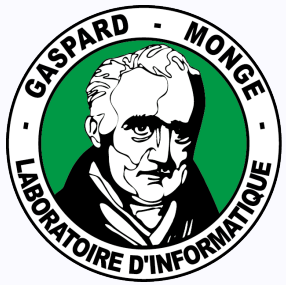
Trust-region methods for Burer-Monteiro factorizations of phase retrieval problems

Clément Royer

PGMO Days

November 17, 2026





Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
UNIVERSITÉ GUSTAVE EIFFEL

Collaborators





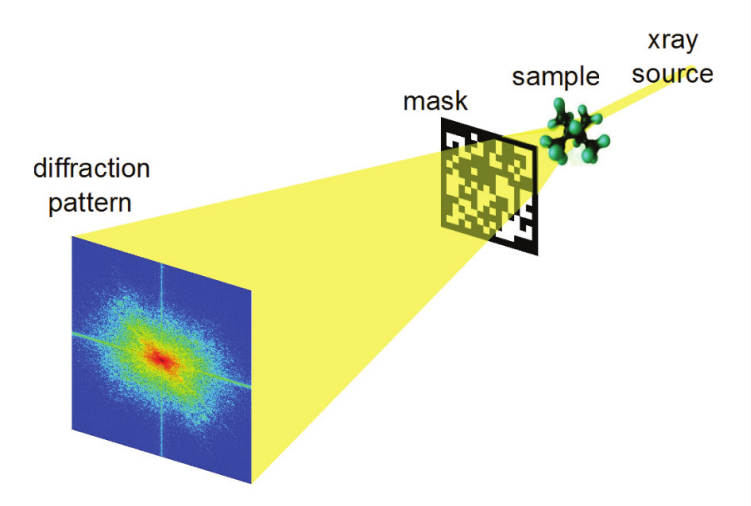
- Paul Caucheteux (ENSAE).
- Florentin Goyens (UCLouvain).
- Irène Waldspurger (CNRS & ENS-PSL).

Motivation: Phase retrieval

Phase retrieval problem

find $x \in \mathbb{C}^n$
such that $|\langle x, a_i \rangle| = b_i, \quad i = 1, \ldots, m.$

- $b_i = |\langle x^*, a_i \rangle|$: Complex moduli.
- a_i: Measurement vectors (Gaussian, STFT, wavelets, etc).



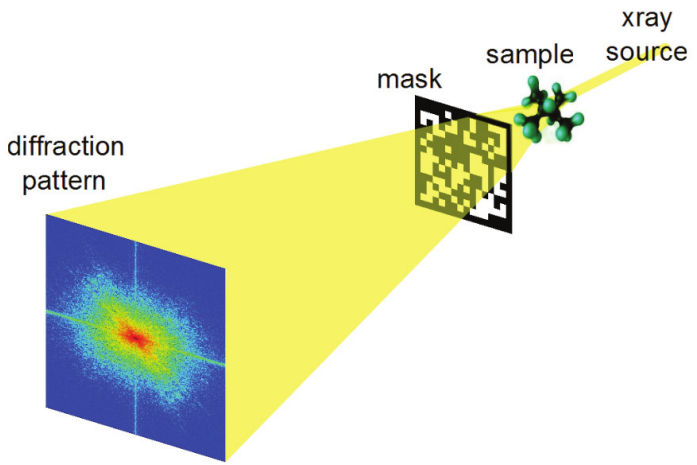
(Picture from Candès et al '15)

Motivation: Phase retrieval

Phase retrieval problem

find $x \in \mathbb{C}^n$
such that $|\langle x, a_i \rangle| = b_i, \quad i = 1, \ldots, m.$

- $b_i = |\langle x^*, a_i \rangle|$: Complex moduli.
- a_i: Measurement vectors (Gaussian, STFT, wavelets, etc).



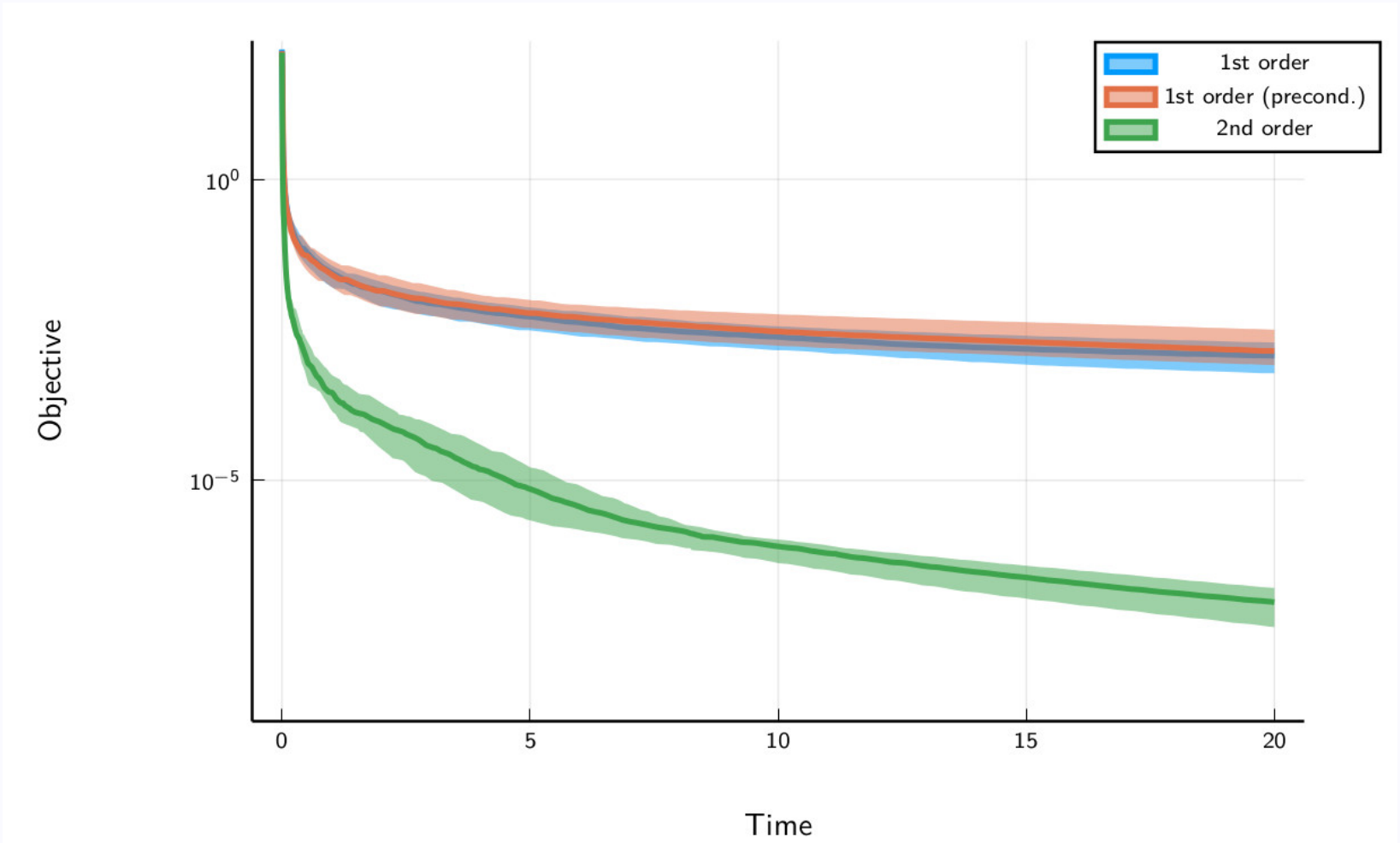
(Picture from Candès et al '15)

Gives rise to convex and **nonconvex** optimization problems!

Numerics on one phase retrieval formulation (using Julia)

Setup:

- $n = 128$, $m = 768$, wavelet measurements.
- **Gradient descent**, **Preconditioned** (Zhang,Fattahi,Zhang '21), **Trust region**.



