A path-following framework on fiber bundle for variational inequalities

Hongbo Sun
shb20@tsinghua.org.cn

Abstract

Developing general-purpose algorithms with global asymptotic convergence guarantees for variational inequalities (VIs) remains a longstanding challenge. This paper proposes a path-following framework for finite-dimensional VIs with arbitrary continuous mappings and polytope domains. The approach first transforms a general VI into a smooth VI on a simplex. Its key innovation is to formulate the resulting problem on a fiber bundle, called the fixed-point bundle, whose geometric structure enables the systematic integration of initialization, path-following, and singularity avoidance. Without imposing monotonicity or similar assumptions, the framework guarantees global linear convergence for nonsingular solutions and global linear convergence up to any prescribed precision for singular solutions. In numerical experiments, the algorithm attained a precision of 10^{-5} on all 6400 randomly generated instances, with dimensions up to 1024. Code and data are available at `https://github.com/shb20tsinghua/FiberBundle_VI`.

Keywords: variational inequality, interior point method, fiber bundle, path-following
MSCcodes: 49J40, 90C33, 90C51, 49M29

Contents

1 Introduction **2**

2 Preprocessing: smooth VI over simplex **3**
2.1 Dirichlet smoothing 4
2.2 Numerical evaluation via Monte Carlo sampling 5
2.3 Approximation error to the original VI 6

3 Three equivalent characterizations regarding the paths **7**
3.1 Linear programming form and path equation 7
3.2 Brouwer fixed-point form and path existence 8
3.3 Mixed complementarity problem form and central path 9

4 Fiber bundle as the solution space of FPKKT **10**
4.1 The fixed-point bundle 10
4.2 Differential properties 12

5 Paths on the fixed-point bundle **14**
5.1 Analytic curves 14
5.2 The path 15
5.3 A compact nonsingular neighborhood for path-following 17

6 Path-following on the fixed-point bundle **19**
6.1 The predictor-corrector framework 19
6.2 Corrector derived from perturbed KKT conditions 20
6.3 Corrector derived from barrier problem 21
6.4 The global linear convergence proof 22

7 Numerical experiments **25**

8 Conclusion **26**

1 Introduction

Variational inequalities, introduced by Fichera [1] and Stampacchia [2] in the 1960s, provide a unifying framework for optimization [3], game theory [4], economics [5], traffic assignment [6], contact mechanics [7], fluid flow [8], machine learning [9], and many other fields [10]. A finite-dimensional variational inequality problem $\mathrm{VI}(H, K)$ is specified by a nonempty compact convex set $K \subseteq \mathbb{R}^n$ and a continuous map $H : K \to \mathbb{R}^n$, and the goal is to find $x^\star \in K$ such that

$$\left\langle H(x^\star), x' - x^\star \right\rangle \geq 0 \quad \text{for all } x' \in K.$$

The variational inequality $\mathrm{VI}(H, K)$ is naturally described by the gap function

$$\mathrm{gap}(x) = \sup_{x' \in K} \left\langle H(x), x - x' \right\rangle,$$

where $\mathrm{gap}(x) \geq 0$ for all $x \in K$. Solutions are precisely the points $x^\star \in K$ with $\mathrm{gap}(x^\star) = 0$. Existence of a solution follows from the Brouwer fixed-point theorem [11]. However, unconditional global convergence of numerical methods remains a challenge. This paper addresses general $\mathrm{VI}(H, K)$ with continuous H and polytope K, where polytopes can approximate general compact convex sets by convex-body approximation.

In terms of iterative methods with asymptotic convergence, existing methods fall into several categories. Projection methods [12, 13] and proximal-point methods [14, 15] reformulate the VI as a fixed-point problem and iterate projections onto K. Merit-function approaches [16, 17] convert the VI into an optimization problem via merit functions such as gap functions and then apply gradient or Newton methods. Interior-point methods [18, 19] treat the VI as a complementarity problem and follow a central path parameterized by a barrier parameter. Homotopy methods [20] deform a simple map into the target map. Operator splitting [21, 22, 23] handles additive decompositions. Modern variants add adaptive step sizes [24, 25], inertial terms [26, 27], or stochastic estimates [28, 28, 29] for large-scale problems.

Most existing algorithms require monotonicity, pseudo-monotonicity, or similar structural assumptions. Algorithms that handle non-monotone problems usually need additional conditions or provide no guarantees. In particular, homotopy methods guarantee the existence of a path from almost every starting point to a solution, but path existence does not yield a globally convergent path-following iteration, especially when the path has singular points. In short, in terms of iterative methods with asymptotic convergence, no existing method guarantees global convergence for every continuous H on a polytope K without additional assumptions.

Among these approaches, interior-point methods are particularly relevant because of their path-following philosophy. They trace a central path parameterized by a barrier parameter μ, each of whose points satisfies the perturbed KKT conditions [30]. Our approach follows the same path-following spirit and is closely related to interior-point methods.

The proposed approach first preprocesses the general VI(H, K) into a smooth VI(F, Δ) on a simplex Δ. Every polytope is a linear image of a simplex, and every continuous map on a simplex is smoothed into a real-analytic function by averaging against a Dirichlet distribution. All subsequent development applies to the resulting smooth problem VI(F, Δ).

We derive three equivalent characterizations of the paths that solve VI(F, Δ), based on linear programming, Brouwer fixed points, and mixed complementarity problems. They yield the fixed-point KKT system (FPKKT), a simultaneous system of perturbed KKT conditions and a fixed-point equation. The paths satisfying FPKKT lead to solutions of VI(F, Δ) and are special central paths, which means that our approach is equivalently a specialized interior-point method that follows these special central paths.

Rather than solve FPKKT directly, we study the geometry of its solution set, which naturally forms a fiber bundle over the simplex Δ with one-dimensional fibers in μ. Two sections of this bundle recover the gap function of VI(F, Δ) and govern the differential properties of the bundle. Analytic curves on the bundle connect a chosen starting point to solutions, and a parity argument implies that the number of nonsingular solutions is odd.

The fixed-point bundle is the central innovation. Its key insight is that the geometry separates motion of $\sigma \in \Delta$ along the base from motion of μ along the fibers. The starting point of the path lies on the fiber over a designated initial point $\sigma_{\text{init}} \in \Delta$. Singular points on each fiber correspond to real eigenvalues of the matrix function $J(\sigma)$, and the bundle structure provides fiber jumps as a simple mechanism to bypass singular points while preserving the path structure.

The fixed-point bundle unifies the selection of a starting point, path-following, and singularity avoidance in a systematic way, yielding a path $\tilde{\Gamma}$ on the bundle. We then construct a compact nonsingular neighborhood $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$ of this path on which the path-following algorithm is well-behaved; here ζ controls the prescribed precision at which path-following terminates before reaching the exact solution if the solution itself is singular.

The path-following algorithm follows the predictor-corrector framework. Using bounds established on $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$, we prove uniform quadratic convergence of the corrector and a uniform bound on the predictor step length. The algorithm then drives $\text{gap}(\sigma)$ to global linear convergence up to a prescribed precision ζ. For nonsingular solutions $\zeta = 0$, and the algorithm is globally linearly convergent; for singular solutions it retains global linear convergence up to $\zeta > 0$.

We provide two different but closely related regularized Newton correctors for practical use. Numerical experiments on 6400 randomly generated instances of dimension up to 1024 show that the algorithm reduces the gap function below 10^{-5} in every case.

The remainder of the paper is organized as follows. Section 2 details the preprocessing. Section 3 presents the three equivalent characterizations. Section 4 constructs the fixed-point bundle and studies its differential properties. Section 5 formalizes the paths on the fixed-point bundle. Section 6 develops the predictor-corrector scheme and proves convergence. Section 7 presents experimental results. Concluding remarks appear in Section 8.

For vectors, $a \circ b$ denotes componentwise multiplication and a/b componentwise division, and the inequality $a \geq b$ is also interpreted componentwise. The limit $\mu \to \mathbf{0}^+$ means $\mu \to \mathbf{0}$ with $\mu > 0$. For a matrix A, componentwise multiplication means $A \circ b = A\text{diag}(b)$ and $b \circ A = \text{diag}(b)A$. Unless otherwise specified, $\|a\|$ and $\|A\|$ denote the L^2 norm of a vector a or a matrix A.

2 Preprocessing: smooth VI over simplex

The subsequent theory is developed for a variational inequality VI(F, Δ) with a smooth mapping and a simplex domain. To place the framework in a sufficiently general setting, we first reduce

a general variational inequality $\mathrm{VI}(H, K)$ with a continuous mapping on a polytope domain to $\mathrm{VI}(F, \Delta)$.

Let $V = (v_1, \ldots, v_k)$ be the vertices of the polytope domain K, and represent any $x \in K$ as $x = V\sigma$ with $\sigma \in \Delta$. The variational inequality $\mathrm{VI}(H, K)$ is then transformed into finding $\sigma^\star \in \Delta$ such that, for all $\sigma \in \Delta$,

$$\langle H(x^\star), x - x^\star \rangle = H(V\sigma^\star)^\top (V\sigma - V\sigma^\star) = \left\langle V^\top H(V\sigma^\star), \sigma - \sigma^\star \right\rangle \geq 0.$$

Thus $\mathrm{VI}(H, K)$ is transformed into a VI with simplex domain Δ and continuous mapping $F(\sigma) = V^\top H(V\sigma)$.

2.1 Dirichlet smoothing

Next we smooth the continuous mapping $F(\sigma) = V^\top H(V\sigma)$ on the simplex. General continuous functions can be approximated by smooth functions using kernel averaging, and since our domain is a simplex, we use the Dirichlet distribution as the kernel. The Dirichlet distribution $\mathrm{Dir}(\alpha)$ on the simplex Δ is given by

$$\mathrm{Dir}(y; \alpha) = \frac{1}{B(\alpha)} \prod_{i=1}^{n} y_i^{\alpha_i - 1}, \quad \alpha > 0, \tag{2.1}$$

where $y \in \Delta$ is the variable, $\alpha > 0$ is a vector of concentration parameters, and $B(\alpha)$ is the Beta function. For a continuous function $F : \Delta \to \mathbb{R}^n$, the smoothed function $F_\epsilon : \Delta \to \mathbb{R}^n$ is given componentwise by

$$F_{\epsilon,i}(\sigma) = \mathbb{E}_{Y \sim \mathrm{Dir}(\alpha)} \left[F_i(Y) \right] = \int_\Delta F_i(y) \mathrm{Dir}(y; \sigma/\epsilon + c)\, dy. \tag{2.2}$$

For the i-th component, $F_{\epsilon,i}(\sigma)$ is the expectation of $F_i(Y)$ under the Dirichlet distribution $\mathrm{Dir}(\alpha)$ with concentration parameter $\alpha = \sigma/\epsilon + c$. Here $\epsilon > 0$ is a scalar that measures the approximation error, and $c > 0$ is a constant vector ensuring that $\sigma/\epsilon + c > 0$ for all $\sigma \in \Delta$.

The smoothing in equation (2.2) is a mollification-type approach [31]. The Dirichlet distribution $\mathrm{Dir}(\sigma/\epsilon + c)$ acts as a mollifier: it is nonnegative on a compact domain, integrates to one, and concentrates its mass at σ as $\epsilon \to 0$. As in mollification, the smoothed function converges uniformly to the original function, satisfying

$$\lim_{\epsilon \to 0} \sup_{\sigma \in \Delta} \| F_\epsilon(\sigma) - F(\sigma) \| = 0.$$

Different parts of this paper require different degrees of smoothness. The computations in the resulting algorithm use the first derivative, the convergence of the algorithm is guaranteed by second-order continuous differentiability, and the compactness of the path is derived from real-analyticity. We therefore prove the strongest smoothness result, namely that $F_\epsilon(\sigma)$ is real-analytic.

Proposition 1. *For every $\epsilon > 0$, $F_\epsilon(\sigma)$ is real-analytic on Δ.*

Proof. A standard theorem on holomorphic parametric integrals in complex analysis can be used to prove this [32]. Specifically, for a complex function $f(z, y)$ with $z \in Z$ and $y \in Y$, if $f(z, y)$ is holomorphic in z and uniformly integrable in y, then the parametric integral $h(z) = \int_Y f(z, y)\, dy$ is holomorphic on Z. Uniform integrability means that there exists an integrable function $g(y)$ such that $|f(z, y)| < g(y)$ for every $z \in Z$.

In our case, we study the complex parametric function $F_i(y)\mathrm{Dir}(y; z/\epsilon + c)$ with $y \in \Delta$ and $z \in U_\sigma$. Here $U_\sigma = \{z \in \mathbb{C}^n | \|z - \sigma\| \le d\}$ is a complex neighborhood of $\sigma \in \Delta$, and we require that $\mathrm{Re}(z/\epsilon + c) \ge m > 0$ for every $z \in U_\sigma$. Such d and m exist for every $\sigma \in \Delta$ because $c > 0$. We then show that $F_i(y)\mathrm{Dir}(y; z/\epsilon + c)$ is holomorphic in z and uniformly integrable in y.

For holomorphy, the Beta function $B(\alpha)$ is known to be holomorphic and nonzero on $\{\alpha \in \mathbb{C}^n | \mathrm{Re}(\alpha) > 0\}$. Hence $1/B(z/\epsilon + c)$ is holomorphic on U_σ since $\mathrm{Re}(z/\epsilon + c) \ge m > 0$. The term $\prod_{i=1}^n y_i^{z_i/\epsilon + c_i - 1}$ is holomorphic in z for every $y > 0$ because it is an exponential function. Thus $F_i(y)\mathrm{Dir}(y; z/\epsilon + c)$ is holomorphic in z on U_σ for every $y > 0$. The set where $\min_i y_i = 0$ has measure zero and does not affect the integral.

For uniform integrability, $F_i(y)$ is continuous on the compact simplex Δ, so it satisfies a bound $|F_i(y)| < M_F$, and $1/B(z/\epsilon + c)$ is continuous on the compact complex neighborhood U_σ, so it satisfies a bound $|1/B(z/\epsilon + c)| < M_B$. Then

$$|F_i(y)\mathrm{Dir}(y; z/\epsilon + c)| \le M_F M_B \left| \prod_{i=1}^n y_i^{z_i/\epsilon + c_i - 1} \right| \le M_F M_B \prod_{i=1}^n y_i^{\mathrm{Re}(z_i/\epsilon + c_i) - 1} \le M_F M_B \prod_{i=1}^n y_i^{m-1}.$$

The function $\prod_{i=1}^n y_i^{m-1}$ is integrable on Δ as the Dirichlet multivariate Beta integral. Hence $F_i(y)\mathrm{Dir}(y; z/\epsilon + c)$ is uniformly integrable in y on Δ.

