints on the RTO's decision, and the function $p_i^{\omega}(x)$ is a response price in the given scenario to the expansion by x. Clearly, the key to solving this problem is to generate a good approximation to this response price since this is the interface to the lower levels of the hierarchy. We believe a lower level equilibrium model involving both generator firms and OPF problem solution in every scenario is one way to generate such a response. Optimization techniques based on derivative free methodology, or noisy function optimization may be a practical way to solve the RTO problem, requesting evaluations of this response price function. Alternatively, automatic differentiation could play a role in generating derivatives for the response price function, but that may require techniques to deal with its inherent nonsmoothness.

As outlined above, the transmission line expansion is likely to foster generator expansion. For each firm f, we denote the generator expansion by y_f and propose that this will be determined by an optimization principle:

$$\min_{y_f \in Y_f} \sum_{\omega} \pi_{\omega} \sum_{j \in F_f} c_j(y_j, q_j^{\omega})$$

Here F_f denotes the generators in firm f's portfolio, and Y_f represents budgetary and other constraints faced by the generator firm. Each firm thus expends its budget to minimize the expected cost of supply. Note that q_j^{ω} is a parameter to this problem - the actual dispatch is determined by a scenario dependent OPF problem:

$$\forall \omega \quad \min_{z,\theta,q} \sum_{j \in F_f} c_j(y_j, q_j)$$

which is subject to flow balance constraints, line data constraints, line capacity constraints, generator capacity constraints and regulatory constraints. The multiplier on the flow balance constraints is $p_i^\omega(x)$. This problem may involve integer variables as well to model switching or commitment features of the problem, but that could raise issues of the integrity of the multipliers.

Note that the collection of all these optimization models (generator expansion and scenario dependent OPF) forms an equilibrium problem, once each optimization model is replaced by its KKT conditions. The specific interface between them is defined by the variables y and q. The equilibrium problem determines all the variables in all the models at one time, whereas the models assume price taking behavior or knowledge of generator expansions in parameteric form. (In fact, this is an example of an embedded complementarity system, details of which follow in the sequel.) In this case, the equilibrium problem may be replaced by a large scale optimization problem. This fact is useful in formally proving convergence of a decomposition algorithm that iteratively solves the small optimization problems and updates the linking variables in a Jacobi sense. All of this then generates the response price $p_i^\omega(x)$ for a given transmission expansion x in a computationally tractable way and allows extension to power system models at the regional or national scale. Other models that can be captured by these concepts include the following: [47, 46, 4, 61].

3 Extended Mathematical Programs

We believe that the design, operation and ennhancement of the Next Generation Electric Grid will rely critically on tools and algorithms from optimization. Accessing these optimization solvers, and many of the other algorithms that have been developed over the past three decades has been made easier by the advent of modeling languages. A modeling language [11, 33] provides a natural, convenient way to represent mathematical programs and provides an interface between a given model and multiple different solvers for its solution. The many advantages of using a modeling language are well known. They typically have efficient automatic procedures to handle vast amounts of data, take advantage of the numerous options for solvers and model types, and can quickly generate a large number of models. For this reason, and the fact that they eliminate many errors that occur without automation, modeling languages are heavily used in practical applications. Although we will use GAMS [13] in our descriptions here, much of what will be said could as well be applied to other algebra based modeling systems like AIMMS [10], AMPL [34], MOSEL, MPL [55] and OPL [75].

While much progress has also been made in developing new modeling paradigms (such as stochastic and robust programming, mixed integer nonlinear optimization, second order cone programming, and optimization of noisy functions), the ability for application experts to utilize these advances from within modeling systems has remained limited. The purpose of this work is to extend the classical problem from the traditional optimization model:

$$\min_x f(x) \text{ s.t. } g(x) \leq 0,\ h(x) = 0, \tag{1}$$

where f, g and h are assumed sufficiently smooth, to a more general format that allows new constraint types and problem features to be specified precisely. The extended mathematical programming (EMP) framework exists to provide these same benefits for problems that fall outside the classical framework [26]. A high-level description of these models in an algebraic modeling language, along with tools to automatically create the different realizations or extensions possible, pass them on to the appropriate solvers, and interpret the results in the context of the original model, makes it possible to model more easily, to conduct experiments with formulations otherwise too time-consuming to consider, and to avoid errors that can make results meaningless or worse.

We believe that further advancements in the application of optimization to electricity grid problems can be best achieved via identification of specific problem structures within planning and operational models, coupled with automatic reformulation techniques that lead to problems that are theoretically better defined and more ameable to rigorous computation. The ability to describe such structures in an application domain context will have benefits on several levels. Firstly, we think this will make the modelers task easier, in that the model can be described more naturally and (for example) soft or probabilistic constraints can be expressed explicitly. Secondly, if an algorithm is given additional structure, it may be able to exploit that in an effective computational manner; indeed, the availability of such structures to a solver may well foster the generation of new features to existing solvers or drive the development of new classes of algorithms. Specific structures that we believe are relevant to this application domain include mathematical programs with equilibrium constraints, second order cone programs (that facilitate the use of "robust optimization" principles), semidefinite programming, bilevel and hierarchical programs, extended nonlinear programs (with richer classes of penalty functions) and embedded optimization models. The EMP framework provides an extensible way to utilize such features.

Some extensions of the traditional format have already been incorporated into modeling systems. There is support for integer, semiinteger, and semicontinuous variables, and some limited support for logical constructs including special ordered sets (SOS). GAMS , AMPL and AIMMS have support for complementarity constraints [29, 28], and there are some extensions that allow the formulation of second-order cone programs within GAMS. AMPL has specific syntax to model piecewise linear functions. Much of this development is tailored to particular constructs within a model. We describe the development of the more general EMP annotation schemes that allow extended mathematical programs to be written clearly and succinctly.

3.1 Complementarity Problems

The EMP framework allows annotation to existing functions and variables within a model. We begin with the example of complementarity, which in its simplest form, is the relationship between nonnegative variables with the additional constraint that at least one must be zero. A variety of models in electricity markets use complementarity at their core, including [43, 44, 45, 54, 58, 69, 81, 84, 83].

A first simple example are the necessary and sufficient optimality conditions for the linear program

$$\begin{aligned} &\min_x \; c^T x \\ &\text{s.t. } \; Ax \geq b, \; x \geq 0 \end{aligned} \tag{2}$$

which state that x and some λ satisfy the complementarity relationships:

$$\begin{aligned} 0 \leq c - A^T\lambda \quad &\perp \quad x \geq 0 \\ 0 \leq Ax - b \quad &\perp \quad \lambda \geq 0. \end{aligned} \tag{3}$$

Here, the "$\perp$" sign signifies (for example) that in addition to the constraints $0 \leq Ax - b$ and $\lambda \geq 0$, each of the products $(Ax - b)_i\lambda_i$ is constrained to be zero. An equivalent viewpoint is that either $(Ax - b)_i = 0$ or $\lambda_i = 0$, a disjunction. Within GAMS, these constraints can be modeled simply as

```
positive variables lambda, x;
model complp / defd.x, defp.lambda /;
```

where defp and defd are the equations that define general primal and dual feasibility constraints ($Ax \geq b$, $c \geq A^T\lambda$) respectively.