From the example to this talk

Study the formulation

- PhaseCut/PhaseLift SDPs.
- Ill-conditioning.

Describe the algorithm

- (Riemannian) Trust Region.
- Numerical behavior.

Analyze the method

- From standard theory...
- ...to specialized convergence rates.

Outline

1. Burer-Monteiro SDPs from phase retrieval
2. Trust-region method and why it works
3. Theoretical motivation

Outline

1. Burer-Monteiro SDPs from phase retrieval
2. Trust-region method and why it works
3. Theoretical motivation

PhaseLift SDP

Recall phase retrieval: Find $x \in \mathbb{C}^n$ such that $|\langle a_i, x\rangle| = b_i \ \forall i$.

PhaseLift SDP relaxation (Candès et al. '11)

$$\begin{array}{ll} \min_{X \in \mathbb{C}^{n\times n}} & \text{trace}(X) \\ \text{subject to} & \mathcal{A}(X) = \langle X, a_i a_i^* \rangle = b_i \ \forall i \\ & X \succeq 0 \end{array}$$

$\rightarrow$ We seek a rank-1 solution.

PhaseLift SDP

Recall phase retrieval: Find $x \in \mathbb{C}^n$ such that $|\langle a_i, x\rangle| = b_i \ \forall i$.

PhaseLift SDP relaxation (Candès et al. '11)

$$\begin{array}{ll} \min_{X \in \mathbb{C}^{n\times n}} & \operatorname{trace}(X) \\ \text{subject to} & \mathcal{A}(X) = \langle X, a_i a_i^* \rangle = b_i \ \forall i \\ & X \succeq 0 \end{array}$$

$\rightarrow$ We seek a rank-1 solution.

Burer-Monteiro PhaseLift ($p \geq 1$)

$$\min_{V \in \mathbb{C}^{n\times p}} f(V) := \frac{1}{2}\|\mathcal{A}(V\,V^{\mathrm{T}}) - b\|_F^2, \qquad \mathcal{A}(X) = [\langle X, a_i\, a_i^*\rangle]_i\,.$$

- Nonconvex structured formulation.
- Favorable landscape in various cases (McRae '26).

Phase Cut

Problem Find $x \in \mathbb{C}^n$ such that $|\langle x, a_i \rangle| = b_i, \quad i = 1, \ldots, m.$

PhaseCut SDP relaxation (Waldspurger et al. '15)

$$\begin{array}{ll} \min_{Y \in \mathbb{C}^{m \times m}} & \langle Y, M \rangle \\ \text{subject to} & Y_{ii} = 1 \ \forall i \\ & Y \succeq 0. \end{array} \quad \text{where} \quad \begin{cases} M &= \operatorname{diag}(b)(I - A\,A^\dagger)\operatorname{diag}(b) \\ A &= \begin{bmatrix} a_1^* \\ \vdots \\ a_m^* \end{bmatrix}. \end{cases}$$

Phase Cut

Problem Find $x \in \mathbb{C}^n$ such that $|\langle x, a_i \rangle| = b_i, \quad i = 1, \ldots, m.$

PhaseCut SDP relaxation (Waldspurger et al. '15)

$$\begin{array}{ll} \min_{Y \in \mathbb{C}^{m \times m}} & \langle Y, M \rangle \\ \text{subject to} & Y_{ii} = 1 \ \forall i \\ & Y \succeq 0. \end{array} \quad \text{where} \quad \begin{cases} M &= \operatorname{diag}(b)(I - A\,A^\dagger)\operatorname{diag}(b) \\ A &= \begin{bmatrix} a_1^* \\ \vdots \\ a_m^* \end{bmatrix}. \end{cases}$$

Burer-Monteiro factorization

$$\begin{array}{ll} \min_{V \in \mathbb{C}^{m \times p}} & \langle V, MV \rangle \\ \text{subject to} & \|V_{i,:}\| = 1 \ \forall i. \end{array}$$

Focus: Solve the problem using $1 < p \ll m$.

Our goal

Develop efficient algorithms for PhaseLift and/or PhaseCut.

Challenges

- Competitors:
  - First-order methods (Gradient descent and preconditioned variants, alternating projections)
  - Low-rank SDP solvers (Chen and Goulart '22, Monteiro, Sujanani and Cifuentes and Monteiro '24-'26).
- Formulations:
  - Nonconvexity may not be benign.
  - Objective function may be ill-conditioned.

Ill-conditioned formulations

Generic formulation

$$\min_{V \in \mathcal{M}} f(V) = g(V\,V^{\mathrm{T}}), \qquad V \in \mathbb{R}^{n \times p} \quad \text{or} \quad V \in \mathbb{R}^{m \times p}.$$

- We seek a rank-1 solution...
- ...and use $p \geq 1$.

Bad conditioning Can arise from:

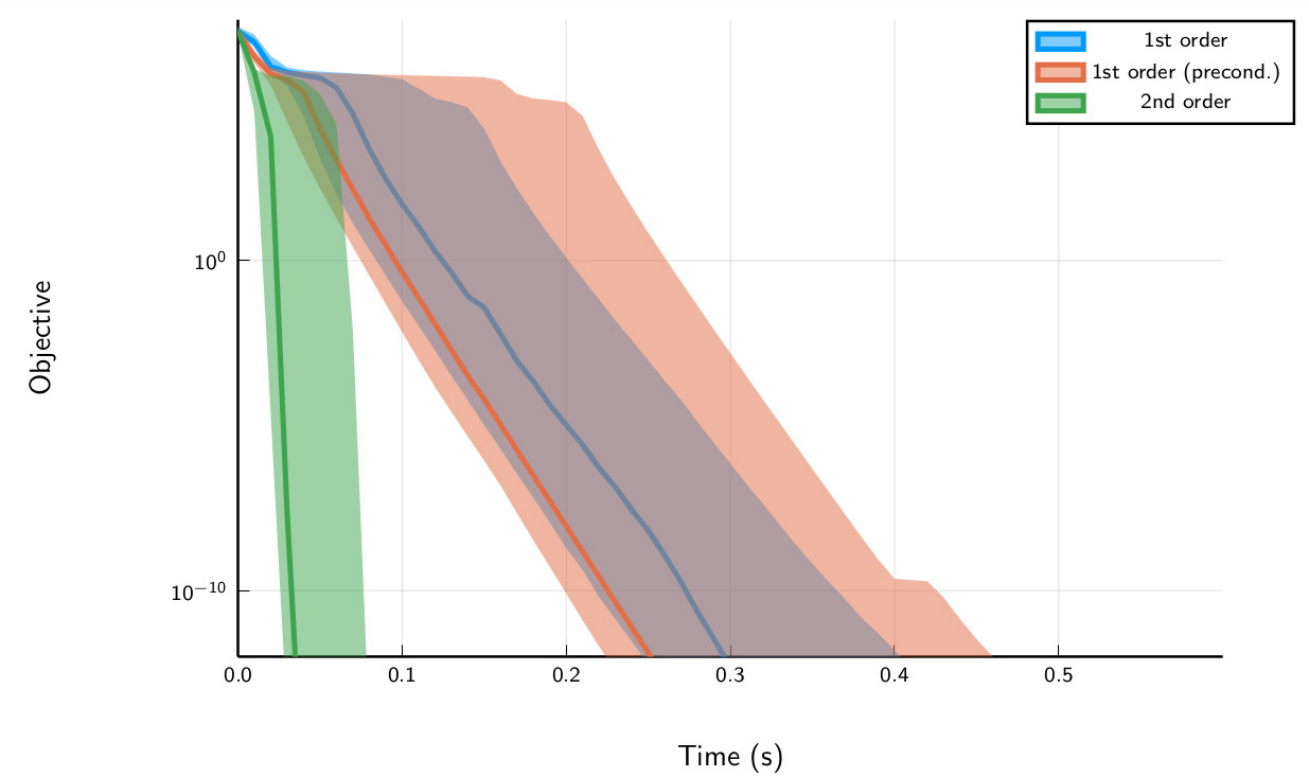
- g being ill-conditioned (due to measurement operators).
- Overparameterization.

