

The fixed-point bundle method over product-of-simplex domains arising from game equilibria

Hongbo Sun*
shb20@tsinghua.org.cn

Abstract

This paper extends the fixed-point bundle framework for finite-dimensional variational inequalities (VIs) from the simplex domain to the product-of-simplex domain, which is directly applicable to solving Nash equilibria. The fixed-point bundle for VIs on the product-of-simplex domain reveals a composite fiber bundle structure. The key innovation is to construct an equivalent VI on the simplex domain and establish the equivalence between the two fixed-point bundle frameworks via a fiber bundle isomorphism. Exploiting this geometric equivalence, the predictor-corrector path-following algorithm for the VI on the product-of-simplex domain is shown to inherit the convergence guarantee of the simplex-domain framework, namely, global convergence with linear gap reduction near solutions. Numerical experiments on 5600 randomly generated instances with dimensions ranging from 2-player 128-action to 128-player 2-action demonstrate robust performance. The algorithm converges in every tested instance.

Keywords: variational inequality, Nash equilibrium, interior point method, fiber bundle, path-following

MSCcodes: 49J40, 90C33, 90C51, 91A06

Contents

1	Introduction	2
2	Approximate reduction to analytic VI on simplex product	3
3	The fixed-point bundle of VI on simplex product	5
3.1	Generalizing the three equivalent characterizations	5
3.2	The composite fixed-point bundle	7
4	Equivalence between VI on simplex product and VI on simplex	9
4.1	Transforming $VI(F, \Delta_n^m)$ into $VI(\bar{F}, \Delta_{mn}^+)$	9
4.2	Isomorphism between fixed-point bundles	11
4.3	Equivalence between characterizations	13
4.4	Equivalence between differential properties	15
5	Path-following on the fixed-point bundles	18
5.1	Transforming the predictor-corrector framework	18
5.2	Transferring the convergence results	20
5.3	The complete path-following algorithm	22

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

6 Experiments	23
7 Conclusion	25

1 Introduction

Game equilibrium problems, introduced by Nash [1], are central in economics [2], engineering [3], artificial intelligence [4], etc. They describe noncooperative settings in which each agent selects a strategy to maximize its own utility given the strategies of others, and an equilibrium is a strategy profile from which no agent has a unilateral incentive to deviate. Under certain conditions, a game equilibrium problem can be equivalently reformulated as a variational inequality (VI) [5]. Let $K \subseteq \mathbb{R}^n$ be a nonempty compact convex set (a convex body) and $H : K \rightarrow \mathbb{R}^n$ be a continuous function. The VI problem $\text{VI}(H, K)$ asks for $x^* \in K$ such that

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

When K has a product structure such that $K = \prod_i K_i$, and H is a block-wise gradient field such that $H = [H_i] = [\nabla_{x_i} f_i]$, the problem $\text{VI}(H, K)$ is a game equilibrium problem with K_i as each player's strategy set and f_i as each player's payoff function.

In terms of iterative methods with asymptotic convergence, existing algorithms for game equilibrium problems are largely the same as those for VIs [6]. VI algorithms fall into several categories, each with its application to game equilibrium problems. Projection [7] and proximal point methods [8] reformulate the equilibrium conditions as a fixed-point equation and iterate projections onto the feasible set. Merit-function approaches [9] convert the game into an optimization problem via merit functions and apply gradient or Newton updates. Interior point methods [10] transform the game into a complementarity problem and follow a central path subject to the perturbed KKT system. Homotopy methods [11] deform a simple operator into the target game operator. Operator splitting methods [12] exploit decompositions such as sums of individual monotonic operators. Modern variants incorporate adaptive step sizes, inertial terms, or stochastic estimates for large-scale strategic interactions [13, 14].

This paper mainly deals with VIs with the product-of-simplex domain, which is the canonical domain for mixed strategies of finite normal-form games. We extend the fixed-point bundle framework in Paper 1 [15] for VIs on simplex domain to VIs on product-of-simplex domain. In Paper 1, the fixed-point bundle framework formulates the solution space of a VI on simplex domain as a fiber bundle, where starting point selection, path-following, and singularity avoidance are systematically integrated. The path-following algorithm applies to VIs with general continuous functions and general compact convex domains after approximate reduction, and it guarantees global convergence with linear gap reduction near solutions. This paper establishes an equivalence between the fixed-point bundle framework over product-of-simplex domain and that over simplex domain, showing that the path-following algorithm with product-of-simplex domain has the same convergence result as in Paper 1.

The proposed approach begins by approximately reducing the general $\text{VI}(H, K)$ to a smooth $\text{VI}(F, \Delta_n^m)$ with domain Δ_n^m (product of m n -dimensional simplices) and a real-analytic operator F , in correspondence to the VI with real-analytic operator and simplex domain that Paper 1 deals with.

The three equivalent characterizations in Paper 1 can be generalized to $\text{VI}(F, \Delta_n^m)$ via the decomposable structure of the gap function of $\text{VI}(F, \Delta_n^m)$. Then, we derive the fixed-point bundle of $\text{VI}(F, \Delta_n^m)$, which turns out to have a composite structure on top of the ordinary fiber bundle structure.

An operator $\bar{F}$ can be constructed based on the homeomorphism $\phi : \Delta_n^m \times \Delta_m^\circ \rightarrow \Delta_{mn}^+$, such that $\text{VI}(\bar{F}, \Delta_{mn}^+)$ and $\text{VI}(F, \Delta_n^m)$ have the same solutions. The central innovation of this paper is to establish the equivalence between the fixed-point bundle frameworks of $\text{VI}(F, \Delta_n^m)$ and $\text{VI}(\bar{F}, \Delta_{mn}^+)$ via a fiber bundle isomorphism. However, there is a gap: $\bar{F}$ is only defined on Δ_{mn}^+ , instead of being a smooth map on the closed simplex Δ_{mn} . Thus, we can only have the fixed-point bundle framework of $\text{VI}(\bar{F}, \Delta_{mn}^+)$ as a formality by substituting $\bar{F}$ into Paper 1, while the assertions concerning this framework do not directly transfer.

The predictor-corrector path-following algorithm for $\text{VI}(F, \Delta_n^m)$ and $\text{VI}(\bar{F}, \Delta_{mn}^+)$ turns out to fix the variable on a specific slice given by $w_{\text{init}} \in \Delta_m^\circ$. The map $\bar{F}$ restricted to this slice can be extended to a map $\tilde{F}$ real-analytic on the closed simplex Δ_{mn} , such that all the results in Paper 1 apply to $\text{VI}(\tilde{F}, \Delta_{mn})$. Then, transferring from $\text{VI}(\tilde{F}, \Delta_{mn})$ to $\text{VI}(\bar{F}, \Delta_{mn}^+)$ and to $\text{VI}(F, \Delta_n^m)$, the path-following algorithm for $\text{VI}(F, \Delta_n^m)$ has the same convergence guarantee as in Paper 1.

In experiments, numerical experiments on 5600 randomly generated instances with dimensions ranging from 2-player 128-action to 128-player 2-action show that the algorithm succeeds in every case.

The remainder of the paper is organized as follows. Section 2 details the approximation reductions. Section 3 generalizes the three equivalent characterizations and the fixed-point bundle. Section 4 constructs a VI on simplex and establishes its equivalence to the VI on product-of-simplex. Section 5 transforms the predictor-corrector scheme and transfers convergence. Section 6 presents experimental results. Concluding remarks appear in Section 7.

For vectors, we use $a \circ b$ for componentwise multiplication and a/b for componentwise division. Inequality $a \geq b$ is also interpreted componentwise. For matrix A , componentwise multiplication is interpreted as $A \circ b = \text{Adiag}(b)$ and $b \circ A = \text{diag}(b)A$. For tensors, $\mathbb{R}^{mn}$ is a vector space with dimension mn , $\mathbb{R}^{m \times n}$ is a tensor space with shape $m \times n$, and the operation $\text{reshape}(\cdot)$ always represents switching between these two shapes. For block stacking, $[a_i]$ represents concatenating small vectors a_i into a large vector, and $\text{Diag}[x_i I]$ represents a block diagonal matrix. The operation $\text{norm}([x_i]) = [x_i]/(\sum_j x_j)$ denotes normalization. The operation $w \odot \pi$ is defined by $w \in \mathbb{R}^m$, $\pi \in \mathbb{R}^{m \times n}$, $w \odot \pi \in \mathbb{R}^{m \times n}$, and $(w \odot \pi)_i = w_i \pi_i$. Finally, $\mathbf{1}$ is the all-one vector; it sometimes has dimension n , sometimes dimension mn , but its dimension can always be inferred from context.

2 Approximate reduction to analytic VI on simplex product

Similar to Paper 1, we first introduce the approximate reduction from general $\text{VI}(H, K)$ to $\text{VI}(F, \Delta_n^m)$.

Approximating convex bodies via simplex product

In Paper 1, the fixed-point bundle method applies to VIs on general convex bodies based on the fact that any convex body can be inner-approximated arbitrarily well by polytopes, which are linear transformations of simplices. In fact, polytopes and simplices are in one-to-one correspondence via linear transformations when they have the same number of vertices. The linear transformation $x = [v_1, \dots, v_n] \sigma$ transforms a simplex Δ to a polytope with vertices $v_1, \dots, v_n$. In this paper, we need to approximate convex bodies with linear transformations of a simplex product, which yield the Minkowski sum of polytopes. The linear transformation

$$x = V \text{reshape}(\pi) = \sum_{i=1}^m V_i \pi_i = \sum_{i=1}^m [v_{i1}, \dots, v_{in}] \pi_i$$

transforms a simplex product Δ_n^m into a Minkowski sum of m polytopes, where each summand has vertices $v_{i1}, \dots, v_{in}$.

The Minkowski sum of polytopes is still a polytope, and if each summand polytope has n_i vertices, the resulting polytope can have at most $\prod_i n_i$ vertices. For instance, each simplex Δ_n has n vertices, and their Cartesian product Δ_n^m , which is a Minkowski sum in higher-dimensional space, has n^m vertices. Thus, the Minkowski sum of polytopes can potentially lead to an enormous compression of the input vertex budget from $\prod_i n_i$ to $\sum_i n_i$, resulting in a dimension reduction for the fixed-point bundle method.

For inner approximation of general convex bodies, since the Minkowski sum of polytopes includes ordinary polytopes as single-summand cases, approximating convex bodies with a Minkowski sum of polytopes is trivially at least as good as approximating with an ordinary polytope. However, the caveat is that this is not true if a specific number of summands is required in the Minkowski sum. Even for a Minkowski sum of n_i -vertex polytopes, $\prod_i n_i$ is only an upper bound on its vertex number; its vertex number can still be less than $\sum_i n_i$.

There are indeed cases where approximation by Minkowski sums has a strict advantage. For d -dimensional convex bodies with some regularity condition, the approximation error using a polytope with n vertices is $O(n^{-2/(d-1)})$ [16]. Then, it can be shown that for a d -dimensional convex body $K = K_1 \times \cdots \times K_m$ that is the Cartesian product of d_i -dimensional convex bodies K_i , the Minkowski sum of individual polytopes approximating K_i has strictly smaller approximation error than directly approximating K with ordinary polytopes under the same input vertex budget.

In this paper, we stop at the trivial guarantee that approximation with a linear image of a simplex product is at least as good as approximation with a linear image of a simplex. We do not study for which kinds of convex bodies the former has a strict advantage, and we do not develop a mechanism for discovering and exploiting nontrivial decomposition structures of a target convex body.

Approximating continuous functions via analytic functions on simplex product

The linear transformation $x = V\text{reshape}(\pi)$ transforms the VI as follows.

$$\langle H(x), x' - x \rangle = \sum_i \langle H_i(\sum_i V_i \pi_i), V_i \pi'_i - V_i \pi_i \rangle = \sum_i \left\langle V_i^\top H_i(\sum_i V_i \pi_i), \pi'_i - \pi_i \right\rangle \geq 0.$$

Similar to Paper 1, we then smooth the continuous mapping $F(\pi) = \text{reshape}(V^\top H(V\text{reshape}(\pi)))$ using Dirichlet distribution averaging. The Dirichlet distribution $\text{Dir}(\alpha)$ on the simplex Δ is given by

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

where $y \in \Delta$ is the variable, $\alpha > 0$ is a vector of concentration parameters, and $B(\alpha)$ is the Beta function. We need a $\text{Dir}(\alpha_i)$ for each $\pi_i \in \Delta_n$, so the map $F(\pi)$ is actually smoothed with the product Dirichlet distribution $\prod_i \text{Dir}(\alpha_i)$. For each component $F_{i,a}$ of F , we define the following operation.

$$\begin{aligned} (T_j F_{i,a})(\pi_1, \dots, \pi_m) &= \mathbb{E}_{Y_j \sim \text{Dir}(\alpha_j)} [F_{i,a}(\pi_1, \dots, Y_j, \dots, \pi_m)] \\ &= \int_{\Delta_n} F_{i,a}(\pi_1, \dots, y_j, \dots, \pi_m) \text{Dir}(y_j; \alpha_j) dy_j \end{aligned} \quad (2.2)$$

Then, the product Dirichlet distribution smoothing of a continuous map $F : \Delta \rightarrow \mathbb{R}^n$ is the map $F_\epsilon : \Delta \rightarrow \mathbb{R}^n$ given componentwise by

$$F_{\epsilon,i,a}(\pi) = \mathbb{E}_{Y \sim \prod_{j=1}^m \text{Dir}(\pi_j/\epsilon + c_j)} [F_{i,a}(Y)] = (T_1 \circ \cdots \circ T_m F_{i,a})(\pi), \quad (2.3)$$

where $\epsilon > 0$ is a scalar that measures approximation error, and $c_i > 0$ is a constant vector ensuring $\pi_i/\epsilon + c_i > 0$ for all $\pi_i \in \Delta_n$.

Following Paper 1, we then prove that $F_\epsilon(\pi)$ is real-analytic on Δ_n^m . Instead of proving it from scratch, we reuse the analyticity result for single-variable Dirichlet distribution averaging in Paper 1. Then, Hartogs' theorem on separate holomorphy [17] from complex analysis can be used to extend this to analyticity for multiple-variable Dirichlet distribution averaging. This theorem states that for a function $g : U \times V \rightarrow \mathbb{C}$ on $U \subset \mathbb{C}^m$ and $V \subset \mathbb{C}^n$, if for every fixed $u \in U$, the map $v \mapsto g(u, v)$ is holomorphic on V , and for every fixed $v \in V$, the map $u \mapsto g(u, v)$ is holomorphic on U , then g is holomorphic on $U \times V$.

Proposition 1. *For every $\epsilon > 0$, $F_\epsilon(\pi)$ is real-analytic on Δ_n^m .*

Proof. In the proof of Proposition 1, it is shown that if $f(\sigma)$ is continuous on the simplex Δ , then for every $\sigma \in \Delta$, $(Tf)(z)$ is holomorphic on the complex neighborhood $U_\sigma = \{z \in \mathbb{C}^n \mid \|z - \sigma\| \leq d\}$.

In our case, for every $\pi \in \Delta_n^m$, we study $(T_1 \circ \dots \circ T_m F_{i,a})(z_1, \dots, z_m)$ on the product complex neighborhood $U_\pi = \prod_j U_{\pi_j}$ with $U_{\pi_j} = \{z_j \in \mathbb{C}^n \mid \|z_j - \pi_j\| \leq d_j\}$. We denote $(y'_k, y_{\hat{k}}) := (y_1, \dots, y_{k-1}, y'_k, y_{k+1}, \dots, y_m)$ and $T_{\hat{k}} := T_1 \circ \dots \circ T_{k-1} \circ T_{k+1} \circ \dots \circ T_m$.

First, the operators T_j are commutative. The operation $(T_1 \circ \dots \circ T_m F_{i,a})(z_1, \dots, z_m)$ is an integral of the continuous $F_{i,a}(y_1, \dots, y_m)$ over the compact region Δ_n^m against the product measure $\prod_{j=1}^m \text{Dir}(y_j; z_j/\epsilon + c_j) dy_j$. By the Fubini-Tonelli theorem [18], if a multiple integral integrates a continuous integrand over a compact region, the order of integration can be interchanged. Thus, for every index k , we can denote $F_{\epsilon,i,a} = T_k \circ T_{\hat{k}} F_{i,a}$.