Complementarity problems do not have to arise as the optimality conditions of a linear program; the optimality conditions of the nonlinear program (1) constitute the following MCP:

$$\begin{array}{lcl} 0 = \nabla f(x) + \lambda^T \nabla g(x) + \mu^T \nabla h(x) & \perp & x \text{ free} \\ 0 \leq -g(x) & \perp & \lambda \geq 0 \\ 0 = -h(x) & \perp & \mu \text{ free.} \end{array} \tag{4}$$

Many examples are no longer simply the optimality conditions of an optimization problem. The paper [30] catalogues a number of other applications both in engineering and economics that can be written in a similar format.

It should be noted that robust large scale solvers exist for such problems; see [29] for example, where a description is given of the PATH solver.

3.2 Disjunctive Constraints

A simple example to highlight disjunctions is the notion of an ordering of tasks, namely that either job i comes before job j or the converse. Such a disjunction can be specified using an annotation:

```
disjuncton * seq(i,j) else seq(j,i)
```

In such an example, one can implement a Big-M method, employ indicator variables or constraints, or utilize a convex hull reformulation.

In fact, there is a growing literature on reformulations of mixed integer nonlinear programs that describe new convex hull descriptions of structured constraint sets. This work includes disjunctive cutting planes [71], Gomory cuts [16] and perpective cuts and reformulations [35, 39].

More complicated (nonlinear) examples make the utility of this approach clearer. The design of a multiproduct batch plan with intermediate storage described in [77] and a synthesis problem involving 8 processes from [74] are included in the EMP model library. As a final example, the gasoline emission model outlined in [36] is precisely in the form that could exploit the features of EMP related to (nonlinear) disjunctive programming. Finally, the work by

Grossmann and colleagues on generalized disjunctive programming [74, 77, 78] involves both nonlinear equations and optimization primitives coupled with pure logic relations; this has been used extensively in the synthesis and design of process networks.

3.3 Mathematical Programs with Complementarity Constraints

A mathematical program with complementarity constraints embeds a parametric MCP into the constraint set of a nonlinear program as indicated in the following problem:

$$\min_{x \in \mathbf{R}^n, y \in \mathbf{R}^m} f(x, y) \tag{5}$$

$$\text{s.t. } g(x, y) \leq 0 \tag{6}$$

$$0 \leq y \;\perp\; h(x, y) \geq 0. \tag{7}$$

The objective function (5) needs no further description, except to state that the solution techniques we are intending to apply require that f (g and h) are at least once differentiable, and for many modern solvers twice differentiable.

The constraints that are of interest here are the complementarity constraints (7). Essentially, these are parametric constraints (parameterized by x) on the variable y, and encode the structure that y is a solution to the nonlinear complementarity problem defined by $h(x, \cdot)$. Within the GAMS modeling system, this can be written simply and directly as:

```
model mpecmod / deff, defg, defh.y /;
option mpec=nlpec;
solve mpecmod using mpec minimizing obj;
```

Here it is assumed that the objective (5) is defined in the equation deff, the general constraints (6) are defined in defg and the function h is described by defh. The complementarity relationship is defined by the bounds on y and the orthogonality relationship shown in the model declaration using ".". AMPL provides a slightly different but equivalent syntax for this, see [28]. The problem is frequently called a mathematical program with complementarity constraints (MPCC). Section 2 provided a specific example.

Some solvers can process complementarity constraints explicitly. In many cases, this is achieved by a reformulation of the constraints (7) into the classical nonlinear programming form given as (1). The paper [37] outlines a variety of ways to carry this out, all of which have been encoded in a solver package called NLPEC. Similar strategies are outlined in section 3 of [6]. While there are large numbers of different reformulations possible, the following parametric approach, coupled with the use of the nonlinear programming solver CONOPT or SNOPT, has proven effective in a large number of applications:

$$\begin{aligned} \min_{x \in \mathbf{R}^n, y \in \mathbf{R}^m, s \in \mathbf{R}^m} \; & f(x,y) \\ \text{s.t. } \; & g(x,y) \le 0 \\ & s = h(x,y) \\ & y \ge 0, s \ge 0 \\ & y_i s_i \le \mu, \quad i = 1, \ldots, m. \end{aligned}$$

Note that a series of approximate problems are produced, parameterized by $\mu > 0$; each of these approximate problems have stronger theoretical properties than the problem with $\mu = 0$ [62]. A solution procedure whereby μ is successively reduced can be implemented as a simple option file to NLPEC, and this has proven very effective. Further details can be found in the NLPEC documentation [37]. The approach has been used to effectively optimize the rig in a sailboat design [79] and to solve a variety of distillation optimization problems [6]. A key point is that other solution methodology may work better with different reformulations – this is the domain of the algorithmic developer and should remain decoupled from the model description. NLPEC is one way to facilitate this.

It is also possible to generalize the above complementarity condition to a mixed complementarity condition; details can be found in [27]. Underlying the NLPEC "solver package" is an automatic conversion of the original problem into a standard nonlinear program which is carried out at a scalar model level. The technology to perform this conversion forms the core of the codes that we use to implement the model extensions herein.

3.4 Variational Inequalities

A variational inequality VI(F, X) is to find $x \in X$:

$$F(x)^T (z - x) \ge 0, \text{ for all } z \in X.$$

Here X is a closed (frequently assumed convex) set, defined for example as

$$X = \{x \mid x \geq 0, h(x) \leq 0\}. \tag{8}$$

Note that the first-order (minimum principle) conditions of a nonlinear program

$$\min_{z \in X} f(z)$$

are precisely of this form with $F(x) = \nabla f(x)$. For a concrete example, note that these conditions are necessary and sufficient for the optimality of a linear programming problem: solving the linear program (2) is equivalent to solving the variational inequality given by

$$F(x) = c, \quad X = \{x \mid Ax \geq b, x \geq 0\}. \tag{9}$$

In this case, F is simply a constant function. While there are a large number of instances of the problem that arise from optimization applications, there are many cases where F is not the gradient of any function f. For example, asymmetric traffic equilibrium problems have this format, where the asymmetry arises for example due to different costs associated with left or right hand turns. A complete treatment of the theory and algorithms in this domain can be found in [25].

