Maximum Stable Set Formulations and Heuristics Based on Continuous Optimization*

Samuel Burer† Renato D.C. Monteiro‡ Yin Zhang§

December 12, 2000

Abstract

The stability number $\alpha(G)$ for a given graph G is the size of a maximum stable set in G. The Lovász theta number provides an upper bound on $\alpha(G)$ and can be computed in polynomial time as the optimal value of the Lovász semidefinite program. In this paper, we show that restricting the matrix variable in the Lovász semidefinite program to be rank-one and rank-two, respectively, yields a pair of continuous, nonlinear optimization problems each having the global optimal value $\alpha(G)$. We propose heuristics for obtaining large stable sets in G based on these new formulations and present computational results indicating the effectiveness of the heuristics.

Keywords: maximum stable set, maximum clique, minimum vertex cover, semidefinite program, semidefinite relaxation, continuous optimization heuristics, nonlinear programming.

AMS 1991 subject classification: 90C06, 90C27, 90C30.

1 Introduction

Let $G = (V, E)$ be a simple, undirected graph. A stable (independent) set S in G is a set of vertices that are mutually nonadjacent, and the size of S is given by its cardinality $|S|$. The stability number of G, denoted by $\alpha(G)$, is the size of a maximum stable set in G. The maximum stable set problem, or MSS problem for short, on G is to find a maximum stable set in G. It is well known that the MSS problem on G is equivalent to the minimum vertex cover problem on G and to the maximum clique problem on the complement graph of G.

The MSS problem is a classical NP-Hard optimization problem which has been studied extensively. Numerous approaches for solving or approximating the MSS problem have been

*This is Technical Report TR00-34, Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77005. Computational results reported in this paper were obtained on an SGI Origin2000 computer at Rice University acquired in part with support from NSF Grant DMS-9872009.

†School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA. This author was supported in part by NSF Grants CCR-9902010 and INT-9910084. (Email: burer@math.gatech.edu).

‡School of ISyE, Georgia Institute of Technology, Atlanta, Georgia 30332, USA. This author was supported in part by NSF Grants CCR-9902010 and INT-9910084. (Email: monteiro@isye.gatech.edu).

§Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77005, USA. This author was supported in part by DOE Grant DE-FG03-97ER25331, DOE/LANL Contract 03891-99-23 and NSF Grant DMS-9973339. (Email: zhang@caam.rice.edu).

proposed. A survey paper [13] by Pardalos and Xue gives a brief overview of progress made on the maximum clique problem or, equivalently, the MSS problem. The authors describe several different formulations of the MSS problem, a number of exact algorithms (such as explicit and implicit enumeration), and a number of heuristic algorithms (such as sequential greedy approaches, local and random searches) for the MSS problem. Though many of these algorithms perform well on certain classes of instances, it seems clear that no single algorithm has demonstrated superiority on all classes of graphs. Hence, new formulations and algorithms are needed to strengthen our ability to solve or approximate the MSS problem in general.

An upper bound on the stability number $\alpha(G)$ of a graph G was defined and studied by Lovász in [11] (see also [8] for further details). This upper bound is called the Lovász theta number $\vartheta(G)$ and can be computed as the optimal value of the following semidefinite program (SDP), called the Lovász theta SDP:

$$\max\{e^T X e : \mathrm{trace}(X) = 1,\ X_{ij} = 0\ \forall\ (i,j) \in E,\ X \succeq 0\}, \tag{1}$$

where X is a symmetric matrix of size $|V| \times |V|$, the constraint $X \succeq 0$ requires that X be positive semidefinite, and e is the $|V|$-length column vector of all ones.

In addition to its theoretical value, the upper bound $\vartheta(G)$ on the stability number of G could be practically useful in an implicit enumeration scheme such as branch-and-bound for solving the MSS problem as long as it can be computed efficiently. There are, however, practical difficulties in applying semidefinite programming to the MSS problem. For a graph having $|V|$ vertices and $|E|$ edges, the number of variables and constraints involved in the corresponding SDP (1) is on the order of $|V|^2+|E|$. Solving such an SDP becomes increasingly expensive as the size of $|V|$ and $|E|$ increase. For example, for $|V| \geq 500$ and $|E| \geq 1,000$, solving (1) via traditional interior-point methods becomes excessively time-consuming and memory-intensive on today's computers. (For further details on semidefinite programming and the classical interior-point algorithms to solve them, we refer the reader to [14].) Even though there have been some recent advances in solving (1) for graphs having more than 1,000 vertices and 100,000 edges using non-traditional approaches (see, for example, [2, 3, 4, 9]), solving (1) for large-scale instances is still a formidable challenge.

Besides providing a high-quality upper bound on $\alpha(G)$, can the Lovász theta SDP (1) be utilized in some way to provide a high-quality lower bound on $\alpha(G)$? More specifically, can (1) be exploited to find large stable sets in G? To the best of our knowledge, no methods to provide such lower bounds have been proposed (though Benson and Ye [1] have solved an alternative SDP formulation for the Lovász theta number $\vartheta(G)$ to generate stable sets in G).

In a recent paper [5], the authors of the present paper have considered another combinatorial optimization problem — the Max-Cut problem on G — in a similar context as we now consider the MSS problem. The SDP relaxation of Max-Cut is well known to provide both a good upper bound on the maximum cut size as well as the ability to obtain guaranteed high-quality cuts in G via the Goemans-Williamson randomization scheme (see [7]). The focus of [5] was to develop fast methods for finding high-quality cuts in G, and so instead of solving the expensive SDP relaxation for Max-Cut, the authors restricted the rank of the matrix variable of the relaxation to be at most two and applied a modified Goemans-Williamson scheme to the "rank-two" problem. They provided strong computational evidence showing that this rank-two problem produces higher quality cuts in G than the SDP relaxation, while taking much less computer time and storage. A disadvantage of the rank-two approach, however, is that it is a nonconvex relaxation of Max-Cut (unlike the SDP relaxation) and hence there are many local maxima which cannot be guaranteed to provide an upper bound on the size of a maximum cut. Nonetheless, [5] has shown that the rank-two Max-Cut relaxation is a powerful tool when one wishes to find high-quality approximate solutions to the Max-Cut problem of large size.

Since the strategy of replacing an expensive convex relaxation by an inexpensive nonconvex relaxation has worked surprisingly well for approximating the Max-Cut problem, it is natural to ask whether or not a similar strategy would also work well for the MSS problem. Hence, in this paper we apply the low-rank restriction strategy to the Lovász theta SDP and study the resulting problems. In particular, we will show that: (i) restricting the matrix variable X in (1) to be of low rank (more precisely, to be either rank one or two) has a meaningful correspondence with the MSS problem; (ii) any feasible solution of the low-rank problems can be used to obtain a stable set in G with size at least as large as the solution's continuous objective value; and (iii) local optima of the low-rank problems can be obtained quickly, taking advantage of graph structure such as sparsity.

This paper is organized as follows. In Section 2, we develop and analyze the nonlinear program formed by restricting the matrix variable X of (1) to be rank one. In particular, we demonstrate how every feasible solution of this rank-one problem naturally leads to a stable set in G and also show that the optimal value of the rank-one problem is exactly the stability number $\alpha(G)$. In Section 3, we show the same results for the problem resulting from restricting X to be at most rank two. In Section 4, we discuss an implementation of several heuristics based on the rank-one and rank-two formulations of MSS. By comparing their performance on a large set of benchmark instances, we conclude that the rank-two heuristic gives better stable sets than the rank-one formulation even though it generally requires more computation time. In Section 5, we conclude with a few final remarks.

2 The Rank-One Problem

As stated in the introduction, an important upper bound on the stability number $\alpha(G)$ for a given graph G is the Lovász theta number $\vartheta(G)$. It was given by Lovász as the optimal value of the Lovász theta SDP:

$$\vartheta(G) \equiv \max\left\{e^T X e : \mathrm{trace}(X) = 1,\ X_{ij} = 0\ \forall\ (i,j) \in E,\ X \succeq 0\right\}. \tag{2}$$

The "rank-one" restriction of the above formulation is to require that X be a rank-one matrix; that is, to require that $X = xx^T$ for some $x \in \Re^n$. Making the substitution $X = xx^T$ in the above formulation yields the following nonlinear program:

$$\alpha_1 \equiv \max\left\{(e^T x)^2 : \|x\|^2 = 1,\ x_i x_j = 0\ \forall\ (i,j) \in E,\ x \in \Re^n\right\}, \tag{3}$$

where $\|\cdot\|$ denotes the Euclidean norm in $\Re^n$. The goal of this subsection is to establish that $\alpha_1 = \alpha(G)$ and to characterize all local maxima of problem (3).

Letting $\mathcal{F}_1$ denote the feasible set of (3), we have the following simple result.

Lemma 2.1 *Given $G = (V, E)$, the following statements hold:*

(a) if $x \in \mathcal{F}_1$, then $S_x \equiv \{i : x_i \neq 0\}$ is a stable set in G;

(b) if S is a stable set, then $\mathcal{F}_S \equiv \{x \in \Re^n : \|x\| = 1,\ x_i = 0\ \forall\ i \notin S\} \subset \mathcal{F}_1$.

Proof. To prove (a), assume that $x \in \mathcal{F}_1$, and let p and q be two arbitrary vertices of S_x. Since $x_p \neq 0$ and $x_q \neq 0$, we have $x_p x_q \neq 0$, and since $x_i x_j = 0$ for all $(i,j) \in E$, we conclude that $(p,q) \notin E$. This shows that S_x is a stable set.

To prove (b), assume that S is a stable set, and let $x \in \mathcal{F}_S$ be given. Since every edge $(i,j) \in E$ has at least one endpoint outside of S, $x_i x_j = 0$ for all $(i,j) \in E$. Combining this with the fact that $\|x\| = 1$, we see that $x \in \mathcal{F}_1$. ∎

Given a set $S \subset V$, we consider the auxiliary problem

$$(P_S): \qquad \max\left\{(e^T x)^2 : x \in \mathcal{F}_S\right\},$$

where $\mathcal{F}_S$ is defined as in Lemma 2.1. Observe that the point $x^S \in \Re^n$ defined by

$$x_i^S = \begin{cases} |S|^{-1/2}, & \text{if } i \in S \\ 0\ , & \text{otherwise} \end{cases}$$

is a feasible solution for (P_S) with objective value equal to $|S|$. In fact, the following result holds.

Lemma 2.2 *The points x^S and $-x^S$ are global maxima of problem (P_S), each having objective value $|S|$. Moreover, these solutions are the only local maxima of (P_S).*

Proof. We prove the lemma by classifying all local maxima of (P_S). It is easy to see that x is not a local maximum if $(e^T x)^2 = 0$, and so we assume that $e^T x \neq 0$. Considering the first-order necessary conditions for (P_S) and noting that the regularity condition holds at $x \neq 0$, we see that any local maximum of (P_S) must satisfy the equations $e^T x = \lambda x_i$ for all $i \in S$ and for some $\lambda \in \Re$. Since $e^T x \neq 0$, we conclude that $\lambda \neq 0$, and so $x_i = e^T x/\lambda$ for all $i \in S$. In other words, x_i is constant over all $i \in S$. The constraint $\|x\| = 1$ thus implies that either $x_i = |S|^{-1/2}$ or $x_i = -|S|^{-1/2}$ for all $i \in S$, hence completing the proof. ∎

We are now ready to state the main result for the rank-one problem (3).

Theorem 2.3 *A point $x \in \mathcal{F}_1$ is a local maximum of (3) if and only if $x = \pm x^S$ for some maximal stable set $S \subset V$, in which case the objective value of x is exactly $|S|$.*

Proof. Assume first that $x \in \mathcal{F}_1$ is a local maximum of (3) and let $S \equiv S_x$. By Lemma 2.1, S is a stable set, and the feasible region $\mathcal{F}_S$ of (P_S) is contained in $\mathcal{F}_1$. Since x is clearly in $\mathcal{F}_S$, it follows that x is also a local maxima of (P_S). By Lemma 2.2, it follows that $x = \pm x^S$ and that the objective value of x is equal to $|S|$. To show that S is a maximal stable set, assume for contradiction that there exists a stable set $\tilde{S} \subset V$ which properly contains S. Clearly $x \in \mathcal{F}_S \subset \mathcal{F}_{\tilde{S}}$, from which it follows that x is also a local maximum of $(P_{\tilde{S}})$. By Lemma 2.2, this implies that the objective value of x is $|\tilde{S}|$, which contradicts the earlier conclusion that the objective value of x is $|S|$ since $|S| < |\tilde{S}|$. We have thus proved the "only if" part of the theorem.

