On the Lovász Theta Function and Some Variants

Laura Galli∗       Adam N. Letchford†

October 2015

Abstract

The Lovász theta function of a graph is a well-known upper bound on the stability number. It can be computed efficiently by solving a semidefinite program (SDP). Actually, one can solve either of two SDPs, one due to Lovász and the other to Grötschel et al. The former SDP is often thought to be preferable computationally, since it has fewer variables and constraints. We derive some new results on these two equivalent SDPs. The surprising result is that, if we weaken the SDPs by aggregating constraints, or strengthen them by adding cutting planes, the equivalence breaks down. In particular, the Grötschel et al. scheme typically yields a stronger bound than the Lovász one.

Key Words: stable set problem, semidefinite programming.

1 Introduction

Consider an undirected graph $G$ with vertex set $V$ and edge set $E$. A set of pairwise non-adjacent (respectively, adjacent) vertices is called a stable or independent set (resp. clique). Given a graph $G$, the maximum cardinality of a stable set (resp. clique) in $G$ is called the stability number (resp. clique number) and denoted by $\alpha(G)$ (resp. $\omega(G)$). The problem of determining $\alpha(G)$ (resp. $\omega(G)$) is called the maximum stable set problem or MSSP (resp. maximum clique problem or MCP).

The complement of $G$, denoted by $\bar{G}$, is the graph with vertex set $V$ and edge set

$$\bar{E} = \{\{i,j\} \subseteq V : \{i,j\} \notin E\}.$$

Since $\alpha(G) = \omega(\bar{G})$, the MSSP and MCP are equivalent. They have a wide range of applications in operational research and computer science. Unfortunately, they are NP-hard in the strong sense, and hard even to approximate (Hastad [13]).

∗Dipartimento di Informatica, Università di Pisa, Largo B. Pontecorvo 3, 56127 Pisa, Italy. E-mail: laura.galli@unipi.it
†Department of Management Science, Lancaster University, Lancaster LA1 4YW, United Kingdom. E-mail: A.N.Letchford@lancaster.ac.uk
In his seminal paper, Lovász [16] defined the so-called theta function of a graph $G$, denoted by $\theta(G)$. He then proved that $\alpha(G) \leq \theta(G) \leq \chi(\overline{G})$, where $\chi(\overline{G})$ denotes the chromatic number of $G$. Grötschel et al. [10] showed that $\theta(G)$ can be computed in polynomial time (to arbitrary fixed precision) by solving a semidefinite program (SDP). This result is remarkable, given that computing the chromatic number of a graph, or even approximating it, is also $\mathcal{NP}$-hard in the strong sense (Feige & Killian [6]).

In fact, there are two alternative SDP characterisations of $\theta(G)$ [11, 17]. The one in [17], having fewer variables and constraints, seems preferable computationally. Gruber & Rendl [12] showed how to map any feasible solution to the larger SDP onto a feasible solution of the smaller SDP with no smaller profit. Yildirim & Fan [24] gave a partial mapping in the reverse direction. In this paper, we give a complete mapping in both directions, and then extend it to the weighted case, in which each node has a positive weight and the goal is to find a maximum weight stable set.

One can define stronger variants of the theta function by adding cutting planes to the SDPs (see, e.g., [5, 7, 12, 17, 23]). We show that, perhaps surprisingly, the two SDPs behave differently in this context. In particular, adding a cutting plane to the smaller SDP can lead to less bound improvement than adding the analogous inequality to the larger SDP.

Going in the opposite direction, Lieder et al. [15] recently proposed to weaken the theta function, by aggregating equations in the larger SDP. We show that here, too, the two SDPs behave differently.

The paper is structured as follows. Section 2 is a literature review. Section 3 presents the results on the theta function and its weighted version. Section 4 presents the results on strengthened and weakened variants of the theta function. Concluding remarks are made in Section 5.

Throughout the paper, $n$ will denote $|V|$, $m$ will denote $|E|$, $e$ will denote the all-ones vector of dimension $n$ and $J = ee^T$ will denote the all-ones square matrix of order $n$. The trace of a square matrix $M$ will be denoted by $\text{Tr}(M)$. Given two symmetric matrices $A, B$ of the same order, $A \bullet B$ will denote $\text{Tr}(AB)$, the usual inner product of $A$ and $B$.

We will also use the following facts. A symmetric matrix $M \in \mathbb{R}^{n \times n}$ is positive semidefinite (psd) if $v^T M v \geq 0$ for all $v \in \mathbb{R}^n$. The set of psd matrices of order $n$ forms a convex cone in $\mathbb{R}^{n \times n}$, which is denoted by $S_n^+$. If $M$ is psd, there exists a matrix $Y \in \mathbb{R}^{n \times n}$ such that $M = Y^T Y$ (Cholesky factorisation). If we let $y_i$ denote the $i$th column of $Y$, then $M_{ij} = y_i \cdot y_j$ for all $i$ and $j$. The vectors $y_1, \ldots, y_n$ form the so-called Gram representation of $M$. Finally, an augmented matrix of the form

$$ \begin{pmatrix} 1 & v^T \\ v & M \end{pmatrix}, $$

where $v \in \mathbb{R}^n$ is psd if and only if the matrix $M - vv^T$ is psd (Schur complement).
2 Literature Review

Now we review the relevant literature. For simplicity of notation, we assume throughout this section that \( G \) contains no isolated nodes.

