A path-following framework on fiber bundle for variational inequalities

Hongbo Sun*
shb20@tsinghua.org.cn

Abstract

Variational inequality (VI) is a fundamental mathematical framework for many classical problems. We present a path-following framework for finite-dimensional VIs with arbitrary continuous functions and compact convex domains. The approach first approximately reduces a general VI to a smooth VI on simplex. Its key innovation is to formulate the smooth VI on simplex on a fiber bundle called the fixed-point bundle. Exploiting this geometric structure, we systematically integrate starting point selection, path-following, and singularity avoidance. Without any assumptions such as monotonicity, the algorithm guarantees global linear convergence to nonsingular solutions. For singular solutions, it retains global linear reduction up to a fixed precision, after which convergence becomes sublinear as the required precision increases. In numerical experiments on 14400 randomly generated VIs of up to 800 dimensions, the algorithm succeeds in every instance, and iteration number increases only mildly with the dimension.

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

Contents

*Experiment codes and results are available at `https://github.com/shb20tsinghua/FiberBundle_VI`.

1 Introduction

1.1 Variational inequality problem

The theory of variational inequalities was established in the 1960s through the pioneering work of Fichera [1] and Stampacchia [2]. Initially applied primarily to mechanics, the versatility of this framework was soon recognized. Over time, variational inequalities have become a fundamental mathematical tool for numerous classical problems, including optimization [3], game theory [4], economics [5], traffic [6], contact mechanics [7], fluid flow [8], machine learning [9], and more [10].

Definition 1 (Variational inequality problem)**.** Let $K \subseteq \mathbb{R}^n$ be a nonempty compact convex set, and let $H : K \to \mathbb{R}^n$ be a continuous map. The variational inequality problem, denoted $\mathrm{VI}(H, K)$ is to find a point $x^\star \in K$ such that

$$\left\langle H(x^\star), x' - x^\star \right\rangle \geq 0$$

holds for all $x' \in K$.

Variational inequalities are described by gap functions

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

where $\mathrm{gap}(x) \geq 0$, and $\mathrm{VI}(H, K)$ requires solutions such that $\mathrm{gap}(x^\star) = 0$.

This paper focuses on the variational inequalities on finite-dimensional compact convex sets. A distinctive feature of this class of VIs is that the existence of their solutions are guaranteed by the Brouwer fixed-point theorem [11].

1.2 Related work

Existing methods for solving variational inequality problems can be broadly classified into several categories based on their underlying mathematical principles.

Projection method [12, 13] and proximal point method [14, 15] leverage the reformulation of variational inequalities as Brouwer fixed-point problems. These approaches iteratively project updated points onto the feasible set K while reducing the residual $\|x - \Pi_K(x - \rho H(x))\|$, where $\rho > 0$ is a step size parameter. The projection method performs updates of the form $x_{t+1} = \Pi_K(x_t - \rho_t H(x_t))$, while the proximal point method performs updates of the form $x_{t+1} = \Pi_K(x_t - (1/\rho_t) H(x_{t+1}))$, where the later is equivalent to solving a regularized variational inequality subproblem of $H(x) - \rho_k(x - x_t)$ to obtain x_{t+1}.

Reformulation-based methods transform variational inequality problems into optimization problems using merit functions [16, 17], such as the gap function $G(x) = \sup_{y \in K} \langle H(x), x - y \rangle$ or dual gap function $D(x) = \inf_{y \in K} \langle H(y), y - x \rangle$. Standard optimization techniques, including gradient descent, Newton methods, and sequential quadratic programming, are then applied to solve the resulting optimization problem.

Interior point methods [18, 19] reformulate variational inequality problems as complementarity problems and solve the perturbed KKT conditions to follow a central path parameterized by a barrier parameter.

Homotopy methods [20] employ continuous deformations to transform complex variational inequalities into simpler ones with known solutions. A typical homotopy $E(x,t) = (1-t)G(x) + tH(x)$ gradually deforms from a simple function $G(x)$ to the target function $H(x)$ as t varies from 0 to 1. Interior point methods can be viewed as a special case of homotopy methods.

Operator splitting methods apply when $H(x) = A(x) + B(x)$ decomposes into simpler components, allowing alternating solutions to subproblems involving $A(x)$ and $B(x)$. Notable examples include Douglas-Rachford splitting [21], forward-backward splitting [22], and the alternating direction method of multipliers (ADMM) [23].

Most existing methods provide convergence guarantees under monotonicity assumptions or weaker variants like pseudo-monotonicity or quasi-monotonicity. While later progress extend these classical paradigms to non-monotone cases, the resulting variants typically require additional assumptions to guarantee convergence.

Modern algorithms for variational inequality problems often incorporate machine learning techniques, such as adaptive methods, inertial methods, and stochastic methods, particularly for non-monotone and large-scale problems. Adaptive methods dynamically adjust step sizes based on local operator properties to improve convergence without monotonicity assumptions, where examples include adaptive proximal methods [24] and adaptive extragradient [25]. Inertial methods introduce momentum-like terms to accelerate convergence, as seen in inertial forward-backward [26] and inertial proximal methods [27]. Stochastic methods address large-scale and data-driven variational inequality problems, with variants like stochastic extragradient [28], stochastic mirror-prox [29], and variance-reduced methods [30] demonstrating robust convergence in high-dimensional settings.

However, no existing method applies to fully general cases with unconditional convergence guarantees. This paper proposes a globally convergent path-following algorithm for finite-dimensional variational inequality problem $\mathrm{VI}(H, K)$ with general compact convex set K and general continuous function H, requiring no assumptions to guarantee convergence. The resulting convergence to nonsingular solutions is linear, and to singular solutions, the algorithm retain linear reduction before reaching a fixed precision.

Our approach follows a predictor-corrector path-following framework [31] resembling the interior point framework [32], but introduces several key innovations: our algorithm follows a special central path that guarantees to lead to solutions of variational inequalities globally, and we formalize the path as a fiber bundle so that starting point selection, path-following, singularity avoidance can be systematically handled on the geometric structure.

1.3 Technical overview and main results

In section 2, we approximately reduce general finite-dimensional $\mathrm{VI}(H, K)$ with continuous functions and compact convex domains to variational inequalities in the form of $\mathrm{VI}(F, \Delta)$, where F is real analytic and Δ is a simplex. General compact convex sets are approximated using inner approximation of convex bodies, and general continuous functions on simplices are approximated using kernel averaging by Dirichlet distribution. Then, all the following developments are based on $\mathrm{VI}(F, \Delta)$.

In section 4, we derive three equivalent characterizations of $\mathrm{VI}(F, \Delta)$. The linear programming form provides an equation called the unbiased KKT conditions that combines the perturbed KKT conditions and a fixed-point condition, which characterizes paths that leads to solutions of $\mathrm{VI}(F, \Delta)$. The Brouwer fixed-point form proves the existence theorem of the solutions of unbiased KKT conditions by the Brouwer fixed-point theorem. The mixed complementarity problem form characterizes the paths subject to the unbiased KKT conditions as special central paths. We use these three equivalent characterizations to study $\mathrm{VI}(F, \Delta)$ in the following developments.

In section 5, we formalize the solution space of the unbiased KKT conditions as a fiber bundle named as the fixed-point bundle. We introduce two sections of the fiber bundle, one characterizes the gap of $\mathrm{VI}(F, \Delta)$, the other characterizes the differential properties of the fixed-point bundle. Then, we analyze a few differential properties, including the differential equation and the manifold of singular points. The analytic curve on the fixed-point bundle proves the oddness theorem of solutions of $\mathrm{VI}(F, \Delta)$ by a parity arguement following from the Newton-Puiseux theorem. The geometric structure of the fixed-point bundle provides a systematic way for path-following and singularity avoidance. Our goal is to follow the analytic curve while avoiding the singular manifold along the fibers.

In section 6, we adopt the predictor-corrector framework as our path-following framework. Combining the analytic curve and singularity avoidance along the fibers, we obtain the path we follow on the fixed-point bundle. We show that there exists a compact neighborhood of the path containing only nonsingular points for the algorithm to work. Then, we introduce a standard Newton corrector solving the fixed-point bundle equation, and use this corrector to prove the convergence result of the overall predictor-corrector path-following framework. Alternatively, we also introduce a gradient corrector optimizing the barrier mixed complementarity problem. These two correctors leads to two different regularized Newton correctors for practical use.

In the convergence rate proof, we show that the algorithm achieves global linear convergence to nonsingular solutions of $\mathrm{VI}(F, \Delta)$. For singular solutions, the algorithm retains global linear reduction up to a fixed precision, but the convergence reduces to sublinear as required precision increases.

In section 7, we provide the pseudocode of the predictor-corrector path-following algorithm in Algorithm 1 and the graph of the fixed-point bundle in Figure (1). In experiment, the algorithm converges to a solution on all 14400 randomly generated $\mathrm{VI}(F, \Delta)$ instances of up to 800 dimensions.

2 Approximate reduction to smooth variational inequalities on simplices

2.1 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 ϵ-balls admits a finite subcover $K \subseteq \bigcup_{i=1}^{k} O(x_i, \epsilon)$ by compactness. The convex hull $S = \mathrm{conv}(\{x_i | 1 \geq i \geq k\})$ 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 constitutes the Minkowski's theorem on inner approximation of the convex body K by $S \subseteq K$ [33]. However, the approximation accuracy $1/\epsilon$ is worst-case exponential to the number k of points [34].

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

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

Thus, general variational inequalities can be approximated by corresponding problems on simplices.

2.2 Approximating continuous functions via analytic functions on simplices

General continuous functions can be approximated by smooth functions using kernel averaging. Our domain is simplex, we use Dirichlet distribution as the kernel to smooth continuous functions on simplices. The Dirichlet distribution $\mathrm{Dir}(\alpha)$ is defined on simplex Δ, such that

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

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

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

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

The smoothing in equation (2.2) can be viewed as a mollification-type approach [35], although the integral is not the classical convolution. However, the Dirichlet distribution $\mathrm{Dir}(\sigma/\epsilon + c)$ is indeed a mollifier, such that it is nonnegative on the compact simplex, its mass integrates to 1, and its mass concentrates to σ as $\epsilon \to 0$. In standard mollification, the smoothed function converges uniformly to the original one. It can be proved that we also have uniform convergence such that

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

We skip the proof for uniform convergence because it is standard and the result is not used in subsequent developments. But for smoothness, we prove that $F_\epsilon(\sigma)$ is real analytic.

Proposition 2. *For every $\epsilon > 0$, $F_\epsilon(\sigma)$ is real analytic on Δ.*

Proof. A standard theorem on holomorphic (complex analytic) parameteric integrals in complex analysis can be used to prove this [36]. 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 $f(z, y)$ is uniformly integrable with respect to y (i.e., there exists a integrable function $g(y)$ such that $|f(z, y)| < g(y)$ for every $z \in Z$), then the parametric integral $h(z) = \int_Y f(z, y)\, dy$ is holomorphic on Z.

