Constrained additive models for global optimization

Marina Cuesta[1], Claudia D'Ambrosio[2], María Durban[1], Vanesa Guerrero [1] and Renan Spencer Trindade

[1]Department of Statistics, Universidad Carlos III de Madrid, Spain
[2]LIX CNRS, École Polytechnique, Institut Polytechnique de Paris, France

2025 SEIO LLEIDA

XLI Congreso Nacional de Estadística e Investigación Operativa

Junio 2025

Outline

Introduction

- Finding the global solution of a Mixed Integer Non-Linear Programming (MINLP) is normally NP-hard [Burer and Letchford, 2012]:
- Motivation for developing methods to improve tractability in MINLPs $\Rightarrow$ surrogate problems.
- Data science enables accurate modeling of complex phenomena when sufficient data is available.

$$\Downarrow$$

Integrate data science into mathematical optimization to address complex MINLPs.

$$\Downarrow$$

Build surrogates that are easier to solve while (hopefully) approximating the global optimum.

Introduction. State of the art

Surrogate MINLPs using data-driven techniques:

- Machine learning approaches [Bertsimas and Öztürk, 2023, Bertsimas and Margaritis, 2025].
- Spline-based approach [Grimstad and Sandnes, 2016].

Surrogate MINLPs using knowledge-driven techniques:

- Piecewise linear approximations [Codsi et al., 2025, Duguet and Ngueveu, 2022, Rebennack and Kallrath, 2015].
- Piecewise quadratic approximations [Göß et al., 2025]

CONS:

- Too simplistic for complex structures in the MINLP.
- Not generic: tailored for specific nonlinearities.
- Difficulty to incorporate expert knowledge into the surrogate models.
- Time-consuming.

Introduction. Mixed-integer Smoothing Surrogate Optimization with Constraints (MiSSOC) algorithm.

Data-driven and knowledge-driven framework to address complex MINLPs by building surrogates that are tractable and accurate.

- The complex objective function is approximated with a smooth additive regression model.
- Incorporate expert knowledge in the estimation of the model in terms of shape, bounds or other desirable properties $\Rightarrow$ results aligns with theoretical or domain expectations [Navarro-García, Guerrero, and Durban, 2023, 2024].
- Build the surrogate problem with an appropriate formulation and solve it.

M. Cuesta, C. D'Ambrosio, M. Durban, V. Guerrero, and R. Spencer Trindade. On leveraging constrained smooth additive regression models for global optimization. *Work in progress*, 2025.

Outline

Smooth additive model

Given a set of n observations $\{(x_{i1}, \ldots, x_{ip}, y_i), i = 1, \ldots, n\}$ from p continuous covariates and a continuous response variable, a smooth additive regression model is:

$$y_i = \alpha + f_1(x_{i1}) + \cdots + f_p(x_{ip}) + \epsilon_i, \quad i = 1, \cdots n.$$

Piecewise smooth polynomial estimation with B-splines: Each univariate function $f_j : [\underline{x}_j, \overline{x}_j] \to \mathbb{R}$ is approximated by a linear combination of its $B-$splines basis functions:

$$\hat{f}_j(x) = \sum_{l_j}^{k_j+d_j} \theta_{l_j} B_{l_j}(x), \quad \forall x \in [\underline{x}_j, \overline{x}_j]$$

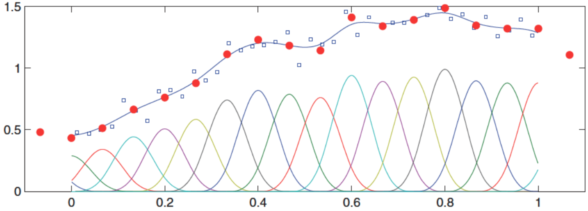


Figure in Eilers et al. [2015].

