Two methods for the maximization of homogeneous polynomials over the simplex

Faizan Ahmed¹ · Georg Still²

Received: 21 December 2020 / Accepted: 29 July 2021 / Published online: 5 August 2021
© The Author(s) 2021

Abstract

The paper deals with the numerical solution of the problem P to maximize a homogeneous polynomial over the unit simplex. We discuss the convergence properties of the so-called replicator dynamics for solving P . We further examine an ascent method, which also makes use of the replicator transformation. Numerical experiments with polynomials of different degrees illustrate the theoretical convergence results.

Keywords Optimization over the simplex · Homogeneous polynomials · Symmetric tensors · Numerical methods · Replicator transformation · Evolutionarily stable strategies · Convergence properties

Mathematics Subject Classification 90C26 · 91A22 · 05C69

1 Introduction

The present paper studies two special methods for computing local or global maximizers of homogeneous polynomials of degree d over the unit simplex in $\mathbb{R}^n$. The maximization problem is referred to as P (cf., (1)). We are especially interested in the so-called replicator iteration to solve P . The mathematical structure and theoretical properties of P have been investigated, e.g., in [5, 6] for $d = 2$ and, e.g., in [1] for $d > 2$. The problem considered has many applications, such as portfolio decisions [15], evolutionary biology (see [3, 21] for $d = 2$, and [20] for $d > 2$), as well as maximum clique problem in (hyper-) graphs [9].

✉ Faizan Ahmed
faizan.ahmed@utwente.nl

Georg Still
g.j.still@utwente.nl

¹ Department of Computer Science, University of Twente, Enschede, The Netherlands

² Department of Mathematics, University of Twente, Enschede, The Netherlands

It is well-known that for $d \geq 2$ the global maximization problem P is NP-hard (cf., [19]). Approximation methods for computing a global maximizer have been investigated in [2] for $d = 2$, and in [12, 13] for $d > 2$. However, in the present article we are less interested in the solution of the global maximization problem but we are looking for an algorithm to compute local maximizers of P . For this purpose we focus our attention to the so-called replicator dynamics (cf., (20) below). The useful monotonicity properties of the replicator transformation are well-known (see [2, 14]). The convergence of the replicator iteration to a (unique) critical point of P has been proven in [16] for $d = 2$. Also for $d = 2$, results on the rate of convergence of the replicator dynamics have been presented in [17].

In the present paper, we are especially interested in the convergence properties of the replicator method for $d > 2$. It appears that (as for $d = 2$) a linear convergence to a critical point $\bar{x}$ of P can only occur if a strict complementarity condition is satisfied at $\bar{x}$. We therefore also examine an appropriate ascent method to increase the convergence speed. Another main aim of the paper is to provide a selection of numerical experiments with many example problems for objective functions of different degrees d .

The paper is organized as follows. Section 2 introduces the maximization problem P , presents properties of the homogeneous objective function $f(x)$ in terms of the corresponding n -dimensional symmetric tensor A of order d , and summarizes useful optimality conditions for P .

In Sect. 3, we describe two important applications of our maximization problem. Section 4 contains the main results of the paper. We study the properties of the replicator method to find local maximizers of P . More precisely, in Sect. 4.1 we discuss the behavior of the replicator transformation $y(x)$ (cf., (16)). It appears that the fixed points of this transformation $y(x)$ coincide with the critical points of P . We summarize known convergence results of the replicator iteration for the case $d = 2$ and present new convergence results for $d \geq 2$. Section 4.3 shows how the replicator transformation $y(x)$ can be used to set up an ascent method, enjoying a faster convergence in comparison with the pure replicator dynamics. Section 5 presents many numerical experiments with the replicator iteration and the special ascent method for example objective functions $f(x)$ of different degrees d . Finally, Sect. 6 is devoted to concluding remarks.

2 Preliminaries

In the present paper, we consider the problem to compute local or global maximizers of an (in general nonconvex) homogeneous polynomial of degree d over the unit simplex in $\mathbb{R}^n$:

$$P : \quad \max f(x) \quad \text{s.t.} \quad x \in \Delta_n := \{x \in \mathbb{R}^n \mid e^T x = 1, x \geq 0\}, \quad (1)$$

where

$$f(x) = \sum_{i_1, \dots, i_d=1}^n a_{i_1, \dots, i_d} x_{i_1} x_{i_2} \cdots x_{i_d}$$

is a homogeneous polynomial of degree d given by the d^{th} order n -dimensional tensor $A = (a_{i_1, \dots, i_d})_{1 \leq i_1, \dots, i_d \leq n}$. As usual, $e \in \mathbb{R}^n$ is the vector with all components equal to 1. We define the index set $N := \{1, \dots, n\}$.

In this paper, we always assume that the tensor A is symmetric, i.e.,

$$a_{i_1, \dots, i_d} = a_{i_{\sigma(1)}, \dots, i_{\sigma(d)}} \quad \text{for all permutations } \sigma \text{ of the numbers } \{1, \dots, d\}.$$

The space of all symmetric tensors of order d in n -variables has dimension $\binom{n+d-1}{d}$ and is denoted by $\mathcal{S}^{n,d}$. For $A, B \in \mathcal{S}^{n,d}$ we define the inner product $A \cdot B$ by,

$$A \cdot B = \sum_{i_1, \dots, i_d=1}^n a_{i_1, \dots, i_d} b_{i_1, \dots, i_d}.$$

For $x \in \mathbb{R}^n$ we introduce the symmetric tensor $x^d = (x_{i_1} x_{i_2} \cdots x_{i_d})_{1 \leq i_1, \dots, i_d \leq n} \in \mathcal{S}^{n,d}$. Then the polynomial $f(x)$ of order d can be written as

$$f(x) = A \cdot x^d.$$

If we make use of the fact that $f(x)$ is a d -linear form we write

$$\underbrace{F(x, x, \dots, x)}_{d \text{ times}} \quad \text{instead of} \quad f(x). \quad (2)$$

We further introduce the tensors of order $d-1$:

$$A^j = (a_{j, i_2, \dots, i_d})_{1 \leq i_2, \dots, i_d \leq n}, \quad j = 1, \dots, n \quad \text{and} \quad x^{d-1} = (x_{i_2} x_{i_3} \cdots x_{i_d})_{1 \leq i_2, \dots, i_d \leq n}.$$

With these abbreviations the following holds for the elements of the gradient $\nabla f(x)$ and the Hessian $\nabla^2 f(x)$,

$$\frac{\partial f(x)}{\partial x_i} = d A^i \cdot x^{d-1} = d \sum_{i_2, \dots, i_d=1}^n a_{i, i_2, \dots, i_d} x_{i_2} \cdots x_{i_d}, \quad (3)$$

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = d(d-1) \sum_{i_3, \dots, i_d=1}^n a_{i, j, i_3, \dots, i_d} x_{i_3} \cdots x_{i_d}. \quad (4)$$

Interested in (local) maximizers of P we summarize some well-known facts needed below. For $\bar{x} \in \Delta_n$ the KKT condition is said to hold if with (unique) Lagrangian multipliers $\lambda \in \mathbb{R}$, $0 \leq \mu \in \mathbb{R}^n$, we have

$$\nabla f(\bar{x}) - \lambda e + \mu = 0 \quad \text{and} \quad \mu_i \bar{x}_i = 0, \quad i \in N. \quad (5)$$

We call such a point $\bar{x}$ a KKT point of P . A feasible point $\bar{x} \in \Delta_n$ satisfying (5) with $\mu \in \mathbb{R}^n$ (without the non-negativity condition) is called a critical point of P . The KKT point $\bar{x}$ (critical point $\bar{x}$) is said to satisfy strict complementarity (SC) if

$$(SC) : \quad \mu_i > 0 \quad (\mu_i \neq 0) \text{ holds for all } i \text{ with } \bar{x}_i = 0. \quad (6)$$

For a KKT point $\bar{x} \in \Delta_n$ with corresponding (unique) multiplier $0 \leq \mu \in \mathbb{R}^n$ the necessary second order (NSOC) and the sufficient second order conditions (SSOC) are given by

$$\begin{aligned} (NSOC) : \quad & s^T \nabla^2 f(\bar{x}) s \leq 0 \quad \forall s \in C_{\bar{x}}, \\ (SSOC) : \quad & s^T \nabla^2 f(\bar{x}) s < 0 \quad \forall 0 \neq s \in C_{\bar{x}}, \end{aligned} \quad (7)$$

where $C_{\bar{x}}$ is the cone of critical directions,

$$C_{\bar{x}} = \{s \in \mathbb{R}^n \mid e^T s = 0, s_i = 0, \text{ if } \mu_i > 0, s_i \geq 0, \text{ if } \mu_i = 0\}.$$

Here are the well-known optimality conditions for P (cf., e.g., [1, Lemma 3.3a,b, Remark 3.3]).

Lemma 1

- (a) [necessary conditions] Let $\bar{x} \in \Delta_n$ be a local maximizer of P . Then the KKT condition holds for $\bar{x}$ with NSOC.
- (b) [sufficient conditions] Let $\bar{x} \in \Delta_n$ be a KKT point of P satisfying SSOC, then $\bar{x}$ is a strict local maximizer of P satisfying the following second order growth condition: with some $\varepsilon, c > 0$ it holds

$$f(\bar{x}) \geq f(x) + c \|x - \bar{x}\|^2 \quad \forall x \in \Delta_n, \|x - \bar{x}\| < \varepsilon. \quad (8)$$

Moreover, $\bar{x}$ is an isolated KKT point.

