Computation of Least Trimmed Squares: A Branch-and-Bound Framework with Hyperplane Arrangement Enhancements

Xiang Meng[1] Andrés Gómez[2] Rahul Mazumder[1,3]

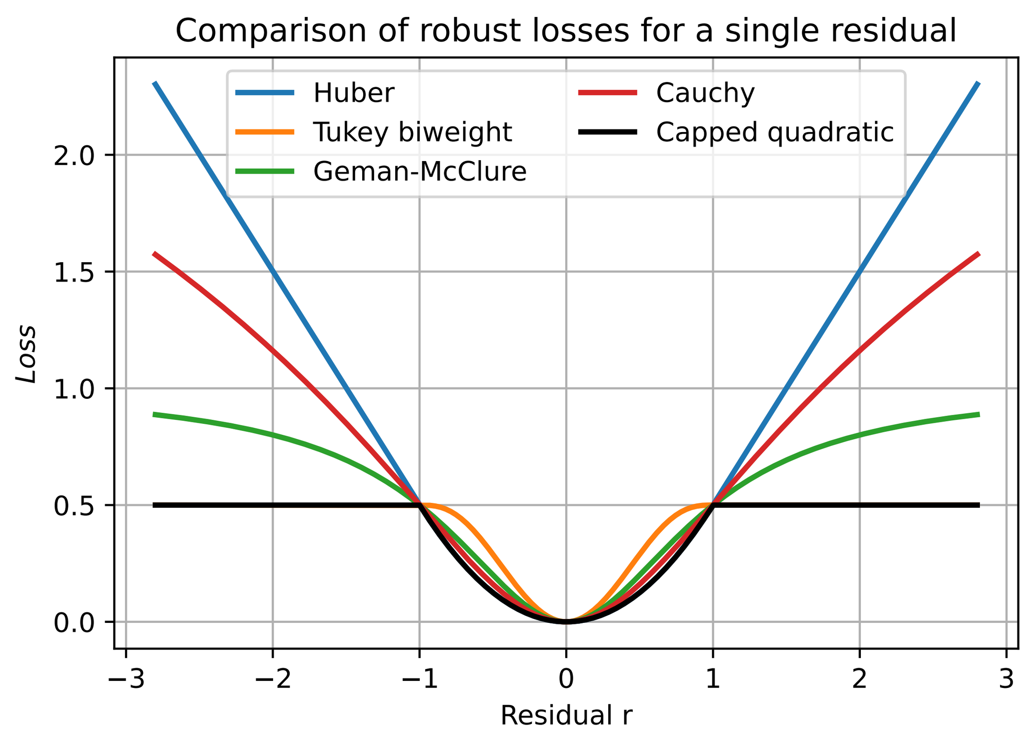
[1]MIT Operations Research Center [2]USC Industrial & Systems Engineering [3]MIT Sloan School of Management

Problem

Given $\boldsymbol{X} \in \mathbb{R}^{n\times p}$, $\boldsymbol{y} \in \mathbb{R}^n$ with $n \gg p$, we seek a linear model that automatically excludes outliers:

$$\min_{\beta, \boldsymbol{z}\in\{0,1\}^n} \frac{1}{2}\sum_i (y_i - \boldsymbol{x}_i^\top \beta)^2 (1-z_i) + \frac{\lambda}{2}\|\beta\|_2^2 + \mu \sum_i z_i$$

- $z_i = 0$: inlier (kept in fit); $z_i = 1$: outlier (excluded).
- $\lambda \geq 0$: ridge regularization; $\mu \geq 0$: outlier penalty.
- A **penalized form of Least Trimmed Squares (LTS)** (Rousseeuw, 1984): $\sqrt{n}$-consistent and asymptotically efficient — but these guarantees **require the global optimum**.
- The problem is **NP-hard** (Bernholt, 2006); heuristics may not inherit any statistical guarantee.



