

Trust-region algorithms for nonconvex optimization and application to phase retrieval

Clément Royer

LIGM, 1er avril 2026

Dauphine
UNIVERSITÉ PARIS

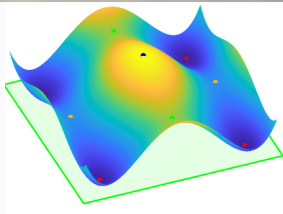
| PSL 

PR[AI]RIE
PaRis Artificial Intelligence Research InstitutE



- Paul Caucheteux (ENSAE),
- Florentin Goyens (UCLouvain),
- Irène Waldspurger (CNRS & Université Paris Dauphine-PSL).

Me and nonconvex optimization



Top: Pic du Midi d'Ossau (Royer '20), Bottom: Nonconvex landscape (Wright & Ma '22)

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} f(\mathbf{x})$$

$f \in \mathcal{C}^2$, bounded below and **nonconvex**.

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} f(\mathbf{x})$$

$f \in \mathcal{C}^2$, bounded below and **nonconvex**.

$\mathbf{x}^* \in \mathbb{R}^n$ is

- a **global minimum** of f if $f(\mathbf{x}^*) \leq f(\mathbf{x}) \forall \mathbf{x} \in \mathbb{R}^n$
→ NP-hard to find/check! (Murti and Kabadi '87)

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} f(\mathbf{x})$$

$f \in \mathcal{C}^2$, bounded below and **nonconvex**.

$\mathbf{x}^* \in \mathbb{R}^n$ is

- a **global minimum** of f if $f(\mathbf{x}^*) \leq f(\mathbf{x}) \forall \mathbf{x} \in \mathbb{R}^n$
→ NP-hard to find/check! (Murti and Kabadi '87)
- a **local minimum** of f if $f(\mathbf{x}^*) \leq f(\mathbf{x}) \forall \mathbf{x}$ close enough to \mathbf{x}^*
→ NP-hard to find/check!

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} f(\mathbf{x})$$

$f \in \mathcal{C}^2$, bounded below and **nonconvex**.

$\mathbf{x}^* \in \mathbb{R}^n$ is

- a **global minimum** of f if $f(\mathbf{x}^*) \leq f(\mathbf{x}) \forall \mathbf{x} \in \mathbb{R}^n$
→ NP-hard to find/check! (Murti and Kabadi '87)
- a **local minimum** of f if $f(\mathbf{x}^*) \leq f(\mathbf{x}) \forall \mathbf{x}$ close enough to \mathbf{x}^*
→ NP-hard to find/check!
- a **first-order stationary point** of f if $\|\nabla f(\mathbf{x}^*)\| = 0$
→ PLS-complete to find (Hollender and Zampetakis '25).

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} f(\mathbf{x})$$

$f \in \mathcal{C}^2$, bounded below and **nonconvex**.

$\mathbf{x}^* \in \mathbb{R}^n$ is

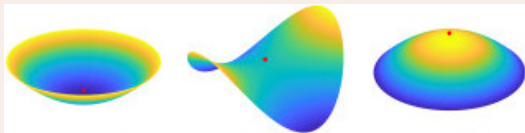
- a **global minimum** of f if $f(\mathbf{x}^*) \leq f(\mathbf{x}) \forall \mathbf{x} \in \mathbb{R}^n$
→ NP-hard to find/check! (Murti and Kabadi '87)
- a **local minimum** of f if $f(\mathbf{x}^*) \leq f(\mathbf{x}) \forall \mathbf{x}$ close enough to \mathbf{x}^*
→ NP-hard to find/check!
- a **first-order stationary point** of f if $\|\nabla f(\mathbf{x}^*)\| = 0$
→ PLS-complete to find (Hollender and Zampetakis '25).
- a **second-order stationary point** if $\|\nabla f(\mathbf{x}^*)\| = 0$ and $\nabla^2 f(\mathbf{x}^*) \succeq 0$
→ Already NP-hard for quartic polynomials.

Approximately solving minimize $_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$

Goal: Reach (ϵ_g, ϵ_H) -points

$$\|\nabla f(\mathbf{x})\| \leq \epsilon_g \quad \text{and} \quad \nabla^2 f(\mathbf{x}) \succeq -\epsilon_H \mathbf{I}.$$

- $\epsilon_g = \epsilon_H = 0$: Second-order stationary point.
- $\epsilon_g = 0$ but second condition false: Saddle points or maxima.

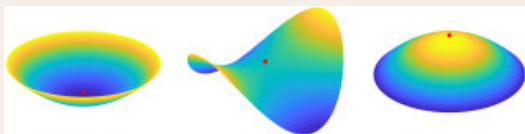


Approximately solving minimize $\mathbf{x} \in \mathbb{R}^n f(\mathbf{x})$

Goal: Reach (ϵ_g, ϵ_H) -points

$$\|\nabla f(\mathbf{x})\| \leq \epsilon_g \quad \text{and} \quad \nabla^2 f(\mathbf{x}) \succeq -\epsilon_H \mathbf{I}.$$

- $\epsilon_g = \epsilon_H = 0$: Second-order stationary point.
- $\epsilon_g = 0$ but second condition false: Saddle points or maxima.



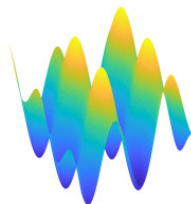
Complexity analysis

Cost of an algorithm to reach an (ϵ_g, ϵ_H) -point:

- Faster means lower cost in terms of ϵ_g/ϵ_H .
- Quality of (ϵ_g, ϵ_H) -points depend on the function **landscape**.

Bad nonconvex instances

- Local, non-global minima.
- High-order saddle points.



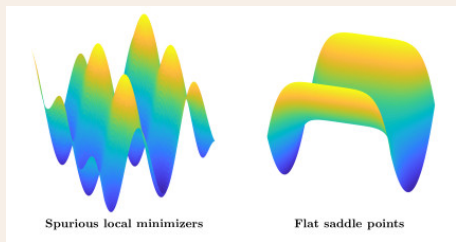
Spurious local minimizers



Flat saddle points

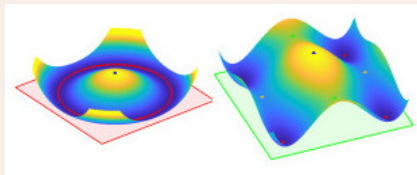
Bad nonconvex instances

- Local, non-global minima.
- High-order saddle points.



Nice instances

- All local minima are global.
- Strict (non-flat) saddle points.



Figures: J. Wright and Y. Ma, *High-Dimensional Data Analysis with Low-Dimensional Models*, 2022.

What this talk is about

Nonconvex optimization

- Go beyond the (hard) general setting.
- Analyze one method (trust region).
- Study one problem (phase retrieval).

What this talk is about

Nonconvex optimization

- Go beyond the (hard) general setting.
- Analyze one method (trust region).
- Study one problem (phase retrieval).

Strict saddle problems

- Nonconvex problems with good landscape.
- Provably faster convergence of trust region.

Nonconvex optimization

- Go beyond the (hard) general setting.
- Analyze one method (trust region).
- Study one problem (phase retrieval).

Strict saddle problems

- Nonconvex problems with good landscape.
- Provably faster convergence of trust region.

Beyond strict saddle

- Ill-conditioned formulations.
- Still fast convergence observed!

- 1 From nonconvex to strict saddle functions
- 2 Algorithms for strict saddle problems
- 3 Beyond strict saddle problems
- 4 Conclusion and perspectives

- 1 From nonconvex to strict saddle functions
- 2 Algorithms for strict saddle problems
- 3 Beyond strict saddle problems
- 4 Conclusion and perspectives

Definition

$f : \mathbb{R}^n \times \mathbb{R} \in \mathcal{C}^2$ has **benign landscape** if

$$\|\nabla f(\mathbf{x})\| = 0 \text{ and } \nabla^2 f(\mathbf{x}) \succeq 0 \iff \mathbf{x} \in \underset{\mathbf{x} \in \mathbb{R}^n}{\operatorname{argmin}} f(\mathbf{x}).$$

- Every local minimum is global.
- All saddle points satisfy $\nabla^2 f(\mathbf{x}) \not\succeq 0 \rightarrow$ Strict saddle points.

Definition

$f : \mathbb{R}^n \times \mathbb{R}^{\mathcal{C}^2}$ has **benign landscape** if

$$\|\nabla f(\mathbf{x})\| = 0 \text{ and } \nabla^2 f(\mathbf{x}) \succeq 0 \iff \mathbf{x} \in \underset{\mathbf{x} \in \mathbb{R}^n}{\operatorname{argmin}} f(\mathbf{x}).$$

- Every local minimum is global.
- All saddle points satisfy $\nabla^2 f(\mathbf{x}) \not\succeq 0 \rightarrow$ Strict saddle points.

