Chordal Decomposition in Rank Minimized SDPs

Jared Miller[1], Yang Zheng[2], Biel Roig-Solvas[1], Mario Sznaier[1], Antonis Papachristodoulou[3]

[1]Northeastern University [2]Harvard University [3]University of Oxford

Motivation

Many problems in machine learning can be equivalently expressed as rank-constrained semidefinite programs (SDPs):

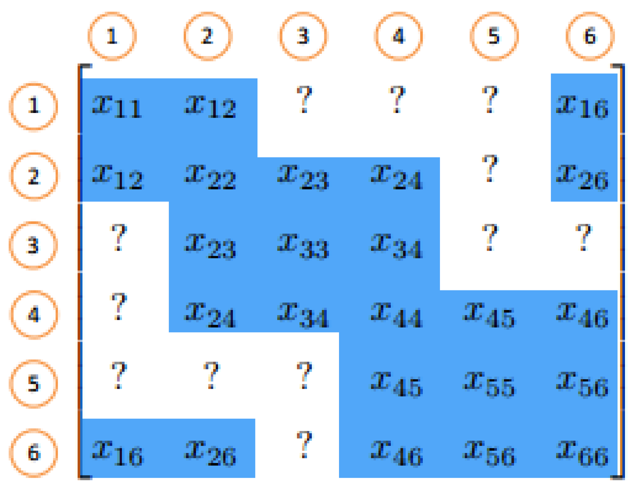
$$\begin{aligned} X^* = \operatorname*{argmin}_X \quad & \langle C, X\rangle \\ \text{subject to} \quad & \langle A_i, X\rangle = b_i \quad i = 1..m \\ & X \succeq 0, \quad \operatorname{rank}(X) \leq t, \end{aligned} \qquad (1)$$

Problem (1) is NP-hard, but the Reweighted Trace Heuristic is a convex relaxation. Rank-approximated SDPs exhibit poor scaling properties as the size of X grows.

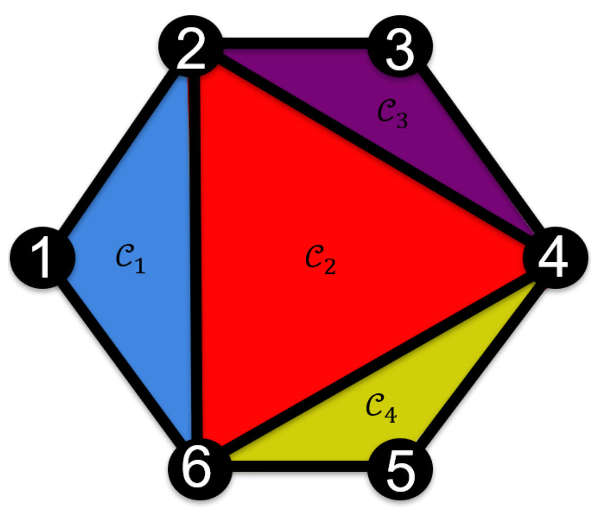
Exploiting chordal sparsity in ML- motivated SDPs often leads to algorithms that scale linearly with the number of data points.

Chordal Graphs and Semidefinite Optimization

In sparse SDPs, only a few entries of X appear in the cost function and equality constraints. All other entries are "free" to choose in order to force $X \succeq 0$. If the structure is chordal, we can take advantage of this to reduce computational complexity.



(a) Variables that appear in (C, A_i) vs. 'free' entries (?).



(b) Associated chordal sparsity pattern $\mathcal{G}(\mathcal{V}, \mathcal{E})$

$\mathbb{S}^n_+(\mathcal{E}, ?)$ is the set of PSD-completable matrices with sparsity $\mathcal{G}$. If $\mathcal{G}(\mathcal{V}, \mathcal{E})$ is chordal (all 4+ length cycles have shortcuts), Grone's theorem gives neccesary and sufficient conditions $X \in \mathbb{S}^n_+(\mathcal{E}, ?)$.

