Modeling Context in Cognition using Variational Inequalities

Ian Gemp and **Sridhar Mahadevan**
School of Computer Science
University of Massachusetts
Amherst, MA 01003
(imgemp,mahadeva)@cs.umass.edu

Abstract

Important aspects of human cognition, like creativity and play, involve dealing with multiple divergent views of objects, goals, and plans. We argue in this paper that the current model of *optimization* that drives much of modern machine learning research is far too restrictive a paradigm to mathematically model the richness of human cognition. Instead, we propose a much more flexible and powerful framework of *equilibration*, which not only generalizes optimization, but also captures a rich variety of other problems, from game theory, complementarity problems, network equilibrium problems in economics, and equation solving. Our thesis is that creative activity involves dealing not with a *single objective function*, which optimization requires, but rather balancing multiple divergent and possibly contradictory goals. Such modes of cognition are better modeled using the framework of variational inequalities (VIs). We provide a brief review of this paradigm for readers unfamiliar with the underlying mathematics, and sketch out how VIs can account for creativity and play in human and animal cognition.

1 Introduction

In this short paper, we briefly outline a novel framework for changing perspectives, such as occur in creative activity and play, using the mathematical framework of *variational inequalities* (VIs) (Nagurney 1999; Facchinei and J. 2003). The symposium goals discuss the motivation of modeling contextual effects in human cognition, whereby the same stimulus can appear in multiple divergent views, depending on the goals of the observer. The argument, made informally in the symposium description, that current AI approaches rely on searching for a single "optimal" answer, and hence fail to accurately model the richness of human cognition finds much resonance in our recent work on using the framework of VIs to introduce a novel framework for machine learning. It is exactly this reliance on strict optimization that motivated our work. In this document, we outline why VIs provide a much richer framework for modeling context dependencies in human cognition. Our treatment of VI theory will of necessity be brief, but we hope to elaborate the main conceptual ideas in sufficient depth to make it clear why this framework is a fundamental extension of the current optimization based approach to AI and machine learning, as elaborated in standard texts.

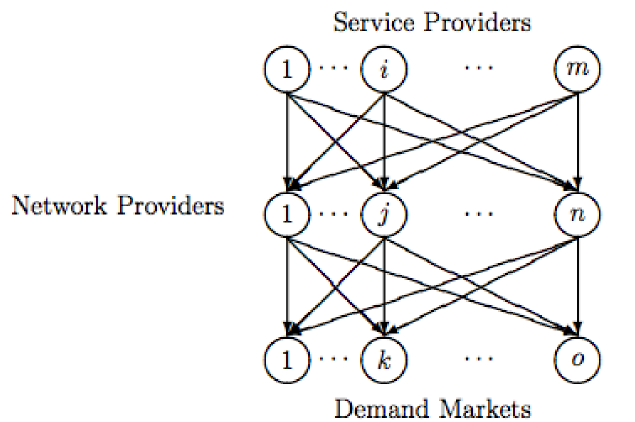

Figure 1: Network model of a service-oriented internet proposed in (Nagurney and Wolf 2014) in which service providers and network providers compete on the basis of quantity, quality, and price to maximize profit in the presence of demand markets.

Figure 1 illustrates our approach with an example involving a complex sort of game played on a recent economic model of the next-generation Internet, whereby content providers like Netflix play a *Cournot-Nash* game and transport providers like Verizon play a *Betrand* game. We have recently developed an algorithm to solve such networked systems, which produces state of the art results[1]. These game-theoretic models are well-established in the economics literature, and seek to reinforce our argument that VIs provide the right framework for modeling complex context-sensitive processes in networked systems like the Internet. In this networked economic model, there is no *single* objective function being optimized, as such a unique criterion does not exist. Each player in the network, whether it be a content provider or a transport provider, has at their disposal a set of levers they can adjust to achieve their unique objectives (e.g., content providers adjust quantity delivered to maximize their profits and transport providers adjust qual-

[1]This work is currently under blind review, and cannot be cited.

ity and price of delivery to maximize their profits). These criteria are often in conflict. The only available solution to such networked systems is an *equilibrium* solution, which the theory of VIs provides a rich framework to represent and compute. Our principal thesis is that this is a desirable mathematical framework for not only networked economic systems, but rather carries over to modeling of human cognition as well.

