

# A path-following framework on fiber bundle for variational inequalities

Hongbo Sun\*  
shb20@tsinghua.org.cn

## Abstract

This paper proposes a path-following framework for finite-dimensional variational inequalities with arbitrary continuous functions and compact convex domains. The approach first approximately reduces a general variational inequality to a smooth variational inequality on a simplex. Its key innovation is to formulate the smooth variational inequality on a simplex on a fiber bundle called the fixed-point bundle. Exploiting this geometric structure, the framework systematically integrates starting point selection, path-following, and singularity avoidance. Without any monotonicity or similar assumptions, the algorithm guarantees global linear convergence to nonsingular solutions. For singular solutions, it maintains global linear reduction up to a prescribed precision, after which convergence becomes sublinear. Numerical experiments on 14400 randomly generated instances with dimensions up to 800 demonstrate robust performance. The algorithm converges in every tested instance, and iteration count grows only mildly with dimension.

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

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Approximate reduction to analytic VIs on a simplex</b>	<b>3</b>
<b>3</b>	<b>Three equivalent characterizations regarding the paths</b>	<b>6</b>
3.1	Linear programming form and path equation . . . . .	6
3.2	Brouwer fixed-point form and path existence . . . . .	7
3.3	Mixed complementarity problem form and central path . . . . .	8
<b>4</b>	<b>Fiber bundle as the solution space of FPKKT</b>	<b>9</b>
4.1	The fixed-point bundle . . . . .	9
4.2	Differential properties . . . . .	11
<b>5</b>	<b>Paths on the fixed-point bundle</b>	<b>13</b>
5.1	Analytic curves . . . . .	13
5.2	The path . . . . .	15
5.3	A compact nonsingular neighborhood for path-following . . . . .	15

---

\*Experiment codes and results are available at [https://github.com/shb20tsinghua/FiberBundle\\_VI](https://github.com/shb20tsinghua/FiberBundle_VI).

<b>6</b>	<b>Path-following on the fixed-point bundle</b>	<b>17</b>
6.1	The predictor-corrector framework . . . . .	17
6.2	Corrector derived from perturbed KKT conditions . . . . .	18
6.3	Corrector derived from barrier problem . . . . .	20
<b>7</b>	<b>Experiments</b>	<b>22</b>
<b>8</b>	<b>Conclusion</b>	<b>22</b>

## 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  $VI(H, K)$  consists of a nonempty compact convex set  $K \subseteq \mathbb{R}^n$  and a continuous map  $H : K \rightarrow \mathbb{R}^n$ . The goal is to find  $x^* \in K$  such that

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

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

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

where  $\text{gap}(x) \geq 0$  for all  $x \in K$ . Solutions are precisely those points  $x^* \in K$  with  $\text{gap}(x^*) = 0$ .

Existence of a solution follows from the Brouwer fixed-point theorem [11]. However, unconditional global convergence of numerical methods remains a challenge. Most existing algorithms require monotonicity or similar structural assumptions. Those that handle non-monotone problems typically need additional conditions or lack guarantees.

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 using 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 incorporate adaptive step sizes [24, 25], inertial terms [26, 27], or stochastic estimates [28, 28, 29] for large-scale problems. However, no existing method guarantees global convergence for all continuous  $H$  on compact convex  $K$  without additional assumptions such as monotonicity or pseudo-monotonicity.

Among these approaches, interior point methods are particularly relevant due to their path-following philosophy. They trace a central path parameterized by a barrier parameter  $\mu$ , with each point on the path satisfying the perturbed KKT conditions [30]. Our approach shares this path-following spirit and is closely related to interior point methods.

The proposed approach begins by approximately reducing the general  $VI(H, K)$  to a smooth  $VI(F, \Delta)$  on a simplex  $\Delta$ . Any compact convex set is approximated from within by a simplex, and any continuous mapping on a simplex is smoothed to a real-analytic function via Dirichlet-kernel averaging. All subsequent developments apply to the resulting smooth problem  $VI(F, \Delta)$ .

We derive three equivalent characterizations of paths solving  $VI(F, \Delta)$  via linear programming, Brouwer fixed points, and mixed complementarity problems. This yields 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  $\text{VI}(F, \Delta)$  and are special central paths, meaning our approach is equivalently a specialized interior point method that follows these special central paths.

Instead of solving FPKKT directly, we study the geometric structure of its solution set, which naturally forms a fiber bundle over the simplex  $\Delta$  with one-dimensional fibers in  $\mu$ . Two sections of this bundle recover the gap function of  $\text{VI}(F, \Delta)$  and govern the differential properties of the bundle. Analytic curves on the bundle connect a chosen starting point to solutions. A parity argument implies an odd number of nonsingular solutions.

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  $\mu$  along fibers. The starting point of the path lies on the fiber over a designated initial  $\sigma_{\text{init}} \in \Delta$ . Singular points on each fiber correspond to real eigenvalues of matrix function  $J(\sigma)$ . The bundle structure provides fiber jumps as a simple mechanism to bypass singular points while preserving the path structure.

The fixed-point bundle unifies starting point selection, path-following, and singularity avoidance in a systematic way, leading to a path  $\tilde{\Gamma}$  on the bundle. We then construct a compact nonsingular neighborhood  $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$  of the path where the path-following algorithm is well-behaved, where  $\zeta$  controls the prescribed precision at which the 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 achieves global linear convergence to the point marked by prescribed precision  $\zeta$ . For nonsingular solutions,  $\zeta = 0$ , and the algorithm is globally linearly convergent. For singular solutions,  $\zeta > 0$ , and the algorithm retains global linear reduction up to the prescribed precision  $\zeta > 0$ , after which the rate becomes sublinear.

We provide two different but deeply related regularized Newton correctors for practical use. Numerical experiments on 14400 randomly generated instances up to dimension 800 show that the algorithm succeeds in every case. Iteration counts grow only mildly with dimension.

The remainder of the paper is organized as follows. Section 2 details the approximation steps. 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, we use  $a \circ b$  for componentwise multiplication and  $a/b$  for componentwise division. Inequality  $a \geq b$  is also interpreted componentwise. The limit  $\mu \rightarrow \mathbf{0}^+$  is interpreted as  $\mu \rightarrow \mathbf{0}$  with  $\mu > 0$ . For matrix  $A$ , componentwise multiplication is interpreted as  $A \circ b = \text{Adiag}(b)$  and  $b \circ A = \text{diag}(b)A$ . For norms,  $\|a\|$  or  $\|A\|$  always means the  $L^2$  norm of vector  $a$  or matrix  $A$  unless otherwise specified.

## 2 Approximate reduction to analytic VIs on a simplex

The subsequent theory is derived for a variational inequality  $\text{VI}(F, \Delta)$  with a smooth mapping and a simplex domain. To place the framework in a sufficiently general setting, we first reduce a general variational inequality  $\text{VI}(H, K)$  with a continuous mapping on a compact convex domain to  $\text{VI}(F, \Delta)$ . We first approximate the compact convex set from within by a simplex. Then we smooth the continuous mapping on the simplex.

