Towards A Conic Bundle Package for Linear Programming over Symmetric Cones

Christoph Helmberg

University of Kaiserslautern
soon: Chemnitz University of Technology

- Lagrangian Relaxation
- Proximal Bundle Method and Wish List
- Convex functions and Cutting Models
- Primal Approximation in Lagrangian Relaxation
- Cutting Models for Linear Programs over Cones
- Preliminary Computational Results

Lagrangian Relaxation of Linear Constraints $Ax = b$

$$\begin{array}{ll} \max & c^T x \\ \text{s.t.} & Ax = b \\ & x \in \operatorname{conv}\Omega \end{array} \quad \Leftrightarrow \quad \max_{x \in \operatorname{conv}\Omega} c^T x + \inf_y (b - Ax)^T y$$

---

regularity assumption (Ω bounded, $\operatorname{conv}\Omega$ closed)

$$\min_y f(y) = b^T y + \max_{x \in \Omega} (c - A^T y)^T x$$

For fixed y the inner max should be easy to solve

Often $\Omega = \Omega_1 \times \cdots \times \Omega_k$, then $f = f_1 + \cdots + f_k$

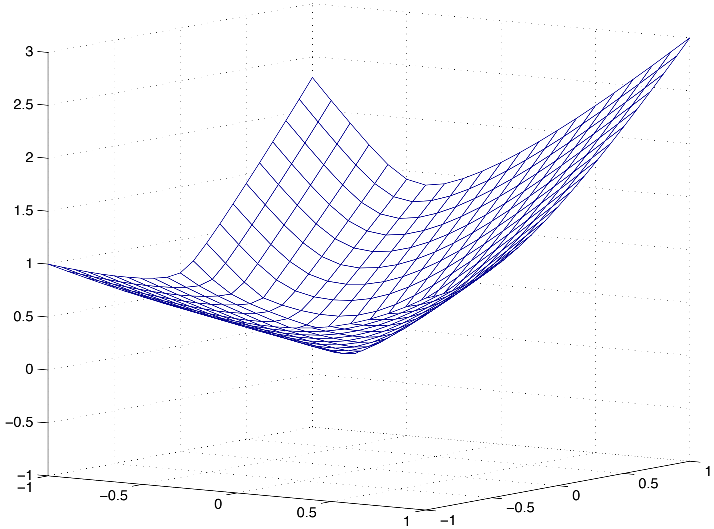
Max over linear functions $\Rightarrow$ f convex

$\Rightarrow$ convex optimization [Hiriart-Urruty Lemaréchal 1993]