2 Brief Review of Variational Inequalities

Variational inequalities (VIs) are a rich and expressive mathematical framework that generalize convex optimization to handle problems with asymmetric Jacobians, while preserving nice properties like convexity, and being able to solve a vast array of diverse equilibrium problems, from game theory to complementarity problems and nonlinear solutions of equations. Originally proposed by Hartman and Stampacchia (Hartman and Stampacchia 1966) in the context of solving partial differential equations in mechanics, VIs gained popularity in the finite-dimensional setting partly as a result of Dafermos (Dafermos 1980), who showed that the traffic network equilibrium problem could be formulated as a finite-dimensional VI. This advance inspired much follow-on research, showing that a variety of equilibrium problems in economics, game theory, manufacturing etc. could also be formulated as finite-dimensional VIs – the books by Nagurney (Nagurney 1999) and Facchinei and Pang (Facchinei and J. 2003) provide a detailed introduction to the theory and applications of finite-dimensional VIs.

As many readers may be unfamiliar with the mathematics of VIs, we begin with a brief review. The formal definition of a VI is as follows:

Definition 1. *The finite-dimensional variational inequality problem VI(F,K) involves finding a vector* $x^* \in K \subset \mathbb{R}^n$ *such that*

$$\langle F(x^*), x - x^* \rangle \geq 0, \ \forall x \in K$$

where $F : K \rightarrow \mathbb{R}^n$ *is a given continuous function and* K *is a given closed convex set, and* $\langle ., . \rangle$ *is the standard inner product in* $\mathbb{R}^n$*.*

Figure 2 provides a geometric interpretation of a variational inequality.[2] The following general result characterizes when solutions to VIs exist:

Theorem 1. *Suppose* K *is compact, and that* $F : K \rightarrow \mathbb{R}^n$ *is continuous. Then, there exists a solution to VI(*F, K*).*

As Figure 2 shows, x^* is a solution to $VI(F, K)$ if and only if the angle between the vectors $F(x^*)$ and $x - x^*$, for any vector $x \in K$, is less than or equal to 90^0. To build up some intuition, the reduction of a few well-known problems to a VI is now provided.

Theorem 2. *Let* x^* *be a solution to the optimization problem of minimizing a continuously differentiable function* $f(x)$*, subject to* $x \in K$*, where* K *is a closed and convex set. Then,* x^* *is a solution to* $VI(\nabla f, K)$*, such that* $\langle \nabla f(x^*), x - x^* \rangle \geq 0, \ \forall x \in K$*.*

[2] In Figure 2, the normal cone $C(x^*)$ at the vector x^* of a convex set K is defined as $C(x^*) = \{y \in \mathbb{R}^n | \langle y, x - x^* \rangle \leq 0, \forall x \in K\}$.

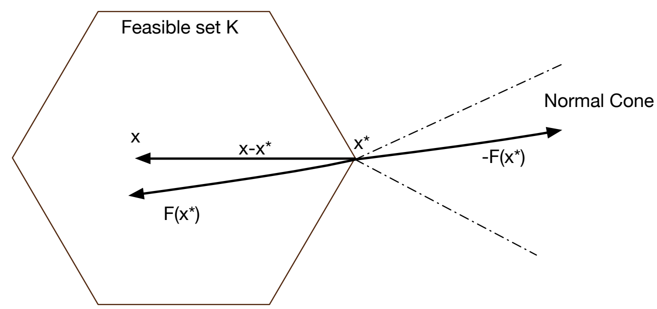

Figure 2: This figure provides a geometric interpretation of the variational inequality $VI(F, K)$. The mapping F defines a vector field over the feasible set K such that at the solution point x^*, the vector field $F(x^*)$ is directed inwards at the boundary, and $-F(x^*)$ is an element of the normal cone $C(x^*)$ of K at x^*.