2.1 LP-based bounds

The standard 0-1 LP formulation of the MSSP is as follows (e.g., [11, 19]). For each \( i \in V \), let \( x_i \) be a binary variable, taking the value 1 if and only if \( i \) is in the stable set. Then:

\[
\begin{align*}
\max & \quad e^T x \\
\text{s.t.} & \quad x_i + x_j \leq 1 \quad (\{i,j\} \in E) \\
& \quad x_i \in \{0,1\} \quad (i \in V).
\end{align*}
\]

To strengthen the LP relaxation one can add valid linear inequalities, such as the following ones, due to Padberg [19]:

- **clique** inequalities, which take the form \( \sum_{i \in C} x_i \leq 1 \), where \( C \) is a maximal clique in \( G \);

- **odd hole** inequalities, which take the form \( \sum_{i \in H} x_i \leq \left\lfloor \frac{|H|}{2} \right\rfloor \), where \( H \) is a set of nodes that induce an ‘odd hole’ (chordless odd cycle) in \( G \).

Other known inequalities are surveyed, e.g., in [2, 7, 11, 17].

The upper bound obtained if one uses all clique (and non-negativity) inequalities is called the **fractional clique covering number** and denoted by \( \chi_f(\bar{G}) \). By definition, \( \alpha(G) \leq \chi_f(\bar{G}) \leq \chi(G) \). Unfortunately, computing \( \chi_f(\bar{G}) \) is also \( \mathcal{NP} \)-hard [11]. Nevertheless, reasonably good computational results have been obtained using LP relaxations (e.g., [1, 3, 8, 9, 18, 21, 22]).

2.2 The theta function

A rather different upper bound on \( \alpha(G) \) was derived by Lovász [16]. Suppose that a vector \( x \in \{0,1\}^n \) is feasible for the 0-1 LP formulation presented in the previous subsection. Define the matrix \( Z = xx^T/(e^T x) \in [0,1]^{n \times n} \). By definition, \( Z \) is psd and \( J \cdot Z = (e^T x)^2/(e^T x) = e^T x \). Moreover, \( \text{diag}(Z) = x/(e^T x) \), which implies that \( \text{Tr}(Z) = 1 \). This leads naturally to the following SDP relaxation of the MSSP, which we will call ‘\( \text{SDP}1 \)’:

\[
\begin{align*}
\max & \quad J \cdot Z \\
\text{s.t.} & \quad Z_{ij} = 0 \quad (\{i,j\} \in E) \\
& \quad \text{Tr}(Z) = 1 \\
& \quad Z \in S^n_+.
\end{align*}
\]
Lovász called the resulting upper bound $\theta(G)$, and proved that $\alpha(G) \leq \theta(G) \leq \chi(\bar{G})$. Nowadays, $\theta(G)$ is called the Lovász theta function.

Shortly after the publication of [16], it was shown that SDPs can be solved in polynomial time (to arbitrary fixed precision) [10]. Thus, $\theta(G)$ can be computed efficiently, at least in theory.

Another characterisation of $\theta(G)$ was derived by Grötschel et al. [11]. Given a feasible vector $x \in \{0, 1\}^n$ as before, consider the matrix $X = xx^T \in \{0, 1\}^{n \times n}$. By definition, $X$ is psd, its main diagonal is equal to $x$, and it has zero entries for all edges. Moreover, the augmented matrix

$$X^+ = \begin{pmatrix} 1 \\ x \\ x^T \\ X \end{pmatrix}$$

is also psd. This leads to the following SDP relaxation, which we call ‘SDP2’:

$$\max e^T x$$

s.t. 

$$X_{ii} = x_i \quad (i \in V)$$

$$X_{ij} = 0 \quad (\{i, j\} \in E)$$

$$(1 \quad x^T \quad X) \in S_+^{n+1}.$$ (8)

The resulting upper bound is again $\theta(G)$.

Note that SDP2 has $n+m$ linear constraints, whereas SDP1 has only $m+1$. On the other hand, SDP2 has a very nice property [11]: if one projects its feasible region into $x$-space, the resulting convex set satisfies all clique and non-negativity inequalities. This implies that

$$\alpha(G) \leq \theta(G) \leq \chi^f(\bar{G}) \leq \chi(\bar{G}),$$

with equality if $G$ is perfect.

The relationship between SDP1 and SDP2 is more complicated than it might appear. Gruber & Rendl [12] proved the following. Let $(x^*, X^*)$ be any feasible solution to SDP2 such that the profit, say $\gamma$, is positive. Then one can obtain a feasible solution to SDP1 whose profit is no smaller than $\gamma$ by setting $Z_{ij}$ to $X_{ij}^*/\gamma$ for all $i$ and $j$. Yildirim & Fan [24] proved a partial result in the other direction. Let $Z^*$ be any feasible solution to SDP1, again with positive profit $\gamma$. Then there exists a feasible solution $(x^*, X^*)$ to SDP2 whose profit is no smaller than $\gamma$, in which:

$$x^*_i = \begin{cases} \left(\sum_{j=1}^n Z^*_{ij}\right)^2 & \text{if } Z^*_{ii} > 0 \\ \gamma Z^*_{ii} & \text{if } Z^*_{ii} > 0 \\ 0 & \text{otherwise.} \end{cases}$$

We remark that a third SDP formulation of $\theta(G)$, with $n + |\bar{E}|$ constraints, was given by Dukanovic & Rendl [5]. This is preferable when $G$ is very dense.
2.3 The weighted theta function

Now consider the weighted version of the stable set problem, mentioned in the introduction. For each vertex $i \in V$, let $w_i > 0$ be the associated weight. Adapting SDP2 to this case is trivial: just change the objective function to $w^T x$. Grötschel et al. [10] showed that to adapt SDP1, it is necessary to replace the objective function with:

$$\sum_{i \in V} \sum_{j \in V} \sqrt{w_i w_j} Z_{ij}. \quad (9)$$

The resulting bound, the weighted theta function, is denoted by $\theta(G, w)$. A proof that both SDPs yield the same bound can be found in [11].