The KKT- and critical point conditions can be written in terms of the tensor A . Componentwise, condition (5) yields using (3),

$$d A^i \cdot \bar{x}^{d-1} - \lambda + \mu_i = 0, \quad \mu_i \bar{x}_i = 0, \quad i \in N. \quad (9)$$

Multiplying with $\bar{x}_i$ and summing up leads to

$$d \sum_{i \in N} \bar{x}_i A^i \cdot \bar{x}^{d-1} - \lambda \sum_{i \in N} \bar{x}_i = 0 \quad \text{or} \quad \lambda = d A \cdot \bar{x}^d.$$

By substituting $\lambda = d A \cdot \bar{x}^d$ into (9) we obtain

$$\frac{\mu_i}{d} = A \cdot \bar{x}^d - A^i \cdot \bar{x}^{d-1}, \quad \mu_i \bar{x}_i = 0, \quad i \in N.$$

So, the critical point condition for $\bar{x}$ is equivalent with

$$A^i \cdot \bar{x}^{d-1} = A \cdot \bar{x}^d \quad \text{for all } i \in N \text{ with } \bar{x}_i > 0, \quad (10)$$

and such a point $\bar{x}$ is a KKT point of P , if in addition we have

$$A^i \cdot \bar{x}^{d-1} \leq A \cdot \bar{x}^d \quad \text{for all } i \in N \text{ with } \bar{x}_i = 0. \quad (11)$$

Moreover, a critical point $\bar{x}$ satisfies the SC condition ($\mu_i \neq 0 \forall i$ such that $\bar{x}_i = 0$) if and only if

$$A^i \cdot \bar{x}^{d-1} \neq A \cdot \bar{x}^d \quad \text{for all } i \in N \text{ with } \bar{x}_i = 0. \quad (12)$$

Remark 1 Note that for any critical point $\bar{x} \in \Delta_n$ of P satisfying $\bar{x} > 0$ ($\bar{x}$ in the relative interior of Δ_n) by definition ($\bar{x}_i \mu_i = 0$ for all i) the multiplier μ (cf., (5)) must be zero. So, any critical point $\bar{x} > 0$ must be a KKT point.

Remark 2 The assumption that a homogeneous polynomial $f(x)$ of degree d ,

$$f(x) = \sum_{i_1, \dots, i_d=1}^n c_{i_1, \dots, i_d} x_{i_1} \cdots x_{i_d}, \quad (13)$$

($c_{i_1, \dots, i_d} \in \mathbb{R}$ arbitrarily) is given in the form

$$f(x) = A \cdot x^d = \sum_{i_1, \dots, i_d=1}^n a_{i_1, \dots, i_d} x_{i_1} \cdots x_{i_d}, \quad (14)$$

with symmetric coefficients $a_{i_1, \dots, i_d}$, i.e., $A = (a_{i_1, \dots, i_d})_{1 \leq i_1, \dots, i_d \leq n} \in S^{n,d}$, does not mean any restriction. Indeed, let be given a polynomial in the form (13). Take any of the subsets $\{j_1, \dots, j_d\} \subset N$ such that $j_1 \leq j_2 \leq \dots \leq j_d$. Let $\mathcal{S} = \mathcal{S}(j_1, \dots, j_d)$ be the set of all permutations $\sigma = \sigma(j_1, \dots, j_d)$ of the numbers in $\{j_1, \dots, j_d\}$ without repetition. Then by

$$a_{i_1, \dots, i_d} = \frac{1}{|\mathcal{S}(j_1, \dots, j_d)|} \sum_{\{k_1, \dots, k_d\} \in \mathcal{S}(j_1, \dots, j_d)} c_{k_1, \dots, k_d}, \quad \{i_1, \dots, i_d\} \in \mathcal{S}(j_1, \dots, j_d),$$

we have defined a symmetric tensor A such that $f(x) = A \cdot x^d$.

Remark 3 In this paper, we study problem P for the case of homogeneous polynomials $f(x)$. But this does not mean a real restriction. Indeed suppose we have given a non-homogeneous polynomial $f(x)$ of degree d . We then can simply transform $f(x)$ into homogeneous form as follows. By multiplying each monomial $\prod_{i=1}^n x_i^{\beta_i}$ with $\sum_{i=1}^n \beta_i < d$ appearing in f by the factor $(\sum_{i=1}^n x_i)^{d - \sum_{i=1}^n \beta_i}$, f becomes a homogeneous polynomial of degree d and in view of $\sum_{i=1}^n x_i = 1$ this does not change the function values of f on Δ_n .

We finally give some notation used throughout the paper. As usual, e_i , $i \in N$, denote the standard basis vectors in $\mathbb{R}^n$, $\|x\|$ is the Euclidean norm of $x \in \mathbb{R}^n$, and

for a tensor $A \in \mathcal{S}^{n,d}$ we define the (Frobenius) norm $\|A\| = \sqrt{A \cdot A}$. This tensor norm satisfies $|A \cdot B| \leq \|A\| \|B\|$ and for $x \in \Delta_n$ we find $\|x^d\| \leq 1$.

3 Applications

Polynomial optimization over the simplex has numerous applications. In this section we describe two of them. For both we present numerical experiments in Sect. 5.

3.1 Evolutionarily stable strategies

The first application depends on the close relation between the maximization problem P and a model in evolutionary biology. For $d = 2$ the model goes back to Maynard Smith and Price [22] and for $d > 2$ we refer to [7, 20].

Let us consider a population of individuals which differ in n distinct features (also called strategies):

- For $x = (x_1, \dots, x_n) \in \Delta_n$, the component x_i gives the percentage of the population having feature i . So, x gives the strategy (state) of the whole population.
- We are given a symmetric fitness tensor $A \in \mathcal{S}^{n,d}$. The value $A \cdot x^d$ represents the (mean) fitness of a population with strategy x .

In evolutionary biology special strategies called evolutionarily stable strategies (ESS) play an important role.

Definition 1 [ESS in biology] A point $x \in \Delta_n$ is called an ESS for $A \in \mathcal{S}^{n,d}$ if for all $x \neq y \in \Delta_n$ the following conditions hold for the d -linear form $f(x) = A \cdot x^d \equiv F(x, \dots, x)$:

1. $F(y, x, \dots, x) \leq F(x, x, \dots, x)$
2. and if $F(y, x, \dots, x) = F(x, x, \dots, x)$ then $F(y, y, x, \dots, x) \leq F(y, x, \dots, x)$
- ⋮
- d. and if $F(y, x, \dots, x) = F(y, y, x, \dots, x) = \dots = F(y, \dots, y, x)$ then $F(y, \dots, y, y) < F(y, \dots, y, x)$.

If only the first relation $F(y, x, \dots, x) \leq F(x, x, \dots, x) \forall y \in \Delta_n$ holds, the strategy x is called a symmetric Nash equilibrium (see, e.g., [1, Remark 3.1]).

The following relation is well-known, see, e.g., [1, Lemma 3.1, Lemma 3.2].

Lemma 2 Let be given $A \in \mathcal{S}^{n,d}$ and $x \in \Delta_n$. Then, x is an ESS for A if and only if x is a strict local maximizer of P . Moreover, x is a symmetric Nash equilibrium if and only if x is a KKT point of P .

3.2 Cliques in a d -graph

The optimization problem P is also connected with the maximum clique problem on graphs ($d = 2$) and on hypergraphs ($d > 2$).

Definition 2 Let $G = (N, E)$ be a hypergraph with set $N = \{1, 2, \dots, n\}$ of vertices and set $E \subset P(N) \setminus \emptyset$ of (hyper) edges $e \in E$. Here, $P(N)$ is the power set of N (set of all subsets of N). Let for $d \geq 2$ the set $P_d(N)$, $P_d(N) \subset P(N)$, denote the set of all subsets in $P(N)$ with exactly d elements. A hypergraph G is called a d -graph if all edges $e \in E$ are in $P_d(N)$. A 2-graph is also called a graph. A subset C of N is a clique in the d -graph G if $P_d(C) \subset E$. A clique C is called maximal if it is not contained in any other clique, and it is said to be a maximum clique if C has maximum cardinality among all cliques. The complement $\bar{G}$ of a d -graph G is defined as the d -graph $\bar{G} = (N, \bar{E})$ where $\bar{E} = P_d(N) \setminus E$.

As we will see now, there is a close relation between the maximal/maximum clique problem in a d -graph and special problems P . For $d = 2$ (graphs) we refer the reader to [4, 18] and for $d > 2$ (hypergraphs) to [9].

Let us introduce the so-called Lagrangean $L_{\bar{G}}$ of the complement $\bar{G}$ of a d -graph G ,

$$L_{\bar{G}}(x) = \sum_{e \in \bar{E}} \left(\prod_{i \in e} x_i \right), \quad x \in \mathbb{R}^n,$$

and for $\tau > 0$ the homogeneous polynomial of degree d ,

$$h_{\bar{G}}^{\tau}(x) = -L_{\bar{G}}(x) - \tau \sum_{i=1}^n x_i^d.$$

Then consider the optimization problem:

$$\max h_{\bar{G}}^{\tau}(x) \quad \text{s.t.} \quad x \in \Delta_n. \quad (15)$$

Define for a subset $S \subset N$ the characteristic vector $x^S \in \Delta_n$ by

$$x_i^S = \begin{cases} \frac{1}{|S|} & \text{if } i \in S \\ 0 & \text{otherwise} \end{cases}.$$

Then the following holds (see [9, Theorem 5] for a proof).