Assume now that $x = \pm x^S$ for some maximal stable set S. Since $x_i \neq 0$ for all $i \in S$, there exists a neighborhood $\mathcal{N}_x \subset \Re^n$ of x such that every $y \in \mathcal{N}_x$ satisfies $y_i \neq 0$ for all $i \in S$. Hence, for every $y \in \mathcal{N}_x$, the set S_y contains S. Now let y be an arbitrary point in $\mathcal{N}_x \cap \mathcal{F}_1$. Since S is a maximal stable set contained in S_y, which in turn is a stable set by Lemma 2.1, it follows that $S_y = S$. Hence, $\mathcal{N}_x \cap \mathcal{F}_1 \subset \mathcal{F}_S$. By Lemma 2.2, this implies that the objective value of any point $y \in \mathcal{N}_x \cap \mathcal{F}_1$ is at most $|S|$. Since the objective value of x is exactly $|S|$, we conclude that x is a local maxima of (3). ∎

As an immediate consequence of the above theorem, we can now state a characterization of the global maxima of problem (3).

Theorem 2.4 *A point $x \in \mathcal{F}_1$ is a global maximum of (3) if and only if $x = \pm x^S$ for some maximum stable set $S \subset V$. Moreover, the optimal value α_1 of (3) is equal to $\alpha(G)$.*

Proof. Clearly, the set of global maxima is nonempty since $\mathcal{F}_1$ is compact. Moreover, the set of global maxima consists of those local maxima which have the largest possible objective function value. By Theorem 2.3, this largest possible value is $\alpha(G)$, and hence the set of global maxima consists of exactly those points of the form $\pm x^S$, where S is a maximum stable set. ∎

3 The Rank-Two Problem

In this section we consider a "rank-two" formulation of MSS obtained by restricting the rank of the matrix variable X of the Lovász theta SDP (2) to be at most two, i.e., we require that X equal $xx^T + yy^T$ for some $x, y \in \Re^n$. Note that the rank-one formulation of the previous section can be obtained from the rank-two formulation by setting y to be a multiple of x. A surprising result which we will show in this section is that the optimal values of the two formulations are equal even though the feasible region of the rank-two problem is strictly larger than that of the rank-one problem. Moreover, we will show how global maxima for the rank-two problem yield maximum stable sets of G, and we will give a partial classification of the local maxima of the rank-two problem.

Making the substitution $X = xx^T + yy^T$ in the SDP problem (2), we obtain the following rank-two problem:

$$\alpha_2 \equiv \max\left\{(e^T x)^2 + (e^T y)^2 : \|x\|^2 + \|y\|^2 = 1,\ x_i x_j + y_i y_j = 0\ \forall\ (i,j) \in E\right\}, \tag{4}$$

where $x, y \in \Re^n$. An alternative form of the above problem is

$$\alpha_2 \equiv \max\left\{\left\|\sum_{i=1}^n z^i\right\|^2 : \sum_{i=1}^n \|z^i\|^2 = 1,\ (z^i)^T z^j = 0\ \forall\ (i,j) \in E\right\}, \tag{5}$$

where

$$z^i = (x_i, y_i)^T \in \Re^2,\ i = 1, \ldots, n.$$

The formulation (5) highlights the fact that the rank-two problem has n vector variables in $\Re^2$ — one for each node in the graph. (Recall that the rank-one problem has a scalar variable for each node.) In the sequel, we will use (x, y) and $(z^1, \cdots, z^n)$ interchangeably to denote the variable of the rank-two problem.

It is interesting to note that the level sets of problem (5) are rotationally invariant in the sense that, if $(z^1, \cdots, z^n)$ is feasible, then so is $(Qz^1, \cdots, Qz^n)$ for any orthogonal matrix $Q \in \Re^{2\times 2}$, and moreover the objective value is unchanged. Hence, the problem does not have any strict local maxima (or minima). Later in this section, we will establish results which show that any non-global local maximum can be shifted to some saddle point having the same objective value from which we may further increase the objective function.

3.1 Global maxima of the rank-two formulation

The main goal of this subsection is to establish that $\alpha_2 = \alpha_1 = \alpha(G)$ and to characterize the global maxima of problem (4). For convenience, we let $f(x, y)$ and $\mathcal{F}_2$ denote the objective function and set of feasible solutions of (4), respectively.

We first explain how a feasible point $(x, y) \in \mathcal{F}_2$ can be used to generate a stable set in G. We first define

$$V_0 \equiv V_0(x, y) \equiv \{i \in V : x_i = y_i = 0\} \equiv \{i \in V : z^i = 0\}. \tag{6}$$

We will adopt the convention that $t/0 = +\infty$ for every nonzero number $t \in \Re$, irrespective of whether it is positive or negative. We can then define

$$\rho_i \equiv x_i/y_i \in (-\infty, \infty], \quad \forall\, i \notin V_0.$$

Now let $\tilde{G} = (\tilde{V}, \tilde{E})$ denote the subgraph induced by the vertex set $V \setminus V_0$, that is, $\tilde{V} \equiv V \setminus V_0$ and $\tilde{E} = E \cap (\tilde{V} \times \tilde{V})$. Lemma 3.1 below shows that $\tilde{G}$ has a bipartite structure.

Recall that a graph $\hat{G} = (\hat{V}, \hat{E})$ is bipartite if there exists a pair of disjoint subsets $\hat{V}_1$ and $\hat{V}_2$ of $\hat{V}$ such that $\hat{V} = \hat{V}_1 \cup \hat{V}_2$ and every edge of $\hat{G}$ has exactly one endpoint in $\hat{V}_1$. In such a case we refer to $(\hat{V}_1, \hat{V}_2)$ as a *bipartition* of $\hat{G}$.

Lemma 3.1 *Let $(x, y) \in \mathcal{F}_2$ be given, and let $V_0 \equiv V_0(x, y)$. Let $\tilde{G} = (\tilde{V}, \tilde{E})$ denote the subgraph induced by the vertex set $\tilde{V} \equiv V \setminus V_0$. Then $\tilde{G}$ is a bipartite graph (hence any of its connected components is also bipartite) with bipartition $(\tilde{V}_1, \tilde{V}_2)$, where*

$$\tilde{V}_1 \equiv \{i \in \tilde{V} : \rho_i \in (0, \infty]\} \qquad \text{and} \qquad \tilde{V}_2 \equiv \{i \in \tilde{V} : \rho_i \in (-\infty, 0]\}.$$

Proof. The identity $\tilde{V} = \tilde{V}_1 \cup \tilde{V}_2$ is obvious. Now, let $(i, j) \in \tilde{E}$, or equivalently $(i, j) \in E$ with both $i, j \in \tilde{V}$, be given. Since $(x, y) \in \mathcal{F}_2$, we have $x_i x_j + y_i y_j = 0$, which clearly implies that $\rho_i = (-1)/\rho_j$. From this identity we immediately conclude that exactly one of the endpoints of (i, j) is in $\tilde{V}_1$. Hence the lemma follows. ∎

It is easy to verify that $\tilde{V}_1$ and $\tilde{V}_2$ can be equivalently defined as:

$$\tilde{V}_1 \equiv \{i \in \tilde{V} : y_i = 0 \;\text{ or }\; x_i y_i > 0\} \qquad \text{and} \qquad \tilde{V}_2 \equiv \{i \in \tilde{V} : x_i = 0 \;\text{ or }\; x_i y_i < 0\}.$$

We now describe how a stable set can be constructed from the graph $\tilde{G} = (\tilde{V}, \tilde{E})$ which is induced by an arbitrary solution $(x, y) \in \mathcal{F}_2$. By the above lemma, the subgraph $\tilde{G}$ is bipartite. The following proposition contains straightforward observations.

Proposition 3.2 *Let the bipartite graph $\tilde{G}$ be induced by $(x, y) = (z^1, \cdots, z^n) \in \mathcal{F}_2$. Let $\tilde{G}_k = (\tilde{V}_k, \tilde{E}_k)$, $k = 1, \ldots, p$, be the connected components of $\tilde{G}$, and (A_k, B_k) be the bipartition of $\tilde{G}_k$ such that $a_k \equiv |A_k| \geq |B_k| \equiv b_k$, where by convention $B_k = \emptyset$ whenever $\tilde{G}_k$ consists of only one vertex. Then:*

(i) the set $A = \cup_{k=1}^{p} A_k$ is a maximal stable set in $\tilde{G}$ of size $a \equiv |A| = a_1 + \ldots + a_p$;

(ii) other maximal stable sets in $\tilde{G}$ of the same size can be obtained by interchanging A_k and B_k whenever $a_k = b_k$.

(iii) for all $i \in A_k$ and $j \in B_k$, $z^i \perp z^j$; hence all vectors z^i for $i \in A_k$ are colinear, and all vectors z^j for $j \in B_k$ are colinear.

In what follows we will refer to a stable set A obtained according to the above proposition as a *stable set induced* by the solution (x, y). The following result gives an important relationship between a feasible solution of problem (4) and its induced stable set.

Proposition 3.3 *Let $(x, y) = (z^1, \cdots, z^n) \in \mathcal{F}_2$ be given and let A denote a stable set induced by (x, y). Then the objective value of (4) at (x, y) is bounded by the size of A, that is, $f(x, y) \leq |A|$. Moreover, the equality $f(x, y) = |A|$ holds if and only if there exists $w \in \Re^2$ of unit-length such that, for $k = 1, 2, \cdots, p$,*

$$z^i = w/\sqrt{a}, \quad z^j = 0, \quad \forall\, i \in A_k, \forall\, j \in B_k, \quad \text{if } a_k > b_k; \tag{7}$$

$$z^i + z^j = w/\sqrt{a}, \quad \forall\, i \in A_k, \forall\, j \in B_k, \quad \text{if } a_k = b_k. \tag{8}$$

The proof of the above result is somewhat lengthy, so we will leave it until Section 3.4. We now state the main result of this subsection.

Theorem 3.4 *For any graph G, the optimal value of problem (4) is exactly the stability number of G, i.e., $\alpha_2 \equiv \alpha(G)$.*

Proof. First observe that (4) has an optimal solution since its feasible region is a compact set. Moreover, since (4) is a relaxation of (3), its optimal value α_2 satisfies $\alpha_2 \geq \alpha(G)$.

Now let (x^*, y^*) be an optimal solution of (4) and let A be a stable set induced by (x^*, y^*). Then, by Proposition 3.3, we have $\alpha_2 = f(x^*, y^*) \leq |A| \leq \alpha(G)$. Hence, it follows that $\alpha_2 = f(x^*, y^*) = |A| = \alpha(G)$. ∎

3.2 Local maxima of the rank-two formulation

In this subsection, we study the local maxima of the rank-two formulation (4). Unlike the results obtained in Theorem 2.3 for the rank-one problem, the results here only provide a partial characterization for local maxima of the rank-two problem. We show, however, that in a certain sense the rank-two problem has many fewer local maxima than the rank-one problem does. In particular, we show that for every maximal stable set S in G, there is an easily computable, "canonical" feasible point (x, y) for the rank-two problem such that S is induced by (x, y), the objective value at (x, y) is $|S|$, and yet (x, y) is not a local maximum of the rank-two problem. This is in direct contrast with the rank-one problem where the canonical solutions $\pm x^S$ associated with S are necessarily local maxima.

