Hybrid Methods for global optimization of large-scale polynomials

J. Aspman [1] G. Bareilles [1] V. Kungurtsev [1] J. Mareček [1] M. Takáč [2]

[1]Czech Technical University [2]Mohamed bin Zayed Univ. of A.I.

Problem & contributions

$$f^\star = \min_x \; f(x) \quad \text{s.t.} \quad g_i(x) \geq 0, \quad i \in \mathcal{I}; \qquad g_j(x) = 0, \quad j \in \mathcal{E} \qquad (\mathcal{POP})$$

where $f, g_i, g_j : \mathbb{R}^n \to \mathbb{R}$ are polynomials.

▹ Problem

Find a *global* minimizer $x^\star$ of large-scale $(\mathcal{POP})$

▹ Contribution

- bring together tools from different math disciplines
- a *hybrid* algorithm
  1. a 1st order method on a convex relaxation (P_r)
  2. a theoretically grounded switch to Newton's method on $(\mathcal{POP})$

Difficulties & approach

The global minimizer of large $(\mathcal{POP})$ is provided by *very large* SDP relaxations (P_r).

▹ Interior Point Methods: go-to method for solving (P_r)

- ✓ fast convergence
- ✗ not applicable (too high memory cost)

▹ First-order methods on (P_r) — fig. 2

- ✓ low per-iteration cost
- ≈ slow convergence: provide rough solutions of (P_r)

▹ Newton on smooth polynomial equations — fig. 1

- ≈ poor convergence away from solution
- ✓ superfast local convergence near solutions

▹ Active constraint identification — fig. 3

- ✓ turns $(\mathcal{POP})$ into smooth poly. eq. near $x^\star$

Illustrative problem

$$\begin{aligned} \min_{x \in \mathbb{R}^2} \; & f(x) = -2.5x_1^2 + 3x_1x_2 - 2.5x_2^2 - 3x_1 + 5x_2 - 2.5 \\ \text{s.t.} \; & g_1(x) = -0.5x_1^3 + x_2 \geq 0 \\ & g_2(x) = -0.05x_1^2 - x_2 + 1.8 \geq 0 \\ & g_3(x) = -0.05x_2^2 + x_1 + 0.1x_2 + 0.35 \geq 0 \end{aligned} \qquad (1)$$

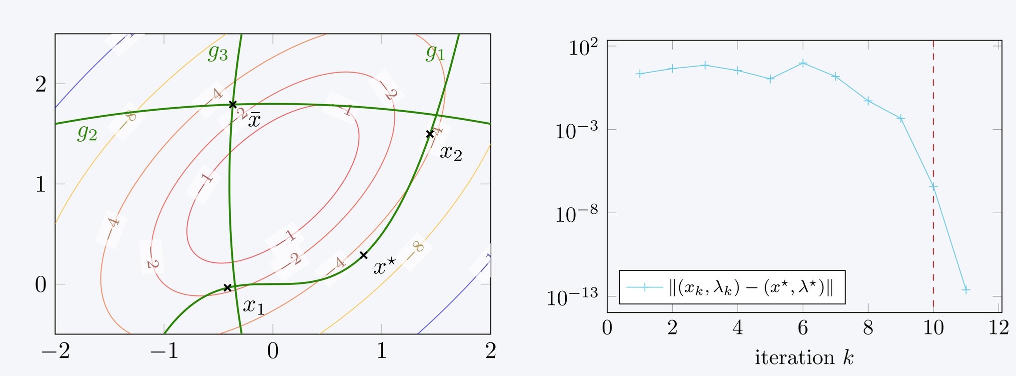

Figure 1. **Left pane**: problem (1). **Right pane**: Newton's method: slow away from $(x^\star, \lambda^\star)$, fast near $(x^\star, \lambda^\star)$.

1. Convex relaxations

TLDR: (P_r) is a sequence of increasingly large SDP problems, whose optimal value converge to $f^\star$ from below.

