SUPPLEMENTARY MATERIALS: Scalable Semidefinite Programming*

Alp Yurtsever†, Joel A. Tropp‡, Olivier Fercoq§, Madeleine Udell¶, and Volkan Cevher†

SM1. Analysis of the CGAL Algorithm. This section contains a complete analysis of the convergence of the CGAL algorithm and its arithmetic costs. For simplicity, we have specialized this presentation to the model problem that is the focus of this paper; many extensions are possible. The convergence results here are adapted from the initial paper [SM27] on the CGAL algorithm. The analysis of the approximate eigenvector computation and the detailed results for the model problem are new. Empirical work suggests that the analysis is still qualitatively suboptimal, which is a direction for future research.

SM1.1. The model problem. We focus on solving the optimization template

$$\text{minimize} \quad \langle \boldsymbol{C},\, \boldsymbol{X} \rangle \quad \text{subject to} \quad \mathcal{A}\boldsymbol{X} = \boldsymbol{b}, \quad \boldsymbol{X} \in \alpha\boldsymbol{\Delta}_n. \tag{SM1.1}$$

The constraint set $\alpha\boldsymbol{\Delta}_n$ consists of $n \times n$ psd matrices with trace α. The objective function depends on a matrix $\boldsymbol{C} \in \mathbb{S}_n$. The linear constraints are determined by the linear map $\mathcal{A} : \mathbb{S}_n \to \mathbb{R}^d$ and the right-hand side vector $\boldsymbol{b} \in \mathbb{R}^d$.

SM1.2. Elements of Lagrangian duality. Introduce the Lagrangian function

$$L(\boldsymbol{X}; \boldsymbol{y}) := \langle \boldsymbol{C},\, \boldsymbol{X} \rangle + \langle \boldsymbol{y},\, \mathcal{A}\boldsymbol{X} - \boldsymbol{b} \rangle \quad \text{for } \boldsymbol{X} \in \alpha\boldsymbol{\Delta}_n \text{ and } \boldsymbol{y} \in \mathbb{R}^d. \tag{SM1.2}$$

We assume that the Lagrangian admits at least one saddle point $(\boldsymbol{X}_\star, \boldsymbol{y}_\star) \in \alpha\boldsymbol{\Delta}_n \times \mathbb{R}^d$:

$$L(\boldsymbol{X}_\star, \boldsymbol{y}) \le L(\boldsymbol{X}_\star, \boldsymbol{y}_\star) \le L(\boldsymbol{X}; \boldsymbol{y}_\star) \quad \text{for all } \boldsymbol{X} \in \alpha\boldsymbol{\Delta}_n \text{ and } \boldsymbol{y} \in \mathbb{R}^d. \tag{SM1.3}$$

This hypothesis is guaranteed under Slater's condition. Write $p_\star$ for the shared extremal value of the dual and primal problems:

$$\max_{\boldsymbol{y} \in \mathbb{R}^d} \min_{\boldsymbol{X} \in \alpha\boldsymbol{\Delta}_n} L(\boldsymbol{X}, \boldsymbol{y}) = p_\star = \min_{\boldsymbol{X} \in \alpha\boldsymbol{\Delta}_n} \sup_{\boldsymbol{y} \in \mathbb{R}^d} L(\boldsymbol{X}, \boldsymbol{y}). \tag{SM1.4}$$

In particular, note that $p_\star = \langle \boldsymbol{C},\, \boldsymbol{X}_\star \rangle$.

*Supplementary material for SIMODS MS#M130504.
https://doi.org/10.1137/19M1305045

†Laboratory for Information and Inference Systems, Department of Electrical Engineering, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland (alp.yurtsever@epfl.ch, volkan.cevher@epfl.ch, https://lions.epfl.ch/).

‡Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, 91125-5000 (jtropp@cms.caltech.edu, http://users.cms.caltech.edu/~jtropp).

§Laboratoire Traitement et Communication d'Information, Télécom Paris, Institut Polytechnique de Paris, Palaiseau, France (olivier.fercoq@telecom-paris.fr, https://perso.telecom-paristech.fr/ofercoq/).

¶Department of Operations Research and Information Engineering, Cornell University, Ithaca, NY, 14850 (udell@cornell.edu).

SM1.3. The CGAL iteration. The CGAL iteration solves the model problem (SM1.1) using an augmented Lagrangian method where the primal step is inspired by the conditional gradient method. See Algorithm SM1.1 for pseudocode.

Let $\beta_0 > 0$ be an initial smoothing parameter. Fix a schedule for the step size parameter η_t and the smoothing parameter β_t:

$$\eta_t := \frac{2}{t+1} \quad \text{and} \quad \beta_t := \beta_0\sqrt{t+1} \quad \text{for } t = 1, 2, 3, \dots. \tag{SM1.5}$$

Define the augmented Lagrangian L_t with smoothing parameter β_t

$$L_t(\boldsymbol{X}; \boldsymbol{y}) := \langle \boldsymbol{C},\, \boldsymbol{X}\rangle + \langle \boldsymbol{y},\, \mathcal{A}\boldsymbol{X} - \boldsymbol{b}\rangle + \frac{1}{2}\beta_t\|\mathcal{A}\boldsymbol{X} - \boldsymbol{b}\|^2. \tag{SM1.6}$$

The CGAL algorithm solves the model problem (SM1.1) by alternating between primal and dual update steps on (SM1.6), while increasing the smoothing parameter.

Fix the initial iterates

$$\boldsymbol{X}_1 = \mathbf{0} \in \mathbb{S}_n \quad \text{and} \quad \boldsymbol{y}_1 = \mathbf{0} \in \mathbb{R}^d. \tag{SM1.7}$$

At each iteration $t = 1, 2, 3, \dots$, we implicitly compute the partial derivative

$$\boldsymbol{D}_t := \partial_{\boldsymbol{X}} L_t(\boldsymbol{X}_t; \boldsymbol{y}_t) = \boldsymbol{C} + \mathcal{A}^*\big(\boldsymbol{y}_t + \beta_t(\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b})\big). \tag{SM1.8}$$

Then we identify a primal update direction $\boldsymbol{H}_t \in \alpha\boldsymbol{\Delta}_n$ that satisfies

$$\langle \boldsymbol{D}_t,\, \boldsymbol{H}_t\rangle \le \min_{\boldsymbol{H}\in\alpha\boldsymbol{\Delta}_n} \langle \boldsymbol{D}_t,\, \boldsymbol{H}\rangle + \frac{\alpha\beta_0}{\beta_t}\|\boldsymbol{D}_t\|. \tag{SM1.9}$$

In other words, the smoothing parameter β_t also controls the amount of inexactness we are willing to tolerate in the linear minimization at iteration t. We construct the next primal iterate via the rule

$$\boldsymbol{X}_{t+1} = \boldsymbol{X}_t + \eta_t(\boldsymbol{H}_t - \boldsymbol{X}_t) \in \alpha\boldsymbol{\Delta}_n. \tag{SM1.10}$$

The dual update takes the form

$$\boldsymbol{y}_{t+1} = \boldsymbol{y}_t + \gamma_t(\mathcal{A}\boldsymbol{X}_{t+1} - \boldsymbol{b}). \tag{SM1.11}$$

We select the largest dual step size parameter $0 \le \gamma_t \le \beta_0$ that satisfies the condition

$$\gamma_t\|\mathcal{A}\boldsymbol{X}_{t+1} - \boldsymbol{b}\|^2 \le \beta_t\eta_t^2\alpha^2\|\mathcal{A}\|^2. \tag{SM1.12}$$

The latter inequality always holds when $\gamma_t = 0$. We will also choose the dual step size to maintain a bounded travel condition:

$$\|\boldsymbol{y}_t\| \le K. \tag{SM1.13}$$

If the bounded travel condition holds at iteration $t-1$, then the choice $\gamma_t = 0$ ensures that it holds at iteration t. In practice, it is not necessary to enforce (SM1.13). The iteration continues until it reaches a maximum iteration count $T_{\max}$.

Algorithm SM1.1 CGAL for the model problem (SM1.1)

Input: Problem data for (SM1.1) implemented via the primitives (2.4); number T of iterations
Output: Approximate solution $\boldsymbol{X}_T$ to (2.2)
Recommendation: Set $T \approx \varepsilon^{-1}$ to achieve ε-optimal solution (SM1.17)

```
function CGAL(T)
    Scale problem data (subsection 7.1.1)                          ▷ [opt] Recommended!
    β₀ ← 1 and K ← +∞                                              ▷ Default parameters
    X ← 0_{n×n} and y ← 0_d
    for t ← 1, 2, 3, ..., T do
        β ← β₀ √(t+1) and η ← 2/(t+1)
        (ξ, v) ← ApproxMinEvec(C + 𝒜*(y + β(𝒜X − b)); q_t)
                                              ▷ Algorithm 4.1 with q_t = t^{1/4} log n
                                              ▷ Implement with primitives (2.4)❶❷!
        X ← (1 − η) X + η (α v v*)
        y ← y + γ(𝒜X − b)              ▷ Step size γ satisfies (SM1.12) and (SM1.13)
```

SM1.4. Distributed computation. Since CGAL builds on the augmented Lagrangian framework, we can apply it even when the problem is too large to solve on one computational node. In particular, when d is large, it may be advantageous to partition the constraint matrices $\boldsymbol{A}_i$ and the associated dual variables y_i among several workers. Distributed CGAL has a structure similar to the alternating directions method of multipliers (ADMM) [SM5].

SM1.5. Theoretical analysis of CGAL. We develop two results on the behavior of CGAL. The first concerns its convergence to optimality, and the second concerns the computational resource usage.

SM1.5.1. Convergence. The first result demonstrates that the CGAL algorithm always converges to a primal optimal solution of the model problem (SM1.1). This result is adapted from [SM27]; a complete proof appears below in subsection SM1.6.

Theorem SM1.1 (CGAL: Convergence). *Define*

$$E := 6\alpha^2 \|\mathcal{A}\|^2 + \alpha(\|\boldsymbol{C}\| + K\|\mathcal{A}\|).$$

The primal iterates $\{\boldsymbol{X}_t : t = 2, 3, 4, \dots\}$ *generated by the CGAL iteration satisfy the a priori bounds*

$$\langle \boldsymbol{C},\, \boldsymbol{X}_t\rangle - p_\star \le \frac{2\beta_0 E}{\sqrt{t}} + K \cdot \|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\|; \tag{SM1.14}$$

$$-\|\boldsymbol{y}_\star\| \cdot \|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\| \le \langle \boldsymbol{C},\, \boldsymbol{X}_t\rangle - p_\star; \tag{SM1.15}$$

$$\|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\| \le \frac{2\beta_0^{-1}(K + \|\boldsymbol{y}_\star\|) + 2\sqrt{E}}{\sqrt{t}}. \tag{SM1.16}$$

The a priori *bounds ensure that*

$$\|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\| \le \varepsilon \quad \textit{and} \quad |\langle \boldsymbol{C},\, \boldsymbol{X}_t\rangle - p_\star| \le \varepsilon \tag{SM1.17}$$

within $O(\varepsilon^{-2})$ iterations. The big-O suppresses constants that depend on the problem data $(\alpha, \|\mathcal{A}\|, \|\boldsymbol{C}\|)$ and the algorithm parameters β_0 and K.

SM1.5.2. Problem scaling. Theorem SM1.1 indicates that it is valuable to scale the model problem (SM1.1) so that $\alpha = \|\boldsymbol{C}\| = \|\mathcal{A}\| = 1$. In this case, a good choice for the smoothing parameter is $\beta_0 = 1$. Nevertheless, the algorithm converges, regardless of the scaling and regardless of the parameter choices β_0 and K. We use a slightly different scaling in practice; see subsection 7.1.1.

SM1.5.3. Bounded travel? Theorem SM1.1 suggests that the optimal choice for the travel bound is $K = 0$. In other words, the dual vector $\boldsymbol{y}_t = \boldsymbol{0}$, and it does not evolve. The algorithm that results from this choice is called HomotopyCGM [SM28]. The numerical work on CGAL, reported in [SM27], makes it clear that updating the dual variable, as CGAL does, allows for substantial performance improvements over HomotopyCGM. Unfortunately, the theory developed in [SM27], and echoed here, does not comprehend the reason for this phenomenon. This is an obvious direction for further research.

SM1.6. Proof of Theorem SM1.1. In this section, we establish the convergence guarantee stated in Theorem SM1.1.

SM1.6.1. Analysis of the primal update. The first steps in the proof address the role of the primal update rule (SM1.10). The arguments parallels the standard convergence analysis [SM12] of CGM, applied to the function

$$f_t(\boldsymbol{X}) := L_t(\boldsymbol{X}; \boldsymbol{y}_t) = \langle \boldsymbol{C},\, \boldsymbol{X}\rangle + \langle \boldsymbol{y}_t,\, \mathcal{A}\boldsymbol{X} - \boldsymbol{b}\rangle + \frac{1}{2}\beta_t \|\mathcal{A}\boldsymbol{X} - \boldsymbol{b}\|^2. \tag{SM1.18}$$

Observe that the gradient $\nabla f_t(\boldsymbol{X}_t)$ coincides with the partial derivative (SM1.8).

To begin, we exploit the smoothness of f_t to control the change in its value at adjacent primal iterates. The function f_t is convex on $\mathbb{S}_n$, and its gradient has Lipschitz constant $\beta_t\|\mathcal{A}\|^2$, so

$$f_t(\boldsymbol{X}_+) - f_t(\boldsymbol{X}) \le \langle \nabla f_t(\boldsymbol{X}),\, \boldsymbol{X}_+ - \boldsymbol{X}\rangle + \frac{1}{2}\beta_t\|\mathcal{A}\|^2\|\boldsymbol{X}_+ - \boldsymbol{X}\|_{\mathrm{F}}^2 \quad \text{for } \boldsymbol{X}, \boldsymbol{X}_+ \in \mathbb{S}_n.$$

In particular, with $\boldsymbol{X}_+ = \boldsymbol{X}_{t+1}$ and $\boldsymbol{X} = \boldsymbol{X}_t$, we obtain

$$\begin{aligned} f_t(\boldsymbol{X}_{t+1}) &\le f_t(\boldsymbol{X}_t) + \langle \boldsymbol{D}_t,\, \boldsymbol{X}_{t+1} - \boldsymbol{X}_t\rangle + \frac{1}{2}\beta_t\|\mathcal{A}\|^2\|\boldsymbol{X}_{t+1} - \boldsymbol{X}_t\|_{\mathrm{F}}^2 \\ &= f_t(\boldsymbol{X}_t) + \eta_t\langle \boldsymbol{D}_t,\, \boldsymbol{H}_t - \boldsymbol{X}_t\rangle + \frac{1}{2}\beta_t\eta_t^2\|\mathcal{A}\|^2\|\boldsymbol{H}_t - \boldsymbol{X}_t\|_{\mathrm{F}}^2 \\ &\le f_t(\boldsymbol{X}_t) + \eta_t\langle \boldsymbol{D}_t,\, \boldsymbol{H}_t - \boldsymbol{X}_t\rangle + \beta_t\eta_t^2\alpha^2\|\mathcal{A}\|^2. \end{aligned} \tag{SM1.19}$$

The second identity follows from the update rule (SM1.10). The bound on the last term depends on the fact that the constraint set $\alpha\boldsymbol{\Delta}_n$ has Frobenius-norm diameter $\alpha\sqrt{2}$.

Next, we use the construction of the primal update to control the linear term in the last display. The update direction $\boldsymbol{H}_t$ satisfies the inequality (SM1.9), so

$$\begin{aligned} \langle \boldsymbol{D}_t,\, \boldsymbol{H}_t - \boldsymbol{X}_t\rangle &\le \min_{\boldsymbol{H}\in\alpha\boldsymbol{\Delta}_n} \langle \boldsymbol{D}_t,\, \boldsymbol{H} - \boldsymbol{X}_t\rangle + \frac{\alpha\beta_0}{\beta_t}\|\boldsymbol{D}_t\| \\ &\le \langle \boldsymbol{D}_t,\, \boldsymbol{X}_\star - \boldsymbol{X}_t\rangle + \frac{\alpha\beta_0}{\beta_t}\|\boldsymbol{D}_t\|. \end{aligned} \tag{SM1.20}$$

The second inequality depends on the fact that $\boldsymbol{X}_\star \in \alpha\boldsymbol{\Delta}_n$.

We can use the explicit formula (SM1.8) for the derivative $\boldsymbol{D}_t$ to control the two terms in (SM1.20). First,

$$\begin{aligned}\langle \boldsymbol{D}_t,\, \boldsymbol{X}_\star - \boldsymbol{X}_t\rangle &= \langle \boldsymbol{C} + \mathcal{A}^*\big(\boldsymbol{y}_t + \beta_t(\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b})\big),\, \boldsymbol{X}_\star - \boldsymbol{X}_t\rangle \\ &= \langle \boldsymbol{C},\, \boldsymbol{X}_\star - \boldsymbol{X}_t\rangle - \langle \boldsymbol{y}_t,\, \mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\rangle - \beta_t\|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\|^2 \\ &= \langle \boldsymbol{C},\, \boldsymbol{X}_\star\rangle - f_t(\boldsymbol{X}_t) - \frac{1}{2}\beta_t\|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\|^2.\end{aligned} \tag{SM1.21}$$

We have invoked the definition of the adjoint $\mathcal{A}^*$ and the fact that $\mathcal{A}\boldsymbol{X}_\star = \boldsymbol{b}$. Last, we used the definition (SM1.18) to identify the quantity $f_t(\boldsymbol{X}_t)$. Second,

$$\begin{aligned}\|\boldsymbol{D}_t\| &= \|\boldsymbol{C} + \mathcal{A}^*\big(\boldsymbol{y}_t + \beta_t(\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b})\big)\| \\ &\le \|\boldsymbol{C}\| + K\|\mathcal{A}\| + \beta_t\|\mathcal{A}\|\|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\|.\end{aligned} \tag{SM1.22}$$

We have used the assumption that $\boldsymbol{y}_t$ satisfies the bounded travel condition (SM1.13).

Combine the last three displays to obtain the estimate

$$\begin{aligned}\langle \boldsymbol{D}_t,\, \boldsymbol{H}_t - \boldsymbol{X}_t\rangle &\le \langle \boldsymbol{C},\, \boldsymbol{X}_\star\rangle - f_t(\boldsymbol{X}_t) \\ &+ \Big(\alpha\beta_0\|\mathcal{A}\|\|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\| - \frac{1}{2}\beta_t\|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\|^2\Big) + \frac{\alpha\beta_0}{\beta_t}(\|\boldsymbol{C}\| + K\|\mathcal{A}\|).\end{aligned} \tag{SM1.23}$$

We can now control the decrease in the function f_t between adjacent primal iterates. Combine the displays (SM1.19) and (SM1.23) to arrive at the bound

$$\begin{aligned}&f_t(\boldsymbol{X}_{t+1}) \le (1-\eta_t)f_t(\boldsymbol{X}_t) + \eta_t\langle \boldsymbol{C},\, \boldsymbol{X}_\star\rangle \\ &+ \Big(\eta_t\alpha\beta_0\|\mathcal{A}\|\|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\| - \frac{1}{2}\beta_t\eta_t\|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\|^2\Big) + \Big(\beta_t\eta_t^2\alpha^2\|\mathcal{A}\|^2 + \frac{\eta_t\alpha\beta_0}{\beta_t}(\|\boldsymbol{C}\| + K\|\mathcal{A}\|)\Big).\end{aligned}$$

Subtract $p_\star = \langle \boldsymbol{C},\, \boldsymbol{X}_\star\rangle$ from both sides to arrive at

$$\begin{aligned}&f_t(\boldsymbol{X}_{t+1}) - p_\star \le (1-\eta_t)\Big(f_t(\boldsymbol{X}_t) - p_\star\Big) \\ &+ \Big(\eta_t\alpha\beta_0\|\mathcal{A}\|\|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\| - \frac{1}{2}\beta_t\eta_t\|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\|^2\Big) + \Big(\beta_t\eta_t^2\alpha^2\|\mathcal{A}\|^2 + \frac{\eta_t\alpha\beta_0}{\beta_t}(\|\boldsymbol{C}\| + K\|\mathcal{A}\|)\Big).\end{aligned}$$

Finally, use the definition (SM1.18) again to pass back to the augmented Lagrangian:

$$\begin{aligned}L_t(\boldsymbol{X}_{t+1}; \boldsymbol{y}_t) - p_\star &\le (1-\eta_t)\Big(L_t(\boldsymbol{X}_t; \boldsymbol{y}_t) - p_\star\Big) \\ &+ \Big(\eta_t\alpha\beta_0\|\mathcal{A}\|\|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\| - \frac{1}{2}\beta_t\eta_t\|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\|^2\Big) \\ &+ \Big(\beta_t\eta_t^2\alpha^2\|\mathcal{A}\|^2 + \frac{\eta_t\alpha\beta_0}{\beta_t}(\|\boldsymbol{C}\| + K\|\mathcal{A}\|)\Big).\end{aligned} \tag{SM1.24}$$

This bound describes the evolution of the augmented Lagrangian as the primal iterate advances. But we still need to include the effects of increasing the smoothing parameter and updating the dual variable.

SM1.6.2. Analysis of the smoothing update. To continue, observe that updates to the smoothing parameter have a controlled impact on the augmented Lagrangian (SM1.6):

$$L_t(\boldsymbol{X}_t; \boldsymbol{y}_t) - L_{t-1}(\boldsymbol{X}_t; \boldsymbol{y}_t) = \frac{1}{2}(\beta_t - \beta_{t-1})\|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\|^2.$$

Add and subtract $L_{t-1}(\boldsymbol{X}_t; \boldsymbol{y}_t)$ in the large parenthesis in the first line of (SM1.24) and invoke the last identity to obtain

$$\begin{aligned} L_t(\boldsymbol{X}_{t+1}; \boldsymbol{y}_t) - p_\star &\le (1-\eta_t)\Big(L_{t-1}(\boldsymbol{X}_t; \boldsymbol{y}_t) - p_\star\Big) \\ &+ \Big(\eta_t \alpha \beta_0 \|\mathcal{A}\| \|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\| + \frac{1}{2}\left[(1-\eta_t)(\beta_t - \beta_{t-1}) - \beta_t \eta_t\right] \|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\|^2\Big) \\ &+ \Big(\beta_t \eta_t^2 \alpha^2 \|\mathcal{A}\|^2 + \frac{\eta_t \alpha \beta_0}{\beta_t}(\|\boldsymbol{C}\| + K\|\mathcal{A}\|)\Big). \end{aligned} \tag{SM1.25}$$

The next step is to develop a uniform bound on the terms in the second line so that we can ignore the role of the feasibility gap $\|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\|$ in the subsequent calculations. The choice (SM1.5) of the parameters ensures that

$$(1-\eta_t)(\beta_t - \beta_{t-1}) - \beta_t \eta_t < \frac{-\beta_0^2}{\beta_t}.$$

Introduce this bound into the second line of (SM1.25) and maximize the resulting concave quadratic function to reach

$$\begin{aligned} &\eta_t \alpha \beta_0 \|\mathcal{A}\| \|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\| + \frac{1}{2}\left[(1-\eta_t)(\beta_t - \beta_{t-1}) - \beta_t \eta_t\right] \|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\|^2 \\ &\qquad \le \eta_t \alpha \beta_0 \|\mathcal{A}\| \|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\| - \frac{\beta_0^2}{2\beta_t}\|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\|^2 \le \beta_t \eta_t^2 \alpha^2 \|\mathcal{A}\|^2. \end{aligned}$$

Substitute the last display into (SM1.25) to determine that

$$\begin{aligned} L_t(\boldsymbol{X}_{t+1}; \boldsymbol{y}_t) - p_\star &\le (1-\eta_t)\Big(L_{t-1}(\boldsymbol{X}_t; \boldsymbol{y}_t) - p_\star\Big) \\ &+ \Big(2\beta_t \eta_t^2 \alpha^2 \|\mathcal{A}\|^2 + \frac{\eta_t \alpha \beta_0}{\beta_t}(\|\boldsymbol{C}\| + K\|\mathcal{A}\|)\Big). \end{aligned} \tag{SM1.26}$$

To develop a recursion, we need to assess how the left-hand side changes when we update the dual variable.

SM1.6.3. Analysis of the dual update. To incorporate the dual update, observe that

$$\begin{aligned} L_t(\boldsymbol{X}_{t+1}; \boldsymbol{y}_{t+1}) &= L_t(\boldsymbol{X}_{t+1}; \boldsymbol{y}_t) + \langle \boldsymbol{y}_{t+1} - \boldsymbol{y}_t,\, \mathcal{A}\boldsymbol{X}_{t+1} - \boldsymbol{b}\rangle \\ &= L_t(\boldsymbol{X}_{t+1}; \boldsymbol{y}_t) + \gamma_t \|\mathcal{A}\boldsymbol{X}_{t+1} - \boldsymbol{b}\|^2 \\ &\le L_t(\boldsymbol{X}_{t+1}; \boldsymbol{y}_t) + \beta_t \eta_t^2 \alpha^2 \|\mathcal{A}\|^2. \end{aligned}$$

The first relation is simply the definition (SM1.6) of the augmented Lagrangian, while the second relation depends on the dual update rule (SM1.11). The last step follows from the selection rule (SM1.12) for the dual step size parameter.