Therefore, by the standard holomorphic parametric integral theorem, $F_{\epsilon,i}(z)$ is holomorphic on the complex neighborhood U_σ of every $\sigma \in \Delta$. It follows that $F_{\epsilon,i}(\sigma)$ is real-analytic in $\sigma \in \Delta$, and the vector function $F_\epsilon(\sigma)$ is real-analytic in $\sigma \in \Delta$ because each of its components is real-analytic. □

Besides $F_\epsilon(\sigma)$, the algorithm also relies on the derivative of $F_\epsilon(\sigma)$. For any distribution $\mathrm{D}(\alpha)$ and function $f(y)$, under some regularity conditions, there is a standard result called the score function identity [33]

$$\frac{\partial}{\partial \alpha} \mathbb{E}_{Y \sim \mathrm{D}(\alpha)} [f(Y)] = \mathrm{Cov}_{Y \sim \mathrm{D}(\alpha)} [f(Y), S(Y; \alpha)],$$

where $S(Y; \alpha) = \partial \log \mathrm{D}(Y; \alpha)/\partial \alpha$ is the score function of the distribution $\mathrm{D}(\alpha)$ and $\mathrm{Cov}_{Y \sim \mathrm{D}(\alpha)}$ denotes covariance. For the Dirichlet distribution, $S_j(Y; \alpha) = \log Y_j - \psi(\alpha_j) + \psi(\mathbf{1}^\top \alpha)$, where ψ is the digamma function. The covariance between a random variable and a constant is zero, so the derivative $F'_\epsilon(\sigma)$ is derived as

$$\frac{\partial F_{\epsilon,i}(\sigma)}{\partial \sigma_j} = \frac{\partial \alpha_j}{\partial \sigma_j} \frac{\partial}{\partial \alpha_j} \mathbb{E}_{Y \sim \mathrm{Dir}(\alpha)} [F_i(Y)] = \frac{1}{\epsilon} \mathrm{Cov}_{Y \sim \mathrm{Dir}(\alpha)} [F_i(Y), \log Y_j]. \tag{2.3}$$

As $\epsilon \to 0$, the derivative $F'_0(\sigma)$ is given by the following equation. Although the derivative does not converge to $F'(\sigma)$, the term $(I - \sigma \mathbf{1}^\top)$ can be absorbed by idempotence in the algorithm, which produces the same effect as $F'(\sigma)$.

$$F'_0(\sigma) = \frac{\partial F_i(\sigma/(\mathbf{1}^\top \sigma))}{\partial \sigma_j} = F'(\sigma) \left(I - \sigma \mathbf{1}^\top \right) \tag{2.4}$$

2.2 Numerical evaluation via Monte Carlo sampling

The function $F_\epsilon(\sigma)$ is given by the expectation of $F(Y)$ with Y distributed as $\mathrm{Dir}(\sigma/\epsilon + c)$. Thus, its value can be evaluated by Monte Carlo sampling. Specifically, we draw N independent samples

$Y_k \sim \mathrm{Dir}(\sigma/\epsilon + c)$ and then estimate the expectation $F_{\epsilon,i}(\sigma)$ by the mean $\hat{F}_{\epsilon,i}(\sigma)$, as equation (2.5) shows.

$$\hat{F}_{\epsilon,i}(\sigma) = \frac{1}{N}\sum_{k=1}^{N} F_i(Y_k) \tag{2.5}$$

The estimation is unbiased, in the sense that $\mathbb{E}[\hat{F}_{\epsilon,i}(\sigma)] = F_{\epsilon,i}(\sigma)$. The estimation error is given by the variance derived as follows, where $|F_i(y)| < M_F$ is the bound that follows from continuity on the compact simplex.

$$\mathrm{Var}\left[\hat{F}_{\epsilon,i}(\sigma)\right] = \frac{\mathrm{Var}[F_i(Y)]}{N} \le \frac{\mathbb{E}[F_i(Y)^2]}{N} \le \frac{M_F^2}{N} \tag{2.6}$$

The derivative $F'_\epsilon(\sigma)$ is also evaluated by Monte Carlo sampling with N independent samples $Y_k \sim \mathrm{Dir}(\sigma/\epsilon + c)$. Since $\mathrm{Cov}[F_i(Y), \log Y_j] = \mathbb{E}[F_i(Y)\log Y_j] - \mathbb{E}[F_i(Y)]\mathbb{E}[\log Y_j]$, the estimate $\hat{F}'_\epsilon(\sigma)$ of $F'_\epsilon(\sigma)$ is given as in equation (2.5).

$$[\hat{F}'_\epsilon(\sigma)]_{ij} = \frac{1}{\epsilon N}\sum_{k=1}^{N} F_i(Y_k)\left(\log Y_{k,j} - \mathbb{E}[\log Y_j]\right),\ \mathbb{E}[\log Y_j] = \psi(\alpha_j) - \psi(\mathbf{1}^\top\alpha) \tag{2.7}$$

The estimation is unbiased, in the sense that $\mathbb{E}[[\hat{F}'_\epsilon(\sigma)]_{ij}] = [F'_\epsilon(\sigma)]_{ij}$. The estimation error is given by the variance derived as follows, where ψ_1 is the trigamma function. In the derivation, apart from $|F_i(y)| < M_F$, we also use $\psi_1(\alpha_j) - \psi_1(\mathbf{1}^\top\alpha) < \psi_1(\alpha_j) \le \psi_1(c_j)$, since ψ_1 is positive and monotonically decreasing on $(0, +\infty)$.

$$\begin{aligned}\mathrm{Var}\left[[\hat{F}'_\epsilon(\sigma)]_{ij}\right] =& \frac{\mathrm{Var}[F_i(Y)(\log Y_j - \mathbb{E}[\log Y_j])]}{\epsilon^2 N} \le \frac{\mathbb{E}[F_i(Y)^2(\log Y_j - \mathbb{E}[\log Y_j])^2]}{\epsilon^2 N} \\ \le& \frac{M_F^2\mathrm{Var}[\log Y_j]}{\epsilon^2 N} = \frac{M_F^2}{\epsilon^2 N}\left(\psi_1(\alpha_j) - \psi_1(\mathbf{1}^\top\alpha)\right) < \frac{M_F^2}{\epsilon^2 N}\psi_1(c_j)\end{aligned} \tag{2.8}$$

Therefore, to achieve a Monte Carlo estimation error ϵ_{MC} and a Dirichlet smoothing error ϵ for both the function value $F_\epsilon(\sigma)$ and its derivative $F'_\epsilon(\sigma)$, the required number of samples N satisfies $N = O(\epsilon_{\mathrm{MC}}^{-2}\epsilon^{-2})$. In other words, compared with evaluating an already smooth mapping, the computational cost of evaluating the Dirichlet-smoothed mapping increases polynomially in the Dirichlet error $1/\epsilon$ and the Monte Carlo error $1/\epsilon_{\mathrm{MC}}$.

2.3 Approximation error to the original VI

To conclude the preprocessing, we derive the approximation error of $\mathrm{VI}(F_\epsilon, \Delta)$ relative to $\mathrm{VI}(H, K)$ in terms of the gap function approximation error

$$\mathrm{gap}_H(V\sigma) - \mathrm{gap}_{F_\epsilon}(\sigma) = \sup_{x'\in K}\left\langle H(V\sigma), V\sigma - x'\right\rangle - \sup_{\sigma'\in\Delta}\left\langle F_\epsilon(\sigma), \sigma - \sigma'\right\rangle.$$

Moreover, we derive the error assuming K to be a general compact convex set, so as to incorporate the error of approximating convex bodies by polytopes. Specifically, Minkowski's theorem on inner approximation [34] shows that for a finite-dimensional convex body K and any required precision ϵ_{set}, there exists a polytope $S \subseteq K$ approximating K to Hausdorff distance $d_H(K, S) = \max\{\sup_{x\in K} d(x, S), \sup_{s\in S} d(s, K)\} \le \epsilon_{\mathrm{set}}$. The Dirichlet smoothing error is denoted by ϵ_{map}.

Let $x'^\star = \arg\max_{x'\in K}\langle H(V\sigma), V\sigma - x'\rangle$. By the inner approximation, there exists $\sigma''^\star \in \Delta$ such that $\|V\sigma''^\star - x'^\star\| < \epsilon_{\mathrm{set}}$. Hence

$$\mathrm{gap}_H(V\sigma) - \mathrm{gap}_{F_\epsilon}(\sigma) \le \left\langle H(V\sigma), V\sigma - x'^\star\right\rangle - \left\langle F_\epsilon(\sigma), \sigma - \sigma''^\star\right\rangle$$

$$= \left\langle H(V\sigma), V\sigma''^{\star} - x'^{\star} \right\rangle + \left\langle V^\top H(V\sigma) - F_\epsilon(\sigma), \sigma - \sigma''^{\star} \right\rangle$$
$$\leq \|H(V\sigma)\| \left\|V\sigma''^{\star} - x'^{\star}\right\| + 2\left\|V^\top H(V\sigma) - F_\epsilon(\sigma)\right\| \leq \|H(V\sigma)\|\,\epsilon_{\text{set}} + 2\epsilon_{\text{map}}.$$

Conversely, let $\sigma'^{\star} = \arg\max_{\sigma' \in \Delta} \langle F_\epsilon(\sigma), \sigma - \sigma' \rangle$. By the inner approximation, there exists $V\sigma'^{\star} \in K$. Hence

$$\text{gap}_H(V\sigma) - \text{gap}_{F_\epsilon}(\sigma) \geq \left\langle H(V\sigma), V\sigma - V\sigma'^{\star} \right\rangle - \left\langle F_\epsilon(\sigma), \sigma - \sigma'^{\star} \right\rangle$$
$$= \left\langle V^\top H(V\sigma) - F_\epsilon(\sigma), \sigma - \sigma'^{\star} \right\rangle \geq -2\left\|V^\top H(V\sigma) - F_\epsilon(\sigma)\right\| \geq -2\epsilon_{\text{map}}.$$

Thus, the gap function approximation error is given by equation (2.9). The coefficient $\|H(V\sigma)\|$ is bounded by continuity on Δ. Therefore the approximation error between the gaps of $\text{VI}(F_\epsilon, \Delta)$ and $\text{VI}(H, K)$ is bounded by the set approximation error ϵ_{set} and the function approximation error ϵ_{map}.

$$-2\epsilon_{\text{map}} \leq \text{gap}_H(V\sigma) - \text{gap}_{F_\epsilon}(\sigma) \leq \|H(V\sigma)\|\,\epsilon_{\text{set}} + 2\epsilon_{\text{map}} \tag{2.9}$$

3 Three equivalent characterizations regarding the paths

The remainder of the paper studies $\text{VI}(F, \Delta)$ with a smooth function F and a simplex domain Δ. The preprocessing is necessary because $\text{VI}(F, \Delta)$ yields a particular system of equations called the fixed-point KKT system (FPKKT), which is the equation of the path we intend to follow. This section develops three equivalent characterizations of FPKKT, and hence of the paths, each of which reveals a different aspect of the framework. The linear programming form derives FPKKT from $\text{VI}(F, \Delta)$, the Brouwer formulation establishes the existence of solutions of FPKKT, and the MCP formulation connects FPKKT to the central path in interior-point methods.

3.1 Linear programming form and path equation

To apply path-following methods, we seek an equation whose solutions coincide with those of $\text{VI}(F, \Delta)$. We begin by introducing an optimization problem associated with $\text{VI}(F, \Delta)$. Recall that $\text{VI}(F, \Delta)$ equivalently requires $\sigma^\star \in \Delta$ such that $\sigma^{\star\top} F(\sigma^\star) \leq \sigma'^\top F(\sigma^\star)$ for all $\sigma' \in \Delta$. Equation (3.1) shows a parameterized linear programming problem (LP) that resembles this requirement, where $\hat{\sigma}$ is the optimization variable and σ is the parameter.

$$\begin{aligned} \min_{\hat{\sigma}} \quad & \hat{\sigma}^\top F(\sigma) \\ \text{s.t.} \quad & \mathbf{1}^\top \hat{\sigma} = 1 \\ & \hat{\sigma} \geq 0 \end{aligned} \tag{3.1}$$

The Lagrangian function of LP (3.1) is

$$L = \hat{\sigma}^\top F(\sigma) + v(\mathbf{1}^\top \hat{\sigma} - 1) - r^\top \hat{\sigma} = \hat{\sigma}^\top (F(\sigma) + v\mathbf{1} - r) - v,$$

where $r \geq 0$ and v are Lagrange multipliers. The $\mu = 0$ case of equation (3.2a), together with $r \geq 0$, constitutes the KKT conditions of LP (3.1). We additionally require the unbiased condition $\hat{\sigma} = \sigma$ to capture the requirement of $\text{VI}(F, \Delta)$, and refer to the simultaneous equation (3.2) as the fixed-point KKT system (FPKKT).

$$\begin{bmatrix} \hat{\sigma} \circ r - \mu \\ r - F(\sigma) - v\mathbf{1} \\ \mathbf{1}^\top \hat{\sigma} - 1 \end{bmatrix} = 0 \tag{3.2a}$$

$$\hat{\sigma} = \sigma \tag{3.2b}$$

The $\mu \geq 0$ case of equation (3.2a) gives the perturbed KKT conditions of LP (3.1), where μ is a vector. Unlike the standard KKT conditions, we do not require $r \geq 0$ explicitly, because $\mu \geq 0$ already implies it if $\hat{\sigma} \in \Delta$.

LP (3.1) and FPKKT (3.2) equivalently characterize $\mathrm{VI}(F, \Delta)$, as Theorem 2 shows. This theorem establishes the equivalence between solutions of FPKKT (3.2) and $\mathrm{VI}(F, \Delta)$ for $\mu = 0$. If FPKKT (3.2) has solutions at every $\mu > 0$, these solutions form a path leading to solutions of $\mathrm{VI}(F, \Delta)$, analogous to the central path.

Theorem 2. *For $\sigma^\star \in \Delta$, the following statements are equivalent.*

(i) *$\sigma^\star$ is a solution of $\mathrm{VI}(F, \Delta)$.*

(ii) *$\sigma^\star$ is an optimizer of LP (3.1) parameterized by $\sigma^\star$.*

(iii) *$\sigma^\star$ satisfies the $\mu = 0$ case of FPKKT (3.2) for some v and $r \geq 0$.*

Proof. (i) $\Leftrightarrow$ (ii): By definition, (i) is equivalent to $\langle F(\sigma^\star), \sigma' - \sigma^\star \rangle \geq 0$ for all $\sigma' \in \Delta$, and (ii) is equivalent to $\sigma^{\star\top} F(\sigma^\star) \leq \sigma'^{\top} F(\sigma^\star)$ for all $\sigma' \in \Delta$. The equivalence follows.

(ii) $\Leftrightarrow$ (iii): For every parameter $\sigma \in \Delta$, LP (3.1) is a linear programming problem. A point $\hat{\sigma}$ is an optimizer of LP (3.1) if and only if it satisfies the KKT conditions, which are given by equation (3.2a) and $r \geq 0$. Substituting $\hat{\sigma} = \sigma = \sigma^\star$ yields the equivalence. □

3.2 Brouwer fixed-point form and path existence

Having established the equivalence for $\mu = 0$, we next address the existence of paths as $\mu \to \mathbf{0}^+$. For path-following, it is essential that FPKKT (3.2) admit solutions for every $\mu > 0$. We therefore study the $\mu > 0$ case.

The correspondence from (σ, μ) to $(\hat{\sigma}, r, v)$ satisfying the perturbed KKT conditions (3.2a) is called the Brouwer function

$$(\hat{\sigma}, r, v) = M(\sigma, \mu), \quad \text{s.t. perturbed KKT conditions (3.2a)}, (\sigma, \mu) \in \Delta \times \{\mu | \mu > 0\}. \tag{3.3}$$

We have the following proposition.

Proposition 3. *$(\hat{\sigma}, r, v) = M(\sigma, \mu)$ is a continuous function on $\Delta \times \{\mu | \mu > 0\}$.*

Proof. The correspondence is single-valued. For every (σ, μ), the corresponding v is the unique root of $q(v) = 0$, where

$$q(v) = \mathbf{1}^\top \frac{\mu}{F(\sigma) + v\mathbf{1}} - 1, \quad \frac{dq(v)}{dv} = -\mathbf{1}^\top \frac{\mu}{(F(\sigma) + v\mathbf{1})^2}.$$

Indeed, $dq(v)/dv < 0$. The limiting behavior is

$$\lim_{v \to (-\min_i F_i(\sigma))^+} q(v) = +\infty \quad \text{and} \quad \lim_{v \to \infty} q(v) = -1.$$

Thus $q(v)$ decreases monotonically from $+\infty$ to -1 on $(-\min_i F_i(\sigma), +\infty)$. Consequently, v is unique for every (σ, μ), hence so is $(\hat{\sigma}, r, v)$.

Moreover, since $q(-\min_i F_i(\sigma) + \mathbf{1}^\top \mu) \leq 0$, the unique solution v belongs to the interval $[-\min_i F_i(\sigma), -\min_i F_i(\sigma) + \mathbf{1}^\top \mu]$.

