Preface

The finite-dimensional nonlinear complementarity problem (NCP) is a system of finitely many nonlinear inequalities in finitely many nonnegative variables along with a special equation that expresses the complementary relationship between the variables and corresponding inequalities. This complementarity condition is the key feature distinguishing the NCP from a general inequality system, lies at the heart of all constrained optimization problems in finite dimensions, provides a powerful framework for the modeling of equilibria of many kinds, and exhibits a natural link between smooth and nonsmooth mathematics. The finite-dimensional variational inequality (VI), which is a generalization of the NCP, provides a broad unifying setting for the study of optimization and equilibrium problems and serves as the main computational framework for the practical solution of a host of continuum problems in the mathematical sciences.

The systematic study of the finite-dimensional NCP and VI began in the mid-1960s; in a span of four decades, the subject has developed into a very fruitful discipline in the field of mathematical programming. The developments include a rich mathematical theory, a host of effective solution algorithms, a multitude of interesting connections to numerous disciplines, and a wide range of important applications in engineering and economics. As a result of their broad associations, the literature of the VI/CP has benefited from contributions made by mathematicians (pure, applied, and computational), computer scientists, engineers of many kinds (civil, chemical, electrical, mechanical, and systems), and economists of diverse expertise (agricultural, computational, energy, financial, and spatial). There are many surveys and special volumes, [67, 240, 243, 244, 275, 332, 668, 687], to name a few.

Written for novice and expert researchers and advanced graduate students in a wide range of disciplines, this two-volume monograph presents a comprehensive, state-of-the-art treatment of the finite-dimensional variational inequality and complementarity problem, covering the basic theory, iterative algorithms, and important applications. The materials presented herein represent the work of many researchers worldwide. In undertaking this ambitious project, we have attempted to include every major aspect of the VI/CP, beginning with the fundamental question of existence and uniqueness of solutions, presenting the latest algorithms and results, extending into selected neighboring topics, summarizing many classical source problems, and including novel application domains. Despite our efforts, there are omissions of topics, due partly to our biases and partly to the scope of the presentation. Some omitted topics are mentioned in the notes and comments.

A Bird's-Eye View of the Subject

The subject of variational inequalities has its origin in the calculus of variations associated with the minimization of infinite-dimensional functionals. The systematic study of the subject began in the early 1960s with the seminal work of the Italian mathematician Guido Stampacchia and his collaborators, who used the variational inequality as an analytic tool for studying free boundary problems defined by nonlinear partial differential operators arising from unilateral problems in elasticity and plasticity theory and in mechanics. Some of the earliest papers on variational inequalities are [333, 512, 561, 804, 805]. In particular, the first theorem of existence and uniqueness of the solution of VIs was proved in [804]. The books by Baiocchi and Capelo [35] and Kinderlehrer and Stampacchia [410] provide a thorough introduction to the application of variational inequalities in infinite-dimensional function spaces; see also [39]. The lecture notes [362] treat complementarity problems in abstract spaces. The book by Glowinski, Lions, and Trémolière [291] is among the earliest references to give a detailed numerical treatment of such VIs. There is a huge literature on the subject of infinite-dimensional variational inequalities and related problems. Since a VI in an abstract space is in many respects quite distinct from the finite-dimensional VI and since the former problem is not the main concern of this book, in this section we focus our introduction on the latter problem only.

The development of the finite-dimensional variational inequality and nonlinear complementarity problem also began in the early 1960s but followed a different path. Indeed, the NCP was first identified in the 1964 Ph.D. thesis of Richard W. Cottle [135], who studied under the supervision of the eminent George B. Dantzig, "father of linear programming." Thus, unlike its infinite-dimensional counterpart, which was conceived in the area of partial differential systems, the finite-dimensional VI/CP was born in the domain of mathematical programming. This origin has had a heavy influence on the subsequent evolution of the field; a brief account of the history prior to 1990 can be found in the introduction of the survey paper [332]; see also Section 1.2 in [331]. In what follows, we give a more detailed account of the evolutionary process of the field, covering four decades of major events and notable highlights.