Introduce the latter display into (SM1.25) to discover that

$$L_t(\boldsymbol{X}_{t+1}; \boldsymbol{y}_{t+1}) - p_\star \le (1-\eta_t)\Big(L_{t-1}(\boldsymbol{X}_t; \boldsymbol{y}_t) - p_\star\Big) + \left(3\beta_t\eta_t^2\alpha^2\|\mathcal{A}\|^2 + \frac{\eta_t\alpha\beta_0}{\beta_t}(\|\boldsymbol{C}\| + K\|\mathcal{A}\|)\right). \tag{SM1.27}$$

We have developed a recursion for the value of the augmented Lagrangian as the iterates and the smoothing parameter evolve.

SM1.6.4. Solving the recursion. Next, we solve the recursion (SM1.27). We assert that

$$\begin{aligned} L_t(\boldsymbol{X}_{t+1}; \boldsymbol{y}_{t+1}) - p_\star &< \frac{2\beta_0}{\sqrt{t+1}}\left[6\alpha^2\|\mathcal{A}\|^2 + \alpha(\|\boldsymbol{C}\| + K\|\mathcal{A}\|)\right] \\ &=: \frac{2\beta_0 E}{\sqrt{t+1}} \quad \text{for } t = 1, 2, 3, \dots. \end{aligned} \tag{SM1.28}$$

For the case $t = 1$, the definition (SM1.5) ensures that $\eta_1 = 1$ and $\beta_1 = \beta_0\sqrt{2}$, so the bound (SM1.28) follows instantly from (SM1.27). When $t > 1$, an inductive argument using the recursion (SM1.27) and the bound (SM1.28) for $t-1$ ensures that

$$\begin{aligned} L_t(\boldsymbol{X}_{t+1}; \boldsymbol{y}_{t+1}) - p_\star &\le \left[\frac{t-1}{t+1}\cdot\frac{2\beta_0}{\sqrt{t}} + \frac{2\beta_0}{(t+1)^{3/2}}\right]\left[6\alpha^2\|\mathcal{A}\|^2 + \alpha(\|\boldsymbol{C}\| + K\|\mathcal{A}\|)\right] \\ &< \frac{2\beta_0}{\sqrt{t+1}}\left[6\alpha^2\|\mathcal{A}\|^2 + \alpha(\|\boldsymbol{C}\| + K\|\mathcal{A}\|)\right]. \end{aligned}$$

We have introduced the stated values (SM1.5) of the step size and smoothing parameters. The induction proceeds, and we conclude that (SM1.28) is valid.

SM1.6.5. Bound for the suboptimality of the objective. We are prepared to develop an upper bound on the suboptimality of the objective of the model problem (SM1.1). The definition (SM1.6) of the augmented Lagrangian directly implies that

$$\langle \boldsymbol{C},\, \boldsymbol{X}_{t+1}\rangle - p_\star = L_t(\boldsymbol{X}_{t+1}; \boldsymbol{y}_{t+1}) - p_\star - \langle \boldsymbol{y}_{t+1},\, \mathcal{A}\boldsymbol{X}_{t+1} - \boldsymbol{b}\rangle - \frac{1}{2}\beta_t\|\mathcal{A}\boldsymbol{X}_{t+1} - \boldsymbol{b}\|^2. \tag{SM1.29}$$

Continuing from here,

$$\langle \boldsymbol{C},\, \boldsymbol{X}_{t+1}\rangle - p_\star \le \frac{2\beta_0 E}{\sqrt{t+1}} + K\cdot\|\mathcal{A}\boldsymbol{X}_{t+1} - \boldsymbol{b}\|.$$

The first identity follows from definition (SM1.6) of the augmented Lagrangian. The bound relies on (SM1.28) and the Cauchy–Schwarz inequality. We have also used the bounded travel condition (SM1.13). This establishes (SM1.14).

SM1.6.6. Bound for the superoptimality of the objective. The CGAL iterates $\boldsymbol{X}_t$ are generally infeasible for (SM1.1), so they can be superoptimal. Nevertheless, we can easily control the superoptimality. By the saddle-point properties (SM1.3) and (SM1.4), the Lagrangian (SM1.2) satisfies

$$p_\star = \max_{\boldsymbol{y}} \min_{\boldsymbol{X}\in\alpha\boldsymbol{\Delta}_n} L(\boldsymbol{X};\boldsymbol{y}) \le L(\boldsymbol{X}_{t+1};\boldsymbol{y}_\star) = \langle \boldsymbol{C},\, \boldsymbol{X}_{t+1}\rangle + \langle \boldsymbol{y}_\star,\, \mathcal{A}\boldsymbol{X}_{t+1} - \boldsymbol{b}\rangle. \tag{SM1.30}$$

Invoke the Cauchy–Schwarz inequality and rearrange to determine that

$$-\|\boldsymbol{y}_\star\| \cdot \|\mathcal{A}\boldsymbol{X}_{t+1} - \boldsymbol{b}\| \le \langle \boldsymbol{C},\, \boldsymbol{X}_{t+1}\rangle - p_\star.$$

This establishes (SM1.15).

SM1.6.7. Bound for the infeasibility of the iterates. Next, we demonstrate that the iterates converge toward the feasible set of (SM1.1). Combine (SM1.29) and (SM1.30) and rearrange to see that

$$\langle \boldsymbol{y}_{t+1} - \boldsymbol{y}_\star,\, \mathcal{A}\boldsymbol{X}_{t+1} - \boldsymbol{b}\rangle \le L_t(\boldsymbol{X}_{t+1};\boldsymbol{y}_{t+1}) - p_\star - \frac{1}{2}\beta_t\|\mathcal{A}\boldsymbol{X}_{t+1} - \boldsymbol{b}\|^2.$$

Bound the left-hand side below using Cauchy–Schwarz and the right-hand side above using (SM1.28):

$$-\|\boldsymbol{y}_{t+1} - \boldsymbol{y}_\star\| \cdot \|\mathcal{A}\boldsymbol{X}_{t+1} - \boldsymbol{b}\| \le \frac{2\beta_0 E}{\sqrt{t+1}} - \frac{1}{2}\beta_t\|\mathcal{A}\boldsymbol{X}_{t+1} - \boldsymbol{b}\|^2.$$

Solve this quadratic inequality to obtain the bound

$$\begin{aligned}
\|\mathcal{A}\boldsymbol{X}_{t+1} - \boldsymbol{b}\| &\le \beta_t^{-1}\left(\|\boldsymbol{y}_{t+1} - \boldsymbol{y}_\star\| + \sqrt{\|\boldsymbol{y}_{t+1} - \boldsymbol{y}_\star\|^2 + \frac{4\beta_t\beta_0 E}{\sqrt{t+1}}}\right) \\
&\le \beta_t^{-1}\Big(2\|\boldsymbol{y}_{t+1} - \boldsymbol{y}_\star\| + \sqrt{4\beta_0^2 E}\Big) \\
&= \frac{2\beta_0^{-1}\|\boldsymbol{y}_{t+1} - \boldsymbol{y}_\star\| + 2\sqrt{E}}{\sqrt{t+1}}.
\end{aligned}$$

The second inequality depends on the definition (SM1.5) of the smoothing parameter β_t and the subadditivity of the square root.

Finally, we control the dual error using the bounded travel condition (SM1.13):

$$\|\boldsymbol{y}_{t+1} - \boldsymbol{y}_\star\| \le \|\boldsymbol{y}_{t+1}\| + \|\boldsymbol{y}_\star\| \le K + \|\boldsymbol{y}_\star\|.$$

The last two displays yield

$$\|\mathcal{A}\boldsymbol{X}_{t+1} - \boldsymbol{b}\| \le \frac{2\beta_0^{-1}(K + \|\boldsymbol{y}_\star\|) + 2\sqrt{E}}{\sqrt{t+1}}.$$

This confirms (SM1.16).

SM1.7. Computable bounds for suboptimality. In this section, we assume that the linear minimization (SM1.9) is performed exactly at iteration t. That is, there is no error depending on $\|\boldsymbol{D}_t\|$. Introduce the duality gap surrogate

$$g_t(\boldsymbol{X}) := \max_{\boldsymbol{H}\in\alpha\boldsymbol{\Delta}_n} \langle \nabla f_t(\boldsymbol{X}),\, \boldsymbol{X} - \boldsymbol{H}\rangle. \tag{SM1.31}$$

The function f_t is defined in (SM1.18). Note that the gap $g(\boldsymbol{X}_t)$ can be evaluated with the information we have at hand:

$$\begin{aligned} g_t(\boldsymbol{X}_t) &= \langle \boldsymbol{D}_t,\, \boldsymbol{X}_t\rangle - \langle \boldsymbol{D}_t,\, \boldsymbol{H}_t\rangle \\ &= \langle \boldsymbol{C},\, \boldsymbol{X}_t\rangle + \langle \boldsymbol{y}_t + \beta_t(\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}),\, \mathcal{A}\boldsymbol{X}_t\rangle - \langle \boldsymbol{D}_t,\, \boldsymbol{H}_t\rangle. \end{aligned} \tag{SM1.32}$$

The last term is just the value of the linear minimization (SM1.9)

The gap gives us computable bounds on the suboptimality of the current iterate $\boldsymbol{X}_t$. Indeed, the convexity of f_t implies that

$$f_t(\boldsymbol{X}_t) - f_t(\boldsymbol{X}_\star) \le \langle \nabla f_t(\boldsymbol{X}_t),\, \boldsymbol{X}_t - \boldsymbol{X}_\star\rangle \le g_t(\boldsymbol{X}_t).$$

Using the definition (SM1.18) and the fact that $\boldsymbol{X}_\star$ is feasible for (SM1.1), we find that

$$f_t(\boldsymbol{X}_t) - p_\star \;=\; f_t(\boldsymbol{X}_t) - f_t(\boldsymbol{X}_\star) \le g_t(\boldsymbol{X}_t).$$

Invoking the definition (SM1.18) again and rearranging, we find that

$$\langle \boldsymbol{C},\, \boldsymbol{X}_t\rangle - p_\star \le g_t(\boldsymbol{X}_t) - \langle \boldsymbol{y}_t,\, \mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\rangle - \frac{1}{2}\beta_t\|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\|^2. \tag{SM1.33}$$

In other words, the suboptimality of the primal iterate $\boldsymbol{X}_t$ is bounded in terms of the dual gap $g_t(\boldsymbol{X}_t)$, the feasibility gap $\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}$, and the dual variable $\boldsymbol{y}_t$.

Remark SM1.2 (Bounds with approximate linear minimization). We can extend the bound (SM1.33) for the approximate linear minimization in (SM1.9) by taking the error term into account. Based on the definition (SM1.31) of $g_t(\boldsymbol{X}_t)$ and the oracle (SM1.9),

$$g_t(\boldsymbol{X}_t) \le \langle \boldsymbol{D}_t,\, \boldsymbol{X}_t\rangle - \langle \boldsymbol{D}_t,\, \boldsymbol{H}_t\rangle + \frac{\alpha\beta_0}{\beta_t}\|\boldsymbol{D}_t\|.$$

This leads to an extended version of (SM1.33):

$$\langle \boldsymbol{C},\, \boldsymbol{X}_t\rangle - p_\star \le g_t(\boldsymbol{X}_t) - \langle \boldsymbol{y}_t,\, \mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\rangle - \frac{1}{2}\beta_t\|\mathcal{A}\boldsymbol{X}_t - \boldsymbol{b}\|^2 + \frac{\alpha\beta_0}{\beta_t}\|\boldsymbol{D}_t\|. \tag{SM1.34}$$

Note that the additional error term converges to zero. We can simplify this bound further by using (SM1.22).

In practice, we have observed that the bound (SM1.34) is less informative than simply using the bound (SM1.33), which assumes that the eigenvalue computation is exact.

Remark SM1.3 (Superoptimality). Note that the suboptimality $\langle \boldsymbol{C},\, \boldsymbol{X}_t\rangle - p_\star$ can attain negative values because the CGAL iterates $\boldsymbol{X}_t$ are generally infeasible. Similarly, the upper bound in (SM1.33) can be negative. In other words, the iterates $\boldsymbol{X}_t$ can be superoptimal. Nevertheless, the superoptimality is controlled by the feasibility gap, as shown in subsection SM1.6.6.

SM2. Sketching and reconstruction of a positive-semidefinite matrix.

This section reviews and gives additional details about the Nyström sketch [SM11, SM9, SM13, SM22]. This sketch tracks an evolving psd matrix and then reports a provably accurate low-rank approximation. The material on error estimation is new.

SM2.1. Sketching and updates.

Consider a psd input matrix $\boldsymbol{X} \in \mathbb{S}_n$. Let R be a parameter that modulates the storage cost of the sketch and the quality of the matrix approximation.

To construct the sketch, we draw and fix a standard normal test matrix $\boldsymbol{\Omega} \in \mathbb{F}^{n \times R}$. The sketch of the matrix $\boldsymbol{X}$ takes the form

$$\boldsymbol{S} = \boldsymbol{X}\boldsymbol{\Omega} \in \mathbb{F}^{n \times R} \quad \text{and} \quad \tau = \operatorname{tr}(\boldsymbol{X}) \in \mathbb{R}. \tag{SM2.1}$$

The sketch supports linear rank-one updates to $\boldsymbol{X}$. Indeed, we can track the evolution

$$\begin{aligned} &\boldsymbol{X} \leftarrow (1-\eta)\,\boldsymbol{X} + \eta\,\boldsymbol{v}\boldsymbol{v}^* \qquad \text{for } \eta \in [0,1] \text{ and } \boldsymbol{v} \in \mathbb{F}^n \\ \text{via} \quad &\boldsymbol{S} \leftarrow (1-\eta)\,\boldsymbol{S} + \eta\,\boldsymbol{v}(\boldsymbol{v}^*\boldsymbol{\Omega}) \\ \text{and} \quad &\tau \leftarrow (1-\eta)\,\tau + \eta\,\|\boldsymbol{v}\|^2. \end{aligned} \tag{SM2.2}$$

The test matrix $\boldsymbol{\Omega}$ and the sketch $(\boldsymbol{S}, \tau)$ require storage of $2Rn + 1$ numbers in $\mathbb{F}$. The arithmetic cost of the linear update (5.2) to the sketch is $\Theta(Rn)$ numerical operations.

Remark SM2.1 (Structured random matrices). We can reduce storage costs by a factor of two by using a structured random matrix in place of $\boldsymbol{\Omega}$. For example, see [SM24, Sec. 3] or [SM20]. This modification requires additional care with implementation (e.g., use of sparse arithmetic or fast trigonometric transforms), but the improvement can be significant for very large problems.

SM2.2. The reconstruction process.

Given the test matrix $\boldsymbol{\Omega}$ and the sketch $\boldsymbol{S} = \boldsymbol{X}\boldsymbol{\Omega}$, we can form a rank-$R$ approximation $\widehat{\boldsymbol{X}}$ of the matrix $\boldsymbol{X}$ contained in the sketch. This approximation is defined by the formula

$$\widehat{\boldsymbol{X}} := \boldsymbol{S}(\boldsymbol{\Omega}^*\boldsymbol{S})^\dagger \boldsymbol{S}^* = (\boldsymbol{X}\boldsymbol{\Omega})(\boldsymbol{\Omega}^*\boldsymbol{X}\boldsymbol{\Omega})^\dagger(\boldsymbol{X}\boldsymbol{\Omega})^*. \tag{SM2.3}$$

This reconstruction is called a *Nyström approximation*. We often truncate $\widehat{\boldsymbol{X}}$ by replacing it with its best rank-r approximation $[\![\widehat{\boldsymbol{X}}]\!]_r$ for some $r \le R$.

See Algorithm SM2.1 for a numerically stable implementation of the Nyström reconstruction process (5.3), including error estimation steps. The algorithm takes $O(R^2 n)$ numerical operations and $\Theta(Rn)$ storage.

SM2.3. The approximation error.

We can easily determine the exact Schatten 1-norm error in the truncated Nyström approximation:

$$\operatorname{err}([\![\widehat{\boldsymbol{X}}]\!]_r) := \|\boldsymbol{X} - [\![\widehat{\boldsymbol{X}}]\!]_r\|_* = \tau - \operatorname{tr}([\![\widehat{\boldsymbol{X}}]\!]_r) \quad \text{for each } r \le R. \tag{SM2.4}$$

Furthermore, we can use (SM2.4) to ascertain whether the unknown input matrix $\boldsymbol{X}$ is (almost) low-rank. Indeed, the best rank-r approximation of $\boldsymbol{X}$ satisfies

$$\|\boldsymbol{X} - [\![\boldsymbol{X}]\!]_r\|_* \le \operatorname{err}([\![\widehat{\boldsymbol{X}}]\!]_r) \quad \text{for each } r \le R. \tag{SM2.5}$$

Thus, large drops in the function $r \mapsto \operatorname{err}([\![\widehat{\boldsymbol{X}}]\!]_r)$ signal large drops in the eigenvalues of $\boldsymbol{X}$. See subsection SM2.9 for the details.

Algorithm SM2.1 NystromSketch implementation (see section SM2)

Input: Dimension n of input matrix, size R of sketch
Output: Rank-R approximation $\widehat{\boldsymbol{X}}$ of sketched matrix in factored form $\widehat{\boldsymbol{X}} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^*$, where $\boldsymbol{U} \in \mathbb{R}^{n\times R}$ has orthonormal columns and $\boldsymbol{\Lambda} \in \mathbb{R}^{R\times R}$ is nonnegative diagonal, and the Schatten 1-norm approximation errors $\mathrm{err}([\![\widehat{\boldsymbol{X}}]\!]_r)$ for $1 \le r \le R$, as defined in (SM2.4)
Recommendation: Choose R as large as possible

function NystromSketch.Init(n, R)
  $\boldsymbol{\Omega} \leftarrow \mathtt{randn}(n, R)$ ▷ Draw and fix random test matrix
  $\boldsymbol{S} \leftarrow \mathtt{zeros}(n, R)$ and $\tau \leftarrow 0$ ▷ Form sketch of zero matrix

function NystromSketch.RankOneUpdate($\boldsymbol{v}$, η) ▷ Implements (5.2)
  $\boldsymbol{S} \leftarrow (1-\eta)\,\boldsymbol{S} + \eta\,\boldsymbol{v}(\boldsymbol{v}^*\boldsymbol{\Omega})$ ▷ Update sketch of matrix
  $\tau \leftarrow (1-\eta)\,\tau + \eta\,\|\boldsymbol{v}\|^2$ ▷ Update the trace

function NystromSketch.Reconstruct( )
  $\sigma \leftarrow \sqrt{n}\,\mathtt{eps}(\mathtt{norm}(\boldsymbol{S}))$ ▷ Compute a shift parameter
  $\boldsymbol{S} \leftarrow \boldsymbol{S} + \sigma\,\boldsymbol{\Omega}$ ▷ Implicitly form sketch of $\boldsymbol{X} + \sigma\,\mathbf{I}$
  $\boldsymbol{L} \leftarrow \mathtt{chol}(\boldsymbol{\Omega}^*\boldsymbol{S})$
  $[\boldsymbol{U}, \boldsymbol{\Sigma}, \sim] \leftarrow \mathtt{svd}(\boldsymbol{S}/\boldsymbol{L})$ ▷ Dense SVD
  $\boldsymbol{\Lambda} \leftarrow \max\{0, \boldsymbol{\Sigma}^2 - \sigma\,\mathbf{I}\}$ ▷ Remove shift
  $\mathbf{err} \leftarrow \tau - \mathtt{cumsum}(\mathtt{diag}(\boldsymbol{\Lambda}))$ ▷ Compute approximation errors

SM2.4. Trace correction. The trace of the Nyström approximation $\widehat{\boldsymbol{X}}$ does not exceed the trace of the input matrix $\boldsymbol{X}$. It can sometimes be helpful to replace the Nyström approximation by another matrix whose trace matches the trace of the input matrix.

Suppose that the Nyström approximation has the form $\widehat{\boldsymbol{X}} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^*$, where $\boldsymbol{U} \in \mathbb{F}^{n\times R}$ is orthonormal and $\boldsymbol{\Lambda} \in \mathbb{S}_R^+$ is diagonal. Assume that $\mathrm{tr}(\boldsymbol{X}) = \alpha$ and $\mathrm{tr}(\widehat{\boldsymbol{X}}) = \widehat{\alpha}$. Since $\widehat{\boldsymbol{X}}$ is a Nyström approximation of $\boldsymbol{X}$, it holds that $\widehat{\alpha} \le \alpha$. Now, we can construct a second approximation

$$\widetilde{\boldsymbol{X}} := \boldsymbol{U}\left(\boldsymbol{\Lambda} + (\alpha - \widehat{\alpha})\mathbf{I}_R/R\right)\boldsymbol{U}^*. \tag{SM2.6}$$

It is evident that $\widehat{\boldsymbol{X}} \preccurlyeq \widetilde{\boldsymbol{X}}$ and that $\mathrm{tr}\,\widetilde{\boldsymbol{X}} = \alpha$.

One might hope that the trace-corrected approximation $\widetilde{\boldsymbol{X}}$ is better than the original Nyström approximation $\widehat{\boldsymbol{X}}$, but this is not necessarily the case. Fortunately, we always have the relation

$$\|\boldsymbol{X} - \widetilde{\boldsymbol{X}}\|_* \le 2\|\boldsymbol{X} - \widehat{\boldsymbol{X}}\|_*.$$

In other words, correcting the trace doubles the error, at worst. See subsection SM2.10 for the proof.

SM2.5. Statistical properties of the Nyström sketch. The truncated Nyström approximation has a number of attractive statistical properties. For a fixed input matrix $\boldsymbol{X}$, the expected approximation error $\mathbb{E}_{\boldsymbol{\Omega}}\,\|\boldsymbol{X} - [\![\widehat{\boldsymbol{X}}]\!]_r\|_*$ is monotone decreasing in both the sketch size

R and the truncation rank r. Furthermore, if $\operatorname{rank}(\boldsymbol{X}) = r$ for $r \le R$, then $\|\boldsymbol{X} - [\![\widehat{\boldsymbol{X}}]\!]_r\|_* = 0$ with probability one. We establish these results below in subsection SM2.11.

SM2.6. A priori error bounds. The Nyström approximation $\widehat{\boldsymbol{X}}$ yields a provably good estimate for the matrix $\boldsymbol{X}$ contained in the sketch [SM22, Thms. 4.1].

Fact SM2.2 (Nyström sketch: Error bound). *Fix a psd matrix $\boldsymbol{X} \in \mathbb{S}_n$. Let $\boldsymbol{S} = \boldsymbol{X\Omega}$ where $\boldsymbol{\Omega} \in \mathbb{F}^{n\times R}$ is standard normal. For each $r < R-1$, the Nyström approximation (SM2.3) satisfies*

$$\mathbb{E}_{\boldsymbol{\Omega}} \|\boldsymbol{X} - \widehat{\boldsymbol{X}}\|_* \le \left(1 + \frac{r}{R-r-1}\right) \|\boldsymbol{X} - [\![\boldsymbol{X}]\!]_r\|_*. \tag{SM2.7}$$

If we replace $\widehat{\boldsymbol{X}}$ with the rank-r truncation $[\![\widehat{\boldsymbol{X}}]\!]_r$, the error bound (SM2.7) remains valid. Similar results hold with high probability.

The truncated Nyström approximations satisfy a stronger error bound when the input matrix $\boldsymbol{X}$ exhibits spectral decay.

Fact SM2.3 (Nyström sketch: Error bound II). *Fix a psd matrix $\boldsymbol{X} \in \mathbb{S}_n$. Let $\boldsymbol{S} = \boldsymbol{X\Omega}$ where $\boldsymbol{\Omega} \in \mathbb{F}^{n\times R}$ is standard normal. For each $r < R-1$, the Nyström approximation (SM2.3) satisfies*

$$\mathbb{E}_{\boldsymbol{\Omega}} \|\boldsymbol{X} - [\![\widehat{\boldsymbol{X}}]\!]_r\|_* \le \|\boldsymbol{X} - [\![\boldsymbol{X}]\!]_r\|_* + \left(1 + \frac{r}{R-r-1}\right) \|\boldsymbol{X} - [\![\boldsymbol{X}]\!]_r\|_*. \tag{SM2.8}$$

If we replace $\widehat{\boldsymbol{X}}$ with the rank-r truncation $[\![\widehat{\boldsymbol{X}}]\!]_r$, the error bound (SM2.8) remains valid. Similar results hold with high probability.

Proof. Combine the proofs of [SM22, Thm. 4.2] and [SM11, Thm. 9.3]. ■

SM2.7. Discussion. In practice, it is best to minimize the error attributable to sketching. To that end, we recommend choosing the sketch size parameter R as large as possible, given resource constraints, so that we can obtain the highest-quality Nyström approximation.

In some problems, e.g., MaxCut with eigenvector rounding, the desired rank r of the truncation is known in advance. In this case, Fact 5.2 offers guidance about how to select R to achieve a specific error tolerance $(1+\zeta)$ in (2.3). For example, when $5r + 1 \le R$, the expected Schatten 1-norm error in the rank-r approximation $[\![\widehat{\boldsymbol{X}}]\!]_r$ is at most $1.25\times$ the error in the best rank-r approximation of $\boldsymbol{X}$.

When the input matrix $\boldsymbol{X}$ has decaying eigenvalues, the error in the truncated approximation may be far smaller than Fact 5.2 predicts; see [SM22, Thm. 4.2]. This happy situation is typical when $\boldsymbol{X}$ is generated by the CGAL iteration.

SM2.8. Representation of the truncated Nyström approximation. The key tool in the analysis is a simple representation for the truncated approximation. These facts are extracted from [SM22, Supp.].

