
Mathematics of Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

# A General Regularized Continuous Formulation for the Maximum Clique Problem

James T. Hungerford, Francesco Rinaldi

MATHEMATICS OF OPERATIONS RESEARCH
Vol. 44, No. 4, November 2019, pp. 1161–1173
ISSN 0364-765X (print), ISSN 1526-5471 (online)
http://pubsonline.informs.org/journal/moor

# A General Regularized Continuous Formulation for the Maximum Clique Problem

**James T. Hungerford,[a] Francesco Rinaldi[b]**

[a] Racetrac Store Support Center, Atlanta, Georgia 30339; [b] Dipartimento di Matematica, Università di Padova, 35121 Padova, Italy
**Contact:** jamesthungerford@gmail.com, http://orcid.org/0000-0002-9898-4911 (JTH); rinaldi@math.unipd.it (FR)

**Received:** September 5, 2017
**Revised:** May 14, 2018
**Accepted:** June 4, 2018
**Published Online in Articles in Advance:** March 19, 2019

**MSC2000 Subject Classification:** Primary: 90C30; secondary: 90C35, 90C20
**OR/MS Subject Classification:** Programming: quadratic; networks/graphs: heuristics

https://doi.org/10.1287/moor.2018.0954



**Abstract.** In this paper, we develop a general regularization-based continuous optimization framework for the maximum clique problem. In particular, we consider a broad class of regularization terms that can be included in the classic Motzkin–Straus formulation, and we develop conditions that guarantee the equivalence between the continuous regularized problem and the original one in both a global and a local sense. We further analyze, from a computational point of view, two different regularizers that satisfy the general conditions.

**Funding:** J. T. Hungerford gratefully acknowledges the financial support of the European Union's Horizon 2020 Marie Curie Network for Initial Training Programme [Grant 316647].

**Keywords:** maximum clique • sparse optimization • support • concave minimization • Motzkin–Straus



## 1. Introduction

Let $G = (\mathcal{V}, \mathcal{E})$ be a simple undirected graph on vertex set $\mathcal{V} = \{1, 2, \ldots, n\}$ and edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. Because $G$ is simple and undirected, $(j, i) \in \mathcal{E}$ whenever $(i, j) \in \mathcal{E}$ and $(i, i) \notin \mathcal{E}$ for any $i \in \mathcal{V}$. A *clique* in $G$ is a subset $C \subseteq \mathcal{V}$, such that $(i, j) \in \mathcal{E}$ for every $i, j \in C$ with $i \neq j$. In this paper, we consider the classical *maximum clique problem* (MCP): find a clique $C \subseteq \mathcal{V}$, such that $|C|$ is maximum.

The MCP has a wide range of applications (see Bomze et al. [5] and Wu and Hao [24] and references therein) in areas such as social network analysis, telecommunication networks, biochemistry, and scheduling. The cardinality of a maximum clique in $G$ is denoted $\omega(G)$. A clique $C$ is said to be *maximal* if it is not contained in any strictly larger clique, that is, if there does not exist a clique $D$, such that $C \subset D$. $C$ is said to be *strictly maximal* if there do not exist vertices $i \in C$ and $j \notin C$ such that $C \cup \{j\} \backslash \{i\}$ is a clique.

The MCP is NP hard (Karp [13]). However, owing in part to its wide applicability, a large variety of both heuristic and exact approaches has been investigated (see Bomze et al. [5] for a thorough overview of formulations and algorithms going up to 1999; a more recent survey of algorithms is given in Wu and Hao [24]). A significant number of the solution methods proposed (for example, Bomze [2], Bomze et al. [6], Gibbons et al. [10], Kuznetsova and Strekalovsky [15], Motzkin and Straus [16], Pelillo [21], and Pelillo and Jagota [22]) are based on solving the following well-known continuous quadratic programming formulation of the MCP due to Motzkin and Straus [16]:

$$\begin{array}{ll} \max & \mathbf{x}^{\mathsf{T}} \mathbf{A} \mathbf{x} \\ \text{subject to} & \mathbf{x} \in \Delta, \end{array} \tag{1}$$

where $\Delta$ is the $n$-dimensional simplex defined by

$$\Delta := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \geq \mathbf{0} \text{ and } \mathbf{1}^{\mathsf{T}} \mathbf{x} = 1\},$$

and $\mathbf{A} = (a_{ij})_{i,j \in \mathcal{V}}$ denotes the adjacency matrix for $G$ defined by

$$a_{ij} = \begin{cases} 1, & (i, j) \in \mathcal{E} \\ 0, & (i, j) \notin \mathcal{E} \end{cases} \quad \forall\, i, j \in \mathcal{V}.$$

For any nonempty set $S \subseteq \mathcal{V}$, we let $\mathbf{x}(S) \in \Delta$ denote the corresponding characteristic vector (defined by $x(S)_i = \frac{1}{|S|}$ whenever $i \in S$ and $x(S)_i = 0$ otherwise). The equivalence between the MCP and (1) is given by the following theorem.

**Theorem 1** (Theorem 1 in Motzkin and Straus [16]). *The optimal objective value of* (1) *is*

$$1-\frac{1}{\omega(G)},$$

*and* $\mathbf{x}(C)$ *is a global maximizer of* (1) *for any maximum clique* $C$.

Solution approaches to the MCP based on solving (1) include nonlinear programming methods (Gibbons et al. [9] and Pardalos and Phillips [19]) and methods based on discrete time replicator dynamics (Bomze [2], Bomze et al. [5, 6], and Pelillo [21]). Because (1) is NP hard (by reduction to the MCP), the computing time required to obtain a global maximizer can grow exponentially with the size of the graph; hence, finding a global maximizer may be impractical in many settings. However, iterative optimization methods will typically converge to a point satisfying the first-order optimality (Karush–Kuhn–Tucker) conditions. In general, verifying whether a first-order point of a quadratic program is even *locally* optimal is an NP-hard problem (Murty and Kabadi [17] and Pardalos and Schnitger [20]). However, it was shown in Gibbons et al. [10] that local optimality of a first-order point (in fact, any feasible point) in (1) can be ascertained in polynomial time.

In Pelillo and Jagota [22, proposition 3], a characteristic vector for a clique was shown to satisfy the standard first-order optimality condition for (1) if and only if the associated clique is maximal. In Gibbons et al. [10, theorem 2], the authors gave a characterization of the local optima of (1) and showed a one-one correspondence between strict local maximizers and strictly maximal cliques. These results suggest the possibility of applying iterative optimization methods to (1) to *approximately* solve the MCP (i.e., to find large maximal cliques). However, one known (Bomze [2], Pelillo [21], and Pelillo and Jagota [22]) drawback of this approach in practice is the presence of "infeasible" or "spurious" local maximizers of (1), which are not characteristic vectors for cliques and from which a clique cannot be recovered through any simple transformation. Such points are an undesirable property of the program, because they can cause continuous-based heuristics to fail by terminating without producing a clique. In Bomze [2], the author addresses this issue by introducing the following regularized formulation (with $\alpha=\frac{1}{2}$):

$$\begin{aligned} \max \quad & \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}+\alpha\|\mathbf{x}\|_2^2 \\ \text{subject to} \quad & \mathbf{x}\in\Delta. \end{aligned} \tag{2}$$

In contrast to (1), the local maximizers of (2) have been shown to be in one-one correspondence with the maximal cliques in $G$ (Bomze [2, theorem 9]), and a replicator dynamics approach to solving (2) was shown to reduce the total number of algorithm failures by 30% compared with a similar approach to solving (1). In Bomze et al. [6], the authors enhanced the algorithm of Bomze [2], adding an annealing heuristic to obtain even stronger results. In addition, it was shown that the correspondence between the local/global optima of (2) and the MCP is maintained for any $\alpha\in(0,1)$. A similar formulation and approach (Bomze et al. [4]) has also been applied successfully to a weighted version of the MCP. A generalization to hypergraphs was introduced in Rota-Bulò and Pelillo [23].

In practice, the numerical performance of an iterative optimization method (in terms of speed and/or solution quality) may depend on the particular regularization term used. In this paper, we consider a broad class of regularization terms and develop conditions under which the regularized program is equivalent to MCP in both a global sense and a local sense. We establish the equivalence in a step-by-step manner that reveals some of the underlying structural properties of (1). We provide two different examples of regularization terms satisfying the general conditions, and we also give some preliminary computational results evaluating their effectiveness in terms of both speed and solution quality. Over the course of our analysis (Section 2), we also correct an (apparently as yet unidentified) erroneous result in the literature linking maximal cliques to local maximizers in (1) by constructing an example of a maximal clique with a characteristic vector that is not a local maximizer of (1).

The paper is organized as follows. In Section 2, we develop a general regularized formulation of MCP and provide conditions under which the global/local maximizers of the regularized program are in one-one correspondence with the maximum/maximal cliques in $G$. In Section 3, we report on some preliminary computational results comparing the performance of two new regularization terms with the one proposed by Bomze in [2]. We conclude in Section 4.