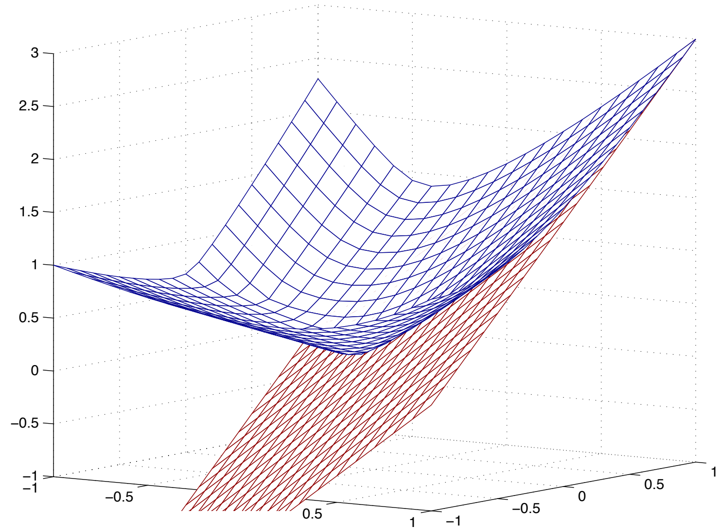
Proximal Bundle Method

Kiwiel 1990

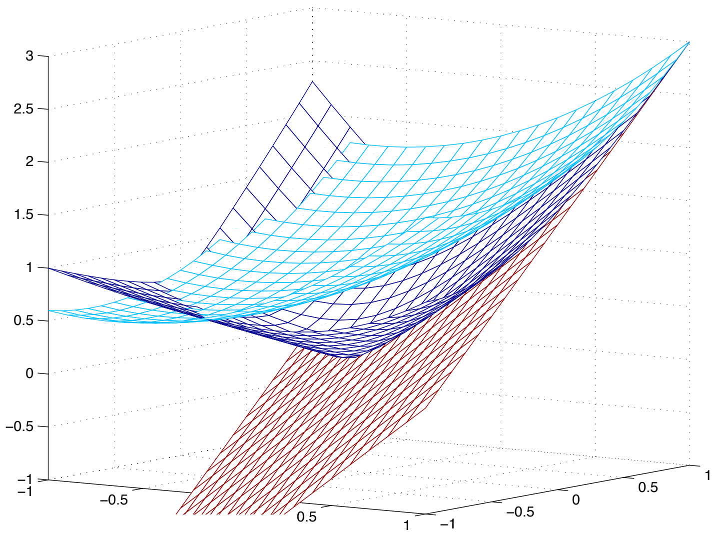
convex function



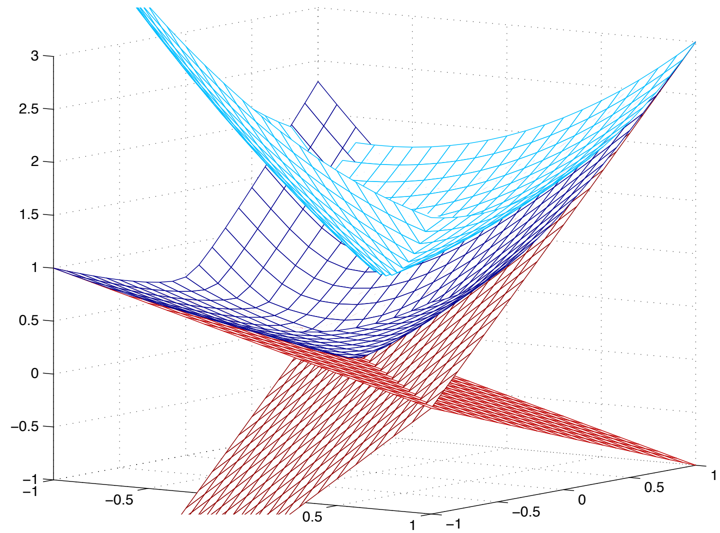
cutting plane model with $g \in \partial f(\hat{y})$



solve augmented model $\to y^+$



improve cutting plane model in y^+



Wish List for a Bundle Method

- general convex functions specified by first order oracle (standard)
- for a sum of convex functions allow the use of separate cutting models
- for Lagrangian relaxation: generate approximate primal solutions
- provide basic building blocks for Lagrangian relaxation
  - linear programs over symmetric cones
    (bounded feasible sets but "unbounded/free" variables,
    exploit block structure)
  - network flow, . . .
- use primal approximations for primal cutting plane algorithms

Convex Function

closed proper convex function $f : \mathbb{R}^n \to \mathbb{R}$
$\quad = \quad$ supremum over its linear minorants $\mathcal{M}$

$$f(y) = \sup_{i \in \mathcal{M}} \gamma_i + g_i^T y$$

Cutting Model

Choose a subset $\widehat{\mathcal{M}} \subset \mathcal{M}$

$$\sup_{i \in \widehat{\mathcal{M}}} \gamma_i + g_i^T y \qquad \le f(y)$$

Examples

Finite $\widehat{\mathcal{M}}$: $$\max_{\xi_i \ge 0,\ \sum \xi_i = 1} \sum \xi_i(\gamma_i + g_i^T y)$$

conic LP: $\begin{array}{l} \gamma_i = c^T x_i \\ g_i = b - A x_i \end{array}$ with $x_i \in \widehat{\Omega} \subset \operatorname{conv} \Omega$, convex, compact

$$\max_{x \in \widehat{\Omega}} c^T x + (b - Ax)^T y$$

Quadratic Subproblem

for finite $\widehat{\mathcal{M}}$

$$\min_y \quad \max_{i \in \widehat{\mathcal{M}}} \quad \gamma_i + g_i^T y \; + \frac{1}{2}\|y - \widehat{y}\|^2$$

equivalently

$$\begin{array}{ll} \max & \sum \xi_i(\gamma_i + g_i^T \widehat{y}) - \frac{1}{2}\|\sum \xi_i g_i\|^2 \\ \text{s.t.} & \xi^T e = 1 \\ & \xi \geq 0. \end{array}$$

Need only two: $(\bar{\gamma}, \bar{g}) = \sum \xi_i^*(\gamma_i, g_i)$ and the new (γ, g) of the oracle

---

Theorem 1 *If* Argmin $f \neq \emptyset$ *(and ++),* [e.g., FK 2000]
the proximal bundle method yields $\sum \xi_i g_i \to 0$ *and* $\sum \xi_i \gamma_i \to f_*$.