Variational inequalities are intimately connected with the concept of a normal cone to a set S, for which a number of authors have provided a rich calculus. Instead of overloading a reader with more notation, however, we simply refer to the seminal work in this area, [67]. While the theoretical development of this area is very rich, the practical application has been somewhat limited. The notable exception to this is in traffic analysis, see for example [40].

It is well known that such problems can be reformulated as complementarity problems when the set X has the representation (8) by introducing multipliers λ on the constraints h:

$$\begin{array}{lcl} 0 \leq F(x) + \lambda^T \nabla h(x) & \perp & x \geq 0 \\ 0 \leq -h(x) & \perp & \lambda \geq 0. \end{array}$$

If X has a different representation, this construction would be modified appropriately. In the linear programming example (9), these conditions are precisely those already given as (3).

When X is the nonnegative orthant, the VI is just an alternative way to state a complementarity problem. However, when X is a more general set, it may be possible to treat it differently than simply introducing multipliers, see [15] for example. In particular, when X is a polyhedral set, algorithms may wish to generate iterates via projection onto X.

Bimatrix games can also be formulated as a variational inequality. In this setting, two players have I and J pure strategies, and p and q (the strategy probabilities) belong to unit simplex $\triangle_I$ and $\triangle_J$ respectively. Payoff matrices $A \in R^{J\times I}$ and $B \in R^{I\times J}$ are defined, where $A_{j,i}$ is the profit received by the first player if strategy i is selected by the first player and j by the second. The expected profit for the first and the second players are then $q^T Ap$ and $p^T Bq$ respectively. A Nash equilibrium is reached by the pair of strategies (p^*, q^*) if and only if

$$p^* \in \arg \min_{p\in\triangle_I} \langle Aq^*, p\rangle \quad \text{and} \quad q^* \in \arg \min_{q\in\triangle_J} \langle B^T p^*, q\rangle$$

Letting x be the combined probability vector (p, q), the (coupled) optimality conditions for the above problems constitute the variational inequality:

$$F\left(\begin{bmatrix} p \\ q \end{bmatrix}\right) = \begin{bmatrix} 0 & A \\ B^T & 0 \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix}, \quad X = \triangle_I \times \triangle_J.$$

Algorithms to solve this problem can exploit the fact that X is a compact set. The thesis [50] contains Newton based algorithms to solve these problems effectively.

3.5 Bilevel Programs

Mathematical programs with optimization problems in their constraints have a long history in operations research including [12, 32, 5]. Hierarchical optimization has recently become important in a number of different applications and new codes are being developed that exploit this structure, at least for simple hierarchies, and attempt to define and implement algorithms for their solution.

The simplest case is that of bilevel programming, where an upper level problem depends on the solution of a lower level optimization. For example:

$$
\begin{aligned}
&\min_{x,y} f(x,y) \\
&\text{s.t. } g(x,y) \leq 0, \\
&\qquad y \text{ solves } \min_{y} v(x,y) \text{ s.t. } \quad h(x,y) \leq 0.
\end{aligned}
$$

This problem can be reformulated as an MPCC by replacing the lower level optimization problem by its optimality conditions:

$$
\begin{aligned}
&\min_{x,y} f(x,y) \\
&\text{s.t. } g(x,y) \leq 0, \\
&\qquad 0 = \nabla_y v(x,y) + \lambda^T \nabla_y h(x,y) \perp x \text{ free} \\
&\qquad 0 \leq -h(x,y) \perp \lambda \geq 0.
\end{aligned}
$$

This approach then allows such problems to be solved using the NLPEC code, for example. However, there are several possible deficiencies that should be noted. Firstly, the optimality conditions encompassed in the complementarity constraints may not have a solution, or the solution may only be necessary (and not sufficient) for optimality. Secondly, the MPCC solver may only find local solutions to the problem. The quest for practical optimality conditions and robust global solvers remains an active area of research. Importantly, the EMP tool will provide the underlying structure of the model to a solver if these advances determine appropriate ways to exploit this.

We can model this bilevel program in GAMS by

```
model bilev /deff,defg,defv,defh/;
solve bilev using emp min f;
```

along with some extra annotations to a subset of the model defining equations. Specifically, within an "empinfo" file we state that the lower level problem involves the objective v which is to be minimized subject to the constraints specified in defv and defh.

```
bilevel x
min v defv defh
```

Note that the variables x are declared to be variables of the upper level problem and that defg will be an upper level constraint. The specific syntax is described in [38]. Having written the problem in this way, the MPCC is generated automatically, and passed on to a solver. In the case where that solver is NLPEC, a further reformulation of the model is carried out to convert the MPCC into an equivalent NLP or a parametric sequence of NLP's. A key extension to the bilevel format allows multiple lower level problems to be specified within the bilevel format.

3.6 Embedded Complementarity Systems

A different type of embedded optimization model that arises frequently in applications is:

$$\begin{array}{lll} \min\limits_{x} & f(x,y) & \\ \text{s.t.} & g(x,y) \leq 0 & (\perp \lambda \leq 0) \\ & & \\ \multicolumn{2}{l}{H(x,y,\lambda) = 0} & (\perp y \text{ free}) \end{array}$$

Note the difference here: the optimization problem is over the variable x, and is parameterized by the variable y. The choice of y is fixed by the (auxiliary) complementarity relationships depicted here by H. Note that the "H" equations are not part of the optimization problem, but are essentially auxiliary constraints to tie down remaining variables in the model.

Within GAMS, this is modeled as:

```
model ecp /deff,defg,defH/;
solve ecp using emp;
```

Again, so this model can be processed correctly as an EMP, the modeler provides additional annotations to the model defining equations in an "empinfo" file, namely that the function H that is defined in defH is complementary to the variable y (and hence the variable y is a parameter to the optimization problem), and furthermore that the dual variable associated with the equation defg in the optimization problem is one and the same as the variable λ used to define H:

```
min f x deff defg
vifunc defH y
dualvar lambda defg
```

Armed with this additional information, the EMP tool automatically creates the following MCP:

$$\begin{array}{lll} 0 = \nabla_x \mathcal{L}(x,y,\lambda) & \perp & x \text{ free} \\ 0 \geq -\nabla_\lambda \mathcal{L}(x,y,\lambda) & \perp & \lambda \leq 0 \\ 0 = H(x,y,\lambda) & \perp & y \text{ free}, \end{array}$$

where the Lagrangian is defined as

$$\mathcal{L}(x,y,\lambda) = f(x,y) - \lambda^T g(x,y).$$

Perhaps the most popular use of this formulation is where competition is allowed between agents. A standard method to deal with such cases is via the concept of Nash Games. In this setting x^* is a Nash Equilibrium if

$$x_i^* \in \arg\min_{x_i \in X_i} \ell_i(x_i, x_{-i}^*, q), \forall i \in \mathcal{I},$$

where x_{-i} are other players decisions and the quantities q are given exogenously, or via complementarity:

$$0 \leq H(x,q) \quad \perp \quad q \geq 0.$$

This mechanism is extremely popular in economics, and Nash famously won the Nobel Prize for his contributions to this literature.

