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

Hongbo Sun
shb20@tsinghua.org.cn

Abstract

Developing general-purpose algorithms with global asymptotic convergence guarantees for Nash equilibria of general-sum games remains a longstanding challenge. This paper extends the fixed-point bundle framework for finite-dimensional variational inequalities (VIs) from simplex domains to product-of-simplex domains, enabling its direct application to Nash equilibrium computation. The resulting fixed-point bundle exhibits a composite fiber bundle structure. The key innovation is to construct an equivalent VI on a simplex and establish the equivalence between the two fixed-point bundle frameworks via a fiber bundle isomorphism. Exploiting this geometric equivalence, we show that the path-following algorithm for VIs on product-of-simplex domains inherits the global linear convergence guarantees of the simplex-domain framework. In numerical experiments, the algorithm attained a precision of 10^{-5} on all 5600 randomly generated game instances, with dimension ranging from 2-player 128-action to 128-player 2-action. Code and data are available at `https://github.com/shb20tsinghua/FiberBundle_VI`.

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

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

Contents

1 Introduction **2**

2 Preprocessing: smooth VI over product-of-simplex **4**

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, Δ_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 21
5.3 The complete path-following algorithm 23

6 Numerical experiments **24**

7 Conclusion **25**

1 Introduction

Game equilibrium problems, introduced by Nash [1], are central in economics [2], engineering [3], artificial intelligence [4], and other fields. 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 some regularity 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 and $H : K \to \mathbb{R}^n$ be a continuous function. The VI problem $\mathrm{VI}(H, K)$ asks for $x^\star \in K$ such that

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

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

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

where $\mathrm{gap}(x) \geq 0$ for all $x \in K$, and $x^\star \in K$ is a solution if and only if $\mathrm{gap}(x^\star) = 0$. 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 $\mathrm{VI}(H, K)$ is a game equilibrium problem, with K_i the strategy set of each player and f_i the payoff function of each player.

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 monotone operators. Modern variants add adaptive step sizes, inertial terms, or stochastic estimates for large-scale strategic interactions [13, 14].

Besides VI algorithms, there are also machine-learning algorithms that operate at the level of individual players, such as self-play dynamics [15, 16] that update each player's strategy with the best-response strategy, and no-regret dynamics [17, 18] that update each player's strategy to decrease the regret relative to each pure action.

Developing general-purpose and efficient algorithms for Nash equilibria remains a longstanding challenge. First, computing a Nash equilibrium of a general-sum normal-form game is PPAD-complete [19], so no polynomial-time algorithm is expected unless PPAD collapses. Moving from polynomial-time guarantees to a broader asymptotic convergence perspective, existing VI algorithms lack global convergence guarantees without additional structural assumptions, and machine-learning algorithms do not guarantee convergence to a Nash equilibrium either. In short, in terms of iterative methods with asymptotic convergence, no existing method guarantees global convergence to a Nash equilibrium of an arbitrary general-sum normal-form game.

This paper mainly deals with VIs with the product-of-simplex domain, which is the canonical domain for mixed strategies of normal-form games. We extend the fixed-point bundle framework in

Paper 1 [20] for VIs on simplex domains to VIs on product-of-simplex domains. In Paper 1, the fixed-point bundle framework formulates the solution space of a VI on a simplex domain as a fiber bundle, in which starting point selection, path-following, and singularity avoidance are systematically integrated. The path-following algorithm applies to VIs with general continuous functions and general polytope domains after preprocessing, and it guarantees global linear convergence for nonsingular solutions and retains global linear convergence up to a prescribed precision for singular solutions. This paper establishes an equivalence between the fixed-point bundle framework over product-of-simplex domains and that over simplex domains, showing that the path-following algorithm over product-of-simplex domains has the same convergence result as in Paper 1.

The proposed approach begins by preprocessing the general $\mathrm{VI}(H, K)$ into $\mathrm{VI}(F, \Delta_n^m)$ with a real-analytic operator F and domain Δ_n^m (the product of m n-dimensional simplices), in correspondence with the preprocessing in Paper 1.

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

Based on the homeomorphism $\phi : \Delta_n^m \times \Delta_m^\circ \to \Delta_{mn}^+$, an operator $\bar{F}$ can be constructed such that $\mathrm{VI}(\bar{F}, \Delta_{mn}^+)$ and $\mathrm{VI}(F, \Delta_n^m)$ have the same solutions. The central innovation of this paper is to establish the equivalence between the fixed-point bundle frameworks of $\mathrm{VI}(F, \Delta_n^m)$ and $\mathrm{VI}(\bar{F}, \Delta_{mn}^+)$ via a fiber bundle isomorphism. The fixed-point bundle framework can be defined for $\mathrm{VI}(\bar{F}, \Delta_{mn}^+)$, but the theoretical guarantees in Paper 1 require an additional extension argument to apply, since $\mathrm{VI}(\bar{F}, \Delta_{mn}^+)$ is not a smooth VI on the closed simplex.

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

In experiments, numerical results on 5600 randomly generated instances, with dimensions ranging from 2-player 128-action to 128-player 2-action, show that the algorithm reduces the gap function below 10^{-5} in every case.

The remainder of the paper is organized as follows. Section 2 details the preprocessing. Section 3 generalizes the three equivalent characterizations and the fixed-point bundle. Section 4 constructs a VI on a simplex and establishes its equivalence to the VI on the 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, $a \circ b$ denotes componentwise multiplication and a/b componentwise division, and the inequality $a \geq b$ is also interpreted componentwise. For a matrix A, componentwise multiplication is interpreted as $A \circ b = A\mathrm{diag}(b)$ and $b \circ A = \mathrm{diag}(b)A$. For tensors, $\mathbb{R}^{mn}$ is a vector space of dimension mn, $\mathbb{R}^{m \times n}$ is a tensor space of shape $m \times n$, and the operation $\mathrm{reshape}(\cdot)$ always represents switching between these two shapes. For block stacking, $[a_i]$ represents the concatenation of small vectors a_i into a large vector, and $\mathrm{Diag}[x_i I]$ represents a block diagonal matrix. The operation $\mathrm{norm}([x_i]) = [x_i]/(\sum_j x_j)$ denotes normalization. The operation $w \odot \pi$ is defined for $w \in \mathbb{R}^m$ and $\pi \in \mathbb{R}^{m \times n}$ by $w \odot \pi \in \mathbb{R}^{m \times n}$ with $(w \odot \pi)_i = w_i \pi_i$. Finally, $\mathbf{1}$ is the all-one vector; it sometimes has dimension n and sometimes dimension mn, but its dimension can always be inferred from the context.

2 Preprocessing: smooth VI over product-of-simplex

As in Paper 1, we first transform the general VI(H, K) with a continuous map and polytope domain into VI(F, Δ_n^m) with a smooth map and a simplex product domain. The subsequent theory is then based on VI(F, Δ_n^m).

