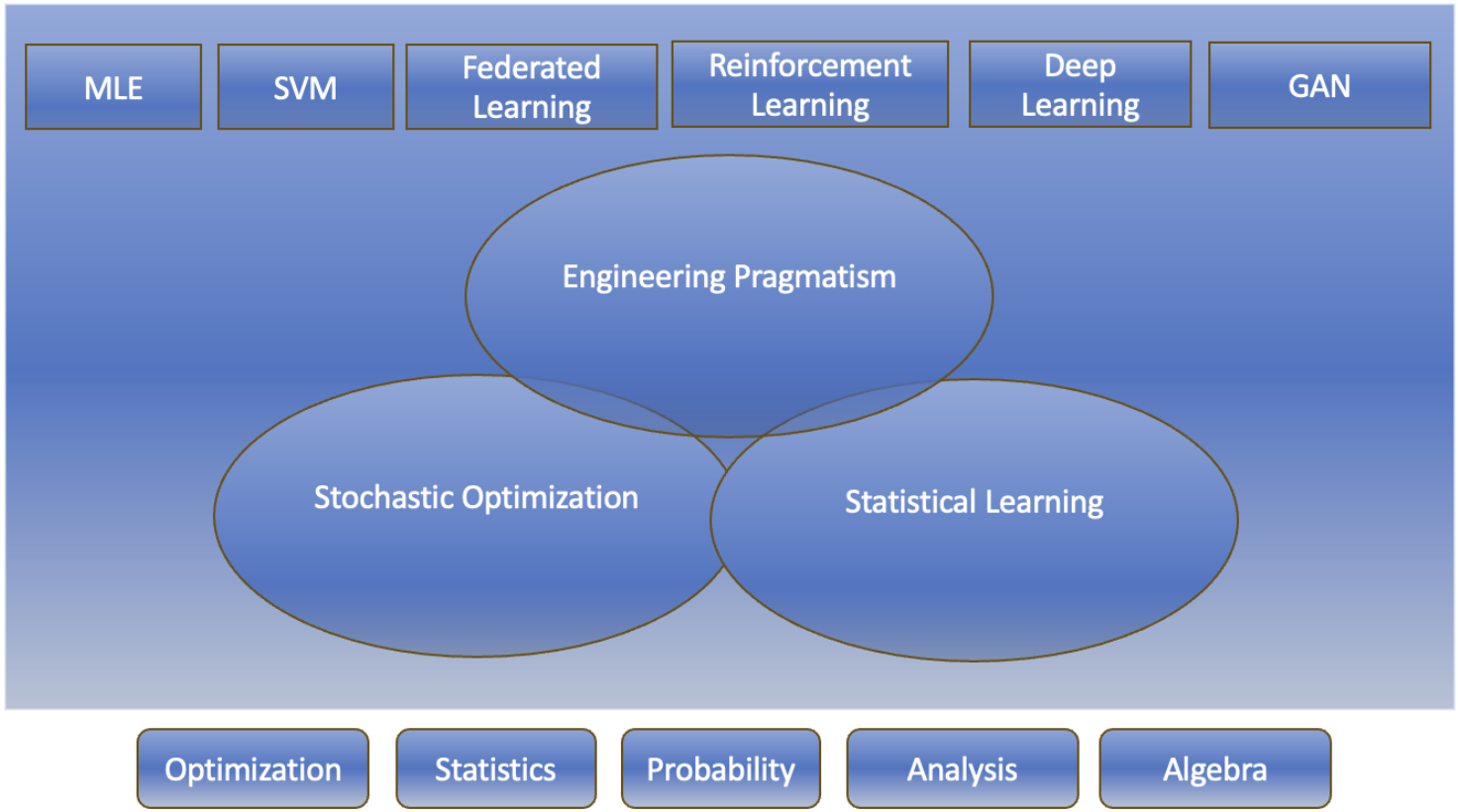
Stochastic Optimization for Machine Learning and Artificial Intelligence

George Lan

H. Milton School of Industrial and Systems Engineering
Georgia Institute of Technology, Atlanta, GA, USA

The 17th ICSP Tutorial Session
July 26, 2025

Stochastic Optimization (SO) for ML/AI



Origin of Stochastic Optimization

Birth of Stochastic Programming
G.B. Dantzig, Linear programming under uncertainty, Management Science 1(3), 197-206, 1955
Foundation for Optimization Involving Uncertain Parameters

Birth of Dynamic Programming
R. Bellman, On the theory of dynamic programming, PNAS 38(8), 716-719, 1952
Foundation for Markov Decision Process, Reinforcement Learning, and Stochastic Optimal Control

Birth of Stochastic Approximation
H. Robbins and S. Monro, A Stochastic Approximation Method, Annals of Mathematical Statistics, 27(3), 400-407, 1951
Foundation for Iterative Stochastic Optimization Methods

