Appendix A: Proofs

A.1. Proof of Proposition 4

For convenience, we repeat the proposition to be proven concerning set

$$\mathcal{X}_{\boldsymbol{d}}^{-} = \{(\boldsymbol{x}, z, t) \in \mathbb{R}^n \times \{0,1\}^n \times \mathbb{R} : t \geq (\boldsymbol{d}^\top \boldsymbol{x})^2,\ x_i(1 - z_i) \leq 0,\ i = 1, \ldots, n\}.$$

PROPOSITION 11. *(a) If $\boldsymbol{d} \geq \boldsymbol{0}$ (equivalently, if $\boldsymbol{d} \leq \boldsymbol{0}$), then*

$$\operatorname{cl}\operatorname{conv}(\mathcal{X}_{\boldsymbol{d}}^{-}) = \left\{(\boldsymbol{x}, z, t) \in \mathbb{R}^n \times [0,1]^n \times \mathbb{R} : t \geq \frac{(\boldsymbol{d}^\top \boldsymbol{x})_+^2}{\min\left\{1, \sum_{i \in \operatorname{supp}(\boldsymbol{d})} z_i\right\}} + (\boldsymbol{d}^\top \boldsymbol{x})_-^2\right\}.$$

(b) Otherwise, if $d_i d_j < 0$ for some indexes $i, j \in [n]$, then

$$\operatorname{cl}\operatorname{conv}(\mathcal{X}_{\boldsymbol{d}}^{-}) = \left\{(\boldsymbol{x}, z, t) \in \mathbb{R}^n \times [0,1]^n \times \mathbb{R} : t \geq (\boldsymbol{d}^\top \boldsymbol{x})^2\right\}.$$

Proof of Proposition 11 *Part (a)* We assume that $\boldsymbol{d} > \boldsymbol{0}$ and discuss the extension to the case $\boldsymbol{d} \geq \boldsymbol{0}$ later. Define

$$\widehat{\mathcal{X}}_{\boldsymbol{d}}^{-} = \left\{(\boldsymbol{x}, z, t) \in \mathbb{R}^n \times [0,1]^n \times \mathbb{R} : t \geq \frac{(\boldsymbol{d}^\top \boldsymbol{x})_+^2}{\min\{1, \sum_{i \in [n]} z_i\}} + (\boldsymbol{d}^\top \boldsymbol{x})_-^2\right\}.$$

We show that for any $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^n, \gamma \in \mathbb{R}$, the two problems

$$\min\ \boldsymbol{\alpha}^\top \boldsymbol{x} + \boldsymbol{\beta}^\top z + \gamma t \text{ s.t. } (\boldsymbol{x}, z, t) \in \mathcal{X}_{\boldsymbol{d}}^{-} \tag{35}$$

and

$$\min\ \boldsymbol{\alpha}^\top \boldsymbol{x} + \boldsymbol{\beta}^\top z + \gamma t \text{ s.t. } (\boldsymbol{x}, z, t) \in \widehat{\mathcal{X}}_{\boldsymbol{d}}^{-} \tag{36}$$

share the same optimal objective value and integer optimal solutions. If $\gamma < 0$, or if $\gamma = 0$ and $\alpha_i \neq 0$ for some i, then both (35) and (36) are unbounded, applying similar arguments as those in Proposition 3. Also, if $\gamma = 0$ and $\boldsymbol{\alpha} = \boldsymbol{0}$, then any solution with $z_i = \mathbf{1}_{\{\beta_i < 0\}}$ is optimal for both problems. Hence, it suffices to consider the case where $\gamma > 0$ and $\boldsymbol{\alpha} \neq \boldsymbol{0}$. In such a case, if there is no $\eta \in \mathbb{R}$ such that $\boldsymbol{\alpha} = \eta \boldsymbol{d}$, then we can always find a vector $\boldsymbol{x} \in \mathbb{R}^n$ such that $\boldsymbol{\alpha}^\top \boldsymbol{x} < 0$ and $\boldsymbol{d}^\top \boldsymbol{x} = 0$, which implies that both (35) and (36) are unbounded. Therefore, the only remaining case to consider is $\boldsymbol{\alpha} = \eta \boldsymbol{d}$ for some $\eta \in \mathbb{R}$.

With a change of variables, we may assume $\boldsymbol{d} = \mathbf{1}$ and $\gamma = 1$. Then (36) can be rewritten as

$$\min\ \eta \mathbf{1}^\top \boldsymbol{x} + \boldsymbol{\beta}^\top z + t \text{ s.t. } (\boldsymbol{x}, z, t) \in \widehat{\mathcal{X}}_{\boldsymbol{d}}^{-}. \tag{37}$$

Let f be defined by

$$f(z) := \boldsymbol{\beta}^\top z + \min_{y = (y)_+ + (y)_-} \eta((y)_+ + (y)_-) + \frac{(y)_+^2}{\min\{1, \sum_{i=1}^n z_i\}} + (y)_-^2,$$

with a change of variables $y = \mathbf{1}^\top \boldsymbol{x}$. Then, (37) reduces to $\min_{z \in [0,1]^n} f(z)$.

If $\eta > 0$, then the minimization above attains an optimal solution $y^* = (y^*)_- = -\eta/2$, that is, $f(z) = \boldsymbol{\beta}^\top z - \eta^2/4$. Hence, (37) reduces to the problem $\min_{z\in[0,1]^n} \boldsymbol{\beta}^\top z - \eta^2/4$, which always achieves an integer optimal solution, say z^*. Setting $x_1^* = y^* = -\eta/2$, $x_i^* = 0$ for all $i \neq 1$, notice that the mixed integer solution $(\boldsymbol{x}^*, z^*, t^*)$ with $t^* := (\mathbf{1}^\top \boldsymbol{x}^*)^2$ is optimal for (37). Also, one can check that the solution $(\boldsymbol{x}^*, z^*, t^*)$ is feasible for (35), which implies that the optimal values of (35) and (37) coincide.

