Operations research meets data science: principles, models and algorithms

Immanuel Bomze, University of Vienna

Dip. Matematica@Univ. Padova 10 November 2025

Lagrangian dual bounds are tighter than LP bounds

For ILP, define the ground set

$$X = \{\mathbf{x} \in \mathbb{R}^n : \mathsf{E}\mathbf{x} \le \mathbf{e}\,,\ x_i \text{ integer for all } i\}$$

so that $M = \{\mathbf{x} \in X : \mathsf{A}\mathbf{x} \le \mathbf{b}\}$.

Catch: $\mathsf{E}\mathbf{x} \le \mathbf{e}$ easy to handle (e.g., variable bounds).

Lagrange function

$$L(\mathbf{x};\mathbf{y}) = \mathbf{c}^\top\mathbf{x} + \mathbf{y}^\top(\mathsf{A}\mathbf{x} - \mathbf{b}) = (\mathbf{c} + \mathsf{A}^\top\mathbf{y})^\top\mathbf{x} + \mathbf{y}^\top\mathbf{b}$$

is easier minimized over $\mathbf{x} \in X$ for any $\mathbf{y} \in \mathbb{R}^m_+$, giving $\Theta(\mathbf{y})$.

$$d^* = \sup_{\mathbf{y}\in\mathbb{R}^m_+} \Theta(\mathbf{y}) \le z^*(P) \quad \text{by weak duality.}$$

In ILPs, Lagrangian dual bound always dominates LP bound:

$$z^*_{\mathsf{LP}} \le d^* \le z^*(P)\,,\ \text{strict inequalities possible.}$$

Gaps example

$$X = \left\{ \mathbf{x} \in \mathbb{R}^2 : x_1 + 2x_2 \leq 1\,,\; x_i \text{ binary} \right\} = \left\{ \mathbf{o}, [1,0]^\top \right\}$$

and

$$M = \left\{ \mathbf{x} \in X : x_1 + x_2 \geq \tfrac{1}{2} \right\} = \left\{ [1,0]^\top \right\}$$

with $\mathbf{c}^\top \mathbf{x} = 2x_1 + x_2$ has obviously $z^*(P) = f([1,0]^\top) = 2$. And

$$L(\mathbf{x}; y) = (2-y)x_1 + (1-y)x_2 + \tfrac{y}{2}$$

so that

$$\Theta(y) = \min\left\{ L(\mathbf{x}; y) : \mathbf{x} \in \left\{ \mathbf{o}, [1,0]^\top \right\} \right\} = \min\left\{ \tfrac{y}{2}, 2 - \tfrac{y}{2} \right\}\,.$$

We find $d^* = \max\limits_{y \geq 0} \Theta(y) = 1$ attained at $y^* = 2$ so that

$$d^* = 1 < 2 = z^*(P)\,.$$

Mind the gaps !

For LP relaxation,

$$F = \left\{ \mathbf{x} \in \mathbb{R}^2_+ : x_1 + 2x_2 \le 1\,,\ x_1 + x_2 \ge \tfrac{1}{2} \right\}$$

has vertices

$$V_F = \left\{ \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \frac{1}{2} \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$$

so $z^*_{\mathsf{LP}} = \min_{\mathbf{v} \in V_F} \mathbf{c}^\top \mathbf{v} = \min_{\mathbf{v} \in V_F} [2v_1 + v_2] = \min \left\{ 1, \frac{1}{2}, 2 \right\} = \frac{1}{2}$. Hence

$$z^*_{\mathsf{LP}} = \tfrac{1}{2} < d^* = 1 < z^*(P) = 2\,.$$

When is there no gap ...

... for LP relaxation ?

Easy answer: if $z^*_{\mathsf{LP}} = \min\left\{\mathbf{c}^\top\mathbf{x} : \mathbf{x} \in F\right\}$ has an integer solution $\mathbf{x}^*$, then $z^*(P_0) = z^*_{\mathsf{LP}}$.

Yes, but when ensured ?

Know: feasible and bounded LP has solution at a vertex $\mathbf{x}^*$ of F.

Write in standard form: A$\mathbf{x} = \mathbf{b}$, $\mathbf{x} \geq \mathbf{o}$. Then either $x^*_i = 0$ (nonbasic var.) or else (basic var.) part of unique soln. to

$$\sum_{j=1}^{m} x^*_j \mathbf{a}_j = \mathbf{b}$$

for m lin.indep. columns $\mathbf{a}_j$ of A. Write $\mathsf{A}_B = [\mathbf{a}_1, \ldots, \mathbf{a}_m]$.

Inverting the basis matrix

At a vertex, we have

$$\mathbf{x}^* = \begin{bmatrix} \mathbf{x}_B^* \\ \mathbf{o} \end{bmatrix} = \begin{bmatrix} \mathsf{A}_B^{-1}\mathbf{b} \\ \mathbf{o} \end{bmatrix}.$$

Obviously this is integer if $\mathbf{b}$ and A_B^{-1} have only integer entries.

Recall: if $(\mathsf{A}_B, \mathbf{b})$ have integer entries and $|\det(\mathsf{A}_B)| = 1$, then also $\mathbf{x}_B = \mathsf{A}_B^{-1}\mathbf{b}$ has integer entries; indeed, Cramer's rule says

$$x_j^* = \det(\mathsf{A}_B|\mathbf{b})_j / \det(\mathsf{A}_B)$$

where $(\mathsf{A}_B|\mathbf{b})_j$ is the matrix where $\mathbf{b}$ replaces column $\mathbf{a}_j$.

And then $\mathbf{x}^*$ is integer, implying that LP relaxation is tight.

Unimodularity

An integer $m \times n$ matrix A is called *unimodular*, if for all $m \times m$-submatrices A_B of A their determinant satisfies

$$|\det A_B| \in \{0, 1\} \ .$$

If in an ILP

$$\min \left\{ \mathbf{c}^\top \mathbf{x} : A\mathbf{x} = \mathbf{b}, \mathbf{x} \in \mathbb{R}^n_+ \text{ integer } \right\}$$

$(A, \mathbf{b})$ are integer and A is unimodular, then the LP relaxation

$$\min \left\{ \mathbf{c}^\top \mathbf{x} : A\mathbf{x} = \mathbf{b}, \mathbf{x} \in \mathbb{R}^n_+ \right\}$$

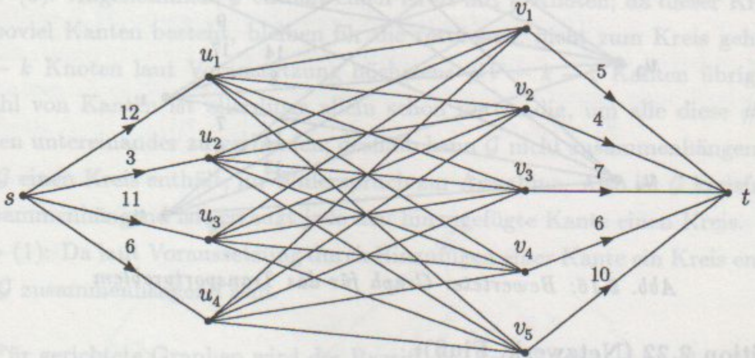
is exact, if feasible and bounded.

Important cases of unimodularity

Incidence matrices A of directed graphs are unimodular; incidence matrices A of undirected bipartite graphs as well.

more generally, the following matrices are unimodular: any matrix with $\{-1, 0, 1\}$-entries with at most two nonzeroes in each column, such that there is a partition of all row indices $\{1, \ldots, m\} = I_1 \cup I_2$ satisfying

(i) if two nonzeroes of a column have same sign, one sits in a row in I_1, the other in a row in I_2;
(ii) if two nonzeroes of a column have opposite sign, they sit in two rows either both in I_1 or both in I_2.



s
t
u_1
u_2
u_3
u_4
v_1
v_2
v_3
v_4
v_5
12
3
11
6
5
4
7
6
10

And if LP solution is fractional ?

Also for fractional optimal LP solution $\mathbf{x}^*$ we have

$$\mathbf{x}^* = \begin{bmatrix} \mathbf{x}_B^* \\ \mathbf{o} \end{bmatrix} = \begin{bmatrix} \mathsf{A}_B^{-1}\mathbf{b} \\ \mathbf{o} \end{bmatrix}.$$

And for *all feasible LP solutions* $\mathbf{x}$ have with $\mathsf{A} = [\mathsf{A}_B \,|\, \mathsf{A}_N]$

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_B \\ \mathbf{x}_N \end{bmatrix} = \begin{bmatrix} \mathsf{A}_B^{-1}(\mathbf{b} - \mathsf{A}_N\mathbf{x}_N) \\ \mathbf{x}_N \end{bmatrix};$$

indeed, $\mathbf{b} = \mathsf{A}\mathbf{x} = \mathsf{A}_B\mathbf{x}_B + \mathsf{A}_N\mathbf{x}_N$ gives $\mathbf{x}_B = \mathsf{A}_B^{-1}(\mathbf{b} - \mathsf{A}_N\mathbf{x}_N)$.