Grone's theorem

Let $\mathcal{G}(\mathcal{V}, \mathcal{E})$ be a chordal graph with a set of maximal cliques $\{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_p\}$. Then, $X \in \mathbb{S}^n_+(\mathcal{E}, ?)$ if and only if

$$X_k = E_{\mathcal{C}_k} X E_{\mathcal{C}_k}^T \in \mathbb{S}_+^{|\mathcal{C}_k|}, \qquad k = 1, \ldots, p$$

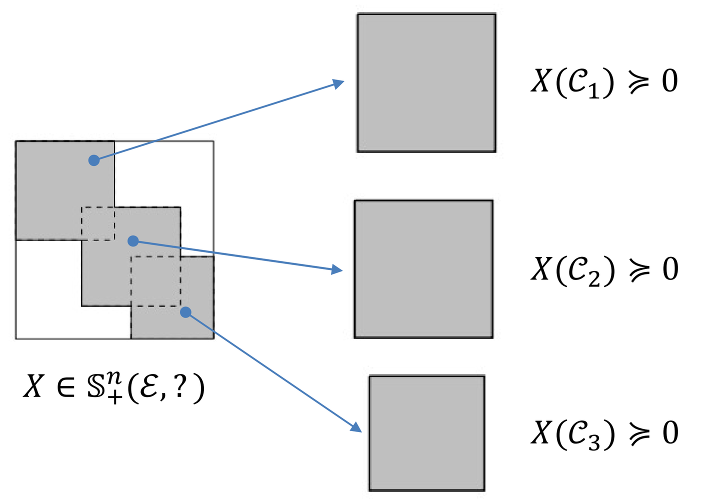


Figure 2: All clique submatrices should be PSD

Minimum Rank Completion

For any $X \in \mathbb{S}^n_+(\mathcal{E}, ?)$, there exists a unique max. determinant completion, and at least one min. rank PSD completion where:

$$\operatorname{rank}(X) = \max_k \operatorname{rank}(E_{\mathcal{C}_k} X E_{\mathcal{C}_k}^T)$$

Minimizing $\operatorname{rank}(X)$ of a large matrix is equivalent to minimizing the maximum $\operatorname{rank}(X_k)$ with smaller matrices.

Chordal Decomposition of Rank-Minimized SDP

The reweighted heuristic is $\operatorname{rank}(X) \approx \langle W, X\rangle$, where $W = (X^* + \delta I)^{-1}$. Penalize $\langle W, X\rangle$, solve for X^*, then update W for the next iteration. The chordalized rank-minimized SDP is:

$$\begin{aligned} \min_X \quad & \langle C + W_{\mathcal{C}}, X\rangle \\ \text{subject to} \quad & \langle A_i, X\rangle = b_i \qquad i = 1..m \\ & X_k = E_{\mathcal{C}_k} X E_{\mathcal{C}_k}^T \succeq 0 \quad k = 1..p \end{aligned} \qquad (2)$$

where $W_{\mathcal{C}} = \sum_{k=1}^p E_{\mathcal{C}_k}^T W_k E_{\mathcal{C}_k}$ is the clique weights, and $W_k = (X_k^* + \delta I)^{-1}$. The new cost $C + W_{\mathcal{C}}$ retains the pattern $\mathcal{E}_{\mathcal{C}}$.