Next, we prove separate holomorphy for every index k . Since $F_{i,a}(y)$ is continuous on the compact set Δ_n^m , it is uniformly continuous, i.e., for any $\epsilon > 0$ there exists $\delta > 0$ such that whenever $\|y_k - y'_k\| < \delta$, we have $|F_{i,a}(y'_k, y_{\hat{k}}) - F_{i,a}(y_k, y_{\hat{k}})| < \epsilon$. Then, given fixed $z_j \in U_{\pi_j}$ for $j \neq k$, we have

$$\begin{aligned} |(T_{\hat{k}} F_{i,a})(y_k, z_{\hat{k}}) - (T_{\hat{k}} F_{i,a})(y'_k, z_{\hat{k}})| &\leq \int_{\Delta_n^{m-1}} |F_{i,a}(y'_k, y_{\hat{k}}) - F_{i,a}(y_k, y_{\hat{k}})| \prod_{j \neq k} \text{Dir}(y_j; z_j/\epsilon + c_j) dy_j \\ &< \epsilon \int_{\Delta_n^{m-1}} \prod_{j \neq k} \text{Dir}(y_j; z_j/\epsilon + c_j) dy_j = \epsilon \end{aligned}$$

Thus, $(T_{\hat{k}} F_{i,a})(y_k, z_{\hat{k}})$ is (uniformly) continuous in y_k on Δ . Then, the analyticity result for single-variable Dirichlet distribution averaging in the proof of Proposition 1 in Paper 1 applies to $F_{\epsilon,i,a} = T_k(T_{\hat{k}} F_{i,a})$. We obtain that for every fixed $z_j \in U_{\pi_j}$ for $j \neq k$, the map $z_k \mapsto F_{\epsilon,i,a}(z_k, z_{\hat{k}})$ is holomorphic in z_k on U_{π_k} .

Then, Hartogs' theorem on separate holomorphy applies, and we obtain that $F_{\epsilon,i,a}(z)$ is holomorphic in z on U_π . Since $\pi \in \Delta_n^m$ is arbitrary, it follows that $F_{\epsilon,i,a}(\pi)$ is real-analytic in π on Δ_n^m . The vector function $F_\epsilon(\pi)$ is real-analytic in π on Δ_n^m because each of its components indexed by (i, a) is real-analytic. $\square$

3 The fixed-point bundle of VI on simplex product

3.1 Generalizing the three equivalent characterizations

As in Paper 1, we deal with $\text{VI}(F, \Delta_n^m)$ assuming F is real-analytic on Δ_n^m throughout the following developments. In Paper 1, the analysis is built on the three equivalent characterizations of $\text{VI}(\bar{F}, \Delta)$. Thus, we begin by transforming these three equivalent characterizations to characterize $\text{VI}(F, \Delta_n^m)$. First, we show that $\text{VI}(F, \Delta_n^m)$ can be decomposed into several VIs on Δ_n via its gap function.

Theorem 2. For every $\pi \in \Delta_n^m$, we have

$$\text{gap}(\pi) = \sum_i \text{gap}_i(\pi_i, \pi_i), \quad \text{where } \text{gap}_i(\pi_i, \pi_i) := \sup_{\pi'_i \in \Delta_n} \langle F_i(\pi_i, \pi_i), \pi_i - \pi'_i \rangle \quad (3.1)$$

Consequently, for every $\pi \in \Delta_n^m$, $\text{gap}(\pi) = 0$ if and only if $\text{gap}_i(\pi_i, \pi_i) = 0$ for every index i .

Proof. First, we derive $\text{gap}(\pi)$ as follows.

$$\begin{aligned} \text{gap}(\pi) &= \sup_{\pi' \in \Delta_n^m} \langle F(\pi), \pi - \pi' \rangle = \sum_i \left(\pi_i^\top F_i(\pi) - \min_a F_{i,a}(\pi) \right) \\ &= \sum_i \sup_{\pi'_i \in \Delta_n} \langle F_i(\pi), \pi_i - \pi'_i \rangle = \sum_i \text{gap}_i(\pi_i, \pi_i) \end{aligned}$$

Thus, we obtain $\text{gap}(\pi) = \sum_i \text{gap}_i(\pi_i, \pi_i)$.

For every $\pi \in \Delta_n^m$, we have $\text{gap}(\pi) \geq 0$ and $\text{gap}_i(\pi_i, \pi_i) \geq 0$ by definition. Then, by $\text{gap}(\pi) = \sum_i \text{gap}_i(\pi_i, \pi_i)$, we have $\text{gap}(\pi) = 0$ if and only if $\text{gap}_i(\pi_i, \pi_i) = 0$ for every index i . $\square$

In Theorem 2, $\text{gap}_i(\pi_i, \pi_i)$ can be viewed as the gap function of a parametric variational inequality $\text{VI}(F_i(\cdot, \pi_i), \Delta_n)$ with π_i as the parameter. The theorem implies that π is a solution of $\text{VI}(F, \Delta_n^m)$ if and only if π_i is a solution of $\text{VI}(F_i(\cdot, \pi_i), \Delta_n)$ for each index i .

Since $\text{VI}(F_i(\cdot, \pi_i), \Delta_n)$ is a VI on simplex, the three equivalent characterizations in Paper 1 apply to it. However, they are compatible over index i only pointwise for individual $\pi \in \Delta_n^m$, at which point all the parameters π_i and variables π_i across index i come from the same π . In Paper 1, the three equivalent characterizations are indeed pointwise. Their definitions consist only of functions, equalities, and inequalities that are essentially pointwise objects assigning values to each individual (π_i, τ_i) . The equivalence is also in pointwise form, stating that for each individual (π_i, τ_i) , the properties of satisfying the three characterizations are equivalent. Thus, stacking the three equivalent characterizations of $\text{VI}(F_i(\cdot, \pi_i), \Delta_n)$ over index i at each individual $\pi \in \Delta_n^m$, we obtain the generalization characterizing $\text{VI}(F, \Delta_n^m)$. We briefly go through them since the generalization merely adds a subscript i to those in Paper 1.

Linear programming form. Equation (3.2) is the fixed-point KKT system (FPKKT). It consists of the perturbed KKT conditions (3.2a) of a parametric linear programming problem and the fixed-point condition $\hat{\pi} = \pi$, where τ is the barrier parameter.

$$\begin{bmatrix} \hat{\pi}_i \circ r_i - \tau_i \\ r_i - F_i(\pi) - v_i \mathbf{1} \\ \mathbf{1}^\top \hat{\pi}_i - \mathbf{1} \end{bmatrix} = 0 \quad (3.2a)$$

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

Brouwer fixed-point form. Equation (3.3) is the Brouwer function. It is given by the correspondence from (π, τ) to $(\hat{\pi}, r, v)$ subject to perturbed KKT conditions (3.2a). It can be shown that this correspondence is a continuous function such that the Brouwer fixed-point theorem applies to it.

$$(\hat{\pi}, r, v) = M(\pi, \tau), \quad \text{s.t. perturbed KKT conditions (3.2a)}. \quad (3.3)$$

The Brouwer function is computed by solving the bisection problem (3.4) for every i .

$$q_i(v_i) = \mathbf{1}^\top \frac{\tau_i}{F_i(\pi) + v_i \mathbf{1}} - 1, \quad v_i \in [-\min_a F_{i,a}(\pi), -\min_a F_{i,a}(\pi) + \mathbf{1}^\top \tau_i] \quad (3.4)$$

Mixed complementarity problem form. Equation (3.5) is the (barriered) mixed complementarity problem (MCP), which comes from FPKKT (3.2) by making the complementarity term $\pi_i \circ r_i$ the objective function. The $\tau = 0$ case is the original problem, and the $\tau > 0$ case is its barrier problem.

$$\begin{aligned} \min_{(\pi, r, v)} \quad & \sum_i \pi_i^\top r_i - \tau_i^\top \ln \pi_i - \tau_i^\top \ln r_i \\ \text{s.t.} \quad & r_i = F_i(\pi) + v_i \mathbf{1} \\ & \mathbf{1}^\top \pi_i = 1 \\ & (\pi_i, r_i) \geq 0 \end{aligned} \tag{3.5}$$

Equivalence. The three characterizations are equivalent in the sense that for each variable-barrier pair (π, τ) , the properties of satisfying the three characterizations are equivalent.

Theorem 3. For $(\pi, \tau) \in \Delta_n^m \times \{\tau \in \mathbb{R}^{m \times n} | \tau > 0\}$, the following statements are equivalent.

- (i) (π, τ) satisfies FPKKT (3.2) for some r, v .
- (ii) π is a fixed point such that $\hat{\pi} = \pi$ of the Brouwer function $(\hat{\pi}, r, v) = M(\pi, \tau)$.
- (iii) (π, τ) satisfies $\tilde{\pi} = \pi$ and the perturbed KKT conditions of MCP (3.5), where $\tilde{\pi}$ is the Lagrangian multiplier associated with the constraint $r \geq 0$.

Analogous to Paper 1, FPKKT (3.2) characterizes paths leading to solutions of $\text{VI}(F, \Delta_n^m)$, the Brouwer function $(\hat{\pi}, r, v) = M(\pi, \tau)$ guarantees the existence of paths subject to FPKKT (3.2) via the Brouwer fixed-point theorem, and MCP (3.5) characterizes the paths as its special central paths. It can be shown that these statements indeed hold for $\text{VI}(F, \Delta_n^m)$, but we only use the formality and the equivalence of the three characterizations. In this paper, we intend to reuse the theoretical guarantees in Paper 1 by establishing an equivalence between $\text{VI}(F, \Delta_n^m)$ and some $\text{VI}(\bar{F}, \Delta)$ to which Paper 1 applies, instead of directly transferring the theoretical guarantees onto $\text{VI}(F, \Delta_n^m)$.

3.2 The composite fixed-point bundle

As in Paper 1, the solution space of FPKKT (3.2) can be constructed as a fiber bundle called the fixed-point bundle, which is the central geometric object of the path-following framework.

For every $\pi \in \Delta_n^m$, the admissible τ value satisfying FPKKT (3.2) can be expressed as

$$\tau_i = \pi_i \circ (F_i(\pi) + v_i \mathbf{1}) = \pi_i \circ F_i(\pi) - (\pi_i^\top F_i(\pi)) \pi_i + (v_i + \pi_i^\top F_i(\pi)) \pi_i.$$

Then, letting $\tilde{\tau}_i(\pi) := \pi_i \circ F_i(\pi) - (\pi_i^\top F_i(\pi)) \pi_i$ and $\bar{v} w_i := v_i + \pi_i^\top F_i(\pi)$, part of the solution space of FPKKT (3.2) has the structure shown in equation (3.6).

$$\begin{aligned} E &= \bigcup_{\pi \in \Delta_n^m} \{\pi\} \times B(\pi) \\ B(\pi) &= \bigcup_{w \in \Delta_m^\circ} \tilde{B}(\pi, w) \\ \tilde{B}(\pi, w) &= \{\tilde{\tau}(\pi) + \bar{v} w \odot \pi | \bar{v} \in \mathbb{R} \setminus \{0\}\} \end{aligned} \tag{3.6}$$

Note that E is not the entire solution space of FPKKT (3.2); it only contains τ such that $\mathbf{1}^\top \tau_i = \bar{v} w_i$ with $\bar{v} \neq 0$ and w ranging over Δ_m° , which indicates that the $\mathbf{1}^\top \tau_i$ have the same sign for all i . We denote this set as

$$\mathbb{R}_\pm^{m \times n} := \{\tau \in \mathbb{R}^{m \times n} | (\mathbf{1}^\top \tau_i)(\mathbf{1}^\top \tau_j) > 0, \forall i, j\},$$

then for every $(\pi, \tau) \in E$, we have $\tau \in \mathbb{R}_{\pm}^{m \times n}$.

As equation (3.6) shows, E is a fiber bundle with projection $\alpha : E \rightarrow \Delta_n^m$ onto its base space Δ_n^m such that $\alpha((\pi, \tau)) = \pi$, where the fiber over $\pi \in \Delta_n^m$ is $\alpha^{-1}(\pi) = \{\pi\} \times B(\pi)$. This resembles the fiber bundle structure in Paper 1. In our case, equation (3.6) additionally shows a composite fiber bundle structure, such that each fiber $B(\pi)$ is itself a fiber bundle.

Lemma 4. *For any $\pi \in \Delta_n^m$ and $w_1, w_2 \in \Delta_m^\circ$, if $w_1 \neq w_2$ then $\tilde{B}(\pi, w_1) \cap \tilde{B}(\pi, w_2) = \emptyset$.*

Proof. Suppose for (π, w_1) and (π, w_2) , there exist $\bar{v}_1 \neq 0$ and $\bar{v}_2 \neq 0$ such that $\pi_i \circ F_i(\pi) + \bar{v}_1 w_{1,i} \pi_i = \pi_i \circ F_i(\pi) + \bar{v}_2 w_{2,i} \pi_i$ for every i . Then $\bar{v}_1 w_{1,i} \pi_i = \bar{v}_2 w_{2,i} \pi_i$. Multiplying by $\mathbf{1}^\top$ and summing over i , we have $\bar{v}_1 w_{1,i} = \bar{v}_2 w_{2,i}$ and $\bar{v}_1 = \bar{v}_2$. Since $\bar{v}_1 \neq 0$ and $\bar{v}_2 \neq 0$, we have $w_1 = w_2$. By contraposition, if $w_1 \neq w_2$, we have $\tilde{B}(\pi, w_1) \cap \tilde{B}(\pi, w_2) = \emptyset$. $\square$

Lemma 4 shows that for every π , $B(\pi)$ is indeed a fiber bundle. Given that $\bar{v} = 0$ is excluded from $\tilde{B}(\pi, w)$, $B(\pi)$ is the disjoint union of $\tilde{B}(\pi, w)$ over $w \in \Delta_m^\circ$, where $\tilde{B}(\pi, w)$ are the fibers of $B(\pi)$. If $\bar{v} = 0$ were not excluded, all the $\tilde{B}(\pi, w)$ over $w \in \Delta_m^\circ$ would share a common point $\tilde{\tau}(\pi)$. Then, $\tilde{B}(\pi, w)$ would no longer be fibers of $B(\pi)$, but $B(\pi)$ could still be fibers of E .

Therefore, E in equation (3.6) is a composite fiber bundle as equation (3.7) shows. We refer to this as the composite fixed-point bundle.

$$\begin{aligned} E &\xrightarrow{\alpha_1} \Delta_n^m \times \Delta_m^\circ \xrightarrow{\alpha_2} \Delta_n^m \\ \alpha_1((\pi, \tau)) &= (\pi, \text{norm}([\mathbf{1}^\top \tau_i])) \\ \alpha_2((\pi, w)) &= \pi \end{aligned} \tag{3.7}$$

There are three fiber bundles in the composite fixed-point bundle $E \rightarrow \Delta_n^m \times \Delta_m^\circ \rightarrow \Delta_n^m$.

- Bundle $E \rightarrow \Delta_n^m \times \Delta_m^\circ$: Each fiber $\{\pi\} \times \tilde{B}(\pi, w)$ is a one-dimensional affine line spanned by $\bar{v} \in \mathbb{R}$ excluding a single point $(\pi, \tilde{\tau}(\pi))$. The projection map α_1 projects each $(\pi, \tau) \in E$ to $(\pi, \text{norm}([\mathbf{1}^\top \tau_i])) \in \Delta_n^m \times \Delta_m^\circ$.
- Bundle $\Delta_n^m \times \Delta_m^\circ \rightarrow \Delta_n^m$: Each fiber Δ_m° is the interior of a simplex. The projection map α_2 projects each $(\pi, w) \in \Delta_n^m \times \Delta_m^\circ$ to $\pi \in \Delta_n^m$.
- Bundle $E \rightarrow \Delta_n^m$: Each fiber $\{\pi\} \times B(\pi)$ is an m -dimensional affine cone spanned by $\bar{v}w$ with $w \in \Delta_m^\circ$ and $\bar{v} \in \mathbb{R}$ excluding a single point $(\pi, \tilde{\tau}(\pi))$. The projection map $\alpha_2 \circ \alpha_1$ projects each $(\pi, \tau) \in E$ to $\pi \in \Delta_n^m$.

In short, $E \rightarrow \Delta_n^m \times \Delta_m^\circ$ has a large base space and small fibers, whereas $E \rightarrow \Delta_n^m$ has a small base space and large fibers. Each fiber $\{\pi\} \times B(\pi)$ of $E \rightarrow \Delta_n^m$ is the union of all the fibers $\{\pi\} \times \tilde{B}(\pi, w)$ of $E \rightarrow \Delta_n^m \times \Delta_m^\circ$ over $w \in \Delta_m^\circ$.