In Paper 1, the fixed-point bundle method applies to VIs over polytope domains, where polytopes and simplices are in one-to-one correspondence via linear transformations when they have the same number of vertices. The linear transformation $x = [v_1, \ldots, v_n]\,\sigma$ transforms a simplex Δ to a polytope with vertices $v_1, \ldots, v_n$. In this paper, a linear transformation

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

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

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, a polytope K with a product structure $K = K_1 \times \cdots \times K_m$ is a Minkowski sum in a higher-dimensional space and has $\prod_i n_i$ vertices. Thus, representing a polytope as a linear transformation of a simplex product can reduce the input vertex budget from $\prod_i n_i$ to $\sum_i n_i$ compared with the plain simplex representation, which may yield a substantial compression and result in a dimension reduction for the fixed-point bundle method.

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

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

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

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

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

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

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

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

where $\epsilon > 0$ is a scalar that measures the approximation error, and $c_i > 0$ is a constant vector ensuring that $\pi_i/\epsilon + c_i > 0$ 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 this from scratch, we reuse the analyticity result for single-variable Dirichlet distribution averaging in

Paper 1. Hartogs' theorem on separate holomorphy [21] from complex analysis can then be used to extend this result to analyticity for multiple-variable Dirichlet distribution averaging. This theorem states that for a function $g : U \times V \to \mathbb{C}$ 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 U, and for every fixed $v \in V$, the map $u \mapsto g(u, v)$ is holomorphic on V, 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 $U_\sigma = \{z \in \mathbb{C}^n | \|z - \sigma\| \le d\}$.

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

First, the operators T_j are commutative. The operation $(T_1 \circ \cdots \circ T_m F_{i,a})(z_1, \ldots, z_m)$ is an integral of the continuous function $F_{i,a}(y_1, \ldots, y_m)$ over the compact region Δ_n^m against the product measure $\prod_{j=1}^m \mathrm{Dir}(y_j; z_j/\epsilon + c_j) dy_j$. By the Fubini-Tonelli theorem [22], 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 write $F_{\epsilon,i,a} = T_k \circ T_{\hat{k}} F_{i,a}$.

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: 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, for fixed $z_j \in U_{\pi_j}$ with $j \neq k$, we have

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

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 of Paper 1 applies to $F_{\epsilon,i,a} = T_k(T_{\hat{k}} F_{i,a})$, and we obtain that for every fixed $z_j \in U_{\pi_j}$ with $j \neq k$, 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. □

3 The fixed-point bundle of VI on simplex product

3.1 Generalizing the three equivalent characterizations

As in Paper 1, we deal with $\mathrm{VI}(F, \Delta_n^m)$ assuming that 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 $\mathrm{VI}(\bar{F}, \Delta)$. We therefore begin by transforming these three characterizations so as to characterize $\mathrm{VI}(F, \Delta_n^m)$. First, we show that $\mathrm{VI}(F, \Delta_n^m)$ can be decomposed into several VIs on Δ_n via its gap function.

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

$$\mathrm{gap}(\pi) = \sum_i \mathrm{gap}_i(\pi_i, \pi_{\hat{i}}), \quad \text{where } \mathrm{gap}_i(\pi_i, \pi_{\hat{i}}) := \sup_{\pi'_i \in \Delta_n} \left\langle F_i(\pi_i, \pi_{\hat{i}}), \pi_i - \pi'_i \right\rangle \tag{3.1}$$

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

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

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

Thus, we obtain $\mathrm{gap}(\pi) = \sum_i \mathrm{gap}_i(\pi_i, \pi_{\hat{i}})$.

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

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

Since $\mathrm{VI}(F_i(\cdot, \pi_{\hat{i}}), \Delta_n)$ is a VI on a simplex, the three equivalent characterizations in Paper 1 apply to it. However, they are compatible over the index i only pointwise for individual $\pi \in \Delta_n^m$, at which point all the parameters $\pi_{\hat{i}}$ and variables π_i across the indices 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), and 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 $\mathrm{VI}(F_i(\cdot, \pi_{\hat{i}}), \Delta_n)$ over the index i at each individual $\pi \in \Delta_n^m$, we obtain the generalization characterizing $\mathrm{VI}(F, \Delta_n^m)$. 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.

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

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

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 the perturbed KKT conditions (3.2a). This correspondence can be shown to be a continuous function, to which the Brouwer fixed-point theorem applies.

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

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

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

Mixed complementarity problem form. Equation (3.5) is the (barriered) mixed complementarity problem (MCP), which is obtained from FPKKT (3.2) by taking the complementarity term

$\pi_i \circ r_i$ as 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 $\check{\pi} = \pi$ and the perturbed KKT conditions of MCP (3.5), where $\check{\pi}$ is the Lagrangian multiplier associated with the constraint $r \geq 0$.*

Analogously to Paper 1, FPKKT (3.2) characterizes the paths leading to solutions of $\mathrm{VI}(F, \Delta_n^m)$, 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 $\mathrm{VI}(F, \Delta_n^m)$, 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 $\mathrm{VI}(F, \Delta_n^m)$ and some $\mathrm{VI}(\bar{F}, \Delta)$ to which Paper 1 applies, instead of transferring the theoretical guarantees directly onto $\mathrm{VI}(F, \Delta_n^m)$.

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 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, setting $\tilde{\tau}_i(\pi) := \pi_i \circ F_i(\pi) - (\pi_i^\top F_i(\pi))\pi_i$ 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 those 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 $\alpha : E \to \Delta_n^m$ 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 reveals a composite fiber bundle structure, in which 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 $\tilde{B}(\pi, w_1) \cap \tilde{B}(\pi, w_2) = \varnothing$.*

Proof. Suppose that 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 $w_1 \neq w_2$, then $\tilde{B}(\pi, w_1) \cap \tilde{B}(\pi, w_2) = \varnothing$. □

Lemma 4 shows that for every π, $B(\pi)$ is indeed a fiber bundle. Since $\bar{v} = 0$ 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 $B(\pi)$, although $B(\pi)$ 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.

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

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

- Bundle $E \to \Delta_n^m \times \Delta_m^\circ$: 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 $(\pi, \operatorname{norm}([\mathbf{1}^\top \tau_i])) \in \Delta_n^m \times \Delta_m^\circ$.
- Bundle $\Delta_n^m \times \Delta_m^\circ \to \Delta_n^m$: 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 $E \to \Delta_n^m$: 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, $E \to \Delta_n^m \times \Delta_m^\circ$ has a large base space and small fibers, whereas $E \to \Delta_n^m$ has a small base space and large fibers. Each fiber $\{\pi\} \times B(\pi)$ of $E \to \Delta_n^m$ is the union of all the fibers $\{\pi\} \times \tilde{B}(\pi, w)$ of $E \to \Delta_n^m \times \Delta_m^\circ$ 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 given by equation (3.8).

