Theory of Computing Systems (2019) 63:1531–1553
https://doi.org/10.1007/s00224-018-9891-0

Opinion Formation Games with Aggregation and Negative Influence

Markos Epitropou[1] · **Dimitris Fotakis**[2,3] · **Martin Hoefer**[4] · **Stratis Skoulakis**[3]

Published online: 18 October 2018

Abstract
In this paper, we study continuous opinion formation games with aggregation aspects. In many domains, expressed opinions of people are not only affected by local interaction and personal beliefs, but also by influences that stem from global properties of the opinions present in the society. To capture the interplay of such global and local effects, we propose a model of opinion formation games with aggregation, where we concentrate on the *average public opinion* as a natural way to represent a global trend in the society. While the average alone does not have good strategic properties as an aggregation rule, we show that with a limited influence of the average public opinion, the good properties of opinion formation models are preserved. More formally, we show that a unique equilibrium exists in average-oriented opinion formation games. Simultaneous best-response dynamics converge to within distance ε of equilibrium in $O(n^2 \ln(n/\varepsilon))$ rounds, even in a model with *outdated information* on the average public opinion. For the Price of Anarchy, we show an upper bound of $9/8 + o(1)$, almost matching the tight bound for games without aggregation. We prove some of the results in the context of a general class of opinion formation games with negative influences, and we extend our results to cases where expressed opinions must come from a restricted domain.

Keywords Opinion formation games · Opinion dynamics · Convergence · Price of anarchy

1 Introduction

The formation and dynamics of opinions are an important aspect in modern society and have been studied extensively for decades (see e.g., [16]). Opinion formation

This article is part of the Topical Collection on *Special Issue on Algorithmic Game Theory (SAGT 2017)*

✉ Stratis Skoulakis
sskoul@corelab.ntua.gr

Extended author information available on the last page of the article.

is based on information exchange, which is often *local* in the sense that socially connected people (e.g., family, friends, colleagues) interact more often and affect each other's opinion more strongly. Moreover, opinion formation is often *dynamic* in the sense that discussions and interactions lead to changes in the expressed opinions. With the advent of the internet and social media, local and dynamic aspects of opinion formation have become ever more dominant. To capture opinion formation on a formal level, several models have been proposed (see e.g., [4, 6, 9, 12, 13, 15] for continuous opinions and [5, 10, 21] for discrete ones). A common assumption, that dates back to DeGroot [9], is that opinions evolve through a form of repeated averaging of local information collected from the agent social neighborhoods.

1.1 Motivation and Opinion Formation Model

Our work builds on the influential model of Friedkin and Johnsen (FJ) [12] for continuous opinion formation and dynamics. In fact, we adopt the game-theoretic viewpoint of [6] to the FJ model. In the FJ model, each agent i holds an *intrinsic belief* $s_i \in [0, 1]$, which is private and invariant over time, and a *public opinion* $z_i \in [0, 1]$. Agent i selects her public opinion so as to minimize the total (weighted) disagreement of z_i to her intrinsic belief and to the public opinions in her social neighborhood. In a dynamic setting, the agents start with their beliefs and in each round $t \geq 1$, update their opinion $z_i(t)$ to the minimizer of their disagreement cost, given the opinions of the others in the previous round.

The FJ model is elegant, extensively studied, and has nice algorithmic properties. It admits a unique equilibrium [6, 12], i.e., a stable state where each agent's public opinion minimizes her disagreement cost, given the public opinions expressed by other agents, and the simultaneous best-response dynamics converges fast to it [13]. The efficiency of the equilibrium is quantified by the Price of Anarchy (PoA), which is the ratio of the total disagreement cost of the agents at equilibrium to the optimal total disagreement cost. For the FJ model, the PoA is $9/8$ for undirected social networks, and $\Omega(n)$ for general directed networks [6]. Moreover, tight PoA bounds can be obtained by an elegant local smoothness argument for both undirected [4] and directed [8] social networks.

Despite these favorable properties, the FJ model disregards influences from global properties of the public opinions, and also the nature of the dynamics of consensus formation. In many domains, public opinions are not only affected by local interaction and personal beliefs, as in e.g., [4, 6, 7, 9, 12, 13], but also by influences that stem from global properties of the opinions present in the society. People are getting exposed to global trends, societal norms, results from voting and polling, etc., which are usually interpreted as the consensus view of the society and may crucially affect opinion formation. Furthermore, groups of people (or networks of agents) often need to agree on a common action, even if their beliefs and/or their expressed opinions are totally different. This might happen e.g., when some networked devices need to implement a common action, when people vote over a set of alternatives, or when a wisdom-of-the-crowd opinion is formed in a social network. In similar situations, an *aggregation rule* maps the public opinions to a *global* opinion that represents the

consensus view on the issue at hand. E.g., in the FJ model, the global opinion might be the average or the median of the equilibrium opinions.

In presence of aggregation, the agents can also anticipate the impact of their public opinions on the global one and might incorporate it in their opinion selection. Then, the disagreement cost should also account for the distance of an agent's intrinsic belief to the global opinion. To address these issues, we consider a variant of the opinion formation game of [6, 12, 13] with opinion aggregation. Each agent i selects her opinion z_i so as to minimize:

$$C_i(\boldsymbol{z}) = w_i(s_i - z_i)^2 + \sum_{j \neq i} w_{ij}(z_j - z_i)^2 + \alpha_i(\mathrm{aggr}(\boldsymbol{z}) - s_i)^2 \ . \tag{1}$$

In (1), $\boldsymbol{z} = (z_1, \ldots, z_n) \in \mathbb{R}^n$ is the vector of public opinions, $s_i \in [0, 1]$ is the belief of agent i , and $\mathrm{aggr}(\boldsymbol{z})$ maps $\boldsymbol{z}$ to a global opinion $\mathrm{aggr}(\boldsymbol{z})$. The weights $w_{ij} \geq 0$ quantify how much the public opinion of agent j influences i, the weight $w_i > 0$ quantifies i's self-confidence, and the weight $\alpha_i > 0$ quantifies the appeal of the global opinion $\mathrm{aggr}(\boldsymbol{z})$ to i.

Motivated by previous work on the wisdom of the crowd (see e.g., [16, Sec. 8.3], [14]), we concentrate on *average-oriented* opinion formation games, where the aggregation rule $\mathrm{aggr}(\boldsymbol{z})$ maps $\boldsymbol{z}$ to its average $\mathrm{avg}(\boldsymbol{z}) = \sum_{j=1}^{n} z_j/n$. Then, the best response of each agent i to a public opinion vector $\boldsymbol{z}$ is:

$$z_i = \frac{\left(w_i + \frac{\alpha_i}{n}\right) s_i + \sum_{j \neq i} \left(w_{ij} - \frac{\alpha_i}{n^2}\right) z_j}{w_i + \frac{\alpha_i}{n^2} + \sum_{j \neq i} w_{ij}} \ . \tag{2}$$

1.2 Contribution

The aggregation rule in (1) might significantly affect both the dynamics and the equilibrium of opinion formation. In this work, we characterize to which extent the nice algorithmic properties of the FJ model are affected by aggregation effects.

A first challenge is evident in (2), where i's influence from some opinions z_j can be negative. Negative influence here models agent competition for dragging the average public opinion close to their intrinsic beliefs. Despite negative influence, we show that if agents admit a certain level of *self-confidence*, simultaneous best-response dynamics in *average-oriented* opinion formation games converges fast to the unique equilibrium of the game. We should highlight that assuming positive self-confidence is necessary for convergence (see e.g., [12, 13]) and that the convergence time decreases as the ratio of w_i to α_i and to $\sum_{j \neq i} w_{ij}$ increases. For clarity, we make the reasonable assumptions that $w_i \geq \alpha_i$ and that $w_i \geq \sum_{j \neq i} w_{ij}/(n-1)$. Namely, we assume that the self-confidence level of each agent is no less than her influence from the average public opinion and no less than her average influence from other agents (this is also consistent with the self-confidence level assumed in [13]). For this self-confidence level, we show (Lemma 1) that simultaneous best-response dynamics in average-oriented opinion formation games converges to the unique equilibrium within distance $\varepsilon > 0$ in $O(n^2 \ln(n/\varepsilon))$ rounds. We highlight that our analysis is general and provides an upper bound on the convergence rate of best

response dynamics in virtually all cases where convergence is guaranteed (see also the discussion in Section 2.1).

For this result, we assume that all agents have access to the average public opinion in each round. Since the average is global information and thus expensive to obtain in large networks, we consider average-oriented opinion dynamics with outdated information. Here the average public opinion is announced to all the agents simultaneously every few rounds (e.g., a polling agency publishes this information in a web page now and then). We prove (Theorem 1) that opinion dynamics with outdated information about the average converges to the unique equilibrium (with full information on the average) within distance $\varepsilon > 0$ after $O(n^2 \ln(n/\varepsilon))$ updates on the average. Both these results are proven for a more general setting with negative influence between the agents and with partially outdated information about the agent public opinions. We essentially prove that negative influence and outdated information do not introduce undesirable oscillating phenomena to opinion dynamics. Our proofs make use of matrix norm properties, which allow us to deal with negative influence between the agents and with the difficulties introduced by outdated information.

In Section 4, we bound the PoA of average-oriented opinion formation games. We restrict our attention to the most interesting case of symmetric games, where $w_{ij} = w_{ji} \geq 0$ for all agent pairs $i \neq j$, all agents have the same self-confidence w and the same influence α from the average. For nonsymmetric games the PoA is $\Omega(n)$, even without aggregation (i.e., when $\alpha = 0$, see [6, Fig. 2]). We show (Theorem 2) that the PoA is at most $9/8 + O(\alpha/(wn^2))$. In general, this bound cannot be improved since for $\alpha = 0$, $9/8$ is a tight bound for the PoA under the FJ model [6]. The proof builds on the elegant local smoothness approach of [4]. However, local smoothness cannot be directly applied to symmetric average-oriented games, because the function $(\text{avg}(z) - s_i)^2$ is not locally smooth. To overcome this difficulty, we carefully combine local smoothness with the fact that the average public opinion at equilibrium is equal to the average intrinsic belief, a consequence of symmetry (Proposition 1).

A frequent assumption in the literature on continuous opinion formation is that agent beliefs and opinions take values in a finite interval of non-negative real numbers. Then, by scaling, one can assume that beliefs and opinions lie in $[0, 1]$. Thus, we always assume that agent beliefs $s_i \in [0, 1]$. On the other hand, an important side-effect of negative influence is that the best-response (and equilibrium) opinions may become polarized and be pushed towards opposite directions, far away from $[0, 1]$. We believe that such opinion polarization is natural and should be allowed when negative influence is considered. Therefore, in Sections 3 and 4, we assume that the public opinions take arbitrary real values. Then, in Section 5, we consider *restricted* average-oriented games, where public opinions are restricted to $[0, 1]$, and study how convergence properties and the price of anarchy are affected.

Existence and uniqueness of equilibrium for restricted opinion formation games follow from [19]. We prove (Lemma 3 and Theorem 3) that the convergence rate of opinion dynamics with negative influence and with outdated information is not affected by restriction of public opinions to $[0, 1]$. The analysis of the convergence rate is similar to that for (unrestricted) opinion formation games. The only difference is a simple case analysis, in the final part of the proofs of Lemma 3 and Theorem 3,

which establishes that the distance of the restricted opinion vector to equilibrium decreases at least as fast as the corresponding distance in the unrestricted case.

For the PoA of restricted symmetric games, we consider the special case where $w = \alpha = 1$ and show that the PoA $\leq 3 + 2\sqrt{2} + O(\frac{1}{n})$ (Theorem 4). The main technical challenge in the PoA analysis of restricted games is that the local smoothness argument of Theorem 2 does not apply, because the function $(\text{avg}(z) - s_i)^2$ is not locally smooth and the average public opinion at equilibrium may be far from the average intrinsic belief. Hence, in the proof of Theorem 4, we need to advance substantially beyond the local smoothness argument of Theorem 2. More specifically, we first show that if all agents only value the distance of their opinion to the average and to their belief (i.e., if $w_{ij} = 0$ for all $i \neq j$), the PoA is at most $1 + 1/n^2$ (Proposition 6). Then we combine the PoA of this simpler game with the local smoothness inequality of [4] and bound the PoA of restricted symmetric games.