Theorem 1 Let $G = (N, E)$ be a d -graph and let $0 < \tau \leq \frac{1}{d(d-1)}$ ($\tau < \frac{1}{d(d-1)}$ for $d = 2$). Then $x \in \Delta_n$ is a strict local (global) maximizer of (15) if and only if $x = x^C$ is the characteristic vector of a maximal (maximum) clique C of G .

4 Numerical methods based on the replicator transformation

In this section, we are going to discuss two methods for computing local maximizers of P which will be based on the following so-called replicator transformation:

For given $A \in \mathcal{S}^{n,d}$ and $x \in \Delta_n$ define the transformation

$$y : \Delta_n \rightarrow \Delta_n, \quad y(x) \text{ has components } y_i(x) = \frac{x_i(A^i \cdot x^{d-1})}{A \cdot x^d}, \quad i \in N. \quad (16)$$

We emphasize that to assure that the image $y(x)$ remains in Δ_n we have to make sure that the values $A \cdot x^d$ and $A^i \cdot x^{d-1}$ remain positive. To do so throughout the paper we assume

$$A > 0, \quad \text{i.e., all elements of } A \text{ are positive.} \quad (17)$$

In Sect. 4.4 lateron, we will see that this assumption does not mean any real restriction.

4.1 Properties of the replicator transformation $y(x)$

As usual a point $\bar{x} \in \Delta_n$ is called a fixed point of the transformation y in (16) if $y(\bar{x}) = \bar{x}$ holds or equivalently if

$$A^i \cdot \bar{x}^{d-1} = A \cdot \bar{x}^d \quad \text{for all } i \in N \text{ with } \bar{x}_i > 0. \quad (18)$$

A comparison with (10) reveals the following key connection between the transformation y and the maximization problem P . For $\bar{x} \in \Delta_n$ it holds,

$$\bar{x} \text{ is a fixed point of } y \quad \Leftrightarrow \quad \bar{x} \text{ is a critical point of } P \quad (19)$$

It is well-known that the transformation $y(x)$ enjoys a monotonicity property.

Lemma 3 (Monotonicity) *For any $x \in \Delta_n$ and $y(x)$ given by (16) it holds for $f(x) = A \cdot x^d$,*

$$f(y(x)) \geq f(x),$$

with equality if and only if $y(x) = x$ (i.e., x is a fixed point of y).

Proof For a proof we refer to [2, 14]. The proofs are valid for general (non-symmetric) tensors. $\square$

In [17] a stronger version of the monotonicity result has been proven for $d = 2$:

$$f(y(x)) - f(x) \geq \sum_{i \in N} x_i (A^i \cdot x^{d-1} - A \cdot x^d)^2.$$

It is not difficult to show that this formula remains true for $d > 2$.

The monotonicity behavior of $y(x)$ gives rise to the following replicator iteration for solving the maximization problem P : Start with some $x^{(0)} \in \Delta_n$ and iterate (cf., (16))

$$x^{(k+1)} = y(x^{(k)}), \quad k = 0, 1, \dots \quad (20)$$

In algorithmic form this iteration leads to Method 1.

Method 1 Replicator dynamics

Input: $x^{(0)} \in \Delta_n, A \in \mathcal{S}^{n,d}, \varepsilon > 0$

- 1: $k = 0$
 - 2: while $\|y(x^{(k)}) - x^{(k)}\| > \varepsilon$ do
 - 3: Update: $x^{(k+1)} = y(x^{(k)})$
 - 4: $k = k + 1$
 - 5: end while
-

By a simple compactness and continuity argument combined with Lemma 3, we obtain (see [16, Proposition 1]).

Lemma 4 Any sequence (20) has a limit point (at least one) and each limit point $\bar{x}$ of this iteration is a fixed point of y , i.e., a critical point of P (cf., (19)).

Unfortunately, limit points $\bar{x}$ of (20) (i.e., critical points of P) need not satisfy the necessary KKT condition for a local maximizer of P (see Lemma 1(a)). If however the iterates $x^{(k)}$ in (20) converge to a fixed point $\hat{x}$ of y which is not a KKT point of P (cf., also Remark 1), then the following lemma could be used to jump away from $\hat{x}$ to better points.

Lemma 5 Let $\hat{x} \in \Delta_n$ be a fixed point of y but not a KKT point of P , so that by (11) there is some index j_0 with $\hat{x}_{j_0} = 0$ and $A^{j_0} \cdot \hat{x}^{d-1} > A \cdot \hat{x}^d$. Then for all $t \geq 0$ the points $\frac{\hat{x} + te_{j_0}}{1+t}$ are in Δ_n and satisfy

$$f\left(\frac{\hat{x} + te_{j_0}}{1+t}\right) > f(\hat{x}) \quad \text{for all } t > 0 \text{ small enough.}$$

Proof Since $e^{T \frac{\hat{x} + te_{j_0}}{1+t}} = \frac{1}{1+t} (e^T \hat{x} + t) = 1$ and $\frac{\hat{x} + te_{j_0}}{1+t} \geq 0$ for all $t \geq 0$, it follows $\frac{\hat{x} + te_{j_0}}{1+t} \in \Delta_n$. To prove the second statement it suffice to show that the differentiable function $g(t) = f\left(\frac{\hat{x} + te_{j_0}}{1+t}\right)$ satisfies $g'(0) > 0$. By our assumptions we actually obtain

$$\begin{aligned} g'(0) &= \nabla f(\hat{x})(e_{j_0} - \hat{x}) = dA^{j_0} \cdot \hat{x}^{d-1} - d \sum_{i \in N} (A^i \cdot \hat{x}^{d-1}) \hat{x}_i \\ &= d [A^{j_0} \cdot \hat{x}^{d-1} - A \cdot \hat{x}^d] > 0. \end{aligned}$$

□

4.2 Convergence properties of the replicator iteration

Let A and thus $f(x) = A \cdot x^d$ be given, and consider for some starting point $x^{(0)}$ the replicator iterates $x^{(k)}$ in (20). Let $S = S(x^{(0)})$ denote the set of all limit points of the sequence $x^{(k)}$. Recall that by Lemma 4 the set S is nonempty and is contained in the set of all fixed points of the mapping y (cf., (16)).

Lemma 6 *The limit point set S is closed (compact), connected, and all points $\bar{x} \in S$ have the same value $f(\bar{x}) = v_S$.*

Proof A proof is, e.g., to be found in [16, Proposition 1], where the case $d = 2$ is considered. The proof is also valid for the general case $d \geq 2$. $\square$

More difficult is to answer the following two questions.

- Is for each starting point $x^{(0)} \in \Delta_n$ the replicator iteration (20) convergent, i.e., is $S = S(x^{(0)}) = \{\bar{x}\}$ a singleton?
- If $x^{(k)}$ converges to (a unique) $\bar{x}$, how fast is this convergence, i.e.,

how fast goes $\|x^{(k)} - \bar{x}\|$ to zero?

For $d = 2$ the first question has been answered positively in [17, Theorem 3.2] and in [16, Convergence Theorem 2]. However, a generalisation of the proofs to the case $d > 2$ seems to be difficult. The arguments in these proofs essentially make use of the fact that (for $d = 2$) the function $f(x)$ is quadratic.

For $d = 2$ also the second question has been answered in [17] as follows. Recall that the convergence of $x^{(k)} \rightarrow \bar{x}$ is said to be linear, if with some constants $C > 0$ and $q, 0 \leq q < 1$, we have

$$\|x^{(k)} - \bar{x}\| \leq C \cdot q^k \quad \forall k \in \mathbb{N}.$$

The constant q is called linear convergence factor.

Theorem 2 *Let $d = 2$ and let the iterates $x^{(k)}$ in (20) converge to $\bar{x}$. (Recall that in this case $\bar{x}$ is a fixed point of y and a critical point of P , see Lemma 4). Then:*

- (a) *With some constant $C_1 > 0$ it holds*

$$\|x^{(k)} - \bar{x}\| \leq C_1 \cdot \frac{1}{\sqrt{k}} \quad \forall k \in \mathbb{N}.$$

- (b) *The convergence is linear if and only if the critical point $\bar{x}$ of P satisfies the SC condition*

$$A^i \cdot \bar{x}^{d-1} - A \cdot \bar{x}^d \neq 0 \quad \text{for all } i \in N \text{ with } \bar{x}_i = 0 \text{ and } x_i^{(0)} > 0. \quad (21)$$

Proof For (a) we refer to [17, Theorem 3.3] and for (b) to [17, Theorem 3.4]. $\square$

Again, we did not see a way to generalize the results of Theorem 2 to the case $d > 2$ (except for the only if direction in Theorem 2(b), cf., Theorem 3 below). The reason is that the proof of the preceding theorem is based on Lemmas 2.3, 2.4 in [17] which again essentially depend on the fact that for $d = 2$ the function $f(x)$ is a quadratic polynomial.

Notice that if the SC condition (21) is not fulfilled the convergence must be sub-linear and according to Theorem 2(a) the convergence might be extremely slow.

Recall that for $d = 2$ (cf., [17, Theorem 3.2]), any iteration (20) converges to a (unique) point $\bar{x}$. Since we did not succeed in a generalization of this statement to $d > 2$, we will at least give a partial convergence result for the general case $d \geq 2$.