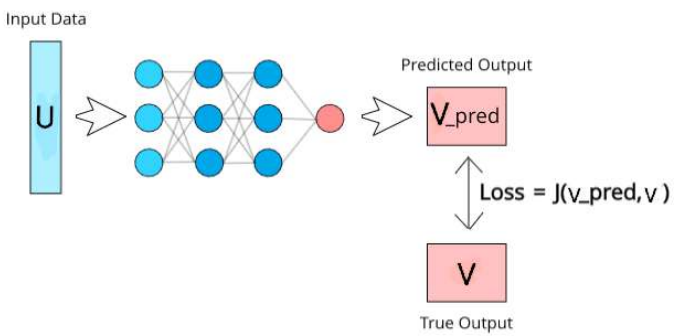
Important SO Models in ML/AI

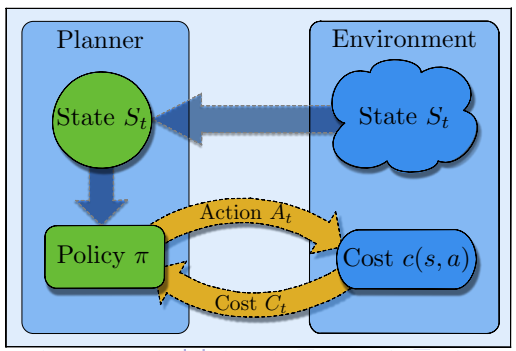
Supervised Learning

- Lasso regression:
  $\min_{\theta} \mathbb{E}[(\langle \theta, u \rangle - v)^2] + \rho \|\theta\|_1$.
- Support vector machine:
  $\min_{\theta} \mathbb{E}_{u,v} [\max\{0, v\langle \theta, u \rangle] + \rho \|\theta\|_2^2$.
- Deep learning:
  $\min_{\Theta} \mathbb{E}_{u,v}[F_d(\Theta_d \ldots F_1(\Theta_1 u)) - v]^2$.

Reinforcement Learning

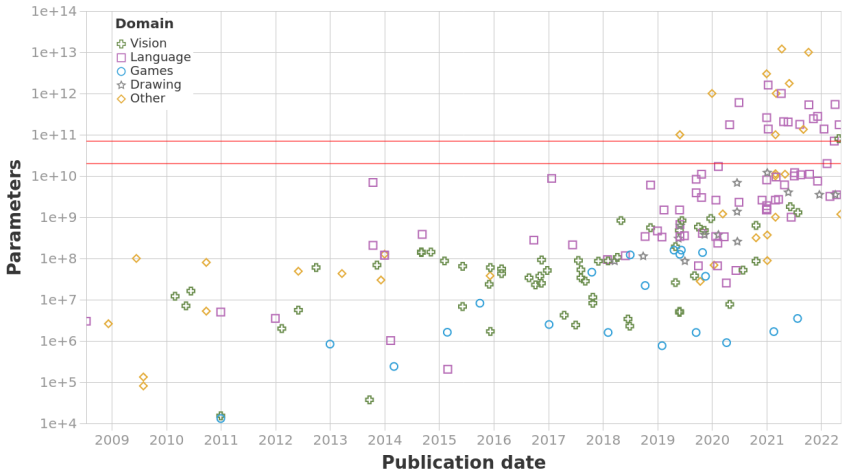
- State space: $\mathcal{S}$
- Action space: $\mathcal{A}$
- Policy: π
- Transition kernel: $\mathcal{P}$
- $\min_{\pi} \mathbb{E}_{\pi, s_0}[\sum_{t=0}^{\infty} \gamma^t c_t | s_0]$.

