Nonlinear Fixed Points and Stationarity: Economic Applications*

Simone Cerreia-Vioglio[a], Roberto Corrao[b], Giacomo Lanzani[b]

[a]Università Bocconi and Igier, [b]MIT

May 2023

Abstract

We consider the fixed points of nonlinear operators that naturally arise in games and general equilibrium models with endogenous networks, dynamic stochastic games, and in models of opinion dynamics with stubborn agents. We study limit cases that correspond to high coordination motives, infinite patience, and vanishing stubbornness in the applications above. Under monotonicity and continuity assumptions, we provide explicit expressions for the limit fixed points. We show that, under differentiability, the limit fixed point is linear in the initial conditions and characterized by the Jacobian of the operator at any constant vector with an explicit and linear rate of convergence. Without differentiability, but under additional concavity properties, the multiplicity of Jacobians is resolved by a representation of the limit fixed point as a maxmin functional evaluated at the initial conditions. In our applications, we use these results to characterize the limit equilibrium actions, prices, and endogenous networks, show the existence of the asymptotic value in a class of zero-sum stochastic games with a continuum of actions, and compute a nonlinear version of the eigenvector centrality of agents in networks.

*We wish to thank Oguzhan Celebi, Joel Flynn, Drew Fudenberg, Ben Golub, Matt Jackson, Peter Klibanoff, Nicolas Lambert, Stephen Morris, Simeon Reich, Karthik Sastry, Lorenzo Stanca, Omer Tamuz, Nicolas Vieille, and Alex Wolitzky for useful comments. Roberto Corrao gratefully acknowledges the financial support of the Gordon Pye fellowship.

1 Introduction

Nonlinear fixed-point equations are ubiquitous in economic models including the ones that characterize general equilibrium prices, Nash equilibria, continuation values in dynamic games (Shapley equation), steady stated under social learning, and recursive preferences. Often, these fixed points are indexed by a key economic parameter $\beta \in (0,1)$ capturing, for example, strength of coordination motives, patience, and stubbornness, with the comparative statics for β close to 1 playing a prominent role. The problem of solving for these nonlinear fixed points has been tackled with different tools across these applications without a unifying approach.

In this paper, we first highlight a few key mathematical properties shared by all these classes of nonlinear fixed-point equations: monotonicity, translation invariance, and normalization. These properties generalize the ones of linear averaging operators for which the structure of corresponding fixed-point equations is well known. In fact, for the linear case it is in general possible to derive a closed-form expression for the fixed point at each β and in particular for the limit as β goes to 1, yielding a rate of convergence as well. These expressions are often interpreted as (Bonacich or eigenvector) centrality measures of agents within the context of, for example, models of production networks (e.g., Acemoglu et al. [2]) or coordination games (e.g., Ballester et al. [6]). However, in all the aforementioned applications nonlinearities naturally arise due to economic forces. For example, in production network models both relaxing the assumption of Cobb-Douglas production functions (e.g., Baqaee and Farhi [8]) and/or allowing for endogenous networks (e.g., Acemoglu and Azar [1] and Kopytov et al. [28]) generate nonlinearities in the equation describing equilibrium prices. Similarly, in coordination games on networks when we relax the assumption of quadratic payoffs and/or allow for endogenous link formation (e.g., Sadler and Golub [32]) the resulting Nash equilibria are characterized by nonlinear fixed points. Moreover, in some other applications such as stochastic games and recursive preferences the maximization defining the value functions already induces nonlinearities (e.g., Sorin [35]). Yet, in all these cases our three properties are still satisfied. Thus, we exploit this common structure to derive properties of the nonlinear fixed point the most important of which is a closed-form expression for the limit as β approaches 1. These expressions admit a natural interpretation as nonlinear versions of the (linear) centrality measures above. Along the way, we derive additional results extending the results obtained for the linear case.

Formally, in this paper we consider an operator $T : \mathbb{R}^k \to \mathbb{R}^k$ where $\mathbb{R}^k$ is endowed with the coordinatewise order $\geq$ and the supnorm $\| \|_\infty$. Let e be the vector whose components are all 1. We assume that:

1. T is *normalized*, that is, $T(he) = he$ for all $h \in \mathbb{R}$;
2. T is *monotone*, that is, $x \geq y$ implies $T(x) \geq T(y)$ for all $x, y \in \mathbb{R}^k$;
3. T is *translation invariant*, that is, $T(x + he) = T(x) + he$ for all $x \in \mathbb{R}^k$ and for all $h \in \mathbb{R}$.

As we already pointed, these three properties are often satisfied in applications in Economics and Computer Science where T is seen as either a best-response map, or a value function, or an opinion aggregator, or a collection of decision criteria. Clearly, for these maps the set of fixed points/equilibria of T, denoted by $E(T)$, contains all the constant vectors, denoted by D, that is, $D \subseteq E(T)$.

Given $x \in \mathbb{R}^k$ and $\beta \in (0,1)$, it is routine to show that the two equations

$$T((1-\beta)x + \beta y) = y \tag{1}$$

and

$$(1-\beta)x + \beta T(z) = z \tag{2}$$

have each a unique solution (cf. Lemma 1). We denote such solutions by x_β and $\tilde{x}_\beta$, respectively, to highlight their dependence on x and β.

The goal of this paper is to provide conditions that guarantee that $\lim_{\beta \to 1} x_\beta$ and $\lim_{\beta \to 1} \tilde{x}_\beta$ exist, characterize their value, and also comment on the rate of convergence. We next introduce the linear case which is well known. This will provide a useful benchmark to which we can compare one of our main results and, more in general, our contributions.

Example 1 We begin by observing that further assuming T linear is equivalent to impose that $T(x) = Wx$ for all $x \in \mathbb{R}^k$ where W is a $k \times k$ stochastic matrix (i.e. the entries of each row are nonnegative and they sum up to 1). Given $x \in \mathbb{R}^k$ and $\beta \in (0,1)$, let $x_{\beta,W}$ and $\tilde{x}_{\beta,W}$ be the (unique) points satisfying:

$$T((1-\beta)x + \beta x_{\beta,W}) = x_{\beta,W} \text{ and } (1-\beta)x + \beta T(\tilde{x}_{\beta,W}) = \tilde{x}_{\beta,W}. \tag{3}$$

By induction and passing to the limit, it is routine to show that

$$x_{\beta,W} = (1-\beta)\sum_{t=0}^{\infty} \beta^t W^{t+1} x \qquad \forall x \in \mathbb{R}^k, \forall \beta \in (0,1). \tag{4}$$

Similarly, one obtains that

$$\tilde{x}_{\beta,W} = (1-\beta)\sum_{t=0}^{\infty} \beta^t W^{t} x \qquad \forall x \in \mathbb{R}^k, \forall \beta \in (0,1).$$

This suggests a simple relation between $x_{\beta,W}$ and $\tilde{x}_{\beta,W}$, that is, $\tilde{x}_{\beta,W} = (1-\beta)x + \beta x_{\beta,W}$. Moreover, by the celebrated Hardy-Littlewood Theorem paired with the Mean Ergodic Theorem,[1] this implies that $\lim_{\beta \to 1} \tilde{x}_{\beta,W}$ and $\lim_{\beta \to 1} x_{\beta,W}$ both exist, belong

[1] See Online Appendix F.1 for a formal proof.

to $E(T)$, and coincide.

Assume now that the unique fixed points of T are the constant vectors, that is, $D = E(T)$. This is equivalent to assume that the matrix W has a unique left Perron-Frobenius eigenvector γ_W, that is, $\gamma_W^{\mathrm{T}} W = \gamma_W^{\mathrm{T}}$ and γ_W is a probability vector. In this case, we can conclude that $\lim_{\beta \to 1} \tilde{x}_{\beta,W} = \lim_{\beta \to 1} x_{\beta,W}$ is a constant vector whose value can be computed by observing that

$$\langle \gamma_W, x_{\beta,W} \rangle = (1-\beta) \sum_{t=0}^{\infty} \beta^t \langle \gamma_W, W^{t+1} x \rangle = \langle \gamma_W, x \rangle \qquad \forall x \in \mathbb{R}^k, \forall \beta \in (0,1), \quad (5)$$

which immediately yields that $\lim_{\beta \to 1} \tilde{x}_{\beta,W} = \lim_{\beta \to 1} x_{\beta,W} = \langle \gamma_W, x \rangle e$.[2]

We conclude by commenting on the rate of convergence of $\{x_{\beta,W}\}_{\beta \in (0,1)}$. Similar considerations hold for $\{\tilde{x}_{\beta,W}\}_{\beta \in (0,1)}$. Fix again $x \in \mathbb{R}^k$ and $\beta \in (0,1)$. Since γ_W is a probability vector, we have that $\min x_{\beta,W} \leq \langle \gamma_W, x_{\beta,W} \rangle \leq \max x_{\beta,W}$. Since $\langle \gamma_W, x_{\beta,W} \rangle = \langle \gamma_W, x \rangle$, this implies that $\min x_{\beta,W} \leq \langle \gamma_W, x \rangle \leq \max x_{\beta,W}$ and, in particular, $\| x_{\beta,W} - \langle \gamma_W, x \rangle e \|_\infty = \max \{ \max x_{\beta,W} - \langle \gamma_W, x \rangle, \langle \gamma_W, x \rangle - \min x_{\beta,W} \} \leq \max x_{\beta,W} - \min x_{\beta,W}$. In other words, bounding the rate of convergence of $x_{\beta,W}$ can be achieved by bounding the range of $x_{\beta,W}$. ▲

Thus, the main takeaways of the linear case are four:

1. $\tilde{x}_\beta = (1-\beta) x + \beta x_\beta$ for all $x \in \mathbb{R}^k$ and for all $\beta \in (0,1)$;

2. The Mean Ergodic and Hardy-Littlewood Theorems imply that $\lim_{\beta \to 1} x_\beta = \lim_{\beta \to 1} \tilde{x}_\beta$ exist;

3. If $E(T) = D$, then we have that

$$\lim_{\beta \to 1} x_\beta = \lim_{\beta \to 1} \tilde{x}_\beta = \langle \gamma_W, x \rangle e \qquad \forall x \in \mathbb{R}^k$$

   where γ_W is the unique left Perron-Frobenius eigenvector of the representing matrix W;

4. In this case, the rate of convergence of x_β is controlled by the rate to which the range of x_β, $\mathrm{Rg}(x_\beta)$, goes to 0.

Our contributions are to generalize these findings well beyond the linear case. We here discuss an important example. To fix ideas, assume that T is *concave*, rather than linear. In this case, the relation in point 1 extends easily (cf. Lemma 1). If $E(T) = D$, we again have that $\lim_{\beta \to 1} x_\beta$ and $\lim_{\beta \to 1} \tilde{x}_\beta$ exist and coincide (cf. Corollary 2). If

[2] As usual, $\langle x, y \rangle = \sum_{i=1}^{k} x_i y_i$ for all $x, y \in \mathbb{R}^k$.

$E(T) = D$ and T is also differentiable around 0 with partial derivatives that are "nicely" bounded away from 0 when nonnull, then

$$\lim_{\beta \to 1} x_\beta = \lim_{\beta \to 1} \tilde{x}_\beta = \langle \gamma, x \rangle e \qquad \forall x \in \mathbb{R}^k$$

where γ is the unique left Perron-Frobenius eigenvector of the Jacobian of T at 0 (cf. Corollary 3). Finally, if T has a Jacobian which is Lipschitz continuous, then the rate of convergence of x_β is controlled by the rate to which $\mathrm{Rg}(x_\beta)$ goes to 0 and $\mathrm{Rg}(x_\beta)$ goes to 0 at least linearly fast (cf. Theorems 2 and 3). In the paper, we go well beyond the concave case, which we actually use to study more general functionals (cf. Theorem 1 as well as Corollary 1).

In the second part of the paper we provide economic applications for these results. First, we consider two models of endogenous network formation applied to general equilibrium in a production economy and a coordination game. In both cases, the parameter β captures the intrinsic coordination motives of the agents and, under standard assumptions such as Cobb-Douglas production functions and quadratic costs of effort, the fixed-point equations characterizing the equilibria would be linear. However, when the agents are allowed to choose their neighbors structure, either in a costly or constrained way, the equilibria fixed-point equations become nonlinear (and in general nondifferentiable) yet still satisfying all of our assumptions. With this, we completely characterize the limit equilibrium as $\beta \to 1$ with respect to (generally nonlinear) measures of centrality of the agents. This allows us to extend some of the comparative statics on the equilibrium from the linear case to the differentiable case and to obtain new ones for the nondifferentiable case. Second, we study the classic issue of existence and characterization of the asymptotic value for zero-sum stochastic games (cf. Sorin [35]). We observe that the Shapley equation characterizing the value of the game for every level of the discount factor is a particular case of our fixed point condition, thereby enabling us to use our abstract results to provide a novel characterization of the asymptotic value in terms of the value of a static zero-sum game. Finally, we apply our results to an extension of the dynamic opinion aggregation model in networks of Cerreia-Vioglio et al. [12] that allow for vanishing stubbornness. In this case, we also consider a sequence of weights $\{\beta_t\}_{t \in \mathbb{N}} \subseteq (0,1)$ such that $\beta_t \to 1$ that represents the time-varying and vanishing stubbornness weight that agents assign to their initial opinions.[3]

[3] In Appendix E, we consider an equilibrium model of interconnected financial institutions that evaluate their losses with respect to coherent risk measures. In this case, the limit $\beta \to 1$ captures the idea of increasing financial interconnectedness and our results imply that the robustness concerns of the banks vanish in this limit, exposing all of them to possible model misspecification and large unforeseen losses.

2 Preliminaries

We start with some additional functional properties. Given T normalized, monotone, and translation invariant, we say that:

1. T is concave if and only if $T(\lambda x + (1-\lambda) y) \geq \lambda T(x) + (1-\lambda) T(y)$ for all $\lambda \in (0,1)$ and for all $x, y \in \mathbb{R}^k$;

2. T is star-shaped if and only if $T(\lambda x) \geq \lambda T(x)$ for all $\lambda \in (0,1)$ and for all $x \in \mathbb{R}^k$;

3. T is positively homogenous if and only if $T(\lambda x) = \lambda T(x)$ for all $\lambda \geq 0$ and for all $x \in \mathbb{R}^k$.

Clearly, if T is either concave or positively homogeneous, then it is star-shaped.

Next, consider a general normalized, monotone, and translation invariant operator T. It is immediate to see that it is Lipschitz continuous of order 1. By Rademacher's Theorem, T is (Frechet) differentiable on a subset $\mathcal{D}$ of $\mathbb{R}^k$ whose complement has (Lebesgue) measure 0. Denote by $T_i : \mathbb{R}^k \to \mathbb{R}$ the i-th component of T. Since T is monotone, we have that $\frac{\partial T_i}{\partial x_j}(x) \geq 0$ for all $i, j \in \{1, ..., k\}$ and for all $x \in \mathcal{D}$. Since T is also translation invariant, we have that $\sum_{j=1}^{k} \frac{\partial T_i}{\partial x_j}(x) = 1$ as well as $\frac{\partial T_i}{\partial x_j}(x) \leq 1$ for all $i, j \in \{1, ..., k\}$ and for all $x \in \mathcal{D}$. With T, we define an adjacency matrix $\underline{A}(T)$ for the operator T, that is, $\underline{a}_{ij} \in \{0,1\}$. Given $i, j \in \{1, ..., k\}$, we set

$$\underline{a}_{ij} = 1 \iff \exists \varepsilon_{ij} \in (0,1) \text{ s.t. } \frac{\partial T_i}{\partial x_j}(x) \geq \varepsilon_{ij} \quad \forall x \in \mathcal{D}. \tag{6}$$

In words, $\underline{a}_{ij}$ is defined to be 1 if and only if the partial derivative $\frac{\partial T_i}{\partial x_j}$ is bounded away from 0, whenever it exists. We say that $\underline{A}(T)$ is regular if and only if it is nontrivial and its essential indices form a single essential class.[4] If we think of $\underline{A}(T)$ as representing a directed graph over k nodes, $\underline{A}(T)$ is regular whenever the graph is strongly connected.

Given $z \in \mathbb{R}^k$, we denote by $\partial_C T_i(z)$ the Clarke differential of the i-th component of T at z. In particular, recall that (see, e.g., [16, Theorem 2.5.1])

$$\partial_C T_i(z) = \text{co}\left\{\gamma \in \mathbb{R}^k : \gamma = \lim_k \nabla T_i\left(z^k\right) \text{ s.t. } z^k \to z \text{ and } z^k \in \mathcal{D}\right\}. \tag{7}$$

By [16, Propositions 2.1.2 and 2.1.5], the correspondence $\partial_C T_i : \mathbb{R}^k \rightrightarrows \mathbb{R}^k$ is nonempty-, convex-, compact-valued, and upper hemicontinuous. If T_i is concave, it is well known that $\partial_C T_i(z)$ coincides with $\partial T_i(z)$ where the latter is the usual superdifferential of

[4] Nontriviality amounts to assume that the matrix $\underline{A}(T)$ does not have a zero row.

convex analysis (see, e.g., [16, Proposition 2.2.7]).[5] Given the above discussion, $\partial_C T_i(z)$ is a collection of probability vectors. We denote by $\partial_C T(z)$ the collection of all $k \times k$ (stochastic) matrices W whose i-th row belongs to $\partial_C T_i(z)$.[6] If T is concave, note that

$$W \in \partial_C T(z) \implies W(y-z) \geq T(y) - T(z) \qquad \forall y \in \mathbb{R}^k. \tag{8}$$

Since T is normalized, monotone, and translation invariant, we have that $\partial_C T_i(z+he) = \partial_C T_i(z)$ for all $i \in \{1, ..., k\}$, for all $z \in \mathbb{R}^k$, and for all $h \in \mathbb{R}$. In particular, we also have that $\partial_C T(he) = \partial_C T(0)$ for all $h \in \mathbb{R}$.

3 Convergence and limit characterization

In this section, we provide our two main results on the existence and characterization of the limit fixed point $\lim_{\beta \to 1} x_\beta$ in terms of the object introduced before. In particular, we will consider two cases: (i) T star-shaped (Theorem 1); (ii) T is continuously differentiable at 0 (Corollary 3).

Theorem 1 *Let T be normalized, monotone, and translation invariant. If T is star-shaped and $\underline{A}(T)$ is regular, then $\lim_{\beta \to 1} \tilde{x}_\beta$ and $\lim_{\beta \to 1} x_\beta$ exist and coincide for all $x \in \mathbb{R}^k$.*

The proof of this result consists of three major steps. In Section 3.1, we first prove Theorem 1 under the assumption that T is concave (a stronger assumption compared to star-shapedness) and assuming $E(T) = D$ (a weaker assumption compared to $\underline{A}(T)$ being regular). This alternative setting allows us to also characterize the limit and show its easy computability. In Section 3.2, we show that convergence of x_β holds also for operators which can be rewritten as the max of a family of normalized, monotone, translation invariant, and concave operators. Finally, we prove Theorem 1 by showing that star-shaped operators can be rewritten as the max of a family of concave operators (see Appendix A). This latter representation result is a refinement of Castagnoli, Cattelan, Maccheroni, Tebaldi, and Wang [11, Theorem 2].

[5] Given a function $\varphi : \mathbb{R}^k \to \mathbb{R}$ and $z \in \mathbb{R}^k$, recall that a vector γ is an element of $\partial\varphi(z)$, that is, a superdifferential of φ at z if and only if

$$\langle \gamma, y - z \rangle \geq \varphi(y) - \varphi(z) \qquad \forall y \in \mathbb{R}^k.$$

We denote by $\partial\varphi : \mathbb{R}^k \rightrightarrows \mathbb{R}^k$ the superdifferential correspondence. In this case, a concave function is Gateaux differentiable if and only if it is Frechet differentiable. Moreover, if φ is differentiable, then $\partial\varphi$ is single-valued and continuous (see, e.g., [9, Theorems 2.2.1 and 2.2.2]).

[6] The notion of generalized gradient we use for real-valued functions coincides with the one of Clarke, but our derived notion of generalized Jacobian for operators is larger than the one of Clarke (see, e.g., [16, Proposition 2.6.2]). With an abuse of notation, we still denote it by $\partial_C T(z)$.

3.1 Concavity and differentiability

Consider $x \in \mathbb{R}^k$ and $\beta \in (0,1)$, the next lemma–a routine application of the Banach contraction principle–shows that both (1) and (2) admit a unique solution, denoted by x_β and $\tilde{x}_\beta$ respectively. Moreover, it proves there is a simple relation between these two points. To this extent, given $T : \mathbb{R}^k \to \mathbb{R}^k$, define $T_{\beta,x}, \tilde{T}_{\beta,x} : \mathbb{R}^k \to \mathbb{R}^k$ by $T_{\beta,x}(y) = T((1-\beta)x + \beta y)$ and $\tilde{T}_{\beta,x}(y) = (1-\beta)x + \beta T(y)$ for all $y \in \mathbb{R}^k$. Clearly, the fixed points of $T_{\beta,x}$ and $\tilde{T}_{\beta,x}$ are the solutions of (1) and (2), respectively.

Lemma 1 *Let T be normalized, monotone, and translation invariant. If $\beta \in (0,1)$ and $x \in \mathbb{R}^k$, then $T_{\beta,x}$ and $\tilde{T}_{\beta,x}$ are β-contractions. In particular, for each $\beta \in (0,1)$ and for each $x \in \mathbb{R}^k$, there exist unique $x_\beta, \tilde{x}_\beta \in \mathbb{R}^k$ such that*

$$T^t_{\beta,x}(y) \to x_\beta \quad \forall y \in \mathbb{R}^k,\ T_{\beta,x}(x_\beta) = x_\beta,\ \text{and}\ \|x_\beta\|_\infty \le \|x\|_\infty \tag{9}$$

and

$$\tilde{T}^t_{\beta,x}(y) \to \tilde{x}_\beta \quad \forall y \in \mathbb{R}^k,\ \tilde{T}_{\beta,x}(\tilde{x}_\beta) = \tilde{x}_\beta,\ \text{and}\ \|\tilde{x}_\beta\|_\infty \le \|x\|_\infty .$$

Moreover, $\tilde{x}_\beta = (1-\beta)x + \beta x_\beta$ for all $\beta \in (0,1)$ and for all $x \in \mathbb{R}^k$.

Consider the set L of limit points of $\{x_\beta\}_{\beta\in(0,1)}$.[7] By construction and since $\{x_\beta\}_{\beta\in(0,1)}$ is bounded, the set L is closed and bounded. We define $\liminf_{\beta\to1} x_\beta = \inf L$ and $\limsup_{\beta\to1} x_\beta = \sup L$ where inf and sup are computed coordinatewise. The next simple lemma yields that the limit points of $\{x_\beta\}_{\beta\in(0,1)}$ are fixed points of T and so are $\liminf_{\beta\to1} x_\beta$ and $\limsup_{\beta\to1} x_\beta$, provided $E(T) = D$.

Lemma 2 *If T is normalized, monotone, and translation invariant, then $L \subseteq E(T)$. Moreover, if $E(T) = D$, then $\liminf_{\beta\to1} x_\beta, \limsup_{\beta\to1} x_\beta \in E(T)$.*

We can now provide a lower bound and an upper bound for $\liminf_{\beta\to1} x_\beta$ and $\limsup_{\beta\to1} x_\beta$. These bounds will be in terms of the left invariant probability vectors of the generalized Jacobian of T at 0. To this extent, we introduce some notation and terminology. Given a stochastic $k \times k$ matrix W, we denote by

$$\Gamma(W) = \left\{\gamma \in \Delta : \gamma^{\mathrm{T}} W = \gamma^{\mathrm{T}}\right\}$$

the collection of all left W-invariant probability vectors where Δ is the set of all probability vectors of $\mathbb{R}^k$. It is routine to show that $\Gamma(W)$ is nonempty, convex, and compact. If W has a unique left Perron-Frobenius eigenvector γ_W, then $\Gamma(W) = \{\gamma_W\}$. Given a subset $\mathcal{M}$ of stochastic matrices, we denote by $\Gamma(\mathcal{M})$ the set $\cup_{W\in\mathcal{M}}\Gamma(W)$. In particular, if $\mathcal{M}$ is closed, then $\Gamma(\mathcal{M})$ is compact.

[7] That is, $\bar{x} \in L$ if and only if there exists $\{x_{\beta_n}\}_{n\in\mathbb{N}} \subseteq \{x_\beta\}_{\beta\in(0,1)}$ such that $\beta_n \to 1$ and $x_{\beta_n} \to \bar{x}$.

Proposition 1 *Let T be normalized, monotone, and translation invariant. If $E(T) = D$, then*

$$\max_{\gamma\in\Gamma(\partial_C T(0))} \langle \gamma, x\rangle e \geq \limsup_{\beta\to 1} x_\beta \geq \liminf_{\beta\to 1} x_\beta \geq \min_{\gamma\in\Gamma(\partial_C T(0))} \langle \gamma, x\rangle e \qquad \forall x\in\mathbb{R}^k.$$

The next result shows that these estimates are enough to establish the existence of $\lim_{\beta\to 1} x_\beta$ and $\lim_{\beta\to 1}\tilde{x}_\beta$ and characterize it, when the map T is continuously differentiable in a neighborhood of the origin and provided $\underline{A}(T)$ is regular. This is intuitive since $\partial_C T(0)$ (given differentiability) and $\Gamma(\partial_C T(0))$ (given regularity) both become singletons.

Corollary 1 *Let T be normalized, monotone, and translation invariant. If T is continuously differentiable in a neighborhood of 0 and $\underline{A}(T)$ is regular, then*

$$\lim_{\beta\to 1}\tilde{x}_\beta = \lim_{\beta\to 1} x_\beta = \langle \gamma, x\rangle e \qquad \forall x\in\mathbb{R}^k$$

where γ is the unique left Perron-Frobenius eigenvector of the Jacobian of T at 0.

In order to dispense with the assumption of differentiability, we provide a sharper upper bound for $\limsup_{\beta\to 1} x_\beta$, under the extra assumption of concavity.

Proposition 2 *Let T be normalized, monotone, and translation invariant. If T is concave and $E(T) = D$, then*

$$\min_{\gamma\in\Gamma(\partial_C T(0))} \langle \gamma, x\rangle \geq \min_{\gamma\in\Gamma(\partial_C T(0))} \langle \gamma, x_\beta\rangle \qquad \forall x\in\mathbb{R}^k, \forall\beta\in(0,1) \tag{10}$$

and

$$\min_{\gamma\in\Gamma(\partial_C T(0))} \langle \gamma, x\rangle e \geq \limsup_{\beta\to 1} x_\beta \qquad \forall x\in\mathbb{R}^k.$$

By combining the bounds of Propositions 1 and 2, we are able to prove convergence in the concave case.

Corollary 2 *Let T be normalized, monotone, and translation invariant. If T is concave and $E(T) = D$, then*

$$\lim_{\beta\to 1}\tilde{x}_\beta = \lim_{\beta\to 1} x_\beta = \min_{\gamma\in\Gamma(\partial_C T(0))} \langle \gamma, x\rangle e \qquad \forall x\in\mathbb{R}^k.$$

Moreover, if $\underline{A}(T)$ is regular, then $\Gamma(\partial_C T(0))$ is the collection of left Perron-Frobenius eigenvectors of the superdifferentials of T at 0.