▹ Moment relaxation, order r

$$\rho_r = \inf_{y \in \mathbb{R}^{\mathbb{N}^n_{2r}}} L_y(f) \quad \text{s.t.} \quad M_r(y) \succeq 0; \quad M_{r-r_j}(g_j y) = 0, \; j \in \mathcal{E}; \quad M_{r-r_j}(g_j y) \succeq 0, \; j \in \mathcal{I}; \quad y_0 = 1 \qquad (P_r)$$

where $r_j = \lceil (\deg g_j)/2 \rceil$, y is a sequence indexed by exponents $\alpha \in \mathbb{N}^n_{2r} = \{\alpha \in \mathbb{N}^n : \sum \alpha_i \leq 2r\}$, $L_y : f = \sum_\alpha f_\alpha X^\alpha \mapsto \sum_\alpha f_\alpha y_\alpha$ is the **Riesz functional**, and $M_d(y)$ and $M_d(g_j y)$ are the **moment** and **localizing matrices**:

$$M_d(y)[\alpha, \beta] = L_y(X^\alpha X^\beta) = y_{\alpha+\beta}$$
$$M_d(g_j y)[\alpha, \beta] = L_y(g_j X^\alpha X^\beta) = \sum_\delta g_{j,\delta} y_{\alpha+\beta+\delta},$$

for all $\alpha, \beta \in \mathbb{N}^n_d$.

▹ **Property** [3]: ρ_r converges to $f^\star$ from below.

If (P_r) has a unique global minimizer $x^\star$, and y^r is a nearly optimal solution of (P_r), then

$$\lim_{r \to \infty} L_{y^r}(X_j) = x_j^\star$$

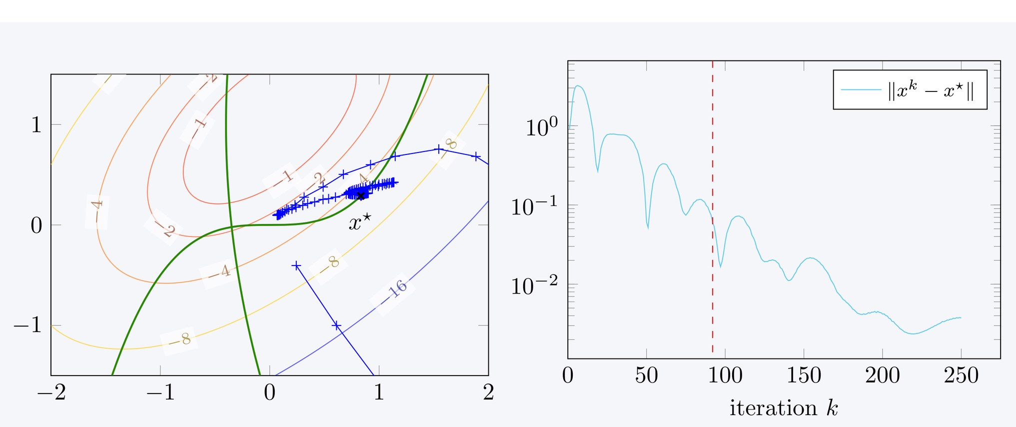

Figure 2. First-order method on the third relaxation (P_3) of (1). **Left pane**: points $x^k = (L_{y^k}(X_1), L_{y^k}(X_2)) \in \mathbb{R}^2$, where $y^k \in \mathbb{R}^{\mathbb{N}^2_6}$ is the k-th iterate of the 1st order method on (P_3). **Right pane**: distance between x^k and the global minimizer $x^\star$. The dashed red line indicates the first time the correct active set is detected.

2. Identifying active constraints

TLDR: at $x^\star$, some constraints are null and others positive.

$$\mathcal{A}^\star = \{i \in \mathcal{I} : g_i(x^\star) = 0\}$$

can be detected from any point x near $x^\star$.

▹ The Lagrangian for $(\mathcal{POP})$ is

$$\mathcal{L}(x, \lambda) = f(x) - \lambda^\top g(x),$$

where $\lambda \in \mathbb{R}^{m_\mathcal{E}} \times \mathbb{R}^{m_\mathcal{I}}_+$ is the Lagrange multiplier vector.

▹ The first-order "KKT" necessary optimality conditions: there exist $\lambda^\star$ such that

$$\nabla_x \mathcal{L}(x^\star, \lambda^\star) = 0, \quad g_\mathcal{E}(x^\star) = 0, \quad 0 \leq g_\mathcal{I}(x^\star) \perp \lambda^\star_\mathcal{I} \geq 0 \qquad (2)$$

where $a \perp b$ means $a^\top b = 0$.

