Revisiting linear programming techniques for nonconvex phase retrieval and unbalanced optimal transport

Clément Royer

MAP5, 3 avril 2026

Collaborators



- Paul Caucheteux (ENSAE)
- Florentin Goyens (UCLouvain)
- Irène Waldspurger (CNRS, Dauphine)


- Gaetano Agazzotti (Dauphine)
- Antonin Chambolle (CNRS, Dauphine)

A Friday afternoon topic

Linear programming (LP)

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \ \boldsymbol{c}^{\mathrm{T}} \boldsymbol{x} \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geq 0.$$

- Most classical optimization problem.
- Numerous applications: OR, Finance, etc.
- Mature general-purpose solvers: Gurobi, CPLEX, HiGHS.

A Friday afternoon topic

Linear programming (LP)

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \ \boldsymbol{c}^{\mathrm{T}} \boldsymbol{x} \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geq 0.$$

- Most classical optimization problem.
- Numerous applications: OR, Finance, etc.
- Mature general-purpose solvers: Gurobi, CPLEX, HiGHS.

What is left to do?

- Classical methods challenged by problem sizes.
- Rise of primal-dual algorithms (nonlinear!) since 2020s.

S. J. Wright, Optimization in theory in practice, 2025

Links between linear and nonlinear programming

Interior-point methods

- Newton-type methods with polynomial guarantees.
- Extend to other convex programming setting such as SDPs:

$$\min_{\boldsymbol{X} \in \mathbb{R}^{n \times n}} \langle \boldsymbol{C}, \boldsymbol{X} \rangle \quad \text{s.t.} \quad \mathcal{A}(\boldsymbol{X}) = \boldsymbol{b}, \boldsymbol{X} \succeq 0.$$

$\rightarrow$ Can we get similar guarantees for **nonconvex** formulations?

Links between linear and nonlinear programming

Interior-point methods

- Newton-type methods with polynomial guarantees.
- Extend to other convex programming setting such as SDPs:

$$\min_{\boldsymbol{X} \in \mathbb{R}^{n \times n}} \langle \boldsymbol{C}, \boldsymbol{X} \rangle \quad \text{s.t.} \quad \mathcal{A}(\boldsymbol{X}) = \boldsymbol{b}, \boldsymbol{X} \succeq 0.$$

$\rightarrow$ Can we get similar guarantees for **nonconvex** formulations?

Simplex algorithms

- Classical algorithm for linear programming.
- Numerous variants but no polynomial guarantees.
- Network simplex: Structured variant for transportation problems?

$\rightarrow$ Can we adapt these methods to **nonlinear settings**?

What this talk is about

Nonlinear algorithms

- Goal: Reach high precision.
- Tools: Modern optimization techniques/perspectives.

What this talk is about

Nonlinear algorithms

- Goal: Reach high precision.
- Tools: Modern optimization techniques/perspectives.

Strict saddle optimization

- Special classes of nonconvex problems.
- Good theory/practice of nonlinear algorithms.
- Focus: Phase retrieval.

What this talk is about

Nonlinear algorithms

- Goal: Reach high precision.
- Tools: Modern optimization techniques/perspectives.

Strict saddle optimization

- Special classes of nonconvex problems.
- Good theory/practice of nonlinear algorithms.
- Focus: Phase retrieval.

Adapted network simplex

- Extension of network simplex methods to nonlinear settings.
- Focus on unbalanced optimal transport.

Outline

Outline

An easy example

$$\min_{\boldsymbol{x}\in\mathbb{R}^2} \; x_1 - x_2 \qquad \text{s.t.} \qquad x_1 + x_2 = 1, \boldsymbol{x} \geq 0.$$

$\rightarrow$ Inequality constraints: difficult part in LP!

$\rightarrow$ Naive/Old approach: Square variables!

An easy example

$$\min_{\boldsymbol{x}\in\mathbb{R}^2} x_1 - x_2 \qquad \text{s.t.} \qquad x_1 + x_2 = 1, \boldsymbol{x} \geq 0.$$

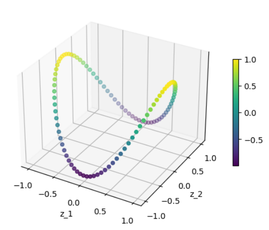
$\rightarrow$ Inequality constraints: difficult part in LP!

$\rightarrow$ Naive/Old approach: Square variables!

Hadamard parametrization

$$\min_{\boldsymbol{z}\in\mathbb{R}^2} z_1^2 - z_2^2 \qquad \text{s.t.} \qquad z_1^2 + z_2^2 = 1.$$

- Nonconvex objective.
- Sphere constraint.



An easy example

$$\min_{\boldsymbol{x}\in\mathbb{R}^2} \; x_1 - x_2 \qquad \text{s.t.} \qquad x_1 + x_2 = 1, \boldsymbol{x} \geq 0.$$

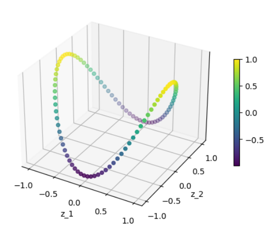
$\rightarrow$ Inequality constraints: difficult part in LP!

$\rightarrow$ Naive/Old approach: Square variables!

Hadamard parametrization

$$\min_{\boldsymbol{z}\in\mathbb{R}^2} \; z_1^2 - z_2^2 \qquad \text{s.t.} \qquad z_1^2 + z_2^2 = 1.$$

- Nonconvex objective.
  But all minima are global!
- Sphere constraint.
  But tractable optimality conditions!