Smooth additive regression models

The approximation of the additive model involves estimating the θ parameters of each f_j and α:

$$\underset{\boldsymbol{\theta} \in \mathbb{R}^{1+\sum_{j=1}^{p} k_j+d_j}}{\text{minimize}} \quad \|\mathbf{y} - \mathbf{B}\boldsymbol{\theta}\|^2 + \boldsymbol{\theta}^\top \mathbf{P}^I \boldsymbol{\theta}$$

where

- $\boldsymbol{\theta}^\top = (\alpha, \boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_p^\top)^\top$ is the parameter vector, where each $\boldsymbol{\theta}_j = (\theta_{j1}, \ldots, \theta_{j(k_j+d_j)})^\top$ contains the coefficients for the $B-$spline basis functions corresponding to covariate X_j,
- $\mathbf{y} = (y_1, \ldots, y_n)^\top$ is the response vector,
- $\mathbf{B} = [\mathbf{1} : \mathbf{B}_1 : \cdots : \mathbf{B}_j : \cdots : \mathbf{B}_p]$ is the design matrix,
- $(\mathbf{B}_j)_{il} = B_{l,d_j,\mathbf{t}_j}(x_{ij})$ is a $n \times (k_j + d_j)$ matrix with the evaluations of the $k_j + d_j$ $B-$spline basis functions of covariate X_j at its n observed values,
- $\boldsymbol{\theta}^\top \mathbf{P}^I \boldsymbol{\theta}$ is a penalty term for identifiability purposes.

Shape-constrained smooth additive regression models

Suppose that it is known that the response variable Y lies within an interval $[L, U]$. Then:

$$L \leq \hat{\alpha} + \hat{f}_1(x_1) + \cdots + \hat{f}_p(x_p) \leq U, \quad \forall (x_1, \ldots, x_p) \in [\underline{x}_1, \overline{x}_1] \times \cdots \times [\underline{x}_p, \overline{x}_p] \Leftrightarrow$$
$$L - \hat{\alpha} \leq \hat{f}_1(x_1) + \cdots + \hat{f}_p(x_p) \leq U - \hat{\alpha}, \quad \forall (x_1, \ldots, x_p) \in [\underline{x}_1, \overline{x}_1] \times \cdots \times [\underline{x}_p, \overline{x}_p]$$

Each univariate estimated function $\hat{f}_j$ is a piecewise polynomial $\Rightarrow$ necessary and sufficient conditions for characterization of non-negative polynomials of Bertsimas and Popescu [2002].

Proposition 1(d) in Bertsimas and Popescu [2002]

A univariate polynomial $g(x) = \sum_{r=0}^{s} c_r x^r$ is non-negative $\forall x \in [a, b]$ if and only if there exists a positive semidefinite matrix $\mathbf{Z} \in \mathbb{S}^{s+1}$ whose entries satisfy *some specific* linear equations determined by the coefficients c_r and the interval endpoints a and b.

PROBLEM: There is not a direct extension of these necessary and sufficient univariate conditions in the the univariate case to the additive setting.

Shape-constrained smooth additive regression models

We address the model-bounded estimation by bounding each individual function $\hat{f}_j$ as follows:

$$\omega_j^L(L-\hat{\alpha}) \leq \hat{f}_j(x) \leq \omega_j^U(U-\hat{\alpha}), \quad j=1,\dots,p, \quad \forall x \in [\underline{x}_j, \overline{x}_j],$$

Data-driven approach to estimate ω_j^L, ω_j^U:

- Estimate each $\hat{f}_j$ with an unconstrained smooth additive regression model and compute $\underline{\hat{f}_j} = \min(\hat{f}_j(x_{ij}))_{i=1}^n$ and $\overline{\hat{f}_j} = \max(\hat{f}_j(x_{ij}))_{i=1}^n$.
- Estimate $\omega_j^L = \dfrac{\underline{\hat{f}_j}}{\sum_{j=1}^p \underline{\hat{f}_j}}$ and $\omega_j^U = \dfrac{\overline{\hat{f}_j}}{\sum_{j=1}^p \overline{\hat{f}_j}}$, $j=1,\cdots,p$.