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

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

through $\mathrm{gap}_i(\pi_i, \pi_{\hat{i}}) = \mathbf{1}^\top \check{\tau}_i(\pi)$. The section $\tilde{\tau}(\pi)$ satisfies $\mathbf{1}^\top \tilde{\tau}_i(\pi) = 0$ and governs the differential properties of the composite fixed-point bundle, as discussed later.

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

4 Equivalence between VI on simplex product and VI on simplex

The equivalence between $\mathrm{VI}(F, \Delta_n^m)$ and $\mathrm{VI}(F_i(\cdot, \pi_{\hat{i}}), \Delta_n)$ in the previous section relies on a fixed parameter $\pi_{\hat{i}}$, so it holds only pointwise for each individual π and does not apply when π varies or in global situations. In this section, we construct a $\mathrm{VI}(\bar{F}, \Delta_{mn}^+)$ that preserves the solutions of $\mathrm{VI}(F, \Delta_n^m)$, and we establish the equivalence between their fixed-point bundle frameworks. The fixed-point bundle framework can be defined for $\mathrm{VI}(\bar{F}, \Delta_{mn}^+)$, but any theoretical guarantee that Paper 1 provides requires an additional extension argument in the next section to transfer.

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

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

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

Thus, for any $w^\star \in \Delta_m^\circ$, $(\pi^\star, w^\star)$ is a solution of $\mathrm{VI}(\check{F}, \Delta_n^m \times \Delta_m^\circ)$ if and only if $\pi^\star$ is a solution of $\mathrm{VI}(F, \Delta_n^m)$.

Viewing each π_i as a conditional distribution and w as a marginal distribution, we can transform (π, w) into a joint distribution σ on a simplex. This yields a real-analytic homeomorphism $\phi : \Delta_n^m \times \Delta_m^\circ \to \Delta_{mn}^+$ defined as follows.

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

We denote

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

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, since they consist only of elementary operations and the denominator $\sum_a \sigma_{i,a} > 0$ on Δ_{mn}^+. Thus $\phi : \Delta_n^m \times \Delta_m^\circ \to \Delta_{mn}^+$ is a real-analytic homeomorphism.

Next, we use the homeomorphism $\phi : \Delta_n^m \times \Delta_m^\circ \to \Delta_{mn}^+$ to transform $\mathrm{VI}(\check{F}, \Delta_n^m \times \Delta_m^\circ)$ into a VI on Δ_{mn}^+, following the standard procedure for changing the variable of a VI [23]. For a general variational inequality $\mathrm{VI}(H, K)$, a standard theorem states that $x^\star \in K$ solves $\mathrm{VI}(H, K)$ if and only if $-H(x^\star) \in N_K(x^\star)$, where $N_K(x) = \{v \in \mathbb{R}^n | \langle v, x' - x \rangle \le 0, \forall x' \in K\}$ is the normal cone. Another standard theorem on the 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 $x^\star \in K$ is a solution of $\mathrm{VI}(H, K)$ if and only

if $z^\star = \psi(x^\star)$ is a solution of $\mathrm{VI}(\bar{H}, \psi(K))$, which is a transformation obtained by changing the variable.

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 $dz = (\partial\psi(x)/\partial x)dx$. We use this equation to find $\bar{F}(\sigma)$ that changes the variable of $\mathrm{VI}(\check{F}, \Delta_n^m \times \Delta_m^\circ)$ via $\sigma = \phi(\pi, w)$, which is derived as follows.

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

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 $\mathrm{VI}(\bar{F}, \Delta_{mn}^+)$ is obtained from $\mathrm{VI}(\check{F}, \Delta_n^m \times \Delta_m^\circ)$ by changing the variable via $\sigma = \phi(\pi, w)$.

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

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

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

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

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

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

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

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

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

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

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

The problem $\mathrm{VI}(\bar{F}, \Delta_{mn}^+)$ 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, the results on the fixed-point bundle framework in Paper 1 do not apply to $\mathrm{VI}(\bar{F}, \Delta_{mn}^+)$. Nevertheless, we can still substitute $\bar{F}(\sigma)$ into Paper 1 to define the fixed-point bundle framework for $\mathrm{VI}(\bar{F}, \Delta_{mn}^+)$, without invoking any theoretical guarantee provided by the theorems in Paper 1.

In the remainder of this section, we establish the equivalence between the fixed-point bundle framework of $\mathrm{VI}(\bar{F}, \Delta_{mn}^+)$ and the composite fixed-point bundle framework of $\mathrm{VI}(F, \Delta_n^m)$, while the transition from $\mathrm{VI}(\bar{F}, \Delta_{mn}^+)$ to an actual smooth VI on a closed simplex is addressed in the next section.

4.2 Isomorphism between fixed-point bundles

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

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

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

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

Note that $\bar{B}(\sigma)$ consists of $\mu = \mathrm{reshape}(\tilde{\tau}(\pi) + \bar{v} w \odot \pi)$, while $\tilde{B}(\pi, w)$ consists of $\tau = \tilde{\tau}(\pi) + \bar{v} w \odot \pi$. This indicates that the bundles $E \to \Delta_n^m \times \Delta_m^\circ$ and $\bar{E} \to \Delta_{mn}^+$ 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.

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

It can be verified that Φ and Φ^{-1} are indeed inverse maps of each other between E and $\bar{E}$, so Φ is a homeomorphism between E and $\bar{E}$ if it is continuous. We aim, however, 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

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

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*

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

Consequently, Φ is a real-analytic homeomorphism such that

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

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

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

The map $\Phi^{-1}\circ\Phi$ is always the identity map, whereas $\Phi\circ\Phi^{-1}$ is the 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 left-hand side (LHS) and the right-hand side (RHS) of 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 $\sum_i\mathbf{1}^\top\tau_i\neq0$ on $\mathbb{R}_{\pm}^{m\times n}$. Therefore, Φ is a real-analytic homeomorphism between the LHS and RHS in equation (4.7). □

Theorem 6 shows that Φ is a homeomorphism beyond E and $\bar{E}$. This is useful for 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 $\bar{E}$ contain only the solutions of the characterizations, whereas with Φ in equation (4.7) we can compare the values of the characterizations both on and off the solutions.

Next, we show that $E\to\Delta_n^m\times\Delta_m^\circ$ and $\bar{E}\to\Delta_{mn}^{+}$ 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 $(\psi:B\to B',\Psi:E\to E')$ such that $\alpha'\circ\Psi=\psi\circ\alpha$.