Rewrite $\mathbf{x}_B = \mathsf{A}_B^{-1}\mathbf{b} - \mathsf{A}_B^{-1}\mathsf{A}_N\mathbf{x}_N = \mathbf{x}_B^* - \mathsf{Z}\mathbf{x}_N$ with $\mathsf{Z} = \mathsf{A}_B^{-1}\mathsf{A}_N$.

Consider a fractional x_i^* with $i \in B$ (all $x_i^* = 0$ if $i \notin B$); then

$$x_i = x_i^* - (\mathsf{Z}\mathbf{x}_N)_i = x_i^* - \sum_{j \in N} z_{ij}x_j \quad \text{or} \quad x_i + \sum_{j \in N} z_{ij}x_j = x_i^*.$$

Enter integrality ...

Use integer part $\lfloor t \rfloor = \max\{m \text{ integer} : m \leq t\}$ of any $t \in \mathbb{R}$:

$$x_i + \sum_{j \in N} \lfloor z_{ij} \rfloor x_j \leq x_i + \sum_{j \in N} z_{ij} x_j = x_i^* \qquad \text{as all } x_j \geq 0\,.$$

Now if $\mathbf{x} \in \mathbb{R}^n_+$ is **integer** with $A\mathbf{x} = \mathbf{b}$, get more:

$$x_i + \sum_{j \in N} \lfloor z_{ij} \rfloor x_j \leq \lfloor x_i^* \rfloor < x_i^*\,,$$

since left-hand side is now integer and x_i^* is not.

Replacing $\mathbf{x}$ with $\mathbf{x}^*$ violates above inequality:

$$x_i^* = x_i^* + 0 = x_i^* + \sum_{j \in N} \lfloor z_{ij} \rfloor x_j^* \leq \lfloor x_i^* \rfloor < x_i^*$$

is a contradiction !

Gomory's cut

Summary:

If $\mathbf{x}^*$ vertex of LP relaxation, then

$$x_i + \sum_{j \in N} \lfloor z_{ij} \rfloor x_j \leq \lfloor x_i^* \rfloor$$

for all feasible solutions $\mathbf{x}$ to ILP.
If x_i^* is fractional, $\mathbf{x}^*$ violates above inequality, if replacing $\mathbf{x}$.

Ralph E. Gomory

(* 07 May 1929)

http://www.ralphgomory.org/

New constraint cuts away $\mathbf{x}^*$ but keeps all ILP-feasible $\mathbf{x}$.

Retry LP relaxation, adding this constraint – tighter bound !

Finiteness of iterative Gomory cuts

> After finitely many Gomory cuts, either an optimal solution to ILP is detected or else infeasibility of this ILP is certified.

But finiteness is not efficiency in general ...

Many other cuts possible ...

... for instance *optimality cuts*:

start from feasible incumbent $\bar{\mathbf{x}}$ (integer) with value $f(\bar{\mathbf{x}}) = \mathbf{c}^\top \bar{\mathbf{x}}$.

If $\delta > 0$ desired improvement, add linear constraint

$$\mathbf{c}^\top \mathbf{x} \leq f(\bar{\mathbf{x}}) - \delta\,.$$

LP-relaxation better than incumbent, but feasible improvement may be impossible (observe that, e.g., ILP-feasible $\bar{\mathbf{x}}$ is cut away).

Does not cut away fractional LP-solution in general.

Cut-and-branch vs. branch-and-cut

Cut-and-branch: (Gomory) cut at root node, then branch-and-bound.

Branch-and-cut: tighten LP bounds at every processed node by cuts.

See example by John Mitchell

$$\begin{aligned} 6x_1 + 5x_2 &\to \text{max} \\ 3x_1 + x_2 &\leq 11 \\ -x_1 + 2x_2 &\leq 5 \\ x_1, \qquad x_2 &\geq 0, \text{integer.} \end{aligned}$$

A Branch-and-Cut Example

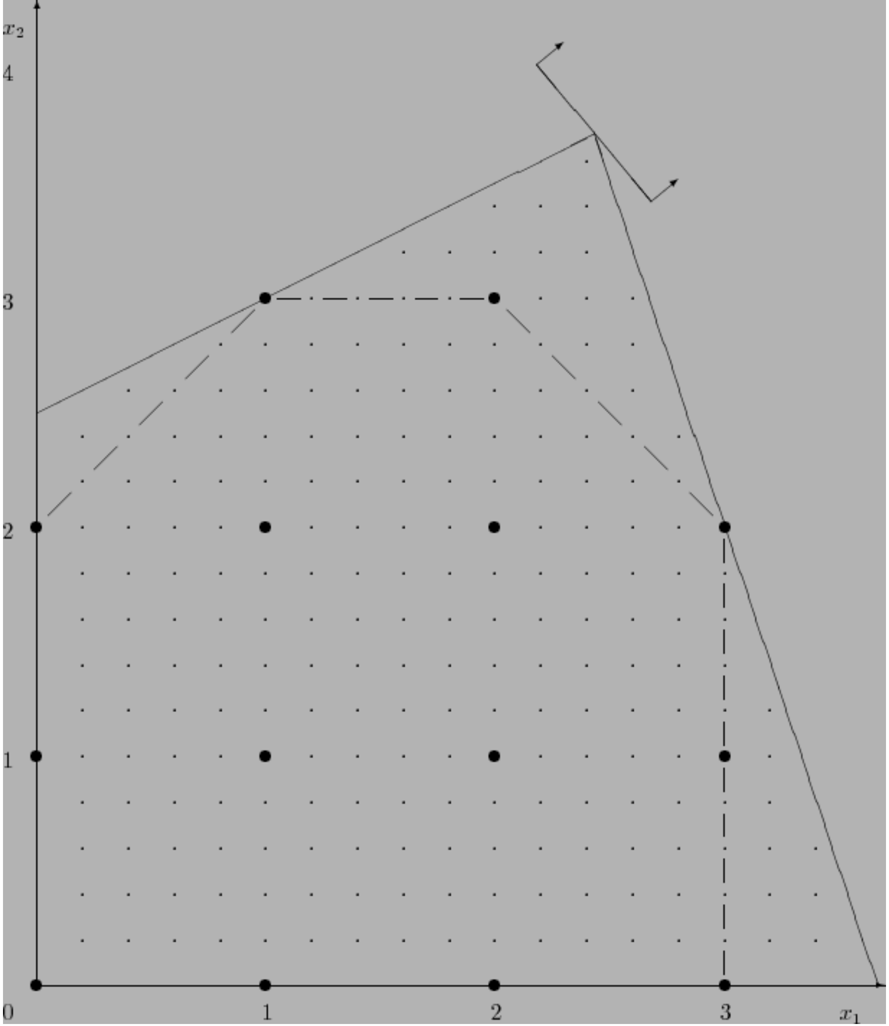
John E. Mitchell

The integer programming problem

$$\begin{array}{llrcrcl} \min & z := & -6x_1 & - & 5x_2 & & \\ \text{subject to} & & 3x_1 & + & x_2 & \leq & 11 \\ & & -x_1 & + & 2x_2 & \leq & 5 \\ & & & & x_1, x_2 & \geq & 0, \text{ integer.} \end{array} \qquad (Eg0)$$

is illustrated in the figure. The feasible integer points are marked. The *linear programming relaxation* (or *LP relaxation*) is obtained by ignoring the integrality restrictions and is indicated by the polyhedron contained in the solid lines.

A branch-and-cut approach first solves the linear programming relaxation, giving the point $(2\frac{3}{7}, 3\frac{5}{7})$, with value $-33\frac{1}{7}$. There is now a choice: should the LP relaxation be improved by adding a cutting plane, for example, $x_1 + x_2 \leq 5$, or should the problem be divided into two by splitting on a variable?