Theorem 3 *For any $d \geq 2$ the following holds.*

- (a) *Let the iterates $x^{(k)}$ given by (20) have a limit point $\bar{x}$ such that $\bar{x}$ is a KKT point with corresponding (unique) multiplier $0 \leq \mu \in \mathbb{R}^n$, satisfying the second order condition SSOC (cf., (7)). Then the sequence $x^{(k)}$ converges to $\bar{x}$, i.e., the point $\bar{x}$ is the only limit point.*
- (b) *Suppose $\bar{x}$ is a KKT point of P satisfying the second order condition SSOC, as well as the SC condition (12). Then there is some $\delta_1 > 0$ such that the following is true: The point $\bar{x}$ is the only fixed point $\hat{x} \in \Delta_n$ of $y(x)$ with $\|\hat{x} - \bar{x}\| < \delta_1$, and for any starting point $x^{(0)} \in \Delta_n$, with $\|x^{(0)} - \bar{x}\| < \delta_1$ the replicator iterates (20) converge to $\bar{x}$.*

Proof By the assumptions the limit point $\bar{x}$ (see Lemma 1(b)) satisfies the relation (8). So let $\varepsilon > 0, c > 0$ be given as in (8). The mapping $y(x)$ is a continuous function of x (recall that by assumption we have $A \cdot x^d \geq m > 0, \forall x \in \Delta_n$). So, in view of $y(\bar{x}) - \bar{x} = 0$, given ε , we can choose some $\delta_0 > 0$ such that

$$\|y(x) - x\| < \frac{\varepsilon}{2} \quad \text{for all } x \text{ satisfying } \|x - \bar{x}\| < \delta_0. \quad (22)$$

Clearly, we can choose $\delta_0 < \frac{\varepsilon}{2}$. For any $\delta, 0 < \delta < \delta_0$, there is some δ_1 such that

$$\frac{f(\bar{x}) - f(x)}{c} < \delta^2 \quad \forall x \text{ with } \|x - \bar{x}\| < \delta_1.$$

Again take $\delta_1 < \delta$. Since $\bar{x}$ is a limit point of $x^{(k)}$ we can find an iterate $x^{(k_\delta)}$ satisfying $\|x^{(k_\delta)} - \bar{x}\| < \delta_1$. By (22) we then have

$$\|y(x^{(k_\delta)}) - \bar{x}\| \leq \|y(x^{(k_\delta)}) - x^{(k_\delta)}\| + \|x^{(k_\delta)} - \bar{x}\| < \frac{\varepsilon}{2} + \delta < \frac{\varepsilon}{2} + \delta_0 < \varepsilon.$$

In view of (8) and the monotonicity condition $f(y(x^{(k_\delta)})) \geq f(x^{(k_\delta)})$ it follows

$$\|x^{(k_\delta+1)} - \bar{x}\|^2 = \|y(x^{(k_\delta)}) - \bar{x}\|^2 \leq \frac{f(\bar{x}) - f(y(x^{(k_\delta)}))}{c} \leq \frac{f(\bar{x}) - f(x^{(k_\delta)})}{c} < \delta^2,$$

i.e., $\|x^{(k_\delta+1)} - \bar{x}\| < \delta$. We can repeat the same arguments and obtain $\|y(x^{(k_\delta+1)}) - \bar{x}\| < \varepsilon$, $\|x^{(k_\delta+2)} - \bar{x}\|^2 \leq \delta^2$ and thus for all $k \geq k_\delta$ the iterates $x^{(k)}$ stay in the neighborhood $\{x \in \Delta_n \mid \|x - \bar{x}\| < \delta\}$. Since δ can be chosen arbitrarily small it follows that the sequence $x^{(k)}$ must converge to $\bar{x}$.

(b) By continuity, using SC, there exists $\delta_2 > 0$ such that (cf., (11), (12))

$$\begin{aligned} A^i \cdot x^{d-1} &< A \cdot x^d \quad \forall i \in N \text{ with } \bar{x}_i = 0 \\ \bar{x}_i > 0 &\Rightarrow x_i > 0 \quad \forall i \in N \end{aligned}, \quad \text{for all } x \text{ with } \|x - \bar{x}\| < \delta_2. \quad (23)$$

It now follows that if $\hat{x} \in \Delta_n$, $\|\hat{x} - \bar{x}\| < \delta_2$, is a fixed point of y , then it must be a KKT point of P . Indeed, by (23) we obtain $\hat{x}_i = 0 \Rightarrow \bar{x}_i = 0$ and thus by the first relation of (23) we find

$$A^i \cdot \hat{x}^{d-1} < A \cdot \hat{x}^d \quad \forall i \in N \text{ with } \hat{x}_i = 0.$$

Consequently, in view of (11) $\hat{x}$ is a KKT point. According to the last statement of Lemma 1(b), by making δ_2 smaller if necessary, $\hat{x}$ must coincide with $\bar{x}$, so that $\bar{x}$ is the only fixed point in a δ_2 -neighborhood of $\bar{x}$.

To prove the convergence statement recall the arguments in the proof of part (a) and choose δ in (a) smaller than δ_2 . Then if we take a starting point $x^{(0)} \in \Delta_n$ satisfying $\|x^{(0)} - \bar{x}\| < \delta_1$ (with δ_1 in (a)) by the arguments above, for all iterates $x^{(k)}$ the relation

$$\|x^{(k)} - \bar{x}\| < \delta < \delta_2 \quad \forall k \in \mathbb{N} \quad (24)$$

holds. To show that $x^{(k)}$ converges to $\bar{x}$, suppose to the contrary that the sequence $x^{(k)}$ has a limit point $\hat{x}$, $\hat{x} \neq \bar{x}$. By (24) we have $\|\hat{x} - \bar{x}\| \leq \delta < \delta_2$ in contradiction to the fact that $\bar{x}$ is the only fixed point satisfying $\|\hat{x} - \bar{x}\| < \delta_2$. $\square$

Fortunately, the proof of the only if part of the statement in Theorem 2(b) remains valid for $d > 2$.

Theorem 4 *For all $d \geq 2$ the following holds. Let the sequence $x^{(k)}$ given by (20) converge linearly to the fixed point $\bar{x}$ of y . Then the following SC conditions must hold:*

$$A^i \cdot \bar{x}^{d-1} < A \cdot \bar{x}^d \quad \text{for all } i \in N \text{ with } \bar{x}_i = 0 \text{ and } x_i^{(0)} > 0. \quad (25)$$

Proof For completeness we reproduce the proof from [17, Theorem 3.3]. Notice first that by definition (cf., (16)) the relation $x_i^{(0)} = 0$ implies $x_i^{(k)} = 0$ for all k and $\bar{x}_i = 0$. Moreover, by (20) it follows

$$x_i^{(0)} > 0 \quad \Rightarrow \quad \frac{x_i^{(k+1)}}{x_i^{(k)}} = \frac{A^i \cdot [x^{(k)}]^{d-1}}{A \cdot [x^{(k)}]^d} > 0,$$

and thus

$$\lim_{k \rightarrow \infty} \frac{x_i^{(k+1)}}{x_i^{(k)}} = \frac{A^i \cdot \bar{x}^{d-1}}{A \cdot \bar{x}^d} \quad \text{if } x_i^{(0)} > 0.$$

By [11, Proposition 2.26] we have

$$\lim_{k \rightarrow \infty} \left(x_i^{(k)} \right)^{1/k} = \lim_{k \rightarrow \infty} \frac{x_i^{(k+1)}}{x_i^{(k)}} = \frac{A^i \cdot \bar{x}^{d-1}}{A \cdot \bar{x}^d}. \quad (26)$$

Now let $x^{(k)}$ converge linearly to $\bar{x}$, i.e., with some $C > 0$, $0 < q < 1$, it holds componentwise

$$|x_i^{(k)} - \bar{x}_i| \leq Cq^k \quad \forall k \in \mathbb{N}.$$

So for any index i such that $x_i^{(0)} > 0$, $\bar{x}_i = 0$ (the nontrivial case), this implies

$$0 < x_i^{(k)} \leq Cq^k \quad \text{and} \quad \lim_{k \rightarrow \infty} \left(x_i^{(k)} \right)^{1/k} \leq q.$$

Consequently using (26) we obtain

$$A^i \cdot \bar{x}^{d-1} \leq q A \cdot \bar{x}^d < A \cdot \bar{x}^d \quad \text{for all } i \in N \text{ with } \bar{x}_i = 0 \text{ and } x_i^{(0)} > 0.$$

□

We emphasize that the SC condition (25) in Theorem 4 is slightly stronger than the SC condition (21) in Theorem 2(b). The former in particular implies that if we start with $x^{(0)} > 0$ then linear convergence can only occur if $x^{(k)}$ converges to a KKT point $\bar{x}$. Recall (cf., Remark 1) that any fixed point $\bar{x}$ of y satisfying $\bar{x} > 0$ must be a KKT point.

4.3 Speeding up the convergence by line maximization

Notice that by Theorem 4, linear convergence of the replicator iteration (20) to a fixed point $\bar{x}$ of y can only happen if the SC condition (25) is met. Keep in mind that this SC condition trivially holds for a fixed point $\bar{x} > 0$ (i.e., $\bar{x}$ is in the relative interior of Δ_n). In other words, for a fixed point $\bar{x}$ where (25) is not fulfilled, a convergence to this point must be sublinear. According to Theorem 2(a) the sublinear convergence might be extremely slow.

