Experiments with Odd Cycle Inequalities in Large Scale Semidefinite Relaxations

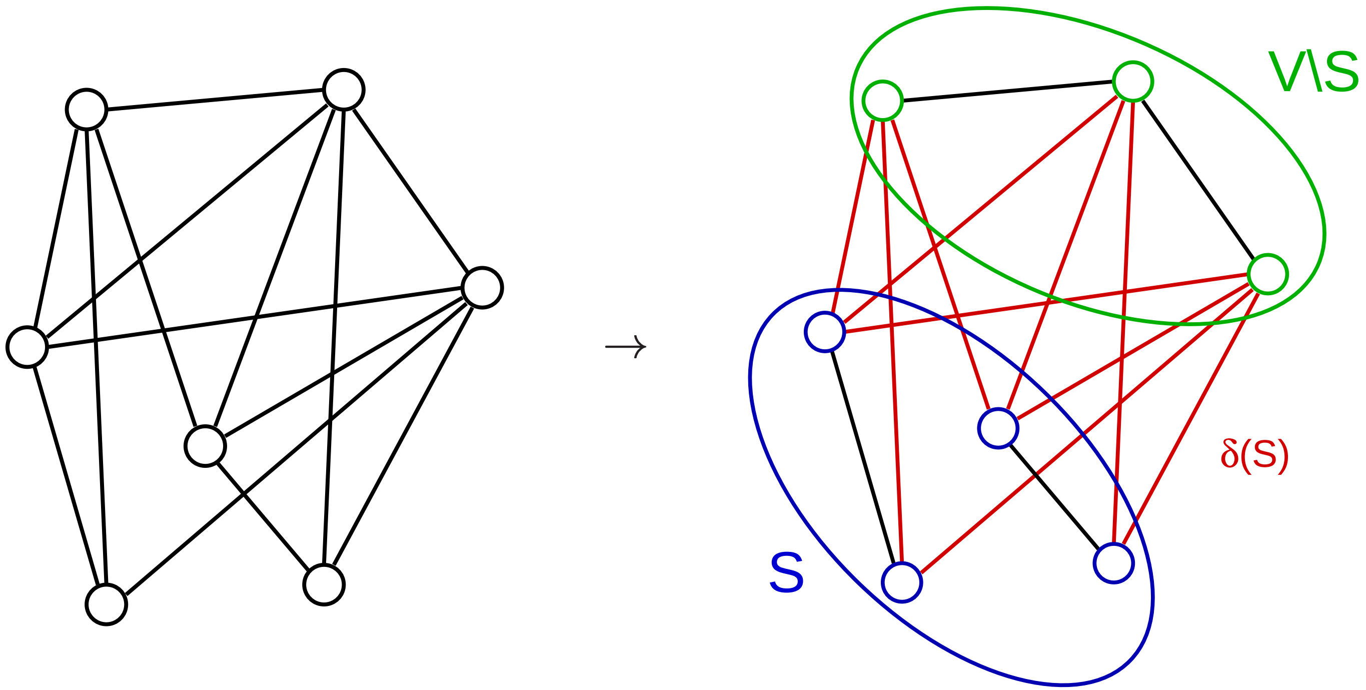
Christoph Helmberg
ZIB-Berlin

- Max-Cut and its Semidefinite Relaxation
- The Cut Polytope and Cutting Planes
- The Spectral Bundle Method
- The Spectral Bundle Method with Cutting Planes
- Toroidal Grid Graphs
- Min Bisection
- Extending the support

thanks to: F. Rendl, M. Jünger, K.C. Kiwiel, F. Liers, S. Filipowski

Max-Cut

given: graph $G = (V, E)$,
$V = \{1, \ldots, n\}$, $E \subset \{ij : i, j \in V, i < j\}$,
edge weights a_{ij} for $ij \in E$



Find partition $(S, S \setminus V)$ that maximizes the sum of the weight of edges running between both sets.

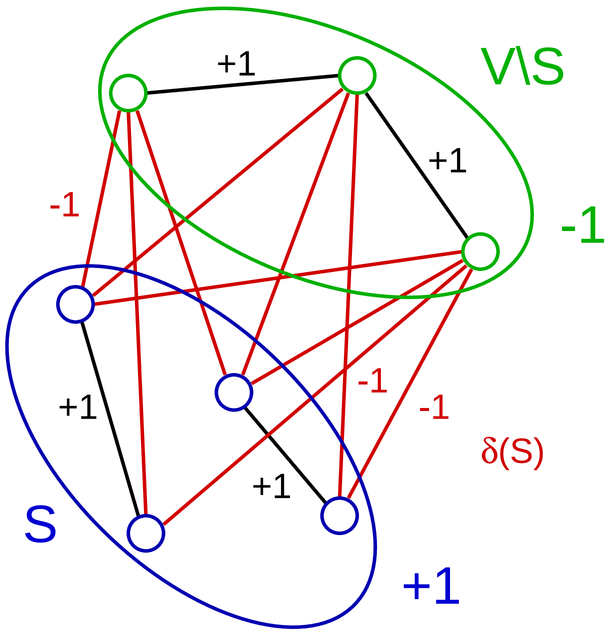
$$\text{(MC)} \qquad \max_{S \subseteq V} \sum_{ij \in \delta(S)} a_{ij}$$

Håstad 1997: cannot be approximated within 0.9412 if $P \neq NP$
Example: toroidal grid graphs are used as a model for Ising spin glasses.

Max-Cut as a Quadratic $\{-1,1\}$ Programming Problem

Represent partition by

$$x \in \{-1,1\}^n \quad \text{with} \quad x_i = \begin{cases} 1 & i \in S \\ -1 & i \in V \setminus S \end{cases}$$



$$x_i x_j = \begin{cases} -1 & ij \in \delta(S) \\ 1 & \text{otherwise} \end{cases}$$

$$\frac{1 - x_i x_j}{2} = \begin{cases} 1 & ij \in \delta(S) \\ 0 & \text{otherwise} \end{cases}$$

$$\max_{S \subseteq V} \sum_{ij \in \delta(S)} a_{ij} = \max_{x \in \{-1,1\}^n} \sum_{ij \in E} a_{ij} \frac{1 - x_i x_j}{2} \qquad \rightarrow \qquad \max_{x \in \{-1,1\}^n} x^T C x$$

for an appropriate symmetric matrix $C \in S_n$.

