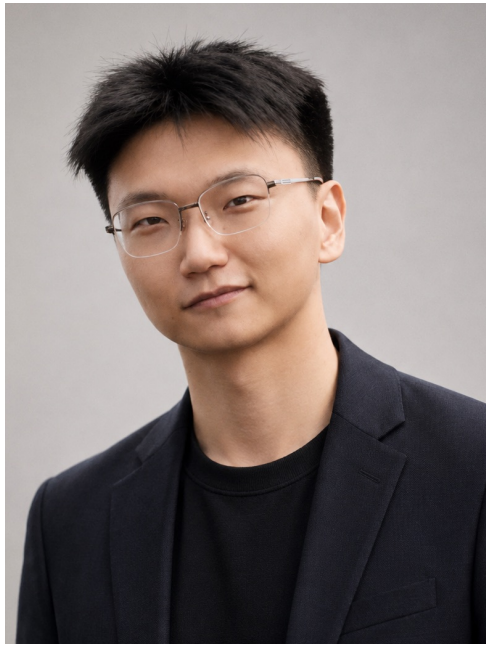
Sparsity-Cone SDP Relaxations and Applications to Variable Fixing for Sparse Quadratic Programs

Di Hou · Thai P.D. Nguyen · Kim-Chuan Toh · Guanyi Wang

National University of Singapore

Peking University · 31 October 2026

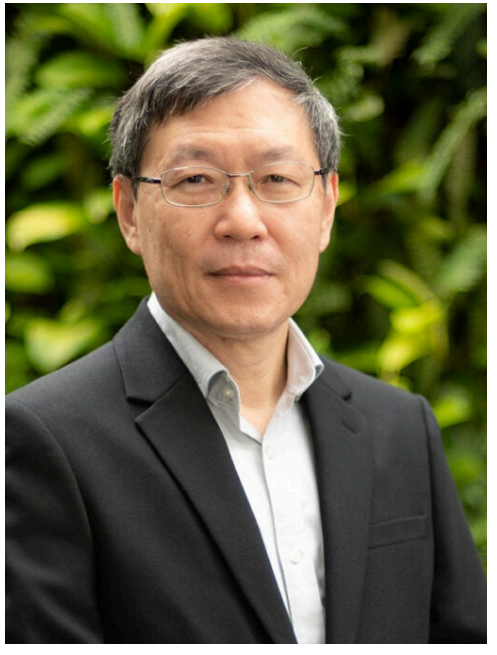
Collaborators



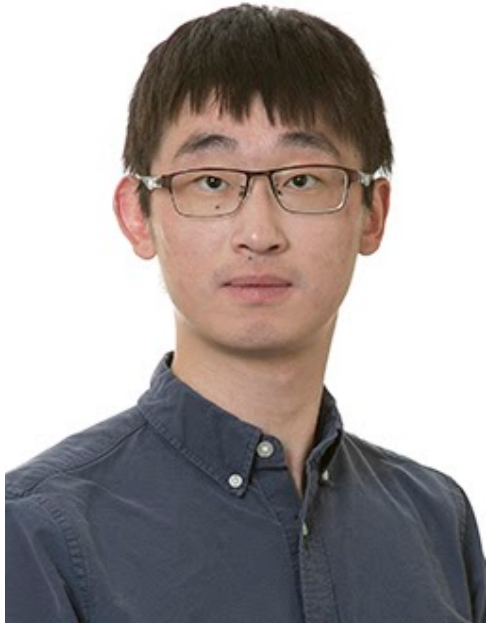
Di Hou



Thai P.D. Nguyen



Kim-Chuan Toh



Guanyi Wang

Outline

Sparsity-cone SDP relaxation

RiNNAL-S: a low-rank augmented Lagrangian method

SDP-based fixing and screening

Numerical experiments

Sparse quadratic programs

$$
\begin{aligned}
v^{\mathrm{P}} := \quad & \min_{x \in \mathbb{R}^n} \quad && x^\top Q x + 2c^\top x \\
& \text{s.t.} && Ax = b, \quad Bx \geq d, \\
& && \|x\|_0 \leq k.
\end{aligned}
$$

- Q can be indefinite.
- $\|x\|_0 := |\{i \in [n] : x_i \neq 0\}|$ counts the nonzero components of x.
- Sparse regression, portfolio optimization, and feature selection.

Objectives

- **Relaxation:** develop a tight relaxation for lower bounds.
- **Algorithm:** design an efficient solver for the relaxation.
- **Presolving:** certify zero variables at optimality from relaxation solutions.

Mixed-binary reformulation

Introduce binary support indicators z:

$$\|x\|_0 \le k \quad \Longleftrightarrow \quad \exists z \in \{0,1\}^n : \quad x \circ (e - z) = 0, \quad e^\top z \le k.$$

Equivalent mixed-binary formulation:

$$\begin{array}{ll} \min\limits_{x,z} & x^\top Q x + 2c^\top x \\ \text{s.t.} & Ax = b, \quad Bx \ge d, \\ & x \circ (e - z) = 0, \quad e^\top z \le k, \quad z \in \{0,1\}^n. \end{array}$$

- Global solvers use branch-and-bound (BnB) to solve this problem.
- Providing tight lower bounds is important for BnB.
- Semidefinite programming (SDP) provides these bounds.

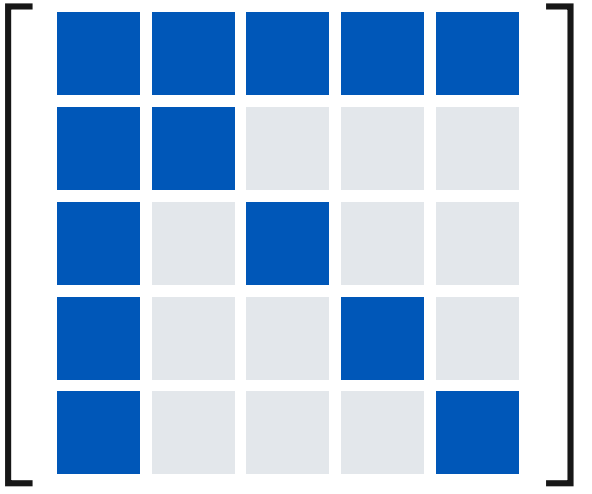
SDP–RLT relaxation

MIQP	Lift and linearize		SDP–RLT
$\min_{x,z} x^\top Q x + 2c^\top x$		$\to$	$\min_{x,X,z,U,Z} \langle Q, X\rangle + 2c^\top x$
s.t. $Ax = b$	$(Ax-b)(Ax-b)^\top = 0$	$\to$	s.t. RLT: $\mathcal{A}(Y) := [-b \;\; A] Y [-b \;\; A]^\top = 0$
$Bx \geq d$	$\begin{bmatrix} 1 \\ Bx-d \end{bmatrix}\begin{bmatrix} 1 \\ Bx-d \end{bmatrix}^\top \geq 0$	$\to$	RLT: $\mathcal{B}(Y) := HYH^\top \geq 0$ (entrywise)
$x \circ (e-z) = 0$	$x_i z_i = x_i$	$\to$	$\operatorname{diag}(U) = x$
$z \in \{0,1\}^n$	$z_i^2 = z_i$	$\to$	$\operatorname{diag}(Z) = z, \quad 0 \leq z \leq e$
$e^\top z \leq k$		$\to$	$e^\top z \leq k$
$\operatorname{rank}(M) = 1$	Drop $\operatorname{rank}(M) = 1$	$\to$	$M \succeq 0$

$$\begin{bmatrix} 1 \\ x \\ z \end{bmatrix}\begin{bmatrix} 1 \\ x \\ z \end{bmatrix}^\top = \begin{bmatrix} 1 & x^\top & z^\top \\ x & xx^\top & xz^\top \\ z & zx^\top & zz^\top \end{bmatrix} \longrightarrow M := \begin{bmatrix} \overbrace{\begin{matrix} 1 & x^\top \\ x & X \end{matrix}}^{Y} & \begin{matrix} z^\top \\ U \end{matrix} \\ \begin{matrix} z & U^\top \end{matrix} & Z \end{bmatrix}, \quad H := \begin{bmatrix} 1 & 0 \\ -d & B \end{bmatrix}, \quad \bar{Q} := \begin{bmatrix} 0 & c^\top \\ c & Q \end{bmatrix}$$

The sparsity cone

$$Y = \begin{bmatrix} a & x^\top \\ x & X \end{bmatrix} \in \begin{bmatrix} \mathbb{R} & \mathbb{R}^{1\times n} \\ \mathbb{R}^n & \mathbb{S}^n \end{bmatrix}$$



used by $\mathcal{D}$ not used by $\mathcal{D}$

Definition (Sparsity cone)

$$\mathcal{K} := \left\{ Y \in \mathbb{S}^{n+1} : \mathcal{D}(Y) := \begin{bmatrix} ka & x^\top \\ x & \operatorname{Diag}(\operatorname{diag} X) \end{bmatrix} \succeq 0 \right\}.$$

Theorem (Rank-one characterization)

For $Y \succeq 0, \quad a = 1 \quad (0/0 := 0)$:

$$Y \in \mathcal{K} \iff \sum_i \frac{x_i^2}{X_{ii}} \leq k \xLeftrightarrow{X=xx^\top} \|x\|_0 \leq k.$$

Equivalent SDP relaxations

SDP–RLT

$$\min_{x,X,z,U,Z} \quad \langle Q, X\rangle + 2c^\top x$$

$$\begin{aligned} \text{s.t.} \quad \mathcal{A}(Y) &= 0, \\ \mathcal{B}(Y) &\geq 0, \end{aligned}$$

$$\begin{aligned} & M \succeq 0, \\ & \operatorname{diag}(U) = x, \\ & \operatorname{diag}(Z) = z, \\ & \\ & 0 \leq z \leq e, \\ & e^\top z \leq k \end{aligned}$$

$\Longleftrightarrow$ U, Z

Reduced SDP–RLT

$$\min_{x,X,z} \quad \langle Q, X\rangle + 2c^\top x$$

$$\begin{aligned} \text{s.t.} \quad \mathcal{A}(Y) &= 0, \\ \mathcal{B}(Y) &\geq 0, \end{aligned}$$

$$\begin{aligned} & Y \succeq 0, \\ & \begin{bmatrix} z_i & x_i \\ x_i & X_{ii} \end{bmatrix} \succeq 0, \\ & i \in [n], \\ & 0 \leq z \leq e, \\ & e^\top z \leq k \end{aligned}$$

$\Longleftrightarrow$ z

SC-SDP

$$\min_{Y} \quad \langle \bar{Q}, Y\rangle$$

$$\begin{aligned} \text{s.t.} \quad \mathcal{A}(Y) &= 0, \\ \mathcal{B}(Y) &\geq 0, \end{aligned}$$

$$\begin{aligned} & Y \succeq 0, \\ & Y_{11} = 1, \\ & Y \in \mathcal{K} \end{aligned}$$

SC-SDP: equivalence and bound strength

Theorem (Relaxation equivalence)

$$v^{\text{SDP-RLT}} = v^{\text{reduced SDP-RLT}} = v^{\text{SC-SDP}} \leq v^{\text{P}}.$$

- Same SDP–RLT bound; lifted matrix dimension $2n + 1 \longrightarrow n + 1$.
- For sparse ridge regression, SC-SDP matches optimal perspective.

