A Low-rank ALM for Doubly Nonnegative Relaxations of Mixed-binary QP

Di Hou

Department of Mathematics, National University of Singapore

Joint work with Tianyun Tang (NUS), Kim-Chuan Toh (NUS)

NUS Graduate Symposium, May 13, 2025

Outline

Problem: DNN relaxations of Mixed-binary QP

Algorithm: RiNNAL

Numerical experiments

Mixed-binary nonconvex quadratic program

$$
\begin{array}{rlr}
\min\limits_{x \in \mathbb{R}^n} & x^\top Q x + 2c^\top x & (\textit{quadratic}) \\
\text{s.t.} & Ax = b & (\textit{linear}) \\
& x \geq 0 & (\textit{nonnegative}) \\
& x_k \in \{0,1\},\ k \in B & (\textit{binary}) \\
& x_i x_j = 0,\ (i,j) \in E & (\textit{complementary})
\end{array}
\qquad \text{(MBQP)}
$$

where $A \in \mathbb{R}^{m \times n}$ has full row rank, $b \geq 0$.

Applications:

- Mixed-integer programming (MIP)
- Quadratic programming (QP)
- Binary quadratic programming (BIQ/QUBO)
- (Extend to) quadratically constrained quadratic program (QCQP)

Doubly nonnegative relaxation

Difficulties: Nonconvex and NP-hard (Pardalos and Vavasis 1991)

E.g., BIQ problem with $n = 200$ cannot be globally solved within 1 hour.

Bottleneck: Solving DNN relaxations to provide lower bound

$$\begin{aligned} &\min \ \langle C, Y\rangle \\ &\text{s.t. } Ax = b,\ \mathrm{diag}(AXA^\top) = b^2,\ \mathrm{diag}_B(X) = x_B,\ X_E = 0, \\ &\qquad Y = \begin{pmatrix} 1 & x^\top \\ x & X \end{pmatrix} \in \mathbb{S}_+^{n+1} \cap \mathbb{N}^{n+1}, \end{aligned} \tag{1}$$

where $\mathbb{N}^{n+1}$ denotes nonnegative cone, $\mathbb{S}_+^{n+1}$ denotes positive semidefinite cone.

Issue: The Slater condition fails, so strong duality may not hold (the duality gap may not be 0), and many solvers fail in finding a solution.

$\longrightarrow$ alleviate by an equivalent reformulation

An equivalent reformulation

$$
\begin{aligned}
&\min \ \langle C, Y \rangle \\
&\text{s.t. } Ax = b,\ \mathrm{diag}(AXA^\top) = b^2 AX = bx^\top,\ \mathrm{diag}_B(X) = x_B,\ X_E = 0, \\
&\qquad Y = \begin{pmatrix} 1 & x^\top \\ x & X \end{pmatrix} \in \mathbb{S}^{n+1}_+ \cap \mathbb{N}^{n+1}.
\end{aligned}
\tag{DNN}
$$

Advantages:

1. Smallest duality gap (Bomze et al. 2017):

$$
\begin{aligned}
&\text{(primal optimal obj)} \quad (1) = (\text{DNN}) \\
&\text{(dual optimal obj)} \quad (1) \leq (\text{DNN})
\end{aligned}
$$

2. Keep the sparsity structure of the constraint matrices.

Difficulty: DNN problems are known to be challenging to solve.

Motivation

Can we design an efficient method for solving (DNN)?

Difficulties:

1. $\Omega(n^2)$ matrix variables. $\longrightarrow$ reduce to $\Omega(nr)$
2. $\Omega(mn)$ constraints $AX = bx^\top$. $\longrightarrow$ reduce to $\Omega(mr)$
3. $\Omega(n^2)$ constraints $X \geq 0$. $\longrightarrow$ projection-friendly

Contribution:

- Solve general DNN problem.
- Augmented Lagrangian method (ALM). $\longrightarrow$ RiNNAL
- Riemannian optimization with low-rank factorization.