On the other hand, even when for larger d (say $d > 4$) linear convergence occurs, the linear convergence factor $q < 1$ might be very close to 1 (see numerical experiments below).

So we may be interested to speed up the convergence by an appropriate ascent method using (some sort of) line maximization. Also here the transformation $y(x)$ can be used to achieve that. Indeed it appears, that the vector $s = y(x) - x$ is a direction of growth for f .

Lemma 7 For all $x \in \Delta_n$ it holds for $y(x)$ in (16),

$$\nabla f(x)^T(y(x) - x) \geq 0 \text{ with equality iff } x \text{ is a fixed point of } y.$$

Proof Using $\sum_i x_i[A^i \cdot x^{d-1}] = A \cdot x^d$, we find for $x \in \Delta_n$,

$$\begin{aligned} \nabla f(x)^T(y(x) - x) &= d \sum_i A^i \cdot x^{d-1} \left(\frac{x_i[A^i \cdot x^{d-1}]}{A \cdot x^d} - x_i \right) \\ &= d \sum_i \left(\frac{x_i[A^i \cdot x^{d-1}]^2}{A \cdot x^d} - x_i[A^i \cdot x^{d-1}] \right) \\ &= d \sum_i \frac{x_i[A^i \cdot x^{d-1}]^2}{A \cdot x^d} - d \sum_i x_i[A^i \cdot x^{d-1}]. \end{aligned}$$

So $f(x)^T(y(x) - x) \geq 0$ holds iff we have

$$\sum_i x_i[A^i \cdot x^{d-1}]^2 \geq \left[\sum_i x_i[A^i \cdot x^{d-1}] \right]^2.$$

The convexity of the real valued function $h(t) = t^2$ yields the inequality

$$\left[\sum_i x_i t_i \right]^2 \leq \sum_i x_i [t_i]^2 \quad \forall x \in \Delta_n, t_i \in \mathbb{R}$$

with equality iff for all $x_i > 0$ the values $t_i = \lambda$ are the same. So taking $t_i := A^i \cdot x^{d-1}$ we have shown that $f(x)^T(y - x) \geq 0$ holds, with equality iff for all $x_i > 0$ the values $A^i \cdot x^{d-1} = \lambda$ are the same. We finally obtain for λ a value

$$\lambda = \sum_i x_i \lambda = \sum_i x_i A^i \cdot x^{d-1} = A \cdot x^d$$

which proves the statement of the lemma. $\square$

The preceding lemma suggests the following ascent method for solving P .

Method 2 Ascent method with search direction $s^{(k)} = y(x^{(k)}) - x^{(k)}$.

Input: $x^{(0)} \in \Delta_n, A \in \mathcal{S}^{n,d}, \varepsilon > 0$

1: $k = 0$

2: while $\|y(x^{(k)}) - x^{(k)}\| > \varepsilon$ do

3: Solve: $t_k = \arg \max_{t \geq 0} \{f(x^{(k)}) + t(y(x^{(k)}) - x^{(k)})\}$ s.t. $x^{(k)} + t(y(x^{(k)}) - x^{(k)}) \geq 0\}$

4: Update: $x^{(k+1)} = x^{(k)} + t_k(y(x^{(k)}) - x^{(k)})$

5: $k = k + 1$

6: end while

It can easily be seen that for the sequence $x^{(k)}$ generated by Method 2 a result as in Lemma 4 remains true. Also Lemma 6 is still valid for the set S of limit points of such a sequence $x^{(k)}$, except for the connectness of S . The reason is that the "pure" line maximization of Method 2 allows the iterates $x^{(k)}$ to jump (infinitely many times) between different (unconnected) fixed points $\bar{x}, \hat{x}$ of S . In our numerical experiments, we however did never see such an event. A zigzagging between different fixed points could however be prevented by an appropriate step-length control in the line maximization.

Efficient implementation of the line maximization step (3): We add some remarks on an efficient implementation of the line maximization step (3) of Method 2. Recall that by Lemma 3 we have for $x = x^{(k)}$ (since $y(x^{(k)}) \neq x^{(k)}$)

$$f(x + t(y(x) - x)) > f(x) \quad \text{for all } t > 0 \text{ small enough.}$$

To find the optimal step-size t we have to solve the line maximization problem (with $x = x^{(k)}$)

$$\max_{0 < t} g(t) := f(x + t(y(x) - x)) \quad \text{s.t.} \quad x + t(y(x) - x) \in \Delta_n. \quad (27)$$

It is easily seen that with $f(x) \equiv F(x, \dots, x)$ (a d -linear form) we have ($y = y(x)$)

$$g(t) = \sum_{i=0}^d \alpha_i t^i \quad \text{where} \quad \alpha_i = \binom{d}{i} F(\underbrace{y-x, \dots, y-x}_{i \text{ times}}, \underbrace{x, \dots, x}_{d-i \text{ times}}). \quad (28)$$

The condition $x + t(y(x) - x) \in \Delta_n$, $t \geq 0$, leads to the restriction

$$t \in [0, t^*] \quad \text{where} \quad t^* := \min_{y_i - x_i < 0} \frac{-x_i}{y_i - x_i}.$$

So we have to solve the line maximization problem

$$\max_{0 < t \leq t^*} g(t),$$

with a polynomial $g(t)$ of degree d . Notice that in view of formula (28) for x close to a fixed point $\bar{x}$ of y the coefficients α_i , $1 \leq i \leq d$, tend to zero according to (recall $y(\bar{x}) - \bar{x} = 0$)

$$\alpha_i = O(\|y(x) - x\|^i) \quad \text{for } x \rightarrow \bar{x}.$$

Consequently, for $x \approx \bar{x}$ the coefficients α_i , $i \geq 3$, will only play a minor role in the line maximization of $g(t)$ and we propose the following efficient procedure.

Approximate line maximization step (3') in Method 2:

- Only compute the coefficients α_1, α_2 ($\alpha_1 > 0$ by Lemma 7), and consider the quadratic approximation $\tilde{g}(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2$ of $g(t)$.
- Compute the solution $\tilde{t} = \frac{-\alpha_1}{2\alpha_2}$ of $\tilde{g}'(t) = 0$, and take as new iterate $x^{(k+1)}$ on the line segment $\{x^{(k)} + t(y(x^{(k)}) - x^{(k)}) \mid t \in (0, t^*]\}$ the point

$$x^{(k+1)} = \begin{cases} x^{(k)} + \bar{t}(y(x^{(k)}) - x^{(k)}) & \text{if } 0 < \bar{t} \leq t^* \\ \arg \max \{f(y(x^{(k)})), f(x^{(k)} + t^*(y(x^{(k)}) - x^{(k)}))\} & \text{otherwise} \end{cases}.$$

We emphasize that for solving this approximate line maximization problem we only have to compute the coefficients α_1, α_2 of $g(t)$. This can be done efficiently, requiring (roughly) the computation time of one function evaluation of f .

Remark 4 The line maximization (27) is particularly easy for $d = 2$. Then (as a simple exercise) we can show: the optimal step-size t_x in (27) is given by

$$t_x = \begin{cases} t^* & \text{if } \alpha_2 := (y - x)^T A x \geq 0 \\ \min\{\frac{-\alpha_1}{2\alpha_2}, t^*\} & \text{if } \alpha_2 < 0 \end{cases}$$

4.4 Dependence of the replicator iteration on a transformation $A \rightarrow A + \alpha E$

We have to recall that the property

$$x \in \Delta_n \quad \Rightarrow \quad y(x) \in \Delta_n$$

and the monotonicity $f(y(x)) \geq f(x)$ for $x \in \Delta_n$ depends on the assumption $f(x) = A \cdot x^d > 0$ and $\frac{\partial}{\partial x_i} f(x) = d A^i \cdot x^{d-1} \geq 0 \forall i$. We have assured these conditions by assuming $A > 0$, i.e., all elements of the tensor A are positive.

In case A has negative elements we can force positivity by applying a transformation $A \rightarrow A + \alpha E$, where $E \in \mathcal{S}^{n,d}$ is the tensor with all elements equal to 1, and $\alpha \in \mathbb{R}$ is chosen such that $A + \alpha E > 0$ holds. So instead of $f(x) = A \cdot x^d$ we consider the function $f^\alpha(x) := (A + \alpha E) \cdot x^d$. Using $E \cdot x^d = 1$ for $x \in \Delta_n$, we obtain the identities

$$f^\alpha(x) = (A + \alpha E) \cdot x^d = f(x) + \alpha(x_1 + \dots + x_n)^d = f(x) + \alpha \quad \text{for } x \in \Delta_n. \quad (29)$$

Fortunately under this transformation the fixed points of $y(x)$ remain the same.