Details in the appendix ↗

Outline

Sparsity-cone SDP relaxation

RiNNAL-S: a low-rank augmented Lagrangian method

SDP-based fixing and screening

Numerical experiments

Augmented Lagrangian reformulation

Keep the PSD and equality constraints in

$$\mathcal{F} := \{Y \in \mathbb{S}_+^{n+1} : \mathcal{A}(Y) = 0,\ Y_{11} = 1\}.$$

SC-SDP

$$\begin{array}{ll} \min\limits_{Y} & \langle \bar{Q}, Y \rangle \\ \text{s.t.} & Y \in \mathcal{F}, \\ & \mathcal{B}(Y) \geq 0, \quad Y \in \mathcal{K}. \end{array}$$

$\Longleftrightarrow$

Slack reformulation

$$\begin{array}{ll} \min\limits_{Y,V,Z} & \langle \bar{Q}, Y \rangle \\ \text{s.t.} & Y \in \mathcal{F}, \quad V \geq 0, \quad Z \in \mathcal{K}, \\ & V - \mathcal{B}(Y) = 0, \quad Z - Y = 0. \end{array}$$

With equality multipliers Λ and T, minimize over V and Z:

$$L_\sigma(Y; \Lambda, T) := \min_{V \geq 0,\, Z \in \mathcal{K}} \left\{ \langle \bar{Q}, Y \rangle + \tfrac{\sigma}{2} \|V - \mathcal{B}(Y) + \sigma^{-1}\Lambda\|_F^2 + \tfrac{\sigma}{2} \|Z - Y + \sigma^{-1} T\|_F^2 \right\}$$

$$= \langle \bar{Q}, Y \rangle + \tfrac{\sigma}{2} \|(\sigma^{-1}\Lambda - \mathcal{B}(Y))_+\|_F^2 + \tfrac{\sigma}{2} \|\Pi_{\mathcal{K}^*}(\sigma^{-1} T - Y)\|_F^2.$$

Completing the squares omits constants independent of the primal variables.

Augmented Lagrangian algorithm

Algorithm (ALM)

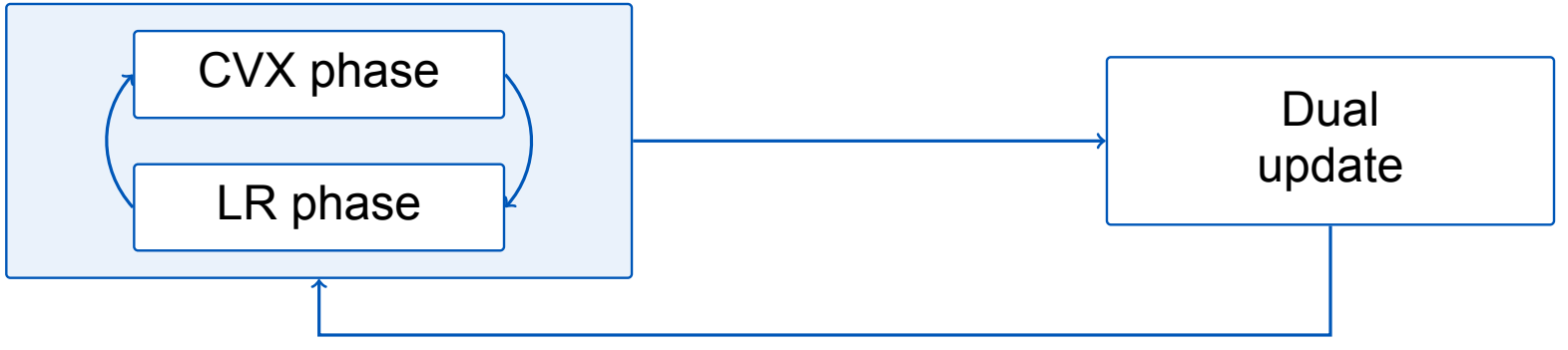
$$Y^{\ell+1} \approx \arg\min_{Y \in \mathcal{F}} L_{\sigma_\ell}(Y; \Lambda^\ell, T^\ell),$$

$$\Lambda^{\ell+1} = (\Lambda^\ell - \sigma_\ell \mathcal{B}(Y^{\ell+1}))_+,$$

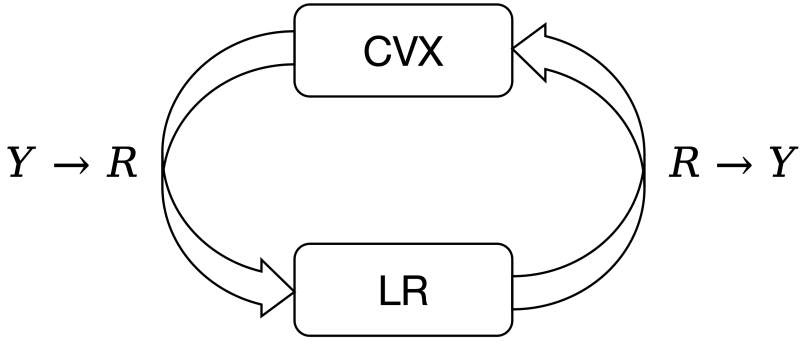
$$T^{\ell+1} = \Pi_{\mathcal{K}^*}(T^\ell - \sigma_\ell Y^{\ell+1}).$$

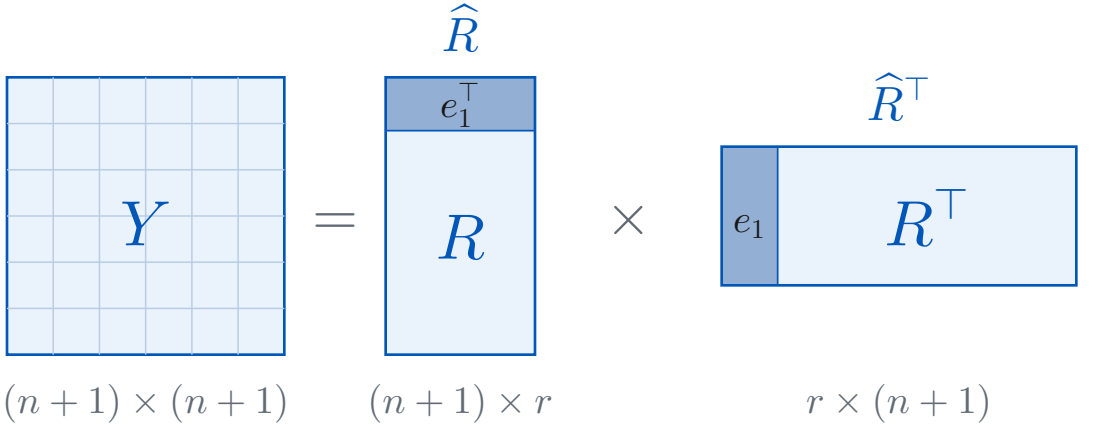
$\sigma_\ell > 0$ is the penalty parameter; solve the inner problem in two phases.

Primal update



A two-phase method for the ALM subproblem





CVX phase

$$\begin{aligned} \min_{Y} \quad & L_\sigma(Y; \Lambda, T) \\ \text{s.t.} \quad & Y \in \mathcal{F} \end{aligned}$$

- Projected-gradient step in Y.
- Global convergence.
- Automatically select rank for LR.

LR phase

$$\begin{aligned} \min_{R} \quad & L_\sigma(\widehat{R}\widehat{R}^\top; \Lambda, T) \\ \text{s.t.} \quad & AR = be_1^\top \end{aligned}$$

- Fewer variables: $R \in \mathbb{R}^{n \times r}$.
- Reduce to affine constraints.
- PSD holds automatically.

See RiNNAL+ [Hou et al., 2026] for more details.

Projection onto the sparsity cone

For a given matrix

$$G = \begin{bmatrix} \bar{a} & \bar{x}^\top \\ \bar{x} & \bar{X} \end{bmatrix}, \qquad \bar{d} := \operatorname{diag}(\bar{X}),$$

the Moreau decomposition gives

$$\Pi_{\mathcal{K}^*}(-G) = \Pi_{\mathcal{K}}(G) - G.$$

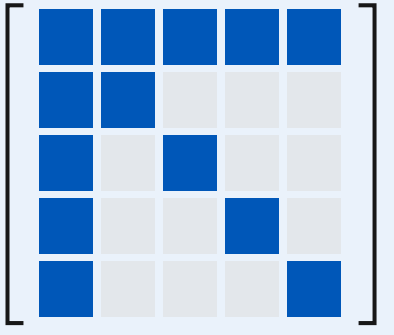
so we only need to compute the projection onto $\mathcal{K}$.

Proposition (Reduced projection)

$$\Pi_{\mathcal{K}}(G) = \begin{bmatrix} a^* & (x^*)^\top \\ x^* & \bar{X} + \operatorname{Diag}(d^* - \bar{d}) \end{bmatrix}$$

where (a^*, x^*, d^*) solve

$$\begin{aligned} \min_{a,x,d} \quad & \tfrac{1}{2}(a-\bar{a})^2 + \|x-\bar{x}\|^2 + \tfrac{1}{2}\|d-\bar{d}\|^2 \\ \text{s.t.} \quad & a \geq 0, \quad d \geq 0, \quad \sum_i x_i^2/d_i \leq ka. \end{aligned}$$



One scalar equation for the projection

If $(\bar{a}_+, \bar{x}, \bar{d}_+)$ is feasible, it is the projection. Otherwise:

For $\lambda > 0$, set $d_i(\lambda) = 0$ if $\bar{d}_i + \bar{x}_i^2/\lambda \leq 0$. Otherwise,

let $d_i(\lambda)$ be the unique positive root of the cubic equation

$$(d_i - \bar{d}_i)(d_i + \lambda)^2 = \lambda \bar{x}_i^2.$$

Proposition (One-dimensional reduction)

$$\psi(\lambda) := \sum_i \frac{\bar{x}_i^2 d_i(\lambda)}{(d_i(\lambda) + \lambda)^2} - k(\bar{a} + k\lambda)_+ = 0.$$

A positive root λ^* determines the projection.

- Solve the scalar equation using Newton and bisection.

Projection efficiency

Cone	n	Ours (s)	MOSEK (s)	Speedup	RelErr
$\mathcal{K}$	50	0.0034	2.6	754	1.28e-06
$\mathcal{K}$	200	0.0086	2448.2	283636	3.26e-07
$\mathcal{K}$	1000	0.0520	$>$3600.0	–	–
$\mathcal{K}$	10000	2.9000	$>$3600.0	–	–

Outline

Sparsity-cone SDP relaxation

RiNNAL-S: a low-rank augmented Lagrangian method

SDP-based fixing and screening

Numerical experiments

Presolving with SC-SDP

Goal: identify variables with $x_p^* = 0$ in every optimal solution.

Let p^* be the optimal value of the original sparse QP.

$$v^{\mathrm{LB}} \leq p^* \leq v^{\mathrm{UB}}.$$

A direct approach: for each index p, impose $z_p = 1$:

$$\underbrace{v^{\mathrm{LB}}(z_p = 1) > v^{\mathrm{UB}}}_{\text{index } p \text{ cannot be selected at optimum}} \quad \Longrightarrow \quad z_p^* = 0 \quad \Longrightarrow \quad x_p^* = 0.$$

- Solving a separate relaxation for every index is expensive.

> **Motivation**
>
> Obtain many fixing and screening decisions from **one solved SC-SDP relaxation**.

SC-SDP: primal and dual

Primal

$$\begin{aligned} \min_{Y} \quad & \langle \bar{Q}, Y \rangle \\ \text{s.t.} \quad & \mathcal{A}(Y) = 0, \quad \mathcal{B}(Y) \geq 0, \quad Y_{11} = 1, \\ & \mathcal{D}(Y) \succeq 0, \quad Y \succeq 0. \end{aligned}$$

Dual

$$\begin{aligned} \max_{\alpha,\Gamma,\Lambda,W,S} \quad & \alpha \\ \text{s.t.} \quad & \bar{Q} - \mathcal{A}^*(\Gamma) - \mathcal{B}^*(\Lambda) - \mathcal{D}^*(W) - S = \alpha E_{11}, \\ & \Lambda \geq 0, \quad W \succeq 0, \quad S \succeq 0. \end{aligned}$$

$v^{\mathrm{LB}} = \alpha$. W carries the support information.

Variable fixing

$$W = \begin{bmatrix} \tau & u^\top \\ u & V \end{bmatrix}, \qquad w_i := \frac{u_i^2}{V_{ii}}, \quad (0/0 := 0), \qquad w_{[1]} \geq \cdots \geq w_{[n]}.$$

Theorem (Dual fixing)

For an SC-SDP dual-feasible solution with value v^{LB}, let $\Delta := v^{\mathrm{UB}} - v^{\mathrm{LB}}$.
If $1 \leq k < n$, every optimal solution (x^*, z^*) of the original sparse QP satisfies

$$z_p^* = \begin{cases} 0, & \text{if } w_{[k]} - w_p > \Delta, \\ 1, & \text{if } w_p - w_{[k+1]} > \Delta. \end{cases}$$

Easy checking

- Scores use only squaring and division; sort once to test every index.