In the notation, 0 and 1 denote column vectors with entries that are all zero and all one, respectively, and $\mathbf{I}$ denotes the identity matrix, where the dimensions should be clear from the context. The gradient of a function $f$, when it exists, is denoted by $\nabla f(\mathbf{x})$, a column vector, and $\nabla^2 f(\mathbf{x})$ denotes the Hessian of $f$. For a set $\mathscr{X}$, $|\mathscr{X}|$ is the number of elements in $\mathscr{X}$. The vector $\mathbf{e}_i \in \mathbb{R}^n$ denotes the $i$th column of the $n\times n$ identity matrix. If $\{s_i\}_{i=1}^n \subset \mathbb{R}$ is a finite sequence of length $n$, then $Diag(\{s_i\}_{i=1}^n)$ is the $n\times n$ diagonal matrix with an $(i,i)$th entry that is $s_i$. If $\mathbf{x}\in\mathbb{R}^n$, then $\operatorname{supp}(\mathbf{x})$ denotes the *support* of $\mathbf{x}$, defined by $\operatorname{supp}(\mathbf{x}) = \{i : x_i \neq 0\}$. Given vectors $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$, $[\mathbf{x},\mathbf{y}] := \{t\mathbf{x} + (1-t)\mathbf{y} : t\in[0,1]\}$. For a given positive integer $n$, we denote the set $\{1,2,\ldots,n\}$ by $[n]$. The inequality $\mathbf{x}\le\mathbf{y}$ denotes the componentwise inequalities $x_i \le y_i \ \forall\, i\in[n]$. $\mathscr{S}_n$ is the set of permutations of $[n]$. If $\mathbf{B}\in\mathbb{R}^{n\times n}$ is a symmetric matrix, we write $\mathbf{B}\succeq\mathbf{0}$ if $\mathbf{B}$ is positive semidefinite, $\mathbf{B}\succ\mathbf{0}$ if $\mathbf{B}$ is positive definite, and $\mathbf{B}\preceq\mathbf{0}$ ($\mathbf{B}\prec\mathbf{0}$) when $-\mathbf{B}\succeq\mathbf{0}$ ($-\mathbf{B}\succ\mathbf{0}$). $\|\mathbf{x}\|_2$ denotes the standard two-norm of $\mathbf{x}$, and $\|\mathbf{B}\|_2$ is the induced matrix two-norm of $\mathbf{B}$. If $\mathbf{X}\subseteq\mathbb{R}^n$ and $f:\mathbf{X}\to\mathbb{R}$, then a point $\mathbf{x}\in\mathbf{X}$ is a *local* (*strict local*) *maximizer* of the problem $\max\{f(\mathbf{x}) : \mathbf{x}\in\mathbf{X}\}$ if there exists some $\epsilon>0$, such that $f(\mathbf{x})\ge f(\tilde{\mathbf{x}})$ ($f(\mathbf{x}) > f(\tilde{\mathbf{x}})$) for every $\tilde{\mathbf{x}}\in\mathbf{X}$ with $0<\|\tilde{\mathbf{x}}-\mathbf{x}\|_2<\epsilon$. $\mathbf{x}$ is an *isolated* local maximizer if there exists some $\epsilon>0$, such that $\tilde{\mathbf{x}}$ is not a local maximizer for any $\tilde{\mathbf{x}}\in\mathbf{X}$ with $0<\|\tilde{\mathbf{x}}-\mathbf{x}\|_2<\epsilon$. The convex hull of $conv(\mathbf{X})$ denotes the convex hull of $\mathbf{X}$ and $span_+(\mathbf{X}) := \{\sum_{i=1}^k \alpha_i \mathbf{x}^i \ : k\in\mathbb{N}_+,\ \alpha_i\ge 0,\ \mathbf{x}^i\in\mathbf{X}\ \forall\, i\in[k]\}$. $\mathbb{N}_+$ denotes the set of positive natural numbers.

## 2. General Regularized Formulation

Consider the following problem:

$$\begin{aligned} \max \quad & f(\mathbf{x}) := \mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} + \Phi(\mathbf{x}) \\ & \text{subject to} \quad \mathbf{x}\in\Delta, \end{aligned} \tag{3}$$

where $\Phi: X\to\mathbb{R}$ is a twice continuously differentiable function defined on some open set $X\subset\Delta$. Throughout this section, we will also assume that $\Phi$ satisfies the following conditions for every $\mathbf{X}\supset\Delta$:

(C1) $\nabla^2\Phi(\mathbf{x}) \succeq \mathbf{0}$,
(C2) $\|\nabla^2\Phi(\mathbf{x})\|_2 < 2$, and
(C3) $\Phi$ is constant on the set

$$\mathscr{P}(\mathbf{x}) := \left\{\tilde{\mathbf{x}}\in\Delta \ : \exists\,\sigma\in\mathscr{S}_n \text{ such that } \tilde{x}_i = x_{\sigma(i)}\ \forall\, i\in[n]\right\},$$

where $\mathscr{S}_n$ is the set of permutations of $[n]$. Note that (C1) amounts to the condition that $\Phi$ be convex over $\Delta$, whereas (C2) is equivalent to $\nabla^2\Phi(\mathbf{x}) - 2\mathbf{I}\prec\mathbf{0}$ (a fact that will be used later). Also, because $\Phi\equiv 0$ satisfies (C1)–(C3) trivially, the results of this section will hold in particular for the original unpenalized formulation (1) when no additional assumptions are made on $\Phi$.

We will establish the global equivalence between (3) and the MCP through a series of intermediate results. For any nonempty set $S\subseteq\mathscr{V}$, let $\Delta(S)$ denote the face of $\Delta$ defined by

$$\Delta(S) := \left\{\mathbf{x}\in\Delta \ : \operatorname{supp}(\mathbf{x})\subseteq S\right\},$$

and consider the following (possibly nonconvex) finite union of faces associated with cliques of $G$:

$$\Delta^0 \ := \bigcup_{C \text{ clique}} \Delta(C) \ = \ \left\{\mathbf{x}\in\Delta \ : \operatorname{supp}(\mathbf{x}) \text{ is a clique}\right\}. \tag{4}$$

**Lemma 1.** *Let* $\mathbf{x}\in\Delta^0$.

1. *For any* $\tilde{\mathbf{x}}\in\mathscr{P}(\mathbf{x})\cap\Delta^0$, *we have*

$$f(\tilde{\mathbf{x}}) = f(\mathbf{x}).$$

2. *For any* $\mathbf{0}\neq\mathbf{d}\in\mathbb{R}^n$ *such that* $\mathbf{x}+t\mathbf{d}\in\Delta^0$ *for all sufficiently small* $t>0$, *we have*

$$\mathbf{d}^\mathsf{T}\nabla^2 f(\mathbf{x})\mathbf{d} < 0.$$

**Proof.** First, we note that, for any $\mathbf{z}\in\mathbb{R}^n$ such that $\operatorname{supp}(\mathbf{z})$ is a clique, we have

$$\begin{aligned} \mathbf{z}^\mathsf{T}\mathbf{A}\mathbf{z} = \sum_{i=1}^n\sum_{j=1}^n z_i a_{ij} z_j \ &= \sum_{i\in\operatorname{supp}(\mathbf{z})}\ \sum_{j\in\operatorname{supp}(\mathbf{z})\setminus\{i\}} z_i z_j \\ &= \sum_{i\in\operatorname{supp}(\mathbf{z})}\left(\sum_{j\in\operatorname{supp}(\mathbf{z})} z_i z_j - z_i^2\right) \\ &= (\mathbf{1}^\mathsf{T}\mathbf{z})^2 - \mathbf{z}^\mathsf{T}\mathbf{z}. \end{aligned} \tag{5}$$

Next, let $\mathbf{x}\in\Delta^0$. We prove the two parts separately.

***Part 1.*** For any $\tilde{\mathbf{x}} \in \mathcal{P}(\mathbf{x}) \cap \Delta^0$, we have

$$\begin{aligned} f(\tilde{\mathbf{x}}) &= \tilde{\mathbf{x}}^\mathsf{T}\mathbf{A}\tilde{\mathbf{x}} + \Phi(\tilde{\mathbf{x}}) \\ &= (\mathbf{1}^\mathsf{T}\tilde{\mathbf{x}})^2 - \tilde{\mathbf{x}}^\mathsf{T}\tilde{\mathbf{x}} + \Phi(\tilde{\mathbf{x}}) \quad (6) \\ &= (\mathbf{1}^\mathsf{T}\mathbf{x})^2 - \mathbf{x}^\mathsf{T}\mathbf{x} + \Phi(\mathbf{x}) \quad (7) \\ &= \mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} + \Phi(\mathbf{x}) \quad (8) \\ &= f(\mathbf{x}), \end{aligned}$$

where (6) and (8) are owing to (5), because $\tilde{\mathbf{x}}, \mathbf{x} \in \Delta^0$, and (7) is owing (C3) and the assumption that $\tilde{\mathbf{x}} \in \mathcal{P}(\mathbf{x})$.

***Part 2.*** Let $\mathbf{0} \neq \mathbf{d} \in \mathbb{R}^n$ be any vector such that $\mathbf{x} + t\mathbf{d} \in \Delta^0$ for all sufficiently small $t > 0$. Then, when $t$ is sufficiently small, $\operatorname{supp}(\mathbf{d}) \subseteq \operatorname{supp}(\mathbf{x} + t\mathbf{d})$, which implies that $\operatorname{supp}(\mathbf{d})$ is a clique; moreover, because $\mathbf{1}^\mathsf{T}\mathbf{x} = \mathbf{1}^\mathsf{T}(\mathbf{x} + t\mathbf{d}) = 1$, we have $\mathbf{1}^\mathsf{T}\mathbf{d} = 0$. Therefore, by (5), we have

$$\begin{aligned} \mathbf{d}^\mathsf{T}\nabla^2 f(\mathbf{x})\mathbf{d} &= 2\mathbf{d}^\mathsf{T}\mathbf{A}\mathbf{d} + \mathbf{d}\nabla^2\Phi(\mathbf{x})\mathbf{d} \quad (9) \\ &= 2[(\mathbf{1}^\mathsf{T}\mathbf{d})^2 - \mathbf{d}^\mathsf{T}\mathbf{d}] + \mathbf{d}^\mathsf{T}\nabla^2\Phi(\mathbf{x})\mathbf{d} \\ &= -2\mathbf{d}^\mathsf{T}\mathbf{d} + \mathbf{d}^\mathsf{T}\nabla^2\Phi(\mathbf{x})\mathbf{d} \\ &= -\mathbf{d}^\mathsf{T}[2\mathbf{I} - \nabla^2\Phi(\mathbf{x})]\mathbf{d} \quad (10) \\ &< 0, \end{aligned}$$

where (10) follows from (C2). □

Now consider the following problem, where $C \subseteq \mathcal{V}$ is any clique:

$$\begin{aligned} &\max \quad f(\mathbf{x}) \\ &\text{subject to} \quad \mathbf{x} \in \Delta(C). \end{aligned} \tag{11}$$