Why conditioning matters

Same setup as intro

- Gradient descent VS Preconditioned VS Trust region.
- $n = 128$, $m = 6n = 768$.

Good conditioning ($p = 1$, random measurements)

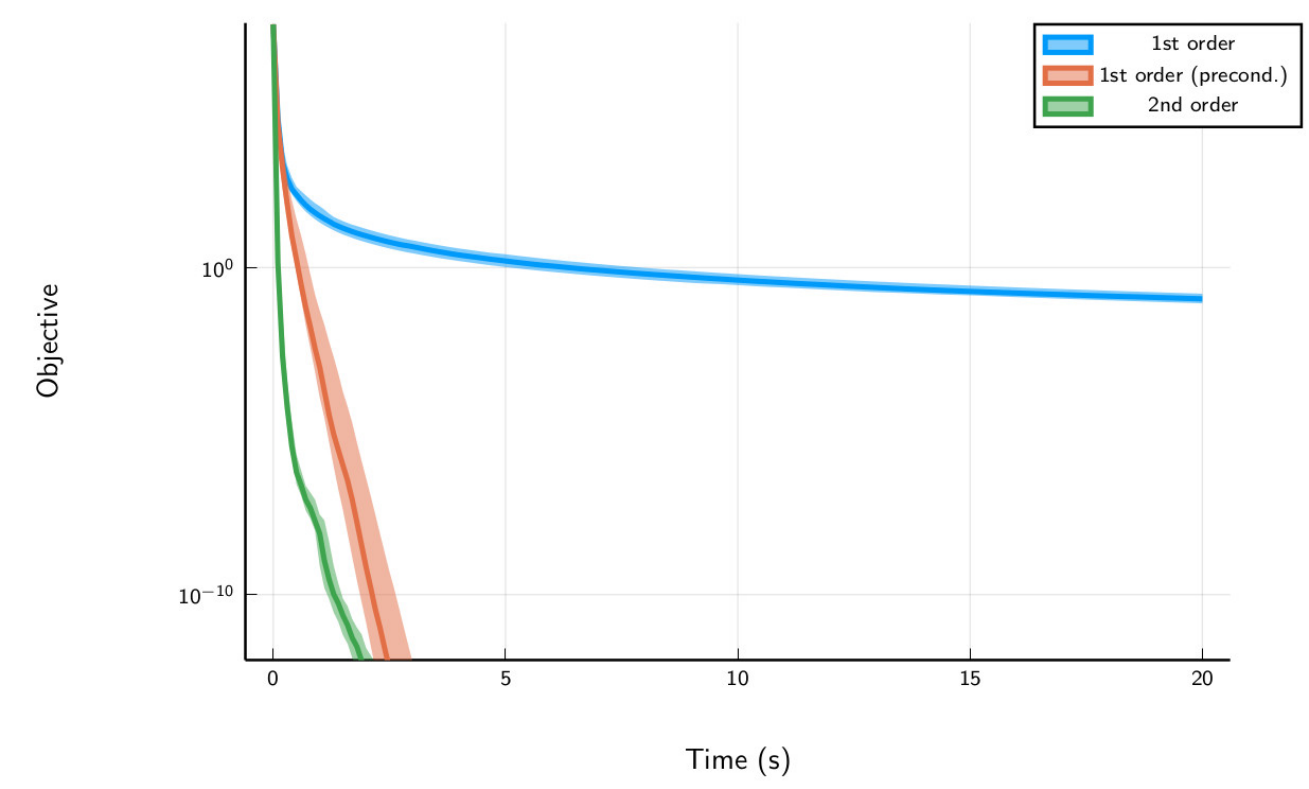


Why conditioning matters

Same setup as intro

- Gradient descent VS Preconditioned VS Trust region.
- $n = 128$, $m = 6n = 768$.

Bad conditioning ($p = 2$, random measurements)

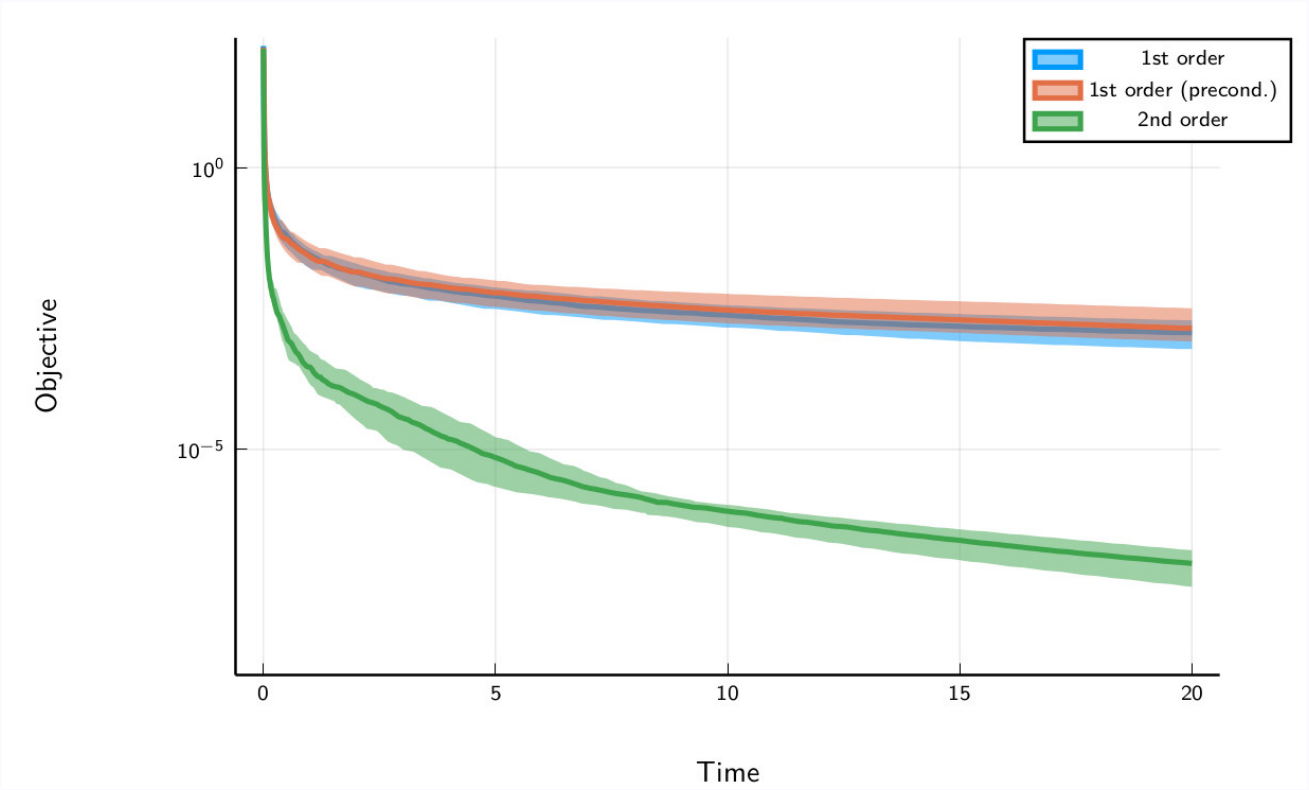


Why conditioning matters

Same setup as intro

- Gradient descent VS Preconditioned VS Trust region.
- $n = 128$, $m = 6n = 768$.

Bad conditioning ($p = 2$+wavelet measurements)



Outline

1. Burer-Monteiro SDPs from phase retrieval
2. Trust-region method and why it works
3. Theoretical motivation

Trust region for $\min_{V \in \mathcal{M}} f(V)$

Inputs: $V_0 \in \mathcal{M}$, $0 < \Delta_0 \leq \Delta_{\text{max}}$, $\eta_2 \geq \eta_1 > 0$.