▹ Measure of the KKT residual at point (x, λ):

$$r(x, \lambda) = \|\nabla_x \mathcal{L}(x, \lambda)\| + \|g_\mathcal{E}(x)\| + \|[-g_\mathcal{I}(x)]_+\| + |\lambda_\mathcal{I}^T g_\mathcal{I}(x)| + \|[-\lambda_\mathcal{I}]_+\|$$

▹ From point x, the active set is chosen as

$$\omega(x) = \min_{\lambda \in \mathbb{R}^{m_\mathcal{E}} \times \mathbb{R}^{m_\mathcal{I}}_+} r(x, \lambda)$$
$$\mathcal{A}(x) = \{i : g_i(x) \leq -1/\log(\omega(x))\}$$

Theorem [1, Th. 3.2]: Consider $x^\star$ that satisfies KKT (2), a qualification condition MFCQ, and is an isolated critical point. Then, there exists $\epsilon > 0$ such that

$$\mathcal{A}(x) = \mathcal{A}^\star \quad \text{for all } x : \|x - x^\star\| < \epsilon.$$

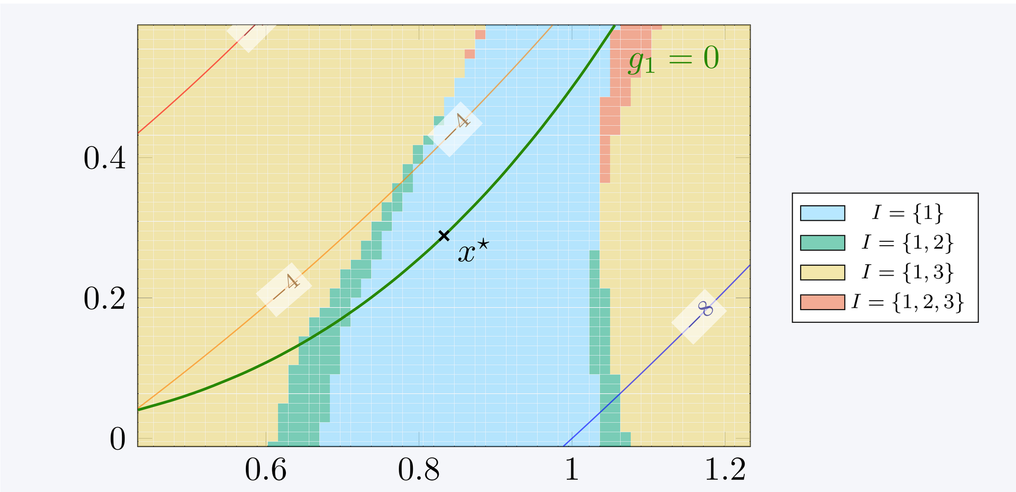

Figure 3. $\mathcal{A}(x) = \mathcal{A}^\star = \{1\}$ for x near $x^\star$, for problem (1).

3. Newton's method, α-β

TLDR: For a polynomial system $F : \mathbb{R}^p \to \mathbb{R}^p$ and $z_0 \in \mathbb{R}^p$, if $\hat\alpha(F, z_0)$ is small enough, there exists a nearby zero of F to which Newton's method

$$z_{i+1} = z_i - \mathrm{D}_F(z_i)^{-1} F(z_i)$$

converge quadratically; see [2].

▹ Some useful quantities

$$\hat\alpha(F, z) = \beta(F, z)\mu(F, z)\frac{1}{2}(\max_{i=1,\dots,p} d_i)^{3/2}\|z\|_\dagger^{-1}$$
$$\beta(F, z) = \left\|\mathrm{D}_F(z)^{-1}F(z)\right\|$$
$$\mu(F, z) = \max\{1, \|F\|_p \|D_F(z)^{-1}\Delta_{(d)}(z)\|\}$$

where $d_i = \deg F_i$, $\|\cdot\|_F$ is a norm on polynomials, and $\Delta_{(d)}(z) = \mathrm{Diag}(d_i^{1/2}\|z\|_\dagger^{d_i-1})$, and $\|z\|_\dagger = \sqrt{1 + \|z\|_2^2}$.