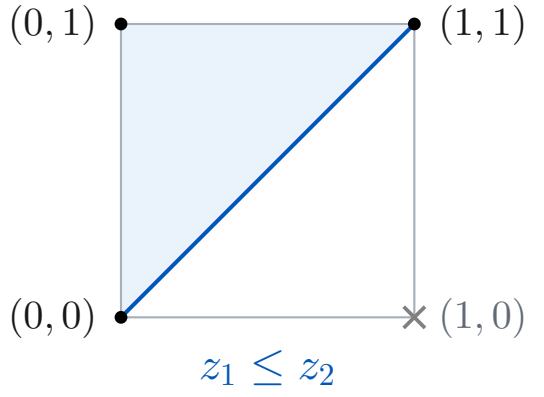
Screening cuts

For disjoint S, N with $|S| \le k$:

$$z_i = 1 \ (i \in S), \qquad z_i = 0 \ (i \in N).$$

Rule out this joint pattern at optimality.

Example: $S = \{1\},\ N = \{2\}$.



Theorem (Screening cuts)

C: indices with the top $k - |S|$ scores outside $S \cup N$, use all if fewer remain.
$C = \varnothing$ if $|S| = k$.

$$\underbrace{\sum_{i=1}^{k} w_{[i]} - \sum_{i \in S} w_i - \sum_{i \in C} w_i > \Delta}_{\text{Rule}} \quad \Longrightarrow \quad \underbrace{\sum_{i \in S} z_i + \sum_{i \in N} (1 - z_i) \le |S| + |N| - 1}_{\text{Cut}}$$

Dual refinement

Motivation

Exploit nonunique dual solutions to improve **presolving ability**, at the same lower bound.

Fix an optimal triple $(\bar{\alpha}, \bar{\Gamma}, \bar{\Lambda})$ and define

$$\bar{C} := \bar{Q} - \mathcal{A}^*(\bar{\Gamma}) - \mathcal{B}^*(\bar{\Lambda}) - \bar{\alpha} E_{11}.$$

$$\begin{aligned} \max_{W} \quad & \underbrace{\sum_{i=1}^{k} w_{[i]}}_{\text{large scores}} - \underbrace{\sum_{i=k+1}^{n} w_{[i]}}_{\text{small scores}} \\ \text{s.t.} \quad & \bar{C} - \mathcal{D}^*(W) \succeq 0, \quad W \succeq 0. \end{aligned}$$

Favor certificates whose scores more clearly separate selected and excluded variables.

An equivalent convex refinement problem

At dual optimality, the leading k scores can be chosen equal to τ:

$$w_{[1]} = \cdots = w_{[k]} = \tau \quad \Longrightarrow \quad \sum_{i=1}^{k} w_{[i]} - \sum_{i=k+1}^{n} w_{[i]} = 2k\tau - \sum_{i} w_i.$$

Otherwise, reallocating the dual matrices would increase the lower bound.

Introduce $w_i \le t_i \le \tau$ to obtain the equivalent **convex** formulation:

$$\begin{aligned} \min_{\tau,u,d,t} \quad & -2k\tau + \sum_{i=1}^{n} t_i \\ \text{s.t.} \quad & \bar{C} - \begin{bmatrix} k\tau & u^\top \\ u & \operatorname{Diag}(d) \end{bmatrix} \succeq 0, \\ & d_i \ge 0, \quad u_i^2 \le t_i d_i, \quad 0 \le t_i \le \tau \quad (i \in [n]). \end{aligned}$$

Few variables

- $3n + 1$ variables; SDP + second-order cones. Same lower bound.

Outline

Sparsity-cone SDP relaxation

RiNNAL-S: a low-rank augmented Lagrangian method

SDP-based fixing and screening

Numerical experiments

Experimental setup

- **Machine:** Intel Xeon E5-2680 v3; 96 GB RAM.
- **Software:** MATLAB R2023b.

Model	Solvers
SC-SDP	RiNNAL-S / MOSEK 11.0
SDP–RLT	RiNNAL+ / SDPNAL+
Original MIQP	Gurobi 13.0.0

- **Stopping (RiNNAL-S):**

$$R_{\max} := \max\{R_p, R_d, R_c\} < \mathrm{tol}, \quad \mathrm{tol} = 10^{-6}.$$

- **Time limit:** 3600 s.

Test problems

- Sparse ridge regression

$$\min_x \left\{ \tfrac{1}{m}\|Mx - y\|^2 + \gamma\|x\|^2 : \ \|x\|_0 \le k \right\}. \tag{SRR}$$

- Equality-constrained SRR

$$\min_x \left\{ \tfrac{1}{m}\|Mx - y\|^2 + \gamma\|x\|^2 : \ e^\top x = e^\top x^*, \ \|x\|_0 \le k \right\}. \tag{SRR-E}$$

- Sparse standard quadratic programming

$$\min_x \left\{ x^\top Q x : \ e^\top x = 1, \ x \ge 0, \ \|x\|_0 \le k \right\}. \tag{StQP}$$

- Sparse binary quadratic programming

$$\min_x \left\{ x^\top Q x + 2c^\top x : \ x \in \{0,1\}^n, \ \|x\|_0 \le k \right\}. \tag{SBQP}$$

- Sparse quadratic knapsack with conflicts

$$\min_x \left\{ x^\top Q x + 2c^\top x : \quad \begin{array}{ll} Gx \le d, & x_i x_j = 0 \ ((i,j) \in E), \\ x \in \{0,1\}^n, & \|x\|_0 \le k \end{array} \right\}. \tag{SQKP-C}$$

Relaxation strength: sparse StQP

Relative gaps (%)

Instance class	Standard SDP	SC-SDP
COP	∞	**10.56**
PSD	1.42	**0.02**
SPN	∞	**0.02**

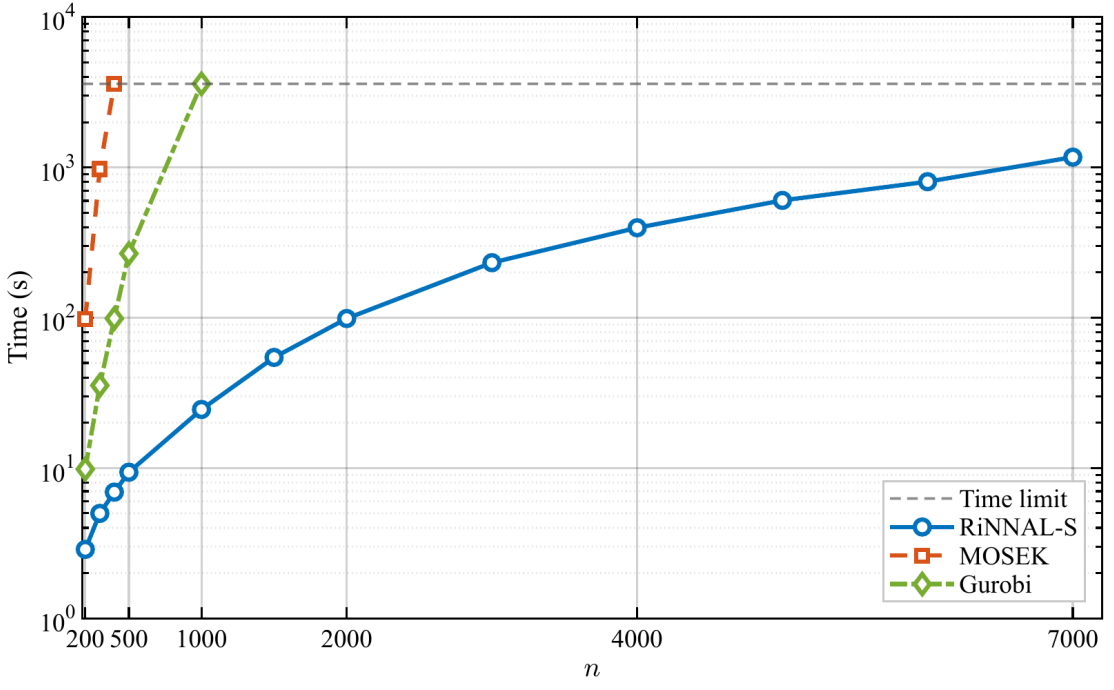
$$\mathrm{gap}(\%) = 100\,\frac{v^{\mathrm{UB}} - v^{\mathrm{LB}}}{\max\{1, |v^{\mathrm{UB}}|\}}.$$

- Instances: COP / PSD / SPN; $n = 100, \quad k = n/8$; MOSEK.
- Same upper bound; ∞ denotes an unbounded relaxation.

- **SC-SDP gives substantially stronger bounds than standard SDP.** Full results ↗

Scalability of RiNNAL-S

SRR: $m = 2n,\ k = 5,\ \rho = 0.1,\ \gamma = 1,\ \mathrm{SNR} = 1.$



- **RiNNAL-S solves instances up to $n = 7000$, beyond commercial solver limits.**
- SC-SDP is exact here, so its solution also solves the original SRR.

Solver speedup summary

Versus MOSEK

Problem	n	Speedup
SRR	1000	**140**
SBQP	250 – 500	**30 – 70**
SQKP-C	500 – 3000	Memory limit*

Versus RiNNAL+ (StQP)

Problem	n	Speedup
COP	100 – 200	**30 – 120**
PSD	100 – 500	**2 – 7**
SPN	100 – 200	**15 – 20**

- **RiNNAL-S delivers large speedups on the comparable benchmarks.**

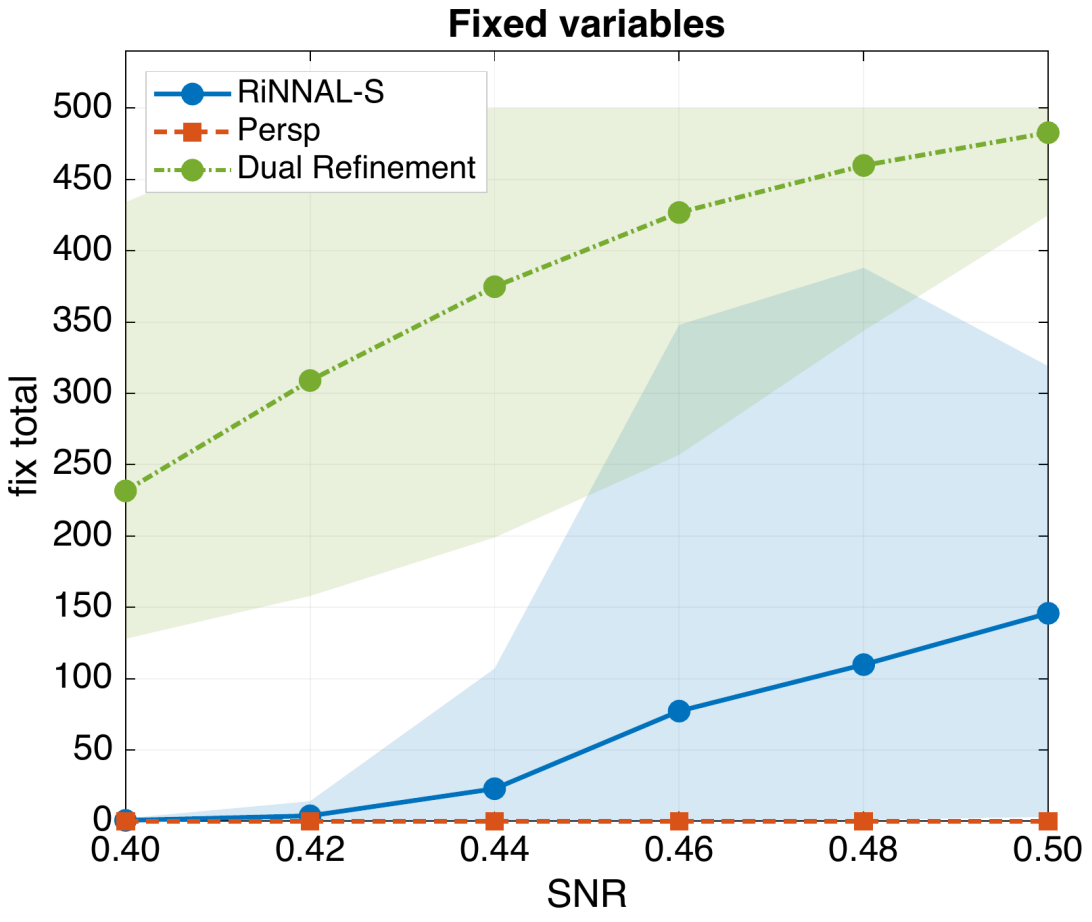
* MOSEK reached its memory limit.