Clearly, there are many alternative ways to model aggregation, which offer interesting directions for future research. For example, a possible aggregation is the *median* instead of the average. The median aggregation rule is prominent in Social Choice (see e.g., [3, 17]). However, it turns out that the FJ model with median aggregation has significantly less favorable properties. There are examples where median-oriented games lack exact equilibria (and, hence, convergence of best-response dynamics), but they can be shown to have approximate equilibria. A study of the median rule is beyond the scope of this paper.

1.3 Further Related Work

To the best of our knowledge, this is the first work to analyze the convergence of simultaneous best-response dynamics of the FJ model with negative influence and outdated information, or the price of anarchy of the FJ model with average opinion aggregation. However, there is some recent work on properties of opinion formation either with global information, or with negative influence, or where consensus is sought. We concentrate here on related previous work most relevant to ours. Discrete opinion formation is considered in [11] in the binary voter model, where each agent i has a certain probability of adopting the opinion of an agent outside i's local neighborhood (this is conceptually equivalent to estimating the average opinion with random sampling). The authors analyze the convergence time and the probability that consensus is reached. In [18], the authors provide a formal model and analyze continuous opinion formation, based on the bounded-confidence model of Deffuant-Weisbuch, in a population with individuals and opinion leaders (i.e., media), where the latter can be regarded opinion aggregators. They mostly investigate conditions under which consensus, polarization, or opinion fragmentation is reached. Games with only two opposite opinions in the network are studied in [1], where necessary and sufficient conditions are derived under which local interaction in social networks with positive and negative influence reaches consensus. Recently, a model of discrete opinion formation was introduced in [2] with generalized social relations, which include positive and negative influence. The authors show that generalized discrete opinion formation games admit a potential function, and thus, best-response dynamics converge to a pure Nash equilibrium.

2 Model and Preliminaries

We define $[n] \equiv \{1, \ldots, n\}$. For a vector $\boldsymbol{z}$, z_i denotes the i-th coordinate of $\boldsymbol{z}$, $\boldsymbol{z}_{-i}$ is $\boldsymbol{z}$ without its i-th coordinate, and $(z, \boldsymbol{z}_{-i})$ is the vector obtained from $\boldsymbol{z}$ if we replace z_i with z. For a vector $\boldsymbol{z}$ (resp. matrix A), $\boldsymbol{z}^T$ (resp. A^T) denotes the transpose. We define $\mathbf{0} \equiv (0, \ldots, 0)$ and $\mathbb{I}$ as the $n \times n$ identity matrix. We use capital letters for matrices and lowercase letters for their elements, with the understanding that a_{ij} is the (i, j) element of a matrix A.

For an $n \times n$ matrix A, $\|A\| = \max_{i\in[n]} \sum_{j=1}^{n} |a_{ij}|$ is the infinity norm of A. Similarly, for an n-dimensional vector $\boldsymbol{z}$, $\|\boldsymbol{z}\| = \max_{i\in[n]} |z_i|$ is the infinity norm of $\boldsymbol{z}$. We use the standard properties of matrix norms without explicitly referring to them. Specifically, we use that (i) for any matrices A and B, $\|AB\| \leq \|A\| \, \|B\|$ and $\|A + B\| \leq \|A\| + \|B\|$; (ii) for any matrix A and any $\lambda \in \mathbb{R}$, $\|\lambda A\| \leq |\lambda| \, \|A\|$; and (iii) for any matrix A and any integer ℓ, $\|A^\ell\| \leq \|A\|^\ell$. Moreover, we use that for any $n \times n$ real matrix A with $\|A\| < 1$, $\sum_{\ell=0}^{\infty} A^\ell = (\mathbb{I} - A)^{-1}$.

2.1 Average-Oriented Opinion Formation

We consider average-oriented opinion formation games with n agents. The model, the agent cost functions and their best response are introduced in Section 1. Next, we give some definitions and state some useful facts.

Without loss of generality, we assume that the vector of agent beliefs $\boldsymbol{s}$ lies in $[0, 1]^n$. As for the public opinions $\boldsymbol{z}$, we initially assume values in $\mathbb{R}$ and then, in Section 5, explain what changes if we restrict them to $[0, 1]$. An average-oriented opinion formation game $\mathcal{G}$ is *symmetric* if $w_{ij} = w_{ji}$ for all $i \neq j$, and $w_i = w$ and $\alpha_i = \alpha$ for all $i \in N$. $\mathcal{G}$ is *nonsymmetric* otherwise. If the game is symmetric, we let $w = 1$, by scaling other weights accordingly. The convergence results hold for nonsymmetric games, while the PoA bounds hold only for symmetric ones.

A vector $\boldsymbol{z}^*$ is an *equilibrium* of an opinion formation game $\mathcal{G}$ if for any agent i and any opinion z, $C_i(\boldsymbol{z}^*) \leq C_i(z, \boldsymbol{z}^*_{-i})$, i.e., the agents cannot improve on their individual cost at $\boldsymbol{z}^*$ by unilaterally changing their opinions. The *social cost* $C(\boldsymbol{z})$ of $\mathcal{G}$ is $C(\boldsymbol{z}) = \sum_{i\in N} C_i(\boldsymbol{z})$. An opinion vector $\boldsymbol{o}$ is *optimal* if for any $\boldsymbol{z}$, $C(\boldsymbol{o}) \leq C(\boldsymbol{z})$. The *Price of Anarchy* of $\mathcal{G}$, or $\mathrm{PoA}(\mathcal{G})$, is $C(\boldsymbol{z}^*)/C(\boldsymbol{o})$, where $\boldsymbol{z}^*$ is the unique equilibrium of $\mathcal{G}$ and $\boldsymbol{o}$ is an optimal vector.

To study the convergence properties of simultaneous best-response dynamics, it is convenient to write (2) in matrix form. Let $S_i = w_i + \frac{\alpha_i}{n^2} + \sum_{j\neq i} w_{ij}$. We define two $n \times n$ matrices A and B. Matrix A has $a_{ii} = 0$, for all $i \in N$, and $a_{ij} = (w_{ij} - \frac{\alpha_i}{n^2})/S_i$, for all $j \neq i$. Matrix B is diagonal and has $b_{ii} = (w_i + \frac{\alpha_i}{n})/S_i$, for all $i \in N$, and $b_{ij} = 0$, for all $j \neq i$. We always assume that $S_i > 0$. This assumption is required so that $\frac{d^2 C_i(\boldsymbol{z})}{dz_i^2} = 2S_i > 0$ and the function $C_i(\boldsymbol{z})$ is strictly convex in z_i, even if some entries of A are negative.

Assuming that agent self-confidence levels w_i are positive is necessary for convergence of opinion formation games (if we do not make any further assumptions on matrix A, e.g., consider an opinion formation game where $w_i = \alpha_i = 0$ for all agents i and the matrix A corresponds to the adjacency matrix of a bipartite

network). Similarly to [13] and for clarity, we assume that $w_i \geq \sum_{j \neq i} w_{ij}/(n-1)$ for all agents i. Namely, we assume that the self-confidence level of each agent is at least as large as her average influence from other agents. Moreover, we assume that for any agent i, $w_i \geq \alpha_i$, i.e., that the self-confidence level of any agent is no less than her influence from the average public opinion. These two assumptions immediately imply that $\alpha_i \leq S_i \leq (n + \frac{1}{n^2})w_i$, for any agent i. Using this inequality on S_i, we obtain the following inequality, which allows for a quasi-quadratic upper bound on the convergence rate of best-response dynamics:

$$\|A\| \leq \frac{S_i - w_i + \frac{\alpha_i(n-2)}{n^2}}{S_i} \leq \frac{S_i - S_i\frac{n^2}{n^3+1} + S_i\frac{(n-2)}{n^2}}{S_i} \leq 1 - \frac{2}{n^2} + \frac{1}{n^3} \leq 1 - \frac{1}{n^2}\,. \quad (3)$$

We use (3) in Corollary 1 and Corollary 2 and show that the best response dynamics converges to equilibrium within distance $\varepsilon > 0$ in $O(n^2 \ln(n/\varepsilon))$ rounds. However, our analysis of the convergence rate is general and can be applied under the weaker assumption that $\|A\| < 1$. Then, the convergence time depends on $1 - \|A\|$ (see also Lemma 1 and Theorem 1).

We usually refer to matrices similar to A, i.e., with infinity norm less than 1 and 0s in their diagonal, as *influence* matrices, and to matrices similar to B, i.e., to diagonal matrices with positive elements, as *self-confidence* matrices.

The simultaneous best-response dynamics of an average-oriented game $\mathcal{G}$ starts with $z(0) = s$ and proceeds in rounds. In each round $t \geq 1$, the public opinion vector $z(t)$ is:

$$z(t) = Az(t-1) + Bs\,. \quad (4)$$

We usually refer to (4) and to similar equations as *opinion formation processes*. We say that an opinion formation process $\{z(t)\}_{t\in\mathbb{N}}$ *converges* to a stable state z^* if for all $\varepsilon > 0$, there is a $t^*(\varepsilon)$, such that for all $t \geq t^*(\varepsilon)$, $\|z(t) - z^*\| \leq \varepsilon$. Iterating (4) over t (see also [13, Sec. 2]), we obtain that for all rounds $t \geq 1$,

$$z(t) = Az(t-1) + Bs = A(Az(t-2) + Bs) + Bs = \cdots = A^t s + \sum_{\ell=0}^{t-1} A^\ell Bs\,. \quad (5)$$

2.2 Average-Oriented Opinion Formation with Outdated Information

We study opinion formation when the agents have outdated information about the average public opinion. We assume an infinite increasing sequence of rounds $0 = \tau_0 < \tau_1 < \tau_2 < \cdots$ that describes an *update schedule* for the average opinion. At the end of round τ_p, the average $\text{avg}(z(\tau_p))$ is announced to the agents. We refer to the rounds between two updates as an *epoch*. Specifically, the rounds $\tau_p + 1, \ldots, \tau_{p+1}$ comprise epoch p. We assume that the length of each epoch p, denoted by $k_p = \tau_{p+1} - \tau_p \geq 1$, is finite. The update schedule is the same for all agents, but the agents do not need to have any information about it. They only need to be aware of the most recent value of the average public opinion provided to them.

We now need to distinguish in (2) and (4) between the influence from social neighbors, for which the most recent opinions $z(t-1)$ are used, and the influence from the average public opinion, where possibly outdated information is used. As such, we

now rely on three different $n \times n$ matrices D, E and B. Self-confidence matrix B is defined as before. Influence matrix D has $d_{ii} = 0$, for all $i \in N$, and $d_{ij} = w_{ij}/S_i$, for all $j \neq i$, and accounts for the influence from social neighbors. Influence matrix E has $e_{ii} = 0$, for all $i \in N$, and $e_{ij} = -\alpha_i/(n^2 S_i)$, for all $j \neq i$, and accounts for the influence from the average public opinion. By definition, $A = D + E$. Moreover, $\|D\| \leq 1 - 1/n$ and that $\|E\| \leq (n-1)/n^2$.

At the beginning of the opinion formation process, $\boldsymbol{z}(0) = \boldsymbol{s}$. For each round t in epoch p, $\tau_p + 1 \leq t \leq \tau_{p+1}$, the agent opinions are updated according to:

$$\boldsymbol{z}(t) = D\boldsymbol{z}(t-1) + E\boldsymbol{z}(\tau_p) + B\boldsymbol{s} \tag{6}$$

We note that at the beginning of each epoch p, every agent i can subtract $z_i(\tau_p)$ from $n \operatorname{avg}(\boldsymbol{z}(\tau_p))$ and compute $E\boldsymbol{z}(\tau_p)$, which is required in (6), as $-\frac{\alpha_i}{n^2 S_i}(n \operatorname{avg}(\boldsymbol{z}(\tau_p)) - z_i(\tau_p))$.

