A Note on the KKT Points for the Motzkin-Straus Program

Guglielmo Beretta[(1),(2),†], Alessandro Torcinovich[(3),†] and Marcello Pelillo[(1)]
[(1)]Ca' Foscari University of Venice, Italy
[(2)]Polytechnic University of Turin, Italy
[(3)]ETH, Switzerland
[†]Equal contribution, authors listed in alphabetical order
guglielmo.beretta@unive.it

preliminary version*
April 28, 2023

Abstract

In a seminal 1965 paper, Motzkin and Straus established an elegant connection between the clique number of a graph and the global maxima of a quadratic program defined on the standard simplex. Since then, the result has been the subject of intensive research and has served as the motivation for a number of heuristics and bounds for the maximum clique problem. Most of the studies available in the literature, however, focus typically on the local/global solutions of the program, and little or no attention has been devoted so far to the study of its Karush-Kuhn-Tucker (KKT) points. In contrast, in this paper we study the properties of (a parameterized version of) the Motzkin-Straus program and show that its KKT points can provide interesting structural information and are in fact associated with certain regular sub-structures of the underlying graph.

Keywords: Standard quadratic optimization, KKT points, clique, regular graphs, replicator dynamics

1 Introduction

In 1965, Motzkin and Straus [14] studied the program:

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{maximize}} \quad f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A} \mathbf{x} \tag{1a}$$

$$\text{subject to} \quad \mathbf{1}^\top \mathbf{x} = 1, \tag{1b}$$

$$\mathbf{x} \geq \mathbf{0}, \tag{1c}$$

where $\mathbf{A}$ denotes the adjacency matrix of an undirected unweighted graph G on a set of n vertices. They proved that the value of (1) equals $\omega(G)^{-1} + 1$, where $\omega(G)$ is the clique number

*Note: this article has not been peer reviewed yet. Guglielmo Beretta's scholarship is funded jointly by Ca' Foscari University of Venice and by Polytechnic University of Turin. The authors have no competing interests to declare that are relevant to the content of this article. All authors contributed to the study conception and design. The first draft of the manuscript was written jointly by Alessandro Torcinovich and Guglielmo Beretta. This study has been performed under the supervision of Marcello Pelillo. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript. Figure 1 and Figure 2 have been produced using TikZ.

of G, and that every maximum clique C corresponds to a global solution for (1). More recently, Pelillo and Jagota [17] studied the "spurious" solutions of the Motzkin-Straus program (namely, solutions that are not associated to any maximum clique) and provided a characterization of its (strict) *local* solutions in terms of (strictly) *maximal* cliques of G. Bomze [1] modified (1) by adding a convex regularization term to its objective function, thereby obtaining a spurious-free version of the program where local (global) solutions are in one-to-one correspondence with maximal (maximum) cliques (and all solutions are strict).[1]

Since its introduction, the Motzkin-Straus program, and its variations, has been the subject of intensive research and has been generalized in various ways [8, 4, 20, 21], motivating a number of heuristics and bounds for the maximum clique problem (see, e.g., [2, 26, 5, 22]). Most of the studies available in the literature, however, focus typically on the properties of the local/global maximizers of (1) and little or no interest has been devoted to its Karush-Kuhn-Tucker (KKT) points. In contrast, in this paper we study the KKT points of (a parametric version of) the program introduced by Bomze [1, 3], in an attempt to obtain structural information on the underlying graph. In particular, we extend some known results about characteristic vectors concerning regular induced subgraphs and discuss how a KKT point is related to the symmetries of the subgraph induced by its support. Using barycentric coordinates [19, 10], we then exploit a suitable representation of KKT points to further analyze the combinatorial structure of its support. To do this, we introduce the novel concept of a partition induced by an element in the standard simplex and that of highly regular families. Finally, the results obtained are applied to the class of generalized star graphs.

2 Notation

Here, we fix the notation that we will adopt throughout this paper. For a positive integer $n \in \mathbb{N}$ we write $[n]$ for the set $\{i \in \mathbb{N} \colon 1 \leq i \leq n\}$. Lowercase bold fonts are reserved to column vectors, whereas uppercase bold fonts will denote matrices and a superscript $\top$ denotes transposition. We denote by $\mathbf{0}$ (resp. $\mathbf{1}$) a vector with every component equal to 0 (resp. 1), with dimension in agreement with the context in which it is used.

The *support* of a vector $\mathbf{x} \in \mathbb{R}^n$ is the set $\operatorname{supp}(\mathbf{x}) = \{i \in [n] \colon x_i \neq 0\}$. The *standard simplex* in $\mathbb{R}^n$ is the set:

$$\Delta_n = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{1}^\top \mathbf{x} = 1, \quad \mathbf{x} \geq \mathbf{0}\},$$

For a non-empty $S \subseteq [n]$, we define:

$$\begin{aligned} \Delta_n(S) &= \{\mathbf{x} \in \Delta_n \colon \operatorname{supp}(\mathbf{x}) \subseteq S\}, \\ \operatorname{int}(\Delta_n(S)) &= \{\mathbf{x} \in \Delta_n \colon \operatorname{supp}(\mathbf{x}) = S\}, \\ \partial(\Delta_n(S)) &= \Delta_n(S) \setminus \operatorname{int} \Delta_n(S), \end{aligned}$$

being $\Delta_n(S)$ the face of Δ_n associated with S, whereas $\operatorname{int}(\Delta_n(S))$ and $\partial(\Delta_n(S))$ are the *(relative) interior* and the *(relative) boundary* of $\Delta_n(S)$ respectively. In this notation $\Delta_n([n])$ is an alias for Δ_n.

As for the graph-related notation, $G = (V, E)$ denotes in the sequel an unweighted undirected graph on a set V and with set of edges $E \subseteq \binom{V}{2}$. We say that two vertices i, $j \in V$ are *adjacent*, and write $i \sim j$, whenever $\{i, j\} \in E$. The *adjacency matrix* of G is the symmetric $n \times n$ matrix with coefficent ij equal to 1 whenever $i \sim j$ and equal to 0 otherwise. Given a non-empty $S \subseteq V$, we use $G[S]$ for the subgraph of G *induced by* S, that is the graph on the set S in which two vertices i, $j \in S$ are adjacent if and only if $\{i, j\} \in E$. A non-empty subset $C \subseteq V$ is called

[1]See Hungerford and Rinaldi [12] for an alternative family of regularizations of the program.

a *clique* if the induced subgraph $G[C]$ is complete, *i.e.*, $i \sim j$ for every distinct i, $j \in C$. The *degree* of a vertex in a graph is the amount of neighbors that vertex has among the vertices in the graph and a graph is said *regular* if every vertex in the graph has the same degree. We also recall that an *automorphism* of a graph is an isomorphism with itself, *i.e.*, a permutation σ of its vertices such that two vertices i and j are neighbors if and only if $\sigma(i)$ is adjacent to $\sigma(j)$.

3 Parametric Motzkin-Straus programs

Consider a graph $G = (V, E)$ on a finite non-empty set V, with $|V| = n$. Without loss of generality, assume $V = [n]$ to simplify the notation. Denote by $\mathbf{A}$ the adjacency matrix of G and by $\mathbf{I}$ the $n \times n$ identity matrix. Fix now a parameter $c \in \mathbb{R}$ and consider the quadratic program:

$$\begin{aligned} \underset{\mathbf{x} \in \mathbb{R}^n}{\text{maximize}} \quad & f_c(\mathbf{x}) = \mathbf{x}^\top(\mathbf{A} + c\mathbf{I})\mathbf{x} && \text{(2a)} \\ \text{subject to} \quad & \mathbf{1}^\top \mathbf{x} = 1, && \text{(2b)} \\ & \mathbf{x} \geq \mathbf{0}, && \text{(2c)} \end{aligned}$$

and the associated *Lagrangian* [13]:

$$\mathcal{L}(\mathbf{x}, \mu_0, \boldsymbol{\mu}) = f_c(\mathbf{x}) + \mu_0(\mathbf{1}^\top \mathbf{x} - 1) + \boldsymbol{\mu}^\top \mathbf{x},$$

which is defined for $\mathbf{x} \in \mathbb{R}^n$ and for the multipliers $\mu_0 \in \mathbb{R}$ and $\boldsymbol{\mu} \in \mathbb{R}^n$. Program (2) is discussed in Bomze [1] for $c = \frac{1}{2}$ and in Bomze *et al.* [3] in its more general formulation. Observe that (1) is precisely (2) in case $c = 0$.

Definition 1. *A point* $\mathbf{x} \in \mathbb{R}^n$ *is a* Karush-Kuhn-Tucker (KKT) point *for* (2) *if some* $(\mu_0, \boldsymbol{\mu}) \in \mathbb{R} \times \mathbb{R}^n$ *exists such that:*

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial \mathbf{x}}(\mathbf{x}, \mu_0, \boldsymbol{\mu}) = \mathbf{0}, \\ \mathbf{1}^\top \mathbf{x} = 1, \\ \mathbf{x} \geq \mathbf{0}, \\ \mu_i x_i = 0 \textit{ for } i \in V, \\ \boldsymbol{\mu} \geq \mathbf{0}. \end{cases} \tag{3}$$

The set $\mathcal{KKT}(c)$ *denotes the set of KKT points for* (2).