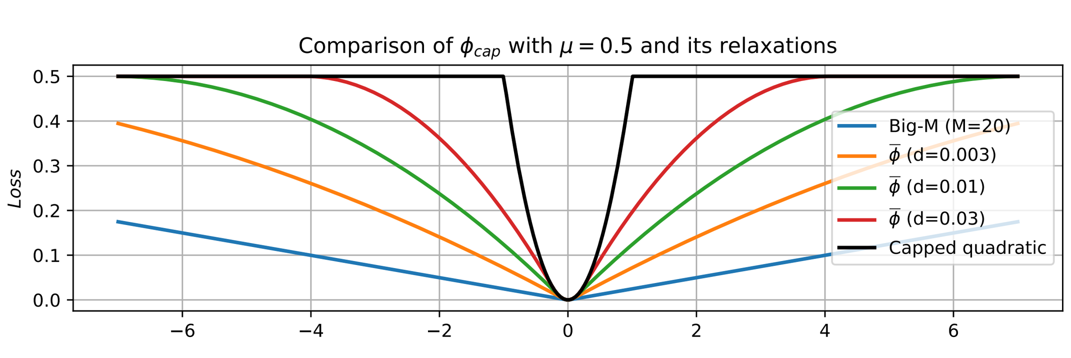
Per-residual loss: capped quadratic plus a constant penalty μ when i is declared an outlier.

Custom BnB Solver

β-space reformulation. We project out the $\mathcal{O}(n)$ auxiliaries $(\boldsymbol{w}, \boldsymbol{z})$, so the node relaxation depends only on $\beta \in \mathbb{R}^p$:

$$\min_{\beta\in\mathbb{R}^p} f(\beta) := \frac{\lambda}{2}\|\beta\|_2^2 + \sum_i \varphi_i(y_i - \boldsymbol{x}_i^\top \beta)$$

Each node-specific loss φ_i has a **closed form** determined by branching $(\varphi_0, \varphi_+, \varphi_-)$.



Closed-form node losses φ_i from the perspective + arrangement formulation.

Ternary branching. Each observation produces three children:

- $z_i = 0$: $|r_i| \leq \sqrt{2\mu}$ (inlier band)
- $z_i = 1, w_i \geq 0$: $r_i \geq \sqrt{2\mu}$ (positive outlier)
- $z_i = 1, w_i \leq 0$: $r_i \leq -\sqrt{2\mu}$ (negative outlier)

ALM with valid dual bounds. Branching adds linear constraints on β, handled via an **Augmented Lagrangian Method**: gradient steps on β alternate with multiplier updates. Strong convexity from the ridge term makes **every** primal iterate produce a **valid** lower bound on the node, so the inner loop need not converge — the branch is pruned the moment this bound exceeds the incumbent.

Implementation: best-bound + most-fractional branching, parent warm starts, alternating-min incumbent (Python + Numba).

Existing Methods & Gap

Heuristics — FAST-LTS (Rousseeuw & Van Driessen, 2006): a fast alternating minimization that gives **no optimality guarantee**; the statistical guarantees of LTS therefore no longer apply.

MIO — Big-M: needs a user-chosen constant M; weak LP relaxation → slow BnB, prohibitive already for $n > 100$.

MIO — Perspective formulation (Gómez & Neto, 2025): Big-M free; perspective terms w_i^2/z_i tighten the relaxation, but it still struggles past $n \approx$ a few hundred.

Hyperplane Arrangement (statistics): solves $\mathcal{O}(n^p)$ ridge regressions — exponential in p. **Theoretical only**; no implementation for $p \geq 3$.

Method	Optimal	Scales in n	$p \geq 3$
FAST-LTS	×	✓	✓
Big-M MIO	✓	$n < 100$	✓
Perspective MIO	✓	limited	✓
HP Arrangement	✓	polynomial	×
This work	✓	✓	✓

Gap. No prior method offers both scalability in n and global optimality. Our BnB closes this gap with up to **100× speedup** over commercial MIO solvers.