▹ **Proposition** [2]: Consider $F : \mathbb{R}^p \to \mathbb{R}^p$ polynomial and $z_0 \in \mathbb{R}^p$. If

$$\hat\alpha(F, z_0) \leq \frac{1}{4}(13 - 3\sqrt{17}) \approx 0.15, \qquad (5)$$

then z_0 is an *approximate zero*:

- $\mathrm{D}_F(z_i)$ are invertible *i.e.*, the sequence is well-defined,
- there exists $\bar z \in \mathbb{R}^p$ s.t. $F(\bar z) = 0$, $\|z_0 - \bar z\| \leq 2\beta(F, z_0)$,

$$\|z_i - \bar z\| \leq \left(\frac{1}{2}\right)^{2^i - 1} \|z_0 - \bar z\|. \qquad (6)$$

The hybrid method

▹ Algorithm outline

1. Partial solve of the convex relaxation (P_r) *e.g.*, with coord. descent on Burer-Monteiro.
2. Extract point x_k and compute its active set $\mathcal{A}_k \subseteq \mathcal{I}$.
3. Reduce the problem to

   $$\min_x f(x) \quad \text{s.t.} \quad g_i(x) = 0, \quad i \in \mathcal{E} \cup \mathcal{A}_k,$$

   and apply Newton to its optimality conditions (2), now *smooth* polynomial equations:

   $$F_{\mathcal{A}_k}(x, \lambda) = \begin{pmatrix} \nabla f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{A}_k} \lambda_i \nabla g_i(x) \\ g_{\mathcal{E} \cup \mathcal{A}_k}(x) \end{pmatrix} = 0,$$

   if the α-β test (5) is satisfied.

   Then, (6) ensures *superfast convergence*: finding $(\hat x, \hat\lambda)$ ϵ-away from $(x^\star, \lambda^\star)$ takes $\lceil 5.82 \rceil = 6$ iterations with $\varepsilon = 2.22 \cdot 10^{-16}$ and $\|z - z^\star\| = 10$.
4. Check global optimality with *e.g.*, the det. hierarchy. Further solve (P_r) or increase r if $\hat x$ is not a global minimizer.

Algorithm 1 Hybrid algorithm

Require: (f, g), y_r^0 initial point for (P_r)
for $k = 0, 1, 2, \dots$ **do**
  $y_r^k \leftarrow$ partial solution of (P_r) from point y_r^{k-1}
  $x^k \leftarrow$ extracted point from SDP iterate y_r^k
  $\mathcal{A}^k \leftarrow \mathcal{A}(x^k)$
  $\lambda^k \leftarrow \lambda_{\mathcal{A}^k}(x^k)$
  if $\hat\alpha\left(F_{\mathcal{A}^k}, (x^k, \lambda^k)\right) \leq 0.15$ **then**
    $\hat x^k, \hat\lambda^k \leftarrow$ limit of Newton's method on $F_{\mathcal{A}^k}$ from point (x^k, λ^k)
    if $(\hat x^k, \hat\lambda^k)$ satisfies a condition for global opt. **then**
      return $\hat x^k, \hat\lambda^k$
    else
      Further solve (P_r) or increase r

Theorem [1, Th. 5.1]: Consider $(\mathcal{POP})$ that admits a unique global minimizer $x^\star$. Then, algorithm 1 generates $(x_k, \mathcal{A}_k)$ such that

- (y_k) converges to the *global* minimizer $x^\star$.

If $x^\star$ is qualified (LICQ and second-order growth), then there exists $\hat k$ such that

- the minimizer active set is identified: $\mathcal{A}_{\hat k} = \mathcal{A}^\star$,
- Newton's method converges *quadratically* (6),
- and it returns the *globally* optimal primal-dual point $(\hat x^{\hat k}, \hat\lambda^{\hat k})$.

Illustrations

▹ Problem: optimal power flow industrial instance

▹ Plots

- fig. 4 shows coordinate descent on (P_r) then Newton's method on $(\mathcal{POP})$ after 2, 4, 8, ... epochs
- fig. 5 shows coordinate descent on (P_r), Newton's method on $(\mathcal{POP})$, and the hybrid method

▹ Observations

- Coordinate descent on its own can be slow — fig. 5.
- Newton's method converges fast eventually
- Newton's method on its own can be slow — fig. 4 (`CD 2` `CD 4` `CD 8`), and fig. 5 (`Newton`)
- The hybrid method, with the α-β test, improves upon coord. descent and Newton's method — fig. 5
- The limit point attained by a first-then-second order algorithm can vary substantially — fig. 4 right; indeed, Newton's method may be attracted to (first-order stationary) points that are not the global minimizer (see x_1, x_2 in fig. 1), and highlights the need for a global optimality check