2.3 Opinion Formation with Negative Influence

An interesting aspect of average-oriented games is that the influence matrix A may contain negative elements. Motivated by this observation, we prove our convergence results for a general domain of opinion formation games that may have negative weights w_{ij}. Similar to [6, 13], the individual cost function of each agent i is $C_i(\boldsymbol{z}) = w_i(z_i - s_i)^2 + \sum_{j \neq i} w_{ij}(z_i - z_j)^2$, and i's best response to $\boldsymbol{z}_{-i}$ is

$$z_i = \frac{w_i s_i + \sum_{j \neq i} w_{ij} z_j}{w_i + \sum_{j \neq i} w_{ij}} \ . \tag{7}$$

The important difference is that now some w_{ij} may be negative. We require that for each agent i, $w_i > 0$ and $S_i = w_i + \sum_{j \neq i} w_{ij} > 0$ (and thus, $C_i(\boldsymbol{z})$ is strictly convex in z_i). The matrices A and B are defined as before. Namely, $a_{ij} = w_{ij}/S_i$, for all $i \neq j$, and B has $b_{ii} = w_i/S_i$ for all i. We always require that $\|A\| < 1 - \beta$, for some $\beta > 0$ (β may depend on n). Simultaneous best-response dynamics is again defined by (4).

3 Convergence of Average-Oriented Opinion Formation

For any nonnegative influence matrix A with $\|A\| \leq 1 - \beta$, (4) converges to the equilibrium point $\boldsymbol{z}^* = (\mathbb{I} - A)^{-1} B\boldsymbol{s}$ within distance ε in $O(\ln(\frac{\|B\|}{\varepsilon\beta})/\beta)$ rounds, as shown in [13, Lemma 3]. The following lemma shows that the same convergence rate holds for average-oriented opinion formation games, where A may contain negative elements. The proof is very similar to the proof of [13, Lemma 3] and we include it for completeness. The only minor difference is that the proof of Lemma 1 uses the infinity norm of A, instead of the largest eigenvalue of A in [13]. Using the infinity norm of A allows for a direct generalization of Lemma 1 to the case of average-oriented opinion formation games with outdated information.

Lemma 1 *Let A be any influence matrix, possibly with negative elements, such that $\|A\| \leq 1-\beta$, for some $\beta > 0$. Then, for any self-confidence matrix B, any $\boldsymbol{s} \in [0,1]^n$ and any $\varepsilon > 0$, the opinion formation process $\boldsymbol{z}(t) = A\boldsymbol{z}(t-1) + B\boldsymbol{s}$ converges to $\boldsymbol{z}^* = (\mathbb{I} - A)^{-1}B\boldsymbol{s}$ within distance ε in $O(\ln(\frac{\|B\|}{\varepsilon\beta})/\beta)$ rounds.*

Proof By (5), we have that for any $t \geq 1$, $\boldsymbol{z}(t) = A^t\boldsymbol{s} + \sum_{\ell=0}^{t-1} A^\ell B\boldsymbol{s}$. Since $\|A\| \leq 1-\beta$, $\|A^t\| \leq (1-\beta)^t$. Therefore, $\lim_{t\to\infty} A^t\boldsymbol{s} = \boldsymbol{0}$ and (5) converges to $\boldsymbol{z}^* = \sum_{\ell=0}^{\infty} A^\ell B\boldsymbol{s}$. Using the identity $\sum_{\ell=0}^{\infty} A^\ell = (\mathbb{I} - A)^{-1}$, we obtain that $\boldsymbol{z}^* = (\mathbb{I} - A)^{-1}B\boldsymbol{s}$. We note that since $\|A\| < 1$, the matrix $\mathbb{I} - A$ is strictly diagonally dominant and thus non-singular. Moreover,

$$\|(\mathbb{I} - A)^{-1}\| \leq \sum_{\ell=0}^{\infty} \|A^\ell\| \leq \sum_{\ell=0}^{\infty} (1-\beta)^\ell = 1/\beta\,.$$

To bound the convergence time to $\boldsymbol{z}^*$, we define $e(t) = \|\boldsymbol{z}(t) - \boldsymbol{z}^*\| = \max_{i\in N} |z_i(t) - z_i^*|$ as the distance of the opinions at time t to equilibrium. We next show that $e(t)$ is decreasing in t and obtain an upper bound on $t^*(\varepsilon) = \min\{t : e(t) \leq \varepsilon\}$. We observe that for any $t \geq 1$,

$$\begin{aligned} e(t) &= \|\boldsymbol{z}(t) - \boldsymbol{z}^*\| \\ &= \|A\boldsymbol{z}(t-1) + B\boldsymbol{s} - A\boldsymbol{z}^* - B\boldsymbol{s}\| \\ &\leq \|A\|\,\|\boldsymbol{z}(t-1) - \boldsymbol{z}^*\| \\ &\leq (1-\beta)e(t-1) \leq (1-\beta)^t e(0)\,. \end{aligned}$$

Since $\boldsymbol{s} \in [0,1]^n$ and $\|(\mathbb{I} - A)^{-1}\| \leq 1/\beta$, we obtain that

$$\|\boldsymbol{z}^*\| \leq \|(\mathbb{I} - A)^{-1}B\boldsymbol{s}\| \leq \|(\mathbb{I} - A)^{-1}\|\,\|B\|\,\|\boldsymbol{s}\| \leq \|B\|/\beta\,.$$

Since $\boldsymbol{z}(0) = \boldsymbol{s}$, we have that $e(0) = \|\boldsymbol{s} - \boldsymbol{z}^*\| \leq 1 + \|B\|/\beta$. Hence, $t^*(\varepsilon) = O(\ln(\frac{\|B\|}{\varepsilon\beta})/\beta)$. □

Since $\mathbb{I} - A$ is nonsingular, $\boldsymbol{z}^*$ is the unique opinion vector that satisfies $\boldsymbol{z}^* = A\boldsymbol{z}^* + B\boldsymbol{s}$. Thus, $\boldsymbol{z}^*$ is the unique equilibrium of the corresponding opinion formation game. Moreover, since for average-oriented games $\|A\| \leq 1 - 1/n^2$, Lemma 1 implies the following:

Corollary 1 *Any average-oriented game admits a unique equilibrium $\boldsymbol{z}^* = (\mathbb{I} - A)^{-1}B\boldsymbol{s}$, and for any $\varepsilon > 0$, (4) converges to $\boldsymbol{z}^*$ within distance ε in $O(n^2 \ln(n/\varepsilon))$ rounds.*

3.1 Convergence with Outdated Information

Next, we extend Lemma 1 to the case where the agents use possibly outdated information about the average public opinion in each round. More generally, we establish convergence for a general domain with negative influence between the agents, which includes average-oriented opinion formation processes as a special case.

Theorem 1 *Let D and E be influence matrices, possibly with negative elements, such that $\|D\| \leq 1-\beta_1$, $\|E\| \leq 1-\beta_2$, for some $\beta_1, \beta_2 \in (0,1)$ with $\beta_1+\beta_2 > 1$. Then, for any self-confidence matrix B, any $\boldsymbol{s} \in [0,1]^n$, any update schedule $0 = \tau_0 < \tau_1 < \tau_2 < \cdots$ and any $\varepsilon > 0$, the opinion formation process (6) converges to $\boldsymbol{z}^* = (\mathbb{I}-(D+E))^{-1}B\boldsymbol{s}$ within distance ε in $O(\ln(\frac{\|B\|}{\varepsilon\beta})/\beta)$ epochs, where $\beta = \beta_1+\beta_2-1 > 0$.*

Proof We observe that $\boldsymbol{z}^* = (\mathbb{I}-(D+E))^{-1}B\boldsymbol{s}$ is the unique solution of $\boldsymbol{z}^* = D\boldsymbol{z}^* + E\boldsymbol{z}^* + B\boldsymbol{s}$ (as in Lemma 1, since $\|E+D\| \leq 1-\beta$, with $\beta > 0$, the matrix $\mathbb{I}-(D+E)$ is non-singular). Hence, if (6) converges, it converges to $\boldsymbol{z}^*$. To show convergence, we bound the distance of $\boldsymbol{z}(t)$ to $\boldsymbol{z}^*$ by a decreasing function of t and show an upper bound on $t^*(\varepsilon) = \min\{t : e(t) \leq \varepsilon\}$.

As in the proof of Lemma 1, for each round $t \geq 1$, we define $e(t) = \|\boldsymbol{z}(t) - \boldsymbol{z}^*\|$ as the distance of the opinions at time t to $\boldsymbol{z}^*$. For convenience, we also define

$$f(\beta_1,\beta_2,k) = (1-\beta_1)^k + (1-\beta_2)\frac{1-(1-\beta_1)^k}{\beta_1}\,.$$

For any fixed value of $\beta_1, \beta_2 \in (0,1)$ with $\beta_1+\beta_2 > 1$, $f(\beta_1,\beta_2,k)$ is a decreasing function of k. Actually, the derivative of f with respect to k is equal to $\ln(1-\beta_1)(1-\beta_1)^k(1-\frac{1-\beta_2}{\beta_1})$, which is negative, because $1 > (1-\beta_2)/\beta_1$, since $\beta_1+\beta_2 > 1$.

We next show that:

Claim (i). For any epoch $p \geq 0$ and any round k, $0 \leq k \leq k_p$, in epoch p,

$$e(\tau_p + k) \leq f(\beta_1,\beta_2,k)e(\tau_p)\,.$$

Claim (ii). In the last round $\tau_{p+1} = \tau_p + k_p$ of each epoch $p \geq 0$, $e(\tau_{p+1}) \leq (1-\beta)e(\tau_p)$.

Claim (i) shows that the distance to equilibrium decreases from each round to the next within each epoch, while Claim (ii) shows that the distance to equilibrium decreases geometrically from the last round of each epoch to the last round of the next epoch. Combining Claim (i) and Claim (ii), we obtain that for any epoch $p \geq 0$ and any round k, $0 \leq k \leq k_p$, in epoch p, $e(\tau_p+k) \leq f(\beta_1,\beta_2,k)(1-\beta)^p e(0)$. Therefore, for any update schedule $\tau_0 < \tau_1 < \tau_2 < \cdots$, the opinion formation process (6) converges to $(\mathbb{I}-(D+E))^{-1}B\boldsymbol{s}$ in $O(\ln(e(0)/\varepsilon)/\beta)$ epochs.

To prove Claim (i), we fix any epoch $p \geq 0$ and apply induction on k. The basis, where $k = 0$, holds because $f(\beta_1,\beta_2,0) = 1$. For any round k, with $1 \leq k \leq k_p$, in p, we have that:

$$\begin{aligned}
e(\tau_p+k) &= \|D\boldsymbol{z}(\tau_p+k-1) + E\boldsymbol{z}(\tau_p) + B\boldsymbol{s} - (D\boldsymbol{z}^* + E\boldsymbol{z}^* + B\boldsymbol{s})\| \\
&\leq \|D\|\,\|\boldsymbol{z}(\tau_p+k-1) - \boldsymbol{z}^*\| + \|E\|\,\|\boldsymbol{z}(\tau_p) - \boldsymbol{z}^*\| \\
&\leq (1-\beta_1)e(\tau_p+k-1) + (1-\beta_2)e(\tau_p) \\
&\leq (1-\beta_1)f(\beta_1,\beta_2,k-1)e(\tau_p) + (1-\beta_2)e(\tau_p) = f(\beta_1,\beta_2,k)e(\tau_p)\,.
\end{aligned}$$

The first inequality follows from the properties of matrix norms. The second inequality holds because $\|D\| \leq 1-\beta_1$ and $\|E\| \leq 1-\beta_2$. The third inequality follows from the induction hypothesis. Finally, we use that for any $k \geq 1$, $(1-\beta_1)f(\beta_1,\beta_2,k-1) + 1 - \beta_2 = f(\beta_1,\beta_2,k)$.

To prove Claim (ii), we fix any epoch $p \geq 0$ and apply claim (i) to the last round $\tau_{p+1} = \tau_p + k_p$, with $k_p \geq 1$, of epoch p. Hence, $e(\tau_{p+1}) = \|z(\tau_p + k_p) - z^*\| \leq f(\beta_1, \beta_2, k_p)e(\tau_p)$.

