NeurIPS 2026

Asymmetric Factorization for Low-Rank PSD Learning

When Is the Relaxation Exact?

Enliang Hu[1] Juho Kannala[2] Guoying Zhao[2]

Quanming Yao[3] Kun Yue[4]

[1]Yunnan Normal University [2]Aalto University

[3]Tsinghua University [4]Yunnan University

Compact factors, coupled optimization

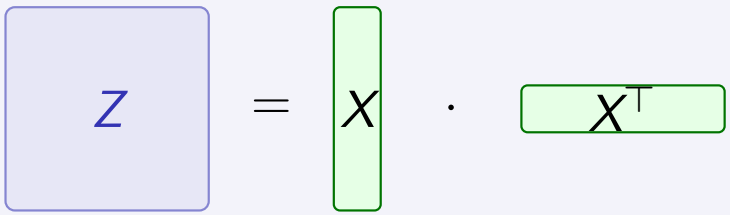
PSD-constrained optimization

$$\min_{Z \succeq 0} f(Z), \qquad Z \in \mathbb{R}^{n \times n}.$$

Symmetric factorization (sBMF)

$$X \in \mathbb{R}^{n \times r}, \qquad r \ll n,$$

$$\min_{X} \; f(XX^\top) + h(X).$$



Storage: $O(n^2)$ to $O(nr)$.

The repeated factor X couples the variables: the factor problem is generally nonconvex even when f is convex.

Existing approach: split the factors

Asymmetric factorization (aBMF)

$$\mathcal{F}(X, Y; \gamma) = f(XY^\top) + \tfrac{1}{2}h(X) + \tfrac{1}{2}h(Y) + \tfrac{\gamma}{2}\|X - Y\|_F^2, \qquad \gamma > 0.$$

For convex f, h, each factor subproblem is strongly convex.

Small γ: the product may not be PSD.

Large γ: slower updates.

The exactness question

When does a finite penalty force factor agreement at critical points?

What do existing guarantees cover?

Work	Setting	Guarantee
Zhu et al. (2018) Li et al. (2023)	Quadratic sNMF; nonnegative factors	Fixed-penalty exactness bounds
Hu & Kwok (2019)	Smooth convex PSD loss; $h = 0$	Update-dependent factor-agreement condition
Han et al. (2024)	Smooth SDP; linear equalities	Approximate KKT transfer

The gap addressed here

Solver-independent exactness for general transpose-symmetric convex losses with shared convex regularizers or constraints.

Three contributions

1. When are the factors equal?

A spectral sufficient condition at joint critical points.

2. How can we choose the penalty?

Sublevel-set analysis gives sufficient bounds from initialization.

3. Does every positive penalty work?

A counterexample answers the sNMF positive-penalty question negatively. Structured cases still admit guarantees.

A spectral condition for exactness

Convex, finite-valued, transpose-symmetric f.
h is proper, closed and σ_h-strongly convex ($\sigma_h \geq 0$).

At a joint critical point $(\bar{X}, \bar{Y})$, choose $\bar{G} \in \partial_Z f(\bar{X}\bar{Y}^\top)$ shared by both stationarity equations. Set $\bar{S} = \frac{1}{2}(\bar{G} + \bar{G}^\top)$.

Theorem 1: a pointwise sufficient bound

$$\boxed{\gamma > \tfrac{1}{2}\lambda_{\max}(\bar{S}) - \tfrac{1}{4}\sigma_h}$$

implies $\bar{X} = \bar{Y}$, hence an sBMF critical point.

Critical-point correspondence

An sBMF critical point $\bar{X}$ lifts to $(\bar{X}, \bar{X})$ in aBMF. The critical sets correspond if the bound holds at every aBMF critical point.

Corollary 1: a sublevel-set bound

Under Theorem 1, let f be coercive and l_f-smooth.

Loss sublevel set determined by $Z_0 = Y_0 Y_0^\top$

$$\mathcal{L}_0 = \{Z : f(Z) \le f(Z_0)\}, \qquad d_0 = \sup_{Z_1, Z_2 \in \mathcal{L}_0} \|Z_1 - Z_2\|_F < \infty.$$

For every $Z \in \mathcal{L}_0$,

$$\|\partial_Z f(Z)\|_F \le \|\partial_Z f(Z_0)\|_F + l_f d_0.$$

One penalty bound for the whole loss sublevel set

Any joint critical point $(\bar{X}, \bar{Y})$ with $\bar{X}\bar{Y}^\top \in \mathcal{L}_0$ satisfies $\bar{X} = \bar{Y}$ if

$$\boxed{\gamma > \tfrac{1}{2}\left(\|\partial_Z f(Z_0)\|_F + l_f d_0\right) - \tfrac{\sigma_h}{4}.}$$

Corollary 1: an initialization bound

Under Theorem 1, let f be coercive, l_f-smooth and σ_f-strongly convex. Set $Z_0 = Y_0 Y_0^\top$ for $Y_0 \in \operatorname{dom} h$.