Proof: Define $\phi(t) = f(x^* + t(x - x^*))$. Since $\phi(t)$ is minimized at $t = 0$, it follows that $0 \leq \phi'(0) = \langle \nabla f(x^*), x - x^* \rangle \geq 0, \ \forall x \in K$, that is x^* solves the VI. □

Theorem 3. *If* $f(x)$ *is a convex function, and* x^* *is the solution of* $VI(\nabla f, K)$*, then* x^* *minimizes* f*.*

Proof: Since f is convex, it follows that any tangent lies below the function, that is $f(x) \geq f(x^*) + \langle \nabla f(x^*), x - x^* \rangle, \ \forall x \in K$. But, since x^* solves the VI, it follows that $f(x^*)$ is a lower bound on the value of $f(x)$ everywhere, or that x^* minimizes f. □

Crucially, VI problems *cannot* be converted back into optimization problems, unless a very restrictive condition is met on the Jacobian of the mapping F.

Theorem 4. *Assume* $F(x)$ *is continuously differentiable on* K *and that the Jacobian matrix* $\nabla F(x)$ *of partial derivatives of* $F_i(x)$ *with respect to (w.r.t) each* x_j *is symmetric and positive semidefinite. Then there exists a real-valued convex function* $f : K \rightarrow \mathbb{R}$ *satisfying* $\nabla f(x) = F(x)$ *with* x^**, the solution of VI(F,K), also being the mathematical programming problem of minimizing* $f(x)$ *subject to* $x \in K$*.*

3 VI Algorithms

Section 3 reviews projection-based algorithms for solving VIs, including the popular extragradient method (Korpelevich 1977).

Projection-Based Algorithms for VIs

The basic projection-based method (Algorithm 1) for solving VIs is one of the simplest methods, which relies on the fixed point characterization of VIs.

Here, $\Pi_{K,D}$ is the projector onto convex set K with respect to the natural norm induced by D, where $\|x\|_D^2 = \langle x, Dx \rangle$. It can be shown that the basic projection algorithm solves any $VI(F, K)$ for which the mapping F is

Algorithm 1 The Basic Projection Algorithm for solving VIs.

INPUT: Given VI(F,K), and a symmetric positive definite matrix D.

1: Set $k = 0$ and $x_k \in K$.
2: **repeat**
3: Set $x_{k+1} \leftarrow \Pi_{K,D}(x_k - D^{-1}F(x_k))$.
4: Set $k \leftarrow k + 1$.
5: **until** $x_k = \Pi_{K,D}(x_k - D^{-1}F(x_k))$.
6: Return x_k

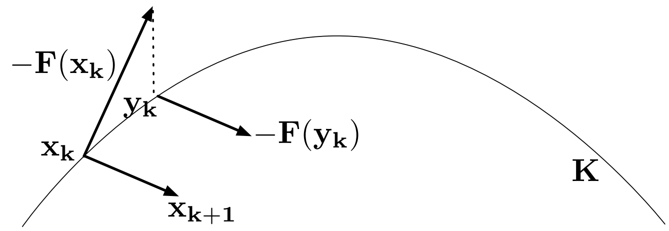

Figure 3: Right: One iteration of the extradient algorithm.

strongly monotone [3] and *Lipschitz*.[4] A simple strategy is to set $D = \alpha I$, where $\alpha > \frac{L^2}{2\mu}$, and L is the Lipschitz smoothness constant, and μ is the strong monotonicity constant. The basic projection-based algorithm has two critical limitations. First, it requires that the mapping F be strongly monotone. If, for example, F is the gradient map of a continuously differentiable function, strong monotonicity implies the function must be strongly convex. Second, setting the parameter α requires knowing the Lipschitz smoothness L and the strong monotonicity parameter μ. The extragradient method of Korpolevich (Korpelevich 1977) addresses some of these concerns, and is defined as Algorithm 2 below.