Primal Approximation in Lagrangian Relaxation

$$\begin{aligned} \gamma_i &= c^T x_i \\ g_i &= b - A x_i \end{aligned} \qquad \text{for } x_i \in \Omega \text{ (or conv}\,\Omega)$$

$$\begin{array}{lcll} \sum \xi_i g_i & = & b - A(\sum \xi_i x_i) & \to 0 \\ c^T(\sum \xi_i x_i) & \to f_* & & \end{array}$$

Accumulation points of $\sum \xi_i^k x_i^k$ (++) are optimal solutions (for conv Ω)

Quadratic Subproblem for convex compact $\widehat{\Omega}$

$$\begin{array}{ll} \max & c^T x + (b - Ax)^T \hat{y} - \frac{1}{2} \|b - Ax\|^2 \\ \text{s.t.} & x \in \widehat{\Omega} \end{array}$$

Need only two in the next $\widehat{\Omega}_+$:

- old subproblem solution $\bar{x} \in \widehat{\Omega}$
- and a new $x \in \Omega$ supplied by the oracle

Primal Approximation in Lagrangian Relaxation

Theorem $\Rightarrow$ for an appropriate subsequence

$$\begin{aligned} b - A\bar{x}^k &\to 0 \\ c^T \bar{x}^k &\to f_* \end{aligned}$$

Accumulation points of $\bar{x}^k$ (++) are optimal solutions (for $\operatorname{conv} \Omega$)

Sum of Convex Functions

sum of separate models

$$\begin{array}{lll} \max & \sum \xi_i(\gamma_i + g_i^T \widehat{y}) \quad + c^T x + (b - Ax)^T \widehat{y} \quad -\frac{1}{2} \left\| \sum \xi_i g_i + b - Ax \right\|^2 \\ \text{s.t.} & \xi^T e = 1 \\ & \xi \geq 0, \qquad\qquad x \in \widehat{\Omega}. \end{array}$$

- $\sum \xi_i(\gamma_i + g_i^T \widehat{y})$ ideal for
  - abstract oracles
  - if the primal set is a polytope
- what can we do with $\widehat{\Omega}$?

Pointed Closed Convex Cone K

Assume: $\exists\, d \in \operatorname{int} K$ so that

- $\mathcal{V} = \{v \in K : \; d^T v = 1\}$ is compact (extreme points are generators)
- $\max_{v\in\mathcal{V}} \; c^T v$ can be computed efficiently for all c

---

Suppose the following program has a bounded feasible set,

$$\begin{array}{ll} \max & c^T x \\ \text{s.t.} & Ax = b \\ & x \in \{\sum \xi_i v_i : \xi_i \geq 0, v_i \in \mathcal{V}\} \end{array}$$

Then for given y the oracle reads

$$\max_{v\in\mathcal{V}} \;\; (c - A^T y)^T v + b^T y$$

Examples: symmetric cones with d being the trace

- $\mathbb{R}^n_+$: $d = e$, $\mathcal{V}$ is the simplex, oracle returns e_i for largest component
- $\mathcal{S}^+_n$: $d = I$, oracle yields $\lambda_{\max}(C - \mathcal{A}^T y)$
- SOC: $d = (1, 0, \ldots, 0)^T$, oracle by “normalizing” $(c - A^T y)_{2\ldots n}$

Symmetric Cone K, bounded trace

$$\begin{array}{lll} \max & c^T x & \\ \text{s.t.} & Ax = b & \text{relax} \\ & \left.\begin{array}{l} d^T x \le \delta \\ x \in K \end{array}\right\} & \Omega \end{array}$$