If the algorithm **splits on** $\mathbf{x_1}$, two new problems are obtained:

$$
\begin{array}{llrcrcll}
\min & z := & -6x_1 & - & 5x_2 & & & \\
\text{subject to} & & 3x_1 & + & x_2 & \leq & 11 & \\
& & -x_1 & + & 2x_2 & \leq & 5 & \qquad (Eg1) \\
& & \mathbf{x_1} & & & \geq & \mathbf{3} & \\
& & & & x_1, x_2 & \geq & 0, \text{ integer.} &
\end{array}
$$

and

$$
\begin{array}{llrcrcll}
\min & z := & -6x_1 & - & 5x_2 & & & \\
\text{subject to} & & 3x_1 & + & x_2 & \leq & 11 & \\
& & -x_1 & + & 2x_2 & \leq & 5 & \qquad (Eg2) \\
& & \mathbf{x_1} & & & \leq & \mathbf{2} & \\
& & & & x_1, x_2 & \geq & 0, \text{ integer.} &
\end{array}
$$

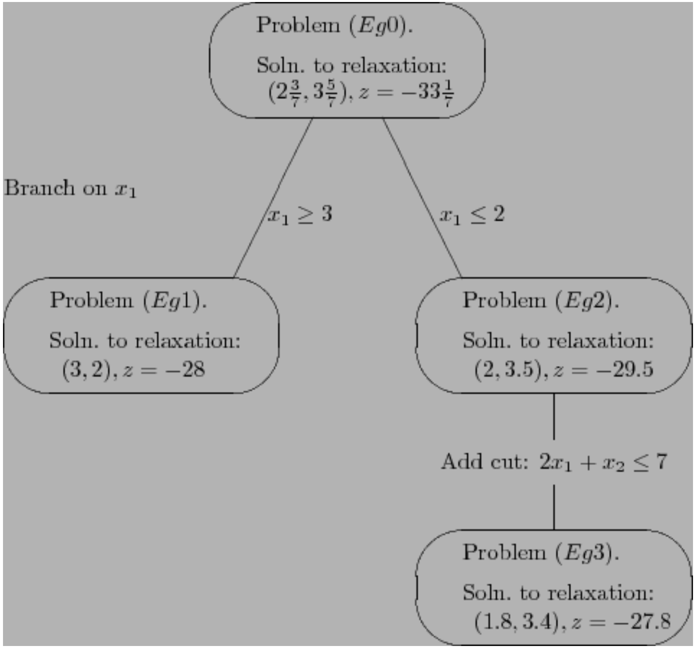
The optimal solution to the original problem will be the better of the solutions to these two subproblems. The solution to the linear programming relaxation of $(Eg1)$ is $(3, 2)$, with value $\underline{-28}$. This solution is integral, so it solves $(Eg1)$, and becomes the incumbent best known feasible solution. The LP relaxation of $(Eg2)$ has optimal solution $(2, 3.5)$, with value $\underline{-29.5}$. This point is nonintegral, so it does not solve $(Eg2)$, and it must be attacked further.

Assume the algorithm **uses a cutting plane approach and adds the inequality** $\underline{\mathbf{2x_1 + x_2 \leq 7}}$ to $(Eg2)$. This is a valid inequality, in that it is satisfied by every integral point that is feasible in $(Eg2)$. Further, this inequality is violated by $(2, 3.5)$, so it is a cutting plane. The resulting subproblem is

$$
\begin{array}{llrcrcll}
\min & z := & -6x_1 & - & 5x_2 & & & \\
\text{subject to} & & 3x_1 & + & x_2 & \leq & 11 & \\
& & -x_1 & + & 2x_2 & \leq & 5 & \qquad (Eg3) \\
& & x_1 & & & \leq & 2 & \\
& & \mathbf{2x_1} & + & \mathbf{x_2} & \leq & \mathbf{7} & \\
& & & & x_1, x_2 & \geq & 0, \text{ integer.} &
\end{array}
$$

The LP relaxation of $(Eg3)$ has optimal solution $(1.8, 3.4)$, with value $\underline{-27.8}$. Notice that the optimal value for this modified relaxation is larger than the value of the incumbent solution. The value of the optimal integral solution to the second subproblem must be at least as large as the value of the relaxation. Therefore, the incumbent solution is better than any feasible integral solution for $(Eg3)$, so it actually solves the original problem.

The progress of the algorithm is illustrated below.



Notice that the cutting plane introduced in the second subproblem is not valid for the first subproblem. This inequality can be modified to make it valid for the first subproblem by using a *lifting* technique.

- Bibliography
- About this document ...

John Mitchell
2009-03-16

Mitchell example: various cuts

LP solution $\mathbf{x}^* = [\frac{17}{7} = 2\frac{3}{7}, \frac{26}{7} = 3\frac{5}{7}]^\top$ with $\mathbf{c}^\top \mathbf{x}^* = -\frac{232}{7} = -33\frac{1}{7}$.

For optimality cut round $\mathbf{x}^*$ to $\bar{\mathbf{x}} = [2, 3]^\top$ with $\mathbf{c}^\top \bar{\mathbf{x}} = -27$.

Try $\delta = 1$ for optimality cut and add

$$-6x_1 - 5x_2 \leq -27 - 1 = -28 \quad \text{or} \quad 6x_1 + 5x_2 \geq 28\,.$$

Gives an LP with still same $\mathbf{x}^*$ as optimal solution.

Note: for $\delta = 2$ get cut $\qquad 6x_1 + 5x_2 \geq 29,$

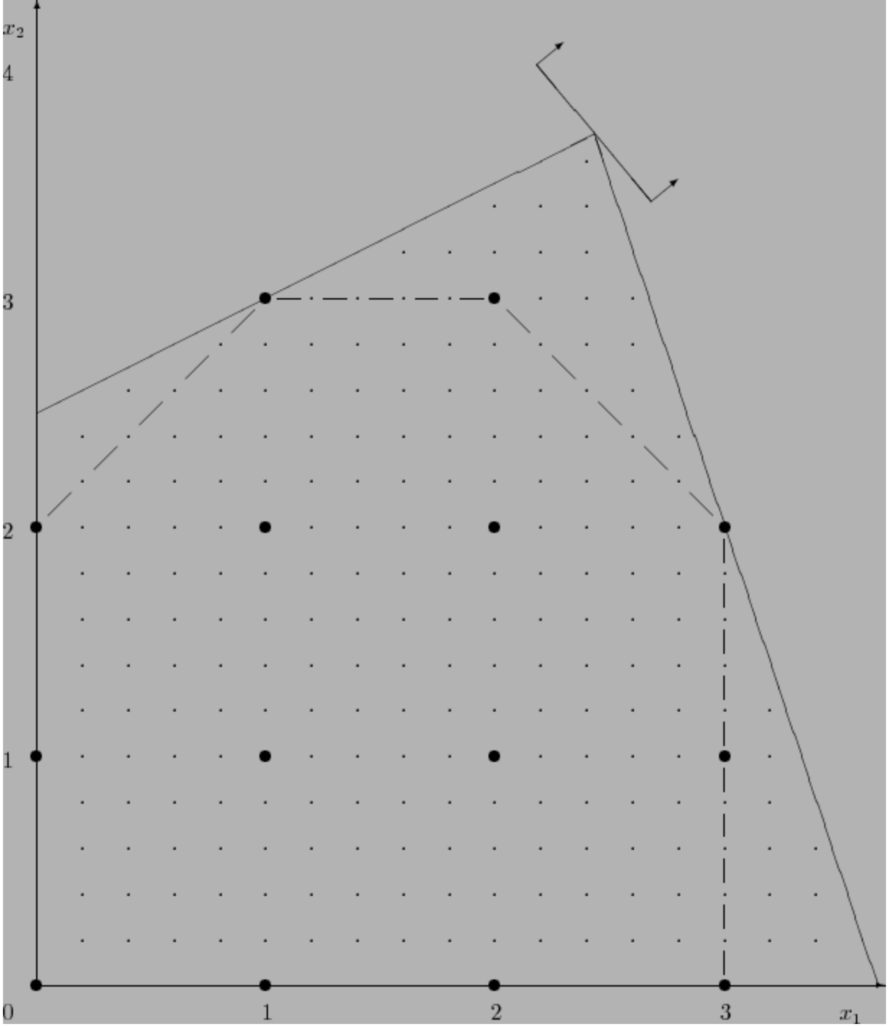
cutting away all ILP-feasible points, for $\delta = 7$ even LP infeasible.

Other cuts ? Mitchell suggests

$$x_1 + x_2 \leq 5\,.$$

Nice, as new LP gives optimal ILP solution
(although LP has also fractional vertices).

But how to get there?