As with the rank-one problem, the analysis of the local maxima of the rank-two problem (4) depends heavily on the properties of a certain auxiliary problem associated with every feasible point $(x, y) \in \mathcal{F}_2$. More precisely, let $(x, y) \in \mathcal{F}_2$, and define V_0 and (A_k, B_k) for $k = 1, \ldots, p$ as in Proposition 3.2. Then the auxiliary problem under consideration is given by (9) below, which in the sequel we will refer to as (P_{V_0}):

$$\begin{array}{lc} \max\limits_{z_i \in \Re^2} & \left\| \sum_{i=1}^{n} z^i \right\|^2 \\ \text{s.t.} & \sum_{i=1}^{n} \|z^i\|^2 = 1;\ z^i = 0,\ \ i \in V_0, \\ & z^i \perp z^j,\ \ \forall\, (i, j) \in A_k \times B_k,\ \ k = 1, \ldots, p. \end{array} \tag{9}$$

Proposition 3.5 *Problem (P_{V_0}) has no local maxima other than its global maxima, at which the objective value is $|A| \equiv | \cup_{k=1}^{p} A_k|$.*

We again defer the proof of this result until later (this time Appendix A) because it is long and technically involved. We are now ready to state the main theorem of this subsection.

Theorem 3.6 *Let $(x, y) \in \mathcal{F}_2$, and let V_0 and (A_k, B_k) for $k = 1, \ldots, p$ be defined as in Proposition 3.2. Let A be a stable set induced by (x, y). If (x, y) is a local maximum of (4), then (x, y) is a global maximum of (P_{V_0}), in which case the objective value at (x, y) is $|A|$.*

Proof. If $(x, y) \in \mathcal{F}_2$ is a local maximum of (4), then it is straightforward to see that (x, y) is a local maximum of (P_{V_0}), in which case Proposition 3.5 implies that (x, y) is a global maximum of (P_{V_0}) with objective value $|A| = \sum_{k=1}^{p} a_k$. Since the objective functions of (P_{V_0}) and (4) coincide at (x, y), we conclude that $f(x, y) = |A|$. ∎

We next show that the converse of Theorem 3.6 does not hold. That is, the global maxima of (P_{V_0}) are not necessarily local maxima of (4), which is in contrast with the "if and only if" characterization for the local maxima of the rank-one problem given in Theorem 2.3. Before we proceed, however, we state a result which gives conditions under which a point $(x, y) \in \mathcal{F}_2$ is not a local maximum.

Lemma 3.7 *Let $(x^0, y^0), (x^1, y^1) \in \mathcal{F}_2$, and suppose that*

$$(x_i^0, y_i^0) = \gamma_i^0 w^0, \quad \forall\, i = 1, \ldots, n, \tag{10}$$

$$(x_i^1, y_i^1) = \gamma_i^1 w^1, \quad \forall\, i = 1, \ldots, n, \tag{11}$$

for some $\gamma_i^0, \gamma_i^1 \in \Re$, $i = 1, \ldots, n$, and some perpendicular, unit vectors $w^0, w^1 \in \Re^2$. In addition, let $f^0 \equiv f(x^0, y^0)$ and $f^1 \equiv f(x^1, y^1)$. Then there exists a feasible path in $\mathcal{F}_2$ from (x^0, y^0) to (x^1, y^1) along which the objective function $f(x, y)$ is strictly monotonic if $f^0 \neq f^1$ or is constant if $f^0 = f^1$.

Proof. For each $t \in [0, 1]$, define $\psi(t) = [(1-t)^2 + t^2]^{-1/2}$ and

$$(x^t, y^t) \equiv \psi(t) \left[(1-t)(x^0, y^0) + t(x^1, y^1) \right]. \tag{12}$$

We will show that the point (x^t, y^t) is feasible with respect to (4) with objective value

$$f(x^t, y^t) \equiv f^t = \psi(t)^2 \left[(1-t)^2 f^0 + t^2 f^1 \right],$$

from which the result of the lemma clearly follows.

To show that (x^t, y^t) is feasible, we first verify that $\|(x^t, y^t)\| = 1$. This follows directly from (12) and the following equalities:

$$
\begin{aligned}
\|(x^t, y^t)\|^2 &= \psi(t)^2 \|(1-t)(x^0, y^0) + t(x^1, y^1)\|^2 \\
&= \psi(t)^2 \left[(1-t)^2 \|(x^0, y^0)\|^2 + t^2 \|(x^1, y^1)\|^2\right] \\
&= \psi(t)^2 \left[(1-t)^2 + t^2\right] = 1.
\end{aligned}
$$

Note that the first equality above follows from (10), (11) and the orthogonality of w^0 and w^1 and that the second equality follows from the feasibility of (x^0, y^0) and (x^1, y^1). Next, we show that (x^t, y^t) satisfies the constraints $x_i^t x_j^t + y_i^t y_j^t = 0$ for all $(i,j) \in E$. Letting $z^{ti} \equiv (x_i^t, y_i^t)$ for all $i = 1, \ldots, n$, the aforementioned constraint can be expressed as $\left(z^{ti}\right)^T z^{tj} = 0$. From (12) and the definition of z^{ti}, it is then easy to see that

$$
\left(z^{ti}\right)^T z^{tj} = \psi(t)^2 \left[(1-t)^2 \left(z^{0i}\right)^T z^{0j} + (1-t)t \left[\left(z^{0i}\right)^T z^{1j} + \left(z^{1i}\right)^T z^{0j}\right] + t^2 \left(z^{1i}\right)^T z^{1j}\right],
$$

where z^{0i} and z^{1i} are analogous to z^{ti} for $i = 1, \ldots, n$. Noting that $\left(z^{0i}\right)^T z^{0j} = \left(z^{1i}\right)^T z^{1j} = 0$ due to the feasibility of (x^0, y^0) and (x^1, y^1) and that $\left(z^{0i}\right)^T z^{1j} = \left(z^{1i}\right)^T z^{0j} = 0$ due to (10), (11), and the orthogonality of w^0 and w^1, we conclude that $\left(z^{ti}\right)^T z^{tj} = 0$, as desired. Hence, $(x^t, y^t) \in \mathcal{F}_2$.

We now compute the objective value f^t at (x^t, y^t). Since $f^t = \|\sum_{i=1}^n z^{ti}\|^2$, it follows from (12), the orthogonality of $\sum_{i=1}^n z^{0i}$ and $\sum_{i=1}^n z^{1i}$ (due to the orthogonality of w^0 and w^1) and the definition of the objective values f^0 and f^1 that

$$
\begin{aligned}
f^t &= \psi(t)^2 \left\| (1-t) \sum_{i=1}^n z^{0i} + t \sum_{i=1}^n z^{1i} \right\|^2 \\
&= \psi(t)^2 \left[(1-t)^2 \left\| \sum_{i=1}^n z^{0i} \right\|^2 + t^2 \left\| \sum_{i=1}^n z^{1i} \right\|^2 \right] \\
&= \psi(t)^2 \left[(1-t)^2 f^0 + t^2 f^1\right],
\end{aligned}
$$

which completes the proof. ∎

The following proposition implies that a global maximum of (P_{V_0}) (where, in the context of the proposition, $V_0 = V \setminus S$) is not necessarily a local maximum of the rank-two problem; hence the converse of Theorem 3.6 does not hold.

Proposition 3.8 *Let S be a stable set in G, and for any $w \in \Re^2$ with unit length, define*

$$
\begin{aligned}
(x_i, y_i) &= \frac{1}{\sqrt{|S|}} w \qquad \forall\ i \in S, \\
(x_i, y_i) &= 0 \qquad\qquad \forall\ i \notin S.
\end{aligned}
$$

Then $(x, y) \in \mathcal{F}^2$, $f(x, y) = |S|$ and S is induced by (x, y). If, in addition, S is not a maximum stable set, then (x, y) is not a local maximum.

Proof. Let the suggested solution (x, y) be called the *canonical solution* associated with (S, w). It is easy to show that the first three conclusions of the proposition hold. The final statement can be shown as follows. Assume S is not maximum, and let S^* be a maximum stable set of G. Then Lemma 3.7 shows that there is a strictly increasing, feasible path between the canonical solution associated with (S, w) and the canonical solution associated with $(S^*, w^\perp)$, where $w^\perp$ is orthogonal to w. Hence, (x, y) is not a local maximum. ∎

It is worthwhile to contrast the results obtained for the local maxima of the rank-two problem with those obtained for the rank-one problem. In the rank-one case, we consider an auxiliary problem (P_S) associated with a feasible solution $x \in \mathcal{F}_1$, where $S \equiv \{i : x_i \neq 0\}$, and show that x is a local maximum of the rank-one problem (3) if and only if S is a maximal stable set and x is a global maximum of (P_S). In the rank-two case, we also consider an auxiliary problem (P_{V_0}) associated with a feasible solution $(x, y) \in \mathcal{F}_2$, where $V_0 \equiv \{i : x_i = y_i = 0\}$, but are only able to show that (x, y) is a global solution of (P_{V_0}) if (x, y) is a local maximum of the rank-two problem (4). Moreover, we give an example in Appendix B showing that the local maximality of a point $(x, y) \in \mathcal{F}_2$ has little to do with the maximality of the stable set induced by (x, y).

On the surface, the above properties seem to be disadvantages of the rank-two formulation, since one would expect that finding a local maximum of (4) should correspond to finding a maximal stable set. However, Proposition 3.8 shows a significant advantage of the rank-two problem over the rank-one problem. In the rank-one problem, if one obtains a local maximum, then the associated stable set S is maximal, but one is "stuck" at the local maximum. In the rank-two problem, on the other hand, if one obtains a local maximum, the associated stable set S may or may not be maximal, but in either case one can easily move to a canonical feasible solution associated with S with the same objective value $|S|$. From this canonical solution which is not a local maximum, it is possible to "escape" to a higher local maximum corresponding to a larger stable set. Indeed, when this feature of the rank-two problem is exploited algorithmically, it allows us to find considerably larger stable sets than does the rank-one problem, as will be demonstrated in the computational results of the next section.

3.3 Extension to higher ranks?

Before closing our discussion of the rank-two problem, we address a question that arises naturally after we see the results for rank-one and rank-two problems. Is the optimal value of the rank-three problem — for variables $x, y, z \in \Re^n$,

$$\begin{array}{ll} \text{maximize} & (e^T x)^2 + (e^T y)^2 + (e^T z)^2 \\ \text{subject to} & \|x\|^2 + \|y\|^2 + \|z\|^2 = 1 \\ & x_i x_j + y_i y_j + z_i z_j = 0, \quad \forall\ (i, j) \in E \end{array} \tag{13}$$

also equal to $\alpha(G)$? Although the answer may be "yes" for special classes of graphs (e.g., perfect graphs), the general answer is "no," as the following example demonstrates. Let

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

be the pentagon graph. It is then straightforward to verify that the point

$$\begin{aligned} x &= (1/3,\ 0,\ 1/3,\ 1/3,\ 0) \\ y &= (0,\ 1/3,\ 0,\ 1/3,\ 1/3) \\ z &= (0,\ 0,\ 1/3,\ -1/3,\ 1/3) \end{aligned}$$

is feasible for the rank-three problem with objective value $2\frac{1}{9}$ which is greater than 2 — the stability number $\alpha(G)$ of the pentagon graph.

3.4 Proof of Proposition 3.3

In this subsection, we provide the proof for Proposition 3.3. We first prove a technical lemma.