In the 1960s, largely as a result of the celebrated almost complementary pivoting algorithm of Lemke and Howson for solving a bimatrix game formulated as a linear complementarity problem (LCP) [491] and the subsequent extension by Lemke to a general LCP [490], much focus was devoted to the study of the latter problem. Cottle, Pang, and Stone presented a comprehensive treatment of the LCP in the 1992 monograph [142]. Among other things, this monograph contains an extensive bibliography of the LCP up to 1990 and also detailed notes, comments, and historical accounts about this fundamental problem. Today, research on the LCP remains active and new applications continue to be uncovered. Since much of the pre-1990 details about the LCP are already documented in the cited monograph, we rely on the latter for most of the background results for the LCP and will touch on the more contemporary developments of this problem where appropriate.

In 1967, Scarf [759] developed the first constructive iterative method for approximating a fixed point of a continuous mapping. Scarf's seminal work led to the development of the entire family of fixed-point methods and of the piecewise homotopy approach to the computation of economic equilibria. The field of *equilibrium programming* was thus born. In essence, the term "equilibrium programming" broadly refers to the modeling, analysis, and computation of equilibria of various kinds via the methodology of mathematical programming. Since the infant days of linear programming, it was clear that complementarity problems have much to do with equilibrium programs. For instance, the primal-dual relation of a linear program provides clear evidence of the interplay between complementarity and equilibrium. Indeed, all the equilibrium problems that were amenable to solution by the fixed-point methods, including the renowned Walrasian problem in general equilibrium theory and variations of this problem [760, 842, 866], were in fact VIs/CPs.

The early research in equilibrium programming was to a large extent a consequence of the landmark discoveries of Lemke and Scarf. In particular, the subject of fixed-point computations via piecewise homotopies dominated much of the research agenda of equilibrium programming in the 1970s. A major theoretical advantage of the family of fixed-point homotopy methods is their global convergence. Attracted by this advantage and the novelty of the methods, many well-known researchers including Eaves, Garcia, Gould, Kojima, Megiddo, Saigal, Todd, and Zangwill all made fundamental contributions to the subject. The flurry of research activities in this area continued for more than a decade, until the occurrence of several significant events that provided clear evidence of the practical inadequacy of this family of methods for solving realistic equilibrium problems. These events, to be mentioned momentarily, marked a turning point whereby the fixed-point/homotopy approach to the computation of equilibria gave way to an alternative set of methods that constitute what one may call a contemporary variational inequality approach to equilibrium programming. For completeness, we mention several prominent publications that contain important works on the subject of fixed-point computation via the homotopy approach and its applications [10, 11, 34, 203, 205, 206, 211, 251, 252, 285, 403, 440, 729, 760, 841, 879]. For a recent paper on this approach, see [883].

In the same period and in contrast to the aforementioned algorithmic research, Karamardian, in a series of papers [398, 399, 400, 401, 402], developed an extensive existence theory for the NCP and its cone generalization. In particular, the basic connection between the CP and the VI, Proposition **1.1.3**, appeared in [400]. The 1970s were a period when many fundamental articles on the VI/CP first appeared. These include the paper by Eaves [202] where the natural map $\mathbf{F}_K^{\text{nat}}$ was used to prove a basic theorem of complementarity, important studies by Moré [623, 624] and Moré and Rheinboldt [625], which studied several distinguished classes of nonlinear functions and their roles in complementarity problems, and the individual and joint work of Kojima and Megiddo [441, 599, 600, 601], which investigated the existence and uniqueness of solutions to the NCP.

Although the initial developments of infinite-dimensional variational inequalities and finite-dimensional complementarity problems had followed different paths, there were attempts to bring the two fields more closely together, with the International School of Mathematics held in Summer 1978 in Erice, Italy, being the most prominent one. The proceedings of this conference were published in [141]. The paper [138] is among the earliest that describes some physical applications of VIs in infinite dimensions solvable by LCP methods.