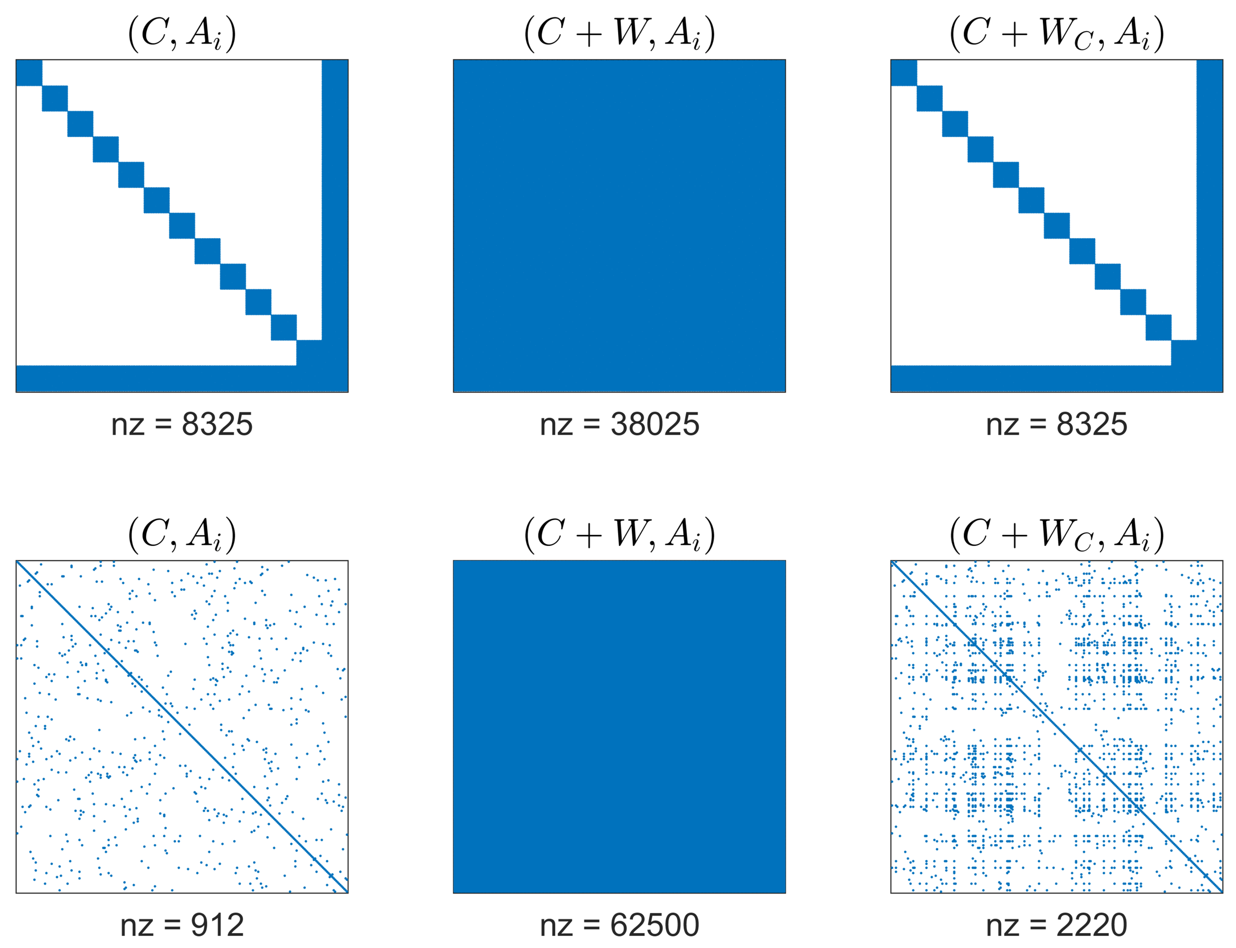


Figure 3: Left column is the pattern of an SDP, solving gives X^*. Center is reweighting by X^*, and right is reweighting by X_k^*. Block-arrow (top) is already chordal, but MCP (bottom) requires a chordal extension.

Experiments

Problem (2) is convex for each reweighting iteration. Tests were run on Subspace Clustering and Maxcut SDPs by algorithms such as interior point methods and ADMM.