If the algorithm **splits on** $\mathbf{x_1}$, two new problems are obtained:

Mitchell example in standard form

LP at root node in standard form needs slacks (s_1, s_2):

$$\begin{bmatrix} 3x_1 & + & x_2 & + & s_1 & \\ -x_1 & + & 2x_2 & + & & s_2 \end{bmatrix} = \mathbf{b} = \begin{bmatrix} 11 \\ 5 \end{bmatrix}$$

has solution $(\mathbf{x}^*; \mathbf{s}^*) = [\frac{17}{7}, \frac{26}{7}; 0, 0]^\top$ with basis/nonbasis matrix

$$\mathsf{A}_B = \begin{bmatrix} 3 & 1 \\ -1 & 2 \end{bmatrix}, \ \mathsf{A}_N = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

so

$$\mathsf{A}_B^{-1} = \frac{1}{7} \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix} \quad \text{and} \quad \mathsf{Z} = \mathsf{A}_B^{-1} \mathsf{A}_N = \begin{bmatrix} \frac{2}{7} & -\frac{1}{7} \\ \frac{1}{7} & \frac{3}{7} \end{bmatrix}$$

Remark: all above is delivered by any simplex-based LP solver, only here done by hand.

Gomory cut for Mitchell example

$$\mathsf{Z} = \mathsf{A}_B^{-1}\mathsf{A}_N = \begin{bmatrix} \frac{2}{7} & -\frac{1}{7} \\ \frac{1}{7} & \frac{3}{7} \end{bmatrix} \text{ rounded down: } \lfloor \mathsf{Z} \rfloor = \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix}.$$

Recall that Z links basic variables (here x_1, x_2) to nonbasic ones (here s_1, s_2) via $\mathbf{x} + \mathsf{Z}\mathbf{s} = \mathbf{x}^*$. Gives Gomory cut for fractional x_1^*:

$$x_1 + 0 * s_1 - 1 * s_2 = x_1 + \lfloor z_{11} \rfloor s_1 + \lfloor z_{12} \rfloor s_2 \leq \lfloor x_1^* \rfloor = \lfloor \tfrac{17}{7} \rfloor = 2,$$

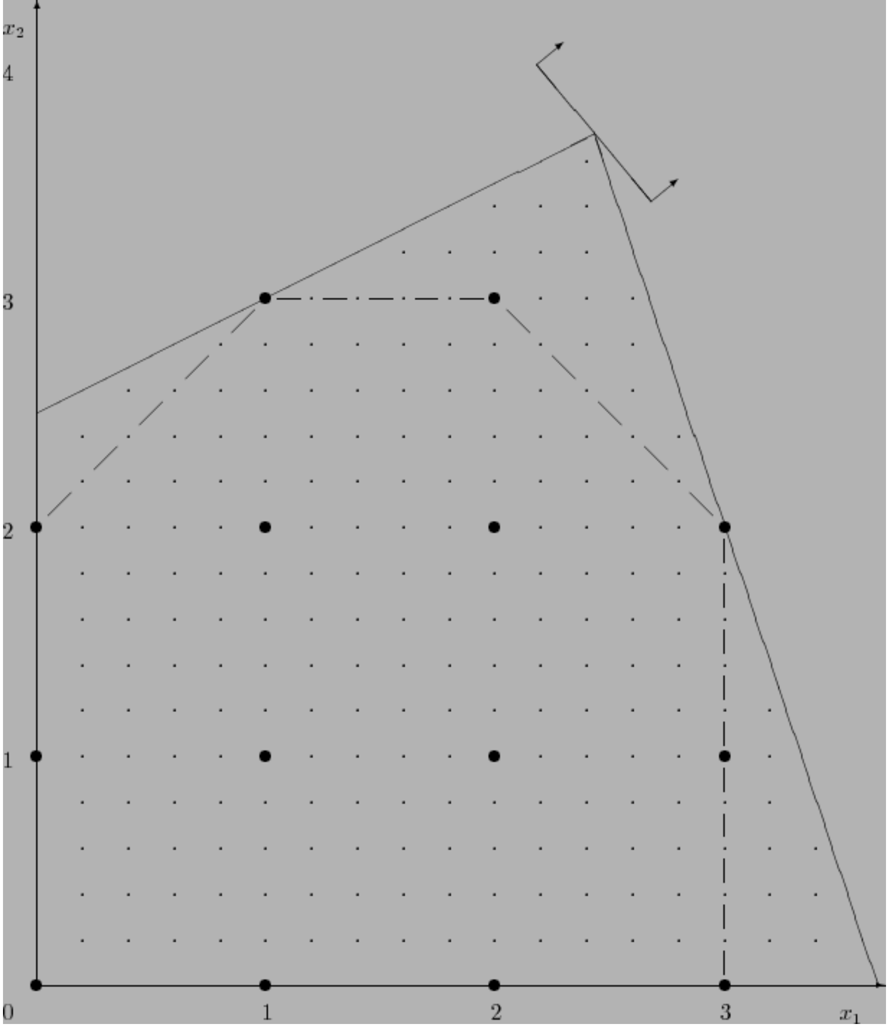
and $s_2 = 5 + x_1 - 2x_2$, which gives

$$2x_2 - 5 = x_1 - (5 + x_1 - 2x_2) \leq 2 \quad \text{or} \quad x_2 \leq \tfrac{7}{2}.$$

But we know that ILP has integer x_2, so can round down $\frac{7}{2}$ again, so we add cut

$$x_2 \leq 3.$$

If we use Gomory cut for x_2^* instead of x_1^*, get $x_2 \leq 3$ again.



If the algorithm **splits on** $\mathbf{x_1}$, two new problems are obtained:

New LP relaxation with Gomory cut for Mitchell example

Arrive at LP relaxation

$$
\begin{aligned}
6x_1 + 5x_2 &\to \text{max} \\
3x_1 + x_2 &\leq 11 \\
-x_1 + 2x_2 &\leq 5 \\
x_2 &\leq 3 \\
x_1, \quad x_2 &\geq 0\,.
\end{aligned}
$$

New optimal solution to this LP: $[\frac{8}{3}, 3]^\top$ still not integral, so iterate: Gomory cut on fractional value $x_1^* = \frac{8}{3}$.

Second Gomory cut for Mitchell example

Now only $x_1^* = \frac{8}{3} = 2\frac{2}{3}$ is fractional. New Z:

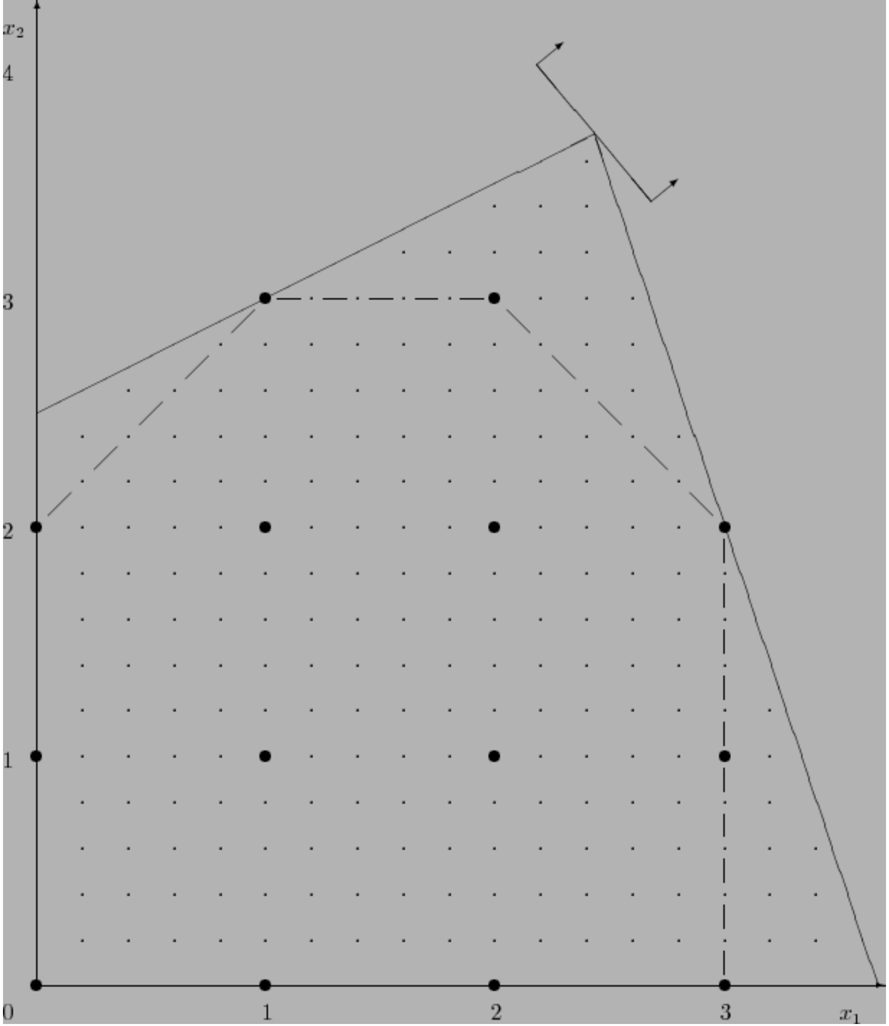
$$Z = \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} \\ 0 & 1 \\ \frac{1}{3} & -\frac{7}{3} \end{bmatrix} \text{ rounded down: } \lfloor Z \rfloor = \begin{bmatrix} 0 & -1 \\ 0 & 1 \\ 0 & -3 \end{bmatrix}.$$

Gomory cut added: $x_1 + (-1) * s_3 = x_1 + x_2 - 3 \leq 2$ or $x_1 + x_2 \leq 5$, leads to ILP solution

$$[3, 2; 0, 4, 1, 0]^\top \quad \text{with slacks } s_1, \ldots, s_4\,,$$

already seen with Mitchell's suggestion.