Algorithm 2 The Extragradient Algorithm for solving VIs.

INPUT: Given VI(F,K), and a scalar α.

1: Set $k = 0$ and $x_k \in K$.
2: **repeat**
3: Set $y_k \leftarrow \Pi_K(x_k - \alpha F(x_k))$.
4: Set $x_{k+1} \leftarrow \Pi_K(x_k - \alpha F(y_k))$.
5: Set $k \leftarrow k + 1$.
6: **until** $x_k = \Pi_K(x_k - \alpha F(x_k))$.
7: Return x_k

The extragradient algoriithm derives its name from the property that it requires an "extra gradient" step (step 4 in Algorithm 2), unlike the basic projection algorithm given earlier as Algorithm 1. The principal advantage of the extragradient method is that it can be shown to converge under a considerably weaker condition on the mapping F, which now has to be merely monotonic: $\langle F(x) - F(y), x - y\rangle \geq 0$. The earlier Lipschitz condition is still necessary for convergence.

The extragradient algorithm has been the topic of much attention in solving VIs since it was proposed, e.g., see (Iusem and Svaiter 1997; Khobotov 1987; Marcotte 1991; Peng and Yao 2008; Nesterov 2007; Solodov and Svaiter 1999). Khobotov (Khobotov 1987) proved that the extragradient method converges under the weaker requirement of *pseudo-monotone* mappings, [5] when the learning rate is automatically adjusted based on a local measure of the Lipschitz constant. Iusem (Iusem and Svaiter 1997) proposed a variant whereby the current iterate was projected onto a hyperplane separating the current iterate from the final solution, and subsequently projected from the hyperplane onto the feasible set. Solodov and Svaiter (Solodov and Svaiter 1999) proposed another hyperplane method, whereby the current iterate is projected onto the intersection of the hyperplane and the feasible set. Finally, the extragradient method was generalized to the non-Euclidean case by combining it with the mirror-descent method (Nemirovksi and Yudin 1983), resulting in the so-called "mirrror-prox" algorithm (Juditsky and Nemirovski 2011).

[3] A mapping F is *strongly monotone* if $\langle F(x) - F(y), x - y\rangle \geq \mu\|x - y\|_2^2, \mu > 0, \forall x, y \in K$.

[4] A mapping F is *Lipschitz* if $\|F(x) - F(y)\|_2 \leq L\|x - y\|_2, \forall x, y \in K$.

[5] A mapping F is *pseudo-monotone* if $\langle F(y), x - y\rangle \geq 0 \Rightarrow \langle F(x), x - y\rangle \geq 0$, $\forall x, y \in K$.

4 Context Modeling using VIs

Finally, we sketch the application of the above VI theory and algorithms to the modeling of human cognition. At the heart of our approach is the reliance on VIs to model multiple interacting goals in a network to achieve equilibrium solutions. This is nicely illustrated by our first example of the service-oriented economic model of the Internet, whereby content providers and transport providers are both optimizing different objectives. In such networked equilibrium problems, there is no notion of an *optimal* solution. Instead, what's possible is to find *equilibrium solutions*, whereby each player can optimize relative to choices made by the other agents in the system. In much the same way, context effects in human cognition, which lie at the heart of goal-oriented perception, curiosity and play can be captured by the VI framework. We briefly sketch out several examples of how the VI framework allows modeling each of these phenomena.

Context Effects in Perception

The organizers ask us to imagine "how might we model divergently unconventional perspectives in a top-down fashion in robotics"? A similar argument made with regard to the networked Internet economic model shown in the first figure carries over to human cognition, where the world appears as a series of interconnected objects to a viewer. An office appears as such not because of a single object, but due to the multitude of objects spatially arranged in a certain networked pattern (e.g., books on shelves, computer on desk, chairs on floor etc.). Human cognition is acutely sensitive to not just the presence of individual objects, but rather to the context of objects appearing in spatial proximity. We can model the process of arriving at divergent views of a given scene, therefore, using the process of finding equilibria in networked systems, a singular ability of the VI framework. Depending on the nature of the equilibrium solution that is found of a networked model of a scene, divergent perspectives of the whole scene emerge. Thus, a house may not be recognized as a “house” because the individual components of the network may “settle” on a divergent view of its parts.