**Proposition 1.** *The unique local (hence, global) maximizer of* (11) *is* $\mathbf{x}(C)$.

**Proof.** Suppose by way of contradiction that there exist distinct local maximizers $\mathbf{x}^1 \neq \mathbf{x}^2$ of (11). Then, by Taylor's Theorem (see, e.g., proposition A.23 of Bertsekas [1]), $\exists\, \xi \in [\mathbf{x}^1, \mathbf{x}^2] \subseteq \Delta(\mathrm{C})$ such that

$$f(\mathbf{x}^2) = f(\mathbf{x}^1) + \nabla f(\mathbf{x}^1)^\mathsf{T}\mathbf{d} + \frac{1}{2}\mathbf{d}^\mathsf{T}\nabla^2 f(\xi)\mathbf{d}, \tag{12}$$

where $\mathbf{d} = \mathbf{x}^2 - \mathbf{x}^1$. Because $\mathbf{x}^1, \mathbf{x}^2 \in \Delta(C)$, we have that $\mathbf{x}^1 + t\mathbf{d} \in [\mathbf{x}^1, \mathbf{x}^2] \subseteq \Delta(C)$ for all sufficiently small $t > 0$. Therefore, by the standard first-order necessary local optimality condition (see, e.g., section 1 of Hager and Hungerford [11]), we have

$$\nabla f(\mathbf{x}^1)^\mathsf{T}\mathbf{d} \leq 0. \tag{13}$$

Moreover, Part 2 of Lemma 1 implies

$$\mathbf{d}^\mathsf{T}\nabla^2 f(\xi)\mathbf{d} < 0. \tag{14}$$

Combining (12)–(14), we obtain $f(\mathbf{x}^2) < f(\mathbf{x}^1)$. However, then interchanging $\mathbf{x}^2$ and $\mathbf{x}^1$, the same argument can be used to show that $f(\mathbf{x}^1) < f(\mathbf{x}^2)$, a contradiction. Therefore, there is a unique local (hence, global) maximizer of (11), say $\mathbf{x}^*$.

Next, we claim that $\mathcal{P}(\mathbf{x}^*) \cap \Delta(C) = \{\mathbf{x}^*\}$. Indeed, suppose by way of contradiction that $\exists\, \tilde{\mathbf{x}} \in \mathcal{P}(\mathbf{x}^*) \cap \Delta(C)$ such that $\tilde{\mathbf{x}} \neq \mathbf{x}^*$. Because $\operatorname{supp}(\mathbf{x}^*) \subseteq C$ and $C$ is a clique, Part 1 of Lemma 1 implies that $f(\tilde{\mathbf{x}}) = f(\mathbf{x}^*)$. However, then $\tilde{\mathbf{x}}$ must be a global maximizer of (11), contradicting the uniqueness of $\mathbf{x}^*$. Hence, we must have $\mathcal{P}(\mathbf{x}^*) \cap \Delta(C) = \{\mathbf{x}^*\}$. Thus, $x_i^* = x_j^*$ for any $i, j \in C$. Because $\mathbf{x}^* \in \Delta(C)$, this implies $\mathbf{x}^* = \mathbf{x}(C)$. □

**Remark 1.** By Part 2 of Lemma 1, (11) is a strictly concave (and smooth) maximization problem. Thus, the uniqueness of the maximizer of (11) may be seen as following from standard results in the theory of convex optimization (for instance, proposition B.4 in Bertsekas [1]).

Next, consider the problem

$$\begin{aligned} &\max \quad f(\mathbf{x}) \\ &\text{subject to} \quad \mathbf{x} \in \Delta^0. \end{aligned} \tag{15}$$

**Proposition 2.** *A point $\mathbf{x} \in \Delta^0$ is a local maximizer of (15) if and only if $\mathbf{x} = \mathbf{x}(C)$ for some maximal clique $C$. Moreover, every local maximizer of (15) is strict.*

**Proof.** First, observe that, for any local maximizer $\mathbf{x}$ of (15), by (4), there exists some maximal clique $C$ such that $\mathbf{x} \in \Delta(C)$; because $\mathbf{x}$ is a local maximizer of (15), it is also a local maximizer of (11), which implies that $\mathbf{x} = \mathbf{x}(C)$ by Proposition 1. Thus, the proof will be complete when we show that every characteristic vector for a maximal clique is a strict local maximizer in (15). To this end, let $C$ be a maximal clique, and suppose by way of contradiction that $\mathbf{x}(C)$ is not a strict local maximizer of (15). Then, for every $k \in \mathbb{N}_+$, there exists some $\mathbf{x}^k \in \Delta^0$ with $0 < \|\mathbf{x}^k - \mathbf{x}(C)\|_2 < 1/k$ such that $f(\mathbf{x}^k) \geq f(\mathbf{x}(C))$. Because there are only finitely many sets in the unions in (4), there must exist some clique $C'$ and some subsequence $(\mathbf{x}^{k_l})_{l=1}^{\infty} \subseteq (\mathbf{x}^k)_{k=1}^{\infty}$ such that $\mathbf{x}^{k_l} \in \Delta(C')$ for each $l \geq 1$, with $\mathbf{x}^{k_l} \to \mathbf{x}(C)$. Hence, $\mathbf{x}(C) \in \overline{\Delta(C')} = \Delta(C')$, which implies $C = \operatorname{supp}(\mathbf{x}(C)) \subseteq C'$. Because $C$ is maximal, we must have that $C = C'$, and thus, $\mathbf{x}^{k_l} \in \Delta(C') = \Delta(C)$ for each $l \geq 1$. Thus, $\mathbf{x}(C)$ is not a strict local maximizer of (11), contradicting Proposition 1. This completes the proof. □

**Proposition 3.** *If $C^1$ and $C^2$ are cliques, then*

$$|C^1| < |C^2| \Leftrightarrow f(\mathbf{x}(C^1)) < f(\mathbf{x}(C^2)).$$

**Proof.** Let $C^1$ and $C^2$ be cliques. Suppose that $|C^1| < |C^2|$. Let $C$ be any clique such that $C \subset C^2$ and $|C| = |C^1|$. Then, $\mathbf{x}(C^1) \in \mathcal{P}(\mathbf{x}(C))$. Therefore, Part 1 of Lemma 1 implies $f(\mathbf{x}(C^1)) = f(\mathbf{x}(C))$. Moreover, by Proposition 1, $f(\mathbf{x}(C)) < f(\mathbf{x}(C^2))$, because $\mathbf{x}(C) \in \Delta(C^2)$. Hence, $f(\mathbf{x}(C^1)) < f(\mathbf{x}(C^2))$. Conversely, suppose that $f(\mathbf{x}(C^1)) < f(\mathbf{x}(C^2))$. Then, by the proof of the forward direction, we must have $|C^1| \leq |C^2|$. Moreover, if $|C^1| = |C^2|$, then $\mathbf{x}(C^1) \in \mathcal{P}(\mathbf{x}(C^2))$, and Part 1 of Lemma 1 implies $f(\mathbf{x}(C^1)) = f(\mathbf{x}(C^2))$, a contradiction. Hence, we must have $|C^1| < |C^2|$. □

**Corollary 1.** *A point $\mathbf{x} \in \Delta^0$ is a global maximizer of (15) if and only if $\mathbf{x} = \mathbf{x}(C)$ for some maximum clique $C$.*

**Proof.** Let $\mathbf{x} \in \Delta^0$. Then, $\mathbf{x}$ is a global maximizer of (15) if and only if $\mathbf{x}$ is a local maximizer and $f(\mathbf{x}) \geq f(\bar{\mathbf{x}})$ for every local maximizer $\bar{\mathbf{x}} \neq \mathbf{x}$, which, by Proposition 2, holds if and only if $\mathbf{x} = \mathbf{x}(C)$ for some maximal clique $C$ and $f(\mathbf{x}(C)) \geq f(\mathbf{x}(\bar{C}))$ for every maximal clique $\bar{C} \neq C$. The corollary then follows from Proposition 3. □

**Proposition 4.** *For every clique $C$, $\mathbf{x}(C)$ is a global maximizer of (3) if and only if $C$ is a maximum clique.*

**Proof.** We will show that there exists a global maximizer of (3) that is in $\Delta^0$. The proof will then follow from Corollary 1. To this end, let $\mathbf{x}$ be any global maximizer of (3). If $\mathbf{x} \in \Delta^0$, then we are done. Therefore, suppose instead that $\mathbf{x} \notin \Delta^0$. Then, $\operatorname{supp}(\mathbf{x})$ is not a clique, and there exist indices $i \neq j \in \operatorname{supp}(\mathbf{x})$ such that $a_{ij} = 0$. Next, for any $t \in [-x_i, x_j]$, let $\mathbf{x}(t) := \mathbf{x} + t(\mathbf{e}_i - \mathbf{e}_j)$, and observe that, by Taylor's Theorem, there exists some $\xi \in [\mathbf{x}, \mathbf{x}(t)]$ such that

$$\begin{aligned} f(\mathbf{x}(t)) &= f(\mathbf{x}) + t\nabla f(\mathbf{x})^{\mathsf{T}}(\mathbf{e}_i - \mathbf{e}_j) + \frac{t^2}{2}(\mathbf{e}_i - \mathbf{e}_j)^{\mathsf{T}}\nabla^2 f(\xi)(\mathbf{e}_i - \mathbf{e}_j) \\ &= f(\mathbf{x}) + \frac{t^2}{2}\left[(2a_{ii} + 2a_{jj} - 4a_{ij}) + (\mathbf{e}_i - \mathbf{e}_j)^{\mathsf{T}}\nabla^2\Phi(\xi)(\mathbf{e}_i - \mathbf{e}_j)\right] && (16) \\ &= f(\mathbf{x}) + \frac{t^2}{2}\left[(\mathbf{e}_i - \mathbf{e}_j)^{\mathsf{T}}\nabla^2\Phi(\xi)(\mathbf{e}_i - \mathbf{e}_j)\right] && (17) \\ &\geq f(\mathbf{x}). && (18) \end{aligned}$$