### Approximating compact convex sets via simplices

Any finite-dimensional compact convex set can be approximated by polytopes with arbitrary precision. For any  $\epsilon > 0$ , the open cover  $K \subseteq \bigcup_{x \in K} O(x, \epsilon)$  of  $\epsilon$ -balls admits a finite subcover  $K \subseteq \bigcup_{i=1}^k O(x_i, \epsilon)$  by compactness. The convex hull  $S = \text{conv}(\{x_i | 1 \leq i \leq k\})$  then approximates  $K$  with Hausdorff distance

$$d_H(K, S) = \max \left\{ \sup_{x \in K} d(x, S), \sup_{s \in S} d(s, K) \right\} \leq \max(\epsilon, 0) = \epsilon.$$

This is Minkowski's theorem on inner approximation of the convex body  $K$  by  $S \subseteq K$  [31].

Let  $X = (x_1, \dots, x_k)$  and represent any  $x \in S$  as  $x = X\sigma$  with  $\sigma \in \Delta$ . The variational inequality  $\text{VI}(H, K)$  is transformed into finding  $\sigma^* \in \Delta$  such that for all  $\sigma \in \Delta$ ,

$$\langle H(x^*), x - x^* \rangle = H(X\sigma^*)^\top (X\sigma - X\sigma^*) = \langle X^\top H(X\sigma^*), \sigma - \sigma^* \rangle \geq 0.$$

Thus  $\text{VI}(H, K)$  is approximated by a corresponding problem on a simplex with continuous mapping  $\hat{F}(\sigma) = X^\top H(X\sigma)$ .

### Approximating continuous functions via analytic functions on simplices

Next we smooth the continuous mapping  $\hat{F}(\sigma) = X^\top H(X\sigma)$  on the simplex. General continuous functions can be approximated by smooth functions using kernel averaging. Our domain is a simplex, so we use the Dirichlet distribution as the kernel. The Dirichlet distribution  $\text{Dir}(\alpha)$  on the simplex  $\Delta$  is given by

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

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

$$F_{\epsilon,i}(\sigma) = \mathbb{E}_{Y \sim \text{Dir}(\sigma/\epsilon + c)} \left[ \hat{F}_i(Y) \right] = \int_{\Delta} \hat{F}_i(y) \text{Dir}(y; \sigma/\epsilon + c) dy. \quad (2.2)$$

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

The smoothing in equation (2.2) is a mollification-type approach [32]. The Dirichlet distribution  $\text{Dir}(\sigma/\epsilon + c)$  acts as a mollifier. It is nonnegative on a compact domain, integrates to one, and its mass concentrates at  $\sigma$  as  $\epsilon \rightarrow 0$ . In mollification, the smoothed function converges uniformly to the original function, satisfying

$$\limsup_{\epsilon \rightarrow 0} \sup_{\sigma \in \Delta} \left\| F_\epsilon(\sigma) - \hat{F}(\sigma) \right\| = 0.$$

We use smoothness of different degrees in different parts of this paper. The computations in the resulting algorithm use the first derivative. The convergence of the algorithm is guaranteed by second-order continuous differentiability. However, we also derive an oddness theorem as a byproduct that relies on analyticity of the underlying mapping. We therefore prove the strongest smoothness result, namely that  $F_\epsilon(\sigma)$  is analytic.

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

*Proof.* A standard theorem on holomorphic parametric integrals in complex analysis can be used to prove this [33]. Specifically, for a complex function  $f(z, y)$  with  $z \in Z$  and  $y \in Y$ , if  $f(z, y)$  is holomorphic with respect to  $z$  and uniformly integrable with respect to  $y$ , then the parametric integral  $h(z) = \int_Y f(z, y) dy$  is holomorphic on  $Z$ . Uniform integrability means 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  $\hat{F}_i(y)\text{Dir}(y; z/\epsilon + c)$  with  $y \in \Delta$  and  $z \in U_\sigma$ . Here  $U_\sigma = \{z \in \mathbb{C}^n \mid \|z - \sigma\| \leq d\}$  is a complex neighborhood of  $\sigma \in \Delta$ . We require that  $\text{Re}(z/\epsilon + c) \geq m > 0$  for every  $z \in U_\sigma$ . Such a pair  $d$  and  $m$  exists for every  $\sigma \in \Delta$  because  $c > 0$ . We then show that  $\hat{F}_i(y)\text{Dir}(y; z/\epsilon + c)$  is holomorphic in  $z$  and uniformly integrable in  $y$ .

For holomorphy, the Beta function  $B(\alpha)$  is known to be holomorphic and nonzero on  $\{\alpha \in \mathbb{C}^n \mid \text{Re}(\alpha) > 0\}$ . Hence  $1/B(z/\epsilon + c)$  is holomorphic on  $U_\sigma$  since  $\text{Re}(z/\epsilon + c) \geq m > 0$ . 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  $\hat{F}_i(y)\text{Dir}(y; z/\epsilon + c)$  is holomorphic in  $z$  on  $U_\sigma$  for every  $y > 0$ . The set where  $\min_i y_i = 0$  has measure zero and does not affect the integral.

For uniform integrability,  $1/B(z/\epsilon + c)$  is continuous on the compact complex neighborhood  $U_\sigma$ , so it attains an upper bound  $|1/B(z/\epsilon + c)| < C_1$ . The function  $\hat{F}_i(y)$  is continuous on the compact simplex  $\Delta$ , so it attains an upper bound  $|\hat{F}_i(y)| < C_2$ . Then

$$\left| \hat{F}_i(y)\text{Dir}(y; z/\epsilon + c) \right| \leq C_1 C_2 \left| \prod_{i=1}^n y_i^{z_i/\epsilon + c_i - 1} \right| \leq C_1 C_2 \prod_{i=1}^n y_i^{\text{Re}(z_i/\epsilon + c_i) - 1} \leq C_1 C_2 \prod_{i=1}^n y_i^{m-1}.$$

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

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

The computations in the algorithm rely on the derivative of  $F_\epsilon(\sigma)$ . In subsequent derivations, we directly assume that  $F(\sigma)$  is smooth and use  $\partial F(\sigma)/\partial \sigma$  in the computation. Although we do not use  $F'_\epsilon(\sigma)$  directly in this paper, we provide its expression for completeness. For any distribution  $D(\alpha)$  and function  $f(y)$ , under some regularity conditions there is a standard result called the score function identity [34]

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

where  $S(Y; \alpha) = \partial \log D(Y; \alpha) / \partial \alpha$  is the score function of the distribution  $D(\alpha)$  and  $\text{Cov}_{Y \sim D(\alpha)}$  denotes covariance. Thus the derivative  $F'_\epsilon(\sigma)$  is derived as

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

### Approximation error

To conclude the approximate reduction, we derive the approximation error of  $\text{VI}(F_\epsilon, \Delta)$  to  $\text{VI}(H, K)$ . For any  $\sigma \in \Delta$ , the gap of  $\text{VI}(F_\epsilon, \Delta)$  approximates the gap of  $\text{VI}(H, K)$ .