Lemma 3.9 *Let V_0 be defined as in (6), (P_{V_0}) be defined as in (9), and (A_k, B_k), (a_k, b_k) for $k = 1, \ldots, p$ be defined as in Proposition 3.2. Then the global maximum value of (P_{V_0}) is bounded above by $a \equiv \sum_{k=1}^{p} a_k$. Moreover, a feasible solution $(z^1, \ldots, z^n)$ of (P_{V_0}) is a global maximum if and only if there exist pairs of vectors $(z^{ak}, z^{bk}) \in \Re^2 \times \Re^2$, $k = 1, \ldots, p$, and a vector $w \in \Re^2$ of unit-length such that the following conditions hold for all $k = 1, \ldots, p$:*

$$z^{ak} + z^{bk} = w/\sqrt{a}, \tag{14}$$

$$a_k > b_k \implies z^{bk} = 0, \tag{15}$$

$$z^i = \begin{cases} z^{ak}, & \forall\ i \in A_k, \\ z^{bk}, & \forall\ i \in B_k. \end{cases} \tag{16}$$

Furthermore, (16), together with feasibility, implies that for all $k = 1, \ldots, p$,

$$z^{ak} \perp z^{bk}. \tag{17}$$

Proof. Let $(z^1, \ldots, z^n)$ be an arbitrary feasible solution of (9) and let $z^i = (x_i, y_i)$ for all $i = 1, \ldots, n$. We define the following quantities for all $k = 1, \ldots, p$:

$$u^k \equiv \sum_{i \in A_k} z^i, \qquad v^k \equiv \sum_{i \in B_k} z^i, \tag{18}$$

$$\theta_{ak} \equiv \sum_{i \in A_k} \|z^i\|^2, \quad \theta_{bk} \equiv \sum_{i \in B_k} \|z^i\|^2, \quad \theta_k \equiv \theta_{ak} + \theta_{bk}. \tag{19}$$

Using the feasibility of $(z^1, \ldots, z^n)$ and the above identities, it is easy to see that $u^k \perp v^k$ for every $k = 1, \ldots, p$, and $\sum_{k=1}^{p} \theta_k = 1$. Furthermore,

$$\|u^k\|^2 = \left\| \sum_{i \in A_k} z^i \right\|^2 = \left(\sum_{i \in A_k} x_i \right)^2 + \left(\sum_{i \in A_k} y_i \right)^2 \le |A_k| \left(\sum_{i \in A_k} (x_i^2 + y_i^2) \right);$$

that is,

$$\|u^k\|^2 \le \ a_k \sum_{i \in A_k} \|z_i\|^2 = a_k\, \theta_{ak}, \tag{20}$$

where the equality holds if and only if the first equation in (16) holds. Similarly, we can show that

$$\|v^k\|^2 \leq \; b_k \sum_{i \in B_k} \|z_i\|^2 = b_k\, \theta_{bk}, \tag{21}$$

where the equality holds if and only if the second equation in (16) holds. Using the fact that $z^i = 0$ for all $i \in V_0$, the relations (18), the triangle inequality for norms and the Pythagorean identity, we obtain

$$\left\| \sum_{i=1}^{n} z^i \right\|^2 = \left\| \sum_{k=1}^{p} (u^k + v^k) \right\|^2 \leq \left(\sum_{k=1}^{p} \|u^k + v^k\| \right)^2 = \left(\sum_{k=1}^{p} \sqrt{\|u^k\|^2 + \|v^k\|^2} \right)^2, \tag{22}$$

where the equality holds if and only if all $u_k + v_k$ are nonnegative multiples of a fixed vector of unit-length, say, w; i.e.,

$$u_k + v_k = \gamma_k w. \tag{23}$$

Using relation (19) and the two inequalities (20) and (21), we obtain

$$\left(\sum_{k=1}^{p} \sqrt{\|u^k\|^2 + \|v^k\|^2} \right)^2 \leq \left(\sum_{k=1}^{p} \sqrt{a_k\, \theta_{ak} + b_k\, \theta_{bk}} \right)^2, \tag{24}$$

where, as mentioned above, the equality holds if and only if (16) holds, which implies (17) given the feasibility of $(z^1, \cdots, z^n)$. Next, using the fact that $a_k \geq b_k$ for all $k = 1, \ldots, p$, we obtain

$$\left(\sum_{k=1}^{p} \sqrt{a_k\, \theta_{ak} + b_k\, \theta_{bk}} \right)^2 \leq \left(\sum_{k=1}^{p} \sqrt{a_k\, \theta_k} \right)^2, \tag{25}$$

where the equality holds if and only if (15) holds. Finally, using the Cauchy-Schwarz inequality and the fact that $\sum_{k=1}^{p} \theta_k = 1$, we obtain

$$\left(\sum_{k=1}^{p} \sqrt{a_k\, \theta_k} \right)^2 \leq \left(\sum_{k=1}^{p} a_k \right) \left(\sum_{k=1}^{p} \theta_k \right) = a, \tag{26}$$

where the equality holds if and only if a_k/θ_k is constant over $k = 1, \ldots, p$. In this case, since $\sum_{k=1}^{p} \theta_k = 1$, we must have

$$\theta_k = a_k/a, \;\; k = 1, \cdots, p. \tag{27}$$

Combining the inequality (22) with (24)–(26), we have thus shown that the objective value of any feasible solution of (9) is bounded above by a and that the upper bound a is attained if and only if all these inequalities hold as equalities, i.e., if and only if (23), (16), (15) and (27) hold.

We now show that under (15) and (16) (hence (17) by feasibility), the relations (23) and (27) hold if and only if (14) holds; therefore, the upper bound a is attained if and only if

(14)–(16) hold. It is easy to verify that when (15) and (16) hold, then (14) implies (23), and also (27) using the definitions (18) and (19). Now assume that (15), (16), (23) and (27) hold. First note from (23) and (16) that for $k = 1, \ldots, p$,

$$\gamma_k\, w = u^k + v^k = a_k\, z^{ak} + b_k\, z^{bk} = a_k\, (z^{ak} + z^{bk}).$$

Hence $z^{ak} + z^{bk} = (\gamma_k/a_k)w$. Now we only need to show that

$$\frac{\gamma_k}{a_k} \equiv \frac{1}{\sqrt{a}} \quad \text{or} \quad \gamma_k^2 = \frac{a_k^2}{a}.$$

Using (15), (16), (17), (27) and the definitions (18) and (19), we have

$$\begin{aligned}
\gamma_k^2 &= \|u^k + v^k\|^2 = \|a_k\, z^{ak} + b_k\, z^{bk}\|^2 = a_k \left(a_k \|z^{ak}\|^2 + b_k \|z^{bk}\|^2 \right) \\
&= a_k \left(\sum_{i \in A_k} \|z^i\|^2 + \sum_{i \in B_k} \|z^i\|^2 \right) = a_k(\theta_{ak} + \theta_{bk}) = a_k \theta_k = a_k^2/a.
\end{aligned}$$

This completes the proof. ∎

Now we are ready to prove Proposition 3.3.

Proof. Let V_0 and (A_k, B_k), $k = 1, \ldots, p$, be the sets associated with the solution $(x, y) = (z^1, \ldots, z^n)$ as defined in (6) and in Proposition 3.2. The feasibility of $(z^1, \cdots, z^n)$ with respect to the rank-two problem (4) clearly implies its feasibility with respect to the auxiliary problem (9). The conclusion of the proposition now follows immediately from Lemma 3.9. ∎

4 Computational Results

In this section, we describe our computational experiences with the rank-one and rank-two formulations of the maximum stable set problem. Since the rank-one problem (3) can be seen as a restriction of the rank-two problem (4) in which the variable y is set to zero, our discussion of algorithmic techniques focuses on (4), from which direct applications to (3) are immediately available. In particular, we discuss an augmented Lagrangian method for obtaining a stationary point of (4) as well as a technique for extracting stable sets of G from points which are "nearly feasible" for (4). Such a technique is necessary since an augmented Lagrangian algorithm will only produce a feasible point in the limit and since the theoretical discussion of Section 3 has only described how to obtain a stable set of G from a feasible solution. We conclude the section with a comparison of three algorithms (one for the rank-one problem and two for the rank-two problem) on a large set of benchmark instances, and we then discuss the relative advantages and disadvantages of each algorithm.

4.1 The augmented Lagrangian algorithm

Since the edge constraints of the rank-two problem (4) are difficult to handle, we employ the standard technique of placing these constraints into the objective function via the augmented Lagrangian function. The unit-norm constraint, on the other hand, is easy to handle by a simple scaling. The augmented Lagrangian method we will consider is based on the following maximization for fixed Lagrangian multipliers $\lambda = (\lambda_{ij})_{(i,j)\in E}$ and fixed penalty parameter $\sigma > 0$:

$$\begin{array}{ll} \text{maximize} & (e^T x)^2 + (e^T y)^2 + \displaystyle\sum_{(i,j)\in E} \left(\lambda_{ij} - \frac{\sigma}{2}(x_i x_j + y_i y_j)\right)(x_i x_j + y_i y_j) \\ \text{subject to} & \|x\|^2 + \|y\|^2 = 1. \end{array} \tag{28}$$

Indeed, our algorithm to obtain a stationary point of (4) proceeds as follows (see [12] for a description of the standard method of multipliers):

0. Set $\lambda^1 = 0$, $\sigma_1 = 1.0$, $v_0 = \infty$, and $k = 1$.
1. Solve (28) with $\lambda = \lambda^k$ and $\sigma = \sigma_k$, obtaining (x^k, y^k). Set $v = \sum_{(i,j)\in E}(x_i^k x_j^k + y_i^k y_j^k)^2$.
2. If $v < 0.25\, v_{k-1}$, then set $\lambda_{ij}^{k+1} = \lambda_{ij}^k - \sigma_k(x_i^k x_j^k + y_i^k y_j^k)$ for all $(i,j) \in E$, $\sigma_{k+1} = \sigma_k$, and $v_{k+1} = v$. Otherwise, set $\lambda^{k+1} = \lambda^k$, $\sigma_{k+1} = 10\,\sigma_k$, and $v_{k+1} = v_k$.
3. Set $k = k + 1$ and go to step 1.

Theory dictates that the augmented Lagrangian algorithm will converge to a stationary point of (4), though in practice one would expect convergence to a local maximum. Indeed, we always observe that the algorithm converges to a point $(\bar{x}, \bar{y})$ which, in accordance with Theorem 3.6, has integer objective value. In fact, it is usually possible to determine the stable set A induced by $(\bar{x}, \bar{y})$ long before the algorithm actually converges to $(\bar{x}, \bar{y})$ (see the next subsection), and in such a situation, a good stopping criterion for the algorithm is a few decimals of agreement between $|A|$ and $f(x^k, y^k)$, where (x^k, y^k) represents an intermediate iterate of the algorithm.

A very important aspect of the algorithm proposed above is the procedure to solve (28). We first note that the single constraint $\|x\|^2 + \|y\|^2 = 1$ is easy to enforce by removing the constraint all together and compensating in the objective function via multiplication by the factor $(\|x\|^2 + \|y\|^2)^{-1}$. Hence, (28) is essentially an unconstrained minimization, and for this, we have implemented a first-order limited-memory BFGS algorithm which uses a strong Wolfe-Powell line search. Our choice of using a first-order approach is motivated by the fact that we do not need highly accurate local solutions of (4) (see the next subsection), that the function and gradient evaluations of the augmented Lagrangian are very fast especially when $|E|$ is small, and that the computation of the Hessian is very expensive relative to the computation of the gradient.

4.2 Extracting stable sets from approximately feasible solutions

By the very nature of the augmented Lagrangian algorithm, a feasible point of (4) will never be readily available. Instead, each point obtained by the algorithm will be approximately feasible; in particular, some or all of the edge constraints will be violated. Of course, as the algorithm progresses, the amount of infeasibility will decrease, but this infeasibility nonetheless complicates the computation of the stable set induced by (x, y) since our definition of the bipartite graph $\tilde{G} = (\tilde{V}, \tilde{E})$ in Section 3 requires $(x, y) \in \mathcal{F}_2$.

So we wish to define an analog $\tilde{G}_\varepsilon$ of $\tilde{G}$ which takes into account the "epsilon" infeasibility of a given point (x, y). First, we define

$$\varepsilon \equiv \max_{(i,j)\in E} |x_i x_j + y_i y_j|$$

so that ε is a measure of the largest constraint violation. (Recall that the augmented Lagrangian algorithm will always maintain the unit-norm constraint of (4).) Next, we define

$$V_\varepsilon(x, y) \equiv \left\{ i : x_i y_i = 0,\ \max(|x_i|, |y_i|) \le \varepsilon^{1/2} \right\} \cup \left\{ i : x_i y_i \ne 0,\ \min(|x_i|, |y_i|) < \varepsilon^{1/2} \right\}.$$

Note that when $\varepsilon = 0$, i.e., when $(x, y) \in \mathcal{F}_2$, the definition of V_ε matches the usual definition of V_0. Finally, we define $\tilde{V}_\varepsilon \equiv V \setminus V_\varepsilon$ and $\tilde{G}_\varepsilon$ to be the graph induced by the vertex set $\tilde{V}_\varepsilon$, that is, $\tilde{G}_\varepsilon = (\tilde{V}_\varepsilon, E \cap (\tilde{V}_\varepsilon \times \tilde{V}_\varepsilon))$. We have the following proposition.