If $\eta \leq 0$, one can verify that for each z, the point $y^*(z) = -\frac{\eta}{2}\min\{1, \sum_{i=1}^n z_i\}$ solves the minimization problem that defines $f(z)$. Thus, $f(z) = \boldsymbol{\beta}^\top z - \frac{\eta^2}{4}\sum_{i=1}^n z_i$. Also, the problem $\min_{z\in[0,1]^n} f(z)$ attains an integer optimal solution $\bar{z}$ as established in Proposition 3. If $\bar{z} = \mathbf{0}$, then we can set $\bar{\boldsymbol{x}} = \mathbf{0}$ and $t = 0$, which is optimal for (35). If $\bar{z}_j = 1$ for some j, we can set $\bar{x}_j = y^*(\bar{z})$ and $\bar{x}_i = 0$ for $i \neq j$. By letting $\bar{t} = (\boldsymbol{d}^\top \bar{\boldsymbol{x}})^2$, the solution $(\bar{\boldsymbol{x}}, \bar{z}, \bar{t})$ is optimal for problem (37) and is also feasible for problem (35). That is, we have shown that (36) and (35) share the same optimal value.

Finally, extending these results to the general case $\boldsymbol{d} \geq \mathbf{0}$ follows from arguments similar to those in Theorem 1. We omit the details for brevity.

Part (b) We assume without loss of generality that $d_1 > 0$ and $d_2 < 0$. We first note that since any point $(\boldsymbol{x}, z, t)$ with $\boldsymbol{x} \leq 0$, $z \in \{0,1\}^n$ and $t \geq (\boldsymbol{d}^\top \boldsymbol{x})^2$ is an element of $\mathcal{X}_{\boldsymbol{d}}^-$, we can easily conclude that points satisfying $\boldsymbol{x} \leq 0$, $\mathbf{0} \leq z \leq \mathbf{1}$ and $t \geq (\boldsymbol{d}^\top \boldsymbol{x})^2$ belong to $\operatorname{cl}\operatorname{conv}(\mathcal{X}_{\boldsymbol{d}}^-)$.

Then, observe that for any point $(\bar{\boldsymbol{x}}, \bar{z}, t)$ with $\bar{t} \geq (\boldsymbol{d}^\top \bar{\boldsymbol{x}})^2$, $\mathbf{0} \leq z \leq \mathbf{1}$ and $\boldsymbol{d}^\top \bar{\boldsymbol{x}} \leq 0$, we find that

$$(\bar{\boldsymbol{x}}, \bar{z}, \bar{t}) = \lim_{\lambda \to 0_+} \lambda \underbrace{\left(\frac{\bar{\boldsymbol{x}}}{\lambda} - \left(\frac{\boldsymbol{d}^\top \bar{\boldsymbol{x}}}{d_1 \lambda}\right)\boldsymbol{e_1}, \mathbf{1}, 0\right)}_{=(\boldsymbol{x^1}, z^1, t^1)} + (1-\lambda) \underbrace{\left(\left(\frac{\boldsymbol{d}^\top \bar{\boldsymbol{x}}}{d_1}\right)\boldsymbol{e_1}, \bar{z}, \bar{t}\right)}_{=(\boldsymbol{x^2}, z^2, t^2)},$$

where $\boldsymbol{e_1}$ is first standard basis vector of $\mathbb{R}^n$. Moreover, $(\boldsymbol{x^1}, z^1, t^1) \in \mathcal{X}_{\boldsymbol{d}}^-$ since $z^1 = \mathbf{1}$ imposes no constraints on $\boldsymbol{x^1}$ and $\boldsymbol{d}^\top \boldsymbol{x^1} = 0$; and $(\boldsymbol{x^2}, z^2, t^2) \in \operatorname{cl}\operatorname{conv}(\mathcal{X}_{\boldsymbol{d}}^-)$ since $\boldsymbol{x^2} \leq \mathbf{0}$ and $\mathbf{0} \leq z^2 \leq \mathbf{1}$. Thus, $(\bar{\boldsymbol{x}}, \bar{z}, \bar{t}) \in \operatorname{cl}\operatorname{conv}(\mathcal{X}_{\boldsymbol{d}}^-)$. If $\boldsymbol{d}^\top \bar{\boldsymbol{x}} > 0$, a similar argument can be repeated using $\boldsymbol{e_2}$ instead of $\boldsymbol{e_1}$. □

A.2. Proof of Proposition 5

We begin by repeating the proposition we want to prove.

PROPOSITION 12. *The following two statements hold true:*

(a) Given a point $(\boldsymbol{x}, z, t) \in \mathbb{R}^{2n+1}$, $(\boldsymbol{x}, z, t) \in \operatorname{conv}(\mathcal{X}_{\boldsymbol{Q}})$ if and only if there exists $\boldsymbol{X} \in \mathbb{R}^{n\times n}$ such that $t \geq \langle \boldsymbol{Q}, \boldsymbol{X}\rangle$ and $(\boldsymbol{x}, \boldsymbol{X}, z) \in \operatorname{conv}(\bar{\mathcal{X}})$.