2.4 Variants of the theta function

As mentioned in the introduction, it is possible to strengthen both SDP1 and SDP2 by adding cutting planes (i.e., valid linear inequalities). Schrijver [23] proposed to strengthen SDP1 simply by adding the non-negativity inequalities:

$$Z_{ik} + Z_{jk} \leq Z_{kk} \quad (\forall \{i, j\} \in E, k \in V \setminus \{i, j\}) \quad (11)$$

$$x_i + x_j + x_k \leq 1 + X_{ik} + X_{jk} \quad (k \in V \setminus \{i, j\}). \quad (12)$$

As for SDP2, Lovász & Schrijver [17] proposed to add:

$$X_{ij} \geq 0 \quad (\forall \{i, j\} \in \bar{E}). \quad (13)$$

$$X_{ik} + X_{jk} \leq x_k \quad (\forall \{i, j\} \in E, k \neq i, j) \quad (14)$$

$$x_i + x_j + x_k \leq 1 + X_{ik} + X_{jk} \quad (\forall \text{ stable } \{i, j, k\}). \quad (15)$$

Gruber & Rendl [12] proposed to add, in addition:

$$X_{ik} + X_{jk} \leq x_k + X_{ij} \quad (\forall \text{ stable } \{i, j, k\}) \quad (16)$$

$$x_i + x_j + x_k \leq 1 + X_{ij} + X_{ik} + X_{jk} \quad (\forall \text{ stable } \{i, j, k\}). \quad (17)$$

For further strengthenings of SDP2, see, e.g., [7, 8, 14, 17]. For computational results, see, e.g., [1, 3, 5, 12]. Generally speaking, adding cutting planes usually yields some improvement in the upper bound, but at the cost of dramatically increased computing times.

Going in the opposite direction, Lieder et al. [15] recently proposed to weaken the theta function, as follows. Take SDP2 and replace the equations (7) with the single equation:

$$\sum_{\{i,j\} \in E} X_{ij} = 0. \quad (18)$$

The resulting bound can be computed much more quickly than $\theta(G)$, but, unfortunately, it is typically significantly weaker.
3 The Theta Function and its Weighted Version

In this section, we consider the theta function and the two associated SDPs. In Subsection 3.1, we show how to map a feasible $Z^*$ onto a feasible $X^*$. In Subsection 3.2, we extend the mappings in both directions to the weighted case.

3.1 The theta function

Yildirim & Fan [24] showed how to map a feasible matrix $Z^*$ to a feasible vector $x^*$ without losing any profit, but they did not show how to construct the associated matrix $X^*$. The following theorem shows how to do this.

Theorem 1 Let $Z^*$ be a feasible solution to SDP1, and let $\gamma = J \bullet Z^* > 0$ be the associated profit. Let $x^*$ be defined according to Yildirim and Fan. If $Z^*_{ij} = 0$, then set $X^*_{ij}$ to 0. Otherwise, set $X^*_{ij}$ to:

$$Z^*_{ij} \sqrt{x^*_i x^*_j} \frac{\sum_{k=1}^{n} Z^*_k}{Z^*_i Z^*_j} = \frac{(\sum_{k=1}^{n} Z^*_k)}{\gamma Z^*_i Z^*_j}.$$

Then the resulting pair $(x^*, X^*)$ forms a feasible solution to SDP2, and its profit, $e^T x^*$, is at least as large as $\gamma$.

Proof. Since the profit of $(x^*, X^*)$ is just $e^T x^*$, it does not depend on $X^*$. Therefore, the fact that the profit is at least as large as $\gamma$ follows from the result in [24]. To complete the proof, it suffices to show that $(x^*, X^*)$ satisfies (6) – (8).

To see that (6) is satisfied, simply note that $X^*_{ii} = Z^*_{ii} x^*_i / Z^*_i$ by construction. To see that (7) is satisfied, simply recall that $Z^*_{ij} = 0$ for all $\{i, j\} \in E$, and note that all $X^*_{ij} = 0$ whenever $Z^*_{ij} = 0$.

To show that (8) is satisfied, it suffices to construct a Gram representation of the augmented matrix

$$(X^+)^* = \begin{pmatrix} 1 & (x^*)^T \\ x^* & X^* \end{pmatrix}.$$

Let $Y^* = Y^T Y$ and let $y_1, \ldots, y_n$ be the columns of $Y$. Note that $||y_i|| = \sqrt{Z^*_i}$ and that $y_i$ is the zero vector whenever $Z^*_i = 0$. We define vectors $v_0, \ldots, v_n$ as follows. We set $v_0$ to $Ye / \sqrt{\gamma}$. For all $i \in V$ such that $Z^*_i > 0$, we set $v_i$ to $y_i / \sqrt{Z^*_i} / ||y_i||$. For all $i \in V$ such that $Z^*_i = 0$, we set $v_i$ to the zero vector.