Try to keep generating structure of K in model $\widehat{\Omega}$

$\mathbb{R}^n_+$:

$$\widehat{\Omega} = \ \{ x = \sum_{i \in \mathcal{I}} \xi_i e_i + \bar{\xi}\bar{x} \ : \ \sum_{i \in \mathcal{I}} \xi_i + \bar{\xi} \le \delta, \xi \ge 0 \}$$

with $\bar{x} \ge 0$, $e^T \bar{x} = 1$, $\mathcal{I} \subset \{1, \ldots, n\}$

$\mathcal{S}^+_n$:

$$\widehat{\Omega} = \ \{ X = PVP^T + \alpha \overline{W} \ : \ \langle I_r, V \rangle + \alpha \le \delta, V \succeq 0, \alpha \ge 0 \}$$

with $\overline{W} \succeq 0$, $\left\langle I, \overline{W} \right\rangle = 1$, and $P^T P = I_r$ with r small

Second Order Cone

$$x = \begin{pmatrix} x_0 \\ \underline{x} \end{pmatrix} \in \mathrm{SOC}_n \qquad \Leftrightarrow \qquad x_0 \geq \|\underline{x}\|$$

Make a small model with the same structure

$P = \begin{bmatrix} 1 & 0 \\ 0 & \underline{P} \end{bmatrix}$ with $\underline{P}^T \underline{P} = I_r$, r small

then $\quad \xi = \begin{pmatrix} \xi_0 \\ \underline{\xi} \end{pmatrix} \in \mathrm{SOC}_{r+1} \quad \Rightarrow \quad x = P\xi \ \in \mathrm{SOC}_n$

because $x_0 = \xi_0$ and $\|\underline{x}\| = \|\underline{P}\underline{\xi}\| = \|\underline{\xi}\|$

Build $\underline{P}_+ = \mathrm{orth}(\underline{P}_{old}, \underline{x}_i^*)$,

Substitute $x = P\xi$ to keep quadratic subproblem small

Use $d^T x = x_0 = \xi_0$

$$\begin{array}{ll} \max & (c^T P - \hat{y}^T AP)\xi + b^T \hat{y} - \frac{1}{2}\|b - AP\xi\|^2 \\ \text{s.t.} & \xi_0 \leq \delta \\ & \xi_0 \geq \|\bar{\xi}\| \end{array}$$

- catches nonpolyhedral structure
- TWO columns in $\underline{P}$ suffice to span optimal solution (NO other aggregate required!) $\rightarrow$ 3 variables
- larger $\underline{P}$ might speed up convergence considerably
- need primal information (v_i, P) and extra oracle routines for computing $A\bar{P}$, $c^T\bar{P}$ in every iteration

Disadvantage:

- NO possibility to generalize to block structure in one model!
  Each block needs an extra $\xi_0 \geq \bar{\xi}$,
  better use polyhedral approximation or separate models

Block Structure in SDP

Can be handled by the oracle:

$$C - \mathcal{A}^T y = \begin{bmatrix} C_1 - \mathcal{A}^T{}_1 y & 0 & & 0 \\ 0 & C_2 - \mathcal{A}^T{}_2 y & & 0 \\ & & \ddots & \\ 0 & 0 & & C_k - \mathcal{A}^T{}_k y \end{bmatrix}$$

Compute $\lambda_{\max}(C_i - \mathcal{A}^T{}_i y)$, use eigenvector $v_i = (0, \ldots, 0, v, 0, \ldots, 0)$.

Sort by eigenvalue and return a few of the largest ones.

- ONE semidefinite quadratic model sufficient,
  NO need to increase its size
- Disadvantage: a block matrix with many small blocks will need many evaluations
- Code: block structure as well as several semidefinite models

Can we extend bounded trace to unbounded trace?

we would like to solve the quadratic subproblem

$$\begin{array}{ll} \max & (c - A^T\widehat{y})x + b^T\widehat{y} - \frac{1}{2}\left\|b - Ax\right\|^2 \\ \text{s.t.} & x = \sum \xi_i v_i \\ & \xi \geq 0 \end{array}$$

no theoretical difficulty because of primal boundedness

but practical difficulties: needs infeasible method

(feasible methods have implementational advantages)

Approach: Solve

$$\begin{array}{ll} \max & (c - A^T\widehat{y})^T x + b^T\widehat{y} - \frac{1}{2}\left\|b - Ax\right\|^2 \\ \text{s.t.} & d^T x \leq \delta \\ & x = \sum \xi_i v_i \\ & \xi \geq 0 \end{array}$$

Whenever $d^T x \geq 0.95 \cdot \delta$, double δ and resolve.