Semidefinite Relaxation of $\max_{x\in\{-1,1\}^n} x^T C x$

$x^T Cx = \langle Cx, x\rangle = \left\langle C, xx^T \right\rangle$

Properties of $xx^T = [x_i x_j]$ für $x \in \{-1, 1\}^n$:

- $x_i^2 = 1 \quad \Rightarrow \quad \mathrm{diag}(xx^T) = e$ $\qquad e = [1, \ldots, 1]^T$
- xx^T is positive semidefinite, $xx^T \succeq 0$
- rank(xx^T)= 1

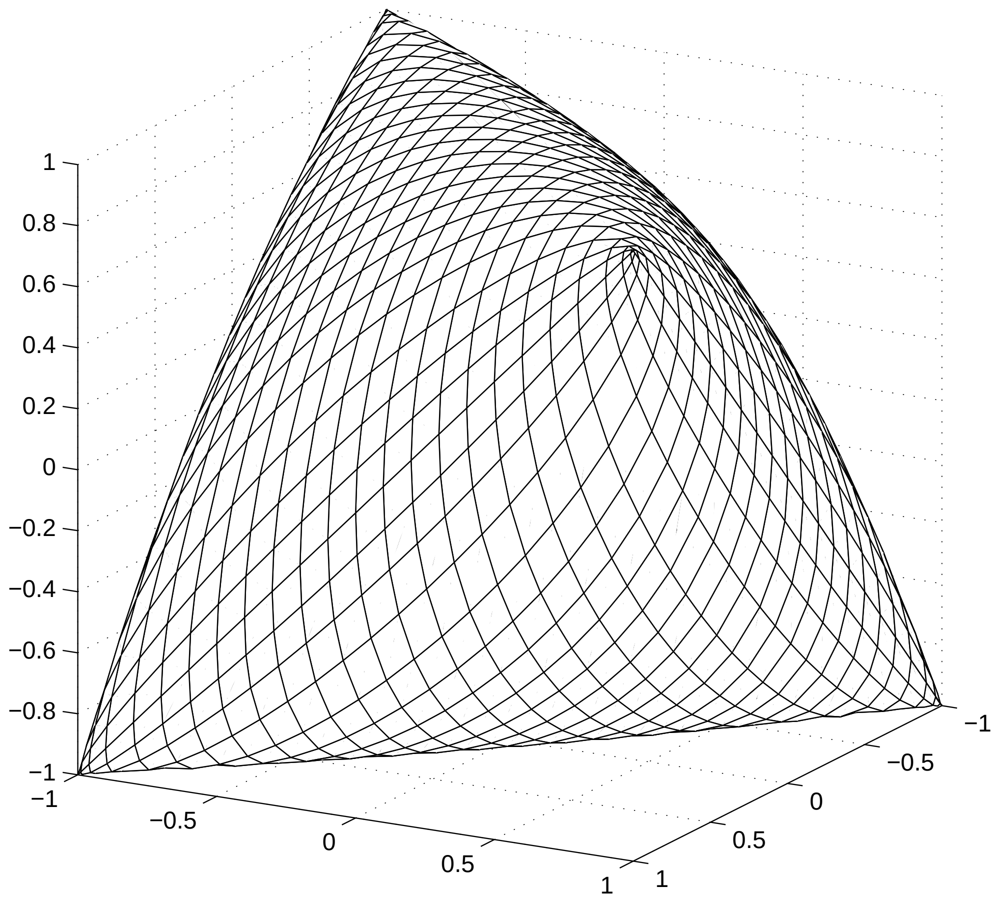
Relax xx^T to a positive semidefinite matrix X

$$\max_{x\in\{-1,1\}^n} x^T Cx \leq \boxed{\begin{array}{rl} \max & \langle C, X\rangle \\ \text{s.t.} & \mathrm{diag}(X) = e \\ & X \succeq 0 \\ & [\mathrm{rank}(X) = 1] \end{array}}$$

Shor 1987; Schrijver; Delorme and Poljak 1993
Goemans and Williamson 1995 (0.878 approximation algorithm for (MC) with $a_{ij} \geq 0$)

The semidefinite feasible set for $n = 3$

The boundary is given by $\quad \det \begin{pmatrix} 1 & x & y \\ x & 1 & z \\ y & z & 1 \end{pmatrix} = 0.$



The Cut Polytope

The semidefinite set is a relaxation of

$$\boxed{P_C = \text{conv}\left\{xx^T : \; x \in \{-1,1\}^n\right\}}$$

"matrix version" of the cut polytope

Hypermetric Inequalities

For $b \in \mathbb{Z}^n$ with $\min\{|b^T x| : x \in \{-1,1\}^n\} = 1$

$$b^T x x^T b \geq 1 \qquad \text{for all } x \in \{-1,1\}^n$$

Therefore

$$\boxed{b^T X b \geq 1}$$

is valid for P_C. $[X \in S_n^+ \Rightarrow b^T X b \geq 0]$

Triangle Inequalities

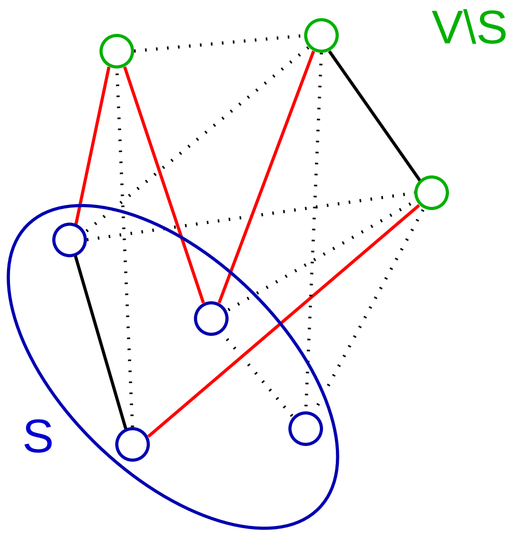
special case: b has exactly three nonzero elements $\in \{-1,1\}$

$$\begin{array}{ccccccc} x_{ij} & + & x_{ik} & + & x_{jk} & \geq & -1 \\ x_{ij} & - & x_{ik} & - & x_{jk} & \geq & -1 \end{array}$$

- define facets of P_C
- form the metric polytope, exact for graphs not contractible to K_5 (Seymour [1981], Barahona [1983])

Odd Cycle Inequalities

A cut has to intersect a cycle in an even number of edges.