Dropping in (3) the sign condition for the multipliers leads to the following generalization of a KKT point:[2]

Definition 2. *A point* $\mathbf{x} \in \mathbb{R}^n$ *is a* generalized KKT point *for* (2) *if some* $(\mu_0, \boldsymbol{\mu}) \in \mathbb{R} \times \mathbb{R}^n$ *exists such that:*

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial \mathbf{x}}(\mathbf{x}, \mu_0, \boldsymbol{\mu}) = \mathbf{0}, \\ \mathbf{1}^\top \mathbf{x} = 1, \\ \mathbf{x} \geq \mathbf{0}, \\ \mu_i x_i = 0 \textit{ for } i \in V, \end{cases} \tag{4}$$

The set g$\mathcal{KKT}(c)$ *is the set of generalized KKT points for* (2).

[2] Definition 2 differs from how Bomze [1] and Bomze *et al.* [3] define generalized KKT points, where the only difference is that they require the condition *p*) $\boldsymbol{\mu}^\top \mathbf{x} = 0$ in place of our *q*) $\mu_i x_i = 0$ for all $i \in V$. If $\boldsymbol{\mu} \geq \mathbf{0}$, then p and q are equivalent, but observe that without requiring $\boldsymbol{\mu} \geq \mathbf{0}$, then p is a weaker constraint then q.

Notice that removing the condition $\boldsymbol{\mu} \geq \mathbf{0}$ amounts to converting active inequality constraints into equality constraints. For this reason, a point $\hat{\mathbf{x}} \in \Delta_n$ with support $S = \operatorname{supp}(\hat{\mathbf{x}})$ satisfies Definition 2 if and only if $\hat{\mathbf{x}}$ is a KKT point for the program:

$$\begin{aligned}
\underset{\mathbf{x} \in \mathbb{R}^n}{\text{maximize}} \quad & f_c(\mathbf{x}) = \mathbf{x}^\top (\mathbf{A} + c\mathbf{I})\mathbf{x} && \text{(5a)} \\
\text{subject to} \quad & \mathbf{1}^\top \mathbf{x} = 1, && \text{(5b)} \\
& \mathbf{x} \geq \mathbf{0}, && \text{(5c)} \\
& x_i \neq 0 \quad \text{for } i \in S, && \text{(5d)} \\
& x_i = 0 \quad \text{for } i \notin S && \text{(5e)}
\end{aligned}$$

i.e., for the program:

$$\underset{\mathbf{x} \in \operatorname{int}(\Delta_n(S))}{\text{maximize}} \quad f_c(\mathbf{x}). \tag{6}$$

The inclusions

$$\mathrm{g}\mathcal{KKT}(c) \cap \operatorname{int}(\Delta_n) \subseteq \mathcal{KKT}(c) \subseteq \mathrm{g}\mathcal{KKT}(c)$$

follow directly from the definitions given. Proposition 1 presents a well known alternative description of $\mathcal{KKT}(c)$ and $\mathrm{g}\mathcal{KKT}(c)$ [8, 3].

Proposition 1. *Let $\mathbf{x} \in \Delta_n$ and set $\lambda = f_c(\mathbf{x})$. For $\mathbf{M} = \mathbf{A} + c\mathbf{I}$, consider the statements:*

1. $(\mathbf{M}\mathbf{x})_i = (\mathbf{M}\mathbf{x})_j$ *for every* $i, j \in \operatorname{supp}(\mathbf{x})$*;*
2. $(\mathbf{M}\mathbf{x})_i = \lambda$ *for every* $i \in \operatorname{supp}(\mathbf{x})$*;*
3. $(\mathbf{M}\mathbf{x})_i \leq \lambda$ *for every* $i \in V \setminus \operatorname{supp}(\mathbf{x})$.

Then:

- *the statements 1. and 2. are equivalent;*
- $\mathbf{x} \in \mathcal{KKT}(c)$ *if and only if 2. and 3. hold;*
- $\mathbf{x} \in \mathrm{g}\mathcal{KKT}(c)$ *if and only if 2. holds.*

Proof. First observe that $f_c(\mathbf{x}) = \sum_{i \in \operatorname{supp}(\mathbf{x})} x_i (\mathbf{M}\mathbf{x})_i$, which entails the equivalence of 1. and 2.

Since:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}}(\mathbf{x}, \mu_0, \boldsymbol{\mu}) = 2\mathbf{M}\mathbf{x} + \mu_0 \mathbf{1} + \boldsymbol{\mu},$$

the implications claimed are a simple restatement of the definitions given under the change of variables $\mu_0 = -2\lambda$ and $\boldsymbol{\mu} = 2(\lambda \mathbf{1} - \mathbf{M}\mathbf{x})$. □

Motzkin and Straus [14], Bomze [1] and Pardalos and Phillips [15] considered, other than some instances of (2), the program obtained by replacing the matrix $\mathbf{A}$ appearing in (2) with the adjacency matrix of the complement graph of G.[3] Looking at the KKT points of these quadratic programs leads to Proposition 2. Let $\overline{G}$ denote the complement graph of G and call $\overline{\mathbf{A}}$ its adjacency matrix.

[3]This is also motivated by the fact that the maximum clique problem for G is the dual problem of finding a maximum independent set for $\overline{G}$ [6].

Proposition 2. *1. $\mathcal{KKT}_{\overline{G}}(c)$ coincides*[4] *with the set of KKT points for the* minimization *program:*

$$\begin{aligned} \underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \quad & \mathbf{x}^\top(\mathbf{A} + (1-c)\mathbf{I})\mathbf{x} && (7a) \\ \text{subject to} \quad & \mathbf{1}^\top \mathbf{x} = 1, && (7b) \\ & \mathbf{x} \geq \mathbf{0}; && (7c) \end{aligned}$$

2.

$$\mathrm{g}\mathcal{KKT}_{\overline{G}}(c) = \mathrm{g}\mathcal{KKT}_G(1-c).$$

Proof. Consider a matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$, the affine transformation $\phi(t) = 1-t$ and a vector $\mathbf{x} \in \Delta_n$. The matrix $\mathbf{N} \in \mathbb{R}^{n \times n}$ with general coefficient $n_{ij} = \phi(m_{ij})$ satisfies also

$$(\mathbf{N}\mathbf{x})_i = \phi((\mathbf{M}\mathbf{x})_i) \quad \text{for every } i \in V \tag{8}$$

since the coordinates of $\mathbf{x}$ sum up to 1.[5] The proof is then a consequence of (8) applied for $\mathbf{M} = \overline{\mathbf{A}} + c\mathbf{I}$ and $\mathbf{M} = \mathbf{A} + (1-c)\mathbf{I}$, using Proposition 1 and the identity $\mathbf{A} + \overline{\mathbf{A}} + \mathbf{I} = \mathbf{1}\mathbf{1}^\top$. □

Proposition 2 has been ispired by Motzkin and Straus's work [14], in which a similar idea shows essentially that the Motzkin-Straus program has the same value as that of a suitable minimization program. Notice that Proposition 2 entails that $\mathrm{g}\mathcal{KKT}_{\overline{G}}(\frac{1}{2}) = \mathrm{g}\mathcal{KKT}_G(\frac{1}{2})$, a curious equality involving the quadratic program studied by Bomze in [1].

3.1 KKT points and replicator dynamics

Let $\mathbf{M} \in \mathbb{R}^{n \times n}$ and consider on $\mathbb{R}^n$ the ordinary differential equation:

$$\dot{x}_i = x_i[(\mathbf{M}\mathbf{x})_i - \mathbf{x}^\top \mathbf{M}\mathbf{x}], \quad i = 1, 2, \ldots, n. \tag{9}$$

It is easy to show [9] that the set Δ_n is invariant under the flux defined by Equation 9. The *replicator dynamics*[6] with payoff-matrix $\mathbf{M}$ is the dynamics defined on Δ_n by 9.

Replicator dynamics denotes a class of continuos-time and discrete-time dynamical systems introduced in Taylor and Jonker [24] to describe the coexistence of interacting self-replicating species [9, 25], and that resulted useful also in economics and social sciences, where behavioural patterns or strategies are studied in place of species, and the concept of replication corresponds to imitation of successful behavior [25].