Let $\boldsymbol{X} \in \mathbb{S}_n$ be a fixed psd matrix, and let $\boldsymbol{\Omega} \in \mathbb{R}^{n\times R}$ be an arbitrary test matrix. Let $\boldsymbol{P} \in \mathbb{S}_n$ be the orthoprojector onto the range of $\boldsymbol{X}^{1/2}\boldsymbol{\Omega}$. Then we can write the Nyström approximation (SM2.3) as

$$\widehat{\boldsymbol{X}} = \boldsymbol{X}^{1/2}\boldsymbol{P}\boldsymbol{X}^{1/2}.$$

For each $r \le R$, define $\boldsymbol{P}_r$ to be the orthoprojector onto the co-range of the matrix $[\![\boldsymbol{X}^{1/2}\boldsymbol{P}]\!]_r$. By construction, $[\![\boldsymbol{X}^{1/2}\boldsymbol{P}]\!]_r = \boldsymbol{X}^{1/2}\boldsymbol{P}_r$. As a consequence,

$$[\![\widehat{\boldsymbol{X}}]\!]_r = ([\![\boldsymbol{X}^{1/2}\boldsymbol{P}]\!]_r)([\![\boldsymbol{X}^{1/2}\boldsymbol{P}]\!]_r)^* = \boldsymbol{X}^{1/2}\boldsymbol{P}_r\boldsymbol{X}^{1/2}.$$

These results allow us to relate the truncated approximations to each other:

SM2.9. Approximation errors: Analysis. We may now obtain explicit formulas for the error in each Nyström approximation. For $r \le R$, note that

$$\boldsymbol{X} - [\![\widehat{\boldsymbol{X}}]\!]_r = \boldsymbol{X}^{1/2}(\mathbf{I} - \boldsymbol{P}_r)\boldsymbol{X}^{1/2} \succcurlyeq \boldsymbol{0}.$$

It follows immediately that

$$\mathrm{err}([\![\widehat{\boldsymbol{X}}]\!]_r) = \|\boldsymbol{X} - [\![\widehat{\boldsymbol{X}}]\!]_r\|_* = \mathrm{tr}(\boldsymbol{X} - [\![\widehat{\boldsymbol{X}}]\!]_r) = \mathrm{tr}(\boldsymbol{X}) - \mathrm{tr}([\![\widehat{\boldsymbol{X}}]\!]_r).$$

This is the relation (SM2.4).

Assume that $r \le r'$. Since $\mathrm{tr}([\![\widehat{\boldsymbol{X}}]\!]_r) \le \mathrm{tr}([\![\widehat{\boldsymbol{X}}]\!]_{r'})$, we have the bound

$$\mathrm{err}([\![\widehat{\boldsymbol{X}}]\!]_r) \ge \mathrm{err}([\![\widehat{\boldsymbol{X}}]\!]_{r'}) \quad \text{for } r \le r'.$$

In other words, for fixed sketch size R, the error in the truncated Nyström approximation is monotone decreasing in the approximation rank.

To obtain (SM2.5), observe that

$$\|\boldsymbol{X} - [\![\boldsymbol{X}]\!]_r\|_* \le \|\boldsymbol{X} - [\![\widehat{\boldsymbol{X}}]\!]_r\|_* = \mathrm{err}([\![\widehat{\boldsymbol{X}}]\!]_r).$$

The inequality holds because $[\![\boldsymbol{X}]\!]_r$ is a best rank-r approximation of $\boldsymbol{X}$ in Schatten 1-norm, while $[\![\widehat{\boldsymbol{X}}]\!]_r$ is another rank-r matrix.

SM2.10. Trace correction: Analysis. Next, we study the trace-corrected Nyström approximation, introduced in subsection SM2.4. Recall that $\boldsymbol{X}$ is a psd matrix with trace α and $\widehat{\boldsymbol{X}}$ is a Nyström approximation of $\boldsymbol{X}$ with $\mathrm{tr}(\widehat{\boldsymbol{X}}) = \widehat{\alpha}$. Owing to the projection representation of $\widehat{\boldsymbol{X}}$ in subsection SM2.8, it must be the case that $\widehat{\alpha} \le \alpha$. The trace-corrected approximation $\widetilde{\boldsymbol{X}}$ is defined in (SM2.6). By construction, this matrix satisfies $\widetilde{\boldsymbol{X}} \succcurlyeq \widehat{\boldsymbol{X}}$ and that $\mathrm{tr}(\widetilde{\boldsymbol{X}}) = \alpha$.

First, we develop a variational interpretation of the trace-corrected approximation:

$$\widetilde{\boldsymbol{X}} \in \arg\min\{\|\boldsymbol{Y} - \widehat{\boldsymbol{X}}\|_* : \mathrm{tr}(\boldsymbol{Y}) = \alpha \text{ and } \boldsymbol{Y} \succcurlyeq \boldsymbol{0}\}. \tag{SM2.9}$$

Indeed, for any feasible point $\boldsymbol{Y}$,

$$\|\boldsymbol{Y} - \widehat{\boldsymbol{X}}\|_* \ge \mathrm{tr}(\boldsymbol{Y} - \widehat{\boldsymbol{X}}) = \alpha - \widehat{\alpha}.$$

In the first relation, equality holds if and only if $\boldsymbol{Y} \succcurlyeq \widehat{\boldsymbol{X}}$. We have already seen that the matrix $\widetilde{\boldsymbol{X}}$ is a feasible point that satisfies the equality condition. There are many solutions to the optimization problem (SM2.9); we have singled out $\widetilde{\boldsymbol{X}}$ as the one that simultaneously minimizes the Frobenius-norm error $\|\boldsymbol{Y} - \widehat{\boldsymbol{X}}\|_{\mathrm{F}}$ over the same feasible set.

Second, we need to argue that trace correction has a controlled impact on the error in the Nyström approximation. This point follows from a standard calculation:

$$\|\widetilde{\boldsymbol{X}} - \boldsymbol{X}\|_* \le \|\widetilde{\boldsymbol{X}} - \widehat{\boldsymbol{X}}\|_* + \|\boldsymbol{X} - \widehat{\boldsymbol{X}}\|_* \le 2\|\boldsymbol{X} - \widehat{\boldsymbol{X}}\|_*.$$

The first relation is the triangle inequality. The second relation holds because $\widetilde{\boldsymbol{X}}$ solves the variational problem (SM2.9), while $\boldsymbol{X}$ is also feasible for this optimization problem.

SM2.11. Statistical properties: Analysis. Next, let us verify the statistical properties of the error. In this section, the test matrix $\boldsymbol{\Omega} \in \mathbb{F}^{n \times R}$ is standard normal.

Assuming that $\operatorname{rank}(\boldsymbol{X}) = r \le R$, let us prove that $\|\boldsymbol{X} - [\![\widehat{\boldsymbol{X}}]\!]_r\|_* = 0$ with probability one. To that end, we observe

$$\operatorname{range}(\boldsymbol{P}) = \operatorname{range}(\boldsymbol{X}^{1/2}\boldsymbol{\Omega}) = \operatorname{range}(\boldsymbol{X}^{1/2}) \quad \text{with probability one.}$$

It follows that

$$\widehat{\boldsymbol{X}} = \boldsymbol{X}^{1/2}\boldsymbol{P}\boldsymbol{X}^{1/2} = \boldsymbol{X} \quad \text{with probability one.}$$

Moreover, $\operatorname{rank}(\boldsymbol{X}^{1/2}\boldsymbol{P}) = r$ with probability one. Conditional on this event,

$$[\![\widehat{\boldsymbol{X}}]\!]_r = \boldsymbol{X}^{1/2}\boldsymbol{P}_r\boldsymbol{X}^{1/2} = ([\![\boldsymbol{X}^{1/2}\boldsymbol{P}]\!]_r)([\![\boldsymbol{X}^{1/2}\boldsymbol{P}]\!]_r)^* = (\boldsymbol{X}^{1/2}\boldsymbol{P})(\boldsymbol{X}^{1/2}\boldsymbol{P})^* = \boldsymbol{X}.$$

This is the stated result.

Next, we show that the expected error in the truncated Nyström approximation is monotone decreasing with respect to the sketch size R. Fix the truncation rank r. Let $\boldsymbol{\Omega}_+ = \begin{bmatrix} \boldsymbol{\Omega} & \boldsymbol{\omega} \end{bmatrix}$, where $\boldsymbol{\omega} \in \mathbb{F}^n$ is a standard normal vector independent from $\boldsymbol{\Omega}$. Define $\boldsymbol{P}_+ \in \mathbb{S}_n$ to be the orthoprojector onto $\operatorname{range}(\boldsymbol{X}^{1/2}\boldsymbol{\Omega}_+)$. It is clear that $\operatorname{range}(\boldsymbol{P}) \subseteq \operatorname{range}(\boldsymbol{P}_+)$, and so

$$\boldsymbol{X}^{1/2}\boldsymbol{P}\boldsymbol{X}^{1/2} \preccurlyeq \boldsymbol{X}^{1/2}\boldsymbol{P}_+\boldsymbol{X}^{1/2}.$$

As a consequence,

$$\operatorname{tr}([\![\boldsymbol{X}^{1/2}\boldsymbol{P}\boldsymbol{X}^{1/2}]\!]_r) \le \operatorname{tr}([\![\boldsymbol{X}^{1/2}\boldsymbol{P}_+\boldsymbol{X}^{1/2}]\!]_r).$$

Equivalently,

$$\|\boldsymbol{X} - [\![\boldsymbol{X}^{1/2}\boldsymbol{P}\boldsymbol{X}^{1/2}]\!]_r\|_* \ge \|\boldsymbol{X} - [\![\boldsymbol{X}^{1/2}\boldsymbol{P}_+\boldsymbol{X}^{1/2}]\!]_r\|_*.$$

Take the expectation with respect to $\boldsymbol{\Omega}_+$. The left-hand side is the average error in the r-truncated Nyström approximation with a standard normal sketch of size R. The right-hand side is the same thing, except the sketch has size $R+1$. This is the required result.

SM3. SketchyCGAL: Additional results. This section contains some additional material about the SketchyCGAL algorithm.

SM3.1. Assessing solution quality. We can develop estimates for the quality of the SketchyCGAL solution by adapted the approach that we used for CGAL.

To do so, we need to track the primal objective value at the current iterate:

$$p_t = \langle \boldsymbol{C},\, \boldsymbol{X}_t \rangle.$$

At each iteration, we can easily update this estimate using the computed approximate eigenvector $\boldsymbol{v}_t$:

$$p_{t+1} = (1-\eta_t)p_t + \eta_t\alpha\langle \boldsymbol{v}_t,\, \boldsymbol{C}\boldsymbol{v}_t\rangle.$$

This update rule is applied with the help of the primitive (2.4)❶.

When we wish to estimate the error, say in iteration t, we can solve the eigenvalue subproblem to very high accuracy:

$$\xi_t = \boldsymbol{v}_t^*\boldsymbol{D}_t\boldsymbol{v}_t = \min_{\|\boldsymbol{v}\|=1} \boldsymbol{v}^*\boldsymbol{D}_t\boldsymbol{v}.$$

Then, we can compute the surrogate duality gap:

$$g_t(\boldsymbol{X}_t) = p_t + \langle \boldsymbol{y}_t + \beta_t(\boldsymbol{z}_t - \boldsymbol{b}),\, \boldsymbol{z}_t \rangle - \xi_t.$$

This expression follows directly from the formula (SM1.32) using the loop invariant that $\boldsymbol{z}_t = \mathcal{A}\boldsymbol{X}_t$. We arrive at a computable error estimate:

$$p_t - p_\star \leq g_t(\boldsymbol{X}_t) - \langle \boldsymbol{y}_t,\, \boldsymbol{z}_t - \boldsymbol{b} \rangle - \frac{1}{2}\beta_t \|\boldsymbol{z}_t - \boldsymbol{b}\|^2.$$

This bound follows directly from (SM1.33).

SM3.2. Convergence theory. In this section, we establish two simple results on the convergence properties of the SketchyCGAL algorithm.

Theorem SM3.1 (SketchyCGAL: Convergence I). *Let $\Psi_\star$ be the solution set of the model problem (2.2). For each $r < R-1$, the iterates $\widehat{\boldsymbol{X}}_t$ computed by SketchyCGAL (subsections 6.1 and 6.2) satisfy*

$$\limsup_{t\to\infty} \mathbb{E}_{\boldsymbol{\Omega}} \operatorname{dist}_*(\widehat{\boldsymbol{X}}_t, \Psi_\star) \leq \left(1 + \frac{r}{R-r-1}\right) \cdot \max_{\boldsymbol{X}\in\Psi_\star} \|\boldsymbol{X} - [\![\boldsymbol{X}]\!]_r\|_*.$$

The same bound holds for the truncated approximations $[\![\widehat{\boldsymbol{X}}_t]\!]_r$. Here, dist_* is the nuclear-norm distance between a matrix and a set of matrices.

Proof. The implicit iterates $\boldsymbol{X}_t$ satisfy the conclusions of Fact 3.1, so they converge toward the compact set $\Psi_\star$. Therefore, we can choose a sequence $\{\boldsymbol{X}_{t\star}\} \subset \Psi_\star$ with the property that $\|\boldsymbol{X}_t - \boldsymbol{X}_{t\star}\|_* \to 0$. By the triangle inequality and (6.7),

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\Omega}} \operatorname{dist}_*(\widehat{\boldsymbol{X}}_t, \Psi_\star) &\leq \mathbb{E}_{\boldsymbol{\Omega}} \|\widehat{\boldsymbol{X}}_t - \boldsymbol{X}_t\|_* + \operatorname{dist}_*(\boldsymbol{X}_t, \Psi_\star)\\ &\leq \left(1 + \frac{r}{R-r-1}\right) \cdot \|\boldsymbol{X}_t - [\![\boldsymbol{X}_t]\!]_r\|_* + \|\boldsymbol{X}_t - \boldsymbol{X}_{t\star}\|_*.\end{aligned}$$

The rank-r approximation error in Schatten 1-norm is 1-Lipschitz with respect to the Schatten 1-norm (cf. [SM24, Sec. SM2.2]), so

$$\big|\|\boldsymbol{X}_t - [\![\boldsymbol{X}_t]\!]_r\|_* - \|\boldsymbol{X}_{t\star} - [\![\boldsymbol{X}_{t\star}]\!]_r\|_*\big| \leq \|\boldsymbol{X}_t - \boldsymbol{X}_{t\star}\|_*.$$

Therefore,

$$\begin{aligned}\|\boldsymbol{X}_t - [\![\boldsymbol{X}_t]\!]_r\|_* &\leq \|\boldsymbol{X}_{t\star} - [\![\boldsymbol{X}_{t\star}]\!]_r\|_* + \|\boldsymbol{X}_t - \boldsymbol{X}_{t\star}\|_*\\ &\leq \max_{\boldsymbol{X}\in\Psi_\star} \|\boldsymbol{X} - [\![\boldsymbol{X}]\!]_r\|_* + \|\boldsymbol{X}_t - \boldsymbol{X}_{t\star}\|_*.\end{aligned}$$

Combine the last two displays, and extract the superior limit. ∎

If the implicit iterates generated by SketchyCGAL happen to converge to a limit, we have a more precise result.

Theorem SM3.2 (SketchyCGAL: Convergence II). *Assume the implicit iterates $\boldsymbol{X}_t$ induced by SketchyCGAL (subsections 6.1 and 6.2) converge to a matrix $\boldsymbol{X}_{\text{cgal}}$. For each $r < R - 1$, the computed iterates $\widehat{\boldsymbol{X}}_t$ satisfy*

$$\limsup_{t\to\infty} \mathbb{E}_{\boldsymbol{\Omega}} \|\mathcal{A}\widehat{\boldsymbol{X}}_t - \boldsymbol{b}\| \leq \left(1 + \frac{r}{R-r-1}\right) \cdot \|\mathcal{A}\| \cdot \|\boldsymbol{X}_{\text{cgal}} - [\![\boldsymbol{X}_{\text{cgal}}]\!]_r\|_*;$$
$$\limsup_{t\to\infty} \mathbb{E}_{\boldsymbol{\Omega}} |\langle \boldsymbol{C}, \widehat{\boldsymbol{X}}_t\rangle - \langle \boldsymbol{C}, \boldsymbol{X}_\star\rangle| \leq \left(1 + \frac{r}{R-r-1}\right) \cdot \|\boldsymbol{C}\| \cdot \|\boldsymbol{X}_{\text{cgal}} - [\![\boldsymbol{X}_{\text{cgal}}]\!]_r\|_*.$$

The same bound holds for the truncated approximations $[\![\widehat{\boldsymbol{X}}_t]\!]_r$. If $\operatorname{rank}(\boldsymbol{X}_{\text{cgal}}) \leq R$, then the computed iterates $\widehat{\boldsymbol{X}}_t$ converge to the solution set of (2.2).

Proof. The implicit iterates $\boldsymbol{X}_t$ satisfy the conclusions of Fact 3.1, so the limit $\boldsymbol{X}_{\text{cgal}}$ solves (2.2). Using the triangle inequality, the operator norm bound, and (6.7), we obtain nonasymptotic error bounds

$$\mathbb{E}_{\boldsymbol{\Omega}} \|\mathcal{A}\widehat{\boldsymbol{X}}_t - \boldsymbol{b}\| \leq \frac{\text{Const}}{\sqrt{t}} + \left(1 + \frac{r}{R-r-1}\right) \cdot \|\mathcal{A}\| \cdot \|\boldsymbol{X}_t - [\![\boldsymbol{X}_t]\!]_r\|_*;$$
$$\mathbb{E}_{\boldsymbol{\Omega}} |\langle \boldsymbol{C}, \widehat{\boldsymbol{X}}_t\rangle - \langle \boldsymbol{C}, \boldsymbol{X}_\star\rangle| \leq \frac{\text{Const}}{\sqrt{t}} + \left(1 + \frac{r}{R-r-1}\right) \cdot \|\boldsymbol{C}\| \cdot \|\boldsymbol{X}_t - [\![\boldsymbol{X}_t]\!]_r\|_*.$$

Extract the limit as $t \to \infty$. The last conclusion follows from the facts outlined in subsection SM2.6. ∎

SM4. Beyond the model problem.

The CGAL algorithm [SM27] applies to a more general problem template than (2.2). Likewise, the SketchyCGAL algorithm can solve a wider class of problems in a scalable fashion. This section outlines some of the opportunities.

SM4.1. A more general template.

Consider the optimization problem

$$\text{(SM4.1)} \qquad \text{minimize} \quad \langle \boldsymbol{C}, \boldsymbol{X}\rangle \quad \text{subject to} \quad \mathcal{A}\boldsymbol{X} \in \mathsf{K} \quad \text{and} \quad \boldsymbol{X} \in \mathsf{X}, \quad \boldsymbol{X} \text{ is psd.}$$

In this expression, $\mathsf{K} \subset \mathbb{R}^d$ is a closed, convex set and $\mathsf{X} \subset \mathbb{S}_n(\mathbb{F})$ is a compact, convex set of matrices. The rest of this section describes some problems that fall within the compass of (SM4.1), as well as new computational challenges that appear. Algorithm SM4.1 contains pseudocode for a version of SketchyCGAL tailored to (SM4.1).

Remark SM4.1 (Other matrix optimization problems). We can also extend SketchyCGAL to optimization problems involving matrices that are symmetric (but not psd) or that are rectangular. For example, matrix completion via nuclear-norm minimization [SM17] falls in this framework. In this case, we need to replace the Nyström sketch with a more general technique, such as [SM23, SM24]. Further extensions are also possible; see [SM27]. We omit these developments.

SM4.2. The convex constraint set.

To handle the convex constraint X that appears in (SM4.1), we must develop a subroutine for the linear minimization problem

$$\text{(SM4.2)} \qquad \underset{\boldsymbol{H}\in\mathbb{S}_n}{\text{minimize}} \quad \langle \boldsymbol{D}_t, \boldsymbol{H}\rangle \quad \text{subject to} \quad \boldsymbol{H} \in \mathsf{X}, \quad \boldsymbol{H} \text{ is psd.}$$

To implement SketchyCGAL efficiently, we need the problem (SM4.2) to admit a structured (e.g., low-rank) approximate solution. Here are some situations where this is possible.

1. **Trace-bounded psd matrices.** $\mathsf{X} := \{\boldsymbol{X} \in \mathbb{S}_n : \operatorname{tr} \boldsymbol{X} \le \alpha \text{ and } \boldsymbol{X} \text{ is psd}\}$. For solving a standard-form SDP, this constraint is more natural than $\mathsf{X} = \alpha \boldsymbol{\Delta}_n$. Given an (approximate) minimum eigenpair $(\xi_t, \boldsymbol{v}_t)$ of $\boldsymbol{D}_t$, the solution of (SM4.2) is

$$\boldsymbol{H}_t = \begin{cases} \alpha \boldsymbol{v}_t \boldsymbol{v}_t^*, & \xi_t < 0, \\ \boldsymbol{0}, & \xi_t \ge 0. \end{cases}$$

   As before, we can solve the eigenvector problem with Algorithm 4.1.
2. **Relaxed orthoprojectors.** $\mathsf{X} := \{\boldsymbol{X} \in \mathbb{S}_n : \operatorname{tr} \boldsymbol{X} = \alpha \text{ and } \boldsymbol{0} \preccurlyeq \boldsymbol{X} \preccurlyeq \mathbf{I}\}$. This is the best convex relaxation of the set of orthogonal projectors with rank α; see [SM16]. When α is small, we can provably solve the linear minimization with randomized subspace iteration [SM11] or randomized block Lanczos methods [SM10, Sec. 10.3.6].

Remark SM4.2 (Standard-form SDP). It is often easy to find an upper bound for the trace of a solution for an SDP. In many applications, the constraint $\mathcal{A}\boldsymbol{X} = \boldsymbol{b}$ already enforces a constant trace (e.g., MaxCut or QAP). In many other applications (e.g., phase retrieval), $\boldsymbol{C}$ is identity matrix, so the objective is to minimize the trace. In this setting, the trace of an arbitrary feasible point is an upper bound for α. If neither situation is in force, we can still solve a standard-form SDP by solving a small number of trace-bounded SDPs:
1. Start with an arbitrary bound $\alpha_0 > 0$, say, twice the trace of an arbitrary feasible point.
2. Solve the trace-bounded SDP with $\operatorname{tr} \boldsymbol{X} \le \alpha_i$ to obtain $X_\star(\alpha_i)$.
3. If the primal objective $\langle \boldsymbol{C}, \boldsymbol{X}_\star(\alpha_i)\rangle = \langle \boldsymbol{C}, \boldsymbol{X}_\star(\alpha_{i-1})\rangle$, then terminate and return $\boldsymbol{X}_\star(\alpha_i)$.
4. Otherwise, set $\alpha_{i+1} = 2\alpha_i$ and return to Step 2.

This procedure terminates in $\tilde{O}(1)$ iterations if a bounded solution exists. It is easy to show by contradiction that there exists no finite $\alpha > \alpha_i$ such that $\langle \boldsymbol{C}, \boldsymbol{X}_\star(\alpha)\rangle < \langle \boldsymbol{C}, \boldsymbol{X}_\star(\alpha_i)\rangle$. Unfortunately, it is nontrivial to extend this claim to the approximate solutions because they are infeasible.

SM4.3. Convex inclusions.

To handle the inclusion in K that appears in (SM4.1), we need an efficient algorithm to perform the Euclidean projection onto K. That is,

$$\operatorname{proj}_{\mathsf{K}}(\boldsymbol{w}) := \arg\min\{\|\boldsymbol{w} - \boldsymbol{u}\| : \boldsymbol{u} \in \mathsf{K}\} \quad \text{for } \boldsymbol{w} \in \mathbb{R}^d.$$

Here are some important examples:

1. **Inequality constraints.** $\mathsf{K} := \{\boldsymbol{u} \in \mathbb{R}^d : \boldsymbol{u} \le \boldsymbol{b}\}$. In this case, the projection takes the form $\operatorname{proj}_{\mathsf{K}}(\boldsymbol{w}) = (\boldsymbol{w} - \boldsymbol{b})_-$, where $(\cdot)_-$ reports the negative part of a vector.
2. **Norm constraints.** $\mathsf{K} := \{\boldsymbol{u} \in \mathbb{R}^d : \|\boldsymbol{u} - \boldsymbol{b}\| \le \delta\}$, where $\|\cdot\|$ is a norm. The projector can be computed easily for many norms, including the ℓ_p norm for $p \in \{1, 2, \infty\}$.

SM4.4. The CGAL iteration for the general template.