Strengthened Formulation

We build on the **perspective formulation**: introduce w_i for the outlier residual, and replace Big-M by perspective terms $d_i w_i^2/z_i$ that force $w_i = 0$ whenever $z_i = 0$:

$$\min \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{w} - \boldsymbol{X}\beta\|_2^2 + \frac{\lambda}{2}\|\beta\|_2^2 + \sum_i \left(\mu z_i + d_i \frac{w_i^2}{z_i} - d_i w_i^2\right)$$

The $d_i > 0$ are chosen so that the first three terms are **jointly convex** in $(\beta, \boldsymbol{w})$ — making the relaxation conic-quadratic representable.

Split outliers by sign through $z_i = z_i^+ + z_i^-$ and $w_i = w_i^+ - w_i^-$, and append the **hyperplane arrangement constraints**:

$$w_i^- \geq \sqrt{2\mu} z_i^-, \quad w_i^+ \geq \sqrt{2\mu} z_i^+$$

$$|y_i - \boldsymbol{x}_i^\top \beta - w_i| \leq \sqrt{2\mu}(1 - z_i^- - z_i^+)$$

The root relaxation is **unchanged**. After branching, fixing $z_i^\pm$ activates the matching inequality and **tightens the relaxation**, so lower bounds improve at every node.

Theorem. BnB has worst-case complexity $\mathcal{O}(\min\{2^n, n^{p+1}\})$ — **polynomial in n** for fixed p. The first MIO formulation to match the $\mathcal{O}(n^p)$ arrangement complexity while keeping BnB practical.

Impact. Even handed to off-the-shelf Gurobi, the strengthened formulation outperforms the perspective formulation by a wide margin. Combined with our custom BnB, it scales to $n = 5000$, $p = 50$ — inaccessible to prior MIO solvers, while preserving exact optimality.

Key Insight: Hyperplane Arrangement

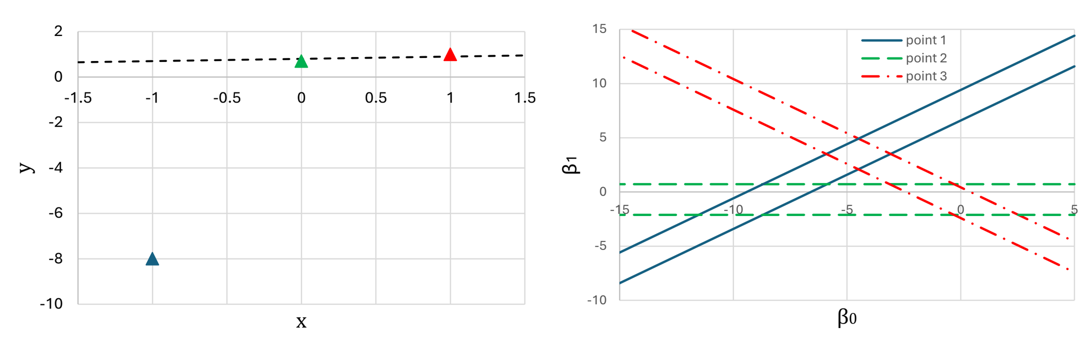
At any optimum $\beta^\star$, the outlier indicators follow a **threshold rule**:

$$z_i^\star = \mathbb{1}\left\{ |y_i - \boldsymbol{x}_i^\top \beta^\star| > \sqrt{2\mu} \right\}$$

Each i induces a **band**

$$B_i = \left\{\beta \in \mathbb{R}^p : |y_i - \boldsymbol{x}_i^\top \beta| \leq \sqrt{2\mu}\right\}$$

in β-space: inside B_i point i is an inlier, outside it is an outlier. The $2n$ bounding hyperplanes partition $\mathbb{R}^p$ into $\mathcal{O}(n^p)$ **regions**; within any region, $\boldsymbol{z}$ is fixed and the problem reduces to a **closed-form ridge regression**.



(a) LTS regression (b) Bands in β-space

Idea. Encode regions through signed binaries z_i^+, z_i^- and add per-point arrangement constraints. They leave the **root relaxation unchanged** but **tighten relaxations after every branching step**.