(b) Given a point $(\boldsymbol{x}, \boldsymbol{X}, z) \in \mathbb{R}^{2n+n^2}$, $(\boldsymbol{x}, \boldsymbol{X}, z) \in \operatorname{cl}\operatorname{conv}(\bar{\mathcal{X}})$ if and only if for all $\boldsymbol{Q} \in \mathcal{S}_+^n$ there exists $t \in \mathbb{R}$ such that $t = \langle \boldsymbol{Q}, \boldsymbol{X}\rangle$ and $(\boldsymbol{x}, z, t) \in \operatorname{cl}\operatorname{conv}(\mathcal{X}_{\boldsymbol{Q}})$.

Proof of Proposition 12 To prove part (a), define

$$\tilde{\mathcal{X}} = \left\{(\boldsymbol{x}, \boldsymbol{X}, z, t) : (\boldsymbol{x}, \boldsymbol{X}, z) \in \bar{\mathcal{X}},\ t \geq \langle \boldsymbol{Q}, \boldsymbol{X}\rangle\right\}.$$

Then one has $\mathcal{X}_{\boldsymbol{Q}} = \operatorname{proj}_{\boldsymbol{x},z,t} \tilde{\mathcal{X}}$ which represents the projection of $\tilde{\mathcal{X}}$ onto the space of $(\boldsymbol{x}, z, t)$. We immediately find that

$$\begin{aligned}
\operatorname{conv}(\mathcal{X}_{\boldsymbol{Q}}) &= \operatorname{conv}\operatorname{proj}_{\boldsymbol{x},z,t} \tilde{\mathcal{X}} \\
&= \operatorname{proj}_{\boldsymbol{x},z,t} \operatorname{conv} \tilde{\mathcal{X}} \\
&= \operatorname{proj}_{\boldsymbol{x},z,t} \left\{(\boldsymbol{x}, \boldsymbol{X}, z, t) : (\boldsymbol{x}, \boldsymbol{X}, z) \in \operatorname{conv}\left(\bar{\mathcal{X}}\right),\ t \geq \langle \boldsymbol{Q}, \boldsymbol{X}\rangle\right\},
\end{aligned}$$

which proves the first statement.

To prove part (b), we consider minimization of a linear function over the two sets involved, that is,

$$\min_{\boldsymbol{x},\boldsymbol{X},z} \boldsymbol{a}^\top \boldsymbol{x} + \boldsymbol{c}^\top z + \langle \boldsymbol{\Sigma}, \boldsymbol{X} \rangle \text{ s.t. } (\boldsymbol{x}, \boldsymbol{X}, z) \in \text{cl conv}(\bar{\mathcal{X}}), \text{ and} \tag{38}$$

$$\min_{\boldsymbol{x},\boldsymbol{X},z} \boldsymbol{a}^\top \boldsymbol{x} + \boldsymbol{c}^\top z + \langle \boldsymbol{\Sigma}, \boldsymbol{X} \rangle \text{ s.t. } (\boldsymbol{x}, z, \langle \boldsymbol{Q}, \boldsymbol{X} \rangle) \in \text{cl conv}(\mathcal{X}_{\boldsymbol{Q}}) \text{ for all } \boldsymbol{Q} \in \mathcal{S}^n_+, \tag{39}$$

and show that both problems have the same optimal solutions and objective values.

First note that if $\boldsymbol{\Sigma} \notin \mathcal{S}^n_+$, then both problems are unbounded. Indeed, it is feasible for both problems to set $\boldsymbol{X} = \boldsymbol{x}\boldsymbol{x}^\top + \gamma \boldsymbol{v}\boldsymbol{v}^\top$ where $\gamma \geq 0$ and $\boldsymbol{v}$ is the eigenvector associated with the minimum eigenvalue of $\boldsymbol{\Sigma}$; letting $\gamma \to \infty$, we find feasible solutions with arbitrarily small objective values. Similarly, if $\boldsymbol{\Sigma} \in \mathcal{S}^n_+$ with minimum eigenvalue 0, and $\boldsymbol{v}$ is any vector satisfying $\boldsymbol{\Sigma}\boldsymbol{v} = \boldsymbol{0}$ and $\boldsymbol{a}^\top \boldsymbol{v} < 0$, both problems can be verified to be unbounded.

Otherwise, if $\boldsymbol{\Sigma} \in \mathcal{S}^n_+$, then any optimal solution of (38) that is also optimal for

$$\min_{\boldsymbol{x} \in \mathbb{R}^n, z \in \{0,1\}} \boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x} + \boldsymbol{a}^\top \boldsymbol{x} + \boldsymbol{c}^\top z \text{ s.t. } x_i z_i \geq 0,\ x_i(1 - z_i) \leq 0,\ \forall i \in [n] \tag{40}$$

with same objective value. Note that problem (39) is a relaxation of (40), since any feasible solution $(\boldsymbol{x}, z)$ of (40) yields a feasible solution for (39) by setting $\boldsymbol{X} = \boldsymbol{x}\boldsymbol{x}^\top$, with the same objective value. Now consider the further relaxation of (39) where constraints for all matrices $\boldsymbol{Q} \neq \boldsymbol{\Sigma}$ are dropped:

$$\begin{aligned} &\min_{\boldsymbol{x},\boldsymbol{X},z} \langle \boldsymbol{\Sigma}, \boldsymbol{X} \rangle + \boldsymbol{a}^\top \boldsymbol{x} + \boldsymbol{c}^\top z \text{ s.t. } (\boldsymbol{x}, z, \langle \boldsymbol{\Sigma}, \boldsymbol{X} \rangle) \in \text{cl conv}(\mathcal{X}_{\boldsymbol{\Sigma}}) \\ = &\min_{(\boldsymbol{x},z,t) \in \text{cl conv}(\mathcal{X}_{\boldsymbol{\Sigma}})} t + \boldsymbol{a}^\top \boldsymbol{x} + \boldsymbol{c}^\top z = \min_{(\boldsymbol{x},z,t) \in \mathcal{X}_{\boldsymbol{\Sigma}}} t + \boldsymbol{a}^\top \boldsymbol{x} + \boldsymbol{c}^\top z, \end{aligned}$$

which reduces to (40). In conclusion, we find that (39) is equivalent to (40) and to (38). □