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

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

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

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

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 $\Phi : E \to \bar{E}$ and $\bar{\alpha} : \bar{E} \to \Delta_{mn}^+$, 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 $\alpha_1 : E \to \Delta_n^m \times \Delta_m^\circ$ and $\phi : \Delta_n^m \times \Delta_m^\circ \to \Delta_{mn}^+$, 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 $E \to \Delta_n^m \times \Delta_m^\circ$ and $\bar{E} \to \Delta_{mn}^+$. □

4.3 Equivalence between characterizations

Having obtained the homeomorphism in equation (4.7), we use it to establish the equivalence between the characterizations of $\text{VI}(F, \Delta_n^m)$ and $\text{VI}(\bar{F}, \Delta_{mn}^+)$. We first introduce the objects whose equivalence is to be studied, replacing the FPKKTs by the fixed-point bundle equations and adding the sections of the fixed-point bundles.

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

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

Fixed-point bundle equations. The fixed-point bundle equation of $E \to \Delta_n^m$ is given as in equation (3.8). The fixed-point bundle equation of $\bar{E} \to \Delta_{mn}^+$ 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 $\mathrm{VI}(F, \Delta_n^m)$ is given as in equation (3.3), and the Brouwer function characterizing $\mathrm{VI}(\bar{F}, \Delta_{mn}^+)$ is taken 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), rather than to the original Brouwer function in Paper 1.

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

Compared with the original, the index-wise variant $(\hat{\sigma}, \bar{r}, \bar{v}) = \bar{M}(\sigma, \mu)$ has two modifications. 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 becomes 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 $\sum_i \mathbf{1}^\top \hat{\sigma}_i$.

MCPs. The two MCPs characterizing $\mathrm{VI}(F, \Delta_n^m)$ and $\mathrm{VI}(\bar{F}, \Delta_{mn}^+)$ are given as in equation (4.11). The term $\mathrm{mcp}(\pi, \tau)$ 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. The term $\mathrm{mcp}(\sigma, \mu)$ 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; we instead modify $(\bar{r}, \bar{v})$ to be constrained by the index-wise variant $(\hat{\sigma}, \bar{r}, \bar{v}) = \bar{M}(\sigma, \mu)$.

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

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

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 $\phi_w^{-1}(\sigma) = \mathrm{norm}([\mathbf{1}^\top \mu_i])$ 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) $\mathrm{mcp}(\sigma, \mu) = \mathrm{mcp}(\pi, \tau)$.

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

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

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

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

(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 $w=\phi_w^{-1}(\sigma)=\mathrm{norm}([\mathbf{1}^\top\mu_i])=\mathrm{norm}([\mathbf{1}^\top\tau_i])$ in the second line, we have

$$\begin{aligned}\bar{G}_i(\sigma,\mu)&=\tilde{\mu}_i(\sigma)-\Big(\mu_i-(\mathbf{1}^\top\mu)\sigma_i\Big)=\tilde{\tau}_i(\pi)-\big(\tau_i-(\textstyle\sum_i\mathbf{1}^\top\tau_i)w_i\pi_i\big)\\&=\tilde{\tau}_i(\pi)-\Big(\tau_i-(\mathbf{1}^\top\tau_i)\pi_i\Big)=G_i(\pi,\tau).\end{aligned}$$

(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).

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

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 $v_i=\bar{v}_iw_i-\pi_i^\top F_i(\pi)$. 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 $\mathrm{mcp}(\sigma,\mu)=\mathrm{mcp}(\pi,\tau)$. □

Proposition 8 will be used to transform the path-following iterations of the fixed-point bundle framework from $\mathrm{VI}(\bar{F},\Delta_{mn}^+)$ to $\mathrm{VI}(F,\Delta_n^m)$. We will also handle 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 $\sigma\in\Delta$, the variable σ is represented by the softmax function as $\sigma=\mathrm{softmax}(\theta)$, and differentiable properties are then 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 likewise use the softmax function to represent π, so as 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*

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

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. □

Theorem 9 shows that if σ,π,w are represented by softmax functions, then $\sigma=\phi(\pi,w)$ holds automatically. If in addition $\mu=\mathrm{reshape}(\tau)$, then $(\sigma,\mu)=\Phi(\pi,\tau)$ holds as long as $w(\theta)=\mathrm{norm}([\mathbf{1}^\top\tau_i])$. Consequently, viewed as functions of θ, the sections $\tilde{\tau}(\pi)$ and $\tilde{\mu}(\sigma)$ in Proposition 8 (i) are in fact the same functions, and, viewed as functions of (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 $w(\theta)=\mathrm{norm}([\mathbf{1}^\top\tau_i])$.

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

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

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

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

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

Note that $\mathrm{softmax}(\theta + k\mathbf{1}) = \mathrm{softmax}(\theta)$, so 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 $d\pi$. We therefore denote $\overline{d\theta}$ and $\widetilde{d\theta}$ as in equation (4.14), so that $\sigma^\top\overline{d\theta} = 0$ and $\pi_i^\top\widetilde{d\theta}_i = 0$. Here $\overline{d\theta}$ has already appeared in Paper 1.

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

As in Paper 1, instead of the plain $d\theta$, we actually use $\overline{d\theta}$ and $\widetilde{d\theta}$ as the differentials measuring the differentiable properties of the fixed-point bundles. We first 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) $\pi_i^\top\overline{d\theta}_i = 0$ *for every index i if and only if* $\widetilde{d\theta} = \overline{d\theta}$, *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 $\pi_i^\top\overline{d\theta}_i$, $\overline{d\theta} - \widetilde{d\theta}$, and dw, we can derive the following three equations.

$$\pi_i^\top\overline{d\theta}_i = \pi_i^\top\left(d\theta_i - (\sigma^\top d\theta)\mathbf{1}\right) = \pi_i^\top d\theta_i - \textstyle\sum_j w_j\pi_j^\top d\theta_j$$
$$\overline{d\theta}_i - \widetilde{d\theta}_i = d\theta_i - (\sigma^\top d\theta)\mathbf{1} - (d\theta_i - (\pi_i^\top d\theta_i)\mathbf{1}) = (\pi_i^\top d\theta_i)\mathbf{1} - (\textstyle\sum_j w_j\pi_j^\top d\theta_j)\mathbf{1}$$
$$dw/w = \left(I-\mathbf{1}w^\top\right)[\pi_i^\top d\theta_i] = \left[\pi_i^\top d\theta_i - \textstyle\sum_j w_j\pi_j^\top d\theta_j\right].$$

Therefore, $\pi_i^\top\overline{d\theta}_i = 0$ for every index i, $\widetilde{d\theta} = \overline{d\theta}$, 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. □

The differentiable properties of the fixed-point bundle are governed by the sections $\tilde{\mu}(\sigma)$ and $\tilde{\tau}_i(\pi)$, whose differentials appear in the singular manifold, the differential equation, the Newton equation, and the gradient of related functions. In Paper 1, $d\tilde{\mu}(\sigma)$ is measured with respect to $\overline{d\theta}$. We next measure $d\tilde{\tau}_i(\pi)$ with respect to $\widetilde{d\theta}$ and derive its 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, with $\delta_{ij} = 1$ 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*

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

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

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

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

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

In Paper 1, $d\tilde{\mu}_i(\sigma) = \sigma_i \circ \bar{J}_i(\sigma)\overline{d\theta}$, and 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 $d\tilde{\tau}_i(\pi) = \sigma_i \circ \bar{J}_i(\sigma)\overline{d\theta} = \sigma_i \circ \bar{J}_i(\sigma)\widetilde{d\theta}$.