Augmented Lagrangian method

$$\begin{aligned} \min\ & \langle C, Y\rangle \\ \text{s.t. } & Ax = b,\ AX = bx^\top,\ \mathrm{diag}_B(X) = x_B,\ X_E = 0, \\ & Y = \begin{pmatrix} 1 & x^\top \\ x & X \end{pmatrix} \in \mathbb{S}^{n+1}_+ \cap \mathbb{N}^{n+1}. \end{aligned} \qquad \text{(DNN)}$$

To apply ALM, consider the reformulation of (DNN):

$$\min \left\{ \langle C, Y\rangle + \delta_{\mathcal{M}}(Y) + \delta_{\mathcal{P}}(Z) :\ Y - Z = 0 \right\}, \qquad \text{(P)}$$

where we split the feasible set into a polyhedral cone $\mathcal{P}$ and a feasible set $\mathcal{M}$:

$$\mathcal{P} := \left\{ \begin{pmatrix} z & x^\top \\ x & X \end{pmatrix} \in \mathbb{S}^{n+1} \cap \mathbb{N}^{n+1} :\ X_{ij} = 0,\ \forall (i,j) \in E \right\},$$

$$\mathcal{M} := \left\{ \begin{pmatrix} z & x^\top \\ x & X \end{pmatrix} \in \mathbb{S}^{n+1}_+ :\ Ax = b,\ AX = bx^\top,\ x_B = \mathrm{diag}_B(X),\ z = 1 \right\}.$$

Given $\sigma > 0$, augmented Lagrangian function:

$$L_\sigma(Y, Z; W) := \langle C, Y\rangle + \delta_{\mathcal{M}}(Y) + \delta_{\mathcal{P}}(Z) - \langle W, Y - Z\rangle + \frac{\sigma}{2} \|Y - Z\|^2.$$

Augmented Lagrangian method

$$
\begin{aligned}
(Y^{k+1}, Z^{k+1}) &= \arg\min \{L_{\sigma_k}(Y, Z; W^k) : \ Y \in \mathcal{M}, \ Z \in \mathcal{P}\}, \\
W^{k+1} &= W^k - \sigma_k (Y^{k+1} - Z^{k+1}),
\end{aligned}
$$

where $\sigma_k \uparrow \sigma_\infty \leq +\infty$ are positive penalty parameters. Let $\widetilde{W} \in \mathbb{S}^{n+1}$ be fixed. The inner subproblem can be expressed as:

$$
\min \left\{ L_\sigma(Y, Z; \widetilde{W}) : \ Y \in \mathcal{M}, \ Z \in \mathcal{P} \right\}.
$$

We can first minimize w.r.t. Z to reduce to convex problem related only to Y:

$$
\min \left\{ \langle C, Y \rangle + \frac{\sigma}{2} \|\Pi_{\mathcal{P}^*}(\sigma^{-1}\widetilde{W} - Y)\|^2 : \ Y \in \mathcal{M} \right\}, \qquad \text{(ALM-sub)}
$$

where $\mathcal{P}^*$ denotes the dual cone of $\mathcal{P}$.

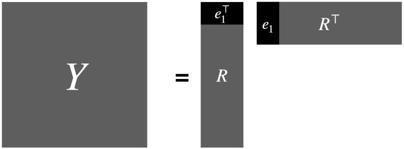
How to solve (ALM-sub) efficiently?