No more branching needed here.



If the algorithm **splits on** $\mathbf{x_1}$, two new problems are obtained:

Overview

1. Growth transformations and convergence principles
2. Baum/Eagon growth transformations for posynomials
3. Convergence for Standard Quadratic Optimization Problems
4. Convexity and power method for posynomial optimization
5. Application: spectral hypergraph bounds for the clique number
6. Multi-Standard Quadratic Optimization Problems (multi-StQPs)
7. Variants, extensions and possible improvements

A familiar example ...

If $Q = [q_{ij}]$ is a symmetric $n \times n$ matrix and $\mathbf{x} \in \mathbb{R}^n$, the problem

$$\gamma_Q = \max\left\{\mathbf{x}^\top Q\mathbf{x} : \mathbf{x}^\top\mathbf{x} = 1\,,\ \mathbf{x} \in \mathbb{R}^n\right\}$$

is easy [Rayleigh]: solution $\mathbf{x} = \bar{\mathbf{x}}$, an eigenvector to $\lambda_{\max}(Q)$. If moreover $q_{ij} \geq 0$ for all i, j, then an eigenvector $\bar{\mathbf{x}} \in \mathbb{R}^n_+$ exists:

$$\max\left\{\mathbf{x}^\top Q\mathbf{x} : \mathbf{x}^\top\mathbf{x} = 1\,,\ \mathbf{x} \in \mathbb{R}^n_+\right\} = \gamma_Q \quad \text{[Perron/Frobenius].}$$

A convergent iterative method is the *power method* (assuming $\lambda_{\max}(Q)$ is simple, otherwise slight complications):

$$\mathbf{x}_t = \frac{1}{\|Q^t\mathbf{x}_0\|}\,Q^t\mathbf{x}_0 \to \bar{\mathbf{x}} \quad \text{as } t \to \infty\,.$$

... behaves nasty !?

For some instances Q, power method is **not** monotone:

$$\mathbf{x}_{t+1}{}^\top Q \mathbf{x}_{t+1} \not\geq \mathbf{x}_t{}^\top Q \mathbf{x}_t \,,$$

e.g. for $n = 4$,

$$Q = \begin{bmatrix} 7 & 0 & 0 & 0 \\ 0 & 0 & 5 & 5 \\ 0 & 5 & 0 & 0 \\ 0 & 5 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_t = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} ,$$

we get

$$\mathbf{x}_t{}^\top Q^3 \mathbf{x}_t - (\mathbf{x}_t{}^\top Q^2 \mathbf{x}_t)(\mathbf{x}_t{}^\top Q \mathbf{x}_t) = 1343 * 4 - 199 * 27 = -1 \ldots$$

found after some time trying to prove monotonicity :-(

Maximization of posynomials under ℓ^p-constraints

Typical problem: for $\mathbf{x} \in \mathbb{R}^n$ let $\|\mathbf{x}\|_p^p = \sum_{i=1}^n |x_i|^p$ and put

$$S_+^p = \left\{\mathbf{x} \in \mathbb{R}_+^n : \|\mathbf{x}\|_p^p = 1\right\} .$$

Generalized monomial: $\mathbf{x}^{\mathbf{k}} = \prod_{i=1}^n x_i^{k_i}$ with real exponents $k_i \geq 0$.

Posynomial is linear combination of such $\mathbf{x}^{\mathbf{k}}$ with $\alpha_{\mathbf{k}} > 0$:

$$f(\mathbf{x}) = \sum_{\mathbf{k} \in K} \alpha_{\mathbf{k}} \mathbf{x}^{\mathbf{k}} = \sum_{\mathbf{k} \in K} \alpha_{\mathbf{k}} \prod_{i=1}^n x_i^{k_i} ,$$

where $K \subset \mathbb{R}_+^n$ is a finite set of exponent vectors $\mathbf{k} = [k_i]_i$.

Degree of f is $d(f) = \max\left\{\sum_{i=1}^n k_i : \mathbf{k} \in K\right\}$.

Optimization problem:

$$\max\left\{f(\mathbf{x}) : \mathbf{x} \in S_+^p\right\} .$$

(Homotopic) Growth Transformations for maximization

Continuous function $f : \Lambda \to \mathbb{R}$ on compact set $\Lambda \subset \mathbb{R}^n$,

consider *growth transformation* (GT) $T : \Lambda \to \Lambda$:

$$f(T(\mathbf{x})) > f(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \Lambda \quad \text{unless} \quad T(\mathbf{x}) = \mathbf{x}\,.$$

Examples: any ascent direction-based (exact) line search;
f convex: even inexact line search;
f concave: T is even *homotopic* GT:

$$f((1-\lambda)\mathbf{x} + \lambda T(\mathbf{x})) > f(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \Lambda \text{ and } 0 \le \lambda \le 1\,,$$

unless $\mathbf{x} \in \mathrm{Fix}\,(T) = \{\mathbf{x} \in \Lambda : T(\mathbf{x}) = \mathbf{x}\}$.

For general f, first-maximum line search yields homotopic GT.

Principle of line search

Unconstrained case $\min\{f(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^n\}$.

Repeat until stop criterion is met

1) at an iterate $\mathbf{x}^{old}$, decide a search direction $\mathbf{d} \in \mathbb{R}^n$;
2) search for optimal step length $t^* > 0$ by looking at $\min\left\{f(\mathbf{x}^{old} + t\mathbf{d}) : t \geq 0\right\}$ — one-dimensional search;
3) update $\mathbf{x}^{new} = \mathbf{x}^{old} + t^*\,\mathbf{d}$ if $f(\mathbf{x}^{new}) < f(\mathbf{x}^{old})$.

output last $\mathbf{x}^{new}$ as (tentatively optimal) solution.

Constrained case: similar, restrictions on $\mathbf{d}$ and step size t.

Convergence principles for Growth Transformations (1)

Choose $\mathbf{x}_0 \in \Lambda$ and generate iteratively sequences $(\mathbf{x}_t)_{t=1}^{\infty}$

$$\mathbf{x}_{t+1} = T(\mathbf{x}_t)\,, \quad t = 1, 2, \ldots .$$

gives increasing objective values $f(\mathbf{x}_t) \nearrow \gamma(\mathbf{x}_0) = \sup_t f(\mathbf{x}_t)$.

Step size is $s_t(\mathbf{x}_0) = \|\mathbf{x}_{t+1} - \mathbf{x}_t\|$, the set of accumulation points

$$\mathrm{Acc}\,(\mathbf{x}_0) = \left\{ z \in \Lambda : \mathbf{x}_{t_j} \to \mathbf{z} \text{ along some subsequence } t_j \nearrow \infty \right\}.$$

Theorem [Baum, Sell '67]: Let T be a GT for f over Λ.

(1) $\mathbf{x}$ is isolated in $\mathrm{Fix}\,(T)$ and a strict local maximizer of f if $\mathbf{x}$ local attractor under T; converse true if T is homotopic GT.

(2) T continuous $\Rightarrow$ stepsize $s_t(\mathbf{x}_0) \to 0$ and $\mathrm{Acc}\,(\mathbf{x}_0) \subseteq \mathrm{Fix}\,(T)$.