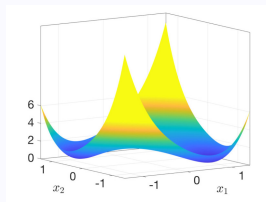
Typical sources of benign nonconvex landscape

- Nonconvex formulations with symmetries.
 - \rightarrow Multiple (redundant) solutions, easy to go from one to another.
- Overparameterized models.
 - \rightarrow Introduces saddle points but those are strict.

Some examples (pictures from Chi et al '19; Wright and Ma '22)

Rank-1 approximation

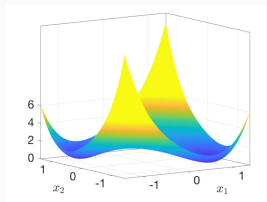
$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \frac{1}{4} \|\mathbf{x}\mathbf{x}^T - \mathbf{M}\|_F^2$$



Some examples (pictures from Chi et al '19; Wright and Ma '22)

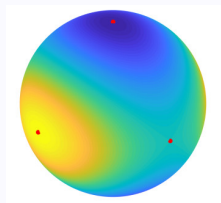
Rank-1 approximation

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \frac{1}{4} \|\mathbf{x}\mathbf{x}^T - \mathbf{M}\|_F^2$$



Minimum Eigenvalue

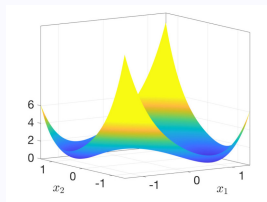
$$\underset{\|\mathbf{x}\|=1}{\min} \mathbf{x}^T \mathbf{A} \mathbf{x}$$



Some examples (pictures from Chi et al '19; Wright and Ma '22)

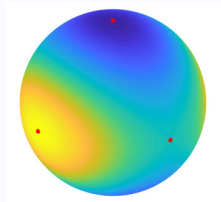
Rank-1 approximation

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \frac{1}{4} \|\mathbf{x}\mathbf{x}^T - \mathbf{M}\|_F^2$$



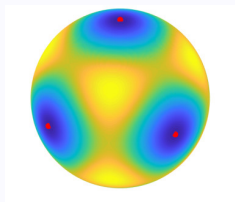
Minimum Eigenvalue

$$\underset{\|\mathbf{x}\|=1}{\min} \mathbf{x}^T \mathbf{A} \mathbf{x}$$



Tensor optimization

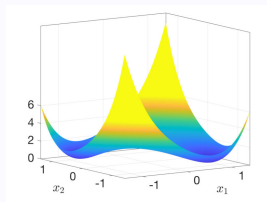
$$\underset{\|\mathbf{x}\|=1}{\min} T(\mathbf{x}, \mathbf{x}, \mathbf{x}, \mathbf{x})$$



Some examples (pictures from Chi et al '19; Wright and Ma '22)

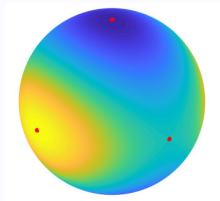
Rank-1 approximation

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \frac{1}{4} \|\mathbf{x}\mathbf{x}^T - \mathbf{M}\|_F^2$$



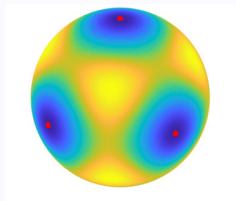
Minimum Eigenvalue

$$\underset{\|\mathbf{x}\|=1}{\min} \mathbf{x}^T \mathbf{A} \mathbf{x}$$



Tensor optimization

$$\underset{\|\mathbf{x}\|=1}{\min} T(\mathbf{x}, \mathbf{x}, \mathbf{x}, \mathbf{x})$$



For more: <https://sunju.org/research/nonconvex/>

What is a “nice” nonconvex problem?

- All local minima are global?
- All saddle points can be escaped?
- Algorithms work well?
- All of the above?

Defining “good nonconvexity” mathematically

What is a “nice” nonconvex problem?

- All local minima are global?
- All saddle points can be escaped?
- Algorithms work well?
- All of the above?

Numerous existing results

- $P\epsilon, K\epsilon$, quadratic growth, error bound (Rebjoek and Boumal '24);
- Regional complexity analysis (Curtis and Robinson '21);
- Phase complexity (Cartis, Gould, Toint '22).

Defining “good nonconvexity” mathematically

What is a “nice” nonconvex problem?

- All local minima are global?
- All saddle points can be escaped?
- Algorithms work well?
- All of the above?

Numerous existing results

- $P\epsilon, K\epsilon$, quadratic growth, error bound (Rebjoek and Boumal '24);
- Regional complexity analysis (Curtis and Robinson '21);
- Phase complexity (Cartis, Gould, Toint '22).

Our focus: Strict saddle property

- Parametric definition.
- Allows for complexity analysis.

Starting point: Strongly convex functions

- $f : \mathbb{R}^n \rightarrow \mathbb{R}$ \mathcal{C}^2 is γ -strongly convex if

$$\nabla^2 f(\mathbf{x}) \succeq \gamma I.$$

- Unique minimum $\mathbf{x}^* \in \mathbb{R}^n$.

One consequence of γ -strong convexity

For any $\alpha > 0$,

$$\|\nabla f(\mathbf{x})\| \leq \alpha \quad \Rightarrow \quad \|\mathbf{x} - \mathbf{x}^*\| \leq \frac{2\alpha}{\gamma}.$$

Definition (Ge et al '17; Goyens, R. '24)

$f : \mathbb{R}^n \rightarrow \mathbb{R}$ \mathcal{C}^2 is $(\alpha, \beta, \gamma, \delta)$ -strict saddle with $\alpha, \beta, \gamma, \delta > 0$ if for any $\mathbf{x} \in \mathbb{R}^n$, one of these properties holds:

- 1 $\|\nabla f(\mathbf{x})\| \geq \alpha$;
- 2 $\lambda_{\min}(\nabla^2 f(\mathbf{x})) \leq -\beta$;
- 3 There exists \mathbf{x}^* local minimum of f such that

$$\|\mathbf{x} - \mathbf{x}^*\| \leq \delta \quad \text{and} \quad \nabla^2 f(\mathbf{y}) \succeq \gamma I \succ 0 \quad \forall \mathbf{y}, \|\mathbf{y} - \mathbf{x}^*\| \leq 2\delta.$$

Definition (Ge et al '17; Goyens, R. '24)

$f : \mathbb{R}^n \rightarrow \mathbb{R} \mathcal{C}^2$ is $(\alpha, \beta, \gamma, \delta)$ -strict saddle with $\alpha, \beta, \gamma, \delta > 0$ if for any $\mathbf{x} \in \mathbb{R}^n$, one of these properties holds:

- 1 $\|\nabla f(\mathbf{x})\| \geq \alpha$;
- 2 $\lambda_{\min}(\nabla^2 f(\mathbf{x})) \leq -\beta$;
- 3 There exists \mathbf{x}^* local minimum of f such that

$$\|\mathbf{x} - \mathbf{x}^*\| \leq \delta \quad \text{and} \quad \nabla^2 f(\mathbf{y}) \succeq \gamma I \succ 0 \quad \forall \mathbf{y}, \|\mathbf{y} - \mathbf{x}^*\| \leq 2\delta.$$

Interpretation: 3 regions in the space

- 1 Large gradient.
- 2 Negative curvature for Hessian.
- 3 Near minimum+strong convexity.

Simple strict saddle functions

Strict saddle: $\{\|\nabla f(x)\| \geq \alpha\}$ or $\{\nabla^2 f(x) \preceq -\beta I\}$ or $\{\|x - x^*\| \leq \delta$ and $\nabla^2 f(y) \succeq \gamma I \forall y: \|y - x^*\| \leq 2\delta\}$.

Strongly convex \Rightarrow Strict saddle!

If f is γ -strongly convex, then it is $\left(\alpha, \beta, \gamma, \frac{2\alpha}{\gamma}\right)$ -strict saddle for any positive α and β .

Simple strict saddle functions

Strict saddle: $\{\|\nabla f(x)\| \geq \alpha\}$ or $\{\nabla^2 f(x) \preceq -\beta I\}$ or $\{\|x - x^*\| \leq \delta$ and $\nabla^2 f(y) \succeq \gamma I \forall y: \|y - x^*\| \leq 2\delta\}$.