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. Substituting $\widetilde{d\theta}$ for $d\theta$ therefore 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).

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

(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. The third line uses $\mathbf{1}^\top d\pi_i = 0$, and the last line uses $d\pi_i/\pi_i = \widetilde{d\theta}_i = (I - \mathbf{1}\pi_i^\top)\widetilde{d\theta}_i$.

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

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 $k w_i w_j \mathbf{1}\pi_j^\top$ 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.

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

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 $w_i(k\mathbf{1}\sigma^\top)_{ij} = w_i(k\mathbf{1}\sigma_j^\top) = k w_i w_j \mathbf{1}\pi_j^\top$, we obtain $w_i \bar{J}_{ij}(\sigma) = J_{ij}(\pi)$. □

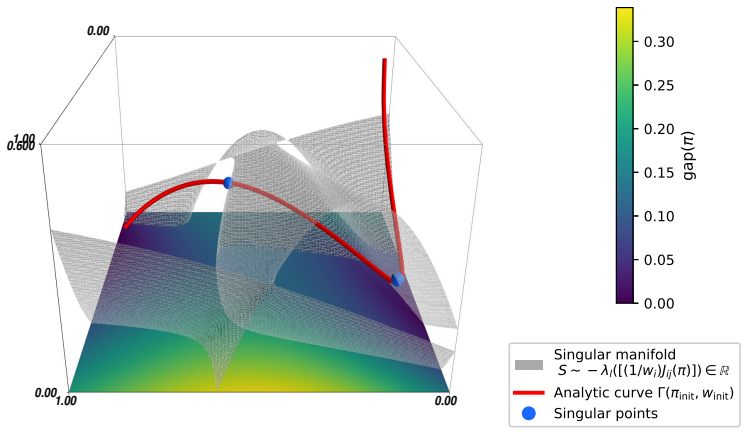

Figure 1: The fixed-point bundle of a $\mathrm{VI}(F, \Delta_n^m)$ instance with $m = n = 2$. The horizontal plane represents the product-of-simplex coordinate π, the vertical axis represents the fiber coordinate $\sum_i \mathbf{1}^\top \tau_i$, and the w coordinate is fixed at w_{init}. The heatmap shows $\mathrm{gap}(\pi)$ and the solutions are the points where $\mathrm{gap}(\pi) = 0$. The surface depicts the singular manifold, determined by the nonzero real eigenvalues of $-\mathrm{Diag}[(1/w_i)I]J(\pi)$, which form a single complicated manifold. The analytic curve $\Gamma(\pi_{\text{init}}, w_{\text{init}}) = \{(\pi, \tau) \in E \mid \tau = \gamma w_{\text{init}} \odot \pi_{\text{init}},\ \gamma \in [0, +\infty)\}$ connects a designated starting point to the solutions as $\tau \to \mathbf{0}^+$. Its intersections with the singular manifold are singular points on it that block path-following.

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. It also yields $w_i \bar{J}_{ij}(\sigma) = J_{ij}(\pi)$ 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

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

Directly verifying $w_i \bar{J}_{ij}(\sigma) = J_{ij}(\pi)$ by symbolic derivation from equation (4.16) would be very complicated, but our numerical tests show that no significant difference can be found between $\mathrm{Diag}[w_i I]\bar{J}(\sigma)$ and $J(\pi)$ with $k = 0$.

5 Path-following on the fixed-point bundles

Having established the equivalence between the two fixed-point bundle frameworks of $\mathrm{VI}(\bar{F}, \Delta_{mn}^+)$ and $\mathrm{VI}(F, \Delta_n^m)$ via a fiber bundle isomorphism, we transform the path-following algorithm in Paper 1 so that it applies to $\mathrm{VI}(F, \Delta_n^m)$. Then the property of the transformed path-following algorithm provides a transition from $\mathrm{VI}(\bar{F}, \Delta_{mn}^+)$ to an actual smooth VI on a closed simplex, so that the theoretical guarantee in Paper 1 transfers.

5.1 Transforming the predictor-corrector framework

Paper 1 adopts the predictor-corrector framework for 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, given by the differential equations of $\bar{G}(\sigma, \mu) = 0$ or $G(\pi, \tau) = 0$, while a corrector step updates onto the fixed-point bundle, given by the Newton

equations of $\bar{G}(\sigma,\mu)=0$ or $G(\pi,\tau)=0$, or by the gradients of $\mathrm{mcp}(\sigma,\mu)$ and $\mathrm{mcp}(\pi,\tau)$. The predictor and corrector for $\bar{G}(\sigma,\mu)=0$ are obtained by substituting $\bar{F}(\sigma)$ into Paper 1. In this subsection, we use the equivalence established above 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).

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

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

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

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

$$\begin{aligned}\mathrm{Diag}\left[w_iI\right]\left(\bar{J}(\sigma)+(\mathbf{1}^\top\mu)I\right)\overline{d\theta}&=\mathrm{Diag}\left[w_iI\right]\left(I-\mathbf{1}\sigma^\top\right)(d\mu/\sigma)\\ \left(J(\pi)+\mathrm{Diag}\left[(\mathbf{1}^\top\tau_i)I\right]\right)\overline{d\theta}&=\left[\left(I-\mathbf{1}\pi_i^\top\right)(d\tau_i/\pi_i)\right]\end{aligned}$$

$$\begin{aligned}\mathrm{Diag}\left[w_iI\right]\left(\bar{J}(\sigma)+(\mathbf{1}^\top\mu)I\right)\overline{d\theta}&=-\mathrm{Diag}\left[w_iI\right]\left(\bar{G}(\sigma,\mu)/\sigma\right)\\ \left(J(\pi)+\mathrm{Diag}\left[(\mathbf{1}^\top\tau_i)I\right]\right)\overline{d\theta}&=-\left[G_i(\pi,\tau)/\pi_i\right]\end{aligned}$$

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 $\pi_i^\top\overline{d\theta}_i=0$. It follows from Lemma 10 (ii) that $\overline{d\theta}=\widetilde{d\theta}$ and $dw=0$. □

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

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

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

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 $-[(1/w_i)J_{ij}(\pi)]$. 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 is consistent with the homeomorphism between (σ,μ) and (π,τ).

