Continuous Optimization

Sensitivity and covariance in stochastic complementarity problems with an application to North American natural gas markets

Sriram Sankaranarayanan [a,*], Felipe Feijoo [b], Sauleh Siddiqui [a,c,d]

[a] Department of Civil Engineering, Johns Hopkins University, United States
[b] Pontificia Universidad Católica de Valparaíso, Chile
[c] Department of Applied Mathematics and Statistics, Johns Hopkins University, United States
[d] German Institute for Economic Research (DIW Berlin), United States

A B S T R A C T

We provide an efficient method to approximate the covariance between decision variables and uncertain parameters in solutions to a general class of stochastic nonlinear complementarity problems. We also develop a sensitivity metric to quantify uncertainty propagation by determining the change in the variance of the output due to a change in the variance of an input parameter. The covariance matrix of the solution variables quantifies the uncertainty in the output and pairs correlated variables and parameters. The sensitivity metric helps in identifying the parameters that cause maximum fluctuations in the output. The method developed in this paper optimizes the use of gradients and matrix multiplications which makes it particularly useful for large-scale problems. Having developed this method, we extend the deterministic version of the North American Natural Gas Model (NANGAM), to incorporate effects due to uncertainty in the parameters of the demand function, supply function, infrastructure costs, and investment costs. We then use the sensitivity metrics to identify the parameters that impact the equilibrium the most.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Complementarity models arise naturally out of various real life problems. A rigorous survey of their application is available in Ferris and Pang (1997). Authors in Abada, Gabriel, Briat, and Massol (2013); Christensen and Siddiqui (2015); Feijoo, Huppmann, Sakiyama, and Siddiqui (2016); Huppmann and Egging (2014); Martín, Smeers, and Aguado (2015); Oke, Huppmann, Marshall, Poulton, and Siddiqui (2016) use complementarity problems to model markets from a game theoretic perspective (Anderson & Xu, 2004; Siddiqui & Christensen, 2016), where the complementarity conditions typically arise between the marginal profit and the quantity produced by the producer. In the field of mechanics, they typically arise in the context of frictional contact problems (Kwak & Lee, 1988), where there is a complementarity relation between the frictional force between a pair of surfaces and the distance of separation between them. With a wide range of applications, understanding the characteristics of solutions to complementarity problems becomes important for advancing the field. In

this paper, we focus on studying the characteristics of solutions to complementarity problems under uncertainty.

The behavior of a solution to a complementarity problem with random parameters was first addressed in Gürkan and Robinson (1999), where such problems were referred to as stochastic complementarity problems (SCP). Authors in Chen and Fukushima (2005); Egging, Pichler, Kalvø, and WalleHansen (2016); Gabriel, Zhuang, and Egging (2009); Jiang and Xu (2008); Shanbhag (2013) define various formulations of SCP for different applications and have devised algorithms to solve the problem. Authors in Lamm, Lu, and Budhiraja (2016) compute confidence intervals for solution of the expected value formulation of the problem, however they do not have efficient methods to find the second-order statistics for large-scale complementarity problems.

Large-scale problems, those with over 10,000 decision variables and uncertain parameters arise naturally out of detailed market models and there is considerable interest in studying, understanding and solving such models. For example, Chen, Cowling, Polack, Remde, and Mourdjis (2017) discuss a case of urban drainage system with large number of variables. Yumashev and Johnson (2017) discuss a case of deciding under large-scale nuclear emergencies. In line with the area of application used in this paper, Gabriel, Kydes, and Whitman (2001) discuss a case of an energy model with large number of variables and parameters. Naturally,

---

developing methods to solve such large-scale problems gained interest. Authors in Kopanos, Méndez, and Puigjaner (2010); Luo, Hong, Nelson, and Wu (2015) discuss various tools ranging from mathematical techniques (decomposition based) to computational techniques (parallel processing) for solving large-scale optimization problems. Ohno, Boh, Nakade, and Tamura (2016) uses an approximate algorithm for a large-scale Markov decision process to optimize production and distribution systems. In this paper, we do not present a new method to solve stochastic complementarity problems, but an efficient algorithm to generate second-order information that is flexible enough to be coupled with any existing algorithm that provides a first-order solution.

The objective of this paper is to efficiently obtain second-order statistical information about solution vectors of large-scale stochastic complementarity problems. This gives us information about variability of the equilibrium obtained by solving a nonlinear complementarity problem (NCP) and the correlation between various variables in the solution. Authors in Hyett, Podosky, Santamaria, and Ham (2007) and Benedetti-Cecchi (2003) provide examples in the area of clinical pathways and ecology respectively, about the utility of understanding the variance of the solution in addition to the mean. They also show that a knowledge of variance aids better understanding and planning of the system. Agrawal, Ding, Saberi, and Ye (2012) emphasize the necessity to understand covariance as a whole rather than individual variances by quantifying "the loss incurred on ignoring correlations" in a stochastic programming model.

In addition, we also introduce a sensitivity metric which quantifies the change in uncertainty in the output due to a perturbation in the variance of uncertain input parameters. This helps us to directly compare input parameters by the amount of uncertainty they propagate to the solution.

In attaining the above objectives, the most computationally expensive step is to solve a system of linear equations. We choose approximation methods over analytical methods, integration, or Monte Carlo simulation because of the computational hurdle involved while implementing those methods for large-scale problems. The method we describe in this paper achieves the following:

- The most expensive step has to be performed just once, irrespective of the covariance of the input parameters. Once the linear system of equations is solved, for each given covariance scenario, we only perform two matrix multiplications.
- Approximating the covariance matrix and getting a sensitivity metric can be obtained by solving the above mentioned linear system just once.

The methods developed in this paper can also be used for nonlinear optimization problems with linear equality constraints. We prove stronger results on error bounds for special cases of quadratic programming.

Having developed this method, we apply it to a large-scale stochastic natural gas model for North America, an extension of the deterministic model developed in Feijoo et al. (2016) and determine the covariance of the solution variables. We then proceed to identify the parameters which have the greatest impact on the solution. A Python class for efficiently storing and operating on sparse arrays of dimension greater than two is created. This is useful for working with high-dimensional problems which have an inherent sparse structure in the gradients.

We divide the paper as follows. Section 2 formulates the problem and mentions the assumptions used in the paper. It then develops the algorithm used to approximate the solution covariance and provides proofs for bounding the error. Section 3 develops a framework to quantify the sensitivity of the solution to each of the random variables. Section 4 shows how the result can be

applied for certain optimization problems with equality constraints. Having obtained the theoretical results, Section 5 gives an example of a oligopoly where this method can be applied and compares the computational time of the approximation method with a Monte-Carlo method showing the performance improvement for large-scale problems. Section 6 describes the Natural Gas Model to which the said method is applied. Section 7 discusses the possible enhancements for the model and its limitations in the current form.

2. Approximation of covariance

For the rest of the paper, all bold quantities are vectors. A subscript i for those quantities refer to the i-th component of the vector in Cartesian representation.

2.1. Definitions

We define a complementarity problem and a stochastic complementarity problem which are central to the results obtained in this paper. We use a general definition of complementarity problems and stochastic complementarity problems as stated below.

Definition 1. (Facchinei & Pang, 2007) Given $\mathbf{F} : \mathbb{R}^{n \times m} \mapsto \mathbb{R}^n$, and parameters $\theta \in \mathbb{R}^m$, the *parametrized nonlinear complementarity problem* (NCP) is to find $\mathbf{x} \in \mathbb{R}^n$ such that

$$\mathbb{K} \ni \mathbf{x} \perp \mathbf{F}(\mathbf{x}; \theta) \in \mathbb{K}^* \quad (2.1)$$

where $\mathbb{K}^*$, the dual cone of $\mathbb{K}$ is defined as

$$\mathbb{K}^* = \left\{ \mathbf{x} \in \mathbb{R}^n : \mathbf{v}^T \mathbf{x} \geq 0 \quad \forall \mathbf{v} \in \mathbb{K} \right\} \quad (2.2)$$

Definition 2. Given a cone $\mathbb{K} \in \mathbb{R}^n$ a random function $\mathbf{F} : \mathbb{K} \times \Omega \mapsto \mathbb{R}^n$, the stochastic complementarity problem (SCP) is to find $x \in \mathbb{R}^n$ such that

$$\mathbb{K} \ni \mathbf{x} \perp \mathbb{E}\mathbf{F}(\mathbf{x}; \omega) \in \mathbb{K}^* \quad (2.3)$$

We assume that we can explicitly evaluate the expectation in (2.3) using its functional form, and that the SCP can be solved using an existing algorithm.

We now make assumptions on the form of $\mathbb{K}$ in (2.1). This form of $\mathbb{K}$ helps in establishing an equivalence between a complementarity problem and a minimization problem which is key to derive the approximation method in this paper.

Assumption 1. $\mathbb{K}$ in (2.1) is a Cartesian product of half spaces and full spaces, *i.e.*, for some $\mathcal{I} \subseteq \{1, 2, \ldots, n\}$

$$\mathbb{K} = \{\mathbf{x} \in \mathbb{R}^n : \quad \mathbf{x}_i \geq 0 \quad \text{if } i \in \mathcal{I}\} \quad (2.4)$$

We now propose a lemma about the form of the dual cone of $\mathbb{K}$ to understand the special form that it has. This will help us convert the complementarity problem into an unconstrained minimization problem.

Lemma 3. *The dual cone $\mathbb{K}^*$ of the set assumed in Assumption 1 is*

$$\mathbb{K}^* = \mathbb{K}' = \begin{cases} \mathbf{x} \in \mathbb{R}^n : \begin{matrix} \mathbf{x}_i \geq 0 & \text{if } i \in \mathcal{I} \\ \mathbf{x}_i = 0 & \text{if } i \notin \mathcal{I} \end{matrix} \end{cases} \quad (2.5)$$

Proof. Check Appendix A □

2.2. Preliminaries for approximation

In this subsection, we prove two preliminary results. Firstly, we prove our ability to pose an NCP as an unconstrained minimization problem. Then we prove results on twice continuous differentiability of the objective function, thus enabling us to use the rich literature available for smooth unconstrained minimization. Following that, Propositions 5 and 6 help in achieving the former while

Proposition 7 along with its corollaries help us in achieving the latter.

We now define C-functions, which are central to pose the complementarity problem into an unconstrained optimization problem. The equivalent formulation as an unconstrained optimization problem assists us in developing the algorithm.