Here, (16) follows from the first-order optimality condition at $\mathbf{x}$, which implies that $\nabla f(\mathbf{x})^{\mathsf{T}}(\mathbf{e}_i - \mathbf{e}_j) = 0$, because $\mathbf{x}(t)$ is feasible for all $t \in [-x_i, x_j]$. Equality (17) follows from the fact that $a_{ij} = 0 = a_{ii} = a_{jj}$. Additionally, (18) follows from (C1). Thus, setting $t = x_j$, we obtain another global maximizer $\mathbf{x}(t) \in \Delta$ such that $\operatorname{supp}(\mathbf{x}(t)) = \operatorname{supp}(\mathbf{x}) \setminus \{j\} \subset \operatorname{supp}(\mathbf{x})$. We may repeat this process, gradually reducing the size of $\operatorname{supp}(\mathbf{x})$ while maintaining global maximality, until $\operatorname{supp}(\mathbf{x})$ is a clique (possibly of size 1), at which point the proof is complete, because then $\mathbf{x} \in \Delta^0$. □

By Proposition 2, a one-one correspondence exists between the local maximizers of (15) and the maximal cliques in $G$. However, as is already well known in the case where $\Phi \equiv 0$ (see the discussions pertaining to infeasible or spurious local optima in Bomze [2], Pelillo [21], and Pelillo and Jagota [22]), if $\Delta^0$ in (15) is relaxed to $\Delta \supseteq \Delta^0$ (in fact, $\Delta = conv(\Delta^0)$, but we need not prove this here), there may exist local maximizers of (3) that are not characteristic vectors for cliques, which may cause iterative optimization methods for solving (3) to fail, terminating without producing a clique.

Mathematics of Operations Research, 2019, vol. 44, no. 4, pp. 1161–1173, © 2019 INFORMS



Conversely, when $\Phi \equiv 0$, there may exist characteristic vectors for maximal cliques that are not local maximizers in (3). Indeed, in the graph $G$ in Figure 1, the sets $C = \{1, 2\}$ and $\hat{C} = \{3, 4, 5\}$ are both maximal cliques; because $\mathbf{x}(C)$, $\mathbf{x}(\hat{C}) \in \Delta$, we have that, for all sufficiently small $t > 0$, $\mathbf{x}(C) + t\mathbf{d} \in \Delta$, where $\mathbf{d} = \mathbf{x}(\hat{C}) - \mathbf{x}(C)$. Moreover, it is easy to check that, by computation, one has that $\nabla f(\mathbf{x})^\mathsf{T}\mathbf{d} = 0$ and $\mathbf{d}^\mathsf{T}\nabla^2 f(\mathbf{x})\mathbf{d} = \frac{1}{6} > 0$, where $\mathbf{x} = \mathbf{x}(C)$. Hence, for all sufficiently small $t > 0$,

$$f(\mathbf{x} + t\mathbf{d}) = f(\mathbf{x}) + t\nabla f(\mathbf{x})^\mathsf{T}\mathbf{d} + \frac{t^2}{2}\mathbf{d}^\mathsf{T}\nabla^2 f(\mathbf{x})\mathbf{d} = f(\mathbf{x}) + \frac{t^2}{2}\mathbf{d}^\mathsf{T}\nabla^2 f(\mathbf{x})\mathbf{d} > f(\mathbf{x}).$$

Thus, $\mathbf{x}(C)$ is not a local maximizer of (3), despite the fact that $C$ is maximal. We note that the above is a counterexample to Gibbons et al. [10, corollary 2].

Hence, there is not necessarily any relationship (in either direction) between the local optima of (1) and the maximal cliques in $G$. However, we will see in the next proposition that the local maximizers of (3) are in *one-one* correspondence with the characteristic vectors for maximal cliques whenever $\Phi$ satisfies (C1) strictly. The proposition is based on three lemmas. The first lemma below is a special case of the result (Hager and Hungerford [11, corollary 2.2]).

**Lemma 2.** *Let $\mathbf{x} \in \Delta$, and let $\mathcal{F}(\mathbf{x})$ denote the set of first-order feasible directions for (3) at $\mathbf{x}$ defined by*

$$\mathcal{F}(\mathbf{x}) = \left\{\mathbf{d} \in \mathbb{R}^n : \mathbf{1}^\mathsf{T}\mathbf{d} = 0 \text{ and } d_i \geq 0 \text{ whenever } x_i = 0\right\}.$$

*Then,*

$$\mathcal{F}(\mathbf{x}) = span_+(\mathcal{F}(\mathbf{x}) \cap \mathcal{D}),$$

*where* $\mathcal{D} = \bigcup_{\substack{i,j=1 \\ i \neq j}}^{n} \{\mathbf{e}_i - \mathbf{e}_j\}$.

**Proof.** Because $\mathcal{F}(\mathbf{x})$ is a cone, the containment $\mathcal{F}(\mathbf{x}) \supseteq span_+(\mathcal{F}(\mathbf{x}) \cap \mathcal{D})$ is trivial. To see the opposite containment, let $\mathbf{d} \in \mathcal{F}(\mathbf{x})$. Then, $\mathbf{1}^\mathsf{T}\mathbf{d} = 0$ and $d_i \geq 0$ for every $i \notin \operatorname{supp}(\mathbf{x})$. Because $\operatorname{supp}(\mathbf{x}) \neq \emptyset$, we may let $j \in \operatorname{supp}(\mathbf{x})$ be arbitrary to obtain the following:

$$\begin{aligned}\mathbf{d} &= \sum_{i=1}^{n} d_i\mathbf{e}_i \\ &= \sum_{i=1}^{n} d_i(\mathbf{e}_i - \mathbf{e}_j) + \sum_{i=1}^{n} d_i\mathbf{e}_j \\ &= \sum_{i \notin \operatorname{supp}(\mathbf{x})} d_i(\mathbf{e}_i - \mathbf{e}_j) + \sum_{i \in \operatorname{supp}(\mathbf{x})} d_i^+(\mathbf{e}_i - \mathbf{e}_j) + \sum_{i \in \operatorname{supp}(\mathbf{x})} d_i^-(\mathbf{e}_j - \mathbf{e}_i) \in span_+(\mathcal{F}(\mathbf{x}) \cap \mathcal{D}),\end{aligned}$$

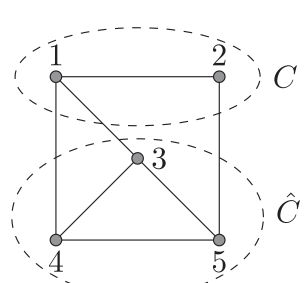
where $d_i^\pm = \max\{0, \pm d_i\}$. □

The next lemma is a restatement in the language of this paper of the result of Pelillo and Jagota [22, proposition 3], which states that a characteristic vector for a maximal clique satisfies the first-order optimality conditions of (1).

**Lemma 3.** *If $C$ is a maximal clique, then*

$$\mathbf{x}(C)^\mathsf{T}\mathbf{A}\mathbf{d} \leq 0 \quad \forall\, \mathbf{d} \in \mathcal{F}(\mathbf{x}(C)).$$

**Figure 1.** An example of a graph $G$ and two maximal cliques $C$ and $\hat{C}$. Here, $\mathbf{x}(C)$ is not a local maximizer of (1), although $C$ is maximal.



**Proof.** Let $C$ be a maximal clique. We claim that, for any $\mathbf{d}^s \in \mathcal{F}(\mathbf{x}(C)) \cap \mathcal{D}$, we have $\mathbf{x}(C)^\mathsf{T}\mathbf{A}\mathbf{d}^s \le 0$. After this is shown, the proof will be complete, because by Lemma 2, $\forall\, \mathbf{d} \in \mathcal{F}(\mathbf{x}(C))\, \exists\, k \in \mathbb{N}_+$, $\alpha_1, \alpha_2, \dots, \alpha_k \ge 0$, $\mathbf{d}^1, \mathbf{d}^2, \dots, \mathbf{d}^k \in \mathcal{F}(\mathbf{x}(C)) \cap \mathcal{D}$ such that $\mathbf{d} = \sum_{s=1}^k \alpha^s \mathbf{d}^s$, and therefore,

$$\mathbf{x}(C)^\mathsf{T}\mathbf{A}\mathbf{d} = \sum_{s\in[k]} \alpha^s \mathbf{x}(C)^\mathsf{T}\mathbf{A}\mathbf{d}^s \le 0\,.$$

Therefore, suppose that $\mathbf{d}^s \in \mathcal{F}(\mathbf{x}(C)) \cap \mathcal{D}$. Then, $\mathbf{d}^s = (\mathbf{e}_i - \mathbf{e}_j)$ for some $i \ne j$. Moreover, by definition of $\mathcal{F}(\mathbf{x}(C))$, we have $x_j(C) > 0$; that is, $j \in C$. Hence,

$$\begin{aligned}\mathbf{x}(C)^\mathsf{T}\mathbf{A}\mathbf{d}^s &= \sum_{t=1}^n \sum_{r=1}^n x_t(C) a_{tr} d_r^s \\ &= \frac{1}{|C|}\sum_{t\in C}(a_{ti} - a_{tj}) \\ &= \frac{1}{|C|}\left[\sum_{t\in C\setminus\{j\}}(a_{ti} - 1) + a_{ji}\right]. \end{aligned} \tag{19}$$

Here, (19) follows from the fact that $C$ is a clique, $j \in C$, and $a_{jj} = 0$. We now consider two cases.