Strongly convex \Rightarrow Strict saddle!

If f is γ -strongly convex, then it is $\left(\alpha, \beta, \gamma, \frac{2\alpha}{\gamma}\right)$ -strict saddle for any positive α and β .

Simple strict saddle functions

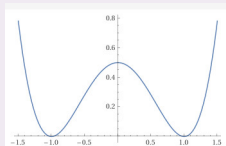
Strict saddle: $\{\|\nabla f(x)\| \geq \alpha\}$ or $\{\nabla^2 f(x) \preceq -\beta I\}$ or $\{\|x - x^*\| \leq \delta$ and $\nabla^2 f(y) \succeq \gamma I \forall y : \|y - x^*\| \leq 2\delta\}$.

Strongly convex \Rightarrow Strict saddle!

If f is γ -strongly convex, then it is $(\alpha, \beta, \gamma, \frac{2\alpha}{\gamma})$ -strict saddle for any positive α and β .

A “truly” strict saddle function

$$f : x \in \mathbb{R} \mapsto \frac{1}{2}(x^2 - 1)^2$$



Example: Matrix completion

$$\underset{\mathbf{U} \in \mathbb{R}^{n \times r}, \mathbf{V} \in \mathbb{R}^{m \times r}}{\text{minimize}} \quad f(\mathbf{U}, \mathbf{V}) := \|\mathcal{P}_\Omega(\mathbf{U}\mathbf{V}^\top - \mathbf{M})\|_F^2, \quad \Omega \subset [n] \times [m].$$

Assume Nice structure for \mathbf{M} (incoherence), probability of sampling \mathbf{M}_{ij} large enough.

Example: Matrix completion

$$\underset{\mathbf{U} \in \mathbb{R}^{n \times r}, \mathbf{V} \in \mathbb{R}^{m \times r}}{\text{minimize}} \quad f(\mathbf{U}, \mathbf{V}) := \|\mathcal{P}_\Omega(\mathbf{U}\mathbf{V}^T - \mathbf{M})\|_F^2, \quad \Omega \subset [n] \times [m].$$

Assume Nice structure for \mathbf{M} (incoherence), probability of sampling \mathbf{M}_{ij} large enough.

Theorem (Ge et al. '17)

Let $(\mathbf{U}, \mathbf{V}) \in \mathbb{R}^{n \times r} \times \mathbb{R}^{m \times r}$. Then, there exists $\alpha > 0$ such that one of these cases occur

- 1 $\|\nabla f(\mathbf{U}, \mathbf{V})\| \geq \alpha$
- 2 The Hessian at \mathbf{U}, \mathbf{V} has negative curvature, i.e.

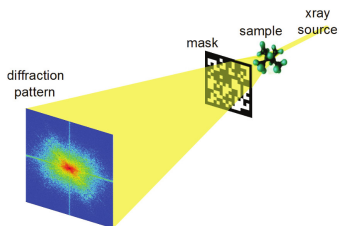
$$\lambda_{\min}(\nabla^2 f(\mathbf{U}, \mathbf{V})) < -\mathcal{O}(\sigma_{\min}(\mathbf{M}))$$

- 3 (\mathbf{U}, \mathbf{V}) is at distance at most $\mathcal{O}\left(\frac{\alpha}{\sigma_{\min}(\mathbf{M})}\right)$ from a global minimum.

Phase retrieval problem

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

- $b_i = |\langle \mathbf{x}^*, \mathbf{a}_i \rangle|$: Complex moduli.
- \mathbf{a}_i : Measurement vectors (Gaussian, wavelets, etc).

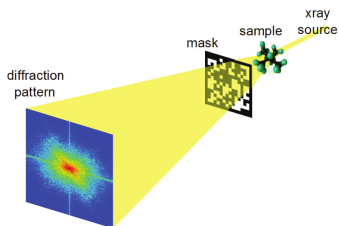


(Picture from Candès et al '15)

Phase retrieval problem

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

- $b_i = |\langle \mathbf{x}^*, \mathbf{a}_i \rangle|$: Complex moduli.
- \mathbf{a}_i : Measurement vectors (Gaussian, wavelets, etc).



(Picture from Candès et al '15)

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

Nonconvex vector formulation (Sun et al '18)

$$\underset{\mathbf{x} \in \mathbb{C}^n}{\text{minimize}} \quad f(\mathbf{x}) = \frac{1}{2m} \sum_{i=1}^m (b_i^2 - |\langle \mathbf{a}_i, \mathbf{x} \rangle|^2)^2$$

with $\{\mathbf{a}_i\}$ Gaussian, $m = \mathcal{O}(n \log^3(n))$.

Nonconvex vector formulation (Sun et al '18)

$$\underset{\mathbf{x} \in \mathbb{C}^n}{\text{minimize}} \quad f(\mathbf{x}) = \frac{1}{2m} \sum_{i=1}^m (b_i^2 - |\langle \mathbf{a}_i, \mathbf{x} \rangle|^2)^2$$

with $\{\mathbf{a}_i\}$ Gaussian, $m = \mathcal{O}(n \log^3(n))$.

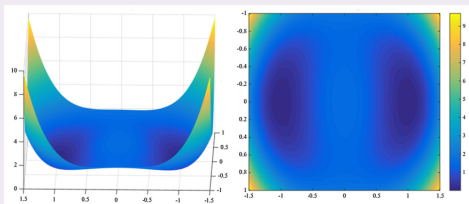
Nonconvex vector formulation (Sun et al '18)

$$\underset{\mathbf{x} \in \mathbb{C}^n}{\text{minimize}} \quad f(\mathbf{x}) = \frac{1}{2m} \sum_{i=1}^m (b_i^2 - |\langle \mathbf{a}_i, \mathbf{x} \rangle|^2)^2$$

with $\{\mathbf{a}_i\}$ Gaussian, $m = \mathcal{O}(n \log^3(n))$.

→ For some universal $c > 0$, w.h.p., f is $\left(\frac{c}{n \log(m)}, c, c, \frac{c}{n \log(m)}\right)$ -strict saddle.

→ Generalized definition of strict saddle (more later).



- 1 From nonconvex to strict saddle functions
- 2 Algorithms for strict saddle problems
- 3 Beyond strict saddle problems
- 4 Conclusion and perspectives

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} f(\mathbf{x})$$

$f \in \mathcal{C}^2$ bounded below and nonconvex.

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} f(\mathbf{x})$$

$f \in \mathcal{C}^2$ bounded below and nonconvex.

Goal: Reach (ϵ_g, ϵ_H) -point

$$\|\nabla f(\mathbf{x})\| \leq \epsilon_g \quad \text{and} \quad \nabla^2 f(\mathbf{x}) \succeq -\epsilon_H \mathbf{I}.$$

- General nonconvex case: Close to second-order stationary point.
- Strict saddle setting: Close to a global minimum!

- **Gradient descent** Can be shown to escape strict saddle points almost surely (Lee et al '19).
- **Perturbed gradient descent** does so in polynomial time (Jin et al '17, Ma et al '25).
- **Line-search algorithms** can also be analyzed in this way (O'Neill and Wright '23).

- **Gradient descent** Can be shown to escape strict saddle points almost surely (Lee et al '19).
- **Perturbed gradient descent** does so in polynomial time (Jin et al '17, Ma et al '25).
- **Line-search algorithms** can also be analyzed in this way (O'Neill and Wright '23).

Our approach

- Use **trust-region** algorithms, originally developed for nonconvex problems.
- Exploit their applicability to **manifold** optimization.

Main problem

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

- $f \in \mathcal{C}^2$ nonconvex
- \mathcal{M} Riemannian manifold ($\mathbb{C}^{n \times m}$, $\mathbb{R}^{n \times m}$, sphere, ...).

Main problem

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

- $f \in \mathcal{C}^2$ nonconvex
- \mathcal{M} Riemannian manifold ($\mathbb{C}^{n \times m}$, $\mathbb{R}^{n \times m}$, sphere, ...).

Goal: Approximate stationary points

$$\|g(\mathbf{x}_k)\| \leq \epsilon_g \quad \text{and} \quad \lambda_{\min}(\mathcal{H}(\mathbf{x}_k)) \geq -\epsilon_H.$$

- $g(\cdot)$ Riemannian gradient.
- $\mathcal{H}(\cdot)$ Riemannian Hessian.