As in Paper 1, we then derive the fixed-point bundle equation and define two sections of the fiber bundle. The fixed-point bundle equation $G(\pi, \tau) = 0$ is obtained by eliminating r and v in FPKKT (3.2) with the projection matrix $I - \pi_i \mathbf{1}^\top$, where $G(\pi, \tau)$ is as equation (3.8) shows.

$$G_i(\pi, \tau) = (I - \pi_i \mathbf{1}^\top) (\pi_i \circ F_i(\pi) - \tau_i), \quad (\pi, \tau) \in \Delta_n^m \times \mathbb{R}_{\pm}^{m \times n} \tag{3.8}$$

The two sections $\check{\tau}(\pi)$ and $\hat{\tau}(\pi)$ are sections of the bundle $E \cup \{(\pi, \tilde{\tau}(\pi)) | \pi \in \Delta_n^m\} \rightarrow \Delta_n^m$. The bundle $E \cup \{(\pi, \tilde{\tau}(\pi)) | \pi \in \Delta_n^m\} \rightarrow \Delta_n^m$ adds back the point excluded by $\bar{v} \neq 0$ in each fiber of $E \rightarrow \Delta_n^m$. Section $\hat{\tau}(\pi)$ satisfies $\hat{\tau}(\pi) \geq 0$, and it characterizes the gap of $\text{VI}(F, \Delta_n^m)$ such that

$\text{gap}_i(\pi_i, \pi_i) = \mathbf{1}^\top \tilde{\tau}_i(\pi)$. Section $\tilde{\tau}(\pi)$ satisfies $\mathbf{1}^\top \tilde{\tau}_i(\pi) = 0$, and it governs the differential properties of the composite fixed-point bundle as will be discussed later.

$$\begin{aligned}\tilde{\tau}_i(\pi) &= \pi_i \circ \left(F_i(\pi) - \left(\min_a F_{i,a}(\pi) \right) \mathbf{1}_a \right) \\ \tilde{\tau}_i(\pi) &= \pi_i \circ \left(F_i(\pi) - \left(\pi_i^\top F_i(\pi) \right) \mathbf{1} \right) = \left(I - \pi_i \mathbf{1}^\top \right) (\pi_i \circ F_i(\pi))\end{aligned}\tag{3.9}$$

4 Equivalence between VI on simplex product and VI on simplex

The equivalence between $\text{VI}(F, \Delta_n^m)$ and $\text{VI}(F_i(\cdot, \pi_i), \Delta_n)$ in the last section relies on a fixed parameter π_i , so the equivalence only holds pointwise for each individual π , and does not apply when π varies or in global situations. In this section, we construct a $\text{VI}(\bar{F}, \Delta_{mn}^+)$ preserving the solutions of $\text{VI}(F, \Delta_n^m)$, and establish the equivalence between their fixed-point bundle frameworks.

4.1 Transforming $\text{VI}(F, \Delta_n^m)$ into $\text{VI}(\bar{F}, \Delta_{mn}^+)$

We begin by constructing a VI on simplex preserving the solutions of $\text{VI}(F, \Delta_n^m)$. First, we lift $\text{VI}(F, \Delta_n^m)$ to a VI defined on $\Delta_n^m \times \Delta_m^\circ$ with operator $\tilde{F}(\pi, w) = (F(\pi), 0)$. Then, the gap function of $\text{VI}(\tilde{F}, \Delta_n^m \times \Delta_m^\circ)$ is derived as follows.

$$\begin{aligned}\text{gap}_{\tilde{F}}(\pi, w) &= \sup_{(\pi', w') \in \Delta_n^m \times \Delta_m^\circ} \langle (F(\pi), 0), (\pi', w') - (\pi, w) \rangle \\ &= \sup_{\pi' \in \Delta_n^m} \langle F(\pi), \pi' - \pi \rangle = \text{gap}(\pi)\end{aligned}$$

Thus, for any $w^* \in \Delta_m^\circ$, (π^*, w^*) is a solution of $\text{VI}(\tilde{F}, \Delta_n^m \times \Delta_m^\circ)$ if and only if π^* is a solution of $\text{VI}(F, \Delta_n^m)$.

Viewing π_i as conditional distributions and w as a marginal distribution, we can transform (π, w) into a joint distribution σ that takes values on a simplex. In fact, this is a real-analytic homeomorphism $\phi : \Delta_n^m \times \Delta_m^\circ \rightarrow \Delta_{mn}^+$ defined as follows.

$$\begin{aligned}\phi(\pi, w) &= \sigma, \quad \text{s.t. } \sigma_{i,a} = w_i \pi_{i,a} \\ \phi^{-1}(\sigma) &= (\pi, w), \quad \text{s.t. } (\pi_{i,a}, w_i) = (\sigma_{i,a} / \sum_a \sigma_{i,a}, \sum_a \sigma_{i,a})\end{aligned}\tag{4.1}$$

We denote

$$\Delta_{mn}^+ = \{\sigma \in \Delta_{mn} \mid \sum_a \sigma_{i,a} > 0\}.$$

The map ϕ is a bijection between $\Delta_n^m \times \Delta_m^\circ$ and Δ_{mn}^+ because ϕ^{-1} as defined is its inverse. Both ϕ and ϕ^{-1} are real-analytic because they consist only of elementary operations and the denominator $\sum_a \sigma_{i,a} > 0$ on Δ_{mn}^+ . Thus, $\phi : \Delta_n^m \times \Delta_m^\circ \rightarrow \Delta_{mn}^+$ is a real-analytic homeomorphism.

Next, we can use the homeomorphism $\phi : \Delta_n^m \times \Delta_m^\circ \rightarrow \Delta_{mn}^+$ to transform $\text{VI}(\tilde{F}, \Delta_n^m \times \Delta_m^\circ)$ into a VI on Δ_{mn}^+ . This is performed using the standard procedure of changing the variable of a VI [19]. For a general variational inequality $\text{VI}(H, K)$, a standard theorem states that $x^* \in K$ solves $\text{VI}(H, K)$ if and only if $-H(x^*) \in N_K(x^*)$, where $N_K(x) = \{v \in \mathbb{R}^n \mid \langle v, x' - x \rangle \leq 0, \forall x' \in K\}$ is the normal cone. Another standard theorem on change of variables for normal cones states that if $z = \psi(x)$ is a diffeomorphism between K and $\psi(K)$, then $N_K(x) = (\partial\psi(x)/\partial x)^\top N_{\psi(K)}(\psi(x))$. Thus, if we define $\bar{H}(z) = (\partial\psi(x)/\partial x)^{-\top} H(x)$ s.t. $x = \psi(z)$, then $x^* \in K$ is a solution of $\text{VI}(H, K)$ if and only if $z^* = \psi(x^*)$ is a solution of $\text{VI}(\bar{H}, \psi(K))$, which is a transformation obtained by change of variables.

Note that $\bar{H}(z) = (\partial\psi(x)/\partial x)^{-\top} H(x)$ is equivalent to $\langle \bar{H}(z), dz \rangle = \langle H(x), dx \rangle$ given $z = \psi(x)$ and $dz = (\partial\psi(x)/\partial x)dx$. Then, we can use this equation to find $\bar{F}(\sigma)$, changing the variable of $\text{VI}(\check{F}, \Delta_n^m \times \Delta_m^\circ)$ based on $\sigma = \phi(\pi, w)$, which is derived as follows.

$$\begin{aligned} \langle \bar{F}(\sigma), d\sigma \rangle &= \langle \check{F}(\pi, w), d(\pi, w) \rangle \\ \sum_i \langle \bar{F}_i(\sigma), w_i d\pi_i + \pi_i dw_i \rangle &= \sum_i \langle F_i(\pi), d\pi_i \rangle \\ \sum_i \langle w_i \bar{F}_i(\sigma), d\pi_i \rangle + \sum_i (\pi_i^\top \bar{F}_i(\sigma)) dw_i &= \sum_i \langle F_i(\pi), d\pi_i \rangle \end{aligned}$$

The coefficients of $d\pi_i$ and dw_i must be identically 0 for the equality to hold. Then, for every index i , $w_i \bar{F}_i(\sigma) - F_i(\pi) = k\mathbf{1}$ and $\pi_i^\top \bar{F}_i(\sigma) = 0$. Therefore, we arrive at the function $\bar{F}(\sigma)$ in equation (4.2), and $\text{VI}(\bar{F}, \Delta_{mn}^+)$ is the transformation of $\text{VI}(\check{F}, \Delta_n^m \times \Delta_m^\circ)$ by changing the variable via $\sigma = \phi(\pi, w)$.

$$\bar{F}_i(\sigma) = \frac{1}{w_i} \left(F_i(\pi) - (\pi_i^\top F_i(\pi))\mathbf{1} \right), \quad \text{s.t. } (\pi, w) = \phi^{-1}(\sigma). \quad (4.2)$$

In fact, we can establish the equivalence between $\text{VI}(\bar{F}, \Delta_{mn}^+)$ and $\text{VI}(F, \Delta_n^m)$ directly, without going through the two-step construction of lifting and change of variables.

Theorem 5. *The gap functions of $\text{VI}(\bar{F}, \Delta_{mn}^+)$ and $\text{VI}(F, \Delta_n^m)$ satisfy the following equation.*

$$\text{gap}_{\bar{F}}(\sigma) = \max_i \frac{\text{gap}_i(\pi_i, \pi_i)}{w_i}, \quad \text{s.t. } \sigma = \phi(\pi, w) \quad (4.3)$$

Consequently, given $\sigma = \phi(\pi, w)$, $\text{gap}_{\bar{F}}(\sigma) = 0$ if and only if $\text{gap}(\pi) = 0$.

Proof. Using $(\pi, w) = \phi^{-1}(\sigma)$, we have the following deduction.

$$\begin{aligned} \langle \bar{F}(\sigma), \sigma' - \sigma \rangle &= \sum_i \left\langle \frac{1}{w_i} \left(F_i(\pi) - (\pi_i^\top F_i(\pi))\mathbf{1} \right), w'_i \pi'_i - w_i \pi_i \right\rangle \\ &= \sum_i \left(\left\langle \frac{1}{w_i} F_i(\pi), w'_i \pi'_i - w_i \pi_i \right\rangle - \frac{\pi_i^\top F_i(\pi)}{w_i} (w'_i - w_i) \right) = \sum_i \frac{w'_i}{w_i} \langle F_i(\pi), \pi'_i - \pi_i \rangle \end{aligned}$$

Then, the gap function of $\text{VI}(\bar{F}, \Delta_{mn}^+)$ is derived as follows. The second line follows from $w'_i/w_i > 0$ for every i .

$$\begin{aligned} \text{gap}_{\bar{F}}(\sigma) &= \sup_{\sigma' \in \Delta_{mn}^+} \langle \bar{F}(\sigma), \sigma' - \sigma \rangle = \sup_{(\pi', w') \in \Delta_n^m \times \Delta_m^\circ} \sum_i \frac{w'_i}{w_i} \langle F_i(\pi), \pi'_i - \pi_i \rangle \\ &= \sup_{w' \in \Delta_m^\circ} \sum_i \frac{w'_i}{w_i} \sup_{\pi'_i \in \Delta_n} \langle F_i(\pi), \pi'_i - \pi_i \rangle = \max_i \frac{\text{gap}_i(\pi_i, \pi_i)}{w_i} \end{aligned}$$

Since $\text{gap}_i(\pi_i, \pi_i) \geq 0$, $w_i > 0$, and $\text{gap}_{\bar{F}}(\sigma) \geq 0$, we have that $\text{gap}_{\bar{F}}(\sigma) = 0$ if and only if $\text{gap}_i(\pi_i, \pi_i) = 0$ for every i , which is equivalently $\text{gap}(\pi) = 0$ by Theorem 2. $\square$

Therefore, for any $w^* \in \Delta_m^\circ$, $\pi^* \in \Delta_n^m$ is a solution of $\text{VI}(F, \Delta_n^m)$ if and only if $\sigma^* = \phi(\pi^*, w^*)$ is a solution of $\text{VI}(\bar{F}, \Delta_{mn}^+)$. Thus, we have transformed $\text{VI}(F, \Delta_n^m)$ to $\text{VI}(\bar{F}, \Delta_{mn}^+)$ while preserving the solutions.

The transformation $\text{VI}(\bar{F}, \Delta_{mn}^+)$ is not a smooth VI on the closed simplex, because Δ_{mn}^+ excludes points where $\mathbf{1}^\top \sigma_i = 0$, and $\bar{F}(\sigma)$ blows up as σ approaches these points. Thus, results about the

fixed-point bundle framework in Paper 1 do not apply to $\text{VI}(\bar{F}, \Delta_{mn}^+)$. Nevertheless, we can still substitute $\bar{F}(\sigma)$ into the fixed-point bundle framework in Paper 1 as a mere formality, without invoking any theoretical guarantee provided by the theorems in Paper 1.

In the following developments of this section, we establish the equivalence between the formality-only fixed-point bundle framework of $\text{VI}(\bar{F}, \Delta_{mn}^+)$ and the composite fixed-point bundle framework of $\text{VI}(F, \Delta_n^m)$, while the final gap from $\text{VI}(\bar{F}, \Delta_{mn}^+)$ to an actual smooth VI on a closed simplex will be filled in the next section.

4.2 Isomorphism between fixed-point bundles

First, we establish the equivalence between the fixed-point bundles of $\text{VI}(F, \Delta_n^m)$ and $\text{VI}(\bar{F}, \Delta_{mn}^+)$. The fixed-point bundle of $\text{VI}(F, \Delta_n^m)$ is the composite fixed-point bundle $E \rightarrow \Delta_n^m \times \Delta_m^\circ \rightarrow \Delta_n^m$ defined in equation (3.6). From Paper 1, the fixed-point bundle of $\text{VI}(\bar{F}, \Delta_{mn}^+)$ has fibers $\bar{B}'(\sigma)$ derived as follows.

$$\begin{aligned} \bar{B}'(\sigma) &= \{ \sigma \circ \bar{F}(\sigma) + v\sigma \mid v \in \mathbb{R} \} \\ &= \left\{ \mu \mid \mu_i = w_i \pi_i \circ \frac{1}{w_i} \left(F_i(\pi) - (\pi_i^\top F_i(\pi)) \mathbf{1} \right) + v w_i \pi_i, (\pi, w) = \phi^{-1}(\sigma), v \in \mathbb{R} \right\} \\ &= \{ \text{reshape}(\tilde{\tau}(\pi) + \bar{v}w \odot \pi) \mid (\pi, w) = \phi^{-1}(\sigma), \bar{v} \in \mathbb{R} \} \end{aligned}$$

We exclude the point given by $\bar{v} = 0$ from each fiber $\bar{B}'(\sigma)$ in correspondence with the composite fixed-point bundle of $\text{VI}(F, \Delta_n^m)$. Then, the fixed-point bundle $\bar{E} \rightarrow \Delta_{mn}^+$ of $\text{VI}(\bar{F}, \Delta_{mn}^+)$ is as equation (4.4) shows.

$$\begin{aligned} \bar{E} &\xrightarrow{\bar{\alpha}} \Delta_{mn}^+ \\ \bar{\alpha}((\sigma, \mu)) &= \sigma \\ \bar{E} &= \bigcup_{\sigma \in \Delta_{mn}^+} \{ \sigma \} \times \bar{B}(\sigma) \\ \bar{B}(\sigma) &= \{ \text{reshape}(\tilde{\tau}(\pi) + \bar{v}w \odot \pi) \mid (\pi, w) = \phi^{-1}(\sigma), \bar{v} \in \mathbb{R} \setminus \{0\} \} \end{aligned} \tag{4.4}$$

Note that $\bar{B}(\sigma)$ consists of $\mu = \text{reshape}(\tilde{\tau}(\pi) + \bar{v}w \odot \pi)$, while $\tilde{B}(\pi, w)$ consists of $\tau = \tilde{\tau}(\pi) + \bar{v}w \odot \pi$. This indicates that the bundles $E \rightarrow \Delta_n^m \times \Delta_m^\circ$ and $\bar{E} \rightarrow \Delta_{mn}^+$ have very similar fibers. We can show that there is a homeomorphism between E and $\bar{E}$, given by Φ as equation (4.5) shows.

$$\begin{aligned} \Phi(\pi, \tau) &= (\sigma, \mu), \quad \text{s.t. } \sigma = \phi(\pi, \text{norm}([\mathbf{1}^\top \tau_i])), \mu_i = \tau_i \\ \Phi^{-1}(\sigma, \mu) &= (\pi, \tau), \quad \text{s.t. } \pi = \phi_\pi^{-1}(\sigma), \tau_i = \mu_i \end{aligned} \tag{4.5}$$

It can be verified that Φ and Φ^{-1} are indeed inverse maps of each other between E and $\bar{E}$. Then, Φ is a homeomorphism between E and $\bar{E}$ if it is continuous. However, we aim to prove further that Φ is a homeomorphism between larger sets containing E and $\bar{E}$. Recall that for every $(\pi, \tau) \in E$, we have $\tau \in \mathbb{R}_\pm^{m \times n}$. Since μ in $\bar{E}$ and τ in E are related by a reshape, we introduce another set

$$\mathbb{R}_\pm^{mn} := \{ \mu \in \mathbb{R}^{mn} \mid (\mathbf{1}^\top \mu_i)(\mathbf{1}^\top \mu_j) > 0, \forall i, j \} = \text{reshape}(\mathbb{R}_\pm^{m \times n}),$$

and study the homeomorphism Φ with $\tau \in \mathbb{R}_\pm^{m \times n}$ and $\mu \in \mathbb{R}_\pm^{mn}$.