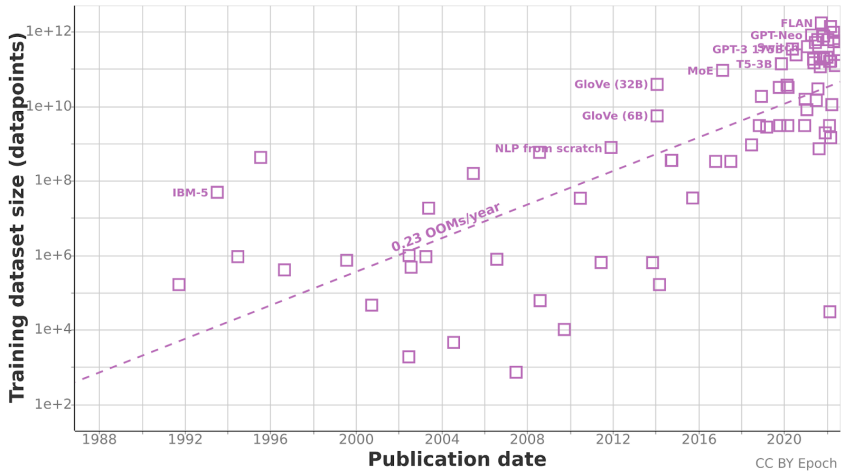




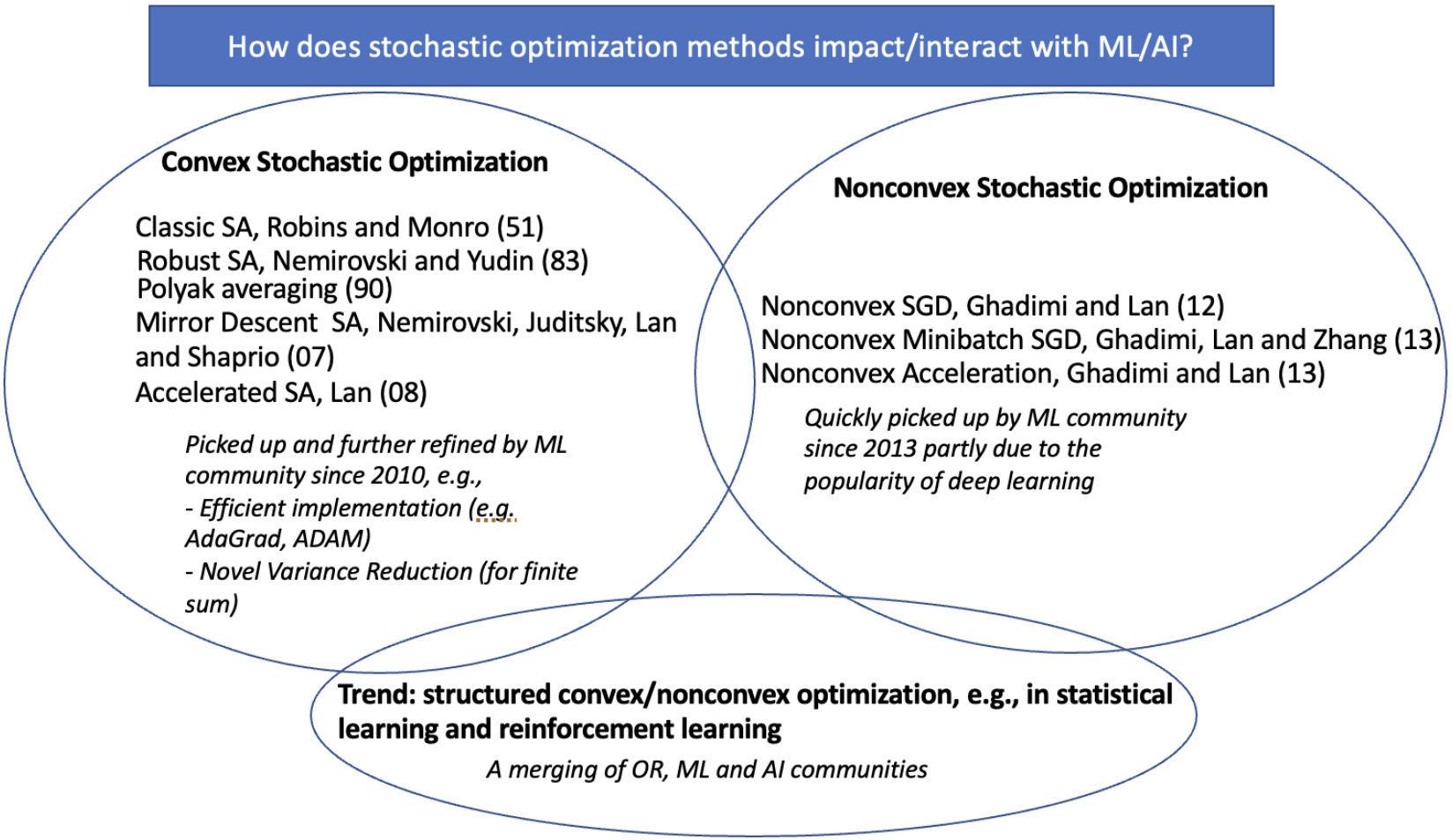
Common Challenges for SO in ML/AI

High Dimension ($\sim 10^{13}$); Large Data volume ($\sim 10^{12}$)





Stochastic Optimization Methods



G. Lan, First-order and Stochastic Optimization Methods for Machine Learning, Springer-Nature, 2020.

Outline

- Deterministic first-order methods
  - Gradient/mirror descent
  - Accelerated gradient descent
- Convex stochastic methods
  - Stochastic gradient/mirror descent
  - Accelerated stochastic gradient descent
- Nonconvex stochastic methods
  - Nonconvex first-order methods
  - Stochastic zeroth-order methods
  - Acceleration for nonconvex problems
- Structured nonconvex optimization
  - Reinforcement learning
  - Policy mirror descent method
- Current/future research directions

(Sub)Gradient descent

Problem

$f^* := \min_{x \in X} f(x)$. Here $\emptyset \neq X \subset \mathbb{R}^n$ is a convex set and f is a convex function.

Basic Idea

Starting from $x_1 \in \mathbb{R}^n$, update x_t by $x_{t+1} = x_t - \gamma_t \nabla f(x_t), t = 1, 2, \ldots$.

Two essential enhancements

- f may be non-differentiable: replace $\nabla f(x_t)$ by $g(x_t) \in \partial f(x_t)$.
- x_{t+1} may be infeasible: project back to X.

$x_{t+1} := \text{argmin}_{x \in X} \|x - (x_t - \gamma_t g(x_t))\|_2, t = 1, 2, \ldots$.

Interpretation

From the proximity control point of view,

$$\begin{aligned} x_{t+1} &= \operatorname{argmin}_{y \in X} \tfrac{1}{2}\|x - (x_t - \gamma_t g(x_t))\|_2^2 \\ &= \operatorname{argmin}_{x \in X} \gamma_t \langle g(x_t), x - x_t \rangle + \tfrac{1}{2}\|x - x_t\|_2^2 \\ &= \operatorname{argmin}_{x \in X} \gamma_t \left[f(x_t) + \langle g(x_t), x - x_t \rangle \right] + \tfrac{1}{2}\|x - x_t\|_2^2 \\ &= \operatorname{argmin}_{x \in X} \gamma_t \langle g(x_t), x \rangle + \tfrac{1}{2}\|x - x_t\|_2^2. \end{aligned}$$

Implication

To minimize the linear approximation $f(x_t) + \langle g(x_t), x - x_t \rangle$ of $f(x)$ over X, without moving too far away from x_t.