Now we show that $v_0, \ldots, v_n$ form a Gram representation of $(X^+)^*$. We do this in three steps. First, we note that $v_0 \cdot v_0 = 1$, since

$$||v_0||^2 = ||Ye||^2 = (e^T Y^T Y e) \frac{e^T Z^* e}{\gamma} = J \bullet Z^* \frac{\gamma}{\gamma} = \gamma = 1.$$
Second, we show that $v_0 \cdot v_i = x_i^*$ for all $i \in V$. If $Z_{ii}^* = 0$, this is trivial, since $v_i$ is the zero vector and $x_i^* = 0$ in this case. On the other hand, if $Z_{ii}^* > 0$, we have:

$$v_0 \cdot v_i = \frac{Ye}{\sqrt{\gamma}} \cdot \frac{y_i \sqrt{x_i^*}}{||y_i||} = \left(e^T Y^T y_i\right) \frac{\sqrt{x_i^*}}{\sqrt{\gamma} ||y_i||} = \sum_{j=1}^{n} Z_{ij}^* \frac{\sqrt{x_i^*}}{\sqrt{Z_{ii}^* ||y_i||}}.$$

Then, since

$$x_i^* = \left(\sum_{j=1}^{n} Z_{ij}^* \right)^2 / \gamma Z_{ii}^*$$

when $Z_{ii}^* > 0$, we have that

$$v_0 \cdot v_i = \sum_{j=1}^{n} Z_{ij}^* \frac{\sum_{j=1}^{n} Z_{ij}^*}{\gamma ||y_i|| \sqrt{Z_{ii}^*}} = x_i^*$$

as desired.

Third, we show that $v_i \cdot v_j = X_{ij}^*$ for all $i$ and $j$. If $Z_{ij}^* = 0$, this is trivial, since $v_i$ and $v_j$ are orthogonal and $X_{ij}^* = 0$ in this case. On the other hand, if $Z_{ij}^* > 0$, we have:

$$v_i \cdot v_j = \frac{(y_i \cdot y_j) \sqrt{x_i^* x_j^*}}{||y_i|| ||y_j||} = \frac{Z_{ij}^* \sqrt{x_i^* x_j^*}}{\sqrt{Z_{ii}^* Z_{jj}^*}} = X_{ij}^*$$

as desired. □

We illustrate Theorem 1 on a small example:

**Example 1:** Let $G$ be the graph on 3 nodes with $E = \{\{1, 2\}, \{2, 3\}\}$. It holds trivially that $\alpha(G) = \theta(G) = 2$. One can check (e.g., with an Eigenvalue calculator) that the following matrix is a feasible solution for SDP1:

$$Z^* = \begin{pmatrix} 1/3 & 0 & 1/3 \\ 0 & 1/3 & 0 \\ 1/3 & 0 & 1/3 \end{pmatrix},$$

We have $\gamma = 5/3$. Applying the procedure of Yildirim and Fan, we obtain $x^* = (4/5, 1/5, 4/5)^T$, with profit $9/5 > \gamma$. Applying our procedure, we obtain:

$$X^* = \begin{pmatrix} 4/5 & 0 & 4/5 \\ 0 & 1/5 & 0 \\ 4/5 & 0 & 4/5 \end{pmatrix}.$$

One can check that the corresponding augmented matrix $(X^+)^*$ is indeed psd. □
The interesting thing about this example is that, if we take the matrix $X^*$ and apply the mapping of Gruber and Rendl, we obtain the following feasible solution of SDP1:

$$
\begin{pmatrix}
\frac{4}{9} & 0 & \frac{4}{9} \\
0 & \frac{1}{9} & 0 \\
\frac{4}{9} & 0 & \frac{4}{9}
\end{pmatrix}
$$

with an even better profit of $17/9$. Further iterations of the mappings yield the profit values $33/17$, $65/33$ and $129/65$. It is apparent that this sequence of profit values is rapidly converging to $\theta(G) = \alpha(G) = 2$.

An interesting question is to find conditions on the graph $G$ and the starting matrix $Z^*$ under which the sequence of profit values is guaranteed to converge to $\theta(G)$. Potentially, this could yield a new (primal) algorithm to compute $\theta(G)$. We do not examine this here, however. Instead, we give a second example, which illustrates how the mappings in both directions work when both $Z^*$ and $(x^*,X^*)$ are already optimal.

\textbf{Example 2:} Let $G$ be the odd hole on 5 nodes. That is, let $n = 5$ and let $E = \{\{1,2\}, \{1,5\}, \{2,3\}, \{3,4\}, \{4,5\}\}$. It holds trivially that $\alpha(G) = 2$, and Lovász [16] showed that $\theta(G) = \sqrt{5} \approx 2.236$. One can check with an SDP solver that the optimal solution $Z^*$ to SDP1 has $Z^*_{ii} = 1/5$ for $i \in V$, $Z^*_{ij} = 0$ for $\{i,j\} \in E$, and $Z^*_{ij} = (\sqrt{5} - 1)/10 \approx 0.1236$ for $\{i,j\} \in \bar{E}$. One can also check that the optimal solution $(x^*,X^*)$ to SDP2 has $x^*_i = X^*_i = 1/\sqrt{5} \approx 0.4472$ for $i \in V$, $X^*_{ij} = 0$ for $\{i,j\} \in E$, and $X^*_{ij} = (1 - 5^{-1/2})/2 \approx 0.2764$ for $\{i,j\} \in \bar{E}$. Finally, one can check that the mappings between $Z^*$ and $(x^*,X^*)$ work in both directions. In particular, (i) $Z^*_{ij} = X^*_{ij}/\gamma$ for all $i$ and $j$, as predicted by Gruber and Rendl, (ii) $x^*_i = \frac{(\sum_{j=1}^{n} Z^*_{ij})^2}{\gamma Z^*_ii}$ for all $i$, as predicted by Yildirim and Fan, and (iii) $X^*_{ij} = \frac{x^*_i x^*_j}{\sqrt{Z^*_ii Z^*_jj}}$ for all $i$ and $j$, as predicted by Theorem 1. \hfill $\square$

### 3.2 The weighted theta function

In this subsection, we extend the mappings to the weighted case. We start by giving an intuitive explanation for the weighted version of SDP1, due to Grötschel et al. [10], that was presented in Subsection 2.3.