Full-objective bound: strongly convex, smooth loss

Let $h \geq \underline{h} > -\infty$ and let $\dot{Z}$ minimize f.
At a joint critical point $(\bar{X}, \bar{Y})$ with $\mathcal{F}(\bar{X}, \bar{Y}; \gamma) \leq \mathcal{F}(Y_0, Y_0; \gamma)$,

$$\gamma > \frac{l_f}{\sqrt{2\sigma_f}} \sqrt{[f(Z_0) + h(Y_0) - \underline{h} - f(\dot{Z})]_+} - \frac{\sigma_h}{4}$$

implies $\bar{X} = \bar{Y}$ under the theorem's symmetry assumptions.

sNMF: a simple sufficient choice

For $f(Z) = \frac{1}{2}\|Z - A\|_F^2$, $h = \delta_+$ and $Y_0 \geq 0$,

$$\boxed{\gamma > \tfrac{1}{2}\left\|Y_0 Y_0^\top - A\right\|_F.}$$

A positive penalty alone is insufficient

The sNMF positive-penalty conjecture fails in general.
Zhu et al. (2018); Li et al. (2023).

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \gamma = \frac{3}{8},$$

$$\bar{X} = \frac{1}{4}\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}, \qquad \bar{Y} = \frac{1}{4}\begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}.$$

An unequal-factor critical point

Both factors are strictly positive. Exact AM starting from $X_0 = Y_0 = \bar{Y}$ reaches $(\bar{X}, \bar{Y})$ in one sweep and stays there.

Factor agreement also fails for HAM and SymHALS.

Corollary 2: structured sNMF cases

For $f(Z) = \frac{1}{2}\|Z - A\|_F^2$, $h = \delta_+$, $A = A^\top$, $\gamma > 0$, any joint critical point $(\bar{X}, \bar{Y})$ satisfies

$$\left\|\bar{X}\right\|_F = \left\|\bar{Y}\right\|_F.$$

Two additional structures ensure $\bar{X} = \bar{Y}$:

Rank one and nonnegative data

If $r = 1$ and $A \geq 0$ entrywise, **every** $\gamma > 0$ **suffices**.

Orthogonal columns

If $\bar{X}$ and $\bar{Y}$ each have pairwise orthogonal columns and

$$\gamma \geq -\lambda_{\min}(A),$$

then $\bar{X} = \bar{Y}$.

Practical adaptive penalty rule

Why adapt? Small penalties may leave unequal factors. Large penalties may slow updates. Reuse the factors as γ changes.
For $D_k = X_k - Y_k \neq 0$, use $G_k \in \partial_Z f(X_k Y_k^\top)$.

Penalty update motivated by Theorem 1

$$\boxed{\gamma_k = \max\{\min\{\gamma_{k-1}, \widehat{\gamma}_k\}, \check{\gamma}_k\}.}$$

$$\check{\gamma}_k = \left[\frac{\operatorname{tr}(D_k^\top G_k D_k)}{2\nu \|D_k\|_F^2} - \frac{\sigma_h}{4}\right]_+ + \varepsilon_0, \qquad \widehat{\gamma}_k = \left[\tfrac{1}{2}\|G_k\| - \tfrac{\sigma_h}{4}\right]_+.$$

Use $0 < \nu < 1$, $\varepsilon_0 > 0$, $[a]_+ = \max(a, 0)$, and a norm with $\|G_k\| \geq \lambda_{\max}\left(\frac{1}{2}(G_k + G_k^\top)\right)$.

Conditional nonincrease ($D_k \neq 0$): $\gamma_k \leq \gamma_{k-1} \iff \check{\gamma}_k \leq \gamma_{k-1}$.
For $D_k = 0$, keep γ_{k-1}.

Algorithm: alternating convex optimization

Initialize $X_0, Y_0 \in \operatorname{dom} h$ and $\gamma_0 > 0$.

For $k = 1, \ldots, K$

1. $X_k \approx \arg\min_X \mathcal{F}(X, Y_{k-1}; \gamma_{k-1})$
2. $Y_k \approx \arg\min_Y \mathcal{F}(X_k, Y; \gamma_{k-1})$
3. Keep γ fixed or apply the dynamic update.
4. Stop when the prescribed tolerance is reached.

$$\text{Return} \quad U = \tfrac{1}{2}(X_k + Y_k), \qquad Z_{\mathrm{PSD}} = UU^\top.$$

AM: exact blocks; HAM: column updates; AAPG: accelerated proximal-gradient steps.

Theorem 2: convergence with fixed γ

Assume f is convex, continuously differentiable and coercive; h is proper, closed and convex; $X_0 = Y_0 \in \operatorname{dom} h$.

Any fixed $\gamma > 0$, with exact AM

The iterates are bounded, and

$$\|X_k - X_{k-1}\|_F + \|Y_k - Y_{k-1}\|_F \longrightarrow 0.$$

Every accumulation point $(\bar{X}, \bar{Y})$ is critical for aBMF.