Concavity allows also to improve Corollary 1. In fact, we can only require differentiability at 0 without explicitly asking the derivative to be continuous near 0.

Corollary 3 *Let T be normalized, monotone, and translation invariant. If T is concave, differentiable at 0, and $\underline{A}(T)$ is regular, then*

$$\lim_{\beta\to 1}\tilde{x}_\beta = \lim_{\beta\to 1} x_\beta = \langle \gamma, x\rangle e \qquad \forall x\in\mathbb{R}^k$$

where γ is the unique left Perron-Frobenius eigenvector of the Jacobian of T at 0.

3.2 Star-shaped operators

In this section, we consider a family of normalized, monotone, translation invariant, and concave operators $\{S_\alpha\}_{\alpha\in\mathcal{A}}$ such that $E(S_\alpha) = D$ for all $\alpha \in \mathcal{A}$. Given $\{S_\alpha\}_{\alpha\in\mathcal{A}}$, we define $T : \mathbb{R}^k \to \mathbb{R}^k$ by $T(x) = \sup_{\alpha\in\mathcal{A}} S_\alpha(x)$ for all $x \in \mathbb{R}^k$.[8] It is immediate to show that T is normalized, monotone, and translation invariant. We say that $\{S_\alpha\}_{\alpha\in\mathcal{A}}$ is *nice* if and only if the previous sup is achieved for all $x \in \mathbb{R}^k$, that is, for each $x \in \mathbb{R}^k$ there exists $\alpha_x \in \mathcal{A}$ such that $T(x) = S_{\alpha_x}(x)$. In such a case, we have that $E(T) = D$.[9]

Given $x \in \mathbb{R}^k$, $\beta \in (0,1)$, and $\alpha \in \mathcal{A}$, we denote by $x_{\beta,\alpha}$ the unique point satisfying $S_\alpha((1-\beta)x + \beta x_{\beta,\alpha}) = x_{\beta,\alpha}$. For each $\alpha \in \mathcal{A}$ we can define $\varphi_{S_\alpha} : \mathbb{R}^k \to \mathbb{R}$ by

$$\varphi_{S_\alpha}(x) = \min_{\gamma\in\Gamma(\partial_C S_\alpha(0))} \langle \gamma, x\rangle \qquad \forall x \in \mathbb{R}^k.$$

The importance of the maps φ_{S_α} is underlined by Corollary 2, since $\varphi_{S_\alpha}(x)e = \lim_{\beta\to 1} x_{\beta,\alpha}$ for all $x \in \mathbb{R}^k$ and for all $\alpha \in \mathcal{A}$. The next result shows that when $\{S_\alpha\}_{\alpha\in\mathcal{A}}$ is nice, the nets $\{x_\beta\}_{\beta\in(0,1)}$ and $\{\tilde{x}_\beta\}_{\beta\in(0,1)}$, defined for the operator T, converge and the limit is given by the sup of the evaluations $\{\varphi_{S_\alpha}(x)\}_{\alpha\in\mathcal{A}}$.

Proposition 3 *If $\{S_\alpha\}_{\alpha\in\mathcal{A}}$ is nice, then $\lim_{\beta\to 1} \tilde{x}_\beta = \lim_{\beta\to 1} x_\beta = \sup_{\alpha\in\mathcal{A}} \varphi_{S_\alpha}(x)e$.*

The importance of the above result is twofold. First, in some applications, a collection $\{S_\alpha\}_{\alpha\in A}$ forms the primitives of the problem and the operator T is derived from it (see, e.g., Section 5). Second, Proposition 3 is useful in proving our main result (Theorem 1). For, a normalized, monotone, translation invariant, and star-shaped operator with regular $\underline{A}(T)$ can always be rewritten as the max of a nice collection $\{S_\alpha\}_{\alpha\in\mathcal{A}}$.

In the next sections, we consider several applications of the results presented in this section. We start with an application to zero-sum stochastic games. In this case, the limit $\beta \to 1$ is interpreted as the limit for infinite patience of the players. Next, we consider an application to an equilibrium model of interconnected financial institutions that use coherent risk measure to evaluate the riskiness of their positions. Here, the limit $\beta \to 1$ captures the idea that the institutions are highly interconnected and that the financial sector dominates the underlying real one. Finally, we apply the results on discrete iterations of Section 7.2 to an extension of the dynamic opinion aggregation

[8] Here the sup is performed coordinatewise. Since S_α is normalized and monotone for all $\alpha \in A$, we also have that $-\|x\|_\infty e \le S_\alpha(x) \le \|x\|_\infty e$ for all $x \in \mathbb{R}^k$ and for all $\alpha \in A$. This implies that each coordinate of $\sup_{\alpha\in A} S_\alpha(x)$ is finite and bounded by the same bounds.

[9] Consider $x \in E(T)$. By construction of T, there exists $\bar{\alpha} \in A$ such that $x = T(x) = S_{\bar{\alpha}}(x)$, yielding that $x \in E(S_{\bar{\alpha}}) = D$. This shows that $E(T) \subseteq D$. The opposite inclusion follows from normalization.

model in networks of Cerreia-Vioglio et al. [12]. In this case, the sequence of weights $\{\beta_t\}_{t\in\mathbb{N}} \subseteq (0,1)$ represents the vanishing stubbornness weight that agents assign to their initial opinions.

4 Application I: Endogenous Networks

In this section, we consider coordination games and competitive equilibria on networks with endogenous links. In both cases, when we fix a network structure, the induced equilibrium map is linear, a feature highly exploited in the literature of coordination games (e.g., Ballester et al. [6] and Golub and Morris [24]) and production networks (e.g., Acemoglu et al. [2]). However, the endogeneity of the network structure introduces nonlinearities in the equilibrium map, thereby complicating the equilibrium analysis. The nonlinear fixed point equation that we have studied in Section 3 implicitly defines the equilibrium maps of both applications, allowing us to characterize it for the limit for high-coordination motives among players and firms respectively.

4.1 Production networks

Following Kopytov et al. [28], we consider a static and frictionless model of *production network* among cost-minimizing firms with Cobb-Douglas production functions and endogenous networks. The endogeneity of the network structure introduces nonlinearities which are not present in the standard linear model (see Acemoglu et al. [2]). This makes the analysis considerably less tractable. Thanks to our results, we completely characterize the equilibrium prices and outputs as the firms' idiosyncratic shocks vanish.

Consider a finite set of firms $\{1,...,k\}$ each of which produces a potentially different output. Each firm can choose a set of weights $w_i \in \Delta$ specifying both the set of inputs from the other firms that are used in production and how these inputs are to be combined. Moreover, each firm uses an external input that is irreproducible by any of the other firms and whose productivity and importance in the production function are fixed. This can be either labor or another factor that is produced outside the economy we analyze.

As in [28], we fix a productivity shifter $S_i : \Delta \to [0,1]$ that depends on the technology w_i selected. Given the level of inputs from the external factor and from the firms in the economy $Q_i = \left(Q_{i0,}\, (Q_{ij})_{j=1}^{k}\right) \in \mathbb{R}_{+}^{k+1}$ and technology $w_i \in \Delta$, the production function of firm i is

$$F_i\left(Q_i, w_i\right) = S_i\left(w_i\right)\xi\left(\beta, w_i\right)\left(Z_i Q_{i0}\right)^{(1-\beta)} \prod_{j=1}^{k} Q_{ij}^{\beta w_{ij}}$$

where $Z_i > 0$ is the productivity relative to the external factor for firm i, $\beta \in (0,1)$ is the common intensity of the external factor, and

$$\xi(\beta, w_i) = (1-\beta)^{-(1-\beta)} \prod_{j=1}^{k} (\beta w_{ij})^{-\beta w_{ij}},$$

is a normalization constant that only depends on the overall technology (β, w_i) of firm i.[10] Each firm selects both a technology $w_i \in \Delta$ and levels of all inputs Q_i needed given the technology selected. For example, if $w_{ij} = 0$ then the input from j is not relevant for i's production.

For a given profile of productivity shifters $S = (S_i)_{i=1}^k$, define the set

$$\operatorname{argmax}(S) = \{W \in \mathcal{W} : S_i(w_i) = 1 \quad \forall i \in \{1, .., k\}\}.$$

The production networks $W \in \operatorname{argmax}(S)$ are the most efficient ones since the production of each firm is not being shifted down by a discount factor. Instead, all those production networks $W \in \mathcal{W}$ such that $S_i(w_i) = 0$ for some $i \in \{1, ..., k\}$ are either extremely inefficient or unfeasible. We maintain the following assumptions on the productivity shifters.

Assumption: The profile of productivity shifters $S = (S_i)_{i=1}^k$ is such that each S_i is upper semicontinuous, log-concave, and

$$\emptyset \neq \operatorname{argmax}(S) \subseteq \mathcal{W}^* \tag{11}$$

Upper semicontinuity is a technical condition that guarantees existence of equilibrium. In turn, equation (11) says that there exist efficient technologies and that every resulting production network admits a unique LPF eigenvector. The next examples illustrate natural settings where our assumption is satisfied.

Example 2 When all the feasible technologies are efficient, we have that the productivity shifter of each i is $S_i = 1_{C_i}$ an indicator function over a nonempty, convex, and compact set $C_i \subseteq \Delta$ of technologies. In this case, $\operatorname{argmax}(S)$ is the set of all stochastic matrices whose i-th row belongs to C_i. The inclusion condition in (11) amounts to assume that each such matrix admits a unique LPF eigenvector. When each C_i is a singleton, $\{w_i^0\}$, for some $W^0 \in \mathcal{W}^*$, we have that $\operatorname{argmax}(S) = \{W^0\}$, that is, the production network is exogenously fixed and we get back the standard Cobb-Douglas model of Acemoglu et al. [2]. Differently, Kopytov et al. [28] consider a continuously

[10]Note that the definition of the normalization constant $\xi_i(\beta, w_i)$ is the same as the one [28] (see their Footnote 8). Differently form [28], our productivity shock Z_i is relative to the external factor Q_{i0} as opposed to be Hicks-neutral.

differentiable and strictly log-concave function productivity shifter. In their leading example, they consider

$$S_i(w_i) = \exp\left(-\sum_{j=1}^{k}\kappa_{ij}\left(w_{ij}-w_{ij}^0\right)^2\right) \qquad \forall w_i \in \Delta \tag{12}$$

where $W^0 \in \mathcal{W}^*$ is the efficient production network for the economy and κ is a positive vector of weights capturing the cost, in terms of productivity, of moving the j-th input share away from its ideal value. Following a parallel logic, we can replace the quadratic distance in equation (12) with another "distance" such as the relative entropy to obtain

$$S_i(w_i) = \exp\left(-\lambda_i R\left(w_i||w_i^0\right)\right). \tag{13}$$

In this case, $R(\cdot||\cdot)$ is the relative entropy while $W^0 \in \mathcal{W}^*$ and $\lambda_i > 0$. In both these smooth cases, we have $\operatorname{argmax}(S) = \{W^0\}$. In general, this is the case every time that each S_i is strictly log-concave (as in [28]). ▲

Next, we proceed with the description of the equilibrium of the economy. We assume that firms are price takers and act in a perfect-competition economy. We normalize the price of the external factor to 1 and, given a vector $P \in \mathbb{R}_+^k$ of inputs' prices and a feasible technology $w_i \in \Delta$, the cost-minimization problem for firm i, producing at least 1 unit of output, is defined by

$$K_i(P, w_i) = \min_{Q_i \in \mathbb{R}_+^{k+1}}\left\{Q_{i0} + \sum_{j=1}^{k} Q_{ij}P_j : F_i(Q_i, w_i) \geq 1\right\} \qquad \forall i \in \{1, ..., k\}.$$

Because each firm can choose its technology w_i so to minimize their unitary cost, the equilibrium zero-profit condition is

$$P_i = \min_{w_i \in \Delta} K_i(P, w_i) = \min_{(w_i, Q_i) \in \Delta \times \mathbb{R}_+^{k+1}}\left\{Q_{i0} + \sum_{j=1}^{k} Q_{ij}P_j : F_i(Q_i, w_i) \geq 1\right\} \quad \forall i \in \{1, ..., k\}. \tag{14}$$

Note that the above equilibria are β dependent. In particular, for each $\beta \in (0, 1)$, an equilibrium is given by a vector of prices $P \in \mathbb{R}_+^k$, a matrix of inputs $Q \in \mathbb{R}_+^{k\times(k+1)}$, and a network structure $W \in \mathcal{W}$. In the triple (P, Q, W), the vector P solves the fixed point equation (14) and the pair (Q_i, w_i) solves the cost-minimization problem in the right hand-side of equation (14).

Following the same steps in [28],[11] the fixed point condition for equilibrium log-prices can be written as

$$p_i = (1-\beta)x_i + \beta \min_{w_i \in \Delta}\left\{\sum_{j=1}^{k} w_{ij}p_j + \frac{1}{\beta}c_i(w_i)\right\} \qquad \forall i \in \{1, ..., k\} \tag{15}$$

[11] A proof is available on request.

where $p_i = \ln(P_i)$, $x_i = \ln(1/Z_i)$, and $c_i(w_i) = \ln(1/S_i(w_i))$. It is standard to show that, for each $\beta \in (0,1)$, there exists a unique vector of log-prices p_β that solves the fixed point equation (15) and therefore a unique vector of equilibrium prices P_β. Given these prices, the equilibrium network and quantities are not unique in general due to the fact that each firm might have multiple optimal technologies, that is,

$$\underset{w_i \in \Delta}{\operatorname{argmin}} \left\{ \sum_{j=1}^{k} w_{ij} p_{\beta,j} + \frac{1}{\beta} c_i(w_i) \right\} \tag{16}$$

might not be single-valued. When S_i is strictly log-concave, as in [28], it follows that c_i is strictly convex and there exists a unique minimizer $w_{\beta,i}$ in equation (16). This in turn uniquely pins down the equilibrium inputs Q_β.

We aim to characterize the vector of equilibrium prices in the limit for a vanishing intensity of the external factor, that is, we aim to compute $\lim_{\beta\to 1} p_\beta$. To this extent, define

$$\Gamma(S) = \left\{\gamma \in \Delta : \exists W \in \operatorname{argmax}(S), \gamma^{\mathrm{T}} = \gamma^{\mathrm{T}} W\right\},$$

the set of all LPF eigenvectors of the efficient production network. With this, we have the following result:

Proposition 4 *The limit equilibrium vector of prices is constant across firms and*

$$\lim_{\beta\to 1} p_{\beta,i} = \min_{\gamma \in \Gamma(S)} \langle \gamma, x \rangle \qquad \forall i \in \{1, ..., k\}.$$

Moreover, if S_i is continuously differentiable and strictly log-concave for all $i \in \{1, ..., k\}$ with $\operatorname{argmax}(S) = \{W^0\}$, *then*

$$\lim_{\beta\to 1} p_{\beta,i} = \langle \gamma_{W^0}, x \rangle, \ \lim_{\beta\to 1} w_{\beta,i} = w_i^0, \ \lim_{\beta\to 1} Q_{\beta,i0} = 0, \text{ and } \lim_{\beta\to 1} Q_{\beta,ij} = w_{ij}^0.$$

So far, we considered endogenous production networks under a Cobb-Douglas production function. Another potential source of nonlinearity recently studied in the literature of production networks comes from generalizing the production function to the class of nested CES (see for example Baqaee and Farhi [8]). It turns out that our method can also be applied in this case. For simplicity, fix a production network $W \in \mathcal{W}^*$, and assume that the production function of firm i is

$$F_i(Q_i) = \hat{\xi}_i(\beta, w_i)(Z_i Q_{i0})^{(1-\beta)} \left(\sum_{j=1}^{k} w_{ij}^{1/\sigma_i} Q_{ij}^{(\sigma_i - 1)/\sigma_i} \right)^{\beta\sigma_i/(\sigma_i - 1)}$$

where σ_i is the elasticity of substitution among the inputs of the firm and $\hat{\xi}_i(\beta, w_i)$ is a normalization constant (potentially different from the one before) that only depends on β and the fixed technology w_i. It is standard to show that in this case the log prices are uniquely characterized by the fixed-point condition

$$p_i = (1-\beta)\, x_i + \beta \frac{1}{1-\sigma_i} \ln \left(\sum_{j=1}^{k} w_{ij} \exp\left((1-\sigma_i)\, p_j\right) \right) \qquad \forall i \in \{1, ..., k\}$$

4.2 Coordination games

We consider a finite set of agents $N = \{1, ..., n\}$ playing a complementary-effort game on an endogenous network. Each agent chooses how much effort to exercise in the partnership with other agents: $a_i \in \mathbb{R}_+$. The benefit of effort for agent i is directly proportional to a linear combination of her ability $x_i \in \mathbb{R}_+$ with a weighted average of the efforts exercised by her neighbors. The cost of effort is instead quadratic, a feature that will guarantee linearity of the best response for a given network structure.

Formally, given a fixed weighted and directed network $W \in \mathcal{W}$, the payoff of agent i for every profile of actions $a = (a_i)_{i=1}^n$ is

$$u_i\,(a, w_i) = a_i \left((1-\beta)\, x_i + \beta \sum_{j=1}^{n} w_{ij} a_j \right) - \frac{a_i^2}{2}$$

where $\beta \in (0,1)$ captures the relative importance of complementary efforts over the personal skills of every agent. In what follows, we consider two different cases of endogenous networks. In both cases, we assume that each feasible network structure W has two features: (i) there is no self-link, that is, $w_{ii} = 0$ and (ii) there exists a unique stationary distribution (i.e., a unique eigenvector centrality). The first assumption is standard in coordination games on networks (cf. [6] and [24]). We discuss the relevance of the second assumption below. Let us denote the set of stochastic matrices satisfying both (i) and (ii) with $\mathcal{W}_0^*$.

Costly link formation Here we assume that, before choosing her effort, each agent i chooses her weighted links $w_i \in \Delta_n$. This is costly and the cost function of agent i is denoted by $c_i : \Delta_n \to [0, \infty]$. In particular, those weighted networks $W \in \mathcal{W}$ such that $c_i\,(w_i) = 0$ for all $i \in N$ are the costless one. Instead, all those weighted networks $W \in \mathcal{W}$ such that $c_i\,(w_i) = \infty$ for some $i \in N$ are unfeasible. We maintain the following assumptions on the cost functions. We let

$$\text{argmin}\,(c) = \left\{ W \in \mathcal{W} : \forall i \in N, w_i \in co\left(c_i^{-1}\,(0)\right) \right\}$$

denote the set of network structures that can be obtained by mixing vector of weights that are costless for all the players.[12]

[12] For every set $K \subseteq \Delta_n$, so for example for $K = c_i^{-1}\,(0)$, we let $co\,(K)$ denote the convex hull of K.

Assumption: The profile of cost functions $c = (c_i)_{i=1}^n$ is such that each c_i is lower semicontinuous and

$$\emptyset \neq \operatorname{argmin}(c) \subseteq \mathcal{W}_0^*. \tag{17}$$

Lower semicontinuity of the cost functions is a technical condition that guarantees existence of a well-defined best response map for the coordination game. In turn, equation (17) says that each agent has at least a free vector of weighted links and that every free weighed network admits a unique stationary distribution. The next example illustrates a setting where our assumption on the cost functions is satisfied

Example 3 Assume that the agents are connected on a baseline unweighted and strongly connected network represented by a graph $G \in \{0,1\}^{n\times n}$ with $g_{ii} = 0$. Maintaining the links specified in G is free for all the agents. However, they can costly form new links in additional to the ones in G. In particular, there is a fixed cost $k > 0$ for each additional link that player i forms on top of the baseline ones. We next show how this particular case of costly link formation can be represented by a profile of cost functions $(c_i)_{i=1}^n$ satisfying our assumption. Let $N_i(G) \subseteq \{1, ..., n\}$ denote the set of neighbors of i in the graph G. For every $i \in N$, define the set of uniform weights

$$D_i(G) = \left\{ \frac{1}{|N_i|} \sum_{j\in N_i} \delta_j \in \Delta_n : N_i(G) \subseteq N_i \subseteq N \backslash \{i\} \right\}$$

and the cost function $c_i : \Delta_n \to [0,\infty]$ as

$$c_i(w_i) = k\left|\{j \in N : w_{ij} > 0\} \backslash N_i(G)\right| + \mathbf{I}_{D_i(G)}(w_i)$$

where $I_{D_i(G)}$ is the equal to 0 if $w_i \in D_i(G)$ and ∞ otherwise. On the one hand, it is easy to see that c_i is lower semicontinuous. On the other hand, the uniform network $W(G)$ defined by $w_i(G) = \frac{1}{|N_i(G)|}\sum_{j\in N_i(G)} \delta_j$ for all $i \in N$ is costless for every player. Moreover, given that $k > 0$, this is the only costless network for all the agents, that is $\arg\min(c) = \{W(G)\} \subseteq W_0^*$, were the last inclusion follows from the properties of G. Therefore $(c_i)_{i=1}^n$ satisfy our assumption. ▲

For a fixed profile of cost functions c, we assume that the total payoff of each player $i \in N$ given a profile of efforts $a \in \mathbb{R}_+^n$ and weighted links $w_i \in \Delta_n$ is $u_i(a, w_i) - a_i c_i(w_i)$. In words, the total cost of forming and maintaining the link is increasing and linear in the effort chosen. The assumption that the effort and the weighted links are complementary in increasing the total cost of the player has been already considered by Sadler and Golub [32] in a context of endogenous link formation. Here, we are adding the linearity assumption which, as we show below, allows us to exploit our previous result to characterize the limit equilibrium of the game and, in Section 7.1, to study the epsilon-equilibria of the game for β away from 1.

We next analyze the best response map of the total game of choosing both the weighted links and the effort. In particular, observe that neither the payoff function u_i nor the cost function c_i of i depend on the links chosen by the other agents. Therefore, given a conjecture $a_{-i} \in \mathbb{R}_+^{n-1}$ about the effort of the other agents, player i solves

$$\max_{a_i \in \mathbb{R}_+} \max_{w_i \in \Delta_n} \left\{ u_i \left(a, w_i \right) - a_i c_i \left(w_i \right) \right\}.$$

Observe that the previous maximization problem can be rewritten as

$$\max_{a_i \in \mathbb{R}_+} \left\{ a_i \max_{\tilde{w}_i \in \Delta_n} \left\{ \beta \sum_{j=1}^{n} \tilde{w}_{ij} a_j - c_i \left(\tilde{w}_i \right) \right\} - \frac{a_i^2}{2} \right\}.$$

Therefore the induced objective function is still quadratic with respect to the choice variable a_i, hence the unique best response can be still characterized by the first-order conditions. In general, this implies that a profile of efforts $a \in \mathbb{R}_+^n$ and a weighted network $W \in \mathcal{W}$ form a Nash equilibrium of the total game if and only if, for every $i \in N$,

$$a_i = (1 - \beta) x_i + \max_{\tilde{w}_i \in \Delta_n} \left\{ \beta \sum_{j=1}^{n} \tilde{w}_{ij} a_j - c_i \left(\tilde{w}_i \right) \right\} \tag{18}$$

and

$$w_i \in \operatorname*{argmax}_{\tilde{w}_i \in \Delta_n} \left\{ \beta \sum_{j=1}^{n} \tilde{w}_{ij} a_j - c_i \left(\tilde{w}_i \right) \right\}.$$

The first condition is a standard fixed-point equation on the profile actions a. The main difference with respect to the game with a fixed weighted network is the nonlinearity of the fixed point equation. However, we show below that it can be still analyzed through the results of the previous sections. The second condition instead requires that the equilibrium network is a best response for each player given the efforts chosen by the others. Notably, the optimal choice of weights does not depend directly on the own effort. This is due to our assumption of multiplicative interaction between the cost of weights and the own effort which makes the latter a shifter that does not alter the incentives in choosing the weighted links.

It is not hard to see that, for every $\beta \in (0, 1)$, there exists a unique equilibrium profile of efforts $a_\beta \in \mathbb{R}_+^n$ that solves the fixed-point equation (18). We aim to characterize the limit for high coordination motives $\lim_{\beta \to 1} a_\beta$. First, define

$$\Gamma(c) = \left\{ \gamma \in \Delta_n : \exists W \in \operatorname{argmin}(c), \gamma^{\mathrm{T}} = \gamma^{\mathrm{T}} W \right\},$$

the set of all the eigenvector centralities of networks that are costless. With this we have the following result.

Proposition 5 *The limit equilibrium profile of efforts is well defined, constant across players, and equal to* $\lim_{\beta \to 1} a_{\beta,i} = \max_{\gamma \in \Gamma(c)} \langle \gamma, x \rangle$ *for every* i.

The main implication is that the most central agent in any of the limit equilibrium networks are those that are at the same time most efficient (higher x_i) and cheaper to link with.

Example 4 As already established, we have $\arg\min(c) = \{W(G)\}$ where $w_i(G) = \frac{1}{|N_i(G)|}\sum_{j\in N_i(G)}\delta_j$ for all $i \in N$. This implies that $W(G)$ is the unique equilibrium network consistent with the limit for $\beta \to 1$. Moreover, it is well known that the eigenvector centrality of $W(G)$ is given by

$$\gamma_i(G) = \frac{|N_i(G)|}{\sum_{j\in N}|N_j(G)|} \qquad \forall i \in N.$$

With this, we have $\Gamma(c) = \{\gamma(G)\}$, hence that

$$\lim_{\beta\to 1} a_{\beta,i} = \frac{\sum_{j\in N}|N_j(G)|\,x_j}{\sum_{j\in N}|N_j(G)|}.$$

Therefore, the common equilibrium effort is relatively higher if the agents who are relatively more efficient (i.e., high x_i) are also those that are more central in the baseline network. ▲

5 Application II: Zero-sum stochastic games

In this section, we consider zero-sum stochastic games with finitely many states and a continuum of actions for both players. We closely follow the textbook formalization of Sorin [35, Chapter 5].

There are two players repeatedly interacting in a zero-sum game under uncertainty. We identify the two player as the *maximizer* and the *minimizer*. Time is discrete $t \in \mathbb{N}$ and at each period the game is at a state drawn from a finite set Ω. At the end of each period, an outcome r from a finite set $R \subseteq \mathbb{R}$ realizes and the maximizer gets payoff r and the minimizer gets $-r$. The set of feasible actions for the maximizer and the minimizer are respectively denoted by S and Q, two compact metric spaces. Both the outcome at period t and the state at period $t+1$ depend on players' actions and the state at period t. Formally, this is described by a *continuous* transition map $\rho : S \times Q \times \Omega \to \Delta(R \times \Omega)$.[13] With a small abuse of notation, we also use ρ to denote its linear extension $\rho : \Delta(S) \times \Delta(Q) \times \Omega \to \Delta(R \times \Omega)$ to mixed actions as

[13]The linear extension of ρ is defined as usual:

$$\rho(\hat{s},\hat{q},\omega)(r,\omega') = \int_S\int_Q \rho(s,q,\omega)(r,\omega')\,d\hat{s}(s)\,d\hat{q}(q),$$

for all $\hat{s} \in \Delta(S)$ and $\hat{q} \in \Delta(Q)$.

well as the corresponding marginal distributions over R and Ω. With this, define the state-dependent one-period expected reward $g : \Delta(S) \times \Delta(Q) \times \Omega \to \mathbb{R}$ as

$$g(\hat{s}, \hat{q}, \omega) = \sum_{r' \in R} r' \rho(\hat{s}, \hat{q}, \omega)(r').$$

This setting is equivalent to the more standard one where there are no outcomes and the primitive objects are a transition function $\rho : S \times Q \times \Omega \to \Delta(\Omega)$ and a one-period reward function $g : S \times Q \times \Omega \to \mathbb{R}$ (e.g., Sorin [35, Chapter 5]). We introduce outcomes so to obtain a cleaner limit characterization using our methods.