Benign landscape property

Definition

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

$$\|\nabla f(\boldsymbol{x})\| = 0 \text{ and } \nabla^2 f(\boldsymbol{x}) \succeq 0 \quad \Leftrightarrow \quad \boldsymbol{x} \in \operatorname*{argmin}_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x}).$$

- Every local minimum is global.
- Non-minima with $\|\nabla f(\boldsymbol{x})\| = 0$ are strict saddle points ($\nabla^2 f(\boldsymbol{x}) \not\succeq 0$).

Benign landscape property

Definition

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

$$\|\nabla f(\boldsymbol{x})\| = 0 \text{ and } \nabla^2 f(\boldsymbol{x}) \succeq 0 \quad \Leftrightarrow \quad \boldsymbol{x} \in \operatorname*{argmin}_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x}).$$

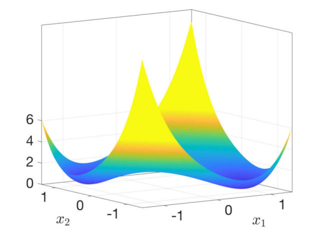
- Every local minimum is global.
- Non-minima with $\|\nabla f(\boldsymbol{x})\| = 0$ are **strict saddle points** ($\nabla^2 f(\boldsymbol{x}) \not\succeq 0$).

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.

Rank-1 approximation

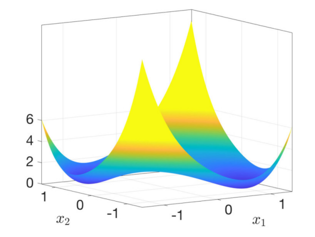
$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \frac{1}{4}\|\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}} - \boldsymbol{M}\|_F^2$$



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

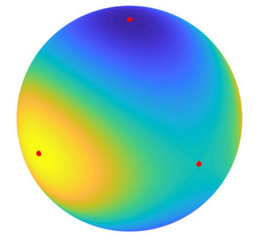
Rank-1 approximation

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



Minimum Eigenvalue problem

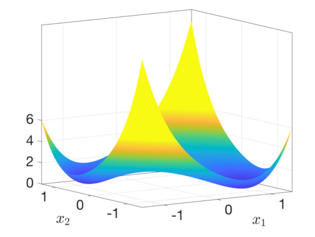
$$\min_{\|\boldsymbol{x}\|=1}\ \boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x}$$



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

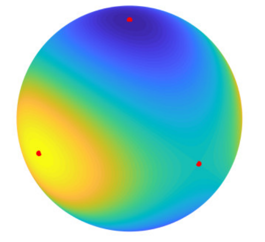
Rank-1 approximation

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



Minimum Eigenvalue problem

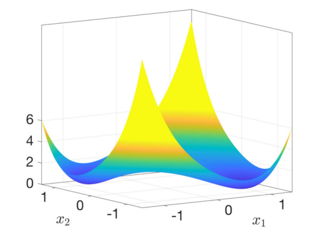
$$\min_{\|\boldsymbol{x}\|=1} \boldsymbol{x}^{\mathrm{T}} A\boldsymbol{x}$$



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

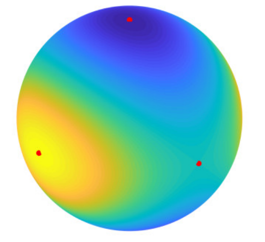
Rank-1 approximation

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



Minimum Eigenvalue problem

$$\min_{\|\boldsymbol{x}\|=1} \boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}$$



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

Motivation: Nonconvex problems are hard

- Finding points such that $\nabla f(\boldsymbol{x}) = 0$ and $\nabla^2 f(\boldsymbol{x}) \succeq 0$ is NP-hard.
- Modern goal: Find approximate stationary points:

$$\|\nabla f(\boldsymbol{x})\| \le \epsilon_{\mathrm{g}} \quad \text{and} \quad \nabla^2 f(\boldsymbol{x}) \succeq -\epsilon_{\mathrm{H}} \boldsymbol{I}.$$

$\rightarrow$ *When are these close to solutions?*

From benign landscape to strict saddle

Motivation: Nonconvex problems are hard

- Finding points such that $\nabla f(\boldsymbol{x}) = 0$ and $\nabla^2 f(\boldsymbol{x}) \succeq 0$ is NP-hard.
- Modern goal: Find approximate stationary points:

$$\|\nabla f(\boldsymbol{x})\| \le \epsilon_{\mathrm{g}} \quad \text{and} \quad \nabla^2 f(\boldsymbol{x}) \succeq -\epsilon_{\mathrm{H}} \boldsymbol{I}.$$

$\rightarrow$ *When are these close to solutions?*

Our tool: Strict saddle property

- Parametric definition.
- Allows for complexity analysis.
- Applies to manifold optimization.

Main problem

$$\operatorname*{minimize}_{\boldsymbol{x} \in \mathcal{M}} f(\boldsymbol{x}),$$

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

Setup: Manifold optimization

Main problem

$$\operatorname*{minimize}_{\boldsymbol{x} \in \mathcal{M}} f(\boldsymbol{x}),$$

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

Goal: Approximate stationary points

$$\|g(\boldsymbol{x}_k)\| \le \epsilon_{\mathrm{g}} \quad \text{and} \quad \lambda_{\min}\left(\mathcal{H}(\boldsymbol{x}_k)\right) \ge -\epsilon_{\mathrm{H}}.$$

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

Strict saddle property on manifold $\mathcal{M}$

Definition (Ge et al '15 → Goyens, R. '24)

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

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

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

Strict saddle property on manifold $\mathcal{M}$

Definition (Ge et al '15 → Goyens, R. '24)

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

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

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

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).