This format is again an EMP, more general than the example given above in two respects. Firstly, there is more than one optimization problem specified in the embedded complementarity system. Secondly, the parameters in each optimization problem consist of two types. Firstly, there are the variables q that are tied down by the auxiliary complementarity condition and hence are treated as parameters by the ith Nash player. Also there are the variables x_{-i} that are treated as parameters by the ith Nash player, but are treated as variables by a different player j. While we do not specify the syntax here for these issues, [38] provides examples that outline how to carry out this matching within GAMS. Finally, two points of note: first it is clear that the resulting model is a complementarity problem and can be solved using PATH, for example. Secondly, performing the conversion from an embedded complementarity system or a Nash Game automatically is a critical step in making such models practically useful.

Many of the energy market pricing models are based in the economic theory of general equilibria. In this context, there are a number of consumers each maximizing some utility function U_k, income is determined by the market price of the endowment and production shares, production is a technology constrained optimization of profit, and the market clears by choosing appropriate prices:

$$
\begin{aligned}
(C) &: \max_{x_k \in X_k} U_k(x_k) \text{ s.t. } p^T x_k \leq i_k(y,p) \\
(I) &: i_k(y,p) = p^T \omega_k + \sum_j \alpha_{kj} p^T g_j(y_j) \\
(P) &: \max_{y_j \in Y_j} p^T g_j(y_j) \\
(M) &: \max_{p \geq 0} p^T \left(\sum_k x_k - \sum_k \omega_k - \sum_j g_j(y_j) \right) \text{ s.t. } \sum_l p_l = 1
\end{aligned}
$$

Note that this model is an example of a Nash Game, with four different types of agents. Note that in each problem, some of the variables are under the control of the agent, and some are given as parameters. One way to solve this problem is to form the KKT conditions of each agent problem, and combine them to make a large scale complementarity problem.

Alternatively, the problem can be reformulated as embedded complementarity system (see [24]) of the form:

$$
\begin{aligned}
\max_{x \in X, y \in Y} \quad & \sum_k \frac{t_k}{\beta_k} \log U_k(x_k) \\
\text{s.t.} \quad & \sum_k x_k \leq \sum_k \omega_k + \sum_j g_j(y_j)
\end{aligned}
$$

$$t_k = i_k(y,p) \text{ where } p \text{ is multiplier on NLP constraint}$$

Note that the consumers and producers have been aggregated together into a large nonlinear program parameterized by t_k, a variable that is updated using the external condition. In practice, such problems can then be solved using the sequential joint maximization algorithm [68].

We note that there is a large literature on discrete-time finite-state stochastic games: this has become a central tool in analysis of strategic interactions among forward-looking players in dynamic environments. The Ericson-Pakes model of dynamic competition [23] in an oligopolistic industry is exactly in the format described above, and has been used extensively in applications such as advertising, collusion, mergers, technology adoption, international trade and finance.

For stylized models of this type, where a game is played over a grid of dimension S, the results of applying the PATH solver to the resulting complementarity problems are as follows:

S	Size	non-zero	dense(%)	Steps	Time (m:s)
20	2568	31536	0.48	5	0 : 03
50	15408	195816	0.08	5	0 : 19
100	60808	781616	0.02	5	1 : 16
200	241608	3123216	0.01	5	5 : 12

Note the number of Newton steps is constant, but the model size is increasing rapidly. For the largest grid size, the residual at each iteration is $1.56 * 10^4, 1.06 * 10^1, 1.34, 2.04 * 10^{-2}, 1.74 * 10^{-5},$ and $2.97 * 10^{-11}$ respectively, demonstrating quadratic convergence. It is clear that it is much easier to generate correctly reformulated models quickly using the automation of the EMP tool.

3.7 Semidefinite Programs

Semidefinite programming is a relatively new optimization format that has found application in many areas of control and signal processing, and is now being more widely utilized due to its inherent modeling power. Excellent survey articles of the background to this area and its applications can be found in [76, 82].

In the context of OPF problems, Lavaei and Low [49] convexify the problem and apply an SDP approach to the Dual OPF. Instead of solving the OPF problem directly, this approach solves the Lagrangian dual problem, and recovers a primal solution from a dual optimal solution. It is proved in the paper that the dual problem is a convex semidenite program and therefore can be solved efficiently using interior point solvers such as those described in [8, 73, 72]. In the general case, the optimal objective value of the dual problem is only a lower bound on the optimal value of the original OPF problem and the lower bound may not be tight (nonzero duality gap). If the primal solution computed from an optimal dual solution indeed satisfies all the constraints of the OPF problem and the resulting objective value equals the optimal dual objective value (zero duality gap), then strong duality holds and the primal solution is indeed (globally) optimal for the original OPF problem. The paper provides a sufficient condition that guarantees zero duality gap and global optimality of the resulting OPF solution.

However, applying the SDP approach outlined above shows that much more development in solution methodology is needed for this to be competitive for practical modeling. For larger OPF models described via [85] with solutions implemented via YALMIP [51] as a modeling tool and SeDuMi [72] as the solver, reported solutions were not feasible for the original nonlinear program and took significantly longer than alternative nonlinear programming approaches (CONOPT, IPOPT or SNOPT). Research is active in this area, however, and it is likely that methods exploiting underlying structure in the SDP will become practical in the near future. Indeed, the first order methods for specially structured SDPs described in [42, 59] have already proven effective in eigenvalue optimization problems.

3.8 Extended Nonlinear Programs

Optimization models have traditionally been of the form (1). Specialized codes have allowed certain problem structures to be exploited algorithmically, for example simple bounds on variables. However, for the most part, assumptions of smoothness of f, g and h are required for many solvers to process these problems effectively.

In a series of papers, Rockafellar and colleagues [64, 65, 63] have introduced the notion of extended nonlinear programming, where the (primal) problem has the form:

$$\min_{x \in X} f_0(x) + \theta(f_1(x), \ldots, f_m(x)). \tag{10}$$

In this setting, X is assumed to be a nonempty polyhedral set, and the functions $f_0, f_1, \ldots, f_m$ are smooth. The function θ can be thought of as a generalized penalty function that may well be nonsmooth. However, when θ has the following form

$$\theta(u) = \sup_{y \in Y} \{y^T u - k(y)\}, \tag{11}$$

a computationally exploitable and theoretically powerful framework can be developed based on conjugate duality. A key point for computation and modeling is that the function θ can be fully described by defining the set Y and the function k. Furthermore, as is detailed in [26], different choices lead to a rich variety of functions θ, many of which are extremely useful for modeling.