Following the standard analysis of zero-sum stochastic games, we consider two different cases: (i) the one-period game is infinitely repeated and the agents maximize their discounted expected payoffs with common discount factor $\beta \in (0, 1)$; (ii) the one-period game is repeated only t times and the agents maximize the time average of their expected payoffs.

In case (i), it is well known that, for each discount factor $\beta \in (0, 1)$, the value of the game $v^{\beta} \in \mathbb{R}^{\Omega}$ exists and is the unique solution of the Shapley equation:

$$v^{\beta}_{\omega} = \max_{\hat{s} \in \Delta(S)} \min_{\hat{q} \in \Delta(Q)} \left\{ (1 - \beta) g(\hat{s}, \hat{q}, \omega) + \beta \sum_{\omega' \in \Omega} v^{\beta}_{\omega'} \rho(\hat{s}, \hat{q}, \omega)(\omega') \right\} \qquad \forall \omega \in \Omega. \tag{19}$$

Similarly, in case (ii), for every length $t \in \mathbb{N}$, the value of the game $v^{t} \in \mathbb{R}^{\Omega}$ exists and satisfies the following recursive equation:

$$v^{t}_{\omega} = \max_{\hat{s} \in \Delta(S)} \min_{\hat{q} \in \Delta(Q)} \left\{ \frac{1}{t} g(\hat{s}, \hat{q}, \omega) + \frac{t-1}{t} \sum_{\omega' \in \Omega} v^{t-1}_{\omega'} \rho(\hat{s}, \hat{q}, \omega)(\omega') \right\} \qquad \forall \omega \in \Omega. \tag{20}$$

We say that the game has an *asymptotic value* (cf. Sorin [35]) if and only if both $\lim_{\beta \to 1} v^{\beta}$ and $\lim_{t} v^{t}$ exist and coincide.[14] Our abstract analysis of nonlinear fixed points yields the existence of the asymptotic value and its explicit form under a minimal connectedness assumption.

We first need some preliminary definitions. Let $\Sigma_S = \Delta(S)^{\Omega}$ and $\Sigma_Q = \Delta(Q)^{\Omega}$ denote the set of stationary mixed strategies of the agents and, for all $\sigma_S \in \Sigma_S$ and $\sigma_Q \in \Sigma_Q$, let $W(\sigma_S, \sigma_Q)$ denote the transition matrix between state-outcome pairs with entries given by

$$w_{(r,\omega),(r',\omega')}(\sigma_S, \sigma_Q) = \rho(\sigma_S(\omega), \sigma_Q(\omega), \omega)(r', \omega') \qquad \forall (r, \omega), (r', \omega') \in R \times \Omega.$$

Let $n = |R \times \Omega|$. Next we state the crucial assumption that allows us to apply our result to the current stochastic-game setting.

[14] The derivations of equations (19) and (20) can be found in Sorin [35, Propositions 5.2 and 5.3].

Assumption A: There exists a regular adjacency matrix $A \in \{0,1\}^{n \times n}$ such that $\underline{A}(W(\sigma_S, \sigma_Q)) \geq A$ for all $\sigma_S \in \Sigma_S$ and $\sigma_Q \in \Sigma_Q$.

In words, we assume that there exist links between outcome-state pairs, the ones prescribed by A, that cannot be shut down by the actions of the players, no matter what. Moreover, these baseline links are nontrivial and such that there exists a unique essential class of essential pairs (r, ω). Importantly, this implies that each $W(\sigma_S, \sigma_Q)$ admits a unique Perron-Frobenius eigenvector, denoted by $\gamma(\sigma_S, \sigma_Q) \in \Delta(R \times \Omega)$. With the same abuse of notation as before we use the same symbol $\gamma(\sigma_S, \sigma_Q)$ for its marginal over outcomes.

Proposition 6 *Under Assumption A, the game has an asymptotic value that is independent of the state and such that*

$$\lim_{\beta \to 1} v^{\beta} = \lim_{t} v^{t} = \left(\sup_{\sigma_S \in \Sigma_S} \min_{\sigma_Q \in \Sigma_Q} \sum_{r \in R} r \gamma(\sigma_S, \sigma_Q)(r) \right) e.$$

This result extends the standard result on the existence and characterization of the asymptotic value of zero-sum stochastic games from the finite case to the class of games considered in the current section (see for example Sorin [35, Propositions 5.12-5.14]). As in the finite case, the asymptotic value coincides with the value of static zero-sum game with expected payoffs given by the stationary distributions generated by the players' strategies.

Whenever Q is a singleton, we obtain a Markov decision process (MDP) where the minimizer is optimally controlling her cost. With this, Proposition 6 collapses to average-cost optimality for MDPs with a continuum of actions and finitely many states, that is,

$$\lim_{\beta \to 1} v^{\beta} = \lim_{t} v^{t} = \left(\sup_{\sigma_S \in \Sigma_S} \sum_{r \in R} r \gamma(\sigma_S)(r) \right) e.$$

Finally, observe that, in this setting, Theorem 2 can be applied and would deliver an estimate on how the value of the game depends on the current state for every $\beta \in (0,1)$.

6 Application III: Opinion aggregation with stubbornness

Cerreia-Vioglio et al. [12] consider a finite set of agents $i \in \{1, ..., k\}$ and let $x \in \mathbb{R}^k$ denote an arbitrary profile of opinions for the agents. An opinion is just a real number that can be interpreted as the estimate of agent i about some fundamental parameter of interest or the intensity with which an individual agrees with a certain policy. Under this interpretation, they assume that the opinions of the agent evolve according to the operator T, that is, if the current profile of opinions is x, then the profile of opinions in the next period is $T(x)$. With this, the sequence of iterates $\{T^t(x)\}_{t=1}^{\infty}$ corresponds to the sequence of profile of opinions in the population over time. For example, when $T = W$ is linear, we obtain the celebrated DeGroot's model [17] of opinion aggregation of experts.[15] In particular, Golub and Jackson [23] interpreted W as a directed and weighted network where the entry w_{ij} represents the weighted link from j to i. In general, motivated by the fact agents may use opinion aggregators reflecting their attraction or aversion for extreme opinions, [12] introduce several classes of nonlinear opinion aggregators T. For example, when

$$T_i(x) = \frac{1}{\lambda_i} \ln \left(\sum_{j=1}^{k} w_{ij} \exp(\lambda_i x_j) \right) \tag{21}$$

for some fixed set of weighted links $w_i \in \Delta$ and a parameter $\lambda_i \in \mathbb{R}$, it is possible to model agents with heterogeneous attractions for high ($\lambda_i > 0$) or low ($\lambda_i < 0$) opinions while maintaining the underlying linear network structure. Alternatively, we can altogether relax the existence of a single network structure and consider opinion aggregators such as

$$T_i(x) = \alpha_i \min_{w_i \in C_i} \langle w_i, x \rangle + (1 - \alpha_i) \max_{w_i \in C_i} \langle w_i, x \rangle \tag{22}$$

where $C_i \subseteq \Delta$ is a compact and convex set of possible weighted links to i and $\alpha_i \in [0, 1]$ is a parameter capturing the relative attraction of i for high or low opinions. It is routine to show that if each element T_i of T is defined as in equations (21) or (22), then T is monotone, normalized, and translation invariant.

Friedkin and Johnsen [19] proposed a variation of the DeGroot's model where the agents have a degree of stubbornness with respect to their initial opinions. Here we extend Friedkin and Johnsen's model of stubbornness by considering nonlinear opinion aggregators T with the functional properties introduced above. Formally, we assume that, for every period $t \in \mathbb{N}$, the profile of opinions in the population is

$$\tilde{x}_i^t = (1 - \beta) x_i + \beta T_i \left(\tilde{x}^{t-1} \right) \qquad \forall i \in \{1, ..., k\} \tag{23}$$

where $\beta \in (0, 1)$ is a fixed parameter capturing the degree of stubbornness in the population and $x = \tilde{x}^0$ is the profile of initial opinions. In words, each agent i aggregates the last-period opinions $\tilde{x}^{t-1}$ with her opinion aggregator T_i and then mixes the resulting aggregate with her original opinion, using the common weight β. When $T = W$ is linear, we exactly obtain Friedkin and Johnsen's model. In general, it is easy to see that the sequence of opinions $\{\tilde{x}^t\}_{t=1}^{\infty}$ converges to the unique fixed point $\tilde{x}_\beta$ defined in

[15] See for example Golub and Jackson [23] for a detailed analysis of this model.

equation (2), which then corresponds to the long-run profile of opinions of the agent under stubbornness β. Provided that it exists, the $\lim_{\beta \to 1} \tilde{x}_\beta$ corresponds to the profile of long-run opinions of the agents as the stubbornness friction is vanishing.[16]

The results of Section 3 can be applied to this setting. For example, consider the standard Friedkin and Johnsen model with linear $T = W$ having a single LPF eigenvector γ_W and compare it to an alternative opinion aggregator $\tilde{T}$ where the agents have the same network structure W but aggregate opinions according to equation (21) for some profile of parameters $(\lambda_1, ..., \lambda_k) \in \mathbb{R}^k$. Because $\tilde{T}$ is continuously differentiable with $J_{\tilde{T}}(0) = W$, Corollary 1 states that, regardless of the value of $(\lambda_1, ..., \lambda_k)$, the opinions of the agents will converge to the same consensus $\langle \gamma_W, x \rangle$ for both T and $\tilde{T}$.

Alternatively, consider the opinion aggregator T defined as in equation (22) such that, for every $i \in \{1, ..., k\}$, we have $\alpha_i = \alpha$ and

$$C_i = \left\{ (1 - \varepsilon) w_i^0 + \varepsilon w_i : w_i \in \Delta \right\}$$

for some $\alpha \in [0, 1]$, $\varepsilon \in [0, 1)$, and stochastic matrix $W \in W$ with $A(W)$ regular. Also fix a vector of initial opinions $x \in \mathbb{R}^k$ and define

$$\hat{x}^\varepsilon = \varepsilon \left[I - (1 - \varepsilon) W \right]^{-1} x.$$

Observe that T can be written as $T = (1 - \varepsilon) W + \varepsilon S$ where

$$S_i(z) = \alpha \min_{j \in \{1,...,k\}} z_j + (1 - \alpha) \max_{j \in \{1,...,k\}} z_j \qquad \forall z \in \mathbb{R}^k, \forall i \in \{1, ..., k\}.$$

Therefore, Proposition 3 implies that the long-run opinions as the stubbornness vanishes converge to the consensus

$$\alpha \min_{j \in \{1,...,k\}} \hat{x}_j^\varepsilon + (1 - \alpha) \max_{j \in \{1,...,k\}} \hat{x}_j^\varepsilon.$$

In words, we first need to compute the vector of opinions of the agents obtained by applying the matrix of ε-weighted Bonacich centralities of W to x and the linearly combine the maximum and the minimum of the opinions so obtained.

7 Additional results

In this section, we provide additional results complementary to our main convergence result and illustrate them by revisiting some of the economic applications proposed.

[16] Banerjee and Compte [7] provided a game-theoretic foundation for this limit by considering a noisy version of the Friedkin and Johnsen's model where the agents choose once and for all the stubbornness weight to assign to their initial opinion so to maximize the accuracy of their long-run opinion. They show that as the noise vanishes, the symmetric equilibrium weight converges to zero, that is $\beta \to 1$, providing an alternative foundation for the limit we study.

7.1 Fit of the approximation

The goal of this section is to provide estimates on the rate of convergence of the nets $\{x_\beta\}_{\beta\in(0,1)}$ and $\{\tilde{x}_\beta\}_{\beta\in(0,1)}$. In order to achieve this, we observe that all our previous results are for operators whose fixed points are the constant vectors. Conceptually, this makes the quantities

$$\max \tilde{x}_\beta - \min \tilde{x}_\beta \text{ and } \max x_\beta - \min x_\beta$$

interesting. In fact, both converge to zero as β goes to 1. We first bound these two quantities and then use them to provide an estimate for the rate of convergence. Perhaps interestingly, computing these bounds does not require to know that $\{x_\beta\}_{\beta\in(0,1)}$ and $\{\tilde{x}_\beta\}_{\beta\in(0,1)}$ converge.

7.1.1 Range

Given a vector $y \in \mathbb{R}^k$, we denote by $\operatorname{Rg} y$ the quantity $\max_{i\in\{1,\dots,k\}} y_i - \min_{i\in\{1,\dots,k\}} y_i$. We define

$$\delta = \min_{i,j:\underline{a}_{ij}=1} \inf_{x\in\mathcal{D}} \frac{\partial T_i}{\partial x_j}(x).$$

Next, consider the adjacency matrix $\underline{A}(T) \vee I$ which coincides with $\underline{A}(T)$ with the possible exception of the diagonal where the diagonal entries of $\underline{A}(T) \vee I$ are all 1. We can define the quantity

$$t_T = \min\left\{t \in \mathbb{N} : \left(\underline{A}(T) \vee I\right)^t \text{ has a strictly positive column}\right\}.$$

It is well known that if $\underline{A}(T)$ is regular, then t_T is well defined. Moreover, if $\underline{A}(T)$ is strongly connected, one can show that $t_T \leq k-1$ where k is the dimension of the space (see, e.g., [26, Theorem 8.5.9]). In proving Theorem 2, we provide a sharper, yet more convoluted, bound compared to the one reported below. Nevertheless, in both cases, the rate to which the ranges of x_β and $\tilde{x}_\beta$ shrink to 0 are linear.

Theorem 2 *Let T be normalized, monotone, and translation invariant. If $\underline{A}(T)$ is regular, then*

$$\operatorname{Rg}(x_\beta) \leq \operatorname{Rg}(\tilde{x}_\beta) \leq (1-\beta)(1+\kappa_T)\operatorname{Rg}(x) \quad \forall x \in \mathbb{R}^k, \forall \beta \in (0,1)$$

where

$$\kappa_T = \frac{1+\delta}{\min\left\{\frac{1}{t_T}, \left(\frac{\delta}{1+\delta}\right)^{2t_T}\right\}}.$$

7.1.2 Rate of convergence

In this section, we prove that x_β converges at least linearly fast to its limit, provided some extra conditions of differentiability hold. The constants that appear in the statement below are the same defined in the section above. We consider maps which are differentiable and their Jacobian is Lipschitz continuous with constant L. More formally, it is natural to view the gradient of each component $T_i : \mathbb{R}^n \to \mathbb{R}$ as an element of the dual of $\mathbb{R}^n$. Therefore, we use $\| \ \|_1$ to compute the norm of the gradient of T_i.[17] We say that the Jacobian of T is Lipschitz continuous (with constant L) if and only if

$$\|\nabla T_i(x) - \nabla T_i(y)\|_1 \le L \|x-y\|_\infty \quad \forall x, y \in \mathbb{R}^k, \forall i \in \{1, ..., k\}.$$

Theorem 3 *Let T be normalized, monotone, and translation invariant. If T has a Lipschitz continuous Jacobian and $\underline{A}(T)$ is regular, then*

$$\|x_\beta - \langle \gamma, x\rangle e\|_\infty \le (1-\beta)(1+\kappa_T)\left(1 + \frac{(1+\delta)^{t_T - 1}}{\delta^{t_T}} t_T L \|x\|_\infty\right) \mathrm{Rg}(x)$$

for all $x \in \mathbb{R}^k$ and for all $\beta \in (0,1)$ where γ is the unique left Perron-Frobenius eigenvector of the Jacobian of T at 0.

In particular, the result above allows us to conclude that convergence happens at a linear rate.

7.2 Discrete iterations

The limit $\lim_{\beta\to 1} x_\beta$ that we have studied so far can also be seen as the result of a double limit. Indeed, observe that, for every $\beta \in (0,1)$, we have $x_\beta = \lim_t x_\beta^t$, where

$$x_\beta^t = T\left((1-\beta)x + \beta x_\beta^{t-1}\right) \qquad \forall t \in \mathbb{N},$$

with $x^0 = x$. Therefore, by taking the limit for $\beta \to 1$, we are implicitly studying $\lim_{\beta\to 1}\lim_t x_\beta^t$. Importantly, here the order of limits matter for the limit value. Whenever we consider the alternative order, we obtain the sequence $\{T^t(x)\}_{t\in\mathbb{N}}$, which was extensively studied in Cerreia-Vioglio et al. [12]. In general, it is easy to see that $\lim_t T^t(x)$ and $\lim_{\beta\to 1} x_\beta$ do not coincide outside the linear case.

An intermediate approach that turns out to be useful for using our results in applications is to consider a single limit that jointly iterates T and let the dependence on x vanish. Formally, we fix an increasing sequence $\{\beta_t\}_{t\in\mathbb{N}} \subseteq (0,1)$ such that $\lim_t \beta_t = 1$, and consider

$$x^{t+1} = T\left((1-\beta_{t+1})x + \beta_{t+1}x^t\right) \qquad \forall t \in \mathbb{N}_0 \tag{24}$$

[17] Recall that $\|x\|_1 = \sum_{i=1}^k |x_i|$ for all $x \in \mathbb{R}^k$.

with $x_0 = x$. Similarly as before, we can consider the alternative iteration

$$\tilde{x}^{t+1} = (1 - \beta_{t+1})\, x + \beta_{t+1} T\left(\tilde{x}^t\right) \qquad \forall t \in \mathbb{N}_0, \tag{25}$$

with $\tilde{x}_0 = x$. These iterations can be seen as the discrete versions of the nonlinear fixed points analyzed so far (i.e., of equations (1) and (2) respectively).

In the next proposition, we show that, whenever $\lim_{\beta \to 1} x_\beta$ exists, the two limits $\lim_t x^t$ and $\lim_t \tilde{x}^t$ exist and coincide with the former, provided that β_t is asymptotically equivalent to $1 - 1/g(t)$ for some function $g : [1, \infty) \to [0, \infty)$ which is strictly increasing, divergent, concave, continuous, and such that $g(z)/z \to 0$ as $z \to \infty$.

Proposition 7 *Let $g : [1, \infty) \to [0, \infty)$ be strictly increasing, divergent, concave, continuous, and such that $g(z)/z \to 0$ as $z \to \infty$, and define $\beta_t = 1 - 1/g(t)$ for all $t \in \mathbb{N}$. For each $x \in \mathbb{R}^k$, if $\lim_{\beta \to 1} x^\beta$ exists, then $\lim_t x^t$ exists and*

$$\lim_{\beta \to 1} \tilde{x}_\beta = \lim_{\beta \to 1} x_\beta = \lim_t x^t = \lim_t \tilde{x}^t.$$

The previous result includes the case $\beta_t = 1 - 1/t^\alpha$ for some $\alpha \in (0, 1)$ but not the case $\beta_t = \frac{t-1}{t}$, which is in turn relevant for some applications. However, under positive homogeneity of T, the equivalence result holds true also for this important case.

Proposition 8 *Let $\beta_t = \frac{t-1}{t}$ for all $t \in \mathbb{N}$ and let T be normalized, monotone, translation invariant, and positively homogeneous. For all $x \in \mathbb{R}^k$, the following are equivalent:*

(i) $\lim_{\beta \to 1} \tilde{x}_\beta$ *exists;*

(ii) $\lim_{\beta \to 1} x_\beta$ *exists;*

(iii) $\lim_t x^t$ *exists;*

(iv) $\lim_t \tilde{x}^t$ *exists.*

Moreover, in this case, we have $\lim_{\beta \to 1} \tilde{x}_\beta = \lim_{\beta \to 1} x_\beta = \lim_t x^t = \lim_t \tilde{x}^t$.

The previous result heavily relies on Theorem 1 in Ziliotto [37]. The latter result is in a sense much more general than Proposition 8 since it deals with the infinite dimensional case and replaces translation invariance and positive homogeneity with a different nonexpansivity assumption on x_β as a function of β. However, the fixed-point condition considered in [37] is different from ours and does not depend on a given point $x \in \mathbb{R}^k$. We show that, under positive homogeneity, the fixed-point condition in equation (1) is equivalent to the one of [37] and that the corresponding nonexpansivity assumption is satisfied. Importantly, this is the crucial observation that allows us to apply our results to zero-sum stochastic games in Section 5.

7.2.1 Application: Social learning and time varying stubbornness

Here, we apply the results of this section to the model of social learning of Section 6 by allowing for time-varying vanishing stubbornness of the agents. Formally, consider n agents $N = \{1, ..., n\}$ aggregating their opinions over time. Initial opinions are represented by a vector $x^0 \in \mathbb{R}^n$ of real numbers so that x_i^0 corresponds to agent i's initial opinion. At every period $t \in \mathbb{N}$, the updated vector of agent's opinions is given by

$$x^t = (1 - \beta_t) x^0 + \beta_t T \left(x^{t-1}\right).$$

Here, $x^{t-1} \in \mathbb{R}^n$ is the last-period vector of opinions, $T : \mathbb{R}^n \to \mathbb{R}^n$ is an opinion aggregator, and $\{\beta_t\}_{t\in\mathbb{N}} \subseteq (0, 1)$ is a sequence of *stubbornness weights* such that $\beta_t \to 1$. At each period t, each agent $i \in N$ first combines the last-period opinions of the group through an individual aggregator $T_i : \mathbb{R}^n \to \mathbb{R}$ and then linearly combines the aggregate opinion $T_i\left(x^{t-1}\right)$ with her original stance x_i^0 with weight β_t. In particular, the common level of stubbornness in the group is vanishing as $t \to \infty$.

We assume that the opinion aggregator T is normalized, monotone, translation invariant, and such that $\underline{A}(T)$ is regular. Normalization and monotonicity of T capture the idea that the agents trust each other opinions and try to coordinate. Moreover, by assuming that T is translation invariant we obtain enough continuity to rule out expansive dynamics. Finally, following [12], we interpret $\underline{A}(T)$ as a network of strong links among the agents. With this, the regularity assumption on $\underline{A}(T)$ amounts to assume that there exists a unique strongly connected and closed group in the network of strong links induced by T.

We start with an irrelevance result of the nonlinearity of T under differentiability and the assumption that the stubbornness is vanishing at a sufficient slow rate, i.e., $(1 - \beta_t)$ is asymptotically equivalent to $1/t^\alpha$ for some $\alpha \in (0, 1)$.

Corollary 4 *If T is continuously differentiable in a neighborhood of 0 and $\lim_t (1 - \beta_t) t^\alpha = 1$ for some $\alpha \in (0, 1)$, then*

$$\lim_t x^t = \left\langle \gamma, x^0 \right\rangle e \qquad \forall x^0 \in \mathbb{R}^k$$

where γ is the unique left Perron-Frobenius eigenvector of the Jacobian of T at 0.

This result easily follows by combining Corollary 1 and Proposition 7. It implies that, under vanishing stubbornness and smoothness, for a large class of opinion aggregation models the long-run outcomes are indistinguishable from the ones of the DeGroot's model (cf. Golub and Jackson)

Next, we show that, without differentiability, the nonlinearity of the aggregator still plays a role for the long-run consensus. Toward this result, we assume that T is star-shaped. By Proposition 10 in the Appendix, there exists a family of normalized, monotone, translation invariant, and concave operators $\{S_\alpha\}_{\alpha\in\mathcal{A}}$ such that $\underline{A}(S_\alpha)$ is regular for all $\alpha \in \mathcal{A}$ and

$$T(x) = \sup_{\alpha\in\mathcal{A}} S_\alpha(x)$$

for all $x \in \mathbb{R}^k$. In particular, we have $\bigcup_{\alpha\in\mathcal{A}} \partial_C S_\alpha(0) \subseteq \mathcal{W}$ and we can interpret each $W \in \bigcup_{\alpha\in\mathcal{A}} \partial_C S_\alpha(0)$ as a weighted undirected network among the agents. Moreover, by construction, each of these networks admits a unique eigenvector centrality $\gamma \in \Delta$ capturing the corresponding long-run influences of the agents. With this, define

$$\Gamma_\alpha = \{\gamma \in \Delta : \exists W \in \partial_C S_\alpha(0), \gamma = \gamma W\} \qquad \forall \alpha \in \mathcal{A}.$$

The next corollary provides a complete characterization of the consensus opinion in terms of all the eigenvector centralities in $\{\Gamma_\alpha\}_{\alpha\in\mathcal{A}}$.

Corollary 5 *If T is star-shaped and $\lim_t(1-\beta_t)t^\alpha = 1$ for some $\alpha \in (0,1)$, then*

$$\lim_t x^t = \left(\sup_{\alpha\in\mathcal{A}} \min_{\gamma\in\Gamma_\alpha} \langle\gamma, x^0\rangle\right) e.$$

This result shows that, in the limit consensus, the nonlinearity of each S_α is greatly simplified to a pessimistic aggregation with respect to all the eigenvector centralities of S_α. In contrast, these pessimistic consensus are aggregated over α in an optimistic fashion, i.e., by taking the maximum over $\alpha \in \mathcal{A}$. In addition, the limit consensus operator is positive homogeneous with respect to the initial opinions. Therefore, for each x^0, there exist $\alpha(x^0) \in \mathcal{A}$ and $\gamma(x^0) \in \Gamma_{\alpha(x^0)}$ such that

$$\lim_t x^t = \langle\gamma(x^0), x^0\rangle,$$

so that we can interpret $\gamma(x^0)$ as a local centrality measure at x^0.

We conclude this section by highlighting that this result has relevant implications for targeting problems in networks. Assume for simplicity that T is concave, that is, $\mathcal{A} = \{\alpha\}$, and that the initial opinions of the agents are binary, that is, $x^0 \in \{0,1\}^n$. Consider a designer optimally choosing $m < n$ agents to endow with the optimistic opinion $x_i^0 = 1$, whereas the rest of the agents j start with the pessimistic opinion $x_j^0 = 0$. The objective of the designer is to obtain the most optimistic long-run consensus possible under vanishing stubbornness. This implies that she needs to take into account the centrality of the agents *given* the seeded initial opinions. In particular, given Corollary 5, the optimal targeting set solves the maxmin problem

$$\max_{M:|M|=m} \inf_{\gamma\in\Gamma_\alpha} \sum_{i\in M} \gamma_i.$$

Therefore, the identity of the first agents targeted can change by changing the number of seeds m due to submodularity, as opposed to the greedy algorithm that would solve the case for linear T.