Bound-constrained additive model estimation (following Navarro-García, Guerrero, and Durban [2023]):

- Non-negativity conditions to each auxiliary function $\hat{f}_{j_L}(x) := \hat{f}_j(x) - \omega_j^L(L-\hat{\alpha})$.
- Non-negativity conditions to each auxiliary function $\hat{f}_{j_U}(x) := -\hat{f}_j(x) + \omega_j^U(U-\hat{\alpha})$.

Shape-constrained smooth additive models

$$
\begin{aligned}
\min \quad & \|\mathbf{y} - \mathbf{B}\boldsymbol{\theta}\|^2 + \boldsymbol{\theta}^\top \mathbf{P}^l \boldsymbol{\theta} \\
& \langle \mathbf{H}_{j\ell_j}, \mathbf{Z}^L_{jq_j} \rangle_F = 0, \ \forall j = 1, \dots, p, \ \forall q_j = d_j + 1, \dots, d_j + k_j, \ \forall \ell_j = 1, \dots, d_j, \\
& \langle \mathbf{H}_{j\ell_j}, \mathbf{Z}^U_{jq_j} \rangle_F = 0, \ \forall j = 1, \dots, p, \ \forall q_j = d_j + 1, \dots, d_j + k_j, \ \forall \ell_j = 1, \dots, d_j, \\
& \begin{pmatrix} \langle \mathbf{H}_{j(d_j+1)}, \mathbf{Z}^L_{jq_j} \rangle_F \\ \vdots \\ \langle \mathbf{H}_{j(2d_j+1)}, \mathbf{Z}^L_{jq_j} \rangle_F \end{pmatrix} = \mathbf{W}_{jq_j} \mathbf{G}_{jq_j} \boldsymbol{\theta_j} - \mathbf{W}_{jq_j} \begin{pmatrix} \omega_j^L (L - \hat{\alpha}) \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \begin{array}{l} \forall j = 1, \dots, p \\ \forall q_j = d_j + 1, \dots, d_j + k_j \end{array}, \\
& \begin{pmatrix} \langle \mathbf{H}_{j(d_j+1)}, \mathbf{Z}^U_{jq_j} \rangle_F \\ \vdots \\ \langle \mathbf{H}_{j(2d_j+1)}, \mathbf{Z}^U_{jq_j} \rangle_F \end{pmatrix} = -\mathbf{W}_{jq_j} \mathbf{G}_{jq_j} \boldsymbol{\theta_j} + \mathbf{W}_{jq} \begin{pmatrix} \omega_j^U (U - \hat{\alpha}) \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \begin{array}{l} \forall j = 1, \dots, p \\ \forall q_j = d_j + 1, \dots, d_j + k_j \end{array}, \\
& \mathbf{Z}^U_{jq_j}, \ \mathbf{Z}^L_{jq_j} \in \mathcal{S}_+^{d_j+1}, \ \forall j = 1, \dots, p, \ \forall q_j = d_j + 1, \dots, d_j + k_j, \\
& \boldsymbol{\theta} = \left(\alpha, \boldsymbol{\theta_1}^\top, \dots, \boldsymbol{\theta_p}^\top \right)^\top \in \mathbb{R}^{1 + \sum_{j=1}^p k_j + d_j},
\end{aligned}
$$

where $\langle \mathbf{U}, \mathbf{V} \rangle_F = \text{Trace}(\mathbf{U}^\top \mathbf{V})$ and matrices $\mathbf{H}$, $\mathbf{W}$, $\mathbf{G}$ and $\mathbf{Z}$ arise from the conditions in Bertsimas and Popescu [2002].