Let $(\sigma_k, \mu_k) \to (\sigma_0, \mu_0)$ with $(\sigma_k, \mu_k) \in \Delta \times \{\mu | \mu > 0\}$ and $(\sigma_0, \mu_0) \in \Delta \times \{\mu | \mu > 0\}$, and let v_k be the unique solution of $q(v) = 0$ for each k. Since each v_k lies in the bounded interval above, the sequence $\{v_k\}$ has a convergent subsequence with limit point v_0. By continuity of $F(\sigma)$ and of $q(\sigma, \mu, v)$ in (σ, μ, v), we have

$$\mathbf{1}^\top \frac{\mu_0}{F(\sigma_0) + v_0 \mathbf{1}} - 1 = 0.$$

Thus v_0 is the unique solution of $q(v) = 0$ for (σ_0, μ_0). Since every convergent subsequence of $\{v_k\}$ converges to this same limit v_0, the entire sequence $\{v_k\}$ converges to v_0. Hence v is continuous with respect to (σ, μ) on $\Delta \times \{\mu | \mu > 0\}$, and continuity of $r = F(\sigma) + v\mathbf{1}$ and $\hat{\sigma} = \mu / r$ then yields continuity of $(\hat{\sigma}, r, v) = M(\sigma, \mu)$. □

By the Brouwer fixed-point theorem, for every $\mu > 0$ there exists a fixed point $\sigma = \hat{\sigma}$ of the Brouwer function $(\hat{\sigma}, r, v) = M(\sigma, \mu)$. Fixed points of $(\hat{\sigma}, r, v) = M(\sigma, \mu)$ are equivalent to solutions of FPKKT (3.2) by definition of the Brouwer function. Theorem 4 therefore guarantees the existence of paths subject to FPKKT (3.2) as $\mu \to \mathbf{0}^+$.

Theorem 4 (Existence theorem)**.** *For any $\mu > 0$, there exists $\sigma \in \Delta$ satisfying FPKKT* (3.2)*.*

The proof of Theorem 4 also suggests that the Brouwer function can be computed by solving the bisection problem (3.4).

$$q(v) = \mathbf{1}^\top \frac{\mu}{F(\sigma) + v\mathbf{1}} - 1, \quad v \in [-\min_i F_i(\sigma), -\min_i F_i(\sigma) + \mathbf{1}^\top \mu] \tag{3.4}$$

3.3 Mixed complementarity problem form and central path

Having established the existence of paths satisfying FPKKT (3.2), we now show that FPKKT (3.2) can be interpreted as the equation of a distinguished central path of the mixed complementarity problem (MCP) in equation (3.5). Deep connections between complementarity problems, variational inequalities, and Brouwer fixed-point problems are well established [35, 36, 37].

$$\begin{aligned} \min_{(\sigma, r, v)} \quad & \sigma^\top r - \mu^\top \ln \sigma - \mu^\top \ln r \\ \text{s.t.} \quad & r = F(\sigma) + v\mathbf{1} \\ & \mathbf{1}^\top \sigma = 1 \\ & (\sigma, r) \geq 0 \end{aligned} \tag{3.5}$$

The $\mu = 0$ case of equation (3.5) is obtained by taking the complementarity pair $\sigma \circ r$ from FPKKT (3.2) as the objective, and the $\mu > 0$ case is the barrier problem of the MCP. The Lagrangian function of MCP (3.5) is

$$L = \sigma^\top r + \bar{\lambda}^\top (r - F(\sigma) - v\mathbf{1}) + \tilde{\lambda}(\mathbf{1}^\top \sigma - 1) - \check{r}^\top \sigma - \check{\sigma}^\top r,$$

where $\bar{\lambda}$, $\tilde{\lambda}$, $\check{\sigma} \geq 0$, and $\check{r} \geq 0$ are Lagrange multipliers. The perturbed KKT conditions of MCP (3.5) are

$$\begin{bmatrix} -\left(\partial F(\sigma)/\partial \sigma\right)^\top \bar{\lambda} + \tilde{\lambda}\mathbf{1} + r - \check{r} \\ \bar{\lambda} + \sigma - \check{\sigma} \\ -\bar{\lambda}^\top \mathbf{1} \\ r \circ \check{\sigma} - \mu \\ \sigma \circ \check{r} - \mu \\ r - F(\sigma) - v\mathbf{1} \\ \mathbf{1}^\top \sigma - 1 \end{bmatrix} = 0. \tag{3.6}$$

Theorem 5 shows that the three characterizations are equivalent representations of the same underlying object.

Theorem 5. *For $(\sigma, \mu) \in \Delta \times \{\mu | \mu > 0\}$, the following statements are equivalent.*

(i) *(σ, μ) satisfies FPKKT (3.2) for some r, v.*

(ii) *σ is a fixed point such that $\sigma = \hat{\sigma}$ of the Brouwer function $(\hat{\sigma}, r, v) = M(\sigma, \mu)$.*

(iii) *(σ, μ) satisfies $\check{\sigma} = \sigma$ and the perturbed KKT conditions of MCP (3.5) for some $r, v, \bar{\lambda}, \tilde{\lambda}, \check{\sigma}, \check{r}$.*

Proof. (i) $\Leftrightarrow$ (ii): The Brouwer function $(\hat{\sigma}, r, v) = M(\sigma, \mu)$ is defined by the perturbed KKT conditions (3.2a). FPKKT (3.2) is the simultaneous system consisting of the perturbed KKT conditions (3.2a) and $\sigma = \hat{\sigma}$, so the equivalence follows.

(i) $\Leftrightarrow$ (iii): Given (σ, μ, r, v) satisfying FPKKT (3.2), the tuple given by $(\sigma, \mu, r, v, \bar{\lambda}, \tilde{\lambda}, \check{\sigma}, \check{r}) = (\sigma, \mu, r, v, \mathbf{0}, 0, \sigma, r)$ satisfies $\check{\sigma} = \sigma$ and the perturbed KKT conditions (3.6). Conversely, if $(\sigma, \mu, r, v, \bar{\lambda}, \tilde{\lambda}, \check{\sigma}, \check{r})$ satisfies $\check{\sigma} = \sigma$ and the perturbed KKT conditions (3.6), then the tuple (σ, μ, r, v) satisfies FPKKT (3.2). □

Central paths are defined by perturbed KKT conditions. Thus the path given by FPKKT (3.2) is precisely the central path of MCP (3.5) on which $\check{\sigma} = \sigma$. As in interior-point methods, these paths lead to solutions of $\mathrm{VI}(F, \Delta)$ as $\mu \to \mathbf{0}^+$.

Note that $(\sigma, \mathbf{0})$ satisfying FPKKT (3.2) is not necessarily a solution of $\mathrm{VI}(F, \Delta)$, since the $\mu = 0$ case additionally requires $r \geq 0$. For $\sigma \in \Delta$, $r \geq 0$ if and only if $\mu \geq 0$, because $\sigma \circ r = \mu$. Hence a point $(\sigma^\star, \mathbf{0})$ satisfying FPKKT (3.2) must be a limit point as $\mu \to \mathbf{0}^+$ in order to be a solution of $\mathrm{VI}(F, \Delta)$. This is proved in the next section.

4 Fiber bundle as the solution space of FPKKT

The previous section established FPKKT (3.2) as an equivalent characterization of $\mathrm{VI}(F, \Delta)$ and characterized the paths subject to it as special central paths. We now introduce the central geometric object of the paper. Instead of viewing FPKKT (3.2) as an equation system, we regard its solution set as a geometric space. This viewpoint reveals a fiber-bundle structure that unifies starting point selection, path-following, singularity avoidance, and convergence to solutions of $\mathrm{VI}(F, \Delta)$. In this section we study the solution space of FPKKT (3.2) with $\mu \in \mathbb{R}^n$ and formally show that the paths as $\mu \to \mathbf{0}^+$ equivalently lead to solutions of $\mathrm{VI}(F, \Delta)$.

4.1 The fixed-point bundle

The solution space of FPKKT (3.2) has a natural fiber-bundle structure. For each σ, the admissible μ values form an affine one-dimensional family given by $\mu = \sigma \circ F(\sigma) + v\sigma$. The disjoint union of these affine lines over all $\sigma \in \Delta$ forms the fiber bundle of equation (4.1). The bundle is denoted $(E, \Delta, \alpha : E \to \Delta)$, where E is the total space, Δ is the base space, $\alpha : E \to \Delta$ is the projection map, and $\alpha^{-1}(\sigma) = \{\sigma\} \times B(\sigma)$ is the fiber over σ.

$$
\begin{aligned}
E &= \bigcup_{\sigma \in \Delta} \{\sigma\} \times B(\sigma) \\
B(\sigma) &= \{\sigma \circ F(\sigma) + v\sigma | v \in \mathbb{R}\} \\
\alpha\left((\sigma, \mu)\right) &= \sigma
\end{aligned}
\tag{4.1}
$$

We refer to this as the fixed-point bundle, and by convention we also use the total space E to refer to the fiber bundle. This geometric structure is not merely a change of terminology: it provides a characterization of the paths that separates motion of σ along the base space from motion of μ along individual fibers, a separation that is crucial for starting point selection and singularity avoidance.

We eliminate r and v in FPKKT (3.2) to obtain a fixed-point bundle equation in (σ, μ) only. Applying the projection matrix $I - \sigma\mathbf{1}^\top$ to FPKKT (3.2) yields the equation $G(\sigma, \mu) = 0$ of equation (4.2). We work primarily with this equivalent formulation in place of FPKKT (3.2); it describes the fixed-point bundle and serves as the fundamental equation throughout the remainder of the paper. We write $(\sigma, \mu) \in E$ to denote points satisfying $G(\sigma, \mu) = 0$.

$$G(\sigma, \mu) = \left(I - \sigma\mathbf{1}^\top\right)(\sigma \circ F(\sigma) - \mu), \quad (\sigma, \mu) \in \Delta \times \mathbb{R}^n \tag{4.2}$$

The projection matrix $I - \sigma\mathbf{1}^\top$ appears throughout this paper, and the following properties are used frequently in the proofs and derivations. It projects along the vector σ, so that $(I - \sigma\mathbf{1}^\top)\sigma = 0$; it projects onto the subspace orthogonal to $\mathbf{1}$, so that $\mathbf{1}^\top(I - \sigma\mathbf{1}^\top) = 0$; and it is idempotent, so that $(I - \sigma\mathbf{1}^\top)(I - \sigma\mathbf{1}^\top) = I - \sigma\mathbf{1}^\top$. The linear equation $(I - \sigma\mathbf{1}^\top)x = b$ is solvable when $\mathbf{1}^\top b = 0$, and its general solution is $x = b + k\sigma$ for $k \in \mathbb{R}$. Moreover, $I - \mathbf{1}\sigma^\top = (I - \sigma\mathbf{1}^\top)^\top$ is another projection matrix with similar properties.

We introduce two sections of the fixed-point bundle. A section of a fiber bundle is a map that sends every σ to a point on the fiber $B(\sigma)$. The nonnegative section $\check{\mu}(\sigma)$ in equation (4.3) satisfies $\check{\mu}(\sigma) \geq 0$ and directly characterizes the gap of $\mathrm{VI}(F, \Delta)$. As a section of the fiber bundle, it satisfies $(\sigma, \check{\mu}(\sigma)) \in E$ for every $\sigma \in \Delta$.

$$\check{\mu}(\sigma) = \sigma \circ \left(F(\sigma) - \left(\min_i F_i(\sigma)\right)\mathbf{1}\right) \tag{4.3}$$

By Theorem 6, the section $\check{\mu}(\sigma)$ characterizes the gap of $\mathrm{VI}(F, \Delta)$; this theorem establishes the fundamental link between the fixed-point bundle E and the variational inequality $\mathrm{VI}(F, \Delta)$.

Theorem 6. *For $(\sigma, \mu) \in \Delta \times \mathbb{R}^n$, the following equation holds.*

$$\mathrm{gap}(\sigma) = \mathbf{1}^\top\check{\mu}(\sigma) = \mathbf{1}^\top\mu - \min_i\left((G(\sigma, \mu) + \mu)/\sigma\right)_i \tag{4.4}$$

Consequently, for $\sigma^\star \in \Delta$, the following statements are equivalent.

(i) *$\sigma^\star$ is a solution of $\mathrm{VI}(F, \Delta)$.*

(ii) *$\sigma^\star$ is a zero point of $\check{\mu}(\sigma)$.*

(iii) *$(\sigma^\star, \mathbf{0})$ is a limit point on $E \cap (\Delta \times \{\mu | \mu > 0\})$.*

Proof. The gap function of $\mathrm{VI}(F, \Delta)$ satisfies

$$\mathrm{gap}(\sigma) = \sup_{\sigma' \in \Delta}\left\langle F(\sigma), \sigma - \sigma'\right\rangle = \sigma^\top F(\sigma) - \min_i F_i(\sigma).$$

The section $\check{\mu}(\sigma)$ satisfies

$$\begin{aligned}&\mathbf{1}^\top\check{\mu}(\sigma) = \sigma^\top F(\sigma) - \min_i F_i(\sigma)\\ =&\sigma^\top F(\sigma) - \min_i\left(\frac{G(\sigma, \mu) + \mu + (\sigma^\top F(\sigma) - \mathbf{1}^\top\mu)\sigma}{\sigma}\right)_i = \mathbf{1}^\top\mu - \min_i\left(\frac{G(\sigma, \mu) + \mu}{\sigma}\right)_i.\end{aligned}$$

This proves the equation.

(i) $\Leftrightarrow$ (ii): Since $\mathrm{gap}(\sigma) \geq 0$ and $\check{\mu}(\sigma) \geq 0$, $\mathrm{gap}(\sigma) = 0$ is equivalent to $\check{\mu}(\sigma) = 0$.

(ii) $\Rightarrow$ (iii): For any $\sigma \in \Delta$ and $k \in \mathbb{R}$, we have $(\sigma, \check{\mu}(\sigma) + k\sigma) \in E$. Continuity of $\check{\mu}(\sigma)$ gives $\check{\mu}(\sigma) \to \check{\mu}(\sigma^\star)$ as $\sigma \to \sigma^\star$. For any $\sigma_t \to \sigma^\star$ with $\sigma_t > 0$ and any $k_t \to 0$ with $k_t > 0$, we have $\mu_t = \check{\mu}(\sigma_t) + k_t\sigma_t > 0$ and $\mu_t \to 0$. Hence $(\sigma^\star, \mathbf{0})$ is a limit point of $(\sigma_t, \mu_t) \in E$ on $\Delta \times \{\mu | \mu > 0\}$.

(iii) $\Rightarrow$ (ii): Given $(\sigma, \mu) \in E$, we have $G(\sigma, \mu) = 0$. If additionally $\mu > 0$, then $\min_i((G(\sigma, \mu) + \mu)/\sigma)_i > 0$, and consequently $0 \leq \mathbf{1}^\top \check{\mu}(\sigma) < \mathbf{1}^\top \mu$. If $(\sigma_t, \mu_t) \to (\sigma^\star, \mathbf{0})$ with $(\sigma_t, \mu_t) \in E$ and $\mu_t > 0$, then $\mathbf{1}^\top \check{\mu}(\sigma_t) \to 0$, hence $\check{\mu}(\sigma_t) \to 0$ since $\check{\mu}(\sigma) \geq 0$. Continuity of $\check{\mu}(\sigma)$ yields $\check{\mu}(\sigma^\star) = 0$. $\square$

Theorem 6 reveals that solutions of $\mathrm{VI}(F, \Delta)$ are equivalently zero points of $\check{\mu}(\sigma)$ and equivalently limit points of the paths satisfying $G(\sigma, \mu) = 0$ as $\mu \to \mathbf{0}^+$. This finally justifies path-following subject to $G(\sigma, \mu) = 0$ (equivalently FPKKT (3.2)) as $\mu \to \mathbf{0}^+$ to solve $\mathrm{VI}(F, \Delta)$.