Proposition 4.1 *The graph $\tilde{G}_\varepsilon$ is bipartite for all $\varepsilon \ge 0$.*

Proof. The case when $\varepsilon = 0$ follows as a result of Lemma 3.1 and the discussion prior to this proposition. So we assume $\varepsilon > 0$.

Note that ρ_i is well-defined for all $i \in \tilde{V}_\varepsilon$ since each i having $x_i = y_i = 0$ is contained in V_ε. Hence, we define $\tilde{V}_1 \equiv \{i \in \tilde{V}_\varepsilon : \rho_i \in (0, \infty]\}$ and $\tilde{V}_2 \equiv \{i \in \tilde{V}_\varepsilon : \rho_i \in (-\infty, 0]\}$. We claim that $(\tilde{V}_1, \tilde{V}_2)$ is a bipartition of $\tilde{G}_\varepsilon$.

We prove the claim by showing that $\tilde{V}_1$ and $\tilde{V}_2$ (by symmetry) are stable sets. So let $i, j \in \tilde{V}_1$; we will show that $(i, j) \notin E$. Using that $\rho_i, \rho_j \in (0, \infty]$, it is straightforward to see that $|x_i x_j + y_i y_j| = |x_i||x_j| + |y_i||y_j|$. Suppose first that $x_i y_i \ne 0$ and $x_j y_j \ne 0$. Then by the comment above and the definition of $V_\varepsilon(x, y)$, we have

$$|x_i x_j + y_i y_j| = |x_i||x_j| + |y_i||y_j| \ge \varepsilon^{1/2} \cdot \varepsilon^{1/2} + \varepsilon^{1/2} \cdot \varepsilon^{1/2} = 2\varepsilon > \varepsilon.$$

Since ε is the largest edge constraint violation, we conclude that $(i, j) \notin E$. Now suppose that either $y_i = 0$ or $y_j = 0$. (Note that $x_i \ne 0$ and $x_j \ne 0$ due to the definitions of $\tilde{V}_1$, ρ_i and ρ_j.) By the definition of $V_\varepsilon(x, y)$, we have

$$|x_i x_j + y_i y_j| = |x_i||x_j| > \varepsilon^{1/2} \varepsilon^{1/2} = \varepsilon.$$

As before, we conclude that $(i, j) \notin E$. This completes the proof that $\tilde{V}_1$ is a stable set and hence that $\tilde{G}_\varepsilon$ is bipartite. ■

Since the graph $\tilde{G}_\varepsilon$ is bipartite and since it is exactly $\tilde{G}$ when (x, y) is feasible, we can easily generalize the notion of an induced stable to an ε-infeasible point (x, y) via use of $\tilde{G}_\varepsilon$. In fact, we use this process of extracting a large stable set from $\tilde{G}_\varepsilon$ within the heuristic presented in the next subsection.

4.3 Comparison of three heuristics

In this subsection, we present a generic heuristic for finding large stable sets in G based on the ideas of the previous two subsections, that is, we combine the augmented Lagrangian algorithm to find a stationary point of (3) or (4) with the technique of extracting stable sets from approximately feasible solutions. After stating the generic heuristic, we then present computational results comparing three specific realizations of the heuristic and discuss the relative advantages and disadvantages of each.

The heuristic we present is motivated by Theorem 3.6 and Proposition 3.8 for the rank-two problem (4). Although the augmented Lagrangian method is only guaranteed to produce a stationary point of (4), it is reasonable to expect in practice that it will most likely produce a stationary point $(\bar{x}, \bar{y})$ at which $f(\bar{x}, \bar{y}) = |\bar{A}|$, where $\bar{A}$ is a stable set induced by $(\bar{x}, \bar{y})$. Such stationary points may or may not be local maxima (for example, it could be a canonical solution associated with $\bar{A}$). As the method converges to $(\bar{x}, \bar{y})$, it is likely that the extraction procedure of the previous subsection applied to an infeasible point (x, y) close to $(\bar{x}, \bar{y})$ will produce exactly $\bar{A}$. Since our overall purpose is to find a large stable set rather than to obtain an exact local maximum of (4), it is reasonable to terminate the augmented Lagrangian method once the objective function value becomes close to the size of the stable set produced by the extraction procedure.

Once the augmented Lagrangian method has terminated with a large stable set $\bar{S}$, it may be possible to improve upon $\bar{S}$ by using ideas from Proposition 3.8. In particular, Proposition 3.8 shows that there is a collection of easily computable *canonical solutions* associated with $\bar{S}$ which have objective value $|\bar{S}|$ but which are not local maxima of (4). Restarting the augmented Lagrangian method at or near one of these points may allow the method to "escape" from the stable set $\bar{S}$ to another stable set of a larger size.

We propose the following heuristic based on the above ideas:

Maximum Stable Set Heuristic:

Input: $G = (V, E)$, $r \in \{1, 2\}$, $K \geq 1$, $\bar{S} = \emptyset$
Output: $\bar{S}$
For $k = 1, \dots, K$

1. Begin the augmented Lagrangian algorithm on the rank-r problem.
   If $k = 1$, initialize with a random starting point, $\lambda = 0$ and $\sigma = 1$.
   If $k > 1$, initialize with a slight perturbation of a canonical solution of $\bar{S}$, the current λ and $\sigma = 1$.
2. As the algorithm runs, extract a stable set S every so often, and terminate the algorithm when the Lagrangian function value agrees with $|S|$ through the first two decimal places.
3. If $|S| > |\bar{S}|$, set $\bar{S} = S$.

End

We remark that the heuristic applies to both the rank-one and rank-two cases but that the technique of escaping from a stable set $\bar{S}$ is not grounded in theory for the rank-one case. Hence, we suggest that the one should take $K = 1$ in the rank-one case. We also comment that there is no theory guaranteeing that an "escape" would always be achievable for the rank-two problem. In fact, although "escapes" were often observed in our experiments, sometimes the algorithm did fail to escape from a non-maximum stable set.

We implemented the above heuristic in an ANSI C code which we call "Max-AO" and have tested it on an SGI Origin2000 with sixteen 300MHz R12000 processors at Rice University, although we note that Max-AO utilizes only one processor. In order to test the difference between the rank-one and rank-two formulations and to establish the effectiveness of the "escaping" procedure, we tested three realizations of the heuristic. In particular, we investigated the heuristics arising from the choice of $(1, 1)$, $(2, 1)$ and $(2, 5)$ for the parameters (r, K). We refer to the resulting heuristics as h11, h21 and h25, respectively.

We ran all three heuristics on a set of 64 graphs obtained from the Center for Discrete Mathematics and Theoretical Computer Science (DIMACS) [6]. These graphs were used as test instances for the maximum clique problem in the Second DIMACS Implementation Challenge. Since the maximum clique problem on a graph is the MSS problem on the complement graph, we actually run Max-AO on the complements of the 64 graphs.

Since each run of Max-AO is randomized, we have run h11, h21 and h22 ten times each on all 64 graphs. In Tables 1 and 2, we report the results of these experiments. Each table has thirteen columns which are divided into four groups. The first group contains information about the test graphs including the name of the graph, the number of vertices and edges, and the value of $\alpha \equiv \alpha(G)$ for the graph. Note that α is reported as -1 if the exact value is not known. The next group of columns gives the size of the largest stable set found by each of the three heuristics over all ten runs. The third group of columns gives the average size (rounded to the nearest integer) of the stable sets found by each heuristic over the ten runs, and the final set of columns gives the average time (in seconds) for each of the heuristics over the ten runs.

A few comments regarding the data in the tables are in order. First and foremost, the data shows that each heuristic is capable of finding large stable sets in a short amount of time. Second, the "average size" column shows that, on average, h21 finds better stable sets than h11, which indicates that the rank-two formulation is more useful than the rank-one formulation, and that h25 finds better stable sets than h21, which indicates that the escaping procedure works well. Third, even though the ranking of the heuristics in terms of quality of solutions is h11 (good), h21 (better) and h25 (best), the average times show that h25 is the most expensive and h11 is the least expensive. Hence, the data demonstrates the standard trade-off between quality of solution and computation time.

Regarding the "max overall" column, we see that h11 found the maximum stable set 22 times out of the 52 instances for which the exact value of α is known. Thus, h11 found a maximum stable set 42% of the time. The percentages for h21 and h25 are 65% and 75%, respectively. From a general perspective, then, we see that the heuristics are highly effective in finding large stable sets, although there is much room for improvement on some graphs — for example, the "brock" instances.

Since there are many varied techniques for the maximum stable set problem (or equivalently, the maximum clique problem) a direct comparison of the heuristics introduced in this paper with other heuristics is not readily available. Instead, we adopt a strategy employed by the Second DIMACS Implementation Challenge which is to provide machine timings of a particular computer algorithm on five different instances of increasing size. Table 3 lists the results for the SGI Origin2000 upon which our code has been run. The purpose of this information is to give a rough estimation of the differences in processor speed over several computers. (Please see [6] for more information and [10] for other heuristics which use the same technique for comparison.)

We do, however, believe that it is worth mentioning that the quality of the stable sets produced by our heuristics compares favorably with those of the heuristics found in [10]. For example, the code developed by Balas and Niehaus, which finds large stable sets using the

GRAPH				MAX OVERALL			AVERAGE SIZE			AVERAGE TIME		
name	$\|V\|$	$\|E\|$	α	h11	h21	h25	h11	h21	h25	h11	h21	h25
MANN-a9.co	45	72	16	16	16	16	16	16	16	0	0	0
MANN-a27.co	378	702	126	118	125	126	100	124	126	0	3	24
MANN-a45.co	1035	1980	345	45	45	344	45	45	343	2	2	390
brock200-1.co	200	5066	21	20	21	21	20	21	21	0	2	13
brock200-2.co	200	10024	12	10	11	11	10	11	11	0	3	16
brock200-3.co	200	7852	15	13	14	14	13	13	14	0	2	13
brock200-4.co	200	6811	17	15	16	16	15	16	16	0	3	12
brock400-1.co	400	20077	27	22	24	25	22	24	24	2	12	75
brock400-2.co	400	20014	29	24	25	25	23	24	24	2	11	81
brock400-3.co	400	20119	31	24	25	25	22	24	24	2	17	65
brock400-4.co	400	20035	33	23	24	25	23	24	24	2	13	70
brock800-1.co	800	112095	23	20	21	21	19	20	20	12	104	490
brock800-2.co	800	111434	24	20	20	20	19	19	20	13	815	476
brock800-3.co	800	112267	25	19	21	21	19	20	20	14	92	476
brock800-4.co	800	111957	26	18	21	21	18	20	20	15	78	500
c-fat200-1.co	200	18366	12	12	12	12	12	12	12	0	1	5
c-fat200-2.co	200	16665	24	24	24	24	24	24	24	0	3	10
c-fat200-5.co	200	11427	58	58	58	58	58	58	58	0	1	9
c-fat500-1.co	500	120291	14	14	14	14	14	14	14	5	13	83
c-fat500-2.co	500	115611	26	26	26	26	26	26	26	3	13	81
c-fat500-5.co	500	101559	64	64	64	64	64	64	64	3	22	75
c-fat500-10.co	500	78123	126	126	126	126	126	126	126	3	18	68
hamming6-2.co	64	192	32	32	32	32	31	32	32	0	0	0
hamming6-4.co	64	1312	4	4	4	4	4	4	4	0	0	0
hamming8-2.co	256	1024	128	128	128	128	124	128	128	0	0	0
hamming8-4.co	256	11776	16	16	16	16	15	16	16	0	1	4
hamming10-2.co	1024	5120	512	512	512	512	512	512	512	2	3	4
hamming10-4.co	1024	89600	-1	40	40	40	34	35	36	47	185	643
johnson8-2-4.co	28	168	4	4	4	4	4	4	4	0	0	0
johnson8-4-4.co	70	560	14	14	14	14	14	14	14	0	0	0
johnson16-2-4.co	120	1680	8	8	8	8	8	8	8	0	0	0
johnson32-2-4.co	496	14880	16	16	16	16	14	16	16	11	2	12

Table 1: Results of ten runs of each heuristic on the first set of 32 graphs.