Low-rank reformulation for (ALM-sub)

Low-rank decomposition:

Suppose (ALM-sub) has a rank r solution Y^*, then

$$Y^* = \begin{pmatrix} e_1^\top \\ R \end{pmatrix} \begin{pmatrix} e_1^\top \\ R \end{pmatrix}^\top := \widehat{R}\widehat{R}^\top.$$



Thus, (ALM-sub) is equivalent to

$$\boxed{\min\left\{\langle C, Y\rangle + \frac{\sigma}{2}\|\Pi_{\mathcal{P}^*}(\sigma^{-1}\widetilde{W} - Y)\|^2 :\ Y \in \mathcal{M}\right\}. \min\left\{f_r(R) := \langle C, \widehat{R}\widehat{R}^\top\rangle + \frac{\sigma}{2}\|\Pi}$$

where

$$\mathcal{N}_r := \left\{R \in \mathbb{R}^{n\times r} :\ ARe_1 = b,\ ARR^\top = b(Re_1)^\top,\ \mathrm{diag}_B(RR^\top) = R_B e_1\right\}.$$

Riemannian optimization

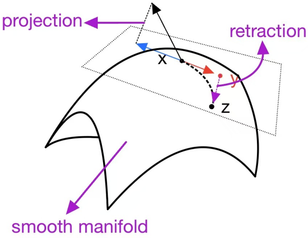
$$\min \{ f_r(R) : \; R \in \mathcal{M} \} . \qquad \text{(Rie-sub)}$$

Riemannian Gradient Descent (RGD):

$$f_r \left(\mathrm{Rtr}_R \left(-t \cdot \mathrm{grad}\, f_r(R) \right) \right) = f_r(R) - t \| \mathrm{grad}\, f_r(R) \|^2 + o(t)$$

Key points:

1. Smoothness
2. Projection $\longrightarrow \mathcal{M}$
3. Retraction
4. Rank
5. Global optimality



Smoothness: reformulation of $\mathcal{N}_r$

$$\mathcal{N}_r := \left\{ R \in \mathbb{R}^{n\times r} : \; ARe_1 = b, \; ARR^\top = b(Re_1)^\top, \; \mathrm{diag}_B(RR^\top) = R_B e_1 \right\}.$$

Issue 1: $\Omega(|B| + mn)$ constraint. $\longrightarrow$ reduce to $|B| + mr$ constraints:

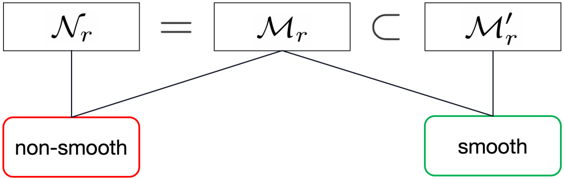
$$\mathcal{M}_r := \left\{ R \in \mathbb{R}^{n\times r} : \; AR = be_1^\top, \; \mathrm{diag}_B(RR^\top) = R_B e_1 \right\},$$

Issue 2: (Nonsmooth) LICQ may not hold. $\longrightarrow$ add slack variables to avoid

$$Ax = b \quad \Longleftrightarrow \quad Ax + s_1 = b, \; Ax - s_2 = b, \; s_1 \geq 0, \; s_2 \geq 0,$$

$$\mathcal{M}_r' := \left\{ R \in \mathbb{R}^{(n+2m)\times r} : \begin{pmatrix} A & I_m & 0 \\ A & 0 & I_m \end{pmatrix} R = \begin{pmatrix} b \\ b \end{pmatrix} e_1^\top, \; \mathrm{diag}_B(RR^\top) = R_B e_1 \right\}.$$

Smoothness: reformulation of $\mathcal{N}_r$



	$\mathcal{N}_r$	$\mathcal{M}_r$	$\mathcal{M}'_r$
constraint number	$mn + m + \lvert B \rvert$	$mr + \lvert B \rvert$	$2mr + \lvert B \rvert$
constraint type	quadratic+spherical	linear + spherical	linear + spherical
LICQ	No	Yes/No	Yes

