Quadratic Optimization with Sign-Switching Constraints

Soobin Choi[1] Valentina Cepeda[2] Andrés Gómez[1] Shaoning Han[3]
[1] University of Southern California
[2] Universidad de los Andes
[3] National University of Singapore

Introduction

Problem. We study the mixed-integer quadratic program (MIQP) of the form:

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{z}} \quad & \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{a}^\top \boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{z} \\ \text{s.t.} \quad & x_i z_i \geq 0,\ x_i(1-z_i) \leq 0,\ i \in [n], \\ & \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{z} \in \{0,1\}^n, \end{aligned} \tag{1}$$

where $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ is positive semi-definite and $\boldsymbol{a}, \boldsymbol{c} \in \mathbb{R}^n$. The constraints in Problem (1) enforce the following logical conditions:

Sign-Switching Constraints

$z_i = 1 \implies x_i \geq 0$
$z_i = 0 \implies x_i \leq 0$

Motivation & Challenges.

- **Complexity: NP-hard**; it encodes the Support Vector Machine (SVM) with 0-1 loss [3], making global optimization significantly challenging.
- **The Research Gap:** While "on-off" constraints ($x_i(1-z_i)=0$) benefit from well-studied **perspective relaxations**, sign-switching logic remains under-explored in the literature.
- **Current Limitations:** Most existing approaches rely on **Big-M formulations**. These often result in weak relaxations.

Application: Robust Classification

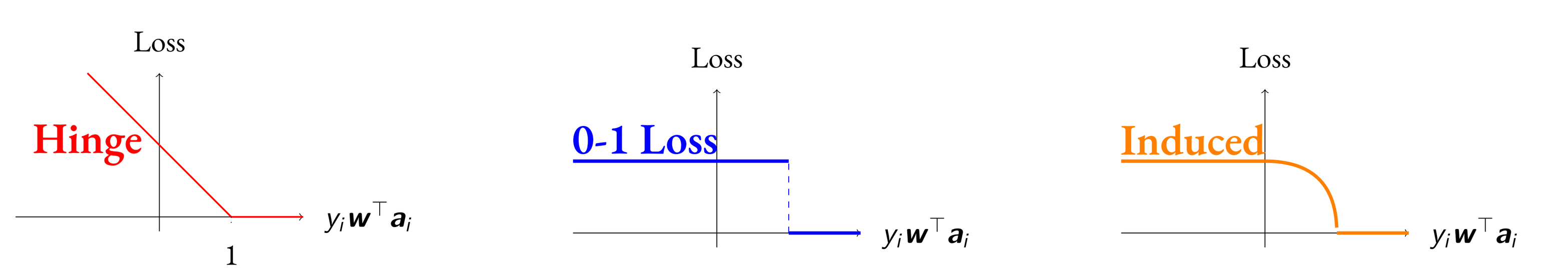
We apply our sign-switching framework to the **SVM with 0-1 loss**, seeking a classifier $\boldsymbol{w}$ that minimizes misclassifications:

$$\min_{\boldsymbol{w},\boldsymbol{z}} \frac{1}{2}\|\boldsymbol{w}\|^2 + \lambda \sum_{i=1}^n z_i \quad \text{s.t.} \begin{cases} z_i = 1 \implies y_i \boldsymbol{w}^\top \boldsymbol{a}_i \leq 1 \\ z_i = 0 \implies y_i \boldsymbol{w}^\top \boldsymbol{a}_i \geq 1 \end{cases}$$

where z_i **indicates misclassification** and $(\boldsymbol{a}_i, y_i) \in \mathbb{R}^p \times \{1,-1\}$, $i \in [n]$ are data points and labels.

Why 0-1 Loss?

- **Outlier Robustness:** Unlike standard hinge loss, 0-1 loss is resilient to noise such as "label flip" noise for rates up to 50% [4].
- **Robust Convex Surrogate:** Our rank-one convexification effectively **induce a tight, bounded loss** that bridges the gap between the discrete 0-1 loss and the loose Hinge loss.



(a) Sensitive to outliers (b) Robust but NP-Hard (c) **Our Convex Relaxation**

Figure 1. Comparison of loss functions. Our induced loss (c) provides a tight, robust surrogate that caps the influence of outliers while remaining computationally tractable via SOCP/SDP.

Rank-One Convexification

We propose **big-M free, convex conic relaxations** for Problem (1), developed in [1, 2].

1. Single Variable Case

For the set

$\mathcal{X} = \{(t,x,z) \in \mathbb{R}^2 \times \{0,1\} : t \geq x^2,\ xz \geq 0,\ x(1-z) \leq 0\}$,

the convex hull is:

$$t \geq \frac{(x)_+^2}{z} + \frac{(x)_-^2}{1-z}, \quad z \in [0,1], \quad x, t \in \mathbb{R}, \tag{2}$$

where $(y)_+ = \max\{y, 0\}$ and $(y)_- = \min\{y, 0\}$ for $y \in \mathbb{R}$.

- **Intuition:** This is the sign-switching analogue to the classic perspective relaxation [5].