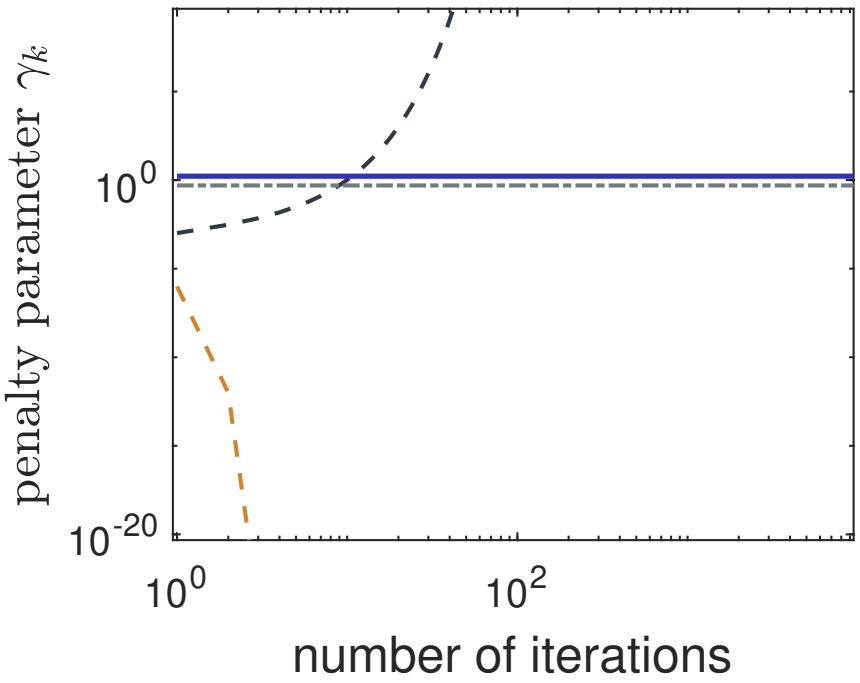
A sufficient fixed penalty also gives exactness

If f is also transpose symmetric, l_f-smooth and σ_f-strongly convex, h is bounded below, and Corollary 1's penalty bound holds,
every accumulation point satisfies $\bar{X} = \bar{Y}$ and yields an sBMF critical point.

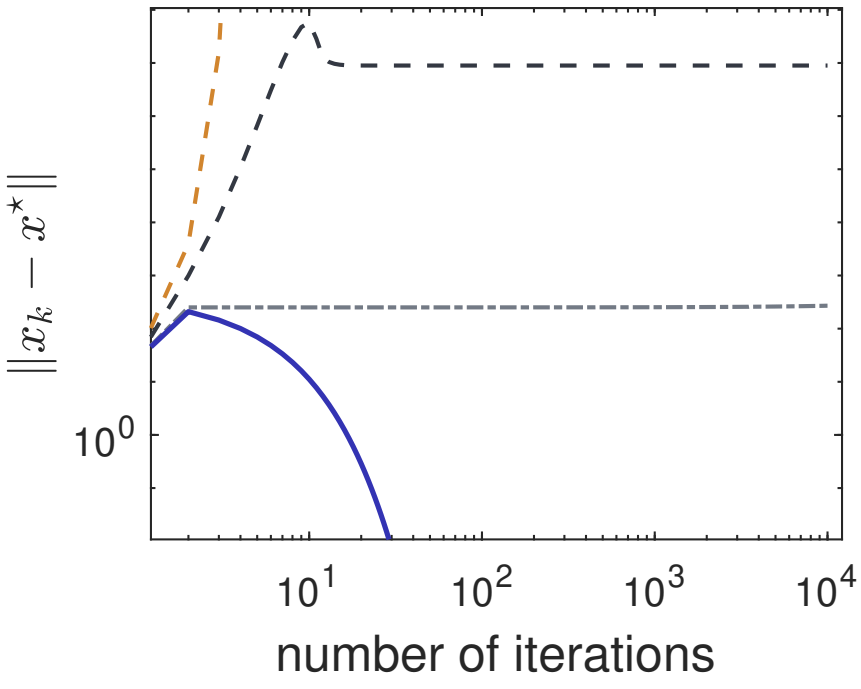
Scalar example 1: penalty and stability

$$\min_x x^2 \quad \rightsquigarrow \quad \mathcal{F}(x, y; \gamma) = xy + \tfrac{\gamma}{2}(x - y)^2.$$

Updated penalty



Solution error



AM_{agd} AM_{adp} AM_{pca} AM_{our}

AM_{our} approaches zero with $\gamma_k \approx 1.6677$.

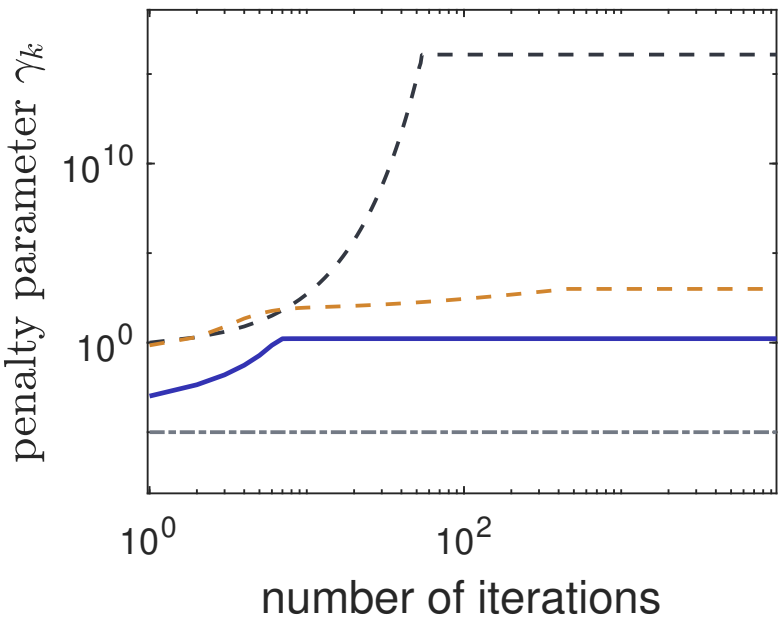
Fixed-γ AM converges to zero iff $\gamma > 1/2$ for $y_0 \neq 0$.