$C \subset E$ cycle, $F \subset C$ odd

then

$$\boxed{x(C \setminus F) - x(F) \leq |C| - 2}$$

$$[x(E) := \sum_{ij \in E} x_{ij}]$$

- separable in polynomial time — Barahona and Mahjoub 1986
- triangle inequalities are a special case
- on complete graph triangle inequalities imply odd cycle inequalities

---

Adding edges to E with zero weight (enlarging the support of the graph) in order to increase the number of cycles

- not useful in linear relaxations — Barahona 1993
- may improve semidefinite relaxations

Combining Semidefinite and Linear Relaxations

Solve

$$\begin{array}{rll} \max & \langle C, X \rangle & \\ \text{s.t.} & \operatorname{diag}(X) = e & \\ & \langle A_i, X \rangle \geq b_i & i \in J_{\leq} \\ & X \succeq 0, & \end{array}$$

Number of constraints m grows quickly with number of nodes n

Interior Point Methods

branch & cut approach H. and Rendl [1998]

- solves max-cut for complete graphs with up to 100 nodes routinely
- interior point methods too slow for much larger sizes, say $n > 500$, because each iteration requires $O(m^3)$ operations

---

Large scale semidefinite programming?

Spectral Bundle Method H. and Rendl [2000], H. and Kiwiel

One can show

$$\begin{array}{rll} \text{max} & \langle C, X\rangle & \\ \text{s.t.} & \langle I, X\rangle = a & \\ & \langle A_i, X\rangle = b_i & i \in J_= \\ & \langle A_i, X\rangle \geq b_i & i \in J_\leq \\ & X \succeq 0, & \end{array} \quad \leq \quad \begin{array}{ll} \text{min} & a\lambda_{\max}(C - \sum_i A_i y_i) + \langle b, y\rangle \\ & y \in \mathbb{R}^m, \;\; y_i \leq 0 \;\text{ for }\; i \in J_\leq \end{array}$$

- The max-cut relaxation is of this form with $a = n$
- $\boxed{\text{Any feasible } y \text{ yields a valid upper bound for max-cut}}$
- The optimal y yields the optimal value of the semidefinite relaxation
- The matrix $C - \sum_i A_i y_i$ inherits structure of cost matrix and constraints

---

Solve the eigenvalue problem

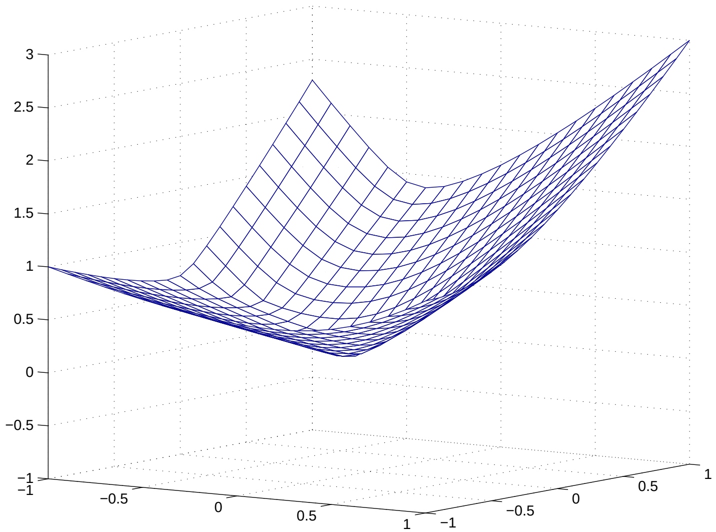
$\longrightarrow$ minimize a nonsmooth convex function