Projection

Projection onto the tangent space $\mathcal{T}_R\mathcal{M}_r$:

$$\boxed{\operatorname{grad} f_r(R) = \operatorname{Proj}_{\mathcal{T}_R\mathcal{M}_r}(\nabla f_r(R)) = \nabla f_r(R) - h_R^*(\lambda, \mu),}$$

where $h_R : \mathbb{R}^{n\times r} \to \mathbb{R}^{m\times r} \times \mathbb{R}^{|B|}$ is the Fréchet differential mapping of the constraints defining $\mathcal{M}_r$ at R:

$$h_R(H) := (AH; 2\,\operatorname{diag}_B(HR^\top) - H_B e_1),$$

and $(\lambda, \mu) \in \mathbb{R}^{m\times r} \times \mathbb{R}^{|B|}$ is the solution to the following linear system:

$$h_R(h_R^*(\lambda, \mu)) = h_R(\nabla f_r(R)). \tag{Proj-sub}$$

Reduced complexisty: (Schur complement and SMW formula)

$$\mathcal{O}\left((|B| + mr)^3\right) \longrightarrow \mathcal{O}\left(\min\left\{|B|^3 + m^2 r + mr|B|, (mr)^2|B| + (mr)^3\right\}\right).$$

Retraction

Projection onto the manifold $\mathcal{M}_r$:

$$\mathrm{Rtr}_{\bar{R}}(H) := \mathrm{Proj}_{\mathcal{M}_r}(\bar{R}+H) = \arg\min \left\{ \|R - (\bar{R}+H)\|_F^2 : \; R \in \mathcal{M}_r \right\},$$

Nonconvex to convex: Define spherical manifold $\mathcal{B}_r$ as

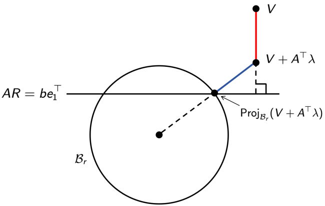
$$\mathcal{B}_r := \left\{ R \in \mathbb{R}^{n\times r} : \mathrm{diag}_B(RR^\top) - R_B e_1 = 0 \right\}$$

Lemma 1. For any $V \in \mathbb{R}^{n\times r}$, if there exists $\lambda \in \mathbb{R}^{m\times r}$ such that

$$A\,\mathrm{Proj}_{\mathcal{B}_r}(V + A^\top \lambda) = be_1^\top,$$

then

$$\mathrm{Proj}_{\mathcal{M}_r}(V) = \mathrm{Proj}_{\mathcal{B}_r}(V + A^\top \lambda).$$



Retraction subproblem

λ is solution to the following unconstrained nonsmooth convex problem:

$$\boxed{\min_{\lambda \in \mathbb{R}^{m \times r}} \sum_{i \in B} \|(V' + A^\top \lambda)_i\| + \sum_{i \in [n] \setminus B} \|(V' + A^\top \lambda)_i\|^2 + \langle b', A^\top \lambda \rangle,}$$

where $V' = V - ee_1^\top/2$, $b' \in \mathbb{R}^{m \times r}$ such that $(Ae - 2b)e_1^\top = Ab'$, and R_i denotes the i-th row of R.

Generalized geometric median problem:

Generalized Weiszfeld algorithm (Weiszfeld 1937) + Newton method

Rank-adaptive strategies

Rank increase: escape from saddle point

For some $\tau \in \mathbb{N}^+$ and KKT solution R of (Rie-sub), we can compute a descent direction U based on S at the point $P := [R, 0_{n\times\tau}]$ such that for some $\beta < 0$,

$$f_{r+\tau}(\mathrm{Rtr}_P(tU)) = f_r(R) + \beta t^2 + o(t^2).$$

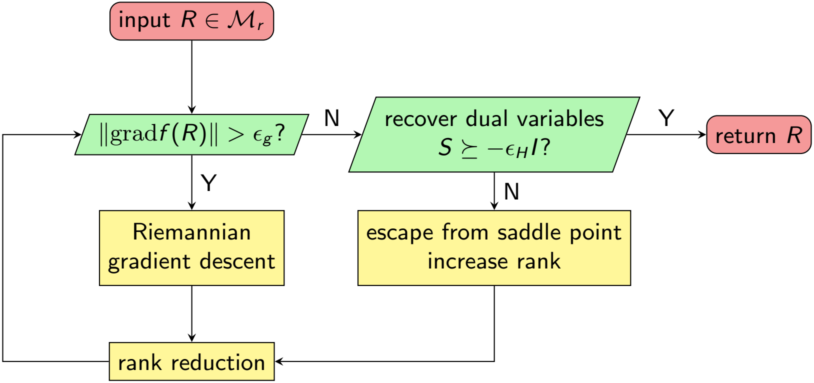
Rank reduction: reduce the computational cost (Gao and Absil 2021)