We next show that $f(\beta_1, \beta_2, k_p) \leq 2 - (\beta_1 + \beta_2) = 1 - \beta$, which concludes the proof of the claim. The inequality holds because for any integer $k \geq 1$, $f(\beta_1, \beta_2, k)$ is a convex function of β_1. For a formal proof, we fix any $k \geq 1$ and any $\beta_2 \in (0, 1)$, and consider the functions $g(x) = (1-x)^k + \frac{1-(1-x)^k}{x}(1-\beta_2)$ and $h(x) = 2-\beta_2-x$, where $x \in [1 - \beta_2, 1]$ (since we assume that $\beta_1 \in (0, 1)$ and that $\beta_1 > 1 - \beta_2$). For any fixed value of $\beta_2 \in (0, 1)$, $h(x)$ is a linear function of x with $h(1 - \beta_2) = 1$ and $h(1) = 1 - \beta_2$. For any fixed value of $k \geq 1$ and $\beta_2 \in (0, 1)$, $g(x)$ is a convex function of x with $g(1-\beta_2) = 1 = h(1-\beta_2)$ and $g(1) = 1-\beta_2 = h(1)$. Therefore, for any $\beta_1 \in [1 - \beta_2, 1]$, $g(\beta_1) \leq h(\beta_1) = 2 - (\beta_1 + \beta_2)$.

To obtain an upper bound on $e(0) = \|s-z^*\|$, we work as in the proof of Lemma 1, using the fact that $\|D+E\| \leq 1-\beta$, and show first that $\|(\mathbb{I}-(D+E))^{-1}\| \leq 1/\beta$ and then that $\|z^*\| \leq \|B\|/\beta$. Since $z(0) = s$, we have that $e(0) = \|s-z^*\| \leq 1+\|B\|/\beta$. Using the fact that for each epoch $p \geq 0$ and for every round k, $0 \leq k \leq k_p$, in p, $e(\tau_p + k) \leq f(\beta_1, \beta_2, k)(1 - \beta)^p e(0)$, we obtain that $t^*(\varepsilon) = O(\ln(\frac{\|B\|}{\varepsilon\beta})/\beta)$ epochs. □

For average-oriented opinion formation games, $D + E = A$, $\|D\| \leq 1 - 1/n$ and $\|E\| \leq (n - 1)/n^2$. Hence, applying Theorem 1 with $\beta \geq 1/n^2$, we obtain the following:

Corollary 2 *For any update schedule and any $\varepsilon > 0$, the opinion formation process (6) with outdated information about* $\mathrm{avg}(z(t))$ *converges to the equilibrium $z^* = (\mathbb{I} - A)^{-1}Bs$ of the corresponding average-oriented game within distance ε in $O(n^2 \ln(n/\varepsilon))$ epochs.*

4 The Price of Anarchy of Symmetric Average-Oriented Games

In this section, we proceed to bound the PoA of average-oriented opinion formation games. We now concentrate on the most interesting case of symmetric games, since nonsymmetric opinion formation games can have a PoA of $\Omega(n)$, even if $\alpha = 0$ (see e.g., [6, Fig. 2]). Recall that for symmetric games, $w_{ij} = w_{ji}$ for all agent pairs i, j, and $w_i = 1$ and $\alpha_i = \alpha$, for all agents i.

Our analysis generalizes a local smoothness argument put forward in [4, Sec. 3.1]. A function $C(z)$ is (λ, μ)-*locally smooth* [20] if there exist $\lambda > 0$ and $\mu \in (0, 1)$, such that for all $z, x \in \mathbb{R}^n$,

$$C(z) + (x - z)^T C'(z) \leq \lambda C(x) + \mu C(z)\,, \tag{8}$$

where $C'(z) = (\frac{dC_1(z)}{dz_1}, \frac{dC_2(z)}{dz_2}, \cdots, \frac{dC_n(z)}{dz_n})$, i.e., the vector of the partial derivatives of $C(z)$ with respect to $z_1, \ldots, z_n$. At the equilibrium z^*, $C'(z^*) = \mathbf{0}$. Hence, applying (8) for the equilibrium z^* and for the optimal solution o, we obtain that PoA $\leq \lambda/(1 - \mu)$. For symmetric games without aggregation, i.e.,

with $\alpha = 0$, it is known [4, Sec. 3.1] that for any $\boldsymbol{s} \in [0,1]^n$, the cost function $\sum_{i=1}^{n}(z_i - s_i)^2 + \sum_{i\in N}\sum_{j\neq i} w_{ij}(z_i - z_j)^2$ is (λ,μ)-locally smooth for any $\lambda \geq \max\{1/(4\mu), 1/(\mu+1)\}$ (see also Proposition 3 and Proposition 4). Then, by selecting $\lambda = 3/4$ and $\mu = 1/3$, the PoA of symmetric opinion formation games without aggregation can be bounded to at most $9/8$ [4]. This is tight as shown in [6, Fig. 1].

This elegant approach cannot be directly generalized to symmetric average-oriented opinion formation games, because the function $\sum_{i\in N}(\text{avg}(\boldsymbol{z}) - s_i)^2$ is not (λ,μ)-locally smooth for any $\mu < 1$. To circumvent this difficulty, we use the local smoothness technique in a more creative way. Observe that finding appropriate values of λ, μ that satisfy (8) for all $\boldsymbol{x}, \boldsymbol{z} \in [0,1]^n$ may be both a hard and a redundant task, because (8) is applied only for $\boldsymbol{z} = \boldsymbol{z}^*$ and $\boldsymbol{x} = \boldsymbol{o}^*$, where $\boldsymbol{z}^*$ denotes the equilibrium and $\boldsymbol{o}^*$ denotes the optimal vector. Next, we derive appropriate values of λ, μ so that (8) holds for all opinion vectors $\boldsymbol{x}, \boldsymbol{z} \in [0,1]^n$ for which $\text{avg}(\boldsymbol{z}) = \text{avg}(\boldsymbol{s})$. In Proposition 1, we show that for symmetric opinion formation games, the average equilibrium opinion is equal to the average belief, which allows us to bound the PoA.

Proposition 1 *Let $\boldsymbol{z}^*$ be the equilibrium and $\boldsymbol{s}$ the agent belief vector of any symmetric average-oriented opinion formation game. Then,* $\text{avg}(\boldsymbol{z}^*) = \text{avg}(\boldsymbol{s})$.

Proof The following holds for the opinion z_i^* of any agent i at equilibrium $\boldsymbol{z}^*$:

$$z_i^* + z_i^* \sum_{j\neq i} w_{ij} = (1 + \alpha/n)s_i + \sum_{j\neq i} w_{ij} z_j^* - (\alpha/n)\text{avg}(\boldsymbol{z}^*) \ .$$

By summing up these inequalities for all agents $i \in [n]$,

$$n\,\text{avg}(\boldsymbol{z}^*) + \sum_{i\in N} z_i^* \sum_{j\neq i} w_{ij} = (n+\alpha)\text{avg}(\boldsymbol{s}) + \sum_{i\in N}\sum_{j\neq i} w_{ij} z_j^* - \alpha\,\text{avg}(\boldsymbol{z}^*) \ .$$

Since the game is symmetric with $w_{ij} = w_{ji}$ for all $i \neq j$,

$$\sum_{i\in N} z_i^* \sum_{j\neq i} w_{ij} = \sum_{i\in N}\sum_{j\neq i} w_{ij} z_j^* = \sum_{i,j:i<j} w_{ij}(z_i^* + z_j^*) \ .$$

Therefore, we obtain that at the equilibrium $\boldsymbol{z}^*$, $(n+\alpha)\text{avg}(\boldsymbol{z}^*) = (n+\alpha)\text{avg}(\boldsymbol{s})$, which directly implies the proposition. □

In the analysis of PoA, we use the following technical proposition repeatedly.

Proposition 2 *For any $\gamma, \lambda, \mu \geq 0$ and $x, z \in \mathbb{R}$ such that $\lambda\mu \geq \gamma^2$, $2\gamma xz \leq \lambda x^2 + \mu z^2$.*

Proof The claim holds trivially if $xz < 0$. In case where $xz \geq 0$, the claim follows from:

$$0 \leq (\sqrt{\lambda}x - \sqrt{\mu}z)^2 = \lambda x^2 + \mu z^2 - 2\sqrt{\lambda\mu}xz \leq \lambda x^2 + \mu z^2 - 2\gamma xz \,.$$

The last inequality holds because $\lambda\mu \geq \gamma^2$ implies that $-\sqrt{\lambda\mu} \leq -\gamma$. □

Based on these properties, we show that the PoA of symmetric average-oriented games tends to 9/8, which is the PoA of symmetric opinion formation games without aggregation. The proof is based on the following technical (and more general) lemma:

Lemma 2 *Let $\mathcal{G}$ be any symmetric average-oriented opinion formation game with n agents, agent belief vector $\boldsymbol{s}$ and influence $\alpha \geq 0$. Then, for all $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^n$ such that* $\text{avg}(\boldsymbol{z}) = \text{avg}(\boldsymbol{s})$,

$$C(\boldsymbol{z}) + (\boldsymbol{x} - \boldsymbol{z})^T C'(\boldsymbol{z}) \leq \nu_1 C(\boldsymbol{x}) + \nu_2 C(\boldsymbol{z}) \,,$$

where $\nu_1 = \max\{3/4 + \mu, \delta\}$ and $\nu_2 = \max\{1/3 + \mu, 1 - \delta + 2\lambda\}$, for all $\lambda > 0$ and $\mu \in (0, 1)$ such that $\lambda\mu \geq \alpha/n^2$ and for all $\delta > 0$.

Proof We recall that the individual cost of each agent i with respect to opinions $\boldsymbol{z}$ is

$$C_i(\boldsymbol{z}) = (s_i - z_i)^2 + \sum_{i \neq j} w_{ij}(z_i - z_j)^2 + \alpha(s_i - \text{avg}(\boldsymbol{z}))^2$$

and that the social cost is $C(\boldsymbol{z}) = \sum_i C_i(\boldsymbol{z})$. We divide agent's i personal cost $C_i(\boldsymbol{z})$ into three parts $C_i(\boldsymbol{z}) = F_i(\boldsymbol{z}) + I_i(\boldsymbol{z}) + A_i(\boldsymbol{z})$, where $F_i(\boldsymbol{z}) = \sum_{j \neq i} w_{ij}(z_i - z_j)^2$, $I_i(\boldsymbol{z}) = (z_i - s_i)^2$ and $A_i(\boldsymbol{z}) = (\text{avg}(\boldsymbol{z}) - s_i)^2$. Following this notation, we have that:

$$\begin{aligned}
F(\boldsymbol{z}) &= \sum_{i \in N} F_i(\boldsymbol{z}) = \sum_{i \in N} \sum_{j \neq i} w_{ij}(z_i - z_j)^2 = 2 \sum_{i,j:i<j} w_{ij}(z_i - z_j)^2 \\
I(\boldsymbol{z}) &= \sum_{i \in N} I_i(\boldsymbol{z}) = \sum_{i \in N} (z_i - s_i)^2 = (\boldsymbol{z} - \boldsymbol{s})^T (\boldsymbol{z} - \boldsymbol{s}) \\
A(\boldsymbol{z}) &= \sum_{i \in N} A_i(\boldsymbol{z}) = \alpha \sum_{i \in N} (\text{avg}(\boldsymbol{z}) - s_i)^2 = \alpha(\text{avg}(\boldsymbol{z}) - \boldsymbol{s})^T (\text{avg}(\boldsymbol{z}) - \boldsymbol{s}) \;\; .
\end{aligned}$$

Consequently, the social cost can be written as $C(\boldsymbol{z}) = F(\boldsymbol{z}) + I(\boldsymbol{z}) + A(\boldsymbol{z})$. We introduce

$$\begin{aligned}
F'(\boldsymbol{z}) &= \left(\frac{dF_1(\boldsymbol{z})}{dz_1}, \cdots, \frac{dF_n(\boldsymbol{z})}{dz_n} \right) \\
I'(\boldsymbol{z}) &= \left(\frac{dI_1(\boldsymbol{z})}{dz_1}, \cdots, \frac{dI_n(\boldsymbol{z})}{dz_n} \right) \\
A'(\boldsymbol{z}) &= \left(\frac{dA_1(\boldsymbol{z})}{dz_1}, \cdots, \frac{dA_n(\boldsymbol{z})}{dz_n} \right)
\end{aligned}$$

We observe that $A'(\boldsymbol{z}) = (2\alpha/n)(\text{avg}(\boldsymbol{z}) - \boldsymbol{s})$. For simplicity and brevity, here and in the proof of Theorem 4, we slightly abuse the notation by letting $\text{avg}(\boldsymbol{z})$ denote a vector with all its coordinates equal to $\text{avg}(\boldsymbol{z})$. The following two propositions are proven in [4, Sec. 3.1] for more general cost functions. We provide their proofs here, for the sake of completeness.