Definition (Goyens, R. '24)

$f : \mathbb{R}^n \rightarrow \mathbb{R}$ is $(\alpha, \beta, \gamma, \delta)$ -strict saddle if for any $\mathbf{x} \in \mathbb{R}^n$, one of these properties holds:

- 1 $\|g(\mathbf{x})\| \geq \alpha$;
- 2 $\lambda_{\min}(\mathcal{H}(\mathbf{x})) \leq -\beta$;
- 3 There exists \mathbf{x}^* local minimum of f such that

$$\|\mathbf{x} - \mathbf{x}^*\| \leq \delta \quad \text{and} \quad \lambda_{\min}(\mathcal{H}(\mathbf{y})) \geq \gamma \quad \forall \mathbf{y}, \|\mathbf{y} - \mathbf{x}^*\| \leq 2\delta.$$

Definition (Goyens, R. '24)

$f : \mathbb{R}^n \rightarrow \mathbb{R}$ is $(\alpha, \beta, \gamma, \delta)$ -strict saddle if for any $\mathbf{x} \in \mathbb{R}^n$, one of these properties holds:

- 1 $\|g(\mathbf{x})\| \geq \alpha$;
- 2 $\lambda_{\min}(\mathcal{H}(\mathbf{x})) \leq -\beta$;
- 3 There exists \mathbf{x}^* local minimum of f such that

$$\|\mathbf{x} - \mathbf{x}^*\| \leq \delta \quad \text{and} \quad \lambda_{\min}(\mathcal{H}(\mathbf{y})) \geq \gamma \quad \forall \mathbf{y}, \|\mathbf{y} - \mathbf{x}^*\| \leq 2\delta.$$

Interpretation: 3 regions in the space

- 1 Large Riemannian gradient g .
- 2 Negative curvature for Riemannian Hessian \mathcal{H} .
- 3 Near minimum+geodesic strong convexity.

N.B. Already studied for special problem classes (Pumir et al '18, Sun et al '16-'18).

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

Inputs: $\mathbf{x}_0 \in \mathcal{M}$, $\Delta_0 > 0$, $\eta > 0$.

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

① Define $m_k(\mathbf{x}_k + \mathbf{s}) := \langle \mathbf{g}(\mathbf{x}_k), \mathbf{s} \rangle + \frac{1}{2} \langle \mathbf{s}, \mathcal{H}(\mathbf{x}_k) \mathbf{s} \rangle$ and compute

$$\mathbf{s}_k \in \underset{\substack{\mathbf{s} \in \mathcal{T}_{\mathbf{x}_k}^{\mathcal{M}} \\ \|\mathbf{s}\| \leq \Delta_k}}{\operatorname{argmin}} m_k(\mathbf{x}_k + \mathbf{s}).$$

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

Inputs: $\mathbf{x}_0 \in \mathcal{M}$, $\Delta_0 > 0$, $\eta > 0$.

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

- 1 Define $m_k(\mathbf{x}_k + \mathbf{s}) := \langle \mathbf{g}(\mathbf{x}_k), \mathbf{s} \rangle + \frac{1}{2} \langle \mathbf{s}, \mathcal{H}(\mathbf{x}_k) \mathbf{s} \rangle$ and compute

$$\mathbf{s}_k \in \underset{\substack{\mathbf{s} \in \mathcal{T}_{\mathbf{x}_k}^{\mathcal{M}} \\ \|\mathbf{s}\| \leq \Delta_k}}{\operatorname{argmin}} m_k(\mathbf{x}_k + \mathbf{s}).$$

- 2 Define $\mathbf{x}_k^{\mathcal{M}}$ as the retraction of $\mathbf{x}_k + \mathbf{s}_k$ onto \mathcal{M} .

Inputs: $\mathbf{x}_0 \in \mathcal{M}$, $\Delta_0 > 0$, $\eta > 0$.

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

- 1 Define $m_k(\mathbf{x}_k + \mathbf{s}) := \langle \mathbf{g}(\mathbf{x}_k), \mathbf{s} \rangle + \frac{1}{2} \langle \mathbf{s}, \mathcal{H}(\mathbf{x}_k) \mathbf{s} \rangle$ and compute

$$\mathbf{s}_k \in \operatorname{argmin} m_k(\mathbf{x}_k + \mathbf{s}).$$

$$\begin{aligned} \mathbf{s} &\in \mathcal{T}_{\mathbf{x}_k}^{\mathcal{M}} \\ \|\mathbf{s}\| &\leq \Delta_k \end{aligned}$$

- 2 Define $\mathbf{x}_k^{\mathcal{M}}$ as the retraction of $\mathbf{x}_k + \mathbf{s}_k$ onto \mathcal{M} .
- 3 Compute $\rho_k = \frac{f(\mathbf{x}_k) - f(\mathbf{x}_k^{\mathcal{M}})}{m_k(\mathbf{x}_k) - m_k(\mathbf{x}_k^{\mathcal{M}})}$.

Inputs: $\mathbf{x}_0 \in \mathcal{M}$, $\Delta_0 > 0$, $\eta > 0$.

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

- 1 Define $m_k(\mathbf{x}_k + \mathbf{s}) := \langle \mathbf{g}(\mathbf{x}_k), \mathbf{s} \rangle + \frac{1}{2} \langle \mathbf{s}, \mathcal{H}(\mathbf{x}_k) \mathbf{s} \rangle$ and compute

$$\mathbf{s}_k \in \underset{\substack{\mathbf{s} \in \mathcal{T}_{\mathbf{x}_k}^{\mathcal{M}} \\ \|\mathbf{s}\| \leq \Delta_k}}{\text{argmin}} m_k(\mathbf{x}_k + \mathbf{s}).$$

- 2 Define $\mathbf{x}_k^{\mathcal{M}}$ as the retraction of $\mathbf{x}_k + \mathbf{s}_k$ onto \mathcal{M} .
- 3 Compute $\rho_k = \frac{f(\mathbf{x}_k) - f(\mathbf{x}_k^{\mathcal{M}})}{m_k(\mathbf{x}_k) - m_k(\mathbf{x}_k^{\mathcal{M}})}$.
- 4 If $\rho_k \geq \eta$, set $\mathbf{x}_{k+1} = \mathbf{x}_k^{\mathcal{M}}$ and $\Delta_{k+1} = 2\Delta_k$.

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

Inputs: $\mathbf{x}_0 \in \mathcal{M}$, $\Delta_0 > 0$, $\eta > 0$.

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

- 1 Define $m_k(\mathbf{x}_k + \mathbf{s}) := \langle \mathbf{g}(\mathbf{x}_k), \mathbf{s} \rangle + \frac{1}{2} \langle \mathbf{s}, \mathcal{H}(\mathbf{x}_k) \mathbf{s} \rangle$ and compute

$$\mathbf{s}_k \in \underset{\substack{\mathbf{s} \in \mathcal{T}_{\mathbf{x}_k}^{\mathcal{M}} \\ \|\mathbf{s}\| \leq \Delta_k}}{\text{argmin}} m_k(\mathbf{x}_k + \mathbf{s}).$$

- 2 Define $\mathbf{x}_k^{\mathcal{M}}$ as the retraction of $\mathbf{x}_k + \mathbf{s}_k$ onto \mathcal{M} .
- 3 Compute $\rho_k = \frac{f(\mathbf{x}_k) - f(\mathbf{x}_k^{\mathcal{M}})}{m_k(\mathbf{x}_k) - m_k(\mathbf{x}_k^{\mathcal{M}})}$.
- 4 If $\rho_k \geq \eta$, set $\mathbf{x}_{k+1} = \mathbf{x}_k^{\mathcal{M}}$ and $\Delta_{k+1} = 2\Delta_k$.
- 5 Otherwise, set $\mathbf{x}_{k+1} = \mathbf{x}_k$ and $\Delta_{k+1} = 0.5\Delta_k$.

Complexity results for trust region

Goal: Compute \mathbf{x}_K such that $\|g(\mathbf{x}_K)\| \leq \epsilon_g$ and $\lambda_{\min}(\mathcal{H}(\mathbf{x}_K)) \geq -\epsilon_H$.

For general nonconvex f (Boumal et al '19)

$$K = \mathcal{O}(\max\{\epsilon_g^{-2}, \epsilon_H^{-3}\}).$$

Complexity results for trust region

Goal: Compute \mathbf{x}_K such that $\|g(\mathbf{x}_K)\| \leq \epsilon_g$ and $\lambda_{\min}(\mathcal{H}(\mathbf{x}_K)) \geq -\epsilon_H$.