$\longrightarrow$ subgradient and bundle methods

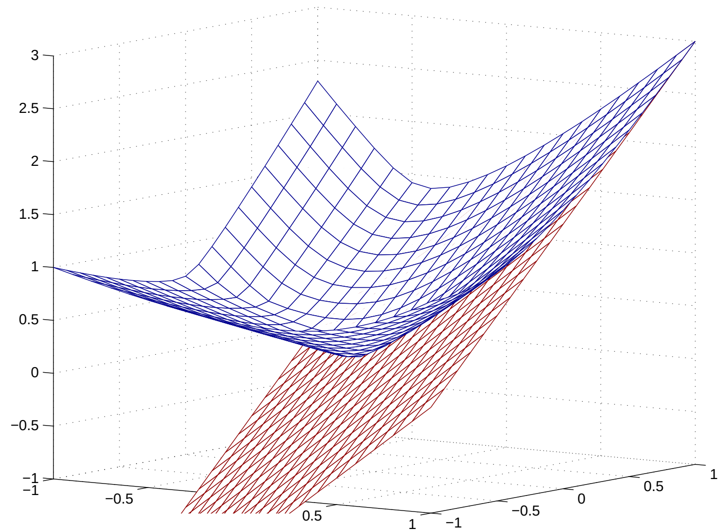
Proximal Bundle Method

Kiwiel 1990

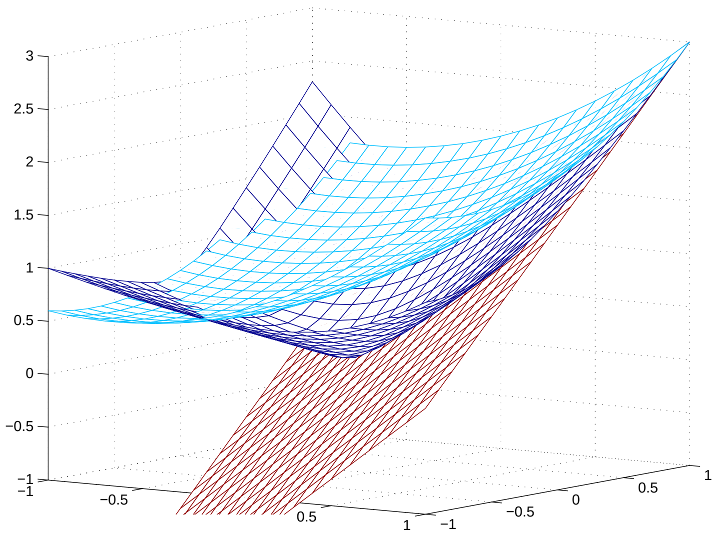
convex function



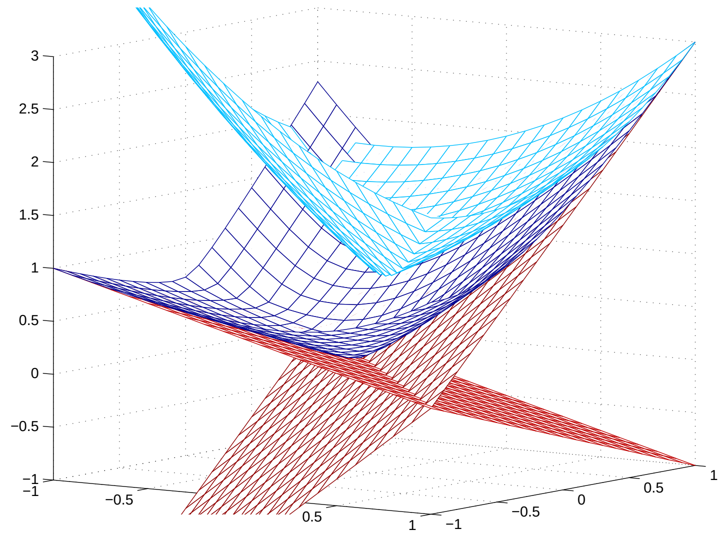
cutting plane model with $g \in \partial f(\hat{x})$



solve augmented model $\to \xi$



improve cutting plane model in ξ



Spectral Bundle Method

determines subgradient via an eigenvector v to $\lambda_{\max}(C - \sum A_i y_i)$

$\longrightarrow$ Lanczos Method allows to exploit structure of the matrix

---

Here, a subgradient at y is a vector $g \in \mathbb{R}^m$ of the form

$$\begin{aligned} g_i &= b_i - \langle A_i, X \rangle & i \in J_= \\ g_i &= b_i + \eta_i - \langle A_i, X \rangle & i \in J_\leq \end{aligned}$$

with

$$\langle I, X \rangle = n, \quad X \succeq 0$$

$$\left\langle C - \sum A_i y_i, X \right\rangle = n\,\lambda_{\max}(C - \sum A_i y_i)$$

$$\eta_i \geq 0 \text{ and } \eta_i = 0 \text{ if } y_i < 0$$

The new subgradient is computed from $X = n\,vv^T$ $[\|v\| = 1]$

In the optimal solution y_*, there is a subgradient g_* of norm zero,

$$\boxed{\|g_*\| = 0 \quad \Longleftrightarrow \quad X_* \text{ is primal feasible and optimal}}$$

In general, $\text{rank}(X_*) > 1$ and X_* cannot be found by the Lanczos method

The Semidefinite Model

The spectral bundle method solves in each iteration the augmented model

$$
\boxed{
\text{(QSP)} \qquad
\begin{array}{ll}
\min & \frac{1}{2u}\|b+\eta-\mathcal{A}X\|^2 - \left\langle X, C-\mathcal{A}^T\widehat{y}\right\rangle - \langle b+\eta, \widehat{y}\rangle \\
\text{s.t.} & X = PVP^T + \alpha\overline{W} \\
 & \operatorname{tr} V + \alpha = n \\
 & V \succeq 0, \alpha \geq 0.
\end{array}
}
$$

- P is an orthonormal matrix, a minimal choice is $P = v$
- $\overline{W}$ is a positive semidefinite matrix of trace 1
  e.g. the last optimal solution of QSP, $\overline{W} = \overline{X}/n$
- X satisfies $X \succeq 0$ and $\langle I, X\rangle = n$
- The new optimal $\overline{X}^+$ of (QSP) determines the next candidate y^+

One can prove

Theorem 1 *H. [2000]*
Let $\{\overline{X}^k\}$ denote the subsequence of optimal solutions of (QSP) that give rise to descent steps to $\widehat{y}^k$.

Then all accumulation points of $\{\overline{X}^k\}$ are primal optimal solutions.

Combining the spectral bundle method with cutting planes

Idea: stop after any descent step and separate with respect to $\overline{X}$

Adding primal constraints = dual variables is no problem, because

- old subgradients can be extended by means of $\overline{W}$
  [$\overline{W}$ is needed only on the support of C and the A_i, sparse!]

$\longrightarrow$ the cutting surface model remains valid, easy to restart

---

Separation is difficult

- $\overline{X}$ is never feasible
- even trivial inequalities may be violated
- the same inequality may be separated again and again

Experiments with 3D toroidal grid graphs

- make $\overline{X}$ feasible for semidefinite relaxation to ensure $-1 \leq x_{ij} \leq 1$:
  - $\overline{X}$ is already positive semidefinite
  - use $X_1 = \overline{X}/\max\{\overline{X}_{ii}\}$, then $\operatorname{diag}(X_1) \leq e$ and $X_1 \succeq 0$
  - or scale by $s = \sqrt{\operatorname{diag}(\overline{X})^{-1}}$, $X_2 = \operatorname{Diag}(s)\overline{X}\operatorname{Diag}(s)$, then $\operatorname{diag}(X_2) = e$ and $X_2 \succeq 0$