Simple strict saddle functions

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

Strongly convex⇒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 $\{\nabla^2 f(x) \not\succeq -\beta I\}$ 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

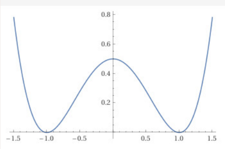
Strict saddle: $\{\|\nabla f(x)\| \geq \alpha\}$ or $\{\nabla^2 f(x) \not\succeq -\beta I\}$ or $\{\|x - x^*\| \leq \delta$ and $\nabla^2 f(y) \succeq \gamma I \ \forall y : \ \|y - x^*\| \leq 2\delta\}$.

Strongly convex⇒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.

A "truly" strict saddle function

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



Example: Matrix completion

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

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

Example: Matrix completion

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

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

Theorem (Ge et al. '17)

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

1. $\|\nabla f(\boldsymbol{U},\boldsymbol{V})\| \geq \alpha$
2. The Hessian at U, V has negative curvature, i.e.
$$\lambda_{\min}\left(\nabla^2 f(\boldsymbol{U},\boldsymbol{V})\right) < -\mathcal{O}(\sigma_{\min}(M))$$
3. $(\boldsymbol{U},\boldsymbol{V})$ is at distance at most $\mathcal{O}(\frac{\alpha}{\sigma_{\min}(\boldsymbol{M})})$ from a global minimum.

Algorithms for strict saddle problems

- **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).

Algorithms for strict saddle problems

- **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 choice: Trust-region methods

- Initially developed for nonconvex problems.
- Use Newton-type steps (like interior-point methods in LP).

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

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

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

$$\boldsymbol{s}_k \in \operatorname*{argmin}_{\substack{\boldsymbol{s}\in\mathcal{T}^{\mathcal{M}}_{\boldsymbol{x}_k} \\ \|\boldsymbol{s}\|\le\Delta_k}} m_k(\boldsymbol{x}_k + \boldsymbol{s}).$$

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

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

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

1. Define $m_k(\boldsymbol{x}_k + \boldsymbol{s}) := \langle g(\boldsymbol{x}_k), \boldsymbol{s} \rangle + \frac{1}{2}\langle \boldsymbol{s}, \mathcal{H}(\boldsymbol{x}_k)\boldsymbol{s} \rangle$ and compute
$$\boldsymbol{s}_k \in \operatorname*{argmin}_{\substack{\boldsymbol{s} \in \mathcal{T}^{\mathcal{M}}_{\boldsymbol{x}_k} \\ \|\boldsymbol{s}\| \le \Delta_k}} m_k(\boldsymbol{x}_k + \boldsymbol{s}).$$
2. Define $\boldsymbol{x}_k^{\mathcal{M}}$ as the retraction of $\boldsymbol{x}_k + \boldsymbol{s}_k$ onto $\mathcal{M}$.

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

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

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

1. Define $m_k(\boldsymbol{x}_k + \boldsymbol{s}) := \langle g(\boldsymbol{x}_k), \boldsymbol{s} \rangle + \frac{1}{2} \langle \boldsymbol{s}, \mathcal{H}(\boldsymbol{x}_k) \boldsymbol{s} \rangle$ and compute
$$\boldsymbol{s}_k \in \underset{\substack{\boldsymbol{s} \in \mathcal{T}^{\mathcal{M}}_{\boldsymbol{x}_k} \\ \|\boldsymbol{s}\| \le \Delta_k}}{\operatorname{argmin}}\, m_k(\boldsymbol{x}_k + \boldsymbol{s}).$$
2. Define $\boldsymbol{x}_k^{\mathcal{M}}$ as the retraction of $\boldsymbol{x}_k + \boldsymbol{s}_k$ onto $\mathcal{M}$.
3. Compute $\rho_k = \frac{f(\boldsymbol{x}_k) - f(\boldsymbol{x}_k^{\mathcal{M}})}{m_k(\boldsymbol{x}_k) - m_k(\boldsymbol{x}_k^{\mathcal{M}})}$.

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

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

1. Define $m_k(\boldsymbol{x}_k + \boldsymbol{s}) := \langle g(\boldsymbol{x}_k), \boldsymbol{s} \rangle + \frac{1}{2}\langle \boldsymbol{s}, \mathcal{H}(\boldsymbol{x}_k)\boldsymbol{s} \rangle$ and compute
$$\boldsymbol{s}_k \in \underset{\substack{\boldsymbol{s} \in \mathcal{T}^{\mathcal{M}}_{\boldsymbol{x}_k} \\ \|\boldsymbol{s}\| \le \Delta_k}}{\operatorname{argmin}}\, m_k(\boldsymbol{x}_k + \boldsymbol{s}).$$
2. Define $\boldsymbol{x}_k^{\mathcal{M}}$ as the retraction of $\boldsymbol{x}_k + \boldsymbol{s}_k$ onto $\mathcal{M}$.
3. Compute $\rho_k = \frac{f(\boldsymbol{x}_k) - f(\boldsymbol{x}_k^{\mathcal{M}})}{m_k(\boldsymbol{x}_k) - m_k(\boldsymbol{x}_k^{\mathcal{M}})}$.
4. If $\rho_k \ge \eta$, set $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k^{\mathcal{M}}$ and $\Delta_{k+1} = 2\Delta_k$.