For general nonconvex f (Boumal et al '19)

$$K = \mathcal{O}(\max\{\epsilon_g^{-2}, \epsilon_H^{-3}\}).$$

For strict saddle f (Goyens and R. '24)

If f is $(\alpha, \beta, \gamma, \delta)$ -strict saddle, $K = K_f + K_\epsilon$, with

$$K_f = \mathcal{O}(\max\{\alpha^{-2}\beta^{-1}, \alpha^{-2}\gamma^{-1}, \beta^{-3}, \gamma^{-3}, \gamma^{-2}\delta^{-1}\})$$

$$K_\epsilon = \log \log [\mathcal{O}(\gamma\epsilon_g^{-1})].$$

Complexity results for trust region

Goal: Compute \mathbf{x}_K such that $\|g(\mathbf{x}_K)\| \leq \epsilon_g$ and $\lambda_{\min}(\mathcal{H}(\mathbf{x}_K)) \geq -\epsilon_H$.

For general nonconvex f (Boumal et al '19)

$$K = \mathcal{O}(\max\{\epsilon_g^{-2}, \epsilon_H^{-3}\}).$$

For strict saddle f (Goyens and R. '24)

If f is $(\alpha, \beta, \gamma, \delta)$ -strict saddle, $K = K_f + K_\epsilon$, with

$$K_f = \mathcal{O}(\max\{\alpha^{-2}\beta^{-1}, \alpha^{-2}\gamma^{-1}, \beta^{-3}, \gamma^{-3}, \gamma^{-2}\delta^{-1}\})$$

$$K_\epsilon = \log \log [\mathcal{O}(\gamma\epsilon_g^{-1})].$$

- log log dependency in ϵ_g (none on ϵ_H)!
- Complexity depends more on **landscape** parameters!

What we have so far

- Trust-region method with good complexity.
- Requires exact step computation.
- Agnostic to the function landscape.

What we have so far

- Trust-region method with good complexity.
- Requires exact step computation.
- Agnostic to the function landscape.

Moving further with inexactness

- Practical trust-region use inexact steps.
- Never store the Hessian matrix (“matrix-free”)!

Can we build an inexact variant that leverages the strict saddle property?

Trust-region subproblem

$$\underset{\mathbf{s} \in \mathcal{T}_{\mathbf{x}_k}^M}{\text{minimize}} \quad \langle \mathbf{g}(\mathbf{x}_k), \mathbf{s} \rangle + \frac{1}{2} \langle \mathbf{s}, \mathcal{H}(\mathbf{x}_k) \mathbf{s} \rangle \quad \text{s.t.} \quad \|\mathbf{s}\| \leq \Delta_k.$$

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

Trust-region subproblem

$$\underset{\mathbf{s} \in \mathcal{T}_{\mathbf{x}_k}^M}{\text{minimize}} \quad \langle \mathbf{g}(\mathbf{x}_k), \mathbf{s} \rangle + \frac{1}{2} \langle \mathbf{s}, \mathcal{H}(\mathbf{x}_k) \mathbf{s} \rangle \quad \text{s.t.} \quad \|\mathbf{s}\| \leq \Delta_k.$$

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

Changes (for complexity)

- Add a cap on the number of CG iterations.
- Guarantee negative curvature detection.

$f : \mathbb{R}^n \rightarrow \mathbb{R}$ is $(\alpha, \beta, \gamma, \delta)$ -strict saddle if for any $\mathbf{x} \in \mathbb{R}^n$, one of these properties holds:

- 1 $\|g(\mathbf{x})\| \geq \alpha$;
- 2 $\lambda_{\min}(\mathcal{H}(\mathbf{x})) \leq -\beta$;
- 3 There exists \mathbf{x}^* local minimum of f such that

$$\|\mathbf{x} - \mathbf{x}^*\| \leq \delta \quad \text{and} \quad \lambda_{\min}(\mathcal{H}(\mathbf{y})) \geq \gamma \quad \forall \mathbf{y}, \|\mathbf{y} - \mathbf{x}^*\| \leq 2\delta.$$

$f : \mathbb{R}^n \rightarrow \mathbb{R}$ is $(\alpha, \beta, \gamma, \delta)$ -strict saddle if for any $\mathbf{x} \in \mathbb{R}^n$, one of these properties holds:

- 1 $\|g(\mathbf{x})\| \geq \alpha$;
- 2 $\lambda_{\min}(\mathcal{H}(\mathbf{x})) \leq -\beta$;
- 3 There exists \mathbf{x}^* local minimum of f such that

$$\|\mathbf{x} - \mathbf{x}^*\| \leq \delta \quad \text{and} \quad \lambda_{\min}(\mathcal{H}(\mathbf{y})) \geq \gamma \quad \forall \mathbf{y}, \|\mathbf{y} - \mathbf{x}^*\| \leq 2\delta.$$

One step per strict saddle case

- 1 $\|g(\mathbf{x}_k)\| \geq \alpha$: Cheap gradient step (along $-g(\mathbf{x}_k)$).

$f : \mathbb{R}^n \rightarrow \mathbb{R}$ is $(\alpha, \beta, \gamma, \delta)$ -strict saddle if for any $\mathbf{x} \in \mathbb{R}^n$, one of these properties holds:

- 1 $\|g(\mathbf{x})\| \geq \alpha$;
- 2 $\lambda_{\min}(\mathcal{H}(\mathbf{x})) \leq -\beta$;
- 3 There exists \mathbf{x}^* local minimum of f such that

$$\|\mathbf{x} - \mathbf{x}^*\| \leq \delta \quad \text{and} \quad \lambda_{\min}(\mathcal{H}(\mathbf{y})) \geq \gamma \quad \forall \mathbf{y}, \|\mathbf{y} - \mathbf{x}^*\| \leq 2\delta.$$

One step per strict saddle case

- 1 $\|g(\mathbf{x}_k)\| \geq \alpha$: Cheap gradient step (along $-g(\mathbf{x}_k)$).
- 2 Otherwise try the last case (CG should work fast!)

$f : \mathbb{R}^n \rightarrow \mathbb{R}$ is $(\alpha, \beta, \gamma, \delta)$ -strict saddle if for any $\mathbf{x} \in \mathbb{R}^n$, one of these properties holds:

- 1 $\|g(\mathbf{x})\| \geq \alpha$;
- 2 $\lambda_{\min}(\mathcal{H}(\mathbf{x})) \leq -\beta$;
- 3 There exists \mathbf{x}^* local minimum of f such that

$$\|\mathbf{x} - \mathbf{x}^*\| \leq \delta \quad \text{and} \quad \lambda_{\min}(\mathcal{H}(\mathbf{y})) \geq \gamma \quad \forall \mathbf{y}, \|\mathbf{y} - \mathbf{x}^*\| \leq 2\delta.$$

One step per strict saddle case

- 1 $\|g(\mathbf{x}_k)\| \geq \alpha$: Cheap gradient step (along $-g(\mathbf{x}_k)$).
- 2 Otherwise try the last case (CG should work fast!)
- 3 If CG does not work fast, then $\lambda_{\min}(\mathcal{H}(\mathbf{x}_k)) \leq -\beta$: Negative curvature step.

Our method: Capped conjugate gradient

Goal:

$$\min_{\mathbf{s} \in \mathcal{T}_{\mathbf{x}_k}^{\mathcal{M}}} \langle \mathbf{g}(\mathbf{x}_k), \mathbf{s} \rangle + \frac{1}{2} \langle \mathbf{s}, (\mathcal{H}(\mathbf{x}_k) + 2\gamma I) \mathbf{s} \rangle \quad \text{s.t.} \quad \|\mathbf{s}\| \leq \Delta.$$

Theorem (Curtis, Robinson, R., Wright '21)

Suppose that we run CG for at most $J^{CG} = \min\{n, \tilde{\mathcal{O}}(\gamma^{-1/2})\}$ iterations/Hessian-vector products. Then,

- Either we compute a good enough step using CG...
- ...or we find a negative curvature direction for H ...
- ...or we know that it exists and we can call a **minimum eigenvalue oracle** to find it.

Our method: Capped conjugate gradient

Goal:

$$\min_{\mathbf{s} \in \mathcal{T}_{\mathbf{x}_k}^{\mathcal{M}}} \langle \mathbf{g}(\mathbf{x}_k), \mathbf{s} \rangle + \frac{1}{2} \langle \mathbf{s}, (\mathcal{H}(\mathbf{x}_k) + 2\gamma I) \mathbf{s} \rangle \quad \text{s.t.} \quad \|\mathbf{s}\| \leq \Delta.$$

Theorem (Curtis, Robinson, R., Wright '21)

Suppose that we run CG for at most $J^{\text{CG}} = \min\{n, \tilde{\mathcal{O}}(\gamma^{-1/2})\}$ iterations/Hessian-vector products. Then,

- Either we compute a good enough step using CG...
- ...or we find a negative curvature direction for H ...
- ...or we know that it exists and we can call a **minimum eigenvalue oracle** to find it.