The role of stepsize

γ_t controls how much we trust the model and depends on which problem classes to be solved.

Convergence of (Sub)Gradient Descent (GD)

Nonsmooth problems

f is M-Lipschitz continuous, i.e., $|f(x) - f(y)| \leq M\|x - y\|_2$.

Theorem

Let $\bar{x}_s^k := \left(\sum_{t=s}^k \gamma_t\right)^{-1} \sum_{t=s}^k (\gamma_t x_t)$. *Then*

$$f(\bar{x}_s^k) - f^* \leq \left(2\sum_{t=s}^k \gamma_t\right)^{-1} \left[\|x^* - x_s\|_2^2 + M^2 \sum_{t=s}^k \gamma_t^2\right].$$

Selection of γ_t

- If $\gamma_t = \sqrt{D_X^2/(kM^2)}$, for some fixed k, then $f(\bar{x}_1^k) - f^* \leq \frac{MD_X}{2\sqrt{k}}$, where $D_X \geq \max_{x_1, x_2 \in X} \|x_1 - x_2\|$.
- If $\gamma_t = \sqrt{D_X^2/(tM^2)}$, then $f(\bar{x}_{\lceil k/2 \rceil}^k) - f^* \leq \mathcal{O}(1)(MD_X/\sqrt{k})$.

Convergence of GD

Smooth problems

f is differentiable, and ∇f is L-Lipschitz continuous, i.e.,

$$|\nabla f(x) - \nabla f(y)| \leq L\|x - y\|_2.$$

$$\exists \mu \geq 0 \text{ s.t. } f(y) \geq f(x) + \langle f'(x), y - x\rangle + \tfrac{1}{2}\mu\|y - x\|^2.$$

Theorem

Let $\gamma \in (0, \frac{2}{L}]$. *If* $\gamma_t = \gamma$, $t = 1, 2, \ldots$, *then*

$$f(x_k) - f^* \leq \frac{\|x_0 - x^*\|^2}{k\gamma(2 - L\gamma)}.$$

Moreover, if $\mu > 0$ *and* $Q_f = L/\mu$, *then*

$$\|x_k - x^*\|^2 \leq \left(\frac{Q_f - 1}{Q_f + 1}\right)^k \|x_0 - x^*\|^2,$$

$$f(x_k) - f^* \leq \frac{L}{2}\left(\frac{Q_f - 1}{Q_f + 1}\right)^{2k} \|x_0 - x^*\|^2.$$

Adaption to Geometry: Mirror Descent (MD)

GD is intrinsically linked to the Euclidean structure of $\mathbb{R}^n$:

- The method relies on the Euclidean projection,
- D_X, L, μ, and M are defined in terms of the Euclidean norm.

Bregman Distance

Let $\|\cdot\|$ be a (general) norm on $\mathbb{R}^n$ and $\|x\|_* = \sup_{\|y\|\leq 1}\langle x, y\rangle$, and ω be a continuously differentiable and strongly convex function with modulus ν with respect to $\|\cdot\|$. Define

$$V(x,z) = \omega(z) - [\omega(x) + \nabla\omega(x)^T(z-x)].$$

Mirror Descent (Nemirovski and Yudin 83, Beck and Teboule 03)

$$x_{t+1} = \operatorname{argmin}_{x\in X}\gamma_t\langle g(x_t), x\rangle + V(x_t, x), t = 1, 2, \ldots.$$

GD is a special case of MD: $V(x_t, x) = \|x - x_t\|_2^2/2$.

Acceleration scheme for smooth problems

Nesterov's Accelerated gradient descent (AGD)

1) Set $\underline{x}_k = (1-\alpha_k)\bar{x}_{k-1} + \alpha_k x_{k-1}$.
2) Compute $f'(\underline{x}_k)$ and set
$$x_k = \operatorname{argmin}_{x\in X}\{\alpha_k\langle f'(\underline{x}_k), x\rangle + \beta_k V(x_{k-1}, x)\},$$
$$\bar{x}_k = (1-\alpha_k)\bar{x}_{k-1} + \alpha_k x_k.$$
3) Set $k \leftarrow k+1$ and go to step 1).

Theorem

If $\beta_k \geq L\alpha_k^2$ *and* $\beta_k = (1-\alpha_k)\beta_{k-1}$ *for* $k \geq 1$, *then*
$$f(\bar{x}_k) - f(x) \leq \frac{1-\alpha_1}{\beta_1}[f(\bar{x}_0) - f(x)] + \beta_k V(x_0, x),\ \forall x \in X.$$
In particular, If $\alpha_k = 2/(k+1)$ *and* $\beta_k = 4L/[k(k+1)]$, *then*
$$f(\bar{x}_k) - f(x^*) \leq \frac{4L}{\nu k(k+1)} V(x_0, x^*).$$

Note: Geometric intuition exists for some precursors in Nemirovski and Yudin (77,83).