Outline

MiSSOC algorithm

Consider a MINLP (P), of the following form:

$$
\begin{aligned}
\min \quad & g_0(\mathbf{x}) \\
& g_m(\mathbf{x}) \leq 0, \quad \forall m = 1, \ldots, \overline{m} \\
& x_j \in \mathbb{Z}, \quad \forall j \in I \subseteq \{1, \ldots, p\} \\
& \underline{x}_j \leq x_j \leq \overline{x}_j, \quad \forall j \in \{1, \ldots, p\}, \\
& \mathbf{x} \in \mathbb{R}^p,
\end{aligned}
$$

where $g_0(\mathbf{x})$ is a complex non-linear function. MiSSOC consists of the following main steps:

1. **Data generation:** Sample training data $\mathcal{T} = \{(\mathbf{x}_i, g_0(\mathbf{x}_i)),\ i = 1, \ldots, n\}$.
2. **Approximation:** Approximate $g_0(x)$ with $\tilde{g}_0(x)$ with shape-constrained additive models using $\mathcal{T}$ and adequate parameters.

MiSSOC algorithm

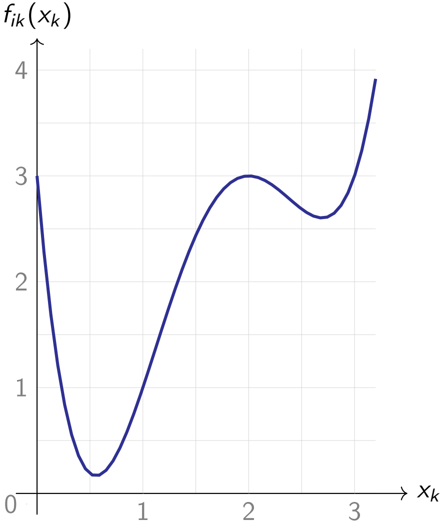
3. **Surrogate problem $\tilde{\mathbf{P}}$:** Build the surrogate problem replacing $g_0(x)$ with $\tilde{g}_0(x)$ though an appropriate reformulation.

4. **Find solution $\tilde{\mathbf{x}}$ to the surrogate ($\tilde{\mathbf{P}}$):** Exploit the separable structure of the surrogate and solve it with the tailored SC-MINLP algorithm [D'Ambrosio et al., 2012, 2019].

5. **Find heuristic solution $\mathbf{x}^*$ to P**: Optimal solution of problem $(\tilde{P})$ is, in general, non-optimal for (P) (optimality loss) $\Rightarrow$ local search in P with $\tilde{\mathbf{x}}$ as starting point $\Rightarrow$ heuristic solution $\mathbf{x}^*$ to P.

SC-MINLP solver

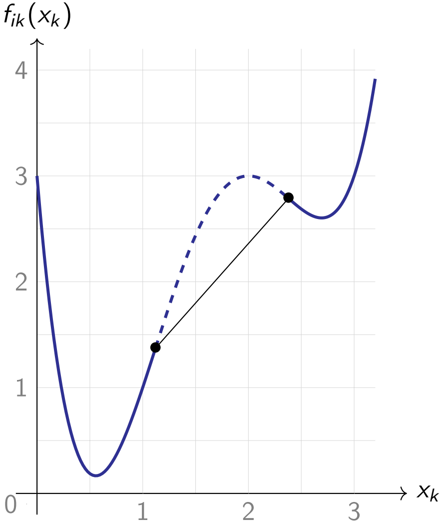
The Sequential Convex - MINLP (SC-MINLP) decomposes the problem into convex subproblems that are easier to solve. It consists of three main steps.



SC-MINLP solver

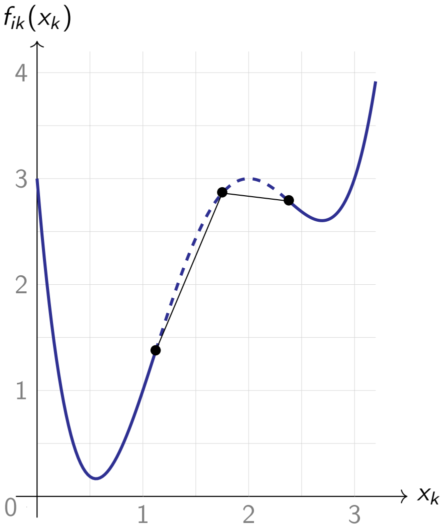
1. Convex relaxation: the concave parts of each univariate function are replaced by linear functions while convex parts remain unchanged $\Rightarrow$ lower bound.