Scalar example 2: agreement and accuracy

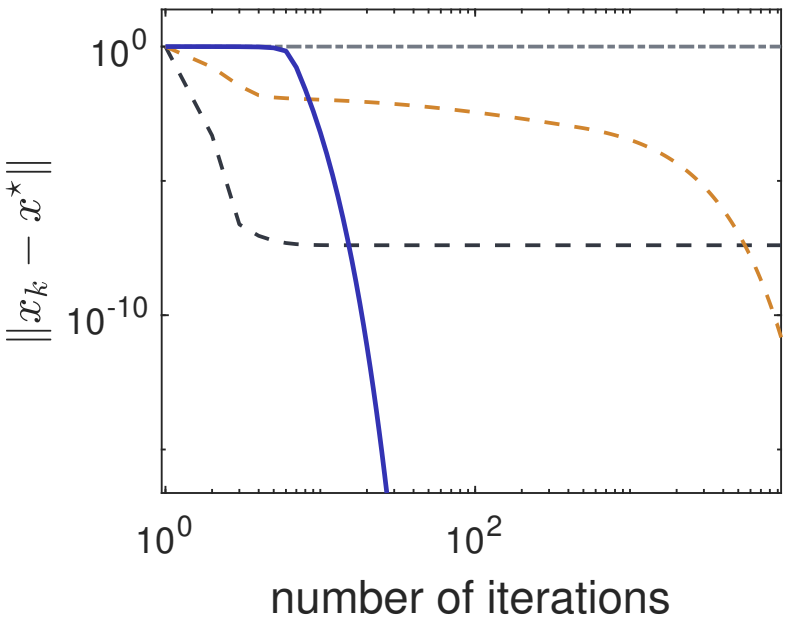
$$\mathcal{F}(x, y; \gamma) = \tfrac{1}{2}(xy+1)^2 + \tfrac{\gamma}{2}(x-y)^2, \quad x^\star = 0.$$

$$x_0 = 1, \; y_0 = -1, \; \gamma_0 = 10^{-5}.$$

Updated penalty



Solution error



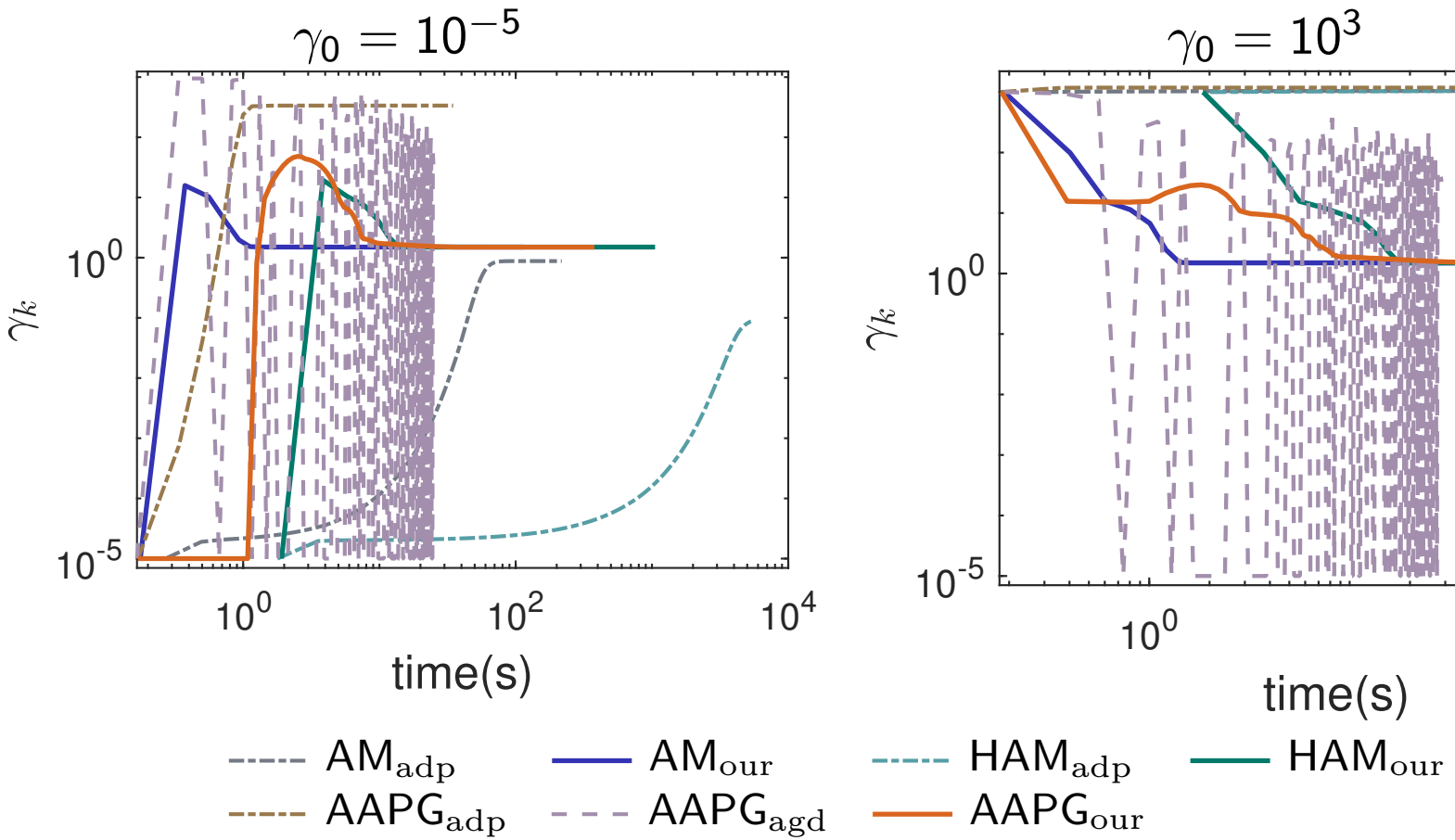
AM$_{\text{agd}}$ AM$_{\text{adp}}$ AM$_{\text{pca}}$ AM$_{\text{our}}$