The EMP model type works in this setting by providing a library of functions θ that specify a variety of choices for k and Y. Once a modeler determines which constraints are treated via which choice of k and Y, the EMP model interface automatically forms an equivalent variational inequality or complementarity problem. There may be alternative formulations that are computationally more appealing; such reformulations can be generated using different options to EMP.

Note that the Lagrangian $\mathcal{L}$ is smooth - all the nonsmoothness is captured in the θ function. The theory is an elegant combination of calculus arguments related to f_i and its derivatives, and variational analysis for features related to θ. Exploitable structure is thus communicated directly to the computational engine that can solve the model.

It is shown in [64] that under a standard constraint qualification, the first-order conditions of (10) are precisely in the form of the following variational inequality:

$$\mathrm{VI}\left(\begin{bmatrix} \nabla_x \mathcal{L}(x,y) \\ -\nabla_y \mathcal{L}(x,y) \end{bmatrix}, X \times Y\right), \tag{12}$$

where the Lagrangian $\mathcal{L}$ is defined by

$$\mathcal{L}(x,y) = f_0(x) + \sum_{i=1}^{m} y_i f_i(x) - k(y)$$
$$x \in X, y \in Y$$

When X and Y are simple bound sets, this is simply a complementarity problem.

Note that EMP exploits this result. In particular, if an extended nonlinear program of the form (10) is given to EMP, then the optimality conditions (12) are formed as a variational inequality problem and can be processed as outlined above. Under appropriate convexity assumptions on this Lagrangian, it can be shown that a solution of the VI (12) is a saddle point for the Lagrangian on $X \times Y$. Furthermore, in this setting, the saddle point generates solutions to the primal problem (10) and its dual problem:

$$\max_{y \in Y} g(y), \quad \text{where } g(y) = \inf_{x \in X} \mathcal{L}(x,y),$$

with no duality gap.

In [26], an alternative solution method is proposed, based on a reformulation as an NLP to solve (12). By communicating the appropriate underlying structure to the solver interface, it is possible to reformulate the nonsmooth problem as smooth optimization problems in a variety of ways. We believe that specifying Y and k is a theoretically sound way to do this. Another example showing formulation of an extended nonlinear program as a complementarity problem within GAMS can be found in [22].

4 Stochastic and Robust Optimization

In order to effectively model many of the problems resulting from electricty grid design, EMP will require new syntax to allow specification of problems such as stochastic recourse programs.

Consider, for example, the two stage stochastic recourse problem:

$$\min c'x + \sum_i p_i Q_i(x) \text{ s.t. } x \in X,$$

where $Q_i(x) = \min_y d_i'y$ s.t. $T_i x + W_i y \geq h_i, y \in Y$. Standard modeling notation allows the specification of both X and Y, and the equations that define the feasible set of the recourse (Q_i) problems. The empinfo file would describe:

- the probability distribution p_i
- what stage is each variable in
- what stage is each constraint in
- what parameters in the original model are random (ie T_i, h_i, etc)
- how to sample these random parameters (using a library of sampling functions)
- what problem to generate and solve (ie the equivalent deterministic linear program, or a format necessary for decomposition approaches).

Within the modeling system, there is no need for the underlying problems to be linear. The automatic system would need to check that no random parameters appear in first stage equations, and that no second stage variables appear in first stage equations (and recursively for multi-stage problems).

An extension to chance constraints is also possible, where the problem is now:

$$\min c'x \text{ s.t. } x \in X, \sum_i p_i \mathcal{I}(T_i x + W_i y_i \geq h_i) \geq 1 - \epsilon,$$

where $\mathcal{I}(\cdot)$ is the indicator function (1 or 0) for its argument. Clearly, not only should the information needed above be generated, but also the annotation must specify the fraction of the constraints that can be violated (ϵ).

By employing variable annotations, it would also be possible to extend EMP to model risk measures such as CVaR. An additional variable (which represents a convex function of the decision variables x) would be used in the appropriate constraint or in the objective to be minimized. Extensions of solvers to perform subgradient optimization would be needed, or alternative decomposition approaches could be implemented "behind the scenes".

4.1 Optimization of noisy functions

Over the past few decades, computer simulation has become a powerful tool for developing predictive outcome of real systems. For example, simulations consisting of dynamic econometric models of travel behavior are used for nationwide demographic and travel demand forecasting. The choice of optimal simulation parameters can lead to improved operation, but configuring them remains a challenging problem. Traditionally, the parameters are chosen by heuristics with expert advice, or by selecting the best from a set of candidate parameter settings. Simulation-based optimization is an emerging field which integrates optimization techniques into the simulation analysis. The corresponding objective function is an associated measurement of an experimental simulation. Due to the complexity of simulation, the objective function may act as a black-box function and be time-consuming to evaluate. Moreover, since the derivative information of the objective function is typically unavailable, many derivative-dependent methods are not applicable. The third example of Section 2 fits nicely into this framework.

We can think of such approaches as a mechanism for coordinating existing optimization technologies. Each simulation evaluation corresponds to the solution of a parameterized problem (maybe in prices, or in variables that link together competing agents) that may be extremely time consuming to compute, and that may incorporate complex domain information, and may be subject to errors due to uncertainties. The noisy function optimization will determine (at a coordination level) what parameters are appropriate and where to concentrate attention in the search space.

Although real world problems have many forms, many optimization strategies consider the following bounded stochastic formulation:

$$\min_{x \in \Omega} f(x) = \mathrm{E}\left[F(x, \xi(\omega))\right], \tag{13}$$

where

$$\Omega = \{x \in \mathbf{R}^n : l \leq x \leq u\}.$$

Here, l and u are the lower and upper bounds for the input parameter x, respectively. The specific application of this framework to the electricity grid arises from considering the variables x to be the "design" or line capacity expansion variables. The function F would then model the response of the underlying system (as a large complex computer simulation) to those design decisions $p_i^{\omega}(x)$ and allow for uncertainties in the operating environment via the set Ω. An approach for solving these problems using Bayesian statistics is outlined in [20]. Such an approach balances the computational needs of the optimization against the need for much more accurate evaluations of the simulation. The facilitation of these extremely time intensive solutions using grid computational resources to couple the underlying optimization problems is critical for efficient solution.