Curiosity and Play

Another example we are asked to consider is how “...in play, an agent deliberately projects a conceptual organization onto an object for which it is not conventionally suited (e.g. think of how children turn boring household objects into exciting props for their flights of fancy). In this playful exploration, backgrounded properties of an object may now become salient, leading to creative insights and solutions for future problems.” Once again, to model such a situation, the use of VIs is especially germane because individual features of an object can now be considered the elements of the network, rather than multiple objects. In the first case above, the entire scene was modeled as a network. Here, an individual object, say a ”chair” or a ”bottle” is modeled as a network of components or features, and the VI framework is used to compute an equilibrium solution of the network “flows”, much as in the service-oriented network example.

Social Cognition

Similar to the argument made above, social cognition may come about from the competition between multiple goals. Defining a single objective function for something as abstract as a social attribute may be impossible, however, a concept like empathy may arise from the interplay of multiple objectives. It’s not only this interplay that could make such complex mechanisms possible, but also the existence of a feasible region that constrains the meanings available to the agent forcing equilibria to exist on boundaries that it would not have typically. This notion of equilibria exhibiting meanings in increasingly complex networks draws from ideas popularized by Douglas Hofstadter in “Gödel, Escher, Bach: An Eternal Golden Braid” and more recently by the study of supernetworks (Nagurney and Dong 2002) in VIs.

Cognition and Divergent Perspectives

Here, we introduce a novel technique for modeling divergent cognitive activity. This idea is motivated by interpretations of the brain as a nonlinear dynamical system, a subject that has undergone deep examination over the past 15 years. Recent dynamical models are able to explain cognitive activity as the “sequential switching between different metastable cognitive states” (Rabinovich et al. 2008). These systems exhibit complex phase space stability and take advantage of stable limit cycles to maintain cognitive states. Stable limit cycles are not fixed points though and hence cannot coincide with the solution to a variational inequality. Although VI’s expressibility in modeling cognition is limited in this way, it’s not an immediate drawback. We should keep the initial transition from optimization to equilibration simple and so shrinking the set of possible dynamics is helpful.

To begin, we need to briefly paint the picture of VI’s sister theory, projected dynamical systems (PDS) - see (Nagurney and Zhang 1996) for details. PDS is essentially VIs from an algorithmic perspective. Imagine initializing the projection based method for VIs with some initial vector $x_0 \in K$. The algorithm will generate a sequence of successive x_k’s forming a trajectory that ultimately converges at a fixed point (assuming x_0 was in the neighborhood of a stable equilibrium). Now, run that algorithm for all the possible initial points in K and you’ll have a clear view of the dynamics enforced by F. Equivalently, you can simply view the vector field, F, and deduce the behavior of the system.

The service-oriented internet example showed how networks of nodes governed by profit maximizing forces could define a vector valued map F with a corresponding equilibrium. Here, we start with the map F and assume a suitable network representation with corresponding governing forces exists. For instance, the network could be a web of neurons competing to emphasize concepts in the mind of the observer by indirectly controlling nerve impulses through neurotransmitter releases at synapses. Although a representation like this one may aid intuition, in the context of machine learning, the existence of a such a network is unnecessary to the validity of the model; it is only a pedagogical tool.

