# Warm-Starting for Sequences of MINLPs

Erik Tamm, PhD Advisor: Jan Kronqvist – KTH Royal Institute of Technology

Solving a Mixed-Integer Nonlinear Program is challenging. Solving a sequence of them is even harder. But if the difference between the problems is known and relatively small there is good hope to improve the performance using warm-starting.

**Contact**
Erik Tamm
etamm@kth.se

## Problem formulation

We consider problems on the form

$$\begin{aligned} \min\ & c^\intercal x + d^\intercal y \\ \text{s.t. } & f(x,y) \le \theta, \\ & g_\lambda(x,y) \le 0, \\ & h(x,y) \le 0, \\ & Ax + By \le b, \\ & x \in \mathbb{R}^n, y \in \mathbb{Z}^m. \end{aligned} \qquad (P_{\theta,\lambda})$$

We assume that $f$, $g_\lambda$ and $h$ are once continuously differentiable and convex functions, that $g_\lambda$ is a monotone function in $\lambda$ for any fixed $(x, y)$, and that the linear constraints defines a bounded set. We finally assume that some suitable constraint qualification holds. This is then a so-called convex Mixed-Integer Nonlinear Program (MINLP) which can be solved with Outer Approximation (OA). We specifically consider the version presented by [5].

The essence of OA is to find points $X^k = \{(x^1, y^1), \dots, (x^k, y^k)\}$ which generate an outer polyhedral approximation $\Omega(X^k, \theta, \lambda)$ described by

$$\begin{cases} f(x^l, y^l) + \langle \nabla f(x^l, y^l), z - z^l \rangle \le \theta, & \forall (x^l, y^l) \in X^k, \\ g_\lambda(x^l, y^l) + \langle \nabla g_\lambda(x^l, y^l), z - z^l \rangle \le 0, & \forall (x^l, y^l) \in X^k, \\ h(x^l, y^l) + \langle \nabla h(x^l, y^l), z - z^l \rangle \le 0, & \forall (x^l, y^l) \in X^k, \end{cases}$$

where $z - z^l = (x - x^l, y - y^l)^\intercal$.

## Example of Outer Approximation

Figure 1 illustrates how OA solves problem (1). The constraints in (1) are plotted in dark blue in Figure 1. Say that we start the algorithm in $P_1$. This generates the first cut, that is the line through $P_1$. If we then solve a MILP, we obtain $P_2$. Since $y = 9$ is never feasible, we obtain the corresponding point $F_2$. Based on this point we obtain the second cut, that is the line $y = 5.22$ in light blue. If we continue similarly we need four cuts to converge to the optimal solution which is $P_1$.

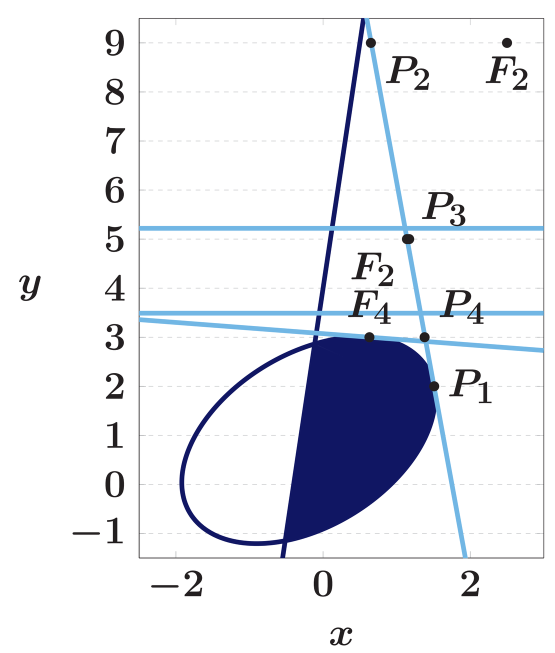



Figure 1: Illustration of Outer Approximation. $P_1$ is the starting point and the optimal solution. $P_2 - P_4$ are the points found by solving a MILP. $F_2 - F_4$ are the points found by solving an NLP.

$$\begin{aligned} \min\ & -2x - y \\ \text{s.t. } & 3x^2 + 2y^2 - 2xy \\ & +3x - 4y \le 5.3, \\ & -10x + y \le 4, \\ & x, y \in [-10, 10], \\ & y \in \mathbb{Z}. \end{aligned} \qquad (1)$$

## Proposed methods

We propose three methods for warm-starting. Note that the sequence of $\theta$ or $\lambda$ is chosen so that the feasible set is tightened for each problem, which is possible due to monotonicity.