Why this is hard. Naive enumeration of $\mathcal{O}(n^p)$ regions is intractable for $p \geq 3$. A custom BnB explores regions **adaptively**, visiting only those that current bounds cannot prune — and the arrangement constraints (next column) make it so.

Computational Results

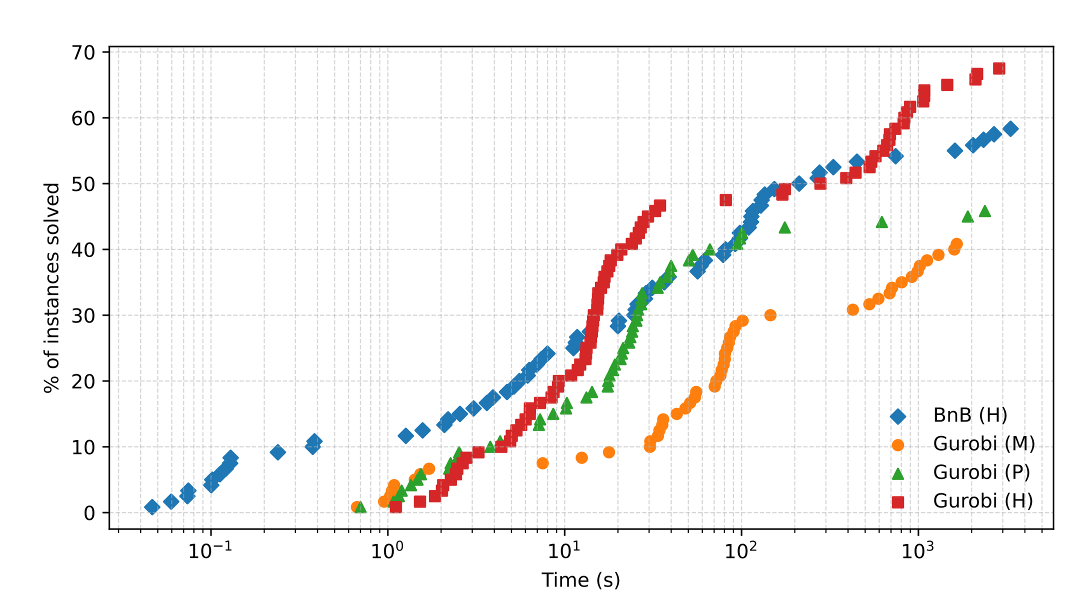
Synthetic (Gaussian $\boldsymbol{X}$; 10 outliers; 1h timeout; 1% gap):

	$n = 1000$		$n = 5000$	
Method	$p = 10$	$p = 50$	$p = 10$	$p = 50$
BnB (H)	**2.5**	**12.9**	**35.7**	**750.6**
Gurobi (H)	358.5	766.4	(14.0%)	(19.6%)
Gurobi (P)	(5.3%)	(4.5%)	(19.2%)	(19.4%)
Gurobi (M)	(33.3%)	(25.6%)	(65.0%)	(64.2%)

Times in s; (gap %) = optimality gap at timeout. (H) strengthened, (P) perspective, (M) Big-M.

BnB (H) solves every instance to optimality, with **up to 100× speedup** over Gurobi (H).

Real datasets (13 benchmarks):



Performance profile: fraction of instances solved to a 1% gap within time t. BnB (H) leads at short and medium runtimes.

Value of exact optimization. FAST-LTS is suboptimal on more than 50% of instances, with gaps reaching 30%.

Take-away. BnB (H) is the only method that reliably attains global optimality at modern scales.

References: [1] Rousseeuw (1984) *JASA* 79:871. [2] Rousseeuw & Van Driessen (2006) *DMKD* 12:29. [3] Bernholt (2006) Tech. Rep., Dortmund. [4] Gómez & Neto (2025) *INFORMS J. Comput.* [5] Hazimeh, Mazumder & Saab (2021) *Math. Program.* [6] Insolia et al. (2022) *Biometrics* 78:1592.
Funding: AFOSR FA9550-24-1-0086, NSF 2346058, ONR N00014-25-1-2504