Proposition 3 [4] *For any symmetric matrix $W = (w_{ij})$, any $\boldsymbol{z}, \boldsymbol{x} \in \mathbb{R}^n$, and any $\lambda > 0$ and $\mu \in (0, 1)$ with $\lambda \geq 1/(4\mu)$,*

$$F(\boldsymbol{z}) + (\boldsymbol{x} - \boldsymbol{z})^T F'(\boldsymbol{z}) \leq \lambda F(\boldsymbol{x}) + \mu F(\boldsymbol{z})$$

Proof To establish the proposition, we consider each agent pair i, j, with $i \neq j$, separately. Since for any agent pair i, j, $w_{ij} = w_{ji}$, we have that for any $\lambda > 0$ and $\mu \in (0, 1)$ with $\lambda\mu \geq 1/4$,

$$\begin{aligned}
F(\boldsymbol{z})+(\boldsymbol{x}-\boldsymbol{z})^T F'(\boldsymbol{z}) &= 2 \sum_{i,j:i\neq j} w_{ij}((z_i-z_j)^2+(x_i-z_i)(z_i-z_j)+(x_j-z_j)(z_j-z_i)) \\
&= 2 \sum_{i,j:i\neq j} w_{ij}((z_i-z_j)^2+(x_i-x_j)(z_i-z_j)-(z_i-z_j)^2) \\
&= 2 \sum_{i,j:i\neq j} w_{ij}(x_i-x_j)(z_i-z_j) \\
&\leq 2\lambda \sum_{i,j:i\neq j} w_{ij}(x_i-x_j)^2 + 2\mu \sum_{i,j:i\neq j} w_{ij}(z_i-z_j)^2 \\
&= \lambda F(\boldsymbol{x}) + \mu F(\boldsymbol{z})\,.
\end{aligned}$$

For the inequality, we apply Proposition 2 with $\gamma = 1/2$. Therefore, for any $x_i, x_j, z_i, z_j \in \mathbb{R}$ and any $\lambda, \mu > 0$ with $\lambda\mu \geq 1/4$, $(x_i - x_j)(z_i - z_j) \leq \lambda(x_i - x_j)^2 + \mu(z_i - z_j)^2$. □

Proposition 4 [4] *For any $\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{s} \in \mathbb{R}^n$, $\lambda > 0$ and $\mu \in (0, 1)$ with $\lambda \geq 1/(\mu + 1)$,*

$$I(\boldsymbol{z}) + (\boldsymbol{x} - \boldsymbol{z})^T I'(\boldsymbol{z}) \leq \lambda I(\boldsymbol{x}) + \mu I(\boldsymbol{z})$$

Proof To establish the proposition, we consider each agent i separately. We have that for any $\lambda > 0$ and $\mu \in (0, 1)$ such that $\lambda(\mu + 1) \geq 1$,

$$\begin{aligned}
I(\boldsymbol{z}) + (\boldsymbol{x} - \boldsymbol{z})^T I'(\boldsymbol{z}) &= \sum_{i\in N}((z_i - s_i)^2 + 2(x_i - z_i)(z_i - s_i)) \\
&= \sum_{i\in N}((z_i - s_i)^2 + 2(x_i - s_i)(z_i - s_i) + 2(s_i - z_i)(z_i - s_i)) \\
&= \sum_{i\in N}((z_i - s_i)^2 + 2(x_i - s_i)(z_i - s_i) - 2(z_i - s_i)^2) \\
&= \sum_{i\in N}(2(x_i - s_i)(z_i - s_i) - (z_i - s_i)^2) \\
&\leq \lambda \sum_{i\in N}(x_i - s_i)^2 + \mu \sum_{i\in N}(z_i - s_i)^2 \\
&= \lambda I(\boldsymbol{x}) + \mu I(\boldsymbol{z})\,.
\end{aligned}$$

For the inequality, we apply Proposition 2 with $\gamma = 1$ and $\mu + 1$ instead of μ. Thus, we obtain that for any $x_i, z_i, s_i \in \mathbb{R}$ and for any $\lambda > 0$ and $\mu \in (0, 1)$ such that $\lambda(\mu + 1) \geq 1$, $2(x_i - s_i)(z_i - s_i) \leq \lambda(x_i - s_i)^2 + (\mu + 1)(z_i - s_i)^2$, which implies the inequality above. □

Next, using Proposition 1, we obtain a similar upper bound on $A(\boldsymbol{z}) + (\boldsymbol{x} - \boldsymbol{z})^T A'(\boldsymbol{z})$.

Proposition 5 *For any $\alpha > 0$, any $\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{s} \in \mathbb{R}^n$ with $\mathrm{avg}(\boldsymbol{z}) = \mathrm{avg}(\boldsymbol{s})$, any $\delta \geq 0$, and any $\lambda > 0$ and $\mu \in (0, 1)$ such that $\lambda\mu \geq \alpha/n^2$,*

$$A(\boldsymbol{z}) + (\boldsymbol{x} - \boldsymbol{z})^T A'(\boldsymbol{z}) \leq \delta A(\boldsymbol{x}) + \mu I(\boldsymbol{x}) + (1 - \delta + 2\lambda)A(\boldsymbol{z}) + \mu I(\boldsymbol{z}) \ . \quad (9)$$

Proof Applying first-order optimality conditions, we obtain that any vector $\boldsymbol{z} \in \mathbb{R}^n$ with $\mathrm{avg}(\boldsymbol{z}) = \mathrm{avg}(\boldsymbol{s})$ minimizes $A(\boldsymbol{z})$. Therefore, for any $\boldsymbol{x} \in \mathbb{R}^n$, $A(\boldsymbol{z}) \leq A(\boldsymbol{x})$, and for any $\delta \geq 0$, $A(\boldsymbol{z}) \leq \delta A(\boldsymbol{x}) + (1 - \delta)A(\boldsymbol{z})$.

To complete the proof of (9), we observe that for any $\lambda > 0$, $\mu \in (0, 1)$ with $\lambda\mu \geq \alpha/n^2$,

$$\begin{aligned}
(\boldsymbol{x} - \boldsymbol{z})^T A'(\boldsymbol{z}) &= \sum_{i \in N} (2\alpha/n)(x_i - z_i)(\mathrm{avg}(\boldsymbol{z}) - s_i) \\
&= \sum_{i \in N} ((2\alpha/n)(x_i - s_i)(\mathrm{avg}(\boldsymbol{z}) - s_i) + (2\alpha/n)(s_i - z_i)(\mathrm{avg}(\boldsymbol{z}) - s_i)) \\
&\leq \sum_{i \in N} (2\lambda\alpha(\mathrm{avg}(\boldsymbol{z}) - s_i)^2 + \mu(x_i - s_i)^2 + \mu(z_i - s_i)^2) \\
&= 2\ \lambda A(\boldsymbol{z}) + \mu I(\boldsymbol{x}) + \mu I(\boldsymbol{z}) \, .
\end{aligned}$$

For the inequality, we apply Proposition 2, with $\gamma = \sqrt{\alpha}/n$, to $(2\alpha/n)(x_i - s_i)(\mathrm{avg}(\boldsymbol{z}) - s_i)$ and to $(2\alpha/n)(s_i - z_i)(\mathrm{avg}(\boldsymbol{z}) - s_i)$. Hence, we obtain that for any $\lambda > 0$ and $\mu \in (0, 1)$ such that $\lambda\mu \geq \alpha/n^2$, $(2\alpha/n)(x_i - s_i)(\mathrm{avg}(\boldsymbol{z}) - s_i) \leq \mu(x_i - s_i)^2 + \lambda\alpha(\mathrm{avg}(\boldsymbol{z}) - s_i)^2$ and $(2\alpha/n)(s_i - z_i)(\mathrm{avg}(\boldsymbol{z}) - s_i) \leq \mu(z_i - s_i)^2 + \lambda\alpha(\mathrm{avg}(\boldsymbol{z}) - s_i)^2$. □

Applying Propositions 3 and 4 with $\lambda = 3/4$ and $\mu = 1/3$, and using (9), we obtain that for any $\delta \geq 0$ and for any $\lambda > 0$ and $\mu \in (0, 1)$ such that $\lambda\mu \geq \alpha/n^2$,

$$\begin{aligned}
C(\boldsymbol{z}) + (\boldsymbol{x} - \boldsymbol{z})^T C'(\boldsymbol{z}) &\leq \frac{3}{4}F(\boldsymbol{x}) + \left(\frac{3}{4} + \mu\right) I(\boldsymbol{x}) + \delta A(\boldsymbol{x}) + \frac{1}{3}F(\boldsymbol{z}) + \\
&\qquad \left(\frac{1}{3} + \mu\right) I(\boldsymbol{x}) + (1 - \delta + 2\lambda)A(\boldsymbol{x}) \\
&\leq \nu_1 C(\boldsymbol{x}) + \nu_2 C(\boldsymbol{z}) \, ,
\end{aligned}$$

where $\nu_1 = \max\{3/4 + \mu, \delta\}$ and $\nu_2 = \max\{1/3 + \mu, 1 - \delta + 2\lambda\}$. □

The main result of this section is an immediate consequence of Lemma 2.

Theorem 2 *Let $\mathcal{G}$ be any symmetric average-oriented opinion formation game with n agents and influence $\alpha \geq 0$. Then,* $\mathrm{PoA}(\mathcal{G}) \leq 9/8 + O(\alpha/n^2)$.

Proof Let $\boldsymbol{z}$ be the equilibrium and let $\boldsymbol{o}$ be the optimal solution. By Proposition 1, $\mathrm{avg}(\boldsymbol{z}) = \mathrm{avg}(\boldsymbol{s})$. Therefore, Lemma 2 implies that

$$C(\boldsymbol{z}) + (\boldsymbol{o} - \boldsymbol{z})^T C'(\boldsymbol{z}) \leq \nu_1 C(\boldsymbol{o}) + \nu_2 C(\boldsymbol{z}) \, ,$$

where $\nu_1 = \max\{3/4 + \mu, \delta\}$ and $\nu_2 = \max\{1/3 + \mu, 1 - \delta + 2\lambda\}$, for all $\lambda > 0$ and $\mu \in (0, 1)$ such that $\lambda\mu \geq \alpha/n^2$ and for all $\delta > 0$. Since $\boldsymbol{z}$ is an equilibrium, $C'(\boldsymbol{z}) = \mathbf{0}$. Hence, for all $\nu_2 \in (0, 1)$, $\mathrm{PoA}(\mathcal{G}) \leq \nu_1/(1 - \nu_2)$, or equivalently,

$$\mathrm{PoA}(\mathcal{G}) \leq \frac{\max\{3/4 + \mu, \delta\}}{1 - \max\{1/3 + \mu, 1 - \delta + 2\lambda\}}\,. \tag{10}$$

If α/n^2 is small enough, e.g., if $\alpha/n^2 \leq 1/2400$, we use $\delta = 3/4$, $\lambda = 1/24$ and $\mu = 24\alpha/n^2$ in (10) and obtain that $\mathrm{PoA}(\mathcal{G}) \leq 9/8 + O(\frac{\alpha}{n^2})$. Otherwise, we use $\mu = 1/3$, $\lambda = 3\alpha/n^2$ and $\delta = 6\alpha/n^2 + 1/3$, and obtain that $\mathrm{PoA}(\mathcal{G}) = O(\frac{\alpha}{n^2})$. □

5 Average-Oriented Games with Restricted Opinions

A frequent assumption in the literature on opinion formation is that agent beliefs come from a finite interval of nonnegative real numbers. Then, by scaling we can assume beliefs $s_i \in [0, 1]$. If the influence matrix A is nonnegative, then since $b_{ii} + \sum_{j=1}^{n} a_{ij} = 1$ for all $i \in [n]$, we have that the equilibrium opinions are $\boldsymbol{z}^* = (\mathbb{I} - A)^{-1}B\boldsymbol{s} \in [0, 1]^n$. In contrast, for the more general domain we treat here, an important side-effect of negative influence is that the best-response (and equilibrium) opinions may not belong to [0, 1]. Motivated by this observation, we consider a *restricted* variant of opinion formation games, where the (best-response and equilibrium) opinions are restricted to [0, 1]. We strive to understand how this restriction of public opinions to [0, 1] affects the convergence properties and the price of anarchy of average-oriented games.