Acceleration scheme (strongly convex problems)

Assumption: $\exists \mu > 0$ s.t. $f(y) \geq f(x) + \langle f'(x), y - x \rangle + \frac{1}{2}\mu \|y - x\|^2$.

AGD for strongly convex problems

Idea: Restart AGD every $N \equiv \left\lceil \sqrt{8L/\mu} \right\rceil$ iterations.

The algorithm.

Input: $p_0 \in X$.

Phase $t = 1, 2, \ldots$:

Set $p_t = \bar{x}_N$, where $\bar{x}_N$ is obtained from AGD with $x_0 = p_{t-1}$.

Theorem

For any $t \geq 1$, we have $\|p_t - x^*\|^2 \leq (\frac{1}{2})^t \|p_0 - x^*\|^2$.

To have $\|p_t - x^*\|^2 \leq \epsilon$, the total number of iterations is bounded by

$$\left\lceil \sqrt{\frac{8L}{\mu}} \right\rceil \log \frac{\|p_0 - x^*\|^2}{\epsilon}$$

Stochastic optimization problems

The Problem: $\min_{x \in X} \{f(x) := \mathbb{E}_\xi[F(x, \xi)]\}$.

Challenges: To compute exact (sub)gradients is computationally prohibitive.

Stochastic oracle

At iteration t, $x_t \in X$ being the input, $\mathcal{SO}$ outputs a vector $G(x_t, \xi_t)$, where $\{\xi_t\}_{t \geq 1}$ are i.i.d. random variables s.t.

$$\mathbb{E}[G(x_t, \xi_t)] \equiv g(x_t) \in \partial \Psi(x_t).$$

Examples:

- $f(x) = \mathbb{E}_\xi[F(x, \xi)]$: $G(x_t, \xi_t) = F'(x_t, \xi_t)$, ξ_t being a random realization of ξ.
- $f(x) = \sum_{i=1}^m f_i(x)/m$: $G(x_t, i_t) = f'_{i_t}(x_t)$, i_t being a uniform random variable on $\{1, \ldots, m\}$.

Stochastic mirror descent

The algorithm: Replace the exact linear model in MD with its stochastic approximation (goes back to Robinson and Monro 51, Nemirovski and Yudin 83).

$$x_{t+1} = \operatorname{argmin}_{x \in X} \gamma_t \langle G_t, x \rangle + V(x_t, x), t = 1, 2, \ldots$$

Theorem (Nemirovski, Juditsky, Lan and Shapiro 07 (09))

Assume $\mathbb{E}[\|G(x,\xi)\|_*^2] \leq M^2$.

$$\mathbb{E}[f(\bar{x}_s^k)] - f^* \leq \left(\sum_{t=s}^k \gamma_t\right)^{-1} \left[\mathbb{E}[V(x_s, x^*)] + (2\nu)^{-1} M^2 \sum_{t=s}^k \gamma_t^2\right].$$

The selection of γ_t

- If $\gamma_t = \sqrt{\Omega^2/(kM^2)}$ for some fixed k, then $\mathbb{E}[f(\bar{x}_1^k) - f^*] \leq \frac{M\Omega}{2\sqrt{k}}$, where $\Omega \geq \max_{x_1, x_2 \in X} V(x_1, x_2)$.
- If $\gamma_t = \sqrt{\Omega^2/(tM^2)}$, then $\mathbb{E}[f(\bar{x}_{\lceil k/2 \rceil}^k) - f^*] \leq \mathcal{O}(1)(M\Omega/\sqrt{k})$.

Complexity?

Stochastic Optimization: $\min_{x\in X} \mathbb{E}[F(x,\xi)]$,

- One $F'(x_t,\xi_t)$ per iteration, totally $\mathcal{O}(1/\epsilon^2)$ iterations.
- Optimal sampling complexity.

Deterministic Finite-sum Optimization:

$$\min_{x\in X} \left\{ f(x) := \frac{1}{m}\sum_{i=1}^m f_i(x) \right\}$$

$$|f_i(x) - f_i(y)| \le M_i\|x-y\|, |f(x)-f(y)| \le \mathcal{M}\|x-y\|, \forall x,y \in X$$

$$\mathcal{M} \le M \equiv \max_i M$$

	Iteration complexity	iteration cost
MD	$\frac{\mathcal{M}^2\Omega^2}{\epsilon^2} \le \frac{M^2\Omega^2}{\epsilon^2}$	m subgradients
SMD	$\frac{M^2\Omega^2}{\epsilon^2}$	1 subgradients

SMD saves up to $\mathcal{O}(m)$ subgradient computations.