Theorem 6. For $(\pi, \tau) \in \Delta_n^m \times \mathbb{R}_\pm^{m \times n}$ and $(\sigma, \mu) \in \Delta_{mn}^+ \times \mathbb{R}_\pm^{mn}$, the maps Φ and Φ^{-1} satisfy

$$\begin{aligned} (\Phi^{-1} \circ \Phi)(\pi, \tau) &= (\pi, \tau), \\ (\Phi \circ \Phi^{-1})(\sigma, \mu) &= \left(\text{reshape} \left(\text{norm}([\mathbf{1}^\top \mu_i]) \odot \phi_\pi^{-1}(\sigma) \right), \mu \right). \end{aligned} \tag{4.6}$$

Consequently, Φ is a real-analytic homeomorphism such that

$$\Delta_n^m \times \mathbb{R}_\pm^{m \times n} \cong (\Delta_{mn}^+ \times \mathbb{R}_\pm^{mn}) \cap \left\{ (\sigma, \mu) \mid \phi_w^{-1}(\sigma) = \text{norm}([\mathbf{1}^\top \mu_i]) \right\}. \quad (4.7)$$

Proof. For each $(\pi, \tau) \in \Delta_n^m \times \mathbb{R}_\pm^{m \times n}$, under the map $\Phi^{-1} \circ \Phi$, we have

$$\begin{aligned} (\pi, \tau) &\mapsto \left(\phi(\pi, \text{norm}([\mathbf{1}^\top \tau_i])), \text{reshape}(\tau) \right) \mapsto \left(\phi_\pi^{-1} \left(\phi(\pi, \text{norm}([\mathbf{1}^\top \tau_i])) \right), \text{reshape}(\text{reshape}(\tau)) \right) \\ &= (\pi, \tau). \end{aligned}$$

For each $(\sigma, \mu) \in \Delta_{mn}^+ \times \mathbb{R}_\pm^{mn}$, under the map $\Phi \circ \Phi^{-1}$, we have

$$\begin{aligned} (\sigma, \mu) &\mapsto (\phi_\pi^{-1}(\sigma), \text{reshape}(\mu)) \mapsto \left(\phi \left(\phi_\pi^{-1}(\sigma), \text{norm}([\mathbf{1}^\top \mu_i]) \right), \text{reshape}(\text{reshape}(\mu)) \right) \\ &= \left(\text{reshape} \left(\text{norm}([\mathbf{1}^\top \mu_i]) \odot \phi_\pi^{-1}(\sigma) \right), \mu \right). \end{aligned}$$

$\Phi^{-1} \circ \Phi$ is always an identity map, whereas $\Phi \circ \Phi^{-1}$ is an identity map if and only if $\text{norm}([\mathbf{1}^\top \mu_i]) = \phi_w^{-1}(\sigma)$. To prove that Φ is a bijection between the LHS (left-hand side) and RHS (right-hand side) in equation (4.7), we further need to show that $\Phi(\text{LHS}) \subset \text{RHS}$ and $\Phi^{-1}(\text{RHS}) \subset \text{LHS}$.

For every $(\sigma, \mu) = \Phi(\pi, \tau)$ with $(\pi, \tau) \in \Delta_n^m \times \mathbb{R}_\pm^{m \times n}$ (belonging to the LHS), by definition of Φ , we have

$$\phi_w^{-1}(\sigma) = \phi_w^{-1} \left(\phi(\pi, \text{norm}([\mathbf{1}^\top \tau_i])) \right) = \text{norm}([\mathbf{1}^\top \tau_i]) = \text{norm}([\mathbf{1}^\top \mu_i]).$$

Then, (σ, μ) belongs to the RHS. Thus, $\Phi(\text{LHS}) \subset \text{RHS}$. For every $(\pi, \tau) = \Phi^{-1}(\sigma, \mu)$ with (σ, μ) belonging to the RHS, we trivially have $(\pi, \tau) \in \Delta_n^m \times \mathbb{R}_\pm^{m \times n}$ (belonging to the LHS). Thus, $\Phi^{-1}(\text{RHS}) \subset \text{LHS}$. Therefore, we have proved that Φ is a bijection between the LHS and RHS in equation (4.7).

The maps ϕ and ϕ^{-1} in equation (4.1) are real-analytic because they consist only of elementary operations and the denominator $\sum_a \sigma_{i,a} > 0$ on Δ_{mn}^+ . The maps Φ and Φ^{-1} in equation (4.5) are also real-analytic because they consist only of elementary operations and the real-analytic ϕ and ϕ^{-1} , and the denominator $\sum_i \mathbf{1}^\top \tau_i > 0$ on $\mathbb{R}_\pm^{m \times n}$. Therefore, Φ is a real-analytic homeomorphism between the LHS and RHS in equation (4.7). $\square$

Theorem 6 shows that Φ is a homeomorphism beyond E and $\bar{E}$. This is useful in establishing the equivalence between the three equivalent characterizations of $\text{VI}(F, \Delta_n^m)$ and $\text{VI}(\bar{F}, \Delta_{mn}^+)$, because E and $\bar{E}$ are only solutions of the characterizations, but with Φ in equation (4.7), we can compare the function values of the characterizations whether on solutions or not.

Next, we show that $E \rightarrow \Delta_n^m \times \Delta_m^\circ$ and $\bar{E} \rightarrow \Delta_{mn}^+$ are isomorphic as fiber bundles. Given general fiber bundles $E \xrightarrow{\alpha} B$ and $E' \xrightarrow{\alpha'} B'$, a bundle isomorphism is a pair of homeomorphisms $(\psi : B \rightarrow B', \Psi : E \rightarrow E')$ such that $\alpha' \circ \Psi = \psi \circ \alpha$.

Theorem 7. (Φ, ϕ) is a bundle isomorphism such that

$$E \rightarrow \Delta_n^m \times \Delta_m^\circ \cong \bar{E} \rightarrow \Delta_{mn}^+.$$

Proof. First, we prove that Φ is a homeomorphism between E and $\bar{E}$. We have $E \subset \Delta_n^m \times \mathbb{R}_\pm^{m \times n}$ trivially. By the definition of $\bar{B}(\sigma)$, every $\mu \in \bar{B}(\sigma)$ for some $\sigma \in \Delta_{mn}^+$ satisfies $\mathbf{1}^\top \mu_i = \bar{w}_i$ with $w = \phi_w^{-1}(\sigma)$. Thus, $\bar{E} \subset (\Delta_{mn}^+ \times \mathbb{R}_\pm^{mn}) \cap \{(\sigma, \mu) \mid \phi_w^{-1}(\sigma) = \text{norm}([\mathbf{1}^\top \mu_i])\}$. We further need to show $\Phi(E) \subset \bar{E}$ and $\Phi^{-1}(\bar{E}) \subset E$.

For every $(\pi, \tau) \in E$, we have $\tau = \tilde{\tau}(\pi) + \bar{v}w \odot \pi$ for some $w \in \Delta_m^\circ$ and $\bar{v} \neq 0$. Then, under the map Φ , we have

$$\begin{aligned} (\pi, \tilde{\tau}(\pi) + \bar{v}w \odot \pi) &\mapsto (\phi(\pi, w), \text{reshape}(\tilde{\tau}(\pi) + \bar{v}w \odot \pi)) \\ &= (\sigma, \text{reshape}(\tilde{\tau}(\pi) + \bar{v}w \odot \pi)) \quad \text{s.t. } (\pi, w) = \phi^{-1}(\sigma), \end{aligned}$$

which indicates that $\Phi(\pi, \tau) \in \bar{E}$. For every $(\sigma, \mu) \in \bar{E}$, we have $\mu = \text{reshape}(\tilde{\tau}(\pi) + \bar{v}w \odot \pi)$ for $(\pi, w) = \phi^{-1}(\sigma)$ and $\bar{v} \neq 0$. Then, under the map Φ^{-1} , we have

$$\begin{aligned} (\sigma, \text{reshape}(\tilde{\tau}(\pi) + \bar{v}w \odot \pi)) &\mapsto (\phi_\pi^{-1}(\sigma), \tilde{\tau}(\pi) + \bar{v}w \odot \pi) \\ &= (\pi, \tilde{\tau}(\pi) + \bar{v}w \odot \pi) \quad \text{s.t. } (\pi, w) = \phi^{-1}(\sigma), \end{aligned}$$

which indicates that $\Phi^{-1}(\sigma, \mu) \in E$. Therefore, we have $\Phi(E) \subset \bar{E}$ and $\Phi^{-1}(\bar{E}) \subset E$, so Φ is a homeomorphism between E and $\bar{E}$.

We already have that ϕ is a homeomorphism between $\Delta_n^m \times \Delta_m^\circ$ and Δ_{mn}^+ . For (Φ, ϕ) to constitute a bundle isomorphism, we additionally need $\bar{\alpha} \circ \Phi = \phi \circ \alpha_1$, where α_1 and $\bar{\alpha}$ are bundle projections.

For each $(\pi, \tau) \in E$, under the maps $\Phi : E \rightarrow \bar{E}$ and $\bar{\alpha} : \bar{E} \rightarrow \Delta_{mn}^+$, we have

$$(\pi, \tau) \mapsto (\phi(\pi, \text{norm}([\mathbf{1}^\top \tau_i])), \text{reshape}(\tau)) \mapsto \phi(\pi, \text{norm}([\mathbf{1}^\top \tau_i])),$$

and under the maps $\alpha_1 : E \rightarrow \Delta_n^m \times \Delta_m^\circ$ and $\phi : \Delta_n^m \times \Delta_m^\circ \rightarrow \Delta_{mn}^+$, we have

$$(\pi, \tau) \mapsto (\pi, \text{norm}([\mathbf{1}^\top \tau_i])) \mapsto \phi(\pi, \text{norm}([\mathbf{1}^\top \tau_i])).$$

Therefore, we obtain $\bar{\alpha} \circ \Phi = \phi \circ \alpha_1$, so (Φ, ϕ) is a bundle isomorphism between $E \rightarrow \Delta_n^m \times \Delta_m^\circ$ and $\bar{E} \rightarrow \Delta_{mn}^+$. $\square$

4.3 Equivalence between characterizations

Having obtained the homeomorphism in equation (4.7), we then use it to establish the equivalence between the characterizations of $\text{VI}(F, \Delta_n^m)$ and $\text{VI}(\bar{F}, \Delta_{mn}^+)$. First, we introduce the objects for which the equivalence is to be studied, where we replace the FPKKTs with the fixed-point bundle equations, and we add sections of the fixed-point bundles into the study.

Sections. The sections $\check{\tau}(\pi)$ and $\tilde{\tau}(\pi)$ of the bundle $E \cup \{(\pi, \tilde{\tau}(\pi)) | \pi \in \Delta_n^m\} \rightarrow \Delta_n^m$ are as equation (3.9) shows. The sections $\check{\mu}(\sigma)$ and $\tilde{\mu}(\sigma)$ of the bundle $\bar{E} \cup \{(\sigma, \text{reshape}(\tilde{\tau}(\phi_\pi^{-1}(\sigma)))) | \sigma \in \Delta_{mn}^+\} \rightarrow \Delta_{mn}^+$ are from Paper 1, shown in equation (4.8).

$$\begin{aligned} \check{\mu}(\sigma) &= \sigma \circ \left(\bar{F}(\sigma) - \left(\min_{i,a} \bar{F}_{i,a}(\sigma) \right) \mathbf{1} \right) \\ \tilde{\mu}(\sigma) &= \sigma \circ \left(\bar{F}(\sigma) - \left(\sigma^\top \bar{F}(\sigma) \right) \mathbf{1} \right) = \left(I - \sigma \mathbf{1}^\top \right) (\sigma \circ \bar{F}(\sigma)) \end{aligned} \tag{4.8}$$

Fixed-point bundle equations. The fixed-point bundle equation of $E \rightarrow \Delta_n^m$ is as equation (3.8) shows. The fixed-point bundle equation of $\bar{E} \rightarrow \Delta_{mn}^+$ is $\bar{G}(\sigma, \mu) = 0$ from Paper 1, where

$$\bar{G}(\sigma, \mu) = \left(I - \sigma \mathbf{1}^\top \right) (\sigma \circ \bar{F}(\sigma) - \mu). \tag{4.9}$$

Brouwer functions. The Brouwer function $(\hat{\pi}, r, v) = M(\pi, \tau)$ characterizing $\text{VI}(F, \Delta_n^m)$ is as equation (3.3) shows. The Brouwer function characterizing $\text{VI}(\bar{F}, \Delta_{mn}^+)$ is from Paper 1. In

fact, $(\hat{\pi}, r, v) = M(\pi, \tau)$ in this paper can only be shown equivalent to an index-wise variant $(\hat{\sigma}, \bar{r}, \bar{v}) = \bar{M}(\sigma, \mu)$ as shown in equation (4.10), instead of the original Brouwer function in Paper 1.

$$(\hat{\sigma}, \bar{r}, \bar{v}) = \bar{M}(\sigma, \mu), \quad \text{s.t.} \quad \begin{bmatrix} \hat{\sigma}_i \circ \bar{r}_i - \mu_i \\ r_i - \bar{F}_i(\sigma) - \bar{v}_i \mathbf{1} \\ \mathbf{1}^\top \hat{\sigma}_i - \mathbf{1}^\top \sigma_i \end{bmatrix} = 0 \quad (4.10)$$

The index-wise variant $(\hat{\sigma}, \bar{r}, \bar{v}) = \bar{M}(\sigma, \mu)$ has two modifications compared to the original one. First, in Paper 1, all the components of $\bar{r}$ depend on a single scalar $\bar{v}$, whereas in equation (4.10), each $\bar{r}_i$ depends on an independent $\bar{v}_i$. Second, the constraint $\sum_i \mathbf{1}^\top \hat{\sigma}_i = 1$ in Paper 1 changes to the index-wise constraint $\mathbf{1}^\top \hat{\sigma}_i - \mathbf{1}^\top \sigma_i$ in equation (4.10). This variant essentially uses the index-wise $\bar{v}_i$ to control each $\mathbf{1}^\top \hat{\sigma}_i$ instead of the entire $\sum_i \mathbf{1}^\top \hat{\sigma}_i$.

MCPs. The two MCPs characterizing $\text{VI}(F, \Delta_n^m)$ and $\text{VI}(\bar{F}, \Delta_{mn}^+)$ are as equation (4.11) shows. $\text{mcp}(\pi, \tau)$ originates from MCP (3.5), and (r, v) is additionally constrained by the Brouwer function $(\hat{\pi}, r, v) = M(\pi, \tau)$ to make π the only optimization variable. $\text{mcp}(\sigma, \mu)$ originates from the MCP in Paper 1, where $(\bar{r}, \bar{v})$ is originally additionally constrained by the Brouwer function in Paper 1 to make σ the only optimization variable, but we modify $(\bar{r}, \bar{v})$ to be additionally constrained by the index-wise variant $(\hat{\sigma}, \bar{r}, \bar{v}) = \bar{M}(\sigma, \mu)$.

$$\begin{aligned} \text{mcp}(\pi, \tau) &= \sum_i \left(\pi_i^\top r_i - \tau_i^\top \ln \pi_i - \tau_i^\top \ln r_i \right), \quad \text{s.t.} \quad (\hat{\pi}, r, v) = M(\pi, \tau) \\ \text{mcp}(\sigma, \mu) &= \sigma^\top \bar{r} - \mu^\top \ln \sigma - \mu^\top \ln \bar{r}, \quad \text{s.t.} \quad (\hat{\sigma}, \bar{r}, \bar{v}) = \bar{M}(\sigma, \mu) \end{aligned} \quad (4.11)$$

Equivalence. The sections $\tilde{\tau}(\pi)$ and $\tilde{\mu}(\sigma)$ do not have a useful relation themselves, but since $\text{gap}_i(\pi_i, \pi_i) = \mathbf{1}^\top \tilde{\tau}_i(\pi)$ and $\text{gap}_{\bar{F}}(\sigma) = \mathbf{1}^\top \tilde{\mu}(\sigma)$, they are related through the gap function relation in Theorem 5. The relations between the remaining characterizations are as the following proposition shows.