AM$_{\text{our}}$: solution error $< 10^{-12}$ by sweep 22.

Nearly equal factors can still leave a larger solution error.

Matrix sensing: penalty trajectories

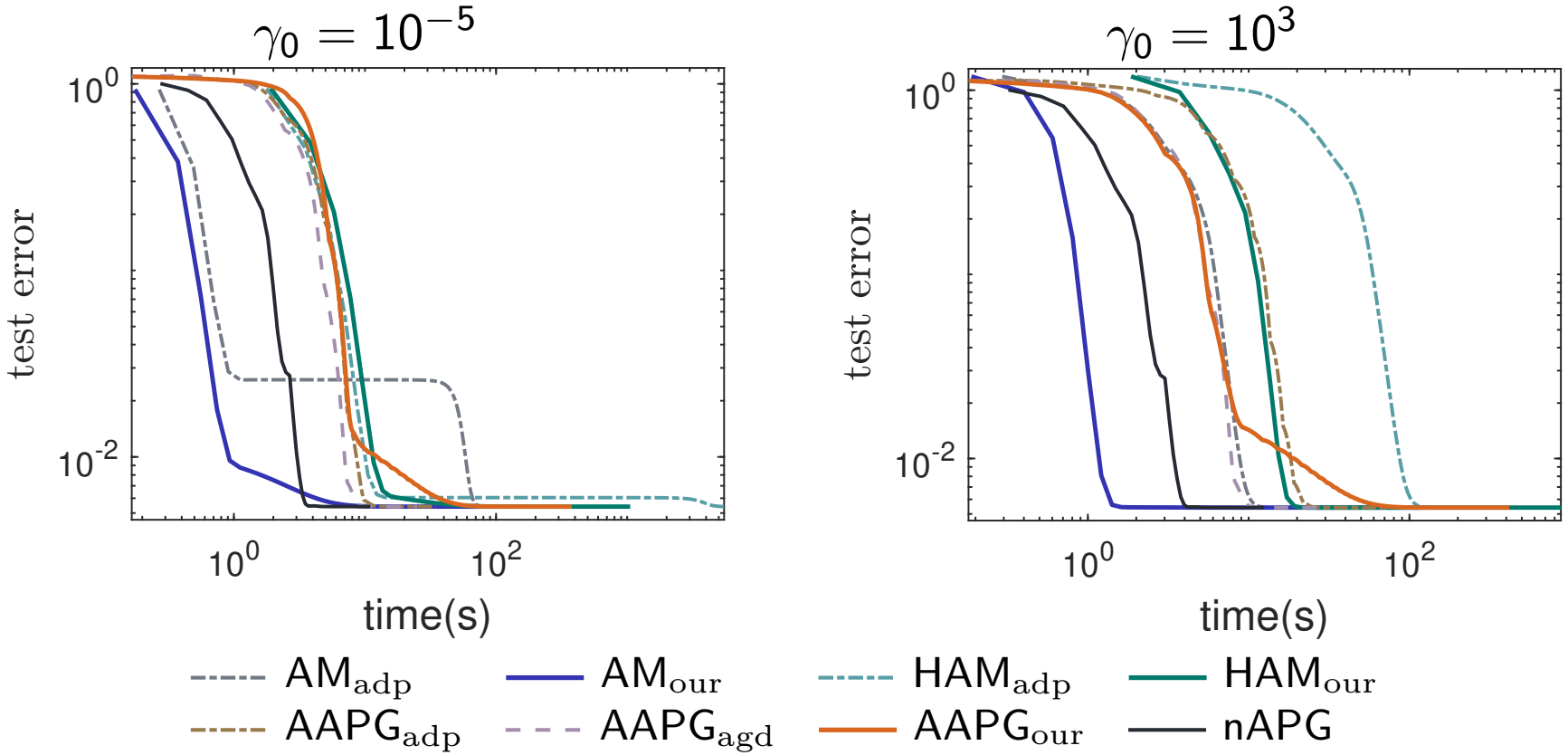
The initial penalties differ by eight orders of magnitude.