Remarks:

- Doubling δ has to stop after finitely many iterations because of primal boundedness.
- Can be thought of as a big M method with dynamically increasing M so as to reduce numerical difficulties
- works reasonably well in practice (dual value jumps up)
- applicable to all proposed cutting models for symmetric cones

Code offers Four Models:
(in all cases primal aggregation is possible)

- convex combinations $e^T\xi = 1$
  - general convex functions, LP over Boxes
  - no primal information required
- conic combinations $e^T\xi \leq \delta$
  - e.g. for $\mathbb{R}^n_+$ or second order cone with blocks
  - no primal information required
- A Single Second Order Cone Block
  - e.g. primal quadratic functions or free variables
  - full primal information available
- The SDP-model
  - single and block structure in one
  - partial primal information available

No need to know a bound on the trace as long as the primal feasible set is bounded!

Free Variables [Jos Sturm]

Splitting into two $\mathbb{R}_+$ variables would destroy primal boundedness

$\rightarrow$ use one “unbounded” second order cone to collect all free variables

---

Additional Features

- inequality constraints $Ax \leq b$ [H., Kiwiel]
  (can be combined with “unbounded” approach)
- Support for primal cutting plane approaches
  (extend old subgradients from primal aggregates)
- Callable library with interfaces for C and $C++$
  (with STL-Classes only)

Preliminary Computational Results

name	opt	val	$\|\|\text{subg}\|\|$	# eval	time	term
toruspm-3-8-50	527.80866	527.80904	$8.7 \cdot 10^{-3}$	133	31	ok
bm1	23.4434	23.4240	$1.7 \cdot 10^{-3}$	5964	1:49:26	aug
filter48	1.4161290	1.3451099	$2.5 \cdot 10^{-3}$	199	14:31	aug
minphase	5.98	5.67	$2.0 \cdot 10^{-5}$	18064	2:56	aug
nb_L2	-1.6289720	-1.6289804	$2.3 \cdot 10^{-5}$	92	1:24	aug
copo14	0	0.10114926	$1.8 \cdot 10^{-3}$	4925	11:43:57	kill
sched_50_50_scaled	7.8520384	7.8520214	$1.2 \cdot 10^{-4}$	709	40:09	ok

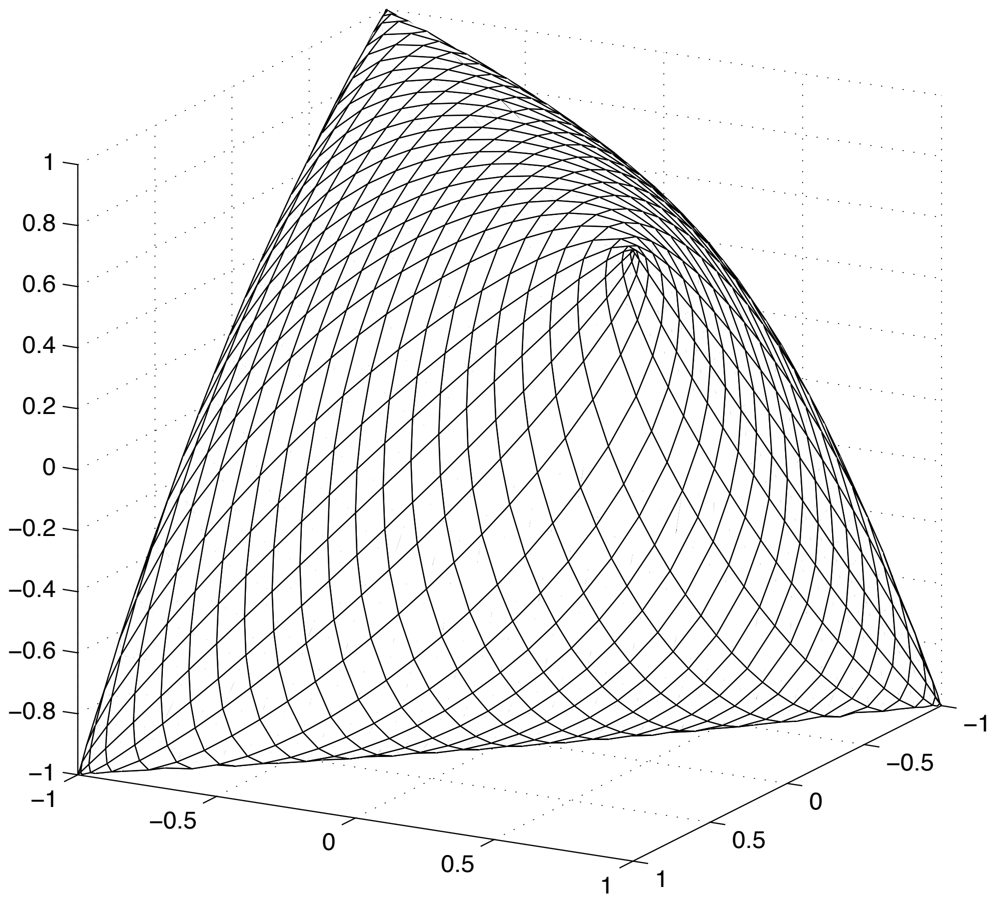
Performs well on same classes as before:
- few and large semidefinite and second order cone blocks
- LP over box constraints (0-1 boxes so far; network flow)
- approximate solutions only