We separate odd cycle inequalities for $\overline{X}$, X_1, and X_2 by

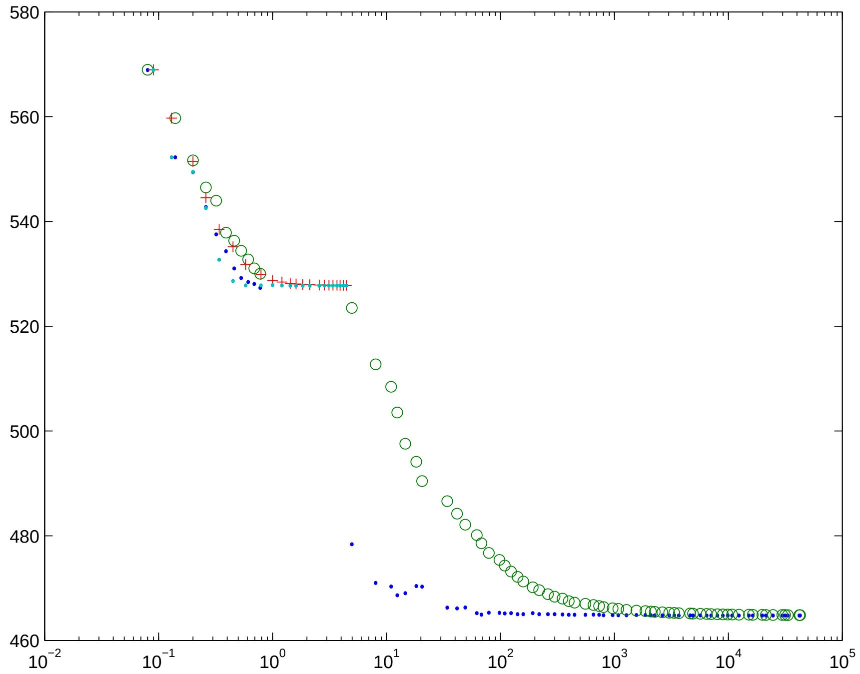
- heuristic 1: run through all small cycles of the grid [successful]
- heuristic 2: maximum weight forest [no success]
- exact separation: shortest path with $3|E|$ edges [worth it]

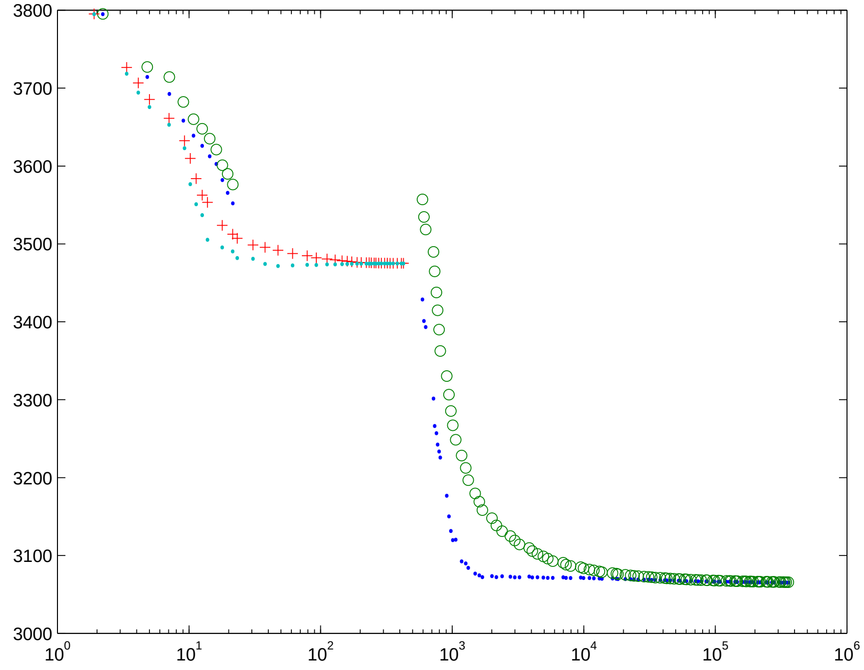
Separation quite successful for all choices of X, we use all three.

Results for the DIMACS set: edge weights in $\{-1, 1\}$

torusp m3-8-50	
$n = 512,\ m = 2300$	
SDP-UB:	527.81
S+C-UB:	464.85
LP-UB:	464.70
LB:	456.00

toruspm3-15-50	
$n = 3375,\ m = 18000$	
SDP-UB:	3475.16
S+C-UB:	3065.70
LP-UB:	u 3066.25
LB:	2944.00





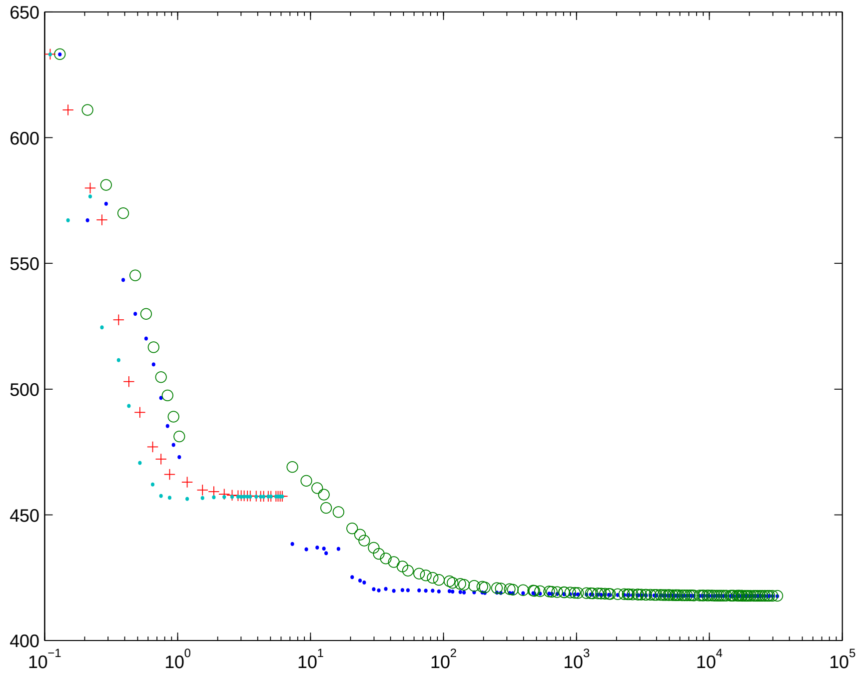
time in logarithmic scale!