Definition 4. (Facchinei and Pang, 2007, pg. 72) A function $\psi :$ $\mathbb{R}^2 \mapsto \mathbb{R}$ is a *C-function* when

$$\psi(x, y) = 0 \Leftrightarrow x \geq 0 \quad y \geq 0 \quad xy = 0 \tag{2.6}$$

We consider the following commonly used C-functions.

$$\psi_{FB}(x, y) = \sqrt{x^2 + y^2} - x - y \tag{2.7}$$

$$\psi_{min}(x, y) = \min(x, y) \tag{2.8}$$

Under our assumptions on $\mathbb{K}$, the following two propositions establish the equivalence of the complementarity problem and an unconstrained minimization problem.

Proposition 5. *Suppose Assumption 1 holds. Then every solution* $\mathbf{x}^*(\boldsymbol{\theta})$ *of the parameterized complementarity problem in (2.1), is a global minimum of the following function* $\mathsf{f}(\mathbf{x}; \theta)$,

$$\Phi_i(\mathbf{x}, \theta; \mathbf{F}) = \begin{cases} \mathbf{F}_i(\mathbf{x}, \theta) & \text{if} \quad i \notin \mathcal{I} \\ \psi_i(\mathbf{x}_i, \mathbf{F}_i(\mathbf{x}, \theta)) & \text{if} \quad i \in \mathcal{I} \end{cases} \tag{2.9}$$

$$\mathsf{f}(\mathbf{x}; \theta) = \frac{1}{2} \|\Phi(\mathbf{x}; \theta; \mathbf{F})\|_2^2 \tag{2.10}$$

with an objective value 0, for some set of not necessarily identical C-functions ψ_i.

Proof. Check Appendix A □

Proposition 6. *Suppose Assumption 1 holds. If a solution to the problem in (2.1) exists and* $\mathbf{x}^*(\boldsymbol{\theta})$ *is an unconstrained global minimizer of* $\mathsf{f}(\mathbf{x}; \theta)$ *defined in (2.10), then* $\mathbf{x}^*(\boldsymbol{\theta})$ *solves the complementarity problem in (2.1).*

Proof. Check Appendix A □

Now given a function $\mathbf{F}$, and a set $\mathbb{K}$ which satisfies Assumption 1, and a solution of the NCP $\mathbf{x}^*(\hat{\theta})$ for some fixed $\theta = \hat{\theta}$, we define a vector valued function $\Phi : \mathbb{R}^{n \times m} \mapsto \mathbb{R}^n$ component-wise as follows.

$$\Phi_i(\mathbf{x}, \theta; \mathbf{F}) = \begin{cases} \mathbf{F}_i(\mathbf{x}, \theta) & \text{if} \quad i \notin \mathcal{I} \\ \psi^2(\mathbf{x}_i, \mathbf{F}_i(\mathbf{x}, \theta)) & \text{if} \quad i \in \mathcal{Z} \\ \psi(\mathbf{x}_i, \mathbf{F}_i(\mathbf{x}, \theta)) & \text{otherwise} \end{cases} \tag{2.11}$$

$$\mathsf{f}(\mathbf{x}; \theta) = \frac{1}{2} \|\Phi(\mathbf{x}; \theta; \mathbf{F})\|_2^2 \tag{2.12}$$

$$\mathcal{Z} = \left\{ i \in \mathcal{I} : \mathbf{x}^*_i(\hat{\theta}) = \mathbf{F}_i(\mathbf{x}^*(\hat{\theta}); \hat{\theta}) = 0 \right\} \tag{2.13}$$

Note that if ψ is a C-function, ψ^2 is also a C-function since $\psi^2 = 0 \iff \psi = 0$. We observe from Propositions 5 and 6 that minimizing $\mathsf{f}(\mathbf{x}; \theta)$ over $\mathbf{x}$ is equivalent to solving the NCP in (2.1).

Now we assume conditions on the smoothness of $\mathbf{F}$ so that the solution to a perturbed problem is sufficiently close to the original solution.

Assumption 2. $\mathbf{F}(\mathbf{x}; \theta)$ is twice continuously differentiable in $\mathbf{x}$ and $\boldsymbol{\theta}$ over an open set containing $\mathbb{K}$.

Given that the rest of the analysis is on the function $\mathsf{f}(\mathbf{x}; \theta)$ defined in (2.12), we prove that a sufficiently smooth $\mathbf{F}$ and a suitable ψ ensure a sufficiently smooth f.

Proposition 7. *With Assumption 2 holding, we state* $\mathsf{f}(\mathbf{x}; \hat{\theta})$ *defined as in (2.12) is a twice continuously differentiable at* $\mathbf{x}$ *satisfying* $\mathsf{f}(\mathbf{x}; \hat{\theta}) = 0$ *for any C function* ψ *provided*

1. ψ *is twice differentiable at* $\{ (a, b) \in \mathbb{R}^2 : \psi(a, b) = 0 \} \setminus \{(0, 0)\}$ *with a finite derivative and finite second derivative.*
2. ψ *vanishes sufficiently fast near the origin. i.e.,*

$$\lim_{(a,b) \to (0,0)} \psi^2(a, b) \frac{\partial^2 \psi(a, b)}{\partial a \partial b} = 0 \tag{2.14}$$

Proof. Given that f is a sum of squares, it is sufficient to prove each term individually is twice continuously differentiable to prove the theorem. Also since we are only interested where f vanishes, it is sufficient to prove the above property for each term where it vanishes.

Consider terms from $i \notin \mathcal{I}$. Since $\mathbf{F}_i$ is twice continuously differentiable, $\mathbf{F}_i^2$ is twice continuously differentiable too.

Consider the case $i \in \mathcal{Z}$. This means $i \in \mathcal{I}$ and $\mathbf{x}^*_i(\hat{\theta}) = \mathbf{F}_i\left(\mathbf{x}^*_i(\hat{\theta}), \hat{\theta}\right) = 0$. These contribute a ψ^4 term to f. With the notation $\psi_i \equiv \psi(\mathbf{x}_i, \mathbf{F}_i(\mathbf{x}; \theta))$, and $\delta_{ij} = 1 \iff i = j$ and 0 otherwise we clearly have,

$$\frac{\partial \psi_i^4}{\partial \mathbf{x}_j} = 4\psi_i^3 \left(\frac{\partial \psi_i}{\partial a} \delta_{ij} + \frac{\partial \psi_i}{\partial b} \frac{\partial \mathbf{F}_i}{\partial \mathbf{x}_j} \right) = 0 \tag{2.15}$$

$$\frac{\partial^2 \psi_i^4}{\partial \mathbf{x}_j \mathbf{x}_k} = 12\psi_i^2 \left(\frac{\partial \psi_i}{\partial a} \delta_{ij} + \frac{\partial \psi_i}{\partial b} \frac{\partial \mathbf{F}_i}{\partial \mathbf{x}_j} \right) \left(\frac{\partial \psi_i}{\partial a} \delta_{ik} + \frac{\partial \psi_i}{\partial b} \frac{\partial \mathbf{F}_i}{\partial \mathbf{x}_k} \right)$$
$$+ 4\psi_i^3 \left(\frac{\partial^2 \psi_i}{\partial a^2} \delta_{ij} \delta_{ik} + \frac{\partial^2 \psi_i}{\partial a \partial b} \frac{\partial \mathbf{F}_i}{\partial \mathbf{x}_k} + \text{ other terms} \right) \tag{2.16}$$

$$= 0 \tag{2.17}$$

For the third case, $i \in \mathcal{I} \setminus \mathcal{Z}$, we have

$$\frac{\partial \psi_i^2}{\partial \mathbf{x}_j} = 2\psi_i \left(\frac{\partial \psi_i}{\partial a} \delta_{ij} + \frac{\partial \psi_i}{\partial b} \frac{\partial \mathbf{F}_i}{\partial \mathbf{x}_j} \right) = 0 \tag{2.18}$$

$$\frac{\partial^2 \psi_i^2}{\partial \mathbf{x}_j \mathbf{x}_k} = \left(\frac{\partial \psi_i}{\partial a} \delta_{ij} + \frac{\partial \psi_i}{\partial b} \frac{\partial \mathbf{F}_i}{\partial \mathbf{x}_j} \right)$$
$$+ 2\psi_i \left(\frac{\partial^2 \psi_i}{\partial a^2} \delta_{ij} \delta_{ik} + \frac{\partial^2 \psi_i}{\partial a \partial b} \frac{\partial \mathbf{F}_i}{\partial \mathbf{x}_k} + \text{ other terms} \right) \tag{2.19}$$

$$= \frac{\partial \psi_i}{\partial a} \delta_{ij} + \frac{\partial \psi_i}{\partial b} \frac{\partial \mathbf{F}_i}{\partial \mathbf{x}_j} \tag{2.20}$$

Continuity of f at the points of interest follow the continuity of the individual terms at the points. □

The following corollaries show the existence of C-functions ψ which satisfy the hypothesis of Proposition 7.

Corollary 8. *With Assumption 2 holding, for the choice of C-function* $\psi = \psi_{min}$ *defined in (2.8), the function f is twice continuously differentiable at its zeros.*

Corollary 9. *With Assumption 2 holding, for the choice of C-function* $\psi = \psi_{FB}$ *defined in (2.7), the function f is twice continuously differentiable.*

We now define an isolated solution to a problem and assume that the problem of interest has this property. This is required to ensure that our approximation is well defined.

Definition 10. (Nocedal & Wright, 2006) A minimum $\mathbf{x}^*$ of a problem is said to be an *isolated minimum*, if there is a neighborhood $\mathcal{B}(\mathbf{x}^*; \epsilon)$ of $\mathbf{x}^*$, where $\mathbf{x}^*$ is the only minimum of the problem.

(a) An example of a function where the
global minimum $x = 0$ is a non-isolated
solution

(b) The intuition behind our approximation for finding where $\nabla f(\mathbf{x}, \theta) = 0$
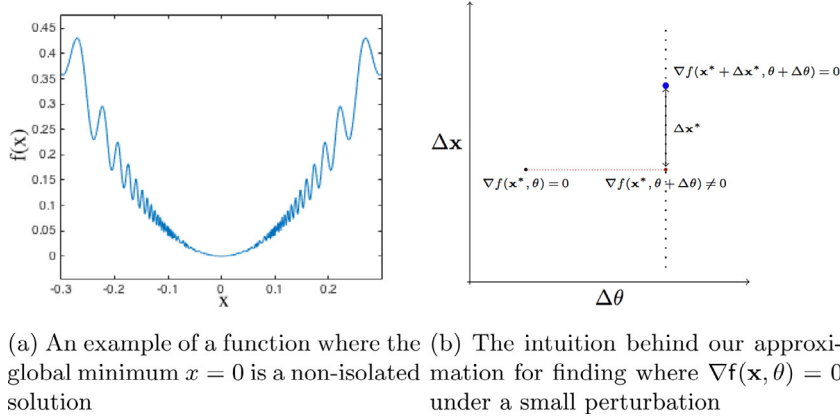under a small perturbation

Fig. 1. Intuitions.