A.3. Proof of Proposition 8

For any cone C, we denote the dual cone of C by $C^* = \{\boldsymbol{X}^* : \langle \boldsymbol{X}, \boldsymbol{X}^* \rangle \geq 0\ \forall \boldsymbol{X} \in C\}$. Define $\mathcal{CP}^n$ as the cone of n-dimensional *completely positive matrices*, that is, $\mathcal{CP} = \text{conv}\left\{\boldsymbol{x}^\top \boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^n_+\right\}$. Note that $\left(C^n_+\right)^* = \mathcal{CP}^n$. Proposition 8 is a direct corollary of Proposition 13.

PROPOSITION 13. *Assume $\boldsymbol{X} \in \mathcal{S}^n_+$. Then $\begin{pmatrix} t & \boldsymbol{x}^\top \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \in C^{n+1}_+$ if and only if $\begin{pmatrix} t & \boldsymbol{y}^\top \\ \boldsymbol{y} & \boldsymbol{X} \end{pmatrix} \succeq 0$ for certain $\boldsymbol{y} \leq \boldsymbol{x}$.*

Proof of Proposition 8 Define $C_L = C^{n+1}_+ \cap \mathcal{R}_L$ and $C_R = \mathcal{S}^{n+1}_+ + \mathcal{R}_R$, where the sum of two sets is understood as Minkowski sum and

$$\mathcal{R}_L = \left\{ \begin{pmatrix} t & \boldsymbol{x}^\top \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} : \boldsymbol{X} \in \mathcal{S}^n_+ \right\} \text{ and } \mathcal{R}_R = \left\{ \begin{pmatrix} 0 & \boldsymbol{x}^\top \\ \boldsymbol{x} & \boldsymbol{0} \end{pmatrix} \in \mathbb{R}^{(n+1)\times(n+1)} : \boldsymbol{x} \in \mathbb{R}^n_+ \right\}.$$

Then Proposition 13 amounts to $C_L = C_R$ which is further equivalent to $C^*_L = C^*_R$ due to the closedness and convexity of C_L and C_R. Since $C_R \subseteq C_L$, one has $C^*_L \subseteq C^*_R$. Thus, it suffices to prove $C^*_R \subseteq C^*_L$.

Note that

$$C_L^* = \left(C_+^{n+1}\right)^* + \mathcal{R}_L^* = C\mathcal{P}^{n+1} + \mathcal{R}_L^*,$$

and

$$C_R^* = \left(\mathcal{S}_+^{n+1}\right)^* \cap \mathcal{R}_R^* = \mathcal{S}_+^{n+1} \cap \mathcal{R}_R^*, \tag{41}$$

where

$$\mathcal{R}_L^* = \left\{\begin{pmatrix} 0 & \mathbf{0}^\top \\ \mathbf{0} & \boldsymbol{X} \end{pmatrix} : \boldsymbol{X} \in \mathcal{S}_+^n\right\} \text{ and } \mathcal{R}_R^* = \left\{\begin{pmatrix} t & \boldsymbol{x}^\top \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} : \boldsymbol{x} \in \mathbb{R}_+^n\right\}.$$

Taking an arbitrary $\hat{\boldsymbol{X}} = \begin{pmatrix} t & \boldsymbol{x}^\top \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \in C_R^*$, one has $t \geq 0$, $\hat{\boldsymbol{X}} \succeq 0$ and $\boldsymbol{x} \geq 0$ by (41). There are two cases:

- Case 1: $t = 0$. In this case, one must have $\boldsymbol{x} = 0$, implying $\hat{\boldsymbol{X}} \in \mathcal{R}_L^* \subseteq \mathcal{R}_L^* + C\mathcal{P}^{n+1} = C_L^*$.
- Case 2: $t > 0$. In this case, define $\hat{\boldsymbol{X}} = \underbrace{\begin{pmatrix} t & \boldsymbol{x}^\top \\ \boldsymbol{x} & \frac{1}{t}\boldsymbol{x}\boldsymbol{x}^\top \end{pmatrix}}_{\hat{\boldsymbol{X}}^1} + \underbrace{\begin{pmatrix} 0 & \mathbf{0}^\top \\ \mathbf{0} & \boldsymbol{X} - \frac{1}{t}\boldsymbol{x}\boldsymbol{x}^\top \end{pmatrix}}_{\hat{\boldsymbol{X}}^2}$. One can see that $\hat{\boldsymbol{X}}^2 \in \mathcal{R}_L^*$ and $\hat{\boldsymbol{X}}^1 = \begin{pmatrix} \sqrt{t} \\ \frac{\boldsymbol{x}}{\sqrt{t}} \end{pmatrix}\begin{pmatrix} \sqrt{t} \\ \frac{\boldsymbol{x}}{\sqrt{t}} \end{pmatrix}^\top \in C\mathcal{P}^{n+1}$, implying $\hat{\boldsymbol{X}} \in \mathcal{R}_L^* + C\mathcal{P}^{n+1} = C_L^*$.