(3) If stepsize $s_t(\mathbf{x}_0) \to 0$ and $\mathrm{Acc}\,(\mathbf{x}_0) \subseteq \mathrm{Fix}\,(T)$, then $\mathrm{Acc}\,(\mathbf{x}_0)$ is *connected* subset of level set $\{\mathbf{x} \in \Lambda : f(\mathbf{x}) = \gamma(\mathbf{x}_0)\}$.

Convergence principles for Growth Transformations (2)

Corollary [by now trivial]: Let T be a continuous GT for f.

(a) If $\mathrm{Acc}\,(\mathbf{x}_0)$ is finite, then $\mathbf{x}_t$ converges,
most probably towards a strict local maximizer of f.

(b) If $\mathrm{Fix}\,(T)$ is finite, then all trajectories $\mathbf{x}_t$ converge, and
almost all such trajectories to strict local maximizers of f.

Remark: not every GT is continuous, classical example: Rosen's projected gradient.

Recent example: Infection-Immunization Dynamics for StQPs [Rota Bulò, B. '09]: discontinuous, but convergent (even in finite time plus solving a final linear system).

Baum-Eagon transformations for posynomials (1)

... formulated in [Baum, Eagon '67] for positive polynomials over $\Lambda = \bigotimes_{i=1}^{m} \Delta_i$ where $\Delta_i = S_+^1 \subset \mathbb{R}^{n_i}$ are standard simplices.

For simplicity just case $m = 1$, i.e., $\Lambda = \left\{\mathbf{x} \in \mathbb{R}^n_+ : \mathbf{e}^\top \mathbf{x} = \sum_i x_i = 1\right\}$.

Consider posynomial $f(\mathbf{x})$ and put $X = \text{diag}\ (\mathbf{x})$.

Theorem [Baum et al. '67]: Assume $\mathbf{x}^\top \nabla f(\mathbf{x}) > 0$ over S_+^1.

Then the transformation

$$T(\mathbf{x}) = \frac{1}{\mathbf{x}^\top \nabla f(\mathbf{x})} X \nabla f(\mathbf{x})$$

is a continuous homotopic GT for f over S_+^1.

Proof for homogeneous posynomials (i.e. $\mathbf{e}^\top \mathbf{k} = d(f)$, all $\mathbf{k} \in K$)

Baum-Eagon transformations for posynomials (2)

Proof ... uses inequality between geometric/arithmetic mean; Hölder's inequality; Euler's identity

$$\mathbf{x}^\top g(\mathbf{x}) = d(f) f(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \mathbb{R}^n\,;$$

as well as the inequality implication for positive a_j, b_j, c_j

$$\sum_j \frac{a_j}{b_j} \le \frac{1}{P} \ \text{ and } \ \sum_j \frac{a_j}{c_j} \le \frac{1}{Q} \quad \Rightarrow \quad \sum_j \frac{a_j}{b_j + c_j} \le \frac{1}{P+Q}\,.$$

In general, homogenize f by replacing $\mathbf{x}^\mathbf{k}$ with $\mathbf{x}^\mathbf{k}(\mathbf{e}^\top \mathbf{x})^{d(f) - \mathbf{e}^\top \mathbf{k}}$. Nowhere needed: integrality of exponents k_i. Therefore true not only for polynomials but also for posynomials.

Corollary [repeat from above]:
If $\mathrm{Acc}\,(\mathbf{x}_0)$ is finite, then $\mathbf{x}_t$ converges, most probably towards a strict local maximizer of f over S^1_+.

Baum-Eagon transformations for polynomials

Observation: we even can relax positivity of $\alpha_{\mathbf{k}}$ if all k_i are integer: just add monomials $2|\alpha_{\mathbf{k}}|(\mathbf{e}^\top\mathbf{x})^{\mathbf{e}^\top\mathbf{k}}$ to f and subtract $2\sum_{\mathbf{k}\in K}|\alpha_{\mathbf{k}}|$ after optimization.

Since $2|\alpha_{\mathbf{k}}|(\mathbf{e}^\top\mathbf{x})^{\mathbf{e}^\top\mathbf{k}} = 2|\alpha_{\mathbf{k}}|\mathbf{x}^{\mathbf{k}} + q_{\mathbf{k}}(\mathbf{x})$ with $q_{\mathbf{k}}$ posynomial, we get

$$\begin{aligned} f(\mathbf{x}) + 2\sum_{\mathbf{k}\in K}|\alpha_{\mathbf{k}}| &= f(\mathbf{x}) + \sum_{\mathbf{k}\in K} 2|\alpha_{\mathbf{k}}|(\mathbf{e}^\top\mathbf{x})^{\mathbf{e}^\top\mathbf{k}} \\ &= \sum_{\mathbf{k}\in K}[\alpha_{\mathbf{k}} + 2|\alpha_{\mathbf{k}}|]\mathbf{x}^{\mathbf{k}} + \sum_{\mathbf{k}\in K} q_{\mathbf{k}}(\mathbf{x}) \end{aligned}$$

for all $\mathbf{x}\in S_+^1$, and coefficients of $\mathbf{x}^{\mathbf{k}}$ are now positive.

Seems to work only on S_+^1 and with polynomials.

Convergence for Standard Quadratic Problems

Special case of above: $f(\mathbf{x}) = \mathbf{x}^\top Q\mathbf{x}$, Q a symmetric $n \times n$ matrix.

Here convergence holds even for infinite Acc $(\mathbf{x}_0)$:

Theorem [Lyubich et al., '78]: Any trajectory converges, i.e., for all $\mathbf{x}_0 \in S^1_+$ there is a $\mathbf{z}(\mathbf{x}_0)$ such that Acc $(\mathbf{x}_0) = \{\mathbf{z}(\mathbf{x}_0)\}$.

Further, convergence speed is sublinear in the worst case:

$$\|\mathbf{x}_t - \mathbf{z}(\mathbf{x}_0)\| \leq Ct^{-1/2} \qquad \text{for all } t\,,$$

where $C = C(Q) > 0$. To be more precise, convergence is linear, i.e., $\|\mathbf{x}_t - \mathbf{z}(\mathbf{x}_0)\| \leq C \exp(-t)$ for all t, if and only if

$$(Qz)_i \neq \mathbf{z}^\top Q\mathbf{z} \quad \text{whenever} \quad z_i = 0 \qquad \text{for } \mathbf{z} = \mathbf{z}(\mathbf{x}_0)\,.$$

If the latter condition holds for all $z \in$ Fix (T), then Fix (T) is finite and convergence for all trajectories is easy from above.

Convexity transformations for posynomial problems (1)

Back to optimization problem studied in [Baratchart et al.'98]:

$$\max\left\{f(\mathbf{x}) : \mathbf{x} \in S_+^p\right\},$$

where f is a posynomial of degree $d(f) \leq p$. Given $p > 0$, we always can homeomorphically transform the problem into

$$\max\left\{f_p(\mathbf{y}) : \mathbf{y} \in S_+^1\right\},$$

where $y_i = [\phi_p(\mathbf{x})]_i = x_i^p$ and $f_p = f \circ \phi_p^{-1}$ is again a posynomial, now of degree $d(f_p) \leq 1$. Therefore

Lemma [Baratchart et al.'98]:
f_p is concave if f is a posynomial of degree $d(f) \leq p$.

Convexity transformations for posynomial problems (2)

Proof: If $x_i > 0$ for all i, then also $y_i > 0$ for all i, and X and $Y = \text{Diag}\ (\mathbf{y})$ are nonsingular. Then $\nabla f_p(\mathbf{y}) = \frac{1}{p} Y^{\frac{1}{p}-1} \nabla f(\phi_p^{-1}(\mathbf{y}))$ and

$$Y^{-1} D^2 f_p(\mathbf{y}) Y = \frac{1}{p^2} \left[X\, D^2 f(\mathbf{x}) - (p-1) \text{Diag}\ (\nabla f(\mathbf{x})) \right].$$

Now use Gershgorin's disc theorem and positivity of $\alpha_{\mathbf{k}}$, as well as Euler's subhomogeneity inequality for $\frac{\partial f}{\partial x_i}(\mathbf{x})$:

$$\sum_{j=1}^{n} x_j \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) \le (p-1) \frac{\partial f}{\partial x_i}(\mathbf{x}) \quad \text{for all } i\,,$$

to get negative-semidefiniteness of $D^2 f_p$.