4.2 Conic Programming

A problem of significant recent interest (due to its applications in robust optimization and optimal control) involves conic constraints [52, 1, 7]:

$$\min_{x \in X} p^T x \text{ s.t. } Ax - b \leq 0, x \in C,$$

where C is a convex cone. For specific cones, such as the Lorentz (ice-cream) cone defined by

$$C = \left\{ x \in \mathbf{R}^n \mid x_1 \geq \sqrt{\sum_{i=2}^{n} x_i^2} \right\},$$

or the rotated quadratic cone, there are efficient implementations of interior point algorithms for their solution [3]. It is also possible to reformulate the problem in the form (1) for example by adding the constraint

$$x_1 \geq \sqrt{\sum_{i=2}^{n} x_i^2}. \tag{14}$$

Annotating the variables that must lie in a particular cone using a "empinfo" file allows solvers like MOSEK [2] to receive the problem as a cone program, while standard NLP solvers would see a reformulation of the problem as a nonlinear program. It is also easy to see that (14) can be replaced by the following equivalent constraints

$$x_1^2 \geq \sum_{i=2}^{n} x_i^2,\ x_1 \geq 0.$$

Such constraints can be added to a nonlinear programming formulation or a quadratically constrained (QCP) formulation. This automatic reformulation allows the interior point algorithms to solve these problems since they can process constraints of the form

$$y^2 \geq x^T Q x,\ y \geq 0, Q \text{ PSD}.$$

Details on the options that implement these approaches can be found in [38].

These approaches have been been adapted to robust optimization models and applied to unit commitment problems [9]. It is straightforward to facilitate the use of stochastic constraints that have become very popular in financial applications. Specifically, we mention the work of [66] on conditional value at risk, and the recent papers by [21], and [53] on stochastic dominance constraints. All of these formulations are easily cast as constraints on decision variables annotated by additional (in this case distributional) information.

5 Computational Needs

It is imperative that we provide a framework for modeling optimization problems for solution on the computing resources that are available to the decision maker. The framework must be easy to adapt to multiple grid engines or cloud computing devices, and should seamlessly integrate evolving mechanisms from particular computing platforms into specific application models.

The design must be flexible and powerful enough for a large variety of optimization applications. The attraction of a grid computing environment is that it can provide an enormous amount of computing resources, many of which are simply commodity computing devices, with the ability to run commercial quality codes, to a larger community of users.

We strongly believe that grid computational resources are not enough to make parallel optimization mainstream. Setting aside the issue of data collection, it is imperative that we provide simple and easy to use tools that allow distributed algorithms to be developed without knowledge of the underlying compute engine. In large scale optimization, there are many techniques that can be used to effectively decompose a problem into smaller computational tasks that can then be controlled by a "coordinator" - essentially a master-worker approach. The above sections have outlined a variety of ways in which this could be accomplished within our optimization framework. We believe that in particular power flow applications, the decomposition approach can be significantly enhanced using specific domain knowledge, but these strategies may be complex to describe for a particular model. This approach is more general that current modeling systems allow, requiring extensions to current languages to facilate interfaces and interactions between model components, their solutions and the resources used to compute in a potentially diverse and distributed environment.

While it is clear that efficiency may well depend on what resources are available and the degree of synchronization required by the algorithm, it must be easy to generate structured large scale optimization problems, and high level implementations of methods to solve them. Stochastic programming (an underlying problem of particular interest within the design of the Next Generation Electric Grid) is perhaps a key example, where in most of the known solution techniques, large numbers of scenario subproblems need to be generated and solved.

A prototype grid facility [14] allows multiple optimization problems to be instantiated or generated from a given set of models. Each of these problems is solved concurrently in a grid computing environment. This grid computing environment can just be a laptop or desktop computer with one or more CPUs. Today's operating systems offer excellent multi-processing scheduling facilities and provide a low cost grid computing environment. Other alternatives include the Condor system, a resource management scheme developed at the University of Wisconsin, or commercial systems such as the cloud or supercomputing facilities avaiable at the National Laboratories. We must facilitate the use of new and evolving modeling paradigms for optimization on the rapidly changing and diverse computing environments that are available to different classes of decision makers.

6 Conclusions

Optimization can provide advice on managing complex systems. Such advice needs to be part of an interactive debate with informed decision makers. Many practical decision problems are carried out in a competitive environment without overall system control. Mechanisms to allow decision making in such circumstances can be informed by game theory and techniques from distributed computing. To answer the major design questions, small dynamic models need to be developed that are “level of detail” specific, and provide interfaces to other “subservient” models that provide appropriate aggregation of data and understanding of underlying complex features.

A number of new modeling formats involving complementarity and variational inequalities have been described in this paper and a framework, EMP, that allows such problems to be specified has been outlined. Such extensions facilitate modeling with competitive agents and automatic problem reformulations. We believe this will make a modeler’s task easier by allowing model structure to be described succinctly in such a setting, and will make model generation more reliable and automatic. In this way, algorithms can exploit model structure to improve solution speed and robustness. Furthermore, models dedicated to well defined decisions can be formulated using new formats such as semidefinite programming and stochastic optimization, and such descriptions carry the potential of global optimality. EMP is only a first step in this vein.

Solution methods for the resulting optimization problems are available within modeling systems, and the electric power industry could exploit these methods in a flexible manner using a combination of different model formats and solution techniques. Recent advances in stochastic optimization and conic programming are readily available within such systems. Treating uncertainties in large scale planning projects will become even more critical over the next decade due to the increase in volatility of the supply side as well as the demand. Optimization models with flexible systems design can help combat these uncertainties in the construction phase, the operational phase of the installed system, and in the long term demand for the provided electricity.

Acknowledgements

The material in this paper draws heavily on discussions with Jim Luedtke, Jesse Holzer, Lisa Tang, Yanchao Liu, Sven Leyffer, Ben Recht, Chris De Marco and Bernie Lesieutre and I acknowledge their insights and contributions to this work. I am grateful to Steven Dirkse, Jan Jagla and Alex Meeraus for implementing the ideas in the EMP framework in the GAMS modeling system. The opinions outlined are my own, however.

References

[1] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Mathematical Programming*, 95:3–51, 2003.

[2] E. D. Andersen and K. D. Andersen. The MOSEK interior point optimizer for linear programming: an implementation of the homogeneous algorithm. In H. Frenk, K. Roos, T. Terlaky, and S. Zhang, editors, *High Performance Optimization*, pages 197–232. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000.

[3] E. D. Andersen, C. Roos, and T. Terlaky. On implementing a primal-dual interior-point method for conic quadratic optimization. *Mathematical Programming*, 95(2):249–277, 2003.

[4] Göran Andersson. *Modelling and Analysis of Electric Power Systems*, 2008.

[5] J. F. Bard. *Practical Bilevel Optimization: Algorithms and Applications*, volume 30 of *Nonconvex Optimization and its Applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.

[6] B. T. Baumrucker, J. G. Renfro, and L. T. Biegler. MPEC problem formulations and solution strategies with chemical engineering applications. *Computers and Chemical Engineering*, 32:2903–2913, 2008.