*Case* 1. $i \in C$. In this case, because $a_{ii} = 0$, the right-hand side of (19) equals

$$\frac{1}{|C|}\left[\sum_{t\in C\setminus\{i,j\}}(a_{ti} - 1) + a_{ji} - 1\right]. \tag{20}$$

However, because $i \in C$ and $C$ is a clique, we have that $a_{ti} = 1$ for every $t \in C\setminus\{i,j\}$. Hence, (20) is equal to

$$\frac{1}{|C|}(a_{ji} - 1) \le 0\,.$$

Thus, $\mathbf{x}(C)^\mathsf{T}\mathbf{A}\mathbf{d}^s \le 0$.

*Case* 2. $i \notin C$. In this case, we have that (19) is equal to

$$\frac{1}{|C|}\left[\sum_{t\in C\setminus\{i,j\}}(a_{ti} - 1) + a_{ji}\right]. \tag{21}$$

However, because $C$ is maximal and $i \notin C$, there must exist some $k \in C$ such that $a_{ik} = 0$. If $k = j$, then (21) is less than or equal to zero, because $a_{ji} = 0$, and each of the terms in the first summation is less than or equal to zero. However, if $k \ne j$, then there exists a term in the first summation of (21) that is equal to $-1$, and hence, because $a_{ji} \le 1$, (21) is less than or equal to zero. Thus, $\mathbf{x}(C)^\mathsf{T}\mathbf{A}\mathbf{d}^s \le 0$. This completes the proof. □

**Lemma 4.** *Let $S \subseteq \mathcal{V}$ be nonempty. Then,*

$$\nabla\Phi(\mathbf{x}(S))^\mathsf{T}\mathbf{d} \le 0 \quad \forall\, \mathbf{d} \in \mathcal{F}(\mathbf{x}(S))\,. \tag{22}$$

*Moreover, if $\nabla^2\Phi(\tilde{\mathbf{x}}) > \mathbf{0}\ \forall\, \tilde{\mathbf{x}} \in \Delta$, then the inequality in (22) is strict whenever* $\operatorname{supp}(\mathbf{d}) \not\subseteq S$.

**Proof.** First, observe that, if (22) holds in the case where $\nabla^2\Phi(\tilde{\mathbf{x}}) > \mathbf{0}\ \forall\, \tilde{\mathbf{x}} \in \Delta$, then it also holds for any $\Phi$ (satisfying (C1)–(C3)). The argument is as follows. If $\Phi$ satisfies (C1)–(C3), then for all sufficiently large $k \in \mathbb{N}$, the regularization function $\Phi^{(k)} := \Phi + \frac{1}{k}\|\cdot\|_2^2$ also satisfies (C1)–(C3), and moreover, for any $\tilde{\mathbf{x}} \in \Delta$, we have that

$$\nabla^2\Phi^{(k)}(\tilde{\mathbf{x}}) = \nabla^2\Phi(\tilde{\mathbf{x}}) + \frac{2}{k}\mathbf{I} > \mathbf{0}\,.$$

Hence,

$$\nabla\Phi(\mathbf{x}(S))^\mathsf{T}\mathbf{d} + \frac{2}{k}\mathbf{x}(S)^\mathsf{T}\mathbf{d} = \nabla\Phi^{(k)}(\mathbf{x}(S))^\mathsf{T}\mathbf{d} \le 0 \quad \forall\, \mathbf{d} \in \mathcal{F}(\mathbf{x}(S))\,. \tag{23}$$

Then, taking the limit of (23) as $k \to \infty$, we obtain (22).

Therefore, suppose that $\nabla^2\Phi(\tilde{\mathbf{x}}) > \mathbf{0}$ for every $\tilde{\mathbf{x}} \in \Delta$. We first prove the following result for any $\mathbf{x} \in \Delta$:

$$\nabla\Phi(\mathbf{x})^\mathsf{T}(\mathbf{e}_i - \mathbf{e}_j) < 0 \quad \forall\, i, j \in [n] \text{ such that } x_i < x_j\,. \tag{24}$$

To see this, let $\mathbf{x} \in \Delta$, suppose that $i, j \in [n]$ are such that $x_i < x_j$, and let $t := x_j - x_i > 0$ and $\bar{\mathbf{x}} := \mathbf{x} + t(\mathbf{e}_i - \mathbf{e}_j)$. Then, $\bar{\mathbf{x}} \in \mathcal{P}(\mathbf{x}) \subseteq \Delta$, because $\bar{x}_i = x_j$, $\bar{x}_j = x_i$, and $\bar{x}_k = x_k\ \forall\, k \neq i, j$. Therefore, by (C3), we have $\Phi(\bar{\mathbf{x}}) = \Phi(\mathbf{x})$. However, taking a Taylor expansion of $\Phi$ about $\mathbf{x}$, we have that, for some $\boldsymbol{\xi} \in [\mathbf{x}, \bar{\mathbf{x}}] \subseteq \Delta$,

$$\Phi(\mathbf{x}) = \Phi(\bar{\mathbf{x}}) = \Phi(\mathbf{x}) + t\nabla\Phi(\mathbf{x})^\mathsf{T}(\mathbf{e}_i - \mathbf{e}_j) + \frac{t^2}{2}(\mathbf{e}_i - \mathbf{e}_j)^\mathsf{T}\nabla^2\Phi(\boldsymbol{\xi})(\mathbf{e}_i - \mathbf{e}_j)\,,$$

and hence,

$$0 = t\nabla\Phi(\mathbf{x})^\mathsf{T}(\mathbf{e}_i - \mathbf{e}_j) + \frac{t^2}{2}(\mathbf{e}_i - \mathbf{e}_j)^\mathsf{T}\nabla^2\Phi(\boldsymbol{\xi})(\mathbf{e}_i - \mathbf{e}_j)\,.$$

Therefore, because $\nabla^2\Phi(\boldsymbol{\xi}) \succ \mathbf{0}$ (by assumption) and $t > 0$, we have that

$$\nabla\Phi(\mathbf{x})^\mathsf{T}(\mathbf{e}_i - \mathbf{e}_j) < 0\,.$$

Thus, (24) is proved. Moreover, note that, by continuity of $\nabla\Phi(\,\cdot\,)^\mathsf{T}(\mathbf{e}_i - \mathbf{e}_j)$, we have in addition that

$$\nabla\Phi(\mathbf{x})^\mathsf{T}(\mathbf{e}_i - \mathbf{e}_j) \leq 0 \quad \forall\, i, j \in [n] \text{ such that } x_i = x_j > 0\,. \tag{25}$$

In fact, by symmetry, it is easy to see that the inequality in (25) must actually be an equality.

Now, we prove the lemma. The case where $\mathbf{d} = \mathbf{0}$ is trivial. Therefore, let $S \subseteq \mathcal{V}$ be nonempty, and let $\mathbf{0} \neq \mathbf{d} \in \mathcal{F}(\mathbf{x}(S))$. Then, by Lemma 2, there exist $k \in \mathbb{N}_+$, $\mathbf{d}^1, \ldots, \mathbf{d}^k \in \mathcal{F}(\mathbf{x}(S)) \cap \mathcal{D}$, and $\alpha^1, \alpha^2, \ldots, \alpha^k > 0$ such that $\mathbf{d} = \sum_{s=1}^k \alpha^s \mathbf{d}^s$. Hence,

$$\nabla\Phi(\mathbf{x}(S))^\mathsf{T}\mathbf{d} = \sum_{s=1}^k \alpha^s \nabla\Phi(\mathbf{x}(S))^\mathsf{T}\mathbf{d}^s\,. \tag{26}$$

Next, note that, by definition of $\mathcal{D}$ for each $s \in [k]$, there exist $i, j \in [n]$ such that $\mathbf{d}^s = (\mathbf{e}_i - \mathbf{e}_j) \in \mathcal{F}(\mathbf{x}(S))$. Hence, $j \in S$, and $x_i(S) \leq \frac{1}{|S|} = x_j(S)$. Hence, by (24) and (25), we have that $\nabla\Phi(\mathbf{x}(S))^\mathsf{T}\mathbf{d}^s \leq 0$ for each $s \in [k]$. Therefore, by (26), we have $\nabla\Phi(\mathbf{x}(S))^\mathsf{T}\mathbf{d} = \sum_{s=1}^k \alpha^s \nabla\Phi(\mathbf{x}(S))^\mathsf{T}\mathbf{d}^s \leq 0$. Moreover, in the case where $\operatorname{supp}(\mathbf{d}) \not\subseteq S$, there must exist some $s \in [k]$ and some $i, j \in [n]$ such that $\mathbf{d}^s = (\mathbf{e}_i - \mathbf{e}_j)$ with $i \notin S$. By (24), this implies $\nabla\Phi(\mathbf{x}(S))^\mathsf{T}\mathbf{d}^s < 0$. Hence, $\nabla\Phi(\mathbf{x}(S))^\mathsf{T}\mathbf{d} < 0$. This completes the proof. □

In the proof of the next proposition, we will use the following standard second-order sufficient optimality condition (see, for instance, Nocedal and Wright [18, theorem 12.6]). A point $\mathbf{x} \in \Delta$ is a strict local maximizer of (3) if

$$\nabla f(\mathbf{x})^\mathsf{T}\mathbf{d} \leq 0 \quad \forall\, \mathbf{d} \in \mathcal{F}(\mathbf{x}) \tag{27}$$

and

$$\mathbf{d}^\mathsf{T}\nabla^2 f(\mathbf{x})\mathbf{d} < 0 \quad \forall\, \mathbf{0} \neq \mathbf{d} \in \mathcal{C}(\mathbf{x})\,, \tag{28}$$

where $\mathcal{C}(\mathbf{x})$ is the critical cone at $\mathbf{x}$ defined by

$$\mathcal{C}(\mathbf{x}) := \left\{\mathbf{d} \in \mathcal{F}(\mathbf{x}) \;:\; \nabla f(\mathbf{x})^\mathsf{T}\mathbf{d} = 0\right\}.$$

For a stronger and more general version of this condition stated in the language of copositivity, see Bomze [3, theorem 1.1].