7.3 Computing the fixed point

Consider a nonexpansive operator $T:\mathbb{R}^k \to \mathbb{R}^k$. Given $\beta \in (0,1)$ and $x \in \mathbb{R}^k$, define $T_{\beta,x}, \tilde{T}_{\beta,x} : \mathbb{R}^k \to \mathbb{R}^k$ by

$$T_{\beta,x}(y) = T((1-\beta)x + \beta y) \text{ and } \tilde{T}_{\beta,x}(y) = (1-\beta)x + \beta T(y) \quad \forall y \in \mathbb{R}^k.$$

Lemma 3 *Let T be nonexpansive. If $\beta \in (0,1)$ and $x \in \mathbb{R}^k$, then $T_{\beta,x}$ and $\tilde{T}_{\beta,x}$ are β-contractions. In particular, for each $\beta \in (0,1)$ and for each $x \in \mathbb{R}^k$, there exist unique $x_\beta, \tilde{x}_\beta \in \mathbb{R}^k$ such that*

$$T^t_{\beta,x}(y) \to x_\beta \quad \forall y \in \mathbb{R}^k,\ T_{\beta,x}(x_\beta) = x_\beta$$

and

$$\tilde{T}^t_{\beta,x}(y) \to \tilde{x}_\beta \quad \forall y \in \mathbb{R}^k,\ \tilde{T}_{\beta,x}(\tilde{x}_\beta) = \tilde{x}_\beta.$$

Proof. Fix $\beta \in (0,1)$ and $x \in \mathbb{R}^k$. We prove that $T_{\beta,x}$ is a β-contraction. A similar argument holds for $\tilde{T}_{\beta,x}$. Since T is nonexpansive, we have that for each $y, z \in \mathbb{R}^k$

$$\begin{aligned}\|T_{\beta,x}(y) - T_{\beta,x}(z)\|_\infty &= \|T((1-\beta)x + \beta y) - T((1-\beta)x + \beta z)\|_\infty \\ &\le \|(1-\beta)x + \beta y - (1-\beta)x - \beta z\|_\infty = \beta\|y - z\|_\infty,\end{aligned}$$

proving that $T_{\beta,x}$ is a β-contraction. By the Banach contraction principle, for each $y \in \mathbb{R}^k$ we have that $T^t_{\beta,x}(y) \to x_\beta$ as well as $T_{\beta,x}(x_\beta) = x_\beta$ where x_β is the unique fixed point of $T_{\beta,x}$. ■

Consider two nonexpansive operators $S, T : \mathbb{R}^k \to \mathbb{R}^k$.If for each $\beta \in (0,1)$ and for each $x \in \mathbb{R}^k$ we define $x_{\beta,S}$ and $x_{\beta,T}$ to be such that

$$x_{\beta,S} = S((1-\beta)x + \beta x_{\beta,S}) \text{ and } x_{\beta,T} = T((1-\beta)x + \beta x_{\beta,T}),$$

then we have the following simple monotonicity result.

Lemma 4 *Let S and T be nonexpansive. If S is monotone and $S \ge T$, then*

$$x_{\beta,S} \ge x_{\beta,T} \qquad \forall \beta \in (0,1), \forall x \in \mathbb{R}^k.$$

Proof. Fix $\beta \in (0,1)$ and $x \in \mathbb{R}^k$. We prove by induction that $S^t_{\beta,x}(x) \ge T^t_{\beta,x}(x)$ for all $t \in \mathbb{N}$. For $t = 1$, note that $S^1_{\beta,x}(x) = S((1-\beta)x + \beta x) = S(x) \ge T(x) = T((1-\beta)x + \beta x) = T^1_{\beta,x}(x)$. Next, we assume the statement is true for t and we prove it holds for $t+1$. Since S is monotone and $S \ge T$, we have that

$$\begin{aligned}S^{t+1}_{\beta,x}(x) &= S_{\beta,x}\left(S^t_{\beta,x}(x)\right) = S\left((1-\beta)x + \beta S^t_{\beta,x}(x)\right) \ge S\left((1-\beta)x + \beta T^t_{\beta,x}(x)\right) \\ &\ge T\left((1-\beta)x + \beta T^t_{\beta,x}(x)\right) = T_{\beta,x}\left(T^t_{\beta,x}(x)\right) = T^{t+1}_{\beta,x}(x),\end{aligned}$$

proving the statement. By Lemma 3 and passing to the limit, $x_{\beta,S} = \lim_t S^t_{\beta,x}(x) \ge \lim_t T^t_{\beta,x}(x) = x_{\beta,T}$. ■

Proposition 9 *Let T be nonexpansive. If there exists a collection $\{S_\alpha\}_{\alpha\in\mathcal{A}}$ such that S_α is monotone and $S_\alpha \geq T$ for all $\alpha \in \mathcal{A}$ and for each $x \in \mathbb{R}^k$ there exists $\alpha_x \in \mathcal{A}$ such that $S_{\alpha_x}(x) = T(x)$, then*

$$x_{\beta,T} = \min_{\alpha\in\mathcal{A}} x_{\beta,S_\alpha} \qquad \forall \beta \in (0,1), \forall x \in \mathbb{R}^k.$$

Proof. Fix $\beta \in (0,1)$ and $x \in \mathbb{R}^k$. By Lemma 4, $x_{\beta,T} \leq x_{\beta,S_\alpha}$ for all $\alpha \in \mathcal{A}$, proving that $x_{\beta,T} \leq \inf_{\alpha\in\mathcal{A}} x_{\beta,S_\alpha}$. Since for each $x \in \mathbb{R}^k$ there exists $\alpha_x \in \mathcal{A}$ such that $S_{\alpha_x}(x) = T(x)$, we have that there exists $\bar{\alpha} \in \mathcal{A}$ such that $S_{\bar{\alpha}}((1-\beta)x + \beta x_{\beta,T}) = T((1-\beta)x + \beta x_{\beta,T}) = x_{\beta,T}$, proving that $\inf_{\alpha\in\mathcal{A}} x_{\beta,S_\alpha} \geq x_{\beta,T} = x_{\beta,S_{\bar{\alpha}}} \geq \inf_{\alpha\in\mathcal{A}} x_{\beta,S_\alpha}$, proving the statement. ■

The leading example of such a collection $\{S_\alpha\}_{\alpha\in\mathcal{A}}$ is when each S_α is such that $S_\alpha(x) = W_\alpha x + h_\alpha$ for all $x \in \mathbb{R}^k$ where $W_\alpha \in \mathcal{W}$ and $h_\alpha \in \mathbb{R}^k$. Recall that $x_{\beta,W_\alpha} = (1-\beta)\sum_{t=0}^{\infty} \beta^t W_\alpha^{t+1} x$ and $W_\alpha((1-\beta)x + \beta x_{\beta,W_\alpha}) = x_{\beta,W_\alpha}$. Define $\hat{x}_{\beta,W_\alpha} = x_{\beta,W_\alpha} + \sum_{t=0}^{\infty} \beta^t W_\alpha^t h_\alpha$. We next show that $\hat{x}_{\beta,W_\alpha} = x_{\beta,S_\alpha}$. Note that

$$\begin{aligned}
S_\alpha((1-\beta)x + \beta\hat{x}_{\beta,W_\alpha}) &= W_\alpha((1-\beta)x + \beta\hat{x}_{\beta,W_\alpha}) + h_\alpha \\
&= W_\alpha\left((1-\beta)x + \beta x_{\beta,W_\alpha} + \sum_{t=0}^{\infty}\beta^{t+1}W_\alpha^t h_\alpha\right) + h_\alpha \\
&= W_\alpha((1-\beta)x + \beta x_{\beta,W_\alpha}) + \sum_{t=0}^{\infty}\beta^{t+1}W_\alpha^{t+1}h_\alpha + h_\alpha \\
&= x_{\beta,W_\alpha} + \sum_{t=0}^{\infty}\beta^t W_\alpha^t h_\alpha = \hat{x}_{\beta,W_\alpha},
\end{aligned}$$

proving that $\hat{x}_{\beta,W_\alpha} = x_{\beta,S_\alpha}$.

A Appendix: A representation result

In this appendix, we consider a functional $g : \mathbb{R}^k \to \mathbb{R}$ which is normalized, monotone, translation invariant, and star-shaped.[18] The objective is to prove that such a functional can be rewritten as the max of a collection $\{g_\alpha\}_{\alpha\in\mathcal{A}}$ of normalized, monotone, translation invariant, and concave functionals. Results of this form have appeared in Decision Theory (see, e.g., Chandrasekher, Frick, Iijima, and Le Yaouanq [14]), Mathematical Finance (see, e.g., Castagnoli, Cattelan, Maccheroni, Tebaldi, and Wang [11, Theorem 2]), and Mathematics (see, e.g., Rubinov and Dzalilov [31]). The version we need for this paper is slightly different from what is available in the literature and it is

[18] With a small abuse of terminology, we use the same name for similar properties that pertain to functionals and operators.

a refinement of [11], whose techniques we also exploit. Compared to their Theorem 5, we obtain a version in which $\{g_\alpha\}_{\alpha\in\mathcal{A}}$ "inherits the derivatives" of g: a property which we badly need for our convergence results. Thus, also to keep the paper self-contained, we prove our version here.

Proposition 10 *Let $g:\mathbb{R}^k\to\mathbb{R}$. The following statements are equivalent:*

(i) The functional g is normalized, monotone, translation invariant, and star-shaped;

(ii) There exists a family $\{g_\alpha\}_{\alpha\in\mathcal{A}}$ of normalized, monotone, translation invariant, and concave functionals such that

$$g(x)=\max_{\alpha\in\mathcal{A}} g_\alpha(x) \qquad \forall x\in\mathbb{R}^k. \tag{26}$$

Moreover, $\{g_\alpha\}_{\alpha\in\mathcal{A}}$ can be chosen to be such that $\bar{\mathrm{co}}\left(\partial_C g_\alpha\left(\mathbb{R}^k\right)\right)\subseteq\bar{\mathrm{co}}\left(\partial_C g\left(\mathbb{R}^k\right)\right)$ for all $\alpha\in\mathcal{A}$.

Before proving the statement, we need to introduce an ancillary object. Given $g:\mathbb{R}^k\to\mathbb{R}$, define the binary relation $\succsim_g^*$ by

$$x\succsim_g^* y \overset{def}{\iff} g(\lambda x+(1-\lambda)z)\geq g(\lambda y+(1-\lambda)z) \qquad \forall\lambda\in(0,1],\forall z\in\mathbb{R}^k.$$

It is immediate to see that $x\succsim_g^* y$ implies that $g(x)\geq g(y)$. By Cerreia-Vioglio, Ghirardato, Maccheroni, Marinacci, and Siniscalchi [13] and if g is normalized, monotone, and continuous, we have that there exists a closed convex set $C_g\subseteq\Delta$ such that

$$x\succsim_g^* y \iff \langle\gamma,x\rangle\geq\langle\gamma,y\rangle \qquad \forall\gamma\in C_g. \tag{27}$$

Moreover, if $\succsim^\circ$ is another conic binary relation such that

$$x\succsim^\circ y \implies g(x)\geq g(y),$$

then $\succsim^\circ$ is a subrelation of $\succsim_g^*$, that is, $x\succsim^\circ y$ implies $x\succsim_g^* y$.[19] Recall that if g is normalized, monotone, and translation invariant $\partial_C g(x)\subseteq\Delta$ for all $x\in\mathbb{R}^k$. By Ghirardato and Siniscalchi [20, Theorem 2], in this case, we have that $C_g=\bar{\mathrm{co}}\left(\partial_C g\left(\mathbb{R}^k\right)\right)$ where $\partial_C g\left(\mathbb{R}^k\right)=\cup_{x\in\mathbb{R}^k}\partial_C g(x)$.

Proof. (i) implies (ii). Define $D=\left\{x\in\mathbb{R}^k : x\succsim_g^* 0\right\}$. It is immediate to see that D is a nonempty, closed, and convex cone. Define $\mathcal{A}=\left\{z\in\mathbb{R}^k\backslash\{0\} : g(z)=0\right\}$. For each $z\in\mathcal{A}$ define $U_z=\mathrm{co}(\{0,z\})+D$.[20] We say that a functional $\tilde{g}:\mathbb{R}^k\to\mathbb{R}$ is $\succsim_g^*$-monotone if and only if $x\succsim_g^* y$ implies $\tilde{g}(x)\geq\tilde{g}(y)$. Since $x\geq y$ implies $x\succsim_g^* y$, we have that $\succsim_g^*$-monotonicity yields standard monotonicity.

Step 1. For each $z\in\mathcal{A}$ the set U_z is a nonempty, convex, and closed set such that

[19] The binary relation $\succsim^\circ$ is conic if and only if there exists a subset $\tilde{C}\subseteq\Delta$ such that $x\succsim^\circ y$ if and only if $\langle\gamma,x\rangle\geq\langle\gamma,y\rangle$ for all $\gamma\in\tilde{C}$.

[20] The construction of [11] differs from ours in that the cone added to $\mathrm{co}(\{0,z\})$ is $\mathbb{R}_+^k$.

1. $0, z \in U_z$;
2. *if* $x \in U_z$, *then* $g(x) \geq 0$*;*
3. *if* $y \succsim_g^* x \in U_z$, *then* $y \in U_z$*;*
4. *if* $h > 0$, *then* $-he \notin U_z$.

Proof of the Step. Since $0, z \in \text{co}(\{0, z\})$ and $0 \in D$ and $\text{co}(\{0, z\})$ is convex and compact and D is convex and closed, we have that $0, z \in U_z = \text{co}(\{0, z\}) + D$ is nonempty, convex, and closed, and, in particular, point 1 holds. If $x \in U_z$, then there exist $\lambda \in [0, 1]$ and $y \in D$ such that $x = \lambda z + (1 - \lambda) 0 + y$. Since g is star-shaped and $g(z) = 0$, we have that $g(\lambda z + (1 - \lambda) 0) = g(\lambda z) \geq \lambda g(z) = 0$. Since $y \in D$, we have that $x = \lambda z + (1 - \lambda) 0 + y \succsim_g^* \lambda z + (1 - \lambda) 0$, yielding that $g(x) \geq g(\lambda z + (1 - \lambda) 0) \geq 0$, proving point 2. Next, consider $x, y \in \mathbb{R}^k$ such that $y \succsim_g^* x \in U_z$, that is, $y - x \succsim_g^* 0$ and $x \in U_z$. Since $x \in U_z$, then there exist $\lambda \in [0, 1]$ and $\hat{y} \in D$ such that $x = \lambda z + \hat{y}$. Since D is a convex cone, it follows that $y = x + (y - x) = \lambda z + (\hat{y} + y - x) \in \text{co}(\{0, z\}) + D = U_z$. Finally, by contradiction, assume that $h < 0$ and $-he \in U_z$. By point 2 and since g is normalized, $0 > -h = g(-he) \geq 0$, a contradiction. □

Step 2. For each $z \in \mathcal{A}$ the functional $g_z : \mathbb{R}^k \to \mathbb{R}$, defined by

$$g_z(x) = \max\{h \in \mathbb{R} : x - he \in U_z\} \qquad \forall x \in \mathbb{R}^k,$$

is well defined, normalized, $\succsim_g^*$-monotone, translation invariant, concave, and such that $g_z(z) = 0$ as well as $g_z(z') \leq 0$ for all $z' \in \mathcal{A}$.

Proof of the Step. Fix $z \in \mathcal{A}$. Consider $x \in \mathbb{R}^k$. Define $I_x = \{h \in \mathbb{R} : x - he \in U_z\}$. Since U_z is convex and closed, I_x is a closed interval. Next we show that I_x is bounded from above. Let $h \geq \|x\|_\infty + \|z\|_\infty$. Since $z \neq 0$ and g is normalized and monotone, note that $g(x - \|x\|_\infty e - \|z\|_\infty e) \leq g(-\|z\|_\infty e) = -\|z\|_\infty < 0$. By point 2 above, we have that $x - \|x\|_\infty e - \|z\|_\infty e \notin U_z$. By point 3 and since $x - (\|x\|_\infty + \|z\|_\infty) e \geq x - he$, we can conclude that $x - he \notin U_z$, proving that I_x is bounded from above. Since $C_g \subseteq \Delta$ is compact, consider $h \in \mathbb{R}$ such that $-h \geq \max_{\gamma \in C}\{\langle \gamma, z \rangle - \langle \gamma, x \rangle\} \in \mathbb{R}$. By points 1 and 3 and the characterization of $\succsim_g^*$, it follows that $x - he \succsim_g^* z \in U_z$, proving that $x - he \in U_z$ and I_x is nonempty. Since I_x is a nonempty, closed, and bounded from above interval, we have that $\sup I_x$ is well defined and attained, proving that g_z is well defined. In particular, $x - g_z(x) e \in U_z$ for all $x \in \mathbb{R}^k$.

Consider $z' \in \mathcal{A}$. By point 2 and since $z' \in \mathcal{A}$ and $z' - g_z(z') e \in U_z$, we have that $g(z') - g_z(z') = g(z' - g_z(z') e) \geq 0$, that is, $0 = g(z') \geq g_z(z')$. By point 1, if $z' = z$, then $0 \in I_z$ and $g_z(z) \geq 0$. Since z' was arbitrarily chosen, we can conclude that $g_z(z') \leq 0$ for all $z' \in \mathcal{A}$ and $g_z(z) = 0$. Consider $x, y \in \mathbb{R}^k$ such that $x \succsim_g^* y$. By point 3, (27), and the definition of g_z, we have that $x - g_z(y) e \succsim_g^* y - g_z(y) e \in U_z$, yielding that $g_z(y) \in I_x$ and $g_z(x) \geq g_z(y)$, that is, g_z is $\succsim_g^*$-monotone. Consider $x \in \mathbb{R}^k$ and $h \in \mathbb{R}$. By definition of g_z, we can conclude that

$$(x + he) - (g_z(x) + h)e = x - g_z(x)e \in U_z.$$

This implies that $g_z(x) + h \in I_{x+he}$ and, in particular, $g_z(x + he) \geq g_z(x) + h$. Since x and h were arbitrarily chosen, we have that

$$g_z(x + he) \geq g_z(x) + h \qquad \forall x \in \mathbb{R}^k, \forall h \in \mathbb{R}.$$

This yields that $g_z(x + he) = g_z(x) + h$ for all $x \in \mathbb{R}^k$ and for all $h \in \mathbb{R}$. Finally, consider $x, y \in \mathbb{R}^k$ and $\lambda \in (0, 1)$. By definition of g_z, we have that $x - g_z(x)e, y - g_z(y)e \in U_z$. Since U_z is convex, this implies that $\lambda x + (1-\lambda)y - (\lambda g_z(x) + (1-\lambda)g_z(y))e \in U_z$, yielding that $g_z(\lambda x + (1-\lambda)y) \geq \lambda g_z(x) + (1-\lambda)g_z(y)$ and proving that g_z is concave.

To sum up, g_z is well defined, $\succsim_g^*$-monotone, translation invariant, and concave. Consider $x = 0$. By point 1, we have that $0 \in U_z$, yielding that $0 \in I_0$ and, in particular, $g_z(0) \geq 0$. By point 4, we have that $I_0 \subseteq (-\infty, 0]$, proving that $g_z(0) \leq 0$, that is, $g_z(0) = 0$. Since g_z is translation invariant and $g_z(0) = 0$, it follows that $g_z(he) = g_z(0 + he) = g_z(0) + h = h$ for all $h \in \mathbb{R}$, that is, g_z is normalized, proving the step. $\square$

We can prove the implication. Consider the family of functionals $\{g_z\}_{z\in\mathcal{A}}$ of Step 2. Each g_z is normalized, $\succsim_g^*$-monotone (in particular, monotone), translation invariant, and concave. Consider $x \in \mathbb{R}^k$. Since g and g_z are normalized for all $z \in \mathcal{A}$, if $x = he$ for some $h \in \mathbb{R}$, we have that $g(x) = h = g_z(x)$ for all $z \in \mathcal{A}$, that is, $g(x) = \max_{z\in\mathcal{A}} g_z(x)$. If x is not a constant vector, define $\bar{z} = x - g(x)e$. Note that $\bar{z} \neq 0$. Since g is translation invariant, we have that $\bar{z} \in \mathcal{A}$. By Step 2, we have that $g_z(\bar{z}) \leq 0 = g_{\bar{z}}(\bar{z}) = 0 = g(\bar{z})$ for all $z \in \mathcal{A}$. Since each g_z is translation invariant, we have that

$$\begin{aligned} g(x) - g(x) &= g(x - g(x)e) = g(\bar{z}) = \max_{z\in\mathcal{A}} g_z(\bar{z}) = \max_{z\in\mathcal{A}} g_z(x - g(x)e) \\ &= \max_{z\in\mathcal{A}} \{g_z(x) - g(x)\} = \max_{z\in\mathcal{A}} g_z(x) - g(x), \end{aligned}$$

proving (26).

(ii) implies (i). It is trivial.

Consider $\{g_z\}_{z\in\mathcal{A}}$ as in the proof of (i) implies (ii). Fix $z \in \mathcal{A}$. By Step 2, we have that g_z is $\succsim_g^*$-monotone. This implies that $x \succsim_g^* y$ implies $x \succsim_{g_z}^* y$. By the Hahn-Banach Theorem, this yields that $\bar{\text{co}}\left(\partial_C g\left(\mathbb{R}^k\right)\right) = C_g \supseteq C_{g_z} = \bar{\text{co}}\left(\partial_C g_z\left(\mathbb{R}^k\right)\right)$, proving the last part of the statement. $\blacksquare$

B Appendix: Proofs of Section 3

In this appendix, we prove all the results and few ancillary lemmas which pertain Section 3. With the exception of Theorem 1, whose proof comes at the end, all the other proofs follow the order in the main text and are divided accordingly to the sections of the main text. Theorem 1 is proved last as a consequence of all the other results.

B.1 Preliminaries

We begin by reporting a few ancillary facts. We start with a result that generalizes to the nonlinear case a well-known fact for stochastic matrices: having a regular adjacency matrix yields that the only fixed points are the constant vectors. Since the property $E(T) = D$ is often used in our results, Proposition 11 provides a condition in terms of the derivatives of T, which guarantees it is satisfied.

Proposition 11 *Let T be normalized, monotone, and translation invariant. If $\underline{A}(T)$ is regular, then $E(T) = D$.*

Proof. Consider $\lambda \in (0,1)$. Given T, define $T_\lambda : \mathbb{R}^k \to \mathbb{R}^k$ by $T_\lambda(x) = \lambda x + (1-\lambda)T(x)$ for all $x \in \mathbb{R}^k$. It is immediate to check that $E(T) = E(T_\lambda)$ and $\underline{A}(T_\lambda) \geq \underline{A}(T), I$ where I is the identity matrix. By [12, Proposition 10] and since $E(T) = E(T_\lambda)$, we have that $E(T) = E(T_\lambda) = D$, proving the statement. ■

The next result shows that, given $z \in \mathbb{R}^k$, the value of $T(x)$ can be calculated alternatively by computing Wz where W is a "replicating" stochastic matrix whose i-th row belongs to the Clarke differential of T_i.

Proposition 12 *Let T be normalized, monotone, and translation invariant. For each $z \in \mathbb{R}^k$ and for each $\hat{h} \in \mathbb{R}$, there exists a stochastic matrix $W_{z,\hat{h}}$ such that $T(z) = W_{z,\hat{h}}z$ where $w^i_{z,\hat{h}} \in \partial_C T_i\left(\lambda_{i,z,\hat{h}}z + \left(1-\lambda_{i,z,\hat{h}}\right)\hat{h}e\right)$ and $\lambda_{i,z,\hat{h}} \in [0,1]$ for all $i \in \{1,...,k\}$.*

Proof. Consider $z \in \mathbb{R}^k$ and $\hat{h} \in \mathbb{R}$. By Lebourg's Mean Value Theorem (see, e.g., Clarke [16, Theorem 2.3.7]) and since T is normalized and Lipschitz continuous, for each $i \in \{1,...,k\}$ there exists $w^i \in \partial_C T_i\left(\lambda_{i,z,\hat{h}}z + \left(1-\lambda_{i,z,\hat{h}}\right)\hat{h}e\right)$ where $\lambda_{i,z,\hat{h}} \in [0,1]$ such that

$$T_i(z) - \hat{h} = T_i(z) - T_i\left(\hat{h}e\right) = \left\langle w^i, z - \hat{h}e\right\rangle = \left\langle w^i, z\right\rangle - \hat{h}.$$

If we define $W_{z,\hat{h}}$ to be such that its i-th row is w^i, then the statement follows. ■

The next and last preliminary result guarantee that the adjacency matrices of the replicating matrices and of the generalized Jacobians of T inherit the property of regularity of $\underline{A}(T)$. Moreover, it provides a quantitative lower bound for the entries of the replicating matrices. The first property will be exploited in proving that $\{x_\beta\}_{\beta\in(0,1)}$ and $\{\tilde{x}_\beta\}_{\beta\in(0,1)}$ converge while the latter will be useful in the proofs that elaborate on the rate of such convergence. Given a stochastic matrix W, we define the adjacency matrix $A(W)$ to be such that $a_{ij}=1$ if and only if $w_{ij}>0$ and $a_{ij}=0$ otherwise.

Proposition 13 *Let T be normalized, monotone, and translation invariant. The following statements are true:*

1. *If $\underline{A}(T)$ is regular, then $A\left(W_{z,\hat{h}}\right)$ is regular for all $z\in\mathbb{R}^k$ and for all $\hat{h}\in\mathbb{R}$. Moreover, we have that*

$$\min_{i,j:\underline{a}_{ij}=1} w_{ij} \geq \min_{i,j:\underline{a}_{ij}=1} \inf_{x\in\mathcal{D}} \frac{\partial T_i}{\partial x_j}(x) \tag{28}$$

where w_{ij} is the ij-th entry of $W_{z,\hat{h}}$.

2. *If $\underline{A}(T)$ is regular, then $A(W)$ is regular for all $W\in\partial_C T(0)$. Moreover, we have that*

$$\min_{i,j:\underline{a}_{ij}=1} w_{ij} \geq \min_{i,j:\underline{a}_{ij}=1} \inf_{x\in\mathcal{D}} \frac{\partial T_i}{\partial x_j}(x).$$

As a consequence, if $\underline{A}(T)$ is regular, we have that $W_{z,\hat{h}}$ has a unique left Perron-Frobenius eigenvector for all $z\in\mathbb{R}^k$ and for all $\hat{h}\in\mathbb{R}$. We denote it by $\gamma_{z,\hat{h}}$.

Proof. 1. By (6) and (7), we have that $\gamma_j\geq\varepsilon_{ij}>0$ for all $\gamma\in\partial_C T_i(z)$, for all $i,j\in\{1,...,k\}$ such that $\underline{a}_{ij}=1$, and for all $z\in\mathbb{R}^k$. By definition of $W_{z,\hat{h}}$ and w_{ij}, this implies that $w_{ij}\geq\varepsilon_{ij}>0$ for all $i,j\in\{1,...,k\}$ such that $\underline{a}_{ij}=1$ and $A\left(W_{z,\hat{h}}\right)\geq\underline{A}(T)$ for all $z\in\mathbb{R}^k$ and for all $\hat{h}\in\mathbb{R}$. Since $\underline{A}(T)$ is regular, we can conclude that $A\left(W_{z,\hat{h}}\right)$ is regular for all $z\in\mathbb{R}^k$ and for all $\hat{h}\in\mathbb{R}$. By (6) and since $w_{ij}\geq\varepsilon_{ij}$ for all $i,j\in\{1,...,k\}$ such that $\underline{a}_{ij}=1$, we have that (28) follows.

