HKU Musketeers Foundation
Institute of Data Science
香港大學同心基金數據科學研究院

IDS Briefing Session

Dr. Yue XIE

Department of Mathematics

HKU Musketeers Foundation Institute of Data Science

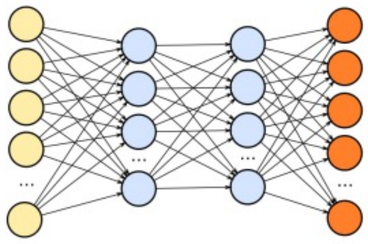
Optimization problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x \in X. \end{aligned} \qquad \text{(OP)}$$

- $f : \mathbb{R}^n \to \mathbb{R}$ is called the objective function.
- $X \subseteq \mathbb{R}^n$ is called the feasible region.

Optimization problems are everywhere in data science.

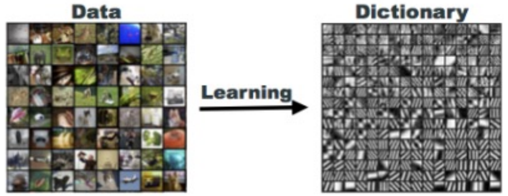
Deep neural networks training.



$$\min_{W_1,\ldots,W_N} \quad \sum_{i=1}^n l(h(x_i), y_i).$$

Unconstrained nonconvex optimization

Dictionary learning.



$$\begin{aligned} \min_{D \in \mathbb{R}^{p\times p}} \quad & \|Y^T D\|_1 \\ \text{subject to} \quad & D^T D = I_p. \end{aligned}$$

Constrained nonconvex optimization

Algorithms

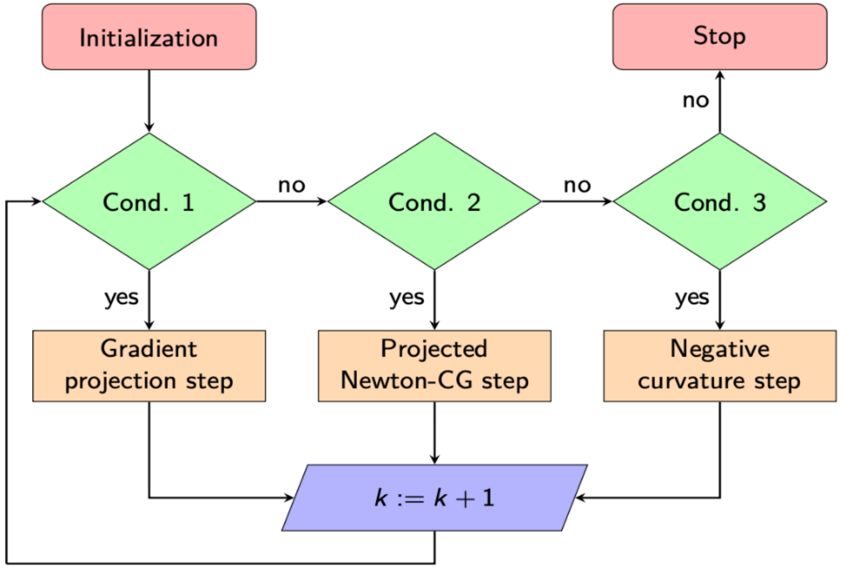
Iterative - $\{ x_k \}_{k\geq 1}, \quad x_{k+1} = T(x_k, \ldots, x_1)$.
Convergence – correctness
Speed - efficiency

My research interest: Algorithm design and analysis to solve nonconvex and stochastic optimization problems with applications in data science.

- **For various subclasses of nonconvex optimization problems, proposed practical second-order algorithms with good theoretical guarantees.**
- **For a class of expected valued convex optimization problems, proposed an efficient stochastic algorithm with ideal sample complexities.**

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x^i \geq 0, \quad i \in \mathcal{I}, \end{aligned}$$

Projected Newton-CG



$$\min \ f(x) \quad \text{subject to} \quad c(x) = 0,$$

$$\min_{x \in \mathcal{X}, y \in \mathcal{Y}} \mathbb{E}[\tilde{f}(x,\xi)] + \mathbb{E}[\tilde{g}(y,\xi)] \quad \text{subject to } Ax + By = b,$$

Algorithm 2 Proximal augmented Lagrangian (Proximal AL)

0. Initialize x_0, λ_0 and $\rho > 0$, $\beta > 0$; Set $k \leftarrow 0$;
1. Update x_k: Find approximate solution x_{k+1} to $\min_x \mathcal{L}_\rho(x, \lambda_k) + \frac{\beta}{2}\|x - x_k\|^2$;
2. Update λ_k: $\lambda_{k+1} \leftarrow \lambda_k + \rho c(x_{k+1})$;
3. If termination criterion is satisfied, STOP; otherwise, $k \leftarrow k+1$ and return to Step 1.