**Proposition 5.** *Suppose that $\nabla^2\Phi(\tilde{\mathbf{x}}) > \mathbf{0}$ for every $\tilde{\mathbf{x}} \in \Delta$. Then, a point $\mathbf{x} \in \Delta$ is a local maximizer of (3) if and only if $\mathbf{x} = \mathbf{x}(C)$ for some maximal clique $C$. Moreover, every local maximizer of (3) is strict.*

**Proof.** Suppose that $\nabla^2\Phi(\tilde{\mathbf{x}}) > \mathbf{0}$ for every $\tilde{\mathbf{x}} \in \Delta$. We claim that every local maximizer of (3) is in $\Delta^0$. To this end, let $\mathbf{x}$ be any local maximizer of (3). If $\operatorname{supp}(\mathbf{x})$ is a clique, then clearly $\mathbf{x} \in \Delta^0$. Therefore, suppose by way of contradiction that $\operatorname{supp}(\mathbf{x})$ is not a clique. By applying an argument similar to the one given in the proof of Proposition 4, there exist indices $i \neq j \in \operatorname{supp}(\mathbf{x})$ such that, for all $0 \neq t \in [-x_i, x_j]$, we have $\mathbf{x}(t) = \mathbf{x} + t(\mathbf{e}_i - \mathbf{e}_j) \in \Delta$ and $f(\mathbf{x}(t)) > f(\mathbf{x})$, where the strict inequality here follows from (17) and the fact that $\nabla^2\Phi(\boldsymbol{\xi}) > \mathbf{0}$ for any $\boldsymbol{\xi} \in [\mathbf{x}, \mathbf{x}(t)] \subseteq \Delta$. However, this contradicts the fact that $\mathbf{x}$ is a local maximizer of (3). Hence, $\operatorname{supp}(\mathbf{x})$ is a clique, and $\mathbf{x} \in \Delta^0$.

Therefore, every local maximizer of (3) is in $\Delta^0$ and is therefore a local maximizer of (15). Thus, by Proposition 2, every local maximizer of (3) is equal to $\mathbf{x}(C)$ for some maximal clique $C$.

We will be done when we show that $\mathbf{x}(C)$ is a strict local maximizer whenever $C$ is a maximal clique. Therefore, suppose that $C$ is a maximal clique, and let $\mathbf{x} := \mathbf{x}(C)$. We will prove that the first-order and second-order conditions (27) and (28) hold. Let $\mathbf{d}^1 \in \mathcal{F}(\mathbf{x})$. Then, by Lemma 3, we have $\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{d}^1 \le 0$, and by Lemma 4, $\nabla\Phi(\mathbf{x})^\mathsf{T}\mathbf{d}^1 \le 0$. Hence,

$$\nabla f(\mathbf{x})^\mathsf{T}\mathbf{d} = 2\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{d}^1 + \nabla\Phi(\mathbf{x})^\mathsf{T}\mathbf{d}^1 \le 0\,, \tag{29}$$

proving (27). To see that the second-order condition (28) holds, let $\mathbf{0} \ne \mathbf{d}^2 \in \mathcal{C}(\mathbf{x})$ be arbitrary. Then,

$$0 = \nabla f(\mathbf{x})^\mathsf{T}\mathbf{d}^2 = 2\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{d}^2 + \nabla\Phi(\mathbf{x})^\mathsf{T}\mathbf{d}^2 \le \nabla\Phi(\mathbf{x})^\mathsf{T}\mathbf{d}^2\,, \tag{30}$$

where the last inequality follows from Lemma 3, because $\mathbf{d}^2 \in \mathcal{F}(\mathbf{x})$. Thus, by the second statement in Lemma 4, we must have $\operatorname{supp}(\mathbf{d}^2) \subseteq C$, implying that $\mathbf{x} + t\mathbf{d}^2 \in \Delta(C) \subseteq \Delta^0$ for all sufficiently small $t > 0$. Therefore, by Part 2 of Lemma 1, we have

$${\mathbf{d}^2}^\mathsf{T}\nabla^2 f(\mathbf{x})\mathbf{d}^2 < 0\,,$$

proving (28). Thus, $\mathbf{x}(C)$ is a strict local maximizer, and the proof is complete. □

## 3. Preliminary Numerical Results

In this section, we conduct some preliminary numerical experiments on three different regularization functions satisfying the conditions outlined in Section 2 to give an indication of the potential impact of different regularization terms on the performance of a local optimization algorithm applied to (3). If, in practice, a maximum clique (rather than merely a maximal clique) is sought, the local optimization algorithm that we use in our experiments would need to be incorporated into a global optimization framework, such as branch and bound, to ensure convergence to a global maximizer.

### 3.1. Regularization Functions

We considered the following three regularization terms with the indicated choices of parameters:

$$\Phi_B(\mathbf{x}) := \frac{1}{2}\|\mathbf{x}\|_2^2, \tag{31}$$

$$\Phi_1(\mathbf{x}) := \alpha_1\|\mathbf{x} + \epsilon\mathbf{1}\|_p^p, \text{ with } \epsilon > 0,\ p > 2, \text{ and } 0 < \alpha_1 < \tfrac{2}{p(p-1)(1+\epsilon)^{p-2}}, \tag{32}$$

$$\Phi_2(\mathbf{x}) := \alpha_2\sum_{i=1}^{n}(e^{-\beta x_i} - 1), \text{ with } \beta > 0 \text{ and } 0 < \alpha_2 < \tfrac{2}{\beta^2}. \tag{33}$$

Here, $\Phi_B$ is the two-norm regularization function introduced by Bomze [2], and $\Phi_1$ is a generalization of $\Phi_B$ to p-norms, where $p > 2$. $\Phi_2$ is a well-known (for instance, see Bradley et al. [7]) approximation of the following nonsmooth function:

$$\tilde{\Phi}(\mathbf{x}) = -\alpha_2\|\mathbf{x}\|_0\,, \tag{34}$$

where $\|\mathbf{x}\|_0 = |\operatorname{supp}(\mathbf{x})|$. The motivation behind the choice of $\Phi_2$ is as follows. By definition of $\Phi_2$, maximizing $\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} + \Phi_2$ over $\Delta$ is closely related to the problem of finding the solution to (1) that has the smallest support (i.e., the maximum sparsity). Following the argument laid out in the Proof of Proposition 4, from any global maximizer of (1) that is not a characteristic vector for a maximum clique, there exists a path leading to another global maximizer that is a characteristic vector for a maximum clique and with support that is strictly smaller than that of the starting point. Hence, the global maximizers of (1) that have the smallest support are necessarily the characteristic vectors for maximum cliques. Thus, $\Phi_2$ is a somewhat natural choice in our context.

Next, we show that each of the regularization functions above satisfies the conditions of Section 2. For any $\delta \in (0, \epsilon)$, where $\epsilon > 0$ is the value used in the definition of $\Phi_1$, let $X := conv(\cup_{i=1}^n \mathcal{B}_\delta(\mathbf{e}_i)) \subseteq \mathbb{R}^n$, where $\mathcal{B}_\delta(\mathbf{e}_i) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{e}_i\|_2 < \delta\}$. Then, $X$ is open; $X \supset \Delta$; and $\Phi_B$, $\Phi_1$, and $\Phi_2$ are each well defined on $X$.

Moreover, it is easy to check that $\Phi_B$, $\Phi_1$, and $\Phi_2$ are each twice continuously differentiable over $X$ and that, for any $\mathbf{x} \in \Delta$, we have

$$\begin{aligned}
&\nabla^2 \Phi_B(\mathbf{x}) = \mathbf{I} > \mathbf{0}\,,\\
&\nabla^2 \Phi_1(\mathbf{x}) = \alpha_1 p(p-1) Diag(\{(x_i+\epsilon)^{p-2}\}_{i=1}^n) > \mathbf{0}\,, \text{ and}\\
&\nabla^2 \Phi_2(\mathbf{x}) = \alpha_2 \beta^2 Diag(\{e^{-\beta x_i}\}_{i=1}^n) > \mathbf{0}\,,
\end{aligned} \tag{35}$$

where the positive definiteness of the Hessians in (35) follows from the choice of parameters. Hence, $\Phi_B$, $\Phi_1$, and $\Phi_2$ each satisfy (C1) strictly. Next, observe that, for any $\mathbf{x} \in \Delta$, we have

$$\begin{aligned}
\|\nabla^2 \Phi_B(\mathbf{x})\|_2 &= 1 < 2\,,\\
\|\nabla^2 \Phi_1(\mathbf{x})\|_2 &= \alpha_1 p(p-1) \max\{(x_i+\epsilon)^{p-2}\}_{i=1}^n\\
&< \frac{2}{(1+\epsilon)^{p-2}} \max\{(x_i+\epsilon)^{p-2}\}_{i=1}^n\\
&\le \frac{2}{(1+\epsilon)^{p-2}} (1+\epsilon)^{p-2} = 2\,, \text{ and}\\
\|\nabla^2 \Phi_2(\mathbf{x})\|_2 &= \alpha_2 \beta^2 \max\{e^{-\beta x_i}\}_{i=1}^n\\
&< 2 \max\{e^{-\beta x_i}\}_{i=1}^n\\
&\le 2\,.
\end{aligned}$$

Thus, (C2) is satisfied for each of $\Phi_B$, $\Phi_1$, and $\Phi_2$. That (C3) holds follows easily from the fact that $\Phi_B$, $\Phi_1$, and $\Phi_2$ are each separable, and the coefficients associated with the terms $x_i$ are independent of $i$.