Results for the DIMACS set: Gaussian distribution

torusg3-8

$n = 512$, $m = 2500$

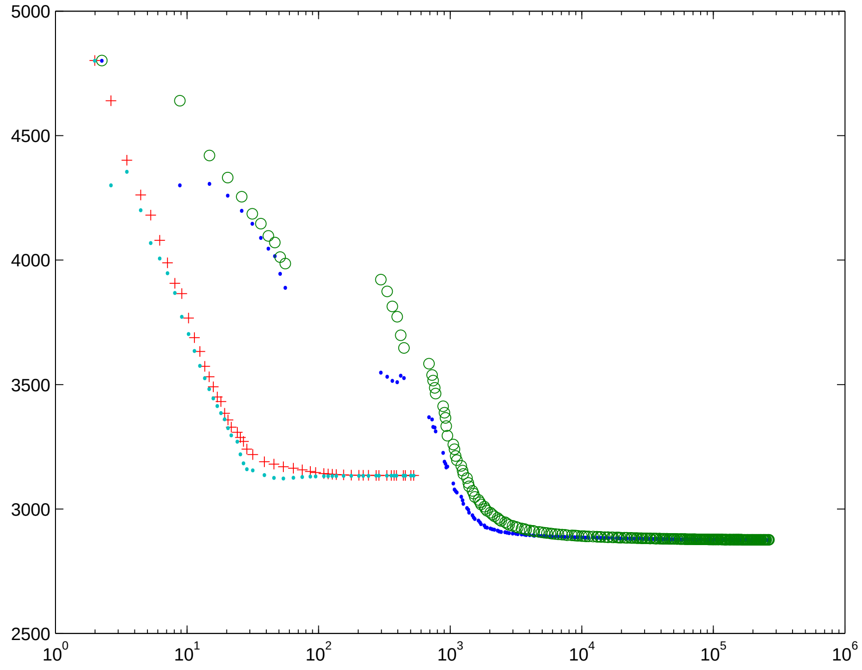
SDP-UB:	457.36
S+C-UB:	417.79
LP-UB:	417.66
LB:	*416.81



torusg3-15

$n = 3375$, $m = 19000$

SDP-UB:	3134.59
S+C-UB:	2875.98
LP-UB:	u 2875.05
LB:	2841.96



time in logarithmic scale!

Min Bisection:

Find partition $(S, S\backslash V)$ with $\Big||S|-|S\backslash V|\Big| \leq \sigma n$ that minimizes the sum of the weight of edges running between both sets.

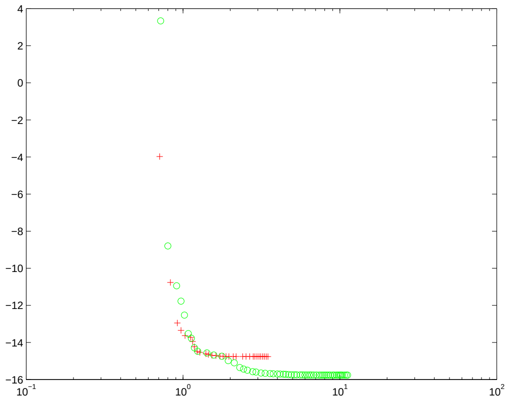
$$\text{(BS)} \qquad \min_{S\subseteq V,\ \big||S|-|S\backslash V|\big|\leq \sigma n} \quad \sum_{ij\in\delta(S)} a_{ij}$$

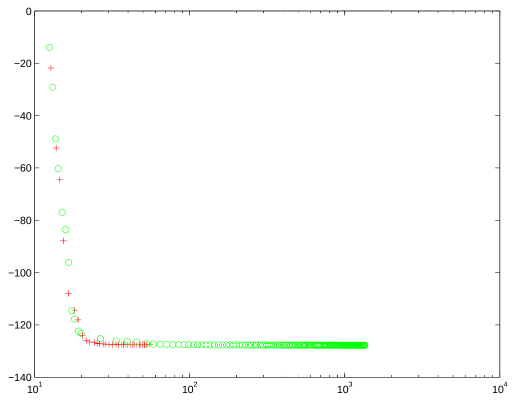
For an appropriate cost matrix C

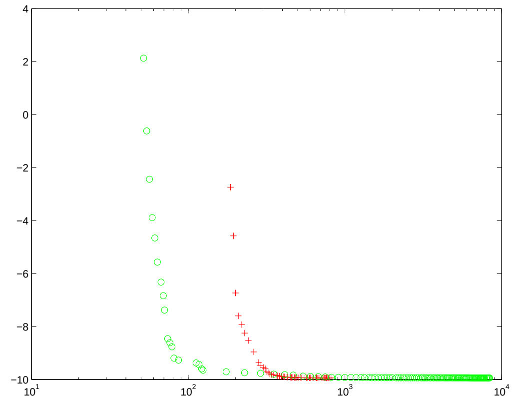
$$\max_{\substack{x \in \{-1,1\}^n \\ (e^T x)^2 \leq \lfloor \sigma n \rfloor^2}} x^T C x \leq \boxed{\begin{array}{rl} \max & \langle C, X\rangle \\ \text{s.t.} & \operatorname{diag}(X) = e \\ & \left\langle ee^T, X \right\rangle \leq \lfloor \sigma n \rfloor^2 \\ & X \succeq 0 \\ & [\operatorname{rank}(X) = 1] \end{array}}$$

Results for Min Bisection: KKT-Instances from Boeing $\sigma = 0.05$

putt01	shut01	heat02
$n = 258$, $nz = 802$	$n = 1704$, $nz = 5779$	$n = 5200$, $nz = 25188$
SDP-UB: -14.76	SDP-UB: -127.61	SDP-UB: -9.93
S+C-UB: -15.76	S+C-UB: -127.81	S+C-UB: -9.94
LB: * -26.00	LB: -154.00	LB: -150.00







time in logarithmic scale!