Explanation and Extension

Examples of favorable geometry

- $X = \{x \in \mathbb{R}^n : \|x\|_2 \leq 1\}$, $\mathbb{E}\left[\|\mathcal{G}(x_t, \xi_t)\|_2^2\right] \leq 1$.
  Sample complexity: $\mathcal{O}(1/\epsilon^2)$
- $X = \{x \in \mathbb{R}^n : \sum_i x_i = 1, x_i \geq 0\}$, $\mathbb{E}\left[\|\mathcal{G}(x_t, \xi_t)\|_\infty^2\right] \leq 1$.
  Sample complexity: $\mathcal{O}(\log(n)/\epsilon^2)$

Extension to saddle point optimization

$\min_{x \in X} \max_{y \in Y} \mathbb{E}[F(x, y, \xi)]$

- Stochastic subgradient: $(F'_x(x, y, \xi), -F'_y(x, y, \xi))$
- Iteration/sample complexity: $\mathcal{O}((M^2 D^2_{X \times Y})/\epsilon^2)$

Accelerated SGD (SGD with momentum)

Consider

$$\Psi^* := \min_{x \in X} \{\Psi(x) := f(x) + \mathcal{X}(x)\}.$$

- $\mathcal{X}(x)$ is a simple convex function with known structure (e.g. $\mathcal{X}(x) = 0$, $\mathcal{X}(x) = \|x\|_1$).
- $\exists\, L \geq 0, M \geq 0$
$$0 \leq f(y) - f(x) - \langle f'(x), y - x \rangle \leq \frac{L}{2}\|y - x\|^2 + M\|y - x\|.$$
- f is represented by an $\mathcal{SO}$, which, given input $x_t \in X$, outputs $G(x_t, \xi_t)$ s.t.
$$\mathbb{E}[G(x_t, \xi_t)] \equiv g(x_t) \in \partial f(x_t), \mathbb{E}\left[\|G(x_t, \xi_t) - g(x_t)\|_*^2\right] \leq \sigma^2.$$
- Covers both non-smooth minimization ($L = 0$) and smooth minimization ($M = 0$);
- Covers both deterministic case ($\sigma = 0$) and stochastic case ($\sigma \neq 0$).

Accelerated SGD (SGD with momentum)

Accelerated SGD

Choose $\bar{x}_0 = x_0 \in X$.

1) Set $\underline{x}_k = (1-\alpha_k)\bar{x}_{k-1} + \alpha_k x_{k-1}$.

2) Compute $G(\underline{x}_k, \xi_k)$ and set

$$x_k = \operatorname{argmin}_{x \in X}\{\alpha_k[\langle G(\underline{x}_k, \xi_k), x\rangle + \mathcal{X}(x)] + \beta_k V(x_k, x)\},$$
$$\bar{x}_k = (1-\alpha_k)\bar{x}_{k-1} + \alpha_k x_k.$$

3) Set $k \leftarrow k+1$ and go to step 1).

Why was it difficult? AGD was originally designed for deterministic smooth problems only.

- Challenge I: a unified analysis for smooth, nonsmooth and stochastic optimization was missing.
- Challenge II: does not exist a proper stepsize policy.
  - Divergence from aggressive stepsize ($\gamma_k \equiv \alpha_k/\beta_k = \mathcal{O}(k)$).

Convergence Results

Theorem

If

$$\alpha_t = \frac{2}{t+1} \quad \text{and} \quad \beta_t = \frac{4\beta}{\nu t(t+1)}, \quad \forall\, t = 1, \ldots, k,$$

for some $\beta \geq 2L$, *then*

$$\mathbb{E}[\Psi(\bar{x}_k) - \Psi^*] \leq \frac{4\beta V(x_0, x^*)}{\nu k(k+1)} + \frac{4(M^2+\sigma^2)(k+2)}{3\beta}.$$

In particular, if k is fixed in advance and

$$\beta = \beta^* = \max\left\{2L, \left[\frac{\nu(M^2+\sigma^2)k(k+1)(k+2)}{3V(x_0, x^*)}\right]^{\frac{1}{2}}\right\},$$

then

$$\mathbb{E}[\Psi(\bar{x}_k) - \Psi^*] \leq \mathcal{O}(1)\left(\frac{L\Omega^2}{k^2} + \frac{(M+\sigma)\Omega}{\sqrt{k}}\right).$$

Note: It is possible to design a stepsize policy without requiring k given in advance by using different stepsize policy or slightly modifying the algorithm.

Remarks

- A universally optimal method for non-smooth, smooth, and composite minimization, allowing stochastic (sub)gradients
- The method can allow a very large Lipschitz constant L without affecting the rate of convergence.

SGD	$\frac{L\Omega^2+(M+Q)\Omega}{\sqrt{k}}$	$L \leq \frac{M+Q}{\Omega}$
Accelerated SGD	$\frac{L\Omega^2}{k^2} + \frac{(M+Q)\Omega}{\sqrt{k}}$	$L \leq \frac{(M+Q)k^{\frac{3}{2}}}{\Omega}$

Strongly convex problems

$\exists \mu > 0$ s.t. $f(y) - f(x) - \langle f'(x), y - x \rangle \geq \frac{\mu}{2}\|y - x\|^2$.

Theorem: Properly restarting the algorithm, we will find an ϵ-solution $\bar{x} \in X$ s.t. $\mathbb{E}[\Psi(\bar{x}) - \Psi^*] \leq \epsilon$ in at most $K := \lceil \log \mathcal{V}_0/\epsilon \rceil$ stages. Moreover, the total number of iterations performed by this algorithm to find such a solution is bounded by

$$\mathcal{O}(1)\left\{\sqrt{\frac{L}{\mu}} \max\left(1, \log \frac{\mathcal{V}_0}{\epsilon}\right) + \frac{M^2+\sigma^2}{\mu\epsilon}\right\}.$$

- The iteration-complexity is optimal for strongly convex stochastic optimization.