To extend the description of the CGAL iteration in subsection SM1.3 for the general template (SM4.1), we consider the following augmented Lagrangian formulation with the slack variable $\boldsymbol{w} \in \mathsf{K}$ instead of (SM1.6):

$$L_t(\boldsymbol{X}; \boldsymbol{y}) := \langle \boldsymbol{C}, \boldsymbol{X} \rangle + \min_{\boldsymbol{w} \in \mathsf{K}} \left\{ \langle \boldsymbol{y}, \mathcal{A}\boldsymbol{X} - \boldsymbol{w} \rangle + \frac{1}{2}\beta_t \|\mathcal{A}\boldsymbol{X} - \boldsymbol{w}\|^2 \right\}.$$

Accordingly, the partial derivative (SM1.8) becomes

$$\boldsymbol{D}_t := \partial_{\boldsymbol{X}} L_t(\boldsymbol{X}_t; \boldsymbol{y}_t) = \boldsymbol{C} + \mathcal{A}^*\big(\boldsymbol{y}_t + \beta_t(\mathcal{A}\boldsymbol{X}_t - \boldsymbol{w}_t)\big) \quad \text{where} \quad \boldsymbol{w}_t := \mathrm{proj}_{\mathsf{K}}(\mathcal{A}\boldsymbol{X}_t + \beta_t^{-1}\boldsymbol{y}_t).$$

We replace the linear minimization subroutine (SM1.9) with (SM4.2). We can still use an inexact variant of (SM4.2) with additive error. We also revise the dual update scheme by modifying (SM1.11) as

$$\boldsymbol{y}_{t+1} = \boldsymbol{y}_t + \gamma_t(\mathcal{A}\boldsymbol{X}_{t+1} - \overline{\boldsymbol{w}}_t) \quad \text{where} \quad \overline{\boldsymbol{w}}_t := \mathrm{proj}_{\mathsf{K}}(\mathcal{A}\boldsymbol{X}_{t+1} + \beta_{t+1}^{-1}\boldsymbol{y}_t).$$

Finally, we replace the dual step size parameter selection rule (SM1.12) with

$$\gamma_t \|\mathcal{A}\boldsymbol{X}_{t+1} - \overline{\boldsymbol{w}}_t\|^2 \le \beta_t \eta_t^2 \alpha^2 \|\mathcal{A}\|^2. \tag{SM4.3}$$

The bounded travel condition (SM1.13) remains the same.

To obtain the extension of SketchyCGAL to the general template, we simply pursue the same program outlined in section 6 to augment CGAL with sketching.

SM5. Details of MaxCut experiments.

This section presents further details about the termination criteria that we used in the MaxCut experiments presented in section 7.

SM5.1. Comparison of SDP solvers.

We begin with a discussion of how we compared the performance of different SDP solvers for the MaxCut problem (1.3).

SM5.1.1. Convention for sign of the dual.

We used the sign convention from the Lagrangian formulation (SM1.2) when describing the quality guarantees of each solver, and also in the definition of DIMACS errors and the dual problem in subsection SM5.3. This convention can be different in other works, so these definitions might have the sign of the dual variable inverted in some references.

SM5.1.2. Details for SketchyCGAL.

We implement the stopping criteria based on the discussion in subsections SM1.7 and SM3.1. We stop the algorithm when both

$$\frac{p_t + \langle \boldsymbol{y}_t, \boldsymbol{b} \rangle + \frac{1}{2}\beta_t \langle \boldsymbol{z}_t - \boldsymbol{b}, \boldsymbol{z}_t + \boldsymbol{b} \rangle - \lambda_{\min}(\boldsymbol{D}_t)}{1 + |p_t|} \le \varepsilon \quad \text{and} \quad \frac{\|\boldsymbol{z}_t - \boldsymbol{b}\|}{1 + \|\boldsymbol{b}\|} \le \varepsilon. \tag{SM5.1}$$

In theory, this bound requires computing $\lambda_{\min}(\boldsymbol{D}_t)$ to high accuracy. In practice, we did not observe a significant difference when using a high accuracy approximation or the inexact computation from step (SM1.9). We fixed the code to use the latter to avoid additional cost.

If SketchyCGAL terminates by (SM5.1), then the implicit variable $\boldsymbol{X}$ provably satisfies

$$\frac{\langle \boldsymbol{C}, \boldsymbol{X} \rangle - \langle \boldsymbol{C}, \boldsymbol{X}_\star \rangle}{1 + |\langle \boldsymbol{C}, \boldsymbol{X} \rangle|} \le \varepsilon \quad \text{and} \quad \frac{\|\mathcal{A}\boldsymbol{X} - \boldsymbol{b}\|}{1 + \|\boldsymbol{b}\|} \le \varepsilon.$$

This is the guarantee that we seek.

Algorithm SM4.1 SketchyCGAL for the general template (SM4.1)

Input: Problem data for (SM4.1) implemented via the primitives (2.4), sketch size R, number T of iterations

Output: Rank-R approximate solution to (SM4.1) in factored form $\widehat{\boldsymbol{X}}_T = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^*$ where $\boldsymbol{U} \in \mathbb{R}^{n \times R}$ has orthonormal columns and $\boldsymbol{\Lambda} \in \mathbb{R}^{R \times R}$ is nonnegative diagonal, and the Schatten 1-norm approximation errors $\mathrm{err}([\![\widehat{\boldsymbol{X}}]\!]_r)$ for $1 \le r \le R$, as defined in (SM2.4)

Recommendation: To achieve (2.3), set R as large as possible, and set $T \approx \varepsilon^{-1}$

function SketchyCGAL(R; T)
  Scale problem data (subsection 7.1.1) ▷ **[opt]** Recommended!
  $\beta_0 \leftarrow 1$ and $K \leftarrow +\infty$ ▷ Default parameters
  NystromSketch.Init(n, R)
  $\boldsymbol{z} \leftarrow \mathbf{0}_d$ and $\boldsymbol{y} \leftarrow \mathbf{0}_d$
  for $t \leftarrow 1, 2, 3, \dots, T$ **do**
    $\beta \leftarrow \beta_0\sqrt{t+1}$ and $\eta \leftarrow 2/(t+1)$
    $\boldsymbol{w} \leftarrow \mathrm{proj}_{\mathsf{K}}(\boldsymbol{z} + \beta^{-1}\boldsymbol{y})$
    $\boldsymbol{D} \leftarrow \boldsymbol{C} + \mathcal{A}^*(\boldsymbol{y} + \beta(\boldsymbol{z} - \boldsymbol{w}))$ ▷ Represent via primitives (2.4)❶❷
    $\boldsymbol{H}$ is a (low-rank) matrix that solves (SM4.2)
    $\boldsymbol{z} \leftarrow (1-\eta)\,\boldsymbol{z} + \eta\,\mathcal{A}(\boldsymbol{H})$ ▷ Use primitive (2.4)❸
    $\beta_+ \leftarrow \beta_0\sqrt{t+2}$
    $\boldsymbol{w} \leftarrow \mathrm{proj}_{\mathsf{K}}(\boldsymbol{z} + \beta_+^{-1}\boldsymbol{y})$
    $\boldsymbol{y} \leftarrow \boldsymbol{y} + \gamma(\boldsymbol{z} - \boldsymbol{w})$ ▷ Step size γ satisfies (SM1.13) and (SM4.3)
    NystromSketch.RankOneUpdate($\sqrt{\alpha}\boldsymbol{v}$, η)
  $[\boldsymbol{U}, \boldsymbol{\Lambda}, \mathbf{err}] \leftarrow$ NystromSketch.Reconstruct()

SM5.1.3. Details for SDPT3.

We used SDPT3 version 4.0 [SM21] in the experiments. This software is designed for solving conic optimization problems by using a primal-dual interior-point algorithm. The algorithm iterates three variables: $\boldsymbol{X}$, $\boldsymbol{y}$, and $\boldsymbol{Z}$. For the MaxCut problem, $\boldsymbol{X}$ and $\boldsymbol{Z}$ are $n \times n$ symmetric positive semidefinite matrices, and $\boldsymbol{y} \in \mathbb{R}^n$.

We can control the desired accuracy by changing the parameter `OPTIONS.gaptol`. When we set this parameter to ε, the outputs ensure

$$\frac{\langle \boldsymbol{X},\, \boldsymbol{Z}\rangle}{1 + |\langle \boldsymbol{C},\, \boldsymbol{X}\rangle| + |\boldsymbol{b}^\top \boldsymbol{y}|} \le \varepsilon, \qquad \frac{\|\mathcal{A}\boldsymbol{X} - \boldsymbol{b}\|}{1 + \|\boldsymbol{b}\|} \le \varepsilon, \qquad \text{and} \qquad \frac{\|\mathcal{A}^*\boldsymbol{y} - \boldsymbol{Z} + \boldsymbol{C}\|_F}{1 + \|\boldsymbol{C}\|_F} \le \varepsilon.$$

These are the required bounds.

SM5.1.4. Details for SeDuMi.

We used SeDuMi 1.3 [SM18] in the experiments. This software implements a self-dual embedding technique [SM26]. The algorithm has two outputs. For the MaxCut problem, they are the $n \times n$ symmetric positive semidefinite primal solution $\boldsymbol{X}$ and n dimensional dual solution $\boldsymbol{y}$. We can control the desired accuracy by changing the parameter `pars.eps`.

For some instances, when we set `pars.eps` very small, even though the algorithm achieves the desired accuracy, we observed that SeDuMi 1.3 overwrites the variables as zeros after solving the problem before returning them, based on a quality control procedure. To prevent this issue, we commented out the section between the lines 638 and 642 in `sedumi.m`.

SM5.1.5. Details for SDPNAL+. We used SDPNAL version 1.0 [SM19] which implements an augmented Lagrangian based method for solving semidefinite programs. When applied to the MAXCUT SDP, the algorithm outputs three variables $\boldsymbol{X}$, $\boldsymbol{y}$ and $\boldsymbol{Z}$ similar to SDPT3. We can control the desired accuracy by changing the parameter `OPTIONS.tol`. When we set this parameter to ε, the outputs ensure

$$\frac{\|\mathcal{A}\boldsymbol{X}-\boldsymbol{b}\|}{1+\|\boldsymbol{b}\|}\le\varepsilon,\qquad \frac{\|\mathcal{A}^*\boldsymbol{y}-\boldsymbol{Z}+\boldsymbol{C}\|_F}{1+\|\boldsymbol{C}\|_F}\le\varepsilon,\qquad\text{and}\qquad \frac{\|\boldsymbol{X}-\mathrm{proj}_{\mathrm{psd}}(\boldsymbol{X}-\boldsymbol{Z})\|_F}{1+\|\boldsymbol{X}\|_F+\|\boldsymbol{Z}\|_F}\le 5\varepsilon,$$

where $\mathrm{proj}_{\mathrm{psd}}:\mathbb{R}^{n\times n}\to\mathbb{S}_n$ is the projection operator onto positive semidefinite cone.

SM5.1.6. Details for Mosek. We used the interior point optimizer for conic optimization from the Mosek Optimization Suite Release 8.1.0.64 [SM15]. This optimizer implements of the so-called homogeneous and self-dual algorithm [SM2]. When applied to the MaxCut SDP, the algorithm returns three variables $\boldsymbol{X}$, $\boldsymbol{y}$, and $\boldsymbol{Z}$. We can control the target accuracy by changing three parameters: `MSK_DPAR_INTPNT_CO_TOL_PFEAS`, `MSK_DPAR_INTPNT_CO_TOL_DFEAS`, and `MSK_DPAR_INTPNT_CO_TOL_RELGAP`. For simplicity, and as suggested in the manual, we relax these parameters together and set each one to ε. This ensures

$$\frac{\|\mathcal{A}\boldsymbol{X}-\boldsymbol{b}\|_\infty}{1+\|\boldsymbol{b}\|_\infty}\le\varepsilon,\qquad \frac{\|\mathcal{A}^*\boldsymbol{y}-\boldsymbol{Z}+\boldsymbol{C}\|_\infty}{1+\|\boldsymbol{C}\|_\infty}\le\varepsilon,$$

$$\frac{\langle\boldsymbol{X},\,\boldsymbol{Z}\rangle}{\max\{1,\min\{|\langle\boldsymbol{C},\,\boldsymbol{X}\rangle|,|\boldsymbol{b}^\top\boldsymbol{y}|\}\}}\le\varepsilon,\qquad\text{and}\qquad \frac{|\langle\boldsymbol{C},\,\boldsymbol{X}\rangle+\boldsymbol{b}^\top\boldsymbol{y}|}{\max\{1,\min\{|\langle\boldsymbol{C},\,\boldsymbol{X}\rangle|,|\boldsymbol{b}^\top\boldsymbol{y}|\}\}}\le\varepsilon.$$

These are the quality guarantees that we need.

SM5.2. Convergence of SketchyCGAL for the MaxCut SDP. This section gives further information about our evaluation of the convergence behavior of the SketchyCGAL algorithm.

Table SM2 presents numerical data from the MaxCut SDP experiment with GSET benchmark. We compare the methods in terms of the

$$\texttt{suboptimality}=\frac{\langle\boldsymbol{C},\,\boldsymbol{X}\rangle-\langle\boldsymbol{C},\,\boldsymbol{X}_\star\rangle}{1+|\langle\boldsymbol{C},\,\boldsymbol{X}_\star\rangle|}\quad\text{and}\quad\texttt{infeasibility}=\frac{\|\mathcal{A}\boldsymbol{X}-\boldsymbol{b}\|}{1+\|\boldsymbol{b}\|},$$

as well as the storage cost and computation time. We use a high-accuracy solution from SDPT3 with default parameters to approximate the optimal point. The storage cost is approximated by monitoring the virtual memory size of the process, hence it includes the memory that is swapped out and it can go beyond 16 GB.

We also compare the weight of cut evaluated after rounding (see subsection 7.2.1 for details of the rounding procedure), relative to the weight obtained by rounding $\boldsymbol{X}_\star$:

$$\texttt{relative cut weight}=\frac{\texttt{cut weight from }\boldsymbol{X}\texttt{ - cut weight from }\boldsymbol{X}_\star}{\texttt{cut weight from }\boldsymbol{X}_\star}.$$

Positive values indicate better cuts with a higher weight. The average of this quantity over all datasets in the benchmark indicates a discrepancy of less than 1.5%, a small price for huge scalability benefits.

SM5.3. Primal–dual convergence. This section presents empirical evidence for the convergence of the primal variables, the dual variables, and the surrogate duality gap generated by SketchyCGAL. We use the MaxCut SDP (1.3) for these tests.

SM5.3.1. DIMACS errors. We evaluate the DIMACS errors [SM1] (with Euclidean scaling) to measure the suboptimality and infeasibilities for primal and dual problems. These measures are commonly used for benchmarking SDP solvers [SM14]. For a given approximate solution triplet $(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{Z})$, we compute

$$\texttt{err}_1 = \frac{\|\mathcal{A}\boldsymbol{X} - \boldsymbol{b}\|}{1 + \|\boldsymbol{b}\|}, \qquad \texttt{err}_2 = \frac{\max\{-\lambda_{\min}(\boldsymbol{X}), 0\}}{1 + \|\boldsymbol{b}\|}, \qquad \texttt{err}_3 = \frac{\|\mathcal{A}^*\boldsymbol{y} - \boldsymbol{Z} + \boldsymbol{C}\|_F}{1 + \|\boldsymbol{C}\|_F},$$

$$\texttt{err}_4 = \frac{\max\{-\lambda_{\min}(\boldsymbol{Z}), 0\}}{1 + \|\boldsymbol{C}\|_F}, \qquad \text{and} \qquad \texttt{err}_5 = \frac{\langle \boldsymbol{C}, \boldsymbol{X}\rangle + \boldsymbol{b}^\top \boldsymbol{y}}{1 + |\langle \boldsymbol{C}, \boldsymbol{X}\rangle| + |\boldsymbol{b}^\top \boldsymbol{y}|}.$$

We also compute the 6th error defined in [SM14] to measure the violation in complementary slackness

$$\texttt{err}_6 = \frac{\langle \boldsymbol{X}, \boldsymbol{Z}\rangle}{1 + |\langle \boldsymbol{C}, \boldsymbol{X}\rangle| + |\boldsymbol{b}^\top \boldsymbol{y}|}.$$

SketchyCGAL and Sedumi do not maintain the third variable $\boldsymbol{Z}$ explicitly. We set $\boldsymbol{Z} = \boldsymbol{C} + \mathcal{A}^*\boldsymbol{y}$ as suggested in the guideline of the 7th DIMACS Implementation Challenge [SM1, SM14].

We use a testbed similar to the MaxCut experiments summarized in subsection 7.2. For ease of comparison, we replace $\operatorname{tr} \boldsymbol{X} = n$ constraint with $\operatorname{tr} \boldsymbol{X} \le 1.05 \cdot n$. We explain the reason of this modification in the next subsection. We also set a stronger target accuracy by choosing $\varepsilon = 10^{-3}$ for each solver. Table SM7 reports the numerical outcomes.

SM5.3.2. The dual of the model problem. The dual function for the model problem (SM1.1) with the trace constraint is

$$\text{(SM5.2)} \qquad \begin{aligned} \phi(\boldsymbol{y}) := \min_{\boldsymbol{X} \in \alpha \boldsymbol{\Delta}_n} L(\boldsymbol{X}, \boldsymbol{y}) &= \min_{\boldsymbol{X} \in \alpha \boldsymbol{\Delta}_n} \langle \boldsymbol{C} + \mathcal{A}^*\boldsymbol{y}, \boldsymbol{X}\rangle - \langle \boldsymbol{y}, \boldsymbol{b}\rangle \\ &= \alpha \cdot \lambda_{\min}(\boldsymbol{C} + \mathcal{A}^*\boldsymbol{y}) - \langle \boldsymbol{y}, \boldsymbol{b}\rangle. \end{aligned}$$

Therefore, the dual problem is an unconstrained nonsmooth concave maximization:

$$\text{(SM5.3)} \qquad \underset{\boldsymbol{y} \in \mathbb{R}^n}{\text{maximize}} \quad \phi(\boldsymbol{y}) := \alpha \cdot \lambda_{\min}(\boldsymbol{C} + \mathcal{A}^*\boldsymbol{y}) - \langle \boldsymbol{y}, \boldsymbol{b}\rangle.$$

Under strong duality, $\phi(\boldsymbol{y}_\star) = p_\star = \langle \boldsymbol{C}, \boldsymbol{X}_\star\rangle$.

Note that (SM5.3) is slightly different from the dual of the standard-form SDP. In the standard-form dual, the maximization in (SM5.2) takes place over the psd cone without any further constraints, and the term $\min_{\{\boldsymbol{X} \text{ is psd}\}} \langle \boldsymbol{C} + \mathcal{A}^*\boldsymbol{y}, \boldsymbol{X}\rangle$ becomes an indicator function that constraints $\boldsymbol{C} + \mathcal{A}^*\boldsymbol{y}$ to the psd cone. We can formulate the standard-form dual problem as

$$\text{(SM5.4)} \qquad \text{maximize} \quad -\langle \boldsymbol{y}, \boldsymbol{b}\rangle \quad \text{subject to} \quad \boldsymbol{C} + \mathcal{A}^*\boldsymbol{y} \text{ is psd}, \quad \boldsymbol{y} \in \mathbb{R}^n.$$

In particular, the problem (SM5.4) is constrained, while (SM5.3) is unconstrained.

DIMACS measures are defined to evaluate convergence to a solution of (SM5.4), hence they do not fit well for evaluating (SM5.3) in general. In our experiments, we empirically observed that $\phi(\boldsymbol{y})$ converges to $p_\star$ for the dual sequence of SketchyCGAL, but $\boldsymbol{y}$ does not converge to a solution of (SM5.4). However, the following lemma from our concurrent work with Ding [SM8] shows that the solution sets of the duals of the standard SDP and the trace-bounded SDP (subsection SM4.2) are the same under some technical conditions.

Lemma SM5.1 (Lemma 6.1 in [SM8]). *Suppose that the standard-form SDP has a solution $\boldsymbol{X}_\star$, satisfies strong duality, and $\boldsymbol{y}_\star$ is the unique solution of the standard-form dual problem (SM5.4). Consider the trace bounded SDP with $\mathsf{X} = \{\boldsymbol{X} \in \mathbb{S}_n : \operatorname{tr} \boldsymbol{X} \leq \alpha \text{ and } \boldsymbol{X} \text{ is psd}\}$ for some $\alpha > \operatorname{tr} \boldsymbol{X}_\star$. The dual problem is*

$$\underset{\boldsymbol{y} \in \mathbb{R}^n}{\text{maximize}} \quad \alpha \cdot \min\{0, \lambda_{\min}(\boldsymbol{C} + \mathcal{A}^* \boldsymbol{y})\} - \langle \boldsymbol{y}, \boldsymbol{b} \rangle. \tag{SM5.5}$$

Suppose $\|\boldsymbol{b}\| \neq 0$. Then, $\boldsymbol{y}_\star$ is the unique solution of (SM5.5).

We refer to [SM8] for the proof.

SM5.4. Failure of the Burer–Monteiro heuristic. This section presents empirical evidence that Burer–Monteiro (BM) factorization methods cannot support storage costs better than $\Omega(n\sqrt{d})$. Our approach is based on the paper of Waldspurger and Waters [SM25], which proves that the BM heuristic (8.1) can produce incorrect results unless $R = \Omega(n\sqrt{d})$.

Waldspurger provided us code that generates a random symmetric $\boldsymbol{C} \in \mathbb{S}_n(\mathbb{R})$. The MaxCut SDP (1.3) with objective $\boldsymbol{C}$ has a unique solution, and the solution has rank 1. If the factorization rank R satisfies $R(R+3) \leq 2n$, then the BM formulation (8.1) has second-order critical points that are not optimal points of the original SDP (1.3). As a consequence, the Burer–Monteiro approach is reliable only if the storage budget is $\Theta(n^{3/2})$. In contrast, for the same problem instances, our analysis (Theorem Theorem 6.3) shows that SketchyCGAL succeeds with factorization rank $R = 2$ and storage budget $\Theta(n)$.

We will demonstrate numerically that Waldspurger and Waters [SM25] have identified a serious obstruction to using the Burer–Monteiro approach. Moreover, we will see that SketchyCGAL resolves the issue. See the code supplement for scripts to reproduce these experiments.

We use the Manopt software [SM3] to solve the Burer–Monteiro formulation of the MaxCut SDP. For $n = 100$, we drew 10 random matrices $\boldsymbol{C}_1, \dots, \boldsymbol{C}_{10}$ using Waldspurger's code. For each instance, we sweep the factorization rank $R = 2, 3, 4, \dots, 13$. (For $R \geq 13$, we anticipate that each second-order critical point of the Burer–Monteiro problem is a solution to the original SDP, owing to the analysis in [SM4].) In each experiment, we ran Manopt with 100 random initializations, and we counted the number of times the algorithm failed. We declared failure if Manopt converged to a second-order stationary point whose objective value is 10^{-3} larger than the true optimal value. See Table SM1 for the statistics.

In contrast, SketchyCGAL can solve all of these instances, even when the sketch size $R = 2$. For these problems, we use the default parameter choices for SketchyCGAL, but we do not pre-scale the data or perform tuning. Figure SM7.1 compares the convergence trajectory of SketchyCGAL and Manopt for one problem instance. The difference is evident.

SM6. Details of phase retrieval experiments. This section presents further details about the phase retrieval experiments presented in section 7.

SM6.1. Synthetic phase retrieval data. This section provides additional details on the construction of synthetic datasets for the abstract phase retrieval SDP.

For each $n \in \{10^2, 10^3, \dots, 10^6\}$, we generate 20 independent datasets as follows. First, draw $\boldsymbol{\chi}_\natural \in \mathbb{C}^n$ from the complex standard normal distribution. We acquire $d = 12n$ phaseless measurements (7.2) using the coded diffraction pattern model [SM7].

To do so, we randomly draw 12 independent modulating waveforms $\boldsymbol{\psi}_j$ for $j = 1, 2, \dots, 12$. Each entry of $\boldsymbol{\psi}_j$ is drawn as the product of two independent random variables, one chosen uniformly from $\{1, i, -1, -i\}$, and the other from $\{\sqrt{2}/2, \sqrt{3}\}$ with probabilities 0.8 and 0.2 respectively. Then, we modulate $\boldsymbol{\chi}_\natural$ with these waveforms and take its Fourier transform. Each $\boldsymbol{a}_i$ corresponds to computing a single entry of this Fourier transform:

$$\boldsymbol{a}_{(j-1)n+\ell} = \boldsymbol{W}_n(\ell, :)\ \mathrm{diag}^*(\boldsymbol{\psi}_j), \qquad 1 \le j \le 12 \quad \text{and} \quad 1 \le \ell \le n,$$

where $\boldsymbol{W}_n(\ell, :)$ is the ℓth row of the $n \times n$ discrete Fourier transform matrix. We use the fast Fourier transform to implement the measurement operator.

SM6.2. Fourier ptychography. We study a more realistic measurement setup, Fourier ptychography (FP), for the phase retrieval problem. In this setup, $\boldsymbol{\chi}_\natural \in \mathbb{C}^n$ corresponds to an unknown high resolution image (vectorized) from a microscopic sample in the Fourier domain. One cannot directly acquire a high resolution image from this sample because of the physical limitations of optical systems. Any measurement is subject to a filter caused by the lens aperture. We can represent this filter by a sparse matrix $\boldsymbol{\Phi} \in \mathbb{C}^{m \times n}$ with $m \le n$, each row of which has only one non-zero coefficient. Because of this filter, we can acquire only low-resolution images, through m-dimensional Fourier transform.

FP enlightens the sample from L different angles using a LED grid. This lets us to to obtain L different aperture matrices Φ_j, $j = 1, 2, \dots, L$. Then, we acquire phaseless measurements from the sample using the following transmission matrices:

$$\boldsymbol{a}_{(j-1)m+\ell} = \boldsymbol{W}_m^*(\ell, :)\, \Phi_j, \qquad 1 \le j \le L \quad \text{and} \quad 1 \le \ell \le m.$$

Here, $\boldsymbol{W}_m^*(\ell, :)$ is the ℓth row of the conjugate transpose of discrete Fourier transform matrix.

The aim in FP is to reconstruct complex valued $\boldsymbol{\chi}_\natural$ from these phaseless measurements. Once we construct $\boldsymbol{\chi}_\natural$, we can generate a high resolution image by taking its inverse Fourier transform.