The proposed variants settle at penalties of order one.

Matrix sensing: PSD recovery

$$f(Z) = \mathrm{tr}(Z) + \tfrac{\eta}{2}\sum\nolimits_i \left(\mathrm{tr}(C_i Z) - b_i\right)^2, \quad h = 0.$$



AM/HAM: faster than the adaptive counterparts, with similar final test errors.

Top-k PSD approximation

$(n, r) = (50{,}000, 1000)$; means over five paired trials.

Method	Time (s)	Rel. objective	Subspace error
sBMF-GD	2332.627	1.99×10^{-5}	5.57×10^{-2}
sBMF-GN	225.145	$\mathbf{2.85 \times 10^{-7}}$	$\mathbf{6.68 \times 10^{-3}}$
$\mathrm{AM}_{\mathrm{adp}}$	2586.158	1.48×10^{-5}	4.82×10^{-2}
$\mathrm{AM}_{\mathrm{our}}$	738.091	2.79×10^{-6}	2.10×10^{-2}

Within AM: better accuracy in less time

3.50× faster than $\mathrm{AM}_{\mathrm{adp}}$ here; 2.29× in the smaller case. sBMF-GN remains best overall.

Low-rank recovery: MovieLens

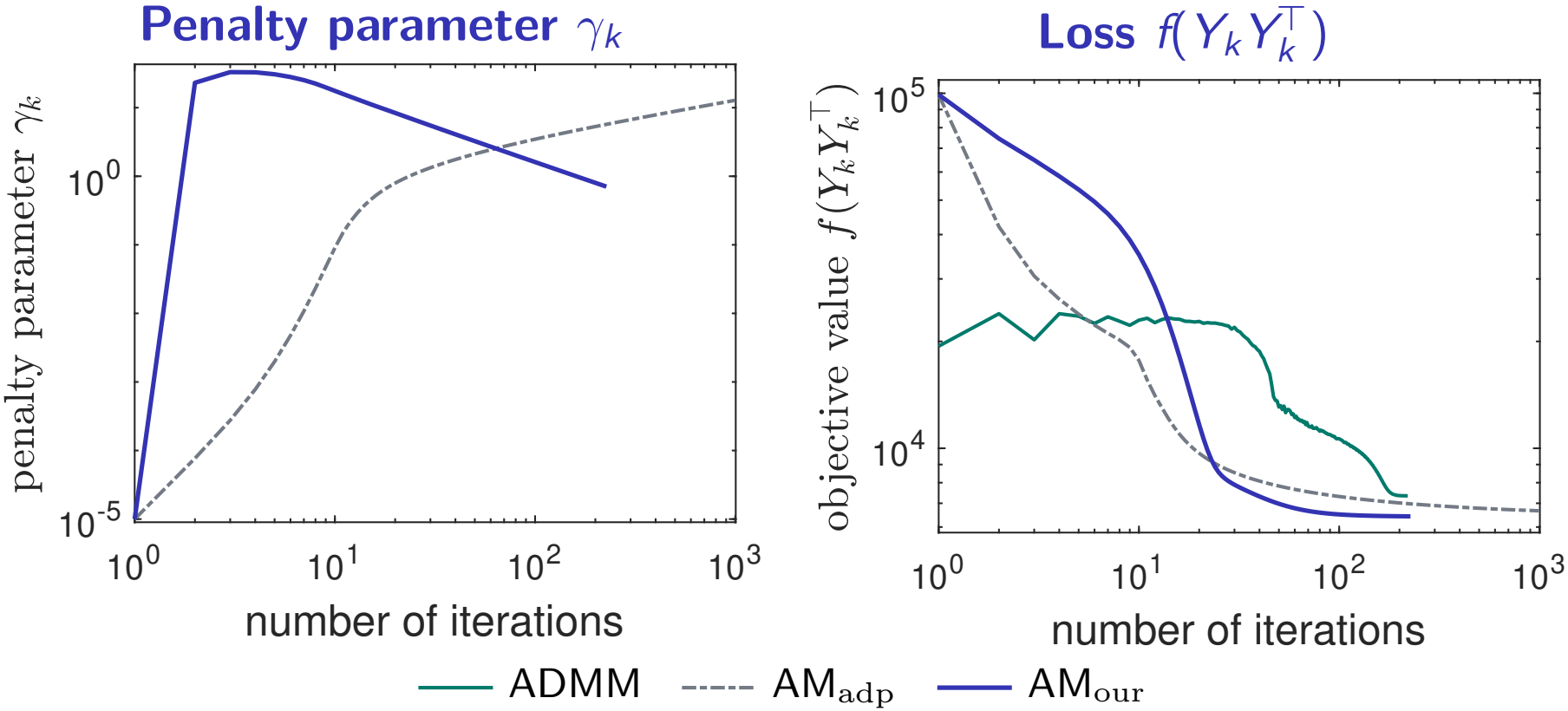
Matrix completion for recommendation; $r = 2$. AM comparisons.
Normalized-rating RMSE; mean $\pm$ SD over 50/50/10 runs.