Strict saddle setting

Suppose that $\|\mathbf{g}(\mathbf{x}_k)\| \leq \alpha$ and run CG for J^{CG} iterations. Then,

- Either the step is accurate enough
- or **we know that** $\lambda_{\min}(\mathcal{H}(\mathbf{x}_k)) \leq -\beta I$ and we call a **minimum eigenvalue oracle** to find negative curvature.

Given $\mathcal{H}(\mathbf{x}_k) \in \mathbb{R}^{n \times n}$, $\beta \in (0, 1)$, and $\xi \in (0, 1)$, output

- 1 A vector \mathbf{s} such that

$$\mathbf{s}^T \mathcal{H}(\mathbf{x}_k) \mathbf{s} \leq -\frac{\beta}{2} \|\mathbf{s}\|^2.$$

- 2 **OR** a certificate that $\mathcal{H}(\mathbf{x}_k) \succ -\beta I$, valid with probability $1 - \xi$.

Minimum eigenvalue oracle (MEO)

Given $\mathcal{H}(\mathbf{x}_k) \in \mathbb{R}^{n \times n}$, $\beta \in (0, 1)$, and $\xi \in (0, 1)$, output

- 1 A vector \mathbf{s} such that

$$\mathbf{s}^T \mathcal{H}(\mathbf{x}_k) \mathbf{s} \leq -\frac{\beta}{2} \|\mathbf{s}\|^2.$$

- 2 **OR** a certificate that $\mathcal{H}(\mathbf{x}_k) \succ -\beta I$, valid with probability $1 - \xi$.

An example of MEO

Run **CG** on $\mathcal{H}(\mathbf{x}_k) \mathbf{s} = \mathbf{b}$, \mathbf{b} uniform on the unit sphere. produces output in $J^{MEO} = \min\{n, \tilde{O}(\beta^{-1/2})\}$ iterations/Hessian-vector products!

Landscape-aware algorithm for minimize $\mathbf{x} \in \mathcal{M} f(\mathbf{x})$

Inputs: $\mathbf{x}_0 \in \mathcal{M}$, $\Delta_0 > 0$, $\eta > 0$, α, β, γ .

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

- 1 If $\|g(\mathbf{x}_k)\| \geq \alpha$, take a (cheap) gradient-type step:

$$\mathbf{s}_k = \underset{\substack{\mathbf{s} \in \mathcal{T}_{\mathbf{x}_k}^{\mathcal{M}} \\ \|\mathbf{s}\| \leq \Delta_k}}{\operatorname{argmin}} \langle g(\mathbf{x}_k), \mathbf{s} \rangle .$$

- 2 Otherwise, apply truncated CG to

$$\underset{\substack{\mathbf{s} \in \mathcal{T}_{\mathbf{x}_k}^{\mathcal{M}} \\ \|\mathbf{s}\| \leq \Delta_k}}{\operatorname{minimize}} \langle g(\mathbf{x}_k), \mathbf{s} \rangle + \frac{1}{2} \langle \mathbf{s}, \mathcal{H}(\mathbf{x}_k) \mathbf{s} \rangle + 2\gamma \|\mathbf{s}\|$$

If it terminates with \mathbf{s} , use $\mathbf{s}_k = \mathbf{s}$.

- 3 Otherwise, call MEO to find \mathbf{s}_k such that $\langle \mathbf{s}, \mathcal{H}(\mathbf{x}_k) \mathbf{s} \rangle \leq -\beta/2 \|\mathbf{s}\|^2$.
- 4 Define $\mathbf{x}_k^{\mathcal{M}}$ as the retraction of $\mathbf{x}_k + \mathbf{s}_k$ onto \mathcal{M} .
- 5 Update \mathbf{x}_{k+1} and Δ_{k+1} as before.

Goal: Compute \mathbf{x}_K such that $\|g(\mathbf{x}_K)\| \leq \epsilon_g$ and $\mathcal{H}(\mathbf{x}_K) \succeq -\epsilon_H I$.

Operation complexity (Goyens and R., '24)

The method reaches an (ϵ_g, ϵ_H) -point in

$$N_{\epsilon_g} = \tilde{\mathcal{O}} \left(\min \left\{ n, \max \{ \beta^{-1/2}, \gamma^{-1/2} \} \right\} \right) \\ \times \left(\max \{ \alpha^{-2} \beta^{-1}, \alpha^{-2} \gamma^{-1}, \beta^{-3}, \gamma^{-3}, \gamma^{-2} \delta^{-1} \} + \log \log [\mathcal{O}(\gamma \epsilon_g^{-1})] \right)$$

gradient/Hessian-vector products with probability $(1 - \xi)^{N_{\epsilon_g}}$.

Goal: Compute \mathbf{x}_K such that $\|g(\mathbf{x}_K)\| \leq \epsilon_g$ and $\mathcal{H}(\mathbf{x}_K) \succeq -\epsilon_H I$.

Operation complexity (Goyens and R., '24)

The method reaches an (ϵ_g, ϵ_H) -point in

$$N_{\epsilon_g} = \tilde{\mathcal{O}} \left(\min \left\{ n, \max \{ \beta^{-1/2}, \gamma^{-1/2} \} \right\} \right) \\ \times \left(\max \{ \alpha^{-2} \beta^{-1}, \alpha^{-2} \gamma^{-1}, \beta^{-3}, \gamma^{-3}, \gamma^{-2} \delta^{-1} \} + \log \log [\mathcal{O}(\gamma \epsilon_g^{-1})] \right)$$

gradient/Hessian-vector products with probability $(1 - \xi)^{N_{\epsilon_g}}$.

- Probability holds for second-order guarantee;
- **Per-iteration cost** does not depend on ϵ_g !

Phase retrieval (Sun et al '18)

$$\underset{\mathbf{x} \in \mathbb{C}^n}{\text{minimize}} \frac{1}{2m} \sum_{i=1}^m (b_i^2 - |\langle \mathbf{a}_i, \mathbf{x} \rangle|^2)^2.$$

If $\{\mathbf{a}_i\}$ are Gaussian and $m = \mathcal{O}(n \log^3(n))$, the objective is $(\frac{c}{n \log(m)}, c, c, \frac{c}{n \log(m)})$ -strict saddle for some absolute constant $c > 0$ w.h.p.

Phase retrieval (Sun et al '18)

$$\underset{\mathbf{x} \in \mathbb{C}^n}{\text{minimize}} \frac{1}{2m} \sum_{i=1}^m (b_i^2 - |\langle \mathbf{a}_i, \mathbf{x} \rangle|^2)^2.$$

If $\{\mathbf{a}_i\}$ are Gaussian and $m = \mathcal{O}(n \log^3(n))$, the objective is $(\frac{c}{n \log(m)}, c, c, \frac{c}{n \log(m)})$ -strict saddle for some absolute constant $c > 0$ w.h.p.

Complexity results

- Basic trust region:
 - $\mathcal{O}(\epsilon_g^{-3/2})$ iterations
 - $\tilde{\mathcal{O}}(\epsilon_g^{-7/4})$ operations.
- Using strict saddle:
 - $\mathcal{O}(n^2) + \log \log(\mathcal{O}(\epsilon_g^{-1}))$ iterations.
 - $\tilde{\mathcal{O}}(n^{5/2}) + \tilde{\mathcal{O}}(n^{1/2}) \log \log(\mathcal{O}(\epsilon_g^{-1}))$ operations.

- 1 From nonconvex to strict saddle functions
- 2 Algorithms for strict saddle problems
- 3 Beyond strict saddle problems**
- 4 Conclusion and perspectives

So far

- Good results for trust region on strict saddle functions.
- Example: Strict saddle phase retrieval.

From strict saddle to good nonconvex?

So far

- Good results for trust region on strict saddle functions.
- Example: Strict saddle phase retrieval.

What's wrong?

- Not all nonconvex problems are strict saddle.
- Yet trust region can work well.

From strict saddle to good nonconvex?

So far

- Good results for trust region on strict saddle functions.
- Example: Strict saddle phase retrieval.

What's wrong?

- Not all nonconvex problems are strict saddle.
- Yet trust region can work well.

The case of phase retrieval

- Popular formulations: Factored SDPs.
- **Goal:** Study trust region on those!