A counter-example for an isolated minimum is shown on Fig. 1(a). It is a plot of the function

$$f(x) = 5x^2 + x^2 \sin\left(\frac{1}{x^2}\right) \tag{2.21}$$

and the global minimum at $x = 0$ is not an isolated minimum as we can confirm that any open interval around $x = 0$ has other minimum contained in it. Unlike this case, in this paper, we assume that if we obtain a global-minimum of f, then it is an isolated minimum. The existence of a neighborhood $\mathcal{B}(\mathbf{x}^*; \epsilon)$ as required in Definition 10 protects the approximation method from returning to a local minimum in the neighborhood for sufficiently small perturbations in problem parameters.

Assumption 3. For some fixed value of $\theta = \hat{\theta}$, there exists a known solution $\mathbf{x}^*(\hat{\theta})$ such that it is an isolated global minimum of $f(\mathbf{x}; \theta)$.

2.3. Approximation algorithm and error bounding

This subsection achieves two primary results as follows. We now propose Algorithm 1 to approximate the covariance of the output given a covariance matrix for the input parameters. Following that Theorem 11 gives a mathematical proof that the algorithm indeed approximates the covariance matrix. The second key result is Theorem 12 which bounds the error in the approximation.

Given a deterministic shift in a parameters value, the sensitivity of the solution has been studied in Castillo, Conejo, Castillo, Mínguez, and Ortigosa (2006); Fiacco (2009) for mathematical programming problems (e.g., nonlinear or large-scale programs). Authors in Castillo, Conejo, and Aranda (2008) used a perturbation technique to study the sensitivities in calculus of variations. We aim to build on the research in Castillo et al. (2006); Fiacco (2009) by proposing an approximation approach for large-scale stochastic complementarity problems. In particular, the sensitivity analysis in Chapter 3 in Fiacco (2009) motivates the method developed in this paper, after converting the complementarity problem into an unconstrained optimization problem with sufficient smoothness properties. Chapters 4 and 5 in Fiacco (2009) provide good context for computational approaches related to the ideas of Chapter 3. Our approach is amenable to extensions towards approximations for large-scale models, as we study the sensitivity of the variance of the solution, to a perturbation in the variance of input random parameters. We invite readers to refer to these works on further context in standard optimization problems. Chapter 2 of Fiacco (2009) provides interesting examples on how small perturbations can lead to large changes in solutions. This is good context in situations where Assumptions 2 (and Corollaries

---

Algorithm 1 Approximating covariance.

Solve the complementarity problem in (2.1) for the mean value of $\theta = \hat{\theta}$, or solve the stochastic complementarity problem in (2.3) and calibrate the value of the parameters $\theta = \hat{\theta}$ for this solution. Call this solution as $\mathbf{x}^*$. Choose a tolerance level τ.

1: Evaluate $\mathbf{F}^* \leftarrow \mathbf{F}(\mathbf{x}^*; \hat{\theta})$, $G_{ij} \leftarrow \frac{\partial \mathbf{F}_i(\mathbf{x}^*; \hat{\theta})}{\partial \mathbf{x}_j}$, $L_{i,j} \leftarrow \frac{\partial \mathbf{F}_i(\mathbf{x}^*; \hat{\theta})}{\partial \theta_j}$.

2: Choose a C-function ψ such that the conditions in proposition 7 are satisfied.

3: Define the function $\psi^a(a, b) = \frac{\partial \psi(a,b)}{\partial a}$, $\psi^b(a, b) = \frac{\partial \psi(a,b)}{\partial b}$.

4: Find the set of indices $\mathcal{Z} = \{z \in \mathcal{I} : |\mathbf{x}^*_z| = |\mathbf{F}^*_z| \leq \tau\}$.

5: Define

$$\mathcal{M}_{ij} \leftarrow \begin{cases} G_{ij} & \text{if } i \notin \mathcal{I} \\ 0 & \text{if } i \in \mathcal{Z} \\ \psi^a(\mathbf{x}^*_i, \mathbf{F}^*_i)\delta_{ij} + \psi^b(\mathbf{x}^*_i, \mathbf{F}^*_i)G_{ij} & \text{otherwise} \end{cases} \tag{2.22}$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

6: Define

$$\mathcal{N}_{ij} \leftarrow \begin{cases} L_{ij} & \text{if } i \notin \mathcal{I} \\ 0 & \text{if } i \in \mathcal{Z} \\ \psi^b(\mathbf{x}^*_i, \mathbf{F}^*_i)L_{ij} & \text{otherwise} \end{cases} \tag{2.23}$$

7: Solve the linear systems of equations for $\mathcal{T}$.

$$\mathcal{M}\mathcal{T} = \mathcal{N} \tag{2.24}$$

If $\mathcal{M}$ is non singular, we have a unique solution. If not, a least square solution or a solution obtained by calculating the Moore Penrose Pseudo inverse 26 can be used.

8: Given $\mathcal{C}$, a covariance matrix of the input random parameters, $\theta(\omega)$, **return** $\mathcal{C}^* \leftarrow \mathcal{T}\mathcal{C}\mathcal{T}^T$.

---

8 and 9) and 3 in our paper do not hold. The research in Castillo et al. (2006) provides good context for standard optimization problems, including situations where we relax assumptions on smoothness and active constraints.

The intuition behind the approximation is shown on Fig. 1(b) and can be summarized as follows. Having posed the NCP as an unconstrained minimization of a function f, we now approximate the change in the solution due to a perturbation of the parameters. Keeping the smoothness properties of f in mind, we say that the gradient of f vanishes at the solution before any perturbation. Following the random perturbation of parameters, we approximate the new value of the gradient of f at the old solution. Then we compute the step to be taken so that the gradient at the new point vanishes. We formalize this idea in Theorem 11 and use that

to build Algorithm 1 with some features for increased efficiency. The analysis builds on Dini's implicit function theorem Krantz and Parks (2012) for deterministic perturbations and extends the results to predict covariance under random perturbations.

Theorem 11. *Algorithm 1 generates Taylor's first-order approximation for the change in solution for a perturbation in parameters and computes the covariance of the solution for a complementarity problem with uncertain parameters with small variances.*

Proof. Consider the function $\mathsf{f}(\mathbf{x}; \theta)$. From Theorem 5, $\mathbf{x}^* \equiv \mathbf{x}^*(\hat{\theta})$ minimizes this function for $\theta = \hat{\theta}$. From Proposition 7, we have $\mathsf{f}(\mathbf{x}; \theta)$ is twice continuously differentiable at all its zeros. Thus we have,

$$\nabla_{\mathbf{x}} \mathsf{f}(\mathbf{x}^*; \hat{\theta}) = 0 \tag{2.25}$$

Now suppose the parameters $\hat{\theta}$ are perturbed by $\Delta\boldsymbol{\theta}$, then the above gradient can be written using the mean value theorem and then approximated up to the first order as follows.

$$\nabla_{\mathbf{x}} \mathsf{f}(\mathbf{x}^*(\hat{\theta}), \hat{\theta} + \Delta\theta) = \nabla_{\mathbf{x}} \mathsf{f}(\mathbf{x}^*(\hat{\theta}), \hat{\theta}) \\ + \nabla_{\hat{\theta}} \nabla_{\mathbf{x}} \mathsf{f}(\mathbf{x}^*(\hat{\theta}), \widetilde{\theta}) \Delta\theta \tag{2.26}$$

$$\nabla_{\mathbf{x}} \mathsf{f}(\mathbf{x}^*(\hat{\theta}), \hat{\theta} + \Delta\theta) - \nabla_{\mathbf{x}} \mathsf{f}(\mathbf{x}^*(\hat{\theta}), \hat{\theta}) \approx \mathcal{J}\Delta\theta \tag{2.27}$$

where,

$$\widetilde{\theta} \quad \in \quad [\theta, \theta + \Delta\theta] \tag{2.28}$$

$$\mathcal{J}_{ij} = [\nabla_{\hat{\theta}} \nabla_{\mathbf{x}} \mathsf{f}(\mathbf{x}^*(\hat{\theta}), \hat{\theta})]_{ij} \tag{2.29}$$

$$= \frac{\partial [\nabla_{\mathbf{x}} \mathsf{f}(\mathbf{x}^*; \hat{\theta})]_i}{\partial \hat{\theta}_j} \tag{2.30}$$

Since $\mathcal{J}\Delta\theta$ is not guaranteed to be 0, we might have to alter $\mathbf{x}$ to bring the gradient back to zero. *i.e.*, we need $\Delta\mathbf{x}$ such that $\nabla_{\mathbf{x}} \mathsf{f}(\mathbf{x}^*(\hat{\theta}) + \Delta\mathbf{x}, \hat{\theta} + \Delta\theta) = \mathbf{0}$. But by the mean value theorem,

$$\nabla_{\mathbf{x}} \mathsf{f}(\mathbf{x}^*(\hat{\theta}) + \Delta\mathbf{x}, \hat{\theta} + \Delta\theta) = \nabla_{\mathbf{x}} \mathsf{f}(\mathbf{x}^*(\hat{\theta}), \hat{\theta} + \Delta\theta) \\ + \nabla_{\mathbf{x}}^2 \mathsf{f}(\widetilde{\mathbf{x}}, \hat{\theta} + \Delta\theta) \Delta\mathbf{x} \tag{2.31}$$

$$\mathbf{0} \approx \mathcal{J}\Delta\theta + \nabla_{\mathbf{x}}^2 \mathsf{f}(\widetilde{\mathbf{x}}, \hat{\theta}) \Delta\mathbf{x} \tag{2.32}$$

$$\approx \mathcal{J}\Delta\theta + \nabla_{\mathbf{x}}^2 \mathsf{f}(\mathbf{x}^*(\hat{\theta}), \hat{\theta}) \Delta\mathbf{x} \tag{2.33}$$

$$\mathcal{H}\Delta\mathbf{x} \approx -\mathcal{J}\Delta\theta \tag{2.34}$$

where,

$$\widetilde{\mathbf{x}} \quad \in \quad [\mathbf{x}^*(\hat{\theta}), \mathbf{x}^*(\hat{\theta}) + \Delta\mathbf{x}] \tag{2.35}$$

$$[\mathcal{H}]_{ij} = [\nabla_{\mathbf{x}}^2 \mathsf{f}(\mathbf{x}^*(\hat{\theta}), \hat{\theta})]_{ij} \tag{2.36}$$

$$= \frac{\partial [\nabla_{\mathbf{x}} \mathsf{f}(\mathbf{x}^*; \hat{\theta})]_i}{\partial \mathbf{x}_j} \tag{2.37}$$