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

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

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

1. Define $m_k(\boldsymbol{x}_k + \boldsymbol{s}) := \langle g(\boldsymbol{x}_k), \boldsymbol{s} \rangle + \frac{1}{2}\langle \boldsymbol{s}, \mathcal{H}(\boldsymbol{x}_k)\boldsymbol{s} \rangle$ and compute
$$\boldsymbol{s}_k \in \operatorname*{argmin}_{\substack{\boldsymbol{s} \in \mathcal{T}^{\mathcal{M}}_{\boldsymbol{x}_k} \\ \|\boldsymbol{s}\| \le \Delta_k}} m_k(\boldsymbol{x}_k + \boldsymbol{s}).$$
2. Define $\boldsymbol{x}_k^{\mathcal{M}}$ as the retraction of $\boldsymbol{x}_k + \boldsymbol{s}_k$ onto $\mathcal{M}$.
3. Compute $\rho_k = \frac{f(\boldsymbol{x}_k) - f(\boldsymbol{x}_k^{\mathcal{M}})}{m_k(\boldsymbol{x}_k) - m_k(\boldsymbol{x}_k^{\mathcal{M}})}$.
4. If $\rho_k \ge \eta$, set $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k^{\mathcal{M}}$ and $\Delta_{k+1} = 2\Delta_k$.
5. Otherwise, set $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k$ and $\Delta_{k+1} = 0.5\Delta_k$.

Complexity results for trust region

Goal: Compute $\boldsymbol{x}_K$ such that $\|g(\boldsymbol{x}_K)\| \le \epsilon_{\mathrm{g}}$ and $\lambda_{\min}(\mathcal{H}(\boldsymbol{x}_K)) \ge -\epsilon_{\mathrm{H}}$.

For general nonconvex f (Boumal et al '19)

$$K = \mathcal{O}\left(\max\{\epsilon_{\mathrm{g}}^{-2}, \epsilon_{\mathrm{H}}^{-3}\}\right).$$

Complexity results for trust region

Goal: Compute $\boldsymbol{x}_K$ such that $\|g(\boldsymbol{x}_K)\| \le \epsilon_{\mathrm{g}}$ and $\lambda_{\min}(\mathcal{H}(\boldsymbol{x}_K)) \ge -\epsilon_{\mathrm{H}}$.

For general nonconvex f (Boumal et al '19)

$$K = \mathcal{O}\left(\max\{\epsilon_{\mathrm{g}}^{-2}, \epsilon_{\mathrm{H}}^{-3}\}\right).$$

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

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

$$\begin{aligned} K_f &= \mathcal{O}\left(\max\left\{\alpha^{-2}\beta^{-1}, \alpha^{-2}\gamma^{-1}, \beta^{-3}, \gamma^{-3}, \gamma^{-2}\delta^{-1}\right\}\right) \\ K_\epsilon &= \log\log\left[\mathcal{O}\left(\gamma\epsilon_{\mathrm{g}}^{-1}\right)\right]. \end{aligned}$$

Complexity results for trust region

Goal: Compute $\boldsymbol{x}_K$ such that $\|g(\boldsymbol{x}_K)\| \le \epsilon_{\mathrm{g}}$ and $\lambda_{\min}\left(\mathcal{H}(\boldsymbol{x}_K)\right) \ge -\epsilon_{\mathrm{H}}$.

For general nonconvex f (Boumal et al '19)

$$K = \mathcal{O}\left(\max\{\epsilon_{\mathrm{g}}^{-2}, \epsilon_{\mathrm{H}}^{-3}\}\right).$$

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

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

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

$$K_\epsilon = \log\log\left[\mathcal{O}\left(\gamma\epsilon_{\mathrm{g}}^{-1}\right)\right].$$

- $\log\log$ dependency in ϵ_{g} (none on ϵ_{H})!
- Complexity depends more on landscape parameters!

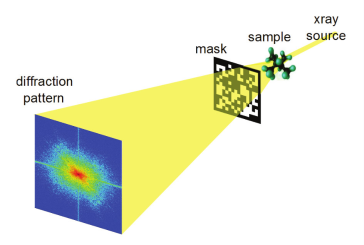
Outline

Phase retrieval

Phase retrieval problem

$$
\begin{array}{ll}
\text{find} & \boldsymbol{x} \in \mathbb{C}^n \\
\text{such that} & |\langle \boldsymbol{x}, \boldsymbol{a}_i \rangle| = b_i, \quad i = 1, \dots, m.
\end{array}
$$

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



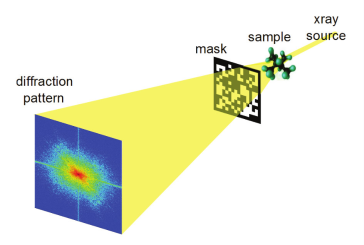
(Picture from Candès et al '15)

Phase retrieval

Phase retrieval problem

$$\begin{array}{ll} \text{find} & \boldsymbol{x} \in \mathbb{C}^n \\ \text{such that} & |\langle \boldsymbol{x}, \boldsymbol{a}_i \rangle| = b_i, \quad i = 1, \dots, m. \end{array}$$

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



(Picture from Candès et al '15)

Gives rise to convex and nonconvex optimization problems!

Strict saddle phase retrieval ('ed)

Nonconvex vector formulation (Sun et al '18)

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

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

Strict saddle phase retrieval ('ed)

Nonconvex vector formulation (Sun et al '18)

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

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

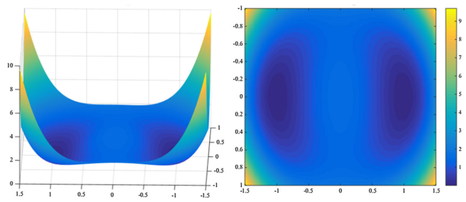
Strict saddle phase retrieval ('ed)

Nonconvex vector formulation (Sun et al '18)

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

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

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



Trust region for phase retrieval

Phase retrieval (Sun et al '18)

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

If $\{\boldsymbol{a}_i\}$ 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 $c > 0$ w.h.p..