Lemma 8 *Under the transformation $f(x) \rightarrow f^\alpha(x)$, $\alpha \in \mathbb{R}$, the fixed points of y and the KKT points of P , and thus the maximizers of P , remain unchanged.*

Proof Consider the transformation $f(x) \rightarrow f^\alpha(x)$ and the corresponding transformation $y(x) \rightarrow y^\alpha(x)$ where $y^\alpha(x)$ is the vector with coefficients

$$y_i^\alpha(x) = \frac{x_i(A^i + \alpha E) \cdot x^{d-1}}{(A + \alpha E) \cdot x^d} = \frac{x_i(A^i \cdot x^{d-1} + \alpha)}{A \cdot x^d + \alpha}.$$

So the fixed point relation $y^\alpha(x) = x$, $x \in \Delta_n$, is equivalent with

$$A^i \cdot x^{d-1} + \alpha = A \cdot x^d + \alpha \quad \text{if } x_i > 0, \quad \text{or} \quad A^i \cdot x^{d-1} = A \cdot x^d \quad \text{if } x_i = 0,$$

which are the fixed point relations for $y(x)$. The same is true for the additional KKT conditions (cf., (11)). $\square$

Note that formally we have $f = f^0$ and $y = y^0$. Evidently, all results in the preceding subsections remain valid for any function f^α , y^α , as long as $(A + \alpha E) > 0$ is satisfied. However the iterates $x^{(k+1)} = y^\alpha(x^{(k)})$ of the replicator dynamics (20) clearly depend on α . So starting at the same point $x^{(0)}$ the limit points of these iterates and the speed of convergence may strongly depend on the choice of α . We give an illustrative example.

Example 1 For $n = d = 2$ consider the function $f^\alpha(x) = x_1^2 + x_2^2 + x_1x_2 + \alpha$ and iteration $x^{(k+1)} = y^\alpha(x^{(k)})$, where

$$y_1^\alpha(x) = \frac{x_1(x_1 + x_2/2 + \alpha)}{x_1^2 + x_2^2 + x_1x_2 + \alpha}, \quad y_2^\alpha(x) = \frac{x_2(x_2 + x_1/2 + \alpha)}{x_1^2 + x_2^2 + x_1x_2 + \alpha},$$

which has a (repelling) fixed point (minimizer) at $(1/2, 1/2)$ and two maximizers at $\bar{x} = (1, 0)$ and $\tilde{x} = (0, 1)$.

Starting from $x^{(0)} = (0.55, 0.45)$ the iterates $x^{(k)}$ converge to $\bar{x}$ with convergence quotients $q^{(k)} := \frac{\|x^{(k)} - \bar{x}\|}{\|x^{(k-1)} - \bar{x}\|}$ given in the following table for different values of α .

	$\alpha = 0$	
k	$x^{(k)}$	$q^{(k)}$
1	(0.5664, 0.4336)	0.96
4	(0.6522, 0.3478)	0.90
8	(0.8639, 0.1361)	0.71
12	(0.9849, 0.0151)	0.53

	$\alpha = 0.5$	
k	$x^{(k)}$	$q^{(k)}$
1	(0.5599, 0.4401)	0.98
4	(0.6022, 0.3978)	0.96
8	(0.6996, 0.3004)	0.91
12	(0.8389, 0.1611)	0.82

It is not difficult to see that depending on the choice of α the linear convergence factors $q^{(k)}$ converge to:

$$q^{(k)} \rightarrow q_\alpha = \frac{0.5 + \alpha}{1 + \alpha} < 1,$$

showing how in this example the linear convergence factor increases with increasing α .

Notice that the SC condition (25) holds at $\bar{x}$ (and $\tilde{x}$). For $i = 2$ with $\bar{x}_2 = 0$ we actually find $A^2 \cdot \bar{x}^{d-1} = \frac{1}{2} < 1 = A \cdot \bar{x}^d$.

We can also give a general formula for the dependence on α of the transformation $y^\alpha(x)$ compared with $y(x) = y^0(x)$:

$$\begin{aligned} y_i^\alpha(x) &= \frac{x_i(A^i \cdot x^{d-1} + \alpha)}{A \cdot x^d + \alpha} = \frac{A \cdot x^d}{A \cdot x^d + \alpha} \left(\frac{x_i(A \cdot x^{d-1})}{A \cdot x^d} + \frac{\alpha}{A \cdot x^d} x_i \right) \\ &= \frac{1}{1 + \frac{\alpha}{A \cdot x^d}} \left(y_i^0(x) + \frac{\alpha}{A \cdot x^d} x_i \right). \end{aligned}$$

So for small $\rho := \frac{\alpha}{A \cdot x^d}$ using $\frac{1}{1+\rho} = 1 - \rho + O(\rho^2)$ we find

$$\begin{aligned} y_i^\alpha(x) &= (1 - \rho)(y_i^0(x) + \rho x_i) + O(\rho^2) \\ &= (1 - \rho)(y_i^0(x) - x_i) + (1 - \rho)\rho x_i + (1 - \rho)x_i + O(\rho^2) \\ &= (1 - \rho)(y_i^0(x) - x_i) + x_i + O(\rho^2). \end{aligned}$$

Rewriting this relation in the form

$$y_i^\alpha(x) = x_i + \left(1 - \frac{\alpha}{A \cdot x^d}\right)(y_i^0(x) - x_i) + O\left(\frac{\alpha}{A \cdot x^d}\right)^2,$$

it appears that compared with $\alpha = 0$, for $\alpha > 0$ the new point $y^\alpha(x)$ is closer to x (i.e., the step is shorter).

This may be seen as an indication that with increasing α the speed of convergence is decreasing, as was shown explicitly in Example 1. All our numerical experiments suggest such a behavior. So as a rule of thumb, to assure $(A + \alpha E) > 0$ we should choose α appropriately small.

5 Numerical experiments

In this section, we report on many examples which illustrate numerically the theoretical convergence results in Section 4.

In all examples, we indicate the numbers n, d , the function $f(x)$ and the choices of α to assure $f^\alpha(x) > 0$ (cf., Subsection 4.4). We also give the starting point $x^{(0)}$ and specify whether the replicator iteration (20) (Method 1) is applied or Method 2 is used. For the former we describe the convergence behavior of the sequence $x^{(k)} \rightarrow \bar{x}$. If linear convergence occurs also information on the linear convergence factors $q^{(k)} := \frac{\|x^{(k+1)} - \bar{x}\|}{\|x^{(k)} - \bar{x}\|}$ is provided.

Example 2 We chose $n = 3, d = 2$ and

$$f(x) = -(x_1 - x_2)^2 - (x_1 - x_3)^2.$$

There is a unique (global) maximizer $\bar{x} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ which trivially satisfies the SC condition, cf., (12). Thus according to Theorem 2 the replicator iteration (20) should always show a linear convergence when starting with $x^{(0)} > 0$.

Replicator iteration:

- $\alpha = 1$, different starting points $x^{(0)} > 0$ led to $x^{(k)} \rightarrow \bar{x}$ and linear convergence factors $q^{(k)}$ converging to $0.66 \dots$. E.g., $x^{(0)} = (0.4, 0.5, 0.1)$ and $\|x^{(20)} - \bar{x}\| \approx 4 \cdot 10^{-4}$.
- $\alpha = 2$ different starting points $x^{(0)} > 0$ led to $x^{(k)} \rightarrow \bar{x}$ and linear convergence factors $q^{(k)}$ converging to $0.83 \dots$. E.g., $x^{(0)} = (0.4, 0.5, 0.1)$ and $\|x^{(20)} - \bar{x}\| \approx 2 \cdot 10^{-2}$.

Method 2:

- $\alpha = 1$, $x^{(0)} = (0.2, 0.3, 0.5)$ led to $x^{(k)} \rightarrow \bar{x}$ with $\|x^{(10)} - \bar{x}\| \approx 3 \cdot 10^{-5}$.
- $\alpha = 2$, $x^{(0)} = (0.2, 0.3, 0.5)$ led to $x^{(k)} \rightarrow \bar{x}$ with $\|x^{(10)} - \bar{x}\| \approx 4 \cdot 10^{-5}$.

Compared with the replicator iteration, Method 2 shows a considerable speeding up of the convergence.

Example 3 Take $n = 3$, $d = 2$ and

$$f(x) = -x_1^2 - x_2^2 - 2x_3^2 + 2x_1x_2 - 2x_1x_3 + 2x_2x_3,$$

with unique maximizer $\bar{x} = \frac{1}{2}(1, 1, 0)$, $f(\bar{x}) = 0$. In view of $\frac{1}{2} \frac{\partial}{\partial x_3} f(\bar{x}) = 0 = f(\bar{x})$ the condition SC is not satisfied. Thus according to Theorem 2 starting with some $x^{(0)} > 0$ the convergence of the replicator dynamics (20) must be sublinear. Replicator iteration:

- $\alpha = 2$, $x^{(0)} = (0.3, 0.2, 0.5)$ led to $x^{(k)} \rightarrow \bar{x}$ and factors $q^{(k)}$ which monotonically increase towards 1 with e.g., $q^{(100)} = 0.989..$, $\|x^{(100)} - \bar{x}\| \approx 3 \cdot 10^{-2}$.

Method 2:

- $\alpha = 2$, $x^{(0)} = (0.3, 0.2, 0.5)$ generated $x^{(k)} \rightarrow \bar{x}$ with $\|x^{(20)} - \bar{x}\| \approx 3 \cdot 10^{-2}$.

Example 4 Consider for $n = 3$, $d = 3$ the function

$$f(x) = 3x_1x_2x_3 - x_1^2x_2 - x_1x_2^2 - x_3^3,$$

with 3 (global) maximizers $\bar{x} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, $\hat{x} = (1, 0, 0)$ and $\tilde{x} = (0, 1, 0)$, where $f(\bar{x}) = f(\hat{x}) = f(\tilde{x}) = 0$. At $\bar{x} > 0$ the SC condition is (trivially) satisfied but not at $\hat{x}, \tilde{x}$.