To distinguish restricted opinion formation processes from their unrestricted counterparts, we use $\boldsymbol{y}(t)$ to denote the opinion vectors restricted to $[0, 1]^n$. For restricted average-oriented games and restricted games with negative influence, the best-response opinion y_i of each agent i to $\boldsymbol{y}_{-i}$ is again computed by (2) and (7), respectively. But now, if the resulting value is $y_i < 0$, we increase it to $y_i = 0$, while if $y_i > 1$, we decrease it to $y_i = 1$. Since the individual cost $C_i(\boldsymbol{y})$ is a strictly convex function of y_i, the restriction of y_i to [0, 1] results in a minimizer $y^* \in [0, 1]$ of $C_i(y, \boldsymbol{y}_{-i})$.

Similarly, the restricted opinion formation process is described by

$$\boldsymbol{y}(t) = [A\boldsymbol{y}(t - 1) + B\boldsymbol{s}\,]_{[0,1]}\,, \tag{11}$$

where $[\cdot]_{[0,1]}$ denotes the restriction of public opinions $\boldsymbol{y}(t)$ to $[0, 1]^n$ described above. The influence matrix A (and the influence matrices D and E for processes with outdated information) and the self-confidence matrix B are computed as for standard (or unrestricted) opinion formation processes.

5.1 Convergence of Restricted Opinion Formation Processes

We show results for restricted opinion formation processes that are equivalent to Lemma 1 and Theorem 1. As in Section 3, we prove our results for the more general setting of negative influence. Using Lemma 3 and Theorem 3, it is straightforward to

obtain the results of Corollary 1 and Corollary 2 also for restricted average-oriented processes.

Lemma 3 *Let A be any influence matrix, possibly with negative elements, such that $\|A\| \leq 1-\beta$, for some $\beta > 0$. Then, for any self-confidence matrix B, any $\boldsymbol{s} \in [0,1]^n$ and any $\varepsilon > 0$, the opinion formation process $\boldsymbol{y}(t) = [A\boldsymbol{y}(t-1) + B\boldsymbol{s}\,]_{[0,1]}$ admits a unique equilibrium $\boldsymbol{y}^*$ and converges to it within distance ε in $O(\ln(\frac{1}{\varepsilon})/\beta)$ rounds.*

Proof In the restricted opinion formation game, the agent opinions lie in the convex set $[0, 1]$. The individual cost $C_i(\boldsymbol{y})$ of each agent i is a continuous function of $\boldsymbol{y}$ and strictly convex in y_i. Hence, according to the results of [19], the restricted game admits a unique equilibrium $\boldsymbol{y}^*$ which satisfies $\boldsymbol{y}^* = [A\boldsymbol{y}^* + B\boldsymbol{s}\,]_{[0,1]}$. Specifically, the existence of an equilibrium $\boldsymbol{y}^*$ follows from [19, Theorem 1], since the restricted opinion formation game is a convex game. The uniqueness of $\boldsymbol{y}^*$ follows from [19, Theorem 2] and from the fact that the function $\sum_{i\in N} C_i(\boldsymbol{y})$ is diagonally strictly convex. The latter holds because the symmetric matrix obtained by adding $2B$ to the Laplacian of $A + A^T$ is positive definite.

Next we bound the convergence time to $\boldsymbol{y}^*$ as in the proof of Lemma 1. For any $t \geq 1$, we define $e(t) = \|\boldsymbol{y}(t) - \boldsymbol{y}^*\|$ as the distance of the opinions at time t to equilibrium. We observe that for any round $t \geq 1$,

$$\begin{aligned} e(t) &= \|\boldsymbol{y}(t) - \boldsymbol{y}^*\| \leq \|A\boldsymbol{y}(t-1) + B\boldsymbol{s} - A\boldsymbol{y}^* - B\boldsymbol{s}\| \\ &\leq \|A\|\,\|\boldsymbol{y}(t-1) - \boldsymbol{y}^*\| \leq (1-\beta)e(t-1) \leq (1-\beta)^t e(0)\,. \end{aligned}$$

For the first inequality, we recall that $\boldsymbol{y}(t)$ (resp. $\boldsymbol{y}^*$) is obtained by computing $A\boldsymbol{y}(t-1) + B\boldsymbol{s}$ (resp. $A\boldsymbol{y}^* + B\boldsymbol{s}$) and then restricting any negative opinions to 0 and any opinions larger than 1 to 1. By a straightforward inspection of all possible 9 cases depending on whether $y_i(t)$ and y_i^* are negative, in $[0, 1]$ or greater than 1, we conclude that opinion restriction to $[0, 1]$ does not increase $|y_i(t) - y_i^*|$ for any i. Since $\boldsymbol{y}(0) = \boldsymbol{s} \in [0,1]^n$ and $\boldsymbol{y}^* \in [0,1]^n$, $e(0) \leq 1$. Hence, after $t^*(\varepsilon) = O(\ln(\frac{1}{\varepsilon})/\beta)$ rounds $\boldsymbol{y}(t)$ is within distance ε to $\boldsymbol{y}^*$. □

The proof of the following theorem is similar to the proof of Theorem 1.

Theorem 3 *Let D and E be influence matrices, possibly with negative elements, such that $\|D\| \leq 1-\beta_1$, $\|E\| \leq 1-\beta_2$, for some $\beta_1, \beta_2 \in (0,1)$ with $\beta_1 + \beta_2 > 1$. Then, for any self-confidence matrix B, any $\boldsymbol{s} \in [0,1]^n$, any update schedule $0 = \tau_0 < \tau_1 < \tau_2 < \cdots$, the restricted opinion formation process $\boldsymbol{y}(t) = [D\boldsymbol{y}(t-1) + E\boldsymbol{y}(\tau_p) + B\boldsymbol{s}\,]_{[0,1]}$ converges to the unique equilibrium point $\boldsymbol{y}^*$ of $\boldsymbol{y}'(t) = [(D+E)\boldsymbol{y}'(t-1) + B\boldsymbol{s}\,]_{[0,1]}$. For any $\varepsilon > 0$, $\boldsymbol{y}(t)$ is within distance ε to $\boldsymbol{y}^*$ after $O(\ln(\frac{1}{\varepsilon})/\beta)$ epochs, where $\beta = \beta_1 + \beta_2 - 1$.*

Proof Lemma 3 shows that for the restricted opinion formation process $\boldsymbol{y}'(t) = [(D+E)\boldsymbol{y}'(t-1) + B\boldsymbol{s}\,]_{[0,1]}$, there is a unique equilibrium point $\boldsymbol{y}^* \in [0,1]^n$ that satisfies $\boldsymbol{y}^* = [(D+E)\boldsymbol{y}^* + B\boldsymbol{s}\,]_{[0,1]}$. Provided that it exists, the equilibrium of the restricted opinion formation process with outdated information $\boldsymbol{y}(t) = [D\boldsymbol{y}(t-1) + E\boldsymbol{y}(\tau_p) + B\boldsymbol{s}\,]_{[0,1]}$ must satisfy $\boldsymbol{y}^* = [D\boldsymbol{y}^* + E\boldsymbol{y}^* + B\boldsymbol{s}\,]_{[0,1]}$, due

to the existence of infinite update points where all agents have accurate information about the current public opinion vector. So, if the process with outdated information admits an equilibrium, it must be unique and equal to $\boldsymbol{y}^*$. We next show that this is indeed the case, by bounding from above the distance of $\boldsymbol{y}(t)$ to $\boldsymbol{y}^*$ by a decreasing function of t and by establishing an upper bound on the convergence time.

For every round $t \geq 1$, we define $e(t) = \|\boldsymbol{y}(t) - \boldsymbol{y}^*\|$ as the distance of the opinions at time t to $\boldsymbol{y}^*$. We proceed similarly to the proof of Theorem 1. As before, we define

$$f(\beta_1, \beta_2, k) = (1-\beta_1)^k + (1-\beta_2)\frac{1-(1-\beta_1)^k}{\beta_1}\,.$$

We recall that for any fixed $\beta_1, \beta_2 \in (0,1)$ with $\beta_1 + \beta_2 > 1$, $f(\beta_1, \beta_2, k)$ is a decreasing function of k.

We next show that:

Claim (i). For every epoch $p \geq 0$ and every round k, $0 \leq k \leq k_p$, in epoch p,

$$e(\tau_p + k) \leq f(\beta_1, \beta_2, k) e(\tau_p)\,.$$

Claim (ii). In the last round $\tau_{p+1} = \tau_p + k_p$ of each epoch $p \geq 0$, $e(\tau_{p+1}) \leq (1-\beta)e(\tau_p)$.

Claims (i) and (ii) imply that for each epoch $p \geq 0$ and every round k, $0 \leq k \leq k_p$, in epoch p, $e(\tau_p + k) \leq f(\beta_1, \beta_2, k)(1-\beta)^p e(0)$. This immediately implies that for any update schedule $\tau_0 < \tau_1 < \tau_2 < \cdots$, the opinion formation process $\boldsymbol{y}(t) = [D\boldsymbol{y}(t-1) + E\boldsymbol{y}(\tau_p) + B\boldsymbol{s}\,]_{[0,1]}$ converges to $\boldsymbol{y}^*$. Moreover, since $e(0) = \|\boldsymbol{s} - \boldsymbol{y}^*\| \leq 1$, $\boldsymbol{y}(t)$ is within distance ε to $\boldsymbol{y}^*$ in $O(\ln(\frac{1}{\varepsilon})/\beta)$ epochs.

The proofs of Claim (i) and Claim (ii) are essentially identical to the proofs of the corresponding claims in the proof of Theorem 1. We include the details for completeness. To prove Claim (i), we fix an epoch $p \geq 0$ and apply induction on k. The basis, where $k = 0$, holds because $f(\beta_1, \beta_2, 0) = 1$. For any round k, with $1 \leq k \leq k_p$, in p, we have that:

$$\begin{aligned}
e(\tau_p + k) &= \|\boldsymbol{y}(\tau_p + k) - \boldsymbol{y}^*\| \\
&= \|[D\boldsymbol{y}(\tau_p + k - 1) + E\boldsymbol{y}(\tau_p) + B\boldsymbol{s}\,]_{[0,1]} - [D\boldsymbol{y}^* + E\boldsymbol{y}^* + B\boldsymbol{s}\,]_{[0,1]}\| \\
&\leq \|(D\boldsymbol{y}(\tau_p + k - 1) + E\boldsymbol{y}(\tau_p) + B\boldsymbol{s}) - (D\boldsymbol{y}^* + E\boldsymbol{y}^* + B\boldsymbol{s})\| \\
&\leq \|D\|\,\|\boldsymbol{y}(\tau_p + k - 1) - \boldsymbol{y}^*\| + \|E\|\,\|\boldsymbol{y}(\tau_p) - \boldsymbol{y}^*\| \\
&\leq (1-\beta_1)e(\tau_p + k - 1) + (1-\beta_2)e(\tau_p) \\
&\leq (1-\beta_1) f(\beta_1, \beta_2, k-1) e(\tau_p) + (1-\beta_2)e(\tau_p) \\
&= f(\beta_1, \beta_2, k) e(\tau_p)\,.
\end{aligned}$$

For the first inequality, we use that opinion restriction to $[0,1]$ does not increase $|y_i(t) - y_i^*|$ for any i, as it is explained in the proof of Lemma 3. The second inequality follows from the properties of matrix norms. The third inequality holds because $\|D\| \leq 1 - \beta_1$ and $\|E\| \leq 1 - \beta_2$. The fourth inequality follows from the induction hypothesis. Finally, we observe that for any integer $k \geq 1$, $(1-\beta_1) f(\beta_1, \beta_2, k-1) + 1 - \beta_2 = f(\beta_1, \beta_2, k)$.