Is terribly slow on most of the smaller DIMACS challenge instances:
- cones that are direct products of many small cones
- numerically difficult
- subproblems too expensive for small problems

Strong dependence on many parameters:
- updating rules for weight and bound on trace
- how to split up sums
- how to fix bundle sizes and updating schemes
- starting point heuristics for general problems

The semidefinite feasible set for $n = 3$

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



Main steps

1. Find candidate by solving quadratic model

2. Evaluate function, determine subgradient

3. Decide on
    - null step
    - descent step

4. Update model and iterate

The Semidefinite Quadratic Model

For fixed slack variable η and center $\hat{y}$ solve

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

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

Theorem 2 *[H. 2001]*
If the eigenvalue problem has an optimal solution then the algorithm generates a subsequence $K \subseteq \mathbb{N}$ so that all cluster points of $\overline{X}^k$, $k \in K$, are primal optimal solutions.

similar to Feltenmark and Kiwiel 2000

Combining the spectral bundle method with cutting planes

Idea: separate with respect to $\overline{X}$

Difficulties

- $\overline{X}$ is ‘never’ feasible for all given constraints

$\longrightarrow$ the same inequalities may be separated again and again

$\longrightarrow$ separation routines can ‘conceal’ certain violated inequalities

---

What kind of separation oracle do we need?

Is it still possible to guarantee convergence to the optimal solution?

Maximum violation oracle with respect to $\mathcal{A}X \leq b$:

- returns inequalities from a finite inequality system

$$\langle A_i, X \rangle \leq b_i, \quad i \in \{1, \ldots, m\}$$

- for a given $\overline{X}$ the oracle either
  - asserts $\overline{X} \in \mathcal{P}$, or
  - returns an inequality $j \in \{1, \ldots, m\}$ with
    $b_j - \left\langle A_j, \overline{X} \right\rangle \leq \min_i b_i - \left\langle A_i, \overline{X} \right\rangle < 0.$

[many separation routines satisfy this]

Cutting plane algorithm 1
[for max $\langle C, X\rangle$ s.t. $X \in \{X \succeq 0 : \langle I, X\rangle = a\} \cap \{X : \mathcal{A}X \leq b\}$]

1. Solve quadratic model $\longrightarrow \overline{X}$

   If oracle($\overline{X}$) returns a new inequality, add it and go to 1

2. Evaluate function, determine subgradient

3. Decide on
   - null step
   - descent step

4. Update model and iterate

Theorem 3 *[H. 2001]*
If the eigenvalue problem (for all m constraints) has an optimal solution then the algorithm converges to an optimal solution and generates a subsequence $K \subseteq \mathbb{N}$ so that all cluster points of $\overline{X}^k$, $k \in K$, are primal optimal solutions.

Idea:

1. Wait till the oracle adds no more inequalities to index set J (finite)
2. Apply Theorem 2 to problem specified by subsystem J

   $\Rightarrow$ there is subsequence K with $\overline{X}^k \to X_J^*$ feasible and optimal for J

   $\Rightarrow$ violation $\to 0$ on inequalities J

   Maximum violation oracle $\Rightarrow$ all are satisfied for X_J^*

Can we eliminate inactive inequalities during runtime?