Enlarging the support

Separating odd cycles on the dense graph helps a lot:

putt01

$n = 258$, $nz = 802$	
SDP-UB:	-14.76
S+C-UB:	-15.76
Dense:	u -25.97
LB:	* -26.00

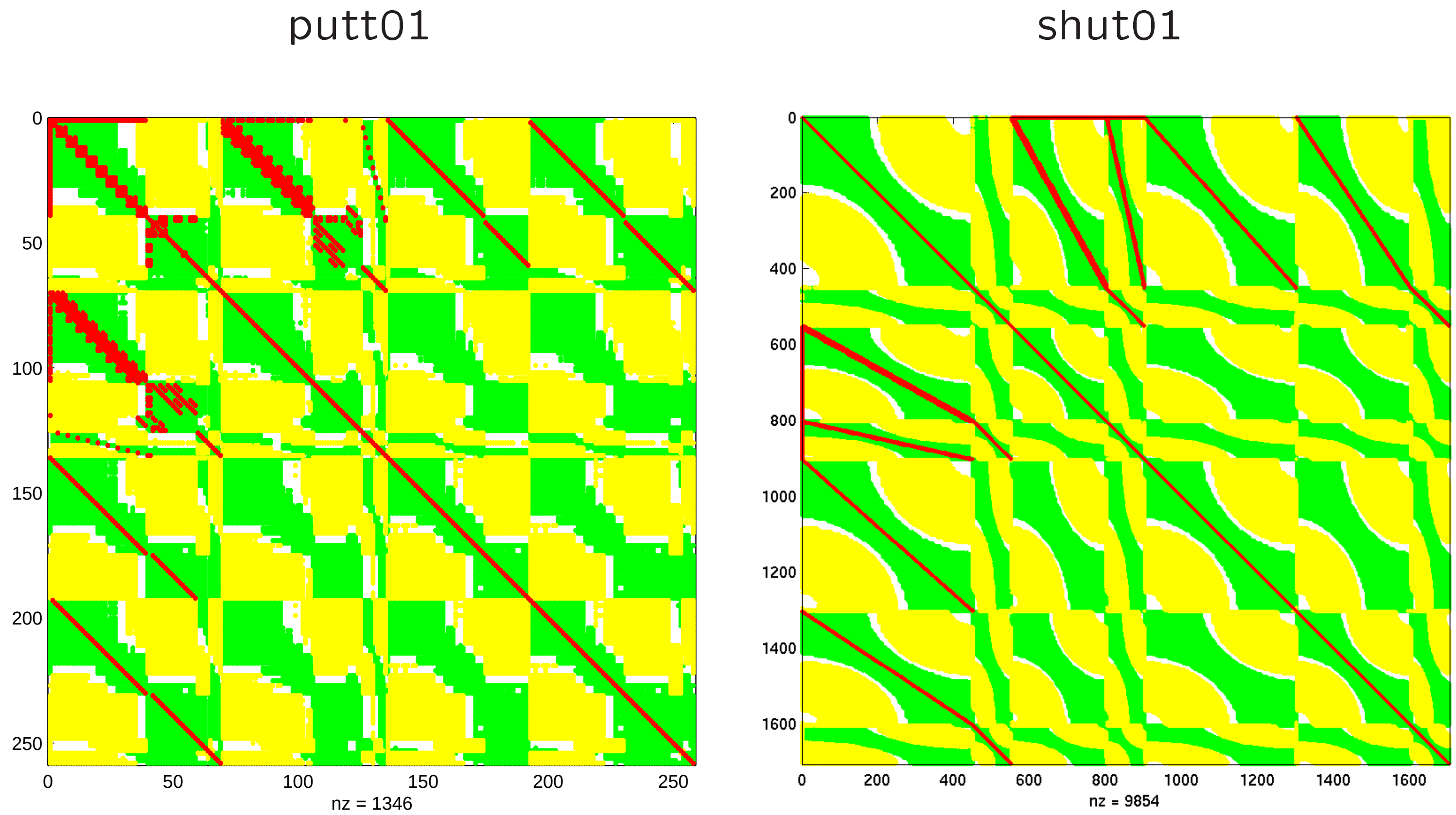
Enlarging the support

Updating a dense $\overline{W}$ (the complete graph) is too expensive for large n.