[7] A. Ben-Tal and A. Nemirovskii. *Lectures on Modern Convex Optimization*. MPS-SIAM Series on Optimization. SIAM, Philadelphia, PA, 2001.

[8] Steven J. Benson and Yinyu Ye. Dsdp3: Dual-scaling algorithm for semidefinite programming. Technical report, Preprint ANL/MCS-P851-1000, 2001.

[9] D. Bertsimas, E. Litvinov, X. Sun, J. Zhao, and T. Zheng. Two-stage adaptive robust optimization for the unit commitment problem. Technical report, MIT, 2011.

[10] J. Bisschop and R. Entriken. *AIMMS – The Modeling System*. Paragon Decision Technology, Haarlem, The Netherlands, 1993.

[11] J. Bisschop and A. Meeraus. On the development of a general algebraic modeling system in a strategic planning environment. *Mathematical Programming Study*, 20:1–29, 1982.

[12] J. Bracken and J. T. McGill. Mathematical programs with optimization problems in the constraints. *Operations Research*, 21:37–44, 1973.

[13] A. Brooke, D. Kendrick, and A. Meeraus. *GAMS: A User's Guide*. The Scientific Press, South San Francisco, California, 1988.

[14] M. R. Bussieck, M. C. Ferris, and A. Meeraus. Grid enabled optimization with GAMS. *INFORMS Journal on Computing*, 21(3):349–362, 2009.

[15] M. Cao and M. C. Ferris. A pivotal method for affine variational inequalities. *Mathematics of Operations Research*, 21:44–64, 1996.

[16] M.T. Cezik and G. Iyengar. Cuts for mixed 0-1 conic programming. *Mathematical Programming*, 104:179–202, 2005.

[17] X. Chen and X. Deng. 3-NASH is PPAD-complete. *Electronic Colloquium on Computational Complexity*, 134, 2005.

[18] Mung Chiang, Stephen H. Low, A. Robert Calderbank, and John C. Doyle. Layering as optimization decomposition: A mathematical theory of network architectures. *Proceedings of the IEEE*, 95(1):255–312, 2007.

[19] C. Daskalakis, P.W. Goldberg, and C.H. Papadimitriou. The complexity of computing a Nash equilibrium. *Electronic Colloquium on Computational Complexity*, 115, 2005.

[20] G. Deng and M. C. Ferris. Variable-number sample-path optimization. *Mathematical Programming*, 117:81–109, 2009.

[21] D. Dentcheva and A. Ruszczyński. Optimization with stochastic dominance constraints. *SIAM Journal on Optimization*, 14:548–566, 2003.

[22] S. P. Dirkse and M. C. Ferris. MCPLIB: A collection of nonlinear mixed complementarity problems. *Optimization Methods and Software*, 5:319–345, 1995.

[23] R. Ericson and A. Pakes. Markov perfect industry dynamics: A framework for empirical analysis. *Review of Economic Studies*, 62:53–82, 1995.

[24] Yuri Ermoliev, Yuri Ermoliev, Gnther Fischer, Gunther Fischer, Vladimir Norkin, and Vladimir Norkin. On the convergence of the sequential joint maximization method for applied equilibrium problems. Technical report, IIASI, Laxenburg, Austria, 1996.

[25] F. Facchinei and J. S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer-Verlag, New York, New York, 2003.

[26] M. C. Ferris, S. P. Dirkse, J.-H. Jagla, and A. Meeraus. An extended mathematical programming framework. *Computers and Chemical Engineering*, 33:1973–1982, 2009.

[27] M. C. Ferris, S. P. Dirkse, and A. Meeraus. Mathematical programs with equilibrium constraints: Automatic reformulation and solution via constrained optimization. In T. J. Kehoe, T. N. Srinivasan, and J. Whalley, editors, *Frontiers in Applied General Equilibrium Modeling*, pages 67–93. Cambridge University Press, 2005.

[28] M. C. Ferris, R. Fourer, and D. M. Gay. Expressing complementarity problems and communicating them to solvers. *SIAM Journal on Optimization*, 9:991–1009, 1999.

[29] M. C. Ferris and T. S. Munson. Complementarity problems in GAMS and the PATH solver. *Journal of Economic Dynamics and Control*, 24:165–188, 2000.

[30] M. C. Ferris and J. S. Pang. Engineering and economic applications of complementarity problems. *SIAM Review*, 39:669–713, 1997.

[31] E. Bartholomew Fisher, R. O'Neill, and M. C. Ferris. Optimal transmission switching. *IEEE Transactions on Power Systems*, 23:1346–1355, 2008.

[32] J. Fortuny-Amat and B. McCarl. A representation and economic interpretation of a two-level programming problem. *Journal of the Operational Research Society*, 32(9):783–792, 1981.

[33] R. Fourer, D. M. Gay, and B. W. Kernighan. A modeling language for mathematical programming. *Management Science*, 36:519–554, 1990.

[34] R. Fourer, D. M. Gay, and B. W. Kernighan. *AMPL: A Modeling Language for Mathematical Programming.* Duxbury Press, Pacific Grove, California, 1993.

[35] A. Frangioni and C. Gentile. Perspective cuts for a class of convex 0-1 mixed integer programs. *Mathematical Programming*, 106:225–236, 2006.

[36] K. Furman and I. P. Androulakis. A novel MINLP-based representation of the original complex model for predicting gasoline emissions. *Computers and Chemical Engineering*, 32:2857–2876, 2008.

[37] GAMS Development Corporation. *NLPEC,* User's Manual, 2008. http://www.gams.com/solvers/nlpec.pdf.

[38] GAMS Development Corporation. *EMP,* User's Manual, 2009. http://www.gams.com/solvers/emp.pdf.

[39] O. Gunluk and J. Linderoth. Perspective reformulation of mixed integer nonlinear programs with indicator variables. Technical report, University of Wisconsin, 2009.

[40] P. T. Harker. *Lectures on Computation of Equilibria with Equation–Based Methods.* CORE Lecture Series. CORE Foundation, Louvain–la–Neuve, Université Catholique de Louvain, 1993.

[41] K. W. Hedman, M. C. Ferris, R. P. O'Neill, E. B. Fisher, and S. S. Oren. Co-optimization of generation unit commitment and transmission switching with n-1 reliability. *IEEE Transactions on Power Systems*, page forthcoming, 2010.

[42] C. Helmberg and F. Rendl. A spectral bundle method for semidefinite programming. *SIAM Journal on Optimization*, 10(3):673–696, 2000.

[43] B.F. Hobbs. Linear complementarity models of Nash-Cournot competition in bilateral and POOLCO power markets. *IEEE Transactions on Power Systems*, 16:194–202, 2001.

[44] B.F. Hobbs, C. B. Metzler, and J.S. Pang. Strategic gaming analysis for electric power systems: An MPEC approach. *IEEE Transactions on Power Systems*, 15:638–645, 2000.