Equation (4.4) also provides a measure of the gap of $\mathrm{VI}(F, \Delta)$. Specifically, for any $(\sigma, \mu) \in \Delta \times \{\mu | \mu > 0\}$, if $|G(\sigma, \mu)| \leq \mu$ componentwise, then $0 \leq \mathbf{1}^\top \check{\mu}(\sigma) \leq \mathbf{1}^\top \mu$. Thus $G(\sigma, \mu)$ measures the proximity to the fixed-point bundle and $\mathbf{1}^\top \mu$ measures the proximity to $\mu = 0$. When both are sufficiently precise, so is $\mathbf{1}^\top \check{\mu}(\sigma)$, which translates precision in approaching the path and its endpoint into precision in solving $\mathrm{VI}(F, \Delta)$.

4.2 Differential properties

Having established that paths on the fixed-point bundle equivalently lead to the desired solutions, we now develop the differential properties needed to construct the path and the algorithm that follows it. To keep $\sigma \in \Delta$, we use the parameterization

$$\sigma = \mathrm{softmax}(\theta) = \frac{\exp(\theta)}{\mathbf{1}^\top \exp(\theta)}. \tag{4.5}$$

The differential $d\sigma$ with respect to $d\theta$ is

$$d\sigma = \frac{\exp(\theta) \circ d\theta}{\mathbf{1}^\top \exp(\theta)} - \frac{\mathbf{1}^\top (\exp(\theta) \circ d\theta)}{(\mathbf{1}^\top \exp(\theta))^2} \exp(\theta) = \sigma \circ d\theta - (\sigma^\top d\theta)\sigma = \sigma \circ \left(I - \mathbf{1}\sigma^\top\right) d\theta. \tag{4.6}$$

For any $d\theta$ we have $\mathbf{1}^\top d\sigma = 0$. Conversely, if $\mathbf{1}^\top d\sigma = 0$, the linear equation $d\sigma/\sigma = (I - \mathbf{1}\sigma^\top)d\theta$ admits the general solution $d\theta = d\sigma/\sigma + k\mathbf{1}$ with $k \in \mathbb{R}$. We denote $\overline{d\theta} := (I - \mathbf{1}\sigma^\top)d\theta$. Since adding a multiple of $\mathbf{1}$ to $d\theta$ does not change $d\sigma$, we use $\overline{d\theta}$ as the differential throughout this paper.

We introduce the differentiable section $\tilde{\mu}(\sigma)$ in equation (4.7), which satisfies $\mathbf{1}^\top \tilde{\mu}(\sigma) = 0$ and governs the differential properties of the fixed-point bundle. As a section of the fiber bundle, it satisfies $(\sigma, \tilde{\mu}(\sigma)) \in E$ for every $\sigma \in \Delta$.

$$\tilde{\mu}(\sigma) = \sigma \circ \left(F(\sigma) - \left(\sigma^\top F(\sigma)\right)\mathbf{1}\right) = \left(I - \sigma\mathbf{1}^\top\right)(\sigma \circ F(\sigma)) \tag{4.7}$$

We use the section $\tilde{\mu}(\sigma)$ and its Jacobian to study the differential properties of the fixed-point bundle. The differential $d\tilde{\mu}(\sigma)$ with respect to $\overline{d\theta}$ is derived as follows; the third line uses $\mathbf{1}^\top d\sigma = 0$, and the last line uses $d\sigma/\sigma = \overline{d\theta} = (I - \mathbf{1}\sigma^\top)\overline{d\theta}$.

$$\begin{aligned} d\tilde{\mu}(\sigma) &= \left(I - \sigma\mathbf{1}^\top\right)\left(\sigma \circ \frac{\partial F(\sigma)}{\partial \sigma} + F(\sigma) \circ I\right) d\sigma - \mathbf{1}^\top (\sigma \circ F(\sigma))\, d\sigma \\ &= \left(I - \sigma\mathbf{1}^\top\right)\left(\sigma \circ \frac{\partial F(\sigma)}{\partial \sigma} + F(\sigma) \circ I - (\sigma^\top F(\sigma))I\right) d\sigma - \left(\sigma^\top F(\sigma)\right)\sigma\mathbf{1}^\top d\sigma \end{aligned}$$

$$= \sigma \circ \left(I - \mathbf{1}\sigma^\top\right) \left(\frac{\partial F(\sigma)}{\partial \sigma} \circ \sigma + \left(\left(I - \mathbf{1}\sigma^\top\right) F(\sigma) \right) \circ I \right) \frac{d\sigma}{\sigma}$$
$$= \sigma \circ \left(I - \mathbf{1}\sigma^\top\right) \left(\frac{\partial F(\sigma)}{\partial \sigma} \circ \sigma + \left(\left(I - \mathbf{1}\sigma^\top\right) F(\sigma) \right) \circ I \right) \left(I - \mathbf{1}\sigma^\top\right) \overline{d\theta}$$

We introduce the matrix $J(\sigma)$ in equation (4.8), where $k \in \mathbb{R}$ is an arbitrary scalar. This Jacobian-like matrix governs the differential properties directly and appears in the singular manifold, the differential equation, and the Newton equation of the fixed-point bundle.

$$J(\sigma) := \left(I - \mathbf{1}\sigma^\top\right) \left(\frac{\partial F(\sigma)}{\partial \sigma} \circ \sigma + \left(\left(I - \mathbf{1}\sigma^\top\right) F(\sigma) \right) \circ I \right) \left(I - \mathbf{1}\sigma^\top\right) + k\mathbf{1}\sigma^\top \tag{4.8}$$

Using $\sigma^\top \overline{d\theta} = 0$, we obtain the differential equation of $\tilde{\mu}(\sigma)$:

$$d\tilde{\mu}(\sigma) = \sigma \circ J(\sigma)\overline{d\theta}. \tag{4.9}$$

The purpose of the arbitrary scalar $k \in \mathbb{R}$ in $J(\sigma)$ is to adjust the eigenvalues. When $k = 0$, $J(\sigma)$ has an intrinsic zero eigenvalue with eigenvector $\mathbf{1}$. Proposition 7 shows how the eigenvalues are adjusted. Recall that a subspace U is an invariant subspace of a matrix A if $Au \in U$ for every $u \in U$, and that if A admits an invariant subspace decomposition, its eigenvalues are the union of the eigenvalues of its restrictions to these invariant subspaces.

Proposition 7. *Matrix $J(\sigma)$ admits the invariant subspace decomposition $\mathbb{R}^n = \mathrm{span}(\mathbf{1}) \oplus \ker(\sigma^\top)$. The eigenvalue corresponding to eigenvectors in $\mathrm{span}(\mathbf{1})$ is k, while eigenvalues corresponding to eigenvectors in $\ker(\sigma^\top)$ are independent of k.*

Proof. First, $J(\sigma)\mathbf{1} = k\mathbf{1}$ since $(I - \mathbf{1}\sigma^\top)\mathbf{1} = 0$ and $\sigma^\top\mathbf{1} = 1$; hence $\mathrm{span}(\mathbf{1})$ is an invariant subspace. Second, $\sigma^\top J(\sigma) = k\sigma^\top$ since $\sigma^\top(I - \mathbf{1}\sigma^\top) = 0$ and $\sigma^\top\mathbf{1} = 1$. Then $\ker(\sigma^\top) = \{u | \sigma^\top u = 0\}$ is invariant because $\sigma^\top u = 0$ implies $\sigma^\top J(\sigma)u = 0$. Since $\mathrm{span}(\mathbf{1})$ and $\ker(\sigma^\top)$ are complementary, this is an invariant subspace decomposition.

For an eigenvector in $\mathrm{span}(\mathbf{1})$, the corresponding eigenvalue is k because $J(\sigma)\mathbf{1} = k\mathbf{1}$. For $u \in \ker(\sigma^\top)$, the equation $J(\sigma)u = \lambda u$ reduces to $J_0(\sigma)u = \lambda u$, where $J_0(\sigma)$ is the $k = 0$ case of $J(\sigma)$, since $\sigma^\top u = 0$. Thus the eigenvalue λ is independent of k. □

Proposition 7 shows that the intrinsic zero eigenvalue of $J(\sigma)$ corresponding to the eigenvector $\mathbf{1}$ is removed, while the remaining eigenvalues are unchanged. This device prevents $J(\sigma)$ from being intrinsically singular, for theoretical simplicity; however, we use $J(\sigma)$ with $k = 0$ in the algorithm for computational simplicity.

Next we derive the differential equation of $G(\sigma, \mu) = 0$. This equation provides the tangent step in θ for a given reduction $d\mu$ along the fixed-point bundle.

$$G'_\theta(\sigma, \mu)d\theta = -G'_\mu(\sigma, \mu)d\mu$$
$$\sigma \circ J(\sigma)\overline{d\theta} + (\mathbf{1}^\top\mu)\sigma \circ \overline{d\theta} = \left(I - \sigma\mathbf{1}^\top\right) d\mu$$

Multiplying both sides by $1/\sigma$, we obtain the differential equation of $G(\sigma, \mu) = 0$ in equation (4.10). We write $J_G := J(\sigma) + (\mathbf{1}^\top\mu)I$ as shorthand.

$$\left(J(\sigma) + (\mathbf{1}^\top\mu)I\right) \overline{d\theta} = \left(I - \mathbf{1}\sigma^\top\right) (d\mu/\sigma) \tag{4.10}$$

A point (σ, μ) on the fixed-point bundle is called singular if $J(\sigma) + (\mathbf{1}^\top\mu)I$ is singular. For any $\sigma \in \Delta$, the fiber $B(\sigma)$ contains at most $n - 1$ singular points, corresponding to the eigenvalues of

$-J(\sigma)$ other than the arbitrary $-k$. The singular manifold of the fixed-point bundle is therefore constructed from the real eigenvalues of $J(\sigma)$. Let $\lambda_j(J(\sigma)) \in \mathbb{R}$ denote the real eigenvalues of $J(\sigma)$ other than k. The set of all singular points of E is formalized as the singular manifold S in equation (4.11).

$$S = \{(\sigma, \tilde{\mu}(\sigma) - \lambda_j\left(J(\sigma)\right)\sigma) | \lambda_j\left(J(\sigma)\right) \in \mathbb{R}, \sigma \in \Delta\}. \tag{4.11}$$

Indeed, S consists of all singular points of the fixed-point bundle. First, for every $\sigma \in \Delta$ and $(\sigma, \mu) \in S$, the point μ lies on the fiber $B(\sigma)$ over σ, because $\tilde{\mu}(\sigma)$ maps σ to a point on $B(\sigma)$ and, if $\mu_0 \in B(\sigma)$, then $\mu_0 + k\sigma \in B(\sigma)$ for any $k \in \mathbb{R}$. Second, for every $\sigma \in \Delta$, we have $(\sigma, \mu) \in S$ if and only if $\mathbf{1}^\top \mu$ is an eigenvalue of $-J(\sigma)$: indeed, $\mathbf{1}^\top \tilde{\mu}(\sigma) = 0$ and $\mathbf{1}^\top \sigma = 1$, so $\mathbf{1}^\top \mu = -\lambda_j(J(\sigma))$, which are precisely the real eigenvalues of $-J(\sigma)$ other than $-k$.

5 Paths on the fixed-point bundle

Having shown that paths on the fixed-point bundle equivalently lead to solutions of $\mathrm{VI}(F, \Delta)$, and having prepared the differential properties of the bundle, we now establish a globally convergent path. We show that from any designated starting point there exists a path leading to solutions of $\mathrm{VI}(F, \Delta)$, and moreover that this path is well-behaved enough to be followed.

5.1 Analytic curves

Consider the curve on the fixed-point bundle E defined by

$$\Gamma(\sigma_{\text{init}}) = \{(\sigma, \mu) \in E | \mu = \gamma\sigma_{\text{init}}, \gamma \in [0, +\infty)\} \tag{5.1}$$

with designated $\sigma_{\text{init}} \in \Delta$. This is a one-dimensional curve parameterized by γ, and every $(\sigma, \mu) \in E$ on it satisfies $G(\sigma, \mu) = 0$. When $F(\sigma)$ is real-analytic, so is $G(\sigma, \mu)$, and $\Gamma(\sigma_{\text{init}})$ is called an analytic curve because it is the zero set of an analytic function.

Since $\sigma \circ F(\sigma) + v\sigma = \gamma\sigma_{\text{init}}$ and $F(\sigma)$ is bounded, $\Gamma(\sigma_{\text{init}})$ starts from a single point $(\sigma_{\text{init}}, \gamma\sigma_{\text{init}})$ as $\gamma \to +\infty$. Thus $\Gamma(\sigma_{\text{init}})$ is a path that starts from any designated $\sigma_{\text{init}} \in \Delta$ as $\gamma \to +\infty$ and ends at solutions of $\mathrm{VI}(F, \Delta)$ as $\gamma \to 0^+$.

The curve $\Gamma(\sigma_{\text{init}})$ is smooth at nonsingular points, where path-following behaves well, but singular points may occur, and traditional path-following methods struggle at such points. The fixed-point bundle structure provides a simple mechanism for singularity avoidance through motion along the fiber $B(\sigma)$. Specifically, at a point (σ, μ) where $\mathbf{1}^\top \mu$ equals an eigenvalue of $-J(\sigma)$, we jump to $(\sigma, \mu + \beta\sigma)$ with β chosen so that $\mu + \beta\sigma > 0$ and $\mathbf{1}^\top \mu + \beta$ is not an eigenvalue of $-J(\sigma)$. Then $(\sigma, \mu + \beta\sigma)$ is no longer singular, and we can proceed to reduce $\mu + \beta\sigma$. This moves us to a new analytic curve $\Gamma((\mu + \beta\sigma)/(\mathbf{1}^\top \mu + \beta))$ with a different starting point.

Analogously to the classical oddness theorem for Nash equilibria [38], there is an oddness theorem for $\mathrm{VI}(F, \Delta)$.

Theorem 8 (Oddness theorem)**.** *For every $\mu > 0$, if $J(\sigma) + (\mathbf{1}^\top \mu)I$ is nonsingular at every solution of $G(\sigma, \mu) = 0$, then there are an odd number of solutions.*

Proof. The proof follows the fixed-point index argument developed in [39].

Let M_μ denote the map from σ to $\hat{\sigma}$ given by the Brouwer function $(\hat{\sigma}, r, v) = M(\sigma, \mu)$. Since M_μ is a continuous function that maps the compact convex set Δ to itself, its global fixed-point index on Δ is $i(M_\mu, \Delta) = 1$.

For a fixed point σ of M_μ, if $I - DM_\mu(\sigma)$ is nonsingular, then the local fixed-point index of σ is given by the Lefschetz-Hopf formula

$$\mathrm{ind}(M_\mu, \sigma) = \mathrm{sign}(\det(I - DM_\mu(\sigma))) \in \{+1, -1\}.$$

As Theorem 5 shows, the fixed-point equation $\hat{\sigma} = \sigma$ subject to $(\hat{\sigma}, r, v) = M(\sigma, \mu)$ and the fixed-point bundle equation $G(\sigma, \mu) = 0$ are equivalent through linear transformations. In addition, since $\mu > 0$, we have $\sigma > 0$ and $r > 0$, so the linear transformations are invertible. Thus, the singularity of $I - DM_\mu(\sigma)$ is equivalent to that of $J(\sigma) + (\mathbf{1}^\top \mu) I$.

Nonsingular solutions are isolated by the inverse function theorem. Since the solutions lie in the compact simplex Δ, there are finitely many of them; otherwise there would be an accumulation point, contradicting isolation.