Now from Nocedal and Wright (2006), the gradient of the least squares function f can be written as

$$\nabla_{\mathbf{x}} \mathsf{f}(\mathbf{x}^*, \hat{\theta}) = \mathcal{M}^T \Phi(\mathbf{x}^*, \hat{\theta}) \tag{2.38}$$

$$[\mathcal{M}]_{ij} = \frac{\partial \Phi_i(\mathbf{x}^*, \hat{\theta})}{\partial \mathbf{x}_j} \tag{2.39}$$

$$= \begin{cases} \dfrac{\partial \mathbf{F}_i(\mathbf{x}^*, \hat{\theta})}{\partial \mathbf{x}_j} & \text{if } i \notin \mathcal{I} \\[2mm] \dfrac{\partial \psi^2(\mathbf{x}_i, \mathbf{F}_i(\mathbf{x}^*, \hat{\theta}))}{\partial \mathbf{x}_j} & \text{if } i \in \mathcal{Z} \\[2mm] \dfrac{\partial \psi(\mathbf{x}_i, \mathbf{F}_i(\mathbf{x}^*, \hat{\theta}))}{\partial \mathbf{x}_j} & \text{otherwise} \end{cases} \tag{2.40}$$

$$= \begin{cases} \dfrac{\partial \mathbf{F}_i(\mathbf{x}^*, \hat{\theta})}{\partial \mathbf{x}_j} & \text{if } i \notin \mathcal{I} \\[2mm] 0 & \text{if } i \in \mathcal{Z} \\[2mm] \dfrac{\partial \psi_i}{\partial \mathbf{x}_j} & \text{otherwise} \end{cases} \tag{2.41}$$

which is the form of $\mathcal{M}$ defined in Algorithm 1. Also

$$\mathcal{H} = \nabla_{\mathbf{x}}^2 \mathsf{f}(\mathbf{x}^*; \hat{\theta}) = \mathcal{M}^T \mathcal{M} + \sum_{i=1}^{n} \Phi_i(\mathbf{x}^*; \hat{\theta}) \nabla_{\mathbf{x}}^2 \Phi_i(\mathbf{x}^*; \hat{\theta}) \tag{2.42}$$

$$= \mathcal{M}^T \mathcal{M} \tag{2.43}$$

where the second term vanishes since we have from Theorem 5 that each term of Φ individually vanishes at the solution. Now

$$\mathcal{J} = \nabla_{\mathbf{x}\theta} \mathsf{f}(\mathbf{x}^*; \hat{\theta}) \tag{2.44}$$

$$\mathcal{J}_{ij} = \frac{\partial [\nabla_{\mathbf{x}} \mathsf{f}(\mathbf{x}^*; \hat{\theta})]_i}{\partial \theta_j} \tag{2.45}$$

$$= \frac{\partial}{\partial \theta_j} \left(\sum_{k=1}^{n} [\nabla_{\mathbf{x}} \Phi(\mathbf{x}^*; \hat{\theta})]_{ki} \Phi_k(\mathbf{x}^*; \hat{\theta}) \right) \tag{2.46}$$

$$= \sum_{k=1}^{n} \left(\frac{\partial [\nabla_{\mathbf{x}} \Phi(\mathbf{x}^*; \hat{\theta})]_{ki}}{\partial \theta_j} \Phi_k(\mathbf{x}^*; \hat{\theta}) + [\nabla_{\mathbf{x}} \Phi(\mathbf{x}^*; \hat{\theta})]_{ki} \frac{\partial \Phi_k(\mathbf{x}^*; \hat{\theta})}{\partial \theta_j} \right) \tag{2.47}$$

$$= \sum_{k=1}^{n} \mathcal{M}_{ki} \mathcal{N}_{kj} = \mathcal{M}^T \mathcal{N} \tag{2.48}$$

where the first term vanished because Φ_i are individually zeros, and we define

$$\mathcal{N}_{ij} = \frac{\partial \Phi_i(\mathbf{x}^*; \hat{\theta})}{\partial \theta_j} \tag{2.49}$$

$$= \begin{cases} \dfrac{\partial \mathbf{F}_i}{\partial \mathbf{x}_j} & \text{if } i \notin \mathcal{I} \\[2mm] 2\psi(\mathbf{x}^*_i; \mathbf{F}^*_i)\psi^b(\mathbf{x}^*_i; \mathbf{F}^*_i)\dfrac{\partial \mathbf{F}_i}{\partial \theta_j} & \text{if } i \in \mathcal{Z} \\[2mm] \psi^b(\mathbf{x}^*_i; \mathbf{F}^*_i)\dfrac{\partial \mathbf{F}_i(\mathbf{x}^*; \hat{\theta})}{\partial \theta_j} & \text{otherwise} \end{cases} \tag{2.50}$$

$$= \begin{cases} \dfrac{\partial \mathbf{F}_i}{\partial \theta_j} & \text{if } i \notin \mathcal{I} \\[2mm] 0 & \text{if } i \in \mathcal{Z} \\[2mm] \psi^b(\mathbf{x}^*_i; \mathbf{F}^*_i)\dfrac{\partial \mathbf{F}_i(\mathbf{x}^*; \hat{\theta})}{\partial \theta_j} & \text{otherwise} \end{cases} \tag{2.51}$$

which is the form of $\mathcal{N}$ defined in Algorithm 1. By Assumption 3, we have a unique minimum in the neighborhood of $\mathbf{x}^*$ where the gradient vanishes. So we have from (2.34), (2.43) and (2.48)

$$\mathcal{H}\Delta\mathbf{x} = -\mathcal{J}\Delta\theta \qquad (2.52)$$

$$\mathcal{M}^T\mathcal{M}\Delta\mathbf{x} = -\mathcal{M}^T\mathcal{N}\Delta\theta \qquad (2.53)$$

$\Delta\mathbf{x}$ solves the above equation, if it solves

$$\mathcal{M}\Delta\mathbf{x} = -\mathcal{N}\Delta\theta \qquad (2.54)$$

By defining $\mathcal{T}$ as the solution to the linear system of equations

$$\mathcal{M}\mathcal{T} = \mathcal{N} \qquad (2.55)$$

$$\Delta\mathbf{x} = -\mathcal{T}\Delta\theta \qquad (2.56)$$

and we have the above first-order approximation. From Seber and Lee (2012), we know that if some vector $\mathbf{x}$ has covariance C, then for a matrix A, the vector $A\mathbf{x}$ will have covariance ACA^T. So we have.

$$\mathcal{C}ov(\Delta\mathbf{x}) \approx \mathcal{T}\mathcal{C}ov(\Delta\theta)\mathcal{T}^T \qquad (2.57)$$

Thus we approximate the covariance of $\Delta\mathbf{x}$ in Algorithm 1. □

The matrix $\mathcal{M}$ could potentially not have full rank, for example, when $\mathcal{Z}$ is non-empty, i.e., when we have weak complementarity terms. But we note that $\mathcal{M}^T\mathcal{M}$ is the Hessian of f and the null-space of the Hessian corresponds to the directions where the gradient does not change for small perturbations. So in a first-order sence, small perturbations in those directions do not move the solution, keeping the method robust even under weak complementarity. This is further confirmed by computational experiments detailed in Section 5.2 and Appendix D where we have cases with weak complementarity terms but no significant error.

Further, when $\mathcal{M}$ is singular, we use Moore–Penrose pseudoinverse, $\mathcal{M}^\dagger$ to solve the system of Eq. (2.55). Among possibly infinite solutions that minimize the error $\|\mathcal{M}\mathcal{T} - \mathcal{N}\|_2$, $\mathcal{M}^\dagger\mathcal{N}$ gives the solution that minimizes $\|\mathcal{T}\|_2$ (Ben-Israel & Greville, 2003). This would lead us to identifying the smallest step $\Delta\mathbf{x}$ that could be taken to reach the perturbed solution, up to first-order approximation and hence give the most conservative estimate of the uncertainty. Uniqueness and existence of $\mathcal{M}^\dagger$ is guaranteed and it can be computed efficiently. For computational purposes the matrix $\mathcal{T}$ in the above equation has to be calculated only once, irrespective of the number of scenarios for which we would like to run for the covariance of θ. Thus if $\mathbf{x} \in \mathbb{R}^n$, $\theta \in \mathbb{R}^m$ and we want to test the output covariance for k different input covariance cases, the complexity is equal to that of solving a system of n linear equations m times as in (2.55), and hence is $O(mn^2)$. i.e., the complexity is quadratic in the number of output variables, linear in the number of input parameters and constant in the number of covariance scenarios we would like to run.

In Theorem 12 below, we prove that the error in the approximation of Theorem 11 can be bounded using the condition number of the Hessian. We need the following assumption that the condition number of the Hessian of f is bounded and the Hessian is Lipschitz continuous.

Assumption 4. At the known solution of the complementarity problem of interest ($\theta = \hat{\theta}$),

1. The condition number of the Hessian of f defined is finite and equals to κ_H.
2. The Hessian of f is Lipschitz continuous with a Lipschitz constant $\mathcal{L}(\mathbf{x}^*; \theta)$.

Theorem 12. With Assumption 4 holding, the error in the linear approximation (2.34) for a perturbation of ϵ is $o(\epsilon)$.