Proposition 8. For $(\pi, \tau) \in \Delta_n^m \times \mathbb{R}_\pm^{m \times n}$ and $(\sigma, \mu) \in \Delta_{mn}^+ \times \mathbb{R}_\pm^{mn}$, let $\phi_w^{-1}(\sigma) = \text{norm}([\mathbf{1}^\top \mu_i])$ and $(\sigma, \mu) = \Phi(\pi, \tau)$, then the following equalities hold.

- (i) $\tilde{\mu}_i(\sigma) = \tilde{\tau}_i(\pi)$.
- (ii) $\bar{G}_i(\sigma, \mu) = G_i(\pi, \tau)$.
- (iii) $\hat{\sigma}_i = w_i \hat{\pi}_i$ and $\bar{r}_i = r_i / w_i$, where $(\hat{\pi}, r, v) = M(\pi, \tau)$ and $(\hat{\sigma}, \bar{r}, \bar{v}) = \bar{M}(\sigma, \mu)$.
- (iv) $\text{mcp}(\sigma, \mu) = \text{mcp}(\pi, \tau)$.

Proof. (i) Substituting $\sigma = \phi(\pi, w)$ and $\bar{F}(\sigma)$, we have

$$(\sigma \circ \bar{F}(\sigma))_i = w_i \pi_i \circ \frac{1}{w_i} \left(F_i(\pi) - (\pi_i^\top F_i(\pi)) \mathbf{1} \right) = \tilde{\tau}_i(\pi).$$

Then, since $\mathbf{1}^\top \tilde{\tau}_i(\pi) = 0$ for every index i , we obtain the equation

$$\tilde{\mu}_i(\sigma) = \left(\sigma \circ \left(\bar{F}(\sigma) - \left(\sigma^\top \bar{F}(\sigma) \right) \mathbf{1} \right) \right)_i = \tilde{\tau}_i(\pi) - \left(\sum_j \mathbf{1}^\top \tilde{\tau}_j(\pi) \right) \sigma_i = \tilde{\tau}_i(\pi).$$

(ii) Substituting $(\sigma, \mu) = \Phi(\pi, \tau)$ and $\sigma = \phi(\pi, w)$, using $\tilde{\mu}_i(\sigma) = \tilde{\tau}_i(\pi)$ in the first line, and using $w = \phi_w^{-1}(\sigma) = \text{norm}([\mathbf{1}^\top \mu_i]) = \text{norm}([\mathbf{1}^\top \tau_i])$ in the second line, we have

$$\bar{G}_i(\sigma, \mu) = \tilde{\mu}_i(\sigma) - \left(\mu_i - (\mathbf{1}^\top \mu) \sigma_i \right) = \tilde{\tau}_i(\pi) - \left(\tau_i - (\sum_i \mathbf{1}^\top \tau_i) w_i \pi_i \right)$$

$$= \tilde{\tau}_i(\pi) - \left(\tau_i - (\mathbf{1}^\top \tau_i) \pi_i \right) = G_i(\pi, \tau).$$

(iii) In $(\hat{\pi}, r, v) = M(\pi, \tau)$, each v_i is given by the zero of $q_i(v_i) = 0$ as equation (3.4) shows. In $(\hat{\sigma}, \bar{r}, \bar{v}) = \bar{M}(\sigma, \mu)$, each $\bar{v}_i$ is given by the zero of $\bar{q}_i(\bar{v}_i) = 0$ via a similar derivation from equation (4.10).

$$\begin{aligned} q_i(v_i) &= \mathbf{1}^\top \frac{\tau_i}{F_i(\pi) + v_i \mathbf{1}} - 1 \\ \bar{q}_i(\bar{v}_i) &= \mathbf{1}^\top \frac{\mu_i}{F_i(\sigma) + \bar{v}_i \mathbf{1}} - \mathbf{1}^\top \sigma_i = w_i \left(\mathbf{1}^\top \frac{\mu_i}{F_i(\pi) - (\pi_i^\top F_i(\pi)) \mathbf{1} + \bar{v}_i w_i \mathbf{1}} - 1 \right) \end{aligned}$$

Comparing $q_i(v_i)$ and $\bar{q}_i(\bar{v}_i)$, it can be directly observed that when $\sigma = \phi(\pi, w)$ and $\mu_i = \tau_i$, their zeros v_i and $\bar{v}_i$ satisfy $v_i = \bar{v}_i w_i - \pi_i^\top F_i(\pi)$. Then, we obtain $\bar{r}_i = r_i / w_i$ at this time from $r_i = F_i(\pi) + v_i \mathbf{1}$ and $\bar{r}_i = \bar{F}_i(\sigma) + \bar{v}_i \mathbf{1}$. Then, $\hat{\sigma}_i = w_i \hat{\pi}_i$ follows from $\hat{\pi}_i = \tau_i / r_i$ and $\hat{\sigma}_i = \mu_i / \bar{r}_i$.

(iv) By $\bar{r}_i = r_i / w_i$ from the Brouwer functions and $\sigma_i = w_i \pi_i$, we have $\sigma_i \circ \bar{r}_i = \pi_i \circ r_i$, then it follows that $\text{mcp}(\sigma, \mu) = \text{mcp}(\pi, \tau)$. $\square$

Proposition 8 will be used to transform the computational formula of the formality-only path-following algorithm for $\text{VI}(\bar{F}, \Delta_{mn}^+)$ into the computational formula of the path-following algorithm for $\text{VI}(F, \Delta_n^m)$, and we will deal with the difference between the index-wise variants and the original Brouwer functions and MCPs.

4.4 Equivalence between differential properties

In Paper 1, to maintain the constraint $\sigma \in \Delta$, σ is represented by the softmax function such that $\sigma = \text{softmax}(\theta)$. Then, the differentiable properties are measured by differentiating with respect to θ instead of σ . In this subsection, we derive the equivalence between the differential properties of the two fixed-point bundles. First, we also use the softmax function to represent π to maintain the constraint $\pi \in \Delta_n^m$.

Theorem 9. *For any $\sigma \in \Delta_{mn}^+$, $\pi \in \Delta_n^m$, and $w \in \Delta_m^\circ$, $\sigma = \phi(\pi, w)$ if and only if there exists θ such that*

$$\sigma = \text{softmax}(\theta) = \frac{\exp(\theta)}{\sum_i \mathbf{1}^\top \exp(\theta_i)}, \quad w_i = \frac{\mathbf{1}^\top \exp(\theta_i)}{\sum_i \mathbf{1}^\top \exp(\theta_i)}, \quad \pi_i = \text{softmax}(\theta_i) = \frac{\exp(\theta_i)}{\mathbf{1}^\top \exp(\theta_i)}. \quad (4.12)$$

Proof. Given $\sigma = \phi(\pi, w)$, we have the softmax representations with $\theta = \ln \sigma$. Given the softmax representations, $\sigma_i = w_i \pi_i$ can be directly verified. $\square$

Theorem 9 shows that if we represent σ, π, w using softmax functions, then $\sigma = \phi(\pi, w)$ is automatically satisfied. If we additionally represent $\mu = \text{reshape}(\tau)$, then $(\sigma, \mu) = \Phi(\pi, \tau)$ is satisfied as long as $w(\theta) = \text{norm}([\mathbf{1}^\top \tau_i])$. Consequently, viewed as functions in θ , the sections $\tilde{\tau}(\pi)$ and $\tilde{\mu}(\sigma)$ in Proposition 8 (i) are in fact the same functions. Viewed as functions in (θ, τ) , the fixed-point bundle functions $\bar{G}(\sigma, \mu)$ and $G(\pi, \tau)$ in Proposition 8 (ii) are in fact the same functions as long as $w(\theta) = \text{norm}([\mathbf{1}^\top \tau_i])$.

The differentiable properties are measured by differentiating with respect to θ , so we derive the differentials of σ, π, w with respect to θ .

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

$$\begin{aligned}
 d\pi_i &= \frac{\exp(\theta_i) \circ d\theta_i}{\mathbf{1}^\top \exp(\theta_i)} - \frac{\mathbf{1}^\top (\exp(\theta_i) \circ d\theta_i)}{(\mathbf{1}^\top \exp(\theta_i))^2} \exp(\theta_i) = \pi_i \circ d\theta_i - (\pi_i^\top d\theta_i) \pi_i \\
 dw_i &= \frac{\mathbf{1}^\top (\exp(\theta_i) \circ d\theta_i)}{\sum_j \mathbf{1}^\top \exp(\theta_j)} - \frac{\sum_j \mathbf{1}^\top (\exp(\theta_j) \circ d\theta_j)}{(\sum_j \mathbf{1}^\top \exp(\theta_j))^2} (\mathbf{1}^\top \exp(\theta_i)) = w_i \pi_i^\top d\theta_i - w_i \sum_j w_j \pi_j^\top d\theta_j
 \end{aligned}$$

The derivation shows that the differentials $d\sigma, d\pi, dw$ are as equation (4.13) shows.

$$d\sigma/\sigma = (I - \mathbf{1}\sigma^\top) d\theta, \quad d\pi_i/\pi_i = (I - \mathbf{1}\pi_i^\top) d\theta_i, \quad dw/w = (I - \mathbf{1}w^\top) [\pi_i^\top d\theta_i] \quad (4.13)$$

Note that $\text{softmax}(\theta + k\mathbf{1}) = \text{softmax}(\theta)$, indicating that softmax is invariant under adding a scaled $\mathbf{1}$, and adding a scaled $\mathbf{1}$ to $d\theta$ does not change the differentials $d\sigma$ or $d\pi$. Thus, we denote $\widetilde{d\theta}$ and $\widetilde{d\theta}$ in equation (4.14) such that $\sigma^\top \widetilde{d\theta} = 0$ and $\pi_i^\top \widetilde{d\theta}_i = 0$, where $\widetilde{d\theta}$ has already appeared in Paper 1.

$$\widetilde{d\theta} := (I - \mathbf{1}\sigma^\top) d\theta, \quad \widetilde{d\theta}_i := (I - \mathbf{1}\pi_i^\top) d\theta_i \quad (4.14)$$

As in Paper 1, instead of using the plain $d\theta$, we actually use $\widetilde{d\theta}$ and $\widetilde{d\theta}$ as the differentials measuring the differentiable properties of the fixed-point bundles. First, we study their relation.

Lemma 10. *Let (σ, π, w) be subject to equation (4.12), then for any $d\theta$, the following statements hold.*

(i) $\sigma^\top \widetilde{d\theta} = 0$.

(ii) $\pi_i^\top \widetilde{d\theta}_i = 0$ for every index i if and only if $\widetilde{d\theta} = \widetilde{d\theta}$, which is equivalently $dw = 0$.

Proof. (i) For $\sigma^\top \widetilde{d\theta}$, we can derive $\sigma^\top \widetilde{d\theta} = \sum_j w_j \pi_j^\top \widetilde{d\theta}_j$. Then, we have $\sigma^\top \widetilde{d\theta} = 0$ since $\pi_j^\top \widetilde{d\theta}_j = 0$.

(ii) For $\pi_i^\top \widetilde{d\theta}_i$, $\widetilde{d\theta} - \widetilde{d\theta}$, and dw , we can derive the following three equations.

$$\begin{aligned}
 \pi_i^\top \widetilde{d\theta}_i &= \pi_i^\top (d\theta_i - (\sigma^\top d\theta) \mathbf{1}) = \pi_i^\top d\theta_i - \sum_j w_j \pi_j^\top d\theta_j \\
 \widetilde{d\theta}_i - \widetilde{d\theta}_i &= d\theta_i - (\sigma^\top d\theta) \mathbf{1} - (d\theta_i - (\pi_i^\top d\theta_i) \mathbf{1}) = (\pi_i^\top d\theta_i) \mathbf{1} - (\sum_j w_j \pi_j^\top d\theta_j) \mathbf{1} \\
 dw/w &= (I - \mathbf{1}w^\top) [\pi_i^\top d\theta_i] = [\pi_i^\top d\theta_i - \sum_j w_j \pi_j^\top d\theta_j].
 \end{aligned}$$

Therefore, $\pi_i^\top \widetilde{d\theta}_i = 0$ for every index i , $\widetilde{d\theta} = \widetilde{d\theta}$, and $dw = 0$ are all equivalent to $\pi_i^\top d\theta_i = \sum_j w_j \pi_j^\top d\theta_j$ for every index i . $\square$

The differentiable properties of the fixed-point bundle are governed by the section $\tilde{\mu}(\sigma)$ and $\tilde{\tau}_i(\pi)$, such that their differentials appear in the singular manifold, differential equation, Newton equation, and gradient of related functions. In Paper 1, $d\tilde{\mu}(\sigma)$ is measured with respect to $\widetilde{d\theta}$. Next, we measure $d\tilde{\tau}_i(\pi)$ with respect to $\widetilde{d\theta}$ and derive the relation to $d\tilde{\mu}(\sigma)$ since $\tilde{\mu}(\sigma) = \tilde{\tau}_i(\pi)$. In the following equations, let δ_{ij} be the Kronecker delta, such that $\delta_{ij} = 1$ when $i = j$ and $\delta_{ij} = 0$ when $i \neq j$.

Theorem 11. *Let $\sigma = \phi(\pi, w)$, then the differentials of $\tilde{\tau}(\pi)$ and $\tilde{\mu}(\sigma)$ satisfy*

$$d\tilde{\tau}_i(\pi) = d\tilde{\mu}_i(\sigma) = \sigma_i \circ \bar{J}_i(\sigma) \widetilde{d\theta} = \sigma_i \circ \bar{J}_i(\sigma) \widetilde{d\theta} = \pi_i \circ J_i(\pi) \widetilde{d\theta}, \quad (4.15)$$

where $J(\pi)$ and $\bar{J}(\sigma)$ are given by the following equations.

$$\begin{aligned} J_{ij}(\pi) &:= \left(I - \mathbf{1}\pi_i^\top \right) \left(\frac{\partial F_i(\pi)}{\partial \pi_j} \circ \pi_j + \delta_{ij} \left(\left(I - \mathbf{1}\pi_i^\top \right) F_i(\pi) \right) \circ I \right) \left(I - \mathbf{1}\pi_j^\top \right) + kw_iw_j\mathbf{1}\pi_j^\top \\ \bar{J}(\sigma) &:= \left(I - \mathbf{1}\sigma^\top \right) \left(\frac{\partial \bar{F}(\sigma)}{\partial \sigma} \circ \sigma + \left(\left(I - \mathbf{1}\sigma^\top \right) \bar{F}(\sigma) \right) \circ I \right) \left(I - \mathbf{1}\sigma^\top \right) + k\mathbf{1}\sigma^\top \end{aligned} \quad (4.16)$$

Consequently, $w_i\bar{J}_{ij}(\sigma) = J_{ij}(\pi)$.

Proof. (1) Proving $d\tilde{\tau}_i(\pi) = d\tilde{\mu}_i(\sigma) = \sigma_i \circ \bar{J}_i(\sigma)\bar{d}\theta$.

In Paper 1, we have $d\tilde{\mu}_i(\sigma) = \sigma_i \circ \bar{J}_i(\sigma)\bar{d}\theta$. Proposition 8 shows that $\tilde{\tau}_i(\pi) = \tilde{\mu}_i(\sigma)$, thus $d\tilde{\tau}_i(\pi) = d\tilde{\mu}_i(\sigma)$.

(2) Proving $d\tilde{\tau}_i(\pi) = \sigma_i \circ \bar{J}_i(\sigma)\bar{d}\theta = \sigma_i \circ \bar{J}_i(\sigma)\widetilde{d\theta}$.