In our case, we need to study the complex parameteric function $F_i(y)\mathrm{Dir}(y; z/\epsilon + c)$ with $y \in \Delta$ and $z \in U_\sigma$, where $U_\sigma = \{z \in \mathbb{C}^n \,|\, \|z - \sigma\| \leq d\}$ is defined as the complex neighborhood of $\sigma \in \Delta$. In particular, we require that $\mathrm{Re}(z/\epsilon + c) \geq m > 0$ for every $z \in U_\sigma$. Such a pair of d and m exists for every $\sigma \in \Delta$ because $c > 0$. Then, we can show that $F_i(y)\mathrm{Dir}(y; z/\epsilon + c)$ is holomorphic in z and uniformly integrable in y.

For holomorphy, the Beta function $B(\alpha)$ is known to be holomorphic and nonzero on $\{\alpha \in \mathbb{C}^n | \mathrm{Re}(\alpha) > 0\}$, then $1/B(z/\epsilon + c)$ is holomorphic on U_σ since $\mathrm{Re}(z/\epsilon + c) \geq m > 0$. And $\prod_{i=1}^n y_i^{z_i/\epsilon + c_i - 1}$ is holomorphic in z for every $y > 0$ since it is an exponential function. Thus, $F_i(y)\mathrm{Dir}(y; z/\epsilon + c)$ is holomorphic in z on U_σ for every $y > 0$, while 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_σ, so it attains an upper bound $|1/B(z/\epsilon + c)| < C_1$, and $F_i(y)$ is continuous on the compact simplex Δ, so it attains an upper bound $|F_i(y)| < C_2$. Then, we have

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

where $\prod_{i=1}^n y_i^{m-1}$ is known to be integrable on Δ as Dirichlet's multivariate Beta integral. Thus, $F_i(y)\mathrm{Dir}(y; z/\epsilon + c)$ is uniformly integrable in y on Δ.

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

Next, we derive the derivative of $F_\epsilon(\sigma)$. For any distribution $\mathrm{D}(\alpha)$ and function $f(y)$, under some regularity conditions, there are the following two identities, where $S(Y; \alpha) = \partial \log \mathrm{D}(Y; \alpha)/\partial\alpha$ is called the score function of distribution $\mathrm{D}(\alpha)$.

$$\frac{\partial}{\partial\alpha}\mathbb{E}_{Y\sim\mathrm{D}(\alpha)}\left[f(Y)\right] = \int_\Delta f(y)\frac{\partial}{\partial\alpha}\mathrm{D}(y;\alpha)\,dy = \int_\Delta f(y)\mathrm{D}(y;\alpha)S(y;\alpha)\,dy = \mathbb{E}_{Y\sim\mathrm{D}(\alpha)}\left[f(Y)S(Y;\alpha)\right]$$

$$\mathbb{E}_{Y\sim\mathrm{D}(\alpha)}\left[S(Y;\alpha)\right] = \int_\Delta S(y;\alpha)\mathrm{D}(y;\alpha)\,dy = \int_\Delta \frac{\partial}{\partial\alpha}\mathrm{D}(y;\alpha)\,dy = \frac{\partial}{\partial\alpha}1 = 0$$

Then, there is a standard result called the score function identity [37]

$$\frac{\partial}{\partial\alpha}\mathbb{E}_{Y\sim\mathrm{D}(\alpha)}\left[f(Y)\right] = \mathrm{Cov}_{Y\sim\mathrm{D}(\alpha)}\left[f(Y), S(Y;\alpha)\right], \tag{2.3}$$

where $\mathrm{Cov}_{Y\sim\mathrm{D}(\alpha)}$ is the covariance of the two random variables. Thus, the derivative $F'_\epsilon(\sigma)$ is derived as

$$\begin{aligned}
\frac{\partial F_{\epsilon,i}(\sigma)}{\partial\sigma_j} =& \frac{\partial\alpha_j}{\partial\sigma_j}\frac{\partial}{\partial\alpha_j}\mathbb{E}_{Y\sim\mathrm{Dir}(\alpha)}\left[F_i(Y)\right] \\
=& \frac{1}{\epsilon}\mathrm{Cov}_{Y\sim\mathrm{Dir}(\alpha)}\left[F_i(Y), \frac{\partial}{\partial\alpha_j}(-\log B(\alpha) + \sum_{k=1}^n(\alpha_k - 1)\log Y_k)\right] \\
=& \frac{1}{\epsilon}\mathrm{Cov}_{Y\sim\mathrm{Dir}(\alpha)}\left[F_i(Y), \log Y_j\right].
\end{aligned} \tag{2.4}$$

3 Notation table

Next, we study variational inequality $\mathrm{VI}(F, \Delta)$ with smooth function and simplex domain, and below are the notations used in the following sections.

Symbol	**Meaning**
$\mathrm{VI}(F,\Delta)$	Variational inequality problem on simplex Δ with smooth function $F:\Delta\to\mathbb{R}^n$, where $n=\dim(\Delta)$
$a\circ b$, $A\circ b$, $b\circ A$	Hadamard (element-wise) product between vector a, vector b and matrix A, where $A\circ b = A\mathrm{diag}(b)$ and $b\circ A=\mathrm{diag}(b)A$
$\mathbf{1}$, $\mathbf{0}$, I	All-ones vector, all-zeros vector, identity matrix
$\|a\|$, $\|A\|$	L^2 norm of vector a or matrix A, equals to the Euclidean norm of a or the largest singular value of A
σ	Vector variable on simplex Δ, s.t. $\sigma\geq 0$ and $\mathbf{1}^\top\sigma=1$
μ	Vector barrier parameter of the same dimension as σ
$(\hat\sigma,r,v)=M(\sigma,\mu)$	Brouwer function such that the Brouwer fixed-point theorem applies to $\sigma\mapsto\hat\sigma$ on simplex Δ
UKKT	Simultaneous equation of perturbed KKT conditions and fixed-point condition $\hat\sigma=\sigma$
$G(\sigma,\mu)=0$	Fixed-point bundle equation, equivalent to UKKT
E	Total space of the fixed-point bundle, consists of (σ,μ)
$B(\sigma)$	Fiber of the fixed-point bundle over σ, consists of μ
$\check\mu(\sigma)$, $\tilde\mu(\sigma)$	Sections of the fixed-point bundle, map from σ to μ
$\mathrm{gap}(\sigma)$	Gap function of $\mathrm{VI}(F,\Delta)$, $\mathrm{gap}(\sigma)=\sup_{\sigma'\in\Delta}\langle F(\sigma),\sigma-\sigma'\rangle=\mathbf{1}^\top\check\mu(\sigma)$
θ	$\sigma=\mathrm{softmax}(\theta)$, where $\mathrm{softmax}(\theta)=\mathrm{softmax}(\theta+k\mathbf{1})$ for any scalar k
$\overline{d\theta}$	$\overline{d\theta}=(I-\mathbf{1}\sigma^\top)d\theta$, removes a scaled $\mathbf{1}$ so that $\sigma\overline{d\theta}=0$
$J(\sigma)$	$\sigma\circ J(\sigma)$ is the Jacobian matrix of $\tilde\mu(\sigma)$
$J_G := J(\sigma)+(\mathbf{1}^\top\mu)I$	$\sigma\circ J_G$ is the Jacobian matrix of $G(\sigma,\mu)=0$
S	Manifold of singular points of the fixed-point bundle, consists of real eigenvalues of $-J(\sigma)$
$\Gamma(\sigma_{\mathrm{init}})$	Analytic curve on the fixed-point bundle with starting point σ_{init}
$\tilde\Gamma$	Path on the fixed-point bundle that avoids singular points, consists of finitely many segments of different $\Gamma(\sigma_{\mathrm{init}})$
$\mathcal{N}_{\tilde\Gamma}$	Compact nonsingular neighborhood of $\tilde\Gamma$, where the algorithm works
$\tilde G(\sigma,\mu)=G(\sigma,\mu)/\sigma$	Affine scaled $G(\sigma,\mu)$, its Euclidean norm $\|\tilde G(\sigma,\mu)\|=\|G(\sigma,\mu)\|_{\mathrm{diag}(1/\sigma^2)}$ is the local norm of $G(\sigma,\mu)$, induced by the Hessian of the log barrier $-\ln\sigma$

4 Three equivalent characterizations

4.1 By linear programming problem

We begin by introducing an optimization problem associated to $\mathrm{VI}(F,\Delta)$. Recall that $\mathrm{VI}(F,\Delta)$ equivalently requires a $\sigma^\star\in\Delta$ such that $\sigma^{\star\top}F(\sigma^\star)\leq\sigma'^\top F(\sigma^\star)$ holds for all $\sigma'\in\Delta$. Then, equation (4.1) shows a parameterized linear programming (LP) resembling this requiement, where $\hat\sigma$ is the optimization variable, σ is the parameter.

$$\begin{aligned}\min_{\hat\sigma}\quad & \hat\sigma^\top F(\sigma)\\ \text{s.t.}\quad & \mathbf{1}^\top\hat\sigma=1\\ & \hat\sigma\geq 0\end{aligned}\tag{4.1}$$

The Lagrangian function of LP (4.1) 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. Then, the $\mu = 0$ case of equation (4.2a) plus $r \geq 0$ is the KKT conditions of LP (4.1). We additionally need an unbiased condition $\hat{\sigma} = \sigma$ to fully capture the requiement of VI(F, Δ), and we refer to the simultaneous equation (4.2) as the unbiased KKT conditions (UKKT).

$$\begin{bmatrix} \hat{\sigma} \circ r - \mu \\ r - F(\sigma) - v\mathbf{1} \\ \mathbf{1}^\top \hat{\sigma} - 1 \end{bmatrix} = 0 \tag{4.2a}$$

$$\hat{\sigma} = \sigma \tag{4.2b}$$

LP (4.1) and UKKT (4.2) equivalently characterizes VI(F, Δ) as the following theorem shows.

Theorem 3. *For $\sigma^\star \in \Delta$, the following statements are equivalent.*

(i) *$\sigma^\star$ is a solution of* VI(F, Δ).