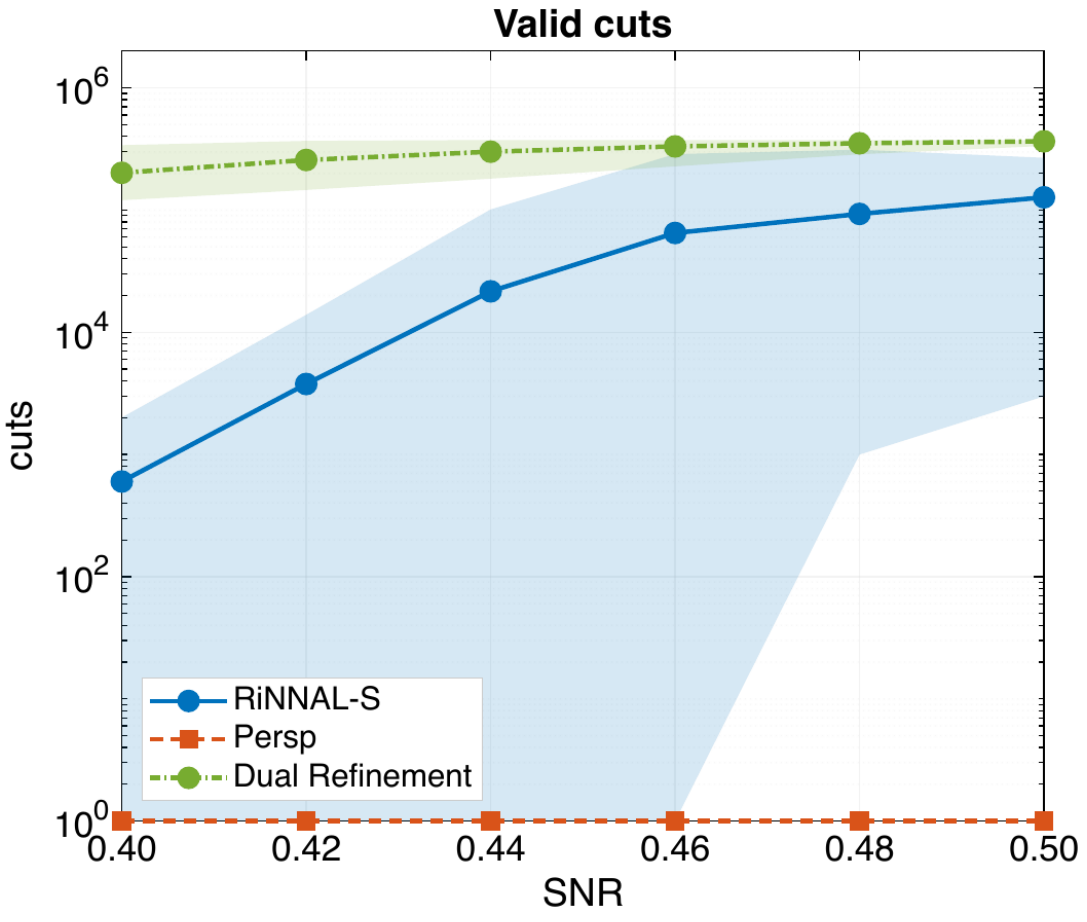
Presolving: experimental settings

- SRR: $n = 500,\ m = 1000,\ k = 5,\ \rho = 0.1,\ \gamma = 0.1$.
- SNR 0.40:0.02:0.50; five instances per setting.

Problem	Solver
Perspective relaxation	MOSEK
SC-SDP	RiNNAL-S
Dual refinement (DR)	MOSEK

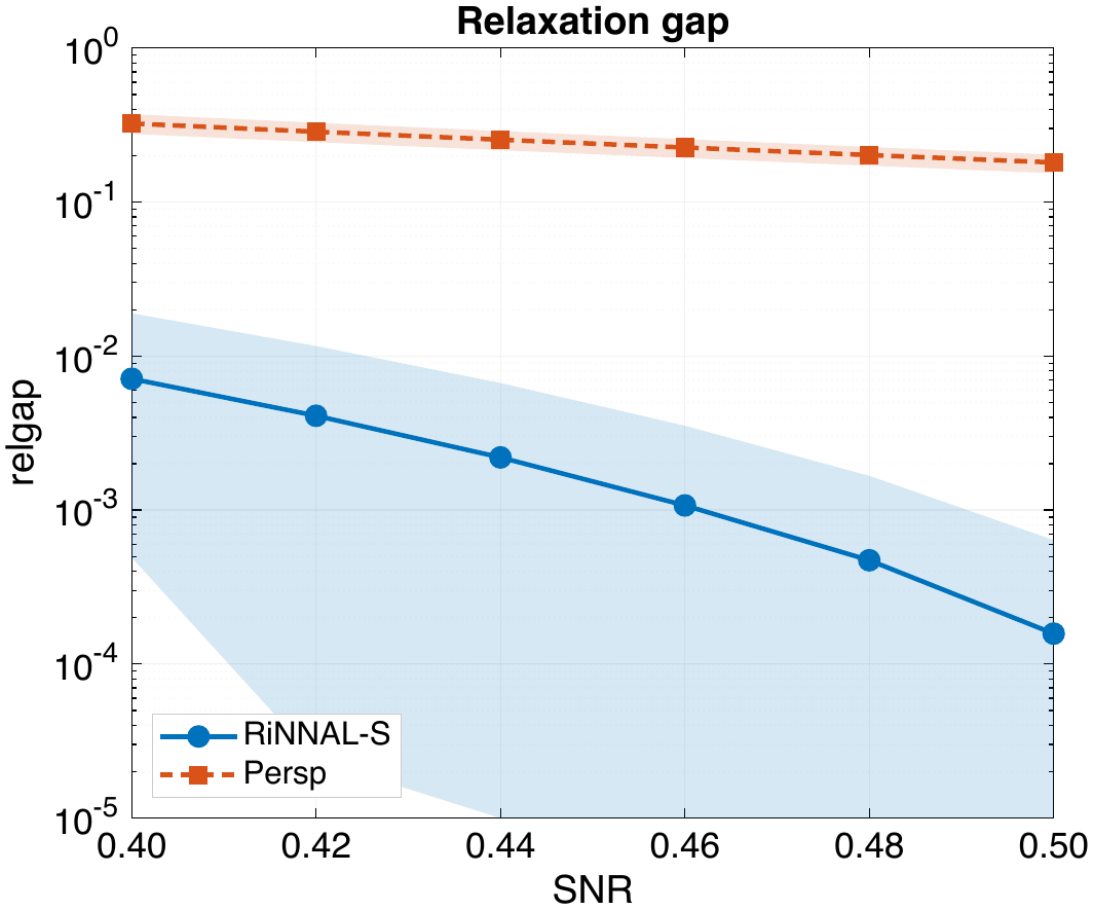
SRR: fixing and screening

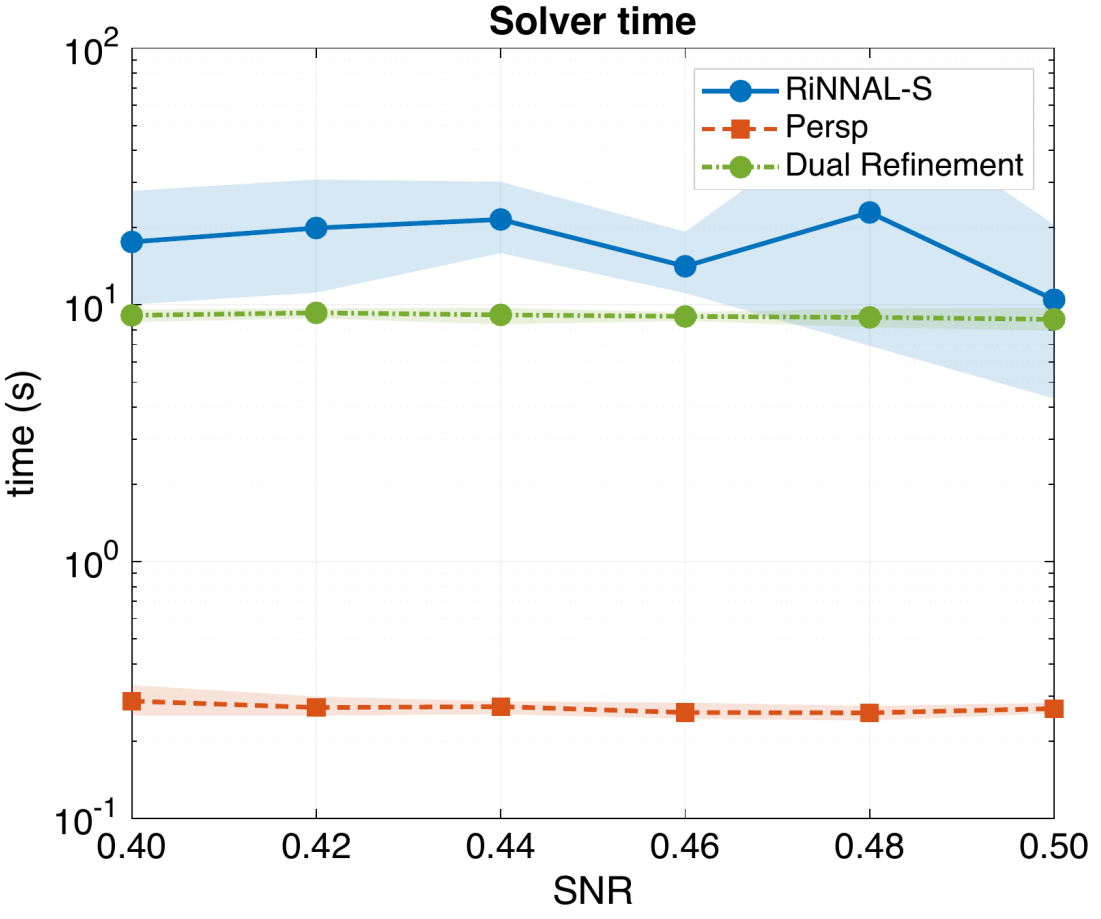




- **SC-SDP fixes more variables and finds more valid cuts; DR strengthens both.**