Algorithm 2 SI-ADMM: A stoch. inexact ADMM scheme

(0) Choose Q, P and sequences $\{T^y_{k+1}, T^x_{k+1}\}_{k \geq 0}$ as number of samples generated for updates (16) and (17); Given positive scalars $\rho, \gamma, \gamma_x, \gamma_y$ and initial points x_0, y_0, λ_0; Let $k := 0$;

(1) Let $x_{k+1}, y_{k+1}, \lambda_{k+1}$ be given by the following:

$$y_{k,j+1} := y_{k,j} - \frac{\gamma_y}{j}[\nabla_y \tilde{\mathcal{L}}_2(x_k, y_{k,j}, \lambda_k, \xi^y_{k,j}) + Q(y_{k,j} - y_k)],$$
$$j = 1, \ldots, T^y_{k+1} - 1, y_{k,1} := y_k,$$
$$\boldsymbol{y_{k+1} := y_{k,T^y_{k+1}}} \tag{16}$$

$$x_{k,j+1} := x_{k,j} - \frac{\gamma_x}{j}[\nabla_x \tilde{\mathcal{L}}_1(x_{k,j}, y_{k+1}, \lambda_k, \xi^x_{k,j}) + P(x_{k,j} - x_k)],$$
$$j = 1, \ldots, T^x_{k+1} - 1, x_{k,1} := x_k,$$
$$\boldsymbol{x_{k+1} := x_{k,T^x_{k+1}}} \tag{17}$$

$$\lambda_{k+1} := \lambda_k - \gamma\rho(Ax_{k+1} + By_{k+1} - b). \tag{18}$$

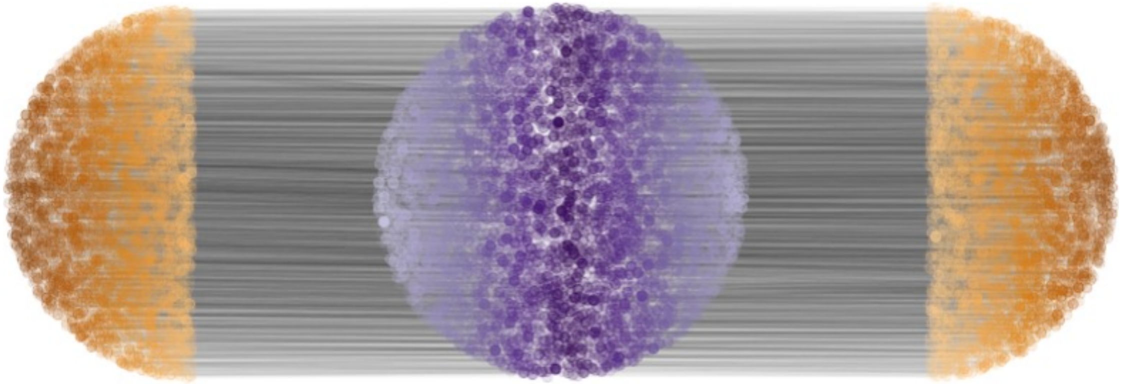
(2) $k := k+1$, return to (1).

✓ **Yue Xie and Stephen J. Wright, Complexity of a Projected Newton-CG Method for Optimization with Bounds, arXiv preprint, arXiv: 2103.15989, Mathematical Programming under 2nd round review.**
✓ **Yue Xie and Stephen J. Wright, Complexity of proximal augmented Lagrangian for nonconvex optimization with nonlinear equality constraints Journal of Scientific Computing, 2021.**
✓ **Yue Xie and Uday V. Shanbhag, SI-ADMM: A stochastic inexact ADMM framework for stochastic convex programs, IEEE Transactions on Automatic Control, vol. 65, no. 6, pp. 2355-2370.**

Optimal Transport

Optimal transport is broadly applied in data science and decision science.

Monge 1781



Gerber, Samuel, and Mauro Maggioni. "Multiscale strategies for computing optimal transport." *arXiv preprint arXiv:1708.02469* (2017).

Linear Programming problem of optimal transport

$$\min_X \langle C, X \rangle \quad \text{subject to} \quad X \cdot \vec{\mathbf{1}} = r_1, X^T \cdot \vec{\mathbf{1}} = r_2, X \geq 0.$$

- Simplex method; Interior point method
- Design algorithms to solve large scale problems with good practical performance and theoretical guarantees.

HKU Musketeers Foundation
Institute of Data Science
香港大學同心基金數據科學研究院

I am hiring Research Assistants and Excellent PhD students. Students with strong mathematical background (linear algebra, mathematical analysis and probability) are especially welcome!

My Webpage: https://yue-xie.github.io/

My Email address: **yxie21@hku.hk**

Dr. Yue Xie

RAP @ HKU

Welcome to my website! Currently I am a research assistant professor in Musketeers Foundation Institute of Data Science and Department of Mathematics at the University of Hong Kong. Before that, I was a postdoc at Wisconsin Intitute for Discovery and lucky to work with Professor Stephen J. Wright. I received my PhD degree in Pennsylvania State University when I was fortunate to have Professor Uday V. Shanbhag as my supervisor and thesis advisor. My research interests are continuous, stochastic and robust optimization, with all types of applications including machine learning and data science. You may find more about me in my CV.

I am hiring postdocs and research assistants at University of Hong Kong. Outstanding candidates for Phd position are also welcome! Feel free to contact me (yxie21@hku.hk) if you are interested!