(ii) *$\sigma^\star$ is an optimizer of LP* (4.1) *parameterized by $\sigma^\star$.*

(iii) *$\sigma^\star$ satisfies the $\mu = 0$ case of UKKT* (4.2) *for some v and $r \geq 0$.*

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

(ii) $\Leftrightarrow$ (iii): For every parameter $\sigma \in \Delta$, since LP (4.1) is a linear programming problem, $\hat{\sigma}$ is an optimizer of LP (4.1) if and only if $\hat{\sigma}$ satisfies the KKT conditions, which is given by equation (4.2a) and $r \geq 0$. Thus, the equivalence is obtained by substituting in $\hat{\sigma} = \sigma = \sigma^\star$. $\square$

4.2 By Brouwer fixed-point problem

The $\mu \geq 0$ case of equation (4.2a) is the perturbed KKT conditions of LP (4.1) in interior point method. Unlike the KKT conditions, we do not need $r \geq 0$ because $\mu \geq 0$, $\hat{\sigma} \geq 0$, and $\hat{\sigma} \circ r = \mu$ already imply that. Similar to the interior point method, there is supposed to be a path leading to the desired solutions as $\mu \to 0$. However, UKKT (4.2) is no longer perturbed KKT conditions, and the existence of $\sigma \in \Delta$ satisfying UKKT (4.2) given $\mu > 0$ is nontrivial. Thus next, we study the $\mu > 0$ case of UKKT (4.2)

There is a correspondence from (σ, μ) to $(\hat{\sigma}, r, v)$ satisfying perturbed KKT conditions (4.2a). We refer to this correspondence as the Brouwer function

$$(\hat{\sigma}, r, v) = M(\sigma, \mu), \quad \text{s.t. perturbed KKT conditions (4.2a)}, (\sigma, \mu) \in \Delta \times \{\mu | \mu > 0\}. \tag{4.3}$$

Then, we have the following proposition.

Proposition 4. *$(\hat{\sigma}, r, v) = M(\sigma, \mu)$ is a continuous function on $\Delta \times \{\mu | \mu > 0\}$.*

Proof. By definition, a function is a correspondence that maps each preimage point to a single image point. First, we show that $(\hat{\sigma}, r, v) = M(\sigma, \mu)$ maps each (σ, μ) to a single $(\hat{\sigma}, r, v)$. For every (σ, μ), the corresponding v satisfies $q(v) = 0$, where $q(v)$ and $dq(v)/dv$ is given by

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

The derivative satisfies $dq(v)/dv < 0$. And we have limiting behavior

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

Then, $q(v)$ monotonically decreases from $+\infty$ to -1 on the interval $(-\min_i F_i(\sigma), +\infty)$. Consequently, there is a unique v corresponding to every (σ, μ), and thus a unique $(\hat{\sigma}, r, v)$.

Next, we use the implicit function theorem to show that $(\hat{\sigma}, r, v) = M(\sigma, \mu)$ is continuous. Consider $q(\sigma, \mu, v) := q(v)$ in both the variable (σ, μ) and v. $q(\sigma, \mu, v)$ is continuously differentiable in (σ, μ, v) near the point where $q(\sigma, \mu, v) = 0$, because $F(\sigma) + v\mathbf{1} > 0$. The Jacobian matrix of $q(\sigma, \mu, v)$ with respect to v is $dq(v)/dv < 0$, which is nonsingular. Thus, by the implicit function theorem, the map from (σ, μ) to v subject to $q(\sigma, \mu, v) = 0$ is continuously differentiable near the point satisfying $q(\sigma, \mu, v) = 0$. By the continuity of $r = F(\sigma) + v\mathbf{1}$ and $\hat{\sigma} = \mu/r$, the continuity of $(\hat{\sigma}, r, v) = M(\sigma, \mu)$ follows. □

The Brouwer fixed-point theorem states that any continuous function mapping a compact convex set to itself has fixed points. Then, there exists fixed points $\sigma = \hat{\sigma}$ of the Brouwer function $(\hat{\sigma}, r, v) = M(\sigma, \mu)$ for every given $\mu > 0$. The fixed points of Brouwer function $(\hat{\sigma}, r, v) = M(\sigma, \mu)$ are equivalently solutions of UKKT (4.2) since $(\hat{\sigma}, r, v) = M(\sigma, \mu)$ is defined by perturbed KKT conditions (4.2a). Therefore, we have Theorem 5, which guarantees that there exist paths subject to UKKT (4.2) as $\mu > 0$ tends to 0.

Theorem 5 (Existence theorem)**.** *For any $\mu > 0$, there exists $\sigma \in \Delta$ satisfying UKKT (4.2).*

Moreover, since $q(-\min_i F_i(\sigma) + \mathbf{1}^\top \mu) < 0$, the Brouwer function $(\hat{\sigma}, r, v) = M(\sigma, \mu)$ is computed via solving the bisection problem (4.4).

$$q(v) = \mathbf{1}^\top \frac{\mu}{F(\sigma) + v\mathbf{1}} - 1, \quad v \in [-\min_i F_i(\sigma), -\min_i F_i(\sigma) + \mathbf{1}^\top \mu] \tag{4.4}$$

4.3 By mixed complementarity problem

We have established the existence of paths subject to UKKT (4.2), where UKKT (4.2) is defined via the perturbed KKT conditions (4.2a) of LP (4.1) in the interior point method. Therefore, our path must align with the central path in interior point method in certain ways. Here, we show that the path defined by UKKT (4.2) are actually a special central path of the mixed complementarity problem (MCP) in equation (4.5). In fact, existing research has built deep connections between complementarity problems, variational inequalities, and Brouwer fixed-point problems [38, 39, 40].

$$\begin{aligned} \min_{(\sigma, r, v)} \quad & \sigma^\top r - \mu^\top \ln \sigma - \mu^\top \ln r \\ \text{s.t.} \quad & r = F(\sigma) + v\mathbf{1} \\ & \mathbf{1}^\top \sigma = 1 \\ & (\sigma, r) \geq 0 \end{aligned} \tag{4.5}$$

The $\mu = 0$ case of equation (4.5) is the MCP directly derived by letting the complementarity pair $\sigma \circ r$ in UKKT (4.2) be the objective function, and the $\mu > 0$ case is the barrier problem of the MCP in interior point method. The Lagrangian function of MCP (4.5) is

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

where $\bar{\lambda}$, $\tilde{\lambda}$, $\check{\sigma} \geq 0$, $\check{r} \geq 0$ are Lagrangian multipliers. Then, the perturbed KKT conditions of MCP (4.5) is

$$\begin{bmatrix} -\left(\partial F(\sigma)/\partial\sigma\right)^{\top}\bar{\lambda} + \tilde{\lambda}\mathbf{1} + r - \check{r} \\ \bar{\lambda} + \sigma - \check{\sigma} \\ -\bar{\lambda}^{\top}\mathbf{1} \\ r \circ \check{\sigma} - \mu \\ \sigma \circ \check{r} - \mu \\ r - F(\sigma) - v\mathbf{1} \\ \mathbf{1}^{\top}\sigma - 1 \end{bmatrix} = 0. \tag{4.6}$$

Then, we have Theorem 6 establishing the equivalence of the three characterizations.

Theorem 6. *For $(\sigma, \mu) \in \Delta \times \{\mu | \mu > 0\}$, the following statements are equivalent.*

(i) *(σ, μ) satisfies UKKT (4.2) for some r, v.*

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

(iii) *(σ, μ) satisfies $\check{\sigma} = \sigma$ and the perturbed KKT conditions of MCP (4.5) for some $r, v, \bar{\lambda}, \tilde{\lambda}, \check{\sigma}, \check{r}$.*

Proof. (i) $\Leftrightarrow$ (ii): Brouwer function $(\hat{\sigma}, r, v) = M(\sigma, \mu)$ is defined by perturbed KKT conditions (4.2a), and UKKT (4.2) is the simultaneous equation of perturbed KKT conditions (4.2a) and $\sigma = \hat{\sigma}$. Thus, the equivalence follows.

(i) $\Leftrightarrow$ (iii): Given (σ, μ, r, v) satisfying UKKT (4.2), $(\sigma, \mu, r, v, \bar{\lambda}, \tilde{\lambda}, \check{\sigma}, \check{r}) = (\sigma, \mu, r, v, \mathbf{0}, 0, \sigma, r)$ satisfies $\check{\sigma} = \sigma$ and perturbed KKT conditions (4.6). Given $(\sigma, \mu, r, v, \bar{\lambda}, \tilde{\lambda}, \check{\sigma}, \check{r})$ satisfying $\check{\sigma} = \sigma$ and perturbed KKT conditions (4.6), (σ, μ, r, v) satisfies UKKT (4.2). □

In the interior point method, central path is defined by perturbed KKT conditions. Therefore, the path given by UKKT (4.2) is the specific central path of MCP (4.5) on which $\check{\sigma} = \sigma$. Similar to the interior point method, these paths lead to solutions of VI(F, Δ) as μ tends to 0 from $\mu > 0$.

It is worth noting that $(\sigma, \mathbf{0})$ satisfying UKKT (4.2) is not always a solution of VI(F, Δ), because $r \geq 0$ is additionally required for the $\mu = 0$ case. For $\sigma \in \Delta$, $r \geq 0$ if and only if $\mu \geq 0$ since $\sigma \circ r = \mu$. Therefore, $(\sigma^{\star}, \mathbf{0})$ satisfying UKKT (4.2) must be a limit point as μ tends to 0 from $\mu > 0$ to constitute a solution of VI(F, Δ), which will be proved in the next section.

5 Fiber bundle as the solution space

5.1 The fixed-point bundle

We have established the existence of paths subject to UKKT (4.2) on $\Delta \times \{\mu | \mu > 0\}$, and they are special central paths in the interior point method, which are supposed to lead to solutions of VI(F, Δ). In this section, we study the solution space of UKKT (4.2) with $\mu \in \mathbb{R}^n$, and formally show that the paths as $\mu \to 0$ with $\mu > 0$ indeed lead to solutions of VI(F, Δ).

First, we eliminate r and equivalently transform UKKT (4.2) into

$$\sigma \circ F(\sigma) + v\sigma = \mu, \quad \sigma \in \Delta. \tag{5.1}$$

Then, equation (5.2) depicts the geometric structure of the solution space of equation (5.1).

$$\begin{aligned} &E = \bigcup_{\sigma \in \Delta} \{\sigma\} \times B(\sigma) \\ &B(\sigma) = \{\sigma \circ F(\sigma) + v\sigma | v \in \mathbb{R}\} \\ &\alpha\left((\sigma, \mu)\right) = \sigma \end{aligned} \tag{5.2}$$

For any $\sigma \in \Delta$, all the μ satisfying equation (5.1) forms a one-dimensional line spanned by $v \in \mathbb{R}$, denoted as $B(\sigma)$. The solution space consisting of (σ, μ) is the disjoint union of all these lines of μ over different $\sigma \in \Delta$, denoted as E. This geometric structure is called a fiber bundle, denoted as $(E, \Delta, \alpha : E \to \Delta)$, where E is the total space, Δ is the base space, $\alpha : E \to \Delta$ is the projection map, and $\alpha^{-1}(\sigma) = \{\sigma\} \times B(\sigma)$ is the fiber over σ. We refer to this fiber bundle as the fixed-point bundle, and conventionally, we also use the total space E to refer to the fiber bundle.