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

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

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 $d\overline{\mathrm{mcp}}(\sigma, \mu)$ to $d\mathrm{mcp}(\sigma, \mu)$, the elimination of $d\bar{v}_i$ remains as shown in the following equation. Thus we obtain the second equality in the equation above, which indicates that $d\mathrm{mcp}(\sigma, \mu)$ yields the same 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. The first line uses $\mathrm{mcp}(\pi, \tau) = \mathrm{mcp}(\sigma, \mu)$ from Proposition 8. The second line replaces $\overline{d\theta} = (I - \mathbf{1}\sigma^\top) d\theta$ by $(I - \mathbf{1}\sigma^\top)\widetilde{d\theta}$ and uses $\sigma^\top \widetilde{d\theta} = 0$ from 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 $d\mathrm{mcp}(\pi, \tau)$. The third line uses $\bar{r}_i = r_i / w_i$ and $\hat{\sigma}_i = w_i \hat{\pi}_i$ from Proposition 8 (iii), $w_i \bar{J}_{ij}(\sigma) = J_{ij}(\pi)$ from Theorem 11, and $\sigma^\top \bar{F}(\sigma) = 0$ and $w_i \bar{F}_i(\sigma) = F_i(\pi) - (\pi_i^\top F_i(\pi))\mathbf{1}$. The fourth line uses $r_i = F_i(\pi) + v_i \mathbf{1}$ and $\hat{\pi}_i^\top F_i(\pi) + v_i = \mathbf{1}^\top \tau_i$.

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

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 $\mathrm{mcp}(\pi, \tau)$ in equation (5.5).

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

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 make the coefficient matrix positive semidefinite and control its condition number through a regularization parameter, so that the iteration is stabilized and the update steps do not blow up near singularities. Here we derive the two corresponding regularized Newton correctors for $\mathrm{VI}(F, \Delta_n^m)$. We write $J_G := J(\pi) + \mathrm{Diag}[(\mathbf{1}^\top \tau_i) I]$ as a shorthand notation.

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

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

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 $\phi_w^{-1}(\sigma) = \mathrm{norm}([\mathbf{1}^\top \mu_i])$, this transformation can be carried further as follows.

$$\frac{\sigma}{\bar{r}} \circ \left(\bar{J}(\sigma) + (\mathbf{1}^\top \mu) I \right) \left(\overline{d\theta} + [v_{gb,i}\mathbf{1}] \right) = - (\sigma - \hat{\sigma})$$

$$\frac{w_i\pi_i}{r_i/w_i} \circ \sum_j \Big((1/w_i) J_{ij}(\pi) + \delta_{ij}(1/w_i)(\mathbf{1}^\top \tau_i) I \Big) \left(\overline{d\theta}_i + v_{gb,i}\mathbf{1} \right) = -\left(w_i\pi_i - w_i\hat{\pi}_i \right)$$

$$\frac{\pi_i}{r_i} \circ \sum_j \Big(J_{ij}(\pi) + \delta_{ij}(\mathbf{1}^\top \tau_i) I \Big) \left(\overline{d\theta}_i + v_{gb,i}\mathbf{1} \right) = -\left(\pi_i - \hat{\pi}_i \right)$$

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).

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

5.2 Transferring the convergence results

So far, we have completed the fixed-point bundle framework of $\mathrm{VI}(\bar{F}, \Delta_{mn}^+)$ and the composite fixed-point bundle framework of $\mathrm{VI}(F, \Delta_n^m)$. We can now address the final transition from $\mathrm{VI}(\bar{F}, \Delta_{mn}^+)$ to an actual smooth VI on a closed simplex, so 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 $\beta\mathrm{norm}([\mathbf{1}^\top\tau_i]) \odot \pi$ 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 $-\mathrm{Diag}[(1/w_i)I]J(\pi)$; (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 $\mathrm{VI}(\bar{F}, \Delta_{mn}^+)$ and $\mathrm{VI}(F, \Delta_n^m)$ 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* $\hat{\tau} = \tau + \beta\mathrm{norm}([\mathbf{1}^\top\tau_i]) \odot \pi$.

(iii) $\sigma_i' = \phi_{w,i}^{-1}(\sigma)\mathrm{softmax}(\ln \phi_{\pi,i}^{-1}(\sigma) + \overline{d\theta}_i)$ *and* $\pi_i' = \mathrm{softmax}(\ln \pi_i + \widetilde{d\theta}_i)$ *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* $w_{\text{init}} = \phi_w^{-1}(\sigma) = \mathrm{norm}([\mathbf{1}^\top\mu_i])$ *hold constantly during the iteration.*

Proof. The first equality $(\sigma,\mu) = \Phi(\pi,\tau)$ holds if and only if $\sigma = \phi(\pi, w)$, $\mu = \mathrm{reshape}(\tau)$, and $w = \mathrm{norm}([\mathbf{1}^\top\mu_i])$. 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 $\sigma = \phi(\pi, w)$, $\mu = \mathrm{reshape}(\tau)$, $\phi_w^{-1}(\sigma) = w_{\text{init}}$, and $\mathrm{norm}([\mathbf{1}^\top\mu_i]) = w_{\text{init}}$. Thus, the conditions $(\sigma,\mu) = \Phi(\pi,\tau)$ and $w_{\text{init}} = \phi_w^{-1}(\sigma) = \mathrm{norm}([\mathbf{1}^\top\mu_i])$ 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 $\mathrm{norm}([\mathbf{1}^\top\mu_i']) = \mathrm{norm}([\mathbf{1}^\top\mu_i])$, so $w = \mathrm{norm}([\mathbf{1}^\top\mu_i'])$ given $w = \mathrm{norm}([\mathbf{1}^\top\mu_i])$. For μ' and τ', we have $\mu' = \mathrm{reshape}(\tau')$ given $\mu = \mathrm{reshape}(\tau)$. 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 $\mathrm{norm}([\mathbf{1}^\top\hat{\mu}_i]) = \mathrm{norm}([\mathbf{1}^\top\mu_i + \beta w_{\text{init},i}]) = w_{\text{init}}$ given $\mathbf{1}^\top\sigma_i = w_i$ and $w = \mathrm{norm}([\mathbf{1}^\top\mu_i])$. For $\hat{\mu}$ and $\hat{\tau}$, we have $\sigma = \mathrm{reshape}(\mathrm{norm}([\mathbf{1}^\top\tau_i]) \odot \pi)$ given $\mu = \mathrm{reshape}(\tau)$ and $\mathrm{norm}([\mathbf{1}^\top\mu_i]) = w$, then we have $\hat{\mu} = \mathrm{reshape}(\hat{\tau})$. Thus, the two conditions are preserved.