Data	Method	RMSE	Time (s)
100K	AM_{adp}	0.20 ± 0.01	12.56 ± 8.86
	AM_{our}	0.19 ± 0.00	7.88 ± 1.44
1M	AM_{adp}	0.18 ± 0.01	174.81 ± 136.86
	AM_{our}	0.18 ± 0.00	75.57 ± 7.53
10M	AM_{adp}	0.17 ± 0.01	10962.04 ± 4695.71
	AM_{our}	0.17 ± 0.00	6723.75 ± 292.08

AM speedups: $1.59\times$, $2.31\times$, $1.63\times$.

Link prediction: penalty and loss

Adult a5a; both horizontal axes show iteration counts.

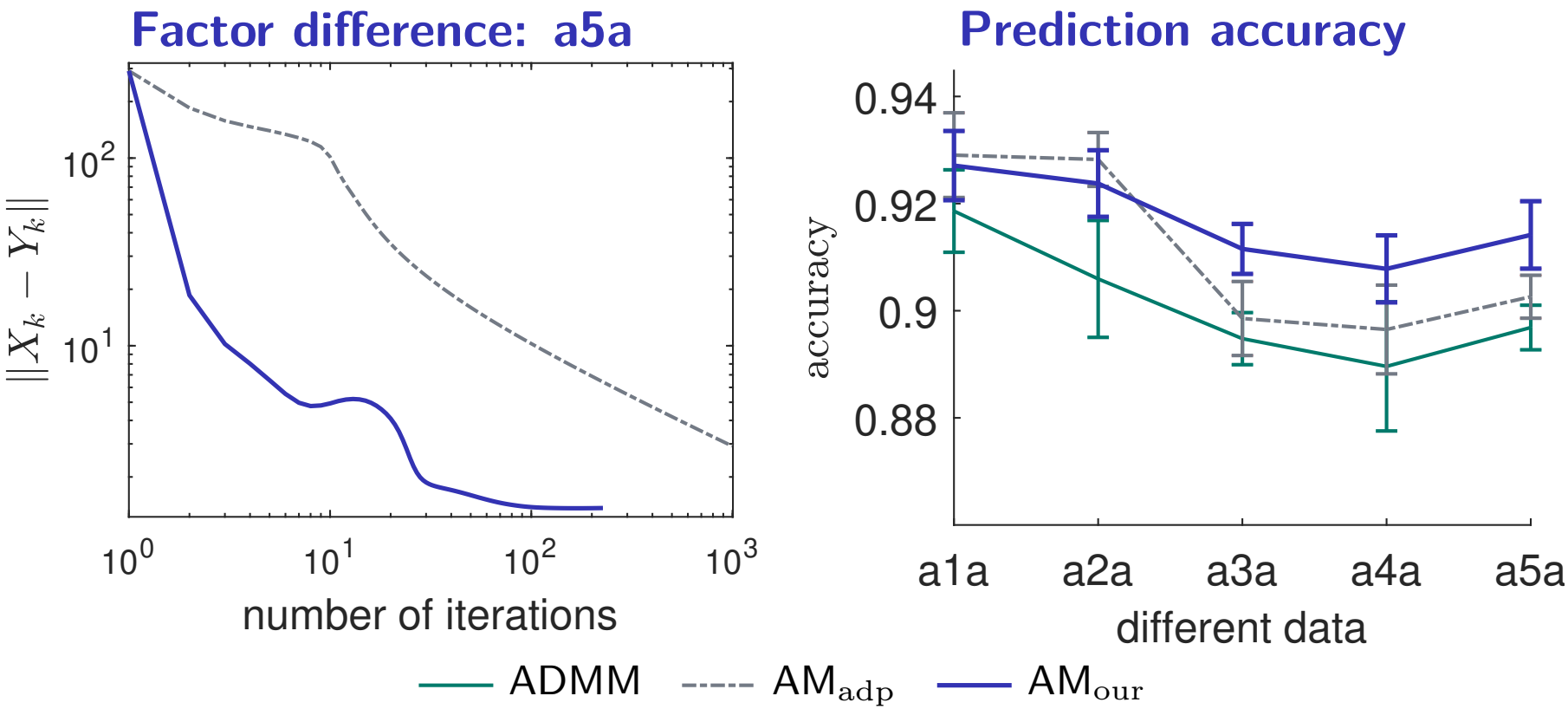


$\mathrm{AM}_{\mathrm{our}}$ raises γ early, then reduces it;

$\mathrm{AM}_{\mathrm{adp}}$ keeps increasing γ.