### 3.2. The Testing Set

In the experiments, we considered different families of widely used maximum clique instances belonging to the DIMACS benchmark (Johnson and Trick [12]):

- C family: Random graphs Cx.y, where x is the number of nodes and y is the edge probability.
- DSJC family: Random graphs DSCJx_y. Here again, x is the number of nodes, and y is the edge probability.
- brock family: Random graphs with cliques hidden among nodes that have a relatively low degree.
- gen family: Artificially generated graphs with large known embedded clique.
- hamming family: The hamminga-b are graphs on a-bit words with an edge if and only if the two words are at least hamming distance b apart.
- keller family: Instances based on the conjecture of Keller [14] on tilings using hypercubes.
- p_hat family: Random graphs generated with the p-hat generator, which is a generalization of the classical uniform random graph generator. Graphs generated with p-hat have wider node degree spread and larger cliques than uniform graphs.

In Table 1, we report the names of the instances used (Instance), the best known solutions (Best known), the numbers of nodes and edges in the instances (Nodes and Edges), and the median and interquartile range (IQR) related to the graph degrees (Median and IQR in column Graph degrees) as well as the median and IQR of the degrees of the nodes lying in the best known solution (Median and IQR in column Best degrees).

### 3.3. Experiments

To conduct our tests, we developed a multistart framework in MATLAB that uses a hybrid algorithm as a local optimizer. It combines the fmincon solver with the Frank–Wolfe method (Frank and Wolfe [8]). For each instance, we ran 100 trials each with a different randomly generated point in $\Delta$ as a starting guess. The same starting guesses were used for all formulations. Because the iterates of the Frank–Wolfe method that we used are only guaranteed to converge to a point satisfying the first-order conditions, we omitted the trials in which the final iterate was not a true local optimizer from the statistical computations below. (Another potential way of dealing with this issue, which we leave to a future work, would be to take a step in an ascent direction whenever the final iterate is not a local maximizer and then rerun the algorithm using the new point as a starting guess.) After a local maximizer $\mathbf{x}^*$ is obtained, the associated clique is constructed by taking $C^* := \operatorname{supp}(\mathbf{x}^*)$. In our experiments, we used the parameter values $p = 3$, $\epsilon = 10^{-9}$, and $\beta = 5$. Furthermore, $\alpha_1$ and $\alpha_2$ were suitably chosen to satisfy condition (32) and (33), respectively. All of the tests were performed on an Intel Core i7-3610QM 2.3-GHz, 8-GB RAM.

**Table 1.** DIMACS instances used in the tests.

| | | | | Graph degrees | | Best degrees | |
|---|---|---|---|---|---|---|---|
| Instance | Best known | Nodes | Edges | Median | IQR | Median | IQR |
| C125.9 | 34 | 125 | 6,963 | 112 | 5 | 114.5 | 4.75 |
| C250.9 | 44 | 250 | 27,984 | 224 | 6 | 227 | 5 |
| C500.9 | 57 | 500 | 112,332 | 449 | 9 | 455 | 9 |
| C1000.9 | 68 | 1,000 | 450,079 | 900 | 13 | 907 | 11.25 |
| C2000.5 | 16 | 2,000 | 999,836 | 999 | 30 | 1,006 | 11.5 |
| C2000.9 | 80 | 2,000 | 1,799,532 | 1,800 | 18 | 1,803 | 15.25 |
| DSJC500_5 | 13 | 500 | 125,248 | 250 | 16 | 259 | 14 |
| DSJC1000_5 | 15 | 1,000 | 499,652 | 500 | 20 | 503 | 23 |
| brock200_2 | 12 | 200 | 9,876 | 99 | 10 | 101 | 11 |
| brock200_4 | 17 | 200 | 13,089 | 131 | 8 | 134 | 6 |
| brock400_2 | 29 | 400 | 59,786 | 299 | 10 | 299 | 9 |
| brock400_4 | 33 | 400 | 59,765 | 299 | 11 | 299 | 9 |
| brock800_2 | 24 | 800 | 208,166 | 521 | 18 | 516.5 | 20.25 |
| brock800_4 | 26 | 800 | 207,643 | 519 | 18.25 | 512 | 20.25 |
| gen200_p0.9_44 | 44 | 200 | 17,910 | 180 | 8 | 179.5 | 4.25 |
| gen200_p0.9_55 | 55 | 200 | 17,910 | 179 | 7.25 | 179 | 5.5 |
| gen400_p0.9_55 | 55 | 400 | 71,820 | 360 | 13.25 | 359 | 6 |
| gen400_p0.9_65 | 65 | 400 | 71,820 | 361 | 14 | 359 | 9 |
| gen400_p0.9_75 | 75 | 400 | 71,820 | 359 | 13 | 359 | 8 |
| hamming8-4 | 16 | 256 | 20,864 | 163 | 0 | 163 | 0 |
| hamming10-4 | 40 | 1,024 | 434,176 | 848 | 0 | 848 | 0 |
| keller4 | 11 | 171 | 9,435 | 110 | 8 | 112 | 17 |
| keller5 | 27 | 776 | 225,990 | 578 | 38 | 578 | 33 |
| keller6 | 59 | 3,361 | 4,619,898 | 2,724 | 50 | 2,724 | 50 |
| p_hat300-1 | 8 | 300 | 10,933 | 73 | 39 | 103 | 20 |
| p_hat300-2 | 25 | 300 | 21,928 | 146.5 | 73 | 213 | 18 |
| p_hat300-3 | 36 | 300 | 33,390 | 224 | 38 | 251 | 15.25 |
| p_hat700-1 | 11 | 700 | 60,999 | 174.5 | 87 | 250 | 22.5 |
| p_hat700-2 | 44 | 700 | 121,728 | 353 | 177.5 | 508 | 31.5 |
| p_hat700-3 | 62 | 700 | 183,010 | 526 | 89 | 602 | 14 |
| p_hat1500-1 | 12 | 1,500 | 284,923 | 383 | 197 | 509 | 82 |
| p_hat1500-2 | 65 | 1,500 | 568,960 | 763 | 387 | 1,100 | 37 |
| p_hat1500-3 | 94 | 1,500 | 847,244 | 1,132.5 | 192 | 1,297.5 | 25.75 |

In Table 2, we report the largest clique size obtained (Max), mean (Mean), standard deviation (Std), and the average CPU time (CPU time) over 100 runs for each instance and each of three formulations. When the largest clique size obtained is the same as the largest known clique size, the result is reported in bold. Observe that, with the exception of the keller instances, the average clique size obtained from using either $\Phi_1$ or $\Phi_2$ was strictly larger than the average obtained from using $\Phi_B$. Overall, $\Phi_1$ performed slightly better than $\Phi_2$, yielding a strictly larger clique in 17 of 33 instances. Next, taking a look at the results related to the C and p_hat families, we notice that, as the number of nodes increases (and also, the number of edges increases), finding a solution close to the best known gets harder and harder for all three formulations. This is likely because of the simplicity of the global optimization approach that we use to solve the problem. However, for the smaller instances in these groups $\Phi_1$ and $\Phi_2$ performed quite well. In particular, the formulation using $\Phi_1$ found the largest known clique size in six of the instances. The DSJC, brock, and gen families all confirm the good behavior of the proposed formulations. Indeed, in all cases, the solutions found were closer (sometimes significantly) to the best known clique than the ones found using $\Phi_B$.

## 4. Conclusions

We described a general regularized continuous formulation for the MCP and developed conditions that guarantee an equivalence between the original problem and the continuous reformulation in both a global and a local sense. We have also proved the results in a step by step manner, which we hope reveals some of the underlying structural properties of the formulation. We further proposed two specific regularizers that satisfy the general conditions given in the paper and compared the two related continuous formulations with the one proposed in Bomze [2] on different families of widely used maximum clique instances belonging to the

**Table 2.** Results obtained on the DIMACS instances. Numbers in bold indicate that the largest clique size obtained is the same as the largest known clique size.