2. By (6) and (7), we have that $\gamma_j\geq\varepsilon_{ij}>0$ for all $\gamma\in\partial_C T_i(0)$, for all $i,j\in\{1,...,k\}$ such that $\underline{a}_{ij}=1$, and for all $z\in\mathbb{R}^k$. By definition of $\partial_C T(0)$, this implies that $A(W)\geq\underline{A}(T)$ for all $W\in\partial_C T(0)$. Since $\underline{A}(T)$ is regular, we can conclude that $A(W)$ is regular. Similarly to before, we can conclude that $\min_{i,j:\underline{a}_{ij}=1} w_{ij}\geq\min_{i,j:\underline{a}_{ij}=1}\inf_{x\in\mathcal{D}}\frac{\partial T_i}{\partial x_j}(x)$. ■

B.2 Convergence

B.2.1 Concavity and differentiability

Proof of Lemma 1. Fix $\beta\in(0,1)$ and $x\in\mathbb{R}^k$. We prove that $T_{\beta,x}$ is a β-contraction. A similar argument holds for $\tilde{T}_{\beta,x}$ (see, e.g., [22, Theorem 11.3]). Since T is Lipschitz continuous of order 1, we have that for each $y, z \in \mathbb{R}^k$

$$\begin{aligned}\|T_{\beta,x}(y) - T_{\beta,x}(z)\|_\infty &= \|T((1-\beta)x + \beta y) - T((1-\beta)x + \beta z)\|_\infty \\ &\leq \|(1-\beta)x + \beta y - (1-\beta)x - \beta z\|_\infty = \beta\|y - z\|_\infty,\end{aligned}$$

proving that $T_{\beta,x}$ is a β-contraction. By the Banach contraction principle, for each $y \in \mathbb{R}^k$ we have that $T^t_{\beta,x}(y) \to x_\beta$ as well as $T_{\beta,x}(x_\beta) = x_\beta$ where x_β is the unique fixed point of $T_{\beta,x}$. Finally, since T is normalized and Lipschitz continuous of order 1, observe that

$$\begin{aligned}\|T_{\beta,x}(y)\|_\infty &= \|T((1-\beta)x + \beta y) - T(0)\|_\infty \leq \|(1-\beta)x + \beta y\|_\infty \\ &\leq (1-\beta)\|x\|_\infty + \beta\|y\|_\infty \quad \forall y \in \mathbb{R}^k.\end{aligned}$$

By induction, this implies that $\left\|T^t_{\beta,x}(x)\right\|_\infty \leq \|x\|_\infty$ for all $t \in \mathbb{N}$. By passing to the limit, (9) follows. By the Banach contraction principle and since $\tilde{T}_{\beta,x}$ is also a β-contraction, for each $y \in \mathbb{R}^k$ we have that $\tilde{T}^t_{\beta,x}(y) \to \tilde{x}_\beta$ as well as $\tilde{T}_{\beta,x}(\tilde{x}_\beta) = \tilde{x}_\beta$ where $\tilde{x}_\beta$ is the unique fixed point of $\tilde{T}_{\beta,x}$. Fix $x \in \mathbb{R}^k$ and $\beta \in (0,1)$. Set $\hat{x}_\beta = (1-\beta)x + \beta x_\beta$. By definition of $\tilde{T}_{\beta,x}$ and $\hat{x}_\beta$ as well as x_β, we have that

$$\tilde{T}_{\beta,x}(\hat{x}_\beta) = (1-\beta)x + \beta T(\hat{x}_\beta) = (1-\beta)x + \beta T((1-\beta)x + \beta x_\beta) = (1-\beta)x + \beta x_\beta = \hat{x}_\beta.$$

Since $\tilde{x}_\beta$ is the unique fixed point of $\tilde{T}_{\beta,x}$ and x and β were arbitrarily chosen, we can conclude that $\tilde{x}_\beta = \hat{x}_\beta = (1-\beta)x + \beta x_\beta$ for all $\beta \in (0,1)$ and for all $x \in \mathbb{R}^k$. By (9), this implies that $\|\tilde{x}_\beta\|_\infty \leq \|x\|_\infty$ for all $\beta \in (0,1)$ and for all $x \in \mathbb{R}^k$, proving the statement. ■

Proof of Lemma 2. Consider $\bar{x} \in L$. By definition of L, there exists $\{x_{\beta_n}\}_{n\in\mathbb{N}} \subseteq \{x_\beta\}_{\beta\in(0,1)}$ such that $\beta_n \to 1$ and $x_{\beta_n} \to \bar{x}$. By definition of x_β and since T is Lipschitz continuous and $\lim_n [(1-\beta_n)x + \beta_n x_{\beta_n}] = \bar{x}$, we have that

$$\bar{x} = \lim_n x_{\beta_n} = \lim_n T((1-\beta_n)x + \beta_n x_{\beta_n}) = T(\bar{x}),$$

proving that $\bar{x} \in E(T)$, that is, $L \subseteq E(T)$. Next, assume that $E(T) = D$. By the previous part of the proof, we have that $L \subseteq E(T) = D$. This implies that there exists a set $H \subseteq \mathbb{R}$ such that $\{he\}_{h\in H} = L$ and, in particular, $\liminf_{\beta\to 1} x_\beta = (\inf H)e$ as well as $\limsup_{\beta\to 1} x_\beta = (\sup H)e$. Since L is closed and bounded, it follows that $(\inf H)e, (\sup H)e \in L \subseteq E(T)$, proving the second part of the statement. ■

Proof of Proposition 1. Consider a sequence $\{x_{\beta_n}\}_{n\in\mathbb{N}} \subseteq \{x_\beta\}_{\beta\in(0,1)}$ such that $\beta_n \to 1$ and $x_{\beta_n} \to \bar{x}$, that is in symbols, $\bar{x}$ is a limit point of $\{x_\beta\}_{\beta\in(0,1)}$ and $\bar{x} \in L$. By Lemma 2 and since $E(T) = D$, we have that $\bar{x} \in L \subseteq E(T)$ and $\bar{x} = \bar{h}e$ for some

$\bar{h} \in \mathbb{R}$. Consider $\tilde{x}_{\beta_n} = (1-\beta_n)\,x + \beta_n x_{\beta_n}$ for all $n \in \mathbb{N}$. For each $n \in \mathbb{N}$ consider also $W_{\tilde{x}_{\beta_n},\bar{h}}$ as in Proposition 12. We have that for each $n \in \mathbb{N}$

$$W_{\tilde{x}_{\beta_n},\bar{h}}\left((1-\beta_n)\,x + \beta_n x_{\beta_n}\right) = W_{\tilde{x}_{\beta_n},\bar{h}}\tilde{x}_{\beta_n} = T\left(\tilde{x}_{\beta_n}\right) = T\left((1-\beta_n)\,x + \beta_n x_{\beta_n}\right) = x_{\beta_n}.$$

By (3) and Example 1, we have that $x_{\beta_n} = x_{\beta_n, W_{\tilde{x}_{\beta_n},\bar{h}}}$ for all $n \in \mathbb{N}$. For each $n \in \mathbb{N}$ consider $\gamma_n \in \Gamma\left(W_{\tilde{x}_{\beta_n},\bar{h}}\right)$. By definition of $\Gamma\left(W_{\tilde{x}_{\beta_n},\bar{h}}\right)$ and (4), we have that

$$\langle \gamma_n, x_{\beta_n}\rangle = \left\langle \gamma_n, x_{\beta_n, W_{\tilde{x}_{\beta_n},\bar{h}}}\right\rangle = \langle \gamma_n, x\rangle \qquad \forall n \in \mathbb{N}. \tag{29}$$

Since $\left\{W_{\tilde{x}_{\beta_n},\bar{h}}\right\}_{n\in\mathbb{N}}$ is a sequence of stochastic matrices, it admits a subsequence $\left\{W_{\tilde{x}_{\beta_{n_l}},\bar{h}}\right\}_{l\in\mathbb{N}}$ such that $W_{\tilde{x}_{\beta_{n_l}},\bar{h}} \to W$. Similarly, since $\{\gamma_{n_l}\}_{l\in\mathbb{N}}$ is a sequence of probability vectors, it admits a subsequence $\left\{\gamma_{n_{l(r)}}\right\}_{r\in\mathbb{N}}$ such that $\gamma_{n_{l(r)}} \to \bar{\gamma} \in \Delta$. Since $\gamma_{n_{l(r)}} \in \Gamma\left(W_{\tilde{x}_{\beta_{n_{l(r)}}},\bar{h}}\right)$ for all $r \in \mathbb{N}$, we can conclude that

$$\bar{\gamma}^{\mathrm{T}} W = \lim_r \gamma_{n_{l(r)}}^{\mathrm{T}} W_{\tilde{x}_{\beta_{n_{l(r)}}},\bar{h}} = \lim_r \gamma_{n_{l(r)}}^{\mathrm{T}} = \bar{\gamma}^{\mathrm{T}},$$

that is, $\bar{\gamma} \in \Gamma(W)$. Denote the i-th row of $W_{\tilde{x}_{\beta_{n_{l(r)}}},\bar{h}}$ (resp., W) by $w^i_{\tilde{x}_{\beta_{n_{l(r)}}},\bar{h}}$ (resp., w^i) for all $i \in \{1,...,k\}$. By Proposition 12, we have that

$$w^i_{\tilde{x}_{\beta_{n_{l(r)}}},\bar{h}} \in \partial_C T_i\left(\lambda_{i,\tilde{x}_{\beta_{n_{l(r)}}},\bar{h}}\tilde{x}_{\beta_{n_{l(r)}}} + \left(1-\lambda_{i,\tilde{x}_{\beta_{n_{l(r)}}},\bar{h}}\right)\bar{h}e\right) \qquad \forall i \in \{1,...,k\}, \forall r \in \mathbb{N}$$

and $w^i_{\tilde{x}_{\beta_{n_{l(r)}}},\bar{h}} \to w^i$ for all $i \in \{1,...,k\}$. By Clarke [16, Proposition 2.1.5], the correspondence $z \mapsto \partial_C T_i(z)$ is upper hemicontinuous for all $i \in \{1,...,k\}$. Since $\tilde{x}_{\beta_{n_{l(r)}}} \to \bar{h}e$ and $\left\{\lambda_{i,\tilde{x}_{\beta_{n_{l(r)}}},\bar{h}}\right\}_{r\in\mathbb{N}} \subseteq [0,1]$, we can conclude that $\lambda_{i,\tilde{x}_{\beta_{n_{l(r)}}},\bar{h}}\tilde{x}_{\beta_{n_{l(r)}}} + \left(1-\lambda_{i,\tilde{x}_{\beta_{n_{l(r)}}},\bar{h}}\right)\bar{h}e \to \bar{h}e$ and, in particular, $w^i \in \partial_C T_i\left(\bar{h}e\right) = \partial_C T_i(0)$ for all $i \in \{1,...,k\}$ as well as $W \in \partial_C T(0)$. Since $\bar{\gamma} \in \Gamma(W)$, this implies that $\bar{\gamma} \in \Gamma(W) \subseteq \Gamma(\partial_C T(0))$. By (29) and since $x_{\beta_{n_{l(r)}}} \to \bar{x} = \bar{h}e$, it follows that

$$\bar{h} = \langle \bar{\gamma}, \bar{x}\rangle = \lim_r \left\langle \gamma_{n_{l(r)}}, x_{\beta_{n_{l(r)}}}\right\rangle = \lim_r \left\langle \gamma_{n_{l(r)}}, x\right\rangle = \langle \bar{\gamma}, x\rangle\,,$$

that is,

$$\sup_{\gamma\in\Gamma(\partial_C T(0))} \langle \gamma, x\rangle\, e \geq \langle \bar{\gamma}, x\rangle\, e = \bar{h}e = \bar{x} = \bar{h}e = \langle \bar{\gamma}, x\rangle\, e \geq \inf_{\gamma\in\Gamma(\partial_C T(0))} \langle \gamma, x\rangle\, e$$

Since $\bar{x}$ was arbitrarily chosen, we can conclude that

$$\sup_{\gamma\in\Gamma(\partial_C T(0))} \langle \gamma, x\rangle\, e \geq \bar{x} \geq \inf_{\gamma\in\Gamma(\partial_C T(0))} \langle \gamma, x\rangle\, e \qquad \forall \bar{x} \in L. \tag{30}$$

Since $E(T) = D$, we have that $\liminf_{\beta \to 1} x_\beta, \limsup_{\beta \to 1} x_\beta \in L$. By (30) applied to $\limsup_{\beta \to 1} x_\beta$ and $\liminf_{\beta \to 1} x_\beta$ and since $\limsup_{\beta \to 1} x_\beta \geq \liminf_{\beta \to 1} x_\beta$, we obtain that $\sup_{\gamma \in \Gamma(\partial_C T(0))} \langle \gamma, x \rangle e \geq \limsup_{\beta \to 1} x_\beta \geq \liminf_{\beta \to 1} x_\beta \geq \inf_{\gamma \in \Gamma(\partial_C T(0))} \langle \gamma, x \rangle e$. Since $\partial_C T(0)$ is closed, we have that $\Gamma(\partial_C T(0))$ is compact, yielding that the above sup and inf are achieved and thus proving the statement. ■

Proof of Corollary 1. Since T is continuously differentiable in a neighborhood of 0, so is each T_i. By [16, p. 32 and Proposition 2.2.4], it follows that $\partial_C T_i(0)$ is a singleton for all $i \in \{1, ..., k\}$. This implies that $\partial_C T(0)$ is a singleton and coincides with the Jacobian of T at 0. By Propositions 11 and 13 and since $\underline{A}(T)$ is regular, so is the Jacobian of T at 0 and $E(T) = D$. In particular, $\Gamma(\partial_C T(0))$ is the singleton given by γ: the unique left Perron-Frobenius eigenvector of the Jacobian of T at 0. By Lemma 1 and Proposition 1 and since $E(T) = D$, we can conclude that

$$\langle \gamma, x \rangle e \geq \limsup_{\beta \to 1} x_\beta \geq \liminf_{\beta \to 1} x_\beta \geq \langle \gamma, x \rangle e,$$

proving that $\lim_{\beta \to 1} \tilde{x}_\beta = \lim_{\beta \to 1} x_\beta = \langle \gamma, x \rangle e$. ■

Proof of Proposition 2. Let $x \in \mathbb{R}^k$. Consider $W \in \partial_C T(0)$ and $\bar{\gamma} \in \Gamma(W)$. Define $S : \mathbb{R}^k \to \mathbb{R}^k$ by $S(y) = Wy$ for all $y \in \mathbb{R}^k$. By (8), it follows that $S(y) \geq T(y)$ for all $y \in \mathbb{R}^k$. By induction, we have that $S^t_{\beta,x} \geq T^t_{\beta,x}$ for all $t \in \mathbb{N}$ and for all $\beta \in (0,1)$. By Lemma 1, if we define by $x_{\beta,W}$ the unique fixed point of $S_{\beta,x}$, this implies that $x_{\beta,W} \geq x_\beta$ for all $\beta \in (0,1)$. By definition of $\Gamma(W)$ and $\Gamma(\partial_C T(0))$ and (4) and since $\bar{\gamma} \in \Gamma(W)$ and $W \in \partial_C T(0)$, we also have that $\bar{\gamma} \in \Gamma(W) \subseteq \Gamma(\partial_C T(0))$ and

$$\langle \bar{\gamma}, x \rangle = \langle \bar{\gamma}, x_{\beta,W} \rangle \geq \langle \bar{\gamma}, x_\beta \rangle \geq \min_{\gamma \in \Gamma(W)} \langle \gamma, x_\beta \rangle \geq \min_{\gamma \in \Gamma(\partial_C T(0))} \langle \gamma, x_\beta \rangle \quad \forall \beta \in (0,1).$$

Since W, $\bar{\gamma}$, and x were arbitrarily chosen, we have that

$$\min_{\gamma \in \Gamma(\partial_C T(0))} \langle \gamma, x \rangle \geq \min_{\gamma \in \Gamma(\partial_C T(0))} \langle \gamma, x_\beta \rangle \quad \forall x \in \mathbb{R}^k, \forall \beta \in (0,1),$$

proving (10). Fix $x \in \mathbb{R}^k$ again. Observe that the function $\varphi : \mathbb{R}^k \to \mathbb{R}$, defined by $\varphi(y) = \min_{\gamma \in \Gamma(\partial_C T(0))} \langle \gamma, y \rangle$ for all $y \in \mathbb{R}^k$, is normalized, monotone, and translation invariant. In particular, φ is Lipschitz continuous. By Lemma 2 and since $E(T) = D$, we have that $\limsup_{\beta \to 1} x_\beta$ is a limit point of $\{x_\beta\}_{\beta \in (0,1)}$, that is, there exists a sequence $\{x_{\beta_n}\}_{n \in \mathbb{N}} \subseteq \{x_\beta\}_{\beta \in (0,1)}$ such that $\beta_n \to 1$ and $\lim_n x_{\beta_n} = \limsup_{\beta \to 1} x_\beta$ and $\limsup_{\beta \to 1} x_\beta = \bar{h}e$ for some $\bar{h} \in \mathbb{R}$. By (10) and since φ is continuous, we can conclude that

$$\min_{\gamma \in \Gamma(\partial_C T(0))} \langle \gamma, x \rangle \geq \lim_n \min_{\gamma \in \Gamma(\partial_C T(0))} \langle \gamma, x_{\beta_n} \rangle = \lim_n \varphi(x_{\beta_n}) = \varphi(\bar{h}e) = \bar{h},$$

proving the statement. ■

Proof of Corollary 2. By Lemma 1 and Propositions 1 and 2, we have that $\liminf_{\beta\to 1} x_\beta \geq \min_{\gamma\in\Gamma(\partial_C T(0))} \langle \gamma, x\rangle e \geq \limsup_{\beta\to 1} x_\beta$ and $\lim_{\beta\to 1} \tilde{x}_\beta = \lim_{\beta\to 1} x_\beta = \min_{\gamma\in\Gamma(\partial_C T(0))} \langle \gamma, x\rangle$, proving the first part of the statement. As for the second one, it follows from Proposition 13. ■

Proof of Corollary 3. If T is differentiable at 0, so is each T_i. Since T is concave, so is each T_i. It follows that $\partial_C T_i(0) = \partial T_i(0)$ is a singleton for all $i \in \{1, ..., k\}$. This implies that $\partial_C T(0)$ is a singleton and coincides with the Jacobian of T at 0. By point 2 of Proposition 13 and since $\underline{A}(T)$ is regular, the Jacobian of T at 0 is regular, yielding that $\Gamma(\partial_C T(0))$ is a singleton given by the unique left Perron-Frobenius eigenvector of the Jacobian of T at 0. By Corollary 2, we can conclude that $\lim_{\beta\to 1} \tilde{x}_\beta = \lim_{\beta\to 1} x_\beta = \langle \gamma, x\rangle e$. ■

B.2.2 Star-shaped operators

In order to prove Proposition 3, we first provide two ancillary lemmas which give bounds on $\liminf_{\beta\to 1} x_\beta$ and $\limsup_{\beta\to 1} x_\beta$. These bounds are in terms of the limits of the operators S_α whose sup gives T.

Lemma 5 *If $\{S_\alpha\}_{\alpha\in\mathcal{A}}$ is nice, then* $\liminf_{\beta\to 1} x_\beta \geq \sup_{\alpha\in\mathcal{A}} \varphi_{S_\alpha}(x) e$ *for all $x \in \mathbb{R}^k$.*

Proof. By construction, we have that $T(y) \geq S_\alpha(y)$ for all $y \in \mathbb{R}^k$ and for all $\alpha \in \mathcal{A}$. By induction and since S_α and T are monotone, this implies that $T^t_{\beta,x}(y) \geq S^t_{\alpha,\beta,x}(y)$ for all $t \in \mathbb{N}$, for all $\beta \in (0,1)$, for all $x, y \in \mathbb{R}^k$, and for all $\alpha \in \mathcal{A}$. By passing to the limit and Lemma 1, this implies that $x_\beta \geq x_{\beta,\alpha}$ for all $\beta \in (0,1)$, for all $x \in \mathbb{R}^k$, and for all $\alpha \in \mathcal{A}$. By Corollary 2 and since $E(T) = D$, it follows that $\liminf_{\beta\to 1} x_\beta \geq \liminf_{\beta\to 1} x_{\beta,\alpha} = \lim_{\beta\to 1} x_{\beta,\alpha} = \varphi_{S_\alpha}(x) e$ for all $x \in \mathbb{R}^k$ and for all $\alpha \in \mathcal{A}$, proving the statement. ■

Lemma 6 *If $\{S_\alpha\}_{\alpha\in\mathcal{A}}$ is nice, then* $\sup_{\alpha\in\mathcal{A}} \varphi_{S_\alpha}(x) e \geq \limsup_{\beta\to 1} x_\beta$ *for all $x \in \mathbb{R}^k$.*

Proof. Fix $x \in \mathbb{R}^k$ and $\beta \in (0,1)$. By construction of T and definition of x_β and since $\{S_\alpha\}_{\alpha\in\mathcal{A}}$ is nice, we have that there exists $\alpha_\beta \in \mathcal{A}$

$$S_{\alpha_\beta}((1-\beta)x + \beta x_\beta) = T((1-\beta)x + \beta x_\beta) = x_\beta.$$

By Lemma 1, it follows that $x_\beta = x_{\beta,\alpha_\beta}$. By Proposition 2 and since β was arbitrarily chosen, we have that for each $\beta \in (0,1)$

$$\sup_{\alpha\in\mathcal{A}} \varphi_{S_\alpha}(x) \geq \varphi_{S_{\alpha_\beta}}(x) = \min_{\gamma\in\Gamma\left(\partial_C S_{\alpha_\beta}(0)\right)} \langle \gamma, x\rangle \geq \min_{\gamma\in\Gamma\left(\partial_C S_{\alpha_\beta}(0)\right)} \left\langle \gamma, x_{\beta,\alpha_\beta}\right\rangle = \min_{\gamma\in\Gamma\left(\partial_C S_{\alpha_\beta}(0)\right)} \langle \gamma, x_\beta\rangle . \tag{31}$$

By Lemma 2 and since $E(T) = D$, we have that there exists a sequence $\{x_{\beta_n}\}_{n\in\mathbb{N}} \subseteq \{x_\beta\}_{\beta\in(0,1)}$ such that $\beta_n \to 1$ and $\limsup_{\beta\to 1} x_\beta = \lim_n x_{\beta_n} = \bar{h}e$ for some $\bar{h} \in \mathbb{R}$. For each $n \in \mathbb{N}$ consider $\gamma_n \in \Gamma\left(\partial_C S_{\alpha_{\beta_n}}(0)\right)$ such that $\langle \gamma_n, x_{\beta_n}\rangle = \min_{\gamma\in\Gamma\left(\partial_C S_{\alpha_{\beta_n}}(0)\right)} \langle \gamma, x_{\beta_n}\rangle$. We have that $\sup_{\alpha\in\mathcal{A}} \varphi_{S_\alpha}(x) \geq \langle \gamma_n, x_{\beta_n}\rangle$ for all $n \in \mathbb{N}$. Since $\{\gamma_n\}_{n\in\mathbb{N}} \subseteq \Delta$, there exists a subsequence $\{\gamma_{n_l}\}_{l\in\mathbb{N}} \subseteq \{\gamma_n\}_{n\in\mathbb{N}}$ such that $\gamma_{n_l} \to \bar{\gamma} \in \Delta$. We can conclude that

$$\sup_{\alpha\in\mathcal{A}} \varphi_{S_\alpha}(x) \geq \lim_l \left\langle \gamma_{n_l}, x_{\beta_{n_l}}\right\rangle = \left\langle \bar{\gamma}, \bar{h}e\right\rangle = \bar{h},$$

proving that $\sup_{\alpha\in\mathcal{A}} \varphi_{S_\alpha}(x)\, e \geq \bar{h}e = \limsup_{\beta\to 1} x_\beta$. Since $x \in \mathbb{R}^k$ was arbitrarily chosen, the statement follows. ■

Proof of Proposition 3. By Lemmas 5 and 6, we have that $\lim_{\beta\to 1} x_\beta = \sup_{\alpha\in\mathcal{A}} \varphi_{S_\alpha}(x)\, e$ for all $x \in \mathbb{R}^k$. By Lemma 1, we can conclude that $\lim_{\beta\to 1} \tilde{x}_\beta = \lim_{\beta\to 1} x_\beta$ for all $x \in \mathbb{R}^k$, proving the statement. ■

B.2.3 Main result

Proof of Theorem 1. Note that T_i is normalized, monotone, translation invariant, and star-shaped for all $i \in \{1, ..., k\}$. By Proposition 10, we have that for each $i \in \{1, ..., k\}$ there exists a family $\{S_{\alpha_i}\}_{\alpha_i\in\mathcal{A}_i}$ of normalized, monotone, translation invariant, and concave functionals such that

$$T_i(x) = \max_{\alpha_i\in\mathcal{A}_i} S_{\alpha_i}(x) \qquad \forall x \in \mathbb{R}^k \tag{32}$$

and $\bar{\text{co}}\left(\partial_C S_{\alpha_i}\left(\mathbb{R}^k\right)\right) \subseteq \bar{\text{co}}\left(\partial_C T_i\left(\mathbb{R}^k\right)\right)$ for all $\alpha_i \in \mathcal{A}_i$. Define $\mathcal{A} = \Pi_{i=1}^k \mathcal{A}_i$ and for each $\alpha \in \mathcal{A}$ define $S_\alpha : \mathbb{R}^k \to \mathbb{R}^k$ to be such that its i-th component coincides with S_{α_i} for all $i \in \{1, ..., k\}$. It is immediate to see that S_α is normalized, monotone, translation invariant, and concave for all $\alpha \in \mathcal{A}$. Since $\bar{\text{co}}\left(\partial_C S_{\alpha_i}\left(\mathbb{R}^k\right)\right) \subseteq \bar{\text{co}}\left(\partial_C T_i\left(\mathbb{R}^k\right)\right)$ for all $\alpha_i \in \mathcal{A}_i$ and for all $i \in \{1, ..., k\}$, it follows that $\underline{A}(S_\alpha) \geq \underline{A}(T)$ for all $\alpha \in \mathcal{A}$. By Proposition 11 and since $\underline{A}(T)$ is regular, this implies that $\underline{A}(S_\alpha)$ is regular and $E(S_\alpha) = D$ for all $\alpha \in \mathcal{A}$. By (32) and since $\mathcal{A}$ has a product structure, we have that

$$T(x) = \sup_{\alpha\in\mathcal{A}} S_\alpha(x) \qquad \forall x \in \mathbb{R}^k$$

and for each $x \in \mathbb{R}^k$ there exists $\alpha_x \in \mathcal{A}$ such that $T(x) = S_{\alpha_x}(x)$. We can conclude that $\{S_\alpha\}_{\alpha\in\mathcal{A}}$ is nice. By Proposition 3, the statement follows. ■

C Appendix: Proofs of Sections 4 and 5

C.1 Zero-sum stochastic games

Proof of Proposition 6. For every $s \in S_M$, define the operator $H(\cdot, s) : \mathbb{R}^{R\times\Omega} \to \mathbb{R}^{R\times\Omega}$ as

$$H_{r,\omega}(z,s) = \min_{\tilde{s}\in S_m} \left\{ \sum_{(r',\omega')\in R\times\Omega} z_{r',\omega'}\rho(s,\tilde{s},\omega)(r',\omega') \right\} \qquad \forall (r,\omega) \in R\times\Omega, \forall z \in \mathbb{R}^{R\times\Omega}.$$

Moreover, define the operator $T : \mathbb{R}^{R\times\Omega} \to \mathbb{R}^{R\times\Omega}$ as

$$T_{r,\omega}(z) = \max_{s\in S_M} H_{r,\omega}(z,s) \qquad \forall (r,\omega) \in R\times\Omega, \forall z \in \mathbb{R}^{R\times\Omega}.$$