This shows $C_R^* \subseteq C_L^*$ and concludes the proof. □

Appendix B: Connections with earlier works

In this section, we discuss connections with earlier works, pointing out similarities and differences with convexifications based on indicator variables.

B.1. Connections for conic quadratic relaxations

As pointed out in §3, the set $\mathcal{X}_{\boldsymbol{Q}}$ is connected to sets with indicator variables previously studied in the literature. Remark 1 below discusses the special case where $n = 1$, where the results of Theorem 1 can indeed be interpreted as a direct application of the perspective reformulation in an extended formulation. However, in Remark 2 we show that, in general, the descriptions of the convex hull are advantageous compared with existing alternatives.

REMARK 1 (CASE $n = 1$). Define set $\mathcal{X}_1 = \{(x, z, t) \in \mathbb{R} \times \{0, 1\} \times \mathbb{R} : t \geq x^2,\ xz \geq 0,\ x(1 - z) \leq 0\}$. Using the extended formulation in (15), set $\mathcal{X}_1$ can be rewritten as

$$\mathcal{X}_1 = \{(x, z, t) \in \mathbb{R} \times \{0, 1\} \times \mathbb{R} : \exists x^+, x^- \in \mathbb{R}_+ \text{ such that } t \geq (x^+ - x^-)^2,\ x^- z = 0,\ x^+(1 - z) = 0, x = x^+ - x^-\}.$$

Then noting that for any feasible point $(x^+ - x^-)^2 = (x^+)^2 + (x^-)^2$ and using the perspective reformulation (18), we obtain the valid inequality $t \geq \frac{(x)_+^2}{z} + \frac{(x)_-^2}{1-z}$ and $0 \leq z \leq 1$. Theorem 1 states that this valid inequality is sufficient to describe the convex hull. ■

REMARK 2 (GENERAL CASE). Using the extended formulation in (15), set $\mathcal{X}_{\boldsymbol{d}}$ can be rewritten as

$$\mathcal{X}_{\boldsymbol{d}} := \Big\{(\boldsymbol{x}, \boldsymbol{z}, t) \in \mathbb{R}^n \times \{0, 1\}^n \times \mathbb{R} : \exists \boldsymbol{x}^-, \boldsymbol{x}^+ \in \mathbb{R}_+^n \text{ such that } t \geq \left(\boldsymbol{d}^\top \boldsymbol{x}^+ - \boldsymbol{d}^\top \boldsymbol{x}^-\right)^2, \\ \boldsymbol{x} = \boldsymbol{x}^+ - \boldsymbol{x}^-,\ x_i^- z_i = 0,\ x_i^+(1 - z_i) = 0,\ i = 1, \ldots, n\Big\}.$$

In other words, set $\mathcal{X}_{\boldsymbol{d}}$ can be written as the epigraph of a rank-one quadratic function with indicator variables and nonnegative constraints, a class of sets that has been studied in the literature. Nonetheless, the resulting strengthening is substantially more complicated and potentially weaker as well:

- The strengthening described in proposition 7 of Atamtürk and Gómez (2023), in a space of variables including $\boldsymbol{x}^+$ and $\boldsymbol{x}^-$ but no additional variables, is described as a piecewise conic quadratic function with a factorial number of pieces ($(2n)!+1$) that cannot be directly implemented with off-the-shelf solvers.
- The conic quadratic strengthenings described in proposition 3 of Han and Gómez (2024) or theorem 5 of Shafiee and Kılınç-Karzan (2024) require introducing additional $4n$ variables, in addition to $\boldsymbol{x}^+$ and $\boldsymbol{x}^-$, and $O(n)$ constraints.

We see that the formulation in Theorem 1 is simpler, as it does not require additional variables (including $\boldsymbol{x}^+$ and $\boldsymbol{x}^-$). ■

B.2. Connections for semidefinite programming relaxations

We briefly discuss how the convexification given in Proposition 9 relates to the results of Atamtürk and Gómez (2025) for a related set with indicator variables.

REMARK 3 (COMPARISON WITH SETS WITH INDICATOR VARIABLES). Consider the set with indicator variables defined as

$$\bar{\mathcal{Y}} := \left\{(\boldsymbol{x}, \boldsymbol{X}, \boldsymbol{z}) \in \mathbb{R}^n \times \mathbb{R}^{n\times n} \times \{0,1\}^n : \boldsymbol{X} \succeq \boldsymbol{x}\boldsymbol{x}^\top,\ x_i(1-z_i) = 0,\ i = 1,\dots,n\right\}.$$

Atamtürk and Gómez (2025) show that valid inequalities for this set, derived from rank-one strengthening, are given by conic constraints of the form $\begin{pmatrix} \sum_{i\in S} z_i & \boldsymbol{x}_S^\top \\ \boldsymbol{x}_S & \boldsymbol{X}_S \end{pmatrix} \in \mathcal{S}_+^{|S|+1} \Leftrightarrow \begin{pmatrix} \sum_{i\in S} z_i & -\boldsymbol{x}_S^\top \\ -\boldsymbol{x}_S & \boldsymbol{X}_S \end{pmatrix} \in \mathcal{S}_+^{|S|+1}$. Proposition 9 shows that to obtain similar inequalities for the relaxation $\bar{\mathcal{X}}^-$, it suffices to relax the inequalities from the positive semidefinite cone to the copositive cone (while being careful with the sign of the continuous variables). ■

Appendix C: Additional insights of the rank-one formulations in SVM problems

We now discuss how to proposed formulation (34) for SVMs can be interpreted as a strengthening based on decomposition of the quadratic term, or as using a non-convex loss function to underestimate the 0-1 loss.