Trust region for phase retrieval

Phase retrieval (Sun et al '18)

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

If $\{\boldsymbol{a}_i\}$ 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 $c > 0$ w.h.p..

Complexity results for trust region (Goyens, R. '24)

- Basic trust region:

$$\mathcal{O}(\max\left\{\epsilon_{\mathrm{g}}^{-2}, \epsilon_{\mathrm{H}}^{-3}\right\}) \quad \text{iterations.}$$

- Using strict saddle:

$$\mathcal{O}(n^2) + \log\log(\mathcal{O}(\epsilon_{\mathrm{g}}^{-1})) \quad \text{iterations.}$$

From strict saddle to good nonconvex?

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!

Phase retrieval SDPs

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

PhaseLift SDP relaxation (Candès et al. '11)

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

→ Convex problem, can be solved in polynomial time (like LP!).

→ In practice, cost of matrix linear algebra prohibitive.

→ Expected: **Rank-1 solution**.

Phase retrieval SDPs

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

PhaseLift SDP relaxation (Candès et al. '11)

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

→ Convex problem, can be solved in polynomial time (like LP!).

→ In practice, cost of matrix linear algebra prohibitive.

→ Expected: Rank-1 solution.

Factored PhaseLift formulation

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

→ Burer-Monteiro factorization (Burer and Monteiro '03).

→ Not strict saddle for small r!

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

Trust region on PhaseLift

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

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

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

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

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

Trust region on PhaseLift

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

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

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

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

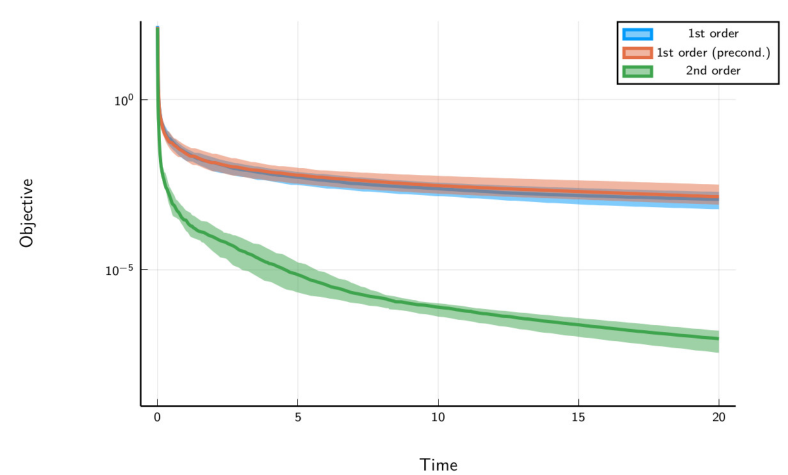
for $K = \mathcal{O}(\log(1/\epsilon))$.

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

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

Setup:

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



Recap: Phase retrieval

Strict saddle phase retrieval

- Amenable to analysis.
- Fast convergence guaranteed for trust region.

SDP-based phase retrieval

- Not always strict saddle.
- But trust region fast anyways!

Outline

Back to linear programming

My main line of research

- Structured nonconvex formulations.
- Can revisit convex algorithms in that space.

Back to linear programming

My main line of research

- Structured nonconvex formulations.
- Can revisit convex algorithms in that space.

What about convex problems?

- Primal-dual methods for solving LPs are booming.
- Rely on convex optimization algorithms!

Back to linear programming

My main line of research

- Structured nonconvex formulations.
- Can revisit convex algorithms in that space.

What about convex problems?

- Primal-dual methods for solving LPs are booming.
- Rely on convex optimization algorithms!

Our approach (w/ G. Agazzotti & A. Chambolle)

- Explore structured LPs from optimal transport.
- Start with classical (simplex) framework.

Optimal transport

$$\min_{\boldsymbol{P} \in \mathcal{U}(\boldsymbol{a},\boldsymbol{b})} \langle \boldsymbol{P}, \boldsymbol{C} \rangle$$

- $\boldsymbol{a}$,$\boldsymbol{b}$ measures on $\mathbb{R}^n$, $\|\boldsymbol{a}\|_1 = \|\boldsymbol{b}\|_1$.
- $\mathcal{U}(\boldsymbol{a}, \boldsymbol{b}) = \{\boldsymbol{P} \geq 0 \,|\boldsymbol{P} 1_n = \boldsymbol{a}, \boldsymbol{P}^{\mathrm{T}} 1_n = \boldsymbol{b}\}$.

Optimal transport

$$\min_{\boldsymbol{P} \in \mathcal{U}(\boldsymbol{a},\boldsymbol{b})} \langle \boldsymbol{P}, \boldsymbol{C} \rangle$$

- $\boldsymbol{a}$,$\boldsymbol{b}$ measures on $\mathbb{R}^n$, $\|\boldsymbol{a}\|_1 = \|\boldsymbol{b}\|_1$.
- $\mathcal{U}(\boldsymbol{a}, \boldsymbol{b}) = \{\boldsymbol{P} \geq 0 \,|\boldsymbol{P} 1_n = \boldsymbol{a}, \boldsymbol{P}^{\mathrm{T}} 1_n = \boldsymbol{b}\}$.

Linear program!

- Transportation problem on bipartite graph.
- Challenging to tackle through interior-point methods.
- State of the art: Network simplex (Flamary et al '21).