Proof. Since $\nabla^2 f$ is Lipschitz continuous on both $\mathbf{x}$ and θ, we can write for $\widetilde{\mathbf{x}}$ near $\mathbf{x}^*$,

$$\left\|\nabla_{\mathbf{x}}^2 f(\mathbf{x}^*, \hat{\theta}) - \nabla_{\mathbf{x}}^2 f(\widetilde{\mathbf{x}}, \hat{\theta})\right\| \le \mathcal{L}(\mathbf{x}^*; \theta)\|\mathbf{x}^* - \widetilde{\mathbf{x}}\| \qquad (2.58)$$

$$\le \mathcal{L}(\mathbf{x}^*; \theta)\|\Delta\mathbf{x}\| \qquad (2.59)$$

$$\widetilde{\mathcal{H}} = \nabla_{\mathbf{x}}^2 f(\widetilde{\mathbf{x}}, \hat{\theta}) \qquad (2.60)$$

$$= \mathcal{H} + \varepsilon_H \qquad (2.61)$$

where $\|\varepsilon_H\| \le \mathcal{L}(\mathbf{x}^*; \theta)\|\Delta\mathbf{x}\|$. Applying the Lipschitz continuity on θ,

$$\left\|\nabla_\theta\nabla_{\mathbf{x}} f(\mathbf{x}^*, \widetilde{\theta}) - \nabla_\theta\nabla_{\mathbf{x}} f(\mathbf{x}^*, \hat{\theta})\right\| \le \mathcal{L}(\mathbf{x}^*; \theta)\left\|\widetilde{\theta} - \hat{\theta}\right\| \qquad (2.62)$$

$$\le \mathcal{L}(\mathbf{x}^*; \theta)\|\Delta\theta\| \qquad (2.63)$$

$$\widetilde{\mathcal{J}} = \nabla_\theta\nabla_{\mathbf{x}} f(\mathbf{x}^*, \widetilde{\theta}) \qquad (2.64)$$

$$= \mathcal{J} + \varepsilon_J \qquad (2.65)$$

where $\|\varepsilon_J\| \le \mathcal{L}(\mathbf{x}^*; \theta)\|\Delta\theta\|$. Thus the equation

$$\widetilde{\mathcal{H}}\Delta\mathbf{x} = \widetilde{\mathcal{J}}\Delta\theta \qquad (2.66)$$

is exact, even if we cannot compute $\widetilde{\mathcal{H}}$ and $\widetilde{\mathcal{J}}$ exactly. Now the error in inverting $\widetilde{\mathcal{H}}$ is bounded by the condition number (Horn and Johnson, 2012, Ch. 5).

$$\frac{\left\|\mathcal{H}^{-1} - \widetilde{\mathcal{H}}^{-1}\right\|}{\left\|\widetilde{\mathcal{H}}^{-1}\right\|} \le \frac{\kappa_H \frac{\|\varepsilon_H\|}{\|\widetilde{\mathcal{H}}\|}}{1 - \kappa_H \frac{\|\varepsilon_H\|}{\|\widetilde{\mathcal{H}}\|}} \qquad (2.67)$$

Assuming $\kappa_H\|\varepsilon_H\| \ll \|\mathcal{H}\|$, the above equation becomes

$$\frac{\left\|\mathcal{H}^{-1} - \widetilde{\mathcal{H}}^{-1}\right\|}{\left\|\widetilde{\mathcal{H}}^{-1}\right\|} \le \kappa_H \frac{\|\varepsilon_H\|}{\|\widetilde{\mathcal{H}}\|} \qquad (2.68)$$

$$\Rightarrow \left\|\widetilde{\mathcal{H}}^{-1} - \mathcal{H}^{-1}\right\| \le \kappa_H \frac{\left\|\widetilde{\mathcal{H}}^{-1}\right\|}{\left\|\widetilde{\mathcal{H}}\right\|}\|\varepsilon_H\| \qquad (2.69)$$

$$\Rightarrow \left\|\widetilde{\mathcal{H}}^{-1}\widetilde{\mathcal{J}} - \mathcal{H}^{-1}\mathcal{J} - \mathcal{H}^{-1}\varepsilon_J\right\|$$
$$\le \kappa_H \frac{\left\|\widetilde{\mathcal{H}}^{-1}\right\|}{\left\|\widetilde{\mathcal{H}}\right\|}\varepsilon_H\|\mathcal{J}\| + \kappa_H \frac{\left\|\widetilde{\mathcal{H}}^{-1}\right\|}{\left\|\widetilde{\mathcal{H}}\right\|}\|\varepsilon_H\|\|\varepsilon_J\| \qquad (2.70)$$

$$\Rightarrow \left\|\widetilde{\mathcal{H}}^{-1}\widetilde{\mathcal{J}} - \mathcal{H}^{-1}\mathcal{J}\right\| \le k_1\|\Delta\mathbf{x}\| + k_2\|\Delta\theta\| \qquad (2.71)$$

with

$$k_1 = \kappa_H\mathcal{L}(\mathbf{x}^*; \theta)\frac{\left\|\widetilde{\mathcal{H}}^{-1}\right\|}{\left\|\widetilde{\mathcal{H}}\right\|}\|\mathcal{J}\| \qquad (2.72)$$

$$k_2 = \mathcal{L}(\mathbf{x}^*; \theta)\left\|\mathcal{H}^{-1}\right\| \qquad (2.73)$$

Thus we have from (2.71), that the error in the approximation done in Algorithm 1 is bounded. □

3. Stochastic sensitivity analyses

In this section, we quantify the sensitivity of the variance of the solution to variance in each of the input parameters. To achieve this, we define total linear sensitivity and how it can be approximated using the matrix $\mathcal{T}$ derived in (2.55). We then proceed to prove that these quantities also bound the maximum increase in uncertainties of the output.

Definition 13. Given a function $f : \mathbb{R}^m \mapsto \mathbb{R}^n$, the *total linear sensitivity*, $\beta_d \in \mathbb{R}_+$ of a dimension $d \leq m$; $d \in \mathbb{N}$ at a point $\mathbf{x} \in \mathbb{R}^m$ is defined for $\delta > 0$, sufficiently small,

$$\beta_d = \inf \left\{ \alpha : \left| \|f(\mathbf{x} + \delta e_d)\|_2 - \|f(\mathbf{x})\|_2 \right| \leq \delta\alpha + o(\delta^2) \right\} \tag{3.1}$$

where e_d is the d-th standard basis vector.

This is a bound on the distance by which the function value can move for a small perturbation in the input. Now we look at the solution to the parametrized complementarity problem in (2.1) as a function from the space of parameter tuples to the space of solution tuples and bound the change in solution for a small perturbation. The next proposition shows how the total linear sensitivity can be calculated from the linear approximation matrix $\mathcal{T}$ derived earlier.

Proposition 14. *Suppose we know, $G \in \mathbb{R}^{n \times m}$ such that $G_{ij} = \frac{\partial f_i(\mathbf{x})}{\partial \mathbf{x}_j}$, then $\beta_d = \sqrt{\left(\sum_{i=1}^{n} G_{id}^2 \right)}$*

Proof. See Appendix A □

The above proposition proves that the $\mathcal{T}$ matrix obtained in (2.55) is sufficient to approximate the total linear sensitivity. The following result suggests how the total linear sensitivity can approximate the total variance in the output variables.

Theorem 15. *Given a function $f : \mathbb{R}^m \mapsto \mathbb{R}^n$ and β_d, the increase in the total uncertainty in the output, i.e., the sum of variances of the output variables, for a small increase of the variance of an input parameter, σ_d^2 of $\mathbf{x}_d$ is approximated by $\beta_d^2 \sigma_d^2$.*

Proof. Let E_d be the matrix of size $m \times m$ with zeros everywhere except the d-th diagonal element, where it is 1. Given $C = Cov(\mathbf{x}(\omega))$, for a small perturbation σ^2 in the variance of $\mathbf{x}_d$, the covariance of $f(\mathbf{x})$ changes as follows.

$$C^* \approx \nabla_{\mathbf{x}} f C \nabla_{\mathbf{x}} f^T \tag{3.2}$$

$$C^* + \Delta C^* \approx \nabla_{\mathbf{x}} f (C + \sigma^2 E_d) \nabla_{\mathbf{x}} f^T \tag{3.3}$$

$$= C^* + \sigma^2 \nabla_{\mathbf{x}} f E_d \nabla_{\mathbf{x}} f^T \tag{3.4}$$

$$[\Delta C^*]_{ij} \approx \sigma^2 [\nabla_{\mathbf{x}} f]_{id} [\nabla_{\mathbf{x}} f]_{jd} \tag{3.5}$$

$$\sum_{i=1}^{n} [\Delta C^*]_{ii} \approx \sigma^2 \beta_d^2 \tag{3.6}$$

which is the total increase in variance. The off-diagonal terms do not affect the total uncertainty in the system because, the symmetric matrix C can be diagonalized as QDQ^T, where Q is a rotation matrix, and the trace is invariant under orthogonal transformations. □

With the above result, we can determine the contribution of each input parameter to the total uncertainty in the output.

4. Application to optimization

To illustrate the application of this method explained in Algorithm 1, we use it to derive an approximation for the covariance of the solution of certain canonical optimization problems. The goal of this section is to walk the reader through a simple application of the method to develop intuition of the analysis and results.

To start with, we assume conditions on the differentiability and convexity of the objective function.

Assumption 5. The objective function $f(\mathbf{x}; \theta)$ is strictly convex in $\mathbf{x}$ and is twice continuously differentiable in $\mathbf{x}$ and θ.

In the theorem below, we approximate the covariance of the decision variables of a convex optimization with uncertainties in the linear term and with only linear equality constraints.