Let $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ be the singular values of R. Let

$$i := \arg\max \left\{\sigma_i/\sigma_{i+1} : \; i \in [r-1]\right\}.$$

If $\sigma_i/\sigma_{i+1} > \kappa_1$, then we truncate all singular values $\sigma_{i+1}, \sigma_{i+2}, \ldots, \sigma_r$ to 0, and reduce r to be i.

RiNNAL subproblem framework



Global convergence: (Rockafellar 1976, Cui et al. 2019).

Numerical Experiments

Machine: All the experiments are run using Matlab R2023b on a workstation with Intel Xeon E5-2680 v3 @ 2.50GHz processor and 96GB RAM.

Tolerance for KKT residue: 10^{-6}. Maxtime: 3600s.

Algorithms comparison:

Besides penalizing $Y \geq 0$ and $X_E = 0$, we may penalize more terms:

additional penalty term	manifold	algorithm	issue
	$\mathcal{N}_r$	-	nonsmooth
	$\mathcal{M}_r$	RiNNAL	-
$Ax = b,\ AX = bx^\top$	$\mathcal{B}_r$	RiNNAL-Diag	-
$Ax = b,\ AX = bx^\top$, $\mathrm{diag}_B(X) = x_B$	$\mathbb{R}^{n\times r}$	SDPLR	slow
$AR = be_1^\top$, $\mathrm{diag}_B(X) = x_B$	$\mathbb{R}^{n\times r}$	-	fixed rank

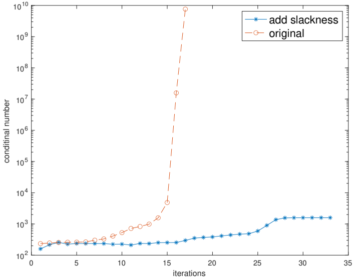
Solvers comparison

algorithm	low rank factorization	handle $Y \geq 0$
ManiSDP (Wang and Hu)	Yes	No
HALLaR (Monteiro et al.)	Yes	No
LoRADS (Han et al.)	Yes	No
SDPF (Tang and Toh)	Yes	No
SketchCGAL (Yurtsever et al.)	No	Yes
SDPNAL+ (Sun, Toh et al.)	No	Yes
SDPLR (Burer and Monteiro)	Yes	Yes
SDPDAL (Wang et al.)	Yes	Yes
RiNNAL-Diag (variant)	Yes	Yes
RiNNAL (ours)	Yes	Yes

- Compare with SDPNAL+ and RiNNAL-Diag.
- Not using SDPDAL: code unavailable and similar to RiNNAL-Diag.
- Not using SDPLR/SketchCGAL: slower than RiNNAL-Diag.

Quadratic assignment problems

$$\min \left\{ \langle Y, WYD \rangle : \; Y \in \{0,1\}^{p\times p}, Ye = e, Y^\top e = e \right\}, \qquad \text{(QAP)}$$



problem	algorithm	time	R_{max}
original	RiNNAL	-	-
	SDPNAL+	20.1	2.1e-7
add slack	RiNNAL	16.7	1.7e-7
	SDPNAL+	66.9	3.2e-7

Avoiding non-smoothness:

- The condition number of $h_R h_R^*$ (projection subproblem) is much better.
- Avoid the failure of projection and retraction.