Let  $x'^* = \arg \max_{x' \in K} \langle H(X\sigma), X\sigma - x' \rangle$ . Then there exists  $\sigma''^* \in \Delta$  such that  $\|X\sigma''^* - x'^*\| < \epsilon_{\text{set}}$  by the inner approximation. Hence

$$\sup_{x' \in K} \langle H(X\sigma), X\sigma - x' \rangle - \sup_{\sigma' \in \Delta} \langle F_\epsilon(\sigma), \sigma - \sigma' \rangle \leq \langle H(X\sigma), X\sigma - x'^* \rangle - \langle F_\epsilon(\sigma), \sigma - \sigma''^* \rangle$$

$$\begin{aligned}
 &= \langle H(X\sigma), X\sigma''^* - x'^* \rangle + \langle X^\top H(X\sigma) - F_\epsilon(\sigma), \sigma - \sigma''^* \rangle \\
 &\leq \|H(X\sigma)\| \|X\sigma''^* - x'^*\| + 2 \left\| X^\top H(X\sigma) - F_\epsilon(\sigma) \right\| \leq \|H(X\sigma)\| \epsilon_{\text{set}} + 2\epsilon_{\text{map}}.
 \end{aligned}$$

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

$$\begin{aligned}
 &\sup_{x' \in K} \langle H(X\sigma), X\sigma - x' \rangle - \sup_{\sigma' \in \Delta} \langle F_\epsilon(\sigma), \sigma - \sigma' \rangle \geq \langle H(X\sigma), X\sigma - X\sigma'^* \rangle - \langle F_\epsilon(\sigma), \sigma - \sigma'^* \rangle \\
 &= \langle X^\top H(X\sigma) - F_\epsilon(\sigma), \sigma - \sigma'^* \rangle \geq 2 \left\| X^\top H(X\sigma) - F_\epsilon(\sigma) \right\| \geq -2\epsilon_{\text{map}}.
 \end{aligned}$$

The coefficient  $\|H(X\sigma)\|$  is bounded by continuity on  $\Delta$ . 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 and the function approximation error.

### 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  $\Delta$ . The approximate reduction is necessary because  $\text{VI}(F, \Delta)$  yields a particular equation system 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 reveals a different aspect of the framework. The linear programming form derives FPKKT from  $\text{VI}(F, \Delta)$ . The Brouwer formulation establishes existence of solutions of FPKKT. 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  $\sigma^* \in \Delta$  such that  $\sigma^{*\top} F(\sigma^*) \leq \sigma'^\top F(\sigma^*)$  for all  $\sigma' \in \Delta$ . Equation (3.1) shows a parameterized linear programming problem (LP) that resembles this requirement. Here  $\hat{\sigma}$  is the optimization variable and  $\sigma$  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 Lagrangian 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  $\text{VI}(F, \Delta)$ . We refer to the simultaneous equation (3.2) as the fixed-point KKT system (FPKKT).

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

$$\hat{\sigma} = \sigma \quad (3.2b)$$

The  $\mu \geq 0$  case of equation (3.2a) are the perturbed KKT conditions of LP (3.1), where  $\mu$  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  $\text{VI}(F, \Delta)$ , as Theorem 2 shows. This theorem establishes the equivalence between solutions of FPKKT (3.2) and  $\text{VI}(F, \Delta)$  for  $\mu = 0$ . If solutions of FPKKT (3.2) exist at every  $\mu > 0$ , they form a path leading to solutions of  $\text{VI}(F, \Delta)$ , analogous to the central path.

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

- (i)  $\sigma^*$  is a solution of  $\text{VI}(F, \Delta)$ .
- (ii)  $\sigma^*$  is an optimizer of LP (3.1) parameterized by  $\sigma^*$ .
- (iii)  $\sigma^*$  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  $\langle F(\sigma^*), \sigma' - \sigma^* \rangle \geq 0$  for all  $\sigma' \in \Delta$ . (ii) is equivalent to  $\sigma^{*\top} F(\sigma^*) \leq \sigma'^\top F(\sigma^*)$  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. The 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  $\hat{\sigma} = \sigma = \sigma^*$  yields the equivalence.  $\square$

### 3.2 Brouwer fixed-point form and path existence

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

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

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

The following proposition holds.

**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  $(\sigma, \mu)$ , 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

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

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

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

Let  $(\sigma_k, \mu_k) \rightarrow (\sigma_0, \mu_0)$  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  $(\sigma, \mu, v)$ , 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  $(\sigma_0, \mu_0)$ . Since every convergent subsequence of  $\{v_k\}$  converges to this same unique limit  $v_0$ , the entire sequence  $\{v_k\}$  converges to  $v_0$ . Hence  $v$  is continuous with respect to  $(\sigma, \mu)$  on  $\Delta \times \{\mu | \mu > 0\}$ . Continuity of  $r = F(\sigma) + v \mathbf{1}$  and  $\hat{\sigma} = \mu/r$  then yields continuity of  $(\hat{\sigma}, r, v) = M(\sigma, \mu)$ .  $\square$

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  $\mu \rightarrow \mathbf{0}^+$ .

**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).

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

### 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). Existing research has established deep connections between complementarity problems, variational inequalities, and Brouwer fixed-point problems [35, 36, 37].

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

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. The  $\mu > 0$  case is the barrier problem of the MCP. The Lagrangian function of MCP (3.5) is

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

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

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

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

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

- (i)  $(\sigma, \mu)$  satisfies FPKKT (3.2) for some  $r, v$ .
- (ii)  $\sigma$  is a fixed point such that  $\sigma = \hat{\sigma}$  of the Brouwer function  $(\hat{\sigma}, r, v) = M(\sigma, \mu)$ .
- (iii)  $(\sigma, \mu)$  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 equation of the perturbed KKT conditions (3.2a) and  $\sigma = \hat{\sigma}$ . The equivalence follows.

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

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  $\text{VI}(F, \Delta)$  as  $\mu \rightarrow \mathbf{0}^+$ .

Note that  $(\sigma, \mathbf{0})$  satisfying FPKKT (3.2) is not necessarily a solution of  $\text{VI}(F, \Delta)$ , 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  $(\sigma^*, \mathbf{0})$  satisfying FPKKT (3.2) must be a limit point as  $\mu \rightarrow \mathbf{0}^+$  to be a solution of  $\text{VI}(F, \Delta)$ . 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  $\text{VI}(F, \Delta)$  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  $\text{VI}(F, \Delta)$ . 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  $\mu \rightarrow \mathbf{0}^+$  equivalently lead to solutions of  $\text{VI}(F, \Delta)$ .

### 4.1 The fixed-point bundle

The solution space of FPKKT (3.2) has a natural fiber-bundle structure. For each  $\sigma$ , the admissible  $\mu$  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 a fiber bundle in equation (4.1). The bundle is denoted as  $(E, \Delta, \alpha : E \rightarrow \Delta)$ , where  $E$  is the total space,  $\Delta$  is the base space,  $\alpha : E \rightarrow \Delta$  is the projection map, and  $\alpha^{-1}(\sigma) = \{\sigma\} \times B(\sigma)$  is the fiber over  $\sigma$ .