Generalized power method for posynomial problems

Suggested in [Baratchart et al.'98]; abbreviate gradient map by $g(\mathbf{x}) = \nabla f(\mathbf{x})$ and put

$$T(\mathbf{x}) = \frac{1}{\|\phi_{p-1}^{-1}[g(\mathbf{x})]\|}\,\phi_{p-1}^{-1}[g(\mathbf{x})]\,, \quad \mathbf{x} \in S_+^p\,.$$

Case $p = 2$ gives for $f(\mathbf{x}) = \mathbf{x}^\top Q\mathbf{x}$ the usual power method $T(\mathbf{x}) = \frac{1}{\|Q\mathbf{x}\|} Q\mathbf{x}$ which converges to the Perron/Frobenius vector of positive matrix Q, if spectral radius $\rho(Q)$ is simple eigenvalue.

Empirically good [Baratchart et al. '98/Rota Bulò, Pelillo '09].

Attention: T is in general not a growth transformation for f !

Wilf's spectral bound for the clique number

Undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $\#\mathcal{V} = n$ vertices and $\mathcal{E} \subseteq \binom{\mathcal{V}}{2}$.

Clique $S \subseteq \mathcal{V}$ is *maximal* if S is not contained in a larger clique.

Clique S^* is a *maximum* clique if $|S^*| = \max\{|T| : T \text{ clique in } \mathcal{G}\}$.

Finding the *clique number* $\omega(\mathcal{G}) = |S^*|$ is an NP-complete combinatorial optimization problem.

Theorem [Motzkin, Straus '65]: For adjac. matrix $A_{\mathcal{G}}$ of $\mathcal{G}$

$$\max\left\{\mathbf{x}^\top A_{\mathcal{G}}\mathbf{x} : \mathbf{x} \in S_+^1\right\} = 1 - 1/\omega(\mathcal{G})\,.$$

Wilf's ['67] bound: $\omega(\mathcal{G}) \leq \rho(A_{\mathcal{G}}) + 1$ with spectral radius of $A_{\mathcal{G}}$.

The k-clique hypergraph and the clique number

A clique $T = \{v_i\}$ with $|T| = k+1 \leq \omega(\mathcal{G})$ generates $k+1$ cliques $S_i = T \setminus \{v_i\}$ with $|S_i| = k$; shrink S_i to vertices s_i of $k+1$-*hypergraph*: with $\mathcal{W}$ all shrunk k-cliques, put

$\mathcal{H}_k = (\mathcal{W}, \mathcal{F})$ with *hyperedge* $F = F_T = \{s_i : v_i \in T\} \in \mathcal{F} \subseteq \binom{\mathcal{W}}{k+1}$.

Theorem [Sós, Straus '82]: Lagrangian polyn. of $\mathcal{H} = (\mathcal{W}, \mathcal{F})$

$$L_{\mathcal{H}}(\mathbf{x}) = \sum_{F \in \mathcal{F}} \prod_{i \in F} x_i \,, \quad \mathbf{x} \in \mathbb{R}^{|\mathcal{W}|} \,.$$

Then $\omega = \omega(\mathcal{G})$ encoded in maximum of $L_{\mathcal{H}_k}$ over S_+^k:

$$\max \left\{ L_{\mathcal{H}_k}(\mathbf{x}) : \mathbf{x} \in S_+^k \right\} = \binom{\omega}{k+1} \Big/ \binom{\omega}{k}^{(k+1)/k} .$$

$k = 1$: Motzkin/Straus; hard for all k as $d = k+1 > k = p$.

Spectral hypergraph bounds for the clique number

$L_{\mathcal{H}_k}$ is given by supersymmetric n^{k+1}-tensor $A_{\mathcal{H}_k}$, the *adjacency tensor*. Now use Qi's ['05] definition of H-eigenvalue to define

$$\rho(A_{\mathcal{H}_k}) = (k+1)!\,\mathsf{max}\left\{L_{\mathcal{H}_k}(\mathbf{x}) : \mathbf{x} \in S^{k+1}\right\}.$$

Since $A_{\mathcal{H}_k} \geq 0$, generalized Perron-Frobenius holds:

$$\rho(A_{\mathcal{H}_k}) = (k+1)!\,\mathsf{max}\left\{L_{\mathcal{H}_k}(\mathbf{x}) : \mathbf{x} \in S_{+}^{k+1}\right\},$$

which is a concave maximization problem by above ($d = p$).

Hence Sós-Straus theorem yields

Theorem [Rota Bulò, Pelillo '09]: For any $k \leq \omega(\mathcal{G})$ we have

$$\omega(\mathcal{G}) \leq \frac{\rho(A_{\mathcal{H}_k})}{k!} + k\,.$$

Multi-Standard Quadratic Optimization Problems

... are of the above form with max $\left\{ f(\mathbf{x}) = \mathbf{x}^\top Q \mathbf{x} : \mathbf{x} \in \Lambda \right\}$ where

$$\Lambda = \left\{ \mathbf{x} = (\mathbf{x}^i)_{i=1}^m \in \mathbb{R}^n : \mathbf{x}^i \in \Delta_i \,,\ i \in [1:m] \right\}$$

with $\Delta_i = S_+^1 \subset \mathbb{R}_+^{n_i}$ and $n = \sum_{i=1}^m n_i$.

Block structure $Q = [Q_{ij}]$ according to $\mathbf{x} = (\mathbf{x}^i)_{i=1}^m$.

Then

$$Q\mathbf{x} = \begin{bmatrix} Q_{11} & \dots & Q_{1m} \\ \vdots & \ddots & \vdots \\ Q_{m1} & \dots & Q_{mm} \end{bmatrix} \begin{bmatrix} \mathbf{x}^1 \\ \vdots \\ \mathbf{x}^m \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^m Q_{1j} \mathbf{x}^j \\ \vdots \\ \sum_{j=1}^m Q_{mj} \mathbf{x}^j \end{bmatrix} = \begin{bmatrix} (Q\mathbf{x})^1 \\ \vdots \\ (Q\mathbf{x})^m \end{bmatrix}.$$

Copositive programming approaches to MStQP [Burer '08] may exhibit complications in dual attainability [Schachinger/B. '09].

Baum/Eagon iteration for multi-StQPs

Suppose Q has no negative entries and a positive diagonal;
Baum-Eagon growth transformation on Λ:

$$T(\mathbf{x})^i_k = x^i_k\, F^i_k\,, \quad \text{all } k, i\,,$$

with F^i_k enhancing/inhibiting factors such that $T^i(\mathbf{x}) \in \Delta_i$ still:

$F^i_k > 1 \quad \Rightarrow \quad T(\mathbf{x})^i_k > x^i_k$, frequency increases with t,

while balance is kept: $\sum_k x^i_k = 1 \;\; \Rightarrow \;\; \sum_k T(\mathbf{x})^i_k = 1$, all i.

'Steepest ascent': $F^i_k = \alpha^i g^i_k$, with $\mathbf{g}^i$ from gradient $(Q\mathbf{x})^i$.

Keep balance: $\alpha^i = \left[\sum_k x^i_k g^i_k\right]^{-1}$.

Gradient variants for Baum/Eagon dynamics

Synchronous/simultaneous update: all $\mathbf{g}^i = (Q\mathbf{x})^i$.

If $\Delta_i = \Delta$, dynamics is relaxation labelling [Hummel, Zucker '83].

Asynchronous/sequential update: use block structure;

first $\mathbf{g}^1 = (Q\mathbf{x})^1$ as above, update $T(\mathbf{x})^1$;

for $\mathbf{g}^2$ use already updated $T(\mathbf{x})^1$ in gradient etc. à la Gauss/Seidel.

General principle:

$$\mathbf{x}_{t+1} = T(\mathbf{x}_t) \quad \text{with} \quad T = T_m \circ \cdots \circ T_1$$

where map $T_i : \Lambda \to \Lambda$ only updates i-th block of $\mathbf{x} = (\mathbf{x}^i)_{i=1}^m$.

Two variants for sequential update: symmetric and asymmetric.
Asymmetric version straightforward but more difficult to analyze.