The content of this paper has an alternative interpretation in the replicator dynamics framework. In facts, imposing $\dot{\mathbf{x}} = 0$ in Equation 9 allows to characterize stationary points under replicator dynamics [9], and it is easy to see that a point $\mathbf{x} \in \Delta_n$ is stationary for the replicator dynamics with payoff-matrix $\mathbf{M}$ if and only if there exists some $\lambda \in \mathbb{R}$ such that $(\mathbf{M}\mathbf{x})_i = \lambda$ for every $i \in \mathrm{supp}(\mathbf{x})$. The reader may notice the resemblance of this with Proposition 1. Indeed, $\mathbf{x} \in \Delta_n$ is stationary for the replicator dynamics with payoff-matrix $\mathbf{A} + c\mathbf{I}$ if and only if $\mathbf{x} \in \mathrm{g}\mathcal{KKT}(c)$ [1].

Moreover, the replicator dynamics with payoff-matrix $\mathbf{A} + c\mathbf{I}$ admits f_c as a Lyapunov function [1],[7] thus motivating the numerical simulation of the dynamics as a means to look

[4] We write here $\mathcal{KKT}_G(c)$ (resp. $\mathrm{g}\mathcal{KKT}_G(c)$) for $\mathcal{KKT}(c)$ (resp. $\mathrm{g}\mathcal{KKT}(c)$) so as to make explicit the dependence on G, which affects the objective function of Program (2).

[5] This is true if ϕ is replaced by any affine transformation of $\mathbb{R}$ into itself.

[6] Indeed, this is not the only possible replicator dynamics having $\mathbf{M}$ as payoff- matrix [9, 18].

[7] The replicator dynamics with a *symmetric* payoff-matrix $\mathbf{M}$ admits $\mathbf{x}^\top \mathbf{M}\mathbf{x}$ as a Lyapunov function, which can be thought as a measure of fitness when it comes to modeling biological systems. The adoption of this framework is hence supported by an intriguing connection with Fisher's Theorem of Natural Selection [1, 7].

for maximizers for f_c in the standard simplex, and even though the dynamics is not bound to converge to a local solution of (2), it can be shown that each trajectory initialized in $\mathrm{int}(\Delta_n)$ converges to an element of $\mathcal{KKT}(c)$ [1, Lemma 4].

4 Characteristic vectors

Let S be a non-empty subset of V. The *characteristic vector* representing S in Δ_n is the vector $\mathbf{x}^S \in \Delta_n$ defined by:

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

In [14] characteristic vectors representing maximum cliques emerge as global solutions to (1), and characteristic vectors representing maximal cliques are among the more interesting local solutions of (1) [14, 17, 23]. For a KKT point for (1) that is a characteristic vector, the subgraph of G induced by its support is not necessarily a complete graph. However, it must be a regular graph. Bomze [1] observed that $\mathbf{x}^S \in \mathrm{g}\mathcal{KKT}(\frac{1}{2})$ if and only if $G[S]$ is a regular graph,[8] and the same proof indeed works also for $c \neq \frac{1}{2}$, as Proposition 3 shows.

Proposition 3. *Let $\mathbf{x}$ be a characteristic vector. Then $\mathbf{x} \in \mathrm{g}\mathcal{KKT}(c)$ if and only if $G[\mathrm{supp}(\mathbf{x})]$ is regular.*

Proof. Let $\mathbf{x}$ be a characteristic vector and set $S = \mathrm{supp}(S)$, so that $\mathbf{x} = \mathbf{x}^S$. For each $i \in S$, let the integer d_i count how many vertices in S are adjacent to i. Then $(\mathbf{A}\mathbf{x}^S)_i = d_i/|S|$, thus $((\mathbf{A} + c\mathbf{I})\mathbf{x}^S)_i = (d_i + c)/|S|$. By Proposition 1, the vector $\mathbf{x}^S$ is in $\mathrm{g}\mathcal{KKT}(c)$ if and only if for some $\lambda \in \mathbb{R}$ the equality $(d_i + c)/|S| = \lambda$ holds for every $i \in S$. This is possible if and only if d_i has the same value for every $i \in S$, that is, if and only if $G[S]$ is regular. □

Thanks to Proposition 3, a characteristic vector is in $\mathrm{g}\mathcal{KKT}(c)$ either for every value of c or for no value of c. In contrast, vectors in the standard simplex that are not characteristic vectors, that is all but $2^n - 1$ elements of Δ_n, exhibit a different behavior. Indeed, as a consequence of the next proposition, which generalizes [3, Proposition 6], each of those vectors is an element of $\mathrm{g}\mathcal{KKT}(c)$ for *one* value of c at most.[9]

Proposition 4. *Let $\mathbf{x} \in \Delta_n$ and suppose two distinct c_1, $c_2 \in \mathbb{R}$ exist such that*[10] *$\mathbf{x} \in \mathcal{KKT}(c_j)$ for $j = 1$, 2. Then $\mathbf{x}$ is a characteristic vector, and $G[\mathrm{supp}(\mathbf{x})]$ is regular.*

Proof. Set $S = \mathrm{supp}(\mathbf{x})$. By Proposition 1, there exist λ_1, $\lambda_2 \in \mathbb{R}$ such that for $j = 1$, 2, the equation $((\mathbf{A} + c_j\mathbf{I})\mathbf{x})_i = \lambda_j$ holds for every $i \in S$, and so every non-zero component of $\mathbf{x}$ equals $(\lambda_1 - \lambda_2)/(c_1 - c_2)$. Then necessarily $\mathbf{x} = \mathbf{x}^S$, and $G[S]$ is regular by Proposition 3. □

5 Automorphisms of induced subgraphs

The elements of $\mathcal{KKT}(c)$ need not be characteristic vectors. For instance, suppose G is the graph on the set of vertices $\{1, 2, 3\}$ and edges $\{\{1, 3\}, \{2, 3\}\}$, sometimes called the *cherry graph*. It is easy to check that the point $\tilde{\mathbf{x}} = (1/4, 1/4, 1/2)$, which clearly is not a characteristic vector, is an element of $\mathcal{KKT}(0)$, as discussed in Pardalos and Phillips [15].

[8] Bomze's proofs contained in [1] work also for $0 < c < 1$, as mentioned in [3].

[9] Let $\mathbf{x} \in \Delta_n$ and assume $\mathbf{x}$ is not a characteristic vector. An analysis on the spectrum of $\mathbf{A} + c\mathbf{I}$ shows that if c lies in a certain range depending on $\mathrm{supp}(\mathbf{x})$ and of $\mathbf{A}$ then $\mathbf{x}$ can't be a stationary point for a replicator dynamics with payoff-matrix $\mathbf{A} + c\mathbf{I}$ [3, 18]. As a consequence of Proposition 3, knowing that $\mathbf{x}$ is a stationary point for the replicator dynamics with payoff-matrix $\mathbf{A} + c\mathbf{I}$, then $\mathbf{x}$ ceases to be stationary in case we perturb c, and this is true for *any* perturbation of c.

[10] The proof requires only $\mathbf{x} \in \mathrm{g}\mathcal{KKT}(c_j)$ for $j = 1$, 2.

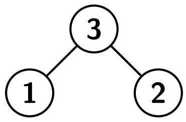

Figure 1: The cherry graph

What kind of information on G can be possibly obtained from $\tilde{\mathbf{x}}$? Observe that both the vector $\tilde{\mathbf{x}}$ and the cherry graph are preserved if vertex 1 and vertex 2 are exchanged. To be rigorous, call σ the permutation on $\{1,2,3\}$ swapping 1 and 2. Then σ is an automorphism for the cherry graph and at the same time $\tilde{\mathbf{x}}$ is invariant under the pull-back by σ, *i.e.*, the vector $\tilde{\mathbf{x}}$ is preserved if its i-th coordinate is replaced with its $\sigma(i)$-th coordinate for every $i \in V$.

Theorem 1 shows that this is an instance of a more general fact.

Theorem 1. *Let* $\mathbf{x} \in \mathrm{g}\mathcal{KKT}(c)$*, set* $S = \mathrm{supp}(\mathbf{x})$ *and let* $\mathcal{G}$ *be a group of automorphisms for the induced subgraph* $G[S]$*. Then there exists a point* $\hat{\mathbf{x}} \in \mathrm{g}\mathcal{KKT}(c)$ *satisfying* $\mathrm{supp}(\hat{\mathbf{x}}) = S$ *and* $\hat{x}_{\sigma(i)} = \hat{x}_i$ *for every* $i \in S$ *and every* $\sigma \in \mathcal{G}$*.*

Proof. For every $\sigma \in \mathcal{G}$, denote by $\sigma^*\mathbf{x}$ the vector in Δ_n satisfying:[11]

$$(\sigma^*\mathbf{x})_i = \begin{cases} x_{\sigma(i)} & \text{for } i \in S \\ 0 & \text{otherwise.} \end{cases}$$