2. Rank-One Hessian Case ($\boldsymbol{Q} = \boldsymbol{d}\boldsymbol{d}^\top$, $\boldsymbol{d} \in \mathbb{R}^n_{++}$)

When the quadratic term is rank-one, the convex hull is given by:

$$t \geq \frac{(\boldsymbol{d}^\top \boldsymbol{x})_+^2}{\min\{1, \sum_{i=1}^n z_i\}} + \frac{(\boldsymbol{d}^\top \boldsymbol{x})_-^2}{\min\{1, \sum_{i=1}^n (1-z_i)\}}, \quad \boldsymbol{z} \in [0,1]^n.$$

- **Strength:** Yields significantly tighter dual bounds than Big-M.
- **Implementation:** Representable via **second-order cone programming**.

3. General PSD Hessian Case ($\boldsymbol{Q} \succeq 0$)

We decompose $\boldsymbol{Q}$ into a rank-one matrix and a residual matrix:

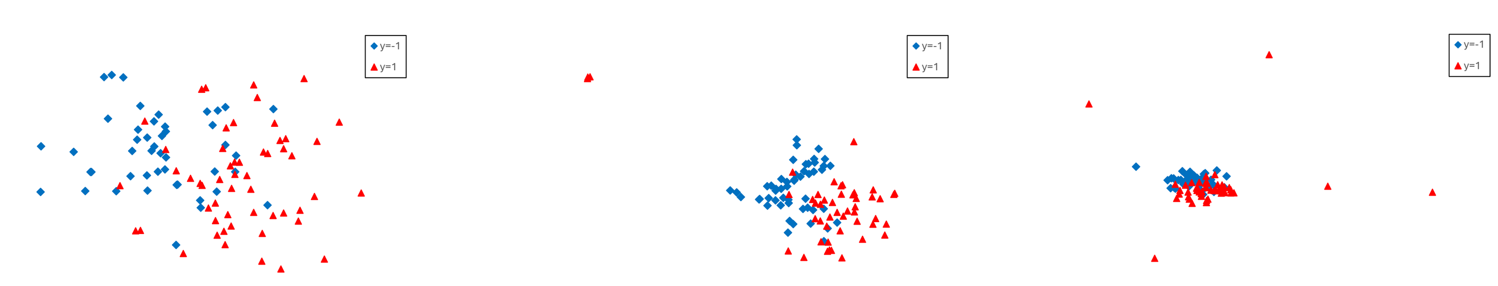
$$\boldsymbol{Q} = \boldsymbol{d}\boldsymbol{d}^\top + \boldsymbol{R}, \quad \boldsymbol{R} \succeq 0.$$

- The **optimal vector $\boldsymbol{d}^*$** giving the **tightest relaxation** for (1) can be found by **semidefinite programming** (see [2] for details).

Experiments

Synthetic Datasets.

We evaluate performance across three data distributions ($\mu \pm \alpha$, variance σ^2): **None** (baseline); **Clustered** (10% leverage points); and **Spread** (10% heavy-tailed outliers).



(a) **none** ($\sigma = 0.5$) (b) **clustered** ($\sigma = 0.5$) (c) **spread** ($\sigma = 0.5$)

Time & Gap. (5 instances from each distribution)

Table 1. $p = 30$, $\sigma = 0.5$, $n = 100$; cardinality $k \in \{10, 20, 30\}$

	big M (Gurobi)		**Single (Mosek)**		**Rank1 (Mosek)**	
k	**Time(s)**	**Gap**	**Time(s)**	**Gap**	**Time(s)**	**Gap**
10	534.6	**33.3%**	**0.2**	73.9%	23.1	33.4%
20	600.0	69.7%	**0.2**	62.3%	23.4	**13.7%**
30	600.0	66.0%	**0.2**	55.6%	21.2	**7.9%**

- Our methods (**Single**, **Rank1**) provide significantly tighter dual bounds and faster solution times than big-M.

Robustness.

While standard hinge loss breaks down (>50% error) for clustered, our relaxations restore predictive power. See [2] for details.

References

[1] Valentina Cepeda, Andrés Gómez, and Shaoning Han. "Robust support vector machines via conic optimization". In: arXiv preprint arXiv:2402.01797 (2024).
[2] Soobin Choi, Valentina Cepeda, Andres Gomez, and Shaoning Han. "Rank-one convexification for quadratic optimization problems with step function penalties". In: arXiv preprint arXiv:2504.16330 (2025).
[3] Corinna Cortes and Vladimir Vapnik. "Support-vector networks". In: Machine Learning 20 (1995), pp. 273–297.
[4] Aritra Ghosh, Naresh Manwani, and PS Sastry. "Making risk minimization tolerant to label noise". In: Neurocomputing 160 (2015), pp. 93–107.
[5] Oktay Günlük and Jeff Linderoth. "Perspective reformulations of mixed integer nonlinear programs with indicator variables". In: Mathematical Programming 124 (2010), pp. 183–205.

Ph.D. Candidate, University of Southern California
soobinch@usc.edu