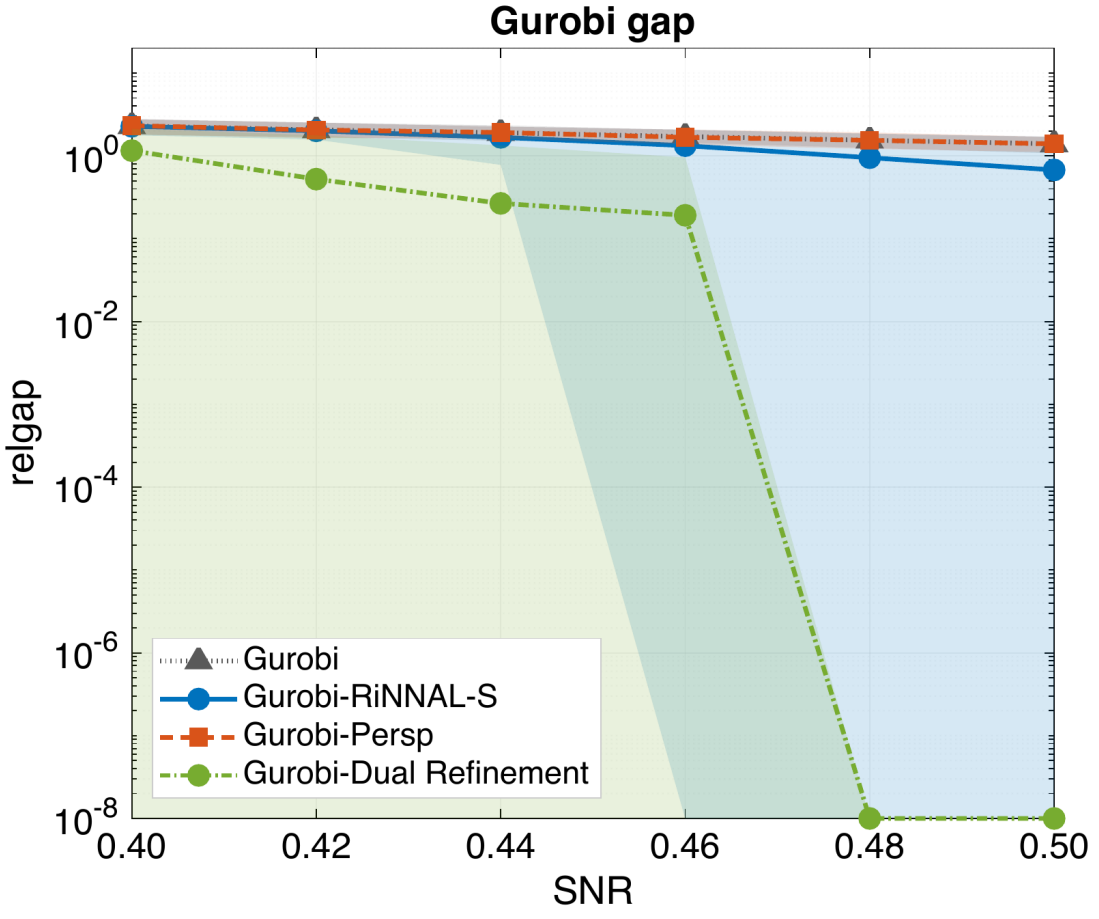
SRR: bounds and time



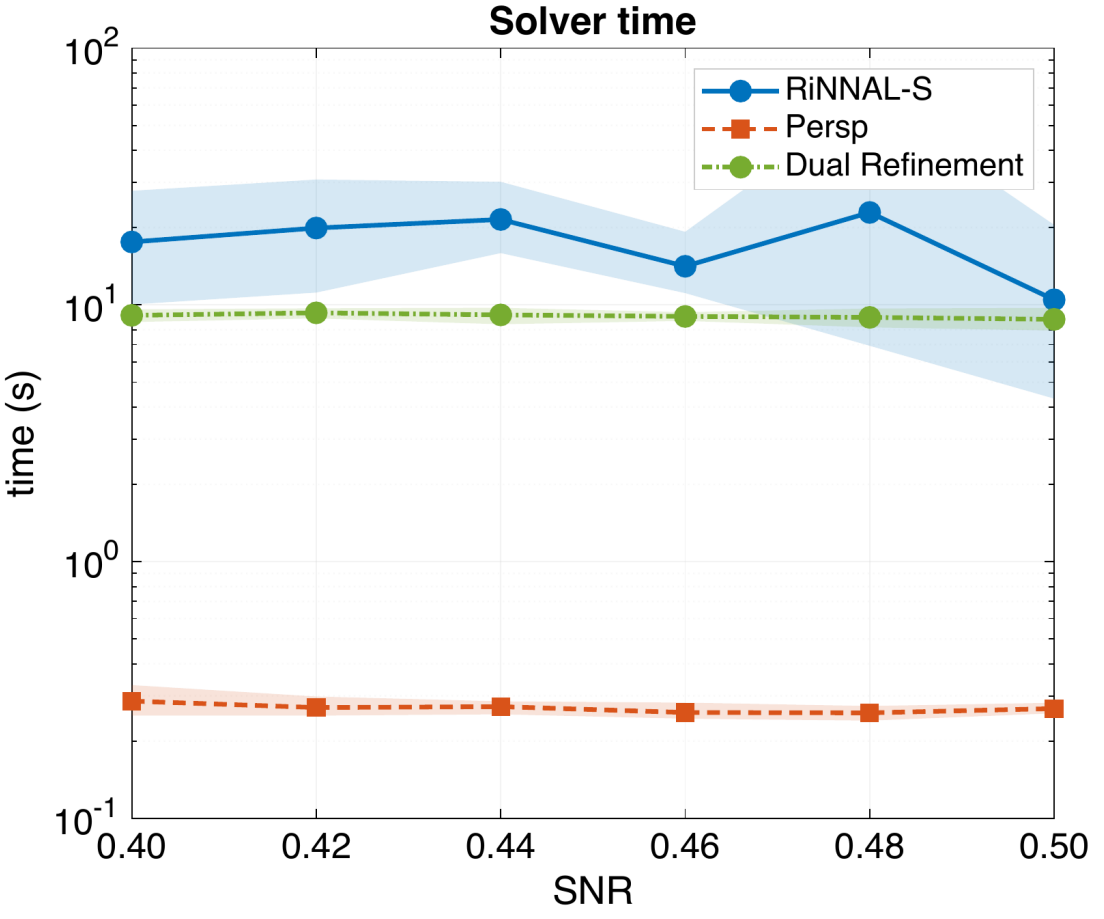


- **SC-SDP yields much smaller gaps than Persp; RiNNAL-S takes more time.**

SRR: MIP gap and refinement time



Additional cost of dual refinement (DR)



- Pass fixed variables and up to 100 filtered cuts.
- **SC-SDP + DR markedly improves Gurobi's MIP gap after 600 s.**

The RiNNAL Series

Problem	Relaxation	Solver	arXiv
Mixed-Binary QP	DNN	RiNNAL	2502.13849
Mixed-Binary QP	SDP–RLT	RiNNAL+	2507.13776
Polynomial OPT	Poly–SDP / Mom–SOS	RiNNAL–POP	2512.06359
Sparse QP	**SC-SDP**	**RiNNAL–S**	**2606.22894**

References

[1] D. Hou, T. Tang, and K.-C. Toh (2025). A low-rank augmented Lagrangian method for doubly nonnegative relaxations of mixed-binary quadratic programs. *Operations Research*, **74**(3), 1626–1647.

[2] D. Hou, T. Tang, and K.-C. Toh (2026). RiNNAL+: a Riemannian ALM solver for SDP-RLT relaxations of mixed-binary quadratic programs. *Mathematical Programming Computation*, **18**(3), 679–735.

[3] D. Hou, T. Tang, and K.-C. Toh (2026). A low-rank augmented Lagrangian method for polyhedral-SDP and moment-SOS relaxations of polynomial optimization. *Mathematical Programming*. doi:10.1007/s10107-026-02395-5.

[4] D. Hou, T. P. D. Nguyen, K.-C. Toh, and G. Wang (2026). Sparsity-cone SDP relaxations and applications to variable fixing for sparse quadratic programs. arXiv preprint arXiv:2606.22894.

Convex lifting phase

Lift the low-rank iterate back to the convex variable:

$$Y = \widehat{R}\widehat{R}^\top, \qquad Y^+ = \Pi_{\mathcal{F}}\big(Y - \eta \nabla_Y L_\sigma(Y; \Lambda, T)\big).$$

- Use the projected-gradient step to improve the convex subproblem.
- Factorize Y^+ to initialize the next low-rank phase:

$$Y^+ = \widehat{R}^+(\widehat{R}^+)^\top, \qquad r^+ = \operatorname{rank}(Y^+).$$

- Alternate the two phases until the ALM accuracy requirement is met.
- Projection onto $\mathcal{F}$ also admits a one-dimensional reduction. Details in the appendix ↗

Scores from the dual multiplier

$$W = \begin{bmatrix} \tau & u^\top \\ u & V \end{bmatrix} \succeq 0, \qquad d_i := V_{ii}.$$

$$\langle \mathcal{D}(Y), W \rangle = k\tau + \sum_i (d_i x_i^2 + 2u_i x_i), \qquad Y = [1; x][1; x]^\top.$$

Definition (Dual scores)

$$w_i := \begin{cases} u_i^2/d_i, & d_i > 0, \\ 0, & d_i = 0, \end{cases} \qquad 0 \le w_i \le \tau.$$

$$d_i x_i^2 + 2u_i x_i \ge -w_i, \qquad x_i = 0 \implies d_i x_i^2 + 2u_i x_i = 0.$$

A support-dependent lower bound

For any feasible (x, z), let $J := \{i : z_i = 1\},\ |J| \le k$. Then

$$f(x) \ge v^{\mathrm{LB}} + k\tau - \sum_{i \in J} w_i$$
$$\ge v^{\mathrm{LB}} + \sum_{i=1}^{k} w_{[i]} - \sum_{i \in J} w_i,$$

where $w_{[1]} \ge \cdots \ge w_{[n]}$. Define $\Delta := v^{\mathrm{UB}} - v^{\mathrm{LB}}$.

$$\sum_{i=1}^{k} w_{[i]} - \sum_{i \in J} w_i > \Delta \quad \Longrightarrow \quad J \text{ cannot be optimal.}$$

- Omitting large scores raises the lower bound for that support.

Screening cuts: full statement

Consider a joint pattern: $z_i = 1\ (i \in S),\ z_i = 0\ (i \in N)$, with $S \cap N = \varnothing,\ |S| \le k$.

$$R := [n] \setminus (S \cup N), \quad c := \min\{k - |S|, |R|\}, \quad C := \{i \in R : w_i \ge (w_R)_{[c]}\}.$$

Set $C = \varnothing$ if $c = 0$; ties are included in C.

Theorem (Screening cuts)

Under the same assumptions as in the dual-fixing theorem,

$$\sum_{i=1}^{k} w_{[i]} - \sum_{i \in S} w_i - \sum_{i \in C} w_i > \Delta$$

$$\Longrightarrow \quad \sum_{i \in S} z_i + \sum_{i \in N} (1 - z_i) \le |S| + |N| - 1.$$

Dual refinement: equivalence

For any feasible dual certificate, define

$$\eta := w_{[k]}, \qquad \delta := k\tau - \sum_{i=1}^{k} w_{[i]} \geq 0, \qquad r_i := \begin{cases} \eta/w_i, & w_i > \eta, \\ 1, & w_i \leq \eta. \end{cases}$$

Set $\widehat{\tau} = \eta,\ \widehat{u}_i = r_i u_i,\ \widehat{d}_i = r_i d_i$. These entries admit a PSD completion $\widehat{W}$.

$$H := \mathcal{D}^*(W) - \mathcal{D}^*(\widehat{W}) - \delta E_{11} \succeq 0.$$

Indeed, its arrow Schur complement is zero:

$$\sum_{w_i > \eta} \frac{((1-r_i)u_i)^2}{(1-r_i)d_i} = \sum_{w_i > \eta} (w_i - \eta) = \sum_{i=1}^{k} w_{[i]} - k\eta.$$

Thus $(\alpha + \delta, \Gamma, \Lambda, \widehat{W}, S + H)$ is dual feasible. At optimality, $\delta = 0$.

$$w_{[1]} = \cdots = w_{[k]} = \tau, \qquad \sum_{i=1}^{k} w_{[i]} - \sum_{i=k+1}^{n} w_{[i]} = 2k\tau - \sum_{i} w_i.$$

Projection onto the affine PSD set

$$P = \begin{bmatrix} b^\top \\ -A^\top \end{bmatrix}, \qquad J = I - P(P^\top P)^{-1}P^\top.$$

Theorem (Projection onto $\mathcal{F}$)

$$\Pi_{\mathcal{F}}(G) = \Pi_{\mathbb{S}_+^{n+1}}\big(J(G + \eta^* E_{11})J\big),$$
$$\eta^* \in \arg\min_{\eta\in\mathbb{R}} \left\{ \tfrac{1}{2}\|\Pi_{\mathbb{S}_+^{n+1}}(J(G+\eta E_{11})J)\|_F^2 - \eta \right\}.$$

- Assume A has full row rank and $Ax = b$ is feasible.