SM7. Details for QAP. This section gives further information about the QAP experiments summarized in subsection 7.5. Tables SM4 and SM6 display the performance of SketchyCGAL, by presenting the upper bound after rounding (subsection 7.5.3), the objective value $\langle \boldsymbol{B} \otimes \boldsymbol{A},\ \boldsymbol{X} \rangle$, feasibility gap, total number of iterations, memory usage, and computation time. Feasibility gap is defined as $\mathrm{dist}_{\mathsf{K}}(\mathcal{A}\boldsymbol{X})$ where K is the Cartesian product between a singleton (for equality constraints) and the nonnegative orthant (for inequality constraints). It is evaluated with respect to the original (not rescaled) problem data and without normalization.

Note that the objective value of SketchyCGAL is not a lower bound for the optimum since the iterates are infeasible. Due to sublinear convergence, SketchyCGAL might be impractical when high accuracy is required, for example within a branch and bound procedure.

Tables SM3 and SM5 compare the relative gap (7.7) obtained by SketchyCGAL with the values for the CSDP method [SM6] with clique size $k = \{2, 3, 4\}$ and the PATH method [SM29] reported in [SM6, Tab. 6]. A graphical view of these results appears in Figure 7.4.

REFERENCES

[1] *7th DIMACS implementation challenge: Semidefinite and related optimization problems*, 2000, http://archive.dimacs.rutgers.edu/Challenges/Seventh/ (accessed July 2020).
[2] E. D. Andersen, C. Roos, and T. Terlaky, *On implementing a primal-dual interior-point method for conic quadratic optimization*, Mathematical Programming, (2003), pp. 249–277.
[3] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, *Manopt, a Matlab toolbox for optimization on manifolds*, J. Mach. Learn. Res., 15 (2014), pp. 1455–1459.
[4] N. Boumal, V. Voroninski, and A. Bandeira, *The non-convex Burer–Monteiro approach works on smooth semidefinite programs*, in Adv. Neural Information Processing Systems 29, 2016, pp. 2757–2765.
[5] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2011), pp. 1–122.
[6] J. F. S. Bravo Ferreira, Y. Khoo, and A. Singer, *Semidefinite programming approach for the quadratic assignment problem with a sparse graph*, Comput. Optim. Appl., 69 (2018), pp. 677–712.
[7] E. J. Candès, X. Li, and M. Soltanolkotabi, *Phase retrieval from coded diffraction patterns*, Appl. Comp. Harmonic Anal., 39 (2015), pp. 277 – 299.
[8] L. Ding, A. Yurtsever, V. Cevher, J. A. Tropp, and M. Udell, *An optimal-storage approach to semidefinite programming using approximate complementarity*, 2019, https://arxiv.org/abs/1902.03373.
[9] A. Gittens, *Topics in Randomized Numerical Linear Algebra*, PhD thesis, California Institute of Technology, Pasadena, CA, USA, 2013.
[10] G. H. Golub and C. F. Van Loan, *Matrix computations*, Johns Hopkins Univ. Press, Baltimore, MD, fourth ed., 2013.
[11] N. Halko, P. G. Martinsson, and J. A. Tropp, *Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions*, SIAM Rev., 53 (2011), pp. 217–288.
[12] M. Jaggi, *Revisiting Frank–Wolfe: Projection-free sparse convex optimization*, in Proc. 30th Int. Conf. Machine Learning, vol. 28(1), 2013, pp. 427–435.
[13] H. Li, G. C. Linderman, A. Szlam, K. P. Stanton, Y. Kluger, and M. Tygert, *Algorithm 971: an implementation of a randomized algorithm for principal component analysis*, ACM Trans. Math. Software, 43 (2017), pp. Art. 28, 14.
[14] H. D. Mittelmann, *An independent benchmarking of SDP and SOCP solvers*, Mathematical Programming, 95 (2003), pp. 407–430.
[15] MOSEK ApS, *MOSEK version 8.1.0.64 (oct. 8, 2018)*, https://www.mosek.com/downloads/8.1.0.64/ (accessed July 2020).
[16] M. L. Overton and R. S. Womersley, *On the sum of the largest eigenvalues of a symmetric matrix*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 41–45, https://doi.org/10.1137/0613006.
[17] N. Srebro, *Learning with matrix factorizations*, PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 2004.
[18] J. Sturm, *SeDuMi 1.3*, http://sedumi.ie.lehigh.edu/?page_id=58 (accessed July 2020).
[19] D. Sun, K.-C. Toh, Y. Yuan, and X.-Y. Zhao, *SDPNAL+: A Matlab software for semidefinite programming with bound constraints (version 1.0)*, Optimization Methods and Software, 35 (2020), pp. 87–115.
[20] Y. Sun, Y. Guo, J. A. Tropp, and M. Udell, *Tensor random projection for low memory dimension reduction*, in NeurIPS Workshop on Relational Representation Learning, 2018.
[21] K.-C. Toh, M. J. Todd, and R. H. Tütüncü, *SDPT3 version 4.0 - a MATLAB software for semidefinite-quadratic-linear programming*, https://blog.nus.edu.sg/mattohkc/softwares/sdpt3/ (ac-

cessed July 2020).
[22] J. A. Tropp, A. Yurtsever, M. Udell, and V. Cevher, *Fixed-rank approximation of a positive-semidefinite matrix from streaming data*, in Adv. Neural Information Processing Systems 30, 2017, pp. 1225–1234.
[23] J. A. Tropp, A. Yurtsever, M. Udell, and V. Cevher, *Practical sketching algorithms for low-rank matrix approximation*, SIAM J. Matrix Anal. Appl., 38 (2017), pp. 1454–1485.
[24] J. A. Tropp, A. Yurtsever, M. Udell, and V. Cevher, *Streaming low-rank matrix approximation with an application to scientific simulation*, SIAM J. Sci. Comput., 41 (2019), pp. A2430–A2463.
[25] I. Waldspurger and A. Waters, *Rank optimality for the Burer–Monteiro factorization*, 2018, https://arxiv.org/abs/1812.03046.
[26] Y. Ye, M. J. Todd, and S. Mizuno, *An $(o(\sqrt{n}l)$-iteration homogeneuos and self-dual linear programming algorithm*, Mathematics of Operations Research, (1994), pp. 53–67.
[27] A. Yurtsever, O. Fercoq, and V. Cevher, *A conditional-gradient-based augmented Lagrangian framework*, in Proc. 36th Int. Conf. Machine Learning, vol. 97, 09–15 Jun 2019, pp. 7272–7281.
[28] A. Yurtsever, O. Fercoq, F. Locatello, and V. Cevher, *A conditional gradient framework for composite convex minimization with applications to semidefinite programming*, in Proc. 35th Intl. Conf. Machine Learning, vol. 80, 2018, pp. 5727–5736.
[29] M. Zaslavskiy, F. Bach, and J. Vert, *A path following algorithm for the graph matching problem*, IEEE Trans. Pattern Anal. Mach. Intel., 31 (2009), pp. 2227–2242.

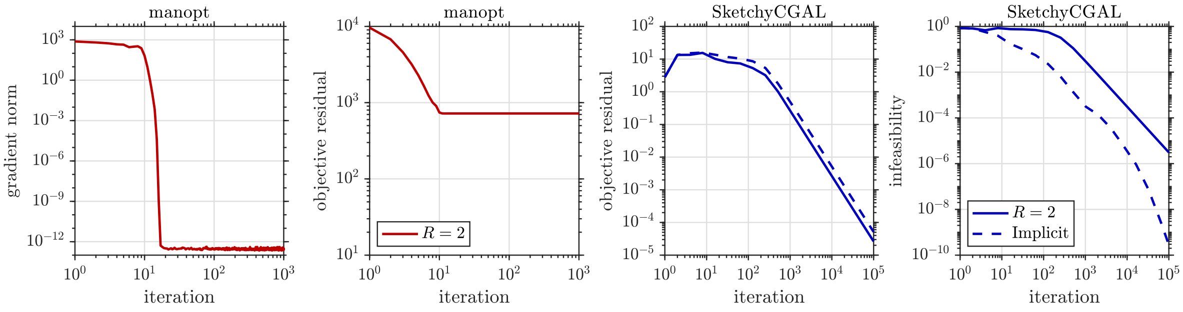
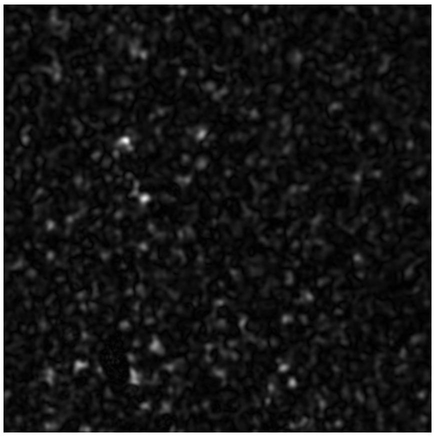


Figure SM7.1. MaxCut SDP: Failure of the Burer–Monteiro heuristic. *We apply Manopt and SketchyCGAL for solving the MaxCut SDP with the dataset $\boldsymbol{C}_1$. The subplots show the gradient norm and suboptimality for Manopt [left] and the suboptimality and infeasibility for SketchyCGAL [right]. Manopt with $R = 2$ converges to a spurious solution, whereas SketchyCGAL successfully computes a rank-1 approximation of the global optimum. The dashed line describes the convergence of the SketchyCGAL implicit iterates. For details, see subsection SM5.4.*

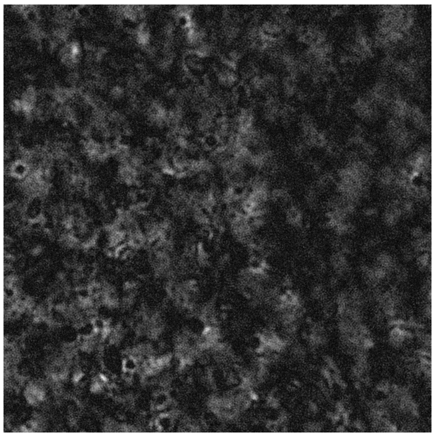
Table SM1

We run Manopt for solving hard instances of the MaxCut SDP. We consider 10 datasets $\boldsymbol{C}_1, \dots, \boldsymbol{C}_{10}$. For each dataset, we run Manopt with 100 random initializations and report the number of failures. We declare failure when Manopt converges to second-order critical point that is not a global optimum. For details, see subsection SM5.4.

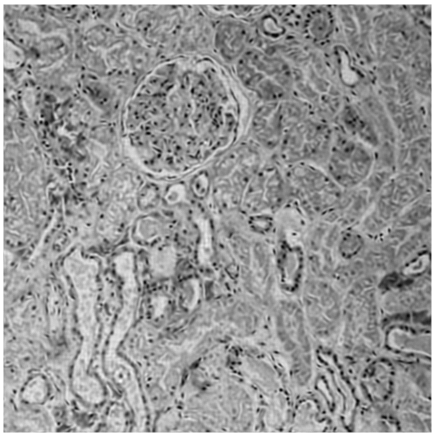
Dataset / R	$R = 2$	3	4	5	6	7	8	9	10	11	12	13
$\boldsymbol{C}_1$	82	69	63	53	35	32	24	12	11	1	4	0
$\boldsymbol{C}_2$	77	56	56	36	19	17	12	2	0	0	0	0
$\boldsymbol{C}_3$	89	65	54	47	44	46	23	11	5	0	3	0
$\boldsymbol{C}_4$	84	69	50	40	27	23	18	17	1	0	9	0
$\boldsymbol{C}_5$	85	68	52	51	43	30	31	20	14	3	4	0
$\boldsymbol{C}_6$	81	68	53	41	23	22	10	10	2	0	1	0
$\boldsymbol{C}_7$	83	76	60	39	19	19	19	3	0	0	1	0
$\boldsymbol{C}_8$	81	73	44	34	41	25	8	12	5	4	10	0
$\boldsymbol{C}_9$	84	64	46	35	25	17	1	10	0	2	4	0
$\boldsymbol{C}_{10}$	83	71	54	50	31	25	24	16	13	0	8	0



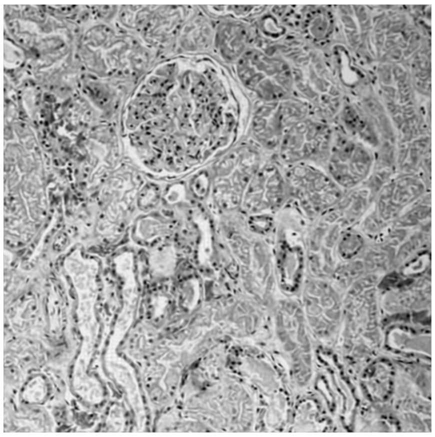
(a) $t = 10$ (6 min)



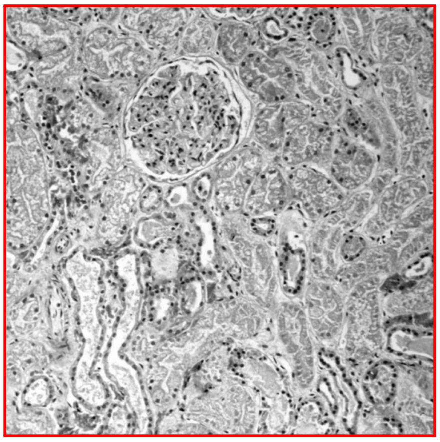
(b) $t = 100$ (78 min)



(c) $t = 1000$ (1'265 min)



(d) $t = 5'000$ (6'248 min)



(e) original

Figure SM7.2. Phase retrieval SDP: Imaging. *Reconstruction of an $n = 640^2$ pixel image from Fourier ptychography data. We solve an $n \times n$ phase retrieval SDP via SketchyCGAL with rank parameter $R = 5$ and show the images obtained at iterations $t = 10, 10^2, 10^3, 5 \cdot 10^3$. The last subfigure is the original. See subsection 7.4.3.*

Table SM2: Numerical outcomes from the MaxCut SDP experiment with Gset Benchmark.

	SketchyCGAL	MoSeK	SDPT3	SDPNAL+	Sedumi
Data Name Size (n)	relative cut weight suboptimality infeasibility storage (MB) time (hh:mm:ss)				
G1 800	-1.0861e-02 1.3811e-02 9.5917e-02 8 00:00:01	1.1387e-03 -3.1293e-02 4.6070e-02 226 00:00:04	-1.0861e-02 1.4254e-01 6.0389e-09 180 00:00:04	-8.4961e-03 8.1450e-02 1.9172e-02 245 00:00:02	2.7152e-03 6.4228e-03 1.4574e-02 288 00:00:11
G2 800	-2.2936e-03 1.4340e-02 9.6277e-02 8 00:00:01	6.1750e-04 -3.3205e-02 4.8907e-02 226 00:00:04	-1.3232e-03 5.5569e-02 1.8859e-10 180 00:00:05	-2.6464e-03 8.2332e-02 1.8822e-02 245 00:00:02	2.0289e-03 6.1814e-03 1.4745e-02 288 00:00:11
G3 800	-9.5622e-03 1.5032e-02 9.6560e-02 8 00:00:01	-4.3863e-04 -3.6397e-02 5.3604e-02 226 00:00:05	-6.4041e-03 5.8755e-02 2.4555e-10 180 00:00:04	-7.8077e-03 8.1880e-02 1.9094e-02 245 00:00:02	-2.0177e-03 9.2430e-03 1.9138e-02 288 00:00:11
G4 800	-5.1115e-03 1.6379e-02 9.6255e-02 8 00:00:01	1.4101e-03 -4.0176e-02 5.9377e-02 226 00:00:05	-1.4101e-03 6.2225e-02 2.9489e-10 180 00:00:08	-3.8777e-03 8.3752e-02 1.9058e-02 180 00:00:02	1.2338e-03 1.5740e-02 2.6591e-02 288 00:00:22
G5 800	-7.3504e-03 1.6157e-02 9.6365e-02 8 00:00:01	-6.1253e-04 -3.9680e-02 5.8582e-02 226 00:00:05	-7.5254e-03 6.2757e-02 3.7160e-10 180 00:00:05	-7.3504e-03 8.3001e-02 1.9091e-02 180 00:00:04	-3.4127e-03 1.0683e-02 2.1958e-02 288 00:00:11
G6 800	-2.8718e-02 3.1505e-02 9.3371e-02 8 00:00:01	-1.7436e-02 3.8635e-01 3.0300e-01 226 00:00:03	-8.2051e-03 1.5088e-01 4.8995e-11 180 00:00:05	-3.0769e-02 1.0956e-01 4.9102e-02 178 00:00:05	3.5897e-03 8.4988e-02 2.7171e-02 352 00:00:12
G7 800	2.8802e-03 2.7527e-02 9.5193e-02 8 00:00:01	-2.3618e-02 4.2087e-01 2.9762e-01 226 00:00:03	-1.7857e-02 1.5116e-01 5.0094e-11 180 00:00:04	-1.2097e-02 1.1725e-01 5.0117e-02 178 00:00:03	-1.4977e-02 6.7074e-02 2.3929e-02 288 00:00:12
G8 800	-7.9320e-03 2.6436e-02 9.4596e-02 8 00:00:01	-1.8130e-02 4.1639e-01 2.9784e-01 226 00:00:02	5.6657e-04 1.5401e-01 4.9526e-11 180 00:00:04	-3.9093e-02 1.2247e-01 4.9573e-02 178 00:00:03	-7.3654e-03 8.2034e-02 2.4421e-02 288 00:00:13

Continued on the next page

Table SM2: MaxCut Benchmark with GSet (cont.).

	SketchyCGAL	MoSeK	SDPT3	SDPNAL+	Sedumi
G9 800	-2.0533e-02	-1.6648e-02	1.4983e-02	-5.0499e-02	1.6648e-02
	2.8791e-02	4.2360e-01	7.6875e-02	1.1126e-01	8.6664e-02
	9.6469e-02	3.0737e-01	3.4834e-12	4.8213e-02	2.7624e-02
	8	226	180	178	286
	00:00:01	00:00:03	00:00:05	00:00:03	00:00:11
G10 800	-3.3409e-02	-5.8324e-02	-3.3975e-03	-7.5878e-02	-1.1891e-02
	3.0233e-02	4.1742e-01	1.4609e-01	1.1787e-01	6.9872e-02
	9.6215e-02	2.9681e-01	5.0049e-11	4.8529e-02	2.3584e-02
	8	226	180	178	286
	00:00:01	00:00:03	00:00:06	00:00:03	00:00:11
G11 800	-1.5444e-02	-2.3166e-02	1.5444e-02	-7.3359e-02	1.5444e-02
	2.6265e-02	6.1699e-01	7.2209e-02	4.4688e-02	7.9842e-02
	9.5640e-02	0	5.3829e-11	3.7060e-02	2.5442e-02
	8	226	185	183	288
	00:00:01	00:00:02	00:00:04	00:00:03	00:00:10
G12 800	-3.9062e-03	-1.9531e-02	3.5156e-02	-7.8125e-03	1.9531e-02
	2.4141e-02	6.5004e-01	8.2364e-02	5.9988e-02	9.3180e-02
	9.5875e-02	0	5.0553e-11	4.8779e-02	2.8291e-02
	8	226	185	183	288
	00:00:01	00:00:01	00:00:04	00:00:03	00:00:10
G13 800	3.7736e-03	-1.8868e-02	2.6415e-02	-4.1509e-02	3.0189e-02
	2.3104e-02	6.3325e-01	8.5197e-02	5.7293e-02	8.2228e-02
	9.5982e-02	0	5.5987e-11	4.9942e-02	2.5854e-02
	8	226	185	178	288
	00:00:01	00:00:01	00:00:03	00:00:03	00:00:10
G14 800	-7.0779e-03	-1.3819e-02	-2.5278e-02	-6.2016e-02	-3.8760e-02
	4.7170e-02	-2.1354e-02	8.2634e-02	8.7609e-02	1.3110e-01
	7.6060e-02	5.6505e-02	4.3439e-11	1.8669e-02	2.6233e-02
	8	226	181	190	288
	00:00:02	00:00:05	00:00:03	00:00:02	00:00:09
G15 800	-1.4137e-02	-1.8849e-02	-2.2215e-02	-7.1693e-02	-4.7122e-02
	4.0370e-02	-2.5681e-02	8.7010e-02	8.8979e-02	1.3943e-01
	6.8233e-02	6.7477e-02	4.1721e-11	1.8486e-02	2.4687e-02
	8	226	186	180	288
	00:00:02	00:00:04	00:00:04	00:00:02	00:00:09
G16 800	-7.0994e-03	-9.4659e-03	-1.9608e-02	-5.0372e-02	-3.9892e-02
	3.8136e-02	-2.2376e-02	8.5358e-02	8.6831e-02	1.3028e-01
	7.0830e-02	5.9677e-02	4.8317e-11	1.8906e-02	2.8522e-02
	8	226	181	180	288
	00:00:02	00:00:04	00:00:04	00:00:02	00:00:09
G17 800	-1.2860e-02	-2.3689e-03	-1.8274e-02	-5.9560e-02	-3.6548e-02
	4.7458e-02	-2.2381e-02	7.8986e-02	8.7610e-02	1.3277e-01
	6.8409e-02	5.9613e-02	3.6482e-11	1.8555e-02	2.4114e-02
	8	226	186	180	286
	00:00:02	00:00:04	00:00:04	00:00:02	00:00:09

Continued on the next page

Table SM2: MaxCut Benchmark with GSet (cont.).

	SketchyCGAL	MoSeK	SDPT3	SDPNAL+	Sedumi
G18 800	-3.6954e-02 4.2249e-02 7.1076e-02 8 00:00:01	-1.0414e-01 3.7409e-01 1.9476e-01 226 00:00:03	-1.0078e-02 1.1502e-01 2.4288e-12 186 00:00:05	-1.1646e-01 1.0608e-01 6.1771e-02 178 00:00:02	-6.7189e-03 1.0740e-01 2.3304e-02 289 00:00:17
G19 800	-2.5031e-02 6.0931e-02 8.0855e-02 8 00:00:01	-9.3867e-02 4.4872e-01 2.0949e-01 226 00:00:03	1.2516e-03 1.4566e-01 2.4037e-12 186 00:00:05	-1.2390e-01 1.1875e-01 6.6121e-02 178 00:00:03	0 1.4665e-01 2.3218e-02 355 00:00:17
G20 800	-3.8232e-02 5.5135e-02 8.5073e-02 8 00:00:01	-1.0036e-01 4.0152e-01 1.9516e-01 226 00:00:03	-1.7921e-02 1.3282e-01 2.2402e-12 186 00:00:05	-1.2067e-01 1.1077e-01 6.4826e-02 178 00:00:02	-8.3632e-03 1.2889e-01 2.9589e-02 289 00:00:17
G21 800	-6.1728e-03 5.7118e-02 8.4940e-02 8 00:00:01	-5.6790e-02 4.1582e-01 1.9936e-01 226 00:00:03	4.4444e-02 1.3161e-01 2.4262e-12 186 00:00:05	-1.0617e-01 1.1080e-01 6.6784e-02 178 00:00:02	4.1975e-02 1.3462e-01 2.7448e-02 289 00:00:17
G22 2000	-1.2115e-02 1.5681e-02 9.6934e-02 9 00:00:01	1.0032e-03 -8.1412e-03 4.1194e-02 984 00:01:14	-9.1056e-03 9.1982e-02 2.6086e-10 670 00:00:32	-1.0417e-02 6.8851e-02 2.8216e-02 492 00:00:24	-6.9450e-04 1.1905e-02 1.2584e-02 1300 00:03:18
G23 2000	-8.0850e-03 1.3379e-02 9.6457e-02 9 00:00:01	-5.3900e-03 -7.8396e-03 3.9683e-02 984 00:01:15	-1.3706e-02 9.2914e-02 3.3246e-10 704 00:00:33	-1.3398e-02 6.8894e-02 2.8258e-02 492 00:00:24	-6.0060e-03 3.0420e-02 3.3485e-02 1300 00:03:43
G24 2000	-1.1172e-02 1.3818e-02 9.6751e-02 9 00:00:01	-1.2328e-03 -7.3304e-03 3.7142e-02 984 00:01:21	-1.1557e-02 9.2312e-02 4.1922e-10 703 00:00:33	-1.2405e-02 6.8781e-02 2.8311e-02 492 00:00:34	-4.0835e-03 2.8684e-02 2.9243e-02 1300 00:03:43
G25 2000	-3.4973e-03 1.3900e-02 9.6988e-02 9 00:00:01	2.4870e-03 -7.7270e-03 3.9683e-02 984 00:01:21	-3.4196e-03 1.0041e-01 7.3474e-10 670 00:00:31	-4.6631e-03 6.9165e-02 2.8242e-02 492 00:00:34	1.6321e-03 1.2113e-02 1.1920e-02 1235 00:04:26
G26 2000	-4.5073e-03 1.6124e-02 9.6396e-02 9 00:00:01	1.0880e-03 -7.7421e-03 3.9605e-02 984 00:01:01	-4.2742e-03 1.0441e-01 9.9323e-10 670 00:00:31	-6.6832e-03 6.8630e-02 2.8347e-02 492 00:00:28	1.5542e-03 3.1191e-02 2.9950e-02 1300 00:07:28

Continued on the next page

Table SM2: MaxCut Benchmark with GSet (cont.).