Link prediction with a PSD affinity matrix

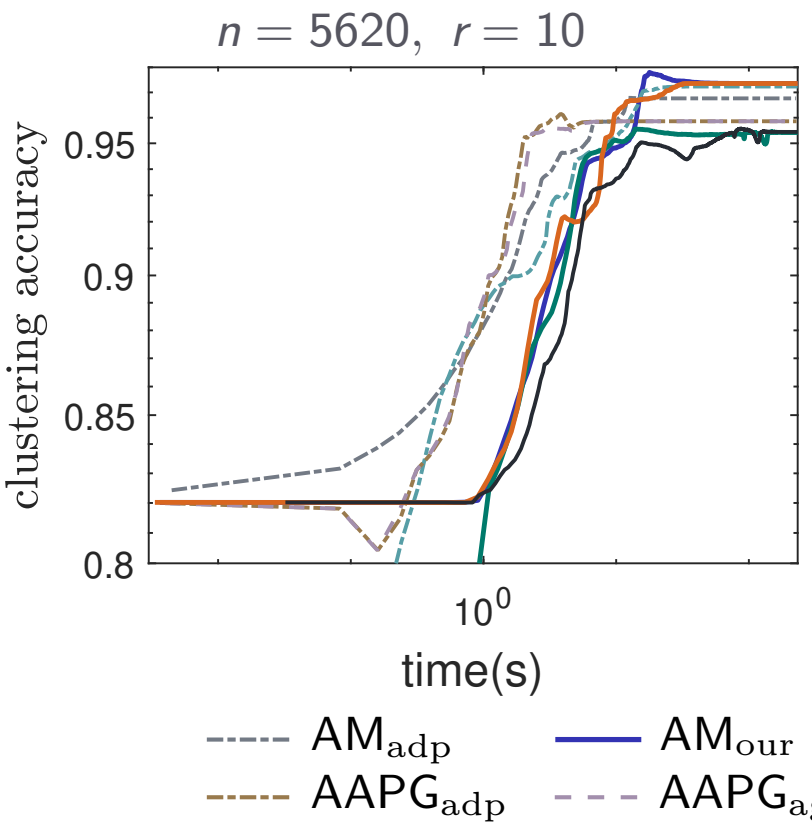
Adult a1a–a5a; mean prediction accuracy over 10 runs.



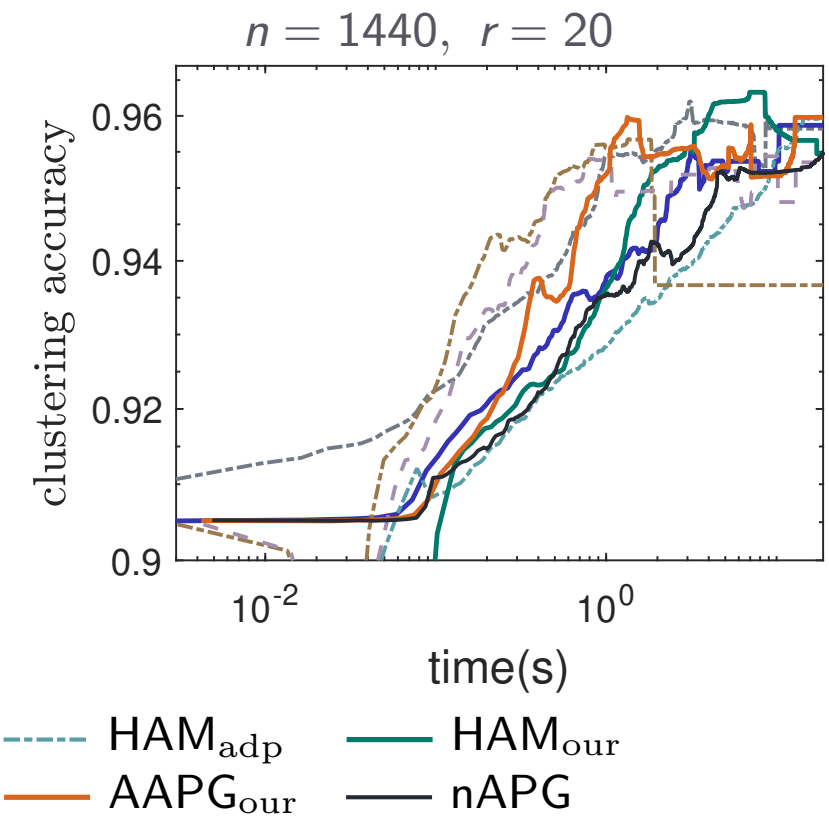
$\mathrm{AM}_{\mathrm{our}}$ vs. $\mathrm{AM}_{\mathrm{adp}}$: higher mean accuracy on a3a–a5a; slightly lower on a1a and a2a.

Clustering with sNMF

OPTDIGITS



COIL-20



AM_{adp} AM_{our} HAM_{adp} HAM_{our}
AAPG_{adp} AAPG_{agd} AAPG_{our} nAPG

AAPG_{our} reaches high accuracy earlier than nAPG and stays higher at late times on both datasets.

Exploration: a neural-network surrogate

Digits 3 vs. 8; lift dimension 29. Means over 120 matched runs.

Method	Objective	PSD violation	Residual pass
sBMF	4.100	2.24×10^{-15}	66.7%
aBMF-fixed	-1.225	2.07×10^{-5}	100%
$\mathrm{AM}_{\mathrm{adp}}$	-1.225	2.24×10^{-5}	100%
$\mathrm{AM}_{\mathrm{our}}$	-1.228	2.21×10^{-5}	100%

Objective excludes the factor penalty. Pass: finite loss and residuals ≤ 0.05.
Disagreement is omitted for sBMF. PSD violation is assessed separately.

Nonconvex surrogate

Outside Theorem 1. Passing does not certify robustness.

Future research directions

1. **More general models**

   Study broader classes of f and h, including nonconvex cases.

2. **Penalty selection**

   Improve parameter rules and analyze convergence with adaptive penalties and inexact updates.

3. **More applications**

   Study larger neural-network verification problems, convexification and other learning tasks.

`github.com/ynelhu/aBMF`

Contact: ynel.hu@gmail.com