Observe that, for every $s \in S$, the operator $H(\cdot, s)$ is monotone, normalized, translation invariant, positive homogeneous, and concave. Moreover, since S_m is compact and ρ is continuous, it follows by the Maximum theorem that, for every $z \in \mathbb{R}^{R\times\Omega}$, the map $s \mapsto H(\cdot, s)$ is continuous. Given that S_M is compact, it follows that $\{H(\cdot,s)\}_{s\in S_M}$ is nice. Moreover, observe that by construction we have

$$H_{r,\omega} = H_{r',\omega} \qquad \forall r, r' \in R, \forall \omega \in \Omega.$$

Therefore, for all $s \in S_M$ we have

$$\begin{aligned}
\partial_C H(0,s) &= \left\{ \begin{array}{c} W \in \mathcal{W}_{R\times\Omega} : \exists \hat{\sigma} \in \Delta(S_m)^{R\times\Omega}, \forall (r,\omega), (r',\omega') \in R\times\Omega, \\ w_{(r,\omega),(r',\omega')} = \int_{S_m} \rho(s,\tilde{s},\omega)(r',\omega')\, d\hat{\sigma}_{r,\omega}(\tilde{s}) \end{array} \right\} \\
&= \left\{ \begin{array}{c} W \in \mathcal{W}_{R\times\Omega} : \exists \tilde{\sigma} \in \Sigma_m, \forall (r,\omega), (r',\omega') \in R\times\Omega, \\ w_{(r,\omega),(r',\omega')} = \int_{S_m} \rho(s,\tilde{s},\omega)(r',\omega')\, d\tilde{\sigma}_{\omega}(\tilde{s}) \end{array} \right\} \\
&= \left\{ W(s,\tilde{\sigma}) \in \mathcal{W}_{R\times\Omega} : \tilde{\sigma} \in \Sigma_m \right\}.
\end{aligned}$$

Next, define $x \in \mathbb{R}^{R\times\Omega}$ as $x_{r,\omega} = r$. Next, observe that, for every $\beta \in (0,1)$, the unique solution v^β of equation (19) does not depend on the realization of r. With this, for every $\beta \in (0,1)$, define $x^\beta \in \mathbb{R}^{R\times\Omega}$ as $x^\beta_{r,\omega} = v^\beta_\omega$, and observe that it is the unique solution of the fixed-point equation: for all $(r,\omega) \in R\times\Omega$,

$$x^\beta_{r,\omega} = T_{r,\omega}\left((1-\beta)x + \beta x^\beta\right).$$

Also, for every $t \in \mathbb{N}$, define $x^t \in \mathbb{R}^{R\times\Omega}$ as $x^t_{r,\omega} = v^t_\omega$, and observe that, for all $t \in \mathbb{N}$ and for all $(r,\omega) \in R\times\Omega$,

$$x^t_{r,\omega} = T_{r,\omega}\left(\frac{1}{t}x + \frac{t-1}{t}x^\beta\right).$$

Therefore, by Propositions 3 and 8, it follows that

$$\lim_{\beta\to 1} x^{\beta} = \lim_{t} x^{t} = \sup_{s\in S_M}\left(\min_{\gamma\in\Gamma(\partial_C H(0,s))}\sum_{(r',\omega')\in R\times\Omega} x_{r',\omega'}\gamma_{r',\omega'}\right)e = \left(\sup_{s\in S_M}\min_{\tilde{\sigma}\in\Sigma_m}\sum_{r\in R} r\gamma\left(s,\tilde{\sigma}\right)(r)\right)e,$$

yielding the result. ■

D Appendix: Proofs of Section 7

D.1 Range

Given the additivity and homogeneity properties of max and min, it is routine to check that

$$\operatorname{Rg}\left(\lambda y+\mu z\right)\leq\lambda\operatorname{Rg}\left(y\right)+\mu\operatorname{Rg}\left(z\right)\quad\forall\lambda,\mu\in\mathbb{R}_{+},\forall y,z\in\mathbb{R}^{k}. \tag{33}$$

In particular, since $\tilde{x}_\beta = (1-\beta)x+\beta x_\beta$ for all $x\in\mathbb{R}^k$ and for all $\beta\in(0,1)$, this implies that

$$\operatorname{Rg}\left(\tilde{x}_{\beta}\right)\leq\left(1-\beta\right)\operatorname{Rg}\left(x\right)+\beta\operatorname{Rg}\left(x_{\beta}\right)\quad\forall x\in\mathbb{R}^{k},\forall\beta\in\left(0,1\right). \tag{34}$$

If T is normalized and monotone, we have that $\left(\min_{i\in\{1,\dots,k\}} y_i\right)e\leq T\left(y\right)\leq\left(\max_{i\in\{1,\dots,k\}} y_i\right)e$ for all $y\in\mathbb{R}^k$, thus

$$\operatorname{Rg}\left(T\left(y\right)\right)\leq\operatorname{Rg}\left(y\right)\qquad\forall y\in\mathbb{R}^{k}. \tag{35}$$

By definition of $\tilde{x}_\beta$ and x_β, we have that $T\left(\tilde{x}_\beta\right)=x_\beta$ for all $x\in\mathbb{R}^k$ and for all $\beta\in(0,1)$ and

$$\operatorname{Rg}\left(x_{\beta}\right)\leq\operatorname{Rg}\left(\tilde{x}_{\beta}\right)\quad\forall x\in\mathbb{R}^{k},\forall\beta\in\left(0,1\right). \tag{36}$$

Thus, for our purposes, (34) and (36) show that we can alternatively either study $\operatorname{Rg}\left(x_\beta\right)$ or $\operatorname{Rg}\left(\tilde{x}_\beta\right)$. Since some results are easier to be derived just focusing on one of the two, we will extensively use these inequalities to go back and forth $\operatorname{Rg}\left(x_\beta\right)$ and $\operatorname{Rg}\left(\tilde{x}_\beta\right)$. We begin with two ancillary lemmas.

Lemma 7 *Let $T:\mathbb{R}^k\to\mathbb{R}^k$ be such that there exist a stochastic matrix W and $\varepsilon\in(0,1]$ such that*

$$T\left(y\right)=\varepsilon Wy+\left(1-\varepsilon\right)S\left(y\right)\qquad\forall y\in\mathbb{R}^{k} \tag{37}$$

where $S:\mathbb{R}^k\to\mathbb{R}^k$ is normalized, monotone, and translation invariant. If there exist $h\in\{1,...,k\}$ and $\hat{t}\in\mathbb{N}$ such that $w_{ih}^{(\hat{t})}>0$ for all $i\in\{1,...,k\}$, then

$$\operatorname{Rg}\left(x_{\beta}\right)\leq\frac{1}{1+\frac{\delta^{\hat{t}}(\beta\varepsilon)^{\hat{t}}(1-\beta\varepsilon)}{(1-\beta)\left(1-(\beta\varepsilon)^{\hat{t}}\right)}}\operatorname{Rg}\left(x\right)\quad\forall\beta\in\left(0,1\right),\forall x\in\mathbb{R}^{k}$$

where $\delta=\min_{i,j:w_{ij}>0} w_{ij}$.

Proof. Recall that $\tilde{x}_\beta = (1-\beta)\,x + \beta x_\beta$ for all $x \in \mathbb{R}^k$ and for all $\beta \in (0,1)$. Given $x \in \mathbb{R}^k$ and $\beta \in (0,1)$, recall also that $T_{\beta,x} : \mathbb{R}^k \to \mathbb{R}^k$ is defined by $T_{\beta,x}(y) = T((1-\beta)\,x + \beta y)$ for all $y \in \mathbb{R}^k$. By Lemma 1, x_β is a fixed point of $T_{\beta,x}$ and so of $T^t_{\beta,x}$ for all $t \in \mathbb{N}$.

Step 1. For each $x \in \mathbb{R}^k$, for each $\beta \in (0,1)$, and for each $t \in \mathbb{N}$

$$T^t_{\beta,x}(x_\beta) = (1-\beta)\,\varepsilon \sum_{\tau=0}^{t-1} (\beta\varepsilon)^\tau W^{\tau+1} x + (\beta\varepsilon)^t W^t x_\beta + (1-\varepsilon) \sum_{\tau=0}^{t-1} (\beta\varepsilon)^\tau W^\tau S(\tilde{x}_\beta).$$

Proof of the Step. We proceed by induction.

Initial Step. If $t = 1$, then

$$\begin{aligned}
T_{\beta,x}(x_\beta) &= T((1-\beta)\,x + \beta x_\beta) \\
&= (1-\beta)\,\varepsilon W x + \beta\varepsilon W x_\beta + (1-\varepsilon)\,S((1-\beta)\,x + \beta x_\beta) \\
&= (1-\beta)\,\varepsilon \sum_{\tau=0}^{t-1} (\beta\varepsilon)^\tau W^{\tau+1} x + (\beta\varepsilon)^t W^t x_\beta + (1-\varepsilon) \sum_{\tau=0}^{t-1} (\beta\varepsilon)^\tau W^\tau S(\tilde{x}_\beta),
\end{aligned}$$

proving the initial step.

Inductive Step. Assume the statement holds for t. We show it holds for $t+1$. By (37) and inductive hypothesis $T^t_{\beta,x}(x_\beta) = x_\beta$, we have that

$$\begin{aligned}
T^{t+1}_{\beta,x}(x_\beta) &= T\left((1-\beta)\,x + \beta T^t_{\beta,x}(x_\beta)\right) \\
&= \varepsilon W\left((1-\beta)\,x + \beta T^t_{\beta,x}(x_\beta)\right) + (1-\varepsilon)\,S\left((1-\beta)\,x + \beta T^t_{\beta,x}(x_\beta)\right) \\
&= (1-\beta)\,\varepsilon W x + \beta\varepsilon W T^t_{\beta,x}(x_\beta) + (1-\varepsilon)\,S((1-\beta)\,x + \beta x_\beta) \\
&= (1-\beta)\,\varepsilon W x \\
&\quad + \beta\varepsilon W\left((1-\beta)\,\varepsilon \sum_{\tau=0}^{t-1} (\beta\varepsilon)^\tau W^{\tau+1} x + (\beta\varepsilon)^t W^t x_\beta + (1-\varepsilon) \sum_{\tau=0}^{t-1} (\beta\varepsilon)^\tau W^\tau S(\tilde{x}_\beta)\right) \\
&\quad + (1-\varepsilon)\,S(\tilde{x}_\beta) \\
&= (1-\beta)\,\varepsilon W x + (1-\beta)\,\varepsilon \sum_{\tau=0}^{t-1} (\beta\varepsilon)^{\tau+1} W^{\tau+2} x + (\beta\varepsilon)^{t+1} W^{t+1} x_\beta \\
&\quad + (1-\varepsilon) \sum_{\tau=0}^{t-1} (\beta\varepsilon)^{\tau+1} W^{\tau+1} S(\tilde{x}_\beta) + (1-\varepsilon)\,S(\tilde{x}_\beta) \\
&= (1-\beta)\,\varepsilon \sum_{\tau=0}^{t} (\beta\varepsilon)^\tau W^{\tau+1} x + (\beta\varepsilon)^{t+1} W^{t+1} x_\beta + (1-\varepsilon) \sum_{\tau=0}^{t} (\beta\varepsilon)^\tau W^\tau S(\tilde{x}_\beta),
\end{aligned}$$

proving the inductive step.

Step 1 follows by induction. □

Step 2. For each $z \in \mathbb{R}^k$

$$\mathrm{Rg}\left(W^{\hat{t}} z\right) \leq \left(1 - \delta^{\hat{t}}\right) \mathrm{Rg}\left(z\right)$$

where $\delta = \min_{i,j: w_{ij} > 0} w_{ij}$*.*

Proof of the Step. Let $z \in \mathbb{R}^k$ and define $y = W^{\hat{t}} z$. Consider $i_1, i_2 \in \{1, ..., k\}$ such that $y_{i_1} = \max_{i \in \{1,...,k\}} y_i$ and $y_{i_2} = \min_{i \in \{1,...,k\}} y_i$. Define also $z^\star = \max_{i \in \{1,...,k\}} z_i$ and $z_\star = \min_{i \in \{1,...,k\}} z_i$. Define $\tilde{\delta} = \min_{i \in \{1,...,h\}} w_{ih}^{(\hat{t})} \in (0,1)$ where $w_{ih}^{(\hat{t})}$ is the ih-th entry of $W^{\hat{t}}$. Note that

$$\begin{aligned}
\mathrm{Rg}\left(W^{\hat{t}} z\right) &= \mathrm{Rg}\left(y\right) = y_{i_1} - y_{i_2} = \sum_{j=1}^{k} w_{i_1 j}^{(\hat{t})} z_j - \sum_{j=1}^{k} w_{i_2 j}^{(\hat{t})} z_j \\
&\leq \left(1 - w_{i_1 h}^{(\hat{t})}\right) z^\star - \left(1 - w_{i_2 h}^{(\hat{t})}\right) z_\star + \left(w_{i_1 h}^{(\hat{t})} - \tilde{\delta}\right) z_h + \tilde{\delta} z_h - \left(w_{i_2 h}^{(\hat{t})} - \tilde{\delta}\right) z_h - \tilde{\delta} z_h \\
&\leq \left(1 - w_{i_1 h}^{(\hat{t})}\right) z^\star - \left(1 - w_{i_2 h}^{(\hat{t})}\right) z_\star + \left(w_{i_1 h}^{(\hat{t})} - \tilde{\delta}\right) z^\star - \left(w_{i_2 h}^{(\hat{t})} - \tilde{\delta}\right) z_\star \\
&\leq \left(1 - \tilde{\delta}\right) \left(z^\star - z_\star\right) = \left(1 - \tilde{\delta}\right) \mathrm{Rg}\left(z\right).
\end{aligned}$$

Next, by induction, it is immediate to see that $\min_{i,j: w_{ij}^{(t)} > 0} w_{ij}^{(t)} \geq \left(\min_{i,j: w_{ij} > 0} w_{ij}\right)^t = \delta^t$ for all $t \in \mathbb{N}$. Since $w_{ih}^{(\hat{t})} > 0$ for all $i \in \{1, ..., k\}$, we can conclude that $\tilde{\delta} \geq \delta^{\hat{t}}$, proving the statement. □

By Steps 1 and 2 as well as (33), (34), and (35) and since $T_{\beta,x}^{\hat{t}}\left(x_\beta\right) = x_\beta$ and the composition of normalized and monotone operators is normalized and monotone, we have that for each $x \in \mathbb{R}^k$ and for each $\beta \in (0,1)$

$$\begin{aligned}
\mathrm{Rg}(x_\beta) &= \mathrm{Rg}\left(T^{\hat{t}}_{\beta,x}(x_\beta)\right) \\
&= \mathrm{Rg}\left[(1-\beta)\varepsilon\sum_{\tau=0}^{\hat{t}-1}(\beta\varepsilon)^\tau W^{\tau+1}x + (\beta\varepsilon)^{\hat{t}} W^{\hat{t}} x_\beta + (1-\varepsilon)\sum_{\tau=0}^{\hat{t}-1}(\beta\varepsilon)^\tau W^\tau S(\tilde{x}_\beta)\right] \\
&\le (1-\beta)\varepsilon\sum_{\tau=0}^{\hat{t}-1}(\beta\varepsilon)^\tau \mathrm{Rg}\left(W^{\tau+1}x\right) + (\beta\varepsilon)^{\hat{t}}\mathrm{Rg}\left(W^{\hat{t}}x_\beta\right) + (1-\varepsilon)\sum_{\tau=0}^{\hat{t}-1}(\beta\varepsilon)^\tau \mathrm{Rg}\left(W^\tau S(\tilde{x}_\beta)\right) \\
&\le (1-\beta)\varepsilon\sum_{\tau=0}^{\hat{t}-1}(\beta\varepsilon)^\tau \mathrm{Rg}(x) + (\beta\varepsilon)^{\hat{t}}\mathrm{Rg}\left(W^{\hat{t}}x_\beta\right) + (1-\varepsilon)\sum_{\tau=0}^{\hat{t}-1}(\beta\varepsilon)^\tau \mathrm{Rg}(\tilde{x}_\beta) \\
&\le (1-\beta)\varepsilon\sum_{\tau=0}^{\hat{t}-1}(\beta\varepsilon)^\tau \mathrm{Rg}(x) + (\beta\varepsilon)^{\hat{t}}\mathrm{Rg}\left(W^{\hat{t}}x_\beta\right) + (1-\varepsilon)(1-\beta)\sum_{\tau=0}^{\hat{t}-1}(\beta\varepsilon)^\tau \mathrm{Rg}(x) \\
&+ (1-\varepsilon)\beta\sum_{\tau=0}^{\hat{t}-1}(\beta\varepsilon)^\tau \mathrm{Rg}(x_\beta) \\
&= (1-\beta)\frac{1-(\beta\varepsilon)^{\hat{t}}}{1-\beta\varepsilon}\mathrm{Rg}(x) + (\beta\varepsilon)^{\hat{t}}\mathrm{Rg}\left(W^{\hat{t}}x_\beta\right) + (1-\varepsilon)\beta\frac{1-(\beta\varepsilon)^{\hat{t}}}{1-\beta\varepsilon}\mathrm{Rg}(x_\beta) \\
&\le (1-\beta)\frac{1-(\beta\varepsilon)^{\hat{t}}}{1-\beta\varepsilon}\mathrm{Rg}(x) + (\beta\varepsilon)^{\hat{t}}\left(1-\delta^{\hat{t}}\right)\mathrm{Rg}(x_\beta) + (1-\varepsilon)\beta\frac{1-(\beta\varepsilon)^{\hat{t}}}{1-\beta\varepsilon}\mathrm{Rg}(x_\beta) \\
&= (1-\beta)\frac{1-(\beta\varepsilon)^{\hat{t}}}{1-\beta\varepsilon}\mathrm{Rg}(x) + \left(\frac{\beta-\beta\varepsilon-\beta(\beta\varepsilon)^{\hat{t}}+(\beta\varepsilon)^{\hat{t}}-\delta^{\hat{t}}(\beta\varepsilon)^{\hat{t}}+\delta^{\hat{t}}(\beta\varepsilon)^{\hat{t}+1}}{1-\beta\varepsilon}\right)\mathrm{Rg}(x_\beta).
\end{aligned}$$

Since $(\beta\varepsilon)^{\hat{t}}\left(1-\delta^{\hat{t}}\right) + (1-\varepsilon)\beta\frac{1-(\beta\varepsilon)^{\hat{t}}}{1-\beta\varepsilon} \in (0,1)$, this implies that

$$\mathrm{Rg}(x_\beta) \le \frac{(1-\beta)\left(1-(\beta\varepsilon)^{\hat{t}}\right)}{(1-\beta)\left(1-(\beta\varepsilon)^{\hat{t}}\right)+\delta^{\hat{t}}(\beta\varepsilon)^{\hat{t}}(1-\beta\varepsilon)}\mathrm{Rg}(x) \quad \forall x \in \mathbb{R}^k, \forall \beta \in (0,1),$$

proving the statement. ■

From the previous lemma, we obtain an estimate on the range of x_β, provided T admits a decomposition as in (37) and W has eventually a strictly positive column. This latter property is achieved whenever $A(W)$ not only is regular, but also "aperiodic". As for the former property, by [12, Proposition 7], we have that if T is normalized, monotone, and translation invariant and $\underline{A}(T)$ is nontrivial, then there there exist a stochastic matrix W and $\varepsilon \in (0,1]$ such that

$$T(y) = \varepsilon W y + (1-\varepsilon)S(y) \qquad \forall y \in \mathbb{R}^k$$

where $S : \mathbb{R}^k \to \mathbb{R}^k$ is normalized, monotone, and translation invariant. Moreover, W can be chosen to be such that $A(W) = \underline{A}(T)$. Thus, if $\underline{A}(T)$ is also regular and aperiodic, so is $A(W)$. Since in our statements we have the property of regularity, but not aperiodicity, we consider an auxiliary operator closely related to T and which will satisfy the property of aperiodicity.

Given $T : \mathbb{R}^k \to \mathbb{R}^k$, it will be thus useful to consider the averaged operator $T_\lambda = \lambda I + (1-\lambda) T$ with $\lambda \in (0,1)$ where I is the identity.[21] Note that $\underline{A}(T_\lambda) \geq \underline{A}(T)$ and, in particular, the only difference between $\underline{A}(T_\lambda)$ and $\underline{A}(T)$ consists in the entries of the diagonal of $\underline{A}(T_\lambda)$ which are all 1, while those of $\underline{A}(T)$ might be 0.

Building on Lemma 7, the next result provides a result on the convergence of $\mathrm{Rg}(x_{\beta,\lambda})$ where for each $x \in \mathbb{R}^k$ and for each $\beta, \lambda \in (0,1)$, $x_{\beta,\lambda}$ is the unique point satisfying

$$T_\lambda((1-\beta)x + \beta x_{\beta,\lambda}) = x_{\beta,\lambda}.$$

Lemma 8 *Let $T : \mathbb{R}^k \to \mathbb{R}^k$ be normalized, monotone, and translation invariant. If $\underline{A}(T)$ is regular, then*

$$\mathrm{Rg}(x_{\beta,\lambda}) \leq \frac{1}{1 + \frac{\hat{\delta}^{\hat{t}} \left(\beta\hat{\delta}\right)^{\hat{t}} \left(1-\beta\hat{\delta}\right)}{(1-\beta)\left(1-\left(\beta\hat{\delta}\right)^{\hat{t}}\right)}} \mathrm{Rg}(x) \qquad \forall x \in \mathbb{R}^k, \forall \beta, \lambda \in (0,1), \tag{38}$$

where $\hat{\delta} = \min\{\lambda, (1-\lambda)\delta\}$, $\delta = \min_{i,j:\underline{a}_{ij}=1} \inf_{x \in \mathcal{D}} \frac{\partial T_i}{\partial x_j}(x)$, and $\hat{t} \in \mathbb{N}$ is the smallest natural number such that $\underline{A}(T_\lambda)^{\hat{t}}$ has one column with all positive entries.

Proof. Since T is normalized, monotone, translation invariant and $\underline{A}(T)$ is nontrivial, we have that $\delta \in (0,1]$. Since $\underline{A}(T_\lambda) \geq \underline{A}(T)$ and $\underline{A}(T)$ is regular, we have that $\underline{A}(T_\lambda)$ is regular. By [12, Proposition 7], we have that there exist a stochastic matrix W and $\varepsilon \in (0,1]$ such that

$$T_\lambda(y) = \varepsilon W y + (1-\varepsilon) S(y) \qquad \forall y \in \mathbb{R}^k$$

where $S : \mathbb{R}^k \to \mathbb{R}^k$ is normalized, monotone, and translation invariant. Moreover, W can be chosen to be such that $A(W) = \underline{A}(T_\lambda)$. By the proof of Proposition 7 in [12], it follows that ε can be chosen to be equal to $\hat{\delta}$ and all the strictly positive entries of W are greater than or equal to $\hat{\delta}$. This implies that $A(W)$ is regular and $A(W) \geq I$. In particular (see, e.g., [33, Exercise 4.13]), the set of natural numbers $t \in \mathbb{N}$ such that $A(W)^{\hat{t}} = \underline{A}(T_\lambda)^{\hat{t}}$ has one column with all positive entries is nonempty and $\hat{t}$ is well defined. By Lemma 7, the statement follows. ■

[21] With a small abuse of notation, we denote with I both the identity matrix and the identity operator.

Remark 1 Note that

$$\frac{1}{1+\frac{\hat{\delta}^{\hat{t}}\left(\beta\hat{\delta}\right)^{\hat{t}}\left(1-\beta\hat{\delta}\right)}{(1-\beta)\left(1-\left(\beta\hat{\delta}\right)^{\hat{t}}\right)}}=\frac{(1-\beta)\left(1-\left(\beta\hat{\delta}\right)^{\hat{t}}\right)}{(1-\beta)\left(1-\left(\beta\hat{\delta}\right)^{\hat{t}}\right)+\hat{\delta}^{\hat{t}}\left(\beta\hat{\delta}\right)^{\hat{t}}\left(1-\beta\hat{\delta}\right)}.$$

Consider the quantity at the denominator of the fraction on the right-hand side. It is immediate to see that it is equal to

$$\begin{aligned}\left(1-\beta\hat{\delta}\right)\left((1-\beta)\frac{1-\left(\beta\hat{\delta}\right)^{\hat{t}}}{1-\beta\hat{\delta}}+\hat{\delta}^{\hat{t}}\left(\beta\hat{\delta}\right)^{\hat{t}}\right)&=\left(1-\beta\hat{\delta}\right)\left((1-\beta)\sum_{\tau=0}^{\hat{t}-1}\left(\beta\hat{\delta}\right)^{\tau}+\hat{\delta}^{\hat{t}}\left(\beta\hat{\delta}\right)^{\hat{t}}\right)\\&\geq\left(1-\beta\hat{\delta}\right)\left(1-\beta+\hat{\delta}^{\hat{t}}\left(\beta\hat{\delta}\right)^{\hat{t}}\right).\end{aligned}$$

Consider now the function $h:\mathbb{R}\to\mathbb{R}$ defined by $h(\beta)=1-\beta+\hat{\delta}^{2\hat{t}}\beta^{\hat{t}}$ for all $\beta\in\mathbb{R}$. Since $\hat{\delta}\in(0,1)$ and $\hat{t}\in\mathbb{N}$, h is convex and differentiable on $\mathbb{R}$ with derivative $h'(\beta)=-1+\hat{t}\hat{\delta}^{2\hat{t}}\beta^{\hat{t}-1}$ for all $\beta\in\mathbb{R}$. Clearly, h' is negative in a neighborhood of 0. We thus have two cases:

1. $h'(\beta)\leq 0$ for all $\beta\in[0,1]$. This happens if and only if $\hat{\delta}\leq\left(\frac{1}{\hat{t}}\right)^{\frac{1}{2\hat{t}}}$ and, in this case, $h(\beta)\geq\hat{\delta}^{2\hat{t}}>0$ for all $\beta\in[0,1]$.