Finally, by the additivity property of the fixed-point index, when the number of fixed points is finite, the global index equals the sum of the local indices. Thus, a finite number of 1 or -1 sums to 1, which implies that the number must be odd. □

By Sard's theorem [40], for almost all $\mu > 0$, all solutions of $G(\sigma, \mu) = 0$ are nonsingular. Thus, Theorem 8 also implies that the analytic curve $\Gamma(\sigma_{\text{init}})$ almost always consists of finitely many branches. By the Newton-Puiseux theorem [41], every branch of $\Gamma(\sigma_{\text{init}})$ has a unique analytic continuation through every point on it. Thus, the only boundary points of $\Gamma(\sigma_{\text{init}})$ occur in the limits $\gamma \to 0^+$ and $\gamma \to +\infty$: the former correspond to solutions of $\mathrm{VI}(F, \Delta)$, while the latter is the unique starting point σ_{init}. Therefore, if the solutions of $\mathrm{VI}(F, \Delta)$ are all nonsingular, the branches of $\Gamma(\sigma_{\text{init}})$ connect the solutions and the unique starting point σ_{init} in pairs.

5.2 The path

The path we follow is derived from the analytic curve $\Gamma(\sigma_{\text{init}})$ and singularity avoidance along the fibers. This is formalized as $\tilde{\Gamma}$ in equation (5.2). The key idea is that $\tilde{\Gamma}$ is obtained by concatenating segments of analytic curves separated by fiber jumps. The starting point $(\sigma_0, \gamma_0\mu_0) \in E$ is given by $\gamma_0\mu_0 = v_{\text{init}}\sigma_{\text{init}}$, with designated σ_{init} and sufficiently large v_{init}. Then $\tilde{\Gamma}$ consists of segments of different analytic curves $\Gamma(\sigma_{\text{init}})$, concatenated in σ but jumping along fibers in μ. It eventually ends at $(\sigma^\star, \mathbf{0})$ with $\mathbf{1}^\top \mu = 0$, where $\sigma^\star$ is a solution of $\mathrm{VI}(F, \Delta)$.

$$\begin{aligned} \tilde{\Gamma} = \bigcup_{q=1}^{Q} \{(\sigma, \mu) \in E | \mu = (1-\gamma)\hat{\mu}_{q-1} + \gamma\mu_q, \gamma \in [0,1]\}, \\ \hat{\mu}_{q-1} = \mu_{q-1} + \beta_q \sigma_{q-1},\ \mu_0 = v_{\text{init}}\sigma_{\text{init}},\ \mu_Q = \mathbf{0} \end{aligned} \tag{5.2}$$

We would like $\tilde{\Gamma} \cap S = \varnothing$ for singularity avoidance, where S is the singular manifold. However, $\mathbf{1}^\top \mu = 0$ is unavoidable at the solution $(\sigma^\star, \mathbf{0})$. If $J(\sigma^\star)$ is singular, then $(\sigma^\star, \mathbf{0})$ is an unavoidable singular point, that is, $\tilde{\Gamma}$ must intersect S. Thus we instead require $\tilde{\Gamma} \cap S \subseteq \{(\sigma^\star, \mathbf{0})\}$ and only avoid the nonzero eigenvalues of $-J(\sigma)$. In fact, this can be achieved with only finitely many segments in $\tilde{\Gamma}$.

Proposition 9. *There exists $\tilde{\Gamma}$ consisting of finitely many segments satisfying $\tilde{\Gamma} \cap S \subseteq \{(\sigma^\star, \mathbf{0})\}$.*

Proof. Every real-analytic curve on a compact domain has only finitely many singular points on it, as long as it is not constantly singular. Thus, there are finitely many singular points on the analytic curve $\Gamma(\sigma_{\text{init}})$, because it is not constantly singular, since it is nonsingular where γ is large. For each singular point $(\sigma_s, \gamma_s\sigma_{\text{init}}) \in \Gamma(\sigma_{\text{init}})$, let d_1, d_2 be two small scalars, and let $(\sigma_1, (1+d_1)\gamma_s\sigma_{\text{init}})$ and $(\sigma_2, (1+d_2)\gamma_s\sigma_{\text{init}})$ be two points on $\Gamma(\sigma_{\text{init}})$ before and after the singular point. We adopt the following path to circumvent the singular point $(\sigma_s, \gamma_s\sigma_{\text{init}})$ on $\Gamma(\sigma_{\text{init}})$.

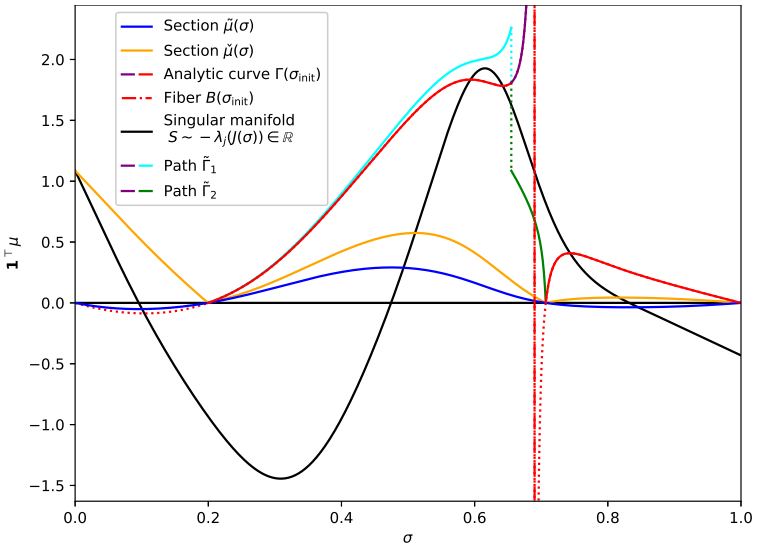


Figure 1: The fixed-point bundle E for a two-dimensional $\mathrm{VI}(F, \Delta)$. The horizontal axis represents the simplex coordinate σ, and the vertical axis represents the fiber coordinate $\mathbf{1}^\top \mu$. The smooth section $\tilde{\mu}(\sigma)$ (the first component plotted since $\mathbf{1}^\top \tilde{\mu}(\sigma) = 0$) induces the differential structure through $J(\sigma)$, while the nonnegative section $\check{\mu}(\sigma)$ (the nonzero component plotted since $\min_i \check{\mu}_i(\sigma) = 0$) equals the gap function of $\mathrm{VI}(F, \Delta)$. Analytic curves $\Gamma(\sigma_{\text{init}})$ connect a designated starting point to solutions as $\mu \to \mathbf{0}^+$. Singular points, where path-following cannot proceed, lie on the singular manifold S, where $\mathbf{1}^\top \mu$ equals a real eigenvalue of $-J(\sigma)$. The path-following trajectory $\tilde{\Gamma}$ bypasses these singularities by moving along fibers, illustrated by a positive jump ($\tilde{\Gamma}_1$) and a negative jump ($\tilde{\Gamma}_2$). After finitely many such fiber jumps, $\tilde{\Gamma}$ reaches a solution of $\mathrm{VI}(F, \Delta)$.

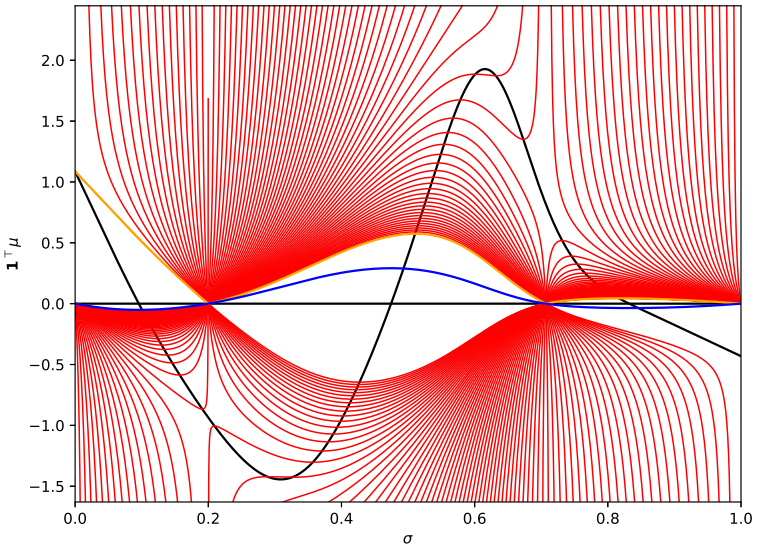


Figure 2: Analytic curves on the fixed-point bundle E with different starting points σ_{init}. For each $\mu > 0$, there is exactly one analytic curve passing through it. For each analytic curve, its intersections with the singular manifold, which is given by the real eigenvalues of $-J(\sigma)$, are singular points. Each fiber jump switches to another analytic curve with a different starting point.

(i) The fiber jump between $(\sigma_1, (1+d_1)\gamma_s\sigma_{\text{init}}) \in \Gamma(\sigma_{\text{init}})$ and $(\sigma_1, (1+d_1)(\gamma_s\sigma_{\text{init}} + \beta\sigma_1))$.

(ii) The segment changing μ from $(1+d_1)(\gamma_s\sigma_{\text{init}} + \beta\sigma_1)$ to $(1+d_2)(\gamma_s\sigma_{\text{init}} + \beta\sigma_2)$.

(iii) The fiber jump between $(\sigma_2, (1+d_2)\gamma_s\sigma_{\text{init}}) \in \Gamma(\sigma_{\text{init}})$ and $(\sigma_2, (1+d_2)(\gamma_s\sigma_{\text{init}} + \beta\sigma_2))$.

For the singular point $(\sigma_s, \gamma_s\sigma_{\text{init}})$, there exists β such that $(\sigma_s, \gamma_s\sigma_{\text{init}} + \beta\sigma_s)$ is a nonsingular point, because there are finitely many singular points on each fiber. Since $(\sigma_s, \gamma_s\sigma_{\text{init}} + \beta\sigma_s)$ is a nonsingular point, it has a nonsingular neighborhood by continuity.

As $d_1 \to 0$, we have $\sigma_1 \to \sigma_s$, and then $(\sigma_1, (1+d_1)(\gamma_s\sigma_{\text{init}} + \beta\sigma_1)) \to (\sigma_s, \gamma_s\sigma_{\text{init}} + \beta\sigma_s)$. As $d_2 \to 0$, we have $\sigma_2 \to \sigma_s$, and then $(\sigma_2, (1+d_2)(\gamma_s\sigma_{\text{init}} + \beta\sigma_2)) \to (\sigma_s, \gamma_s\sigma_{\text{init}} + \beta\sigma_s)$. Thus, there exist sufficiently small $|d_1| > 0$ and $|d_2| > 0$ such that the segment in step (ii) lies in the nonsingular neighborhood of $(\sigma_s, \gamma_s\sigma_{\text{init}} + \beta\sigma_s)$.

Therefore, it takes one segment and two fiber jumps to circumvent a singular point on $\Gamma(\sigma_{\text{init}})$ while sticking to $\Gamma(\sigma_{\text{init}})$ before and after the singular point. Since there are finitely many singular points on $\Gamma(\sigma_{\text{init}})$, we obtain a path consisting of finitely many segments satisfying $\tilde{\Gamma} \cap S \subseteq \{(\sigma^\star, \mathbf{0})\}$. □

Therefore, $\tilde{\Gamma} \cap S \subseteq \{(\sigma^\star, \mathbf{0})\}$ with finitely many segments, where $\tilde{\Gamma} \cap S = \varnothing$ if $J(\sigma^\star)$ is nonsingular, and $\tilde{\Gamma} \cap S = \{(\sigma^\star, \mathbf{0})\}$ if $J(\sigma^\star)$ is singular. In other words, except possibly at the solution, singular points can always be bypassed through finitely many segments, so singularities do not obstruct global convergence.

In particular, the path in Proposition 9 follows the analytic curve $\Gamma(\sigma_{\text{init}})$ except for circumventing finitely many singular points on it. Note that $\mathbf{1}^\top\mu$ is not monotone along $\Gamma(\sigma_{\text{init}})$, since the direction in which it varies may change after crossing a singular point. At each $(\sigma, \mu) \in \Gamma(\sigma_{\text{init}})$, the direction of variation of $\mathbf{1}^\top\mu$ along $\Gamma(\sigma_{\text{init}})$ is determined by $\text{sign}(\det(J(\sigma) + (\mathbf{1}^\top\mu)I))$. That is, to follow $\Gamma(\sigma_{\text{init}})$, $\mathbf{1}^\top\mu$ must decrease when $\text{sign}(\det(J_G)) = 1$ and must increase when $\text{sign}(\det(J_G)) = -1$.

Finally, Proposition 9 is an existence result. In fact, we can fully exploit the geometric freedom of the fixed-point bundle without necessarily following the path constructed in its proof. As long as the path leads to a solution with finitely many segments, the subsequent analysis remains valid.

5.3 A compact nonsingular neighborhood for path-following

In addition to $\tilde{\Gamma}$ avoiding S, we also need a neighborhood of $\tilde{\Gamma}$ that avoids S, because the corrector is a subiteration that converges onto the path $\tilde{\Gamma}$ within its neighborhood. We introduce the neighborhood $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$ in equation (5.3). Its purpose is to provide a compact nonsingular neighborhood on which path-following is well-behaved, allowing us to bound the relevant quantities.

$$\mathcal{N}_{\tilde{\Gamma}}(\zeta) = \text{Closure}\left(\left\{(\text{softmax}(\theta), \mu) \,|\, \|\theta - \ln\sigma_\mu\| \le d, (\sigma_\mu, \mu) \in \tilde{\Gamma}, \mathbf{1}^\top\mu > \zeta\right\}\right) \tag{5.3}$$

The set $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$ is the union, over all $(\sigma_\mu, \mu) \in \tilde{\Gamma}$ with $\mathbf{1}^\top\mu > \zeta$, of d-radius neighborhoods of σ_μ in the $\ln\sigma$-coordinate. Here $\zeta \ge 0$ is the prescribed precision, which determines whether path-following ends exactly at $(\sigma^\star, \mathbf{0})$ or slightly before it if the solution is singular. Taking the closure ensures compactness.

The $\ln\sigma$-coordinate in the definition of $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$ aligns with the parameterization $\sigma = \text{softmax}(\theta)$, yielding a neighborhood of σ_μ that remains well-behaved near the simplex boundary. Note that $\|\theta - \ln\sigma_\mu\| \le d$ is equivalent to $\mathrm{e}^{-d} - 1 \le (\sigma_\mu - \sigma)/\sigma \le \mathrm{e}^d - 1$. The $\ln\sigma$-coordinate strongly resembles the local norm $\|d\sigma\|_{\text{diag}(1/\sigma^2)} = d\sigma^\top \text{diag}(1/\sigma^2) d\sigma$ used in interior-point methods, where $\text{diag}(1/\sigma^2)$ is the Hessian of the log barrier $-\ln\sigma$. In interior-point methods, the neighborhood defined by the Hessian-induced local norm is called the Dikin ellipsoid [42].

We prove $\mathcal{N}_{\tilde{\Gamma}}(\zeta) \cap S = \varnothing$ using compactness arguments. In the subsequent developments, we repeatedly use that the image of a compact set under a continuous function is compact, and that compact sets are bounded.

Proposition 10. *Let $\zeta \neq 0$ if $J(\sigma^\star)$ is singular. Then for any $\zeta \geq 0$, there exists $d > 0$ such that $\mathcal{N}_{\tilde{\Gamma}}(\zeta) \cap S = \varnothing$.*

Proof. Denote $\tilde{\Gamma}' = \{(\sigma, \mu) \in \tilde{\Gamma} | \mathbf{1}^\top \mu \geq \zeta\}$. If $J(\sigma^\star)$ is nonsingular, $\zeta = 0$ and $\tilde{\Gamma}' = \tilde{\Gamma}$. If $J(\sigma^\star)$ is singular, $\zeta > 0$ and $\tilde{\Gamma}'$ is the truncation of $\tilde{\Gamma}$ at $\mathbf{1}^\top \mu = \zeta$. Thus $\tilde{\Gamma}' \cap S = \varnothing$; then, if $\tilde{\Gamma}'$ and S are compact, there is a positive gap between them.

Each analytic curve segment in $\tilde{\Gamma}'$ is compact as the image of a compact interval of γ under a continuous map. A finite union of compact sets is compact, so $\tilde{\Gamma}'$ is compact. The map $\sigma \mapsto (\sigma, \tilde{\mu}(\sigma) - \lambda_j(J(\sigma))\sigma)$ is continuous because $\tilde{\mu}(\sigma)$ is continuous and the eigenvalue mapping $\lambda_j(J(\sigma))$ is continuous. Hence S is compact as the image of the compact set Δ under a continuous function.