Replicator iteration:

- $\alpha = 0.5$, $x^{(0)} = (0.2, 0.6, 0.2)$ led to a linear convergence $x^{(k)} \rightarrow \bar{x}$ with $q^{(k)} \rightarrow 0.925..$, and $\|x^{(100)} - \bar{x}\| \approx 3 \cdot 10^{-4}$.
- $\alpha = 1$, $x^{(0)} = (0.5, 0.2, 0.3)$ led to a linear convergence $x^{(k)} \rightarrow \bar{x}$ with $q^{(k)} \rightarrow 0.963..$, and $\|x^{(100)} - \bar{x}\| \approx 8 \cdot 10^{-2}$.
- $\alpha = 1$, $x^{(0)} = (0.4, 0.4, 0.2)$ led to a linear convergence $x^{(k)} \rightarrow \bar{x}$ with $q^{(k)} \rightarrow 0.667..$, and $\|x^{(30)} - \bar{x}\| \approx 1 \cdot 10^{-6}$.
- $\alpha = 1$, $x^{(0)} = (0.8, 0.1, 0.1)$ led to a sublinear convergence $x^{(k)} \rightarrow \hat{x}$ with $q^{(k)} \uparrow 1$, and $q^{(200)} \approx 0.997..$, $\|x^{(200)} - \hat{x}\| \approx 9 \cdot 10^{-2}$.

Note that for the same value $\alpha = 1$ different starting points led to convergence to $\bar{x}$ with different linear convergence factors 0.963..., resp., 0.667.. .

Method 2:

- $\alpha = 1$, $x^{(0)} = (0.5, 0.2, 0.3)$ led to $x^{(k)} \rightarrow \bar{x}$ with $\|x^{(10)} - \bar{x}\| \approx 4 \cdot 10^{-3}$.
- $\alpha = 2$, $x^{(0)} = (0.8, 0.1, 0.1)$ led to $x^{(k)} \rightarrow \hat{x}$ with $x^{(2)} \approx (0.91, 0, 0.09)$ on subface, $x_2^{(2)} = 0$, and $x^{(11)} = \hat{x}$ exactly.

Example 5 For $n = 3$, $d = 4$ the function

$$f(x) = -(x_1 - x_2)^2 \cdot (x_2 + x_3)^2 - (x_1 - x_3)^2 \cdot x_2^2$$

has the (global) maximizers $\bar{x} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, $\hat{x} = (1, 0, 0)$ and $\tilde{x} = (0, 0, 1)$, where at $\bar{x}$ SC (trivially) holds but not at $\hat{x}, \tilde{x}$.

Replicator iteration:

- $\alpha = 0.5$, $x^{(0)} = (0.2, 0.3, 0.5)$ led to a linear convergence $x^{(k)} \rightarrow \bar{x}$ with $q^{(k)} \rightarrow 0.948\dots$, and $\|x^{(200)} - \bar{x}\| \approx 3 \cdot 10^{-5}$.
- $\alpha = 1$, $x^{(0)} = (0.2, 0.3, 0.5)$ led to a linear convergence $x^{(k)} \rightarrow \bar{x}$ with $q^{(k)} \rightarrow 0.974\dots$, and $\|x^{(200)} - \bar{x}\| \approx 4 \cdot 10^{-3}$.
- $\alpha = 1$, $x^{(0)} = (0.8, 0.1, 0.1)$ led to a sublinear convergence $x^{(k)} \rightarrow \hat{x}$ with $q^{(k)} \uparrow 1$, and $q^{(200)} \approx 0.995\dots$, $\|x^{(200)} - \hat{x}\| \approx 1 \cdot 10^{-2}$.

Method 2:

- $\alpha = 1$, $x^{(0)} = (0.2, 0.3, 0.5)$ led to $x^{(k)} \rightarrow \bar{x}$ with $\|x^{(20)} - \bar{x}\| \approx 5 \cdot 10^{-4}$.
- $\alpha = 1$, $x^{(0)} = (0.8, 0.1, 0.1)$ led to $x^{(k)} \rightarrow \hat{x}$ with $\|x^{(20)} - \hat{x}\| \approx 8 \cdot 10^{-3}$.

Example 6 Consider for $n = 3$, $d = 6$ the function

$$f(x) = 3x_1^2x_2^2x_3^2 - x_1^4x_2^2 - x_1^2x_2^4 - x_3^6.$$

The (global) maximizers are $\bar{x} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, $\hat{x} = (1, 0, 0)$ and $\tilde{x} = (0, 1, 0)$, with $f(\bar{x}) = f(\hat{x}) = f(\tilde{x}) = 0$. Again at $\bar{x}$ the condition SC holds but not at $\hat{x}, \tilde{x}$.

Replicator iteration:

- $\alpha = 0.5$, $x^{(0)} = (0.2, 0.5, 0.3)$ led to a linear convergence $x^{(k)} \rightarrow \bar{x}$ with $q^{(k)} \rightarrow 0.994\dots$, and $\|x^{(500)} - \bar{x}\| \approx 3 \cdot 10^{-2}$.
- $\alpha = 1$, $x^{(0)} = (0.2, 0.5, 0.3)$ led to a linear convergence $x^{(k)} \rightarrow \bar{x}$ with $q^{(k)} \rightarrow 0.997\dots$, and $\|x^{(500)} - \bar{x}\| \approx 1.2 \cdot 10^{-1}$.
- $\alpha = 1$, $x^{(0)} = (0.8, 0.1, 0.1)$ led to a sublinear convergence $x^{(k)} \rightarrow \hat{x}$ with $q^{(k)} \uparrow 1$, and $q^{(500)} \approx 0.999\dots$, $\|x^{(500)} - \hat{x}\| \approx 3 \cdot 10^{-1}$.

Method 2:

- $\alpha = 0.5$, $x^{(0)} = (0.2, 0.5, 0.3)$ led to $x^{(k)} \rightarrow \bar{x}$ with $\|x^{(100)} - \bar{x}\| \approx 4 \cdot 10^{-6}$.
- $\alpha = 1$, $x^{(0)} = (0.2, 0.5, 0.3)$ led to $x^{(k)} \rightarrow \bar{x}$ with $\|x^{(100)} - \bar{x}\| \approx 2 \cdot 10^{-3}$.
- $\alpha = 1$, $x^{(0)} = (0.8, 0.1, 0.1)$ led to $x^{(k)} \rightarrow \hat{x}$ with $\|x^{(100)} - \hat{x}\| \approx 3 \cdot 10^{-1}$.

Table 1 Results for the 1000 runs in Example 6

Maximum	Method 1	Method 2
(1, 0, 0)	Convergence occurred: 1 time $\max : 701, \min : 701$ $-\pm -$	Convergence occurred: 92 times $\max : 132, \min : 3$ 11.18 ± 15.43
(0, 1, 0)	Convergence occurred: 0 times $\max : -, \min : -$ $-\pm -$	Convergence occurred: 90 times $\max : 25, \min : 3$ 7.92 ± 4.63
$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	Convergence occurred: 395 times $\max : 1000, \min : 219$ 865.25 ± 214.72	Convergence occurred: 462 times $\max : 452, \min : 18$ 181.73 ± 65.32

Fig. 1 Method 1 applied to Example 6 (NC = did not meet convergence criterion)

For this example problem we have added some more experiments with $\alpha = 1$. For both methods, 1000 random initial points $x^{(0)}$ have been generated. The stopping criterion for convergence is $\min\{\|x^{(k)} - \bar{x}\|, \|x^{(k)} - \hat{x}\|, \|x^{(k)} - \tilde{x}\|\} < 10^{-3}$ and we allowed a maximum number of 1000 iterations in each run.

In both methods, for $x^{(0)}$ near $\bar{x}$ the replicator iteration converges to $\bar{x}$. For Method 1 only one of the 1000 initial points converged to $\hat{x}$ within the 1000 iterations and none converged to $\tilde{x}$ (cf., Figure 1, Table 1). The convergence to these points was too slow (sublinear). Method 2 performed better (see Figure 2 and Table 1). All maximizers $\bar{x}, \hat{x}, \tilde{x}$ have been detected.

Fig. 2 Method 2 applied to Example 6 (NC = did not meet convergence criterion)

Table 2 Results for the experiments in Example 7

Maximizer	Method 1	Method 2
$\left(\frac{1}{4}, \frac{3}{4}\right)$	$\max : 57, \min : 31$ 32.14 ± 7.05	$\max : 36, \min : 3$ 10.76 ± 6.25
$\left(\frac{3}{4}, \frac{1}{4}\right)$	$\max : 56, \min : 11$ 31.92 ± 6.59	$\max : 34, \min : 3$ 10.36 ± 6.05

Table 1 presents the results of the experiments for both methods. It contains the number of cases out of the 1000 runs where convergence occurred, as well as the corresponding maximum, minimum and mean numbers (with $\pm$ standard deviation) of iterations for the random initial point $x^{(0)}$.