For k=$0, 1, 2, \ldots$

1. Compute
   $S_k \in \operatorname{argmin}_{\substack{S \in \mathcal{T}_{V_k}\mathcal{M} \\ \|S\| \leq \Delta_k}} m_k(V_k + s) = \langle g(V_k), S \rangle + \frac{1}{2}\langle S, \mathcal{H}(V_k)S$.
2. Define $V_k^{\mathcal{M}}$ as the retraction of $V_k + S_k$ onto $\mathcal{M}$.
3. Compute $\rho_k = \frac{f(V_k) - f(V_k^{\mathcal{M}})}{m_k(V_k) - m_k(V_k^{\mathcal{M}})}$.
4. If $\rho_k \geq \eta_1$, set $V_{k+1} = V_k^{\mathcal{M}}$ and

$$\Delta_{k+1} = \begin{cases} \min(2\Delta_k, \Delta_{\text{max}}) & \text{if } \rho_k > \eta_2 \text{ and } \|S_k\| = \Delta_k \\ \Delta_k & \text{otherwise.} \end{cases}$$

5. Otherwise, set $V_{k+1} = V_k$ and $\Delta_{k+1} = \frac{1}{4}\Delta_k$.

Solving the trust-region subproblem

Trust-region subproblem

$$\min_{S \in \mathcal{T}_{V_k}\mathcal{M}} < g(V_k), s > + \tfrac{1}{2} < S, \mathcal{H}(V_k)S > \quad \text{s.t.} \quad \|S\| \leq \Delta_k.$$

- Apply conjugate gradient (CG) to the linear system $\mathcal{H}(V_k)S = -g(V_k)$
- Stop when residual $\|\mathcal{H}(V_k)S + g(V_k)\|$ is small enough or $\|S\| = \Delta_k$;
- For $\mathcal{H}(V_k) \not\succeq 0$: if **negative curvature** is encountered, take a negative curvature step such that $\|S\| = \Delta_k$.

Solving the trust-region subproblem

Trust-region subproblem

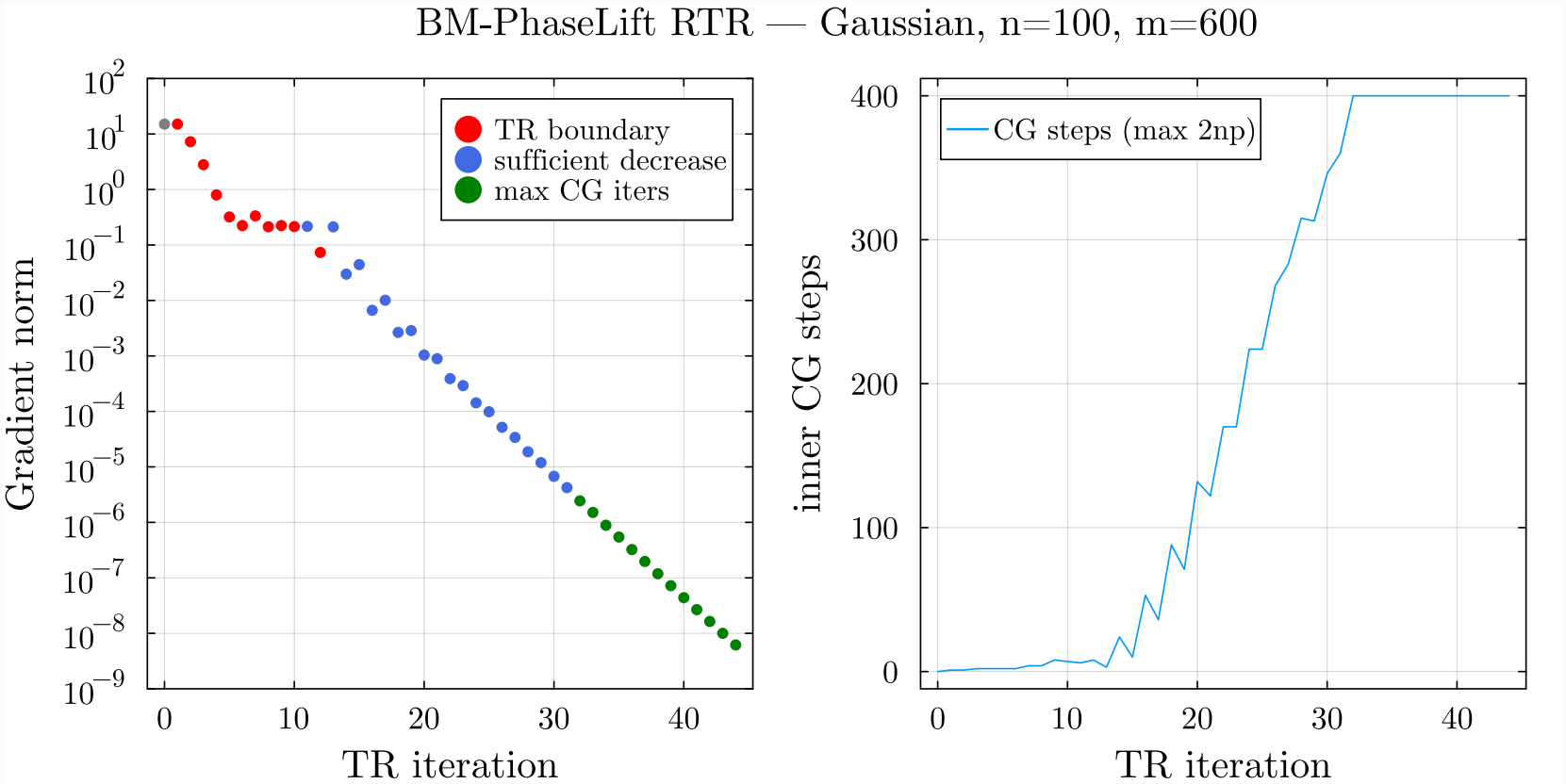
$$\min_{S \in \mathcal{T}_{V_k}^{\mathcal{M}}} < g(V_k), s > + \tfrac{1}{2} < S, \mathcal{H}(V_k)S > \quad \text{s.t.} \quad \|S\| \leq \Delta_k.$$

- Apply conjugate gradient (CG) to the linear system $\mathcal{H}(V_k)S = -g(V_k)$
- Stop when residual $\|\mathcal{H}(V_k)S + g(V_k)\|$ is small enough or $\|S\| = \Delta_k$;
- For $\mathcal{H}(V_k) \not\succeq 0$: if **negative curvature** is encountered, take a negative curvature step such that $\|S\| = \Delta_k$.

Convergence of Trust Region with CG

- Influenced by Hessian eigenvalues (i.e. conditioning).
- Small eigenvalues are a typical issue.