Using projection matrix $I - \sigma\mathbf{1}^\top$, we can further eliminate v in equation (5.1). Thus, we arrive at $G(\sigma, \mu) = 0$, where $G(\sigma, \mu)$ is defined in equation (5.3). $G(\sigma, \mu) = 0$ is the fixed-point bundle equation, and it is equivalent to UKKT (4.2). We also use $(\sigma, \mu) \in E$ to refer to points that satisfy $G(\sigma, \mu) = 0$.

$$G(\sigma, \mu) = \left(I - \sigma\mathbf{1}^\top\right)(\sigma \circ F(\sigma) - \mu), \quad (\sigma, \mu) \in \Delta \times \mathbb{R}^n \tag{5.3}$$

The matrix $I - \sigma\mathbf{1}^\top$ is a projection. It projects along vector σ such that $(I - \sigma\mathbf{1}^\top)\sigma = 0$, and it projects onto the subspace orthogonal to vector $\mathbf{1}$ 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 that has similar properties.

Next, we introduce two sections $\tilde{\mu}(\sigma)$ and $\check{\mu}(\sigma)$ of the fixed-point bundle as shown in equation (5.4). As sections of a fiber bundle, they map every σ to a point on the fiber $B(\sigma)$ over it. Thus, we have $(\sigma, \tilde{\mu}(\sigma)) \in E$ and $(\sigma, \check{\mu}(\sigma)) \in E$ for any $\sigma \in \Delta$. In addition, we have $\mathbf{1}^\top\tilde{\mu}(\sigma) = 0$ and $\check{\mu}(\sigma) \geq 0$.

$$\begin{aligned} \tilde{\mu}(\sigma) &= \sigma \circ \left(F(\sigma) - \left(\sigma^\top F(\sigma)\right)\mathbf{1}\right) = \left(I - \sigma\mathbf{1}^\top\right)(\sigma \circ F(\sigma)) \\ \check{\mu}(\sigma) &= \sigma \circ \left(F(\sigma) - \left(\min_i F_i(\sigma)\right)\mathbf{1}\right) \end{aligned} \tag{5.4}$$

$\tilde{\mu}(\sigma)$ is differentiable, thus it is used to study the differential properties of the fixed-point bundle in subsequent developments. As for $\check{\mu}(\sigma)$, we have the following theorem.

Theorem 7. *For $(\sigma, \mu) \in \Delta \times \mathbb{R}^n$, the following equation holds.*

$$\mathrm{gap}(\sigma) = \mathbf{1}^\top\check{\mu}(\sigma) = \mathbf{1}^\top\mu - \min_i\left((G(\sigma, \mu) + \mu)/\sigma\right)_i \tag{5.5}$$

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

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

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

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

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

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

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

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

Thus, we obtain the equation.

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

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

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

Therefore, the solutions of $\text{VI}(F, \Delta)$ are equivalently the zero points of $\check{\mu}(\sigma)$, and they are equivalently limit points of the paths subject to UKKT (4.2) as $\mu \to 0$ with $\mu > 0$. Furthermore, equation (5.5) also provides a measurement of the gap of $\text{VI}(F, \Delta)$. Specifically, for any $(\sigma, \mu) \in \Delta \times \{\mu | \mu > 0\}$, if $|G(\sigma, \mu)| < \mu$ component-wise, then $0 \leq \mathbf{1}^\top \check{\mu}(\sigma) < \mathbf{1}^\top \mu$. $G(\sigma, \mu)$ measures the proximity to the fixed-point bundle, and $\mathbf{1}^\top \mu$ measures the proximity of $\mu \to 0$, when they both have sufficient precision, $\mathbf{1}^\top \check{\mu}(\sigma)$ has sufficient precision. This transforms the precision approaching the path and the precision approaching the path endpoint to the precision solving $\text{VI}(F, \Delta)$.

It is worth noting that $\check{\mu}(\sigma)$ and $\tilde{\mu}(\sigma)$ do not have the same zero points. For example, all the vertices of the simplex are zero points of $\tilde{\mu}(\sigma)$, but not all of them are zero points of $\check{\mu}(\sigma)$ since they are not necessarily solutions of $\text{VI}(F, \Delta)$. The reason is that we have the nonnegativity $F(\sigma) - (\min_i F_i(\sigma))\mathbf{1} \geq 0$, while $F(\sigma) - (\sigma^\top F(\sigma))\mathbf{1}$ does not have this nonnegativity. Thus, despite that $(\sigma, \check{\mu}(\sigma))$ and $(\sigma, \tilde{\mu}(\sigma))$ both satisfy UKKT (4.2), the corresponding r of $(\sigma, \check{\mu}(\sigma))$ satisfies $r \geq 0$, whereas the corresponding r of $(\sigma, \tilde{\mu}(\sigma))$ may have $r_i < 0$ for some indices. That is why zero points of $\check{\mu}(\sigma)$ are equivalently solutions of $\text{VI}(F, \Delta)$ while zero points of $\tilde{\mu}(\sigma)$ are not necessarily.

5.2 Differential properties

To maintain $\sigma \in \Delta$, we use the parameterization

$$\sigma = \text{softmax}(\theta) = \frac{\exp(\theta)}{\mathbf{1}^\top \exp(\theta)}. \tag{5.6}$$

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

$$d\sigma = \frac{\exp(\theta) \circ d\theta}{\mathbf{1}^\top \exp(\theta)} - \frac{\mathbf{1}^\top \left(\exp(\theta) \circ d\theta\right)}{(\mathbf{1}^\top \exp(\theta))^2} \exp(\theta) = \sigma \circ d\theta - (\sigma^\top d\theta)\sigma = \sigma \circ \left(I - \mathbf{1}\sigma^\top\right) d\theta. \tag{5.7}$$

Note that $\mathbf{1}^\top d\sigma = 0$ for any $d\theta$. Conversely, when $\mathbf{1}^\top d\sigma = 0$, the linear equation $d\sigma/\sigma = (I - \mathbf{1}\sigma^\top)d\theta$ admits the general solution $d\theta = d\sigma/\sigma + k\mathbf{1}$ with $k \in \mathbb{R}$. We denote $\overline{d\theta} := (I - \mathbf{1}\sigma^\top)d\theta$.

Next, we derive the differential $d\tilde{\mu}(\sigma)$ with respect to $\overline{d\theta}$. In the 3rd line, we use $\mathbf{1}^\top d\sigma = 0$. In the last line, we use $d\sigma/\sigma = \overline{d\theta} = (I - \mathbf{1}\sigma^\top)\overline{d\theta}$.

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

Here, we introduce the matrix $J(\sigma)$ in equation (5.8), where $k \in \mathbb{R}$ is an arbitrary scalar.

$$J(\sigma) := \left(I - \mathbf{1}\sigma^\top\right)\left(\frac{\partial F(\sigma)}{\partial \sigma} \circ \sigma + \left(\left(I - \mathbf{1}\sigma^\top\right)F(\sigma)\right) \circ I\right)\left(I - \mathbf{1}\sigma^\top\right) + k\mathbf{1}\sigma^\top \tag{5.8}$$

By $\sigma^\top \overline{d\theta} = 0$, we arrive at the following differential equation of $\tilde{\mu}(\sigma)$.

$$d\tilde{\mu}(\sigma) = \sigma \circ J(\sigma)\overline{d\theta} \tag{5.9}$$

Next, we have Proposition 8 that classifies the eigenvalues of $J(\sigma)$ using invariant subspace decomposition. A subspace U is an invariant subspace of matrix A if there is $Au \in U$ for any $u \in U$. If A admits an invariant subspace decomposition, then the eigenvalues of A consist of the union of the eigenvalues of the restrictions of A to these invariant subspaces.

Proposition 8. *Matrix $J(\sigma)$ admits invariant subspace decomposition $\mathbb{R}^n = \operatorname{span}(\mathbf{1}) \oplus \ker(\sigma^\top)$. The eigenvalue corresponding to eigenvectors in $\operatorname{span}(\mathbf{1})$ is k, and eigenvalues corresponding to eigenvectors in $\ker(\sigma^\top)$ are independent of k.*

Proof. First, we have $J(\sigma)\mathbf{1} = k\mathbf{1}$ since $(I - \mathbf{1}\sigma^\top)\mathbf{1} = 0$ and $\sigma^\top\mathbf{1} = 1$. Then, $\operatorname{span}(\mathbf{1})$ is an invariant subspace of $J(\sigma)$. Second, we have $\sigma^\top J(\sigma) = k\sigma^\top$ 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 an invariant subspace of $J(\sigma)$. Since $\operatorname{span}(\mathbf{1})$ and $\ker(\sigma^\top)$ are complementary, $\mathbb{R}^n = \operatorname{span}(\mathbf{1}) \oplus \ker(\sigma^\top)$ is an invariant subspace decomposition.

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

Proposition 8 reveals that our intention adding an arbitrary scalar $k \in \mathbb{R}$ in $J(\sigma)$ is to adjust the eigenvalue corresponding to eigenvector $\mathbf{1}$. Specifically, the inherent zero eigenvalue corresponding to eigenvector $\mathbf{1}$ is removed, and the rest of the eigenvalues remain unchanged. This is a trick preventing $J(\sigma)$ 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 $G(\sigma, \mu) = 0$, i.e., the differential equation of the fixed-point bundle.

$$\begin{aligned} G'_\theta(\sigma,\mu)d\theta &= -\,G'_\mu(\sigma,\mu)d\mu \\ \sigma \circ J(\sigma)\overline{d\theta} + (\mathbf{1}^\top\mu)\sigma \circ \overline{d\theta} &= \left(I - \sigma\mathbf{1}^\top\right)d\mu \end{aligned} \tag{5.10}$$

Multiplying both sides with $1/\sigma$, we arrive at the differential equation of $G(\sigma, \mu) = 0$ in equation (5.11). This differential equation provides the tangent step in θ with respect to the reduction $d\mu$ to follow the fixed-point bundle. We denote $J_G := J(\sigma) + (\mathbf{1}^\top\mu)I$.

$$\left(J(\sigma) + (\mathbf{1}^\top\mu)I\right)\overline{d\theta} = \left(I - \mathbf{1}\sigma^\top\right)(d\mu/\sigma) \tag{5.11}$$

The point (σ, μ) on the fixed-point bundle is called a singular point if $J(\sigma) + (\mathbf{1}^\top\mu)I$ is singular. Apparently, 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 $-k$. Therefore, we can construct the singular manifold of the fixed-point bundle using the real eigenvalues of $J(\sigma)$. Denoting the real eigenvalues of $J(\sigma)$ except k as $\lambda_j^\perp(J(\sigma)) \in \mathbb{R}$, the set of all the singular points of fixed-point bundle E is formalized as the singular manifold S in equation (5.12).

$$S = \left\{(\sigma, \tilde{\mu}(\sigma) - \lambda_j^\perp\left(J(\sigma)\right)\sigma)|\lambda_j^\perp\left(J(\sigma)\right) \in \mathbb{R}, \sigma \in \Delta\right\}. \tag{5.12}$$

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

5.3 Analytic curves

We study the curve on the fixed-point bundle E defined by

$$\Gamma(\sigma_{\text{init}}) = \{(\sigma, \mu) \in E | \mu = \gamma\sigma_{\text{init}}, \gamma \in [0, +\infty)\} \tag{5.13}$$

with designated $\sigma_{\text{init}} \in \Delta$. $\Gamma(\sigma_{\text{init}})$ is a one-dimensional curve parameterized by γ, 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. As the zero set of the analytic function $G(\sigma, \mu) = 0$, $\Gamma(\sigma_{\text{init}})$ is called an analytic curve.