Set $\hat{\mathbf{x}} = \frac{1}{|\mathcal{G}|}\sum_{\sigma\in\mathcal{G}} \sigma^*\mathbf{x}$ and all is left is to check that $\hat{\mathbf{x}}$ satisfies the desired properties.
Observe first that $\hat{\mathbf{x}} \in \Delta_n$ by convexity of Δ_n, and that for every $\sigma \in \mathcal{G}$ we have $\mathrm{supp}(\sigma^*\mathbf{x}) = S$, thus $\mathrm{supp}(\hat{\mathbf{x}}) = S$ by construction.
To prove that $\hat{\mathbf{x}}$ is a generalized KKT point, recall that by hypothesis on $\mathbf{x}$ some λ exists such that $((\mathbf{A} + c\mathbf{I})\mathbf{x})_i = \lambda$ for every $i \in S$. For every $\sigma \in \mathcal{G}$ and every i, $j \in S$ we have $i \sim j$ if and only if $\sigma(i) \sim \sigma(j)$, which is equivalent to $a_{ij} = a_{\sigma(i)\sigma(j)}$. Then:

$$\begin{aligned} ((\mathbf{A} + c\mathbf{I})\sigma^*\mathbf{x})_i &= \sum_{j\in S} a_{ij}x_{\sigma(j)} + cx_{\sigma(i)} = \sum_{j\in S} a_{\sigma(i)\sigma(j)}x_{\sigma(j)} + cx_{\sigma(i)} \\ &= ((\mathbf{A} + c\mathbf{I})\mathbf{x})_{\sigma(i)} = \lambda \end{aligned}$$

and so

$$((\mathbf{A} + c\mathbf{I})\hat{\mathbf{x}})_i = \frac{1}{|\mathcal{G}|}\sum_{\sigma\in\mathcal{G}}((\mathbf{A} + c\mathbf{I})\sigma^*\mathbf{x})_i = \lambda,$$

showing that $\hat{\mathbf{x}} \in \mathrm{g}\mathcal{KKT}(c)$.
Since $\mathcal{G}$ is a group, for every $\tau \in \mathcal{G}$ we have $\mathcal{G} = \tau^{-1}\mathcal{G}$, thus:

$$\tau^*\hat{\mathbf{x}} = \frac{1}{|\mathcal{G}|}\sum_{\sigma\in\mathcal{G}}\sigma^*(\tau^*\hat{\mathbf{x}}) = \frac{1}{|\mathcal{G}|}\sum_{\sigma\in\tau^{-1}\mathcal{G}}\sigma^*(\tau^*\hat{\mathbf{x}}) = \frac{1}{|\mathcal{G}|}\sum_{\sigma\in\mathcal{G}}\sigma^*\hat{\mathbf{x}} = \hat{\mathbf{x}}.$$

□

Given a non-empty $S \subseteq V$, Theorem 1 allows to infer information about the automorphism $G[S]$ provided we are able to find $\mathbf{x} \in \mathrm{g}\mathcal{KKT}(c)$ with support equal to S. For $|S| = 1$, 2 it is trivial to check that $\mathbf{x}^S \in \mathrm{g}\mathcal{KKT}(c)$. However, it is sometimes impossible to find such an $\mathbf{x}$ for $|S| \geq 3$.

[11] The reader may notice a subtle abuse of notation for the pull-back: we are identifying σ, which is a permutation on S, and the permutation on V extending σ to V so that it keeps fixed every vertex in $V \setminus S$.

Proposition 5. *Suppose three distinct vertices i_1, i_2, $i_3 \in V$ satisfy $i_1 \not\sim i_3$ and $i_2 \sim i_3$. Let $S' \subseteq V$ be such that every vertex in S' that is adjacent to i_1 is also adjacent to i_2 and set $S = S' \cup \{i_1, i_2, i_3\}$.*

(a) If $i_1 \sim i_2$, then no element of $\mathrm{g}\mathcal{KKT}(1)$ has support equal to S;

(b) If $i_1 \not\sim i_2$, then no element of $\mathrm{g}\mathcal{KKT}(0)$ has support equal to S.

Proof. (a) Set $\mathbf{M} = \mathbf{A} + \mathbf{I}$. By hypothesis $m_{i_2 j} - m_{i_1 j} \geq 0$ for every $j \in S$, and the inequality is strict for $j = i_3$. Suppose now some $\mathbf{x} \in \mathrm{g}\mathcal{KKT}(1)$ satisfies $\mathrm{supp}(\mathbf{x}) = S$. By Proposition 1, every $i \in \mathrm{supp}(\mathbf{x})$ yields the same value for the quantity $\sum_{j \in V} m_{ij} x_j$, hence:

$$0 = \sum_{j \in V} m_{i_2 j} x_j - \sum_{j \in V} m_{i_1 j} x_j = \sum_{j \in S} (m_{i_2 j} - m_{i_1 j}) x_j \geq (m_{i_2 i_3} - m_{i_1 i_3}) x_{j_0} > 0,$$

and this is absurd.

(b) This time, set $\mathbf{M} = \mathbf{A}$. Arguing as before, a vector $\mathbf{x} \in \mathrm{g}\mathcal{KKT}(0)$ such that $\mathrm{supp}(\mathbf{x}) = S$ leads to a contradiction. □

The two conclusions of Proposition 5 are indeed equivalent in light of Proposition 2. Observe that in Proposition 5 the graph $G[S]$ is isomorphic to either the cherry graph or its complement graph in case $S = \emptyset$.

Even though it is possible that no element of $\mathrm{g}\mathcal{KKT}(c)$ has support equal to S, this does not depend solely on S. In fact, $\mathrm{g}\mathcal{KKT}(c) \neq \emptyset$ in case c lies outside a suitable bounded subset of $\mathbb{R}$.

Proposition 6. *Let S be a non-empty subset of V. Then there exists a bounded interval $I \subset \mathbb{R}$ such that for every $c \in \mathbb{R} \setminus I$ at least an element of $\mathrm{g}\mathcal{KKT}(c)$ has support equal to S.*

Proof. Call $s = |S|$ and assume $s > 1$, for otherwise the proof is trivial. Observe that:

$$0 < \min_{\mathbf{x} \in \Delta_n(S)} \mathbf{x}^\top \mathbf{x} = \frac{1}{s} < \frac{1}{s-1} = \min_{\mathbf{x} \in \partial(\Delta_n(S))} \mathbf{x}^\top \mathbf{x}$$

thus:

$$\max_{\mathbf{x} \in \Delta_n(S)} f_c(\mathbf{x}) \sim \frac{c}{s}, \quad \max_{\mathbf{x} \in \partial(\Delta_n(S))} f_c(\mathbf{x}) \sim \frac{c}{s-1}$$

as $c \to -\infty$. Then, for c negative and with modulus sufficently big, the function f_c restricted to $\Delta_n(S)$ admits a maximum $\mathbf{z} \in \operatorname{int} \Delta_n(S)$. By construction, $\mathbf{z} \in \mathrm{g}\mathcal{KKT}(c)$. To complete the proof, apply the same idea to $\overline{G}$ in place of G and use Proposition 2. □

As an immediate application of Theorem 1, we can show that in case $-c$ is not an eigenvalue of $\mathbf{A}$, then a KKT point $\mathbf{x}$ reveals additional information about its support, since in this case for *no* automorphism σ of $G[\mathrm{supp}(\mathbf{x})]$ two vertices i and j lying in the same orbit under σ may satisfy $x_i \neq x_j$.

Recall that for the matrix $\mathbf{A}$ and a non-empty subset $S \subset V$ the *principal submatrix* $\mathbf{A}[S, S]$ is the submatrix of $\mathbf{A}$ having entries in the rows and columns of $\mathbf{A}$ indexed by S [11].[12] The concept of induced partition is the final ingredient for Corollary 1.

Definition 3. *Given $\mathbf{x} \in \Delta_n$, define on* $\mathrm{supp}(\mathbf{x})$ *the equivalence relation $\sim_{\mathbf{x}}$ such that $i \sim_{\mathbf{x}} j$ if and only if $x_i = x_j$. The* partition induced by $\mathbf{x}$ *is the family of the equivalence classes of $\sim_{\mathbf{x}}$.*

[12] A tighter bound on the interval I in Proposition 6 can be derived from the spectral radius of $\mathbf{A}[S, S]$, as shown in [16, Theorem 1].

Corollary 1. *Let* $\mathbf{x} \in \mathcal{KKT}(c)$*, set* $S = \mathrm{supp}(\mathbf{x})$ *and suppose* $-c$ *is* not *an eigenvalue of* $\mathbf{A}[S, S]$*. Every class of the partition induced by* $\mathbf{x}$ *is invariant under* every *automorphism of* $G[S]$*.*

Proof. The thesis is that $x_{\sigma(i)} = x_i$ for every $i \in S$ and automorphism σ for the induced subgraph $G[S]$. Let σ be an automorphism for $G[S]$. Apply Theorem 1 for the group of automorphisms generated by σ to get $\hat{\mathbf{x}} \in \mathrm{g}\mathcal{KKT}(c)$ satisfying $\mathrm{supp}(\hat{\mathbf{x}}) = S$ and $\hat{x}_{\sigma(i)} = \hat{x}_i$ for every $i \in S$. The hypothesis on the spectrum of $\mathbf{A}[S, S]$ entails that $\mathrm{g}\mathcal{KKT}(c)$ contains one point at most with support equal to S. Consequently, $\mathbf{x} = \hat{\mathbf{x}}$. □

6 KKT points and convex hulls of characteristic vectors

We are about to discuss representations of elements in the standard simplex as convex combinations of characteristic vectors. Such representations turn out to be interesting for KKT points of (5).