To prove Claim (ii), we fix any epoch $p \geq 0$ and apply claim (i) to the last round $\tau_{p+1} = \tau_p + k_p$ of epoch p, where $k_p \geq 1$. Hence, we obtain that:

$$e(\tau_{p+1}) = \|\boldsymbol{y}(\tau_p + k_p) - \boldsymbol{y}^*\| \leq f(\beta_1, \beta_2, k_p) e(\tau_p) \leq (2 - \beta_1 - \beta_2) e(\tau_p) = (1 - \beta) e(\tau_p),$$

where $\beta = \beta_1 + \beta_2 - 1$. The last inequality follows from convexity and has already been proven in the corresponding part of the proof of Theorem 1. □

5.2 The Price of Anarchy of Restricted Average-Oriented Games

We proceed to bound the PoA of restricted symmetric average-oriented games. Due to opinion restriction to [0, 1], the average opinion at equilibrium may be far from $\text{avg}(\boldsymbol{s})$. Therefore, we cannot rely on Proposition 5 anymore. Moreover, the PoA of restricted games increases fast with α (e.g., if $\boldsymbol{s} = (0, \ldots, 0, 1/n)$, $w_{ij} = 0$ for all $i \neq j$, and $\alpha = n^2$, $\text{PoA} = \Omega(n)$). Therefore, we here restrict our attention to the case where $\alpha = w = 1$ and show that the PoA of restricted symmetric average-oriented games remains constant. An interesting intermediate result of our analysis is that if all agents only value the distance of their opinion to their belief and to the average, i.e., if $w_{ij} = 0$ for all $i \neq j$, the PoA of such games is at most $1 + 1/n^2$.

As in the proofs of Lemma 2 and Theorem 2, we use a generalized local smoothness argument. In this case, however, the function $\sum_{i=1}^{n} (\text{avg}(\boldsymbol{y}) - s_i)^2$ is not (λ, μ)-locally smooth and $\text{avg}(\boldsymbol{y}^*)$ at the equilibrium $\boldsymbol{y}^*$ may be far from $\text{avg}(\boldsymbol{s})$. Hence, to bound the PoA, we need to advance substantially beyond the local smoothness argument of [4, Sec. 3.1]. The rest of this section is devoted to the proof of the following:

Theorem 4 *Let $\mathcal{G}$ be any symmetric average-oriented opinion formation game with $w = \alpha = 1$, $n \geq 2$ agents and opinions restricted to* [0, 1]. *Then,* $\text{PoA}(\mathcal{G}) \leq 3 + \sqrt{2} + O(\frac{1}{n})$.

Proof As in the proofs of Lemma 2 and Theorem 2, we seek to find appropriate parameters $\lambda > 0$ and $\mu \in (0, 1)$ such that for all $\boldsymbol{x}, \boldsymbol{y} \in [0, 1]^n$,

$$C(\boldsymbol{y}) + (\boldsymbol{x} - \boldsymbol{y})^T C'(\boldsymbol{y}) \leq \lambda C(\boldsymbol{x}) + \mu C(\boldsymbol{y}) . \tag{12}$$

where $C'(\boldsymbol{y}) = (\frac{dC_1(\boldsymbol{y})}{dy_1}, \cdots, \frac{dC_n(\boldsymbol{y})}{dy_n})$.

Next, we show that (12) indeed implies $\text{PoA}(\mathcal{G}) \leq \lambda/(1-\mu)$. To this end, we show that at the equilibrium $\boldsymbol{y}^*$ of a restricted game, $(\boldsymbol{x} - \boldsymbol{y}^*)^T C'(\boldsymbol{y}^*) \geq 0$. By definition $\boldsymbol{y}^* \in [0, 1]^n$. In case where $y_i^* \in (0, 1)$, due to first-order optimality conditions, $\frac{dC_i(\boldsymbol{y}^*)}{dy_i} = 0$ and $(x_i - y_i^*)\frac{dC_i(\boldsymbol{y}^*)}{dy_i} = 0$. If $y_i^* = 0$ then $\frac{dC_i(\boldsymbol{y}^*)}{dy_i} \geq 0$. Otherwise, agent i could decrease her cost by increasing y_i^*. Since $x_i \in [0, 1]$, $(x_i - y_i^*)\frac{dC_i(\boldsymbol{y}^*)}{dy_i} \geq 0$. By a symmetric argument, if $y_i^* = 1$, $\frac{dC_i(\boldsymbol{y}^*)}{dy_i} \leq 0$ and $(x_i - y_i^*)\frac{dC_i(\boldsymbol{y}^*)}{dy_i} \geq 0$. Applying (12) for $\boldsymbol{y} = \boldsymbol{y}^*$ and $\boldsymbol{x} = \boldsymbol{o}^*$ (we recall that the optimal solution $\boldsymbol{o}^* \in [0, 1]^n$) yields

$$C(\boldsymbol{y}^*) \leq C(\boldsymbol{y}^*) + (\boldsymbol{o}^* - \boldsymbol{y}^*)^T C'(\boldsymbol{y}^*) \leq \lambda C(\boldsymbol{o}^*) + \mu C(\boldsymbol{y}^*) .$$

Therefore, $\text{PoA}(\mathcal{G}) = C(\boldsymbol{y}^*)/C(\boldsymbol{o}^*) \leq \lambda/(1 - \mu)$.

We proceed to establish (12). As in Section 4, in order to find appropriate values for λ and μ, we divide the individual cost of each agent i into two parts, writing $C_i(\boldsymbol{y}) = F_i(\boldsymbol{y}) + M_i(\boldsymbol{y})$, and analyze each part separately. We again have that:

$$\begin{aligned} F(\boldsymbol{y}) &= \sum_{i=1}^{n} F_i(y) = \sum_{i\in N}\sum_{j\neq i} w_{ij}(y_i - y_j)^2 \\ M(\boldsymbol{y}) &= \sum_{i=1}^{n} M_i(y) = \sum_{i\in N}((y_i - s_i)^2 + (\mathrm{avg}(\boldsymbol{y}) - s_i)^2) \\ &= (\boldsymbol{y}-\boldsymbol{s})^T(\boldsymbol{y}-\boldsymbol{s}) + (\mathrm{avg}(\boldsymbol{y}) - \boldsymbol{s})^T(\mathrm{avg}(\boldsymbol{y}) - \boldsymbol{s})\,. \end{aligned}$$

We again denote

$$\begin{aligned} F'(\boldsymbol{y}) &= \left(\frac{dF_1(\boldsymbol{y})}{dy_1}, \cdots, \frac{dF_n(\boldsymbol{y})}{dy_n}\right) \\ M'(\boldsymbol{y}) &= \left(\frac{dM_1(\boldsymbol{y})}{dy_1}, \cdots, \frac{dM_n(\boldsymbol{y})}{dy_n}\right) \end{aligned}$$

We also recall that $M'(\boldsymbol{y}) = 2(\boldsymbol{y}-\boldsymbol{s}) + (2/n)(\mathrm{avg}(\boldsymbol{y}) - \boldsymbol{s})$.

Proposition 3 provides an appropriate upper bound on the term $F(\boldsymbol{y}) + (\boldsymbol{x} - \boldsymbol{y})^T F'(\boldsymbol{y})$. So, we next focus on finding appropriate values of λ and μ so that we can bound from above the term $M(\boldsymbol{y}) + (\boldsymbol{x} - \boldsymbol{y})^T M'(\boldsymbol{y})$.

To this end, we first observe that:

$$M(\boldsymbol{y}) + (\boldsymbol{x}-\boldsymbol{y})^T M'(\boldsymbol{y}) = M(\boldsymbol{y}) + (\boldsymbol{s}-\boldsymbol{y})^T M'(\boldsymbol{y}) + (\boldsymbol{x}-\boldsymbol{s})^T M'(\boldsymbol{y})\ .$$

We first bound $M(\boldsymbol{y}) + (\boldsymbol{s} - \boldsymbol{y})^T M'(\boldsymbol{y})$ from above using the following proposition. Intuitively, the proposition holds because the left-hand side of (13) is a strictly concave function of $\boldsymbol{y}$.

Proposition 6 *For any $\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{s} \in [0,1]^n$,*

$$M(\boldsymbol{y}) + (\boldsymbol{s}-\boldsymbol{y})^T M'(\boldsymbol{y}) \leq (1 + \frac{1}{n^2})M(\boldsymbol{s}) \leq (1 + \frac{1}{n^2})M(\boldsymbol{x})\,. \tag{13}$$

Proof Let $\mathbb{K}_n$ denote the $n \times n$ matrix with all its entries equal to $1/n$. Recall that $\mathbb{I}$ is the $n \times n$ identity matrix. Clearly, $\mathbb{K}_n \boldsymbol{y}$ is the vector with all its coordinates equal to $\mathrm{avg}(\boldsymbol{y})$. Moreover, we observe that $\mathbb{K}_n\mathbb{K}_n = \mathbb{K}_n$. Using matrix notation, we obtain that:

$$\begin{aligned} M(\boldsymbol{y}) + (\boldsymbol{s}-\boldsymbol{y})^T M'(\boldsymbol{y}) &= (\mathbb{K}_n\boldsymbol{y} - \boldsymbol{s})^T(\mathbb{K}_n\boldsymbol{y} - \boldsymbol{s}) + (\boldsymbol{y}-\boldsymbol{s})^T(\boldsymbol{y}-\boldsymbol{s}) \\ &\quad +2(\boldsymbol{s}-\boldsymbol{y})^T(\boldsymbol{y}-\boldsymbol{s}) + (2/n)(\boldsymbol{s}-\boldsymbol{y})^T(\mathbb{K}_n\boldsymbol{y} - \boldsymbol{s}) \\ &= \boldsymbol{y}^T((1-\frac{2}{n})\mathbb{K}_n - \mathbb{I})\boldsymbol{y} + 2\,\boldsymbol{y}^T((1+\frac{1}{n})\mathbb{I} - (1-\frac{1}{n})\mathbb{K}_n)\boldsymbol{s} \\ &\quad -\frac{2}{n}\boldsymbol{s}^T\boldsymbol{s}\ . \end{aligned}$$

We observe that the matrix $\mathbb{I} - (1 - \frac{2}{n})\mathbb{K}_n$ is strictly diagonally dominant, and thus positive definite. So, the matrix $(1 - \frac{2}{n})\mathbb{K}_n - \mathbb{I}$ is negative definite. Thus, $M(\boldsymbol{y}) + (\boldsymbol{s} - \boldsymbol{y})^T M'(\boldsymbol{y})$ is strictly concave in $\boldsymbol{y}$ and has a unique maximum in $\mathbb{R}$.

We next show that $M(\boldsymbol{y}) + (\boldsymbol{s} - \boldsymbol{y})^T M'(\boldsymbol{y})$ is maximized at $\boldsymbol{y}^* = (1 + \frac{1}{n})\boldsymbol{s} - \text{avg}(\boldsymbol{s})/n$. To find the unique maximizer $\boldsymbol{y}^*$ of $M(\boldsymbol{y}) + (\boldsymbol{s} - \boldsymbol{y})^T M'(\boldsymbol{y})$, we apply first-order optimality conditions. The gradient of $M(\boldsymbol{y}) + (\boldsymbol{s} - \boldsymbol{y})^T M'(\boldsymbol{y})$ with respect to $y_1, \ldots, y_n$ is equal to

$$2((1 - \frac{2}{n})\mathbb{K}_n - \mathbb{I})\boldsymbol{y} + 2((1 + \frac{1}{n})\mathbb{I} - (1 - \frac{1}{n})\mathbb{K}_n)\boldsymbol{s} \ .$$

So the unique maximizer $\boldsymbol{y}^*$ of $M(\boldsymbol{y}) + (\boldsymbol{s} - \boldsymbol{y})^T M'(\boldsymbol{y})$ satisfies

$$y_i^* = (1 + \frac{1}{n})s_i + (1 - \frac{2}{n})\text{avg}(\boldsymbol{y}^*) - (1 - \frac{1}{n})\text{avg}(\boldsymbol{s}) \ .$$

Summing up these equations for all $i \in N$, we obtain that

$$n\,\text{avg}(\boldsymbol{y}^*) = (n+1)\text{avg}(\boldsymbol{s}) + (n-2)\text{avg}(\boldsymbol{y}^*) - (n-1)\text{avg}(\boldsymbol{s})\,,$$

which implies that $\text{avg}(\boldsymbol{y}^*) = \text{avg}(\boldsymbol{s})$. Therefore, the maximizer $\boldsymbol{y}^*$ of $M(\boldsymbol{y}) + (\boldsymbol{s} - \boldsymbol{y})^T M'(\boldsymbol{y})$ has $y_i^* = (1 + \frac{1}{n})s_i - \text{avg}(\boldsymbol{s})/n$ (note in particular that y_i^* does not need to belong to $[0, 1]$).