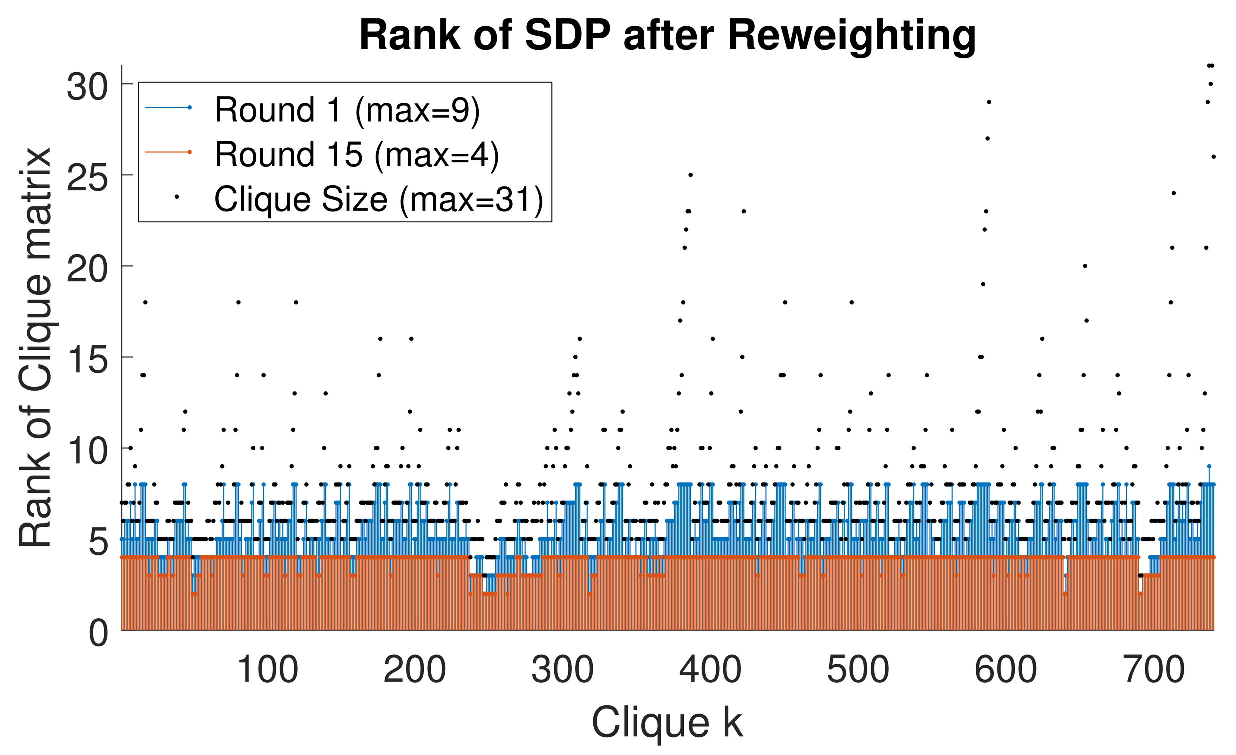
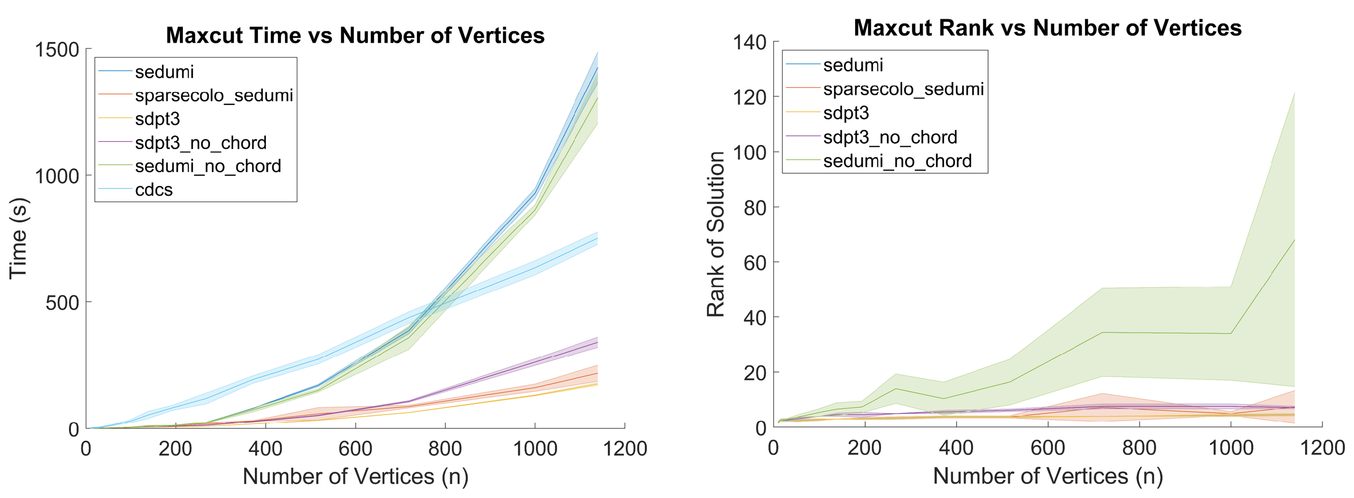


Figure 4: Chordal Rank SDP on an 1000-vertex Maxcut problem. Maximum clique rank starts at 9 (blue), and drops to 4 (orange) after 15 rounds.