Definition 4. *Consider a family* $\mathcal{F} = \{V_1, V_2, \ldots, V_k\}$ *of pairwise disjoint non-empty subsets of* V*. Given* $\mathbf{x} \in \mathrm{conv}(\mathbf{x}^{V_1}, \mathbf{x}^{V_2}, \ldots, \mathbf{x}^{V_k})$*, the barycentric coordinates of* $\mathbf{x}$ *with respect to (the characteristic vectors representing the classes of)* $\mathcal{F}$ *is the unique*[13] *vector* $\mathbf{y} = \mathrm{bary}_{\mathcal{F}}(\mathbf{x})$ *in* Δ_k *such that* $\mathbf{x} = \sum_{\ell=1}^{k} y_\ell \mathbf{x}^{V_\ell}$*.*

For instance, in case $\mathcal{F} = \{\{i\} \mid i \in V\}$, it is trivial to check that for every $\mathbf{x}$ in Δ_n the equality $\mathrm{bary}_{\mathcal{F}}(\mathbf{x}) = \mathbf{x}$. Observe that in the setting of Definition 5 we must have $x_i = x_j = y_\ell/|V_\ell|$ in case $i, j \in V_\ell$. Moreover, it is easy to see that $\mathrm{bary}_{\mathcal{F}}(\mathbf{x})$ lies in $\mathrm{int}(\Delta_k)$ if and only if $\mathrm{supp}(\mathbf{x}) = \cup_{\ell=1}^{k} V_\ell$.

Definition 5. *Let* $\mathbf{x} \in \Delta_n$*. A partition* $\mathcal{P}$ *of* $\mathrm{supp}(\mathbf{x})$ separates distinct values *of* $\mathbf{x}$ *if for every* $i, j \in \mathrm{supp}(\mathbf{x})$ *the relation* $x_i \neq x_j$ *implies that the vertices* i *and* j *belong to distinct classes of* $\mathcal{P}$*.*

In other words, a partition $\mathcal{P}$ of $\mathrm{supp}(\mathbf{x})$ separates distinct values if and only if it is finer than the partition induced by $\mathbf{x}$.

Choose $\mathbf{x} \in \Delta_n$ and consider a partition $\mathcal{P} = \{V_1, V_2, \ldots, V_k\}$ of $\mathrm{supp}(\mathbf{x})$ separating distinct values of $\mathbf{x}$. Then $\mathbf{x} \in \mathrm{span}(\mathbf{x}^{V_1}, \mathbf{x}^{V_2}, \ldots, \mathbf{x}^{V_k})$ and thanks to Proposition 7 we get $\mathbf{x} \in \mathrm{conv}(\mathbf{x}^{V_1}, \mathbf{x}^{V_2}, \ldots, \mathbf{x}^{V_k})$, hence it makes sense to consider $\mathrm{bary}_{\mathcal{P}}(\mathbf{x})$. In particular, in the setting of Theorem 1, such a partition for $\mathrm{supp}(\hat{\mathbf{x}})$ is given by the orbits of $\mathrm{supp}(\hat{\mathbf{x}})$ under the action of $\mathcal{G}$.

Proposition 7. *Consider a family* $\{V_1, V_2, \ldots V_k\}$ *of pairwise disjoint non-empty subsets of* V*. Then:*

$$\mathrm{conv}(\mathbf{x}^{V_1}, \mathbf{x}^{V_2}, \ldots, \mathbf{x}^{V_k}) = \Delta_n \cap \mathrm{span}(\mathbf{x}^{V_1}, \mathbf{x}^{V_2}, \ldots, \mathbf{x}^{V_k}).$$

Proof. The vectors $\mathbf{x}^{V_1}, \mathbf{x}^{V_2}, \ldots, \mathbf{x}^{V_k}$ are elements of $\mathrm{span}(\mathbf{x}^{V_1}, \mathbf{x}^{V_2}, \ldots, \mathbf{x}^{V_k})$ and of Δ_n, which are convex sets, hence $\mathrm{conv}(\mathbf{x}^{V_1}, \mathbf{x}^{V_2}, \ldots, \mathbf{x}^{V_k})$ is included in their intersection. The trivial inclusion $\mathrm{conv}(\mathbf{x}^{V_1}, \mathbf{x}^{V_2}, \ldots, \mathbf{x}^{V_k}) \subseteq \Delta_n \cap \mathrm{span}(\mathbf{x}^{V_1}, \mathbf{x}^{V_2}, \ldots, \mathbf{x}^{V_k})$ is thus proved.

[13] Strictly speaking, the uniqueness depends also on the enumeration of the classes forming the partition $\mathcal{F}$. Many of the results contained in this section depend indeed on the enumeration of the classes, we omit writing it explicitly to simplify the notation. Barycentric coordinates are widely employed in finite element method and computer graphics [19, 10].

To prove the reversed inclusion, consider some real coefficients $a_1, a_2, \ldots, a_n$ such that $\mathbf{x} = \sum_{\ell=1}^{k} a_\ell \mathbf{x}^{V_\ell}$ is an element of the standard simplex Δ_n. Every component of $\mathbf{x}$ is non-negative, and since the sets $V_1, V_2, \ldots, V_k$ are pairwise disjoint[14] this means that $a_\ell \geq 0$ for all $\ell \in [k]$. Using again that $\mathbf{x} \in \Delta_n$ we obtain:

$$1 = \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \left(\sum_{\ell=1}^{k} a_\ell \mathbf{x}^{V_\ell} \right)_i = \sum_{\ell=1}^{k} \sum_{i=1}^{n} \left(a_\ell \mathbf{x}^{V_\ell} \right)_i = \sum_{\ell=1}^{k} a_\ell.$$

Then $\mathbf{x}$ is a convex combination of $\mathbf{x}^{V_1}, \mathbf{x}^{V_2}, \ldots, \mathbf{x}^{V_k}$. □

Some additional graph theory tools [6] can help recognizing some properties of barycentric coordinates for KKT points.

For every non-empty $S_1, S_2 \subseteq V$, let $e_G(S_1, S_2)$ count the ordered[15] pairs of adjacent vertices in the set $S_1 \times S_2$:

$$e_G(S_1, S_2) = |\{(i, j) \in S_1 \times S_2 \mid i \sim j\}|$$

and call *edge density* between S_1 and S_2 the ratio:

$$d_G(S_1, S_2) = \frac{e_G(S_1, S_2)}{|S_1||S_2|}.$$

For a finite family $\mathcal{F} = \{V_1, V_2, \ldots V_k\}$ of distinct non-empty subsets of V that are not necessarily pairwise disjoint, call *density matrix* of $\mathcal{F}$ the matrix $\mathbf{D} \in \mathbb{R}^{k \times k}$ with general coefficient $d_{\ell,m} = d_G(V_\ell, V_m)$.[16] By definition, $\mathbf{D}$ is a symmetric matrix. The following lemma will be useful to prove Theorem 2 and Theorem 3.

Lemma 1. *Let $\mathbf{x} \in \Delta_n$, let $\mathcal{P} = \{V_1, V_2, \ldots V_k\}$ be a partition of* $\operatorname{supp}(\mathbf{x})$ *separating distinct values of $\mathbf{x}$, and set $\mathbf{y} = \operatorname{bary}_{\mathcal{P}}(\mathbf{x})$. Then for every vertex $i \in V$:*

$$(\mathbf{A}\mathbf{x})_i = \sum_{m=1}^{k} d_G(\{i\}, V_m) y_m.$$

Proof. Since $\mathbf{x} = \sum_\ell y_\ell \mathbf{x}^{V_\ell}$, then:

$$\begin{aligned}(\mathbf{A}\mathbf{x})_i &= \sum_{j=1}^{n} a_{ij} x_j = \sum_{m=1}^{k} \sum_{j \in V_m} a_{ij} x_j = \sum_{m=1}^{k} \sum_{j \in V_m} a_{ij} (y_m / |V_m|) \\ &= \sum_{m=1}^{k} \left(e_G(\{i\}, V_m) / |V_m| \right) y_m = \sum_{m=1}^{k} d_G(\{i\}, V_m) y_m.\end{aligned}$$

□

We are going to prove the main result of this section, namely that for $\mathbf{x} \in \mathcal{KKT}(c)$, under a suitable choice of $\mathcal{F}$ separating distinct values of $\mathbf{x}$, also $\operatorname{bary}_{\mathcal{F}}(\mathbf{x})$ is a KKT point for a quadratic program having as many variables as $|\mathcal{F}|$.