Using that $\boldsymbol{y}^* = (1 + \frac{1}{n})\boldsymbol{s} - \text{avg}(\boldsymbol{s})/n$ and $\text{avg}(\boldsymbol{y}^*) = \text{avg}(\boldsymbol{s})$, we obtain:

$$\begin{aligned} M(\boldsymbol{y}^*) + (\boldsymbol{s} - \boldsymbol{y}^*)^T M'(\boldsymbol{y}^*) &= -(\boldsymbol{y}^* - \boldsymbol{s})^T(\boldsymbol{y}^* - \boldsymbol{s}) + (\text{avg}(\boldsymbol{s}) - \boldsymbol{s})^T(\text{avg}(\boldsymbol{s}) - \boldsymbol{s}) \\ &\quad +(2/n)(\boldsymbol{y}^* - \boldsymbol{s})^T(\text{avg}(\boldsymbol{s}) - \boldsymbol{s})^T \\ &= -(1/n^2)(\text{avg}(\boldsymbol{s}) - \boldsymbol{s})^T(\text{avg}(\boldsymbol{s}) - \boldsymbol{s}) \\ &\quad +(\text{avg}(\boldsymbol{s}) - \boldsymbol{s})^T(\text{avg}(\boldsymbol{s}) - \boldsymbol{s}) \\ &\quad +(2/n^2)(\text{avg}(\boldsymbol{s}) - \boldsymbol{s})^T(\text{avg}(\boldsymbol{s}) - \boldsymbol{s})^T \\ &= (1 + \frac{1}{n^2})(\text{avg}(\boldsymbol{s}) - \boldsymbol{s})^T(\text{avg}(\boldsymbol{s}) - \boldsymbol{s}) \ . \end{aligned}$$

The proposition follows from the following observations: (i) for any $\boldsymbol{y} \in [0, 1]^n$, $M(\boldsymbol{y}) + (\boldsymbol{s} - \boldsymbol{y})^T M'(\boldsymbol{y}) \le M(\boldsymbol{y}^*) + (\boldsymbol{s} - \boldsymbol{y}^*)^T M'(\boldsymbol{y}^*)$, since $\boldsymbol{y}^* \in \mathbb{R}^n$ is the unique maximizer of the strictly concave function $M(\boldsymbol{y}) + (\boldsymbol{s} - \boldsymbol{y})^T M'(\boldsymbol{y})$; and (ii) for any $\boldsymbol{x} \in [0, 1]^n$,

$$\begin{aligned} M(\boldsymbol{y}^*) + (\boldsymbol{s} - \boldsymbol{y}^*)^T M'(\boldsymbol{y}^*) &= (1 + \frac{1}{n^2})(\text{avg}(\boldsymbol{s}) - \boldsymbol{s})^T(\text{avg}(\boldsymbol{s}) - \boldsymbol{s}) \\ &= (1 + \frac{1}{n^2})M(\boldsymbol{s}) \le (1 + \frac{1}{n^2})M(\boldsymbol{x})\,, \end{aligned}$$

where the last inequality holds because $\boldsymbol{s}$ is a minimizer of $M(\boldsymbol{y})$. □

Remark 1 If $w_{ij} = 0$ for all $i \neq j$, the cost of each agent i becomes $C_i(\boldsymbol{y}) = (y_i - s_i)^2 + (\text{avg}(\boldsymbol{y}) - y_i)^2$. For this interesting class of restricted symmetric average-oriented games, Proposition 6 implies that the PoA is at most $1 + 1/n^2$.

We proceed to show an upper bound on $(\boldsymbol{x} - \boldsymbol{s})^T M'(\boldsymbol{y})$.

Proposition 7 *For any* $\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{s} \in [0,1]^n$, *and for any* $\lambda_1, \lambda_2 > 0$ *and* $\mu_1, \mu_2 \in (0,1)$ *such that* $\lambda_1\mu_1 \geq 1$ *and* $\lambda_2\mu_2 \geq 1/n^2$,

$$(\boldsymbol{x}-\boldsymbol{s})^T M'(\boldsymbol{y}) \leq (\lambda_1+\lambda_2)M(\boldsymbol{x}) + \max\{\mu_1,\mu_2\}M(\boldsymbol{y})\,. \tag{14}$$

Proof We observe that

$$(\boldsymbol{x}-\boldsymbol{s})^T M'(\boldsymbol{y}) = 2(\boldsymbol{x}-\boldsymbol{s})^T(\boldsymbol{y}-\boldsymbol{s}) + (2/n)(\boldsymbol{x}-\boldsymbol{s})^T(\mathrm{avg}(\boldsymbol{y})-\boldsymbol{s})\ .$$

Applying Proposition 2, with $\gamma = 1$, for each term $2(x_i - s_i)(y_i - s_i)$ of $2(\boldsymbol{x} - \boldsymbol{s})^T(\boldsymbol{y} - \boldsymbol{s})$, we obtain that for any $\lambda_1 > 0$ and $\mu_1 \in (0,1)$ with $\lambda_1\mu_1 \geq 1$,

$$2(\boldsymbol{x}-\boldsymbol{s})^T(\boldsymbol{y}-\boldsymbol{s}) \leq \lambda_1(\boldsymbol{x}-\boldsymbol{s})^T(\boldsymbol{x}-\boldsymbol{s}) + \mu_1(\boldsymbol{y}-\boldsymbol{s})^T(\boldsymbol{y}-\boldsymbol{s})\ .$$

Similarly, applying Proposition 2, with $\gamma = 1/n$, for each term $(2/n)(x_i - s_i)(\mathrm{avg}(\boldsymbol{y}) - s_i)$ of $(2/n)(\boldsymbol{x} - \boldsymbol{s})^T(\mathrm{avg}(\boldsymbol{y}) - \boldsymbol{s})$, we obtain that for any $\lambda_2 > 0$ and $\mu_2 \in (0,1)$ with $\lambda_2\mu_2 \geq 1/n^2$,

$$(2/n)(\boldsymbol{x}-\boldsymbol{s})^T(\mathrm{avg}(\boldsymbol{y})-\boldsymbol{s}) \leq \lambda_2(\boldsymbol{x}-\boldsymbol{s})^T(\boldsymbol{x}-\boldsymbol{s}) + \mu_2(\mathrm{avg}(\boldsymbol{y})-\boldsymbol{s})^T(\mathrm{avg}(\boldsymbol{y})-\boldsymbol{s})\ .$$

Inequality (14) follows from summing up the two inequalities above and using that $M(\boldsymbol{x}) \geq (\boldsymbol{x}-\boldsymbol{s})^T(\boldsymbol{x}-\boldsymbol{s})$ and that $M(\boldsymbol{y}) = (\boldsymbol{y}-\boldsymbol{s})^T(\boldsymbol{y}-\boldsymbol{s}) + (\mathrm{avg}(\boldsymbol{y})-\boldsymbol{s})^T(\mathrm{avg}(\boldsymbol{y})-\boldsymbol{s})$. □

Using Proposition 6 and Proposition 7, we obtain that for all $\boldsymbol{x}, \boldsymbol{y} \in [0,1]^n$, and for all $\lambda_1, \lambda_2 > 0$ and $\mu_1, \mu_2 \in (0,1)$ such that $\lambda_1\mu_1 \geq 1$ and $\lambda_2\mu_2 \geq 1/n^2$,

$$M(\boldsymbol{y}) + (\boldsymbol{x}-\boldsymbol{y})^T M'(\boldsymbol{y}) \leq \left(1 + \frac{1}{n^2} + \lambda_1 + \lambda_2\right) M(\boldsymbol{x}) + \max\{\mu_1,\mu_2\}M(\boldsymbol{y})\,. \tag{15}$$

Applying Proposition 3 with $\lambda = 1$ and $\mu = \sqrt{2} - 1$, and (15) with $\lambda_1 = \sqrt{2} + 1$, $\lambda_2 = 1/n$, $\mu_1 = \sqrt{2} - 1$ and $\mu_2 = 1/n$, and summing up the corresponding inequalities, we obtain that (12) holds with $\lambda = 2 + \sqrt{2} + \frac{n+1}{n^2}$ and $\mu = \sqrt{2} - 1$. Hence, we conclude that

$$\mathrm{PoA} \leq (2+\sqrt{2})^2/2 + (\sqrt{2}+1)\frac{n+1}{n^2}\,. \qquad \square$$

Acknowledgments Supported by DFG grant EXC 284 (Cluster of Excellence MMCI at Saarland University).

References

1. Altafini, C.: Consensus problems on networks with antagonistic interactions. IEEE Trans. Autom. Control **58**(4), 935–946 (2013)
2. Auletta, V., Caragiannis, I., Ferraioli, D., Galdi, C., Persiano, G.: Generalized discrete preference games. In: Proc. of the 25th International Joint Conference on Artificial Intelligence (IJCAI '16), pp. 53–59 (2016)
3. Barberà, S.: An introduction to strategyproof social choice functions. Soc. Choice Welf. **18**, 619–653 (2001)

4. Bhawalkar, K., Gollapudi, S., Munagala, K.: Coevolutionary opinion formation games. In: Proc. of the 45th ACM Symposium on Theory of Computing (STOC '13), pp. 41–50 (2013)
5. Bilò, V., Fanelli, A., Moscardelli, L.: Opinion formation games with dynamic social influences. In: Proc. of the 12th Conference on Internet and Network Economics (WINE '16), volume 10123 of LNCS, pp. 444–458 (2016)
6. Bindel, D., Kleinberg, J.M., Oren, S.: How bad is forming your own opinion? In: Proc. of the 52nd IEEE Symposium on Foundations of Computer Science (FOCS '11), pp. 57–66 (2011)
7. Chazelle, B., Wang, C.: Inertial Hegselmann-Krause systems. IEEE Trans. Autom. Control **62**, 3905–3913 (2017)
8. Chen, P.-A., Chen, Y.-L., Lu, C.-J.: Bounds on the price of anarchy for a more general class of directed graphs in opinion formation games. Oper. Res. Lett. **44**(6), 808–811 (2016)
9. DeGroot, M.H.: Reaching a consensus. J. Am. Stat. Assoc. **69**, 118–121 (1974)
10. Ferraioli, D., Goldberg, P., Ventre, C.: Decentralized dynamics for finite opinion games. Theor. Comput. Sci. **648**, 96–115 (2016)
11. Fotouhi, B., Rabbat, M.G.: The effect of exogenous inputs and defiant agents on opinion dynamics with local and global interactions. IEEE J. Selected Topics Signal Process. **7**(2), 347–357 (2013)
12. Friedkin, N.E., Johnsen, E.C.: Social influence and opinions. J. Math. Sociol. **15**(3-4), 193–205 (1990)
13. Ghaderi, J., Srikant, R.: Opinion dynamics in social networks with stubborn agents: equilibrium and convergence rate. Automatica **50**, 3209–3215 (2014)
14. Golub, B., Jackson, M.O.: Naï. Am. Econ. J. Microecon. **2**(1), 112–149 (2010)
15. Hegselmann, R., Krause, U.: Opinion dynamics and bounded confidence models, analysis, and simulation. J. Artif. Societ. Soc. Simul, 5 (2002)
16. Jackson, M.O.: Social and Economic Networks. Princeton University Press (2008)
17. Moulin, H.: On strategy-proofness and single-peakedness. Public Choice **35**, 437–455 (1980)
18. Quattrociocchi, W., Caldarelli, G., Scala, A.: Opinion dynamics on interacting networks: media competition and social influence. Sci. Rep. **4**, 4938 (2014)
19. Rosen, J.B.: Existence and uniqueness of equilibrium points in concave n-person games. Econometrica **33**, 520–534 (1965)
20. Roughgarden, T., Schoppmann, F.: Local smoothness and the price of anarchy in splittable congestion games. J. Econ. Theory **156**, 317–342 (2015)
21. Yildiz, E., Ozdaglar, A., Acemoglu, D., Saberi, A., Scaglione, A.: Binary opinion dynamics with stubborn agents. ACM Trans. Econ. Comput. **1**(4), 19,1–19,30 (2013)

Affiliations

Markos Epitropou[1] · Dimitris Fotakis[2,3] · Martin Hoefer[4] · Stratis Skoulakis[3]

Markos Epitropou
mep@seas.upenn.edu

Dimitris Fotakis
dfotakis@oath.com; fotakis@cs.ntua.gr

Martin Hoefer
mmhoefer@cs.uni-frankfurt.de

1 Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA

2 Yahoo Research, New York, NY, USA

3 School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece

4 Institut für Informatik, Goethe-Universität Frankfurt am Main, Frankfurt, Germany