- **Base method**: A standard technique is to start the next problem in the sequence with the optimal solution from the last problem.
- **Cut-tightening method**: Since each iteration tightens the feasible set of the problem, the points $X^k$ can be reused to derive a valid outer polyhedral approximation $\Omega(X^k, \theta, \lambda)$ by only updating the values of $\theta$ and $\lambda$.
- **Point-based method**: The integer combinations $y^i$ in $X^k$ will probably be good integer combinations for the next iteration. We can therefore find $\hat{x}^i$ for each $y^i$ by solving an NLP to form $\hat{X}^k$. These new points then define a valid outer polyhedral approximation $\Omega(\hat{X}^k, \theta, \lambda)$.

## Biobjective MINLPs

One application is to solve biobjective MINLPs. They can be solved using the $\varepsilon$-method to estimate the set of nondominated points [3]. The set of efficient solutions is not convex which means that the classic weighted-sum method is not applicable.

For the method, we solve $(P_\varepsilon)$ for a decreasing sequence of $\varepsilon$ where

$$\begin{aligned} \min\ & t \\ \text{s.t. } & f_1(x,y) \le t, \\ & f_2(x,y) \le \varepsilon, \\ & h(x,y) \le 0, \\ & Ax + By \le 0, \\ & x \in \mathbb{R}^n, y \in \mathbb{Z}^m, t \in \mathbb{R}. \end{aligned} \qquad (P_\varepsilon)$$

As an example, we consider problem (2) (Problem (TI4) from [4])

$$\begin{aligned} \min\ & \{x_1 + x_3 + y_1 + y_3, x_2 + x_4 + y_2 + y_4\} \\ \text{s.t. } & x_1^2 + x_2^2 \le 1, \\ & x_3^2 + x_4^2 \le 1, \\ & (y_1 - 2)^2 + (y_2 - 5)^2 \le 10, \\ & (y_3 - 3)^2 + (y_4 - 8)^2 \le 10, \\ & x \in [-20, 20]^4, y \in ([-20, 20]^4 \cap \mathbb{Z}^4). \end{aligned} \qquad (2)$$

## Regularized Sparse Linear Regression

In regularized sparse linear regression (SLR) [1], we consider the problem

$$\begin{aligned} \min\ & ||Ax - b||_2^2 + \lambda ||x||_2^2, \\ \text{s.t. } & ||x||_0 \le \kappa. \end{aligned}$$