Binary integer nonconvex quadratic programming

$$\min \left\{ x^\top Q x + 2c^\top x : \; x \in \{0,1\}^n \right\}. \qquad \text{(BIQ)}$$

problem	algorithm	it	itsub	r/itA	R$_\mathrm{p}$	R$_\mathrm{d}$	R$_\mathrm{c}$	obj	time
$n = 1000$	RiNNAL	11	911	55	9.04e-07	3.17e-07	3.80e-07	-3.9849472e+05	9.40e+00
	SDPNAL+	118	172	2752	8.92e-07	9.69e-07	5.54e-07	-3.9849494e+05	2.61e+02
$n = 2500$	RiNNAL	9	817	97	6.11e-07	4.53e-07	5.57e-08	-1.6354913e+06	1.03e+02
	SDPNAL+	-	-	-	-	-	-	-	-
$n = 5000$	RiNNAL	6	1091	147	5.52e-07	9.46e-08	1.51e-07	-4.7435656e+06	5.61e+02
	SDPNAL+	-	-	-	-	-	-	-	-
$n = 10000$	RiNNAL	5	1078	216	4.80e-07	1.80e-07	9.93e-07	-1.3832829e+07	2.32e+03
	SDPNAL+	-	-	-	-	-	-	-	-

Maximum stable set problems

$$\max\left\{x^\top x:\ x_i x_j = 0,\ \forall (i,j)\in E,\ x\in\{0,1\}^n\right\}. \qquad (\theta_+)$$

problem (n, m)	algorithm	it	itsub	r/itA	R_p	R_d	R_c	obj	time
G43	RiNNAL	10	795	80	7.48e-07	4.18e-08	2.09e-08	-2.7973625e+02	9.61e+00
(1000,9990)	SDPNAL+	48	61	1250	4.58e-07	6.61e-07	9.64e-14	-2.7973595e+02	1.43e+02
G34	RiNNAL	10	2188	11	3.79e-07	1.48e-08	5.58e-07	-9.9999198e+02	9.91e+01
(2000,4000)	SDPNAL+	-	-	-	-	-	-	-	-
G48	RiNNAL	8	1497	21	9.25e-07	4.67e-08	3.09e-07	-1.4999238e+03	2.44e+02
(3000,6000)	SDPNAL+	-	-	-	-	-	-	-	-
G55	RiNNAL	20	2130	353	9.96e-07	1.85e-07	4.84e-08	-2.3230485e+03	1.45e+03
(5000,12498)	SDPNAL+	-	-	-	-	-	-	-	-

Quadratic knapsack problems

$$\max \left\{ x^\top Q x : \ a^\top x \le \tau, \ x \in \{0,1\}^n \right\}. \qquad \text{(QKP)}$$

problem	algorithm	it	itsub	r/itA	R_p	R_d	R_c	obj	time
$n = 500$	RiNNAL	4	184	21	6.15e-07	3.34e-07	2.57e-08	-1.0261057e+07	1.47e+00
	SDPNAL+	56	102	3375	6.79e-07	9.82e-07	5.01e-07	-1.0261059e+07	1.19e+02
	RiNNAL-Diag	50	163983	19	4.83e-07	4.16e-07	2.50e-15	-1.0261052e+07	4.49e+02
$n = 1000$	RiNNAL	4	198	24	5.23e-07	2.17e-09	2.02e-09	-4.0961669e+07	6.79e+00
	SDPNAL+	113	310	7179	8.21e-07	9.83e-07	2.91e-07	-4.0934695e+07	1.13e+03
	RiNNAL-Diag	-	-	-	-	-	-	-	-
$n = 5000$	RiNNAL	1	177	143	7.97e-07	4.54e-08	1.96e-09	-1.0209621e+09	1.21e+02
	SDPNAL+	-	-	-	-	-	-	-	-
	RiNNAL-Diag	-	-	-	-	-	-	-	-
$n = 10000$	RiNNAL	1	226	202	3.36e-07	3.41e-08	1.73e-10	-4.0846437e+09	6.33e+02
	SDPNAL+	-	-	-	-	-	-	-	-
	RiNNAL-Diag	-	-	-	-	-	-	-	-