	SketchyCGAL	MoSeK	SDPT3	SDPNAL+	Sedumi
G27 2000	-1.4187e-02	-3.9792e-02	-6.9204e-04	-3.5294e-02	-1.0035e-02
	3.6489e-02	3.4331e-01	1.3416e-01	1.2784e-01	5.9977e-02
	9.7233e-02	2.2480e-01	5.3716e-12	4.7806e-02	1.7503e-02
	9	984	769	558	1300
	00:00:01	00:00:48	00:00:42	00:00:29	00:03:50
G28 2000	-9.4970e-03	-2.2511e-02	2.8139e-03	-2.5677e-02	-4.5726e-03
	3.7008e-02	3.7142e-01	1.2557e-01	1.2923e-01	1.2780e-01
	9.7469e-02	2.3601e-01	5.4840e-12	4.7381e-02	3.6643e-02
	9	984	735	561	1300
	00:00:01	00:00:56	00:00:48	00:00:54	00:03:17
G29 2000	-2.0168e-02	-3.1261e-02	-1.6807e-03	-3.1933e-02	-9.0756e-03
	3.4436e-02	3.6747e-01	1.2288e-01	1.2817e-01	1.1352e-01
	9.7664e-02	2.4240e-01	5.1238e-12	4.7327e-02	3.7082e-02
	9	984	704	558	1300
	00:00:01	00:00:56	00:00:48	00:00:22	00:03:17
G30 2000	-2.6605e-02	-3.3588e-02	9.9767e-04	-3.7912e-02	-3.9907e-03
	3.1683e-02	3.5612e-01	1.3726e-01	1.2752e-01	5.3960e-02
	9.7498e-02	2.3418e-01	5.7873e-12	4.7109e-02	1.5929e-02
	9	984	671	558	1300
	00:00:01	00:00:56	00:00:48	00:00:24	00:05:14
G31 2000	-1.2170e-02	-4.3811e-02	-1.7385e-03	-3.0250e-02	7.9972e-03
	3.2478e-02	4.1374e-01	1.2072e-01	1.2853e-01	1.3652e-01
	9.7291e-02	2.5613e-01	4.4749e-12	4.7889e-02	4.0526e-02
	9	984	704	558	1204
	00:00:01	00:00:47	00:00:48	00:00:24	00:04:19
G32 2000	-4.7022e-03	-2.5078e-02	2.5078e-02	-9.2476e-02	3.2915e-02
	2.3071e-02	6.3823e-01	7.5100e-02	4.6005e-02	8.7684e-02
	9.7150e-02	0	3.3059e-11	4.2067e-02	2.7141e-02
	9	984	558	624	1235
	00:00:01	00:00:30	00:00:21	00:00:36	00:02:56
G33 2000	2.5974e-02	-2.7597e-02	4.5455e-02	-3.0844e-02	5.5195e-02
	2.4605e-02	6.4997e-01	7.3805e-02	4.5305e-02	8.8476e-02
	9.6924e-02	0	3.2314e-11	4.2924e-02	2.6904e-02
	9	984	595	561	1300
	00:00:01	00:02:12	00:00:17	00:01:00	00:02:57
G34 2000	-1.6000e-03	-1.7600e-02	3.3600e-02	-2.4000e-02	4.6400e-02
	3.1015e-02	6.6216e-01	7.5042e-02	5.0275e-02	9.5898e-02
	9.7385e-02	0	2.8472e-11	4.1950e-02	2.9024e-02
	9	984	595	624	1300
	00:00:01	00:00:32	00:00:18	00:00:27	00:02:53
G35 2000	-5.9347e-03	-2.4278e-03	-2.6976e-02	-7.4049e-02	-5.6110e-02
	3.6621e-02	-2.8394e-02	1.0806e-01	9.4208e-02	1.5892e-01
	5.3605e-02	7.6066e-02	5.0420e-11	1.7448e-02	2.3757e-02
	9	984	682	592	1300
	00:00:13	00:01:26	00:00:29	00:00:19	00:02:44

Continued on the next page

Table SM2: MaxCut Benchmark with GSet (cont.).

	SketchyCGAL	MoSeK	SDPT3	SDPNAL+	Sedumi
G36 2000	-1.1345e-02 3.6011e-02 4.9253e-02 9 00:00:11	-3.2415e-03 -1.4296e-02 3.8655e-02 984 00:01:43	-1.9044e-02 8.0139e-02 1.0676e-11 591 00:00:29	-6.9692e-02 8.9393e-02 8.3391e-03 592 00:00:24	-6.3614e-02 1.6699e-01 1.4397e-02 1300 00:03:40
G37 2000	-1.6125e-02 4.7562e-02 5.7891e-02 9 00:00:16	-2.4187e-03 -1.2850e-02 3.5075e-02 984 00:01:29	-3.3593e-02 1.1191e-01 4.7745e-11 703 00:00:27	-7.8877e-02 9.4806e-02 1.7306e-02 592 00:00:19	-6.7052e-02 1.6740e-01 1.9096e-02 1269 00:03:40
G38 2000	-5.4039e-03 4.3352e-02 5.3739e-02 9 00:00:14	-1.7563e-03 -1.2564e-02 3.4143e-02 984 00:01:42	-2.8641e-02 1.1024e-01 4.8473e-11 639 00:00:27	-7.3629e-02 9.4460e-02 1.7351e-02 594 00:00:17	-5.6336e-02 1.6723e-01 2.2529e-02 1300 00:03:39
G39 2000	-6.1445e-02 6.6946e-02 9.6734e-02 9 00:00:01	-1.2899e-01 4.4741e-01 2.3043e-01 984 00:00:55	-1.6417e-02 1.2226e-01 6.2560e-13 661 00:00:50	-1.4071e-01 1.2104e-01 6.8073e-02 627 00:00:31	-3.2364e-02 2.9167e-01 3.4601e-02 1498 00:06:45
G40 2000	-5.1092e-02 6.0931e-02 8.1869e-02 9 00:00:01	-1.9368e-01 4.8275e-01 2.4701e-01 984 00:00:55	-2.8333e-02 1.4461e-01 6.1378e-13 692 00:00:40	-1.8346e-01 1.0003e-01 6.7156e-02 624 00:00:33	-3.1119e-02 1.9028e-01 2.8995e-02 1564 00:07:50
G41 2000	-3.2640e-02 5.3490e-02 9.7598e-02 9 00:00:01	-1.3907e-01 4.6215e-01 2.3489e-01 984 00:00:43	-2.3179e-02 1.2991e-01 5.6737e-13 742 00:00:40	-1.6509e-01 1.1735e-01 6.9128e-02 624 00:00:34	-2.2233e-02 1.9720e-01 3.7320e-02 1367 00:06:28
G42 2000	-3.9796e-02 5.3904e-02 9.7783e-02 9 00:00:01	-1.2772e-01 4.4115e-01 2.3361e-01 984 00:01:06	8.7922e-03 1.2281e-01 6.2319e-13 607 00:00:41	-1.4715e-01 1.1432e-01 6.9205e-02 624 00:00:34	-7.8667e-03 2.7490e-01 3.3145e-02 1498 00:06:44
G43 1000	-1.2117e-02 1.5552e-02 9.6716e-02 8 00:00:01	-1.6871e-03 -7.0064e-03 3.3450e-02 299 00:00:12	-1.7025e-02 8.6541e-02 6.1918e-11 245 00:00:07	-2.1012e-02 6.5582e-02 2.7835e-02 246 00:00:06	-4.7546e-03 1.7807e-02 2.1330e-02 417 00:00:20
G44 1000	1.7025e-03 1.3966e-02 9.6710e-02 8 00:00:01	-1.2382e-03 -6.4682e-03 3.1378e-02 299 00:00:13	-8.0483e-03 8.5723e-02 6.1629e-11 311 00:00:07	-1.5942e-02 6.4779e-02 2.8081e-02 246 00:00:05	-2.0121e-03 1.7696e-02 1.9844e-02 417 00:00:20

Continued on the next page

Table SM2: MaxCut Benchmark with GSet (cont.).

	SketchyCGAL	MoSeK	SDPT3	SDPNAL+	Sedumi
G45 1000	-8.0695e-03 1.5134e-02 9.6247e-02 8 00:00:01	1.0863e-03 -6.2559e-03 3.0095e-02 299 00:00:13	-8.6903e-03 8.7887e-02 6.1268e-11 311 00:00:06	-1.1173e-02 6.4124e-02 2.8040e-02 180 00:00:04	7.7592e-04 2.0988e-02 2.6194e-02 417 00:00:29
G46 1000	-9.3516e-04 1.5542e-02 9.6680e-02 8 00:00:01	1.4027e-03 -6.0780e-03 2.9203e-02 299 00:00:11	-2.6496e-03 8.5848e-02 6.0013e-11 311 00:00:06	-4.2082e-03 6.4099e-02 2.8459e-02 188 00:00:04	2.1820e-03 1.6751e-02 2.0716e-02 417 00:00:29
G47 1000	-3.4098e-03 1.5588e-02 9.6696e-02 8 00:00:01	2.0149e-03 -6.5081e-03 3.1407e-02 299 00:00:08	-8.6795e-03 8.7395e-02 6.0498e-11 311 00:00:08	-6.6646e-03 6.5387e-02 2.8090e-02 261 00:00:11	-9.2994e-04 1.7486e-02 2.2294e-02 417 00:00:27
G48 3000	-1.3667e-02 1.0092e-02 9.4775e-02 10 00:00:01	0 4.7282e-02 2.2182e-15 2087 00:03:26	0 1.2901e-01 3.5505e-11 1024 00:00:44	0 4.1637e-02 6.8384e-02 879 00:01:27	0 1.5852e-02 1.5935e-02 2339 00:12:26
G49 3000	-1.4000e-02 9.7028e-03 9.7780e-02 10 00:00:01	0 4.7281e-02 1.8776e-15 2087 00:03:14	0 1.2905e-01 3.5624e-11 958 00:00:43	0 4.1763e-02 6.8213e-02 879 00:01:24	0 4.7003e-02 4.7125e-02 2273 00:09:31
G50 3000	-1.3946e-02 8.0999e-03 9.7044e-02 10 00:00:01	0 4.2954e-02 1.5870e-15 2087 00:03:51	0 1.2508e-01 3.5298e-11 958 00:00:46	0 3.9432e-02 6.8811e-02 879 00:01:25	0 3.2078e-02 3.6382e-02 2414 00:09:45
G51 1000	-1.6319e-02 4.4961e-02 6.1555e-02 8 00:00:04	-9.3633e-03 -2.4666e-02 6.4965e-02 299 00:00:08	-2.6752e-02 8.9620e-02 4.7564e-11 311 00:00:07	-7.7582e-02 8.9975e-02 1.7660e-02 196 00:00:03	-5.1899e-02 1.3953e-01 2.5262e-02 417 00:00:22
G52 1000	-8.8901e-03 3.5672e-02 6.1392e-02 8 00:00:03	-3.7716e-03 -3.0821e-02 7.7417e-02 299 00:00:08	-2.1013e-02 8.8583e-02 3.9779e-11 245 00:00:10	-5.4149e-02 8.7957e-02 1.8136e-02 196 00:00:03	-3.8793e-02 1.4374e-01 2.4383e-02 417 00:00:15
G53 1000	-4.8767e-03 4.9023e-02 7.5129e-02 8 00:00:04	-2.9802e-03 -2.9221e-02 7.3965e-02 299 00:00:11	-5.1477e-03 8.7032e-02 4.9598e-11 245 00:00:06	-6.0688e-02 8.8963e-02 1.7903e-02 196 00:00:03	-3.7930e-02 1.3194e-01 2.6304e-02 417 00:00:16

Continued on the next page

Table SM2: MaxCut Benchmark with GSet (cont.).

	SketchyCGAL	MoSeK	SDPT3	SDPNAL+	Sedumi
G54 1000	-1.0243e-02	0	-1.7251e-02	-6.9542e-02	-3.6927e-02
	4.7776e-02	-2.3638e-02	8.2065e-02	8.7252e-02	1.2880e-01
	7.5128e-02	6.1915e-02	4.3089e-11	1.8259e-02	2.5147e-02
	8	299	245	196	417
	00:00:06	00:00:10	00:00:06	00:00:03	00:00:16
G55 5000	-1.5392e-02	-7.2911e-03	-9.2152e-03	-1.9949e-02	-1.7215e-03
	3.2784e-02	3.8322e-02	9.8695e-02	7.7692e-02	5.5184e-02
	9.8225e-02	5.9028e-02	2.7354e-11	3.7645e-02	2.9373e-02
	12	6541	2544	2444	6080
	00:00:01	00:18:23	00:04:29	00:05:04	00:59:13
G56 5000	-1.3081e-02	-7.5146e-02	-8.3496e-04	-4.4531e-02	1.1133e-03
	4.2961e-02	6.0115e-01	1.5219e-01	1.4778e-01	1.3967e-01
	9.8497e-02	2.4789e-01	6.0728e-12	3.5298e-02	3.1123e-02
	12	6476	2736	2432	6080
	00:00:02	00:09:19	00:06:02	00:05:35	00:59:12
G57 5000	-1.5873e-02	-3.6190e-02	2.0317e-02	-2.6984e-01	2.5397e-02
	3.4139e-02	6.5144e-01	9.2966e-02	4.9293e-02	1.0633e-01
	9.8470e-02	0	3.3474e-11	4.2377e-02	3.2602e-02
	12	6476	2652	2349	6080
	00:00:01	00:09:20	00:04:52	00:11:27	00:44:54
G58 5000	-9.9085e-03	-6.0851e-03	-3.7641e-02	-7.7868e-02	-9.0361e-02
	4.0681e-02	-1.5515e-02	1.0405e-01	7.7087e-02	2.1977e-01
	4.6279e-02	4.2906e-02	1.3232e-11	7.2850e-03	2.5553e-02
	12	6541	2749	2354	6080
	00:02:31	00:34:53	00:09:44	00:08:31	00:28:48
G59 5000	-4.4515e-02	-2.1826e-01	-1.1458e-02	-1.9591e-01	-1.5026e-02
	4.7775e-02	5.1949e-01	1.3418e-01	1.0923e-01	1.9235e-01
	9.8420e-02	3.0113e-01	3.7377e-14	6.7124e-02	3.1820e-02
	79	6541	2827	2349	6862
	00:00:02	00:14:27	00:14:07	00:06:30	02:04:37
G60 7000	-1.5879e-02	-5.9546e-03	-6.1016e-03	-2.0290e-02	1.3968e-03
	3.4044e-02	4.0277e-02	1.0393e-01	7.8355e-02	6.5477e-02
	9.8489e-02	5.9479e-02	3.0375e-11	3.7698e-02	3.6393e-02
	17	12682	5017	4756	11705
	00:00:02	00:38:37	00:11:47	00:15:47	02:41:03
G61 7000	-1.3506e-02	-8.9138e-02	-1.2541e-02	-3.8009e-02	6.5599e-03
	3.7879e-02	5.7665e-01	1.9524e-01	1.4471e-01	1.5498e-01
	9.8261e-02	2.4786e-01	2.5011e-11	3.5055e-02	3.7988e-02
	17	12682	5329	4734	11705
	00:00:03	00:19:08	00:12:54	00:17:20	03:49:14
G62 7000	-2.0464e-02	-6.0482e-02	1.0459e-02	-2.9013e-01	2.5466e-02
	3.5993e-02	6.5473e-01	9.1622e-02	3.0213e-02	9.8260e-02
	9.8628e-02	0	3.8092e-11	4.5815e-02	3.0209e-02
	17	12682	5155	4571	11705
	00:00:02	00:15:42	00:26:48	00:21:52	02:36:22

Continued on the next page

Table SM2: MaxCut Benchmark with GSet (cont.).

	SketchyCGAL	MoSeK	SDPT3	SDPNAL+	Sedumi
G63 7000	-6.6179e-03	-4.5402e-03	-3.6553e-02	-7.7299e-02	-9.0535e-02
	3.3992e-02	-1.5066e-02	1.0744e-01	7.8815e-02	2.2130e-01
	3.5469e-02	4.2551e-02	1.3603e-11	7.5063e-03	2.5199e-02
	17	12682	5236	4678	11705
	00:04:40	00:59:53	00:14:56	00:14:15	01:42:50
G64 7000	-6.0457e-02	-2.4390e-01	-1.8344e-02	-2.2181e-01	-1.8861e-02
	5.5067e-02	5.2060e-01	1.1120e-01	1.0273e-01	1.5301e-01
	9.8677e-02	3.1977e-01	1.4076e-14	6.8433e-02	3.3139e-02
	17	12682	5458	4714	13237
	00:00:04	00:25:34	00:25:26	00:24:07	06:54:36
G65 8000	-4.4426e-03	-4.1599e-02	2.4637e-02	-2.7746e-01	3.6349e-02
	3.1221e-02	6.5182e-01	9.4792e-02	2.9541e-02	1.0447e-01
	9.8691e-02	0	3.8620e-11	4.6184e-02	3.2221e-02
	17	15496	6863	6164	15221
	00:00:02	00:27:01	00:11:39	00:40:02	03:27:59
G66 9000	-1.8544e-02	-4.0588e-02	1.5395e-02	-3.1176e-01	–
	3.3867e-02	6.4642e-01	8.6972e-02	2.8440e-02	–
	9.8820e-02	0	3.7667e-11	4.6642e-02	–
	17	21970	8340	7755	–
	00:00:04	00:42:07	00:16:41	00:55:07	–
G67 10000	-1.8971e-02	–	3.2797e-02	-3.2990e-01	–
	3.0792e-02	–	4.8610e-02	3.3673e-02	–
	9.8756e-02	–	8.0771e-12	4.4400e-02	–
	17	–	10269	9540	–
	00:00:03	–	00:25:29	01:13:18	–

Table SM3

We solve SDP relaxations of QAP instances from QAPLIB using SketchyCGAL. We compute the relative gap and compare it with the values for the CSDP method [SM6] with clique size $k = \{2, 3, 4\}$ and the PATH method [SM29] reported in [SM6, Tab. 4]. Smaller is better. See subsection 7.5.

Dataset	Optimum	SketchyCGAL	CSDP2	CSDP3	CSDP4	PATH
chr12a	9552	0	34.5	6	0	42.7
chr12b	9742	0	38.9	25.4	11.9	38.1
chr12c	11156	0	5.8	2.3	2.3	18.6
chr15a	9896	0.4	2.1	2.1	2.1	52
chr15b	7990	0	26.3	34.5	29.2	158.6
chr15c	9504	0	0	0	0	63.3
chr18a	11098	1.5	69.8	0.2	0.2	76.3
chr18b	1534	8.7	8.9	22.9	29.5	99.3
chr20a	2192	6.6	122.5	76.1	43.8	95.4
chr20b	2298	0	62.9	9.3	9.3	82.2
chr20c	14142	24.7	173	100.1	111.5	88.9
chr22a	6156	0	17.2	7.6	3	38.3
chr22b	6194	7.1	7.3	2.3	1	40.4
chr25a	3796	21.8	107	49.2	25.2	69.9
esc16a	68	2.9	8.8	11.8	11.8	11.8
esc16b	292	0	0	0.7	0	2.7
esc16c	160	5	5	7.5	8.7	6.3
esc16d	16	12.5	12.5	50	25	75
esc16e	28	0	14.3	7.1	14.3	21.4
esc16g	26	7.7	7.7	0	15.4	15.4
esc16h	996	0	1.6	0	1.6	16.9
esc16i	14	0	0	0	0	57.1
esc16j	8	0	0	0	0	75
esc32a	130	52.3	115.4	124.6	113.8	93.8
esc32b	168	57.1	109.5	114.3	111.9	88.1
esc32c	642	2.2	12.8	15.9	13.7	7.8
esc32d	200	14	38	36	39	21
esc32e	2	0	0	0	0	600
esc32g	6	0	0	0	0	366.7
esc32h	438	5	24.7	26.9	22.8	18.3
esc64a	116	6.9	60.3	53.4	60.3	106.9
esc128	64	28.1	250	206.3	175	221.9
ste36a	9526	15.6	70.2	74.7	74.2	76.3
ste36b	15852	15.7	188.8	204.3	211.9	158.6
ste36c	8239110	9	66	62.8	63.7	83.2

Table SM4

We run SketchyCGAL for (the first of) 10^6 iterations or 72 hours of runtime, for solving SDP relaxations of QAP instances from QAPLIB. We report the upper bound after rounding (subsection 7.5.3), the objective value, the feasibility gap, the number of iterations, the memory usage (in MB), and the cpu time ('hh:mm:ss'). See subsection 7.5.

Dataset	Optimum	Upper bnd.	Objective	Feas. gap	Iteration	Memory	Time
chr12a	9552	9552	9 576,13	5.56e−5	1000000	201	05 : 21 : 44
chr12b	9742	9742	9 759,34	2.53e−5	1000000	201	05 : 22 : 35
chr12c	11156	11156	11 021,45	2.04e−4	1000000	201	05 : 24 : 09
chr15a	9896	9936	9 519,53	1.92e−4	1000000	214	09 : 20 : 38
chr15b	7990	7990	7 469,92	1.73e−4	1000000	214	09 : 58 : 04
chr15c	9504	9504	9 517,71	3.55e−5	1000000	214	08 : 40 : 31
chr18a	11098	11262	10 700,65	2.38e−4	1000000	214	14 : 04 : 57
chr18b	1534	1668	1 534,29	7.64e−5	1000000	214	13 : 52 : 37
chr20a	2192	2336	2 183,68	1.13e−4	1000000	201	18 : 59 : 05
chr20b	2298	2298	2 288,94	1.08e−4	1000000	201	18 : 47 : 35
chr20c	14142	17630	13 316,48	1.91e−4	1000000	201	20 : 20 : 08
chr22a	6156	6156	6 152,80	9.99e−5	1000000	193	27 : 47 : 48
chr22b	6194	6634	6 198,39	1.05e−4	1000000	193	27 : 35 : 40
chr25a	3796	4624	3 692,50	2.08e−4	1000000	197	45 : 30 : 14
esc16a	68	70	60,36	5.96e−5	1000000	201	15 : 36 : 34
esc16b	292	292	287,64	9.65e−4	1000000	198	24 : 47 : 46
esc16c	160	168	145,01	4.71e−5	1000000	197	17 : 28 : 31
esc16d	16	18	13,00	6.16e−5	1000000	201	11 : 37 : 42
esc16e	28	28	25,41	6.11e−5	1000000	201	11 : 32 : 58
esc16g	26	28	22,46	7.03e−5	1000000	201	12 : 00 : 45
esc16h	996	996	975,41	2.10e−4	1000000	198	26 : 53 : 14
esc16i	14	14	11,37	7.43e−5	1000000	201	10 : 42 : 41
esc16j	8	8	7,11	9.00e−5	1000000	201	10 : 05 : 17
esc32a	130	198	99,61	3.62e−4	569710	262	72 : 00 : 00
esc32b	168	264	118,36	2.86e−4	449002	197	72 : 00 : 00
esc32c	642	656	610,84	7.01e−4	402329	197	72 : 00 : 00
esc32d	200	228	187,33	5.92e−4	494696	273	72 : 00 : 00
esc32e	2	2	1,90	4.16e−4	1000000	197	60 : 25 : 27
esc32g	6	6	5,83	4.37e−4	1000000	197	66 : 37 : 07
esc32h	438	460	417,79	3.48e−4	363879	197	72 : 00 : 00
esc64a	116	124	97,85	7.28e−3	126620	288	72 : 00 : 00
esc128	64	82	53,95	4.06e−2	23882	587	72 : 00 : 00
ste36a	9526	11012	8 995,09	1.03e−3	256265	263	72 : 00 : 00
ste36b	15852	18336	15 344,24	1.03e−3	257781	263	72 : 00 : 00
ste36c	8239110	8983468	7 984 574,51	1.05e−3	257657	263	72 : 00 : 00

Table SM5

We solve SDP relaxations of QAP instances from TSPLIB using SketchyCGAL. We compute the relative gap and compare it with the values for the CSDP method [SM6] with clique size $k = \{2, 3, 4\}$ and the PATH method [SM29] reported in [SM6, Tab. 6]. Smaller is better.