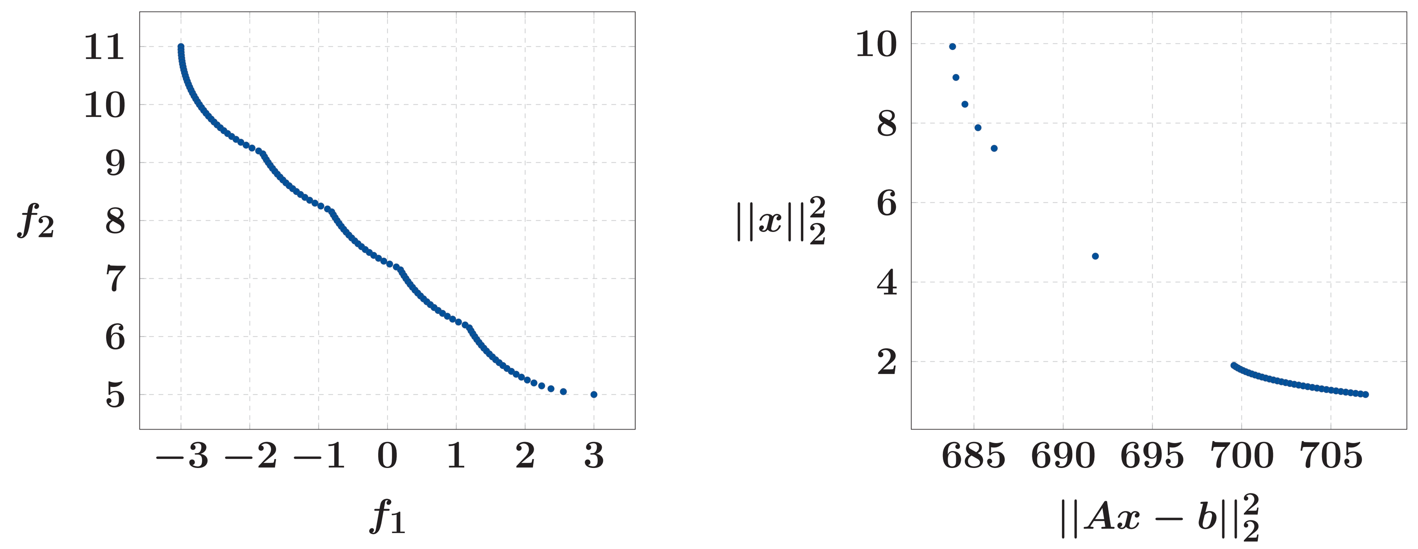
In practical application it is unclear which value of $\lambda$ to choose. If we can solve the problem for a sequence of values we can evaluate which value to choose. The problem can be formulated on the form of $(P_{\theta,\lambda})$ as

$$\begin{aligned} \min\ & t \\ \text{s.t. } & ||Ax - b||_2^2 + \lambda ||x||_2^2 \le t, \\ & l_i y_i \le x_i \le u_i y_i,\ i = 1, 2, \dots, n, \\ & \textstyle\sum_{i=1}^n y_i \le \kappa, \\ & y_i \in \{0, 1\},\ i = 1, 2, \dots, n. \end{aligned} \qquad (3)$$

## Numerical Results

The examples have been solved in Julia. The results for the biobjective problem (2) are presented in Figure 2a and Table 1.

The regularized sparse linear regression problem (3) was solved for a dataset on the quality of Portuguese red wine [2]. The data contains 11 features and 1599 data points. The problem was solved with $\kappa = 5$ and $0 \le \lambda \le 10$ with a step size of 0.5. The numerical results are presented in Figure 2b and Table 1.




(a) Efficient solutions of (2)

(b) Solutions of SLR with $\kappa = 5$ and $0 \le \lambda \le 10$.

Figure 2: Plots of the solutions to each problem.

| | | Base | Cut-tightening | Point-based |
|---|---|---|---|---|
| Biobjective | NLP | 1461 | 146 | 1766 |
| | MILP | 820 | 136 | 142 |
| SLR | NLP | 7395 | 258 | 7916 |
| | MILP | 7395 | 258 | 272 |

Table 1: Number of subproblems solved for each method and problem

## Acknowledgment

The ideas for this work were developed in collaboration with Prof. Gabriele Eichfelder, TU Ilmenau, Germany. This works has been funded by a grant from the Swedish Research Council (2022-03502).

## References

[1] D Bertsimas and B. Van Parys. “Sparse high-dimensional regression: Exact scalable algorithms and phase transitions”. In: *The Annals of Statistics* 48.1 (2020), pp. 300–323.

[2] P. Cortez et al. *Wine Quality*. UCI Machine Learning Repository. 2009. URL: https://doi.org/10.24432/C56S3T.

[3] G Eichfelder. “Twenty years of continuous multiobjective optimization in the twenty-first century”. In: *EJCO* 9 (2021), p. 100014.

[4] G Eichfelder, T Gerlach, and L Warnow. *Test Instances For Multiobjective Mixed-Integer Nonlinear Optimization*. Mar. 24, 2023. URL: https://optimization-online.org/?p=22458.

[5] R Fletcher and S Leyffer. “Solving mixed integer nonlinear programs by outer approximation”. In: *Math. Prog.* 66.3 (1994), pp. 327–349.