2. Non-convex NLP restriction: fixing the integer variables and locally solving the non-convex NLP subproblem$\Rightarrow$ upper bound



SC-MINLP solver

3. Refinement: The convex relaxation is improved at each iteration to update the lower bound $\Rightarrow$ reduce the gap between the lower and upper bounds.



Outline

MINLPlib benchmark experiments

instance	solver	Original problem P		MiSSOC surrogate $\tilde{P}$	
		obj.	time	obj.	time
ex6_2_13	SC-MINLP	-	-	-0.216	**0.018**
	Gurobi 12	-0.216	600.205	-0.216	0.174
	BARON	-0.216	603.305	-0.216	3.416
	COUENNE	-0.216	607.255	-0.216	8.526
ex6_2_5	SC-MINLP	-	-	-70.558	**0.071**
	Gurobi 12	-70.599	600.278	-70.558	0.479
	BARON	**-70.752**	603.024	-70.558	22.652
	COUENNE	**-70.752**	607.002	-70.558	46.899
ex6_2_7	SC-MINLP	-	-	-0.113	**0.652**
	Gurobi 12	**-0.161**	600.633	-0.113	5.772
	BARON	**-0.161**	602.975	**-0.161**	603.215
	COUENNE	**-0.161**	607.277	**-0.161**	607.752

Outline

A real case study: Hydro Unit Commitment

- Find the optimal schedule for operating hydroelectric units over a given time horizon to maximize revenue subject to technical and operational constraints [Borghetti et al., 2015].
- A feasible solution with objective function value of **14533.1** is reported.
- Complex nonlinearities are in the objective function with the power generation function that depends on water flow q and reservoir volume v.

$$p_{jt} = p(q_{jt}, v_{jt}) = 9.81 q_{jt} \sum_{h=0}^{6} \left(L_h q_{jt}^h \left(\sum_{k=0}^{6} K_k v_{jt}^k - \underline{L} - R_0 q_{jt}^2 \right) \right)$$

- Expert knowledge indicates that the power output lies between 2.595 and 24.089 $\Rightarrow$ function estimation with lower and upper bounds.

A real case study: Hydro Unit Commitment

Problem	Metric	SC-MINLP	Gurobi 12	BARON	COUENNE
Original	obj.	-	-	3692.91	-
	time	-	0	600.00	0.22
MiSSOC surrogate degree 2	obj.	**14538.77**	**14538.87**	**14538.87**	10383.09
	time	**1.52**	**0.92**	156.41	604.46
MiSSOC surrogate degree 3	obj.	**14540.08**	**14540.09**	**14540.09**	12697.24
	time	**1.55**	**3.82**	413.79	601.67
MiSSOC surrogate degree 4	obj.	**14540.05**	**14539.94**	**14540.08**	11548.79
	time	**1.58**	**2.94**	446.69	601.65

Outline

Conclusions

- Innovative integration of statistical modeling into mathematical optimization.
- New approach to address complex MINLPs by building tractable and accurate surrogates.
- Mix of data-driven and knowledge-driven.
- Incorporate domain knowledge into the problem by estimating with contraints $\Rightarrow$ enhancing global optimization algorithms.
- New approach to estimate additive regression models with shape constraints.

Future work

- Selecting the number and location of knots in additive models estimation [Goepp et al., 2025] (with Nicolás Carrizosa, Vanesa Guerrero and María Durban).
- Consider complex functions in constraints $\Rightarrow$ relaxation.
- Impose sparsity in the surrogate piecewise polynomial.
- Extension to black-box optimization with constraint learning.

References I

D. Bertsimas and G. Margaritis. Global optimization: a machine learning approach. *Journal of Global Optimization*, pages 1–37, 2025.

D. Bertsimas and B. Öztürk. Global optimization via optimal decision trees. *Journal of Global Optimization*, pages 1–41, 2023.