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

We refer to this as the fixed-point bundle. Conventionally, 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 movement of  $\sigma$  along the base space from movement of  $\mu$  along individual fibers. This separation 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  $(\sigma, \mu)$  only. Apply the projection matrix  $I - \sigma \mathbf{1}^\top$  to FPKKT (3.2). The resulting equation is  $G(\sigma, \mu) = 0$  given in 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 use  $(\sigma, \mu) \in E$  to denote points satisfying  $G(\sigma, \mu) = 0$ .

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

The projection matrix  $I - \sigma \mathbf{1}^\top$  appears throughout this paper. The following properties are frequently used in proofs and derivations. It projects along vector  $\sigma$  such that  $(I - \sigma \mathbf{1}^\top)\sigma = 0$ . It projects onto the subspace orthogonal to  $\mathbf{1}$  such that  $\mathbf{1}^\top(I - \sigma \mathbf{1}^\top) = 0$ . It is idempotent such that  $(I - \sigma \mathbf{1}^\top)(I - \sigma \mathbf{1}^\top) = I - \sigma \mathbf{1}^\top$ . The linear equation  $(I - \sigma \mathbf{1}^\top)x = b$  has a solution when  $\mathbf{1}^\top b = 0$ , and the general solution is  $x = b + k\sigma$  for  $k \in \mathbb{R}$ . Additionally,  $I - \mathbf{1}\sigma^\top = (I - \sigma \mathbf{1}^\top)^\top$  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  $\sigma$  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  $\text{VI}(F, \Delta)$ . As a section of the fiber bundle, it satisfies  $(\sigma, \check{\mu}(\sigma)) \in E$  for every  $\sigma \in \Delta$ .

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

Section  $\check{\mu}(\sigma)$  characterizes the gap of  $\text{VI}(F, \Delta)$  by Theorem 6. This theorem establishes the fundamental link between the fixed-point bundle  $E$  and the variational inequality  $\text{VI}(F, \Delta)$ .

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

$$\text{gap}(\sigma) = \mathbf{1}^\top \check{\mu}(\sigma) = \mathbf{1}^\top \mu - \min_i ((G(\sigma, \mu) + \mu)/\sigma)_i \quad (4.4)$$

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

- (i)  $\sigma^*$  is a solution of  $\text{VI}(F, \Delta)$ .
- (ii)  $\sigma^*$  is a zero point of  $\check{\mu}(\sigma)$ .
- (iii)  $(\sigma^*, \mathbf{0})$  is a limit point on  $E \cap (\Delta \times \{\mu | \mu > 0\})$ .

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

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

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  $\text{gap}(\sigma) \geq 0$  and  $\check{\mu}(\sigma) \geq 0$ ,  $\text{gap}(\sigma) = 0$  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  $\check{\mu}(\sigma) \rightarrow \check{\mu}(\sigma^*)$  as  $\sigma \rightarrow \sigma^*$ . For any  $\sigma_t \rightarrow \sigma^*$  with  $\sigma_t > 0$  and any  $k_t \rightarrow 0$  with  $k_t > 0$ , we have  $\mu_t = \check{\mu}(\sigma_t) + k_t\sigma_t > 0$  and  $\mu_t \rightarrow 0$ . Hence  $(\sigma^*, \mathbf{0})$  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  $(\sigma_t, \mu_t) \rightarrow (\sigma^*, \mathbf{0})$  with  $(\sigma_t, \mu_t) \in E$  and  $\mu_t > 0$ , then  $\mathbf{1}^\top \check{\mu}(\sigma_t) \rightarrow 0$ , hence  $\check{\mu}(\sigma_t) \rightarrow 0$  since  $\check{\mu}(\sigma) \geq 0$ . Continuity of  $\check{\mu}(\sigma)$  yields  $\check{\mu}(\sigma^*) = 0$ .  $\square$

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

Equation (4.4) also provides a measurement of the gap of  $\text{VI}(F, \Delta)$ . 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 proximity to the fixed-point bundle, and  $\mathbf{1}^\top \mu$  measures proximity to  $\mu = 0$ . When both are sufficiently precise, so is  $\mathbf{1}^\top \check{\mu}(\sigma)$ . This translates precision approaching the path and its endpoint into precision solving  $\text{VI}(F, \Delta)$ .

## 4.2 Differential properties

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

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

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

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

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

We introduce the differentiable section  $\tilde{\mu}(\sigma)$  in equation (4.7), which satisfies  $\mathbf{1}^\top \tilde{\mu}(\sigma) = 0$  and characterizes 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  $\sigma \in \Delta$ .

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

We use 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. In the third line we use  $\mathbf{1}^\top d\sigma = 0$ . In the last line we use  $d\sigma/\sigma = \overline{d\theta} = (I - \sigma \mathbf{1}^\top) \overline{d\theta}$ .

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

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

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

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

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

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

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 inherent zero eigenvalue with eigenvector  $\mathbf{1}$ . Proposition 7 shows how the eigenvalues are adjusted. A subspace  $U$  is an invariant subspace of matrix  $A$  if  $Au \in U$  for every  $u \in U$ . If  $A$  admits an invariant subspace decomposition, then 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  $\mathbb{R}^n = \text{span}(\mathbf{1}) \oplus \ker(\sigma^\top)$ . The eigenvalue corresponding to eigenvectors in  $\text{span}(\mathbf{1})$  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  $\text{span}(\mathbf{1})$  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  $\text{span}(\mathbf{1})$  and  $\ker(\sigma^\top)$  are complementary, this is an invariant subspace decomposition.

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

Proposition 7 shows that the inherent zero eigenvalue of  $J(\sigma)$  corresponding to eigenvector  $\mathbf{1}$  is removed, while the remaining eigenvalues are unchanged. This trick prevents  $J(\sigma)$  from being inherently 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  $\theta$  with respect to the reduction  $d\mu$  for following the fixed-point bundle.

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

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

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

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

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

$$S = \left\{ (\sigma, \tilde{\mu}(\sigma) - \lambda_j^\perp(J(\sigma))\sigma) \mid \lambda_j^\perp(J(\sigma)) \in \mathbb{R}, \sigma \in \Delta \right\}. \quad (4.11)$$

Indeed,  $S$  consists of all singular points of the fixed-point bundle. First, for every  $\sigma \in \Delta$  and  $(\sigma, \mu) \in S$ ,  $\mu$  lies on the fiber  $B(\sigma)$  over  $\sigma$  because  $\tilde{\mu}(\sigma)$  maps  $\sigma$  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  $\sigma \in \Delta$ ,  $(\sigma, \mu) \in S$  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  $\mathbf{1}^\top \mu = -\lambda_j^\perp(J(\sigma))$ , which are precisely the real eigenvalues of  $-J(\sigma)$  except  $-k$ .

## 5 Paths on the fixed-point bundle

Having shown that paths on the fixed-point bundle equivalently lead to solutions of  $\text{VI}(F, \Delta)$ , 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  $\text{VI}(F, \Delta)$ , and moreover this path is well-behaved enough to follow.