Disjunctive quadratic knapsack problems

$$\max\left\{x^\top Qx:\ a^\top x \le \tau,\ x_i x_j = 0,\ (i,j)\in E,\ x\in\{0,1\}^n\right\}. \qquad \text{(DQKP)}$$

problem	algorithm	it	itsub	r/itA	R_p	R_d	R_c	obj	time
$n=1000$	RiNNAL	20	524	50	6.36e-07	1.39e-08	1.34e-07	-2.5519045e+06	1.54e+01
$d=2$	SDPNAL+	99	715	5949	2.98e-07	9.97e-07	1.27e-06	-2.5519060e+06	2.69e+03
$n=1000$	RiNNAL	35	8270	116	6.80e-07	9.44e-07	1.53e-10	-1.2078151e+06	2.48e+02
$d=5$	SDPNAL+	-	-	-	-	-	-	-	-
$n=2000$	RiNNAL	15	1164	77	7.06e-07	3.15e-07	4.33e-08	-9.5394855e+06	1.00e+02
$d=2$	SDPNAL+	-	-	-	-	-	-	-	-
$n=2000$	RiNNAL	27	5059	148	7.04e-07	9.38e-07	2.59e-07	-4.8842259e+06	4.61e+02
$d=5$	SDPNAL+	-	-	-	-	-	-	-	-
$n=5000$	RiNNAL	10	642	148	9.41e-07	4.52e-07	3.89e-09	-5.6073030e+07	4.86e+02
$d=2$	SDPNAL+	-	-	-	-	-	-	-	-
$n=5000$	RiNNAL	14	2727	178	7.52e-07	2.76e-08	3.19e-09	-2.7536080e+07	1.81e+03
$d=5$	SDPNAL+	-	-	-	-	-	-	-	-

Gromov-Wasserstein distance

$$\min\left\{-\langle D_X \Pi D_Y, \Pi\rangle :\ \Pi e^k = a,\ \Pi^\top e^l = b,\ \Pi \in \mathbb{R}_+^{l\times k}\right\}. \qquad \text{(GWD)}$$

problem	algorithm	it	itsub	r/itA	R_p	R_d	R_c	obj	time
Cat	RiNNAL	42	10862	42	4.66e-07	3.09e-07	6.89e-07	1.3986652e+05	2.26e+02
$n = 1225$	SDPNAL+	572	1002	10988	5.30e-07	1.85e-07	8.72e-15	1.3990409e+05	3.54e+03
David	RiNNAL	76	12694	14	2.72e-07	9.36e-07	1.77e-07	2.8688769e+05	2.32e+02
$n = 1225$	SDPNAL+	-	-	-	-	-	-	-	-
Cat	RiNNAL	108	21820	15	6.26e-07	1.39e-07	9.49e-07	3.4775634e+05	1.33e+03
$n = 2025$	SDPNAL+	-	-	-	-	-	-	-	-
David	RiNNAL	64	14851	37	5.97e-07	9.46e-07	7.85e-07	2.4304548e+05	9.35e+02
$n = 2025$	SDPNAL+	-	-	-	-	-	-	-	-

Conclusion

Summary:

- RiNNAL: a low-rank ALM for solving DNN relaxations.
- Efficient for problems with many constraints and low-rank solutions.
- Following work: extend to general SDP-RLT relaxations.

Thank you for your attention!

References:

D. Hou, T. Tang, and K.-C. Toh, "A low-rank augmented Lagrangian method for doubly nonnegative relaxations of mixed-binary quadratic programs," under review, *arXiv preprint* arXiv:2502.13849, 2025.

D. Hou, T. Tang, and K.-C. Toh, "RiNNAL+: a Riemannian ALM solver for SDP-RLT relaxations of mixed-binary quadratic programs," under review, 2025.