Note that $d\theta$ and $\widetilde{d\theta}$ lead to the same $d\pi$ in equation (4.13) since $I - \mathbf{1}\pi_i^\top$ is idempotent. Then, substituting $d\theta$ with $\widetilde{d\theta}$ does not change $d\tilde{\tau}_i(\pi)$ since $\tilde{\tau}_i(\pi)$ depends only on π . Thus, we have the following derivation, where $(I - \mathbf{1}\sigma^\top)\widetilde{d\theta} = \widetilde{d\theta}$ by Lemma 10 (i).

$$d\tilde{\tau}_i(\pi) = \sigma_i \circ \bar{J}_i(\sigma)\bar{d}\theta = \sigma_i \circ \bar{J}_i(\sigma) \left(I - \mathbf{1}\sigma^\top \right) d\theta = \sigma_i \circ \bar{J}_i(\sigma) \left(I - \mathbf{1}\sigma^\top \right) \widetilde{d\theta} = \sigma_i \circ \bar{J}_i(\sigma)\widetilde{d\theta}$$

(3) Proving $d\tilde{\tau}_i(\pi) = \pi_i \circ J_i(\pi)\widetilde{d\theta}$.

The differential $d\tilde{\tau}(\pi)$ is derived as follows. In the third line, we use $\mathbf{1}^\top d\pi_i = 0$. In the last line, we use $d\pi_i/\pi_i = \widetilde{d\theta}_i = (I - \mathbf{1}\pi_i^\top)\widetilde{d\theta}_i$.

$$\begin{aligned} d\tilde{\tau}_i(\pi) &= \left(I - \pi_i\mathbf{1}^\top \right) \left(\pi_i \circ \sum_j \frac{\partial F_i(\pi)}{\partial \pi_j} d\pi_j + F_i(\pi) \circ d\pi_i \right) - \mathbf{1}^\top (\pi_i \circ F_i(\pi)) d\pi_i \\ &= \left(I - \pi_i\mathbf{1}^\top \right) \left(\pi_i \circ \sum_j \frac{\partial F_i(\pi)}{\partial \pi_j} d\pi_j + \left(F_i(\pi) \circ I - (\pi_i^\top F_i(\pi))I \right) d\pi_i \right) - \left(\pi_i^\top F_i(\pi) \right) \pi_i\mathbf{1}^\top d\pi_i \\ &= \pi_i \circ \left(I - \mathbf{1}\pi_i^\top \right) \left(\sum_j \left(\frac{\partial F_i(\pi)}{\partial \pi_j} \circ \pi_j \right) \frac{d\pi_j}{\pi_j} + \left(\left(I - \mathbf{1}\pi_i^\top \right) F_i(\pi) \right) \circ \frac{d\pi_i}{\pi_i} \right) \\ &= \pi_i \circ \left(I - \mathbf{1}\pi_i^\top \right) \left(\sum_j \left(\frac{\partial F_i(\pi)}{\partial \pi_j} \circ \pi_j \right) \left(I - \mathbf{1}\pi_j^\top \right) \widetilde{d\theta}_j + \left(\left(I - \mathbf{1}\pi_i^\top \right) F_i(\pi) \right) \circ \left(I - \mathbf{1}\pi_i^\top \right) \widetilde{d\theta}_i \right) \end{aligned}$$

Substituting $J_{ij}(\pi)$ and using $\pi_i^\top \widetilde{d\theta}_i = 0$, we obtain $d\tilde{\tau}_i(\pi) = \pi_i \circ J_i(\pi)\widetilde{d\theta}$.

(4) Proving $w_i\bar{J}_{ij}(\sigma) = J_{ij}(\pi)$.

Let $\bar{J}_{0ij}(\sigma)$ and $J_{0ij}(\pi)$ be the corresponding matrices with the arbitrary scalar k set to 0. Then, we have the following derivation, where $k\mathbf{1}\sigma^\top$ and $kw_iw_j\mathbf{1}\pi_j^\top$ are eliminated by $(I - \mathbf{1}\sigma^\top)$ and $(I - \mathbf{1}\pi_j^\top)$, and $(I - \mathbf{1}\sigma^\top)$ and $(I - \mathbf{1}\pi_j^\top)$ are absorbed into $\bar{J}_{0ij}(\sigma)$ and $J_{0ij}(\pi)$ by idempotence.

$$\begin{aligned} \sigma_i \circ \bar{J}_i(\sigma)\widetilde{d\theta} &= \sigma_i \circ \bar{J}_i(\sigma) \left(I - \mathbf{1}\sigma^\top \right) d\theta = \sigma_i \circ \sum_j \bar{J}_{0ij}(\sigma) d\theta_j \\ \pi_i \circ J_i(\pi)\widetilde{d\theta} &= \pi_i \circ \sum_j J_{ij}(\pi) \left(I - \mathbf{1}\pi_j^\top \right) d\theta_j = \pi_i \circ \sum_j J_{0ij}(\pi) d\theta_j \end{aligned}$$

Then, by $\sigma_i \circ \bar{J}_i(\sigma)\widetilde{d\theta} = \pi_i \circ J_i(\pi)\widetilde{d\theta}$ for every $d\theta$, we have $w_i\bar{J}_{0ij}(\sigma) = J_{0ij}(\pi)$. From $w_i(k\mathbf{1}\sigma^\top)_{ij} = w_i(k\mathbf{1}\sigma_j^\top) = kw_iw_j\mathbf{1}\pi_j^\top$, we obtain $w_i\bar{J}_{ij}(\sigma) = J_{ij}(\pi)$. $\square$

Theorem 11 establishes the equivalence between the differentials of $\tilde{\tau}(\pi)$ and $\tilde{\mu}(\sigma)$, which govern the differential properties of the fixed-point bundles. Furthermore, we obtain $w_i \bar{J}_{ij}(\sigma) = J_{ij}(\pi)$ between the two Jacobian-like matrices, which play a central role in the predictor-corrector algorithm. The derivative of $\bar{F}(\sigma)$ is given by

$$\frac{\partial \bar{F}_i(\sigma)}{\partial \sigma_j} = \frac{1}{w_i w_j} \left(I - \mathbf{1} \pi_i^\top \right) \frac{\partial F_i(\pi)}{\partial \pi_j} \left(I - \mathbf{1} \pi_i^\top \right)^\top - \delta_{ij} \frac{1}{w_i} \left(\bar{F}_i(\sigma) \mathbf{1}^\top + \mathbf{1} \bar{F}_i(\sigma)^\top \right). \quad (4.17)$$

It would be very complicated to directly verify $w_i \bar{J}_{ij}(\sigma) = J_{ij}(\pi)$ via symbolic derivation from equation (4.16), but we have verified in numerical experiments that no significant difference can be found between $\text{Diag}[w_i I] \bar{J}(\sigma)$ and $J(\pi)$ with $k = 0$.

5 Path-following on the fixed-point bundles

5.1 Transforming the predictor-corrector framework

Paper 1 adopts the predictor-corrector framework to perform path-following on the fixed-point bundle. This framework alternates between predictor steps and corrector steps. A predictor step updates along the tangent direction of the fixed-point bundle, which is given by the differential equations of $\bar{G}(\sigma, \mu) = 0$ or $G(\pi, \tau) = 0$. A corrector step updates onto the fixed-point bundle, which can be given by the Newton equations of $\bar{G}(\sigma, \mu) = 0$ or $G(\pi, \tau) = 0$, or by the gradients of $\text{mcp}(\sigma, \mu)$ and $\text{mcp}(\pi, \tau)$. We have the formality-only predictor and corrector for $\bar{G}(\sigma, \mu) = 0$ from Paper 1. In this subsection, we use the equivalence established previously to obtain the predictor and corrector for $G(\pi, \tau) = 0$.

Differential equations and Newton equations of $\bar{G}(\sigma, \mu) = 0$ and $G(\pi, \tau) = 0$. In Paper 1, the differential equation and Newton equation of $\bar{G}(\sigma, \mu) = 0$ are given by equations (5.1) and (5.2).

$$\left(\bar{J}(\sigma) + (\mathbf{1}^\top \mu) I \right) \bar{d}\theta = \left(I - \mathbf{1} \sigma^\top \right) (d\mu / \sigma) \quad (5.1)$$

$$\left(\bar{J}(\sigma) + (\mathbf{1}^\top \mu) I \right) \bar{d}\theta = -\bar{G}(\sigma, \mu) / \sigma \quad (5.2)$$

Proposition 12. *Let $(\sigma, \mu) = \Phi(\pi, \tau)$ and $\phi_w^{-1}(\sigma) = \text{norm}([\mathbf{1}^\top \mu_i])$. In both differential equation (5.1) and Newton equation (5.2), we have $\pi_i^\top \bar{d}\theta_i = 0$ for every index i . Consequently, $\bar{d}\theta = \widetilde{d}\theta$ and $dw = 0$.*

Proof. Differential equation (5.1) and Newton equation (5.2) can be derived as follows. We first multiply both sides by $\text{Diag}[w_i I]$ in the first lines, and then substitute $w_i \bar{J}_{ij}(\sigma) = J_{ij}(\pi)$ and $w = \text{norm}([\mathbf{1}^\top \tau_i])$ in the second lines.

$$\text{Diag}[w_i I] \left(\bar{J}(\sigma) + (\mathbf{1}^\top \mu) I \right) \bar{d}\theta = \text{Diag}[w_i I] \left(I - \mathbf{1} \sigma^\top \right) (d\mu / \sigma)$$

$$\left(J(\pi) + \text{Diag}[(\mathbf{1}^\top \tau_i) I] \right) \bar{d}\theta = \left[\left(I - \mathbf{1} \pi_i^\top \right) (d\tau_i / \pi_i) \right]$$

$$\text{Diag}[w_i I] \left(\bar{J}(\sigma) + (\mathbf{1}^\top \mu) I \right) \bar{d}\theta = -\text{Diag}[w_i I] (\bar{G}(\sigma, \mu) / \sigma)$$

$$\left(J(\pi) + \text{Diag}[(\mathbf{1}^\top \tau_i) I] \right) \bar{d}\theta = -[G_i(\pi, \tau) / \pi_i]$$

Then, multiplying the i -th row blocks on both sides by $\pi_i^\top$ and using $\pi_i^\top J_i(\pi) = 0$, we obtain $\pi_i^\top \bar{d}\theta_i = 0$. Then, $\bar{d}\theta = \widetilde{d}\theta$ and $dw = 0$ follow from Lemma 10 (ii). $\square$

Proposition 8 already shows that $\bar{G}(\sigma, \mu) = 0$ and $G(\pi, \tau) = 0$ are in fact the same function with the same variable and the same value. Then, from the proof of Proposition 12 and $\bar{d}\theta = \widetilde{d\theta}$, it immediately follows that the differential equation and Newton equation of $G(\pi, \tau) = 0$ are given by equations (5.3) and (5.4).

$$\left(J(\pi) + \text{Diag} \left[(\mathbf{1}^\top \tau_i) I \right] \right) \widetilde{d\theta} = \left[(I - \mathbf{1}\pi_i^\top) (d\tau_i / \pi_i) \right] \quad (5.3)$$

$$\left(J(\pi) + \text{Diag} \left[(\mathbf{1}^\top \tau_i) I \right] \right) \widetilde{d\theta} = \left[(I - \mathbf{1}\pi_i^\top) (F_i(\pi) - \tau_i / \pi_i) \right] \quad (5.4)$$

These equations also indicate that $(\pi, \tau) \in E$ is a singular point if and only if $\sum_i \mathbf{1}^\top \tau_i$ is an eigenvalue of $-\left[(1/w_i)J_{ij}(\pi)\right]$. In Paper 1, $(\sigma, \mu) \in \bar{E}$ is a singular point if and only if $\mathbf{1}^\top \mu$ is an eigenvalue of $-\bar{J}(\sigma)$. This aligns with the homeomorphism between (σ, μ) and (π, τ) .

Gradients of $\text{mcp}(\sigma, \mu)$ and $\text{mcp}(\pi, \tau)$. Recall that $\text{mcp}(\sigma, \mu)$ is a variant of the original MCP characterizing $\text{VI}(\bar{F}, \Delta_{mn}^+)$ in Paper 1. Denote the original MCP as $\overline{\text{mcp}}(\sigma, \mu)$. In Paper 1, the gradient of $\overline{\text{mcp}}(\sigma, \mu)$ is derived as the first equality in the following equation.

$$d\overline{\text{mcp}}(\sigma, \mu) = (\sigma - \hat{\sigma})^\top \left(\bar{J}(\sigma) + \left(\bar{r} - (I - \mathbf{1}\sigma^\top) \bar{F}(\sigma) \right) \circ I \right) \bar{d\theta} = d\text{mcp}(\sigma, \mu)$$

In the first equality, $(\sigma - \hat{\sigma})^\top \mathbf{1} d\bar{v} = 0$ is used to eliminate $d\bar{v}$ in the derivation in Paper 1. From $d\overline{\text{mcp}}(\sigma, \mu)$ to $d\text{mcp}(\sigma, \mu)$, the elimination of $d\bar{v}_i$ remains as the following equation shows. Thus, we have the second equality in the above equation, indicating that $d\text{mcp}(\sigma, \mu)$ also yields that expression.

$$(\sigma - \mu/\bar{r})^\top [d\bar{v}_i \mathbf{1}] = \sum_i (\sigma_i - \hat{\sigma}_i)^\top \mathbf{1} d\bar{v}_i = \sum_i (w_i \pi_i - w_i \hat{\pi}_i)^\top \mathbf{1} d\bar{v}_i = 0$$

Then, we have the following derivation from that expression. In the first line, we use $\text{mcp}(\pi, \tau) = \text{mcp}(\sigma, \mu)$ in Proposition 8. In the second line, we replace $\bar{d\theta} = (I - \mathbf{1}\sigma^\top) d\theta$ with $(I - \mathbf{1}\sigma^\top) \widetilde{d\theta}$, and use $\sigma^\top \widetilde{d\theta} = 0$ in Lemma 10 (i). The substitution is valid because $d\theta$ and $\widetilde{d\theta}$ lead to the same $d\pi$ in equation (4.13) since $I - \mathbf{1}\pi_i^\top$ is idempotent, which in turn leads to the same $d\text{mcp}(\pi, \tau)$. In the third line, we use $\bar{r}_i = r_i/w_i$ and $\hat{\sigma}_i = w_i \hat{\pi}_i$ in Proposition 8 (iii), $w_i \bar{J}_{ij}(\sigma) = J_{ij}(\pi)$ in Theorem 11, as well as $\sigma^\top \bar{F}(\sigma) = 0$ and $w_i \bar{F}_i(\sigma) = F_i(\pi) - (\pi_i^\top F_i(\pi)) \mathbf{1}$. In the fourth line, we use $r_i = F_i(\pi) + v_i \mathbf{1}$ and $\hat{\pi}_i^\top F_i(\pi) + v_i = \mathbf{1}^\top \tau_i$.

$$\begin{aligned} d\text{mcp}(\pi, \tau) &= d\text{mcp}(\sigma, \mu) \\ &= (\sigma - \hat{\sigma})^\top \left(\bar{J}(\sigma) + \left(\bar{r} - (I - \mathbf{1}\sigma^\top) \bar{F}(\sigma) \right) \circ I \right) \widetilde{d\theta} \\ &= \sum_{i,j} (\pi_i - \hat{\pi}_i)^\top \left(J_{ij}(\pi) + \delta_{ij} \left(r_i - (F_i(\pi) - (\pi_i^\top F_i(\pi)) \mathbf{1}) \right) \right) \circ I \widetilde{d\theta}_j \\ &= \sum_{i,j} (\pi_i - \hat{\pi}_i)^\top \left(J_{ij}(\pi) + \delta_{ij} \left(\mathbf{1}^\top \tau_i + (\pi_i - \hat{\pi}_i)^\top F_i(\pi) \right) \right) I \widetilde{d\theta}_j \end{aligned}$$

Ignoring the higher-order term $((\pi_i - \hat{\pi}_i)^\top F_i(\pi))(\pi_i - \hat{\pi}_i)$, we obtain the inexact gradient of $\text{mcp}(\pi, \tau)$ in equation (5.5).

$$\nabla \text{mcp}(\pi, \tau) = \left(J(\pi) + \text{Diag} \left[(\mathbf{1}^\top \tau_i) I \right] \right)^\top [\pi_i - \hat{\pi}_i] \quad (5.5)$$

Regularized Newton equations. In Paper 1, the standard Newton corrector is used to prove the convergence rate of the overall predictor-corrector path-following algorithm, and two regularized

Newton correctors are provided for practical use. The regularized Newton equations turn the coefficient matrix positive semidefinite and control its condition number using a regularization parameter, so that the iteration is stabilized and the update steps do not blow up when approaching singularities. Here, we derive the two corresponding regularized Newton correctors for $\text{VI}(F, \Delta_n^m)$. We denote $J_G := J(\pi) + \text{Diag}[(\mathbf{1}^\top \tau_i)I]$ as a shorthand notation.