Update (iii): The update $\sigma_i' = \phi_{w,i}^{-1}(\sigma)\mathrm{softmax}(\ln \phi_{\pi,i}^{-1}(\sigma) + \overline{d\theta}_i)$ 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 $\mu = \mathrm{reshape}(\tau)$ and $w = \mathrm{norm}([\mathbf{1}^\top \mu_i])$. Proposition 12 shows that $\overline{d\theta} = \widetilde{d\theta}$ for either the differential equations or the Newton equations, then $\mathrm{softmax}(\ln \phi_{\pi,i}^{-1}(\sigma) + \overline{d\theta}_i) = \pi_i'$ 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. □

In Theorem 13 (iii), σ is updated by $\sigma_i' = \phi_{w,i}^{-1}(\sigma)\mathrm{softmax}(\ln \phi_{\pi,i}^{-1}(\sigma) + \overline{d\theta}_i)$, instead of the Paper 1 version $\sigma' = \mathrm{softmax}(\ln \sigma + \overline{d\theta})$. In fact, the former is a second-order correction of the latter, enforcing 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 the second line uses $d\pi_i = \pi_i \circ \widetilde{d\theta}_i$ from equation (4.13) and $\overline{d\theta} = \widetilde{d\theta}$ from Proposition 12.

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

A second-order correction $o(\overline{d\theta}_i)$ at each update does not change the convergence rate of the original predictor-corrector framework [24].

Theorem 13 shows that the predictor-corrector path-following algorithms on $\bar{E} \to \Delta_{mn}^+$ and $E \to \Delta_m^\circ \times \Delta_n^m \to \Delta_n^m$ generate sequences of (σ, μ) and (π, τ) that correspond exactly via the homeomorphism Φ, and that w remains constant throughout the iteration. This can be used for the final transition of constructing a smooth $\mathrm{VI}(\tilde{F}, \Delta_{mn})$ on the closed simplex such that all results in Paper 1 apply. Specifically, we want a map $\tilde{F}$ that is real-analytic on the closed simplex Δ_{mn}, such that $\tilde{F}$ and $\bar{F}$ 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 $\mathbf{1}^\top \tilde{\pi}_i(\sigma) = 1$ follows from $\mathbf{1}^\top \sigma_i = \phi_{w,i}^{-1}(\sigma)$ and $\mathbf{1}^\top \pi_{\text{init},i} = 1$, and $\tilde{\pi}_i(\sigma) \geq 0$ follows from $\tilde{w}(\sigma) \geq \phi_w^{-1}(\sigma)$. Since, in addition, 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 $\tilde{\pi}(\sigma) = \phi_\pi^{-1}(\sigma)$, so that $\tilde{F}(\sigma) = \bar{F}(\sigma)$. Thus, $\tilde{F}$ and $\bar{F}$ have the same value on the subset $W(w_{\text{init}})$.

Besides their function values, we also need to compare the derivatives of $\tilde{F}$ and $\bar{F}$.

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

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

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

Theorem 13 shows that the update of the predictor-corrector path-following algorithm keeps σ on $W(w_{\text{init}})$ at all times. Proposition 14 shows that on $W(w_{\text{init}})$ the differential equations and Newton equations are the same for the 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, so all the results in Paper 1 apply to it. The theoretical guarantees of the path-following on the fixed-point bundle then 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

As in Paper 1, in practice we exploit the geometric freedom of the fixed-point bundle without necessarily following the path constructed in the proofs, so as to improve the robustness of the algorithm. Now that we have shown that w can remain unchanged during the iteration, we can add a mechanism to modify w arbitrarily in singularity avoidance. Specifically, we reallocate $\mathbf{1}^\top\tau_i$ to change $w=\text{norm}([\mathbf{1}^\top\tau_i])$.

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

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

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

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

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}
repeat
  repeat
    Compute $(\hat{\pi}_k,r_k,v_k)=M(\pi_k,\tau_t)$ by solving bisection problem (3.4)
    Construct the matrix $J(\pi_k)$ in equation (4.16)
    Solve regularized Newton equation (5.6) or (5.7) for $\widetilde{d\theta}_k$ with $\delta_H=\|G(\pi_k,\tau_t)\|/(mn)$
    Update $\pi_{k+1,i}=\text{softmax}(\ln\pi_{k,i}+\widetilde{d\theta}_{k,i})$
  until $\|G(\pi_k,\tau_t)\|\le\epsilon/(mn)$
  Set $\hat{\tau}_{t,i}=\tilde{\tau}_i(\pi_k)+(\mathbf{1}^\top\hat{\tau}_{t,i})\pi_k$ with $\mathbf{1}^\top\hat{\tau}_{t,i}$ given by equation (5.9), such that $\sum_i\mathbf{1}^\top\tau_{t,i}+\beta_t$ avoids the nonzero eigenvalues of $-\text{Diag}[(1/w_i)I]J(\pi_k)$
  Update $\tau_{t+1}=(\hat{\tau}_t-\eta_t u_t).\text{clip}(\min=\epsilon/(mn))$ with a small η_t and a direction $u_t=\hat{\tau}_t$
  Solve differential equation (5.3) for $\widetilde{d\theta}_t$ with $d\tau_t=\tau_{t+1}-\hat{\tau}_t$
  Set $\pi_{0,i}=\text{softmax}(\ln\pi_{k,i}+\widetilde{d\theta}_{t,i})$
until $\sum_i\mathbf{1}^\top\check{\tau}_i(\pi_k)\le\epsilon$
return π_k as an approximate solution of $\text{VI}(F,\Delta_n^m)$

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 $-\text{Diag}[(1/w_i)I]J(\pi)$, where $w_i=\mathbf{1}^\top\hat{\tau}_i/(\sum_j\mathbf{1}^\top\hat{\tau}_j)$ is given by equation (5.9). The value of $\hat{\tau}$ after

singularity avoidance is then $\hat{\tau} = \tilde{\tau}(\pi) + (\sum_i \mathbf{1}^\top \tau_i + \beta) w \odot \pi$. Our experimental results are based on this singularity avoidance mechanism with $\tilde{w} \in \Delta_m^\circ$ chosen randomly, which performs better than keeping w fixed.

The path-following algorithm for $\mathrm{VI}(F, \Delta_n^m)$ 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.

The corrector can be either regularized Newton equation (5.6) or regularized Newton equation (5.7). As in Paper 1, the former mixes the standard Newton equation (5.4) with 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) with the inexact gradient $J_G^\top[\pi_i - \hat{\pi}_i]$ of $\mathrm{mcp}(\pi,\tau)$.

6 Numerical experiments

The problem $\mathrm{VI}(F, \Delta_n^m)$ contains $\mathrm{VI}(F, \Delta_n)$ of Paper 1 as the special case $m = 1$. Likewise, Algorithm 1 reduces to Algorithm 1 of Paper 1 when $m = 1$, because $\tilde{w}$ can only take value 1, and the singularity avoidance in equation (5.9) degenerates into $\mathbf{1}^\top \hat{\tau} = \mathbf{1}^\top \tau + \beta$ used in Paper 1. The implementation below likewise contains that of Paper 1 as the case $m = 1$.