Dataset	Optimum	SketchyCGAL	CSDP2	CSDP3	CSDP4	PATH
att48	10628	73.8	213	236.5	233.6	329.8
bayg29	1610	38	114.3	115.8	114.3	210.1
bays29	2020	44	107.6	118.3	115.4	164.8
berlin52	7542	52	175	127.2	127.2	280.6
bier127	118282	73.2	216.4	193.8	193.8	234.2
brazil58	25395	86.5	248	200.8	200.8	337
burma14	3323	10.8	24.6	28.4	32.3	95.5
ch130	6110	102.8	352.4	380.6	380.6	621.3
ch150	6528	121.1	346.9	318.2	318.2	689.3
dantzig42	699	73.5	193.1	174	174	82
eil101	629	70.6	227.3	235.3	235.3	437.7
eil51	426	57.3	203.6	205.4	205.5	244.4
eil76	538	65.4	282.9	183	183	328.2
fri26	937	32.7	91.6	39.4	39.4	41.6
gr120	6942	123.6	445.2	261.6	261.6	617.6
gr137	69853	147.5	264.6	220.3	220.3	38.9
gr17	2085	15.3	46.8	32.4	44.9	86.9
gr21	2707	18.8	94.5	69.7	66.3	185.7
gr24	1272	30	89.2	86.2	73.9	129.4
gr48	5046	59.7	210.2	187.4	187.4	270.4
gr96	55209	119.3	228.9	201.7	201.7	46
hk48	11461	43.9	222.4	207.7	207.7	281.6
kroA100	21282	125	469.6	469	469	720.2
kroA150	26524	168.4	411	467.4	467.4	945.8
kroB100	22141	164.3	411.9	313.6	313.6	624.2
kroB150	26130	166.8	417.3	353.7	353.7	844.7
kroC100	20749	163.4	507.4	445.1	445.1	763
kroD100	21294	118.5	504.2	349.8	349.8	654.4
kroE100	22068	137.1	489.5	346.3	346.3	684.2
lin105	14379	167.2	303.1	234.8	234.8	248.4
pr107	44303	195.3	181.5	207.9	207.9	41.6
pr124	59030	192.4	293.8	180.2	180.2	67.6
pr136	96772	148.6	325.5	164.7	164.7	196.6
pr144	58537	270.4	255	283.7	283.7	59.8
pr76	108159	81	192.2	194	194	39.4
rat99	1211	95.3	236.4	161.5	161.5	444.1
rd100	7910	89	438.4	375.3	375.3	506.5
st70	675	70.8	300.9	320	317.9	387.9
swiss42	1273	35.5	163.2	190.4	190.8	194
ulysses16	6859	11.5	23.6	20.2	23.2	82.7
ulysses22	7013	26.4	64.5	57	59.7	126.3

Table SM6

We run SketchyCGAL for (the first of) 10^6 iterations or 72 hours of runtime, for solving SDP relaxations of QAP instances from TSPLIB. We report the upper bound after rounding (subsection 7.5.3), the objective value, the feasibility gap, the number of iterations, the memory usage (in MB), and the cpu time ('hh:mm:ss'). See subsection 7.5.

Dataset	Optimum	Upper bnd.	Objective	Feas. gap	Iteration	Memory	Time
att48	10628	18474	9 076,89	8.18e−4	274725	207	72 : 00 : 00
bayg29	1610	2222	1 498,61	5.43e−4	1000000	197	69 : 06 : 57
bays29	2020	2908	1 855,26	6.49e−4	1000000	193	69 : 03 : 56
berlin52	7542	11463	6 662,82	2.10e−3	228895	267	72 : 00 : 00
bier127	118282	204819	112 696,76	6.14e−2	19905	1006	72 : 00 : 00
brazil58	25395	47362	18 379,06	1.87e−3	173449	262	72 : 00 : 00
burma14	3323	3682	3 153,62	5.15e−4	1000000	219	08 : 22 : 03
ch130	6110	12389	7 874,20	6.79e−2	15373	922	72 : 00 : 00
ch150	6528	14432	13 316,42	2.36e−1	10047	1592	72 : 00 : 00
dantzig42	699	1213	589,86	5.57e−4	350934	262	72 : 00 : 00
eil101	629	1073	618,22	1.60e−2	34377	529	72 : 00 : 00
eil51	426	670	407,33	4.52e−3	213654	202	72 : 00 : 00
eil76	538	890	534,11	1.05e−2	77730	287	72 : 00 : 00
fri26	937	1243	858,27	6.59e−4	1000000	197	56 : 04 : 52
gr120	6942	15525	7 409,45	2.75e−2	23936	742	72 : 00 : 00
gr137	69853	172902	101 276,79	6.71e−2	16900	1206	72 : 00 : 00
gr17	2085	2403	1 800,49	3.84e−4	1000000	214	12 : 35 : 03
gr21	2707	3217	2 570,13	5.52e−4	1000000	193	22 : 26 : 13
gr24	1272	1653	1 136,69	3.47e−4	1000000	197	33 : 36 : 35
gr48	5046	8058	4 460,17	1.09e−3	289217	197	72 : 00 : 00
gr96	55209	121079	51 885,99	1.16e−2	42664	451	72 : 00 : 00
hk48	11461	16491	10 520,26	1.02e−3	257800	197	72 : 00 : 00
kroA100	21282	47891	20 403,88	1.24e−2	38581	529	72 : 00 : 00
kroA150	26524	71197	54 112,35	1.90e−1	10528	1592	72 : 00 : 00
kroB100	22141	58529	20 638,06	1.08e−2	40757	557	72 : 00 : 00
kroB150	26130	69717	46 897,55	1.57e−1	11642	1592	72 : 00 : 00
kroC100	20749	54643	20 017,67	1.15e−2	40813	529	72 : 00 : 00
kroD100	21294	46518	19 866,95	1.48e−2	40801	491	72 : 00 : 00
kroE100	22068	52316	20 819,92	1.77e−2	32288	464	72 : 00 : 00
lin105	14379	38417	13 473,25	1.67e−2	29375	506	72 : 00 : 00
pr107	44303	130840	40 106,77	2.44e−2	24450	512	72 : 00 : 00
pr124	59030	172577	87 856,31	5.25e−2	15894	885	72 : 00 : 00
pr136	96772	240538	151 852,46	1.27e−1	12598	1183	72 : 00 : 00
pr144	58537	216827	172 290,76	2.21e−1	10624	1383	72 : 00 : 00
pr76	108159	195718	87 972,52	6.71e−3	68843	287	72 : 00 : 00
rat99	1211	2365	1 217,21	1.40e−2	33688	529	72 : 00 : 00
rd100	7910	14948	7 460,67	1.03e−2	44344	491	72 : 00 : 00
st70	675	1153	597,10	4.59e−3	114035	267	72 : 00 : 00
swiss42	1273	1725	1 115,10	7.29e−4	416682	262	72 : 00 : 00
ulysses16	6859	7650	5 968,50	8.01e−4	1000000	206	10 : 47 : 06
ulysses22	7013	8865	5 999,69	1.03e−3	1000000	198	26 : 07 : 38

Table SM7: Primal–dual errors from the MaxCut SDP experiment with GSET Benchmark.

	SketchyCGAL	MoSeK	SDPT3	SDPNAL+	Sedumi
Data Name Size (n)	err1 err2 err3 err4 err5 err6 time (dd:hh:mm:ss)				
G1 800	9.2240e-04 0 0 8.8644e-06 4.3895e-05 1.1514e-04 00:00:00:20	1.6451e-04 0 7.4893e-05 0 -9.0090e-05 3.0968e-05 00:00:00:08	5.6194e-13 0 6.3556e-10 0 1.3200e-04 1.3200e-04 00:00:00:06	9.3543e-04 1.6028e-14 2.3076e-04 0 -5.8070e-04 0 00:00:00:56	1.0055e-04 0 0 2.2546e-06 -6.5269e-08 5.1987e-05 00:00:00:35
G2 800	9.6153e-04 0 0 6.8492e-06 -4.9289e-05 1.5580e-04 00:00:00:19	1.6864e-04 0 7.6757e-05 0 -9.2305e-05 3.1789e-05 00:00:00:08	6.3048e-13 0 6.5615e-10 0 1.1758e-04 1.1758e-04 00:00:00:06	2.0071e-05 1.4523e-14 3.4207e-04 0 -5.2680e-04 0 00:00:00:59	9.7060e-05 0 0 2.1165e-06 -5.9330e-08 5.0184e-05 00:00:00:34
G3 800	9.6575e-04 0 0 1.4693e-05 2.2749e-04 1.2655e-04 00:00:00:15	2.4802e-04 0 1.1287e-04 0 -1.3581e-04 4.6700e-05 00:00:00:08	5.0941e-13 0 6.3521e-10 0 1.6284e-04 1.6284e-04 00:00:00:07	2.8275e-04 1.7220e-14 4.6705e-04 0 1.3902e-03 -3.3075e-15 00:00:01:14	8.5287e-05 0 0 2.0824e-06 -5.8110e-08 4.4091e-05 00:00:00:22
G4 800	8.9262e-04 0 0 2.3408e-06 2.0978e-04 2.3739e-04 00:00:00:18	2.5869e-04 0 1.1772e-04 0 -1.4133e-04 4.8998e-05 00:00:00:08	5.5903e-13 0 6.6670e-10 0 1.3819e-04 1.3819e-04 00:00:00:07	1.8571e-04 2.9142e-14 2.0840e-04 0 2.6590e-04 0 00:00:01:03	9.9246e-05 0 0 2.7813e-06 -1.3404e-07 5.1241e-05 00:00:00:27
G5 800	9.6447e-04 0 0 1.9874e-06 1.0526e-04 2.1343e-04 00:00:00:15	1.6316e-04 0 7.4251e-05 0 -8.9228e-05 3.0827e-05 00:00:00:08	5.4488e-13 0 6.2467e-10 0 1.1856e-04 1.1856e-04 00:00:00:06	6.7925e-04 1.3025e-14 8.3802e-05 0 1.6789e-04 0 00:00:01:16	1.2531e-04 0 0 3.1253e-06 -8.5141e-08 6.4780e-05 00:00:00:36

Continued on the next page

Table SM7: Primal–dual errors from the MaxCut SDP experiment with GSET Benchmark (cont.).

	SketchyCGAL	MoSeK	SDPT3	SDPNAL+	Sedumi
G6 800	9.3303e-04	1.0866e-03	5.5518e-13	7.5832e-05	1.2792e-04
	0	0	0	1.5318e-14	0
	0	1.4492e-03	2.8708e-09	7.9668e-04	0
	3.1301e-05	0	0	0	2.1511e-05
	2.1382e-04	-3.3339e-05	7.7761e-04	-1.3520e-03	-4.4192e-07
	1.1763e-04	1.4312e-03	7.7761e-04	3.9961e-15	6.5767e-05
	00:00:00:24	00:00:00:09	00:00:00:09	00:00:00:55	00:00:00:31
G7 800	8.5782e-04	1.2372e-03	5.5790e-13	6.0123e-05	9.4592e-05
	0	0	0	2.6558e-14	0
	0	1.5941e-03	2.5753e-09	7.8940e-04	0
	3.1960e-05	0	0	0	1.3177e-05
	-3.7838e-05	3.7809e-05	9.3721e-04	-1.4522e-03	-2.7617e-07
	6.9057e-05	1.7315e-03	9.3721e-04	4.2082e-15	4.8682e-05
	00:00:00:26	00:00:00:07	00:00:00:06	00:00:00:56	00:00:00:24
G8 800	9.3260e-04	1.0938e-03	5.5102e-13	1.7077e-04	1.2127e-04
	0	0	0	1.4665e-14	0
	0	1.4105e-03	2.4743e-09	7.9436e-04	0
	4.1592e-05	0	0	0	2.0611e-05
	-3.1418e-05	3.7726e-05	8.5330e-04	-1.4241e-03	-4.5087e-07
	-8.7893e-05	1.5324e-03	8.5331e-04	3.0468e-15	6.2313e-05
	00:00:00:23	00:00:00:08	00:00:00:10	00:00:01:01	00:00:00:28
G9 800	9.5542e-04	1.2435e-03	2.2281e-14	1.0045e-04	1.2441e-04
	0	0	0	1.9571e-14	0
	0	3.1113e-04	2.5145e-10	8.5002e-04	0
	5.0016e-05	0	0	0	2.0214e-05
	-1.2117e-04	1.5484e-05	5.3311e-05	-1.6243e-03	-4.2149e-07
	3.3033e-06	1.6621e-03	5.3311e-05	-1.0117e-14	6.3972e-05
	00:00:00:26	00:00:00:07	00:00:00:07	00:00:00:49	00:00:00:27
G10 800	9.6338e-04	1.1629e-03	8.5789e-13	1.0744e-04	3.7687e-04
	0	0	0	8.3036e-14	0
	0	1.5074e-03	2.7911e-09	4.4804e-04	0
	3.6980e-05	0	0	0	5.6003e-05
	7.1590e-05	3.8077e-05	6.5403e-04	3.4374e-04	-1.2148e-06
	-1.0780e-05	1.6306e-03	6.5403e-04	-2.8520e-14	1.9384e-04
	00:00:00:23	00:00:00:07	00:00:00:06	00:00:01:06	00:00:00:21
G11 800	9.4272e-04	2.5133e-15	1.0470e-13	2.0571e-05	2.2880e-04
	0	0	0	1.1073e-14	0
	0	1.9802e-03	5.0100e-10	2.2129e-04	0
	9.4851e-06	0	0	0	3.1321e-05
	6.8661e-04	-6.8224e-05	3.2120e-04	2.3467e-05	-7.2874e-07
	5.9825e-04	4.6042e-03	3.2120e-04	9.4460e-15	1.1762e-04
	00:00:00:17	00:00:00:06	00:00:00:04	00:00:01:03	00:00:00:18

Continued on the next page

Table SM7: Primal–dual errors from the MaxCut SDP experiment with GSET Benchmark (cont.).

	SketchyCGAL	MoSeK	SDPT3	SDPNAL+	Sedumi
G12 800	9.6198e-04	1.8094e-15	3.7075e-14	1.7431e-05	2.3199e-04
	0	0	0	1.7518e-14	0
	0	2.3384e-03	5.0100e-10	1.5710e-04	0
	4.1334e-05	0	0	0	3.2627e-05
	2.5989e-04	7.0095e-06	4.1058e-04	-6.5189e-05	-7.3963e-07
	3.2035e-04	6.2413e-03	4.1058e-04	1.7019e-14	1.1926e-04
	00:00:00:16	00:00:00:09	00:00:00:09	00:00:01:16	00:00:00:18
G13 800	9.0704e-04	1.6417e-15	1.9662e-14	2.4604e-04	2.3859e-04
	0	0	0	2.3089e-14	0
	0	3.3393e-03	7.4695e-10	2.2677e-04	0
	4.6549e-05	0	0	0	3.3410e-05
	1.9049e-04	-1.2000e-04	1.6121e-04	-5.2514e-04	-7.2136e-07
	3.1058e-04	7.9402e-03	1.6121e-04	1.4203e-14	1.2269e-04
	00:00:00:11	00:00:00:09	00:00:00:08	00:00:00:55	00:00:00:20
G14 800	8.7483e-04	6.8671e-04	0	2.7915e-05	1.1104e-04
	0	0	0	2.9734e-14	0
	0	4.2273e-04	1.6712e-10	3.0881e-04	0
	2.1202e-05	0	0	0	2.9799e-05
	3.8840e-04	-3.4869e-04	6.5529e-04	-7.4014e-04	-8.6246e-07
	4.1983e-04	2.4403e-04	6.5529e-04	0	5.6610e-05
	00:00:00:16	00:00:00:08	00:00:00:06	00:00:01:27	00:00:00:23
G15 800	9.4137e-04	2.0233e-04	1.1360e-15	8.3307e-04	2.8862e-04
	0	0	0	5.5046e-14	0
	0	1.2268e-04	1.2565e-11	9.7876e-05	0
	3.8869e-05	0	0	0	8.9568e-05
	3.0556e-04	-1.0261e-04	4.5321e-04	-4.7425e-04	-2.9107e-06
	2.8753e-04	7.2245e-05	4.5321e-04	1.0909e-14	1.4647e-04
	00:00:00:19	00:00:00:12	00:00:00:06	00:00:02:16	00:00:00:23
G16 800	9.4043e-04	6.7439e-04	2.7704e-15	1.7499e-04	1.8725e-04
	0	0	0	9.7405e-14	0
	0	4.1500e-04	1.6493e-10	2.1924e-04	0
	2.2862e-05	0	0	0	4.7086e-05
	3.7868e-04	-3.4259e-04	6.8970e-04	-7.9071e-04	-1.5078e-06
	4.2238e-04	2.4011e-04	6.8970e-04	-3.4307e-15	9.5411e-05
	00:00:00:16	00:00:00:08	00:00:00:05	00:00:01:12	00:00:00:24
G17 800	9.3256e-04	7.6583e-04	1.1136e-14	8.0914e-04	1.3170e-04
	0	0	0	6.2389e-14	0
	0	4.6857e-04	1.7499e-10	2.2869e-04	0
	2.9837e-05	0	0	0	3.7355e-05
	3.4252e-04	-3.8909e-04	6.2508e-04	-9.5150e-04	-1.3348e-06
	3.9942e-04	2.7278e-04	6.2508e-04	0	6.6833e-05
	00:00:00:15	00:00:00:08	00:00:00:07	00:00:01:48	00:00:00:25

Continued on the next page

Table SM7: Primal–dual errors from the MaxCut SDP experiment with GSET Benchmark (cont.).

	SketchyCGAL	MoSeK	SDPT3	SDPNAL+	Sedumi
G18 800	9.4244e-04	3.8563e-03	1.6886e-15	2.8812e-05	8.5070e-05
	0	0	0	1.8694e-13	0
	0	1.1277e-03	6.0189e-11	3.1932e-04	0
	9.0768e-05	0	0	0	1.6922e-05
	-3.7838e-04	-1.1201e-04	3.9986e-04	4.5379e-04	-3.5880e-07
	-2.4600e-04	6.1459e-03	3.9986e-04	6.9407e-14	4.3661e-05
	00:00:00:22	00:00:00:08	00:00:00:06	00:00:01:11	00:00:00:25
G19 800	9.4555e-04	1.5963e-03	1.4078e-15	1.5129e-04	1.3630e-04
	0	0	0	4.0056e-14	0
	0	4.3819e-04	5.7485e-11	6.4481e-04	0
	1.1372e-04	0	0	0	3.6206e-05
	-3.2867e-04	8.5952e-05	7.3135e-04	-1.4201e-03	-7.6293e-07
	-3.3669e-04	2.8085e-03	7.3135e-04	8.2191e-15	6.9764e-05
	00:00:00:25	00:00:00:08	00:00:00:06	00:00:01:19	00:00:00:27
G20 800	9.3800e-04	1.0881e-03	2.5029e-15	4.4890e-04	2.6035e-04
	0	0	0	3.3210e-13	0
	0	2.0899e-03	6.0636e-11	2.3547e-04	0
	9.7939e-05	0	0	0	4.8009e-05
	-1.7534e-04	2.2938e-05	7.0134e-04	-6.7662e-05	-1.0140e-06
	-1.3401e-04	1.9626e-03	7.0134e-04	1.0248e-13	1.3370e-04
	00:00:00:20	00:00:00:08	00:00:00:06	00:00:01:22	00:00:00:27
G21 800	9.2523e-04	4.6484e-03	2.4371e-15	6.9572e-05	1.4658e-04
	0	0	0	2.8251e-14	0
	0	1.3595e-03	6.1223e-11	7.2098e-04	0
	1.0570e-04	0	0	0	3.0686e-05
	-2.7449e-04	1.4565e-04	5.4134e-04	-1.7175e-03	-6.6934e-07
	-3.0912e-04	7.8067e-03	5.4134e-04	3.8137e-14	7.5175e-05
	00:00:00:19	00:00:00:08	00:00:00:07	00:00:01:00	00:00:00:24
G22 2000	9.4106e-04	4.2850e-04	2.8563e-13	7.2831e-05	6.1785e-05
	0	0	0	5.0064e-14	0
	0	5.9683e-04	3.9451e-09	1.6903e-04	0
	6.3403e-06	0	0	0	1.8727e-06
	4.2900e-05	-3.0984e-04	7.3429e-04	-6.1514e-04	-2.4181e-08
	1.5465e-04	1.9025e-04	7.3429e-04	2.6991e-15	3.1558e-05
	00:00:00:37	00:00:01:55	00:00:00:54	00:00:13:00	00:00:08:59
G23 2000	9.4571e-04	4.5428e-04	2.8472e-13	2.1546e-05	4.5267e-04
	0	0	0	4.9131e-14	0
	0	6.3328e-04	4.0224e-09	1.7164e-04	0
	5.7582e-06	0	0	0	1.3276e-05
	8.9313e-05	-3.2834e-04	6.9341e-04	-6.0588e-04	-1.6808e-07
	1.0244e-04	2.0180e-04	6.9342e-04	1.2644e-14	2.3122e-04
	00:00:00:39	00:00:03:05	00:00:01:24	00:00:12:47	00:00:06:57

Continued on the next page

Table SM7: Primal–dual errors from the MaxCut SDP experiment with GSET Benchmark (cont.).

	SketchyCGAL	MoSeK	SDPT3	SDPNAL+	Sedumi
G24 2000	9.6919e-04	4.4984e-04	2.9846e-13	1.6084e-05	1.4380e-04
	0	0	0	4.4865e-14	0
	0	6.2656e-04	3.9715e-09	1.9607e-04	0
	4.4040e-06	0	0	0	4.4677e-06
	1.5659e-04	-3.2516e-04	7.0366e-04	-6.9019e-04	-6.4201e-08
	1.0454e-04	1.9981e-04	7.0366e-04	-9.7938e-15	7.3443e-05
	00:00:00:32	00:00:02:05	00:00:00:54	00:00:22:12	00:00:07:04
G25 2000	9.7551e-04	4.2290e-04	2.9191e-13	1.0653e-04	5.6048e-04
	0	0	0	2.3325e-14	0
	0	5.8914e-04	3.9520e-09	8.8316e-04	0
	5.0085e-06	0	0	0	1.8061e-05
	9.7544e-05	-3.0561e-04	7.3658e-04	-2.6564e-03	-2.2954e-07
	9.9098e-05	1.8791e-04	7.3658e-04	0	2.8627e-04
	00:00:00:34	00:00:08:45	00:00:00:53	00:00:29:46	00:00:07:34
G26 2000	9.4268e-04	3.8322e-04	3.0824e-13	2.6009e-05	1.4346e-04
	0	0	0	5.1837e-14	0
	0	5.3371e-04	4.0921e-09	1.8849e-04	0
	5.5424e-06	0	0	0	4.7426e-06
	1.2342e-04	-2.7715e-04	7.7996e-04	-6.7260e-04	-6.8321e-08
	1.3050e-04	1.7010e-04	7.7997e-04	7.0813e-15	7.3263e-05
	00:00:00:43	00:00:02:07	00:00:01:05	00:00:12:41	00:00:07:32
G27 2000	9.6526e-04	2.0634e-03	8.3648e-15	3.4442e-04	2.7095e-04
	0	0	0	4.8075e-14	0
	0	4.7227e-04	9.4357e-09	5.7485e-04	0
	1.5259e-05	0	0	0	2.8739e-05
	3.5349e-05	1.0351e-05	2.5244e-04	2.0859e-05	-3.8409e-07
	4.7595e-05	3.0249e-03	2.5245e-04	1.7521e-13	1.3810e-04
	00:00:00:48	00:00:01:56	00:00:01:12	00:00:16:57	00:00:05:47
G28 2000	9.7225e-04	1.9942e-03	1.1612e-14	2.9845e-04	2.3205e-04
	0	0	0	4.8939e-14	0
	0	4.6056e-04	9.6511e-09	6.0724e-04	0
	1.8238e-05	0	0	0	2.5077e-05
	1.3091e-04	2.5565e-05	2.7169e-04	2.5404e-05	-3.3463e-07
	1.0975e-04	2.9542e-03	2.7169e-04	-2.1276e-14	1.1827e-04
	00:00:01:03	00:00:02:53	00:00:01:13	00:00:18:03	00:00:06:29
G29 2000	9.4325e-04	2.1648e-03	1.0602e-14	5.1373e-05	2.3064e-04
	0	0	0	7.7832e-14	0
	0	4.8413e-04	9.5579e-09	1.3324e-04	0
	1.1656e-05	0	0	0	2.2448e-05
	7.8858e-05	-2.1517e-05	2.6210e-04	5.2998e-05	-2.8169e-07
	1.0417e-04	3.1277e-03	2.6210e-04	6.1585e-14	1.1760e-04
	00:00:01:00	00:00:02:53	00:00:01:12	00:00:24:10	00:00:04:31

Continued on the next page

Table SM7: Primal–dual errors from the MaxCut SDP experiment with GSET Benchmark (cont.).

	SketchyCGAL	MoSeK	SDPT3	SDPNAL+	Sedumi
G30 2000	9.5286e-04	2.4869e-03	8.3719e-15	2.4693e-04	2.4866e-04
	0	0	0	1.0288e-13	0
	0	5.6779e-04	9.4828e-09	2.4961e-04	0
	1.7223e-05	0	0	0	2.6003e-05
	-4.1280e-05	-3.0774e-05	2.8462e-04	6.6465e-05	-3.2359e-07
	1.3337e-05	3.5902e-03	2.8463e-04	2.7779e-13	1.2677e-04
	00:00:00:49	00:00:01:53	00:00:01:48	00:00:24:38	00:00:05:44
G31 2000	9.6727e-04	1.7409e-03	1.1184e-14	2.8342e-05	2.1105e-04
	0	0	0	9.7757e-14	0
	0	4.0157e-04	3.3898e-09	2.4348e-04	0
	1.8430e-05	0	0	0	2.2272e-05
	-3.6136e-05	1.7021e-05	2.4462e-04	-1.1642e-04	-2.8744e-07
	2.5746e-06	2.5715e-03	2.4462e-04	-2.6184e-13	1.0759e-04
	00:00:00:57	00:00:02:36	00:00:01:49	00:00:33:26	00:00:05:48
G32 2000	9.5091e-04	2.7476e-15	5.4729e-14	3.9317e-04	3.6993e-04
	0	0	0	1.1256e-14	0
	0	2.7594e-03	4.5010e-09	2.8428e-04	0
	6.8408e-06	0	0	0	3.2460e-05
	3.4139e-04	-2.7464e-05	6.0739e-04	-4.6221e-04	-4.6052e-07
	3.8942e-04	6.9408e-03	6.0739e-04	-3.4964e-15	1.8858e-04
	00:00:00:38	00:00:01:41	00:00:00:37	00:00:19:32	00:00:10:27
G33 2000	9.4110e-04	2.5624e-15	3.1971e-14	2.2252e-04	3.8940e-04
	0	0	0	1.0459e-14	0
	0	2.6593e-03	4.5010e-09	3.8443e-04	0
	1.1287e-05	0	0	0	3.4800e-05
	3.2243e-04	3.3035e-05	6.0105e-04	-7.4134e-04	-4.9724e-07
	3.6409e-04	6.8252e-03	6.0106e-04	-3.9643e-15	1.9849e-04
	00:00:00:29	00:00:01:40	00:00:00:31	00:00:16:18	00:00:10:27
G34 2000	9.6895e-04	2.2269e-15	1.0497e-15	8.0679e-05	1.2211e-04
	0	0	0	1.5713e-14	0
	0	2.8937e-03	4.5010e-09	2.6039e-04	0
	1.1397e-05	0	0	0	1.0901e-05
	2.9852e-04	6.0226e-05	8.6747e-04	-5.4601e-04	-1.5148e-07
	3.5070e-04	7.6479e-03	8.6747e-04	-7.3622e-15	6.2250e-05
	00:00:00:29	00:00:01:52	00:00:00:35	00:00:15:19	00:00:10:21
G35 2000	8.9543e-04	4.3815e-04	2.0904e-15	6.0172e-05	1.5840e-04
	0	0	0	6.7384e-14	0
	0	2.5141e-04	1.1770e-11	1.8310e-04	0
	1.4946e-05	0	0	0	3.7964e-05
	3.8998e-04	-2.1941e-04	1.3746e-04	4.7801e-04	-7.6303e-07
	4.6150e-04	1.5373e-04	1.3746e-04	1.6827e-15	8.0201e-05
	00:00:00:50	00:00:02:16	00:00:00:53	00:00:18:38	00:00:13:48

Continued on the next page

Table SM7: Primal–dual errors from the MaxCut SDP experiment with GSET Benchmark (cont.).