Optimality conditions

If $\boldsymbol{P}$ solution, $\exists(\boldsymbol{u}, \boldsymbol{v})$ such that

$$\boldsymbol{P} \in \mathcal{U}(\boldsymbol{a}, \boldsymbol{b}), \quad \boldsymbol{C} \geq \boldsymbol{u} \oplus \boldsymbol{v}, \quad \boldsymbol{P} \odot (\boldsymbol{C} - \boldsymbol{u} \oplus \boldsymbol{v}) = 0.$$

Network simplex

Optimality conditions

If $\boldsymbol{P}$ solution, $\exists(\boldsymbol{u}, \boldsymbol{v})$ such that

$$\boldsymbol{P} \in \mathcal{U}(\boldsymbol{a}, \boldsymbol{b}), \quad \boldsymbol{C} \geq \boldsymbol{u} \oplus \boldsymbol{v}, \quad \boldsymbol{P} \odot (\boldsymbol{C} - \boldsymbol{u} \oplus \boldsymbol{v}) = 0.$$

Network simplex idea

- Maintain $(\boldsymbol{P}, \boldsymbol{u}, \boldsymbol{v})$ such that $\boldsymbol{P} \in \mathcal{U}(\boldsymbol{a}, \boldsymbol{b})$ and $\boldsymbol{P} \odot (\boldsymbol{C} - \boldsymbol{u} \oplus \boldsymbol{v}) = 0$.
- If $\boldsymbol{C} \geq \boldsymbol{u} \oplus \boldsymbol{v}$, optimal. Otherwise, find $u_i + v_j > C_{ij}$ and update $(\boldsymbol{P}, \boldsymbol{u}, \boldsymbol{v})$ to satisfy $C_{ij} \geq u_i + v_j$ + other two conditions.

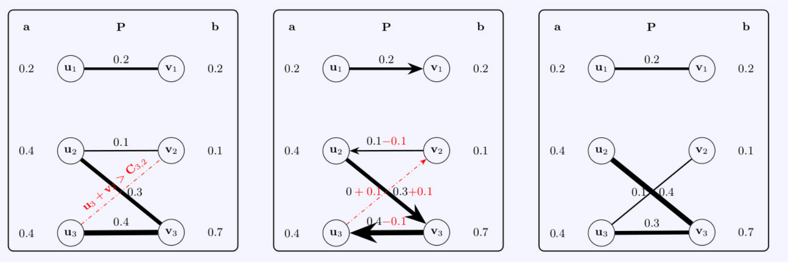
Optimality conditions

If $\boldsymbol{P}$ solution, $\exists(\boldsymbol{u}, \boldsymbol{v})$ such that

$$\boldsymbol{P} \in \mathcal{U}(\boldsymbol{a}, \boldsymbol{b}), \quad \boldsymbol{C} \geq \boldsymbol{u} \oplus \boldsymbol{v}, \quad \boldsymbol{P} \odot (\boldsymbol{C} - \boldsymbol{u} \oplus \boldsymbol{v}) = 0.$$

Network simplex idea

- Maintain $(\boldsymbol{P}, \boldsymbol{u}, \boldsymbol{v})$ such that $\boldsymbol{P} \in \mathcal{U}(\boldsymbol{a}, \boldsymbol{b})$ and $\boldsymbol{P} \odot (\boldsymbol{C} - \boldsymbol{u} \oplus \boldsymbol{v}) = 0$.
- If $\boldsymbol{C} \geq \boldsymbol{u} \oplus \boldsymbol{v}$, optimal. Otherwise, find $u_i + v_j > C_{ij}$ and update $(\boldsymbol{P}, \boldsymbol{u}, \boldsymbol{v})$ to satisfy $C_{ij} \geq u_i + v_j$ + other two conditions.



Unbalanced (L2) optimal transport

Optimal transport

$$\min_{\boldsymbol{P} \in \mathcal{U}(\boldsymbol{a},\boldsymbol{b})} \langle \boldsymbol{P}, \boldsymbol{C} \rangle, \quad \mathcal{U}(\boldsymbol{a},\boldsymbol{b}) = \{\boldsymbol{P} \geq 0, \boldsymbol{P}1_n = \boldsymbol{a}, \boldsymbol{P}^{\mathrm{T}}1_n = \boldsymbol{b}\}.$$

Unbalanced (L2) optimal transport

Optimal transport

$$\min_{\boldsymbol{P}\in\mathcal{U}(\boldsymbol{a},\boldsymbol{b})}\langle \boldsymbol{P},\boldsymbol{C}\rangle, \quad \mathcal{U}(\boldsymbol{a},\boldsymbol{b})=\{\boldsymbol{P}\geq 0, \boldsymbol{P}1_n=\boldsymbol{a}, \boldsymbol{P}^{\mathrm{T}}1_n=\boldsymbol{b}\}.$$

Unbalanced optimal transport (UOT)

$$\min_{\boldsymbol{P}\geq 0}\langle \boldsymbol{P},\boldsymbol{C}\rangle+\frac{1}{2\tau}\|\boldsymbol{P}1_n-\boldsymbol{a}\|^2+\frac{1}{2\tau}\|\boldsymbol{P}^{\mathrm{T}}1_n-\boldsymbol{b}\|^2.$$

Optimal transport

$$\min_{\boldsymbol{P} \in \mathcal{U}(\boldsymbol{a},\boldsymbol{b})} \langle \boldsymbol{P}, \boldsymbol{C} \rangle, \quad \mathcal{U}(\boldsymbol{a},\boldsymbol{b}) = \{\boldsymbol{P} \geq 0, \boldsymbol{P} 1_n = \boldsymbol{a}, \boldsymbol{P}^{\mathrm{T}} 1_n = \boldsymbol{b}\}.$$

Unbalanced optimal transport (UOT)

$$\min_{\boldsymbol{P} \geq 0} \langle \boldsymbol{P}, \boldsymbol{C} \rangle + \frac{1}{2\tau} \|\boldsymbol{P} 1_n - \boldsymbol{a}\|^2 + \frac{1}{2\tau} \|\boldsymbol{P}^{\mathrm{T}} 1_n - \boldsymbol{b}\|^2.$$

- $\tau > 0$: Relaxes marginal constraints.
- Quadratic (nonlinear) optimization problem.