Since $\tilde{\Gamma}' \cap S = \varnothing$ and both sets are compact, there exists $d_0 > 0$ such that for any $(\sigma_\mu, \mu) \in \tilde{\Gamma}'$ and $(\sigma_s, \mu_s) \in S$,

$$\|(\sigma_\mu, \mu) - (\sigma_s, \mu_s)\| \geq d_0.$$

Let $\mathcal{N}'_{\tilde{\Gamma}}(\zeta)$ be the set that differs from $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$ only in that the closure operation is not taken. For any $(\sigma_n, \mu) \in \mathcal{N}'_{\tilde{\Gamma}}(\zeta)$, there exists $(\sigma_\mu, \mu) \in \tilde{\Gamma}'$ such that

$$\begin{aligned} &\|(\sigma_n, \mu) - (\sigma_\mu, \mu)\| = \|\mathrm{softmax}(\theta_n) - \mathrm{softmax}(\ln \sigma_\mu)\| \\ &\leq \left(\sup_{\sigma \in \Delta} \left\| \sigma \circ (I - \mathbf{1}\sigma^\top) \right\| \right) \|\theta_n - \ln \sigma_\mu\| \leq d/2, \end{aligned}$$

where $\sigma \circ (I - \mathbf{1}\sigma^\top)$ is the Jacobian of softmax, whose norm is bounded by $1/2$ on Δ.

For any $(\sigma_n, \mu) \in \mathcal{N}'_{\tilde{\Gamma}}(\zeta)$ and $(\sigma_s, \mu_s) \in S$,

$$\|(\sigma_n, \mu) - (\sigma_s, \mu_s)\| \geq \|(\sigma_\mu, \mu) - (\sigma_s, \mu_s)\| - \|(\sigma_n, \mu) - (\sigma_\mu, \mu)\| \geq d_0 - d/2.$$

Choose $d \in (0, 2d_0)$. Then $\mathcal{N}'_{\tilde{\Gamma}}(\zeta)$ and S have positive gap $d_0 - d/2 > 0$. If two sets have positive gap, their closures also have positive gap. Hence $\mathcal{N}_{\tilde{\Gamma}}(\zeta) \cap S = \varnothing$. □

The main purpose of constructing $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$ is to bound two quantities that will appear in the algorithm. Let $p_\Delta(\mathcal{N}_{\tilde{\Gamma}}(\zeta))$ denote the projection of $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$ onto the σ-space.

Lemma 11. *Let $\zeta \neq 0$ if $J(\sigma^\star)$ is singular. Then for any $\zeta \geq 0$, the following quantities are bounded on $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$.*

(i) *The smallest singular value $s_{\min}(J(\sigma) + (\mathbf{1}^\top \mu)I)$ attains a lower bound on $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$.*

(ii) *$\|\mu/\sigma\|$ attains an upper bound on $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$.*

Proof. (i) Given that $\partial F/\partial \sigma$ is continuous on $p_\Delta(\mathcal{N}_{\tilde{\Gamma}}(\zeta))$, the matrix $J(\sigma) + (\mathbf{1}^\top \mu)I$ is continuous on $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$. Then $s_{\min}(J(\sigma) + (\mathbf{1}^\top \mu)I)$ is continuous on the compact set $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$, hence attains a minimum.

(ii) We have

$$\mu/\sigma = (\sigma_\mu/\sigma) \circ r_\mu = (\sigma_\mu/\sigma) \circ (F(\sigma_\mu) + v_\mu \mathbf{1}).$$

The ratio is bounded by $\mathrm{e}^{-d} \leq \sigma_\mu/\sigma \leq \mathrm{e}^d$ on $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$. The term $F(\sigma_\mu)$ is bounded by continuity on the compact set $p_\Delta(\mathcal{N}_{\tilde{\Gamma}}(\zeta))$. The term v_μ is bounded by continuity of the Brouwer function

$(\hat{\sigma}, r, v) = M(\sigma, \mu)$ on the compact set $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$. Indeed, $(\hat{\sigma}, r, v) = M(\sigma, \mu)$ is continuous on the part of $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$ where $\mu > 0$, by the same argument as in Proposition 3, and the limit as $\mu \to \mathbf{0}^+$ exists and yields $v \to -\min_i F_i(\sigma)$. □

In Lemma 11, $p_\Delta(\mathcal{N}_{\tilde{\Gamma}}(\zeta))$ denotes the projection of $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$ onto the σ-space. Since we only need bounds on $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$, continuous differentiability of F on $p_\Delta(\mathcal{N}_{\tilde{\Gamma}}(\zeta))$ suffices.

Denote the lower bound of $s_{\min}(J(\sigma) + (\mathbf{1}^\top \mu)I)$ on $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$ by $s_{\min}(J_G)$. If $J(\sigma^\star)$ is nonsingular, this bound is independent of ζ and uniform over the entire path-following process until reaching $(\sigma^\star, \mathbf{0})$. If $J(\sigma^\star)$ is singular, the bound is uniform only over the part where $\mathbf{1}^\top \mu \geq \zeta > 0$, and $s_{\min}(J_G) \to 0$ as $\zeta \to 0$. Thus, for a singular solution, the subsequent convergence result applies only to path-following that ends at $\mathbf{1}^\top \mu = \zeta$ with a prescribed $\zeta > 0$.

6 Path-following on the fixed-point bundle

6.1 The predictor-corrector framework

The fixed-point bundle provides a path connecting an arbitrary starting point to a solution of $\mathrm{VI}(F, \Delta)$. We now translate this geometric structure into a predictor-corrector algorithm, shown in Algorithm 1. The predictor follows the tangent direction of the bundle, the corrector restores feasibility with respect to the bundle equation, and singularities are avoided by moving along bundle fibers.

Algorithm 1 Path-following on the fixed-point bundle

Require: A smooth map $F : \Delta \to \mathbb{R}^n$ and its derivative $\partial F / \partial \sigma$, a designated starting point $\sigma_{\text{init}} \in \Delta$, and a desired precision $\epsilon > 0$

Set $(\sigma_0, \mu_0) = (\sigma_{\text{init}}, v_{\text{init}} \sigma_{\text{init}})$ for a sufficiently large v_{init}
repeat
  repeat
    Compute $(\hat{\sigma}_k, r_k, v_k) = M(\sigma_k, \mu_t)$ by solving bisection problem (3.4)
    Construct the matrix $J(\sigma_k)$ in equation (4.8)
    Solve regularized Newton equation (6.2) or (6.5) for $\overline{d\theta}_k$ with $\delta_H = \|G(\sigma_k, \mu_t)\|/n$
    Update $\sigma_{k+1} = \mathrm{softmax}(\ln \sigma_k + \overline{d\theta}_k)$
  until $\|G(\sigma_k, \mu_t)\| \leq \epsilon/n$
  Set $\hat{\mu}_t = \mu_t + \beta_t \sigma_k > 0$ such that $\mathbf{1}^\top \mu_t + \beta_t$ avoids the nonzero eigenvalues of $-J(\sigma_k)$
  Update $\mu_{t+1} = (\hat{\mu}_t - \eta_t u_t).\mathrm{clip}(\min = \epsilon/n)$ with a small η_t and a direction u_t
  Solve differential equation (4.10) for $\overline{d\theta}_t$ with $d\mu_t = \mu_{t+1} - \hat{\mu}_t$
  Set $\sigma_0 = \mathrm{softmax}(\ln \sigma_k + \overline{d\theta}_t)$
until $\mathbf{1}^\top \breve{\mu}(\sigma_k) \leq \epsilon$
return σ_k as an approximate solution of $\mathrm{VI}(F, \Delta)$

Predictor: The predictor first performs singularity avoidance. When a singular point is approached, the predictor moves along the fiber by adding $\beta\sigma$ to μ, jumping to a nonsingular point. It then moves in the tangent direction given by differential equation (4.10) with $d\mu = -\eta u$, where u is the direction vector parallel to $\hat{\mu}_{q-1} - \mu_q$ from the finitely many segments of the path $\tilde{\Gamma}$. In practice, we simply use $u_t = \pm \hat{\mu}_t$, following whichever analytic curve the fiber jump switches to, without strictly following the proof of Proposition 9 to return to the original analytic curve; this works well in our experiments. In the subsequent convergence analysis, however, we still use the free directions as defined by $\tilde{\Gamma}$.

Corrector: The corrector solves the fixed-point bundle equation $G(\sigma, \mu) = 0$ for a fixed μ. In interior-point methods, the perturbed KKT conditions and the barrier problem are derived from the same original problem. However, here they do not entirely align, so we derive two different but closely related correctors from them. In the following subsections, we introduce two regularized Newton correctors for practical use and prove the convergence of the overall algorithm with the standard Newton corrector.

Stopping criterion: The outer iteration terminates when $\mathbf{1}^\top \check{\mu}(\sigma)$ falls below a prescribed tolerance ϵ, which guarantees precision on the gap of VI(F, Δ). During the predictor steps, we enforce $\min_i \mu_i \geq \epsilon/n$, so that $\|G(\sigma, \mu)\| \leq \epsilon/n$ in the inner iteration suffices to imply $|G(\sigma, \mu)| \leq \mu$ in the corrector loop. When $\mathbf{1}^\top \mu$ falls below ϵ, Theorem 6 guarantees $\mathbf{1}^\top \check{\mu}(\sigma) \leq \epsilon$.

6.2 Corrector derived from perturbed KKT conditions

By perturbed KKT conditions, we mean the fixed-point bundle equation $G(\sigma, \mu) = 0$ that is equivalent to FPKKT (3.2). The Newton equation of $G(\sigma, \mu) = 0$ is derived as follows, where $G'_\theta(\sigma, \mu)d\theta$ was obtained in differential equation (4.10).

$$\begin{aligned} G'_\theta(\sigma, \mu)d\theta &= -\, G(\sigma, \mu) \\ \sigma \circ \Big(J(\sigma) + (\mathbf{1}^\top \mu)I\Big)\, \overline{d\theta} &= -\, \Big(I - \sigma \mathbf{1}^\top\Big)\, (\sigma \circ F(\sigma) - \mu) \end{aligned}$$

Multiplying both sides by $1/\sigma$, we obtain the Newton equation of $G(\sigma, \mu) = 0$ in equation (6.1). In addition, we denote the term on the right-hand side as $\tilde{G}(\sigma, \mu) := G(\sigma, \mu)/\sigma$.

$$\Big(J(\sigma) + (\mathbf{1}^\top \mu)I\Big)\, \overline{d\theta} = -\, \Big(I - \mathbf{1}\sigma^\top\Big)\, (F(\sigma) - \mu/\sigma) \tag{6.1}$$

The standard Newton iteration is simple for theoretical analysis, but its performance in practice is not ideal. The step length is affected by the smallest singular value $s_{\min}(J_G)$, which makes the iteration appear unstable if $s_{\min}(J_G)$ is too small. A classical remedy is the regularized Newton equation (6.2), derived from the standard Newton equation (6.1) by multiplying both sides by $J_G^\top$ and adding the regularization term $\delta_H I$ with $\delta_H > 0$. The coefficient then becomes positive semidefinite and is controlled by δ_H. Indeed, the smallest singular value of the coefficient is now $s^2_{\min}(J_G) + \delta_H$, and the step length can be controlled with a mild δ_H. However, convergence reduces from quadratic to linear with rate $\delta_H/(s^2_{\min}(J_G) + \delta_H)$, so regularization trades convergence speed for iteration stability.

$$\begin{aligned} \Big(J_G^\top J_G + \delta_H I\Big)\, d\theta &= -\, J_G^\top \tilde{G}(\sigma, \mu) \\ \overline{d\theta} &= \Big(I - \mathbf{1}\sigma^\top\Big)\, d\theta \end{aligned} \tag{6.2}$$

Regularized Newton equation (6.2) is approximately the regularized Gauss-Newton equation of the least squares problem (6.3) in terms of the Euclidean norm of $\tilde{G}(\sigma, \mu)$, which is equivalently the local norm of $G(\sigma, \mu)$. Since the local norm is induced by the Hessian matrix $\mathrm{diag}(1/\sigma^2)$ of the log barrier $-\ln \sigma$, problem (6.3) is actually a log barrier problem. Thus gradient descent with its inexact gradient $J_G^\top \tilde{G}(\sigma, \mu)$ also serves as a corrector for solving $G(\sigma, \mu) = 0$. Regularized Newton equation (6.2) mixes Newton and gradient steps: it reduces to the standard Newton step (6.1) if $\delta_H = 0$, and approximates the gradient step $-(1/\delta_H) J_G^\top \tilde{G}(\sigma, \mu)$ if δ_H is large.

$$\min_\sigma \frac{1}{2} \left\| \tilde{G}(\sigma, \mu) \right\| = \min_\sigma \frac{1}{2} \|G(\sigma, \mu)\|_{\mathrm{diag}(1/\sigma^2)} \tag{6.3}$$

6.3 Corrector derived from barrier problem

Since the fixed-point bundle is also a special central path of the MCP (3.5), we can also derive a gradient from the barrier MCP (3.5) for the corrector. First, we additionally require that (r, v) satisfy the Brouwer function $(\hat{\sigma}, r, v) = M(\sigma, \mu)$. Then σ becomes the only optimization variable of the barrier MCP (3.5), with (r, v) as an intermediate variable. The gradient is taken only with respect to $d\sigma$, and hence only with respect to $\overline{d\theta}$. In addition, this enables us to use $\mathbf{1}^\top(\sigma - \hat{\sigma}) = 0$ to eliminate dv from the differential.