GRAPH				MAX OVERALL			AVERAGE SIZE			AVERAGE TIME		
name	$\|V\|$	$\|E\|$	α	h11	h21	h25	h11	h21	h25	h11	h21	h25
keller4.co	171	5100	11	7	11	11	7	10	11	0	3	10
keller5.co	776	74710	27	16	24	27	16	21	25	560	391	1306
p-hat300-1.co	300	33917	8	7	8	8	7	8	8	2	19	64
p-hat300-2.co	300	22922	25	25	25	25	25	25	25	1	22	66
p-hat300-3.co	300	11460	36	35	36	36	35	36	36	1	11	46
p-hat500-1.co	500	93181	9	9	9	9	8	9	9	11	55	192
p-hat500-2.co	500	61804	36	36	36	36	36	36	36	6	83	338
p-hat500-3.co	500	30950	-1	48	50	50	48	49	50	4	33	188
p-hat700-1.co	700	183651	11	9	11	11	8	9	10	28	139	1175
p-hat700-2.co	700	122922	-1	44	44	44	44	44	44	18	193	850
p-hat700-3.co	700	61640	-1	60	62	62	60	62	62	10	117	524
p-hat1000-1.co	1000	377247	-1	9	10	10	9	9	10	45	371	1410
p-hat1000-2.co	1000	254701	-1	45	46	46	45	46	46	45	599	2580
p-hat1000-3.co	1000	127754	-1	63	68	68	63	68	68	21	348	1445
p-hat1500-1.co	1500	839327	-1	10	11	11	10	10	10	204	2214	6237
p-hat1500-2.co	1500	555290	-1	64	65	65	64	65	65	148	1687	7098
p-hat1500-3.co	1500	277006	-1	93	94	94	93	93	94	81	1155	4917
san200-0.7-1.co	200	5970	30	15	30	30	15	24	30	0	1	2
san200-0.7-2.co	200	5970	18	12	18	18	12	15	18	0	2	12
san200-0.9-1.co	200	1990	70	70	70	70	67	70	70	0	0	0
san200-0.9-2.co	200	1990	60	36	60	60	35	60	60	0	0	0
san200-0.9-3.co	200	1990	44	44	44	44	39	40	44	0	1	5
san400-0.5-1.co	400	39900	13	7	9	13	6	8	11	10	22	68
san400-0.7-1.co	400	23940	40	20	40	40	19	22	40	226	72	17
san400-0.7-2.co	400	23940	30	15	19	30	15	16	30	3	11	51
san400-0.7-3.co	400	23940	22	12	18	22	12	15	18	3	185	98
san400-0.9-1.co	400	7980	100	52	100	100	52	100	100	1	2	2
san1000.co	1000	249000	15	8	9	10	7	7	9	391	3401	3090
sanr200-0.7.co	200	6032	18	17	18	18	17	17	18	0	4	18
sanr200-0.9.co	200	2037	-1	41	42	42	41	41	42	0	1	8
sanr400-0.5.co	400	39816	13	12	13	13	12	12	13	2	22	102
sanr400-0.7.co	400	23931	-1	20	21	21	20	21	21	2	17	86

Table 2: Results of ten runs of each heuristic on the second set of 32 graphs.

instance	r100.5	r200.5	r300.5	r400.5	r500.5
time	0.00	0.16	1.36	8.30	31.57

Table 3: DIMACS processor speed benchmarks (in seconds)

idea of maximum matchings in bipartite subgraphs of G, achieves the same size stable set as our heuristic h25 on essentially all test instances.

In addition, as mentioned in the introduction, Benson and Ye [1] have developed and implemented a heuristic for finding stable sets based on solving a different SDP relaxation of the MSS problem. In their paper, they report the size of the best stable set found by their heuristic on twelve of the graphs that are listed in Tables 2 and 3. We remark that our heuristic performed at least as well as theirs on all twelve instances. In particular, our heuristic h25 found stable sets which matched theirs on nine of the twelve instances and exceeded theirs on the remaining three (sanr200-0.7, 18 versus 11; sanr200-0.9, 42 versus 34; brock200-1, 21 versus 14).

5 Final Remarks

In this paper, we have extended the path laid in [5] by providing yet another example in which low-rank, nonconvex formulations serve as efficient tools for obtaining high-quality approximate (and often exact) solutions to NP-hard combinatorial optimization problems. As is the case in [5] with the Max-Cut problem, our experimental results with the maximum stable set problem indicate that the semidefinite program (1), or its equivalents, is unlikely to be an effective vehicle for finding stable sets in the graph G because of the high computational costs associated with solving such a semidefinite program. Instead, the rank-one and rank-two formulations are more attractive alternatives for that task. We stress, however, that the upper bound $\vartheta(G)$ that (1) provides on $\alpha(G)$ can be highly valuable in its own right as mentioned in the introduction.

It is worth noting that the ideas developed in this paper extend in a straightforward manner to the maximum weight stable set problem. In particular, if each node i in the graph G has an associated weight $w_i > 0$, then the problem of finding a stable set with maximum total weight on its nodes can be formulated as (3) and (4) by simply replacing the vector e by the vector $(\sqrt{w_1}, \ldots, \sqrt{w_n})$.

The formulations and heuristics detailed in this paper add novel techniques to those already available for solving or approximating the MSS problem. We believe that these techniques are of particular interest because they are among a handful of approaches based on continuous formulations of the MSS problem. Nonetheless, there are still many interesting avenues for further improvement. For example, can we devise a more efficient local optimization method than the augmented Lagrangian method? Can we obtain better heuristics by combining the continuous heuristics with some discrete heuristics? Is it possible to escape more reliably from the saddle points corresponding to sub-optimal stable sets? We believe that these questions deserve further investigations.

References

[1] S. Benson and Y. Ye. Approximating Maximum Stable Set and Minimum Graph Coloring Problems with the Positive Semidefinite Relaxation. Working paper. Department of Management Sciences, College of Business Administration, The University of Iowa, Iowa City, Iowa, U.S.A.

[2] S. Burer, R. D. C. Monteiro and Y. Zhang. Solving semidefinite programs via nonlinear programming. Part I: Transformations and derivatives. Technical Report TR99-17, Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77005, USA, September 1999.

[3] S. Burer, R. D. C. Monteiro and Y. Zhang. Solving semidefinite programs via nonlinear programming. Part II: Interior point methods for a subclass of SDPs. Technical Report TR99-23, Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77005, USA, October 1999.

[4] S. Burer, R. D. C. Monteiro and Y. Zhang. Interior-point algorithms for semidefinite programming based on a nonlinear programming formulation. Technical Report TR99-27, Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77005, USA, December 1999.

[5] S. Burer, R. D. C. Monteiro and Y. Zhang. Rank-two relaxation heuristics for max-cut and other binary quadratic programs. Manuscript, School of Industrial and Systems Engineering, Atlanta, GA, USA, November 2000.

[6] See the website: `http://dimacs.rutgers.edu/Challenges/`.

[7] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of ACM*, 42:1115–1145, 1995.

[8] M. Grötschel, L. Lovász and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Springer-Verlag, 1988.

[9] C. Helmberg and F. Rendl. An spectral bundle method for semidefinite programming. *SIAM Journal on Optimization*, 10:673–696, 2000.

[10] D. Johnson and M. Trick, eds. *Cliques, Coloring, and Satisfiability: Second DIMACS Implementation Challenge*. AMS, Providence, RI, 1996.

[11] L. Lovász. On the Shannon capacity of a graph. *IEEE Transactions of Information Theory*, IT-25(1):1–7, January 1979.

[12] J. Nocedal and S. Wright. *Numerical Optimization*. Springer-Verlag, 1999.

[13] P. Pardalos and J. Xue. The maximum clique problem. *Journal of Global Optimization*, 4:301–328, 1994.

[14] H. Wolkowicz, R. Saigal and L. Vandenberghe, eds. *Handbook of Semidefinite Programming: Theory, Algorithms and Applications*. Kluwer, 2000.

A Proof of Proposition 3.5

For notational convenience, we also let the variable Z collectively denote the variables in (P_{V_0}), that is, Z represents $(z^1, \ldots, z^n)$, and we let $f(Z)$ denote the objective value at Z. Our first result establishes some characteristics of every local maximum of (P_{V_0}).

Lemma A.1 *Let $\bar{Z} \equiv (\bar{z}^1, \ldots, \bar{z}^n)$ be a local maximum of (P_{V_0}). Then $f(\bar{Z}) > 0$, and for each $k = 1, \ldots, p$, there exist $\bar{z}^{ak}$ and $\bar{z}^{bk}$, at least one of which is nonzero, such that $\bar{z}^i = \bar{z}^{ak}$ for all $i \in A_k$ and $\bar{z}^j = \bar{z}^{bk}$ for all $j \in B_k$.*

Proof. It is not difficult to see that $\bar{f} \equiv f(\bar{Z}) > 0$ since otherwise $\bar{Z}$ would both be a local maximum and a global minimum of (P_{V_0}). Using this fact, we proceed to prove the second statement of the lemma.

We first show that, for each $k = 1, \ldots, p$, there exist $\bar{z}^{ak}$ and $\bar{z}^{bk}$ such that $\bar{z}^i = \bar{z}^{ak}$ for all $i \in A_k$ and $\bar{z}^j = \bar{z}^{bk}$ for all $j \in B_k$. Note that this claim is obviously true when $b_k = 0$ since then $a_k = 1$. So assume that $b_k > 0$. Since $\bar{Z}$ is a local maximum, the point $(\bar{z}^i)_{i \in C_k}$ is a local maximum of the problem

$$\max \left\{ \left\| \sum_{i \in C_k} z^i + \hat{z} \right\|^2 : z^i \perp z^j \ \forall \, (i,j) \in A_k \times B_k, \ \sum_{i \in C_k} \|z^i\|^2 = \delta \right\}, \tag{29}$$

where $C_k \equiv A_k \cup B_k$, $\hat{z} \equiv \sum_{i \notin C_k} \bar{z}^i$ and $\delta \equiv 1 - \sum_{i \notin C_k} \|\bar{z}^i\|^2$. We now examine (29) in the three cases in which the span of $\{\bar{z}^i : i \in A_k\}$ is either zero-, one- or two-dimensional.

If the span is zero-dimensional, then $\bar{z}^i = 0$ for all $i \in A_k$, in which case $(\bar{z}^j)_{j \in B_k}$ is a local maximum of the problem

$$\max \left\{ \left\| \sum_{j \in B_k} z^j + \hat{z} \right\|^2 : \sum_{j \in B_k} \|z^j\|^2 = \delta \right\}. \tag{30}$$

Analyzing the first-order necessary conditions of (30) at $(\bar{z}^j)_{j \in B_k}$, we see that there exists $\lambda \in \Re$ such that $\sum_{j \in B_k} \bar{z}^j + \hat{z} = \lambda \bar{z}^j$ for all $j \in B_k$. Suppose that $\lambda = 0$. Then it must be the case that $\sum_{j \in B_k} \bar{z}^j + \hat{z} = 0$, in which case the objective value of (30) is 0, but this contradicts the fact that $\bar{f} > 0$. Hence, $\lambda \neq 0$ from which it follows from the first-order conditions that $\bar{z}^j$ is constant over $j \in B_k$.

If the span is two-dimensional, then $\bar{z}^j = 0$ for all $j \in B_k$. Hence, a similar argument as in the zero-dimensional case shows that $\bar{z}^i$ is constant over all $i \in A_k$.