Sparse StQP: Full Gap Table

Family / method	$n = 25$	$n = 50$	$n = 75$	$n = 100$
COP / SDP	∞	∞	∞	∞
COP / SC-SDP	10.56	10.56	10.56	10.56
PSD / SDP	1.80	2.21	2.33	1.42
PSD / SC-SDP	0.00	0.00	0.09	0.02
SPN / SDP	∞	∞	∞	∞
SPN / SC-SDP	0.00	0.00	0.09	0.02

Relative gaps in percent. Values of 0.00 are rounded numerical results.

Relation to Perspective Relaxations

Sparse ridge regression · no additional continuous constraints

$$\begin{aligned} v_{\mathrm{PR}}(\mu) = \quad & \min_{x,z} \quad && x^\top (Q - \operatorname{Diag}\mu)x + 2c^\top x + \sum_i \mu_i \frac{x_i^2}{z_i} \\ & \text{s.t.} && 0 \le z \le e, \quad e^\top z \le k. \end{aligned}$$

$$Q = \tfrac{1}{m} M^\top M + \gamma I, \qquad c = -\tfrac{1}{m} M^\top y$$

Theorem (Optimal perspective relaxation)

$$v_{\mathrm{PR}}(\gamma e) \;\le\; \max_{\substack{\mu \ge 0 \\ Q - \operatorname{Diag}\mu \succeq 0}} v_{\mathrm{PR}}(\mu) \;=\; v_{\mathrm{SC-SDP}}$$

Common constant $\|y\|^2/m$ omitted; closed perspective at $z_i = 0$.

Presolving: full experimental settings

SRR and SRR-E: $n = 500,\ m = 1000,\ k = 5,\ \rho = 0.1,\ \gamma = 0.1$.

Problem	Certificates	Test range
SRR	Perspective / SC-SDP / SC-SDP + DR	SNR 0.40:0.02:0.50
SRR-E	SC-SDP / SC-SDP + DR	SNR 0.32:0.04:0.56
StQP	MOSEK dual solutions; no DR	COP / PSD / SPN

We evaluate three aspects for each problem:

- **Bounds and time:** relaxation gap and certificate cost.
- **Fixing and screening:** fixed variables and valid pairwise cuts.
- **Additional cuts and MIP:** cuts beyond fixing and the 600 s MIP gap.

SRR bands: min–max over five instances. DR = dual refinement.

Sparse StQP: COP solver results (1/2)

RiNNAL-S solves SC-SDP; RiNNAL+ and SDPNAL+ use stronger SDP–RLT reference models.

Problem	Algorithm	Iteration	Rank	$R_{\max}$	Objective	Time (s)
COP 100	RiNNAL-S	10, 550, 2	3	5.04e-07	-1.0195884e-01	1.0
COP 100	RiNNAL+	125, 6300, 25	102	6.64e-07	-1.0190589e-01	30.7
COP 100	SDPNAL+	76, 112, 1170	102	8.59e-07	-1.0196083e-01	40.1
COP 200	RiNNAL-S	12, 650, 3	3	9.63e-07	-1.0444021e-01	2.0
COP 200	RiNNAL+	300, 15050, 60	202	9.99e-07	-1.0402158e-01	230.5
COP 200	SDPNAL+	103, 145, 1511	202	9.24e-07	-1.0445084e-01	170.7
COP 500	RiNNAL-S	15, 800, 3	4	9.46e-07	-1.0493236e-01	5.4
COP 500	RiNNAL+	672, 33650, 135	621	1.20e-04†	-4.8722936e-01	3600.0
COP 500	SDPNAL+	200, 227, 2350	502	9.28e-07	-1.0466603e-01	1779.7

Tolerance 10^{-6}; limit 3600 s. †: tolerance not reached; —: no result at the target accuracy.

Sparse StQP: COP solver results (2/2)

RiNNAL-S solves SC-SDP; RiNNAL+ and SDPNAL+ use stronger SDP–RLT reference models.

Problem	Algorithm	Iteration	Rank	$R_{\max}$	Objective	Time (s)
COP 1000	RiNNAL-S	20, 1050, 4	3	8.45e-07	-1.0530038e-01	58.0
COP 1000	RiNNAL+	—	—	—	—	$>$3600.0
COP 1000	SDPNAL+	—	—	—	—	$>$3600.0
COP 2000	RiNNAL-S	37, 1900, 8	3	4.21e-07	-1.0542925e-01	373.3
COP 2000	RiNNAL+	—	—	—	—	$>$3600.0
COP 2000	SDPNAL+	—	—	—	—	$>$3600.0

Tolerance 10^{-6}; limit 3600 s. †: tolerance not reached; —: no result at the target accuracy.

Sparse StQP: PSD solver results

RiNNAL-S solves SC-SDP; RiNNAL+ and SDPNAL+ use stronger SDP–RLT reference models.

Problem	Algorithm	Iteration	Rank	$R_{\max}$	Objective	Time (s)
PSD 100	RiNNAL-S	44, 4400, 1	7	9.82e-07	1.4002781e-02	11.7
PSD 100	RiNNAL+	374, 18750, 75	5	8.59e-07	1.4174794e-02	80.7
PSD 100	SDPNAL+	801, 1058, 17306	2	8.50e-07	1.4174999e-02	468.1
PSD 200	RiNNAL-S	322, 32200, 0	44	9.82e-07	8.9018590e-03	137.0
PSD 200	RiNNAL+	360, 18050, 72	108	9.50e-07	9.0349366e-03	231.0
PSD 200	SDPNAL+	801, 905, 36057	89	1.84e-06†	9.3529559e-03	3600.0
PSD 500	RiNNAL-S	347, 34700, 0	99	9.16e-07	2.9152494e-03	450.1
PSD 500	RiNNAL+	507, 25400, 102	234	9.88e-07	2.9188026e-03	1838.0
PSD 500	SDPNAL+	426, 430, 4726	160	7.73e-06†	1.0912535e-02	3600.0

Tolerance 10^{-6}; limit 3600 s. †: tolerance not reached; —: no result at the target accuracy.

Sparse StQP: SPN solver results

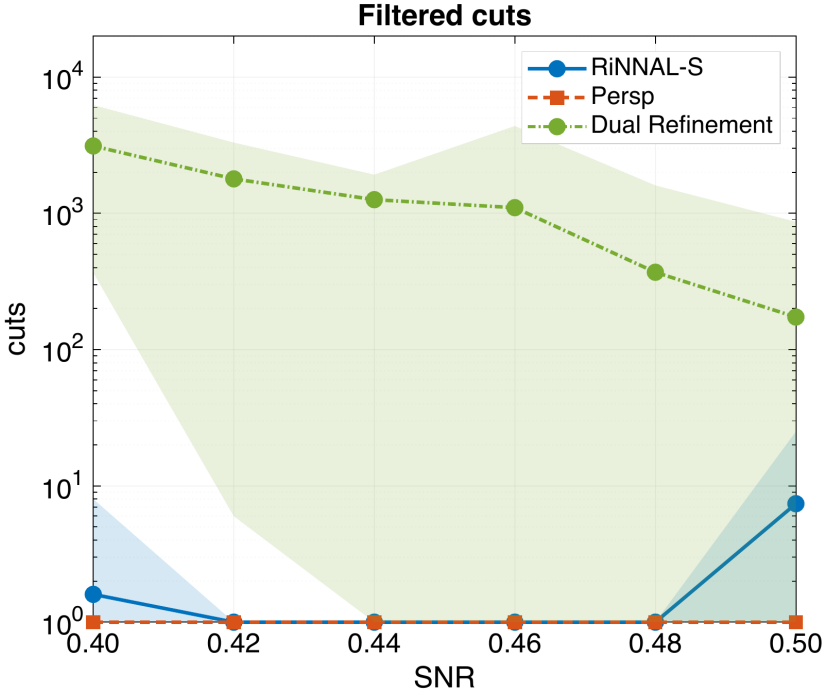
RiNNAL-S solves SC-SDP; RiNNAL+ and SDPNAL+ use stronger SDP–RLT reference models.

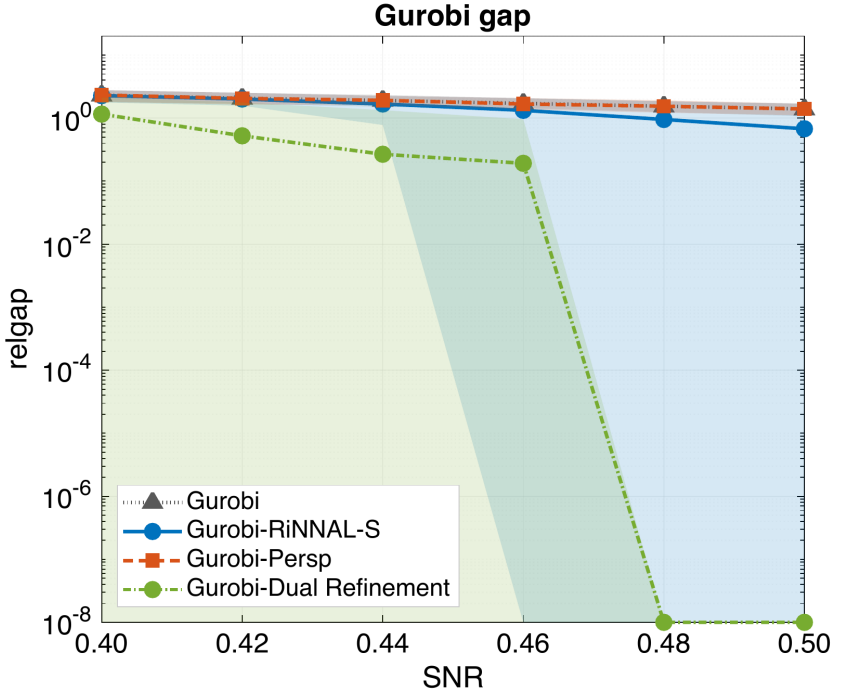
Problem	Algorithm	Iteration	Rank	$R_{\max}$	Objective	Time (s)
SPN 100	RiNNAL-S	119, 6000, 1	4	8.60e-07	1.4007862e-02	13.8
SPN 100	RiNNAL+	747, 37400, 150	107	9.31e-07	1.3592384e-02	204.0
SPN 100	SDPNAL+	801, 1032, 29187	12	9.94e-07	1.4887842e-02	866.1
SPN 200	RiNNAL-S	391, 19600, 0	8	9.28e-07	8.9032518e-03	74.3
SPN 200	RiNNAL+	1812, 90650, 363	248	9.97e-07	-1.2530973e-02	1556.1
SPN 200	SDPNAL+	801, 896, 38599	25	1.02e-06†	1.3190619e-02	3600.0
SPN 500	RiNNAL-S	462, 167810, 0	15	9.78e-07	2.9193141e-03	3552.5
SPN 500	RiNNAL+	—	—	—	—	$>$3600.0
SPN 500	SDPNAL+	—	—	—	—	$>$3600.0

Tolerance 10^{-6}; limit 3600 s. †: tolerance not reached; —: no result at the target accuracy.

SRR: additional cuts and MIP

Perspective, SC-SDP, and SC-SDP + DR; SNR 0.40–0.50.