Cutting plane algorithm 2

[for max $\langle C, X\rangle$ s.t. $X \in \{X \succeq 0 : \langle I, X\rangle = a\} \cap \{X : \mathcal{A}X \leq b\}$]

1. Solve quadratic model $\longrightarrow \overline{X}$

   If oracle($\overline{X}$) returns a new inequality, add it and go to 1

2. Evaluate function, determine subgradient

3. Decide on
   - null step
   - descent step: delete inequalities inactive for $\overline{X}$

4. Update model and iterate

Theorem 4 *[H. 2001]*
If the primal has a strictly feasible solution then the upper bound converges to the optimal value and the algorithm generates a subsequence $K \subseteq \mathbb{N}$ *so that all cluster points of* $\overline{X}^k$, $k \in K$, *are primal optimal solutions.*

The strictly feasible primal solution ensures boundedness of dual iterates

---

??? It would be nice to have: If the primal is feasible . . .

What I do in practice

1. Solve quadratic model $\longrightarrow \overline{X}$

2. Evaluate function, determine subgradient

3. Decide on
    - null step
    - descent step: if relative error $\leq$ 0.05
      delete inequalities significantly inactive for $\overline{X}$
      separate for $\overline{X}$, add new inequalities.

4. Update model and iterate

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.

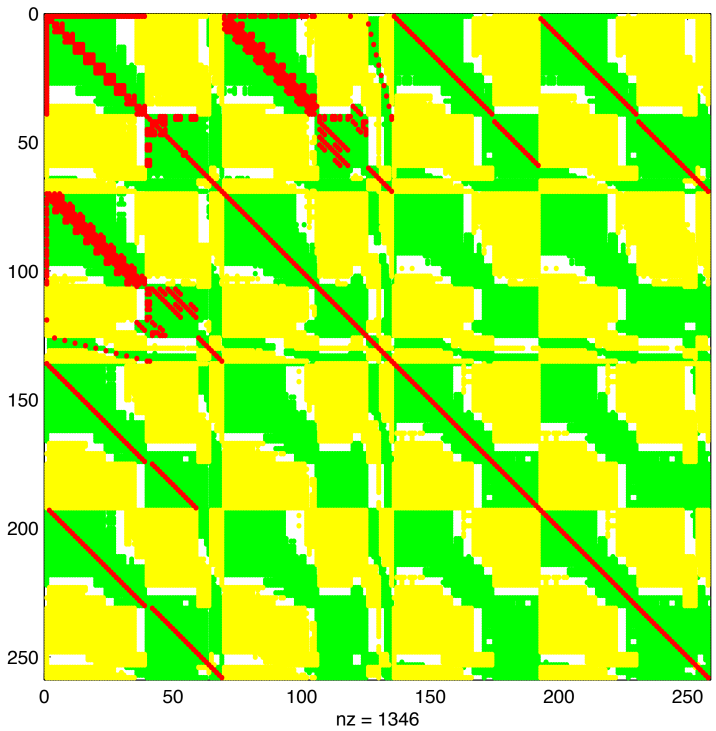
$$\text{(BS)} \qquad \min_{S\subseteq V,\ \big||S|-|S\backslash V|\big|\leq\sigma n} \quad \sum_{ij\in\delta(S)} a_{ij}$$

For an appropriate cost matrix C

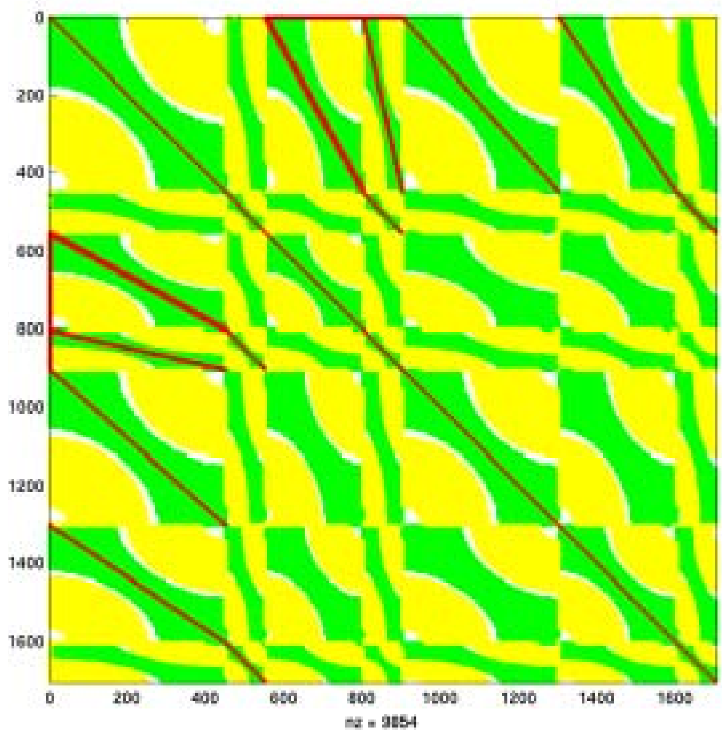
$$\max_{\substack{x\in\{-1,1\}^n \\ (e^T x)^2 \leq \lfloor \sigma n \rfloor^2}} x^T C x \leq \boxed{\begin{array}{rl} \max & \langle C, X\rangle \\ \text{s.t.} & \text{diag}(X) = e \\ & \left\langle ee^T, X \right\rangle \leq \lfloor \sigma n \rfloor^2 \\ & X \succeq 0 \\ & [\text{rank}(X) = 1] \end{array}}$$