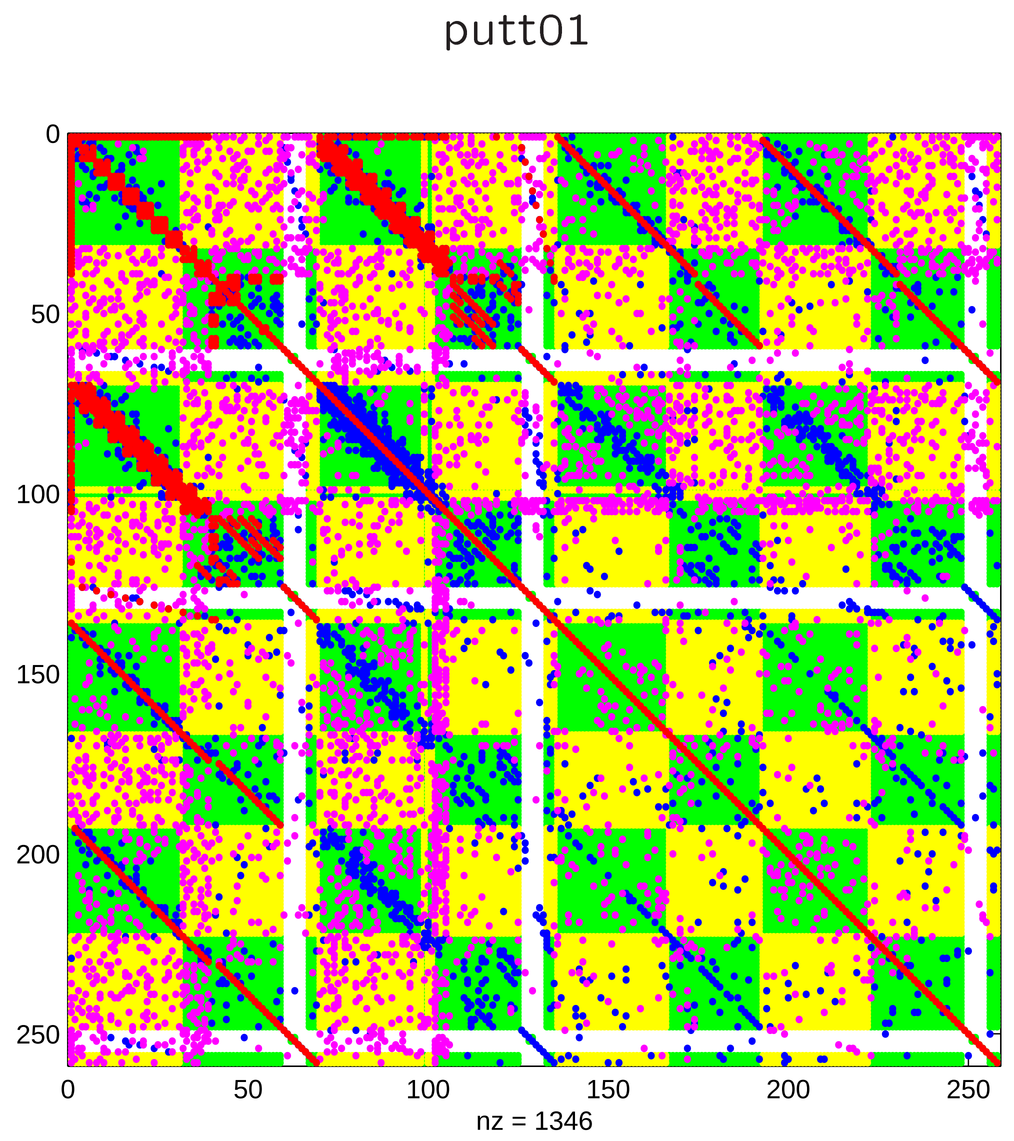
Is it possible to obtain good improvement by adding just a few edges?

How can one locate "important" edges?

Structure of X



Structure of Z



Cheap heuristic

odd cycle inequality: $C \subset E$ cycle, $F \subset C$ odd, then

$$\boxed{2 \leq |C| - x(C \setminus F) + x(F)} \qquad [x(E) := \sum_{ij \in E} x_{ij}]$$

Update $\overline{W}$ on support E

For each node v add two edges:

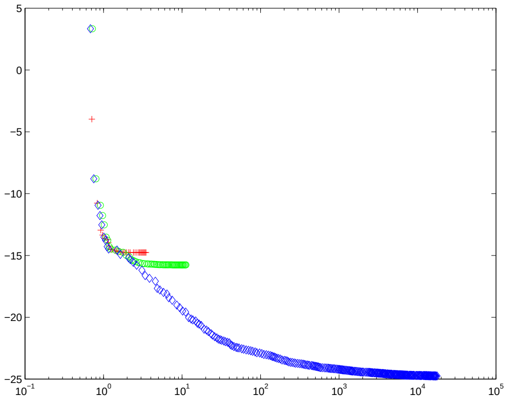
- Compute shortest path tree in E rooted at v for weights $1 - |x_{ij}|$
- For $vw \notin E$ use bundle to approximate weight, $[PVP^T]_{vw} \to$ cycle
- Add the edge vw whose cycle seems most violated
- Furthermore, add the edge $v\bar{w}$ whose absolute weight is smallest

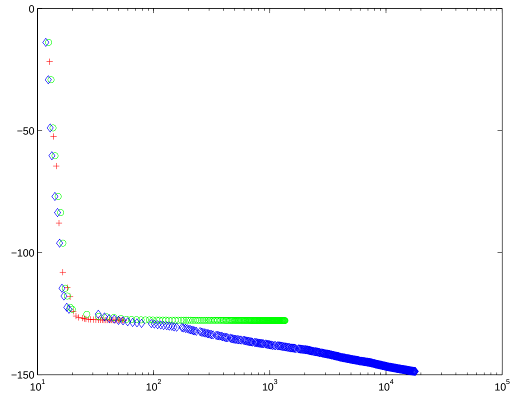
---

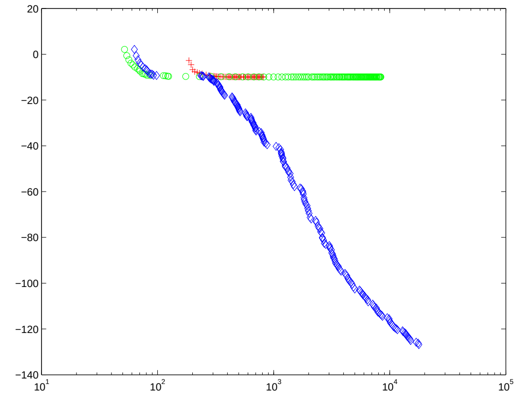
Each 10^{th} separation step remove unused edges and determine further new edges with respect to all edges currently in use.

Results for Min Bisection: KKT-Instances from Boeing $\sigma = 0.05$

	putt01	shut01	heat02
	$n = 258$, $nz = 802$	$n = 1704$, $nz = 5779$	$n = 5200$, $nz = 25188$
SDP-UB:	-14.76	-127.61	-9.93
S+C-UB:	-15.76	-127.81	-9.94
S+EC-UB:	u -24.75	u -148.63	u -126.83
	2529	12987	42005
LB:	*-26.00	-154.00	-150.00







time in logarithmic scale!

Remarks

Bad news:

- For (MC) on grid graphs, the SDP-relaxation with cutting planes seems "not better" than LP, extending support did not help.

Good news:

- The cutting plane approach for the spectral bundle method works
- Improvements with respect to simple bound may be considerable
- For some bisection problems increasing the support slightly improves the bound considerably (gap $< 5\%$).
- Support heuristic may work independent of cost matrix

Extensions:

- VLSI-Design: weighted nodes ("no problem"), quite dense
- Various other heuristics for support selection conceivable
  Are there theoretical justifications for certain choices?
  Why does it (currently?) not work for Max-Cut on toroidal grids?
- Approach applicable to constrained quadratic binary programming