Details for sequential variants $T_m \circ \cdots \circ T_1$

Asymmetric updates: map $\widetilde{T}_i : \Lambda \to \Lambda$ with

$$(\widetilde{T}_i\,\mathbf{x})^i = \frac{X^i \widetilde{Q}_i(\mathbf{x})\mathbf{x}^i}{(\mathbf{x}^i)^\top \widetilde{Q}_i(\mathbf{x})\mathbf{x}^i}, \quad \text{while} \quad (\widetilde{T}_i\,\mathbf{x})^j = \mathbf{x}^j \text{ if } j \neq i\,,$$

where $\widetilde{Q}_i(\mathbf{x})$ are $n_i \times n_i$-matrices with $\widetilde{Q}_i(\mathbf{x})\mathbf{x}^i = (Q\mathbf{x})^i$:

$$\widetilde{Q}_i(\mathbf{x}) = Q_{ii} + \sum_{j \neq i} (Q_{ij}\mathbf{x}^j)[1, \ldots, 1] \neq \left[\widetilde{Q}_i(\mathbf{x})\right]^\top .$$

Symmetric updates: map $T_i : \Lambda \to \Lambda$ with

$$(T_i\,\mathbf{x})^i = \frac{X^i Q_i(\mathbf{x})\mathbf{x}^i}{(\mathbf{x}^i)^\top Q_i(\mathbf{x})\mathbf{x}^i} \quad \text{while} \quad (T_i\,\mathbf{x})^j = \mathbf{x}^j \text{ if } j \neq i\,,$$

where $Q_i(\mathbf{x}) = \widetilde{Q}_i(\mathbf{x}) + \left[\widetilde{Q}_i(\mathbf{x})\right]^\top - Q_{ii}$ ... symmetric $n_i \times n_i$, so the Baum/Eagon theorem applies to T_i and thus to T.

Regularity and genericity

Consider $m \times n$ matrix such that $\Lambda = \left\{ \mathbf{x} \in \mathbb{R}^n_+ : A\mathbf{x} = \mathbf{e} \right\}$:

$$A = \begin{bmatrix} 1...1 & 0...0 & \cdots & 0...0 \\ 0...0 & 1...1 & \cdots & 0...0 \\ \vdots & & \ddots & \vdots \\ 0...0 & 0...0 & \cdots & 1...1 \end{bmatrix} \quad \text{and} \quad \mathbf{e} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^m .$$

If Λ_I denotes a face of Λ, denote by Q_I restriction of Q to Λ_I; similarly, A_I is obtained from A by dropping columns not in I.

Theorem [B., Schachinger '09]: Abbreviate by $\mathcal{F} = \mathrm{Fix}\,(T)$.

If $\det Q_I \neq 0$ and $\det(A_I Q_I^{-1} A_I^\top) \neq 0$, then $|\mathcal{F} \cap \ \mathrm{relint}\,\Lambda_I| \leq 1$; if above holds for all I, then $\mathcal{F}$ is finite.

... satisfied for all matrices Q in an open dense subset of the set of all symmetric $n \times n$-matrices. Hence regularity is generic.

Variants of Baum/Eagon; applications of MStQP

Alert: possibly different basins of attraction in three variants !

Example: Let $Q_{ii} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $Q_{ij} = \begin{bmatrix} 0 & 10 \\ 10 & 0 \end{bmatrix}$ for $i \neq j$, to build $Q_0 = [Q_{ij}]$ for MStQP on $\Lambda = (S^1_+)^3 \subseteq \mathbb{R}^6$. This has 6 global solutions; pick one $\mathbf{x}^*$, perturb Q_0 to $Q^* = Q_0 + \varepsilon\,\mathrm{Diag}\,\mathbf{x}^*$. Then $\mathbf{x}^*$ is the unique global solution of MStQP with Q^*. But we end up in different points if starting from $\mathbf{x}_0 \approx (0,1,0,1,0,1)$:

simultaneous: switch in last block, $f : 3 \nearrow 43$, $(0,1,0,1,1,0)$;

asymmetric sequential: switch in middle block, $(0,1,1,0,0,1)$;

symmetric sequential: switch in last two blocks, $(0,1,1,0,1,0)$.

By perturbation, any of three variants may miss global solution !

Applications: supervised and unsupervised learning, relaxation labelling, pattern recognition, image processing, game theory ...

Beyond QP: first-order method with inexact line search

Interpret increment = step-size α_t^i × search direction $\mathbf{d}_t$:

$$T(\mathbf{x}_t)_k^i - (\mathbf{x}_t)_k^i = \alpha_t^i \, (d_t)_k^i \, , \quad \text{all } k, i \, ,$$

with $\alpha_t^i = \left[\sum_k (x_t)_k^i (g_t)_k^i\right]^{-1}$, a scaling factor constant across k, and $(d_t)_k^i = (\mathbf{x}_t)_k^i \left((g_t)_k^i - \frac{1}{\alpha_t^i} \right)$, with $(g_t)_k^i$ from gradient $\nabla f(\mathbf{x}_t)$.

This extends to general non-QP problem $\quad \max \{ f(\mathbf{x}) : \mathbf{x} \in \Lambda \}$.
Instead of above α^i get step-size by (inexact) Armijo line search: for largest feasible step size $\alpha_t^0 > 0$ choose constants $\beta, \sigma < 1$. Then used step size α_t is largest $\alpha \in \{\beta^k \alpha_t^0 : k \in \mathbb{N}_0\}$ satisfying

$$f\,(\mathbf{x}_t + \alpha \mathbf{d}_t) \geq f\,(\mathbf{x}_t) + \sigma \alpha \, \mathbf{g}_t{}^\top \mathbf{d}_t \, .$$

Parametric deflection of the gradient

Instead of $d_k^i = x_k^i \left(g_k^i - \frac{1}{\alpha^i} \right)$ now further deflection of gradient:

let $\Lambda = \left\{ \mathbf{x} \in \mathbb{R}^n_+ : A\mathbf{x} = e \right\}$, choose parameter $\gamma > 0$ and put

$$(\mathbf{d}_t)_k^i = [(\mathbf{x}_t)_k^i]^{2\gamma} \left(P_\gamma(\mathbf{x}_t)\mathbf{g}_t{}^i \right)_k , \quad \text{all } k, i ,$$

where [Tseng et al. 09]

$$P_\gamma(\mathbf{x}_t) = I - A^\top (A(X_t)^{2\gamma} A^\top)^{-1} A(X_t)^{2\gamma} .$$

For $\gamma = \frac{1}{2}$: Armijo search unstable, replicator dynamics better.

But $\gamma \approx 1$ seems to work well (inexact line search, even StQP).

Similar theory (conv.rates), good experience ($n = 1000$).

Selected references (1)

[Baratchart/Berthod/Pottier '98] Optimization of positive generalized polynomials under ℓ^p constraints, Journal of Convex Analysis **5**, 353–379.

[Baum/Eagon '67] An inequality with applications to statistical prediction for function of Markov processes and to a model of ecology, Bulletin of the American Mathematical Society **73**, 360–363.

[Baum/Sell '67] Growth transformations for functions on manifolds, Pacific Journal of Mathematics **27**, 211–227.

[B./Schachinger '10] Multi-standard quadratic optimization problems, Computational Optimization and Applications **45**, 237–256.

[B./Rinaldi/Rota Buló '19] Infection and immunization for partnership games: a correction, Games Economic Behaviour **114**, 315–317 (2019).

[B./Mertikopoulos/Schachinger/Staudigl '19] A Hessian barrier algorithm for linearly constrained optimization problems, *SIAM Journal on Optimization* **29**, 2100-2127.

Selected references (2)

[Burer '09] On the copositive representation of binary and continuous non-convex quadratic programs, Mathematical Programming **120**, 479-495.

[Hummel/Zucker '83] On the foundations of relaxation labeling processes, IEEE Trans. Pattern Analysis and Machine Intelligence **5**, 267-287.

[Lyubich/Maistrovskii/Ol'chovskii '78] Selection-induced convergence to equilibrium in a single-locus autosomal population,
Problems of Information Transmission **16**, 93–104.

[Motzkin/Straus '65] Maxima for graphs and a new proof of a theorem of Turán, Canadian Journal of Mathematics **17**, 533-540.

[Qi '05] Eigenvalues of a real supersymmetric tensor,
Journal of Symbolic Computing **40**, 1302-1324.

[Rota Buló/B. '11] Infection and immunization: a new class of evolutionary game dynamics, Games Economic Behaviour **71**, 193–211.

Selected references (3)

[Rota Buló/Pelillo '09] A Generalization of the Motzkin-Straus Theorem to hypergraphs, Optimization Letters **3**, 287–295.

[Schachinger/B. '09] A conic duality Frank–Wolfe type theorem via exact penalization in quadratic optimization,
Mathematics of Operations Research **34**, 83–91.

[Sós/Straus '82] Extrema of functions on graphs with applications to graphs and hypergraphs, Journal of Combinatorial Theory Ser. **B 63**, 189-207.

[Tseng/B./Schachinger '11] A first-order interior-point method for linearly constrained smooth optimization,
Mathematical Programming **127**, 399–424.

[Wilf '67] The eigenvalues of a graph and its chromatic number, Journal of the London Mathematical Society **42**, 330–332.