We will now give a simplistic example of how VIs might be used to model diverging perspectives. Let $F \in \mathbb{R}^2$ be defined as follows where A is a square 2×2 matrix and b is a 2×1 vector. F and x are then 2×1. The feasible set K is a unit box in the second quadrant of $\mathbb{R}^2$.

$$\langle F(x^*), x - x^* \rangle \geq 0,\ \forall x \in K$$
$$F(x) = Ax + b$$
$$A = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}$$

For simplicity, we’ll take b to be a vector of zeros. The entries in A will correspond to features a_i, or activation levels, gathered from data. The vector x corresponds to a phase space that is abstracted a level above the features a_i. For instance, the features could be indicator variables for the presence of objects in a scene weighted by their inherent importance relative to the other objects as well as the confidence levels that those objects have, in fact, been identified correctly. For our example, let a_1, a_2, a_3, and a_4 function in this way and represent a frying pan, a bowl of vegetables, a burglar, and a bottle of oil respectively. Furthermore, let the elements in x represent the concepts “cooking” and “burglary” on the x-axis and y-axis respectively. A high value means that concept is active in the scene while a low value means that concept is dormant.

In the matrix displayed in Figure 4, features a_1, a_2, and a_4 are all nonzero meaning they are activated in the current sample. The burglar, feature a_3, however, is not present. The resulting dynamics of this system are such that the bottom right corner ($x = [0, 0]^T$) is a stable fixed point meaning the concept “cooking” is active.

In Figure 5, feature a_3 is slightly activated. The features a_1, a_2, and a_4 are still the same. The overall dynamics of

$$A = \begin{bmatrix} -1 & -2 \\ 0 & -0.5 \end{bmatrix}$$

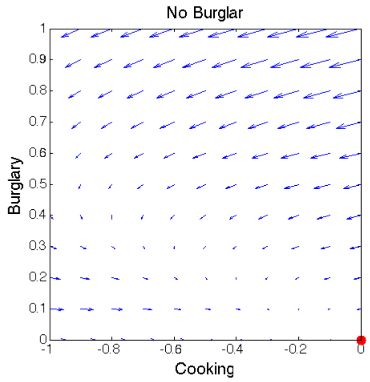

Figure 4: Dynamics of the F mapping created by the appearance of various objects in the scene. The corresponding matrix A is displayed below the phase-space plot. In this example, the burglar is not present ($a_3 = 0$).

$$A = \begin{bmatrix} -1 & -2 \\ -0.1 & -0.5 \end{bmatrix}$$

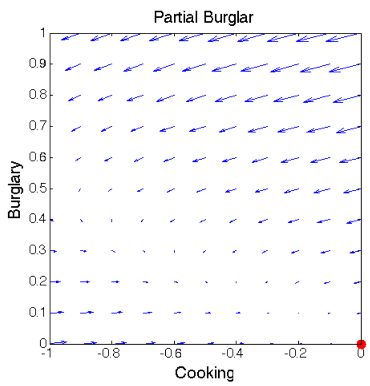

Figure 5: Dynamics of the F mapping created by the appearance of various objects in the scene. The corresponding matrix A is displayed below the phase-space plot. In this example, the burglar feature is marginally active ($a_3 = -0.1$).

the system have not changed and the vector x still implies the scene is "cooking" related.

In the final sample, Figure 6, the burglar feature is highly active and has completely changed the dynamics of the system rendering the original fixed point unstable. The vector x follows a trajectory diverging away from the original fixed point across the phase space and settles at a new fixed point in the top left corner against the boundary. The location of this new fixed point in the phase space is highly associated with "burglary".

This is a very basic example. One can imagine a larger phase space with multiple equilibria and more complex dynamics. Ingrained perspectives could become divergent once active features cross critical thresholds. Once that happens, x wanders the phase space (creativity and play) through possibly meaningless regions before converging on a new cognitive state (unconventional perspective).