Using CG in Trust Region



Numerical comparison

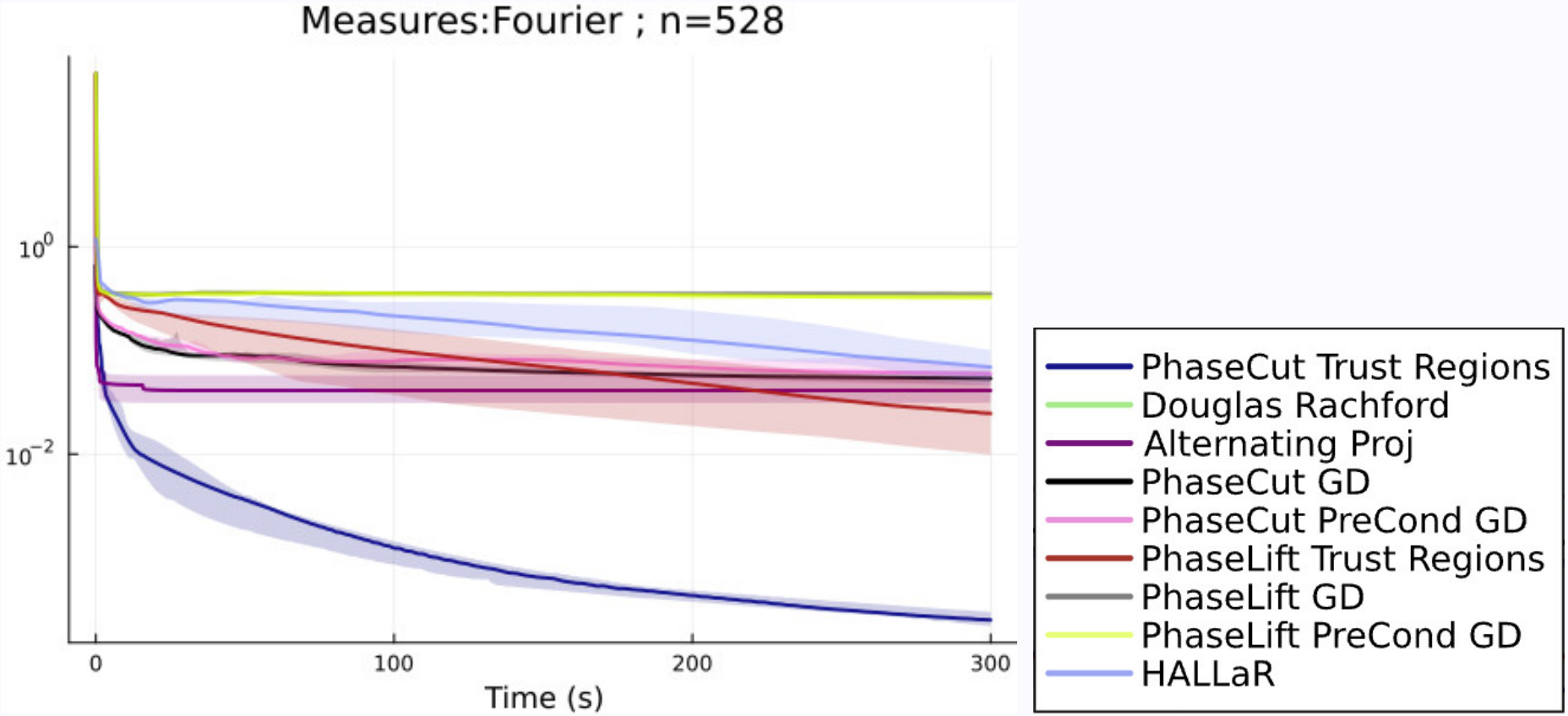
Competitors (non-exhaustive!)

- Gradient descent.
- Preconditioned gradient descent, two versions (Tong, Ma, Chi '21, Zhang, Fattahi, Zhang '21).
- Alternating methods: Douglas-Rachford, Alternating projections.
- L-BFGS.
- Mixing method (Wang, Chang, Kolter '17)
- BM-ADMM (Chen and Goulart '23).
- HALLAR (Monteiro, Sujanani, Cifuentes '24).

→ Use variants for both PhaseLift and PhaseCut when possible.
→ We had to implement most of those!

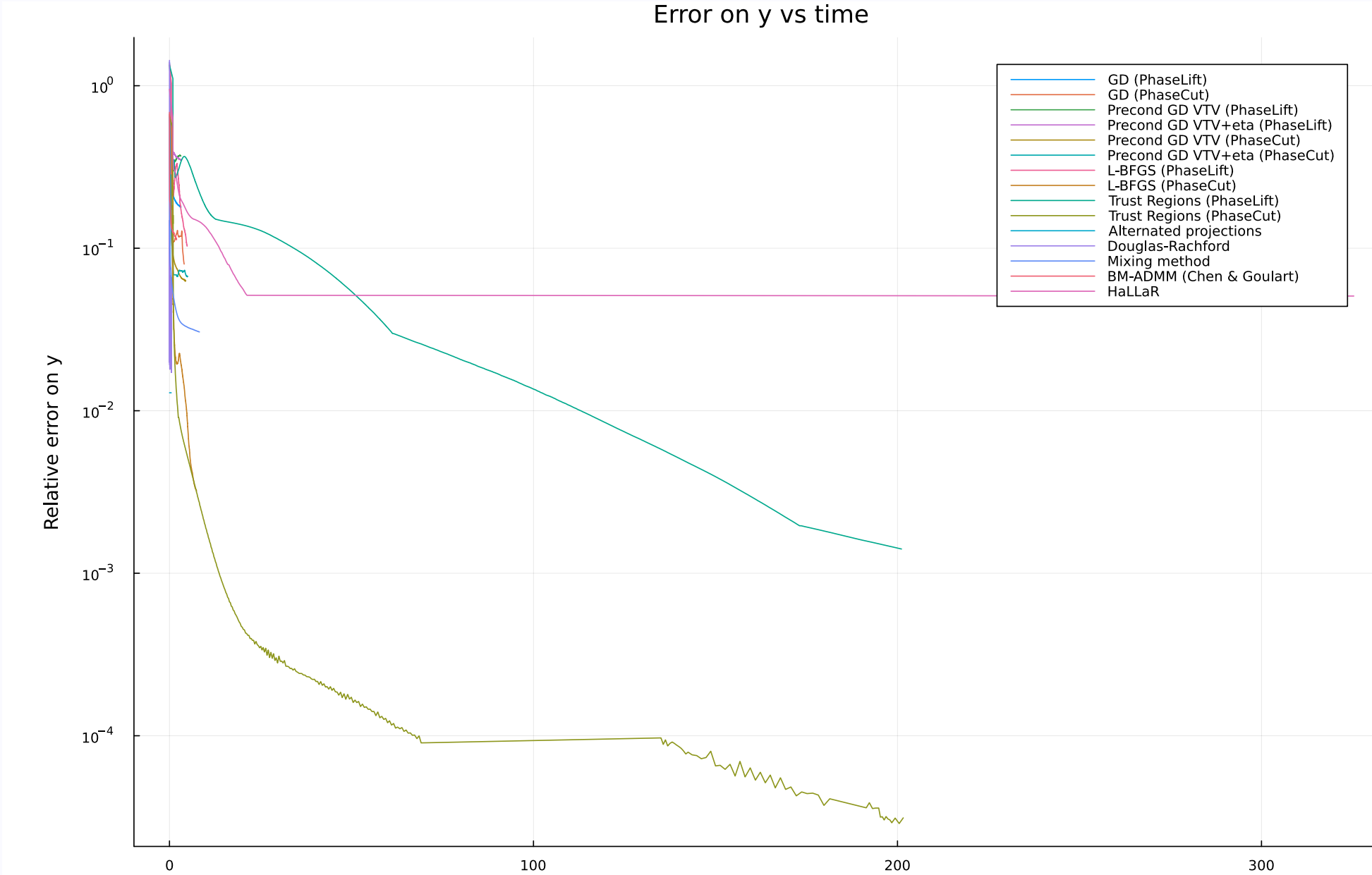
Implementation In Julia, based on `Manopt.jl` and `Manifold.jl`.