[14] This hypothesis can be relaxed to $V_\ell \not\subseteq \cup_{m \neq \ell} V_m$ for every ℓ, and the conclusion $a_\ell \geq 0$, which is what we are interested in, would follow as well.

[15] The definition of $e_G(S_1, S_2)$ can be regarded as a way to count the edges crossing S_1 and S_2, keeping in mind that each edge with both endnodes in $S_1 \cap S_2$ is counted twice.

[16] Also $\mathbf{D}$ depends on the enumeration of the sets in the family $\mathcal{F}$.

Theorem 2. *Let* $\mathbf{x} \in \Delta_n$ *and let the partition* $\mathcal{P} = \{V_1, V_2, \dots V_k\}$ *of* $\operatorname{supp}(\mathbf{x})$ *separate distinct values of* $\mathbf{x}$*. Call* $\mathbf{D}$ *the density matrix associated with* $\mathcal{F}$ *and set:*

$$\mathbf{\Lambda} = \operatorname{diag}(|V_1|, |V_2|, \dots, |V_k|) = \begin{pmatrix} |V_1| & & & \\ & |V_2| & & \\ & & \ddots & \\ & & & |V_k| \end{pmatrix}$$

If $\mathbf{x} \in \mathcal{KKT}(c)$*,*[17] *then* $\operatorname{bary}_{\mathcal{F}}(\mathbf{x})$ *is a KKT point for the program:*

$$\begin{array}{cc} \text{maximize} & \mathbf{y}^\top (\mathbf{D} + c\mathbf{\Lambda}^{-1})\mathbf{y}. \\ \mathbf{y} \in \operatorname{int}(\Delta_k) & \end{array} \tag{10}$$

Proof. Set $\mathbf{y} = \operatorname{bary}_{\mathcal{F}}(\mathbf{x})$. We may write $\mathbf{x} = \sum_\ell y_\ell \mathbf{x}^{V_\ell}$ by definition of $\mathbf{y}$. By Proposition 1 and Lemma 1, there exists λ such that for every i in the support of $\mathbf{x}$:

$$\lambda = \left((\mathbf{A} + c\mathbf{I})\, \mathbf{x} \right)_i = \sum_{m=1}^{k} d_G(\{i\}, V_m) y_m + c x_i.$$

Consider now $\ell \in [k]$. Computing the arithmetic mean of the previous expression as i varies in V_ℓ we get

$$\begin{aligned} \lambda &= \frac{1}{|V_\ell|} \sum_{i \in V_\ell} \left(\sum_{m=1}^{k} d_G(\{i\}, V_m) y_m + c x_i \right) \\ &= \frac{1}{|V_\ell|} \sum_{i \in V_\ell} \left(\sum_{m=1}^{k} e_G(\{i\}, V_m)(y_m / |V_m|) + c(y_\ell / |V_\ell|) \right) \\ &= \sum_{m=1}^{k} d_G(V_\ell, V_m) y_m + (c/|V_\ell|) y_\ell \\ &= \left(\left(\mathbf{D} + c\mathbf{\Lambda}^{-1} \right) \mathbf{y} \right)_\ell . \end{aligned}$$

Then $\mathbf{y}$ is a KKT point for (10).

$\square$

What one could hope is that a converse of Theorem 2 holds. Still, suppose G is the graph on $V = [4]$ with adjacency matrix:

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \tag{11}$$

and consider the characteristic vector $\mathbf{x} = \mathbf{x}^V$. For $V_1 = \{1, 2\}$ and $V_2 = \{2, 3\}$ the family $\mathcal{P} = \{V_1, V_2\}$ partitions the support of $\mathbf{x}$, and in this case:

$$\mathbf{D} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{\Lambda} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}. \tag{12}$$

[17] The proof requires only that $\mathbf{x} \in \mathrm{g}\mathcal{KKT}(c)$.

Observe that $\mathrm{bary}_{\mathcal{P}}(\mathbf{x}) = \left(\dfrac{1}{2}, \dfrac{1}{2}\right)$ is a KKT point for (10) regardless of the value of c, whereas $\mathbf{x} \notin \mathcal{KKT}(c)$.

Theorem 3 allows to obtain a partial converse of Theorem 2 in case stronger hypotheses hold.

Definition 6. *Consider a finite family $\mathcal{F} = \{V_1, V_2, \dots V_k\}$ of pairwise disjoint non-empty subsets of V. We call $\mathcal{F}$* highly regular *for G if:*[18]

$$d_G(V_\ell, V_m) = d_G(\{i\}, V_m) \quad \text{for every } \ell,\, m \in [k] \text{ and every } i \in V_\ell.$$

A finite family $\mathcal{F} = \{V_1, V_2, \dots V_k\}$ of subsets of V is highly regular for G if and only if the following two conditions hold:

(a) for every $\ell \in [k]$ the set V_ℓ is non-empty and the induced subgraph $G[V_\ell]$ is regular;

(b) for every distinct ℓ, $m \in [k]$ the sets V_ℓ, V_m are disjoint and each vertex in V_ℓ has the same amount of neighbors in V_m.

In fact, suppose $\mathcal{F}$ is highly regular. Then every vertex in V_ℓ has $d_G(V_\ell, V_m)|V_m|$ neighbors in V_m for every ℓ, $m \in [k]$, and this proves (a) and (b).
Conversely, assume (a) and (b). Fix ℓ, $m \in [k]$ and some $i \in V_\ell$. It's easy to see that the equality $\sum_{j \in V_\ell} e_G(\{j\}, V_m) = e_G(V_\ell, V_m)$ holds true and that every term appearing in the summation is equal to $e_G(\{i\}, V_m)$, thanks to (a) in case $\ell = m$ and to (b) in case $\ell \neq m$. Hence dividing both sides of the equality by $|V_\ell||V_m|$ we get $d_G(\{i\}, V_m) = d_G(V_\ell, V_m)$, and this value is independent of the choice of i in V_ℓ.

The following proposition gives some examples of highly regular families.

Proposition 8. *Consider some non-empty $S \subseteq V$. Then:*

1. *The family $\{\{i\} \mid i \in S\}$ is highly regular for G.*

2. *The induced subgraph $G[S]$ is regular if and only if the family $\{S\}$ is highly regular for G.*

3. *Suppose S is not an independent set. Then S is a clique if and only if every partition of S is highly regular for G.*

Proof. (1) Trivial. (2) It follows from the equivalent formulation of highly regular families. (3) Every partition of a clique is trivially highly regular.
To prove the other implication, assume there exist two adjacent vertices in i_1, $i_2 \in S$ and that every partition of C is highly regular. Then the partition $\{\{i_1\}, S \setminus \{i_1\}\}$ is highly regular, and we get $d_G(\{i_0\}, \{i_1 \;= d_G(\{i_2\}, \{i_1 \;= 1$ for every vertex $i_0 \in S \setminus \{i_1\}$, which means that the degree of i_1 in $G[S]$ is $|S| - 1$. Also $\{S\}$ is highly regular, thus by (2) each vertex of the induced subgraph $G[S]$ has degree $|S| - 1$, *i.e.*, S is a clique. □

We are now in the position to prove Theorem 3.

Theorem 3. *Let $\mathbf{x} \in \Delta_n$ and let the partition $\mathcal{P} = \{V_1, V_2, \dots V_k\}$ of* $\mathrm{supp}(\mathbf{x})$ *separate distinct values of $\mathbf{x}$. Call $\mathbf{D}$ the density matrix associated with $\mathcal{P}$ and set $\mathbf{\Lambda} = \mathrm{diag}(|V_1|, |V_2|, \dots, |V_k|)$. If $\mathcal{P}$ is highly regular for G, then $\mathbf{x} \in \mathrm{g}\mathcal{KKT}(c)$ if and only if* $\mathrm{bary}_{\mathcal{P}}(\mathbf{x})$ *is a KKT point for* (10).

[18]Equivalently, for every non-empty $X \subseteq V_\ell$ and every non-empty $Y \subseteq V_m$ we have $d_G(X, V_m) = d_G(V_\ell, V_m) = d_G(V_\ell, Y)$.