From the standard Newton equation (5.4), by multiplying both sides by J_G and adding a regularization parameter δ_H to the coefficient, we obtain the regularized Newton equation (5.6).

$$\begin{aligned} (J_G^\top J_G + \delta_H I) d\theta &= -J_G^\top \left[(I - \mathbf{1}\pi_i^\top) (F_i(\pi) - \tau_i/\pi_i) \right] \\ \widetilde{d\theta}_i &= (I - \mathbf{1}\pi_i^\top) d\theta_i \end{aligned} \quad (5.6)$$

In Paper 1, the right-hand side of the standard Newton equation (5.4) can be transformed into $\sigma - \hat{\sigma}$ as the first line of the following equation shows, where $v_{gb}\mathbf{1}$ is changed to $[v_{gb,i}\mathbf{1}]$ to reflect the change from the original Brouwer function to the index-wise variant $(\hat{\sigma}, \bar{r}, \bar{v}) = \bar{M}(\sigma, \mu)$. Given $(\sigma, \mu) = \Phi(\pi, \tau)$ and $\phi_w^{-1}(\sigma) = \text{norm}([\mathbf{1}^\top \mu_i])$, this transformation can be further transformed as follows.

$$\begin{aligned} \frac{\sigma}{\bar{r}} \circ (\bar{J}(\sigma) + (\mathbf{1}^\top \mu)I) (\bar{d\theta} + [v_{gb,i}\mathbf{1}]) &= -(\sigma - \hat{\sigma}) \\ \frac{w_i \pi_i}{r_i / w_i} \circ \sum_j \left((1/w_i) J_{ij}(\pi) + \delta_{ij} (1/w_i) (\mathbf{1}^\top \tau_i) I \right) (\bar{d\theta}_i + v_{gb,i}\mathbf{1}) &= -(w_i \pi_i - w_i \hat{\pi}_i) \\ \frac{\pi_i}{r_i} \circ \sum_j \left(J_{ij}(\pi) + \delta_{ij} (\mathbf{1}^\top \tau_i) I \right) (\bar{d\theta}_i + v_{gb,i}\mathbf{1}) &= -(\pi_i - \hat{\pi}_i) \end{aligned}$$

Then, by multiplying both sides by J_G and adding a regularization parameter δ_H to the coefficient, we obtain the regularized Newton equation (5.7).

$$\begin{aligned} \left(J_G^\top \circ \left[\frac{\pi_i}{r_i} \right] \circ J_G + \delta_H I \right) d\theta &= -J_G^\top [\pi_i - \hat{\pi}_i] \\ \widetilde{d\theta}_i &= (I - \mathbf{1}\pi_i^\top) d\theta_i \end{aligned} \quad (5.7)$$

5.2 Transferring the convergence results

So far, we have completed the formality-only fixed-point bundle framework of $\text{VI}(\bar{F}, \Delta_{mn}^+)$ and the composite fixed-point bundle framework of $\text{VI}(F, \Delta_n^m)$. Now, we are finally ready to fill the final gap from $\text{VI}(\bar{F}, \Delta_{mn}^+)$ to an actual smooth VI on a closed simplex, such that the theoretical guarantees in Paper 1 apply.

As in Paper 1, the path-following iteration consists of four parts: (i) reducing $\mu > 0$ or $\tau > 0$ toward 0; (ii) avoiding singular points along the fibers by adding $\beta\sigma$ or $\beta\text{norm}([\mathbf{1}^\top \tau_i]) \odot \pi$ such that $\mathbf{1}^\top \mu + \beta$ avoids the eigenvalues of $-\bar{J}(\sigma)$, or $\sum_i \mathbf{1}^\top \tau_i + \beta$ avoids the eigenvalues of $-\text{Diag}[(1/w_i)I]J(\pi)$; (iii) a predictor step along the tangent direction of $\bar{G}(\sigma, \mu) = 0$ or $G(\pi, \tau) = 0$; (iv) a corrector step toward $\bar{G}(\sigma, \mu) = 0$ or $G(\pi, \tau) = 0$. First, we establish the equivalence between the path-following algorithms for $\text{VI}(\bar{F}, \Delta_{mn}^+)$ and $\text{VI}(F, \Delta_n^m)$ with standard Newton correctors.

Theorem 13. *Let the starting points be $(\sigma_{\text{init}}, v_{\text{init}}\sigma_{\text{init}})$ and $(\pi_{\text{init}}, v_{\text{init}}w_{\text{init}} \odot \pi_{\text{init}})$ such that $\sigma_{\text{init}} = \phi(\pi_{\text{init}}, w_{\text{init}})$. Let the iteration consist of the following three kinds of updates.*

- (i) $\mu' = (1 - \eta)\mu$ and $\tau' = (1 - \eta)\tau$.

(ii) $\hat{\mu} = \mu + \beta\sigma$ and $\hat{\tau} = \tau + \beta\text{norm}([\mathbf{1}^\top \tau_i]) \odot \pi$.

(iii) $\sigma'_i = \phi_{w,i}^{-1}(\sigma)\text{softmax}(\ln \phi_{\pi,i}^{-1}(\sigma) + \overline{d\theta}_i)$ and $\pi'_i = \text{softmax}(\ln \pi_i + \widetilde{d\theta}_i)$ with $\overline{d\theta}$ and $\widetilde{d\theta}$ subject to the differential equations or the Newton equations of $\bar{G}(\sigma, \mu) = 0$ and $G(\pi, \tau) = 0$.

Then, $(\sigma, \mu) = \Phi(\pi, \tau)$ and $w_{\text{init}} = \phi_w^{-1}(\sigma) = \text{norm}([\mathbf{1}^\top \mu_i])$ hold constantly during the iteration.

Proof. The first equality $(\sigma, \mu) = \Phi(\pi, \tau)$ holds if and only if $\sigma = \phi(\pi, w)$, $\mu = \text{reshape}(\tau)$, and $w = \text{norm}([\mathbf{1}^\top \mu_i])$. Given the first equality, the second equality holds if and only if $w = w_{\text{init}}$. Thus, we use these four equalities to verify that the two conditions are preserved during the iteration.

At the starting points $(\sigma_{\text{init}}, v_{\text{init}}\sigma_{\text{init}})$ and $(\pi_{\text{init}}, v_{\text{init}}w_{\text{init}} \odot \pi_{\text{init}})$, from $\sigma_{\text{init}} = \phi(\pi_{\text{init}}, w_{\text{init}})$, we can verify $\sigma = \phi(\pi, w)$, $\mu = \text{reshape}(\tau)$, $\phi_w^{-1}(\sigma) = w_{\text{init}}$, and $\text{norm}([\mathbf{1}^\top \mu_i]) = w_{\text{init}}$. Thus, the conditions $(\sigma, \mu) = \Phi(\pi, \tau)$ and $w_{\text{init}} = \phi_w^{-1}(\sigma) = \text{norm}([\mathbf{1}^\top \mu_i])$ hold at the starting points. Next, it remains to prove that given that the two conditions hold, the updates preserve them.

Update (i): Since σ and π remain unchanged, we have $\sigma = \phi(\pi, w)$ and $w = w_{\text{init}}$. For μ' , we have $\text{norm}([\mathbf{1}^\top \mu'_i]) = \text{norm}([\mathbf{1}^\top \mu_i])$, so $w = \text{norm}([\mathbf{1}^\top \mu'_i])$ given $w = \text{norm}([\mathbf{1}^\top \mu_i])$. For μ' and τ' , we have $\mu' = \text{reshape}(\tau')$ given $\mu = \text{reshape}(\tau)$. Thus, the two conditions are preserved.

Update (ii): Since σ and π remain unchanged, we have $\sigma = \phi(\pi, w)$ and $w = w_{\text{init}}$. For $\hat{\mu}$, we have $\text{norm}([\mathbf{1}^\top \hat{\mu}_i]) = \text{norm}([\mathbf{1}^\top \mu_i + \beta w_{\text{init},i}]) = w_{\text{init}}$ given $\mathbf{1}^\top \sigma_i = w_i$ and $w = \text{norm}([\mathbf{1}^\top \mu_i])$. For $\hat{\mu}$ and $\hat{\tau}$, we have $\sigma = \text{reshape}(\text{norm}([\mathbf{1}^\top \tau_i]) \odot \pi)$ given $\mu = \text{reshape}(\tau)$ and $\text{norm}([\mathbf{1}^\top \mu_i]) = w$, then we have $\hat{\mu} = \text{reshape}(\hat{\tau})$. Thus, the two conditions are preserved.

Update (iii): The update $\sigma'_i = \phi_{w,i}^{-1}(\sigma)\text{softmax}(\ln \phi_{\pi,i}^{-1}(\sigma) + \overline{d\theta}_i)$ preserves $\phi_w^{-1}(\sigma) = w$ such that $\phi_w^{-1}(\sigma') = w$ as well, and then $w = w_{\text{init}}$ is also preserved. Since μ and τ remain unchanged, we have $\mu = \text{reshape}(\tau)$ and $w = \text{norm}([\mathbf{1}^\top \mu_i])$. Proposition 12 shows that $\overline{d\theta} = \widetilde{d\theta}$ for either the differential equations or the Newton equations, then $\text{softmax}(\ln \phi_{\pi,i}^{-1}(\sigma) + \overline{d\theta}_i) = \pi'_i$ given $\sigma = \phi(\pi, w)$, which indicates $\sigma' = \phi(\pi', w)$. Thus, the two conditions are preserved.

Therefore, the conditions hold at the starting points, and the updates preserve them, so they hold constantly during the iteration. $\square$

In Theorem 13 (iii), σ is updated by $\sigma'_i = \phi_{w,i}^{-1}(\sigma)\text{softmax}(\ln \phi_{\pi,i}^{-1}(\sigma) + \overline{d\theta}_i)$, instead of the Paper 1 version $\sigma' = \text{softmax}(\ln \sigma + \overline{d\theta})$. In fact, the former is a second-order correction of the latter to enforce w as an exact constant on top of the $dw = 0$ implied by Proposition 12. The difference between the two updates for σ is derived as follows, where $d\pi_i = \pi_i \circ \widetilde{d\theta}_i$ in equation (4.13) and $\overline{d\theta} = \widetilde{d\theta}$ in Proposition 12 are used in the second line.

$$\begin{aligned} & \text{softmax}_i(\ln \sigma + \overline{d\theta}) - \phi_{w,i}^{-1}(\sigma)\text{softmax}(\ln \phi_{\pi,i}^{-1}(\sigma) + \overline{d\theta}_i) \\ &= \sigma_i + \sigma_i \circ \overline{d\theta}_i + o(\overline{d\theta}_i) - \phi_{w,i}^{-1}(\sigma)(\phi_{\pi,i}^{-1}(\sigma) + \phi_{\pi,i}^{-1}(\sigma) \circ \overline{d\theta}_i + o(\overline{d\theta}_i)) = o(\overline{d\theta}_i) \end{aligned}$$

A second-order correction $o(\overline{d\theta}_i)$ at each update does not change the convergence rate of the original predictor-corrector framework [20].

Theorem 13 shows that the predictor-corrector path-following algorithms on $\bar{E} \rightarrow \Delta_{mn}^+$ and $E \rightarrow \Delta_m^\circ \times \Delta_n^m \rightarrow \Delta_n^m$ generate sequences of (σ, μ) and (π, τ) that correspond exactly via the homeomorphism Φ . Furthermore, w remains constant during the entire iteration. This can be used to fill the final gap by constructing a smooth $\text{VI}(\bar{F}, \Delta_{mn})$ on the closed simplex such that all results in Paper 1 apply. Specifically, we want a map $\bar{F}$ that is real-analytic on the closed simplex Δ_{mn} , such that $\bar{F}$ and $\tilde{F}$ have the same value on the subset $W(w_{\text{init}}) = \{\sigma \in \Delta_{mn} | \phi_{w,i}^{-1}(\sigma) = w_{\text{init}}\}$, which

is given by equation (5.8).

$$\begin{aligned}\tilde{F}_i(\sigma) &= \frac{1}{w_{\text{init},i}} \left(F_i(\tilde{\pi}) - (\tilde{\pi}_i^\top F_i(\tilde{\pi})) \mathbf{1} \right) \\ \text{s.t. } \tilde{\pi}_i(\sigma) &= \left(1 - \frac{\phi_{w,i}^{-1}(\sigma)}{\tilde{w}_i(\sigma)} \right) \pi_{\text{init},i} + \frac{\sigma_i}{\tilde{w}_i(\sigma)} \\ \tilde{w}(\sigma) &= \phi_w^{-1}(\sigma) + (w_{\text{init}} - \phi_w^{-1}(\sigma))^2\end{aligned}\tag{5.8}$$

For every $\sigma \in \Delta_{mn}$, we have $\tilde{\pi}(\sigma) \in \Delta_n^m$, because there is $\mathbf{1}^\top \tilde{\pi}_i(\sigma) = 1$ from $\mathbf{1}^\top \sigma_i = \phi_{w,i}^{-1}(\sigma)$ and $\mathbf{1}^\top \pi_{\text{init},i} = 1$, and $\tilde{\pi}_i(\sigma) \geq 0$ from $\tilde{w}(\sigma) \geq \phi_w^{-1}(\sigma)$. Considering also that F is real-analytic on Δ_n^m and the denominators $w_{\text{init}} > 0$ and $\tilde{w}(\sigma) > 0$, we obtain that $\tilde{F}(\sigma)$ is real-analytic on the simplex Δ_{mn} . Furthermore, when $\phi_w^{-1}(\sigma) = w_{\text{init}}$, we have $\tilde{w}(\sigma) = w_{\text{init}}$ and $\tilde{\pi}(\sigma) = \phi_\pi^{-1}(\sigma)$, then $\tilde{F}(\sigma) = \bar{F}(\sigma)$. Thus, $\tilde{F}$ and $\bar{F}$ have the same value on the subset $W(w_{\text{init}})$.

We also need to compare the derivatives of $\tilde{F}$ and $\bar{F}$ in addition to their function values.

Proposition 14. *On $W(w_{\text{init}})$, the differential equation (5.1) or Newton equation (5.2) remains the same when the operator $\bar{F}$ is replaced by $\tilde{F}$.*

Proof. Proposition 12 shows that $dw = 0$ in differential equation (5.1) and Newton equation (5.2). For the differentials $d\tilde{F}(\sigma)$ and $d\bar{F}(\sigma)$, if $\sigma \in W(w_{\text{init}})$ and $dw = 0$, we have $d\tilde{F}(\sigma) = d\bar{F}(\sigma)$ since $\tilde{F}(\sigma)$ and $\bar{F}(\sigma)$ have the same value on $W(w_{\text{init}})$. Then, expanding $d\tilde{F}(\sigma)$ and $d\bar{F}(\sigma)$, we have

$$\frac{\partial \tilde{F}(\sigma)}{\partial \sigma} \circ \sigma \circ d\bar{\theta} = \frac{\partial \bar{F}(\sigma)}{\partial \sigma} \circ \sigma \circ d\bar{\theta}.$$

Note that the differential equations and Newton equations depend on the operator $\bar{F}$ only through its value $\bar{F}(\sigma)$ and its differential $(\partial \bar{F}(\sigma)/\partial \sigma) \circ \sigma \circ d\bar{\theta}$. Thus, we obtain that on $W(w_{\text{init}})$, the differential equations or Newton equations are the same for operators $\bar{F}$ and $\tilde{F}$. $\square$

Theorem 13 shows that the update of the predictor-corrector path-following algorithm keeps σ constantly on $W(w_{\text{init}})$. Theorem 14 shows that on $W(w_{\text{init}})$, the differential equations or Newton equations are the same for operators $\bar{F}$ and $\tilde{F}$. Therefore, the predictor-corrector path-following algorithms applied to $\text{VI}(\bar{F}, \Delta_{mn}^+)$ and $\text{VI}(\tilde{F}, \Delta_{mn})$ produce exactly the same iteration sequence.