If the span is one-dimensional, then there exist perpendicular unit vectors $w^a, w^b \in \Re^2$ and scalars $\bar{\gamma}_i$ for all $i \in C_k$ such that $\bar{z}^i = \bar{\gamma}_i w^a$ for all $i \in A_k$ and $\bar{z}^j = \bar{\gamma}_j w^b$ for all $j \in B_k$. It then follows that $(\bar{\gamma}_i)_{i \in C_k}$ is a local maximum of the problem that originates from (29) by adding the extra constraints $z^i = \gamma_i w^a$ for all $i \in A_k$ and $z^j = \gamma_j w^b$ for all $j \in B_k$, or equivalently, the following maximization problem whose variables are $(\gamma_i)_{i \in C_k}$:

$$\max \left\{ \left\| \sum_{i \in A_k} \gamma_i w^a + \sum_{j \in B_k} \gamma_j w^b + \hat{z} \right\|^2 : \sum_{i \in A_k} \gamma_i^2 + \sum_{j \in B_k} \gamma_j^2 = \delta \right\}. \tag{31}$$

Analyzing the first-order necessary conditions of the above maximization, we see that there exists $\lambda \in \Re$ such that

$$
\begin{aligned}
(w^a)^T \left(\sum_{i \in A_k} \bar{\gamma}_i w^a + \sum_{j \in B_k} \bar{\gamma}_j w^b + \hat{z} \right) &= \lambda \bar{\gamma}_i \quad \forall\, i \in A_k, \\
\left(w^b\right)^T \left(\sum_{i \in A_k} \bar{\gamma}_i w^a + \sum_{j \in B_k} \bar{\gamma}_j w^b + \hat{z} \right) &= \lambda \bar{\gamma}_j \quad \forall\, j \in B_k.
\end{aligned}
$$

Suppose $\lambda = 0$. Then the vector $\sum_{i \in A_k} \bar{\gamma}_i w^a + \sum_{j \in B_k} \bar{\gamma}_j w^b + \hat{z}$ is perpendicular to both w^a and w^b. Since w^a and w^b are themselves perpendicular nonzero vectors in $\Re^2$, it follows that $\sum_{i \in A_k} \bar{\gamma}_i w^a + \sum_{j \in B_k} \bar{\gamma}_j w^b + \hat{z} = 0$ in which case the objective value of (31) at $(\bar{\gamma}_i)_{i \in C_k}$ is 0. This contradicts the fact that $\bar{f} > 0$, and so we conclude in fact that $\lambda \neq 0$ from which it easily follows that $\bar{\gamma}_i$ is constant over $i \in A_k$ and $\bar{\gamma}_j$ is constant over $j \in B_k$.

Hence, in each of the three cases pertaining to the assumption that $b_k > 0$, we have shown that $\bar{z}^i$ is constant over $i \in A_k$ and $\bar{z}^j$ is constant over $j \in B_k$, and so the first claim above follows.

We now show the claim that for all $k = 1, \ldots, p$, at least one of the vectors $\bar{z}^{ak}$ or $\bar{z}^{bk}$ is nonzero. Assume for contradiction that $\bar{z}^{al} = \bar{z}^{bl} = 0$ for some $l \in \{1, \ldots, p\}$. Define

$$
u \equiv \sum_{k \neq l} (a_k \bar{z}^{ak} + b_k \bar{z}^{bk}) \tag{32}
$$

and note that $\bar{f} \equiv f(\bar{Z}) = \|u\|^2$. Let r be a unit-length vector such that $u = \|u\| r$ and define the path $Z : [0, 1/\sqrt{a_l}] \to \Re^{2n}$ whose two-dimensional vector components are:

$$
z^i(t) = \begin{cases} tr, & \text{if } i \in A_l \\ 0, & \text{if } i \in B_l \end{cases}, \qquad z^i(t) = \begin{cases} \sqrt{1 - a_l t^2}\ \bar{z}^{ak}, & \text{if } i \in A_k,\ k \neq l \\ \sqrt{1 - a_l t^2}\ \bar{z}^{bk}, & \text{if } i \in B_k,\ k \neq l. \end{cases}
$$

Using the fact that $\bar{z}^i = 0$ for all $i \in C_l$, and hence that $\sum_{i \notin C_l} \|\bar{z}^i\|^2 = 1$, and the above relations, we see by means of a simple verification that $Z(t)$ is feasible for all $t \in [0, 1/\sqrt{a_l}]$. Also,

$$
\begin{aligned}
f(Z(t)) &= \left\| \sum_{k=1}^{p} \left(a_k z^{ak}(t) + b_k z^{bk}(t) \right) \right\|^2 \\
&= \left\| \sqrt{1 - a_l t^2}\ \sum_{k \neq l} \left(a_k \bar{z}^{ak} + b_k \bar{z}^{bk} \right) + a_l t r \right\|^2 \\
&= \left\| \sqrt{1 - a_l t^2}\ u + a_l t r \right\|^2 = \left\| \sqrt{1 - a_l t^2}\ \|u\| r + a_l t r \right\|^2 \\
&= \left(\sqrt{(1 - a_l t^2) \bar{f}} + a_l t \right)^2.
\end{aligned}
$$

A simple calculation reveals that the derivative of $f(Z(t))$ is positive for all $t \in (0, 1/\sqrt{\bar{f}+a_l})$. This implies that $\bar{Z} = Z(0)$ is not a local maximum of (29), which gives the required contradiction. ∎

In view of the above lemma, when we identify z^i with $i \in A_k$ (respectively, $i \in B_k$) with a single variable z^{ak} (respectively, z^{bk}), any local maximum of problem (P_{V_0}) becomes a local maximum of the following problem

$$\begin{array}{ll} \text{maximize} & g(Z^{ab}) \equiv \left\| \sum_{k=1}^{p} \left(a_k z^{ak} + b_k z^{bk} \right) \right\|^2 \\ \text{subject to} & \sum_{k=1}^{p} \left(a_k \|z^{ak}\|^2 + b_k \|z^{bk}\|^2 \right) = 1, \\ & z^{ak} \perp z^{bk}, \; \forall\, k = 1, \ldots, p, \end{array} \tag{33}$$

whose variables are $Z^{ab} \equiv (z^{a1}, z^{b1}, \ldots, z^{ap}, z^{bp})$.

Lemma A.2 *Let $Z^{ab} = (z^{a1}, z^{b1}, \ldots, z^{ap}, z^{bp}) \in \Re^{2p}$ be a point such that for all $k = 1, \ldots, p$:*

$$z^{ak} + z^{bk} = \gamma_k w, \qquad z^{ak} \perp z^{bk}, \tag{34}$$

$$a_k > b_k \implies \bar{z}^{ak} = 0 \text{ or } \bar{z}^{bk} = 0, \tag{35}$$

where $w \in \Re^2$ is a vector of unit length and $\gamma_1, \ldots, \gamma_p$ are scalars. Let

$$\mathcal{K} = \mathcal{K}(Z^{ab}) \equiv \{k : a_k > b_k,\ \bar{z}^{bk} \neq 0\}. \tag{36}$$

Then,

$$\sum_{i=1}^{p} (a_k z^{ak} + b_k z^{bk}) = \left(\sum_{k \notin \mathcal{K}} \gamma_k a_k + \sum_{k \in \mathcal{K}} \gamma_k b_k \right) w, \tag{37}$$

$$g(Z^{ab}) = \left(\sum_{k \notin \mathcal{K}} \gamma_k a_k + \sum_{k \in \mathcal{K}} \gamma_k b_k \right)^2. \tag{38}$$

Moreover, Z^{ab} is a feasible solution for problem (33) if and only if

$$\sum_{k \notin \mathcal{K}} a_k \gamma_k^2 + \sum_{k \in \mathcal{K}} b_k \gamma_k^2 = 1.$$

Proof. Using (34), (35) and the definition of $\mathcal{K}$, it is easy to see that for all $k = 1 \ldots, p$,

$$a_k \|z^{ak}\|^2 + b_k \|z^{bk}\|^2 = \begin{cases} a_k \|z^{ak} + z^{bk}\|^2, & \text{if } k \notin \mathcal{K}, \\ b_k \|z^{ak} + z^{bk}\|^2, & \text{if } k \in \mathcal{K}, \end{cases}$$

$$a_k z^{ak} + b_k z^{bk} = \begin{cases} a_k (z^{ak} + z^{bk}), & \text{if } k \notin \mathcal{K}, \\ b_k (z^{ak} + z^{bk}), & \text{if } k \in \mathcal{K}. \end{cases}$$

Hence, we have

$$\begin{aligned}
\sum_{i=1}^{p} a_k\|z^{ak}\|^2 + b_k\|z^{bk}\|^2 &= \sum_{k\notin\mathcal{K}} a_k\|z^{ak}+z^{bk}\|^2 + \sum_{k\in\mathcal{K}} b_k\|z^{ak}+z^{bk}\|^2 \\
&= \sum_{k\notin\mathcal{K}} a_k\|\gamma_k w\|^2 + \sum_{k\in\mathcal{K}} b_k\|\gamma_k w\|^2 \\
&= \sum_{k\notin\mathcal{K}} a_k\gamma_k^2 + \sum_{k\in\mathcal{K}} b_k\gamma_k^2
\end{aligned}$$

and

$$\sum_{i=1}^{p}(a_k z^{ak} + b_k z^{bk}) = \sum_{k\notin\mathcal{K}} a_k(z^{ak}+z^{bk}) + \sum_{k\in\mathcal{K}} b_k(z^{ak}+z^{bk}) = \left(\sum_{k\notin\mathcal{K}} a_k\gamma_k + \sum_{k\in\mathcal{K}} b_k\gamma_k\right) w.$$

The expression for $g(Z^{ab})$ and the if and only if statement of the lemma now follow immediately from these two identities. ∎

Now we give the proof of Proposition 3.5.

Proof. Let $\bar{Z} \equiv (\bar{z}^1, \ldots, \bar{z}^n)$ be a local maximum of (P_{V_0}). We will show that $\bar{f} \equiv f(\bar{Z}) = a_1 + \ldots + a_p$, from which the proposition immediately follows due to Lemma 3.9. Define the point $\bar{Z}^{ab} \equiv (\bar{z}^{a1}, \bar{z}^{b1}, \ldots, \bar{z}^{ap}, \bar{z}^{bp})$ whose two-dimensional vector components satisfy the conditions of Lemma A.1. Using the fact that $\bar{Z}$ is a local maximum of (P_{V_0}), one can easily verify that $\bar{Z}^{ab}$ is a local maximum of problem (33). Now define $\bar{u} \equiv \sum_{k=1}^{p}\left(a_k\bar{z}^{ak} + b_k\bar{z}^{bk}\right)$ and note that $\bar{f} = g(\bar{Z}^{ab}) = \|\bar{u}\|^2$. The first-order optimality conditions that hold at $\bar{Z}^{ab}$ are:

$$a_k\bar{u} = \lambda a_k z^{ak} + \lambda_k z^{bk}, \tag{39}$$

$$b_k\bar{u} = \lambda b_k z^{bk} + \lambda_k z^{ak}, \tag{40}$$

for all $k = 1, \ldots, p$, where $\lambda, \lambda_1, \ldots, \lambda_p \in \Re$ are the Lagrange multipliers. (The constraint qualification that the gradients of the constraints of (33) at the point $\bar{Z}^{ab}$ are linearly independent can be easily verified.) By taking the dot product of (39) and (40) with $\bar{z}^{ak}$ and $\bar{z}^{bk}$, respectively, then adding the two resultant equations, and finally summing all such equations over the index k, we easily obtain that $\lambda = \bar{f}$. Hence, in what follows, we replace λ by $\bar{f}$. Now multiplying (39) by b_k and (40) by a_k, and subtracting the two resulting equations, we obtain

$$\bar{f}a_k b_k\left(\bar{z}^{ak} - \bar{z}^{bk}\right) = \lambda_k\left(a_k\bar{z}^{ak} - b_k\bar{z}^{bk}\right).$$

Taking the dot product of the last equation with $\bar{z}^{ak}$ and $\bar{z}^{bk}$, and using the fact that $\bar{z}^{ak} \perp \bar{z}^{bk}$, we obtain respectively that

$$a_k(b_k\bar{f} - \lambda_k)\|\bar{z}^{ak}\|^2 = 0, \tag{41}$$

$$b_k(a_k\bar{f} - \lambda_k)\|\bar{z}^{bk}\|^2 = 0. \tag{42}$$

By Lemma A.1, we know that for all $k = 1, \ldots, p$, at least one of the vectors $\bar{z}^{ak}$ and $\bar{z}^{bk}$ is nonzero. If $\bar{z}^{ak} \neq 0$ then by (41) we have $\lambda_k = b_k\bar{f}$, which together with (40) implies that