Our first comparison

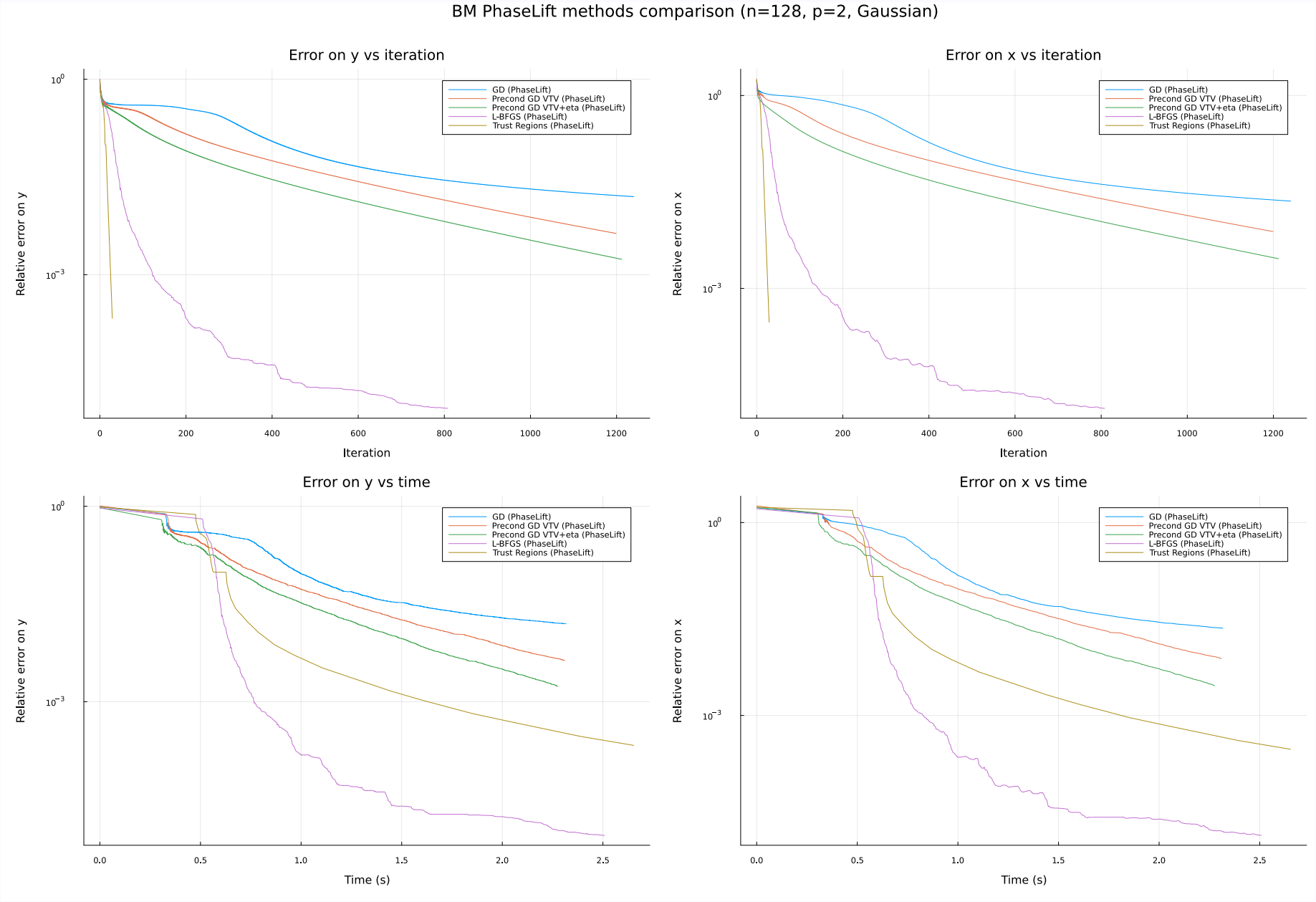


- PhaseCut + Trust Region often yields best performance.
- PhaseLift interesting too!

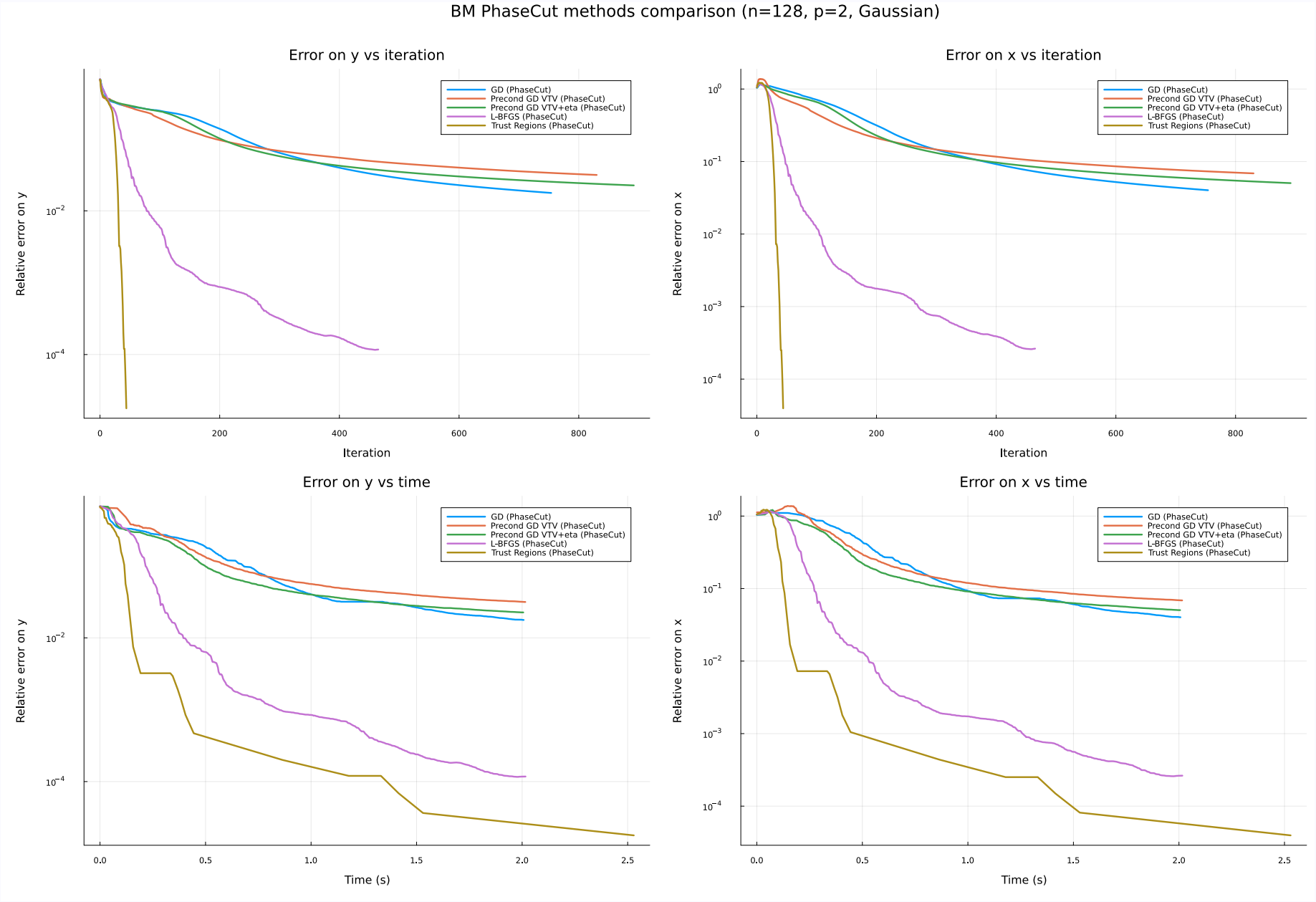
Our broadest comparison (Gaussian measurements)