Unbalanced (L2) optimal transport

Optimal transport

$$\min_{\boldsymbol{P} \in \mathcal{U}(\boldsymbol{a},\boldsymbol{b})} \langle \boldsymbol{P}, \boldsymbol{C} \rangle, \quad \mathcal{U}(\boldsymbol{a},\boldsymbol{b}) = \{\boldsymbol{P} \geq 0, \boldsymbol{P}1_n = \boldsymbol{a}, \boldsymbol{P}^{\mathrm{T}}1_n = \boldsymbol{b}\}.$$

Unbalanced optimal transport (UOT)

$$\min_{\boldsymbol{P} \geq 0} \langle \boldsymbol{P}, \boldsymbol{C} \rangle + \frac{1}{2\tau}\|\boldsymbol{P}1_n - \boldsymbol{a}\|^2 + \frac{1}{2\tau}\|\boldsymbol{P}^{\mathrm{T}}1_n - \boldsymbol{b}\|^2.$$

- $\tau > 0$: Relaxes marginal constraints.
- Quadratic (nonlinear) optimization problem.

Goals

- Adapt network simplex to UOT.
- Get efficient implementation.

Network simplex for UOT

Network simplex for OT

- Uses optimality conditions

$$\boldsymbol{P} \in \mathcal{U}(\boldsymbol{a},\boldsymbol{b}), \quad \boldsymbol{C} \geq \boldsymbol{u} \oplus \boldsymbol{v}, \quad \boldsymbol{P} \odot (\boldsymbol{C} - \boldsymbol{u} \oplus \boldsymbol{v}) = 0.$$

- Exploits bipartite graph structure.
- Complexity guarantees depend on the variant.

Network simplex for UOT

- Optimality conditions still easy to check:

$$\boldsymbol{P} \in \mathcal{U}(\boldsymbol{a} - \tau \boldsymbol{u}, \boldsymbol{b} - \tau \boldsymbol{v}), \quad \boldsymbol{C} \geq \boldsymbol{u} \oplus \boldsymbol{v}, \quad \boldsymbol{P} \odot (\boldsymbol{C} - \boldsymbol{u} \oplus \boldsymbol{v}) = 0.$$

- Key: Explicit formulas for updating transport graph!

$\rightarrow$ Can prove complexity results!

Extensions

- Optimal transport:

$$\min_{\boldsymbol{P} \in \mathcal{U}(\boldsymbol{a},\boldsymbol{b})} \langle \boldsymbol{P}, \boldsymbol{C} \rangle$$

- Unbalanced optimal transport:

$$\min_{\boldsymbol{P} \geq 0} \langle \boldsymbol{P}, \boldsymbol{C} \rangle + \frac{1}{2\tau} \|\boldsymbol{P} 1_n - \boldsymbol{a}\|^2 + \frac{1}{2\tau} \|\boldsymbol{P}^{\mathrm{T}} 1_n - \boldsymbol{b}\|^2$$

- Semi-unbalanced optimal transport (SUOT)

$$\min_{\substack{\boldsymbol{P} 1_n = \boldsymbol{a} \\ \boldsymbol{P} \geq 0}} \langle \boldsymbol{P}, \boldsymbol{C} \rangle + \frac{1}{2\tau} \left\| \boldsymbol{P}^{\mathrm{T}} 1_n - \boldsymbol{b} \right\|^2 .$$

Extensions

- Optimal transport:

$$\min_{\boldsymbol{P} \in \mathcal{U}(\boldsymbol{a},\boldsymbol{b})} \langle \boldsymbol{P}, \boldsymbol{C} \rangle$$

- Unbalanced optimal transport:

$$\min_{\boldsymbol{P} \geq 0} \langle \boldsymbol{P}, \boldsymbol{C} \rangle + \frac{1}{2\tau} \|\boldsymbol{P} 1_n - \boldsymbol{a}\|^2 + \frac{1}{2\tau} \|\boldsymbol{P}^{\mathrm{T}} 1_n - \boldsymbol{b}\|^2$$

- Semi-unbalanced optimal transport (SUOT)

$$\min_{\substack{\boldsymbol{P} 1_n = \boldsymbol{a} \\ \boldsymbol{P} \geq 0}} \langle \boldsymbol{P}, \boldsymbol{C} \rangle + \frac{1}{2\tau} \left\|\boldsymbol{P}^{\mathrm{T}} 1_n - \boldsymbol{b}\right\|^2.$$

- Can use strictly convex functions/Bregman divergences:

$$\min_{\boldsymbol{P} \geq 0} \langle \boldsymbol{P}, \boldsymbol{C} \rangle + \frac{1}{2\tau} \psi(\boldsymbol{P} 1_n, \boldsymbol{a}) + \frac{1}{2\tau} \psi(\boldsymbol{P}^{\mathrm{T}} 1_n, \boldsymbol{b}).$$

Ex) Kullback-Leibler $\psi(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} x_i \log\left(\frac{x_i}{y_i}\right) - x_i + y_i$.