$\bar z^{ak} + \bar z^{bk} = \bar u/\bar f$. In a similar manner, using (42) and (39) we conclude that if $\bar z^{bk} \neq 0$ and $b_k > 0$ then we have $\lambda_k = a_k \bar f$ and $\bar z^{ak} + \bar z^{bk} = \bar u/\bar f$. These two observations clearly imply that for all $k = 1, \ldots, p$, we have

$$\begin{aligned} \bar z^{ak} + \bar z^{bk} &= \frac{1}{\bar f^{1/2}} \frac{\bar u}{\|\bar u\|}, \\ a_k > b_k > 0 &\implies \text{ either } \bar z^{ak} = 0 \text{ or } \bar z^{bk} = 0. \end{aligned}$$

Therefore, letting $\bar{\mathcal{K}} \equiv \{k : a_k > b_k > 0,\ \bar z_b^k \neq 0\}$, it follows from Lemma A.2 with $\gamma_k = \bar f^{-1/2}$ for all $k = 1, \ldots, p$ that

$$\bar f = \sum_{k \notin \mathcal{K}} a_k + \sum_{k \in \mathcal{K}} b_k. \tag{43}$$

To conclude the proof of the proposition, it remains to show that $\mathcal{K} = \emptyset$. Indeed, assume for contradiction that $\mathcal{K} \neq \emptyset$ and note that in this case $\bar f < a$. Consider the point $\hat Z^{ab} = (\hat z^{a1}, \hat z^{b1}, \ldots, \hat z^{ap}, \hat z^{bp})$ defined as

$$\hat z^{ak} = \begin{cases} (\bar f/a)^{1/2} Q \bar z^{ak}, & \text{if } k \notin \mathcal{K} \\ (a \bar f)^{-1/2} Q \bar u, & \text{if } k \in \mathcal{K} \end{cases}, \qquad \hat z^{bk} = \begin{cases} (\bar f/a)^{1/2} Q \bar z^{bk}, & \text{if } k \notin \mathcal{K} \\ 0, & \text{if } k \in \mathcal{K} \end{cases},$$

where $Q \in \Re^{2\times 2}$ is defined as

$$Q \equiv \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

It is a simple verification to see that, for all $k = 1, \ldots, p$, we have $\hat z^{ak} \perp \hat z^{bk}$ and $\hat z^{ak} + \hat z^{bk} = w/\sqrt{a}$, where w is the unit length vector $w \equiv Q\bar u/\|Q \bar u\| = Q\bar u/\bar f^{1/2}$. Moreover, we have that $\hat{\mathcal{K}} \equiv \mathcal{K}(\hat Z^{ab}) = \{1, \ldots, p\}$, and so it follows that $\hat Z^{ab}$ is a feasible solution of (33). Using all these facts along with Lemma A.2, we conclude that $g(\hat Z^{ab}) = a$. Consider now the path $Z^{ab} : [0,1] \to \Re^{2p}$ defined as

$$Z^{ab}(t) \equiv \psi(t) \left[(1-t) \bar Z^{ab} + t \hat Z^{ab} \right]$$

for all $t \in [0,1]$, where $\psi(t) \equiv [(1-t)^2 + t^2]^{-1/2}$. We claim that $Z^{ab}(t)$ is feasible for problem (33) for all $t \in [0,1]$. Indeed, writing $Z^{ab}(t) = (z^{a1}(t), z^{b1}(t), \ldots, z^{ap}(t), z^{bp}(t))$ and noting that $\hat z^{ak} \perp \hat z^{bk}$ for all $k = 1, \ldots, p$, we have for every $k \notin \mathcal{K}$ that

$$\begin{aligned} \left(z^{ak}(t)\right)^T z^{bk}(t) &= \psi(t)^2 \left((1-t)\bar z^{ak} + t \hat z^{ak}\right)^T \left((1-t)\bar z^{bk} + t\hat z^{bk}\right) \\ &= \psi(t)^2 t(1-t) \left(\left(\bar z^{ak}\right)^T \hat z^{bk} + \left(\hat z^{ak}\right)^T \bar z^{bk} \right) \\ &= \psi(t)^2 t (1-t) \left(\frac{\bar f}{a}\right)^{1/2} \left(\bar z^{ak}\right)^T \left(Q + Q^T\right) \bar z^{bk} \ = \ 0, \end{aligned}$$

since $Q + Q^T = 0$. Noting that $\bar z^{bk} = \bar u/\bar f$ for $k \in \mathcal{K}$ and using the relation $\bar u^T Q \bar u = 0$, we easily see that the above inner product is also zero when $k \in \mathcal{K}$. Hence, it follows that

$z^{ak}(t) \perp z^{bk}(t)$ for all $t \in [0,1]$. We also have

$$\begin{aligned}
&\sum_{k=1}^{p} \left(a_k \|z^{ak}(t)\|^2 + b_k \|z^{bk}(t)\|^2 \right) \\
&= \psi(t)^2 \sum_{k=1}^{p} \left(a_k \|(1-t)\bar{z}^{ak} + t\hat{z}^{ak}\|^2 + b_k \|(1-t)\bar{z}^{bk} + t\hat{z}^{bk}\|^2 \right) \\
&= \psi(t)^2 \sum_{k=1}^{p} \left(a_k \left[(1-t)^2 \|\bar{z}^{ak}\|^2 + t^2 \|\hat{z}^{ak}\|^2 \right] + b_k \left[(1-t)^2 \|\bar{z}^{bk}\|^2 + t^2 \|\hat{z}^{bk}\|^2 \right] \right) \\
&= \psi(t)^2 \left[(1-t)^2 \sum_{k=1}^{p} \left(a_k \|\bar{z}^{ak}\|^2 + b_k \|\bar{z}^{bk}\|^2 \right) + t^2 \sum_{k=1}^{p} \left(a_k \|\hat{z}^{ak}\|^2 + b_k \|\hat{z}^{bk}\|^2 \right) \right] \\
&= \psi(t)^2 \left[(1-t)^2 + t^2 \right] \; = \; 1.
\end{aligned}$$

We have thus established the feasibility of $Z^{ab}(t)$ for all $t \in [0,1]$. We now claim that $g(Z^{ab}(t))$ is a strictly increasing function of $t \in [0,1]$. Note that this claim implies that $\bar{Z}^{ab} = Z^{ab}(0)$ is not a local maximum of (33), thereby giving the desired contradiction. To prove the claim, note that

$$\begin{aligned}
g(Z^{ab}(t)) &= \left\| \sum_{k=1}^{p} \left(a_k z^{ak}(t) + b_k z^{bk}(t) \right) \right\|^2 \\
&= \psi(t)^2 \left\| \sum_{k=1}^{p} \left(a_k \left((1-t)\bar{z}^{ak} + t\hat{z}^{ak} \right) + b_k \left((1-t)\bar{z}^{bk} + t\hat{z}^{bk} \right) \right) \right\|^2 \\
&= \psi(t)^2 \left\| (1-t) \sum_{k=1}^{p} \left(a_k \bar{z}^{ak} + b_k \bar{z}^{bk} \right) + t \sum_{k=1}^{p} \left(a_k \hat{z}^{ak} + b_k \hat{z}^{bk} \right) \right\|^2 \\
&= \psi(t)^2 \left[(1-t)^2 \left\| \sum_{k=1}^{p} \left(a_k \bar{z}^{ak} + b_k \bar{z}^{bk} \right) \right\|^2 + t^2 \left\| \sum_{k=1}^{p} \left(a_k \hat{z}^{ak} + b_k \hat{z}^{bk} \right) \right\|^2 \right] \\
&= \frac{(1-t)^2 \bar{f} + t^2 a}{(1-t)^2 + t^2}.
\end{aligned}$$

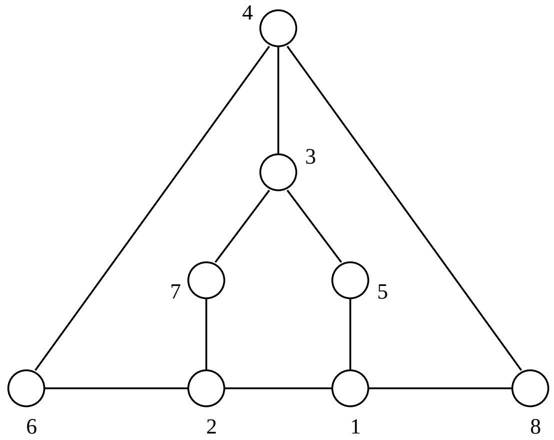
The fourth equality follows from the orthogonality between the vectors $\sum_{k=1}^{p} \left(a_k \bar{z}^{ak} + b_k \bar{z}^{bk} \right)$ and $\sum_{k=1}^{p} \left(a_k \hat{z}^{ak} + b_k \hat{z}^{bk} \right)$, which is guaranteed by Lemma A.2 and the fact that $\bar{u}$ is orthogonal to $Q\bar{u}$. ■

B An Example

We provide an example in which (x, y) is a local maximum in (4) but the induced stable set A is not maximal in G. Consider the example having $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$ and

$$E = \{(1, 2), (1, 5), (1, 8), (2, 6), (2, 7), (3, 4), (3, 5), (3, 7), (4, 6), (4, 8)\},$$

i.e., G is the following graph:



Let $w \in \Re^2$ be any vector of unit length, and consider any feasible (x, y) satisfying

$$(x_1, y_1) + (x_2, y_2) = (x_3, y_3) + (x_4, y_4) = \frac{1}{\sqrt{2}} w,$$
$$(x_5, y_5) = (x_6, y_6) = (x_7, y_7) = (x_8, y_8) = 0,$$

and the following: (i) none of x_i, y_i, $i = 1, 2, 3, 4$, is zero; and (ii) $x_i y_j \neq x_j y_i$ for all non-edges (i, j) in the collection $(1, 3), (2, 4), (2, 3), (1, 4)$. For example, one could take $w = (1/\sqrt{2}, 1/\sqrt{2})$ and

$$(x_1, y_1) = \left(\frac{1}{4}, \frac{1}{4} - \frac{1}{2\sqrt{2}}\right), \qquad (x_2, y_2) = \left(\frac{1}{4}, \frac{1}{4} + \frac{1}{2\sqrt{2}}\right),$$
$$(x_3, y_3) = \left(\frac{1}{4} + \frac{1}{2\sqrt{2}}, \frac{1}{4}\right), \qquad (x_4, y_4) = \left(\frac{1}{4} - \frac{1}{2\sqrt{2}}, \frac{1}{4}\right).$$

In such a case, it can be easily seen that $f(x, y) = 2$. In addition, $V_0 = \{5, 6, 7, 8\}$, which implies $p = 2$, $A_1 = \{1\}$, $B_1 = \{2\}$, $A_2 = \{3\}$ and $B_2 = \{4\}$.

By continuity, any nearby feasible point $(\hat{x}, \hat{y})$ satisfies the following: (a) none of $\hat{x}_i, \hat{y}_i$, $i = 1, 2, 3, 4$, is zero; and (b) $\hat{x}_i \hat{y}_j \neq \hat{x}_j \hat{y}_i$ for all non-edges (i, j) in the collection $(1, 3), (2, 4), (2, 3)$, and $(1, 4)$. In fact, because of items (a) and (b) as well as the orthogonality constraints on the edges in the set $E \setminus \{(1, 2), (3, 4)\}$, it is straightforward to see that $(\hat{x}_5, \hat{y}_5)$, $(\hat{x}_6, \hat{y}_6)$, $(\hat{x}_7, \hat{y}_7)$ and $(\hat{x}_8, \hat{y}_8)$ must all equal 0. Thus, $(\hat{x}, \hat{y})$ gives rise to the same set V_0 as (x, y), and so $f(\hat{x}, \hat{y}) \leq |A_1| + |A_2| = 2$. This implies that $f(\hat{x}, \hat{y}) \leq f(x, y)$ for all nearby feasible points $(\hat{x}, \hat{y})$. We conclude that (x, y) is a local maximum. Even though (x, y) is a local maximum, however, its induced stable set $A = A_1 \cup A_2$ is not maximal in G. In fact, none of its alternate induced stable sets — $A_1 \cup A_2$, $A_1 \cup B_2$, $B_1 \cup A_2$ or $B_1 \cup B_2$ — are maximal.