Structure of X

putt01

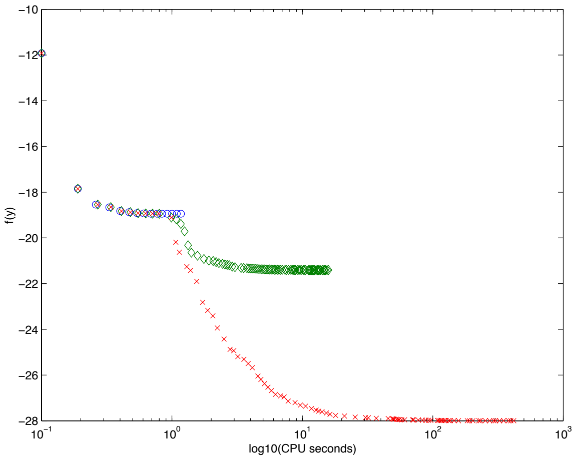


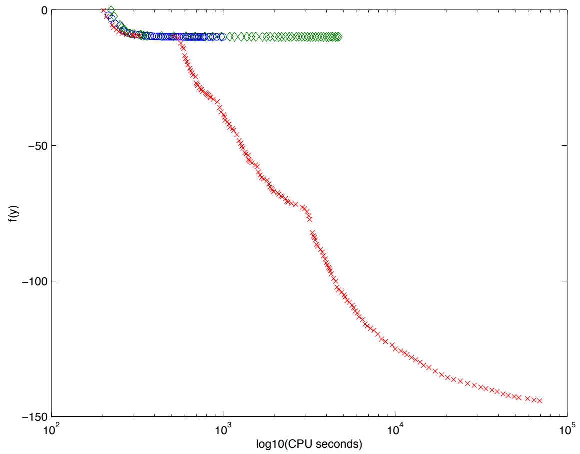
shut01

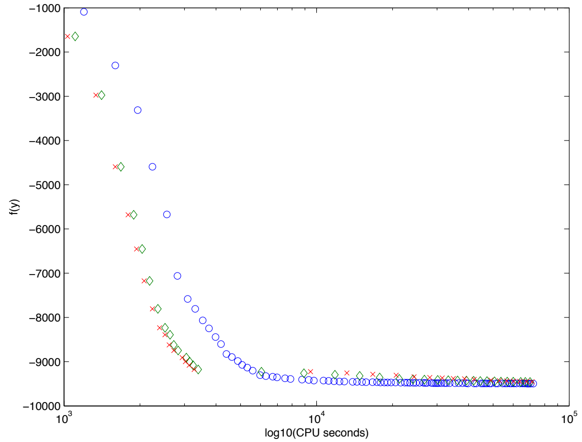


Results for Min Bisection: KKT-Instances from Boeing $\sigma = 0.05$

	putt01	heat02	traj33
	$n = 258$, $nz = 548$	$n = 5200$, $nz = 25056$	$n = 20006$, $nz = 261953$
SDP-UB:	-18.94562	-9.940345	-9496.117
S+C-UB:	-21.40906	-9.940555	-9460.441
S+EC-UB:	-27.99941	-144.1614	-9454.076
LB:	-28	-150	-9593







time in logarithmic scale!