C.1. Interpretation as decomposition

We now provide some high-level intuition and interpretation of the formulation proposed in Proposition 10. Consider an alternative application of the formulations proposed, consisting of using the valid inequalities in Proposition 4, without introducing matrix variables $\boldsymbol{X}$ or $\boldsymbol{W}$. Then, given any matrix $\boldsymbol{F} \in \mathbb{R}_+^{q\times n}$ (for some $q \in \mathbb{Z}_+$), we can rewrite (33) as

$$\min_{\boldsymbol{w},\boldsymbol{x},\boldsymbol{X},\boldsymbol{z}} \|\boldsymbol{w}\|_2^2 - \|\boldsymbol{F}(\boldsymbol{1} - \boldsymbol{A}\boldsymbol{w})\|_2^2 + \|\boldsymbol{F}\boldsymbol{x}\|_2^2 + \lambda \sum_{i=1}^n z_i \tag{42a}$$

$$\text{s.t. } \boldsymbol{x} = \boldsymbol{1} - \boldsymbol{A}\boldsymbol{w} \tag{42b}$$

$$x_i(1-z_i) \le 0 \quad \forall i \in [n] \tag{42c}$$

$$\boldsymbol{w} \in \mathbb{R}^p,\ \boldsymbol{x} \in \mathbb{R}^n,\ \boldsymbol{z} \in \{0,1\}^n. \tag{42d}$$

Then, letting $\boldsymbol{f_j} \in \mathbb{R}^n$ denote the j-th row of $\boldsymbol{F}$, we find from Proposition 4 that

$$t \geq \|\boldsymbol{Fx}\|_2^2 \Leftrightarrow t \geq \sum_{j=1}^{q} \left(\boldsymbol{f_j}^\top \boldsymbol{x}\right)^2 \implies t \geq \sum_{j=1}^{q} \left(\frac{(\boldsymbol{f_j}^\top \boldsymbol{x})_+^2}{\min\{1, \sum_{i \in \text{supp}(\boldsymbol{f_j})} z_i\}} + (\boldsymbol{f_j}^\top \boldsymbol{x})_-^2 \right). \tag{43}$$

Thus, if term $\|\boldsymbol{w}\|_2^2 - \|\boldsymbol{F}(\boldsymbol{1} - \boldsymbol{Aw})\|_2^2$ is convex (that is, $\boldsymbol{I}_{p\times p} \succeq \boldsymbol{A}^\top \boldsymbol{F}^\top \boldsymbol{F} \boldsymbol{A}$) and term $\|\boldsymbol{Fx}\|_2^2$ in the objective (42a) is replaced using (43), we obtain a stronger reformulation of (33). However, the quality of relaxation depends on $\boldsymbol{F}$, which must be chosen in an informed way. As we now show, the formulation proposed in Proposition 10 can be shown to be at least as strong as formulations obtained using representation (43).

PROPOSITION 14. *If* $\text{supp}(\boldsymbol{f_j}) \in \mathcal{L}$ *for all* $j \in [q]$*, then* (34) *is at least as strong as the formulation obtained from using strengthening* (43)*.*

Proof of Proposition 14 Our proof technique is based on constructing successive relaxations of (34) and showing that the last relaxation in the sequence is still stronger than using (43). We begin by reformulating the constraints (34b) using the Schur complement for copositive constraints (see Corollary 2), as the (infinite) set of constraints

$$\langle \boldsymbol{f}\boldsymbol{f}^\top, \boldsymbol{X} \rangle \geq \frac{(\boldsymbol{f}^\top \boldsymbol{x})_+^2}{\min\{1, \sum_{i \in L} z_i\}} + (\boldsymbol{f}^\top \boldsymbol{x})_-^2 \quad \forall L \in \mathcal{L}, \forall \boldsymbol{f} \in \mathbb{R}^n_+ : \text{supp}(\boldsymbol{f}) \subseteq L, \tag{44}$$

where $\boldsymbol{X} = \boldsymbol{1}\boldsymbol{1}^\top - \boldsymbol{A}\boldsymbol{w}\boldsymbol{1}^\top - \boldsymbol{1}\boldsymbol{w}^\top \boldsymbol{A}^\top + \boldsymbol{A}\boldsymbol{W}\boldsymbol{A}^\top$ and $\boldsymbol{x} = \boldsymbol{1} - \boldsymbol{A}\boldsymbol{w}$. Consider a relaxation of the set of constraints (44) where: *(i)* constraints associated with sets $L \neq \text{supp}(\boldsymbol{f_j})$ for some $j \in [q]$ are removed, and *(ii)* if $L = \text{supp}(\boldsymbol{f_j})$ for some $j \in [q]$, then constraints associated with any vector $\boldsymbol{f} \neq \boldsymbol{f_j}$ are also removed. In other words, we obtain the relaxed system

$$\begin{aligned} 0 \geq & - \langle \boldsymbol{f_j}\boldsymbol{f_j}^\top, \boldsymbol{1}\boldsymbol{1}^\top - \boldsymbol{A}\boldsymbol{w}\boldsymbol{1}^\top - \boldsymbol{1}\boldsymbol{w}^\top \boldsymbol{A}^\top + \boldsymbol{A}\boldsymbol{W}\boldsymbol{A}^\top \rangle \\ & + \frac{(\boldsymbol{f_j}^\top (\boldsymbol{1} - \boldsymbol{A}\boldsymbol{w}))_+^2}{\min\{1, \sum_{i \in \text{supp}(\boldsymbol{f_j})} z_i\}} + (\boldsymbol{f_j}^\top (\boldsymbol{1} - \boldsymbol{A}\boldsymbol{w}))_-^2 \quad \forall j \in [q]. \end{aligned} \tag{45}$$