One could argue that the final years of the 1970s marked the beginning of the contemporary chapter on the finite-dimensional VI/CP. During that time, the U.S. Department of Energy was employing a market equilibrium system known as the Project Independent Evaluation System (PIES) [350, 351] for energy policy studies. This system is a large-scale variational inequality that was solved on a routine basis by a special iterative algorithm known as the PIES algorithm, yielding remarkably good computational experience. For a detailed account of the PIES model, see the monograph by Ahn [5], who showed that the PIES algorithm was a generalization of the classical Jacobi iterative method for solving system of nonlinear equations [652]. For the convergence analysis of the PIES algorithm, see Ahn and Hogan [6]; for a recent update of the PIES model, which has become the National Energy Modeling System (NEMS), see [278].

The original PIES model provided a real-life economic model for which the fixed-point methods mentioned earlier were proved to be ineffective. This experience along with several related events inspired a new wave of research into iterative methods for solving VIs/CPs arising from various applied equilibrium contexts. One of these events is an important algorithmic advance, namely, the introduction of Newton's method for solving generalized equations (see below).

At about the same time as the PIES model appeared, Smith [793] and Dafermos [151] formulated the traffic equilibrium problem as a variational inequality. Parallel to the VI formulation, Aashitiani and Magnanti [1] introduced a complementarity formulation for Wardrop's user equilibrium principle [868] and established existence and uniqueness results of traffic equilibria using fixed-point theorems; see also [20, 253]. Computationally, the PIES algorithm had served as a model approach for the design of iterative methods for solving the traffic equilibrium problem [2, 254, 259]. More broadly, the variational inequality approach has had a significant impact on the contemporary point of view of this problem and the closely related spatial price equilibrium problem.

In two important papers [594, 595], Mathiesen reported computational results on the application of a sequential linear complementarity (SLCP) approach to the solution of economic equilibrium problems. These results firmly established the potential of this approach and generated substantial interest among many computational economists, including Manne and his (then Ph.D.) students, most notably, Preckel, Rutherford, and Stone. The volume edited by Manne [581] contains the papers [697, 814], which give further evidence of the computational efficiency of the SLCP approach for solving economic equilibrium problems; see also [596].

The SLCP method, as it was called in the aforementioned papers, turned out to be Newton's method developed and studied several years earlier by Josephy [389, 390, 391]; see also the later papers by Eaves [209, 210]. While the results obtained by the computational economists clearly established the practical effectiveness of Newton's method through sheer numerical experience, Josephy's work provided a sound theoretical foundation for the fast convergence of the method. In turn, Josephy's results were based on the seminal research of Robinson, who in several landmark papers [728, 730, 732, 734] introduced the *generalized equations* as a unifying mathematical framework for optimization problems, complementarity problems, variational inequalities, and related problems. As we explain below, in addition to providing the foundation for the convergence theory of Newton's method, Robinson's work greatly influenced the modern development of sensitivity analysis of mathematical programs.

While Josephy's contributions marked a breakthrough in algorithmic advances of the field, they left many questions unanswered. From a computational perspective, Rutherford [754] recognized early on the lack of robustness in Newton's method applied to some of the most challenging economic equilibrium problems. Although ad hoc remedies and specialized treatments had lessened the numerical difficulty in solving these problems, the heuristic aids employed were far from satisfactory in resolving the practical deficiency of the method, which was caused by the lack of a suitable stabilizing strategy for global convergence. Motivated by the need for a computationally robust Newton method with guaranteed global convergence, Pang [663] developed the B-differentiable Newton method with a line search and established that the method is globally convergent and locally superlinearly convergent. While this is arguably the first work on global Newton methods for solving nonsmooth equations, Pang's method suffers from a theoretical drawback in that its convergence requires a Fréchet differentiability assumption at a limit point of the produced sequence.

Newton's method for solving nondifferentiable equations had been investigated before Pang's work. Kojima and Shindo [454] discussed such a method for PC^1 functions. Kummer [466] studied this method for general nondifferentiable functions. Both papers dealt with the local convergence but did not address the globalization of the method. Generalizing the class of semismooth functions of one variable defined by Mifflin [607], Qi and Sun [701] introduced the class of vector semismooth functions and established the local convergence of Newton's method for this class of functions. The latter result of Qi and Sun is actually a special case of the general theory of Kummer. Since its introduction, the class of vector semismooth functions has played a central role throughout the subsequent algorithmic developments of the field. Although focused mainly on the smooth case, the two recent papers [282, 878] present an enlightening summary of the historical developments of the convergence theory of Newton's method.