Nonconvex SP problems

Issue: All previous stochastic algorithms require the convexity assumption to show convergence.
Problem: Consider $f^* := \min_{x\in\mathbb{R}^n} f(x)$.

- f is differentiable and bounded below, $f \in C_L^{1,1}(\mathbb{R}^n)$: $\|\nabla f(y) - \nabla f(x)\| \leq L\|y - x\|, \forall x, y \in \mathbb{R}^n$.
- f is not necessarily convex.
- Goal: computing an approximate stationary point.
- Well-understood in the deterministic case.

Only stochastic gradient $G(x_k, \xi_k)$ is available at x_k.

a) $\mathbb{E}[G(x_k, \xi_k)] = g(x_k) \equiv \nabla f(x_k)$,
b) $\mathbb{E}\left[\|G(x_k, \xi_k) - g(x_k)\|^2\right] \leq \sigma^2$.

No convergence analysis for nonconvex SGD before our work.

Motivating examples

- Nonconvex loss or regularization (e.g., deep learning):
  $$f(x) = \int_{\Xi} L(x, \xi) dP(\xi) + r(x),$$
  where either the loss function $L(x, \xi)$ or the regularization term $r(x)$ is nonconvex, $G(x, \xi) = \nabla_x L(x, \xi) + \nabla r(x)$.
- Simulation-based optimization
  $$f(x) = \mathbb{E}_{\xi}[F(x, \xi)].$$
  Here $F(x, \xi)$ is given in a black box. Usually only noisy function values available. No convexity information.
- Endogenous uncertainty:
  $$f(x) = \int_{\Xi(x)} F(x, \xi) dP_x(\xi).$$
  Here $f(x)$ is usually nonconvex even though $F(x, \xi)$ is convex w.r.t. x.

The randomized stochastic gradient method

Algorithm 1 A randomized stochastic gradient (RSG) method

Input: Initial point x_1, iteration limit N, stepsizes $\{\gamma_k\}_{k\geq 1}$ and probability mass function $P_R(\cdot)$ supported on $\{1,\ldots,N\}$.

Step 0. Let R be a random variable with probability mass function P_R.

Step $k = 1,\ldots,R$. Compute $G(x_k,\xi_k)$ and set

$$x_{k+1} = x_k - \gamma_k G(x_k,\xi_k).$$

Output x_R.

How would randomization help?

By the smoothness of f:

$$f(x_{k+1}) \leq f(x_k) - \gamma_k \langle g(x_k), G(x_k, \xi_k) \rangle + \frac{L\gamma_k^2}{2} \|G(x_k, \xi_k)\|^2.$$

Taking expectation on both sides w.r.t. ξ_k, conditioning on $(\xi_1, \ldots, \xi_{k-1})$:

$$\mathbb{E}[f(x_{k+1})] \leq f(x_k) - \left(\gamma_k - \tfrac{L}{2}\gamma_k^2\right) \|g(x_k)\|^2 + \tfrac{L}{2}\gamma_k^2 \sigma^2.$$

Applying the above relation inductively, we obtain:

$$\begin{aligned} &\tfrac{1}{L} \textstyle\sum_{k=1}^N \left(\gamma_k - \tfrac{L}{2}\gamma_k^2\right) \mathbb{E}[\|g(x_k)\|^2] \\ &\quad \leq \tfrac{1}{L} [f(x_1) - \mathbb{E}[f(x_{N+1})]] + \tfrac{1}{2}\sigma^2 \textstyle\sum_{k=1}^N \gamma_k^2 \\ &\quad \leq \underbrace{\tfrac{1}{L}[f(x_1) - f^*]}_{D_f} + \tfrac{1}{2}\sigma^2 \textstyle\sum_{k=1}^N \gamma_k^2. \end{aligned}$$

The randomized stochastic gradient method

Theorem: Suppose that $P_R(\cdot)$ and $\{\gamma_k\}$ are set to

$$P_R(k) := \text{Prob}\{R = k\} = \frac{2\gamma_k - L\gamma_k^2}{\sum_{k=1}^N (2\gamma_k - L\gamma_k^2)}, \quad k = 1, ..., N,$$

$$\gamma_k = \min\left\{\frac{1}{L}, \frac{\tilde{D}}{\sigma\sqrt{N}}\right\}, k = 1, \ldots, N,$$

for some $\tilde{D} > 0$. Then,

$$\frac{1}{L}\mathbb{E}[\|\nabla f(x_R)\|^2] \leq \frac{LD_f^2}{N} + \left(\tilde{D} + \frac{D_f^2}{\tilde{D}}\right)\frac{\sigma}{\sqrt{N}}.$$

If, in addition, f is convex with an optimal solution x^*, then

$$\mathbb{E}[f(x_R) - f^*] \leq \frac{LD_X^2}{N} + \left(\tilde{D} + \frac{D_X^2}{\tilde{D}}\right)\frac{\sigma}{\sqrt{N}},$$

where $D_X := \|x_1 - x^*\|$.