### 5.1 Analytic curves

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

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

with designated  $\sigma_{\text{init}} \in \Delta$ . This is a one-dimensional curve parameterized by  $\gamma$ , where every  $(\sigma, \mu) \in E$  on it satisfies  $G(\sigma, \mu) = 0$ . When  $F(\sigma)$  is real-analytic,  $G(\sigma, \mu)$  is real-analytic, so  $\Gamma(\sigma_{\text{init}})$  is called an analytic curve as 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  $\gamma \rightarrow +\infty$ . Thus  $\Gamma(\sigma_{\text{init}})$  constitutes a path starting from any designated  $\sigma_{\text{init}} \in \Delta$  as  $\gamma \rightarrow +\infty$  and ending at solutions of  $\text{VI}(F, \Delta)$  as  $\gamma \rightarrow 0^+$ .

The curve  $\Gamma(\sigma_{\text{init}})$  is smooth at nonsingular points, and path-following is well-behaved there. However, singular points may occur. 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  $(\sigma, \mu)$  where  $\mathbf{1}^\top \mu$  equals an eigenvalue of  $-J(\sigma)$ , we jump to  $(\sigma, \mu + \beta\sigma)$  with  $\beta$  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 to a new analytic curve  $\Gamma((\mu + \beta\sigma)/(\mathbf{1}^\top \mu + \beta))$  with a different starting point.

For path-following, continuous differentiability suffices. Nonetheless, analyticity of  $\Gamma(\sigma_{\text{init}})$  yields an oddness theorem for  $\text{VI}(F, \Delta)$  as a byproduct, analogous to the classical oddness theorem for Nash equilibria [38]. This theorem shows that when all solutions are nonsingular, they are connected in pairs together with a designated starting point, implying an odd number of solutions. Since singular solutions are nongeneric, this conclusion holds for almost all  $\text{VI}(F, \Delta)$  instances.

**Theorem 8** (Oddness theorem). *For  $\text{VI}(F, \Delta)$  with  $F(\sigma)$  real-analytic on  $\Delta$ , if  $J(\sigma)$  is nonsingular at every solution, then there are an odd number of solutions.*

Since the argument closely follows [38], with the essential difference that the polynomial equilibrium equation is replaced by the real-analytic fixed-point bundle equation  $G(\sigma, \mu) = 0$ , we only sketch the main points.

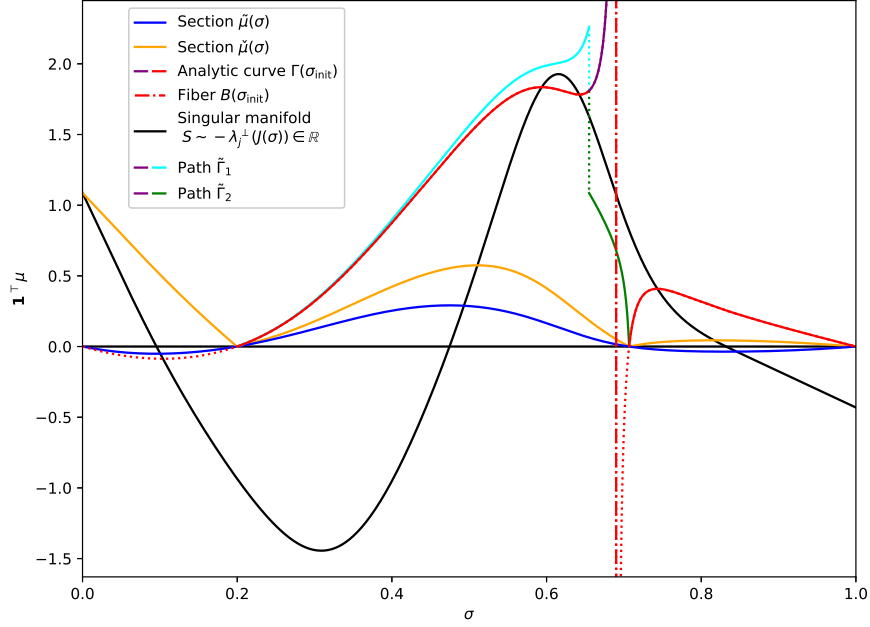


Figure 1: The fixed-point bundle  $E$  for a two-dimensional  $\text{VI}(F, \Delta)$ . The horizontal axis represents the simplex coordinate  $\sigma$ , and the vertical axis represents the fiber coordinate  $\mathbf{1}^\top \mu$ . The smooth section  $\tilde{\mu}(\sigma)$  induces the differential structure through  $J(\sigma)$ , while the nonnegative section  $\hat{\mu}(\sigma)$  equals the gap function of  $\text{VI}(F, \Delta)$ . Analytic curves  $\Gamma(\sigma_{\text{init}})$  connect a designated starting point to solutions as  $\mu \rightarrow \mathbf{0}^+$ . 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  $\text{VI}(F, \Delta)$ .

For any fixed  $\mu$ , the solution set of  $G(\sigma, \mu) = 0$  is finite. Indeed, compactness of  $\Delta$  implies that an infinite solution set would have an accumulation point, which would force the real-analytic function  $G(\sigma, \mu)$  to vanish identically. Consequently, the analytic curve  $\Gamma(\sigma_{\text{init}})$  consists of finitely many branches. By the Newton-Puiseux theorem [39], every local branch of the real-analytic equation  $G(\sigma, \gamma\sigma_{\text{init}}) = 0$  admits a convergent Puiseux expansion near every point on it, whether singular or nonsingular. Therefore each branch has a unique analytic continuation through every point.

It follows that the only boundary points of  $\Gamma(\sigma_{\text{init}})$  occur in the limits  $\gamma \rightarrow 0^+$  and  $\gamma \rightarrow +\infty$ . The former correspond to solutions of  $\text{VI}(F, \Delta)$ , while the latter is the unique starting point  $\sigma_{\text{init}}$ . If  $J(\sigma)$  is nonsingular at every solution, the implicit function theorem implies that exactly one branch terminates at each solution. Hence the branches of  $\Gamma(\sigma_{\text{init}})$  pair the solutions with the unique starting point  $\sigma_{\text{init}}$ . The parity argument yields an odd number of solutions.

Our path-following framework does not rely on analyticity of  $F$  or the oddness theorem, but the theorem provides additional evidence that the fixed-point bundle naturally generates global solution paths.

## 5.2 The path

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

$$\tilde{\Gamma} = \bigcup_{i=1}^q \{(\sigma, \gamma \mu_i) \in E \mid \mu_i = \gamma_{i-1} \mu_{i-1} + \beta_i \sigma_{i-1}, \gamma \in [\gamma_i, 1]\}, \quad (5.2)$$

$$\gamma_0 \mu_0 = v_{\text{init}} \sigma_{\text{init}}, \quad \mu_i > 0, \quad \gamma_q \mathbf{1}^\top \mu_q = 0$$

We would like  $\tilde{\Gamma} \cap S = \emptyset$  for singularity avoidance, where  $S$  is the singular manifold. However,  $\mathbf{1}^\top \mu = 0$  is unavoidable at the solution  $(\sigma^*, \mathbf{0})$ . If  $J(\sigma^*)$  is singular, then  $(\sigma^*, \mathbf{0})$  is an unavoidable singular point, that is,  $\tilde{\Gamma}$  must intersect  $S$ . Thus we instead require  $\tilde{\Gamma} \cap S \subseteq \{(\sigma^*, \mathbf{0})\}$  and only avoid 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  $\tilde{\Gamma} \cap S \subseteq \{(\sigma^*, \mathbf{0})\}$ .*

*Proof.* Suppose we adopt a singularity avoidance procedure that always uses  $\beta_i < 0$  to avoid the nonzero eigenvalues of  $-J(\sigma)$ . This procedure is admissible. Indeed, for  $(\sigma, \mu) \in E$  with  $\mu > 0$ , we have  $\mu = \sigma \circ r$  with  $r > 0$ . Then there always exists  $\beta_i < 0$  such that  $\mu_i = \gamma_{i-1} \mu_{i-1} + \beta_i \sigma_{i-1} = \sigma_{i-1} \circ (\gamma_{i-1} r_{i-1} + \beta_i \mathbf{1}) > 0$  as long as  $\gamma_{i-1} \mu_{i-1} > 0$ . Thus the procedure is admissible, and it generates a decreasing sequence of  $\gamma_i \mathbf{1}^\top \mu_i$ .