Theorem 16. *With Assumption 5 holding, the covariance of the primal and dual variables at the optimum of the problem,*

$$\underset{\mathbf{x}}{Minimize} \quad f(\mathbf{x}; \theta) = g(\mathbf{x}) + c(\theta)^T \mathbf{x} \tag{4.1}$$

subject to $Ax = b(\theta)$ $\qquad\qquad$ **(y)** $\qquad\qquad$ (4.2)

where $\theta = \theta(\omega)$ are random parameters with covariance C, is first-order approximated by $\mathcal{T}C\mathcal{T}^T$ where

$$\mathcal{T} = \begin{pmatrix} \nabla_{\mathbf{x}}^2 g(\mathbf{x}^*) & A^T \\ A & 0 \end{pmatrix}^{-1} \begin{pmatrix} -\nabla_\theta c(\theta) \\ \nabla_\theta b(\theta) \end{pmatrix} \tag{4.3}$$

Proof. For the given optimization problem, because of Assumption 5 and linear independence constraint qualification (LICQ), the KKT conditions are necessary and sufficient for optimality. The KKT condition satisfied at a solution $(\mathbf{x}^*, \mathbf{y}^*)$ for the problem are given by

$$\nabla_{\mathbf{x}} g(\mathbf{x}^*) + c(\theta) + A^T \mathbf{y}^* = 0 \tag{4.4}$$

$$A\mathbf{x}^* = b(\theta) \tag{4.5}$$

for some vector $\mathbf{y}$ so that the equation is well defined. Suppose from there, θ is perturbed by $\Delta\theta$, we have

$$\nabla_{\mathbf{x}} g(\mathbf{x}^*) + c(\theta + \Delta\theta) + A^T \mathbf{y}^* \approx \nabla_\theta c(\theta)\Delta\theta \tag{4.6}$$

$$A\mathbf{x}^* - b(\theta + \Delta\theta) \approx -\nabla_\theta b(\theta)\Delta\theta \tag{4.7}$$

Now we need to find $\Delta\mathbf{x}$ and $\Delta\mathbf{y}$ such that

$$\nabla_{\mathbf{x}} g(\mathbf{x}^* + \Delta\mathbf{x}) + c(\theta + \Delta\theta) + A^T (\mathbf{y}^* + \Delta\mathbf{y}) \approx 0 \tag{4.8}$$

$$A(\mathbf{x}^* + \Delta\mathbf{x}) - b(\theta + \Delta\theta) \approx 0 \tag{4.9}$$

$$\nabla_{\mathbf{x}}^2 g(\mathbf{x}^*)\Delta\mathbf{x} + A^T \Delta\mathbf{y} \approx \nabla_\theta c(\theta)\Delta\theta \tag{4.10}$$

$$A\Delta\mathbf{x} \approx -\nabla_\theta b(\theta)\Delta\theta \tag{4.11}$$

The above conditions can be compactly represented as

$$\begin{pmatrix} \nabla_{\mathbf{x}}^2 g(\mathbf{x}^*) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta\mathbf{x} \\ \Delta\mathbf{y} \end{pmatrix} = \begin{pmatrix} \nabla_\theta c(\theta) \\ -\nabla_\theta b(\theta) \end{pmatrix} \Delta\theta \tag{4.12}$$

If A has full rank, then the above matrix is non-singular. So the change in the decision variables $\mathbf{x}$ and the duals $\mathbf{y}$ can be written

as a linear transformation of the perturbation in the random parameters. And we now have

$$Cov\left(\begin{array}{c}\Delta\mathbf{x}\\\Delta\mathbf{y}\end{array}\right) = \mathcal{T}Cov(\theta)\mathcal{T}^T \tag{4.13}$$

$$\mathcal{T} = \begin{pmatrix}\nabla_{\mathbf{x}}^2 g(\mathbf{x}^*) & A^T\\A & 0\end{pmatrix}^{-1}\begin{pmatrix}-\nabla_\theta c(\theta)\\\nabla_\theta b(\theta)\end{pmatrix} \tag{4.14}$$

□

In the corollary below, we show that the method suggested is accurate (i.e. has zero error) for an unconstrained quadratic optimization problem with uncertainty in the linear term.

Corollary 17. *For an optimization problem with uncertainty of objectives of the form,*

$$f(\mathbf{x};\theta) = \frac{1}{2}\mathbf{x}^T G\mathbf{x} + \theta(\omega)^T\mathbf{x} \tag{4.15}$$

where G is positive definite, the approximation method has zero error. In other words, the obtained covariance matrix is exact.

Proof. See Appendix A □

5. Application to a general oligopoly market

We now present an example of a complementarity problem in a natural gas oligopoly and show how the methods developed in this paper can be applied.

5.1. Problem formulation and results

Consider k producers competitively producing natural gas in a Nash-Cournot game. Let the random unit costs of production be $\gamma_i(\omega)$, $i \in \{1, \ldots, k\}$. Also, let us assume that the consumer behavior is modeled by a linear demand curve $P(\tilde{Q})$ as follows.

$$P = a(\omega) + b(\omega)\tilde{Q} \tag{5.1}$$

where P is the price the consumer is willing to pay, $\tilde{Q}$ is the total quantity of the natural gas produced and random variables $a(\omega) > 0, b(\omega) < 0 \,\forall \omega \in \Omega$. Suppose the producers are maximizing their profits, then the Nash equilibrium can be obtained by solving the following complementarity problem (Cottle, Pang, & Stone, 2009; Gabriel, Conejo, Fuller, Hobbs, & Ruiz, 2012).

$$0 \le Q_i \perp \mathbf{F}_i(\mathbf{Q}) = \gamma_i - a - b\left(\sum_{j=1}^k Q_k\right) - bQ_i \ge 0 \tag{5.2}$$

In this formulation, a, b, γ_i correspond to $\boldsymbol{\theta}$ and Q_i correspond to $\mathbf{x}$ in (2.1) with $\mathcal{I} = \{1, 2, \ldots, k\}$. In the current numerical example, let us consider a duopoly where $k = 2$. Let

$$\mathbb{E}\begin{pmatrix}\gamma_1 & \gamma_2 & a & b\end{pmatrix}^T = \begin{pmatrix}2 & 1 & 15 & -1\end{pmatrix}^T \tag{5.3}$$

Solving the complementarity problem deterministically with the above parameter values, we get Q_1 and Q_2 to be 4 and 5, respectively. We use the C-function $\psi_{min}(x, y) = \min(x, y)$ for this example to get

$$\mathcal{M} = \begin{pmatrix}2 & 1\\1 & 2\end{pmatrix} \quad \mathcal{N} = \begin{pmatrix}1 & 0 & -1 & -13\\0 & 1 & -1 & -14\end{pmatrix} \tag{5.4}$$

Now we have from (2.52)

$$\mathcal{T} = \mathcal{M}^{-1}\mathcal{N} = \frac{1}{3}\begin{pmatrix}2 & -1 & -1 & -12\\-1 & 2 & -1 & -15\end{pmatrix} \tag{5.5}$$

Having obtained $\mathcal{T}$, we attempt to get insight on how uncertainties in various input parameters propagate through the model causing uncertainty in the equilibrium quantities. If we assume that all

these parameters, viz. γ_1, γ_2, a, b have a 10% coefficient of variation and are all uncorrelated, then the covariance matrix of the input is

$$C_1 = \begin{pmatrix}0.04 & 0 & 0 & 0\\0 & 0.01 & 0 & 0\\0 & 0 & 2.25 & 0\\0 & 0 & 0 & 0.01\end{pmatrix} \tag{5.6}$$

Then the covariance matrix of the solution would be

$$C_1^* = \mathcal{T}C_1\mathcal{T}^T = \begin{pmatrix}0.4289 & 0.4389\\0.4389 & 0.5089\end{pmatrix} \tag{5.7}$$

The standard deviation of the produced quantities are $0.65(=\sqrt{0.4289})$ and $0.71(=\sqrt{0.5089})$, respectively. The produced quantities also have about 95% positive correlation as an increase in demand will cause both producers to produce more and a decrease in demand will cause both producers to produce less.

If we assume that we have perfect knowledge about the demand curve, and if the uncertainty is only in the production costs, then the new parameter covariance C_2 has the third and fourth diagonal term of C_1 as zero. In such a scenario, we would expect the decrease in the quantity of production of one player to cause an increase in the quantity of production of the other and vice versa, caused by re-adjustment of market share. We can see this effect by computing the covariance of the solution as $\mathcal{T}C_2\mathcal{T}$. The solution thus obtained shows that the produced quantities are negatively correlated with a correlation of -85%. The uncertainties in the produced quantities are 3% and 2% respectively of the quantity produced by each producer. We also note that the variances are smaller now, as we no longer have uncertainties stemming from the demand side of the problem.

Now if we assume a more realistic scenario of the production costs being correlated (60% correlation), then we note that the produced quantity are negatively correlated with -62% correlation. The standard deviations in the produced quantities have also dropped to about 2.9% and 1.2%. Thus we not only obtain insight about the uncertainties in the output, but also the correlation between the output parameters. From an energy market policy maker's perspective this is crucial information as it helps identifying the regions where increase or decrease in production, consumption, price, pipeline flows and infrastructural expansions occur simultaneously and where they change asynchronously. Now we calculate the sensitivity of each of the input parameters to identify the parameter that causes maximum uncertainty in the output. The values for β for each of the four parameters γ_1, γ_2, a, b are calculated below.

$$\beta = \frac{1}{3}\begin{pmatrix}\sqrt{5} & \sqrt{5} & \sqrt{2} & \sqrt{369}\end{pmatrix}^T$$
$$= \begin{pmatrix}0.745 & 0.745 & 0.471 & 6.40\end{pmatrix}^T \tag{5.8}$$

Thus we see that the solution is more sensitive to the slope of the demand curve than to say production cost. Strictly speaking, if we define the variance in equilibrium as the sum of the variance of all output variables, this says, a unit increase in variance of the slope of the demand curve will be magnified about 41 times (6.4^2) variance in the equilibrium. However, a unit increase in the variance of the production cost only increases the variance in equilibrium by 0.556 (0.745^2) units.

5.2. Computational complexity

We used a Monte-Carlo based method as a comparison against our approximation method to compute covariance in the decision variables. To achieve this, we modeled the oligopoly complementarity problem mentioned in (5.2) varying the number of players, and hence the number of random parameters and the decision variables. For the Monte-Carlo simulation based approach,
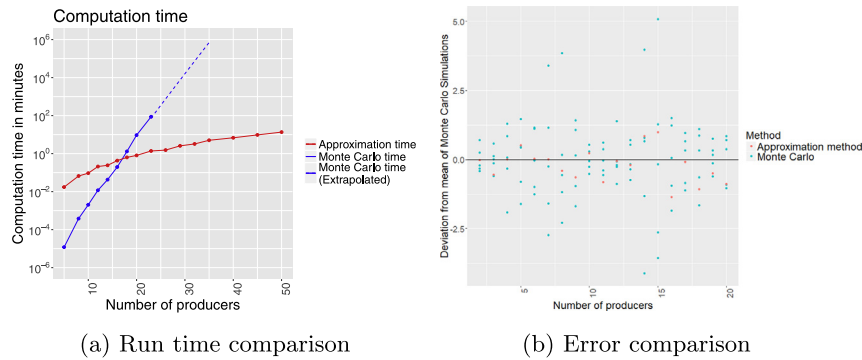
(a) Run time comparison

(b) Error comparison

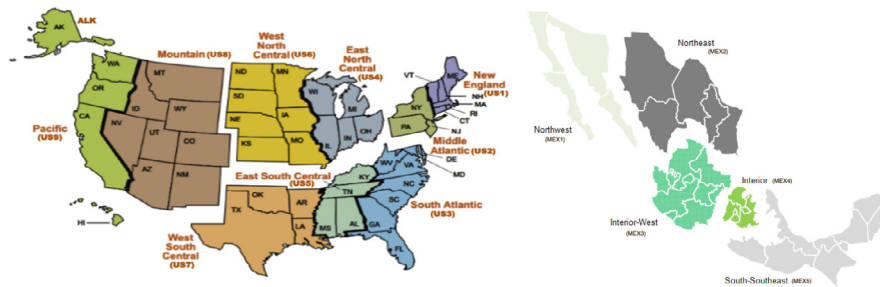**Fig. 2.** Computational experiments comparing Monte-Carlo methods and First-order approximation method.