Furthermore, the features need not be completely independent of the observer. If, for instance, the observer is sus-

$$A = \begin{bmatrix} -1 & -2 \\ -3 & -0.5 \end{bmatrix}$$

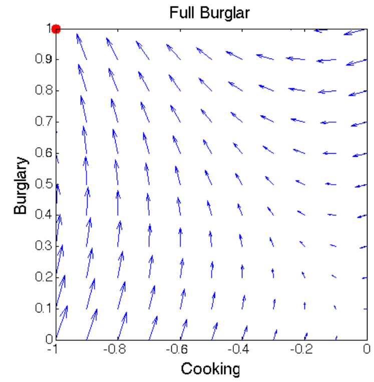

Figure 6: Dynamics of the F mapping created by the appearance of various objects in the scene. The corresponding matrix A is displayed below the phase-space plot. In this example, the burglar feature is sufficiently active ($a_3 = -3$).

picious of an intruder or naturally cautious in general, the "burglar" feature, a_3, may be weighted more heavily. This would cause the observer to change their perception of the scene at a lower critical threshold. We could then create goal-oriented perception by subscribing to various feature weights representing different mental lenses.

5 Conclusion

We argued that the current model of optimization is too restrictive a framework to capture human cognition. It considers only a single objective function, and from a dynamical systems standpoint, it admits only a small family of possible dynamical behaviors. The theory of variational inequalities, on the other hand, provides additional flexibility to support multiple possibly conflicting objectives as well as an extended family of dynamical behaviors enabling a richer paradigm to model cognitive activity. We also presented several explanations including a basic example for how VIs might be used to model curiosity, play, social cognition, divergent perspectives, and goal-oriented perception.

References

Dafermos, S. 1980. Traffic equilibria and variational inequalities. *Transportation Science* 14:42–54.

Facchinei, F., and J., P. 2003. *Finite-Dimensional Variational Inequalities and Complimentarity Problems*. Springer.

Hartman, P., and Stampacchia, G. 1966. On some nonlinear elliptic differential functional equations. *Acta Mathematica* 115:271–310.

Iusem, A., and Svaiter, B. 1997. A variant of Korpelevich's method for variational inequalities with a new search strategy. *Optimization* 42:309–321.

Juditsky, A., and Nemirovski, A. 2011. First order methods for nonsmooth convex large-scale optimization, i: General purpose methods. In *Optimization in Machine Learning*. MIT Press.

Khobotov, E. 1987. Modification of the extragradient method for solving variational inequalities of certain optimization problems. *USSR Computational Mathematics and Mathematical Physics* 27:120–127.

Korpelevich, G. 1977. The extragradient method for finding saddle points and other problems. *Matekon* 13:35–49.

Marcotte, P. 1991. Application of Khobotov's algorithm to variational inequalities and network equilibrium problems. *INFORM* 29.

Nagurney, A., and Dong, J. 2002. Supernetworks: Decision-making for the information age, 2002.

Nagurney, A., and Wolf, T. 2014. A Cournot–Nash–Bertrand game theory model of a service-oriented internet with price and quality competition among network transport providers. *Computational Management Science* 11(4):475–502.

Nagurney, A., and Zhang, D. 1996. *Projected Dynamical Systems and Variational Inequalities with Applications*. Kluwer Academic Press.

Nagurney, A. 1999. *Network Economics: A Variational Inequality Approach, second and revised edition*. Kluwer Academic Press.

Nemirovksi, A., and Yudin, D. 1983. *Problem Complexity and Method Efficiency in Optimization*. John Wiley Press.

Nesterov, Y. 2007. Dual extrapolation and its application to solving variational inequalities and related problems. *Mathematical Programming Series B.* 109:319–344.

Peng, J., and Yao, J. 2008. A new hybrid extragradient method for generalized mixed equilibrium problems, fixed point problems, and variational inequality problems. *Taiwanese Journal of Mathematics* 12:1401–1432.

Rabinovich, M. I.; Huerta, R.; Varona, P.; and Afraimovich, V. S. 2008. Transient cognitive dynamics, metastability, and decision making. *PLoS computational biology* 4(5):e1000072.

Solodov, M., and Svaiter, B. 1999. A new projection method for variational inequality problems. *SIAM Journal of Control and Optimization* 37(3):756–776.