Suppose, for contradiction, that infinitely many segments in  $\tilde{\Gamma}$  are needed to achieve  $\tilde{\Gamma} \cap S \subseteq \{(\sigma^*, \mathbf{0})\}$ . The infinite sequence  $\gamma_i \mathbf{1}^\top \mu_i$  converges because it is bounded below by 0. The limit must be 0 because there always exists  $\beta_i < 0$  for singularity avoidance that ensures decreasing as long as  $\gamma_{i-1} \mu_{i-1} > 0$ . Hence infinitely many singularity avoidance events would occur in an arbitrarily small neighborhood of  $(\sigma^*, \mathbf{0})$ . This yields a contradiction because we only avoid nonzero eigenvalues of  $-J(\sigma)$ .  $\square$

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

## 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.

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

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  $\sigma_\mu$  in the  $\ln \sigma$ -coordinate. Here  $\zeta \geq 0$  is the prescribed precision marking whether path-following ends exactly at  $(\sigma^*, \mathbf{0})$  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  $\sigma_\mu$  that remains well-behaved near the simplex boundary. Note that  $\|\theta - \ln \sigma_\mu\| \leq d$  is equivalent to  $e^{-d} - 1 \leq (\sigma_\mu - \sigma)/\sigma \leq e^d - 1$ . 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 [40].

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

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

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

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

Since  $\tilde{\Gamma}' \cap S = \emptyset$  and both 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 by not taking the closure operation. For any  $(\sigma_n, \mu) \in \mathcal{N}'_{\tilde{\Gamma}}(\zeta)$ , there exists  $(\sigma_\mu, \mu) \in \tilde{\Gamma}'$  such that

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

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

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  $\mathcal{N}_{\tilde{\Gamma}}(\zeta) \cap S = \emptyset$ .  $\square$

The main purpose of constructing  $\mathcal{N}_{\tilde{\Gamma}}(\zeta)$  is to bound two quantities that will appear in the algorithm.

**Lemma 11.** *Let  $\zeta \neq 0$  if  $J(\sigma^*)$  is singular. Assume that  $F \in C^1(p_\Delta(\mathcal{N}_{\tilde{\Gamma}}(\zeta)))$ . 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  $p_\Delta(\mathcal{N}_{\bar{\Gamma}}(\zeta))$ , the matrix  $J(\sigma) + (\mathbf{1}^\top \mu)I$  is continuous on  $\mathcal{N}_{\bar{\Gamma}}(\zeta)$ . Then  $s_{\min}(J(\sigma) + (\mathbf{1}^\top \mu)I)$  is continuous on the compact  $\mathcal{N}_{\bar{\Gamma}}(\zeta)$ , 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  $e^{-d} \leq \sigma_\mu/\sigma \leq e^d$  on  $\mathcal{N}_{\bar{\Gamma}}(\zeta)$ . The term  $F(\sigma_\mu)$  is bounded by continuity on the compact  $p_\Delta(\mathcal{N}_{\bar{\Gamma}}(\zeta))$ . The term  $v_\mu$  is bounded by continuity of the Brouwer function  $(\hat{\sigma}, r, v) = M(\sigma, \mu)$  on the compact  $\mathcal{N}_{\bar{\Gamma}}(\zeta)$ . Indeed,  $(\hat{\sigma}, r, v) = M(\sigma, \mu)$  is continuous on the  $\mu > 0$  part of  $\mathcal{N}_{\bar{\Gamma}}(\zeta)$  by the same argument as in Proposition 3, and the limit as  $\mu \rightarrow \mathbf{0}^+$  exists and yields  $v \rightarrow -\min_i F_i(\sigma)$ .  $\square$

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

Denote the lower bound of  $s_{\min}(J(\sigma) + (\mathbf{1}^\top \mu)I)$  on  $\mathcal{N}_{\bar{\Gamma}}(\zeta)$  by  $s_{\min}(J_G)$ . If  $J(\sigma^*)$  is nonsingular, this bound is independent of  $\zeta$  and uniform over the entire path-following process until reaching  $(\sigma^*, \mathbf{0})$ . If  $J(\sigma^*)$  is singular, the bound is uniform only over the part where  $\mathbf{1}^\top \mu \geq \zeta > 0$ , and  $s_{\min}(J_G) \rightarrow 0$  as  $\zeta \rightarrow 0$ . 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 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. Singularities are avoided by moving along bundle fibers.

---

#### Algorithm 1 Path-following on the fixed-point bundle

---

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

- 1: Set  $(\sigma_0, \mu_0) = (\sigma_{\text{init}}, v_{\text{init}}\sigma_{\text{init}})$  for a sufficiently large  $v_{\text{init}}$
  - 2: **repeat**
  - 3:     **repeat**
  - 4:         Compute  $(\hat{\sigma}_k, r_k, v_k) = M(\sigma_k, \mu_t)$  by solving bisection problem (3.4)
  - 5:         Construct the matrix  $J(\sigma_k)$  in equation (4.8)
  - 6:         Solve regularized Newton equation (6.4) or (6.7) for  $\bar{d}\theta_k$  with  $\delta_H = \|G(\sigma_k, \mu_t)\|/n$
  - 7:         Update  $\sigma_{k+1} = \text{softmax}(\ln \sigma_k + \bar{d}\theta_k)$
  - 8:     **until**  $\|G(\sigma_k, \mu_t)\| \leq \epsilon/n$
  - 9:     Set  $\check{\mu}_t = \mu_t + \beta_t \sigma_k > 0$  such that  $\mathbf{1}^\top \mu_t + \beta_t$  avoids the nonzero eigenvalues of  $-J(\sigma_k)$
  - 10:     Update and truncate  $\mu_{t+1} = ((1 - \eta_t)\check{\mu}_t).\text{clip}(\min = \epsilon/n)$  with a sufficiently small  $\eta_t$
  - 11:     Solve differential equation (4.10) for  $\bar{d}\theta_t$  with  $d\mu_t = \mu_{t+1} - \check{\mu}_t$
  - 12:     Set  $\sigma_0 = \text{softmax}(\ln \sigma_k + \bar{d}\theta_t)$
  - 13: **until**  $\mathbf{1}^\top \check{\mu}(\sigma_k) \leq \epsilon$
  - 14: **return**  $\sigma_k$  as an approximate solution of 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  $\mu$ , jumping to another analytic

curve segment that bypasses the singular point. It then moves in the tangent direction given by differential equation (4.10) while reducing  $\mu$  geometrically.