D. Bertsimas and I. Popescu. On the relation between option and stock prices: a convex optimization approach. *Operations Research*, 50(2):358–374, 2002.

A. Borghetti, C. D'Ambrosio, A. Lodi, and S. Martello. Optimal scheduling of a multiunit hydro power station in a short-term planning horizon. In *Case Studies in Operations Research*, pages 167–181. Springer, 2015.

S. Burer and A. N. Letchford. Non-convex mixed-integer nonlinear programming: A survey. *Surveys in Operations Research and Management Science*, 17(2):97–106, 2012. doi: 10.1016/j.sorms.2012.08.001.

Julien Codsi, Sandra Ulrich Ngueveu, and Bernard Gendron. Lina: a faster approach to piecewise linear approximations using corridors and its application to mixed-integer optimization. *Mathematical Programming Computation*, 17(2):265–306, 6 2025. ISSN 1867-2957.

Aloïs Duguet and Sandra Ulrich Ngueveu. Piecewise linearization of bivariate nonlinear functions: Minimizing the number of pieces under a bounded approximation error. In Ivana Ljubić, Francisco Barahona, Santanu S. Dey, and A. Ridha Mahjoub, editors, *Combinatorial Optimization*, pages 117–129, Cham, 2022. Springer International Publishing.

C. D'Ambrosio, J. Lee, and A. Wächter. An algorithmic framework for MINLP with separable non-convexity. In J. Lee and S. Leyffer, editors, *Mixed Integer Nonlinear Programming*, volume 154 of *The IMA Volumes in Mathematics and its Applications*, pages 315–347. Springer, New York, 2012.

C. D'Ambrosio, A. Frangioni, and C. Gentile. Strengthening the sequential convex MINLP technique by perspective reformulations. *Optimization Letters*, 13:673–684, 2019.

P. H. C. Eilers, B. D. Marx, and M. Durban. Twenty years of P-splines. *Statistics and Operations Research Transactions*, 39(2): 149–186, 2015.

V. Goepp, O. Bouaziz, and G. Nuel. Spline regression with automatic knot selection. *Computational Statistics and Data Analysis*, 202:108043, 2025.

References II

B. Grimstad and A. Sandnes. Global optimization with spline constraints: a new branch-and-bound method based on B-splines. *Journal of Global Optimization*, 65:401–439, 2016.

Adrian Göß, Robert Burlacu, and Alexander Martin. Parabolic approximation & relaxation for minlp, 2025. URL `https://arxiv.org/abs/2407.06143`.

M. Navarro-García, V. Guerrero, and M. Durban. On constrained smoothing and out-of-range prediction using p-splines: a conic optimization approach. *Applied Mathematics and Computation*, 441:127679, 2023.

M. Navarro-García, V. Guerrero, and M. Durban. A mathematical optimization approach to shape-constrained generalized additive models. *Expert Systems with Applications*, 255(C):124654, 2024.

Steffen Rebennack and Josef Kallrath. Continuous piecewise linear delta-approximations for bivariate and multivariate functions. *Journal of Optimization Theory and Applications*, 167(1):102–117, 10 2015.

Constrained additive models for global optimization

Marina Cuesta[1], Claudia D'Ambrosio[2], María Durban[1], Vanesa Guerrero [1] and Renan Spencer Trindade

[1]Department of Statistics, Universidad Carlos III de Madrid, Spain
[2]LIX CNRS, École Polytechnique, Institut Polytechnique de Paris, France

This research is part of the I+D+i projects CNS2023-144260, PDC2022-133359-I00 and PID2022-137240B-I00 funded by MICIU/AEI/10.13039/501100011033, first and second also by European Union Next GenerationEU/PRTR and third by FEDER/UE, and by the PGMO program of the Fondation Mathématique Jaques Hadamard.

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
AGENCIA ESTATAL DE INVESTIGACIÓN
PGMO
FONDATION MATHÉMATIQUE JACQUES HADAMARD