As an alternative to Pang's line search globalization strategy, Ralph [710] presented a path search algorithm that was implemented by Dirkse and Ferris in their highly successful PATH solver [187], which was awarded the 1997 Beale-Orchard-Hays Prize for excellence in computational mathematical programming; the accompanying paper [186] contains an extensive collection of MiCP test problems. In an important paper that dealt with an optimization problem [247], Fischer proposed the use of what is now called the Fischer-Burmeister function to reformulate the Karush-Kuhn-Tucker conditions arising from an inequality constrained optimization problem as a system of nonsmooth equations. Collectively, these works paved the way for an outburst of activities that started with De Luca, Facchinei, and Kanzow [162]. The latter paper discussed the application of a globally convergent semismooth Newton method to the Fischer-Burmeister reformulation of the nonlinear complementarity problem; the algorithm described therein provided a model approach for many algorithms that followed. The semismooth Newton approach led to algorithms that are conceptually and practically simpler than the B-differentiable Newton method and the path Newton method, and have, at the same time, better convergence properties.

The attractive theoretical properties of the semismooth methods and their good performance in practice spurred much research to investigate further this class of methods and inspired much of the subsequent studies. In the second half of the 1990s, a large number of papers was devoted to the improvement, extension, and numerical testing of semismooth algorithms, bringing these algorithms to a high level of sophistication. Among other things, these developments made it clear that the B-differentiable Newton method is intimately related to the semismooth Newton method applied to the min reformulation of the complementarity problem, thus confirming the breadth of the new approach.

The above overview gives a general perspective on the evolution of the VI/CP and documents several major events that have propelled this subject to its modern status as a fruitful and exciting discipline within mathematical programming. There are many other interesting developments, such as sensitivity and stability analysis, piecewise smooth functions, error bounds, interior point methods, smoothing methods, methods of the projection family, and regularization, as well as the connections with new applications and other mathematical disciplines, all of which add to the richness and vitality of the field and form the main topics in our work. The notes and comments of these developments are contained at the end of each chapter.

A Synopsis of the Book

Divided into two volumes, the book contains twelve main chapters, followed by an extensive bibliography, a summary of main results and key algorithms, and a subject index. The first volume consists of the first six chapters, which present the basic theory of VIs and CPs. The second volume consists of the remaining six chapters, which present algorithms of various kinds for solving VIs and CPs. Besides the main text, each chapter contains (a) an extensive set of exercises, many of which are drawn from published papers that supplement the materials in the text, and (b) a set of notes and comments that document historical accounts, give the sources for the results in the main text, and provide discussions and references on related topics and extensions. The bibliography contains more than 1,300 publications in the literature up to June 2002. This bibliography serves two purposes: one purpose is to give the source of the results in the chapters, wherever applicable; the other purpose is to give a documentation of papers written on the VI/CP and related topics.

Due to its comprehensiveness, each chapter of the book is by itself quite lengthy. Among the first six sections in Chapter **1**, Sections **1.1**, **1.2**, **1.3**, and **1.5** make up the basic introduction to the VI/CP. The source problems in Section **1.4** are of very diverse nature; they fall into several general categories: mathematical programming, economics, engineering, and finance. Depending on an individual's background, a reader can safely skip those subsections that are outside his/her interests; for instance, an economist can omit the subsection on frictional contact problems, a contact mechanician can omit the subsection on Nash-Cournot production models. Section **1.6** mainly gives the definition of several extended problems; except for (1.6.1), which is re-introduced and employed in Chapter **11**, this section can be omitted at first reading.

Chapters **2** and **3** contain the basic theory of existence and multiplicity of solutions. Several sections contain review materials of well-known topics; these are included for the benefit of those readers who are unfamiliar with the background for the theory. Section **2.1** contains the review of degree theory, which is a basic mathematical tool that we employ throughout the book; due to its powerful implications, we recommend this to a reader who is interested in the theoretical part of the subject. Sections **2.2**, **2.3** (except Subsection **2.3.2**), **2.4**, and **2.5** (except Subsection **2.5.3**) contain fundamental results. While Sections **2.6** and **2.8** can be skipped at first reading, Section **2.7** contains very specialized results for the discrete frictional contact problem and is included herein only to illustrate the application of the theory developed in the chapter to an important class of mechanical problems.