Proof. If $\mathbf{x} \in \mathrm{g}\mathcal{KKT}(c)$ then $\mathrm{bary}_{\mathcal{P}}(\mathbf{x})$ is a KKT point for (10) by Theorem 2. Set now $\mathbf{y} = \mathrm{bary}_{\mathcal{P}}(\mathbf{x})$ and assume $\mathcal{P}$ is higly regular and that $\mathbf{y}$ is a KKT point for (10). Then $\mathrm{supp}(\mathbf{y}) = [k]$, and there exists $\lambda \in \mathbb{R}$ such that for every $\ell \in [k]$:

$$\big((\mathbf{D} + c\mathbf{\Lambda}^{-1})\mathbf{y}\big)_\ell = \lambda.$$

Pick any i in the support of $\mathbf{x}$. The vertex i is in V_ℓ for some $\ell \in [k]$ and $d_G(\{i\}, V_m) = d_G(V_\ell, V_m)$ since $\mathcal{P}$ is highly regular. By Lemma 1:

$$\begin{aligned}((\mathbf{A} + c\mathbf{I})\mathbf{x})_i &= \sum_{m=1}^{k} d_G(\{i\}, V_m) y_m + c x_i \\ &= \sum_{m=1}^{k} d_G(V_\ell, V_m) y_m + (c/|V_\ell|) y_\ell = \lambda.\end{aligned}$$

Then $\mathbf{x} \in \mathrm{g}\mathcal{KKT}(c)$ thanks to Proposition 1. □

Specializing Theorem 3 for a family $\{V_1, V_2\}$ that is highly regular for G we get Corollary 2 and Corollary 3. The two corollaries differ in the hypothesis on the regularity of $G[V_1 \cup V_2]$, and this produces different behaviors on how $\mathrm{g}\mathcal{KKT}(c)$ intersects the set $[\mathbf{x}^{V_1}, \mathbf{x}^{V_2}] = \mathrm{conv}(\mathbf{x}^{V_1}, \mathbf{x}^{V_2})$.

Corollary 2. *Let $\{V_1, V_2\}$ be highly regular for G and assume $G[V_1 \cup V_2]$ is a regular graph. There exists $c^* \in \mathbb{R}$ such that:*

- *If $c = c^*$, then $\mathrm{g}\mathcal{KKT}(c) \cap [\mathbf{x}^{V_1}, \mathbf{x}^{V_2}] = [\mathbf{x}^{V_1}, \mathbf{x}^{V_2}]$;*
- *If $c \neq c^*$, then $\mathrm{g}\mathcal{KKT}(c) \cap [\mathbf{x}^{V_1}, \mathbf{x}^{V_2}] = \{\mathbf{x}^{V_1}, \mathbf{x}^{V_2}, \mathbf{x}^{V_1 \cup V_2}\}$.*

Proof. Let $\mathbf{D} \in \mathbb{R}^{2\times 2}$ be the density matrix associated with $\{V_1, V_2\}$ and set $\alpha = |V_2|(d_{12} - d_{22})$ and $\beta = |V_1|(d_{21} - d_{11})$.

Observe first that $G[V_1 \cup V_2]$ is a regular graph if and only if $\alpha = \beta$. Indeed, for $i = 1$, 2 every vertex in V_i is adjacent to $d_{i1}|V_1| + d_{i2}|V_2|$ vertices of $V_1 \cup V_2$. This means that $G[V_1 \cup V_2]$ is regular if and only if $d_{11}|V_1| + d_{12}|V_2| = d_{21}|V_1| + d_{22}|V_2|$, which is equivalent to $\alpha = \beta$.

Both $\mathbf{x}^{V_1}$ and $\mathbf{x}^{V_2}$ are in $\mathrm{g}\mathcal{KKT}(c)$ regardless of the value of c as a consequence of Proposition 3. By Theorem 3, we can find the remaining elements of $\mathrm{g}\mathcal{KKT}(c)$ within $[\mathbf{x}^{V_1}, \mathbf{x}^{V_2}]$ by looking for points of the form $y_1\mathbf{x}^{V_1} + y_2\mathbf{x}^{V_2}$, where $(y_1, y_2)^\top \in \mathrm{int}(\Delta_2)$ satisfies for some parameter λ:

$$\left(\mathbf{D} + c\,\mathrm{diag}\left(|V_1|^{-1}, |V_2|^{-1}\right)\right) \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \lambda \\ \lambda \end{pmatrix}$$

By eliminating λ, this means that:

$$(c/|V_1| + d_{11_{y_1 + d_{12}}})\, y_2 = d_{21} y_1 + (d_{22} + c/|V_2|) y_2,$$

which is equivalent to:

$$(c - \beta) y_1/|V_1| = (c - \alpha) y_2/|V_2|. \tag{13}$$

Set $c^* = \alpha = \beta$. Then for $c = c^*$ every $(y_1, y_2)^\top \in \mathrm{int}(\Delta_2)$ satisfies (13). For $c \neq c^*$, dividing both sides in (13) by $c - c^*$ yields $y_1/|V_1| = y_2/|V_2|$, leading to the solution $\mathbf{x}^{V_1 \cup V_2}$. □

Corollary 3. *Let $\{V_1, V_2\}$ be highly regular for G and assume $G[V_1 \cup V_2]$ is* not *a regular graph. There exists an interval $[a, b] \subset \mathbb{R}$ such that:*

- *If $c \in [a, b]$, then $\mathrm{g}\mathcal{KKT}(c) \cap [\mathbf{x}^{V_1}, \mathbf{x}^{V_2}] = \{\mathbf{x}^{V_1}, \mathbf{x}^{V_2}\}$;*

- *If $c \notin [a,b]$, then* $\mathrm{g}\mathcal{KKT}(c) \cap [\mathbf{x}^{V_1}, \mathbf{x}^{V_2}] = \{\mathbf{x}^{V_1}, \mathbf{x}^{V_2}, \mathbf{x}_c\}$ *for some* $\mathbf{x}_c \in \Delta_n$ *depending on* c *that is not a characteristic vector.*

Proof. Define α and β as in the proof of Corollary 2. Then that proof shows that $\alpha \neq \beta$ in this case. Set $a = \min(\alpha, \beta)$ and $b = \max(\alpha, \beta)$. Looking for generalized KKT points in $[\mathbf{x}^{V_1}, \mathbf{x}^{V_2}]$ leads to $\mathbf{x}^{V_1}$, $\mathbf{x}^{V_2}$, and to points of the form $y_1\mathbf{x}^{V_1} + y_2\mathbf{x}^{V_2}$, where $(y_1, y_2)^\top \in \mathrm{int}(\Delta_2)$ solves (13). For positive y_1, y_2 a solution to (13) is possible only in case $c \notin [a,b]$, and if that occurs then the unique solution is $\mathbf{x}_c = y_1\mathbf{x}^{V_1} + y_2\mathbf{x}^{V_2}$, where:

$$\begin{aligned} y_1 &= \frac{(c-\alpha)|V_1|}{(c-\alpha)|V_1| + (c-\beta)|V_2|} \\ y_2 &= \frac{(c-\beta)|V_2|}{(c-\alpha)|V_1| + (c-\beta)|V_2|}. \end{aligned}$$

□

Corollary 3, which is applicable to the cherry graph, is also useful for a broader class of graphs. Recall that a *star* is a complete bipartite graph in which one vertex, called *center* of the star, is adjacent to every edge of the graph [6]. The cherry graph (Fig. 1) is trivially a star, with center $z = 3$.

Definition 7. *A graph $G = (V, E)$ is a* generalized star *with core H if there exists a graph $S = (V', E')$ and a surjection $\phi\colon V \to V'$ such that:*

- *S is a star with center $z \in V'$ and $H = \phi^{-1}(z)$;*
- *Every node in H is adjacent to every node in $V \setminus H$;*
- *G is not complete, whereas the induced subgraph $G[H]$ is complete;*
- *The induced subgraph $G[V \setminus H]$ is regular.*

Theorem 4. *Let H, P be disjoint subsets of V such that $G[H \cup P]$ is a generalized star with core H. There exists an integer $b > 1$ such that, if $c \notin [1,b]$, then* $\mathrm{g}\mathcal{KKT}(c)$ *contains a vector with support $H \cup P$.*

Proof. Set $h = |H|$, $p = |P|$ and assume $G[P]$ is a d-regular graph. By hypothesis, the family $\{P, H\}$ is highly regular for G and the associated density matrix is:

$$\mathbf{D} = \begin{pmatrix} d/p & 1 \\ 1 & 1 - 1/h \end{pmatrix}.$$

The integer $b = p - d$ satisfies $1 < b < p$ since $G[P]$ is not complete. By Corollary 3, for $c \in \mathbb{R} \setminus [1,b]$ and

$$y_1 = \frac{(c-1)p}{(c-1)p + (c-b)h} \tag{14}$$

$$y_2 = \frac{(c-b)h}{(c-1)p + (c-b)h} \tag{15}$$

the point $\mathbf{x} = y_1\mathbf{x}^P + y_2\mathbf{x}^H$ is an element of $\mathrm{g}\mathcal{KKT}(c)$. □

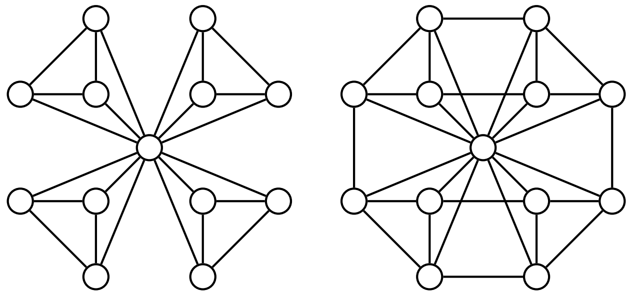

Figure 2: Generalized stars

In [17, Theorem 10], a configuration of cliques $C_1, C_2, \ldots, C_q$ is exhibited such that the convex hull $\mathrm{conv}(\mathbf{x}^{C_1}, \ldots, \mathbf{x}^{C_q})$ is entirely contained in $\mathrm{g}\mathcal{KKT}(0)$, due to the fact that every point of that convex hull is a local solution to the parametric Motzkin-Straus program for $c = 0$ (recently, this has been generalized by Tang *et al.* in [23]). Theorem 4 allows to exhibit a particular configuration of cliques $C_1, C_2, \ldots, C_q$ and conditions on c such that the set $\mathrm{g}\mathcal{KKT}(c) \setminus \mathrm{conv}(\mathbf{x}^{C_1}, \ldots, \mathbf{x}^{C_q})$ contains a vector with support $\cup_\ell C_\ell$.