	SketchyCGAL	MoSeK	SDPT3	SDPNAL+	Sedumi
G36 2000	6.8005e-04	7.4971e-04	1.6344e-15	4.2775e-05	1.9278e-04
	0	0	0	1.7356e-13	0
	0	4.3186e-04	1.0779e-11	2.3336e-04	0
	1.1880e-05	0	0	0	7.3947e-05
	4.9472e-04	-3.7549e-04	5.9490e-04	4.9405e-04	-3.0945e-06
	5.0098e-04	2.6313e-04	5.9490e-04	-1.8953e-14	9.5444e-05
	00:00:01:19	00:00:02:23	00:00:00:52	00:00:19:36	00:00:12:52
G37 2000	7.2413e-04	5.7490e-04	2.1943e-15	3.0272e-04	2.2910e-04
	0	0	0	1.1485e-13	0
	0	3.2790e-04	1.1286e-11	2.0440e-04	0
	1.1905e-05	0	0	0	7.5591e-05
	5.5250e-04	-2.8794e-04	2.3988e-04	7.1309e-04	-2.6385e-06
	5.6725e-04	2.0161e-04	2.3988e-04	0	1.1446e-04
	00:00:01:05	00:00:02:37	00:00:00:53	00:00:25:50	00:00:09:27
G38 2000	6.9024e-04	5.9235e-04	1.3810e-15	5.2776e-04	1.1645e-04
	0	0	0	1.4494e-13	0
	0	3.3869e-04	1.0665e-11	1.5038e-04	0
	1.2879e-05	0	0	0	3.3208e-05
	4.9102e-04	-2.9665e-04	4.6436e-04	3.9093e-04	-6.5520e-07
	4.7173e-04	2.0780e-04	4.6436e-04	-8.2949e-15	5.8869e-05
	00:00:01:08	00:00:02:43	00:00:00:49	00:00:25:53	00:00:09:05
G39 2000	9.6781e-04	2.1901e-03	3.1846e-15	5.5948e-05	1.7300e-04
	0	0	0	9.9559e-14	0
	0	4.9220e-04	1.2180e-10	3.8443e-04	0
	4.7806e-05	0	0	0	4.1089e-05
	-1.3034e-04	-1.1277e-05	7.8432e-04	5.2063e-04	-5.2393e-07
	-5.0727e-05	3.6071e-03	7.8432e-04	-9.6073e-15	8.7894e-05
	00:00:01:46	00:00:02:30	00:00:01:03	00:00:24:44	00:00:10:02
G40 2000	9.6841e-04	3.2031e-03	1.6078e-14	5.5471e-04	1.0523e-04
	0	0	0	1.0969e-13	0
	0	7.2687e-04	1.3897e-11	7.4617e-04	0
	5.6218e-05	0	0	0	1.9443e-05
	-1.1338e-04	5.5916e-05	1.4461e-04	1.5733e-03	-2.4104e-07
	-1.0103e-04	5.3231e-03	1.4461e-04	9.3855e-14	5.3539e-05
	00:00:01:40	00:00:01:49	00:00:01:05	00:01:28:45	00:00:10:28
G41 2000	9.4959e-04	2.9907e-03	1.8685e-14	1.4761e-04	3.6781e-04
	0	0	0	7.8040e-14	0
	0	6.7766e-04	3.7649e-11	6.7791e-04	0
	7.0674e-05	0	0	0	5.4200e-05
	-1.5249e-04	8.6809e-06	8.1046e-05	1.2410e-03	-6.9022e-07
	-9.7060e-05	4.9304e-03	8.1046e-05	-1.8967e-15	1.8730e-04
	00:00:01:21	00:00:02:35	00:00:01:38	00:01:30:45	00:00:08:27

Continued on the next page

Table SM7: Primal–dual errors from the MaxCut SDP experiment with GSET Benchmark (cont.).

	SketchyCGAL	MoSeK	SDPT3	SDPNAL+	Sedumi
G42 2000	9.7592e-04	2.0139e-03	2.6746e-15	8.3607e-05	1.0746e-04
	0	0	0	1.4588e-13	0
	0	4.6025e-04	3.9803e-11	6.1700e-04	0
	6.1789e-05	0	0	0	2.5823e-05
	-1.2744e-04	-4.2775e-05	7.4182e-04	1.2828e-03	-3.2331e-07
	-1.7927e-04	3.2573e-03	7.4182e-04	1.4930e-13	5.4599e-05
	00:00:01:33	00:00:02:18	00:00:00:52	00:00:18:50	00:00:09:19
G43 1000	9.3475e-04	2.1860e-04	5.9232e-13	4.1395e-04	1.2988e-04
	0	0	0	4.8906e-14	0
	0	3.0691e-04	1.5921e-09	2.1816e-04	0
	1.0919e-05	0	0	0	5.1650e-06
	2.6443e-04	-1.6022e-04	2.1419e-04	3.3428e-04	-9.9318e-08
	1.6891e-04	9.7363e-05	2.1419e-04	-8.3658e-15	6.6887e-05
	00:00:00:14	00:00:00:24	00:00:00:10	00:00:02:00	00:00:00:53
G44 1000	9.5960e-04	1.8448e-04	5.2090e-13	6.6220e-04	9.1836e-05
	0	0	0	5.4161e-14	0
	0	2.5894e-04	1.4687e-09	1.8707e-04	0
	1.1915e-05	0	0	0	3.8575e-06
	1.4291e-04	-1.3529e-04	1.9904e-04	3.7178e-05	-7.2956e-08
	9.0926e-05	8.2099e-05	1.9904e-04	-8.1424e-15	4.7294e-05
	00:00:00:14	00:00:00:16	00:00:00:09	00:00:01:36	00:00:00:56
G45 1000	9.3935e-04	1.7449e-04	5.7531e-13	8.8850e-04	1.2776e-04
	0	0	0	5.3692e-14	0
	0	2.4507e-04	1.5085e-09	2.0479e-04	0
	8.5894e-06	0	0	0	4.9143e-06
	9.6997e-05	-1.2802e-04	2.8376e-04	-1.1397e-04	-9.9553e-08
	1.4242e-04	7.7608e-05	2.8376e-04	9.6044e-15	6.5796e-05
	00:00:00:15	00:00:00:15	00:00:00:10	00:00:01:45	00:00:00:54
G46 1000	9.5919e-04	1.7429e-04	5.7220e-13	1.2383e-04	9.9527e-05
	0	0	0	3.2220e-14	0
	0	2.4488e-04	1.5276e-09	3.0758e-04	0
	1.0933e-05	0	0	0	3.8482e-06
	1.1956e-04	-1.2778e-04	1.5929e-04	7.7609e-04	-7.2593e-08
	8.8945e-05	7.7594e-05	1.5929e-04	-7.0508e-15	5.1261e-05
	00:00:00:15	00:00:00:21	00:00:00:10	00:00:02:19	00:00:00:55
G47 1000	9.5642e-04	2.2700e-04	5.3453e-13	9.5704e-04	1.1095e-04
	0	0	0	5.1509e-14	0
	0	3.1874e-04	1.4995e-09	1.6724e-04	0
	8.0600e-06	0	0	0	4.1527e-06
	2.1827e-04	-1.6626e-04	1.9451e-04	-2.2679e-04	-7.9365e-08
	1.3089e-04	1.0118e-04	1.9451e-04	-9.8917e-15	5.7148e-05
	00:00:00:14	00:00:00:15	00:00:00:15	00:00:02:42	00:00:00:38

Continued on the next page

Table SM7: Primal–dual errors from the MaxCut SDP experiment with GSET Benchmark (cont.).

	SketchyCGAL	MoSeK	SDPT3	SDPNAL+	Sedumi
G48 3000	9.6092e-04	1.1491e-14	2.7859e-14	1.7221e-05	4.5896e-08
	0	0	0	1.1588e-14	0
	0	4.7998e-04	1.1881e-07	8.7379e-04	0
	7.2356e-05	0	0	0	2.1067e-09
	-8.1065e-04	-6.6175e-05	1.2623e-04	-1.5367e-03	-9.9465e-09
	-6.5493e-04	1.3226e-04	1.2627e-04	-2.1276e-14	1.3419e-08
	00:00:00:13	00:00:06:07	00:00:01:23	00:01:47:16	00:00:17:05
G49 3000	8.7711e-04	8.6008e-15	4.2404e-13	1.6246e-04	1.1879e-04
	0	0	0	1.2076e-14	0
	0	1.0779e-04	5.4479e-07	2.1347e-04	0
	3.1797e-05	0	0	0	3.7620e-06
	-4.0582e-04	-1.4857e-05	9.4205e-04	-1.8609e-03	-2.0698e-08
	-1.6411e-04	2.9698e-05	9.4220e-04	1.7790e-15	6.0453e-05
	00:00:00:14	00:00:07:37	00:00:01:26	00:00:34:06	00:00:19:11
G50 3000	9.0458e-04	1.1434e-14	4.0840e-13	1.3191e-04	7.0325e-04
	0	0	0	1.7726e-14	0
	0	1.8344e-03	5.3630e-07	6.7731e-04	0
	1.4867e-05	0	0	0	2.2364e-05
	2.8789e-05	-2.5364e-04	6.1843e-04	1.0880e-03	-1.2859e-07
	6.1924e-05	5.0555e-04	6.1859e-04	4.5250e-14	3.5788e-04
	00:00:00:15	00:00:05:51	00:00:01:13	00:01:03:59	00:00:19:15
G51 1000	9.4758e-04	1.8096e-04	0	9.4846e-05	3.2323e-04
	0	0	0	8.4322e-14	0
	0	1.0685e-04	1.2125e-11	6.6263e-05	0
	2.2294e-05	0	0	0	8.7008e-05
	5.4910e-04	-9.1780e-05	4.7349e-04	2.3462e-05	-2.8725e-06
	4.9088e-04	6.3721e-05	4.7349e-04	-5.7231e-15	1.6383e-04
	00:00:00:25	00:00:00:17	00:00:00:13	00:00:02:32	00:00:00:52
G52 1000	9.3422e-04	2.3942e-04	0	3.6346e-05	2.7082e-04
	0	0	0	8.7588e-15	0
	0	1.4337e-04	1.2359e-11	3.3592e-04	0
	2.0451e-05	0	0	0	7.9081e-05
	4.9357e-04	-1.2145e-04	3.8698e-04	-1.9403e-05	-2.8396e-06
	4.8520e-04	8.4252e-05	3.8698e-04	0	1.3683e-04
	00:00:00:22	00:00:00:20	00:00:00:10	00:00:03:19	00:00:00:51
G53 1000	8.9247e-04	2.9963e-04	0	2.7250e-05	7.1987e-05
	0	0	0	2.6300e-14	0
	0	1.7934e-04	1.2351e-11	1.1149e-04	0
	2.0807e-05	0	0	0	1.7338e-05
	4.4112e-04	-1.5194e-04	3.8368e-04	-2.4948e-04	-5.1172e-07
	4.6495e-04	1.0550e-04	3.8368e-04	2.8611e-15	3.6615e-05
	00:00:00:20	00:00:00:20	00:00:00:10	00:00:02:14	00:00:00:45

Continued on the next page

Table SM7: Primal–dual errors from the MaxCut SDP experiment with GSET Benchmark (cont.).

	SketchyCGAL	MoSeK	SDPT3	SDPNAL+	Sedumi
G54 1000	8.2442e-04	9.6590e-04	0	6.3149e-04	2.3269e-04
	0	0	0	7.8280e-14	0
	0	5.7862e-04	1.2365e-11	1.2067e-04	0
	1.9272e-05	0	0	0	5.3841e-05
	4.7339e-04	-4.9058e-04	4.3025e-04	-1.4418e-04	-2.0052e-06
	4.2979e-04	3.3945e-04	4.3025e-04	-5.4878e-15	1.1801e-04
	00:00:00:28	00:00:00:17	00:00:00:15	00:00:02:28	00:00:01:17
G55 5000	9.8552e-04	2.4798e-04	2.8764e-15	7.7904e-04	1.9044e-04
	0	0	0	3.4424e-14	0
	0	5.6271e-04	2.5720e-09	1.1896e-04	0
	5.0552e-06	0	0	0	7.9203e-06
	4.1808e-04	-1.4241e-04	7.4066e-05	-4.3644e-04	-5.9389e-08
	4.2793e-04	2.2303e-04	7.4067e-05	2.0123e-14	9.6502e-05
	00:00:05:54	00:00:42:38	00:00:10:21	00:23:04:37	00:03:04:42
G56 5000	9.7842e-04	9.7935e-04	2.5583e-15	8.0068e-05	1.5314e-04
	0	0	0	6.4741e-14	0
	0	2.6628e-03	1.3289e-08	1.2807e-04	0
	8.9039e-06	0	0	0	1.2366e-05
	1.1780e-04	5.5033e-06	5.5448e-04	-4.1735e-04	-9.7127e-08
	1.0618e-04	2.0459e-03	5.5448e-04	7.5565e-15	7.7549e-05
	00:00:03:35	00:00:31:46	00:00:08:57	00:23:11:02	00:03:00:38
G57 5000	9.7494e-04	3.1018e-15	3.7205e-15	3.8694e-04	2.5653e-04
	0	0	0	9.8838e-15	0
	0	3.5240e-03	4.0510e-09	2.9910e-04	0
	4.7070e-06	0	0	0	1.4371e-05
	2.1366e-04	2.1686e-05	5.9892e-04	-4.2607e-04	-1.2404e-07
	2.7322e-04	8.6078e-03	5.9892e-04	-3.4877e-14	1.2994e-04
	00:00:01:40	00:00:23:26	00:00:05:45	00:19:20:50	00:02:30:43
G58 5000	4.0780e-04	2.6793e-04	2.0718e-15	5.3746e-04	1.8343e-04
	0	0	0	6.6770e-14	0
	0	1.4522e-04	9.0746e-12	3.6710e-04	0
	4.9369e-06	0	0	0	6.9383e-05
	4.9332e-04	-1.3301e-04	9.4745e-04	2.4708e-04	-1.9770e-06
	4.9184e-04	9.3117e-05	9.4745e-04	1.7301e-14	9.1031e-05
	00:00:10:32	00:00:47:06	00:00:10:57	00:23:08:17	00:03:35:20
G59 5000	9.8314e-04	2.2738e-03	4.8543e-15	2.3647e-04	2.6866e-04
	0	0	0	1.4684e-13	0
	0	4.0049e-04	1.8306e-11	9.6139e-05	0
	2.6813e-05	0	0	0	2.8799e-05
	-7.4885e-05	-1.5350e-05	4.0915e-04	-1.4136e-04	-2.2324e-07
	3.1707e-05	3.6954e-03	4.0915e-04	9.3195e-15	1.3600e-04
	00:00:09:36	00:00:37:11	00:00:11:56	00:16:04:30	00:03:46:58

Continued on the next page

Table SM7: Primal–dual errors from the MaxCut SDP experiment with GSET Benchmark (cont.).

	SketchyCGAL	MoSeK	SDPT3	SDPNAL+	Sedumi
G60 7000	9.4222e-04	2.9488e-04	2.7808e-15	9.3599e-05	4.3747e-04
	0	0	0	4.3064e-14	0
	0	6.7677e-04	4.2801e-09	1.3113e-04	0
	1.9291e-06	0	0	0	1.5191e-05
	7.1170e-05	-1.6815e-04	1.6532e-04	-2.2004e-04	-8.6697e-08
	4.0690e-05	2.6756e-04	1.6533e-04	-4.5629e-14	2.2125e-04
	00:00:22:47	00:01:20:38	00:00:24:05	00:16:47:49	00:08:17:19
G61 7000	7.4498e-04	1.1123e-03	3.0992e-15	2.7651e-04	1.4060e-04
	0	0	0	4.9482e-14	0
	0	3.0721e-03	6.3286e-09	2.7933e-04	0
	5.7313e-06	0	0	0	9.0226e-06
	5.2060e-05	-3.0045e-05	1.8713e-04	-1.1403e-03	-6.6698e-08
	5.4104e-05	2.2545e-03	1.8713e-04	-9.7085e-14	7.1066e-05
	00:00:07:42	00:01:29:46	00:00:23:22	00:16:39:51	00:08:25:12
G62 7000	9.8167e-04	3.2885e-15	1.0568e-15	2.6213e-04	2.8277e-04
	0	0	0	1.4285e-14	0
	0	3.2915e-03	6.7510e-09	1.5326e-04	0
	3.3956e-06	0	0	0	1.3510e-05
	2.2564e-04	3.1229e-05	5.9666e-04	-2.4200e-04	-9.7909e-08
	2.1179e-04	8.1609e-03	5.9666e-04	4.1265e-14	1.4296e-04
	00:00:03:09	00:01:18:18	00:00:14:27	00:10:27:25	00:07:20:45
G63 7000	9.2555e-04	2.9923e-04	2.8620e-15	1.0476e-04	1.6669e-04
	0	0	0	9.4341e-14	0
	0	1.6003e-04	4.9605e-12	3.8707e-04	0
	5.6376e-06	0	0	0	5.4369e-05
	4.8804e-04	-1.4816e-04	4.8130e-04	1.3827e-05	-1.4318e-06
	7.6415e-04	1.0374e-04	4.8130e-04	-8.7045e-15	8.2908e-05
	00:00:17:25	00:02:10:33	00:00:31:58	00:14:22:49	00:10:26:11
G64 7000	9.8308e-04	3.8323e-03	3.6898e-14	8.8455e-04	2.1453e-04
	0	0	0	6.5761e-13	0
	0	6.5016e-04	1.7381e-11	1.3625e-04	0
	2.6912e-05	0	0	0	1.5266e-05
	3.6805e-05	-9.8637e-05	1.4654e-04	-4.6901e-05	-9.7600e-08
	-5.6791e-05	6.0347e-03	1.4654e-04	-3.2217e-13	1.0844e-04
	00:00:17:25	00:01:28:39	00:00:33:48	00:18:37:55	00:08:04:41
G65 8000	9.8854e-04	3.4940e-15	3.8001e-15	5.3185e-04	–
	0	0	0	7.4499e-15	–
	0	3.4543e-03	6.7510e-09	3.1610e-04	–
	3.8799e-06	0	0	0	–
	1.9849e-04	2.8883e-05	8.5582e-04	-5.2525e-04	–
	2.1539e-04	8.3666e-03	8.5582e-04	1.3871e-14	–
	00:00:02:49	00:01:34:04	00:00:49:44	00:22:46:02	–

Continued on the next page

Table SM7: Primal–dual errors from the MaxCut SDP experiment with GSET Benchmark (cont.).

	SketchyCGAL	MoSeK	SDPT3	SDPNAL+	Sedumi
G66 9000	9.8472e-04	3.5663e-15	6.7137e-15	4.6474e-04	–
	0	0	0	1.1480e-14	–
	0	3.4548e-03	6.7510e-09	2.2088e-04	–
	3.8099e-06	0	0	0	–
	1.9103e-04	-2.5418e-05	8.9009e-04	-5.4521e-04	–
	1.9216e-04	8.2468e-03	8.9009e-04	-1.4191e-14	–
	00:00:02:55	00:02:18:24	00:01:15:51	01:07:25:38	–
G67 10000	9.8870e-04	–	1.0890e-15	9.7122e-04	–
	0	–	0	1.5106e-14	–
	0	–	2.0260e-09	4.3451e-05	–
	3.7053e-06	–	0	0	–
	2.5490e-04	–	4.5646e-04	-2.3706e-04	–
	1.8960e-04	–	4.5646e-04	-2.6790e-15	–
	00:00:03:46	–	00:00:45:40	01:20:08:29	–

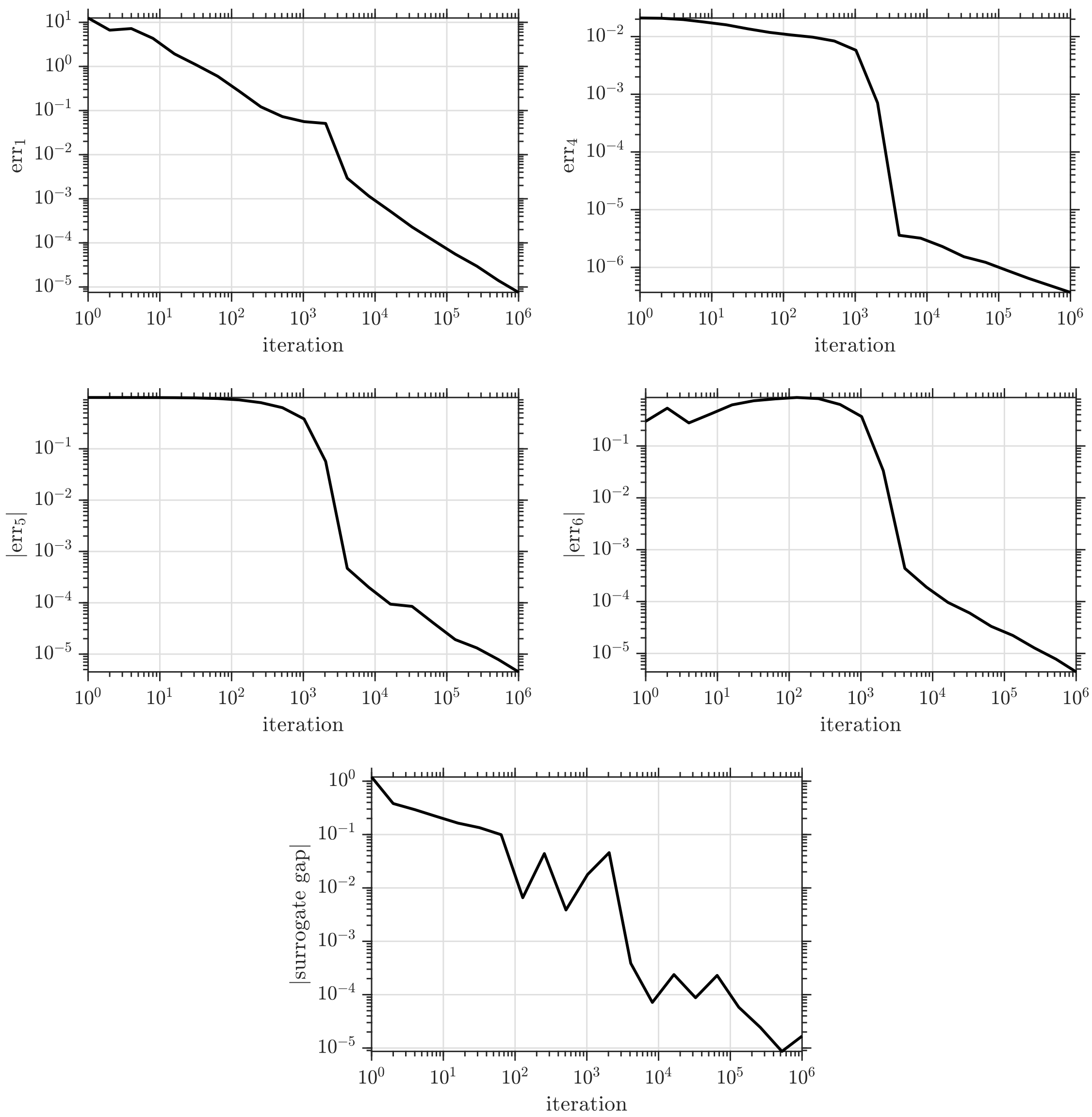


Figure SM7.3. MaxCut SDP: Primal–dual convergence. *We solve the* ***MaxCut*** *SDP for the* ***G67*** *dataset ($n = 10\,000$) with* ***SketchyCGAL****. The subplots show the magnitudes of the* DIMACS *errors and the surrogate gap of the implicit primal iterate and the dual iterate. We omit* err_2 *and* err_3 *since they are zero by construction. See subsection* SM5.3*.*