Section **3.1** in Chapter **3** introduces the class of B-differentiable functions that plays a fundamental role throughout the book. With the exception of the nonstandard SBCQ, Section **3.2** is a review of various well-known CQs in NLP. Except for the last two subsections in Section **3.3** and Subsection **3.5.1**, which may be omitted at first reading, the remainder of this chapter contains important properties of solutions to the VI/CP.

Chapter **4** serves two purposes: One, it is a technical precursor to the next chapter; and two, it introduces the important classes of PA functions (Section **4.2**) and PC^1 functions (Section **4.6**). Readers who are not interested in the sensitivity and stability theory of the VI/CP can skip most of this and the next chapter. Nevertheless, in order to appreciate the class of semismooth functions, which lies at the heart of the contemporary algorithms for solving VIs/CPs, and the regularity conditions, which are key to the convergence of these algorithms, the reader is advised to become familiar with certain developments in this chapter, such as the basic notion of coherent orientation of PA maps (Definition **4.2.3**) and its matrix-theoretic characterizations for the special maps $\mathbf{M}_K^{\mathrm{nor}}$ and $\mathbf{M}_K^{\mathrm{nat}}$ (Proposition **4.2.7**) as well as the fundamental role of this notion in the globally unique solvability of AVIs (Theorem **4.3.2**). The inverse function Theorem **4.6.5** for PC^1 functions is of fundamental importance in nonsmooth analysis. Subsections **4.1.1** and **4.3.1** are interesting in their own right; but they are not needed in the remainder of the book.

Chapter **5** focuses on the single topic of sensitivity and stability of the VI/CP. While stability is the cornerstone to the fast convergence of Newton's method, readers who are not interested in this specialized topic or in the advanced convergence theory of the mentioned method can skip this entire chapter. Notwithstanding this suggestion, Section **5.3** is of classical importance and contains the most basic results concerning the local analysis of an isolated solution.

Chapter **6** contains another significant yet specialized topic that can be omitted at first reading. From a computational point of view, an important goal of this chapter is to establish a sound basis for understanding the connection between the exact solutions to a given problem and the computed solutions of iterative methods under prescribed termination criteria used in practical implementation. As evidenced throughout the chapter and also in Section **12.6**, the theory of error bounds has far-reaching consequences that extend beyond this goal. For instance, since the publication of the paper [222], which is the subject of discussion in Section **6.7**, there has been an increasing use of error bounds in designing algorithms that can identify active constraints accurately, resulting in enhanced theoretical properties of algorithms and holding promise for superior computational efficiency.

Of independent interest, Chapters **7** and **8** contain the preparatory materials for the two subsequent chapters. While Sections **7.1** and **7.4** are both concerned with the fundamentals of nonsmooth functions, the former pertains to general properties of nonsmooth functions, whereas the latter focuses on the semismooth functions. As far as specific algorithms go, Algorithms **7.3.1** and **7.5.1** in Sections **7.3** and **7.5**, respectively, are the most basic and strongly recommended for anyone interested in the subsequent developments. The convergence of the former algorithm depends on the (strong) stability theory in Chapter **5**, whereas that of the latter is rather simple, provided that one has a good understanding of semismoothness. In contrast to the previous two algorithms, Algorithm **7.2.17** is closest to a straightforward application of the classical Newton method for smooth systems of equations to the NCP.

The path search Newton method **8.1.9** is the earliest algorithm to be coded in the highly successful PATH computer software [187]. Readers who are already familiar with the line search and/or trust region methods in standard nonlinear programming may wish to peruse Subsection **8.3.3** and skip the rest of Chapter **8** in order to proceed directly to the next chapter. When specialized to C^1 optimization problems, as is the focus in Chapters **9** and **10**, much of the material in Sections **8.3** and **8.4** is classical; these two sections basically offer a systematic treatment of known techniques and results and present them in a way that accommodates nonsmooth objective functions.