2. $h'(\beta)>0$ for some $\beta\in[0,1]$. Since $h'(\beta)>0$ for some $\beta\in[0,1]$, we have that $\hat{\delta}>\left(\frac{1}{\hat{t}}\right)^{\frac{1}{2\hat{t}}}$. Since h is convex and $\hat{\delta}>\left(\frac{1}{\hat{t}}\right)^{\frac{1}{2\hat{t}}}$, that is $1>1/\hat{t}\hat{\delta}^{2\hat{t}}>0$, this implies that h is minimized at $\beta_\star\in(0,1)$ where $\beta_\star=\sqrt[\hat{t}-1]{1/\hat{t}\hat{\delta}^{2\hat{t}}}\in(0,1)$ and

$$\begin{aligned}h(\beta_\star)&=1-\left(\frac{1}{\hat{t}\hat{\delta}^{2\hat{t}}}\right)^{\frac{1}{\hat{t}-1}}+\hat{\delta}^{2\hat{t}}\left(\frac{1}{\hat{t}\hat{\delta}^{2\hat{t}}}\right)^{\frac{\hat{t}}{\hat{t}-1}}\\&=1-\left(\frac{1}{\hat{t}\hat{\delta}^{2\hat{t}}}\right)^{\frac{1}{\hat{t}-1}}+\hat{\delta}^{2\hat{t}}\left(\frac{1}{\hat{t}\hat{\delta}^{2\hat{t}}}\right)\left(\frac{1}{\hat{t}\hat{\delta}^{2\hat{t}}}\right)^{\frac{1}{\hat{t}-1}}\\&=1-\left(\frac{1}{\hat{t}\hat{\delta}^{2\hat{t}}}\right)^{\frac{1}{\hat{t}-1}}\left(1-\frac{1}{\hat{t}}\right)\geq 1-\left(1-\frac{1}{\hat{t}}\right)\geq\frac{1}{\hat{t}}>0.\end{aligned}$$

We can conclude that

$$\frac{1}{1+\frac{\hat{\delta}^{\hat{t}}\left(\beta\hat{\delta}\right)^{\hat{t}}\left(1-\beta\hat{\delta}\right)}{(1-\beta)\left(1-\left(\beta\hat{\delta}\right)^{\hat{t}}\right)}}\leq(1-\beta)\frac{1}{\left(1-\hat{\delta}\right)\min\left\{\frac{1}{\hat{t}},\hat{\delta}^{2\hat{t}}\right\}}.\tag{39}$$

Finally, since $\hat{\delta}=\min\{\lambda,(1-\lambda)\delta\}$ and λ can be arbitrarily chosen, $\hat{\delta}$ is maximized for $\lambda=\delta/(1+\delta)$. In this case, $\hat{\delta}=\delta/(1+\delta)$. We will use (39) with this choice of $\hat{\delta}$ later on. ▲

Lemma 8, paired with Remark 1, is instrumental in proving Theorem 2. In fact, it only provides an estimate for the range of the fixed points of the averaged operator T_λ with $\lambda = \delta / (1+\delta)$. The next formula describes the relation between the points $\tilde{x}_\beta$ which solve (2) for the operator T and the points $\tilde{x}_{\beta,\lambda}$ which solve the same equation, but for the operator T_λ. In turn, this provides a relation between $\mathrm{Rg}(\tilde{x}_\beta)$ and $\mathrm{Rg}(\tilde{x}_{\beta,\lambda})$, and via (34) and (36), between $\mathrm{Rg}(x_\beta)$ and $\mathrm{Rg}(x_{\beta,\lambda})$.

Lemma 9 *If T is normalized, monotone, and translation invariant, then*

$$\tilde{x}_\beta = \tilde{x}_{\frac{\beta}{(1-\lambda)+\lambda\beta},\lambda} \qquad \forall x \in \mathbb{R}^k, \forall \beta, \lambda \in (0,1).$$

Moreover, for each $\lambda \in (0,1)$ the function $f_\lambda : (0,1) \to (0,1)$, defined by $f_\lambda(\beta) = \beta / [(1-\lambda) + \lambda\beta]$ for all $\beta \in (0,1)$, is strictly increasing and $\lim_{\beta \to 1} f_\lambda(\beta) = 1$.

Proof. Define the averaged operator $T_\lambda = \lambda I + (1-\lambda) T$ with $\lambda \in (0,1)$. By definition of $\tilde{x}_{\beta,\lambda}$, note that

$$\begin{aligned}
&(1-\beta) x + \beta\lambda\tilde{x}_{\beta,\lambda} + \beta(1-\lambda) T(\tilde{x}_{\beta,\lambda}) \\
&= (1-\beta) x + \beta[\lambda\tilde{x}_{\beta,\lambda} + (1-\lambda) T(\tilde{x}_{\beta,\lambda})] \\
&= (1-\beta) x + \beta T_\lambda(\tilde{x}_{\beta,\lambda}) = \tilde{x}_{\beta,\lambda} \quad \forall x \in \mathbb{R}^k, \forall \beta, \lambda \in (0,1),
\end{aligned}$$

that is,

$$\frac{1-\beta}{1-\beta\lambda} x + \frac{\beta(1-\lambda)}{1-\beta\lambda} T(\tilde{x}_{\beta,\lambda}) = \tilde{x}_{\beta,\lambda} \quad \forall \beta, \lambda \in (0,1), \forall x \in \mathbb{R}^k,$$

yielding that $\tilde{x}_{\beta,\lambda}$ solves equation (2) for the operator T with weight $\frac{\beta(1-\lambda)}{1-\beta\lambda}$. By the uniqueness of the solution, we can conclude that $\tilde{x}_{\frac{\beta(1-\lambda)}{1-\beta\lambda}} = \tilde{x}_{\beta,\lambda}$ for all $x \in \mathbb{R}^k$ and for all $\beta, \lambda \in (0,1)$. Fix $\lambda \in (0,1)$. If we define $g_\lambda : (0,1) \to (0,1)$ by $g_\lambda(\beta) = \beta(1-\lambda) / (1-\beta\lambda)$ for all $\beta \in (0,1)$, then g_λ is well defined and $g'_\lambda > 0$. The inverse of g_λ is f_λ and shares the same properties and, in particular, $\lim_{\beta \to 1} f_\lambda(\beta) = 1$. Since λ was arbitrarily chosen, it follows that

$$\tilde{x}_{f_\lambda(\beta),\lambda} = \tilde{x}_{g_\lambda(f_\lambda(\beta))} = \tilde{x}_\beta \qquad \forall x \in \mathbb{R}^k, \forall \lambda, \beta \in (0,1),$$

proving the statement. ■

Proof of Theorem 2. Set $\bar{\lambda} = \delta / (1+\delta) \in (0,1)$. By Lemma 8 and Remark 1 and since $\underline{A}(T_{\bar{\lambda}}) = \underline{A}(T) \vee I$, we have that for each $x \in \mathbb{R}^k$ and for each $\beta \in (0,1)$

$$\mathrm{Rg}\left(x_{\beta,\bar{\lambda}}\right) \leq (1-\beta) \frac{1}{\left(1 - \frac{\delta}{1+\delta}\right) \min\left\{\frac{1}{\hat{t}}, \left(\frac{\delta}{1+\delta}\right)^{2\hat{t}}\right\}} \mathrm{Rg}(x) \leq (1-\beta) \kappa_T \mathrm{Rg}(x).$$

By (34), we have that

$$\mathrm{Rg}\left(\tilde{x}_{\beta,\bar{\lambda}}\right) \leq (1-\beta)(1+\beta\kappa_T) \mathrm{Rg}(x) \quad \forall x \in \mathbb{R}^k, \forall \beta \in (0,1).$$

By Lemma 9, recall that

$$\tilde{x}_\beta = \tilde{x}_{\frac{\beta}{(1-\bar{\lambda})+\bar{\lambda}\beta},\bar{\lambda}} = \tilde{x}_{f_{\bar{\lambda}}(\beta),\bar{\lambda}} \qquad \forall \beta \in (0,1), \forall x \in \mathbb{R}^k.$$

We can conclude that

$$\mathrm{Rg}(\tilde{x}_\beta) = \mathrm{Rg}\left(\tilde{x}_{f_{\bar{\lambda}}(\beta),\bar{\lambda}}\right) \leq (1 - f_{\bar{\lambda}}(\beta))(1 + f_{\bar{\lambda}}(\beta)\kappa_T)\,\mathrm{Rg}(x) \quad \forall x \in \mathbb{R}^k, \forall \beta \in (0,1).$$

By (36), we have that

$$\mathrm{Rg}(x_\beta) \leq \mathrm{Rg}(\tilde{x}_\beta) \leq (1 - f_{\bar{\lambda}}(\beta))(1 + f_{\bar{\lambda}}(\beta)\kappa_T)\,\mathrm{Rg}(x) \quad \forall x \in \mathbb{R}^k, \forall \beta \in (0,1).$$

Finally, observe that $\left(1-\bar{\lambda}\right) + \bar{\lambda}\beta \in (\beta, 1)$, that is, $1 > \frac{\beta}{(1-\delta)+\delta\beta} > \beta$ for all $\beta \in (0,1)$. This implies that $1 > f_{\bar{\lambda}}(\beta) > \beta > 0$ and $0 < 1 - f_{\bar{\lambda}}(\beta) < 1 - \beta$ for all $\beta \in (0,1)$. Since $\kappa_T > 0$, we can conclude that

$$(1 - f_{\bar{\lambda}}(\beta))(1 + f_{\bar{\lambda}}(\beta)\kappa_T) \leq (1-\beta)(1+\kappa_T) \quad \forall \beta \in (0,1),$$

yielding that

$$\mathrm{Rg}(x_\beta) \leq \mathrm{Rg}(\tilde{x}_\beta) \leq (1-\beta)(1+\kappa_T)\,\mathrm{Rg}(x) \quad \forall x \in \mathbb{R}^k, \forall \beta \in (0,1),$$

proving the statement. ■

D.2 Rate of convergence

Proof of Theorem 3. Consider $x \in \mathbb{R}^k$ and $\beta \in (0,1)$. As usual, we have that $\tilde{x}_\beta = (1-\beta)x + \beta x_\beta$. By point 2 of Proposition 13 and since T is continuously differentiable and $\underline{A}(T)$ is regular, $\Gamma(\partial_C T(0))$ consists of only one element, denoted by γ. Set $\bar{h} = \langle \gamma, x \rangle$. By definition of $W_{\tilde{x}_\beta,\bar{h}}$ (see proof of Proposition 1), we have that

$$W_{\tilde{x}_\beta,\bar{h}}((1-\beta)x + \beta x_\beta) = W_{\tilde{x}_\beta,\bar{h}}\tilde{x}_\beta = T(\tilde{x}_\beta) = T((1-\beta)x + \beta x_\beta) = x_\beta.$$

By (3), we have that $x_\beta = x_{\beta, W_{\tilde{x}_\beta,\bar{h}}}$. By point 1 of Proposition 13 and since $\underline{A}(T)$ is regular, $\Gamma\left(W_{\tilde{x}_\beta,\bar{h}}\right)$ consists of only one element which we denote by γ_β. It follows that

$$\begin{aligned}
\|x_\beta - \langle \gamma, x \rangle e\|_\infty &\leq \|x_\beta - \langle \gamma_\beta, x \rangle e\|_\infty + \|\langle \gamma_\beta, x \rangle e - \langle \gamma, x \rangle e\|_\infty \\
&= \left\|x_{\beta, W_{\tilde{x}_\beta,\bar{h}}} - \langle \gamma_\beta, x \rangle e\right\|_\infty + |\langle \gamma_\beta - \gamma, x \rangle| \\
&\leq \left\|x_{\beta, W_{\tilde{x}_\beta,\bar{h}}} - \langle \gamma_\beta, x \rangle e\right\|_\infty + \|\gamma_\beta - \gamma\|_1 \|x\|_\infty.
\end{aligned}$$

We next bound the two terms on the right-hand side.

a By Example 1, we have that

$$\left\| x_{\beta,W_{\tilde{x}_\beta,\bar{h}}} - \langle \gamma_\beta, x\rangle e \right\|_\infty \leq \mathrm{Rg}\left(x_{\beta,W_{\tilde{x}_\beta,\bar{h}}}\right) = \mathrm{Rg}\left(x_\beta\right).$$

b We next bound $\|\gamma_\beta - \gamma\|_1 = \|\gamma - \gamma_\beta\|_1$. Consider $W = J_T(0)$. By Proposition 13, we have that $\min_{i,j:\underline{a}_{ij}=1} w_{ij} \geq \delta$. Set $\tilde{W} = \lambda I + (1-\lambda) W$ and $\hat{W} = \lambda I + (1-\lambda) W_{\tilde{x}_\beta,\bar{h}}$ where λ can be arbitrarily chosen in $(0,1)$. It is immediate to see that $\gamma^{\mathrm{T}}\tilde{W}^t = \gamma^{\mathrm{T}}$ and $\gamma_\beta^{\mathrm{T}}\hat{W}^t = \gamma_\beta^{\mathrm{T}}$ for all $t \in \mathbb{N}$. Note that $A\left(\tilde{W}\right)$ and $A\left(\hat{W}\right)$ are both regular and such that $A\left(\tilde{W}\right), A\left(\hat{W}\right) \geq \underline{A}(T) \vee I$. It follows that there exist $h \in \{1,...,k\}$ and $\hat{t} \in \mathbb{N}$ such that $\tilde{w}_{ih}^{(\hat{t})} > 0$ for all $i \in \{1,...,k\}$ and $\hat{t}$ can be chosen to be t_T. Since $\min_{i,j:\underline{a}_{ij}=1} w_{ij} \geq \delta$, $\min_{i\in\{1,...,k\}} \tilde{w}_{ih}^{(t_T)} \geq \tilde{\delta}^{t_T}$ where $\tilde{\delta} = \min\{\lambda, (1-\lambda)\delta\}$. By Seneta [34], this implies that

$$\|\gamma - \gamma_\beta\|_1 \leq \frac{1}{\tilde{\delta}^{t_T}} \left\| \tilde{W}^{t_T} - \hat{W}^{t_T} \right\|_\infty .$$

Since Since λ can be arbitrarily chosen, we then choose $\lambda = \delta/(1+\delta)$ so to maximize $\tilde{\delta}$, yielding that

$$\|\gamma - \gamma_\beta\|_1 \leq \frac{(1+\delta)^{t_T}}{\delta^{t_T}} \left\| \tilde{W}^{t_T} - \hat{W}^{t_T} \right\|_\infty . \tag{40}$$

Next, by induction and since the space of matrices endowed with $\| \ \|_\infty$ is a Banach algebra and $\left\|\bar{W}\right\|_\infty = 1$ for all stochastic matrices $\bar{W}$,[22] we have that

$$\left\| \tilde{W}^{t_T} - \hat{W}^{t_T} \right\|_\infty \leq t_T \left\| \tilde{W} - \hat{W} \right\|_\infty = \left(1 - \frac{\delta}{1+\delta}\right) t_T \left\| W - W_{\tilde{x}_\beta,\bar{h}} \right\|_\infty . \tag{41}$$

[22] First, recall that, given a $k \times k$ matrix E,

$$\|E\|_\infty = \max_{i\in\{1,...k\}} \sum_{j=1}^k |e_{ij}| .$$

In other words, $\|E\|_\infty$ is the dual norm of the operator $x \mapsto Ex$ when $\mathbb{R}^k$ is endowed with $\| \ \|_\infty$. For this reason, it can be computed by calculating the $\| \ \|_1$ of each row of E and then take the maximum of these values. By induction, we prove that

$$\left\| \tilde{W}^t - \hat{W}^t \right\|_\infty \leq t \left\| \tilde{W} - \hat{W} \right\|_\infty \quad \forall t \in \mathbb{N}.$$

The statement is trivial for $t = 1$. Assume it holds for t, we show it holds for $t+1$. Observe that

$$\begin{aligned}
\left\| \tilde{W}^{t+1} - \hat{W}^{t+1} \right\|_\infty &= \left\| \tilde{W}\left(\tilde{W}^t - \hat{W}^t\right) + \left(\tilde{W} - \hat{W}\right)\hat{W}^t \right\|_\infty \\
&\leq \left\| \tilde{W}\left(\tilde{W}^t - \hat{W}^t\right) \right\|_\infty + \left\| \left(\tilde{W} - \hat{W}\right)\hat{W}^t \right\|_\infty \\
&\leq \left\|\tilde{W}\right\|_\infty \left\| \tilde{W}^t - \hat{W}^t \right\|_\infty + \left\| \tilde{W} - \hat{W} \right\|_\infty \left\|\hat{W}^t\right\|_\infty \\
&= \left\| \tilde{W}^t - \hat{W}^t \right\|_\infty + \left\| \tilde{W} - \hat{W} \right\|_\infty \\
&\leq t\left\| \tilde{W} - \hat{W} \right\|_\infty + \left\| \tilde{W} - \hat{W} \right\|_\infty = (t+1)\left\| \tilde{W} - \hat{W} \right\|_\infty ,
\end{aligned}$$

the statement follows by induction.

Consider the i-th row of $W - W_{\tilde{x}_\beta,\bar{h}}$. By definition of $W_{\tilde{x}_\beta,\bar{h}}$ and since $\nabla T_i(he) = \nabla T_i(0)$ for all $h \in \mathbb{R}$, we have that the i-th row of $W - W_{\tilde{x}_\beta,\bar{h}}$ is equal to $\nabla T_i\left(\hat{h}e\right) - \nabla T_i\left(\lambda_{i,\tilde{x}_\beta,\bar{h}}\tilde{x}_\beta + \left(1 - \lambda_{i,\tilde{x}_\beta,\bar{h}}\right)\bar{h}e\right)$ where $\lambda_{i,\tilde{x}_\beta,\bar{h}} \in [0,1]$ and $\hat{h}$ we chose it to be an element of

$$\left[\min\left(\lambda_{i,\tilde{x}_\beta,\bar{h}}\tilde{x}_\beta + \left(1 - \lambda_{i,\tilde{x}_\beta,\bar{h}}\right)\bar{h}e\right), \max\left(\lambda_{i,\tilde{x}_\beta,\bar{h}}\tilde{x}_\beta + \left(1 - \lambda_{i,\tilde{x}_\beta,\bar{h}}\right)\bar{h}e\right)\right].$$

Since the Jacobian of T is Lipschitz continuous, we also have that

$$\begin{aligned}
&\left\|\nabla T_i\left(\hat{h}e\right) - \nabla T_i\left(\lambda_{i,\tilde{x}_\beta,\bar{h}}\tilde{x}_\beta + \left(1 - \lambda_{i,\tilde{x}_\beta,\bar{h}}\right)\bar{h}e\right)\right\|_1 \\
&\leq L\left\|\hat{h}e - \left(\lambda_{i,\tilde{x}_\beta,\bar{h}}\tilde{x}_\beta + \left(1 - \lambda_{i,\tilde{x}_\beta,\bar{h}}\right)\bar{h}e\right)\right\|_\infty.
\end{aligned}$$

By (33) and given our choice of $\hat{h}$, we can conclude that

$$\begin{aligned}
\left\|\hat{h}e - \left(\lambda_{i,\tilde{x}_\beta,\bar{h}}\tilde{x}_\beta + \left(1 - \lambda_{i,\tilde{x}_\beta,\bar{h}}\right)\bar{h}e\right)\right\|_\infty &= \left\|\left(\lambda_{i,\tilde{x}_\beta,\bar{h}}\tilde{x}_\beta + \left(1 - \lambda_{i,\tilde{x}_\beta,\bar{h}}\right)\bar{h}e\right) - \hat{h}e\right\|_\infty \\
&\leq \mathrm{Rg}\left(\lambda_{i,\tilde{x}_\beta,\bar{h}}\tilde{x}_\beta + \left(1 - \lambda_{i,\tilde{x}_\beta,\bar{h}}\right)\bar{h}e\right) \\
&\leq \lambda_{i,\tilde{x}_\beta,\bar{h}}\,\mathrm{Rg}\left(\tilde{x}_\beta\right) + \left(1 - \lambda_{i,\tilde{x}_\beta,\bar{h}}\right)\mathrm{Rg}\left(\bar{h}e\right) \\
&\leq \mathrm{Rg}\left(\tilde{x}_\beta\right)
\end{aligned}$$

By definition of $\left\|W - W_{\tilde{x}_\beta,\bar{h}}\right\|_\infty$ and (40) and (41) and since i was arbitrarily chosen, this implies that

$$\begin{aligned}
\left\|\gamma - \gamma_\beta\right\|_1 &\leq \frac{(1+\delta)^{t_T}}{\delta^{t_T}}\left\|\tilde{W}^{t_T} - \hat{W}^{t_T}\right\|_\infty \leq \frac{(1+\delta)^{t_T}}{\delta^{t_T}}\left(1 - \frac{\delta}{1+\delta}\right)t_T\left\|W - W_{\tilde{x}_\beta,\bar{h}}\right\|_\infty \\
&\leq \frac{(1+\delta)^{t_T}}{\delta^{t_T}}\left(1 - \frac{\delta}{1+\delta}\right)t_T L\,\mathrm{Rg}\left(\tilde{x}_\beta\right).
\end{aligned}$$

By points a and b and Theorem 2 an since $\mathrm{Rg}\left(\tilde{x}_\beta\right) \leq \mathrm{Rg}\left(x_\beta\right)$, we can conclude that

$$\begin{aligned}
\left\|x_\beta - \langle\gamma, x\rangle e\right\|_\infty &\leq \mathrm{Rg}\left(x_\beta\right) + \frac{(1+\delta)^{t_T}}{\delta^{t_T}}\left(1 - \frac{\delta}{1+\delta}\right)t_T L\left\|x\right\|_\infty \mathrm{Rg}\left(\tilde{x}_\beta\right) \\
&\leq (1-\beta)\left(1+\kappa_T\right)\left(1 + \frac{(1+\delta)^{t_T}}{\delta^{t_T}}\left(1 - \frac{\delta}{1+\delta}\right)t_T L\left\|x\right\|_\infty\right)\mathrm{Rg}\left(x\right),
\end{aligned}$$

proving the statement. ■

D.3 Discrete iterations

Lemma 10 *For all $x \in \mathbb{R}^k$, $\lim_t x^t$ exists if and only if $\lim_t \tilde{x}^t$ exists and, in this case, the two limits coincide.*

Proof. Fix $x \in \mathbb{R}^k$. First, we prove by induction that

$$\tilde{x}^t = (1-\beta_{t+1})\, x + \beta_{t+1} x^t \qquad \forall t \in \mathbb{N}_0. \tag{42}$$

For $t = 0$, we have $x_0 = \tilde{x}_0 = x$. Assume that (42) holds true for all $\tau \leq t$. Next, observe that

$$\tilde{x}^{t+1} = (1-\beta_{t+1})\, x + \beta_{t+1} T\left(\tilde{x}^t\right) = (1-\beta_{t+1})\, x + \beta_{t+1} T\left((1-\beta_{t+1})\, x + \beta_{t+1} x^t\right)$$

that is

$$\frac{\tilde{x}^{t+1} - (1-\beta_{t+1})\, x}{\beta_{t+1}} = T\left((1-\beta_{t+1})\, x + \beta_{t+1} x^t\right) = x^{t+1}$$

yielding (42) for $t+1$. Finally, given that $\beta_t \to 1$, (42) immediately yields the statement. ■

Lemma 11 *Let T be nonexpansive and normalized. If $s, l \in \mathbb{N}$ are such that $s \geq l$, then*

$$\|x_{s+m} - x_{\beta_l}\|_\infty \leq \beta_l^m \|x_s - x_{\beta_l}\|_\infty + 2\|x\|_\infty \sum_{r=1}^{m} \beta_l^{m-r} (\beta_{s+r} - \beta_l) \quad \forall m \in \mathbb{N}. \tag{43}$$

Proof. We start with a preliminary observation. By induction and since T is nonexpansive and normalized, it is obvious that $\|x^t\|_\infty \leq \|x\|_\infty$ for all $t \in \mathbb{N}$. Note that for each $l, t \in \mathbb{N}$

$$\begin{aligned}
\left\|x^{t+1} - x_{\beta_l}\right\|_\infty &= \left\|T\left((1-\beta_{t+1})\, x + \beta_{t+1} x^t\right) - T\left((1-\beta_l)\, x + \beta_l x_{\beta_l}\right)\right\|_\infty \\
&\leq \left\|T\left((1-\beta_{t+1})\, x + \beta_{t+1} x^t\right) - T\left((1-\beta_l)\, x + \beta_l x^t\right)\right\|_\infty \\
&+ \left\|T\left((1-\beta_l)\, x + \beta_l x^t\right) - T\left((1-\beta_l)\, x + \beta_l x_{\beta_l}\right)\right\|_\infty \\
&\leq \left\|(1-\beta_{t+1})\, x + \beta_{t+1} x^t - (1-\beta_l)\, x - \beta_l x^t\right\|_\infty \\
&+ \left\|(1-\beta_l)\, x + \beta_l x^t - (1-\beta_l)\, x - \beta_l x_{\beta_l}\right\|_\infty \\
&= \left\|(\beta_l - \beta_{t+1})\left(x - x^t\right)\right\|_\infty + \beta_l \left\|x^t - x_{\beta_l}\right\|_\infty \\
&\leq \beta_l \left\|x^t - x_{\beta_l}\right\|_\infty + |\beta_{t+1} - \beta_l| \left\|x - x^t\right\|_\infty \\
&\leq \beta_l \left\|x^t - x_{\beta_l}\right\|_\infty + 2\|x\|_\infty |\beta_{t+1} - \beta_l|.
\end{aligned}$$

Since $\{\beta_t\}_{t\in\mathbb{N}}$ is an increasing sequence, we have that for each $t \geq l$

$$\left\|x^{t+1} - x_{\beta_l}\right\|_\infty \leq \beta_l \left\|x^t - x_{\beta_l}\right\|_\infty + 2\|x\|_\infty (\beta_{t+1} - \beta_l). \tag{44}$$

We next prove (43) by induction. By (44) and setting $s = t$, the statement is true for $m = 1$. Assume (43) holds for m. We show it holds for $m+1$. By (44) and inductive hypothesis, we have that

$$\begin{aligned}\|x_{s+m+1}-x_{\beta_l}\|_\infty &\le \beta_l \|x_{s+m}-x_{\beta_l}\|_\infty + 2\|x\|_\infty (\beta_{s+m+1}-\beta_l)\\ &\le \beta_l^{m+1}\|x_s - x_{\beta_l}\|_\infty + 2\|x\|_\infty \sum_{r=1}^{m}\beta_l^{m+1-r}(\beta_{s+r}-\beta_l) + 2\|x\|_\infty(\beta_{s+m+1}-\beta_l)\\ &= \beta_l^{m+1}\|x_s - x_{\beta_l}\|_\infty + 2\|x\|_\infty \sum_{r=1}^{m+1}\beta_l^{m+1-r}(\beta_{s+r}-\beta_l),\end{aligned}$$

proving the inductive step. ■

Proposition 14 *Let T be nonexpansive and normalized and assume that $\lim_{\beta\to 1} x_\beta = \bar{x}$ is well defined for all $x \in \mathbb{R}^k$. If there exists $f : \mathbb{N} \to \mathbb{N}$ increasing and such that $\beta_l^{f(l)} \to 0$ as well as $\frac{\beta_{l+f(l)}-\beta_l}{1-\beta_l} \to 0$, then*

$$\lim_{\beta\to 1}\tilde{x}_\beta = \lim_{\beta\to 1} x_\beta = \lim_t x^t = \lim_t \tilde{x}^t \quad \forall x \in \mathbb{R}^k.$$

Proof. Fix $x \in \mathbb{R}^k$. By Lemma 11 and since $\{\beta_l\}_{l\in\mathbb{N}}$ is an increasing sequence, we have that

$$\begin{aligned}\left\|x_{l+f(l)}-x_{\beta_l}\right\|_\infty &\le \beta_l^{f(l)}\|x_l - x_{\beta_l}\|_\infty + 2\|x\|_\infty \sum_{r=1}^{f(l)}\beta_l^{f(l)-r}(\beta_{l+r}-\beta_l)\\ &\le 2\|x\|_\infty \beta_l^{f(l)} + 2\|x\|_\infty\left(\beta_{l+f(l)}-\beta_l\right)\sum_{r=1}^{f(l)}\beta_l^{f(l)-r}\\ &\le 2\|x\|_\infty\left(\beta_l^{f(l)} + \frac{\beta_{l+f(l)}-\beta_l}{1-\beta_l}\right) \qquad \forall l \in \mathbb{N},\end{aligned}$$

yielding that $\lim_l \left\|x_{l+f(l)}-x_{\beta_l}\right\|_\infty = 0$. This implies that for each $t \in \mathbb{N}$ such that $t \ge 1+f(1)$ there exists a unique $l_t \in \mathbb{N}$ such that $l_t + f(l_t) \le t < l_t + 1 + f(l_t+1)$ and $l_t \to \infty$ as $t \to \infty$.[23] Consider $t \ge 1+f(1)$. By (43) and since $l_t + 1 + f(l_t+1) > t \ge l_t + f(l_t)$ and $f(l_t) \ge 1$ for all $t \ge 1+f(1)$, if we set $s = l_t + f(l_t)$, $l = l_t + 1$, and

[23] Define N to be the set of all natural numbers $\ge 1+f(1)$. For each $t \in N$ define $D(t) = \{l \in \mathbb{N} : l + f(l) \le t\}$. It is immediate to see that if $t' \ge t$, then $D(t') \supseteq D(t) \ni 1$. At the same time, for each $t \ge 1+f(1)$ the set $D(t)$ is bounded above. If we define $g : N \to \mathbb{N}$ by $g(t) = \max D(t)$, then g is well defined. Moreover, we have that if $t' \ge t$, then $g(t') \ge g(t)$ as well as $g(l+f(l)) = l + f(l)$ for all $l \in \mathbb{N}$. This implies that l_t can be set to be equal to $g(t)$ and $\{l_t\}_{t\in\mathbb{N}} \subseteq \mathbb{N}$ is an increasing sequence which diverges.