Corollary 4. *Consider $q \geq 2$ distinct cliques $C_1, C_2, \ldots, C_q$ such that:*

- *the set $H = \cap_\ell C_\ell$ is not empty and $C_\ell \cap C_m = H$ for every distinct ℓ, $m \in [q]$;*
- *the set $\cup_\ell C_\ell \setminus H$ is not empty and the induced subgraph $G[\cup_\ell C_\ell \setminus H]$ is regular but not complete.*

There exist $c_0 < 1$ and $2 < b < |\cup_\ell C_\ell|$ such that, if $c \notin \{c_0\} \cup [1, b]$, then the set $\mathrm{g}\mathcal{KKT}(c) \setminus \mathrm{conv}(\mathbf{x}^{C_1}, \ldots, \mathbf{x}^{C_q})$ contains a vector with support $\cup_\ell C_\ell$.

Proof. Call $P = |\cup_\ell C_\ell \setminus H|$ and observe that $G[H \cup P]$ is a generalized star. By Theorem 4 the point $\mathbf{x} = y_1 \mathbf{x}^P + y_2 \mathbf{x}^H$ is an element of $\mathrm{g}\mathcal{KKT}(c)$ if we set:

$$y_1 = \frac{(c-1)p}{(c-1)p + (c-b)h} \tag{16}$$

$$y_2 = \frac{(c-b)h}{(c-1)p + (c-b)h} \tag{17}$$

$h = |H|$, $p = |P|$, $b = p - h$ under the assumptions that $G[P]$ is a d-regular graph and $c \in \mathbb{R} \setminus [1, b]$. However, nothing so far proved excludes that $\mathbf{x} \in \mathrm{conv}(\mathbf{x}^{C_1}, \ldots, \mathbf{x}^{C_q})$. Suppose it is possible to write $\mathbf{x}$ as a convex combination of $\mathbf{x}^{C_1}, \mathbf{x}^{C_2}, \ldots, \mathbf{x}^{C_q}$. Equations (16) and (17) yield the equality $(y_2/h)/(y_1/p) = (c-b)/(c-1)$.
There is an alternative way to compute the ratio $(y_2/h)/(y_1/p)$. In fact, notice that $\mathbf{x}_i = \mathbf{x}_j$ whenever $i, j \in \cup_\ell C_\ell \setminus H$, and by hypothesis the cliques $C_1, \ldots, C_q$ have the same cardinality. Therefore, $\mathbf{x}$ must be the arithmetic mean of $\mathbf{x}^{C_1}, \mathbf{x}^{C_2}, \ldots, \mathbf{x}^{C_q}$, and so:

$$\begin{cases} y_1/p = 1/(p + qh) \\ y_2/h = q/(p + qh), \end{cases}$$

thus $(y_2/h)/(y_1/p) = q$. The two equalities obtained for $(y_2/h)/(y_1/p)$ give $(c-b)/(c-1) = q$, which solved for c gives $c = (q-b)/(q-1)$. Set $c_0 = (q-b)/(q-1)$, and observe that $b \geq 2$ implies $c_0 < 1$. $\square$

7 Conclusion

In this article, we have discussed some properties of the KKT points of the parametric Motzkin-Straus programs introduced by Bomze *et al.*. We would like to mention that Theorem 2 and Theorem 3 have a nice interpretation in the replicator dynamics setting, as they may provide a correspondence between stationary points of distinct dynamics running on simplices of distinct dimension. In this regard, it would be interesting to fruitfully apply the topic discussed to the replicator dynamics and use the resulting stationary point, that are generalized KKT points for the parametric programs mentioned, as a means to probe a given graph, so as to detect symmetries and regular structures therein, besides cliques. This would be especially useful in computer science applications.

Acknowledgments

The authors would like to thank Morteza H. Chehreghani for his insightful advice and Sebastiano Smaniotto for his help in the initial stages of this work.

References

[1] I. M. Bomze. Evolution towards the maximum clique. *J. Glob. Optim.*, 10(2):143–164, 1997.

[2] I. M. Bomze, M. Budinich, P. M. Pardalos, and M. Pelillo. The Maximum Clique Problem. In D.-Z. Du and P. M. Pardalos, editors, *Handbook of Combinatorial Optimization*, pages 1–74. Springer US, Boston, MA, 1999.

[3] I. M. Bomze, M. Budinich, M. Pelillo, and C. Rossi. Annealed replication: a new heuristic for the maximum clique problem. *Discret. Appl. Math.*, 121(1-3):27–49, 2002.

[4] I. M. Bomze, M. Pelillo, and V. Stix. Approximating the maximum weight clique using replicator dynamics. *IEEE Trans. Neural Networks*, 11(6):1228–1241, 2000.

[5] M. Budinich. Exact bounds on the order of the maximum clique of a graph. *Discrete Applied Mathematics*, 127(3):535–543, May 2003.

[6] R. Diestel. *Graph theory.* Springer, Heidelberg, Germany, 2005.

[7] R. A. Fisher. *The genetical theory of natural selection.* Oxford University Press, Oxford, UK, 1999.

[8] L. E. Gibbons, D. W. Hearn, P. M. Pardalos, and M. V. Ramana. Continuous characterizations of the maximum clique problem. *Math. Oper. Res.*, 22(3):754–768, 1997.

[9] J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics.* Cambridge University Press, Cambridge, UK, 1998.

[10] K. Hormann and N. Sukumar. *Generalized barycentric coordinates in computer graphics and computational mechanics.* CRC Press, Boca Raton, Florida, 2017.

[11] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, New York, NY, 2nd edition, 2013.

[12] J. Hungerford and F. Rinaldi. A general regularized continuous formulation for the maximum clique problem. *Math. Oper. Res.*, 44(4):1161–1173, 09 2017.

[13] D. G. Luenberger and Y. Ye. *Linear and Nonlinear Programming*, volume 228. Springer International Publishing, Cham, Switzerland, 2016.

[14] T. S. Motzkin and E. G. Straus. Maxima for graphs and a new proof of a theorem of Turán. *Can. J. Math.*, 17:533–540, 1965.

[15] P. M. Pardalos and A. T. Phillips. A global optimization approach for solving the maximum clique problem. *Int. J. Comput. Math.*, 33:209–216, 1990.

[16] M. Pavan and M. Pelillo. Dominant sets and hierarchical clustering. In *9th IEEE International Conference on Computer Vision (ICCV), Nice, France*, pages 362–369, 2003.

[17] M. Pelillo and A. Jagota. Feasible and infeasible maxima in a quadratic program for maximum clique. *J. Artif. Neural Netw.*, 2(4):411–420, 1996.

[18] M. Pelillo and A. Torsello. Payoff-monotonic game dynamics and the maximum clique problem. *Neural Comput.*, 18(5):1215–1258, 2006.

[19] A. Quarteroni. *Numerical Models for Differential Problems.* Springer, Cham, Switzerland, 3rd edition, 2017.

[20] S. Rota Bulò and M. Pelillo. A generalization of the Motzkin–Straus theorem to hypergraphs. *Optim. Lett.*, 3(2):287–295, 2009.

[21] S. Rota Bulò and M. Pelillo. Dominant-set clustering: A review. *Eur. J. Oper. Res.*, 262(1):1–13, 2017.

[22] V. Stozhkov, G. Pastukhov, V. Boginski, and E. L. Pasiliao. New analytical lower bounds on the clique number of a graph. *Optimization Methods and Software*, 32(2):336–368, Mar. 2017.

[23] Q. Tang, X. Zhang, C. Zhao, and P. Zhao. On the maxima of Motzkin-Straus programs and cliques of graphs. *J. Glob. Optim.*, 84(4):989–1003, 2022.

[24] P. D. Taylor and L. B. Jonker. Evolutionarily stable strategies and game dynamics. *Math. Biosci.*, 40:145–156, 1978.

[25] J. Weibull. *Evolutionary game theory.* MIT Press, Cambridge, Massachusetts, 1995.

[26] H. S. Wilf. Spectral bounds for the clique and independence numbers of graphs. *Journal of Combinatorial Theory, Series B*, 40(1):113–117, Feb. 1986.