Adapting the proof of the oddness theorem of Nash equilibrium [41], we have the oddness theorem of VI(F, Δ) in Theorem 9. Theorem 9 shows that when all the solutions are nonsingular, they are connected in pairs with the designated starting point, implying that there are an odd number of solutions. Furthermore, the solutions of VI(F, Δ) are almost always all nonsingular, i.e., the VI(F, Δ) instances that have at least one singular solution have measure zero.

Theorem 9 (Oddness theorem)**.** *For* VI(F, Δ) *with* $F(\sigma)$ *being real analytic on* Δ*, if* $J(\sigma)$ *is nonsingular on all the solutions, then there are an odd number of solutions.*

Proof. We prove this using the analytic curve $\Gamma(\sigma_{\text{init}})$, such that there is exactly one solution connected to the starting point σ_{init}, and the rest of the solutions are connected in pairs, which is called the parity arguement.

First, for any μ, there are at most finitely many solutions of $G(\sigma, \mu) = 0$. Because Δ is a compact set, if there are infinitely many zero points, there must be a convergent zero point sequence. If an analytic function takes the value 0 on a convergent point sequence, the analytic function is constantly 0. Thus, there are always finitely many solutions of $G(\sigma, \mu) = 0$ for any μ. Then, the analytic curve $\Gamma(\sigma_{\text{init}})$ consists of finitely many branches.

Second, the Newton-Puiseux theorem states that for a real analytic function $G(\sigma, \gamma\sigma_{\text{init}})$ that is not constantly zero, in the neighborhood of every (σ_0, γ_0) such that $G(\sigma_0, \gamma_0\sigma_{\text{init}}) = 0$, the solution of $G(\sigma, \gamma\sigma_{\text{init}}) = 0$ can be locally expanded as a convergent Puiseux series

$$\sigma - \sigma_0 = \sum_{i=1}^{\infty} c_i (\gamma - \gamma_0)^{i/m}$$

with some positive integer m. As a consequence, there is a unique analytic continuation beyond every point $(\sigma_0, \gamma_0\sigma_{\text{init}})$ on $\Gamma(\sigma_{\text{init}})$, whether $(\sigma_0, \gamma_0\sigma_{\text{init}})$ is singular or nonsingular.

Third, since $\Gamma(\sigma_{\text{init}})$ has a unique analytic continuation beyond every point on it, the boundary of $\Gamma(\sigma_{\text{init}})$ can only be at where $\gamma \to 0^+$ and $\gamma \to +\infty$, and the branches of $\Gamma(\sigma_{\text{init}})$ connect them. The ending points at where $\gamma \to 0^+$ are equivalently the solutions of VI(F, Δ). The starting point at where $\gamma \to +\infty$ is a single point σ_{init}, since $\sigma \circ F(\sigma) + v\sigma = \gamma\sigma_{\text{init}}$ and $F(\sigma)$ is bounded.

Finally, when $J(\sigma)$ is nonsingular on all the solutions of VI(F, Δ), there is exactly one branch of $\Gamma(\sigma_{\text{init}})$ connecting to each solution due to the implicit function theorem. Thus, the branches of analytic curve $\Gamma(\sigma_{\text{init}})$ connect the solutions of VI(F, Δ) and a single starting point σ_{init} in pairs.

Therefore, we obtain the parity arguement when $J(\sigma)$ is nonsingular on all the solutions of VI(F, Δ), it follows that there are an odd number of solutions. □

The parity arguement reveals that there is an analytic curve branch connecting any designated starting point σ_{init} and a solution of $\text{VI}(F, \Delta)$. If we follow this path, we can achieve global convergence to solutions of $\text{VI}(F, \Delta)$. However, there can be singular points on the analytic curve, and path-following cannot go beyond a singular point, thus they must be avoided. On the fixed-point bundle, singularity avoidance is achievable simply by moving along the fiber $B(\sigma)$. Specifically, on a point (σ, μ) where $\mathbf{1}^\top \mu$ equals an eigenvalue of $-J(\sigma)$, we can simply jump to $(\sigma, \mu + \beta\sigma)$ with some β such that $\mu + \beta\sigma > 0$ and $\mathbf{1}^\top \mu + \beta$ is not an eigenvalue of $-J(\sigma)$. Then, $(\sigma, \mu + \beta\sigma)$ is not a singular point, and we can proceed to reduce $\mu + \beta\sigma$. This actually moves to a new analytic curve $\Gamma((\mu + \beta\sigma)/(\mathbf{1}^\top \mu + \beta))$ with a different starting point.

6 Predictor-corrector path-following on the fixed-point bundle

We adopt the predictor-corrector framework to perform the path-following on the fixed-point bundle. Given a starting point on the path, the predictor-corrector path-following framework alternates between predictor steps and corrector steps. The predictor step moves a small step along the tangent direction of the path, and the corrector step is a subiteration that locally converges back onto the path.

6.1 The path and its nonsingular neighborhood

The path we follow is derived from the analytic curve $\Gamma(\sigma_{\text{init}})$ and singularity avoidance along the fiber, which can be formalized as $\tilde{\Gamma}$ in equation (6.1). The starting point $(\sigma_0, \gamma_0\mu_0) \in E$ is given by $\gamma_0\mu_0 = v_{\text{init}}\sigma_{\text{init}}$ with designated σ_{init} and a large enough v_{init}. Then, $\tilde{\Gamma}$ consists of segments of different analytic curve $\Gamma(\sigma_{\text{init}})$, which are continuous in σ, but jump along the fibers in μ. $\tilde{\Gamma}$ eventually ends at the point where $\mathbf{1}^\top \mu = \zeta$. With $\zeta = 0$, the ending point of $\tilde{\Gamma}$ is $(\sigma^\star, \mathbf{0})$, where $\sigma^\star$ is a solution of $\text{VI}(F, \Delta)$.

$$\begin{aligned} \tilde{\Gamma} = \bigcup_{i=1}^{q} \left\{ (\sigma, \gamma\mu_i) \in E | \mu_i = \gamma_{i-1}\mu_{i-1} + \beta_i\sigma_{i-1}, \gamma \in [\gamma_i, 1] \right\}, \\ \gamma_0\mu_0 = v_{\text{init}}\sigma_{\text{init}},\ \mu_i > 0,\ \gamma_q \mathbf{1}^\top \mu_q = \zeta \end{aligned} \tag{6.1}$$

We wish to have $\tilde{\Gamma} \cap S = \varnothing$ for singularity avoidance, where S is the singular manifold. However, $\mathbf{1}^\top \mu = 0$ is unavoidable at the solution $(\sigma^\star, \mathbf{0})$. If $J(\sigma^\star)$ is singular, it is unavoidable that $\tilde{\Gamma}$ intersects with S. Thus, we introduce a parameter ζ to control whether $\tilde{\Gamma}$ ends exactly at $(\sigma^\star, \mathbf{0})$ or ends before $(\sigma^\star, \mathbf{0})$ at some $\mathbf{1}^\top \mu = \zeta > 0$. In fact, we can have $\tilde{\Gamma} \cap S = \varnothing$ with only finitely many segments in $\tilde{\Gamma}$.

Proposition 10. *If at the ending point $(\sigma_q, \gamma_q\mu_q)$ of $\tilde{\Gamma}$, $\gamma_q \mathbf{1}^\top \mu_q = \zeta$ is not an eigenvalue of $-J(\sigma_q)$, then there exists $\tilde{\Gamma}$ that consists of finitely many segments satisfying $\tilde{\Gamma} \cap S = \varnothing$.*

Proof. Suppose we adopt a specific singularity avoidance procedure that always uses $\beta_i < 0$. Note that 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, this singularity avoidance procedure is admissible, and it generates a nonincreasing sequence of $\gamma_i \mathbf{1}^\top \mu_i$.

Suppose, for contradiction, that there have to be infinitely many segments in $\tilde{\Gamma}$ to achieve $\tilde{\Gamma} \cap S = \varnothing$. Then, the infinite sequence of $\gamma_i \mathbf{1}^\top \mu_i$ converges since it is bounded from below by ζ. The limit must be ζ because there always exists $\beta_i < 0$ for singularity avoidance that ensures decreasing as long as $\gamma_{i-1}\mu_{i-1} > 0$. Then, there have to be infinitely many times of singularity avoidance in arbitrarily small neighborhood of $(\sigma_q, \gamma_q\mu_q)$. This yields a contradiction, because $\gamma_q \mathbf{1}^\top \mu_q$ is not an eigenvalue of $-J(\sigma_q)$. $\square$

Therefore, we set $\zeta = 0$ for nonsingular $(\sigma^\star, \mathbf{0})$ and set $\zeta > 0$ for singular $(\sigma^\star, \mathbf{0})$, then we can always have $\tilde{\Gamma} \cap S = \varnothing$ with finitely many segments in $\tilde{\Gamma}$.

In addition to $\tilde{\Gamma} \cap S = \varnothing$, we also need a neighborhood of $\tilde{\Gamma}$ that is disjoint to S, because the corrector is a subiteration that converges onto the path $\tilde{\Gamma}$ in its neighborhood. We introduce the neighborhood $\mathcal{N}_{\tilde{\Gamma}}$ of $\tilde{\Gamma}$ in equation (6.2). $\mathcal{N}_{\tilde{\Gamma}}$ is the union of all the d-radius $\ln \sigma$-coordinate neighborhood of σ_μ for each $(\sigma_\mu, \mu) \in \tilde{\Gamma}$. We require $\mu > 0$ because $\sigma_\mu > 0$ is needed to guarantee that $\ln \sigma_\mu$ is well-defined, this excludes the ending point of $\tilde{\Gamma}$ where $\mathbf{1}^\top \mu = 0$ if $\zeta = 0$. Then, we take closure to make sure that $\mathcal{N}_{\tilde{\Gamma}}$ is compact, this includes the neighborhood of the ending point of $\tilde{\Gamma}$ where $\mathbf{1}^\top \mu = 0$ if $\zeta = 0$.

$$\mathcal{N}_{\tilde{\Gamma}} = \text{Closure}\left(\left\{(\text{softmax}(\theta), \mu) \,|\, \|\theta - \ln \sigma_\mu\| \le d, (\sigma_\mu, \mu) \in \tilde{\Gamma}, \mu > 0\right\}\right) \tag{6.2}$$

The $\ln \sigma$-coordinate in the definition of $\mathcal{N}_{\tilde{\Gamma}}$ aligns with our parameterization $\sigma = \text{softmax}(\theta)$, resulting in a neighborhood of σ_μ that always lies in the simplex even as σ_μ approaches the boundary. Note that $\|\theta - \ln \sigma_\mu\| \le d$ is equivalently $\mathrm{e}^{-d} - 1 \le (\sigma_\mu - \sigma)/\sigma \le \mathrm{e}^{d} - 1$. In fact, the $\ln \sigma$-coordinate strongly resembles the local norm $\|d\sigma\|_{\text{diag}(1/\sigma^2)} = d\sigma^\top \text{diag}(1/\sigma^2) d\sigma$ in interior point method, where $\text{diag}(1/\sigma^2)$ is the Hessian matrix of the log barrier $-\ln \sigma$. In interior point method, the neighborhood defined by the Hessian-induced local norm is called the Dikin ellipsoid [42].