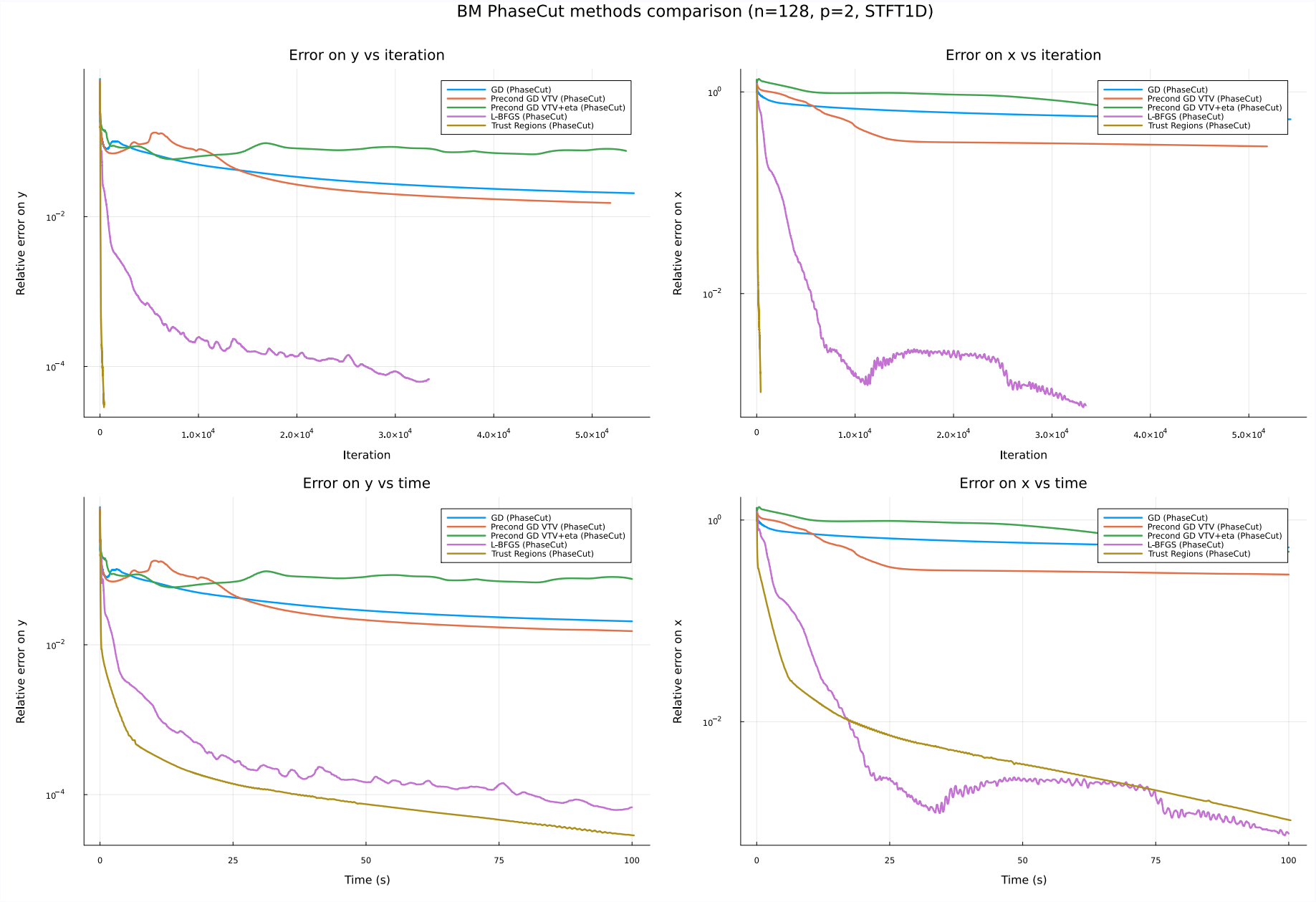
On PhaseLift



On PhaseCut (Gaussian)



On PhaseCut (STFT)



Outline

1. Burer-Monteiro SDPs from phase retrieval
2. Trust-region method and why it works
3. Theoretical motivation

Setup: Manifold optimization

Main problem

$$\underset{V \in \mathcal{M}}{\text{minimize}}\, f(V),$$

- f smooth ($\mathcal{C}^2$), nonconvex function.
- $\mathcal{M}$ Riemannian manifold.

Setup: Manifold optimization

Main problem

$$\underset{V \in \mathcal{M}}{\text{minimize}}\, f(V),$$

- f smooth ($\mathcal{C}^2$), nonconvex function.
- $\mathcal{M}$ Riemannian manifold.

Two possible goals

- Find approximate stationary points (complexity/global convergence rate).
- Fast convergence near a minimizer (local convergence).

Convergence results for trust region

Complexity results

Goal: Bound on K to satisfy $\|g(V_K)\| \leq \epsilon_{\mathrm{g}}$ and $\lambda_{\min}\left(\mathcal{H}(V_K)\right) \geq -\epsilon_{\mathrm{H}}$.

- Standard result (Boumal et al '19): $K = \mathcal{O}\left(\max\{\epsilon_{\mathrm{g}}^{-2}, \epsilon_{\mathrm{H}}^{-3}\}\right)$.
- Improved variants/optimal (Curtis et al '21, Goyens and R. '24): $\mathcal{O}\left(\max\{\epsilon_{\mathrm{g}}^{-3/2}, \epsilon_{\mathrm{H}}^{-3}\}\right)$.

Convergence results for trust region

Complexity results

Goal: Bound on K to satisfy $\|g(V_K)\| \leq \epsilon_{\mathrm{g}}$ and $\lambda_{\min}\left(\mathcal{H}(V_K)\right) \geq -\epsilon_{\mathrm{H}}$.

- Standard result (Boumal et al '19): $K = \mathcal{O}\left(\max\{\epsilon_{\mathrm{g}}^{-2}, \epsilon_{\mathrm{H}}^{-3}\}\right)$.
- Improved variants/optimal (Curtis et al '21, Goyens and R. '24): $\mathcal{O}\left(\max\{\epsilon_{\mathrm{g}}^{-3/2}, \epsilon_{\mathrm{H}}^{-3}\}\right)$.

Local convergence rates

If f strongly convex near minimizer, then

- Superlinear local convergence holds (Absil et al '07).
- In presence of benign landscape, the convergence can even be quadratic (Goyens and R. '24).

But this does not apply in our setup!

What we want

Burer-Monteiro PhaseCut

$$\min_{V \in \mathbb{C}^{m \times p}} \quad \langle V, MV \rangle \qquad \text{subject to} \qquad \|V_{i,:}\| = 1 \quad \forall i = 1, \ldots, m.$$

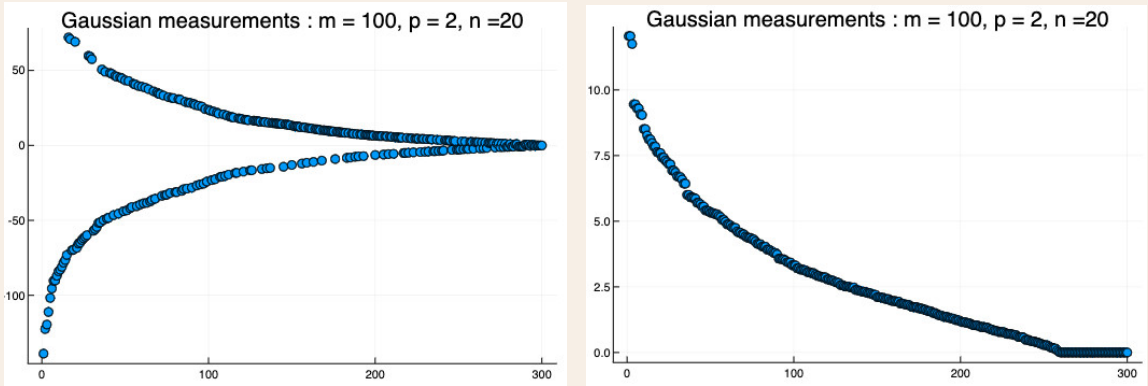
What we want

Burer-Monteiro PhaseCut

$$\min_{V \in \mathbb{C}^{m \times p}} \quad \langle V, MV \rangle \qquad \text{subject to} \qquad \|V_{i,:}\| = 1 \quad \forall i = 1, \ldots, m.$$

PhaseCut and critical points

- Only solutions do not have negative curvature.
- The function is not strongly convex around minima!



Hessian eigenvalues at random point (left) and solution (right).

A theorem

Toy PhaseLift $\min_{V \in \mathbb{C}^{n \times p}} \tilde{f}(V) := \frac{1}{4} \| V\, V^{\mathrm{T}} - e_1 e_1^{\mathrm{T}} \|_F^2.$

A theorem

Toy PhaseLift $\min_{V \in \mathbb{C}^{n \times p}} \tilde{f}(V) := \frac{1}{4} \| V\, V^{\mathrm{T}} - e_1 e_1^{\mathrm{T}} \|_F^2$.