Recall that $w \in \mathbb{R}^n_+$ is the vector of node weights. Let $\sqrt{w}$ denote the vector in $\mathbb{R}^n$ whose $i$th component is $\sqrt{w_i}$. Then, the objective function in the weighted version of SDP1, namely (9), can be written in the simpler form $\sqrt{w}^T Z \sqrt{w}$. Now, let $\bar{x} \in \{0,1\}^n$ be the incidence vector of a stable set, whose weight $\gamma = w^T \bar{x}$ is positive. Let $\bar{z} \in \mathbb{R}^n_+$ be the vector whose $i$th
component is $\sqrt{w_i} x_i / \sqrt{\gamma}$, and let $\tilde{Z} = \tilde{z} \tilde{z}^T$. We then have:

$$\tilde{Z} = \text{Diag}(\sqrt{w}) x x^T \text{Diag}(\sqrt{w}) / \gamma.$$ 

One can check that $\tilde{Z}$ satisfies (2)–(4), and is therefore feasible for SDP1. Moreover, we have:

$$\sqrt{w}^T \tilde{Z} \sqrt{w} = \frac{\sqrt{w}^T \text{Diag}(\sqrt{w}) x x^T \text{Diag}(\sqrt{w}) \sqrt{w}}{\gamma} = \frac{w^T x x^T w}{\gamma} = \frac{\gamma^2}{\gamma} = \gamma,$$

which explains the form of the objective function (9).

We are now in a position to extend the mappings to the weighted case. This is accomplished in the following two theorems.

**Theorem 2** Let $(x^*, X^*)$ be a feasible solution to the weighted version of SDP2, such that $\gamma = w^T x > 0$. We can obtain a feasible solution $Z^*$ to the weighted version of SDP1, whose profit is at least as large, by setting $Z^*_{ij}$ to $\sqrt{w_i w_j} x^*_i / \gamma$ for all $i$ and $j$.

**Proof.** By construction, $Z^*_{ij} = 0$ whenever $X^*_ij = 0$, and therefore $Z^*$ satisfies (2). Also, we have:

$$\text{Tr}(Z^*) = \sum_{i \in V} \frac{\sqrt{w_i} w_i X^*_i}{\gamma} = \frac{1}{\gamma} \sum_{i \in V} w_i x^*_i = \gamma / \gamma = 1,$$

which shows that (3) is satisfied. Moreover, by construction, we have

$$Z^* = \frac{\text{Diag}(\sqrt{w}) X^* \text{Diag}(\sqrt{w})}{\gamma}.$$

Since $X^*$ is psd, it can be factorised as $Y^T Y$. This gives:

$$Z^* = \frac{\text{Diag}(\sqrt{w}) Y^T Y \text{Diag}(\sqrt{w})}{\gamma} = \left( \frac{Y \text{Diag}(\sqrt{w})}{\sqrt{\gamma}} \right)^T \left( \frac{Y \text{Diag}(\sqrt{w})}{\sqrt{\gamma}} \right),$$

which shows that (4) is satisfied.

Finally, we have to show that the profit of $Z^*$ is at least $\gamma$. This profit is equal to:

$$\sqrt{w}^T Z^* \sqrt{w} = \frac{\sqrt{w}^T \text{Diag}(\sqrt{w}) X^* \text{Diag}(\sqrt{w}) \sqrt{w}}{\gamma} = \frac{w^T X^* w}{\gamma}.$$ 

Due to (8) and the Schur complement, $X^* - x^* (x^*)^T$ must be psd. This implies that $w^T X^* w - (w^T x^*)^2 \geq 0$, and therefore $(w^T X^* w) / \gamma \geq w^T x^*$. □
Theorem 3 Let $Z^*$ be a feasible solution to the weighted version of SDP1, whose profit $\gamma$ is positive. We can obtain a feasible solution $(x^*, X^*)$ to the weighted version of SDP2, whose profit is at least as large, as follows. If $Z^*_{ii} = 0$, set $x_i^* = 0$. Otherwise, set $x_i^*$ to

$$
\left( \sum_{j=1}^{n} \sqrt{w_j} Z^*_{ij} \right)^2 \over \gamma Z^*_{ii}.
$$

If $Z^*_{ij} = 0$, set $X^*_{ij} = 0$. Otherwise, set $X^*_{ij}$ to

$$
Z^*_{ij} \sqrt{x_i^* x_j^*} \over Z^*_{ii} Z^*_{jj} = Z^*_{ij} \left( \sum_{k=1}^{n} \sqrt{w_k} Z^*_{ik} \right) \left( \sum_{k=1}^{n} \sqrt{w_k} Z^*_{jk} \right) \over \gamma Z^*_{ii} Z^*_{jj}.
$$

Then the resulting pair $(x^*, X^*)$ forms a feasible solution to SDP2, and its profit, $w^T x^*$, is at least as large as $\gamma$.

Proof. The proof that $(x^*, X^*)$ is feasible is similar to the proof of Theorem 1. The only difference is that, in the Gram representation of the augmented matrix $(X^+)^*$, we need to set $v_0$ to $Y \sqrt{w}/\sqrt{\gamma}$.