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

PhaseLift SDP relaxation (Candès et al. '11)

$$\begin{aligned} \min_{\mathbf{X} \in \mathbb{C}^{n \times n}} \quad & \text{trace}(\mathbf{X}) \\ \text{subject to} \quad & \langle \mathbf{X}, \mathbf{a}_i \mathbf{a}_i^* \rangle = b_i^2 \quad \forall i \\ & \mathbf{X} \succeq 0 \end{aligned}$$

- Convex problem, can be solved in polynomial time.
- In practice, cost of matrix linear algebra prohibitive.
- Expected: **Rank-1 solution**.

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

PhaseLift SDP relaxation (Candès et al. '11)

$$\begin{aligned} \min_{\mathbf{X} \in \mathbb{C}^{n \times n}} \quad & \text{trace}(\mathbf{X}) \\ \text{subject to} \quad & \langle \mathbf{X}, \mathbf{a}_i \mathbf{a}_i^* \rangle = b_i^2 \quad \forall i \\ & \mathbf{X} \succeq 0 \end{aligned}$$

- Convex problem, can be solved in polynomial time.
- In practice, cost of matrix linear algebra prohibitive.
- Expected: **Rank-1 solution**.

Factored PhaseLift formulation

$$\min_{\mathbf{U} \in \mathbb{C}^{n \times r}} f(\mathbf{U}) := \frac{1}{2} \|\mathcal{A}(\mathbf{U} \mathbf{U}^T) - \mathbf{b}\|_F^2, \quad \mathcal{A}(\mathbf{X}) = [\langle \mathbf{X}, \mathbf{a}_i \mathbf{a}_i^* \rangle]_i.$$

- Burer-Monteiro factorization (Burer and Monteiro '03).
- Benign landscape for relatively small r (McRae '24).
- But not strict saddle for small r !

Simplified PhaseLift instance $\min_{\mathbf{U} \in \mathbb{C}^{n \times r}} f(\mathbf{U}) := \frac{1}{2} \|\mathbf{U} \mathbf{U}^T - \mathbf{b}\|_F^2.$

Simplified PhaseLift instance $\min_{\mathbf{U} \in \mathbb{C}^{n \times r}} f(\mathbf{U}) := \frac{1}{2} \|\mathbf{U} \mathbf{U}^T - \mathbf{b}\|_F^2$.

Theorem (Goyens, R., Waldspurger '25+)

If $r = 2$, \mathbf{U}_0 close enough to a solution, then

$$f(\mathbf{U}_K) - \min_{\mathbf{U}} f(\mathbf{U}) \leq \epsilon \left(f(\mathbf{U}_0) - \min_{\mathbf{U}} f(\mathbf{U}) \right)$$

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

Simplified PhaseLift instance $\min_{\mathbf{U} \in \mathbb{C}^{n \times r}} f(\mathbf{U}) := \frac{1}{2} \|\mathbf{U} \mathbf{U}^T - \mathbf{b}\|_F^2$.

Theorem (Goyens, R., Waldspurger '25+)

If $r = 2$, \mathbf{U}_0 close enough to a solution, then

$$f(\mathbf{U}_K) - \min_{\mathbf{U}} f(\mathbf{U}) \leq \epsilon \left(f(\mathbf{U}_0) - \min_{\mathbf{U}} f(\mathbf{U}) \right)$$

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

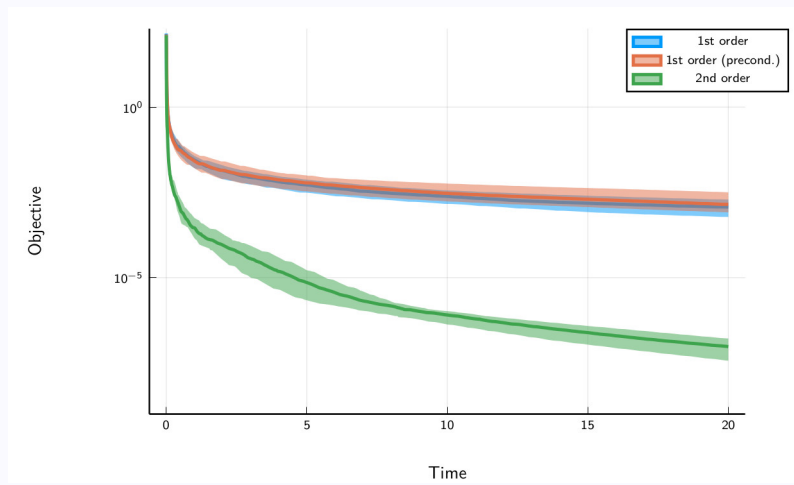
- Worse than the $\log(\log(1/\epsilon))$ strict saddle rate.
- Local result but no need for PL/ strong convexity near solution.
- On par with other analyses (Rebjoek and Boumal '24-'25)

Ongoing: Extension to more PhaseLift problems, in line with numerics.

Numerics on Burer-Monteiro PhaseLift (using Julia)

Setup:

- $n = 128$, $m = 768$, wavelet measurements, $r = 2$.
- Gradient descent, Preconditioned (Zhang, Fattahi, Zhang '21), Trust region.



Another Phase retrieval SDP

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

PhaseCut SDP relaxation (Waldspurger et al. '15)

$$\begin{array}{ll} \min_{\mathbf{Y} \in \mathbb{C}^{m \times m}} & \langle \mathbf{Y}, \mathbf{M} \rangle \\ \text{subject to} & Y_{ii} = 1 \quad \forall i \\ & \mathbf{Y} \succeq 0. \end{array} \quad \text{where} \quad \left\{ \begin{array}{l} \mathbf{M} = \text{diag}(\mathbf{b})(\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\text{diag}(\mathbf{b}) \\ \mathbf{A} = \begin{bmatrix} \mathbf{a}_1^* \\ \vdots \\ \mathbf{a}_m^* \end{bmatrix}. \end{array} \right.$$

Another Phase retrieval SDP

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

PhaseCut SDP relaxation (Waldspurger et al. '15)

$$\begin{array}{ll} \min_{\mathbf{Y} \in \mathbb{C}^{m \times m}} & \langle \mathbf{Y}, \mathbf{M} \rangle \\ \text{subject to} & Y_{ii} = 1 \quad \forall i \\ & \mathbf{Y} \succeq 0. \end{array} \quad \text{where} \quad \left\{ \begin{array}{l} \mathbf{M} = \text{diag}(\mathbf{b})(\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\text{diag}(\mathbf{b}) \\ \mathbf{A} = \begin{bmatrix} \mathbf{a}_1^* \\ \vdots \\ \mathbf{a}_m^* \end{bmatrix}. \end{array} \right.$$

Burer-Monteiro factorization (rank $r \leq m$)

$$\begin{array}{ll} \min_{\mathbf{U} \in \mathbb{C}^{m \times r}} & \langle \mathbf{U}, \mathbf{MU} \rangle \\ \text{subject to} & \|\mathbf{U}_{i,:}\| = 1 \quad \forall i. \end{array}$$

Goal: Solve the problem using $r \ll m$.

Is PhaseCut strict saddle?

Strict saddle: $\{\|\nabla f(x)\| \geq \alpha\}$ or $\{\nabla^2 f(x) \preceq -\beta I\}$ or $\{\|x - x^*\| \leq \delta \text{ and } \nabla^2 f(y) \succeq \gamma I \forall y : \|y - x^*\| \leq 2\delta\}$.

Burer-Monteiro PhaseCut

$$\min_{U \in \mathbb{C}^{m \times r}} \langle U, MU \rangle \quad \text{subject to} \quad \|U_{i,:}\| = 1 \quad \forall i = 1, \dots, m.$$

Is PhaseCut strict saddle?

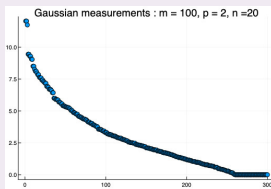
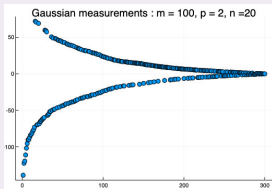
Strict saddle: $\{\|\nabla f(x)\| \geq \alpha\}$ or $\{\nabla^2 f(x) \not\geq -\beta I\}$ or $\{\|x - x^*\| \leq \delta \text{ and } \nabla^2 f(y) \geq \gamma I \forall y : \|y - x^*\| \leq 2\delta\}$.