Theorem (Goyens, R., Waldspurger '26+)

Suppose $p \geq 2$. If V_0 close enough to a solution, then

$$\tilde{f}(V_K) - \min_V \tilde{f}(V) \leq \epsilon \left(\tilde{f}(V_0) - \min_V \tilde{f}(V) \right)$$

for $K = \mathcal{O}(\log(1/\epsilon))$.

A theorem

Toy PhaseLift $\min_{V \in \mathbb{C}^{n \times p}} \tilde{f}(V) := \frac{1}{4} \| V\, V^{\mathrm{T}} - e_1 e_1^{\mathrm{T}} \|_F^2$.

Theorem (Goyens, R., Waldspurger '26+)

Suppose $p \geq 2$. If V_0 close enough to a solution, then

$$\tilde{f}(V_K) - \min_V \tilde{f}(V) \leq \epsilon \left(\tilde{f}(V_0) - \min_V \tilde{f}(V) \right)$$

for $K = \mathcal{O}(\log(1/\epsilon))$.

- Local result but no need for PŁ/ strong convexity near solution.
- On par with other analyses (Rebjock and Boumal '24-'25)

Tales from the proof

- **Model decrease** At every iteration,

$$m_k(V_k) - m_k(V_k + S_k) \geq C \min\left(\frac{\Delta_k}{\sigma_k}, 1\right) \tilde{f}(V_k),$$

where σ_k measures distance to optimal solution.
$\rightarrow$ Replaces classical Cauchy decrease in trust region.

Tales from the proof

- **Model decrease** At every iteration,

$$m_k(V_k) - m_k(V_k + S_k) \geq C \min\left(\frac{\Delta_k}{\sigma_k}, 1\right) \tilde{f}(V_k),$$

  where σ_k measures distance to optimal solution.
  $\rightarrow$ Replaces classical Cauchy decrease in trust region.
- **Lyapunov function** Complicated formula depending on $\frac{\Delta_k}{\sigma_k}$, Δ_k and $\tilde{f}(V_k)$.
  $\rightarrow$ Resembles those used in stochastic trust region (Chen, Menickelly, Scheinberg '18).

Tales from the proof

- **Model decrease** At every iteration,

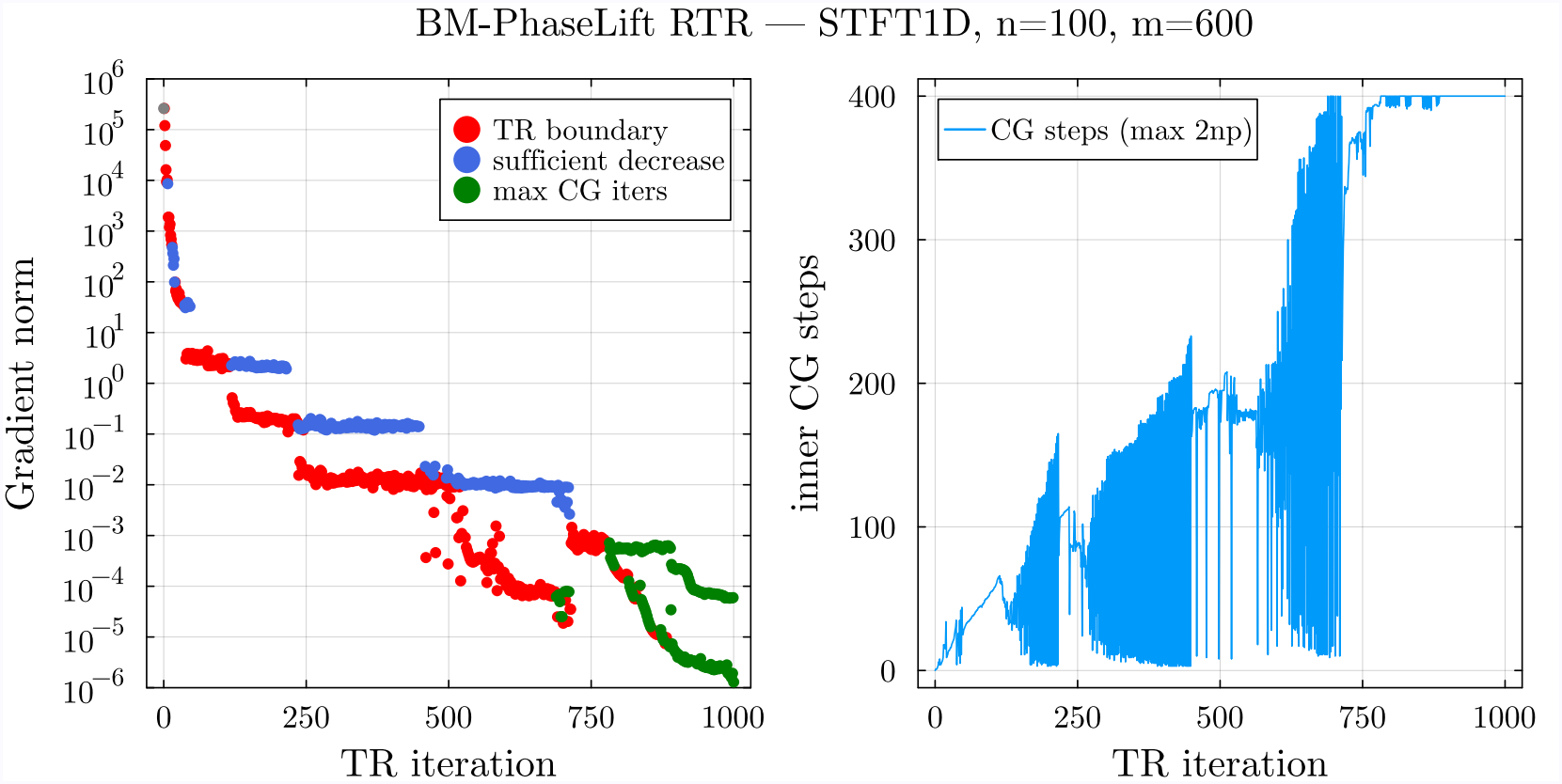
$$m_k(V_k) - m_k(V_k + S_k) \geq C \min\left(\frac{\Delta_k}{\sigma_k}, 1\right) \tilde{f}(V_k),$$

  where σ_k measures distance to optimal solution.
  $\rightarrow$ Replaces classical Cauchy decrease in trust region.
- **Lyapunov function** Complicated formula depending on $\frac{\Delta_k}{\sigma_k}$, Δ_k and $\tilde{f}(V_k)$.
  $\rightarrow$ Resembles those used in stochastic trust region (Chen, Menickelly, Scheinberg '18).

What we believe

- The result is true for well-conditioned measurements.
- It can inform us on what will go wrong in the ill-conditioned case!

An open question: CG behavior



Can we quantify what is happening in terms of CG iterations?

Time to conclude

Phase retrieval

- We like PhaseLift and PhaseCut!
- But there are badly conditioned instances.

Trust Region

- Easy to implement via `Manopt`.
- Good results on.

Guarantees

- Limited assumptions.
- Some (local) convergence rates.

That's it!

References

- F. Goyens and C. W. Royer, *Riemannian trust-region methods for strict saddle functions with complexity guarantees*, Mathematical Programming, 2024.
- P. Caucheteux, F. Goyens, C. W. Royer and I. Waldspurger, *Trust-region methods for Burer-Monteiro factorizations of phase retrieval problems*, working paper, 2026.

That's it!

References

- F. Goyens and C. W. Royer, *Riemannian trust-region methods for strict saddle functions with complexity guarantees*, Mathematical Programming, 2024.
- P. Caucheteux, F. Goyens, C. W. Royer and I. Waldspurger, *Trust-region methods for Burer-Monteiro factorizations of phase retrieval problems*, working paper, 2026.

Merci!

`clement.royer@univ-eiffel.fr`