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

Proposition 11. *There exists $d > 0$ such that $\mathcal{N}_{\tilde{\Gamma}} \cap S = \varnothing$.*

Proof. If two compact sets are disjoint, there must be a positive gap between them. We first show that $\tilde{\Gamma}$ and S are compact, and then we can derive $d > 0$ since $\tilde{\Gamma} \cap S = \varnothing$.

Each analytic curve segment in $\tilde{\Gamma}$ is compact, because it is the image of a compact interval of γ under the continuous analytic curve mapping. Then, $\tilde{\Gamma}$ is compact because the union of finitely many compact sets is compact. The mapping from σ to $(\sigma, \tilde{\mu}(\sigma) - \lambda_j^\perp(J(\sigma))\sigma)$ is continuous, because the section $\tilde{\mu}(\sigma)$ is continuous and the eigenvalue mapping $\lambda_j^\perp(J(\sigma))$ is continuous. Then, S is compact because it is the image of compact Δ under a continuous function.

Since $\tilde{\Gamma} \cap S = \varnothing$ and they are both compact, there exists a positive gap between them. Specifically, there exists $d_0 > 0$, such that for any $(\sigma_\mu, \mu) \in \tilde{\Gamma}$ and $(\sigma_s, \mu_s) \in S$, we have

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

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

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

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

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

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

Therefore, let $d \in (0, 2d_0)$, then $\mathcal{N}'_{\tilde{\Gamma}}$ and S have a positive gap $d_0 - d/2 > 0$. If two sets have a positive gap, their closure also have a positive gap. Therefore, we obtain $\mathcal{N}_{\tilde{\Gamma}} \cap S = \varnothing$. ☐

Finally, we bound two quantities on the compact nonsingular path neighborhood $\mathcal{N}_{\tilde{\Gamma}}$.

Lemma 12. *The following quantities are bounded on $\mathcal{N}_{\tilde{\Gamma}}$.*

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

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

Proof. (i) $s_{\min}(J(\sigma)+(\mathbf{1}^\top\mu)I)$ is bounded because it is continuous on the compact set $\mathcal{N}_{\tilde{\Gamma}}$.
(ii) μ/σ satisfies

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

First, the radio is bounded by $\mathrm{e}^{-d}\le\sigma_\mu/\sigma\le\mathrm{e}^{d}$ on $\mathcal{N}_{\tilde{\Gamma}}$. Second, $F(\sigma_\mu)$ is bounded by continuity of $F(\sigma)$ on the compact Δ. Third, v_μ is bounded by continuity of Brouwer function $(\hat{\sigma},r,v)=M(\sigma,\mu)$ on the compact $\mathcal{N}_{\tilde{\Gamma}}$. Brouwer function $(\hat{\sigma},r,v)=M(\sigma,\mu)$ is continuous on $\mathcal{N}_{\tilde{\Gamma}}$ because it is continuous on every $\sigma\in\Delta$ and $\mu>0$, and the limit as $\mu\to0$ with $\mu>0$ exists and is $v\to-\min_i F_i(\sigma)$. □

Denote the lower bound of $s_{\min}(J(\sigma)+(\mathbf{1}^\top\mu)I)$ on $\mathcal{N}_{\tilde{\Gamma}}$ as $s_{\min}(J_G)$. If $J(\sigma^\star)$ is nonsingular, the bound $s_{\min}(J_G)$ is uniform over the entire path-following process till reaching $(\sigma^\star,\mathbf{0})$. However, if $J(\sigma^\star)$ is singular, the bound $s_{\min}(J_G)$ is uniform only over the part where $\mathbf{1}^\top\mu\ge\zeta>0$. In addition, $s_{\min}(J_G)$ is dependent to ζ such that $s_{\min}(J_G)\to0$ as $\zeta\to0$. Thus, for a singular solution, the subsequent convergence result only holds for path-following that ends at $\mathbf{1}^\top\mu=\zeta$ with a fixed ζ.

6.2 Corrector derived from perturbed KKT conditions

In the interior point method, 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 can derive two different but deeply related correctors from them. We start with the Newton method solving fixed-point bundle equation $G(\sigma,\mu)=0$, which is equivalently solving UKKT (4.2). The Newton equation of $G(\sigma,\mu)=0$ is derived as follows, where $G'_\theta(\sigma,\mu)d\theta$ is previously obtained in differential equation (5.11).

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

Multiplying both sides with $1/\sigma$, we arrive at the Newton equation of $G(\sigma,\mu)=0$ in equation (6.3).

$$\Big(J(\sigma)+(\mathbf{1}^\top\mu)I\Big)\,\overline{d\theta} = -\,\Big(I-\mathbf{1}\sigma^\top\Big)\,(F(\sigma)-\mu/\sigma) \tag{6.3}$$

We denote the term appears in the right-hand side of Newton equation (6.3) as $\tilde{G}(\sigma,\mu)=G(\sigma,\mu)/\sigma$. $\tilde{G}(\sigma,\mu)$ is bounded on the neighborhood $\mathcal{N}_{\tilde{\Gamma}}$, because $I-\mathbf{1}\sigma^\top$ and $F(\sigma)$ are both bounded on the compact simplex by continuity, and μ/σ is bounded on $\mathcal{N}_{\tilde{\Gamma}}$ by Lemma 12. The first derivative of $\tilde{G}(\sigma,\mu)$ is derived as follows.

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

The second derivative of $\tilde{G}(\sigma,\mu)$ has the following bound.

$$\Big\|\tilde{G}''(\sigma,\mu)\Big\| = \Big\|J'(\sigma)-\Big(\tilde{G}(\sigma,\mu)\circ I\Big)'\Big\| \le \big\|J'(\sigma)\big\| + \Big\|\tilde{G}'(\sigma,\mu)\Big\|$$

$$\leq \left\|J'(\sigma)\right\| + \|J(\sigma)\| + \mathbf{1}^\top \mu + \left\|\tilde{G}(\sigma, \mu)\right\|$$

We analyze the convergence rate of the predictor-corrector path-following framework with the standard Newton corrector in equation (6.3). Adapting the standard quadratic convergence proof of the standard Newton iteration [43], we have the following proposition stating that $\tilde{G}(\sigma, \mu)$ achieves uniform quadratic convergence to 0 under the corrector.

Proposition 13. *Let $\theta_{k+1} = \theta_k + \overline{d\theta}_k$, where $\overline{d\theta}_k$ is given by Newton equation* (6.3)*. Then, for any $(\sigma_k, \mu) \in \mathcal{N}_{\tilde{\Gamma}}$, there exists $C_1 > 0$ such that*

$$\left\|\tilde{G}(\sigma_{k+1}, \mu)\right\| \leq C_1 \left\|\tilde{G}(\sigma_k, \mu)\right\|^2. \tag{6.4}$$

Proof. We have the following derivation. In the 3rd line and 4th line, we use the fundamental theorem of calculus to substitute the residuals, where $\sigma_u = \mathrm{softmax}(\theta_k + u(\theta_{k+1} - \theta_k))$ and $\sigma_w = \mathrm{softmax}(\theta_u + w(\theta_u - \theta_k))$. In the 5th line, we substitute $\theta_{k+1} - \theta_k$ with $\overline{d\theta}_k$. L_1 in the last line follows from the bound of the second derivative $\tilde{G}''(\sigma, \mu)$.

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

In the above equation, $\|(J(\sigma) + (\mathbf{1}^\top \mu) I)^{-1}\|$ is uniformly bounded by $1/s_{\min}(J_G)$ on the neighborhood $\mathcal{N}_{\tilde{\Gamma}}$. L_1 is also uniformly bounded, because $\|J'(\sigma)\|$ and $\|J(\sigma)\|$ are bounded on the compact simplex by continuity, and $\mathbf{1}^\top \mu$ and $\|\tilde{G}(\sigma, \mu)\|$ are bounded on the compact neighborhood $\mathcal{N}_{\tilde{\Gamma}}$. Therefore, we obtain the inequality. ☐

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

Proposition 14. *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* (5.11) *with $d\mu_t = -\eta \mu_t$. Then, for any $(\sigma_t, \mu_t) \in \mathcal{N}_{\tilde{\Gamma}}$, there exists $C_2 > 0$ such that*

$$\left\|\tilde{G}(\sigma_{t+1}, \mu_{t+1}) - \tilde{G}(\sigma_t, \mu_t)\right\| \leq C_2 \eta. \tag{6.5}$$

Proof. We have the following derivation. In the 2nd line, we use the fundamental theorem of calculus to substitute the residual, where $\sigma_u = \mathrm{softmax}(\theta_t + u(\theta_{t+1} - \theta_t))$. In the 3rd line, we substitute $\theta_{t+1} - \theta_t$ with $\overline{d\theta}_t$. L_2 in the last line follows from the bound of the first derivative $\tilde{G}'(\sigma, \mu)$.

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

$$
\begin{aligned}
&\leq \int_0^1 \left\| \tilde{G}'(\sigma_u, \mu_{t+1}) \right\| \|\theta_{t+1} - \theta_t\| \, du + \left\| \left(I - \mathbf{1}\sigma_t^\top \right) (d\mu_t / \sigma_t) \right\| \\
&\leq \left(L_2 \left\| \left(J(\sigma_t) + (\mathbf{1}^\top \mu_t) I \right)^{-1} \right\| + 1 \right) \left\| I - \mathbf{1}\sigma_t^\top \right\| \eta \, \|\mu_t/\sigma_t\| \\
&L_2 := \sup_{\sigma \in \Delta} \|J(\sigma)\| + \mathbf{1}^\top \mu_{t+1} + \sup_{(\sigma,\mu) \in \mathcal{N}_{\tilde{\Gamma}}} \left\| \tilde{G}(\sigma, \mu) \right\|
\end{aligned}
$$

In the above equation, $\|(J(\sigma) + (\mathbf{1}^\top \mu) I)^{-1}\|$ is uniformly bounded by $1/s_{\min}(J_G)$ on the neighborhood $\mathcal{N}_{\tilde{\Gamma}}$. L_2 is also uniformly bounded, because $\|J(\sigma)\|$ is bounded on the compact simplex by continuity, and $\mathbf{1}^\top \mu$ and $\|\tilde{G}(\sigma, \mu)\|$ are bounded on the compact neighborhood $\mathcal{N}_{\tilde{\Gamma}}$. $\|I - \mathbf{1}\sigma^\top\|$ is bounded on the compact simplex by continuity. $\|\mu/\sigma\|$ is bounded on $\mathcal{N}_{\tilde{\Gamma}}$ by Lemma 12. Therefore, we obtain the inequality. □

In conclusion, for the predictor-corrector path-following algorithm with corrector steps in Proposition 13 and predictor steps in Proposition 14: (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 ζ under the overall iteration.