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

**Stopping criterion:** The outer iteration terminates when  $\mathbf{1}^\top \tilde{\mu}(\sigma)$  falls below a prescribed tolerance  $\epsilon$ , which guarantees precision on the gap of  $\text{VI}(F, \Delta)$ . 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  $\epsilon$ , Theorem 6 guarantees  $\mathbf{1}^\top \tilde{\mu}(\sigma) \leq \epsilon$ .

We now introduce two different correctors. One is used to prove the convergence result of the overall algorithm. The other is an alternative for practical use.

## 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).

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

Multiplying both sides by  $1/\sigma$ , we obtain the Newton equation of  $G(\sigma, \mu) = 0$  in equation (6.1).

$$\left( J(\sigma) + (\mathbf{1}^\top \mu)I \right) \bar{d}\theta = - \left( I - \mathbf{1}\sigma^\top \right) (F(\sigma) - \mu/\sigma) \quad (6.1)$$

Denote  $\tilde{G}(\sigma, \mu) = G(\sigma, \mu)/\sigma$ , the term appearing in the right-hand side of Newton equation (6.1). Assuming  $F \in C^0(p_\Delta(\mathcal{N}_{\bar{\Gamma}}(\zeta)))$ ,  $\tilde{G}(\sigma, \mu)$  is bounded on the neighborhood  $\mathcal{N}_{\bar{\Gamma}}(\zeta)$ , because  $I - \mathbf{1}\sigma^\top$  and  $F(\sigma)$  are bounded on the compact  $p_\Delta(\mathcal{N}_{\bar{\Gamma}}(\zeta))$  by continuity, and  $\mu/\sigma$  is bounded on  $\mathcal{N}_{\bar{\Gamma}}(\zeta)$  by Lemma 11. The first derivative of  $\tilde{G}(\sigma, \mu)$  is derived as follows.

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

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

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

Assuming  $F \in C^2(p_\Delta(\mathcal{N}_{\bar{\Gamma}}(\zeta)))$ ,  $\tilde{G}'(\sigma, \mu)$  is bounded on  $\mathcal{N}_{\bar{\Gamma}}(\zeta)$ , because  $J'(\sigma)$  and  $J(\sigma)$  are bounded on the compact  $p_\Delta(\mathcal{N}_{\bar{\Gamma}}(\zeta))$  by continuity, and  $\mathbf{1}^\top \mu$  and  $\tilde{G}(\sigma, \mu)$  are bounded on  $\mathcal{N}_{\bar{\Gamma}}(\zeta)$ .

We analyze the convergence rate of the predictor-corrector path-following framework with the standard Newton corrector (6.1). To prove the 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 difference of  $\tilde{G}(\sigma, \mu)$  due to predictor steps.

Adapting the standard quadratic convergence proof of the Newton iteration [41], we have the following proposition stating 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). Assume that  $F \in C^2(p_\Delta(\mathcal{N}_{\bar{\Gamma}}(\zeta)))$ . Then for any  $(\sigma_k, \mu) \in \mathcal{N}_{\bar{\Gamma}}(\zeta)$ , there exists  $C_1 > 0$  such that*

$$\left\| \tilde{G}(\sigma_{k+1}, \mu) \right\| \leq C_1 \left\| \tilde{G}(\sigma_k, \mu) \right\|^2. \quad (6.2)$$

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

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

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

For the predictor step, we have the following proposition stating that the difference in  $\tilde{G}(\sigma, \mu)$  due to the predictor is uniformly bounded with step size  $\eta$  as coefficient.

**Proposition 13.** *Let  $\mu_{t+1} = (1 - \eta)\mu_t$  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  $d\mu_t = -\eta\mu_t$ . Assume that  $F \in C^1(p_\Delta(\mathcal{N}_{\bar{\Gamma}}(\zeta)))$ . Then for any  $(\sigma_t, \mu_t) \in \mathcal{N}_{\bar{\Gamma}}(\zeta)$ , there exists  $C_2 > 0$  such that*

$$\left\| \tilde{G}(\sigma_{t+1}, \mu_{t+1}) - \tilde{G}(\sigma_t, \mu_t) \right\| \leq C_2 \eta. \quad (6.3)$$

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

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

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

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

In conclusion, assuming  $F \in C^2(p_\Delta(\mathcal{N}_{\tilde{\Gamma}}(\zeta)))$ , for the predictor-corrector path-following algorithm with corrector steps in Proposition 12 and predictor steps in Proposition 13: (1)  $\tilde{G}(\sigma, \mu)$  achieves uniform quadratic convergence to 0 under corrector subiterations, (2) there exists  $\eta \in (0, 1)$  such that  $\tilde{G}(\sigma, \mu)$  remains in the convergence region of the corrector after each predictor step, and (3)  $\mathbf{1}^\top \mu$  achieves linear reduction to  $\zeta$  under the overall iteration.

Therefore, the predictor-corrector algorithm achieves global linear convergence along the path  $\tilde{\Gamma}$  to the ending point where  $\mathbf{1}^\top \mu = \zeta$ . For a nonsingular solution  $(\sigma^*, \mathbf{0})$ , path-following ends exactly at  $(\sigma^*, \mathbf{0})$  where  $\mathbf{1}^\top \mu = 0$ , and the algorithm achieves global linear convergence. For a singular solution  $(\sigma^*, \mathbf{0})$ , path-following ends before  $(\sigma^*, \mathbf{0})$  at  $\mathbf{1}^\top \mu = \zeta > 0$ , and the algorithm retains global linear reduction up to a prescribed precision  $\zeta > 0$ . In the singular case, the algorithm does not maintain global linear reduction to  $(\sigma^*, \mathbf{0})$ . Convergence reduces to sublinear as the required precision increases, because any convergent process is trivially at least sublinear.

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  $s_{\min}(J_G)$ , making the iteration appear unstable if  $s_{\min}(J_G)$  is too small. A classical remedy is the regularized Newton equation (6.4), derived from 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 becomes positive semidefinite and controlled by  $\delta_H$ . Indeed, the smallest singular value of the coefficient is now  $s_{\min}^2(J_G) + \delta_H$ , and the step length can be controlled with a mild  $\delta_H$ . However, convergence reduces from quadratic to linear with rate  $\delta_H/(s_{\min}^2(J_G) + \delta_H)$ . Thus regularization trades convergence speed for iteration stability.

$$\begin{aligned}
 \left( J_G^\top J_G + \delta_H I \right) d\theta &= -J_G^\top \tilde{G}(\sigma, \mu) \\
 \overline{d\theta} &= \left( I - \mathbf{1}\sigma^\top \right) d\theta
 \end{aligned} \tag{6.4}$$

Regularized Newton equation (6.4) is approximately the regularized Gauss-Newton equation of the least square problem (6.5) 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  $\text{diag}(1/\sigma^2)$  of the log barrier  $-\ln \sigma$ , problem (6.5) is actually a log barrier problem. Thus gradient descent with its inexact gradient  $J_G^\top \tilde{G}(\sigma, \mu)$  works as a corrector solving  $G(\sigma, \mu) = 0$  as well. Regularized Newton equation (6.4) 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  $\delta_H$  is large.

$$\min_{\sigma} \frac{1}{2} \left\| \tilde{G}(\sigma, \mu) \right\| = \min_{\sigma} \frac{1}{2} \|G(\sigma, \mu)\|_{\text{diag}(1/\sigma^2)} \tag{6.5}$$

### 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  $(r, v)$  is

subject to the Brouwer function  $(\hat{\sigma}, r, v) = M(\sigma, \mu)$ . Then  $\sigma$  becomes the only optimization variable of 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  $\overline{d\theta}$ . Additionally, this enables us to use  $\mathbf{1}^\top(\sigma - \hat{\sigma}) = 0$  to eliminate  $dv$  in the differential.