The problem $\text{VI}(\tilde{F}, \Delta_{mn})$ is a real-analytic VI on the closed simplex such that all the results in Paper 1 apply to it. Then, the theoretical guarantees of the path-following on the fixed-point bundle transfer from $\text{VI}(\tilde{F}, \Delta_{mn})$ to $\text{VI}(\bar{F}, \Delta_{mn}^+)$, and then to $\text{VI}(F, \Delta_n^m)$.

5.3 The complete path-following algorithm

In the last subsection, we adopt a singularity avoidance mechanism for $\text{VI}(F, \Delta_n^m)$ corresponding to the one for $\text{VI}(\bar{F}, \Delta_{mn}^+)$ in Paper 1, which is setting $\hat{\tau} = \tau + \beta w \odot \pi$ with $w = \text{norm}([\mathbf{1}^\top \tau_i])$ such that $\sum_i \mathbf{1}^\top \tau_i + \beta$ avoids the eigenvalues of $-\text{Diag}[(1/w_i)I]J(\pi)$. Now that we have shown that w can remain unaffected during the iteration process, we can add a mechanism to arbitrarily modify w in singularity avoidance. Specifically, we reallocate $\mathbf{1}^\top \tau_i$ to change $w = \text{norm}([\mathbf{1}^\top \tau_i])$.

Similar to Paper 1, τ must satisfy $\tau > 0$ to ensure path-following leading to solutions of $\text{VI}(F, \Delta_n^m)$. Since $\tilde{\tau}(\pi)$ is a section of the fixed-point bundle, for any $(\pi, \tau) \in E$, we have $\tau_i = \tilde{\tau}_i(\pi) + k_i \pi_i$. Considering $\min_a \tilde{\tau}_{i,a}(\pi) = 0$, we have $\tau > 0$ if and only if $\mathbf{1}^\top \tau_i > \mathbf{1}^\top \tilde{\tau}_i(\pi)$. Thus, we have $\hat{\tau} > 0$ and $\sum_i \mathbf{1}^\top \hat{\tau}_i = \sum_i \mathbf{1}^\top \tau_i + \beta$ if and only if there exists $\tilde{w}_i \in \Delta_m^o$ such that $\hat{\tau}$ satisfies

$$\mathbf{1}^\top \hat{\tau}_i = \tilde{w}_i \left(\sum_i \mathbf{1}^\top \tau_i + \beta - \sum_i \mathbf{1}^\top \tilde{\tau}_i(\pi) \right) + \mathbf{1}^\top \tilde{\tau}_i(\pi).\tag{5.9}$$

In singularity avoidance, we select $\tilde{w} \in \Delta_m^\circ$ and β such that $\sum_i \mathbf{1}^\top \tau_i + \beta$ avoids the eigenvalues of $-\text{Diag}[(1/w_i)I]J(\pi)$ with $w_i = \mathbf{1}^\top \hat{\tau}_i / (\sum_j \mathbf{1}^\top \hat{\tau}_j)$ given by equation (5.9). Then, the $\hat{\tau}$ after singularity avoidance is $\hat{\tau} = \tilde{\tau}(\pi) + (\sum_i \mathbf{1}^\top \tau_i + \beta)w \odot \pi$. Our experimental results are based on this singularity avoidance mechanism with $\tilde{w} \in \Delta_m^\circ$ randomly selected, which has better performance than keeping w fixed.

The path-following algorithm for $\text{VI}(F, \Delta_n^m)$ is given in Algorithm 1. Apart from its singularity avoidance step, it corresponds to Algorithm 1 in Paper 1 via the equivalence established in this paper.

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

Require: A smooth map $F : \Delta_n^m \rightarrow \mathbb{R}^{m \times n}$ and its derivative $[\partial F_i / \partial \pi_j]$, a designated starting point $\pi_{\text{init}} \in \Delta_n^m$, and a desired precision $\epsilon > 0$

- 1: Set $(\pi_0, \tau_0) = (\pi_{\text{init}}, v_{\text{init}} w_{\text{init}} \odot \pi_{\text{init}})$ for a $w_{\text{init}} \in \Delta_m^\circ$ and a sufficiently large v_{init}
- 2: **repeat**
- 3: **repeat**
- 4: Compute $(\hat{\pi}_k, r_k, v_k) = M(\pi_k, \tau_t)$ by solving bisection problem (3.4)
- 5: Construct the matrix $J(\pi_k)$ in equation (4.16)
- 6: Solve regularized Newton equation (5.6) or (5.7) for $\tilde{d\theta}_k$ with $\delta_H = \|G(\pi_k, \tau_t)\|/(mn)$
- 7: Update $\pi_{k+1,i} = \text{softmax}(\ln \pi_{k,i} + \tilde{d\theta}_{k,i})$
- 8: **until** $\|G(\pi_k, \tau_t)\| \leq \epsilon/(mn)$
- 9: Set $\hat{\tau} = \tilde{\tau}(\pi) + (\sum_i \mathbf{1}^\top \tau_i + \beta)w \odot \pi$ with $w_i = \mathbf{1}^\top \hat{\tau}_i / (\sum_j \mathbf{1}^\top \hat{\tau}_j)$ given by equation (5.9), such that $\sum_i \mathbf{1}^\top \tau_{t,i} + \beta_t$ avoids the nonzero eigenvalues of $-\text{Diag}[(1/w_i)I]J(\pi_k)$
- 10: Update and truncate $\tau_{t+1} = ((1 - \eta_t)\hat{\tau}_t) \cdot \text{clip}(\min = \epsilon/(mn))$ with a sufficiently small η_t
- 11: Solve differential equation (5.3) for $\tilde{d\theta}_t$ with $d\tau_t = \tau_{t+1} - \hat{\tau}_t$
- 12: Set $\pi_{0,i} = \text{softmax}(\ln \pi_{k,i} + \tilde{d\theta}_{t,i})$
- 13: **until** $\sum_i \mathbf{1}^\top \tilde{\tau}_i(\pi_k) \leq \epsilon$
- 14: **return** π_k as an approximate solution of $\text{VI}(F, \Delta_n^m)$

The corrector can be either regularized Newton equation (5.6) or regularized Newton equation (5.7). Similar to Paper 1, the former mixes the standard Newton equation (5.4) and the inexact gradient $J_G^\top[G_i(\pi, \tau)/\pi_i]$ of $\|G_i(\pi, \tau)/\pi_i\|$, while the latter mixes the standard Newton equation (5.4) and the inexact gradient $J_G^\top[\pi_i - \hat{\pi}_i]$ of $\text{mcp}(\pi, \tau)$.

A difference from Paper 1 is that the performance of regularized Newton corrector (5.6) is not ideal for solving $\text{VI}(F, \Delta_n^m)$. This is because the τ/π term in corrector (5.6) frequently causes overflow in problems with large dimension, whereas the similar term τ/r in corrector (5.7) does not, because it is subject to the bisection problem (3.4). The overflow could cause $G(\pi, \tau)$ to fail to be corrected to 0, triggering singularity avoidance that practically lets the algorithm try to approach the solution in another direction where π has more even components. This leads to much higher iteration cost for $\text{VI}(F, \Delta_n^m)$ on simplex product than for VIs on simplex in Paper 1, because $\text{VI}(F, \Delta_n^m)$ has much more nonlinearity, which can be seen from the rapid increase in the number of singularity encounters as m increases. Thus, we only test the algorithm with corrector (5.7) in experiments.

6 Experiments

We test the algorithm on randomly generated instances of $\text{VI}(F, \Delta_n^m)$ with F represented by two different models: a neural network and a payoff tensor. The neural network representation models

VI problems with a general real-analytic operator. In our experiments, we use a neural network with architecture $[mn, 50, mn]$ and tanh activation to represent F , with parameters generated from a uniform distribution.

The payoff tensor representation models finite normal-form games. For m -player n -action finite normal-form games, the payoff function is a multilinear function given by an m -th order payoff tensor of shape $(n, \dots, n)$, which consists of n^m elements, growing exponentially with m . We use the classical low-rank CP (canonical polyadic) representation as a practical alternative to the full tensor representation [21]. Specifically, F is represented by $F_i(\pi) = \nabla_{\pi_i} f_i(\pi)$, where

$$f_i(\pi) = \sum_r \prod_{j=1}^m (U_{r,i,j} \pi_j) \quad (6.1)$$

is the low-rank CP payoff function. Each U_i is a third-order tensor of shape (R, m, n) , reducing the total storage requirement from n^m to Rmn . In our experiments, each element $u \in U$ is generated from a normal distribution with mean 1 and variance $1/m$, so that $\prod_{j=1}^m u_j$ has mean 1 and variance near $e - 1$, preventing overflow or underflow for very large m . We use $R = 50$ in the experiments.

Table 1: Statistics (mean / median) solving VI(F, Δ_n^m)

m, n	Neural network F		Payoff F	
	Iteration count	Singularity avoidance times	Iteration count	Singularity avoidance times
8, 2	375 / 353	0.9 / 0.0	370 / 359	0.3 / 0.0
8, 4	500 / 409	5.0 / 0.0	451 / 440	2.1 / 0.0
8, 8	658 / 553	10.8 / 7.0	507 / 436	3.8 / 0.0
8, 16	839 / 704	18.4 / 13.5	579 / 511	6.3 / 1.0
8, 32	1191 / 1006	34.8 / 28.0	735 / 641	12.3 / 9.0
2, 8	381 / 363	1.0 / 0.0	422 / 413	0.3 / 0.0
4, 8	442 / 377	3.5 / 0.0	446 / 415	1.4 / 0.0
8, 8	658 / 553	10.8 / 7.0	507 / 436	3.8 / 0.0
16, 8	1666 / 1301	46.3 / 33.5	634 / 552	8.9 / 5.5
32, 8	7481 / 4902	218.1 / 147.5	1124 / 937	27.7 / 21.0
2, 128	480 / 401	5.3 / 0.0	492 / 455	1.7 / 0.0
4, 64	701 / 626	15.5 / 11.5	575 / 482	5.4 / 0.0
8, 32	1191 / 1006	34.8 / 28.0	735 / 641	12.3 / 9.0
16, 16	3178 / 2013	101.2 / 58.0	964 / 841	21.7 / 17.0
32, 8	7481 / 4902	218.1 / 147.5	1124 / 937	27.7 / 21.0
64, 4	9676 / 5811	274.5 / 171.0	1387 / 1064	37.7 / 26.0
128, 2	5217 / 3934	152.8 / 117.0	839 / 712	17.8 / 12.5

The starting point π_{init} is also randomly generated, and the required precision is $\epsilon = 10^{-5}$. We solve the problems until $\text{gap}_i(\pi) < \epsilon$ rather than $\text{gap}(\pi) = \sum_i \text{gap}_i(\pi) < \epsilon$. For dimensions, we fix the player number $m = 8$ and vary the action number n from 2 to 32, and fix the action number $n = 8$ and vary the player number m from 2 to 32. Since the algorithm operates in dimension mn , we also fix $mn = 256$ and test (m, n) ranging from (2, 128) to (128, 2). For each data point, we run 200 instances, totaling 5600 instances. We test the algorithm only with the regularized Newton

corrector (5.7). Table 1 reports the mean and median iteration counts and singularity avoidance times.

The algorithm succeeds on all 5600 instances. When m is fixed, the iteration count and singularity avoidance count increase mildly as n increases; whereas when n is fixed, they increase dramatically as m increases. When mn is fixed, the peak counts appear in the middle rather than at the endpoints $(2, 128)$ or $(128, 2)$, indicating that both m and n affect problem difficulty for a fixed total dimension mn . In addition, solving $\text{VI}(F, \Delta_n^m)$ with payoff F is significantly easier than with neural network F , because neural networks produce greater nonlinearity and a more twisted singular manifold, causing much more singularity avoidance.

7 Conclusion

This paper generalizes the fixed-point bundle framework in Paper 1 for VIs with simplex domain to VIs with product-of-simplex domain. The fixed-point bundle $E \rightarrow \Delta_n^m \times \Delta_m^\circ \rightarrow \Delta_n^m$ of $\text{VI}(F, \Delta_n^m)$ has a composite fiber bundle structure.

We construct a $\text{VI}(\bar{F}, \Delta_{mn}^+)$ with simplex domain from $\text{VI}(F, \Delta_n^m)$ that preserves the solutions. The central innovation of this paper is to establish the equivalence between their fixed-point bundle frameworks, based on $\bar{E} \rightarrow \Delta_{mn}^+$ and $E \rightarrow \Delta_n^m \times \Delta_m^\circ$, via the bundle isomorphism $(\Phi : E \rightarrow \bar{E}, \phi : \Delta_n^m \times \Delta_m^\circ \rightarrow \Delta_{mn}^+)$.

The path-following algorithms on the fixed-point bundles keep w constant. The map $\bar{F}$ restricted to the slice $w = w_{\text{init}}$ can be extended to a real-analytic map $\tilde{F}$ on the closed simplex Δ_{mn} . Then, all the results in Paper 1 apply to $\text{VI}(\tilde{F}, \Delta_{mn})$. Thus, the path-following algorithm for $\text{VI}(F, \Delta_n^m)$ has the same convergence guarantee as in Paper 1. Numerical experiments on 5600 randomly generated problems with dimensions ranging from 2-player 128-action to 128-player 2-action confirm universal success.

The singularity avoidance in the framework involves arbitrarily selecting a new w , and we currently adopt the most basic random selection method. Although simple and effective in experiments, computing w based on the state of the path-following process could further improve practical performance.

References

- [1] John Nash. Non-cooperative games. *Annals of mathematics*, 54(2):286–295, 1951.
- [2] David M Kreps. *Game theory and economic modelling*. Oxford University Press, 1990.
- [3] Zhu Han. *Game theory in wireless and communication networks: theory, models, and applications*. Cambridge university press, 2012.
- [4] Tanmoy Hazra and Kushal Anjaria. Applications of game theory in deep learning: a survey. *Multimedia Tools and Applications*, 81(6):8963–8994, 2022.
- [5] Francisco Facchinei and Jong-Shi Pang. *Finite-dimensional variational inequalities and complementarity problems*. Springer, 2003.
- [6] Francisco Facchinei, Andreas Fischer, and Veronica Piccialli. On generalized nash games and variational inequalities. *Operations Research Letters*, 35(2):159–164, 2007.

- [7] Giuseppe Belgioioso and Sergio Grammatico. Projected-gradient algorithms for generalized equilibrium seeking in aggregative games are preconditioned forward-backward methods. *arXiv preprint arXiv:1803.10441*, 2018.
- [8] Peng Yi and Lacra Pavel. Distributed generalized nash equilibria computation of monotone games via double-layer preconditioned proximal-point algorithms. *IEEE Transactions on Control of Network Systems*, 6(1):299–311, 2018.
- [9] Axel Dreves, Anna von Heusinger, Christian Kanzow, and Masao Fukushima. A globalized newton method for the computation of normalized nash equilibria. *Journal of Global Optimization*, 56(2):327–340, 2013.
- [10] Axel Dreves, Francisco Facchinei, Christian Kanzow, and Simone Sagratella. On the solution of the kkt conditions of generalized nash equilibrium problems. *SIAM Journal on Optimization*, 21(3):1082–1108, 2011.
- [11] P Jean-Jacques Herings and Ronald Peeters. Homotopy methods to compute equilibria in game theory. *Economic Theory*, 42(1):119–156, 2010.
- [12] Peng Yi and Lacra Pavel. An operator splitting approach for distributed generalized nash equilibria computation. *Automatica*, 102:111–121, 2019.
- [13] Nguyen The Vinh and Le Dung Muu. Inertial extragradient algorithms for solving equilibrium problems. *Acta Mathematica Vietnamica*, 44(3):639–663, 2019.
- [14] Barbara Franci and Sergio Grammatico. Stochastic generalized nash equilibrium seeking under partial-decision information. *Automatica*, 137:110101, 2022.
- [15] Hongbo Sun. A path-following framework on fiber bundle for variational inequalities, 2026.
- [16] Peter M Gruber. *Convex and discrete geometry*. Springer, 2007.
- [17] Steven George Krantz. *Function theory of several complex variables*, volume 340. American Mathematical Soc., 2001.
- [18] Walter Rudin. *Real and complex analysis*. McGraw-Hill Education, 1974.
- [19] R Tyrrell Rockafellar and Roger JB Wets. *Variational analysis*. Springer, 1998.
- [20] Eugene L Allgower and Kurt Georg. *Introduction to numerical continuation methods*. SIAM, 2003.
- [21] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.