[45] W. W. Hogan. Energy policy models for Project Independence. *Computers & Operations Research*, 2:251–271, 1975.

[46] William W. Hogan. *Financial Transmission Right Formulations*, 2002.

[47] William W. Hogan, Scott M. Harvey, and Susan L. Pope. *Transmission Capacity Reservations and Transmission Congestion Contracts*, 1997.

[48] Tarjei Kristiansen and Juan Rosellón. A merchant mechanism for electricity transmission expansion. *Journal of Regulatory Economics*, 29:167–193, 2006.

[49] Javad Lavaei and Steven H. Low. Convexification of optimal power flow problem. In *Forty-Eighth Annual Allerton Conference*, 2010. Available from `https://www.cds.caltech.edu/~lavaei/Allerton_2010.pdf`.

[50] Q. Li. *Large-Scale Computing for Complementarity and Variational Inequalities*. PhD thesis, University of Wisconsin, Madison, Wisconsin, January 2010.

[51] J. Löberg. YALMIP : A toolbox for modeling and optimization in MATLAB. In *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.

[52] M. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra and its Applications*, 284:193–228, 1998.

[53] J. Luedtke. New formulations for optimization under stochastic dominance comstraints. *SIAM Journal on Optimization*, 19:1433–1450, 2008.

[54] A. S. Manne and T. F. Rutherford. International trade in oil, gas and carbon emission rights: An intertemporal general equilibrium model. *The Energy Journal*, 14:1–20, 1993.

[55] Maximal Software. MPL. http://www.maximal-usa.com/mpl/what.html, 2009.

[56] Robert H. Miller and James H. Malinowski. *Power System Operation*. McGraw-Hill Professional, third edition, 1994.

[57] A. Monticelli, M.V.F. Pereira, and S. Granville. Security-constrained optimal power flow with post-contingency corrective rescheduling. *IEEE Transactions on Power Systems*, PWRS-2(1), 1987.

[58] Frederic H. Murphy, Hanif D. Sherali, and Allen L. Soyster. A mathematical programming approach for determining oligopolistic market equilibrium. *Mathematical Programming*, 24:92–106, 1982.

[59] Yurii Nesterov. Smoothing technique and its applications in semidefinite optimization. *Mathematical Programming, Series A*, 110(2):245–259, 2006.

[60] Office of Integrated Analysis and U.S. Department of Energy Forecasting. Energy information administration, international energy outlook 2007, 2007.

[61] Thomas J. Overbye, Xu Cheng, and Yan Sun. A comparison of the ac and dc power flow models for lmp calculations. In *Proceedings of the 37th Hawaii International Conference on System Sciences*, January 2004.

[62] D. Ralph and S. J. Wright. Some properties of regularization and penalization schemes for MPECs. *Optimization Methods and Software*, 19(5):527–556, 2004.

[63] R. T. Rockafellar. Linear-quadratic programming and optimal control. *SIAM Journal on Control and Optimization*, 25:781–814, 1987.

[64] R. T. Rockafellar. Lagrange multipliers and optimality. *SIAM Review*, 35:183–238, 1993.

[65] R. T. Rockafellar. Extended nonlinear programming. In G. Di Pillo and F. Giannnessi, editors, *Nonlinear Optimization and Related Topics*, pages 381–399. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999.

[66] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *The Journal of Risk*, 2:21–41, 2000.

[67] R. T. Rockafellar and R. J. B. Wets. *Variational Analysis*. Springer-Verlag, 1998.

[68] T.F. Rutherford. Sequential joint maximization. In John Weyant, editor, *Energy and Environmental Policy Modeling*, volume 18 of *International Series in Operations Research and Management Science*. Kluwer, 1999.

[69] Y. Smeers. Computable equilibrium models and the restructuring of the European electricity and gas markets. *The Energy Journal*, 1997.

[70] Brian Stott, Ongun Alsac, and Alcir J. Monticelli. Security analysis and optimization. *Proceedings of the IEEE*, 75(12), 1987.

[71] R. Stubbs and S. Mehrotra. A branch-and-cut method for 0-1 mixed convex programming. *Mathematical Programming*, 86:515–532, 1999.

[72] Jos F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11-12:625–653, 1999.

[73] K. C. Toh, M.J. Todd, and R. H. Tütüncü. *SDPT3: A MATLAB software package for semidefinite-quadratic-linear programming*. Available from `http://www.math.nus.edu.sg/~mattohkc/sdpt3.html`.

[74] M. Turkay and I. E. Grossmann. Logic-based MINLP algorithms for the optimal synthesis of process networks. *Computers and Chemical Engineering*, 20:959–978, 1996.

[75] P. Van Hentenryck. *The OPL Optimization Programming Language*. MIT Press, 1999.

[76] Lieven Vandenberghe and Stephen Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, 1996.

[77] A. Vecchietti and I. E. Grossmann. LOGMIP: a disjunctive 0-1 nonlinear optimizaer for process system models. *Computers and Chemical Engineering*, 23:555–565, 1999.

[78] A. Vecchietti, S. Lee, and I. E. Grossmann. Modeling of discrete/continuous optimization problems: Characterizaton and formulations of disjunctions and their relaxations. *Computers and Chemical Engineering*, 27:433–448, 2003.

[79] J. Wallace, A. Philpott, M. O'Sullivan, and M. Ferris. Optimal rig design using mathematical programming. In *2nd High Performance Yacht Design Conference, Auckland, 14–16 February, 2006*, pages 185–192, 2006.

[80] Hongye Wang, Carlos E. Murillo-Sánchez, Ray D. Zimmerman, and Robert J. Thomas. On computational issues of market-based optimal power flow. *IEEE Transactions On Power Systems*, 22(3):1185–1193, 2007.

[81] J.-Y. Wei and Y. Smeers. Spatial oligopolistic electricity models with Cournot firms and regulated transmission prices. *Operations Research*, 47:102–112, 1999.

[82] H. Wolkowitz, R. Saigal, and L. Vandenberghe, editors. *Handbook of semidefinite programming: theory, algorithms and applications*. Kluwer, 2000.

[83] J. Yao, B. Willems, S. Oren, and I. Adler. Cournot equilibrium in price-capped two-settlement electricity markets. In *Proceedings of the 38th Hawaii International Conference on Systems Sciences HICSS38*, Big Island, Hawaii, January 3-6 2005.

[84] Jian Yao, Shmuel S. Oren, and Ilan Adler. Computing cournot equilibria in two settlement electricity markets. In *Proceedings of the 37th Hawaii International Conference on System Sciences*, January 2004.

[85] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas. MATPOWER's extensible optimal power flow architecture. In *Power and Energy Society General Meeting*. IEEE, July 26-30 2009.