As in Paper 1, we use a single-level derivative-free implementation for numerical efficiency, where both features are standard in path-following. Experiments with the single-level derivative-based and the two-level derivative-based implementations were also performed, with results stored in the GitHub repository accompanying this paper.

We test the algorithm on randomly generated instances of $\mathrm{VI}(F, \Delta_n^m)$, with F represented by two different models: a neural network model and a payoff tensor model. The neural network model represents VI problems with a general real-analytic operator. We use tanh-activated neural networks of architecture $[mn, mn, mn]$, with parameters generated by Xavier normal initialization, which prevents the tanh activations from saturating and hence keeps the complexity of F from collapsing.

The payoff tensor model represents finite normal-form games. For an m-player n-action finite normal-form game, the payoff function is multilinear and is given by an m-th order payoff tensor of shape $(n, \ldots, n)$. The tensor has n^m entries, which grows exponentially in m. As a practical alternative to storing the full tensor, we use the classical low-rank CP (canonical polyadic) representation [25]. Specifically, F is given by $F_i(\pi) = \nabla_{\pi_i} f_i(\pi)$, where

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

is the low-rank CP payoff function. Each U_i is a third-order tensor of shape (R, m, n), which reduces the storage requirement from n^m to Rmn. We use $R = mn$ in our experiments. The entries of U are drawn from a normal distribution with mean 1 and variance $1/m$, so that $\prod_{j=1}^m u_j$ has mean 1 and variance close to $e - 1$, which prevents overflow or underflow for large m.

The starting point π_{init} is also randomly generated, and the required precision is $\epsilon = 10^{-5}$. We solve the problems until $\mathrm{gap}_i(\pi) < \epsilon$ rather than $\mathrm{gap}(\pi) = \sum_i \mathrm{gap}_i(\pi) < \epsilon$. 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), which is faster in large dimensions as reported in Paper 1. Table 1 reports the results.

Table 1: Iteration counts (mean/median) for solving $\mathrm{VI}(F, \Delta_n^m)$.

(a) $m = 8$ varying n; $n = 8$ varying m

(m, n)	Neural network F	Payoff tensor F
$(8, 2)$	175 / 173	173 / 173
$(8, 4)$	181 / 176	179 / 174
$(8, 8)$	198 / 188	187 / 178
$(8, 16)$	251 / 228	210 / 201
$(8, 32)$	409 / 342	310 / 289
$(2, 8)$	176 / 174	174 / 174
$(4, 8)$	181 / 175	178 / 174
$(8, 8)$	198 / 188	187 / 178
$(16, 8)$	264 / 235	218 / 203
$(32, 8)$	494 / 427	374 / 333

(b) $mn = 256$

(m, n)	Neural network F	Payoff tensor F
$(2, 128)$	320 / 280	271 / 266
$(4, 64)$	371 / 313	283 / 270
$(8, 32)$	409 / 342	310 / 289
$(16, 16)$	469 / 386	344 / 314
$(32, 8)$	494 / 427	374 / 333
$(64, 4)$	438 / 393	369 / 339
$(128, 2)$	297 / 248	251 / 225

The algorithm succeeds on all 5600 instances. The iteration count increases faster when m increases with n fixed than when n increases with m fixed. When mn is fixed, the peak iteration count occurs in the middle rather than at the endpoints $(2, 128)$ or $(128, 2)$, indicating that both m and n affect the problem difficulty for a fixed total dimension mn. In addition, solving $\mathrm{VI}(F, \Delta_n^m)$ with the payoff F is easier than with the neural network F, because neural networks produce greater nonlinearity and more twisted path and singular manifold.

Table 2: Statistics (mean/median) for solving $\mathrm{VI}(F, \Delta_n^m)$ under different implementations, with $(m, n) = (16, 16)$.

Implementation	Neural network F		Payoff tensor F	
	Iteration count	Singularity avoidance count	Iteration count	Singularity avoidance count
single-level derivative-free	469 / 386	1.2 / 1.0	344 / 314	0.7 / 1.0
single-level derivative-based	439 / 372	1.2 / 1.0	327 / 306	0.6 / 1.0
two-level derivative-based	2128 / 1518	70.8 / 44.5	1054 / 929	23.3 / 19.0

Table 2 reports the iteration counts and singularity avoidance counts of the different implementations with $(m, n) = (16, 16)$. The pattern found in Paper 1 reappears here. The single-level iteration needs several times fewer iterations than the two-level iteration, and the derivative-free mode needs nearly the same number of iterations as the derivative-based mode. The single-level derivative-free implementation is therefore the most practical choice.

7 Conclusion

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

We construct $\mathrm{VI}(\bar{F}, \Delta^+_{mn})$ with simplex domain from $\mathrm{VI}(F, \Delta^m_n)$ that preserves the solutions. The central innovation of this paper is to establish the equivalence between their fixed-point bundle frameworks, based on $\bar{E} \to \Delta^+_{mn}$ and $E \to \Delta^m_n \times \Delta^\circ_m$, via the bundle isomorphism $(\Phi : E \to \bar{E}, \phi : \Delta^m_n \times \Delta^\circ_m \to \Delta^+_{mn})$.

The path-following algorithms on the fixed-point bundles keep w constant. The restriction of the map $\bar{F}$ to the slice $w = w_{\text{init}}$ can be extended to a real-analytic map $\tilde{F}$ on the closed simplex Δ_{mn}, and then all the results in Paper 1 apply to $\mathrm{VI}(\tilde{F}, \Delta_{mn})$. Thus, the path-following algorithm for $\mathrm{VI}(F, \Delta^m_n)$ 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 from 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] George W Brown. Iterative solution of games by fictitious play. *Activity Analysis of Production and Allocation*, 13(1):374, 1951.

[16] Johannes Heinrich, Marc Lanctot, and David Silver. Fictitious self-play in extensive-form games. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, pages 805–813, Lille, France, 2015. PMLR.

[17] Sergiu Hart and Andreu Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127–1150, 2000.

[18] Martin Zinkevich, Michael Johanson, Michael Bowling, and Carmelo Piccione. Regret minimization in games with incomplete information. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems*, volume 20. Curran Associates, Inc., 2007.

[19] Christos H Papadimitriou. On the complexity of the parity argument and other inefficient proofs of existence. *Journal of Computer and system Sciences*, 48(3):498–532, 1994.

[20] Hongbo Sun. A path-following framework on fiber bundle for variational inequalities, 2026.

[21] Steven George Krantz. *Function theory of several complex variables*, volume 340. American Mathematical Soc., 2001.

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

[23] R Tyrrell Rockafellar and Roger JB Wets. *Variational analysis*. Springer, 1998.

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

[25] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.