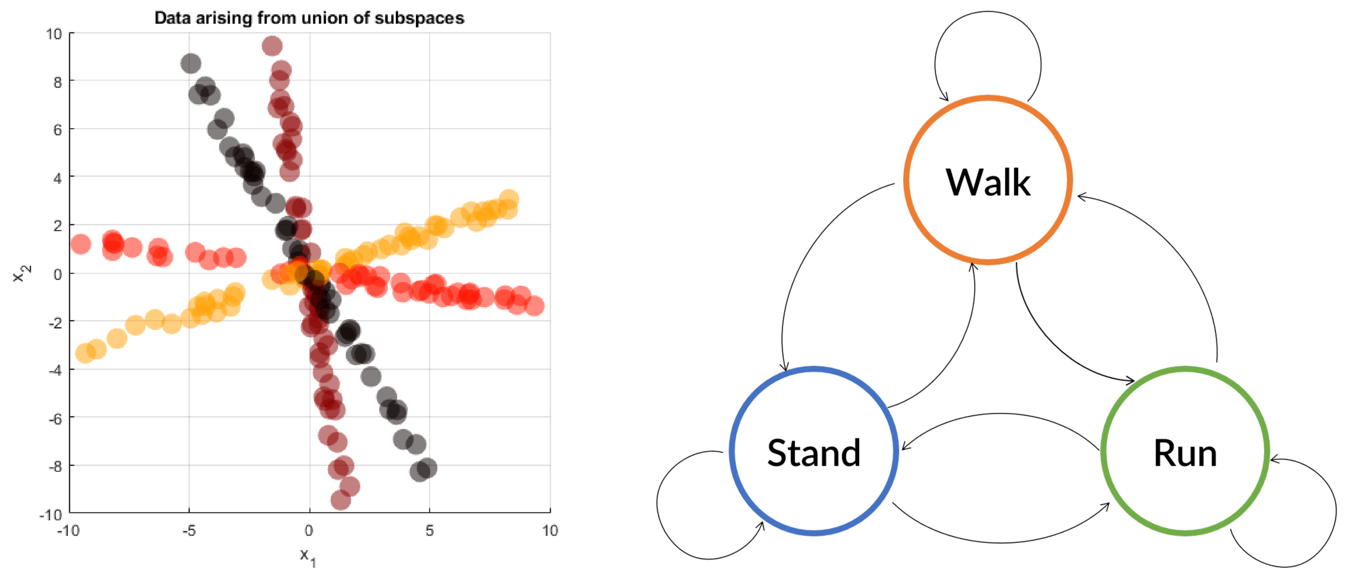


(a) Time vs. $|\mathcal{V}|$ (b) Rank vs. $|\mathcal{V}|$

Subspace Clustering (Switched Sysid)

Given N_p points $x_j \in \mathbb{R}^D$ and a N_s subspaces with normals $r_i \in \mathbb{R}^D$, subspace clustering aims to determine if point x_j came from subspace r_i (binary labels s_{ij}). This occurs if $r_i^T x_j = 0$, relaxed to $|r_i^T x_j| \leq \epsilon$ under bounded noise. Switched System Identification is an application of subspace clustering (SARX model).

These algorithms allow subspace clustering to scale linearly with number of points and subspaces. Chordal sparsity patterns are preserved throughout.



(a) Points from subspaces (b) Example of a switched system

Finding (r_i, s_{ij}) is a nonconvex quadratic feasibility problem:

$$\begin{aligned} \operatorname*{find}_{r,s} \quad & s_{ij} |r_i^T x_j| \leq s_{ij}\epsilon & s_{ij} = s_{ij}^2 \\ & \sum_{i=1}^{N_s} s_{ij} = 1 & r_i^T r_i = 1 \end{aligned} \qquad (3)$$

Given $X = [1, r_i, s_{ij}][1, r_i, s_{ij}]^T$, this is a rank-1 SDP in X. **We improve Cheng *et. al.* [2016]'s chordal sparsity (grey) by using a reduced chordal extension (red) with smaller cliques.**