The hard part is to show that the profit of $(x^*, X^*)$ is at least as large as that of $Z^*$. We adapt the proof in [24]. Let $V' = \{ i \in V : Z^*_{ii} > 0 \}$. By construction, we have the following for all $i \in V'$:

$$
x_i^* = v_0 \cdot v_i = \sqrt{x_i^*} \over \sqrt{\gamma} ||y_i|| \left( y_i^T Y \sqrt{w} \right) = \sqrt{x_i^*} \over \sqrt{\gamma} ||y_i|| \sum_{j \in V} \sqrt{w_j} Z^*_{ij}.
$$

Multiplying both sides by $||y_i||/\sqrt{x_i^*}$, we have:

$$
||y_i|| \sqrt{x_i^*} = 1 \over \sqrt{\gamma} \sum_{j \in V} \sqrt{w_j} Z^*_{ij}.
$$

Now, multiplying these equations by $\sqrt{w_i}$ and summing over all $i \in V'$, we obtain:

$$
\sum_{i \in V'} ||y_i|| \sqrt{w_i x_i^*} = \sqrt{w^T Z^+ \sqrt{w}} \over \sqrt{\gamma} = \gamma / \sqrt{\gamma} = \sqrt{\gamma}.
$$

Equivalently,

$$
\gamma = \left( \sum_{i \in V'} ||y_i|| \sqrt{w_i x_i^*} \right)^2.
$$

By the Cauchy-Schwarz inequality, this cannot exceed:

$$
\left( \sum_{i \in V'} ||y_i||^2 \right) \left( \sum_{i \in V'} w_i x_i^* \right) = \text{Tr}(Z^+) \ w^T x^*.
$$

The result then follows from the fact that $\text{Tr}(Z^+) = 1$. \qed
Together, Theorems 2 and 3 provide an alternative proof that both SDPs yield the same upper bound, namely $\theta(G, w)$. They are illustrated on the following example.

**Example 3:** Let $G$ be the odd hole on 5 nodes, as in Example 2, but now suppose that nodes 1, . . . , 4 have weight 2 and node 5 has weight 3. There are two optimal stable sets, each of weight 5. One can check with an SDP solver that the optimal solution to the weighted version of SDP1 is as follows (to 3 d.p.):

$$Z^* = \begin{pmatrix}
0.074 & 0 & 0.057 & 0.057 & 0 \\
0 & 0.194 & 0 & 0.057 & 0.198 \\
0.057 & 0 & 0.194 & 0 & 0.198 \\
0.057 & 0.057 & 0 & 0.074 & 0 \\
0 & 0.198 & 0.198 & 0 & 0.464
\end{pmatrix}.$$ 

One can also check that the optimal solution to the weighted version of SDP2 is:

$$(X^+)^* = \begin{pmatrix}
1 & 0.188 & 0.494 & 0.494 & 0.188 & 0.788 \\
0.188 & 0.188 & 0 & 0.145 & 0.145 & 0 \\
0.494 & 0 & 0.494 & 0 & 0.145 & 0.412 \\
0.494 & 0.145 & 0 & 0.494 & 0 & 0.412 \\
0.188 & 0.145 & 0.145 & 0 & 0.188 & 0 \\
0.788 & 0 & 0.412 & 0.412 & 0 & 0.788
\end{pmatrix}.$$ 

These solutions both yield $\theta(G, w) \approx 5.091 > 5$. Finally, one can check that the mappings described in Theorems 2 and 3 work in this case. □

### 4 Variants of the Theta Function

Now we consider the variants of the theta function mentioned in Subsection 2.4. In Subsection 4.1, we consider stronger variants that are obtained by the addition of cutting planes. In Subsection 4.2, we consider the weakened variant of Lieder et al., along with another weakened variant that, to our knowledge, is new.

#### 4.1 Strengthening the function with cutting planes

It is interesting to observe that SDP1 and SDP2 behave rather differently when it comes to cutting planes. For example, whereas the inequalities (10), (11) and (12) are very similar to the inequalities (13), (14) and (16), respectively, the remaining inequalities for SDP2, i.e., (15) and (17), have not been adapted to SDP1. To explain this fact, we will need the following definition:
Definition 1 (Padberg [20]) The Boolean quadric polytope of order $n$, denoted by $BQP_n$, is:

$$\text{conv}\left\{(x, y) \in \{0, 1\}^{n+\binom{n}{2}} : y_{ij} = x_i x_j \ (1 \leq i < j \leq n)\right\}.$$  

This family of polytopes has been studied in great depth, and many families of valid and facet-defining inequalities are known (see Deza & Laurent [4] for a survey).

We have the following two propositions:

**Proposition 1** If the inequality $\lambda^T x + \mu^T y \leq \nu$ is valid for the Boolean quadric polytope of order $n$, then the inequality

$$\sum_{i \in V} \lambda_i Z_{ii} + \sum_{\{i,j\} \in E} \mu_{ij} Z_{ij} \leq \nu$$

can be added to SDP2 as a cutting plane. Moreover, no other cutting planes are necessary.

**Proof.** For the sake of brevity, we only sketch the proof. A feasible solution $(x^*, X^*)$ to SDP2 represents a stable set if and only if $x \in \{0, 1\}^n$, $x_i x_j = 0$ for all $\{i,j\} \in E$, and $X = xx^T$. Let $P(x, X)$ denote the convex hull of all such solutions. For any such solution $X^*$, we construct a point $(x^*, y^*)$ by setting $y_{ij}^*$ to $X_{ij}^*$ for $1 \leq i < j \leq n$. Let $P(x, y)$ denote the convex hull of all such points. Note that $P(x, X)$ and $P(x, y)$ are affinely congruent.

Now, Padberg [20] showed that the non-negativity inequalities $y_{ij} \geq 0$ define facets of $BQP_n$ for $1 \leq i < j \leq n$. He also noted that $P(x, y)$ is nothing but the face of $BQP_n$ induced by the non-negativity inequalities $y_{ij} \geq 0$ for $\{i,j\} \in E$. The result follows easily. □

**Proposition 2** If the inequality $\lambda^T x + \mu^T y \leq 0$ is valid for the Boolean quadric polytope of order $n$, then the inequality

$$\sum_{i \in V} \lambda_i Z_{ii} + \sum_{\{i,j\} \in E} \mu_{ij} Z_{ij} \leq 0 \quad (19)$$

can be added to SDP1 as a cutting plane. Moreover, no other cutting planes are necessary.