The gradient of the barrier MCP (3.5) is derived as follows. In the third line, we use $\mu/r = \hat{\sigma}$, $dv\mathbf{1}^\top(\sigma - \hat{\sigma}) = 0$, and $d\sigma/\sigma = \overline{d\theta}$, and we also use $(I - \sigma\mathbf{1}^\top)(\sigma - \hat{\sigma}) = (\sigma - \hat{\sigma})$ and $\overline{d\theta} = (I - \mathbf{1}\sigma^\top)\overline{d\theta}$ there; we use them again in the fifth line to reverse it. In the fourth line, we isolate $J(\sigma)$ and use $k\sigma\mathbf{1}^\top(\sigma - \hat{\sigma}) = 0$ in it. In the last line, we use $r = F(\sigma) + v\mathbf{1}$ and $\hat{\sigma}^\top F(\sigma) + v = \mathbf{1}^\top\mu$.

$$
\begin{aligned}
& d\left(\sigma^\top r - \mu^\top \ln\sigma - \mu^\top \ln r\right) = (\sigma - \mu/r)^\top dr + (r - \mu/\sigma)^\top d\sigma \\
&= (\sigma - \mu/r)^\top \left(\frac{\partial F(\sigma)}{\partial \sigma} \circ \sigma \circ \frac{d\sigma}{\sigma} + dv\mathbf{1}\right) + (\sigma - \mu/r)^\top \left(r \circ \frac{d\sigma}{\sigma}\right) \\
&= (\sigma - \hat{\sigma})^\top \left(I - \mathbf{1}\sigma^\top\right) \left(\frac{\partial F(\sigma)}{\partial \sigma} \circ \sigma + r \circ I\right) \left(I - \mathbf{1}\sigma^\top\right) \overline{d\theta} \\
&= (\sigma - \hat{\sigma})^\top \left(J(\sigma) + \left(I - \mathbf{1}\sigma^\top\right) \left(\left(r - (I - \mathbf{1}\sigma^\top)F(\sigma)\right) \circ I\right) \left(I - \mathbf{1}\sigma^\top\right)\right) \overline{d\theta} \\
&= (\sigma - \hat{\sigma})^\top \left(J(\sigma) + \left(r - (I - \mathbf{1}\sigma^\top)F(\sigma)\right) \circ I\right) \overline{d\theta} \\
&= (\sigma - \hat{\sigma})^\top \left(J(\sigma) + \left(\mathbf{1}^\top\mu + (\sigma - \hat{\sigma})^\top F(\sigma)\right) I\right) \overline{d\theta}
\end{aligned}
$$

Ignoring the higher-order term $((\sigma - \hat{\sigma})^\top F(\sigma))(\sigma - \hat{\sigma})$, we obtain the inexact gradient of MCP (3.5) in equation (6.4).

$$
\nabla\left(\sigma^\top r - \mu^\top \ln\sigma - \mu^\top \ln r\right) = \left(J(\sigma) + (\mathbf{1}^\top\mu)I\right)^\top (\sigma - \hat{\sigma}) \tag{6.4}
$$

Next we derive a regularized Newton equation from the standard Newton equation (6.1), with the gradient $J_G^\top(\sigma - \hat{\sigma})$ on the right-hand side. The derivation is as follows. In the first line, denote $v_g = \sigma^\top F(\sigma) - \mathbf{1}^\top\mu$ such that $\sigma \circ F(\sigma) + v_g\sigma - \mu = G(\sigma, \mu)$. In the second line, denote $(\hat{\sigma}, r, v_b) = M(\sigma, \mu)$ such that $\hat{\sigma} \circ (F(\sigma) + v_b\mathbf{1}) = \mu$. In the third line, denote $v_{gb} = (v_g - v_b)/(\mathbf{1}^\top\mu + k)$ and use $J(\sigma)\mathbf{1} = k\mathbf{1}$.

$$
\begin{aligned}
\sigma \circ \left(J(\sigma) + (\mathbf{1}^\top\mu)I\right) \overline{d\theta} &= -\left(\sigma \circ F(\sigma) + v_g\sigma - \mu\right) \\
\sigma \circ \left(J(\sigma) + (\mathbf{1}^\top\mu)I\right) \overline{d\theta} + (v_g - v_b)\sigma &= -\left(\sigma \circ F(\sigma) + v_b\sigma - \mu\right) \\
\sigma \circ \left(J(\sigma) + (\mathbf{1}^\top\mu)I\right) \left(\overline{d\theta} + v_{gb}\mathbf{1}\right) &= -\left(\sigma \circ F(\sigma) + v_b\sigma - \mu\right) \\
\frac{\sigma}{r} \circ \left(J(\sigma) + (\mathbf{1}^\top\mu)I\right) \left(\overline{d\theta} + v_{gb}\mathbf{1}\right) &= -\left(\sigma - \hat{\sigma}\right)
\end{aligned}
$$

This yields the regularized Newton equation (6.5). Similar to regularized Newton equation (6.2), regularized Newton equation (6.5) mixes Newton and gradient steps: it reduces to the standard Newton step (6.1) if $\delta_H = 0$, and approximates the gradient step $-(1/\delta_H)J_G^\top(\sigma - \hat{\sigma})$ if δ_H is large.

$$
\begin{aligned}
\left(J_G^\top \circ \frac{\sigma}{r} \circ J_G + \delta_H I\right) d\theta &= -J_G^\top(\sigma - \hat{\sigma}) \\
\overline{d\theta} &= \left(I - \mathbf{1}\sigma^\top\right) d\theta
\end{aligned} \tag{6.5}
$$

Similar to regularized Newton equation (6.2), regularized Newton equation (6.5) also corresponds to a least squares problem (6.6) with a local norm. It can be verified that $J_G^\top(\sigma - \hat{\sigma})$ is the inexact gradient of $(\sigma - \mu/r)^\top(r - \mu/\sigma)$ by a derivation similar to that for the barrier MCP (3.5).

$$\min_{\sigma} \|\sigma \circ r - \mu\|_{\mathrm{diag}(1/(\sigma \circ r))} = \min_{\sigma} (\sigma - \mu/r)^\top (r - \mu/\sigma) \tag{6.6}$$

6.4 The global linear convergence proof

We analyze the convergence rate of the predictor-corrector path-following framework with the standard Newton corrector (6.1). First, we bound the function $\tilde{G}(\sigma, \mu)$ and its derivatives on the neighborhood $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$. The function $\tilde{G}(\sigma, \mu)$ is bounded on the neighborhood $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$, because $I - \mathbf{1}\sigma^\top$ and $F(\sigma)$ are bounded on the compact set $p_\Delta(\mathcal{N}_{\tilde{\Gamma}}(\zeta))$ by continuity, and μ/σ is bounded on $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$ by Lemma 11. The first derivative of $\tilde{G}(\sigma, \mu)$ is derived as follows.

$$\begin{aligned} d\tilde{G}(\sigma, \mu) =& d\frac{G(\sigma, \mu)}{\sigma} = \frac{1}{\sigma} \circ \frac{\partial G(\sigma, \mu)}{\partial \sigma} \circ \sigma \circ \overline{d\theta} - G(\sigma, \mu) \circ (1/\sigma^2) \circ \sigma \circ \overline{d\theta} \\ =& \left(J(\sigma) + (\mathbf{1}^\top \mu) I\right) \overline{d\theta} - \tilde{G}(\sigma, \mu) \circ \overline{d\theta} \end{aligned}$$

$\tilde{G}'(\sigma, \mu)$ is bounded on $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$, because $J(\sigma)$ is bounded on the compact set $p_\Delta(\mathcal{N}_{\tilde{\Gamma}}(\zeta))$ by continuity, and $\mathbf{1}^\top \mu$ and $\tilde{G}(\sigma, \mu)$ are bounded on $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$. The second derivative satisfies the following bound.

$$\begin{aligned} \left\|\tilde{G}''(\sigma, \mu)\right\| &= \left\|J'(\sigma) - \left(\tilde{G}(\sigma, \mu) \circ I\right)'\right\| \leq \|J'(\sigma)\| + \left\|\tilde{G}'(\sigma, \mu)\right\| \\ &\leq \|J'(\sigma)\| + \|J(\sigma)\| + \mathbf{1}^\top \mu + \left\|\tilde{G}(\sigma, \mu)\right\| \end{aligned}$$

$\tilde{G}'(\sigma, \mu)$ is bounded on $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$, because $J'(\sigma)$ and $J(\sigma)$ are bounded on the compact set $p_\Delta(\mathcal{N}_{\tilde{\Gamma}}(\zeta))$ by continuity, and $\mathbf{1}^\top \mu$ and $\tilde{G}(\sigma, \mu)$ are bounded on $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$.

To prove the overall convergence rate, we prove the convergence rate of the corrector subiteration and show that each predictor step does not leave the corrector convergence region. This involves proving a uniform bound on the contraction of $\tilde{G}(\sigma, \mu)$ under corrector subiterations and a uniform bound on the change in $\tilde{G}(\sigma, \mu)$ caused by predictor steps, which leads to the following two propositions.

Adapting the standard quadratic convergence proof of the Newton iteration [43], we obtain the following proposition, which states that $\tilde{G}(\sigma, \mu)$ achieves uniform quadratic convergence to 0 under the corrector.

Proposition 12. *Let $\theta_{k+1} = \theta_k + \overline{d\theta}_k$, where $\overline{d\theta}_k$ is given by Newton equation* (6.1)*. Then for any $(\sigma_k, \mu) \in \mathcal{N}_{\tilde{\Gamma}}(\zeta)$, there exists $C_1 > 0$ such that*

$$\left\|\tilde{G}(\sigma_{k+1}, \mu)\right\| \leq C_1 \left\|\tilde{G}(\sigma_k, \mu)\right\|^2. \tag{6.7}$$

Proof. We have the following derivation. In the third and fourth lines, we use the fundamental theorem of calculus to substitute the residuals, where $\sigma_u = \mathrm{softmax}(\theta_k + u(\theta_{k+1} - \theta_k))$ and $\sigma_w = \mathrm{softmax}(\theta_u + w(\theta_u - \theta_k))$. In the fifth line, we replace $\theta_{k+1} - \theta_k$ with $\overline{d\theta}_k$. The constant L_1 in the last line follows from the bound on the second derivative $\tilde{G}''(\sigma, \mu)$.

$$\left\|\tilde{G}(\sigma_{k+1}, \mu)\right\| = \left\|\tilde{G}(\sigma_{k+1}, \mu) - \tilde{G}(\sigma_k, \mu) - \left(J(\sigma_k) + (\mathbf{1}^\top \mu) I\right)(\theta_{k+1} - \theta_k)\right\|$$

$$
\begin{aligned}
&= \left\| \tilde{G}(\sigma_{k+1}, \mu) - \tilde{G}(\sigma_k, \mu) - \left(\tilde{G}'(\sigma_k, \mu) + \tilde{G}(\sigma_k, \mu) \circ I \right) (\theta_{k+1} - \theta_k) \right\| \\
&\leq \int_0^1 \left\| \tilde{G}'(\sigma_u, \mu) - \tilde{G}'(\sigma_k, \mu) \right\| \|\theta_{k+1} - \theta_k\| \, du + \left\| \tilde{G}(\sigma_k, \mu) \right\| \|\theta_{k+1} - \theta_k\| \\
&\leq \int_0^1 \left\| \tilde{G}''(\sigma_w, \mu) \right\| u \, \|\theta_{k+1} - \theta_k\|^2 \, du + \left\| \tilde{G}(\sigma_k, \mu) \right\| \|\theta_{k+1} - \theta_k\| \\
&\leq \left(\frac{1}{2} L_1 \left\| \left(J(\sigma_k) + (\mathbf{1}^\top \mu) I \right)^{-1} \right\| + 1 \right) \left\| \left(J(\sigma_k) + (\mathbf{1}^\top \mu) I \right)^{-1} \right\| \left\| \tilde{G}(\sigma_k, \mu) \right\|^2
\end{aligned}
$$

In the above, $\|(J(\sigma) + (\mathbf{1}^\top \mu) I)^{-1}\|$ is uniformly bounded by $1/s_{\min}(J_G)$ on $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$, and the constant L_1 is uniformly bounded by the bound on $\tilde{G}''(\sigma, \mu)$ on $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$. This proves the inequality. □

For the predictor step along the path $\tilde{\Gamma}$, μ is updated along linear segments such as $\mu = (1-\gamma)\hat{\mu}_{q-1} + \gamma \mu_q$. A natural update is $-\eta u_t$ with $\eta \in (0,1)$ and a direction vector u_t; for a point μ_t on some segment, we take u_t given by

$$
u_t := \begin{cases} \mu_t - \mu_q, & \text{if } \mu_q = 0 \\ \hat{\mu}_{q-1} - \mu_q, & \text{if } \mu_q \neq 0 \end{cases}. \tag{6.8}
$$

This setting has two important properties: (1) finitely many iterations suffice to reach the last segment of $\tilde{\Gamma}$, because u_t has constant length except on the last segment where $\mu_q = 0$; (2) u_t/μ_t is bounded along $\tilde{\Gamma}$, because $u_t = \mu_t$ when μ_t is on the last segment, where $\mu_t \to 0$, and $u_t = \hat{\mu}_{q-1} - \mu_q$ has only finitely many choices on the segments where μ_t does not approach 0.

Then we have the following proposition, which states that the change in $\tilde{G}(\sigma, \mu)$ caused by the predictor is uniformly bounded with the step size η as coefficient.

Proposition 13. *Let $\mu_{t+1} = \mu_t - \eta u_t$ and $\theta_{t+1} = \theta_t + \overline{d\theta}_t$, where $\eta \in (0,1)$, and $\overline{d\theta}_t$ is given by differential equation (4.10) with $d\mu_t = -\eta u_t$. Then for any $(\sigma_t, \mu_t) \in \mathcal{N}_{\tilde{\Gamma}}(\zeta)$, there exists $C_2 > 0$ such that*

$$
\left\| \tilde{G}(\sigma_{t+1}, \mu_{t+1}) - \tilde{G}(\sigma_t, \mu_t) \right\| \leq C_2 \eta. \tag{6.9}
$$

Proof. We have the following derivation. In the second line, we use the fundamental theorem of calculus to substitute the residual, where $\sigma_u = \mathrm{softmax}(\theta_t + u(\theta_{t+1} - \theta_t))$. In the third line, we replace $\theta_{t+1} - \theta_t$ with $\overline{d\theta}_t$. The constant L_2 in the last line follows from the bound on the first derivative $\tilde{G}'(\sigma, \mu)$.

$$
\begin{aligned}
&\left\| \tilde{G}(\sigma_{t+1}, \mu_{t+1}) - \tilde{G}(\sigma_t, \mu_t) \right\| \leq \left\| \tilde{G}(\sigma_{t+1}, \mu_{t+1}) - \tilde{G}(\sigma_t, \mu_{t+1}) \right\| + \left\| \tilde{G}(\sigma_t, \mu_{t+1}) - \tilde{G}(\sigma_t, \mu_t) \right\| \\
&\leq \int_0^1 \left\| \tilde{G}'(\sigma_u, \mu_{t+1}) \right\| \|\theta_{t+1} - \theta_t\| \, du + \left\| \left(I - \mathbf{1}\sigma_t^\top \right) (d\mu_t / \sigma_t) \right\| \\
&\leq \left(L_2 \left\| \left(J(\sigma_t) + (\mathbf{1}^\top \mu_t) I \right)^{-1} \right\| + 1 \right) \left\| I - \mathbf{1}\sigma_t^\top \right\| \eta \, \|u_t/\mu_t\| \, \|\mu_t/\sigma_t\|
\end{aligned}
$$

In the above, $\|(J(\sigma) + (\mathbf{1}^\top \mu) I)^{-1}\|$ is uniformly bounded by $1/s_{\min}(J_G)$ on $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$, and the constant L_2 is uniformly bounded by the bound on $\tilde{G}'(\sigma, \mu)$ on $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$. The term $\|I - \mathbf{1}\sigma^\top\|$ is bounded on the compact set $p_\Delta(\mathcal{N}_{\tilde{\Gamma}}(\zeta))$ by continuity. The term $\|u_t/\mu_t\|$ is bounded by our setting of u_t, and $\|\mu/\sigma\|$ is bounded on $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$ by Lemma 11. This proves the inequality. □

By Proposition 12, $\tilde{G}(\sigma,\mu)$ achieves uniform quadratic convergence to 0 under the inner iteration. By Proposition 13, there exists a uniform $\eta \in (0,1)$ such that $\tilde{G}(\sigma,\mu)$ remains in the convergence region of the inner iteration after each predictor step. Under the outer iteration, μ achieves linear convergence to 0 along the last segment of the path $\tilde{\Gamma}$, where the remaining segments do not change the convergence rate since there are finitely many of them. Thus, we obtain the convergence rate of the overall predictor-corrector iteration.

Proposition 14. *Let the predictor-corrector iteration be given as in Propositions 12 and 13. For each t, let the Newton corrector perform sufficiently many iterations, and denote the iteration result by σ_t. Then, on the neighborhood $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$, there exist $C > 0$ and $\eta \in (0,1)$ such that*

$$\mathrm{gap}(\sigma_t) < C(1-\eta)^t$$

holds until $\mathrm{gap}(\sigma_t) \le \zeta$. In particular, it is sufficient that each Newton corrector step take $O(\log t)$ subiterations.

Proof. We first study the Newton corrector subiteration in Proposition 12. By the contraction inequality $\|\tilde{G}(\sigma_{k+1},\mu)\| \le (C_1\|\tilde{G}(\sigma_k,\mu)\|)\|\tilde{G}(\sigma_k,\mu)\|$, the convergence region is $\|\tilde{G}(\sigma_0,\mu)\| < 1/C_1$. By induction, the error bound is $C_1\|\tilde{G}(\sigma_k,\mu)\| \le (C_1\|\tilde{G}(\sigma_0,\mu)\|)^{2^k}$. Thus, the number of iterations needed to achieve ϵ_G is

$$O\left(\log_2 \frac{\ln(C_1\epsilon_G)}{\ln(C_1\|\tilde{G}(\sigma_0,\mu)\|)}\right) = O\left(\log\log\frac{1}{\epsilon_G}\right).$$

Then we study the predictor step in Proposition 13. By

$$\left\|\tilde{G}(\sigma_{t+1},\mu_{t+1})\right\| \le \left\|\tilde{G}(\sigma_t,\mu_t)\right\| + \left\|\tilde{G}(\sigma_{t+1},\mu_{t+1}) - \tilde{G}(\sigma_t,\mu_t)\right\| \le \epsilon_G + C_2\eta,$$

we have that, for predictor steps with $\eta < (1/C_1 - \epsilon_G)/C_2$, $\tilde{G}(\sigma,\mu)$ remains in the convergence region of the Newton corrector subiteration.