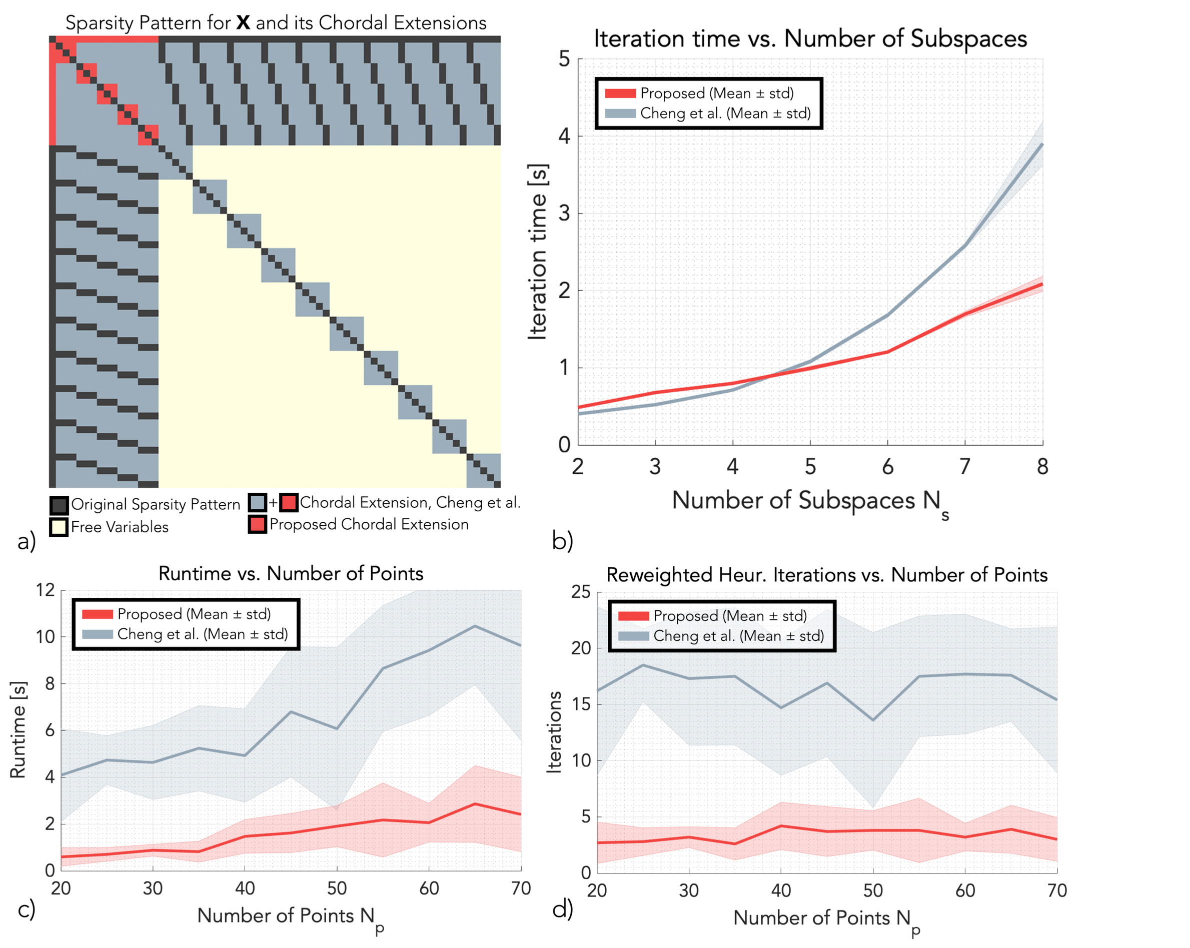


Figure 7: a) X structure and chordal extensions. b, c, d) Runtime analysis.

A summary of the clique matrix sizes $|\mathcal{C}_k|$ are:

Problem	Rank 1 PSD # Cliques	Rank 1 PSD Size Cliques	Other PSD # Cliques	Other PSD Size Cliques	% edges
Full X	1	$1 + N_s(D + N_p)$	$\varnothing$	$\varnothing$	1637%
Cheng	1	$1 + N_s D$	N_p	$1 + N_s(D+1)$	350%
Ours	N_s	$D+1$	$N_p N_s$	$D+2$	13%

Table 1: Sizes of cliques in subspace clustering. %edges measures size of chordal extension over baseline $\mathcal{E}$ (variables in (3)) ($D = 3, N_p = 10, N_s = 5$).

Full paper available at: `https://arxiv.org/abs/1904.10041`