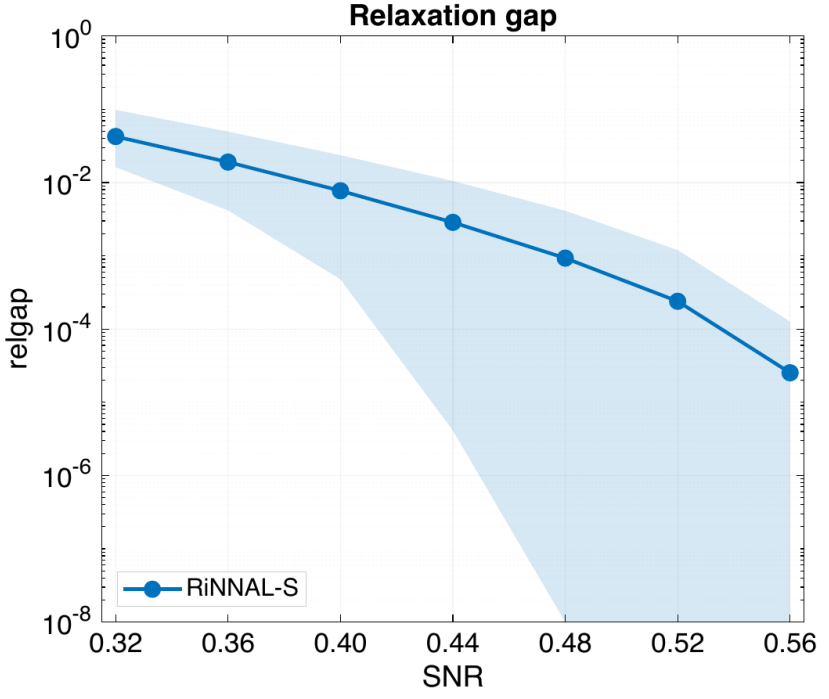


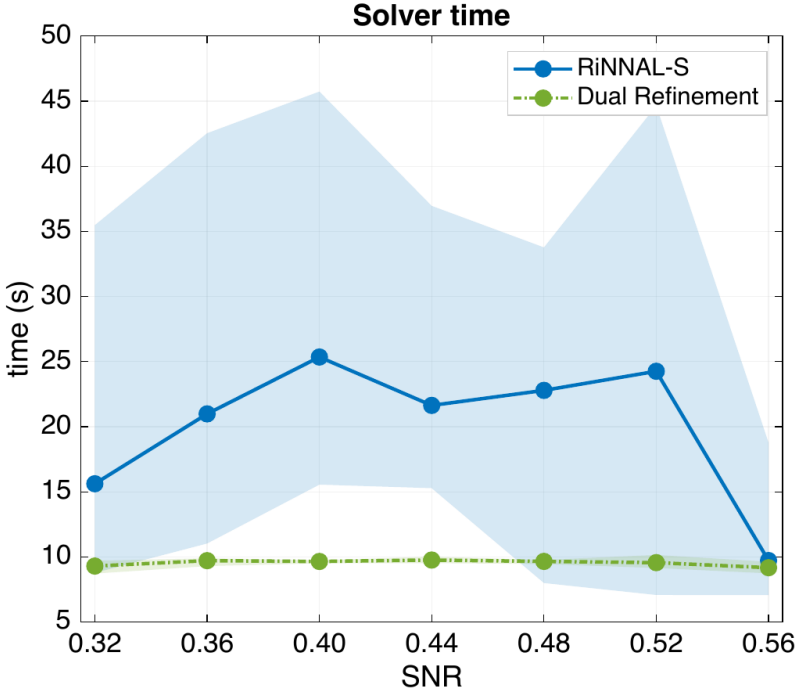


Pass fixed variables and up to 100 filtered cuts to Gurobi; MIP time excludes presolving.

SRR-E: bounds and time

SC-SDP and SC-SDP + DR; SNR 0.32–0.56.

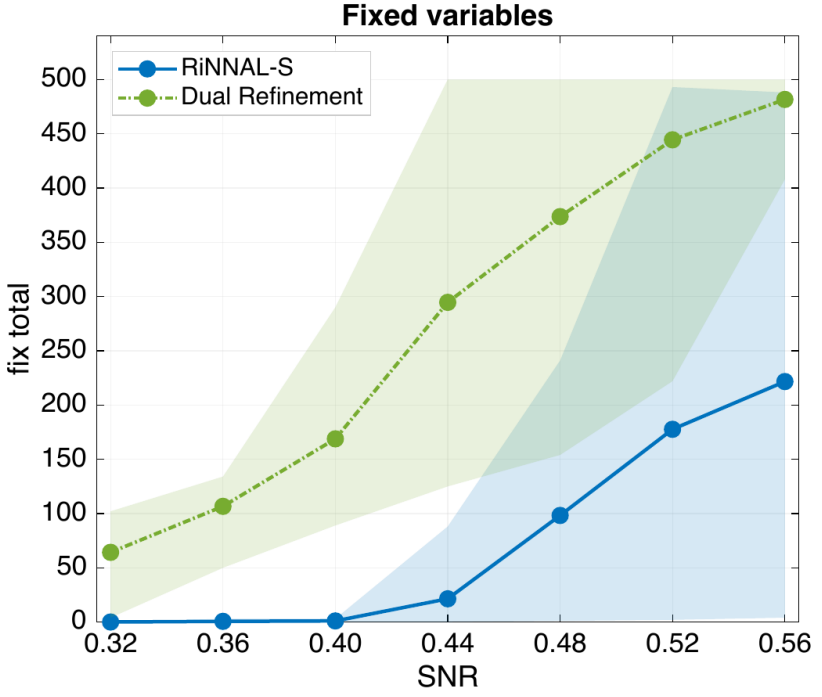


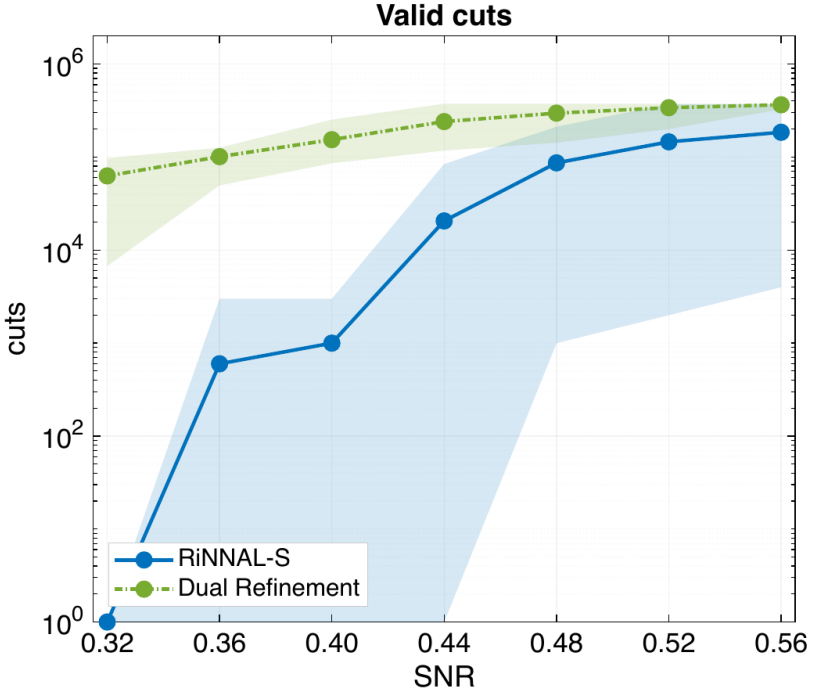


Relaxation gap measures bound quality; the time plot shows certificate cost.

SRR-E: fixing and screening

SC-SDP and SC-SDP + DR; SNR 0.32–0.56.

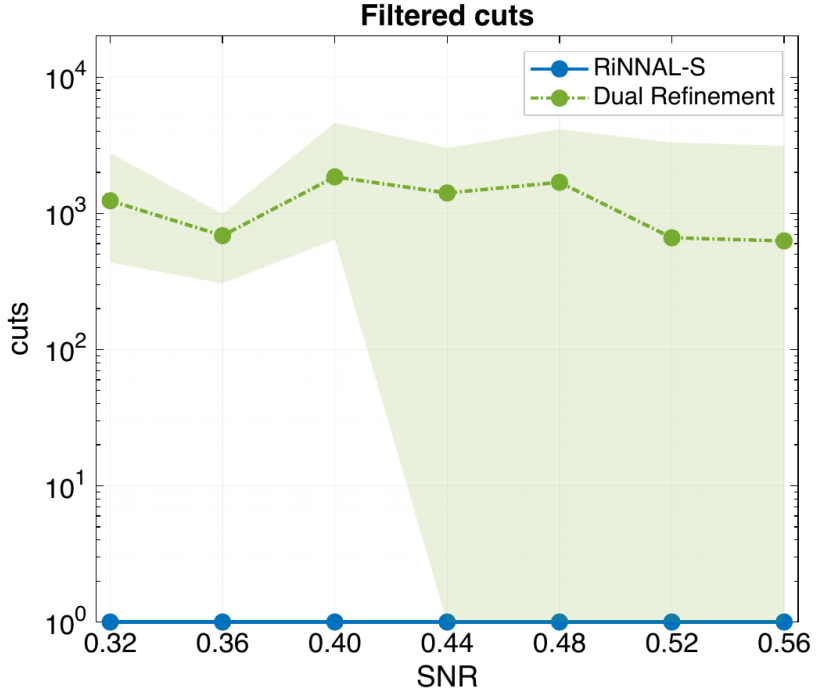


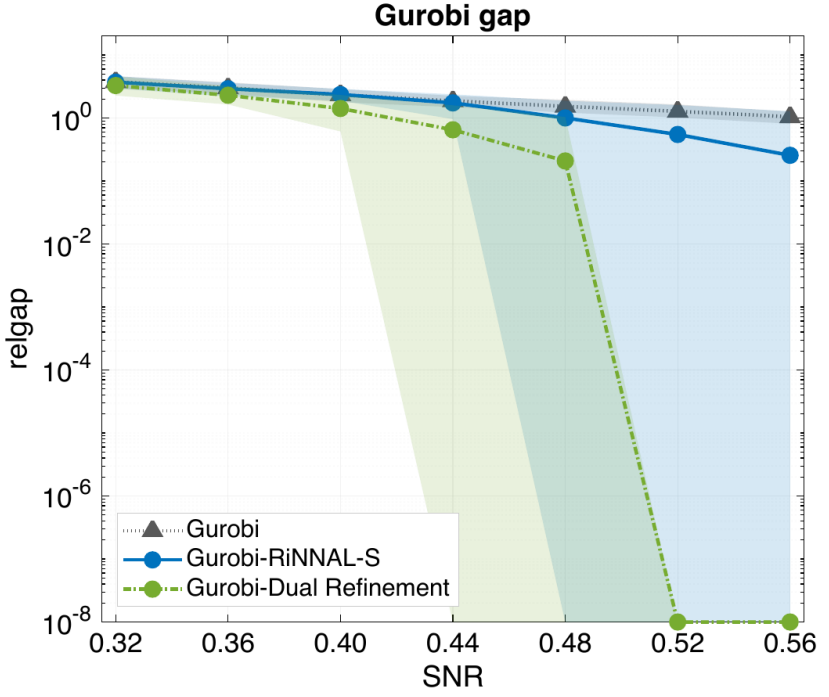


Count all valid pairwise cuts here, before removing those implied by variable fixing.

SRR-E: additional cuts and MIP

SC-SDP and SC-SDP + DR; SNR 0.32–0.56.