**Proof.** Again, we only sketch the proof. A feasible solution $Z^*$ to SDP1 represents a stable set if and only if $Z = xx^T / e^T x$ for some non-zero vector $x \in \{0, 1\}^n$ such that $x_i x_j = 0$ for all $\{i,j\} \in E$. Let $P(Z)$ denote the convex hull of all such solutions. For any such solution $Z^*$, we construct a point $(x^*, y^*)$ by setting $y_{ij}^*$ to $Z_{ij}^*$ for $1 \leq i < j \leq n$. Let $P'(x, y)$ denote the convex hull of all such points. Again, $P(Z)$ and $P'(x, y)$ are affinely congruent.
Now, one shows that $P'(x, y)$ can instead be obtained as follows. Take all extreme points of $BQP_n$ that satisfy $y_{ij} = 0$ for $\{i, j\} \in E$, construct the cone of these points, and then intersect that cone with the hyperplane defined by the equation $e^T x = 1$. From this, it follows that $P'(x, y)$ is completely described by the equation $e^T x = 1$, the equations $y_{ij} = 0$ for $\{i, j\} \in E$, and all valid homogeneous inequalities for $BQP_n$. The result follows easily.

As an illustration of these results, we recall that Padberg [20] showed that the following triangle inequalities define facets of $BQP_n$:

$$y_{ik} + y_{jk} \leq x_k + y_{ij} \quad (\{i, j, k\} \subset \{1, \ldots, n\}) \quad (20)$$

$$x_i + x_j + x_k \leq 1 + y_{ij} + y_{ik} + y_{jk} \quad (1 \leq i < j < k \leq n). \quad (21)$$

It is apparent that the inequalities (20) are the source of the inequalities (11) and (12) for $SDP1$ and the inequalities (14) and (16) for $SDP2$. The inequalities (21) are the source of the inequalities (15) and (17) for $SDP2$, but, being inhomogeneous, do not yield any useful inequalities for $SDP1$.

Another issue that turns out to be worth considering is the effect that cutting planes have on the quality of the upper bound. We start with the simplest cutting planes, the non-negativity inequalities (10) and (13). One can check that the mappings given by Gruber & Rendl [12], Yildirim & Fan [24] and ourselves all preserve non-negativity. Thus, the bound obtained by imposing non-negativity on $X$ in $SDP2$ is the same as the one obtained by imposing it on $Z$ in $SDP1$, i.e., is nothing but Schrijver’s bound.

More generally, suppose we strengthen $SDP2$ by adding an arbitrary collection of homogeneous valid inequalities, such as some or all of (13), (14) and (16). Let $(\tilde{x}, \tilde{X})$ be the optimal solution to the strengthened relaxation. The mapping of Gruber and Rendl, which merely divides $\tilde{X}$ by a constant, yields a feasible solution $\tilde{Z}$ to $SDP1$ that satisfies the analogous homogeneous inequalities, and has at least as much profit. This implies that the bound obtained by adding the inequalities to $SDP2$ is at least as strong as the one obtained by adding the analogous inequalities to $SDP1$. Surprisingly, it can be strictly stronger in some cases. This is shown in the following example.

**Example 2 (cont.)**: Again, let $G$ be the odd hole on 5 nodes. The optimal solution to $SDP1$ violates the homogeneous inequality $Z_{13} + Z_{14} \leq Z_{11}$, which is of the form (11). One can check with an SDP solver that, if this inequality is added to $SDP1$ as a cutting plane, the optimal solution becomes (to 3 d.p.):

$$\begin{pmatrix}
0.272 & 0 & 0.136 & 0.136 & 0 \\
0 & 0.185 & 0 & 0.113 & 0.113 \\
0.136 & 0 & 0.179 & 0 & 0.113 \\
0.136 & 0.113 & 0 & 0.179 & 0 \\
0 & 0.113 & 0.113 & 0 & 0.185
\end{pmatrix}.$$
The associated upper bound is 2.224.

The optimal solution to $\text{SDP}_2$ violates the analogous homogeneous inequality $X_{13} + X_{14} \leq x_1$, which is of the form (14). One can check that, if this is added to $\text{SDP}_2$, the optimal matrix $X^+$ becomes:

$$
\begin{pmatrix}
1 & 0.586 & 0.293 & 1/2 & 1/2 & 0.293 \\
0.586 & 0.586 & 0 & 0.293 & 0.293 & 0 \\
0.293 & 0 & 0.293 & 0 & 0.293 & 1/2 \\
1/2 & 0.293 & 0 & 1/2 & 0 & 0 \\
1/2 & 0.293 & 0.293 & 0 & 1/2 & 0 \\
0.293 & 0 & 0.121 & 0.293 & 0 & 0.293
\end{pmatrix}.
$$

The associated upper bound is 2.172. □

Thus, if one wishes to add cutting planes, the choice between $\text{SDP}_1$ and $\text{SDP}_2$ is not at all clear-cut. It is true that we reduce the number of variables and constraints by using $\text{SDP}_1$, but this can come at the cost of a weakened upper bound.

We remark that Proposition 2 is not valid in the weighted case. In that case, one has to modify the inequality (19) by dividing each $\lambda_i$ by $w_i$ and dividing each $\mu_{ij}$ by $\sqrt{w_i w_j}$. It remains true, however, that adding homogeneous cutting planes to $\text{SDP}_2$ always leads to at least as much bound improvement as adding the analogous cutting planes to $\text{SDP}_1$. (This can be shown by a modification of the proof of Theorem 2.)