Model (Blondel et al '18)

$$\min_{\boldsymbol{P} 1_n = 1_n/K} \langle \boldsymbol{P}, \boldsymbol{C} \rangle + \frac{1}{2\tau} \left\| \boldsymbol{P}^{\mathrm{T}} 1_n - \frac{1_n}{K} \right\|^2$$

- Uniform measure on K uniform points for each image.
- $\boldsymbol{C}_{ij} = \|\boldsymbol{x}_i - \boldsymbol{y}_j\|^2$, $\boldsymbol{x}_i$ and $\boldsymbol{y}_j$ sampled points.

Model (Blondel et al '18)

$$\min_{\boldsymbol{P}1_n = 1_n/K} \langle \boldsymbol{P}, \boldsymbol{C} \rangle + \frac{1}{2\tau} \left\| \boldsymbol{P}^{\mathrm{T}} 1_n - \frac{1_n}{K} \right\|^2$$

- Uniform measure on K uniform points for each image.
- $\boldsymbol{C}_{ij} = \|\boldsymbol{x}_i - \boldsymbol{y}_j\|^2$, $\boldsymbol{x}_i$ and $\boldsymbol{y}_j$ sampled points.

(a) *Femme au béret et à la robe quadrillée*

(b) *La nuit étoilée*

(c) τ = 1e-4

(d) τ = 1e-1

(e) τ = 1e1

(f) τ = 1e4

Benchmark Random, uniform instances.

Competitors High-accuracy solvers ($<10^{-12}$ accuracy).

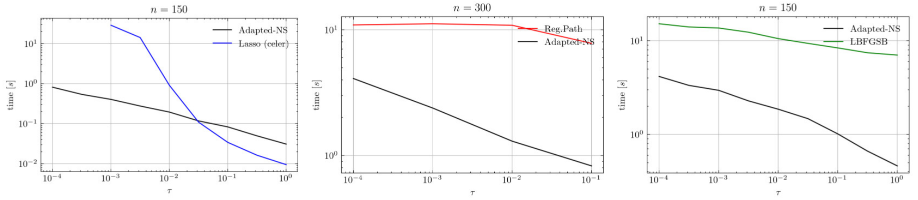
- UOT: `celer` (Massias et al '18).
- SUOT: Regularization path (Lee et al '19).
- UOT+KL: LBFGSB from Scipy.

Comparison with other solvers

Benchmark Random, uniform instances.

Competitors High-accuracy solvers ($<10^{-12}$ accuracy).

- UOT: `celer` (Massias et al '18).
- SUOT: Regularization path (Lee et al '19).
- UOT+KL: LBFGSB from Scipy.



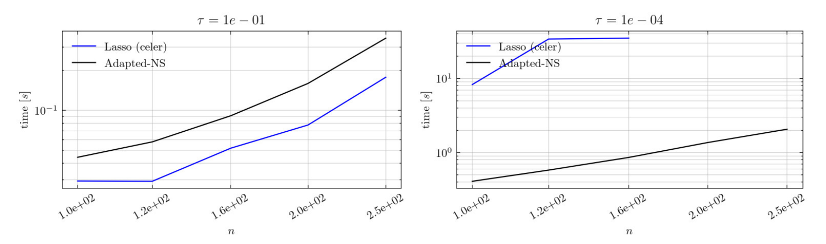
(a) UOT-L^2 (b) SUOT-L^2 (c) UOT-KL

Comparison with other solvers ('ed)

Focus: UOT

$$\min_{\boldsymbol{P}\geq 0} \langle \boldsymbol{P}, \boldsymbol{C} \rangle + \frac{1}{2\tau}\|\boldsymbol{P}1_n - \boldsymbol{a}\|^2 + \frac{1}{2\tau}\|\boldsymbol{P}^{\mathrm{T}}1_n - \boldsymbol{b}\|^2$$

- celer/LASSO approach more efficient with large τ.
- Network simplex better when $\tau \to 0$.



Summary

Summary

Linear programming

- Currently being revisited.
- Algorithmic focus shifts.

Summary

Linear programming

- Currently being revisited.
- Algorithmic focus shifts.

Strict saddle programming

- Tractable nonconvex instances (phase retrieval).
- Efficient algorithms akin to Newton-type interior point.

Summary

Linear programming

- Currently being revisited.
- Algorithmic focus shifts.

Strict saddle programming

- Tractable nonconvex instances (phase retrieval).
- Efficient algorithms akin to Newton-type interior point.

Unbalanced optimal transport

- Convex (nonlinear) problems with graph structure.
- Can adapt network simplex algorithms!

Perspectives

Next steps

- Get those papers out (+code).
- Study nonconvex formulations of unbalanced optimal transport?

- G. Agazzotti, A. Chambolle and C. W. Royer, *Adapted network simplex algorithms for unbalanced optimal transport*, working paper.
- P. Caucheteux, F. Goyens, C. W. Royer and I. Waldspurger, *Trust-region methods for Burer-Monteiro factorizations of phase retrieval problems*, working paper.
- F. Goyens and C. W. Royer, *Riemannian trust-region methods for strict saddle functions with complexity guarantees*, Mathematical Programming, 2024.

Next steps

- Get those papers out (+code).
- Study nonconvex formulations of unbalanced optimal transport?

- G. Agazzotti, A. Chambolle and C. W. Royer, *Adapted network simplex algorithms for unbalanced optimal transport*, working paper.
- P. Caucheteux, F. Goyens, C. W. Royer and I. Waldspurger, *Trust-region methods for Burer-Monteiro factorizations of phase retrieval problems*, working paper.
- F. Goyens and C. W. Royer, *Riemannian trust-region methods for strict saddle functions with complexity guarantees*, Mathematical Programming, 2024.

Merci !

clement.royer@lamsade.dauphine.fr