The gradient of 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  $d\sigma/\sigma = \overline{d\theta} = (I - \mathbf{1}\sigma^\top)\overline{d\theta}$ . In the fourth line, we use  $(I - \sigma\mathbf{1}^\top)(\sigma - \hat{\sigma}) = (\sigma - \hat{\sigma})$ , and also  $r = F(\sigma) + v\mathbf{1}$  and  $\hat{\sigma}^\top F(\sigma) + v = \mathbf{1}^\top\mu$ . In the fifth line, we use  $k\sigma\mathbf{1}^\top(\sigma - \hat{\sigma}) = 0$  in the first additive  $J(\sigma)$  and  $(I - \mathbf{1}\sigma^\top)(I - \mathbf{1}\sigma^\top) = (I - \mathbf{1}\sigma^\top)$  in the second additive. In the sixth line, we use  $(I - \sigma\mathbf{1}^\top)(\sigma - \hat{\sigma}) = (\sigma - \hat{\sigma})$ .

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

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 (6.6).

$$\nabla \left( \sigma^\top r - \mu^\top \ln \sigma - \mu^\top \ln r \right) = \left( J(\sigma) + (\mathbf{1}^\top\mu)I \right)^\top (\sigma - \hat{\sigma}) \quad (6.6)$$

Next we derive a regularized Newton equation from standard Newton equation (6.1) with 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  $J(\sigma)\mathbf{1} = k\mathbf{1}$ .

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

This yields the regularized Newton equation (6.7). Similar to regularized Newton equation (6.4), regularized Newton equation (6.7) 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  $\delta_H$  is large.

$$\begin{aligned}
 \left( J_G^\top \circ \frac{\sigma}{r} \circ J_G + \delta_H I \right) d\theta &= -J_G^\top (\sigma - \hat{\sigma}) \\
 \overline{d\theta} &= \left( I - \mathbf{1}\sigma^\top \right) d\theta
 \end{aligned} \quad (6.7)$$

Similar to regularized Newton equation (6.4), regularized Newton equation (6.7) also corresponds to a least square problem (6.8) with local norm. It can be verified that  $J_G^\top(\sigma - \hat{\sigma})$  is the inexact gradient of  $(\sigma - \mu/r)^\top(r - \mu/\sigma)$  via a derivation similar to that of barrier MCP (3.5).

$$\min_{\sigma} \|\sigma \circ r - \mu\|_{\text{diag}(1/(\sigma \circ r))} = \min_{\sigma} (\sigma - \mu/r)^\top (r - \mu/\sigma) \quad (6.8)$$

## 7 Experiments

We test the algorithm on randomly generated instances of  $\text{VI}(F, \Delta)$  with dimensions ranging from 3 to 800. The map  $F$  is a neural network with architecture  $[n, 50, n]$  and tanh activations, hence real analytic. Its derivative is computed by backpropagation. The starting point  $\sigma_{\text{init}}$  is also randomly generated, and the precision is  $\epsilon = 10^{-5}$ . For each dimension, we run 1000 instances, except dimensions 400 and 800, for which we use 100 instances each. Table 1 reports the mean and median iteration counts for two variants of the regularized Newton corrector, equation (6.4) and equation (6.7).

Table 1: Iteration numbers (mean / median) solving  $\text{VI}(F, \Delta)$

Dimension	Regularized Newton (6.4)	Regularized Newton (6.7)
[3, 50, 3]	331 / 329	347 / 342
[6, 50, 6]	362 / 339	379 / 356
[12, 50, 12]	372 / 347	389 / 368
[25, 50, 25]	390 / 360	402 / 372
[50, 50, 50]	437 / 370	440 / 380
[100, 50, 100]	438 / 378	431 / 382
[200, 50, 200]	539 / 475	539 / 487
[400, 50, 400]	655 / 528	640 / 546
[800, 50, 800]	645 / 550	667 / 582

The algorithm succeeds on every one of the 14400 instances. The iteration count increases only mildly with dimension, demonstrating the framework’s scalability. The variant based on regularized Newton corrector (6.7) is slightly slower but more stable, as its mean and median are closer.

Comparing the two regularized Newton equations (6.4) and (6.7), both reduce to the standard Newton step (6.1) when  $\delta_H = 0$ , but differ for  $\delta_H \neq 0$ . Equation (6.4) corresponds to a least-squares problem with local norm  $\text{diag}(1/\sigma^2)$ , which penalizes only  $\sigma \rightarrow 0$ . Equation (6.7) corresponds to local norm  $\text{diag}(1/(\sigma \circ r))$ , penalizing both  $\sigma \rightarrow 0$  and  $r \rightarrow 0$ . Its Hessian  $J_G^\top \circ (\sigma/r) \circ J_G$  vanishes in components as  $\sigma \rightarrow 0$ , which is controlled by  $\delta_H$ , and blows up in components as  $r \rightarrow 0$ , which induces cautious steps. Since  $\sigma \circ r = \mu$  along the path,  $r \rightarrow 0$  can cause  $\sigma$  to blow up under fixed  $\mu > 0$ , whereas  $r$  is subject to a function as  $\sigma \rightarrow 0$ . Thus regularized Newton (6.7) is more stable near the boundary but overall slower.

## 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 an approximate reduction.

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 monotonicity or other structural assumptions. It achieves global linear convergence to nonsingular solutions and retains global linear reduction for singular solutions up to any prescribed precision. Numerical experiments on 14400 randomly generated problems up to dimension 800 confirm universal success and only mild growth in iteration count with dimension.

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] Rolf Schneider. *Convex bodies: the Brunn–Minkowski theory*, volume 151. Cambridge university press, 2013.
- [32] Serge Lang. *Real and functional analysis*. Springer Science & Business Media, 2012.
- [33] Walter Rudin. *Real and complex analysis*. McGraw-Hill Education, 1974.
- [34] Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.
- [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] Evgenii Mikhailovich Chirka. *Complex analytic sets*. Springer Science & Business Media, 2012.
- [40] Iliya Iosiphovich Dikin. Iterative solution of problems of linear and quadratic programming. In *Soviet Math. Dokl.*, volume 8, pages 674–675, 1967.
- [41] Jorge Nocedal and Stephen J Wright. *Numerical optimization*. Springer, 2006.