Escaping Saddle Points in Constrained Optimization

Aryan Mokhtari, Asuman Ozdaglar, and Ali Jadbabaie

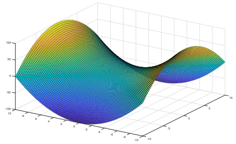
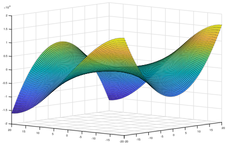
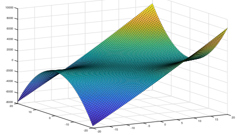
Laboratory for Information and Decision Systems (LIDS)
Institute for Data, Systems, and Society (IDSS)
Massachusetts Institute of Technology

The 32nd Annual Conference on Neural Information Processing Systems

December 5, 2018

This work is supported by DARPA Lagrange Program under grant No. FA 8650-18-2-7838

Nonconvex optimization

- Consider general optimization program $\quad \min_{x \in \mathcal{C}} \ f(x)$

  $\Rightarrow$ $\mathcal{C} \subseteq \mathbb{R}^d$ is a convex compact closed set

- In the nonconvex setting (f is nonconvex)

  $\Rightarrow$ Saddle points exist $\Rightarrow$ First-order optimality condition is not enough

  $\Rightarrow$ Check higher order derivatives to escape from saddle points

  $\Rightarrow$ Search for a second-order stationary point (SOSP)

- In several cases, all saddle points are escapable and all local minima are global

  $\Rightarrow$ Convergence to an SOSP implies convergence to a global minimum!

  $\Rightarrow$ Eigenvector problem[a], phase retrieval[b], dictionary learning[c], ...

---

[a] [Absil et al., '10] [b] [Sun et al., '16] [c] [Sun et al., '17]

Unconstrained optimization

- Consider the unconstrained nonconvex setting ($\mathcal{C} = \mathbb{R}^d$)
- x^* is a second-order stationary point if

$$\underbrace{\|\nabla f(x^*)\| = 0}_{\text{first-order optimality condition}} \quad \text{and} \quad \underbrace{\nabla^2 f(x^*) \succeq \mathbf{0}}_{\text{second-order optimality condition}}$$

- Various attempts to design algorithms converging to an SOSP
  - ⇒ Perturbing iterates by injecting noise[a]
  - ⇒ Finding the eigenvector of the smallest eigenvalue of the Hessian[b]
- Overall comput. cost to find an (ε, γ)-SOSP ⇒ Polynomial in ε^{-1} and γ^{-1}
- However, not applicable to the convex constrained setting!
  - ⇒ **Question**: In the constrained case, can we find an SOSP in poly-time?

[a] [Ge et al., '15], [Jin et al., '17a], [Jin et al., '17b], [Daneshmand et al., '18]
[b] [Carmon et al., '16], [Allen-Zhu, '17], [Xu & Yang, '17], [Royer & Wright, '17], [Agarwal et al., '17], [Reddi et al., '18]

Unconstrained optimization

- Consider the unconstrained nonconvex setting ($\mathcal{C} = \mathbb{R}^d$)
- x^* is an approximate (ε, γ)-second-order stationary point if

$$\underbrace{\|\nabla f(x^*)\| \leq \varepsilon}_{\text{first-order optimality condition}} \quad \text{and} \quad \underbrace{\nabla^2 f(x^*) \succeq -\gamma \mathbf{I}}_{\text{second-order optimality condition}}$$

- Various attempts to design algorithms converging to an SOSP
  - $\Rightarrow$ Perturbing iterates by injecting noise[a]
  - $\Rightarrow$ Finding the eigenvector of the smallest eigenvalue of the Hessian[b]
- Overall comput. cost to find an (ε, γ)-SOSP $\Rightarrow$ Polynomial in ε^{-1} and γ^{-1}
- However, not applicable to the convex constrained setting!
  - $\Rightarrow$ **Question**: In the constrained case, can we find an SOSP in poly-time?

[a] [Ge et al., '15], [Jin et al., '17a], [Jin et al., '17b], [Daneshmand et al., '18]
[b] [Carmon et al., '16], [Allen-Zhu, '17], [Xu & Yang, '17], [Royer & Wright, '17], [Agarwal et al., '17], [Reddi et al., '18]

Constrained optimization: Second-order stationary point

- How should we define an SOSP for the constrained setting?
- Optimality conditions for the constrained setting

  $\Rightarrow$ (i) $\nabla f(x^*)^T(x - x^*) \geq 0 \quad \text{for all } x \in \mathcal{C}$

  $\Rightarrow$ (ii) $(x-x^*)^T\nabla^2 f(x^*)(x-x^*) \geq 0 \quad \text{for all } x \in \mathcal{C} \text{ s.t. } \nabla f(x^*)^T(x-x^*)=0$

- (ii) should hold only on the subspace on which the function could be increasing

Constrained optimization: Second-order stationary point

- How should we define an SOSP for the constrained setting?
- $x^* \in \mathcal{C}$ is an approximate (ε, γ)-second-order order stationary point if

$$\Rightarrow \nabla f(x^*)^T(x - x^*) \geq -\varepsilon \quad \text{for all } x \in \mathcal{C}$$

$$\Rightarrow (x - x^*)^T \nabla^2 f(x^*)(x - x^*) \geq -\gamma \quad \text{for all } x \in \mathcal{C} \text{ s.t. } \nabla f(x^*)^T(x - x^*) = 0$$

- (ii) should hold only on the subspace on which the function could be increasing

Constrained optimization: Second-order stationary point

- How should we define an SOSP for the constrained setting?
- $x^* \in \mathcal{C}$ is an approximate (ε, γ)-second-order order stationary point if

  $\Rightarrow$ (i) $\nabla f(x^*)^T(x - x^*) \geq -\varepsilon$ for all $x \in \mathcal{C}$

  $\Rightarrow$ (ii) $(x-x^*)^T \nabla^2 f(x^*)(x-x^*) \geq -\gamma$ for all $x \in \mathcal{C}$ s. t. $\nabla f(x^*)^T(x-x^*) = 0$
- (ii) should hold only on the subspace on which the function could be increasing
- Setting $\varepsilon = \gamma = 0$ gives the necessary conditions for a local min
- We propose a framework that finds an (ε, γ)-SOSP in poly-time
  - $\Rightarrow$ If optimizing a quadratic loss over $\mathcal{C}$ up to a constant factor is tractable
  - $\Rightarrow$ Using recent advances in solving nonconvex QCQPs

Main result

Theorem *If $\mathcal{C}$ is defined by a single quadratic constraint, then our algorithm finds an (ε, γ)-SOSP after at most $\mathcal{O}(\max\{\varepsilon^{-2}, d^3\gamma^{-3}\})$ arithmetic operations where d is the problem dimension.*

Theorem *If $\mathcal{C}$ is defined as a set of m quadratic constraints $(m > 1)$, and the objective function Hessian satisfies $\max_{x \in \mathcal{C}} x^T \nabla^2 f(x) x \leq \mathcal{O}(\gamma)$, then our algorithm finds an (ε, γ)-SOSP after at most $\mathcal{O}(\max\{\varepsilon^{-2}, d^3 m^7 \gamma^{-3}\})$ arithmetic operations.*

Poster Information

Date/Time: Wed Dec 5th 05:00 – 07:00

Room: 210 & 230 AB

Poster number: 47