$m = t - l_t - f(l_t)$, then we have that $s \geq l$ and

$$\begin{aligned}
&\left\|x^t - x_{\beta_{l_t+1}}\right\|_\infty \\
&\leq \begin{cases} \left\|x_{l_t+f(l_t)} - x_{\beta_{l_t+1}}\right\|_\infty & \text{if } t = l_t + f(l_t) \\ \beta_{l_t+1}^{t-l_t-f(l_t)} \left\|x_{l_t+f(l_t)} - x_{\beta_{l_t+1}}\right\|_\infty & \\ +2\|x\|_\infty \sum_{r=1}^{t-l_t-f(l_t)} \beta_{l_t+1}^{t-l_t-f(l_t)-r} \left(\beta_{l_t+f(l_t)+r} - \beta_{l_t+1}\right) & \text{if } t > l_t + f(l_t) \end{cases} \\
&\leq \begin{cases} \left\|x_{l_t+f(l_t)} - x_{\beta_{l_t+1}}\right\|_\infty & \text{if } t = l_t + f(l_t) \\ \left\|x_{l_t+f(l_t)} - x_{\beta_{l_t+1}}\right\|_\infty + 2\|x\|_\infty \frac{\beta_t - \beta_{l_t+1}}{1-\beta_{l_t+1}} & \text{if } t > l_t + f(l_t) \end{cases} \\
&\leq \begin{cases} \left\|x_{l_t+f(l_t)} - x_{\beta_{l_t+1}}\right\|_\infty & \text{if } t = l_t + f(l_t) \\ \left\|x_{l_t+f(l_t)} - x_{\beta_{l_t+1}}\right\|_\infty + 2\|x\|_\infty \frac{\beta_{l_t+1+f(l_t+1)} - \beta_{l_t+1}}{1-\beta_{l_t+1}} & \text{if } t > l_t + f(l_t) \end{cases}
\end{aligned}$$

Since t was arbitrarily chosen, we have that

$$\left\|x^t - x_{\beta_{l_t+1}}\right\|_\infty \leq \left\|x_{l_t+f(l_t)} - x_{\beta_{l_t+1}}\right\|_\infty + \varepsilon_t$$

where $\varepsilon_t = 2\|x\|_\infty \left(\beta_{l_t+1+f(l_t+1)} - \beta_{l_t+1}\right) / (1 - \beta_{l_t+1})$ for all $t \geq 1 + f(1)$. By assumption and since $l_t \to \infty$ as $t \to \infty$, it follows that $\varepsilon_t \to 0$ as $t \to \infty$. By Theorem 1, $\bar{x} \overset{\text{def}}{=} \lim_t x_{\beta_t} = \lim_t \tilde{x}_{\beta_t}$ is well defined. We have that for each $t \geq 1 + f(1)$

$$\begin{aligned}
\left\|x^t - \bar{x}\right\|_\infty &\leq \left\|x^t - x_{\beta_{l_t+1}}\right\|_\infty + \left\|x_{\beta_{l_t+1}} - \bar{x}\right\|_\infty \\
&\leq \left\|x_{l_t+f(l_t)} - x_{\beta_{l_t+1}}\right\|_\infty + \varepsilon_t + \left\|x_{\beta_{l_t+1}} - \bar{x}\right\|_\infty \\
&\leq \left\|x_{l_t+f(l_t)} - x_{\beta_{l_t}}\right\|_\infty + \left\|x_{\beta_{l_t}} - x_{\beta_{l_t+1}}\right\|_\infty + \varepsilon_t + \left\|x_{\beta_{l_t+1}} - \bar{x}\right\|_\infty \to 0,
\end{aligned}$$

proving the statement. ■

Proof of Proposition 7. We begin by making an observation on g. Since g is strictly increasing and concave, we have that $g(y) - g(x) \leq g'_+(x)(y - x)$ for all $y \in [1, \infty)$ and for all $x \in (1, \infty)$, yielding that

$$0 < g'_+(x) \leq (g(x) - g(1)) / (x - 1) \qquad \forall x \in (1, \infty). \tag{45}$$

In this case, the two conditions of Proposition 14 become:

1. $\left(1 - \frac{1}{g(l)}\right)^{f(l)} \to 0$ as $l \to \infty$;
2. $\frac{g(l+f(l)) - g(l)}{g(l+f(l))} = \frac{1 - \frac{1}{g(l+f(l))} - 1 + \frac{1}{g(l)}}{\frac{1}{g(l)}} = \frac{\beta_{l+f(l)} - \beta_l}{1 - \beta_l} \to 0$ as $l \to \infty$.

Define $f : \mathbb{N} \to \mathbb{N}$ by $f(1) = \left\lceil g(2) \frac{1}{g(2) - g(1)} \right\rceil$ and $f(l) = \left\lceil g(l) \left(\frac{l-1}{g(l) - g(1)}\right)^{\frac{1}{2}} \right\rceil$ for all $l \geq 2$. We make three observations:

a. Since g is concave and positive, it follows that, on $\mathbb{N}\backslash\{1\}$, $l \mapsto \frac{g(l)-g(1)}{l-1}$ is positive and decreasing and so is $l \mapsto \left(\frac{g(l)-1}{l-1}\right)^{\frac{1}{2}}$. This implies that $l \mapsto \left(\frac{l-1}{g(l)-1}\right)^{\frac{1}{2}}$ is well defined, positive, and increasing on $\mathbb{N}\backslash\{1\}$. Since g is positive and strictly increasing, $l \mapsto g(l)\left(\frac{l-1}{g(l)-g(1)}\right)^{\frac{1}{2}}$ is increasing on $\mathbb{N}\backslash\{1\}$ and so is f on $\mathbb{N}$.

b. Since g is divergent, we have that $\frac{g(l)-g(1)}{l-1} \sim \frac{g(l)}{l}$. Since $\frac{g(l)}{l} \to 0$ as $l \to \infty$ and g is positive, this implies that $\frac{f(l)}{g(l)} \geq g(l)\left(\frac{l-1}{g(l)-g(1)}\right)^{\frac{1}{2}}\frac{1}{g(l)} = \left(\frac{l-1}{g(l)-g(1)}\right)^{\frac{1}{2}} \to \infty$. Since $\frac{f(l)}{g(l)} \to \infty$ and for each $l \in \mathbb{N}\backslash\{1\}$

$$\left(1-\frac{1}{g(l)}\right)^{f(l)} = \left(\left(1-\frac{1}{g(l)}\right)^{g(l)}\right)^{\frac{f(l)}{g(l)}},$$

we can conclude that $\lim_l \left(1-\frac{1}{g(l)}\right)^{f(l)} = \left(e^{-1}\right)^{\infty} = 0$, yielding that condition 1 holds.

c. Since g is increasing, concave, and positive and f is positive, we have that

$$\begin{aligned} 0 &\leq \frac{g(l+f(l))-g(l)}{g(l+f(l))} \leq \frac{g'_+(l)f(l)}{g(l+f(l))} \leq \frac{g'_+(l)f(l)}{g(l)} \\ &\leq \frac{g'_+(l)}{g(l)}\left(g(l)\left(\frac{l-1}{g(l)-g(1)}\right)^{\frac{1}{2}}+1\right) = g'_+(l)\left(\frac{l-1}{g(l)-g(1)}\right)^{\frac{1}{2}} + \frac{g'_+(l)}{g(l)}. \end{aligned}$$

By (45) and since $\frac{g(l)-g(1)}{l-1} \sim \frac{g(l)}{l}$ and $\frac{g(l)}{l} \to 0$ as $l \to \infty$, we have that $g'_+(l) \to 0$ as $l \to 0$ and, in particular, $\frac{g'_+(l)}{g(l)} \to 0$. By (45) and since $\frac{g(l)-g(1)}{l-1} > 0$ for all $l \geq 2$ as well as $\frac{g(l)-g(1)}{l-1} \to 0$ as $l \to 0$, we have that

$$g'_+(l)\left(\frac{l-1}{g(l)-g(1)}\right)^{\frac{1}{2}} \leq \left(\frac{g(l)-g(1)}{l-1}\right)^{\frac{1}{2}} \to 0.$$

We can conclude that $\frac{g(l+f(l))-g(l)}{g(l+f(l))} \to 0$ as $l \to 0$, proving that condition 2 holds.

By Points a–c and Proposition 14, the statement follows. ■

Proof of Proposition 8. Fix $x \in \mathbb{R}^k$. (i) implies (ii). This immediately follows from Lemma 1. (ii) implies (iii). Define $\Psi : \mathbb{R}^k \to \mathbb{R}^k$ as

$$\Psi(z) = T(x+z) \qquad \forall z \in \mathbb{R}^k.$$

Observe that, for every $\beta \in [0,1)$, we have

$$x_\beta = T((1-\beta)x+\beta x_\beta) = (1-\beta)T\left(x+\frac{\beta}{1-\beta}x_\beta\right) = (1-\beta)\Psi\left(\frac{\beta}{1-\beta}x_\beta\right).$$

For all $\beta, \beta' \in [0,1)$ and for all $z \in \mathbb{R}^k$, we have

$$\left\| (1-\beta)\Psi\left(\frac{z}{1-\beta}\right) - (1-\beta')\Psi\left(\frac{z}{1-\beta'}\right) \right\|_\infty = \|T((1-\beta)x+z) - T((1-\beta')x+z)\|_\infty$$
$$\leq \|(\beta'-\beta)x\|_\infty = |\beta'-\beta| \, \|x\|_\infty .$$

This shows that Ψ satisfies Assumption 1 in [37]. We next show by induction that, for every $t \in \mathbb{N}$,

$$x^t = \frac{1}{t}\Psi^t(0). \tag{46}$$

First, observe that $x^1 = T(x) = \Psi(0)$, proving that (46) holds for $t = 1$. Next, assume that (46) holds for all $\tau \leq t$. We have

$$x^{t+1} = T\left(\frac{1}{t+1}x + \frac{t}{t+1}\frac{1}{t}\Psi^t(0)\right) = \frac{1}{t+1}T\left(x + \Psi^t(0)\right) = \frac{1}{t+1}\psi^{t+1}(0),$$

yielding (46) for $t+1$. With this, (iii) follows from [37, Theorem 1]. (iii) implies (iv). It follows by Lemma 10. (iv) implies (i). Lemma 10 implies that $\lim_t x^t = \lim_t \tilde{x}^t$ exists. By [37, Theorem 1], it follows that $\lim_{\beta\to 1} x_\beta$ exists. By Lemma 1 it follows that $\lim_{\beta\to 1} \tilde{x}_\beta = \lim_{\beta\to 1} x_\beta$ exists.

The second part of the proposition immediately follows by Lemmas 1 and 10. ■

E Appendix: Additional application on financial networks

In this section, we consider an equilibrium model of systemic risk where a group of financial institutions are exposed to idiosyncratic losses and hold cross-capital interdependencies. Following the approach of Adrian and Brunnermeier [3] in modeling systemic risk, we assume that the banks' expected losses conditional on each macroeconomic scenario are given by conditional risk measures (see also Detlefsen and Scandolo [18]).

Consider a finite number of banks $K = \{1, ..., |K|\}$ and states of the world $\Omega = \{1, ..., |\Omega|\}$. We interpret each state $\omega \in \Omega$ as a possible macroeconomic scenario that determines the nature of the idiosyncratic losses of the banks. Formally, fix $x \in \mathbb{R}^{K\times\Omega}$ and interpret $x_{k,\omega}$ as the realization of the idiosyncratic loss in real-economy assets for bank k in state ω. Each bank k is endowed with a partition $\Pi_k \subseteq 2^\Omega$ of the states representing the coarsened scenarios considered by bank k. In particular, we assume that Π_k is finer than the partition induced by the random variable $x_k \in \mathbb{R}^\Omega$ representing the idiosyncratic loss of k, that is, each bank is able to discern the scenarios determining their own losses in real-economy assets. Moreover, we assume that $\{\Pi_k\}_{k\in K}$ are common knowledge. Finally, the banks are connected in a financial network represented by a strongly connected stochastic matrix $M \in [0,1]^{K\times K}$ whose entries correspond to the financial interdependencies among the banks: $m_{k,k'} \in [0,1]$ is the exposure of bank k to bank k' and $m_{k,k} = 0$ for all $k \in K$.

Each bank k has to declare their expected total loss in each of the considered scenario, that is, for each cell of Π_k. The total loss of each bank is a combination of the realized idiosyncratic loss and the estimate loss induced by the exposures to the other banks. For every $k, k' \in K$, let $y_{k,k'} \in \mathbb{R}^\Omega$ denote the state-contingent loss of bank k' conjectured by bank k. In particular, $y_{k,k'}$ is a random variable that is measurable with respect to $\Pi_{k'}$. With this, the total conjectured loss of bank k for state ω is

$$(1-\beta)\, x_{k,\omega} + \beta \sum_{k'\in K} m_{k,k'} y_{k,k',\omega}, \tag{47}$$

where $\beta \in (0,1)$ captures the intensity of cross exposure of the banks.

The total conjectured loss in equation (47) is still a random variable from the point of view of bank k since each $y_{k,k'}$ is only measurable with respect to $\Pi_{k'}$. We endow each bank k with a *conditional risk measure* $V_k : \Omega \times \mathbb{R}^\Omega \to \mathbb{R}$ that quantifies each possible uncertain prospect in terms of monetary loss, conditional to each scenario considered by bank k. Following Detlefsen and Scandolo [18], we assume that each V_k is measurable with respect to Π_k and such that, for every $\omega \in \Omega$, the functional $V_k(\omega, \cdot)$ is normalized, monotone decreasing, convex, cash invariant, that is

$$V_k(\omega, \ell + ke) = V(\omega, \ell) - k \qquad \forall \omega \in \Omega, \forall \ell \in \mathbb{R}^\Omega, \forall k \in \mathbb{R},$$

and information regular, that is,

$$V_k\left(\omega, \ell 1_{\Pi(\omega)} + h 1_{\Pi(\omega)^c}\right) = V(\omega, \ell) \quad \forall \omega \in \Omega, \forall \ell, h \in \mathbb{R}^\Omega.$$

By [18, Theorem 1], this conditional risk measure admits the following representation

$$V_k(\omega, \ell) = \max_{p\in\Delta(\Omega)} \left\{ -\sum_{\tilde\omega\in\Omega} \ell_{\tilde\omega} p_{\tilde\omega} - c_{k,\omega}(p) \right\} \qquad \forall \omega \in \Omega, \forall \ell \in \mathbb{R}^\Omega,$$

where, for every $\omega \in \Omega$, the function $c_{k,\omega} : \Delta(\Omega) \to [0,\infty]$ is grounded, convex, lower semicontinuous, and such that $c_{k,\omega}(p) < \infty$ implies that $p \in \Delta(\Pi_k(\omega))$.

Given conjectures $\{y_{k,k'}\}_{k'\in K\setminus\{k\}}$, the risk of bank k in state ω is given by

$$\begin{aligned} &V_k\left(\omega, (1-\beta)\, x_k + \beta \sum_{k'\in K} m_{k,k'} y_{k,k'}\right) \\ &= -(1-\beta)\, x_{k,\omega} + \beta \max_{p\in\Delta(\Omega)} \left\{ -\sum_{(k',\omega')\in K\times\Omega} m_{k,k'} p_{\omega'} y_{k,k',\omega'} - \frac{1}{\beta} c_{k,\omega}(p) \right\}, \end{aligned}$$

where the equality follows from the fact that x_k is Π_k-measurable and information regularity.

In equilibrium, each bank has correct conjectures about the loss declared by all the other banks in every state. Formally, for every level of interconnectedness β, the vector of losses $x^{\beta} \in \mathbb{R}^{K \times \Omega}$ is an equilibrium if and only if

$$x^{\beta}_{k,\omega} = -V_k \left(\omega, (1-\beta) x_k + \beta \sum_{k' \in K} m_{k,k'} x^{\beta}_{k'} \right) \qquad \forall (k,\omega) \in K \times \Omega. \tag{48}$$

Fixed-point conditions such as the one in equation (48) are pervasive in equilibrium analysis of financial network (see for example the survey Jackson and Pernoud [27]). In particular, as $\beta \to 1$, the losses from financial interdependencies dominate the idiosyncratic losses from own real assets.

Next, define the concave operator $T : \mathbb{R}^{K \times \Omega} \to \mathbb{R}^{K \times \Omega}$ as

$$T_{(k,\omega)}(z) = \min_{p \in \Delta(\Omega)} \left\{ \sum_{(k',\omega') \in K \times \Omega} m_{k,k'} p_{\omega'} z_{k',\omega'} + c_{k,\omega}(p) \right\} \qquad \forall z \in \mathbb{R}^{K \times \Omega}.$$

Under the mild connectedness assumption that $E(T) = D \subseteq \mathbb{R}^{K \times \Omega}$, Corollary 2 implies that the limit risk $\lim_{\beta \to 1} x^{\beta}$ exists and is independent of the realized fundamental state as well as of the bank's identity.

This result has particularly strong implications for the case of *smooth divergence risk measures* with respect to a common ex-ante probabilistic model. Formally, we assume that the banks share the same *full support* probabilistic model $p^0 \in \Delta(\Omega)$ in the ex-ante stage and then update conditional on their private information. For example, bank k in state ω has interim belief $p^0(\cdot|\Pi_k(\omega))$. Therefore, in the interim stage, the conditional risk measure of bank k in state ω is

$$V_k(\omega, \ell) = \max_{p \in \Delta(\Omega)} \left\{ -\sum_{\tilde{\omega} \in \Omega} \ell_{\tilde{\omega}} p_{\tilde{\omega}} - D_k\left(p || p^0(\cdot|\Pi_k(\omega))\right) \right\} \qquad \forall \ell \in \mathbb{R}^{\Omega}$$

where $D_k(\cdot||\cdot) : \Delta(\Omega) \times \Delta(\Omega) \to [0,\infty]$ is a divergence that is essentially strictly convex (cf. Maccheroni et al. [30]). The standard example of such divergences is the relative entropy.

We are now ready for the main result of this section. Let $\mu \in \Delta(K)$ denote the unique left Perron-Frobenius eigenvector of M.

Corollary 6 *We have that*

$$\lim_{\beta \to 1} x^{\beta} = \left(\sum_{k \in K} \mu_k \left(\sum_{\omega \in \Omega} p^0_{\omega} x_{k,\omega} \right) \right) e.$$

This result follows by Corollary 3 and Golub and Morris [24, Proposition 3]. It shows that the limit equilibrium exists, is independent on the state-bank index, and coincides with convex linear combination of the ex-ante linear expectation of the banks' losses with weights given by the eigenvector centrality of the network. Therefore, as $\beta \to 1$, the losses declared by all the banks tend to ignore completely their concern for robustness converging to an aggregated probabilistic evaluation of the losses. This result is even more surprising when we observe that the concern for robustness, indexed by the divergences $D_k(\cdot||\cdot)$, can be heterogeneous across the banks.

The result has also important implications whenever the common ex-ante probabilistic model p^0 of the banks is highly misspecified. Indeed, suppose that the banks are aware of the possibility of misspecification and evaluate their losses with robust risk measures such as the divergence ones. Even in this case, high interconnectedness and equilibrium reasoning can offset the caution used in the evaluations and lead the banks to declare losses that become closer and closer to their original misspecified expectations.

References

[1] D. Acemoglu and P.D. Azar, Endogenous production networks, *Econometrica*, 88, 33-82, 2020.

[2] D. Acemoglu, V.M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi, The network origins of aggregate fluctuations, *Econometrica*, 80, 1977-2016, 2010.

[3] T. Adrian and M.K. Brunnermeier, CoVaR, *The American Economic Review*, 106, 1705, 2016.

[4] C. D. Aliprantis and K. C. Border, *Infinite Dimensional Analysis*, 3rd ed., Springer-Verlag, Berlin, 2006.

[5] N.I. Al-Najjar and E. Shmaya, Recursive utility and parameter uncertainty, *Journal of Economic Theory*, 181, 274-288, 2019.

[6] C. Ballester, A. Calvó-Armengol, and Y. Zenou, Who's who in networks. Wanted: The key player, *Econometrica*, 74, 1403–1417, 2006.

[7] A. Banerjee and O. Compte, Consensus and Disagreement: Information Aggregation under (not so) Naive Learning, National Bureau of Economic Research, 2022.

[8] D.R. Baqaee and E. Farhi, The macroeconomic impact of microeconomic shocks: Beyond Hulten's theorem, *Econometrica*, 87, 1155-1203, 2019.

[9] J. M. Borwein and J. D. Vanderwerff, *Convex Functions: Constructions, Characterizations, and Counterexamples*, Cambridge University Press, Cambridge, 2010.

[10] V.M. Carvalho and A. Tahbaz-Salehi, Production networks: A primer, *Annual Review of Economics*, 11, 2019.

[11] E. Castagnoli, G. Cattelan, F. Maccheroni, C. Tebaldi, and R. Wang, Star-shaped risk measures, *Operations Research*, forthcoming.

[12] S. Cerreia-Vioglio, R. Corrao, and G. Lanzani, Dynamic Opinion Aggregation: Long-run Stability and Disagreement, mimeo, 2021.

[13] S. Cerreia-Vioglio, P. Ghirardato, F. Maccheroni, M. Marinacci, and M. Siniscalchi, Rational preferences under ambiguity, *Economic Theory*, 48, 341–375, 2011.

[14] M. Chandrasekher, M. Frick, R. Iijima, and Y. Le Yaouanq, Dual-self representations of ambiguity preferences, *Econometrica*, forthcoming.

[15] F. H. Clarke, Generalized Gradients and Applications, *Transactions of the American Mathematical Society*, 205, 247–262, 1975.

[16] F. H. Clarke, *Optimization and Nonsmooth Analysis*, SIAM, Philadelphia, 1990.

[17] M. H. DeGroot, Reaching a consensus, *Journal of the American Statistical Association*, 69, 118–121, 1974.

[18] K. Detlefsen and G. Scandolo, Conditional and dynamic convex risk measures, *Finance and Stochastics*, 9, 539-561, 2005.

[19] N. E. Friedkin and E. C. Johnsen, Social Influence and Opinions, *The Journal of Mathematical Sociology*, 15, 193–205, 1990.

[20] P. Ghirardato and M. Siniscalchi, Ambiguity in the small and in the large, *Econometrica*, 80, 2827–2847, 2012.

[21] P. Ghirardato, F. Maccheroni, and M. Marinacci, Differentiating ambiguity and ambiguity attitude, *Journal of Economic Theory*, 118, 133–173, 2004.

[22] K. Goebel and W. A. Kirk, *Topics in Metric Fixed Point Theory*, Cambridge University Press, Cambridge, 1990.

[23] B. Golub and M. O. Jackson, Naïve learning in social networks and the wisdom of crowds, *American Economic Journal: Microeconomics*, 2, 112–149, 2010.

[24] B. Golub and S. Morris, Expectations, Networks, and Conventions, mimeo, 2018.

[25] B. Golub and E. Sadler, Learning in social networks, in *The Oxford Handbook of the Economics of Networks* (Y. Bramoullé, A. Galeotti, and B. Rogers, eds.), Oxford University Press, New York, 2016.

[26] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed., Cambridge University Press, Cambridge, 2013.

[27] M.O. Jackson, and A. Pernoud, Systemic risk in financial networks: A survey, *Annual Review of Economics*, 13, 171-202, 2021.

[28] A. Kopytov, K. Nimark, B. Mishra, and M. Taschereau-Dumouchel, Endogenous production networks under supply chain uncertainty, mimeo, 2022.

[29] M. Marinacci and L. Montrucchio, Unique solutions for stochastic recursive utilities, *Journal of Economic Theory*, 145, 1776-1804, 2010.

[30] F. Maccheroni, M. Marinacci, and A. Rustichini, Ambiguity aversion, robustness, and the variational representation of preferences, *Econometrica*, 74, 1447–1498, 2006.

[31] A. Rubinov and Z. Dzalilov, Abstract convexity of positively homogeneous functions, *Journal of Statistics and Management Systems*, 5, 1–20, 2002.

[32] E. Sadler and B. Golub, Games on endogenous networks, arXiv preprint arXiv:2102.01587, 2021.

[33] E. Seneta, *Non-Negative Matrices and Markov Chains*, 3rd ed., Springer-Verlag, New York, 2006.

[34] E. Seneta, Perturbation of the stationary distribution measured by ergodicity coefficients, *Advances in Applied Probability*, 20, 228–230, 1988.

[35] S. Sorin, *A first course on zero-sum repeated games*, Springer Science & Business Media, Berlin, 2002.

[36] M. Taschereau-Dumouchel, Cascades and fluctuations in an economy with an endogenous production network. Available at SSRN 3115854, 2020.

[37] B. Ziliotto, A Tauberian theorem for nonexpansive operators and applications to zero-sum stochastic games, *Mathematics of Operations Research*, 41, 1522-1534, 2016.

F Online appendix: omitted proofs

In this section, we first report the proofs of some of the secondary results in the main text that were omitted by the main appendix. Then, we report the proofs of the ancillary results stated in the main appendix and whose proofs were omitted.

F.1 Proofs of additional claims in the main text

Proof of Claim in Example 1. By the Mean Ergodic Theorem (see, e.g., Aliprantis and Border [4, Corollary 20.20]) and since W is a stochastic matrix, we have that $\lim_T \frac{1}{T+1} \sum_{t=0}^{T} W^t x$ exists for all $x \in \mathbb{R}^k$. By Hardy-Littlewood's Theorem and since $\{W^t x\}_{t\in\mathbb{N}_0}$ is norm-bounded by $\|x\|_\infty$, we have that

$$\lim_{\beta\to 1} \tilde{x}_{\beta,W} = \lim_{\beta\to 1} (1-\beta) \sum_{t=0}^{\infty} \beta^t W^t x = \lim_T \frac{1}{T+1} \sum_{t=0}^{T} W^t x \quad \forall x \in \mathbb{R}^k.$$

It is then routine to check that $\lim_{\beta\to 1} \tilde{x}_{\beta,W}$ is a fixed point of T.