Example 7 (cf. [8, Example 9.8]) In this example we consider the ESS problem with $n = 2, d = 4$. The tensor $A \in \mathcal{S}^{2,4}$ of the function $f(x) = A \cdot x^d$ is given by:

$$a_{1111} = -\frac{3}{32}; a_{1112} = 0; a_{1122} = -\frac{13}{96}; a_{1222} = 0; a_{2222} = -\frac{3}{32};$$

The function f has two (global) maximizers $\bar{x} = (\frac{1}{4}, \frac{3}{4})$ and $\hat{x} = (\frac{3}{4}, \frac{1}{4})$. We choose $\alpha = \frac{13}{96} + \frac{1}{100}$ and generate 1000 random instances for the initial vector $x^{(0)}$. In this example for all initial vectors the iterations met the convergence criterion

Table 3 Result of experiment in Example 8

Maximizer	Convergence occurred	Iterations	
		Method 1	Method 2
$(0, 0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	74 times	$\max : 190, \min : 27$ $\text{mean} \pm SD : 110.13 \pm 21.20$	$\max : 80, \min : 13$ $\text{mean} \pm SD : 19.87 \pm 2.63$
$(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0)$	447 times	$\max : 353, \min : 33$ $\text{mean} \pm SD : 171.02 \pm 29.95$	$\max : 118, \min : 12$ $\text{mean} \pm SD : 20 \pm 2.86$
$(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, \frac{1}{4})$	479 times	$\max : 290, \min : 46$ $\text{mean} \pm SD : 167.55 \pm 25.71$	$\max : 120, \min : 12$ $\text{mean} \pm SD : 19.87 \pm 2.91$

$\min\{\|x^{(k)} - \bar{x}\|, \|x^{(k)} - \hat{x}\|\} < 10^{-3}$ and around 50% of the initial points converged to $\bar{x}$. The others converged to $\hat{x}$. The details of the experiment are presented in Table 2, where again for both methods the maximum, minimum and mean numbers (with $\pm$ standard deviation) of iterations are given.

Example 8 (cf. [10]) Consider the 3–hypergraph with $V = \{1, 2, 3, 4, 5\}$ and

$$E = \{(1, 2, 3), (1, 2, 4), (1, 3, 4), (2, 3, 4), (1, 2, 5), (1, 3, 5), (2, 3, 5), (3, 4, 5)\}.$$

So we have (see Section 3.2), $n = 5, d = 3$

$$h_{\bar{G}(x)}^-(x) = -(x_1x_4x_5 + x_2x_4x_5 + \tau(x_1^3 + x_2^3 + x_3^3 + x_4^3 + x_5^3))$$

The 3–hypergraph has three maximum cliques, $\{1, 2, 3, 4\}, \{1, 2, 3, 5\}, \{3, 4, 5\}$ with corresponding global maximizers $\bar{x} = (0, 0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}), \tilde{x} = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0), \hat{x} = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, \frac{1}{4})$ of $h_G^-(x)$. The corresponding tensor $A \in S^{5,3}$ is given by,

$$\begin{aligned} a_{1,1,1} &= a_{2,2,2} = a_{3,3,3} = a_{4,4,4} = a_{5,5,5} = -\tau \\ a_{1,4,5} &= a_{2,4,5} = -\frac{1}{6} \\ a_{1,2,3} &= a_{1,2,4} = a_{1,3,4} = a_{2,3,4} = a_{1,2,5} = a_{1,3,5} = a_{2,3,5} = a_{3,4,5} = 0; \end{aligned}$$

We have chosen $\tau = \frac{1}{12}$ (cf. Theorem 1) and have applied a transformation $A \rightarrow A + \alpha E$ with $\alpha = \frac{1}{12} + \frac{1}{100}$. The stopping criterion for convergence is $\min\{\|x^{(k)} - \bar{x}\|, \|x^{(k)} - \hat{x}\|, \|x^{(k)} - \tilde{x}\|\} < 10^{-3}$ and we allowed a maximum number of 1000 iterations in each run. For all 1000 random starting vectors $x^{(0)}$ the iterations converged to one of the maximizers. The details are to be found in Table 3.

Table 4 Results for experiment in Example 4

Instance	Max Clique size	Best known	n	$ E $	Clique sizes found
C125.9	34	34	125	6963	28:10, 29:40, 30:85, 31:186, 32:259, 33:361, 34:59
C250.9	44	44	250	27984	31:2, 32:1, 33:14, 34:77, 35:171, 36:237, 37:229, 38:157, 39:81, 40:19, 41:10, 42:1, 43:1
C500.9	≥ 57	57	500	112332	38:1, 39:1, 40:4, 41:30, 42:71, 43:109, 44:212, 45:225, 46:159, 47:117, 48:46, 49:17, 50:5, 51:3
C1000.9	≥ 68	68	1000	450079	45:1, 46:5, 47:10, 48:39, 49:94, 50:186, 51:197, 52:158, 53:127, 54:101, 55:44, 56:26, 57:10, 58:2
C2000.9	≥ 80	80	2000	1799532	51:2, 52:6, 53:16, 54:66, 55:94, 56:142, 57:191, 58:171, 59:145, 60:88, 61:48, 62:21, 63:6, 64:3, 66:1

Example 9 We apply Method 2 to detect the maximal cliques for some of the graphs in the well-known Second DIMACS Implementation Challenge. Note that in this case we have $d = 2$ and line maximization step is simplified (see Remark 4).¹ For each graph we apply the method for 1000 initial points $x^{(0)}$. The results are shown in Table 4. Column 4 and 5 give the number n of nodes and the number $|E|$ of edges. In the last column the notation $c : \#$ means that a maximal clique of size c has been detected $\#$ times out of the 1000 runs with stopping criterion $\|x^{k+1} - x^k\| \leq 10^{-6}$.

6 Conclusion

In this paper we are interested in the problem to maximize homogeneous polynomials of degree d over the simplex. Two methods for the computation of local (global) maximizers are investigated. Both are based on the replicator dynamics.

New convergence results for the replicator iteration are presented for the case $d > 2$. Linear convergence to a maximizer $\bar{x}$ can only occur if strict complementarity holds at $\bar{x}$. Many numerical experiments illustrate the theoretical convergence results.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Ahmed, F., Still, G.: Maximization of homogeneous polynomials over simplex and sphere: structure, stability, and generic behavior. *J. Optim. Theory Appl.* **181**, 972–996 (2019)
2. Baum, L., Eagon, J.: An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. *Bull. Am. Math. Soc.* **73**, 360–363 (1967)
3. Bomze, I.M.: Non-cooperative two-person games in biology: a classification. *Int. J. Game Theory* **15**(1), 31–57 (1986)
4. Bomze, I.M.: Evolution towards the maximum clique. *J. Glob. Optim.* **10**(2), 143–164 (1997)
5. Bomze, I.M.: On standard quadratic optimization problems. *J. Glob. Optim.* **13**(4), 369–387 (1998)
6. Bomze, I.M., De Klerk, E.: Solving standard quadratic optimization problems via linear, semidefinite and copositive programming. *J. Glob. Optim.* **24**(2), 163–185 (2002)
7. Broom, M., Cannings, C., Vickers, G.: Multi-player matrix games. *Bull. Math. Biol.* **59**(5), 931–952 (1997)
8. Broom, M., Rychtár, J.: *Game-Theoretical Models in Biology*. CRC Press, Boca Raton (2013)
9. Bulò, S.R., Pelillo, M.: A continuous characterization of maximal cliques in k -uniform hypergraphs. In: *International conference on learning and intelligent optimization*, pp. 220–233. Springer (2007)
10. Bulò, S.R., Pelillo, M.: A generalization of the Motzkin–Straus theorem to hypergraphs. *Optim. Lett.* **3**(2), 287–295 (2009)

¹ <http://dimacs.rutgers.edu/archive/Challenges/>.

11. Clark, P.: Sequences and series; a source book. <http://www.math.uga.edu/~pete/3100supp.pdf>
12. De Klerk, E., Laurent, M., Sun, Z., Vera, J.C.: On the convergence rate of grid search for polynomial optimization over the simplex. *Optim. Lett.* **11**(3), 597–608 (2017)
13. Faybusovich, L.: Global optimization of homogeneous polynomials on the simplex and on the sphere. In: *Frontiers in Global Optimization*, pp. 109–121. Springer (2004)
14. Kingman, J.F.: On an inequality in partial averages. *Q. J. Math.* **12**, 78–80 (1961)
15. Kleniati, P.-M., Parpas, P., Rustem, B.: Partitioning procedure for polynomial optimization. *J. Glob. Optim.* **48**, 549–567 (2010)
16. Losert, V., Akin, E.: Dynamics of games and genes: discrete versus continuous time. *J. Math. Biol.* **17**, 241–251 (1983)
17. Lyubich, Y., Maistrovskii, G., Ol'khovskii, Y.: Selection-induced convergence to equilibrium in a single-locus autosomal population. *Probl. Peredachi Inf.* **16**(1), 93–104 (1980). ((Russian))
18. Motzkin, T., Strauss, E.: Maxima for graphs and a new proof of a theorem of Turán. *Can. J. Math.* **17**, 533–540 (1965)
19. Murty, K., Kabadi, S.: Some NP-complete problems in quadratic and nonlinear programming. *Math. Program.* **39**(2), 117–129 (1987)
20. Palm, G.: Evolutionary stable strategies and game dynamics for n -person games. *J. Math. Biol.* **19**(3), 329–334 (1984)
21. Smith, J.M.: The theory of games and the evolution of animal conflicts. *J. Theor. Biol.* **47**(1), 209–221 (1974)
22. Smith, J.M., Price, G.R.: The logic of animal conflict. *Nature* **246**(5427), 15–18 (1973)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.