Fig. 3. Regional disaggregation of United States and Mexico. Source: AEO (2015) and U.S. Energy Information Administration http://www.eia.gov/todayinenergy/detail.php?id=16471.

a symmetrically balanced stratified design (Shields, Teferra, Hapij, & Daddazio, 2015) is used with each dimension divided into two strata. With increasing number of random parameters and equilibrium variables, Monte-Carlo methods become increasingly inefficient as the number of simulations required grows exponentially. A comparison of the time taken in an *8 gigabyte RAM 1600 megahertz DDR3 2.5 gigahertz Intel Core i5* processor to solve the above oligopoly problem with varying number of players is shown in Fig. 2(a). Despite developments in algorithms to solve complementarity problems, the said exponential growth in the number of sample points required in a Monte-Carlo based approach deters the computational speed. A problem with as few as 25 uncertain variables takes about 2 hours to solve and one with 30 uncertain variables takes about seven days to solve using Monte-Carlo based approaches while it takes few seconds to minutes in the first-order approximation method. Fig 2(b) compares the error between 5 rounds of Monte-Carlo simulation and the first-order approximation method. More details on these computational experiments are provided in Appendix D.

6. Application to North American natural gas market

In Mexico, motivation to move from coal to cleaner energy sources creates an increasing trend in natural gas consumption, particularly in the power sector. Technology change, including fracking, has made natural gas available at a low cost. This resulted in increased production and higher proven reserves in the U.S. Therefore, the US is expected to become a net exporter of Natural Gas (increasing pipelines exports to Mexico and LNG) during the next years (EIA, 2016; Feijoo et al., 2016). The North American Natural Gas Model (NANGAM) developed in Feijoo et al. (2016) analyzes the impacts of cross border trade with Mexico. NANGAM models the equilibrium under various scenarios by competitively maximizing the profits of suppliers and pipeline operators, and the utility of consumers, resulting in a complementarity problem. The model also uses the Golombek function (Golombek, Gjelsvik,

& Rosendahl, 1995; Huppmann, 2013) to model the increase in marginal cost of production when producing close to capacity. The formal description of the model is provided in Appendix B.

For the model in this paper, which is motivated by NANGAM, we have disaggregated the United States into 9 census regions (US1-9) and Alaska (AEO, 2015). Mexico is divided into 5 regions (MEX1-5). A map showing this regional disaggregation is shown in Fig. 3. Further Canada is divided into two zones, Canada East (CAE) and Canada West (CAW). The model has 13 suppliers, 17 consumers, 17 nodes, and 7 time-steps. This amounts to 12,047 variables (primal and dual) and 2023 parameters. The gradient matrix of the complementarity function would contain $12,047^2$ elements and a Hessian matrix will have $12,047^3$ elements which is more than 1700 trillion floating point variables. We need efficient methods to handle these large objects. We observe, however, that the dependence of each component of the complementarity function is limited to few variables, thus making the gradient matrix sparse. Efficient sparse matrix tools in scipy (Jones, Oliphant, Peterson, & Others, 2015) are used along with a python class we specially built to handle a sparse multi-dimensional array. The details of this class are given in Appendix C. This model is calibrated to match the region-wise production and consumption data by adjusting the parameters of the demand curve, supply curve and the transportation cost. The source for the projected numbers are the same as the ones in Table 2 of Feijoo et al. (2016). The parameters of the demand curve were chosen in such a way that an elasticity of 0.29 is maintained at the solution to be consistent with EIA (2012).

6.1. Covariance matrix calibration

We used the method developed in Algorithm 1 to understand the propagation of uncertainty in the model. The covariance for each parameter across years is obtained by fitting a Wiener process to the parameter value. This is chosen to mimic the Markovian and independent increment properties of market parameters. Thus we
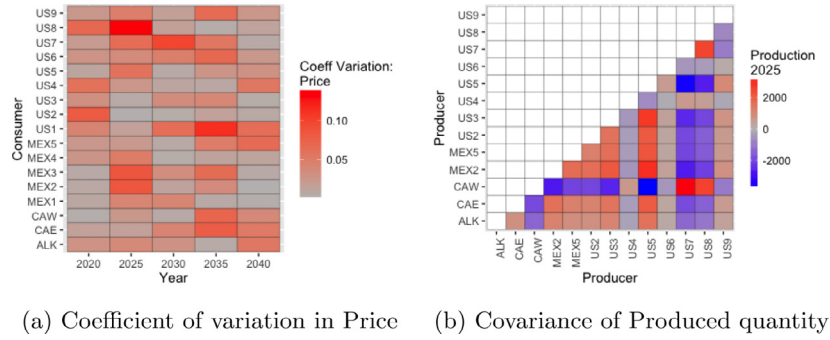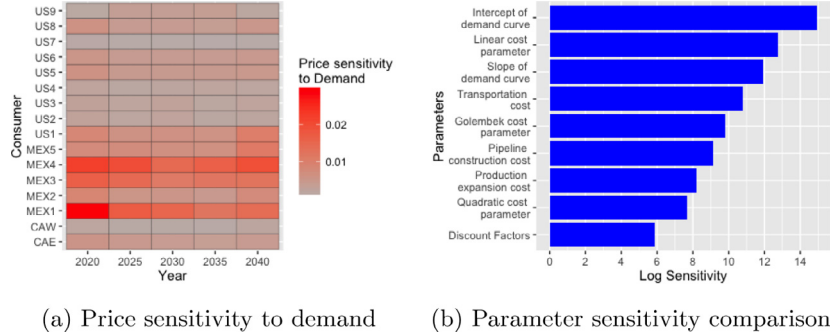
(a) Coefficient of variation in Price   (b) Covariance of Produced quantity

Fig. 4. Covariance results.

(a) Price sensitivity to demand   (b) Parameter sensitivity comparison

Fig. 5. Sensitivity results.

have for any parameter

$$d\theta(t) = d\mu_\theta(t) + \sigma_\theta dB(t) \qquad (6.1)$$

where μ_θ is calibrated, σ_θ is chosen to be 1% of the average value of μ_θ in the analyzed period and $B(t)$ is the standard Brownian motion. The diffusion parameter σ_θ is assumed to be independent of time. Additionally to understand the effect of greater uncertainty in US7, that accounts for about 40% of the total production in the continent, the parameters of production cost are assumed to have 5 times the variance than in any other region.

6.2. Results

The deterministic version of the problem is solved using the PATH algorithm (Dirkse & Ferris, 1995) by assuming a mean value for all random parameters. Following this, Algorithm 1 was applied and the $\mathcal{T}$ matrix defined in (2.52) is obtained by solving the linear system of equations using a Moore-Penrose pseudoinverse (Horn & Johnson, 2012). In the following paragraph, we discuss some of the results obtained in this study.

The heat map on Fig. 4(a) shows the coefficient of variation (standard deviation divided by mean) in consumer price in each year caused by the uncertainty in parameters as mentioned in Section 6.1. We notice that this large uncertainty in production costs of US7 caused relatively small uncertainties in the consumer price. This is partially due to the availability of resources in US8 and CAW to compensate for the large uncertainty in US7. The fact that it is actually US8 and CAW that compensate for this uncertainty is known by looking at the covariance plot on Fig. 4(b) which shows large correlation between US7 and US8 and also between US7 and CAW.

Fig. 5 shows the sensitivity of the solution to various input parameters. The graph on Fig. 5(a) shows the sum total change in uncertainty in price for a 1% fluctuation in the demand curve of consumers. We notice that the price is particularly sensitive to changes in demand in Mexico. This reflects the increasing concern

about growing exports (both LNG and pipeline) that are likely to result in higher consumer prices in the U.S. We also note that fluctuations in demand at nodes where production facilities are not available (MEX1, MEX3, MEX4) cause greater uncertainty in price. This is because, for regions with a production facility in the same node, the production facility produces more to cater the demand at that node and there is little effect in the flows and in the prices at other nodes. This is also contingent to sufficient pipeline capacity. Larger changes in demand for regions with limited pipeline capacity (e.g. MEX1) may result in major changes in price. However a perturbation to the demand at a node with no production unit causes the flows to alter to have its demand catered. This affects natural gas availability elsewhere and causes larger fluctuations in price. The tornado plot on Fig. 5(b) sorts the parameters in decreasing order of their effect on the uncertainty of the solution.

The plot Fig. 5(b) shows the total change of the equilibrium if a parameter (e.g., the demand intercept) is shifted by 1% from its original value. Note that the results are plotted in logarithmic scale and are sorted in decreasing order. In general, our results suggest that parameters of consumers and producers play a major role on the equilibrium and hence a small perturbation have a large effect on the solution. In particular, the intercept and slope of the demand curve and the linear cost parameter are the three most significant parameters. Interestingly, the expansion cost parameters (pipeline as well as production expansion) have a lower effect on the solution equilibrium. As it was described on Fig 5(a), uncertainties in demand significantly affect regions with no or low production capacities. Fig 5(b) corroborates that changes to the demand affects the equilibrium the most. The results also indicate that the infrastructure expansion happens independently of changes in expansion cost. Natural gas prices paid by consumers account for cost expansions and transportation. Therefore, the level of expansion is then driven by changes on demand. Hence, if policy changes need to be implemented in order to, for instance, to increase economic activity or reduce carbon emissions, respectively subsidizes or taxes the downstream or upstream ends of market

rather than the mid-stream players to have larger impacts. However, a policy maker who is interested in generating revenue without much impacts on the equilibrium should tax fuel transportation or infrastructure expansion for the greatest benefit.