The last four chapters are the core of the algorithmic part of this book. While Chapter **9** focuses on the NCP, Chapter **10** is devoted to the VI. The first section of the former chapter presents a detailed exposition of algorithms based on the FB merit function and their convergence theory. The most basic algorithm, **9.1.10**, is described in Subsection **9.1.1** and is accompanied by a comprehensive analysis. Algorithm **9.2.3**, which combines the min function and the FB merit function in a line search method, is representative of a mixture of algorithms in one overall scheme. Example **9.3.3** contains several C-functions that can be used in place of the FB C-function. The box-constrained VI in Subsection **9.4.3** unifies the generalized problems in Section **9.4**.

The development in Section **10.1** is very similar to that in Section **9.1.1**; the only difference is that the analysis of the first section in Chapter **10** is tailored to the KKT system of a finitely representable VI. The other major development in this chapter is the D-gap function in Section **10.3**, which is preceded by the preparatory discussion of the regularized gap function in Subsection **10.2.1**. The implicit Lagrangian function presented in Subsection **10.3.1** is an important merit function for the NCP.

Chapter **11** presents interior and smoothing methods for solving CPs of different kinds, including KKT systems. Developed in the abstract setting of constrained equations, the basic interior point method for the implicit MiCP, Algorithm **11.5.1**, is presented in Section **11.5**. An extensive theoretical study of the latter problem is the subject of the previous Section **11.4**; in which the important mixed P_0 property is introduced (see Definition **11.4.1**). A Newton smoothing method is outlined in Subsection **11.8.1**; this method is applicable to smoothed reformulations of CPs using the smoothing functions discussed in Subsection **11.8.2**, particularly those in Example **11.8.11**.

The twelveth and last chapter discusses various specialized methods that are applicable principally to (pseudo) monotone VIs and NCPs of the P_0 type. The first four sections of the chapter contain the basic methods and their convergence theories. The theory of maximal monotone operators in Subsection **12.3.1** plays a central role in the proximal point method that is the subject of Subsection **12.3.2**. Bregman-based methods in Subsection **12.7.2** are well researched in the literature, whereas the interior/barrier methods in Subsection **12.7.4** are recent entrants to the field.

Acknowledgments

Writing a book on this subject has been the goal of the second author since he and Harker published their survey paper [332] in 1990. This goal was not accomplished and ended with Harker giving a lecture series at the Université Catholique de Louvain in 1992 that was followed by the lecture notes [331]. The second author gratefully acknowledges Harker for the fruitful collaboration and for his keen interest during the formative stage of this book project.

The first author was introduced to optimization by Gianni Di Pillo and Luigi Grippo, who did much to shape his understanding of the discipline and to inspire his interest in research. The second author has been very lucky to have several pioneers of the field as his mentors during his early career. They are Richard Cottle, Olvi Mangasarian, and Stephen Robinson. To all these individuals we owe our deepest gratitude. Both authors have benefitted from the fruitful collaboration with their doctoral students and many colleagues on various parts of the book. We thank them sincerely. Michael Ferris and Stefan Scholtes have provided useful comments on a preliminary version of the book that help to shape its final form. We wish to thank our Series Editor, Achi Dosanjh, the Production Editor, Louise Farkas, and the staff members at Springer-New York, for their skillful editorial assistance.

Facchinei's research has been supported by grants from the Italian Research Ministry, the National Research Council, the European Commission, and the NATO Science Committee. The U.S. National Science Foundation has provided continuous research grants through several institutions to support Pang's work for the last twenty-five years. The joint research with Monteiro was supported by the Office of Naval Research as well. Pang's students have also benefited from the financial support of these two funding agencies, to whom we are very grateful.

Finally, the text of this monograph was typeset by the authors using LaTeX, a document preparation system based on Knuth's TeX program. We have used the document style files of the book [142] that were prepared by Richard Cottle and Richard Stone, and based on the LaTeX book style.

Rome, Italy — Francisco Facchinei
Baltimore, Maryland, U.S.A. — Jong-Shi Pang
February 24, 2003