Therefore, we can draw the conclusion that the predictor-corrector algorithm achieves global linear convergence along the path $\tilde{\Gamma}$ to its ending point where $\mathbf{1}^\top \mu = \zeta$. For a nonsingular solution $(\sigma^\star, \mathbf{0})$, $\tilde{\Gamma}$ ends exactly at $(\sigma^\star, \mathbf{0})$ where $\mathbf{1}^\top \mu = 0$, and thus the algorithm achieves global linear convergence. For a singular solution $(\sigma^\star, \mathbf{0})$, $\tilde{\Gamma}$ ends before $(\sigma^\star, \mathbf{0})$ at where $\mathbf{1}^\top \mu = \zeta > 0$, and thus the algorithm retains global linear reduction up to a fixed precision $\zeta > 0$. For the singular solution case, the algorithm does not maintain global linear reduction to $(\sigma^\star, \mathbf{0})$, but it is still convergent, thus the convergence reduces to sublinear as required precision increases.

Standard Newton iteration is simple for theoretical analysis, however, its performance in practice is not ideal. Because the step length is affected by the smallest singular value $s_{\min}(J_G)$, such that the iteration appears to be unstable if $s_{\min}(J_G)$ is too small. A classic method addressing this issue is regularized Newton equation (6.6), derived from standard Newton equation (6.3) by multiplying both sides with $J_G^\top$ and adding the regularization term $\delta_H I$ with $\delta_H > 0$, so that the coefficient is positive semidefinite and controlled by δ_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 δ_H. But the convergence reduces from quadratic convergence to linear convergence with rate $\delta_H / (s_{\min}^2(J_G) + \delta_H)$. Thus, regularization actually trades convergence speed for iteration stability.

$$
\begin{aligned}
\left(J_G^\top J_G + \delta_H I \right) d\theta &= - \, J_G^\top \tilde{G}(\sigma, \mu) \\
\overline{d\theta} &= \left(I - \mathbf{1}\sigma^\top \right) d\theta
\end{aligned}
\tag{6.6}
$$

Regularized Newton equation (6.6) is approximately the regularized Gauss-Newton equation of the least square problem (6.7) in terms of the Euclidean norm of $\tilde{G}(\sigma, \mu)$, which is equivalently the local norm of $G(\sigma, \mu)$. Since the local norm is induced by the Hessian matrix $\mathrm{diag}(1/\sigma^2)$ of the log barrier $-\ln \sigma$, problem (6.7) is actually a log barrier problem. Thus, the 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 (6.6) mixes Newton steps and gradient steps, such that it reduces to standard Newton step (6.3) if $\delta_H = 0$, and it approximates gradient step $-(1/\delta_H) J_G^\top \tilde{G}(\sigma, \mu)$ if δ_H is large.

$$
\min_\sigma \frac{1}{2} \left\| \tilde{G}(\sigma, \mu) \right\| = \min_\sigma \frac{1}{2} \|G(\sigma, \mu)\|_{\mathrm{diag}(1/\sigma^2)} \tag{6.7}
$$

6.3 Corrector derived from barrier problem

Since the fixed-point bundle is also a special central path of the MCP (4.5), we can also derive a gradient from the barrier MCP (4.5) for corrector. First, we additionally require that (r, v) is subject to the Brouwer function $(\hat{\sigma}, r, v) = M(\sigma, \mu)$. Then, we can view σ as the only optimization variable of barrier MCP (4.5), whereas (r, v) is an intermediate variable, so the gradient is only with respect to $d\sigma$, and thus only with respect to $\overline{d\theta}$. Additionally, this also enables us to use $\mathbf{1}^\top(\sigma - \hat{\sigma}) = 0$ to eliminate dv in the differential.

The gradient of barrier MCP (4.5) is derived as follows. In the 3rd 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 4th 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 5th line, we use $k\sigma\mathbf{1}^\top(\sigma - \hat{\sigma}) = 0$ in the first additive $J(\sigma)$ and use $(I - \mathbf{1}\sigma^\top)(I - \mathbf{1}\sigma^\top) = (I - \mathbf{1}\sigma^\top)$ in the second additive. In the 6th line, we use $(I - \sigma\mathbf{1}^\top)(\sigma - \hat{\sigma}) = (\sigma - \hat{\sigma})$.

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

Ignoring the higher-order term $((\sigma - \hat{\sigma})^\top F(\sigma))(\sigma - \hat{\sigma})$, we arrive at the inexact gradient of MCP (4.5) in equation (6.8).

$$
\nabla_\theta\left(\sigma^\top r - \mu^\top \ln\sigma - \mu^\top \ln r\right) = \left(J(\sigma) + (\mathbf{1}^\top\mu)I\right)^\top (\sigma - \hat{\sigma}) \tag{6.8}
$$

Next, we can also derive a regularized Newton equation from standard Newton equation (6.3) with gradient $J_G^\top(\sigma - \hat{\sigma})$ on the right-hand side. The derivation is as follows. In the 1st line, we 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 2nd line, we denote $(\hat{\sigma}, r, v_b) = M(\sigma, \mu)$ such that $\hat{\sigma} \circ (F(\sigma) + v_b\mathbf{1}) = \mu$. In the 3rd line, we denote $v_{gb} = (v_g - v_b)/(\mathbf{1}^\top\mu + k)$ and use $J(\sigma)\mathbf{1} = k\mathbf{1}$.

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

Then, we arrive at the regularized Newton equation (6.9). Similar to regularized Newton equation (6.6), regularized Newton equation (6.9) mixes Newton steps and gradient steps, such that it reduces to standard Newton step (6.3) if $\delta_H = 0$, and it approximates gradient step $-(1/\delta_H)J_G^\top(\sigma - \hat{\sigma})$ if δ_H is large.

$$\begin{aligned} \left(J_G^\top \circ \frac{\sigma}{r} \circ J_G + \delta_H I\right) d\theta &= -\, J_G^\top\left(\sigma - \hat{\sigma}\right) \\ \overline{d\theta} &= \left(I - \mathbf{1}\sigma^\top\right) d\theta \end{aligned} \tag{6.9}$$

Similar to regularized Newton equation (6.6), regularized Newton equation (6.9) also corresponds to a least square problem (6.10) 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 similar derivation as that of barrier MCP (4.5).

$$\min_\sigma \|\sigma \circ r - \mu\|_{\mathrm{diag}(1/(\sigma\circ r))} = \min_\sigma \left(\sigma - \mu/r\right)^\top\left(r - \mu/\sigma\right) \tag{6.10}$$

7 The unified framework

The predictor-corrector path-following algorithm on the fixed-point bundle is as Algorithm 1 shows. The algorithm follows the path $\tilde{\Gamma}$ from the starting point near $(\sigma_{\text{init}}, v_{\text{init}}\sigma_{\text{init}})$ with a designated $\sigma_{\text{init}} \in \Delta$ and a sufficiently large v_{init}. The predictor step consists of three parts: avoiding singular points along the fibers, reducing μ, and updating σ along the tangent vector. The corrector step is a subiteration using either regularized Newton equation (6.6) or (6.9), in which obtaining $(\hat{\sigma}, r, v)$ involves solving bisection problem (4.4). The stopping criteria is $\mathbf{1}^\top\check{\mu}(\sigma) < \epsilon$, because $\mathbf{1}^\top\check{\mu}(\sigma)$ equals to the gap of $\mathrm{VI}(F, \Delta)$. By Theorem 7, we have $\mathbf{1}^\top\check{\mu}(\sigma) < \epsilon$ when $\mathbf{1}^\top\mu < \epsilon$ and $G(\sigma, \mu) < \mu$ component-wise. Thus, if we truncate μ such that $\min_i \mu_i \geq \epsilon/(n+1)$, then $\|G(\sigma, \mu)\| < \epsilon/(n+1)$ suffices to imply $G(\sigma, \mu) < \mu$ component-wise, and μ can still decrease down below $\mathbf{1}^\top\mu < \epsilon$.

Algorithm 1 Path-following on the fixed-point bundle

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

```
Set (σ_0, μ_0) = (σ_init, v_init σ_init) for a sufficiently large v_init
repeat
    repeat
        Compute (σ̂_k, r_k, v_k) = M(σ_k, μ_t) by solving bisection problem (4.4)
        Construct the matrix J(σ_k) in equation (5.8)
        Solve regularized Newton equation (6.6) or (6.9) for dθ̄_k with δ_H = ‖G(σ_k, μ_t)‖/n
        Update σ_{k+1} = softmax(ln σ_k + dθ̄_k)
    until ‖G(σ_k, μ_t)‖ < ε/(n+1)
    Set μ̌_t = μ_t + β_t σ_k > 0 such that 1ᵀμ_t + β_t avoids the nonzero eigenvalues of −J(σ_k)
    Update and truncate μ_{t+1} = ((1 − η_t)μ̌_t).clip(min = ε/(n+1)) with a sufficiently small η_t
    Solve differential equation (5.11) for dθ̄_t with dμ_t = μ_{t+1} − μ̌_t
    Set σ_0 = softmax(ln σ_k + dθ̄_t)
until 1ᵀμ̌(σ_k) < ε
return σ_k as an approximate solution of VI(F, Δ)
```

Comparing the two regularized Newton equations (6.6) and (6.9), despite that they are both equivalent to the standard Newton equation (6.3) when $\delta_H = 0$, they are different when $\delta_H \neq 0$. First, the former corresponds to least square problem (6.7) with local norm $\mathrm{diag}(1/\sigma^2)$ that only penalizes σ approaching 0, whereas the latter corresponds to least square problem (6.10) with local norm $\mathrm{diag}(1/(\sigma \circ r))$ that penalizes both σ approaching 0 and r approaching 0.