Burer-Monteiro PhaseCut

$$\min_{U \in \mathbb{C}^{m \times r}} \langle U, MU \rangle \quad \text{subject to} \quad \|U_{i,:}\| = 1 \quad \forall i = 1, \dots, m.$$

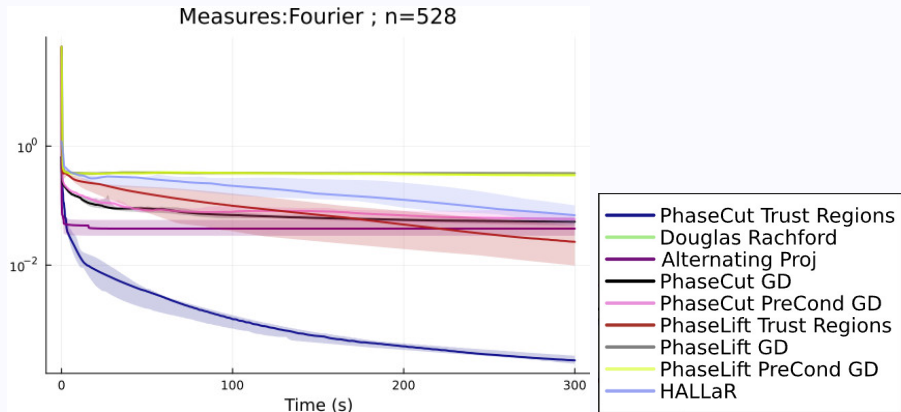
PhaseCut not strict saddle

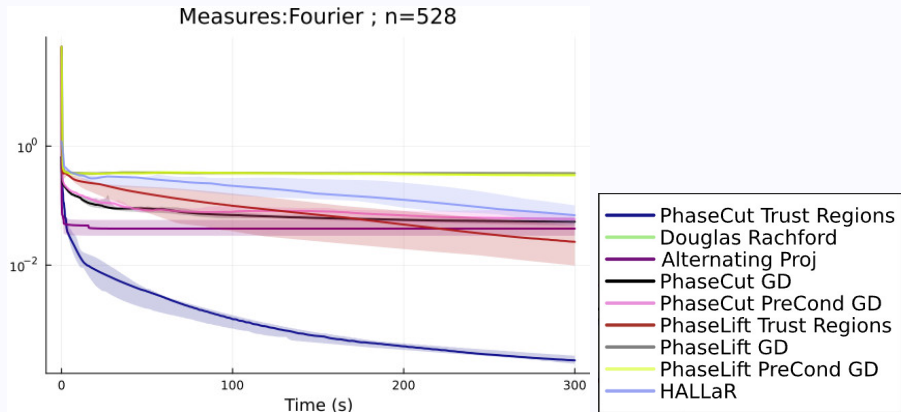
- Only solutions do not have negative curvature.
- The function is not strongly convex around minima!



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

More numerics (ft. Paul Caucheteux)





- PhaseCut + Trust Region often yields best performance.
- Proving this remains difficult!

- 1 From nonconvex to strict saddle functions
- 2 Algorithms for strict saddle problems
- 3 Beyond strict saddle problems
- 4 Conclusion and perspectives

Nonconvex problems can be (relatively) easy!

- Instances in data science have favorable landscape.
- No bad local minima, non-minima points easy to escape from.

Nonconvex problems can be (relatively) easy!

- Instances in data science have favorable landscape.
- No bad local minima, non-minima points easy to escape from.

Strict saddle functions are nice

- Guaranteed convergence to local minima.
- Faster convergence rates for trust region.

Nonconvex problems can be (relatively) easy!

- Instances in data science have favorable landscape.
- No bad local minima, non-minima points easy to escape from.

Strict saddle functions are nice

- Guaranteed convergence to local minima.
- Faster convergence rates for trust region.

Nice functions exist beyond strict saddle

- Trust region efficient on Burer-Monteiro SDPs.
- Numerical and theoretical arguments.

- E. M. Achour, F. Malgouyres and S. Gerchinovitz, *The loss landscape of deep linear neural networks: a second-order analysis*, Journal of Machine Learning Research, 2024.
- N. Boumal, P.-A. Absil and C. Cartis, *Global rates of convergence for nonconvex optimization on manifolds*, IMA Journal of Numerical Analysis, 2018.
- S. Burer and R. D. C. Monteiro, *A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization*, Mathematical Programming, 2003.
- C. Cartis, N. I. M. Gould and Ph. L. Toint, *Evaluation complexity of algorithms for nonconvex optimization: Theory, computation and perspectives*, SIAM, 2022.
- P. Caucheteux, F. Goyens, C. W. Royer and I. Waldspurger, *Trust-region methods for Burer-Monteiro factorizations of phase retrieval problems*, working paper, 2025.
- Y. Chi, Y. M. Lu and Y. Chen, *Nonconvex optimization meets low-rank matrix factorization: An overview*, IEEE Transactions on Signal Processing, 2019.
- F. E. Curtis, D. P. Robinson, C. W. Royer and S. J. Wright. *Trust-region Newton-CG with strong second-order complexity guarantees for nonconvex optimization*, SIAM Journal on Optimization, 2021.
- R. Ge, C. Jin and Y. Zheng, *No spurious local minima in nonconvex low rank problems: A unified geometric analysis*, International Conference on Machine Learning, 2017.
- F. Goyens and C. W. Royer, *Riemannian trust-region methods for strict saddle functions with complexity guarantees*, Mathematical Programming, 2024.
- C. Jin, P. Netrapalli, R. Ge, S. M. Kakade and M. I. Jordan, *How to escape saddle points efficiently*, International Conference on Machine Learning, 2017.
- J. D. Lee, I. Panageas, G. Piliouras, M. Simchowitz, M. I. Jordan and B. Recht, *First-order methods almost always avoid strict saddle points*, Mathematical Programming, 2019.
- J. Ma, G. Liang and S. Fattahi, *Implicit regularization of infinitesimally-perturbed gradient descent toward low-dimensional solutions*, arXiv:2505.17304, 2025.
- A. D. McRae, *Phase retrieval and matrix sensing via benign and overparametrized nonconvex optimization*, arXiv:2505.02636, 2025.
- Q. Rebjock and N. Boumal, *Fast convergence to non-isolated minima: four equivalent conditions for C2 functions*, Mathematical Programming, 2025.
- J. Wright and Y. Ma, *High-Dimensional Data Analysis with Low-Dimensional Models*, 2022.

- E. M. Achour, F. Malgouyres and S. Gerchinovitz, *The loss landscape of deep linear neural networks: a second-order analysis*, Journal of Machine Learning Research, 2024.
- N. Boumal, P.-A. Absil and C. Cartis, *Global rates of convergence for nonconvex optimization on manifolds*, IMA Journal of Numerical Analysis, 2018.
- S. Burer and R. D. C. Monteiro, *A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization*, Mathematical Programming, 2003.
- C. Cartis, N. I. M. Gould and Ph. L. Toint, *Evaluation complexity of algorithms for nonconvex optimization: Theory, computation and perspectives*, SIAM, 2022.
- P. Caucheteux, F. Goyens, C. W. Royer and I. Waldspurger, *Trust-region methods for Burer-Monteiro factorizations of phase retrieval problems*, working paper, 2025.
- Y. Chi, Y. M. Lu and Y. Chen, *Nonconvex optimization meets low-rank matrix factorization: An overview*, IEEE Transactions on Signal Processing, 2019.
- F. E. Curtis, D. P. Robinson, C. W. Royer and S. J. Wright, *Trust-region Newton-CG with strong second-order complexity guarantees for nonconvex optimization*, SIAM Journal on Optimization, 2021.
- R. Ge, C. Jin and Y. Zheng, *No spurious local minima in nonconvex low rank problems: A unified geometric analysis*, International Conference on Machine Learning, 2017.
- F. Goyens and C. W. Royer, *Riemannian trust-region methods for strict saddle functions with complexity guarantees*, Mathematical Programming, 2024.
- C. Jin, P. Netrapalli, R. Ge, S. M. Kakade and M. I. Jordan, *How to escape saddle points efficiently*, International Conference on Machine Learning, 2017.
- J. D. Lee, I. Panageas, G. Piliouras, M. Simchowitz, M. I. Jordan and B. Recht, *First-order methods almost always avoid strict saddle points*, Mathematical Programming, 2019.
- J. Ma, G. Liang and S. Fattahi, *Implicit regularization of infinitesimally-perturbed gradient descent toward low-dimensional solutions*, arXiv:2505.17304, 2025.
- A. D. McRae, *Phase retrieval and matrix sensing via benign and overparametrized nonconvex optimization*, arXiv:2505.02636, 2025.
- Q. Rebjock and N. Boumal, *Fast convergence to non-isolated minima: four equivalent conditions for C2 functions*, Mathematical Programming, 2025.
- J. Wright and Y. Ma, *High-Dimensional Data Analysis with Low-Dimensional Models*, 2022.

Merci !

clement.royer@lamsade.dauphine.fr