| | $\Phi_B$ | | | | $\Phi_1$ | | | | $\Phi_2$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Instances | Max | Mean | Std | CPU time | Max | Mean | Std | CPU time | Max | Mean | Std | CPU time |
| C125.9 | **34** | 32.83 | 0.92 | 0.13 | **34** | 33.17 | 0.38 | 0.69 | **34** | 33.22 | 0.42 | 0.25 |
| C250.9 | 40 | 37.08 | 1.30 | 0.43 | **44** | 40.79 | 1.09 | 1.31 | **44** | 40.77 | 1.13 | 0.77 |
| C500.9 | 49 | 45.51 | 1.46 | 1.80 | 54 | 51.03 | 1.57 | 5.14 | 54 | 50.92 | 1.68 | 3.36 |
| C1000.9 | 60 | 54.40 | 2.26 | 10.01 | 63 | 58.90 | 1.91 | 23.57 | 63 | 59.16 | 1.94 | 15.76 |
| C2000.5 | 13 | 11.46 | 0.70 | 56.09 | 15 | 12.90 | 1.08 | 101.40 | 15 | 12.95 | 0.94 | 66.69 |
| C2000.9 | 69 | 61.75 | 1.86 | 56.67 | 67 | 63.35 | 1.61 | 106.22 | 68 | 63.89 | 1.96 | 76.71 |
| DSJC500_5 | 10 | 9.43 | 0.61 | 2.20 | 11 | 10.45 | 0.55 | 4.59 | 12 | 10.19 | 0.44 | 2.36 |
| DSJC1000_5 | 13 | 10.68 | 0.97 | 10.29 | 14 | 11.86 | 0.79 | 25.53 | 14 | 11.53 | 1.20 | 13.69 |
| brock200_2 | 9 | 8.00 | 0.58 | 0.25 | 10 | 9.30 | 0.67 | 0.64 | 10 | 9.04 | 0.21 | 0.37 |
| brock200_4 | 14 | 12.60 | 0.87 | 0.30 | 15 | 13.44 | 0.63 | 1.35 | 15 | 13.40 | 0.64 | 0.43 |
| brock400_2 | 23 | 19.84 | 1.00 | 1.16 | 24 | 22.14 | 1.26 | 3.49 | 24 | 21.80 | 1.23 | 1.60 |
| brock400_4 | 22 | 19.77 | 1.33 | 1.10 | 24 | 21.47 | 1.15 | 2.62 | 24 | 21.46 | 1.05 | 1.51 |
| brock800_2 | 18 | 15.60 | 1.12 | 6.14 | 20 | 17.32 | 0.96 | 15.58 | 20 | 17.28 | 0.94 | 7.81 |
| brock800_4 | 17 | 15.22 | 0.94 | 6.53 | 20 | 17.12 | 1.05 | 17.79 | 19 | 17.08 | 1.00 | 8.83 |
| gen200_p0.9_44 | 36 | 33.37 | 1.25 | 0.31 | 40 | 37.51 | 1.05 | 1.45 | 40 | 37.43 | 1.12 | 0.49 |
| gen200_p0.9_55 | 40 | 36.93 | 1.02 | 0.31 | 41 | 38.95 | 1.14 | 1.06 | 41 | 38.98 | 1.20 | 0.41 |
| gen400_p0.9_55 | 48 | 44.19 | 1.57 | 0.96 | 51 | 49.01 | 0.87 | 3.22 | 51 | 49.03 | 0.86 | 1.64 |
| gen400_p0.9_65 | 48 | 43.97 | 2.28 | 1.13 | 51 | 48.05 | 1.75 | 4.14 | 51 | 48.25 | 1.90 | 1.79 |
| gen400_p0.9_75 | 46 | 42.93 | 1.41 | 1.05 | 49 | 47.76 | 1.04 | 5.69 | 50 | 47.65 | 1.38 | 2.91 |
| hamming8-4 | **16** | 13.61 | 2.27 | 0.43 | **16** | 15.73 | 1.00 | 0.66 | **16** | 15.73 | 1.01 | 0.51 |
| hamming10-4 | 34 | 30.54 | 1.23 | 10.69 | **40** | 33.45 | 1.37 | 35.23 | **40** | 33.47 | 1.42 | 18.99 |
| keller4 | 8 | 7.17 | 0.38 | 0.19 | 7 | 7.00 | 0.00 | 0.73 | 9 | 7.02 | 0.20 | 0.18 |
| keller5 | 16 | 15.02 | 0.15 | 4.40 | 15 | 15.00 | 0.00 | 7.54 | 15 | 15.00 | 0.00 | 4.39 |
| keller6 | 34 | 32.86 | 1.46 | 388.05 | 34 | 33.83 | 0.61 | 290.79 | 34 | 32.80 | 1.48 | 222.96 |
| p_hat300-1 | 7 | 7.00 | 0.00 | 0.54 | **8** | 8.00 | 0.00 | 0.88 | **8** | 8.00 | 0.00 | 0.52 |
| p_hat300-2 | 24 | 24.00 | 0.00 | 0.55 | **25** | 24.01 | 0.10 | 1.89 | 24 | 24.00 | 0.00 | 0.44 |
| p_hat300-3 | 33 | 31.15 | 0.76 | 0.62 | **36** | 33.39 | 0.70 | 1.86 | **36** | 33.20 | 0.64 | 0.90 |
| p_hat700-1 | 9 | 7.24 | 0.62 | 3.97 | 9 | 8.15 | 0.43 | 13.31 | 9 | 7.94 | 0.55 | 6.15 |
| p_hat700-2 | 43 | 41.53 | 0.75 | 4.49 | **44** | 43.61 | 0.70 | 10.07 | **44** | 43.50 | 0.72 | 4.74 |
| p_hat700-3 | 60 | 58.71 | 0.87 | 4.77 | 61 | 58.99 | 0.69 | 10.55 | 61 | 59.21 | 0.80 | 7.73 |
| p_hat1500-1 | 11 | 9.00 | 0.86 | 23.26 | 11 | 9.36 | 1.12 | 41.27 | 11 | 9.83 | 0.65 | 25.40 |
| p_hat1500-2 | 62 | 59.09 | 1.10 | 29.06 | 62 | 60.79 | 0.87 | 44.76 | 64 | 62.30 | 1.15 | 30.42 |
| p_hat1500-3 | 87 | 81.90 | 1.80 | 31.76 | 88 | 86.99 | 0.39 | 225.12 | 92 | 88.21 | 1.98 | 38.19 |

DIMACS benchmark. The numerical results, albeit still preliminary, seem to confirm the effectiveness of the proposed regularizers; that is, when a local optimization method is applied to the new regularized formulations, cliques of high quality can often be obtained in reasonable computational times.

## Acknowledgments

This work was partly developed while the first author was an Experienced Researcher of the "MINO: Mixed-Integer Nonlinear Optimization" Program funded by the European Union. In addition, the authors wish to thank two anonymous reviewers for their thoughtful reading of the manuscript and one of these reviewers in particular for providing the straightforward proof of Lemma 2.

## References

[1] Bertsekas DP (1999) *Nonlinear Programming* (Athena Scientific, Belmont, MA).
[2] Bomze IM (1997) Evolution towards the maximum clique. *J. Global Optim.* 10(2):143–164.
[3] Bomze IM (2016) Copositivity for second-order optimality conditions in general smooth optimization problems. *Optimization* 65(4):779–795.
[4] Bomze IM, Pelillo M, Stix V (2000) Approximating the maximum weight clique using replicator dynamics. *IEEE Trans. Neural Network* 11(6): 1228–1241.
[5] Bomze IM, Budinich M, Pardalos PM, Pelillo M (1999) *The Maximum Clique Problem* (Springer, Boston), 1–74
[6] Bomze IM, Budinich M, Pelillo M, Rossi C (2002) Annealed replication: A new heuristic for the maximum clique problem. *Discrete Appl. Math.* 121(13):27–49.
[7] Bradley P, Mangasarian O, Rosen J (1998) Parsimonious least norm approximation. *Comput. Optim. Appl.* 11(1):5–21.
[8] Frank M, Wolfe P (1956) An algorithm for quadratic programming. *Naval Res. Logist. Quart.* 3(1/2):95–110.
[9] Gibbons LE, Hearn DW, Pardalos PM (1993) A continuous based heuristic for the maximum clique problem. Johnson DS, Trick M, eds. *Second DIMACS Implementation Challenge*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, vol. 26 (American Mathematical Society, Providence, RI), 103–124.

[10] Gibbons LE, Hearn DW, Pardalos PM, Ramana MV (1997) Continuous characterizations of the maximum clique problem. *Math. Oper. Res.* 22(3):754–768.
[11] Hager WW, Hungerford JT (2014) Optimality conditions for maximizing a function over a polyhedron. *Math. Programming* 145(1/2): 179–198.
[12] Johnson DS, Trick MA (1996) *Cliques, Coloring, and Satisfiability: Second DIMACS Implementation Challenge, October 11–13, 1993*, vol. 26 (American Mathematical Society, Providence, RI).
[13] Karp R (1972) Reducibility among combinatorial problems. Miller RE, Thatcher JW, eds. *Complexity of Computer Computations: Proc. Sympos. Complexity Comput. Comput.* (Plenum Press, New York), 85–103.
[14] Keller OH (1930) Über die lückenlose Erfüllung des Raumes mit Würfeln. *J. Die Reine Angew Math.* 1930(163):231–248.
[15] Kuznetsova A, Strekalovsky A (2001) On solving the maximum clique problem. *J. Global Optim.* 21(3):265–288.
[16] Motzkin TS, Straus EG (1965) Maxima for graphs and a new proof of a theorem of Turán. *Canadian J. Math.* 17:533–540.
[17] Murty KG, Kabadi SN (1987) Some NP-complete problems in quadratic and linear programming. *Math. Programming* 39(2):117–129.
[18] Nocedal J, Wright SJ (2006) *Numerical Optimization*, 2nd ed. (Springer, New York).
[19] Pardalos PM, Phillips AT (1990) A global optimization approach for solving the maximum clique problem. *Internat. J. Comput. Math.* 33(3-4): 209–216.
[20] Pardalos PM, Schnitger G (1988) Checking local optimality in constrained quadratic programming can be NP-hard. *Oper. Res. Lett.* 7(1): 33–35.
[21] Pelillo M (1996) Relaxation labeling networks for the maximum clique problem. *J. Artificial Neural Networks* 2(4):313–328.
[22] Pelillo M, Jagota A (1996) Feasible and infeasible maxima in a quadratic program for maximum clique. *J. Artificial Neural Networks* 2(4): 411–420.
[23] Rota-Bulò S, Pelillo M (2009) A generalization of the Motzkin-Straus theorem to hypergraphs. *Optim. Lett.* 3(2):287–295.
[24] Wu Q, Hao JK (2015) A review on algorithms for maximum clique problems. *Eur. J. Oper. Res.* 242(3):693–709.