7. Conclusion and future work

In this paper, we developed a method to approximate the covariance of the output of a large-scale nonlinear complementarity problem with random input parameters using first-order approximation methods. We extended this method to general optimization problems with equality constraints. We then developed sensitivity metrics for each of the input parameters quantifying their contribution to the uncertainty in the output. We used these tools to understand the covariance in the equilibrium of the North American natural gas Market. The method gave insights into how production, consumption, pipeline flows, prices would vary due to large uncertainties. While the variances identified the regions that are affected the most, the covariance gave information about whether the quantity will increase or decrease due to perturbation in the input. We also obtained results on the sensitivity of price uncertainty to demand uncertainty in various nodes. We then quantified the contribution of each input parameter to the uncertainty in the output. This in turn, helps in identifying the regions that can have large impacts on equilibrium.

We note that the method is particularly useful for large-scale nonlinear complementarity problems with a large number of uncertain parameters, which make Monte-Carlo simulations intractable. It is robust in approximating the solution covariance for small uncertainty in the inputs. It is also good in quantifying the sensitivity of the output (and its variance) to the variance of input parameters. However since all the above are obtained as an approximation based on first-order metrics, there is a compromise in the accuracy if the variances of the input are large. The method works the best for problems involving a large number of decision variables and random parameters with small variance.

We foresee expanding this work by using progressively higher order terms of the Taylor series to capture the nonlinearities more efficiently. To ensure computational feasibility, this would typically require us to have stronger assumptions on the sparsity of the Hessian and the higher-order derivatives. This will also require analysis and stronger assumptions about higher-order moments of the random parameters.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.ejor.2017.11.003.

References

Abada, I., Gabriel, S., Briat, V., & Massol, O. (2013). A generalized Nash–Cournot model for the Northwestern European natural gas markets with a fuel substitution demand function: The Gammes model. *Networks and Spatial Economics, 13*(1), 1–42. doi:10.1007/s11067-012-9171-5.

AEO, A. E. O. (2015). US Energy Information Administration, Annual Energy Outlook (2015) URL:https://www.eia.gov/outlooks/aeo/pdf/0383(2015).

Agrawal, S., Ding, Y., Saberi, A., & Ye, Y. (2012). Price of correlations in stochastic optimization. *Operations Research, 60*(1), 150–162. doi:10.1287/opre.1110.1011.

Anderson, E. J., & Xu, H. (2004). Nash equilibria in electricity markets with discrete prices. *Mathematical Methods of Operations Research, 60*(2), 215–238.

Ben-Israel, A., & Greville, T. N. (2003). *Generalized inverses: Theory and applications*: vol. 15. Springer Science Business Media.

Benedetti-Cecchi, L. (2003). The importance of the variance around the mean effect size of ecological processes. *Ecology, 84*(9), 2335–2346.

Castillo, E., Conejo, A., Castillo, C., Mínguez, R., & Ortigosa, D. (2006). Perturbation approach to sensitivity analysis in mathematical programming. *Journal of Optimization Theory and Applications, 128*(1), 49–74.

Castillo, E., Conejo, A. J., & Aranda, E. (2008). Sensitivity analysis in calculus of variations. some applications. *SIAM Review, 50*(2), 294–312.

Chen, X., & Fukushima, M. (2005). Expected residual minimization method for stochastic linear complementarity problems. *Mathematics of Operations Research, 30*(4), 1022–1038. doi:10.1287/moor.1050.0160.

Chen, Y., Cowling, P., Polack, F., Remde, S., & Mourdjis, P. (2017). Dynamic optimisation of preventative and corrective maintenance schedules for a large scale urban drainage system. *European Journal of Operational Research, 257*(2), 494–510.

Christensen, A., & Siddiqui, S. (2015). A mixed complementarity model for the US biofuel market with federal policy interventions. *Biofuels, Bioproducts and Biorefining, 9*(4), 397–411.

Cottle, R. W., Pang, J. S., & Stone, R. E. (2009). *The linear complementarity problem*: vol. 60. SIAM.

Dirkse, S. P., & Ferris, M. C. (1995). The path solver: A nommonotone stabilization scheme for mixed complementarity problems. *Optimization Methods and Software, 5*(2), 123–156. doi:10.1080/10556789508805606.

Egging, R., Pichler, A., Kalvø, Ø. I., & WalleHansen, T. M. (2016). Risk aversion in imperfect natural gas markets. *European Journal of Operational Research*. doi:10.1016/j.ejor.2016.10.020.

EIA, E. I. A. (2012). Fuel competition in power generation and elasticities of substitution. https://www.eia.gov/analysis/studies/fuelelasticities/pdf/eia-fuelelasticities.pdf.

EIA, E. I. A. (2016). U.S natural gas exports to Mexico continue to grow. http://www.eia.gov/todayinenergy/detail.php?id=28932&src=email.

Facchinei, F., & Pang, J. S. (2007). *Finite-dimensional variational inequalities and complementarity problems*. Springer Science & Business Media.

Feijoo, F., Huppmann, D., Sakiyama, L., & Siddiqui, S. (2016). North American natural gas model: Impact of cross-border trade with Mexico. *Energy, 112*, 1084–1095. doi:10.1016/j.energy.2016.06.133.

Ferris, M. C., & Pang, J. S. (1997). Engineering and economic applications of complementarity problems. *SIAM Review, 39*(4), 669–713. doi:10.1137/S0036144595285963.

Fiacco, A. V. (2009). Sensitivity and stability in NLP: approximation Sensitivity and Stability in NLP: Approximation. In C. A. Floudas, & P. M. Pardalos (Eds.), *Encyclopedia of Optimization* (pp. 3454–3467). Boston, MA: Springer US.

Gabriel, S. A., Conejo, A. J., Fuller, J. D., Hobbs, B. F., & Ruiz, C. (2012). *Complementarity modeling in energy markets*: vol. 180. Springer Science & Business Media.

Gabriel, S. A., Kydes, A. S., & Whitman, P. (2001). The national energy modeling system: A large-scale energy-economic equilibrium model. *Operations Research, 49*(1), 14–25. doi:10.1287/opre.49.1.14.11195.

Gabriel, S. A., Zhuang, J., & Egging, R. (2009). Solving stochastic complementarity problems in energy market modeling using scenario reduction. *European Journal of Operational Research, 197*(3), 1028–1040. doi:10.1016/j.ejor.2007.12.046.

Golombek, R., Gjelsvik, E., & Rosendahl, K. E. (1995). Effects of liberalizing the natural gas markets in Western Europe. *The Energy Journal*, 85–111.

Gürkan, G., & Robinson, S. M. (1999). Sample-path solution of stochastic variational inequalities. *Mathematical Programming, Series B, 84*(2), 313–333. doi:10.1007/s10107980038a.

Horn, R. A., & Johnson, C. R. (2012). *Matrix analysis*. Cambridge University Press.

Huppmann, D. (2013). Endogenous production capacity investment in natural gas market equilibrium models. *European Journal of Operational Research, 231*(2), 503–506.

Huppmann, D., & Egging, R. (2014). Market power, fuel substitution and infrastructure–a large-scale equilibrium model of global energy markets. *Energy, 75*, 483–500.

Hyett, K. L., Podosky, M., Santamaria, N., & Ham, J. C. (2007). Valuing variance: The importance of variance analysis in clinical pathways utilisation. *Australian Health Review, 31*(4), 565–570.

Jiang, H., & Xu, H. (2008). Stochastic approximation approaches to the stochastic variational inequality problem. *IEEE Transactions on Automatic Control, 53*(6), 1462–1475. doi:10.1109/TAC.2008.925853.

Jones, E., Oliphant, T., Peterson, P., Others (2015). {SciPy}: Open source scientific tools for {Python}. http://www.scipy.org/.

Kopanos, G. M., Méndez, C. A., & Puigjaner, L. (2010). MIP-based decomposition strategies for large-scale scheduling problems in multiproduct multistage batch plants: A benchmark scheduling problem of the pharmaceutical industry. *European journal of operational research, 207*(2), 644–655.

Krantz, S. G., & Parks, H. R. (2012). *The implicit function theorem: History, theory, and applications*. Springer Science Business Media.

Kwak, B. M., & Lee, S. S. (1988). A complementarity problem formulation for two-dimensional frictional contact problems. *Computers & structures, 28*(4), 469–480.

Lamm, M., Lu, S., & Budhiraja, A. (2016). Individual confidence intervals for solutions to expected value formulations of stochastic variational inequalities. *Mathematical Programming*, 1–46. doi:10.1007/s10107-016-1046-y.

Luo, J., Hong, L. J., Nelson, B. L., & Wu, Y. (2015). Fully sequential procedures for large-scale ranking-and-selection problems in parallel computing environments. *Operations Research, 63*(5), 1177–1194. doi:10.1287/opre.2015.1413.

Martín, S., Smeers, Y., & Aguado, J. A. (2015). A stochastic two settlement equilibrium model for electricity markets with wind generation. *IEEE Transactions on Power Systems, 30*(1), 1–13. doi:10.1109/TPWRS.2014.2326334.

Nocedal, J., & Wright, S. (2006). *Numerical optimization*. Springer Science & Business Media.

Ohno, K., Boh, T., Nakade, K., & Tamura, T. (2016). New approximate dynamic programming algorithms for large-scale undiscounted Markov decision processes and their application to optimize a production and distribution system. *European Journal of Operational Research, 249*(1), 22–31.

Oke, O., Huppmann, D., Marshall, M., Poulton, R., Siddiqui, S. (2016). Mitigating environmental and public-safety risks of United States crude-by-rail transport. DIW Discussion Papers, No. 1575.

Seber, G. A. F., & Lee, A. J. (2012). *Linear regression analysis*: vol. 936. John Wiley & Sons.

Shanbhag, U. V. (2013). *Stochastic variational inequality problems : Applications, analysis, and algorithms*. INFORMS.

Shields, M. D., Teferra, K., Hapij, A., & Daddazio, R. P. (2015). Refined stratified sampling for efficient Monte Carlo based uncertainty quantification. *Reliability Engineering & System Safety, 142*, 310–325.

Siddiqui, S., & Christensen, A. (2016). Determining energy and climate market policy using multiobjective programs with equilibrium constraints. *Energy, 94*, 316–325.

Yumashev, D., & Johnson, P. (2017). Flexible decision making in the wake of large scale nuclear emergencies: Long-term response. *European Journal of Operational Research*.