Next, we dualize constraints (45) by introducting the dual variables d_j for each $j \in [q]$. We obtain the Lagrangian dual

$$\begin{aligned} \max_{\boldsymbol{d} \in \mathbb{R}^n_+} \min_{\boldsymbol{w}, \boldsymbol{W}, \boldsymbol{z}} & \left\langle \boldsymbol{I}_{p\times p} - \sum_{j=1}^{q} d_j \boldsymbol{A}^\top \boldsymbol{f_j}\boldsymbol{f_j}^\top \boldsymbol{A}, \boldsymbol{W} \right\rangle + \left(2 d_j \sum_{j=1}^{q} \boldsymbol{A}^\top \boldsymbol{f_j}\boldsymbol{f_j}^\top \boldsymbol{1} \right)^\top \boldsymbol{w} - \sum_{j=1}^{q} d_j \boldsymbol{1}^\top \boldsymbol{f_j}\boldsymbol{f_j}^\top \boldsymbol{1} \\ & + \sum_{j=1}^{q} d_j \frac{(\boldsymbol{f_j}^\top (\boldsymbol{1} - \boldsymbol{A}\boldsymbol{w}))_+^2}{\min\{1, \sum_{i \in \text{supp}(\boldsymbol{f_j})} z_i\}} + \sum_{j=1}^{q} d_j (\boldsymbol{f_j}^\top (\boldsymbol{1} - \boldsymbol{A}\boldsymbol{w}))_-^2 + \lambda \sum_{i=1}^{n} z_i, \end{aligned} \tag{46}$$

which, by weak duality, is a relaxation of (34). A further relaxation of (46) is obtained by simply setting $\boldsymbol{d} = \boldsymbol{1}$. This final relaxation is at least as strong as the formulation obtained from (43) if

$$\left\langle \boldsymbol{I}_{p\times p} - \sum_{j=1}^{q} \boldsymbol{A}^\top \boldsymbol{f_j}\boldsymbol{f_j}^\top \boldsymbol{A}, \boldsymbol{W} \right\rangle \geq \boldsymbol{w}^\top \left(\boldsymbol{I}_{p\times p} - \sum_{j=1}^{q} \boldsymbol{A}^\top \boldsymbol{f_j}\boldsymbol{f_j}^\top \boldsymbol{A} \right) \boldsymbol{w},$$

which holds since $\boldsymbol{W} \succeq \boldsymbol{w}\boldsymbol{w}^\top$ and $\boldsymbol{I}_{p\times p} \succeq \boldsymbol{A}^\top \boldsymbol{F}^\top \boldsymbol{F} \boldsymbol{A}$, concluding the proof. □

Proposition 14 ascertains that relaxation (34) is indeed stronger than any direct application of the rank-one convexifications discussed in §4, *provided that the collection $\mathcal{L}$ is rich enough*. However, this strength may come at a prohibitive cost, as the extreme case where $\mathcal{L} = 2^{[n]}$ requires exponentially many constraints, involving cones of dimension up to $p+1$. In practice, we suggest using $\mathcal{L} = \{L \subseteq [n] : |L| \leq \kappa\}$ for some small enough $\kappa \in \mathbb{Z}_+$: the number $O(n^\kappa)$ of constraints added is polynomial in the number of datapoints n, and the dimension of the cones used is bounded by $\kappa + 1$ – except for one $(p+1)$-dimensional cone used to express (34c). While this compromise comes at the expense of relaxation quality, as we show in the next section even setting $\kappa = 1$ may result in substantial gains compared with alternatives proposed in the statistics and machine learning literature.

C.2. Interpretation as regularization

Consider the case where set $\mathcal{L}$ is taken to be the collection of all singletons. In this case, from Proposition 14, we find that relaxation (34) can be interpreted as the relaxation where the objective is "optimally" expressed as

$$\|\boldsymbol{w}\|_2^2 - \sum_{i=1}^n d_i(1 - y_i \boldsymbol{a_i}^\top \boldsymbol{w})^2 + \sum_{i=1}^n d_i \left(\frac{(1 - y_i \boldsymbol{a_i}^\top \boldsymbol{w})_+^2}{z_i} + (1 - y_i \boldsymbol{a_i}^\top \boldsymbol{w})_-^2 \right) + \lambda \sum_{i=1}^n z_i, \tag{47}$$

where "optimally" indicates that vector $\boldsymbol{d}$ is chosen to ensure the strongest relaxation while guaranteeing that $\boldsymbol{I} - \boldsymbol{A}^\top \mathrm{Diag}(\boldsymbol{d}) \boldsymbol{A} \in \mathcal{S}_+^p$. We now retrieve a representation in the original space of variables of (7), i.e., without involving indicator variables z. To do so, we define a class of one-dimensional functions ϕ parametrized by two parameters $d, \lambda \in \mathbb{R}_+$ given by

$$\phi(x; d, \lambda) := \min_{0 \leq z \leq 1} \lambda z - d x^2 + d \frac{(x)_+^2}{z_i} + d(x)_-^2. \tag{48}$$

Observe then that the proposed relaxation of SVM problems (7), with objective value given by (47), can be written as

$$\min_{\boldsymbol{w} \in \mathbb{R}^p} \underbrace{\|\boldsymbol{w}\|_2^2}_{\text{margin}} + \sum_{i=1}^n \underbrace{\phi\left(1 - y_i \boldsymbol{a_i}^\top \boldsymbol{w}; d_i, \lambda\right)}_{\text{misclassification loss for } i}, \tag{49}$$

that is, function ϕ controls the penalty for misclassifying datapoint i depending on the violation $1 - y_i \boldsymbol{a_i}^\top \boldsymbol{w}$. We now provide an explicit description of function ϕ.