Then we study the iteration of μ_t. By the setting $u_t = \mu_t - \mu_q$ when $\mu_q = 0$ and $u_t = \hat{\mu}_{q-1} - \mu_q$ otherwise, finitely many iteration steps suffice to reach the last segment where $\mu_q = 0$. Let μ_T lie on the last segment; then $\mu_t = \mu_T(1-\eta)^{t-T}$ for $t \ge T$ by the predictor steps. Denote

$$C_i^- = \min_{0\le t\le T} \frac{\mu_{t,i}}{(1-\eta)^t},\ C_i^+ = \max_{0\le t\le T} \frac{\mu_{t,i}}{(1-\eta)^t},$$

then $C_i^-(1-\eta)^t \le \mu_{t,i} \le C_i^+(1-\eta)^t$ for every t. It follows that $\min_i \mu_{t,i} \ge (\min_i C_i^-)(1-\eta)^t$ and $\mathbf{1}^\top\mu_t \le (\sum_i C_i^+)(1-\eta)^t$.

Finally, by $\mathrm{gap}(\sigma) = \mathbf{1}^\top\mu - \min_i((G(\sigma,\mu)+\mu)/\sigma)_i$ in Theorem 6, if $\|G(\sigma_k,\mu_t)\| < \min_i \mu_{t,i}$, then $\mathrm{gap}(\sigma_k) < \mathbf{1}^\top\mu_t \le (\sum_i C_i^+)(1-\eta)^t$, where a sufficient condition for the former is $\|\tilde{G}(\sigma_k,\mu_t)\| < (\min_i C_i^-)(1-\eta)^t$.

Note that Propositions 12 and 13 are proved on the neighborhood $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$, where the path $\tilde{\Gamma}$ is truncated at $\mathbf{1}^\top\mu = \zeta$. This implies that $\mathbf{1}^\top\mu_t \ge \zeta$, so $\mathrm{gap}(\sigma_k) < \mathbf{1}^\top\mu_t \le (\sum_i C_i^+)(1-\eta)^t$ holds until $\mathrm{gap}(\sigma_t) < \zeta$.

Therefore, if for every outer iteration index t the inner iteration iterates σ_k to

$$\|\tilde{G}(\sigma_k,\mu_t)\| < \min\left(1/C_1, (\min_i C_i^-)(1-\eta)^t\right),$$

then, with the inner iteration result $\sigma_t := \sigma_k$, the outer iteration satisfies $\mathrm{gap}(\sigma_t) < (\sum_i C_i^+)(1-\eta)^t$. Then, substituting $\epsilon_G = \min(1/C_1, (\min_i C_i^-)(1-\eta)^t)$, the number of subiterations needed is $O(\log t)$ for sufficiently large t. □

In conclusion, as the predictor-corrector path-following algorithm with Newton corrector solves $\mathrm{VI}(F, \Delta)$, $\mathrm{gap}(\sigma)$ achieves global linear convergence up to the prescribed precision ζ on the neighborhood $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$, where $\zeta = 0$ when the target solution is nonsingular and $\zeta > 0$ when it is singular.

When the algorithm is applied to the general $\mathrm{VI}(H, K)$ with continuous mapping and polytope domain via $\mathrm{VI}(F_\epsilon, \Delta)$ after preprocessing, the gap function satisfies the following equation, where the bound on $\mathrm{gap}_H(V\sigma) - \mathrm{gap}_{F_\epsilon}(\sigma)$ is from equation (2.9) with $\epsilon_{\mathrm{set}} = 0$ and $\epsilon_{\mathrm{map}} = \epsilon$.

$$\mathrm{gap}_H(V\sigma) \leq \mathrm{gap}_{F_\epsilon}(\sigma) + \mathrm{gap}_H(V\sigma) - \mathrm{gap}_{F_\epsilon}(\sigma) \leq C(1-\eta)^t + 2\epsilon \tag{6.10}$$

Therefore, applying the algorithm to $\mathrm{VI}(H, K)$, the gap function also achieves global linear convergence up to a prescribed precision ζ. In particular, it can further be shown that if H is continuously differentiable at the target solution, we can have $\zeta = 0$.

7 Numerical experiments

We test the algorithm on randomly generated instances of $\mathrm{VI}(F, \Delta)$, with F represented by tanh-activated neural networks. We evaluate two neural network architectures: (i) architecture $[n, n, n]$ with dimension n ranging from 2 to 1024; (ii) architecture $[256, \ldots, 256]$ with depth ranging from 2 to 8. The parameters of the neural networks are generated by Xavier normal initialization, which prevents the tanh activations from being saturated and hence decreases the complexity of F.

The starting point σ_{init} is also randomly generated, and the required precision is $\epsilon = 10^{-5}$. For each data point, we run 200 instances, totaling 6400 instances. Table 1 reports the results for two variants of the regularized Newton corrector, equation (6.2) and equation (6.5).

Table 1: Statistics (mean / median) solving $\mathrm{VI}(F, \Delta)$

	Regularized Newton corrector (6.2)		Regularized Newton corrector (6.5)	
Dimension	Iteration count	Singularity avoidance times	Iteration count	Singularity avoidance times
$[2] \times 3$	280 / 291	0.1 / 0.0	294 / 309	0.1 / 0.0
$[4] \times 3$	353 / 327	1.2 / 0.0	360 / 339	1.3 / 0.0
$[8] \times 3$	386 / 334	2.8 / 0.0	377 / 338	2.0 / 0.0
$[16] \times 3$	436 / 341	4.5 / 0.0	410 / 332	4.1 / 0.0
$[32] \times 3$	554 / 359	10.7 / 0.0	511 / 335	10.8 / 0.0
$[64] \times 3$	804 / 521	24.8 / 8.0	723 / 501	24.8 / 8.5
$[128] \times 3$	1194 / 726	46.2 / 20.0	1009 / 656	43.6 / 20.5
$[256] \times 3$	1062 / 648	40.4 / 18.5	954 / 598	41.1 / 17.0
$[512] \times 3$	1062 / 634	40.3 / 17.5	927 / 579	39.8 / 17.5
$[1024] \times 3$	948 / 553	35.0 / 10.0	892 / 452	38.3 / 8.5
$[256] \times 2$	1650 / 981	72.9 / 34.5	1258 / 837	60.9 / 33.5
$[256] \times 3$	1062 / 648	40.4 / 18.5	954 / 598	41.1 / 17.0
$[256] \times 4$	902 / 571	31.1 / 12.0	870 / 512	35.5 / 11.5
$[256] \times 5$	823 / 543	26.7 / 9.0	717 / 474	25.4 / 8.5
$[256] \times 6$	685 / 442	19.1 / 5.0	605 / 406	18.2 / 4.5
$[256] \times 7$	757 / 424	23.8 / 4.5	621 / 396	19.7 / 3.5
$[256] \times 8$	717 / 422	21.0 / 3.5	584 / 376	16.8 / 3.0

The algorithm succeeds on every one of the 6400 instances, demonstrating the robustness and scalability of the framework. The variant based on the regularized Newton corrector (6.2) is slightly slower, because the term $\widetilde{G}(\sigma, \mu)$ on its right-hand side involves dividing by σ and may cause overflow when σ is near the boundary. For architecture $[n, n, n]$, the iteration count first increases and then decreases as the dimension n increases. This is probably because, although the complexity of F increases with the dimension n, the number of solutions also increases, so the path length to a solution does not necessarily increase. For architecture $[256] \times d$, the iteration count decreases as the neural network depth d increases. This stems from Xavier normal initialization. Specifically, for tanh-activated neural networks with Xavier normal initialization, the function complexity actually decreases as the depth increases.

8 Conclusion

This paper has introduced a path-following framework for finite-dimensional VIs with general continuous functions and compact convex domains. The framework directly handles smooth VIs on simplices after a preprocessing.

The central innovation is the fixed-point bundle, a fiber-bundle structure that geometrizes the solution set of the fixed-point KKT system. This system characterizes paths that solve the smooth VIs on simplices. The geometry separates motion along the base from motion along fibers in path-following, unifying starting point selection, path-following, and singularity avoidance in a systematic way.

The path-following algorithm requires no assumptions. It achieves global linear convergence to nonsingular solutions and retains global linear convergence up to any prescribed precision for singular solutions. Numerical experiments on 6400 randomly generated problems up to dimension 1024 confirm universal success.

The current framework uses an inner simplex approximation and analytic smoothing. Although theoretically controllable, tighter integration of these preprocessing steps with the path-following phase could further improve practical performance.

References

[1] Gaetano Fichera. *Problemi elastostatici con vincoli unilaterali: il problema di Signorini con ambigue condizioni al contorno*. Accademia nazionale dei Lincei, 1964.

[2] Guido Stampacchia. Formes bilineaires coercitives sur les ensembles convexes. *Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences*, 258(18):4413, 1964.

[3] Gue Myung Lee, Byung Soo Lee, Nguyen Dong Yen, et al. Vector variational inequality as a tool for studying vector optimization problems. *Nonlinear Analysis: Theory, Methods & Applications*, 34(5):745–765, 1998.

[4] Gesualdo Scutari, Daniel P Palomar, Francisco Facchinei, and Jong-Shi Pang. Convex optimization, game theory, and variational inequality theory. *IEEE Signal Processing Magazine*, 27(3):35–49, 2010.

[5] Anna Nagurney. *Network economics: A variational inequality approach*, volume 10. Springer Science & Business Media, 2013.

[6] Dimitri P Bertsekas and Eli M Gafni. Projection methods for variational inequalities with application to the traffic assignment problem. In *Nondifferential and variational techniques in optimization*, pages 139–159. Springer, 2009.

[7] Mircea Sofonea and Andaluzia Matei. *Variational inequalities with applications: a study of antiplane frictional contact problems*, volume 18. Springer Science & Business Media, 2009.

[8] Michel Chipot. *Variational inequalities and flow in porous media*, volume 52. Springer Science & Business Media, 2012.

[9] Gauthier Gidel, Hugo Berard, Gaëtan Vignoud, Pascal Vincent, and Simon Lacoste-Julien. A variational inequality perspective on generative adversarial networks. *arXiv preprint arXiv:1802.10551*, 2018.

[10] David Kinderlehrer and Guido Stampacchia. *An introduction to variational inequalities and their applications*. SIAM, 2000.

[11] Philip Hartman and Guido Stampacchia. On some nonlinear elliptic differential-functional equations. *Acta Mathematica*, 115:271–310, 1966.

[12] Galina M Korpelevich. The extragradient method for finding saddle points and other problems. *Matecon*, 12:747–756, 1976.

[13] Yair Censor, Aviv Gibali, and Simeon Reich. The subgradient extragradient method for solving variational inequalities in hilbert space. *Journal of Optimization Theory and Applications*, 148(2):318–335, 2011.

[14] R Tyrrell Rockafellar. Monotone operators and the proximal point algorithm. *SIAM journal on control and optimization*, 14(5):877–898, 1976.

[15] Regina S Burachik and Alfredo N Iusem. A generalized proximal point algorithm for the variational inequality problem in a hilbert space. *SIAM journal on Optimization*, 8(1):197–216, 1998.

[16] Torbjörn Larsson and Michael Patriksson. A class of gap functions for variational inequalities. *Mathematical Programming*, 64(1):53–79, 1994.

[17] Masao Fukushima. Merit functions for variational inequality and complementarity problems. In *Nonlinear Optimization and Applications*, pages 155–170. Springer, 1996.

[18] Yair Censor, Alfredo N Iusem, and Stavros A Zenios. An interior point method with bregman functions for the variational inequality problem with paramonotone operators. *Mathematical Programming*, 81(3):373–400, 1998.

[19] Yurii Nesterov and Arkadii Nemirovskii. *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.

[20] Xiaojun Chen and Yinyu Ye. On homotopy-smoothing methods for box-constrained variational inequalities. *SIAM Journal on Control and Optimization*, 37(2):589–616, 1999.

[21] Pierre-Louis Lions and Bertrand Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.

[22] George HG Chen and R Tyrrell Rockafellar. Convergence rates in forward–backward splitting. *SIAM Journal on Optimization*, 7(2):421–444, 1997.

[23] Bingsheng He, Li-Zhi Liao, Deren Han, and Hai Yang. A new inexact alternating directions method for monotone variational inequalities. *Mathematical Programming*, 92(1):103–118, 2002.

[24] Alexander Vladimirovich Gasnikov, Pavel Evgenievich Dvurechensky, Fedor Sergeevich Stonyakin, and Aleksandr Aleksandrovich Titov. An adaptive proximal method for variational inequalities. *Computational Mathematics and Mathematical Physics*, 59(5):836–841, 2019.

[25] Alina Ene, Huy L Nguyen, and Adrian Vladu. Adaptive gradient methods for constrained convex optimization and variational inequalities. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 7314–7321, 2021.

[26] Dirk A Lorenz and Thomas Pock. An inertial forward-backward algorithm for monotone inclusions. *Journal of Mathematical Imaging and Vision*, 51(2):311–325, 2015.

[27] Caihua Chen, Shiqian Ma, and Junfeng Yang. A general inertial proximal point algorithm for mixed variational inequality problem. *SIAM Journal on Optimization*, 25(4):2120–2142, 2015.

[28] Anatoli Juditsky, Arkadi Nemirovski, and Claire Tauvel. Solving variational inequalities with stochastic mirror-prox algorithm. *Stochastic Systems*, 1(1):17–58, 2011.

[29] Ahmet Alacaoglu and Yura Malitsky. Stochastic variance reduction for variational inequality methods. In *Conference on Learning Theory*, pages 778–816. PMLR, 2022.

[30] Masakazu Kojima, Shinji Mizuno, and Akiko Yoshise. *A primal-dual interior point algorithm for linear programming*, pages 29–47. Springer, 1989.

[31] Serge Lang. *Real and functional analysis*. Springer Science & Business Media, 2012.

[32] Walter Rudin. *Real and complex analysis*. McGraw-Hill Education, 1974.

[33] Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

[34] Rolf Schneider. *Convex bodies: the Brunn–Minkowski theory*, volume 151. Cambridge university press, 2013.

[35] Stephen Karamardian. The complementarity problem. *Mathematical programming*, 2(1):107–129, 1972.

[36] Francisco Facchinei and Jong-Shi Pang. *Finite-dimensional variational inequalities and complementarity problems*. Springer, 2003.

[37] Carlton E Lemke and Joseph T Howson, Jr. Equilibrium points of bimatrix games. *Journal of the Society for Industrial and Applied Mathematics*, 12(2):413–423, 1964.

[38] John C Harsanyi. Oddness of the number of equilibrium points: a new proof. *International Journal of Game Theory*, 2(1):235–250, 1973.

[39] Andrzej Granas and James Dugundji. *Fixed point theory*. Springer Science & Business Media, 2013.

[40] Arthur Sard. The measure of the critical values of differentiable maps. *Bulletin of the American Mathematical Society*, 48(12):883–890, 1942.

[41] Evgenii Mikhailovich Chirka. *Complex analytic sets*. Springer Science & Business Media, 2012.

[42] Iliya Iosiphovich Dikin. Iterative solution of problems of linear and quadratic programming. In *Soviet Math. Dokl.*, volume 8, pages 674–675, 1967.

[43] Jorge Nocedal and Stephen J Wright. *Numerical optimization*. Springer, 2006.