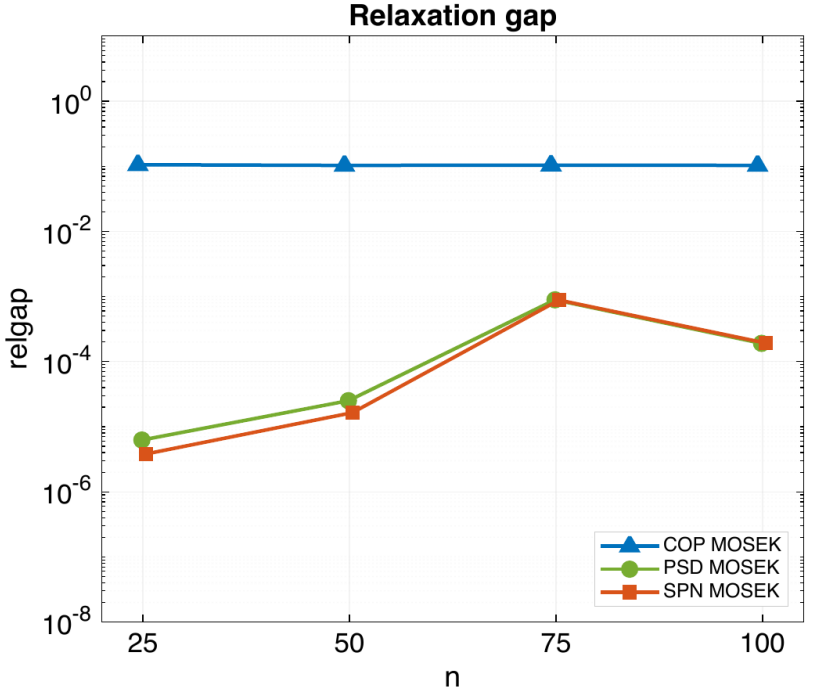


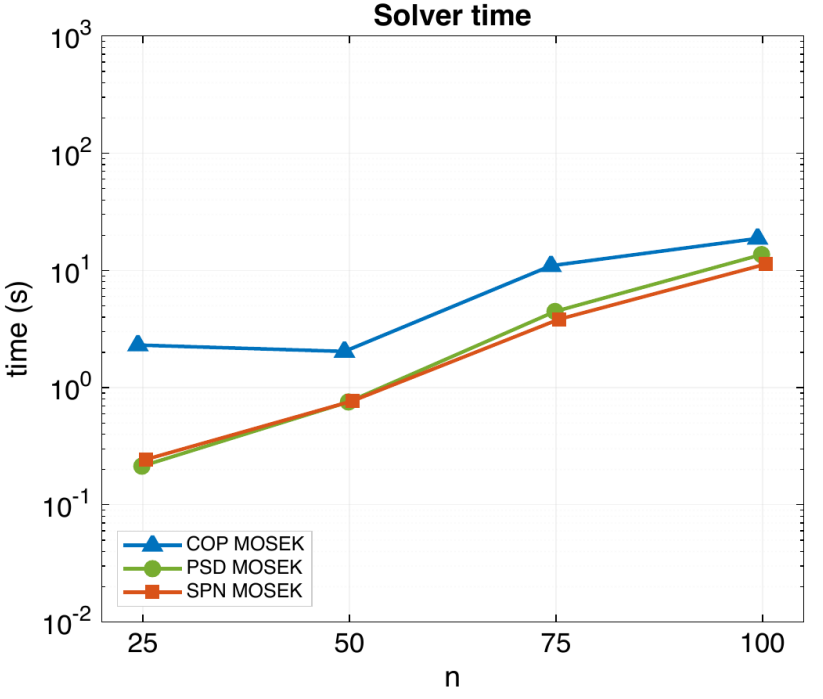


Pass fixed variables and up to 100 filtered cuts to Gurobi; MIP time excludes presolving.

StQP: bounds and time

MOSEK certificates at tolerance 10^{-8}; no dual refinement.

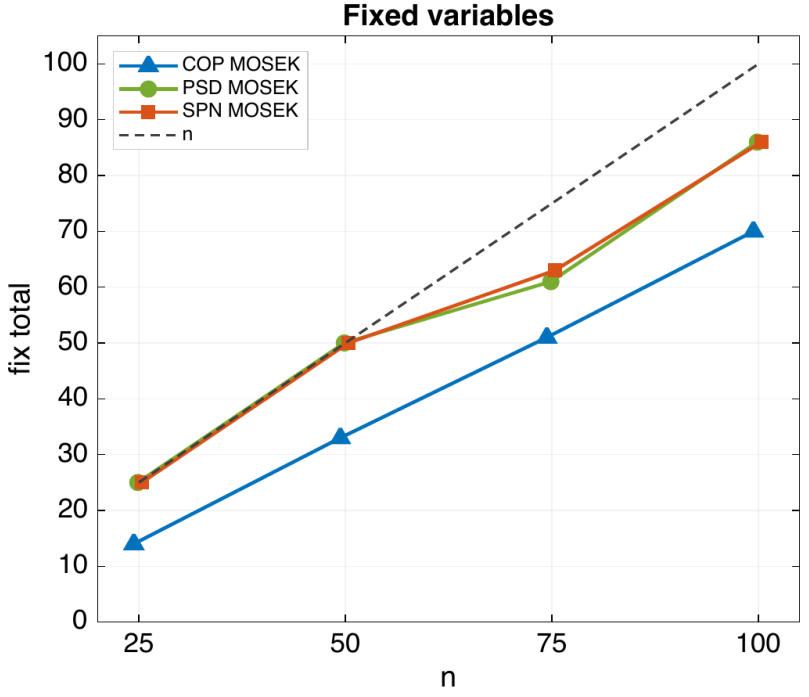


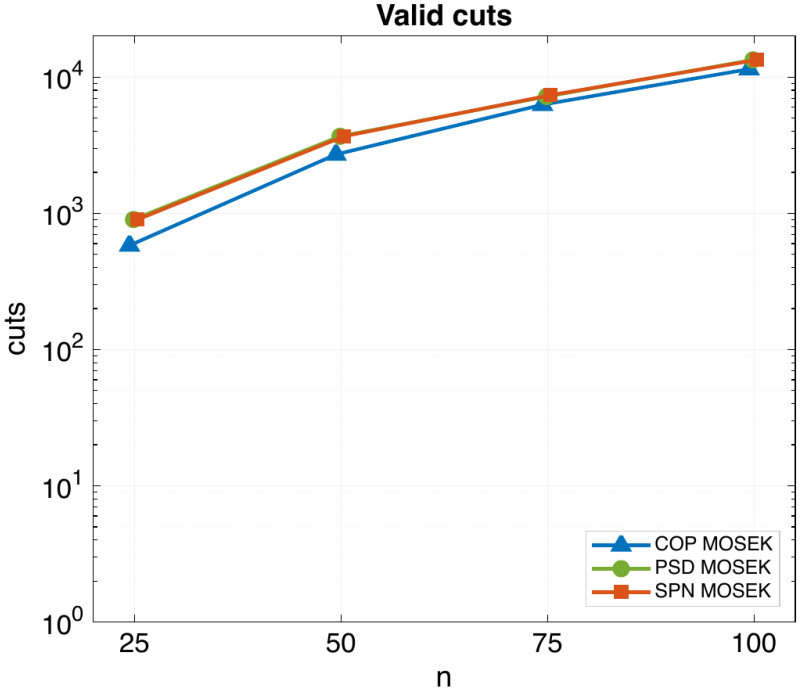


Relaxation gap measures bound quality; the time plot shows certificate cost.

StQP: fixing and screening

MOSEK certificates at tolerance 10^{-8}; no dual refinement.

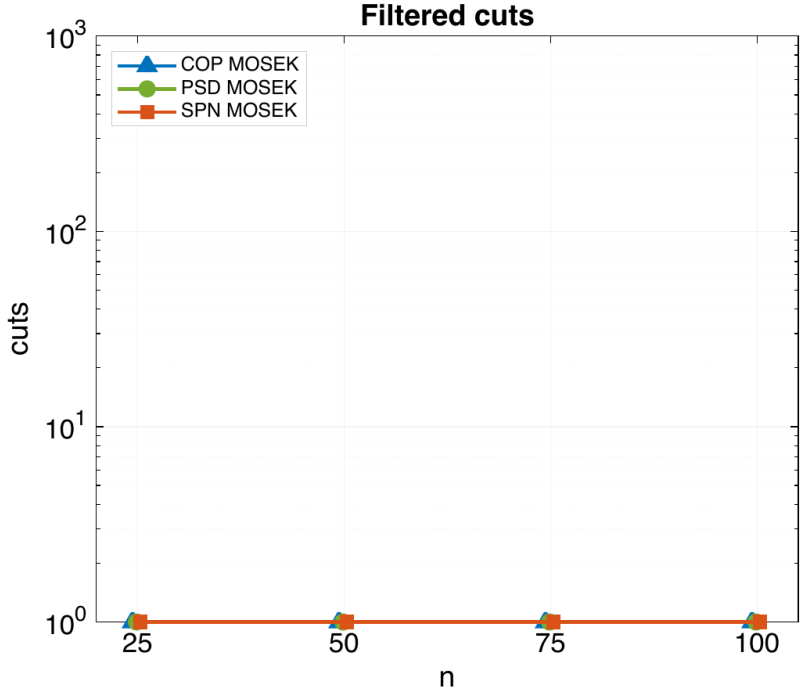


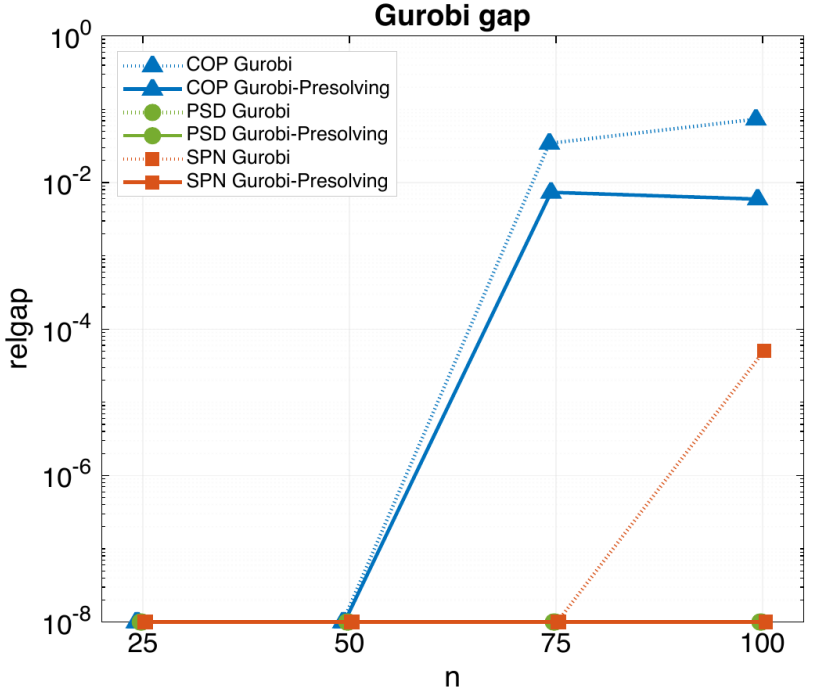


Count all valid pairwise cuts here, before removing those implied by variable fixing.

StQP: additional cuts and MIP

MOSEK certificates at tolerance 10^{-8}; no dual refinement.





No cuts remain beyond fixing; MIP gains here come from variable fixing.

SBQP: solver results

ORLIB: first instance at each size; $k = n/5$. Add $\mathrm{diag}(X) = x$ and $Y \geq 0$ entrywise.

n	Algorithm	Iteration	Rank	$R_{\max}$	Objective	Time (s)
250	RiNNAL-S	198, 4040, 52	31	9.21e-07	-2.0241801e+04	30.3
250	MOSEK	—	38	6.55e-07	-2.0241970e+04	1033.2
500	RiNNAL-S	130, 2680, 32	57	2.99e-07	-5.6538511e+04	46.9
500	MOSEK	—	—	—	—	>3600.0
1000	RiNNAL-S	189, 3860, 50	89	8.32e-07	-1.7015018e+05	325.0
1000	MOSEK	—	—	—	—	>3600.0
2500	RiNNAL-S	198, 4040, 57	167	6.61e-07	-7.0768612e+05	2768.4
2500	MOSEK	—	—	—	—	>3600.0

Tolerance 10^{-6}; limit 3600 s. MOSEK iteration counts are not reported.

SQKP-C: solver results

10 knapsacks; $k = \lceil 0.2n \rceil$; quadratic density 0.25; conflict density ρ.

ρ	n	Algorithm	Iteration	Rank	$R_{\max}$	Objective	Time (s)
0.1	1000	RiNNAL-S	78, 870, 16	198	9.33e-07	-1.0932744e+05	142.9
0.1	2000	RiNNAL-S	108, 1170, 22	388	6.40e-07	-2.2755262e+05	940.1
0.1	3000	RiNNAL-S	102, 1110, 21	580	9.95e-07	-3.4644352e+05	3196.9
0.01	500	RiNNAL-S	789, 7980, 158	33	9.99e-07	-1.5698783e+05	536.7
0.01	1000	RiNNAL-S	360, 3690, 72	52	9.64e-07	-4.9705848e+05	616.8
0.01	1500	RiNNAL-S	408, 4170, 82	76	9.00e-07	-9.5741504e+05	1339.6

All reported RiNNAL-S runs reach 10^{-6} within 3600 s.