PROPOSITION 15. *Assuming $d, \lambda > 0$, the optimal value of problem (48) is the non-convex loss function*

$$\phi(x; d, \lambda) = \begin{cases} 0 & \text{if } x \leq 0 \\ 2(\sqrt{\lambda d})x - d x^2 & \text{if } 0 < x \leq \sqrt{\lambda/d} \\ \lambda & \text{if } x > \sqrt{\lambda/d}. \end{cases} \tag{50}$$

Proof of Proposition 15 To determine the optimal value of z in (48), we consider two cases depending on the value of x.

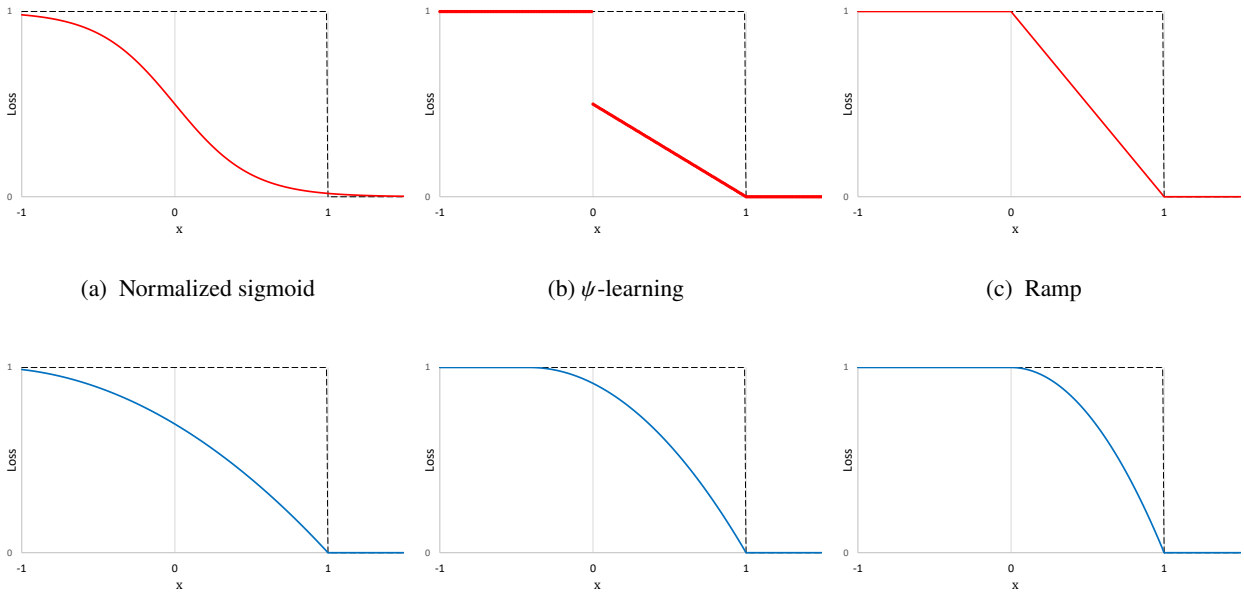
• *Case 1:* $x \leq 0$. In this case, the point is classified on the correct side of the margin. We find that the optimal value is $z^* = 0$ as the objective is minimized to zero.

• *Case 2:* $x \geq 0$. In this case,setting the derivative of (48) with respect to z to zero, we find that the optimal solution is

$$z^* = \min\left\{\sqrt{\frac{d}{\lambda}}x, 1\right\},$$

with objective value $2(\sqrt{\lambda d})x - dx^2$ if $0 < x \leq \sqrt{\lambda/d}$, and λ if $x > \sqrt{\lambda/d}$.

Figure 3 depicts the loss function ϕ along with alternatives from the literature. The parameter d controls the tradeoff between non-convexity and approximation to the 0-1 loss. In problems with uncertainty and outliers, the concave downward shape of the loss function redistributes the influence from outliers to points near the boundary. Note that for $d = 1$, the derived loss is a better approximation of the misclassification loss than the normalized sigmoid, ψ-learning, and ramp losses. This reduces the sensitivity to noise when compared to hinge and other non-convex losses. More importantly, while alternatives Mason et al. (1999), Shen et al. (2003b), Wu and Liu (2007) require tackling non-convex problems (often settling for local minimizer instead of global), the proposed formulation in Proposition 10 retains convexity and can be solved to global optimality.



(a) Normalized sigmoid (b) ψ-learning (c) Ramp

(d) $\phi(\cdot; 0.2, 1)$ (e) $\phi(\cdot; 0.5, 1)$ (f) $\phi(\cdot; 1.0, 1)$

Figure 3 Non-convex robust losses approximating the 0-1 loss, as a function of $x = y_i \boldsymbol{a}_i^\top \boldsymbol{w}$. Top row: loss functions from the literature: the normalized sigmoid loss (Mason et al. 1999), the ψ-learning loss (Shen et al. 2003a) and the ramp loss (Wu and Liu 2007). Bottom row: the derived ϕ loss in Proposition 15 for different values of hyperparameters d (with $\lambda = 1$) . By solving the conic optimization problem (34), d is chosen automatically to ensure the convexity of the ensuing learning problem.