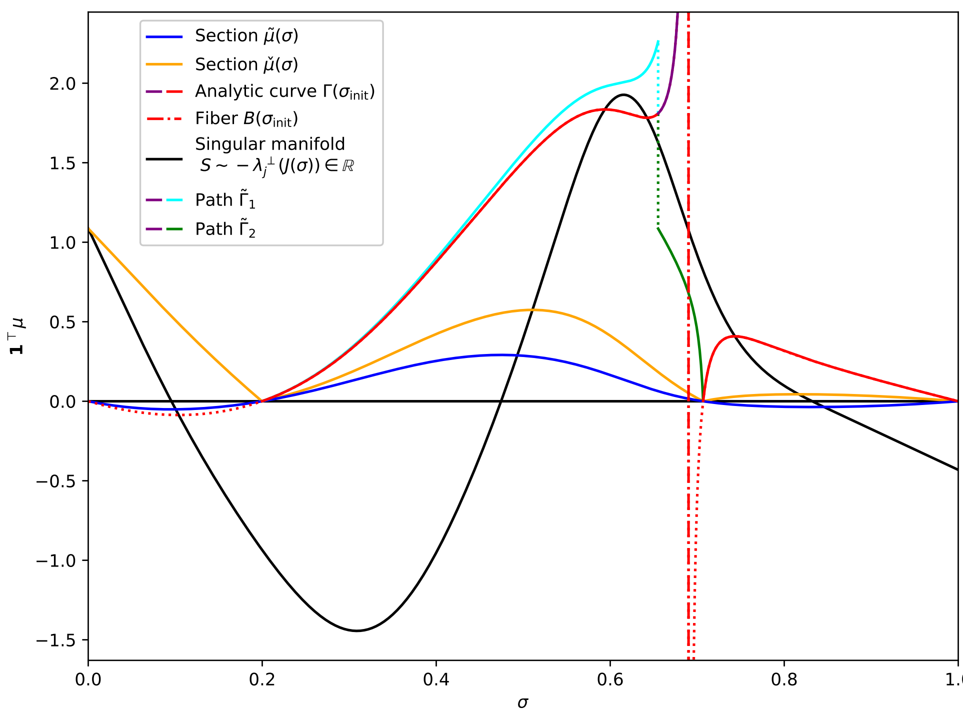

Figure 1: Graph of the fixed-point bundle. This graph shows the fixed-point bundle E of a 2-dimensional $\text{VI}(F, \Delta)$. The horizontal axis in σ represents the simplex, and the vertical axis in $\mathbf{1}^\top \mu$ represents the fiber. Section $\tilde{\mu}(\sigma)$ is smooth and derives the matrix $J(\sigma)$ that measures the differential properties of E, whereas section $\check{\mu}(\sigma)$ is nonnegative and measures the gap of $\text{VI}(F, \Delta)$. The analytic curve $\Gamma(\sigma_{\text{init}})$ is the subcurve of $\{(\sigma, \gamma\sigma_{\text{init}}) \in E | \gamma \in \mathbb{R}\}$ where $\mathbf{1}^\top \mu \geq 0$, setting aside the negative part represented by the red dashed curve. As $\mu \to 0$ with $\mu > 0$, σ tends to the zero points of $\check{\mu}(\sigma)$, which are equivalently solutions of $\text{VI}(F, \Delta)$. Whereas as $\mu \to 0$ with $\mu \in \mathbb{R}^n$, σ tends to the zero points of $\tilde{\mu}(\sigma)$, which are not always solutions of $\check{\mu}(\sigma)$. The analytic curve $\Gamma(\sigma_{\text{init}})$ consists of several branches, one of the branches connects σ_{init} with one solution, and other branches connect the rest of the solutions in pairs, which is the parity arguement. The branch of $\Gamma(\sigma_{\text{init}})$ that connects σ_{init} starts from a known point infinitely high on the fiber $B(\sigma_{\text{init}})$, so we let path $\tilde{\Gamma}$ starts from σ_{init} and follows along $\Gamma(\sigma_{\text{init}})$ as its first segment. However, $\Gamma(\sigma_{\text{init}})$ may have singular points on it, and those are where it intersects with the singular manifold S. The singular manifold S is the manifold of real eigenvalues of $-J(\sigma)$. In this 2-dimensional case, S has a simple shape since $-J(\sigma)$ always has an eigenvalue $-k$ and another real eigenvalue, but S could have much more complex shapes in higher-dimensional cases where there could be complex eigenvalues. On a singular point of $\Gamma(\sigma_{\text{init}})$, the direction that $\mathbf{1}^\top \mu$ moves reverses. As is shown, singular points are branch points of $\Gamma(\sigma_{\text{init}})$ as $\mathbf{1}^\top \mu$ varies, and they are critical points of $\Gamma(\sigma_{\text{init}})$ in variable σ. Singular points must be avoided because path-following cannot go beyond them, and this is done by moving μ along the fiber by adding $\beta\sigma$. There are two paths $\tilde{\Gamma}_1$ and $\tilde{\Gamma}_2$ in this graph, representing singularity avoidance with a positive β and a negative β. Singularity avoidance produces a segment of a different $\Gamma(\sigma_{\text{init}})$ as the next segment of path $\tilde{\Gamma}$. At a suitable σ with a suitable β, the next segment no longer intersects with the singular manifold. Finally, after finitely many segments, $\tilde{\Gamma}$ leads to a solution of $\text{VI}(F, \Delta)$.

Second, the Hessian $J_G^\top J_G$ of regularized Newton (6.6) remains bounded as σ approaches 0, whereas the Hessian $J_G^\top \circ (\sigma/r) \circ J_G$ of regularized Newton (6.9) is more skewed. As σ approaches 0, some components of $J_G^\top \circ (\sigma/r) \circ J_G$ vanishes, but this is controlled by regularization parameter δ_H. As r approaches 0, some components of $J_G^\top \circ (\sigma/r) \circ J_G$ blows up, making step length to vanish in those components. Thus, regularized Newton (6.9) takes extra cautious steps as r approaches 0. This is a useful feature, because there is $\sigma \circ r = \mu$ along the path, and unlike σ approaching 0 but r is subject to a function, r approaching 0 can cause σ to blow up under iteration with fixed $\mu > 0$. Thus, regularized Newton equation (6.9) appears to be more stable approaching the boundary but overall slower than regularized Newton equation (6.6).

Table 2: Iteration numbers solving VI(F, Δ)

Dimension	Regularized Newton (6.6)	Regularized Newton (6.9)
$[3, 50, 3]$	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
$[400, 50, 400]$[1]	655 / 528	640 / 546
$[800, 50, 800]$[1]	645 / 550	667 / 582

[1] Each entry in these two rows is based on 100 VI instances, while entries in the remaining rows are based on 1000 VI instances.

[2] Each entry is of the form mean / median over the iteration numbers solving instances.

[3] $[n, 50, n]$ is the architecture of the neural network modeling F.

The algorithm is tested on randomly generated instances of VI(F, Δ) in different dimensions. Specifically, F is a neural network with a $[n, 50, n]$ architecture and tanh activation functions, initialized with random weights and biases. This neural network model is real analytic, and its derivative is computed via backpropagation [44]. The initial point σ_{init} is also randomly generated, and the required precision is set to 10^{-5}. We test the algorithm using both the two different regularized Newton correctors and record the number of iterations. For dimensions 3 to 200, we test 1000 instances, and for dimensions 400 and 800, we test 100 instances. The mean and median of the iteration numbers are reported in Table 7. It can be seen that the algorithm with regularized Newton corrector (6.9) is slower, but it is more stable since the mean and median iteration numbers are closer. Across all the 14400 test instances, the algorithm converges to a solution in every single instance, demonstrating its robustness in practice. Moreover, the iteration number does not increase too drastically with the dimension, illustrating the scalability of the framework.

8 Conclusion

In this paper, we present a path-following framework on the fixed-point bundle for general finite-dimensional variational inequalities. First, we approximately reduce general variational inequality VI(H, K) to variational inequality VI(F, Δ) with analytic function F and simplex domain Δ.

Then, we present three equivalent characterizations of VI(F, Δ). The linear programming form produces UKKT (4.2), which leads to solutions of VI(F, Δ) as $\mu \to 0$ with $\mu > 0$. The Brouwer fixed-point form is Brouwer function $(\hat{\sigma}, r, v) = M(\sigma, \mu)$, such that Brouwer fixed-point theorem applies to it to guarantee the existence of paths subject to UKKT (4.2). The mixed complementarity problem form is MCP (4.5), where the paths subject to UKKT (4.2) are special central paths of it.

Next, we formalize the solution space of UKKT (4.2) as the fixed-point bundle E and equivalently transform UKKT (4.2) to the fixed-point bundle equation $G(\sigma, \mu) = 0$. We introduce two sections of the fixed-point bundle, $\breve{\mu}(\sigma)$ measures the gap of VI(F, Δ), and $\tilde{\mu}(\sigma)$ derives matrix $J(\sigma)$ that characterizes the differential properties of the fixed-point bundle. Through matrix $J(\sigma) + (\mathbf{1}^\top \mu)I$, we formalize the singular manifold S of the fixed-point bundle as the manifold of real eigenvalues of $-J(\sigma)$. By analyzing the analytic curve $\Gamma(\sigma_{\text{init}})$, we establish the parity arguement for VI(F, Δ), which guarantees the oddness of the number of solutions of VI(F, Δ).

Finally, we arrive at our predictor-corrector path-following framework on the fixed-point bundle. Combining the analytic curve $\Gamma(\sigma_{\text{init}})$ and singularity avoidance along the fibers, we formalize the path we follow as $\tilde{\Gamma}$, which consists of finitely many segments of different $\Gamma(\sigma_{\text{init}})$. We show that there exists a compact neighborhood $\mathcal{N}_{\tilde{\Gamma}}$ of $\tilde{\Gamma}$ satisfying $\mathcal{N}_{\tilde{\Gamma}} \cap S = \varnothing$ for the algorithm to work. Then, we introduce standard Newton corrector (6.3) and prove the convergence result of the overall algorithm. Alternatively, we also introduce a gradient corrector (6.8). The standard Newton equation and the gradient equation lead to regularized Newton equation (6.6) and (6.9), which both work as effective correctors for practical use.

The final convergence result is proved with the standard Newton corrector (6.3). The algorithm achieves global linear convergence to nonsingular solutions of VI(F, Δ). For singular solutions, the algorithm retains global linear reduction before reaching a fixed precision measured by $\mathbf{1}^\top \mu = \zeta > 0$, but the convergence reduces to sublinear as required ζ tends to 0.

In experiment, the algorithm is tested on 14400 randomly generated VIs of up to 800 dimensions using both the two regularized Newton correctors. It works in every single case, and iteration number increases only mildly with the dimension.

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] Yu-Guan Hsieh, Franck Iutzeler, Jérôme Malick, and Panayotis Mertikopoulos. Explore aggressively, update conservatively: Stochastic extragradient methods with variable stepsize scaling. *Advances in Neural Information Processing Systems*, 33:16223–16234, 2020.

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

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

[31] Eugene L Allgower and Kurt Georg. *Introduction to numerical continuation methods*. SIAM, 2003.

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

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

[34] Efim M Bronshteyn and LD Ivanov. The approximation of convex sets by polyhedra. *Siberian Mathematical Journal*, 16(5):852–853, 1975.

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

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

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

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

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

[40] 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.

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

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

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

[44] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.