4.2 Weakening the function with constraint aggregation

Recall that the weakened theta function of Lieder et al. [15] is obtained by taking the constraints (7) in $\text{SDP}_2$ and replacing them with the single weaker constraint (18). We will denote this bound by $\theta^{LRJ}(G)$. A natural question is whether one obtains the same bound if one weakens $\text{SDP}_1$ by replacing the constraints (2) with the following single constraint:

$$
\sum_{\{i,j\} \in E} Z_{ij} = 0. \tag{22}
$$

The following proposition shows that the bound obtained in this way cannot be better than $\theta^{LRJ}(G)$.

**Proposition 3** If we weaken $\text{SDP}_1$ by replacing (2) with (22), the resulting upper bound on $\alpha(G)$ is at least as large as $\theta^{LRJ}(G)$.

**Proof.** We follow the strategy used in [12]. Suppose the pair $(x^*, X^*)$ satisfies (6), (8) and (18), and let $\gamma = e^T x^* > 0$ be the associated objective value. Consider the matrix $Z^* = X^*/\gamma$. It satisfies (3), since

$$
\text{Tr}(Z^*) = \text{Tr}(X^*)/\gamma = e^T x^*/\gamma = \gamma/\gamma = 1.
$$

14
It satisfies (4), since $X^*$ is psd; and it satisfies (22), since
\[
\sum_{(i,j) \in E} Z_{ij}^* = \left( \sum_{(i,j) \in E} X_{ij}^* \right) / \gamma = 0 / \gamma = 0.
\]
Now, since $X^* - (x^*)^T x^*$ is psd, we have $e^T X^* e - e^T (x^*)^T e \geq 0$. This implies that $J \cdot X^* \geq \gamma^2$, which in turn implies that $J \cdot Z^* \geq \gamma$. □

The following example shows that, in fact, the bound obtained by weakening SDP1 can be larger than $\theta^{LRJ}(G)$. For this reason we will denote it by $\theta_{LRJ}^+(G)$.

**Example 4:** Let $n = 4$ and $E = \{\{1, 2\}, \{1, 3\}, \{2, 3\}, \{3, 4\}\}$. One can check with an SDP solver that $\alpha(G) = \theta(G) = \theta^{LRJ}(G) = 2$, while $\theta_{LRJ}^+(G) \approx 2.3855$. □

To close this section, we introduce two more natural weakened variants of the theta function, which to our knowledge are new. The first one, which we will call $\theta^{GL}(G)$, is obtained by taking SDP2 and replacing the $m$ constraints (7) with the following $n$ constraints:
\[
\sum_{j \in V \setminus \{i\}: (i,j) \in E} X_{ij} = 0 \quad (i \in V).
\]
The second, which we will call $\theta^{GL}_+(G)$, is obtained by taking SDP1 and replacing the $m$ constraints (2) with the $n$ constraints
\[
\sum_{j \in V \setminus \{i\}: (i,j) \in E} Z_{ij} = 0 \quad (i \in V).
\]
By definition, we have $\theta(G) \leq \theta^{GL}(G) \leq \theta^{LRJ}(G)$ and $\theta(G) \leq \theta^{GL}_+(G) \leq \theta^{LRJ}_+(G)$. Moreover, the technique used to prove Proposition 3 can be used to show that $\theta^{GL}(G) \leq \theta^{GL}_+(G)$. The following example shows that all of the inequalities can be strict, and also that $\theta^{LRJ}(G)$ can be larger than $\theta^{GL}_+(G)$:

**Example 5:** Let $n = 6$ and
\[
E = \{\{1, 2\}, \{1, 5\}, \{2, 3\}, \{3, 4\}, \{3, 6\}, \{4, 5\}, \{4, 6\}, \{5, 6\}\}.
\]
One can check that $\alpha(G) = 2$, $\theta(G) = \sqrt{5} \approx 2.2361$, $\theta^{LRJ}(G) \approx 2.3176$, $\theta^{LRJ}_+(G) \approx 2.6440$, $\theta^{GL}(G) \approx 2.2836$, $\theta^{GL}_+(G) \approx 2.2880$. □

Finally, the following example shows that $\theta^{GL}_+(G)$ can be larger than $\theta^{LRJ}(G)$.
Example 6: Let \( n = 6 \) and
\[
E = \{\{1, 2\}, \{1, 5\}, \{1, 6\}, \{2, 3\}, \{2, 6\}, \{3, 4\}, \{3, 6\}, \{4, 5\}, \{4, 6\}, \{5, 6\}\}.
\]
(This graph is called the “5-wheel” [11].) One can check that \( \alpha(G) = 2 \), \( \theta(G) = \theta_{LRJ}^L(G) = \theta_{GL}^L(G) = \sqrt{5} \approx 2.2361 \), \( \theta_{LRJ}^R(G) = 8/3 \approx 2.6667 \) and \( \theta_{GL}^R(G) \approx 2.2429 \).

5 Concluding Remarks

It has been known since the 1980s [11] that the Lovász theta function can be computed by solving either of two equivalent SDPs, which we have labelled SDP1 and SDP2. We have shown, however, that the equivalence between SDP1 and SDP2 breaks down when they are either strengthened (via cutting planes) or weakened (via constraint aggregation).

Our results have implications for the development of exact algorithms for the maximum stable set problem. In a branch-and-bound algorithm, a trade-off must be made between the quality of the upper bound and the time taken to compute it. If one wishes to use the theta function itself as the upper bound, then SDP1 is to be preferred to SDP2. If, however, one wishes to use a strengthened or weakened variant of the theta number, then the choice between SDP1 and SDP2 is no longer obvious.

References