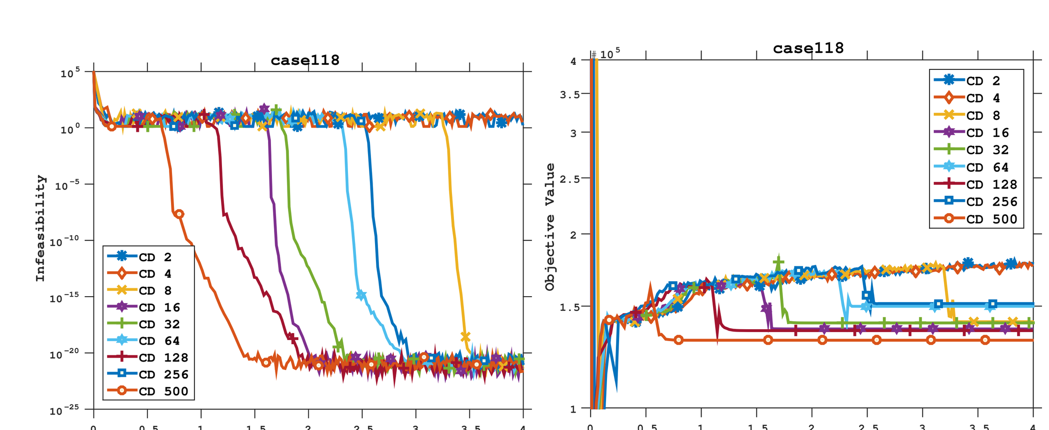

Figure 4. Infeasibility and objective function over time for coordinate descent on (P_r) then Newton's method on $(\mathcal{POP})$ after 2, 4, 8, ... epochs.

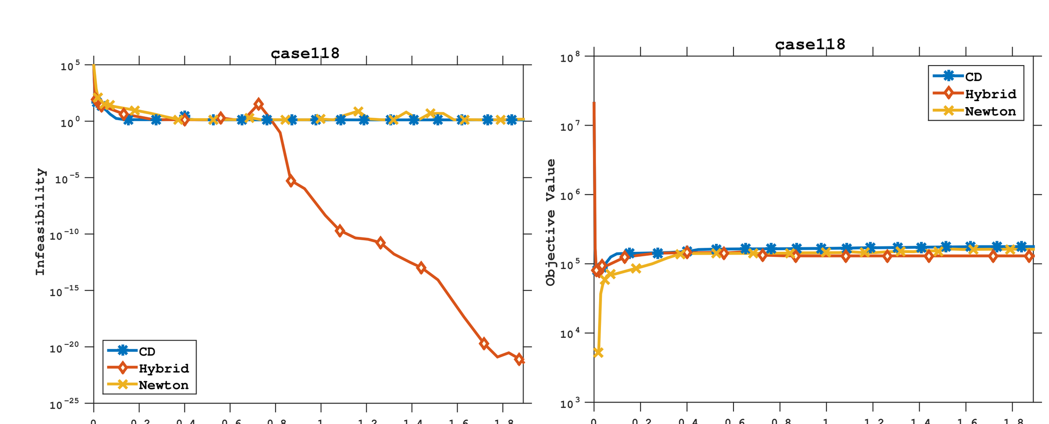

Figure 5. infeasibility and objective function over time for coordinate descent on (P_r), Newton's method on $(\mathcal{POP})$, and the hybrid method.

More details in: `arxiv.org/abs/2305.16122`

References

[1] Johannes Aspman, Gilles Bareilles, Vyacheslav Kungurtsev, Jakub Mareček, and Martin Takáč. Hybrid methods in polynomial optimisation, 2023.

[2] Felipe Cucker and Steve Smale. Complexity estimates depending on condition and round-off error. *Journal of the ACM*, 46(1):113–184, January 1999.

[3] Jean Bernard Lasserre. *An Introduction to Polynomial and Semi-Algebraic Optimization*. Cambridge University Press, first edition, February 2015.

[4] Christina Oberlin and Stephen J. Wright. Active Set Identification in Nonlinear Programming. *SIAM Journal on Optimization*, 17(2):577–605, January 2006.