Large-deviation properties for the RSG method

Goal: To find an (ϵ, Λ)-solution $\bar{x} \in \mathcal{E}$ in a single run of the RSG algorithm s.t. $\text{Prob}\{\|\nabla f(\bar{x})\|^2 \leq \epsilon\} \geq 1 - \Lambda$ for some $\Lambda \in (0, 1)$.

- By Markov's inequality, iteration-complexity of finding an (ϵ, Λ)-solution: $\mathcal{O}\left\{\frac{L^2 D_f^2}{\Lambda \epsilon} + \frac{L^2}{\Lambda^2}\left(\tilde{D} + \frac{D_f^2}{\tilde{D}}\right)^2 \frac{\sigma^2}{\epsilon^2}\right\}$.

A two-phase randomized stochastic gradient method

Algorithm 2 A two-phase RSG (2-RSG) method

Input: Initial point x_1, iteration limit N, sample size T, and confidence level $\Lambda \in (0,1)$.

Initialize: Set $S = \lceil \log 2/\Lambda \rceil$.

Optimization phase:

For $s = 1, \ldots, S$:

Call the RSG method with input x_1, iteration limit N, stepsizes $\{\gamma_k\}$ and probability mass function P_R same as in RSG method. Let $\bar{x}_s$ be the output of this procedure.

Post-optimization phase: Choose $\bar{x}^*$ from $\{\bar{x}_1, \ldots, \bar{x}_S\}$ s.t. $\|g(\bar{x}^*)\| = \min_{s=1,\ldots,S} \|g(\bar{x}_s)\|$, $g(\bar{x}_s) := \frac{1}{T}\sum_{k=1}^{T} G(\bar{x}_s, \xi_k)$.

A two-phase randomized stochastic gradient method

- Iteration complexity of 2-RSG method for finding (ϵ, Λ)-solution:
$$\mathcal{O}\left\{\frac{L^2 D_f^2 \log(1/\Lambda)}{\epsilon} + L^2\left(\tilde{D} + \frac{D_f^2}{\tilde{D}}\right)^2 \log(1/\Lambda)\frac{\sigma^2}{\epsilon^2} + \frac{\log^2(1/\Lambda)\sigma^2}{\Lambda\epsilon}\right\}.$$
- When the second terms are the dominating ones in both bounds, considerably smaller than the previous one up to a factor of
$$\mathcal{O}\left\{1/(\Lambda^2 \log^2(1/\Lambda))\right\}.$$
- The third term can be improved by a factor of $1/\Lambda$ under certain light-tail assumptions about ξ_k.

Stochastic zeroth-order oracle

Only stochastic function values are available:

- At iteration k, $x_k \in \mathcal{E}$ being the input, one can compute $F(x_k, \xi_k)$ s.t.

$$\mathbb{E}[F(x_k, \xi_k)] = f(x_k)$$

- Derivative-free methods (Conn et al. 09, Rios and Sahinidis 13)
  - Limited study on the complexity issues (Nemirovski and Yudin 83, Nesterov 11) focused on the convex case only.

Gaussian smooth approximation

- Use a approximation of f whose gradient is computable by function values.
- Gaussian smooth approximation of f is given by

$$f_\mu(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathcal{E}} f(x+\mu u) e^{-\frac{1}{2}\|u\|^2}\, du = \mathbb{E}_u[f(x+\mu u)].$$

- Gradient of f_μ is given by

$$\nabla f_\mu(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathcal{E}} \frac{f(x+\mu u)-f(x)}{\mu} u e^{-\frac{1}{2}\|u\|^2}\, du.$$

More on Gaussian smoothing

- f_μ has a Lipschitz continuous gradient with constant $L_\mu \leq L$,
- $\|\nabla f(x)\|^2 \leq 2\|\nabla f_\mu(x)\|^2 + \frac{\mu^2}{2} L^2 (n+4)^2$,
- $|f_\mu(x) - f(x)| \leq \frac{\mu^2}{2} Ln$,

A randomized stochastic gradient free (RSGF) method

Algorithm 3 A two-phase RSG (2-RSG) method

Input: Initial point x_1, iteration limit N, stepsizes $\{\gamma_k\}_{k\geq 1}$, probability mass function $P_R(\cdot)$ supported on $\{1,\ldots,N\}$.

Step 0. Let R be a random variable with probability mass function P_R.

Step $k = 1,\ldots,R$. Generate a Gaussian random vector u_k and call the stochastic zeroth-order oracle for computing $G_\mu(x_k,\xi_k)$ given by

$$G_\mu(x_k,\xi_k) = \frac{F(x_k+\mu u_k,\xi_k)-F(x_k,\xi_k)}{\mu} u_k,$$

set

$$x_{k+1} = x_k - \gamma_k G_\mu(x_k,\xi_k)$$

Output x_R.

Convergence of the RSGF method

Theorem: Suppose that the stepsizes $\{\gamma_k\}$, the probability mass function $P_R(\cdot)$ and the smoothing parameter μ are set to

$$\gamma_k = \frac{1}{\sqrt{n+4}} \min\left\{\frac{1}{4L\sqrt{n+4}}, \frac{\bar{D}}{\sigma\sqrt{N}}\right\},$$

$$P_R(k) := \mathrm{Prob}\{R = k\} = \frac{\gamma_k - 2L(n+4)\gamma_k^2}{\sum_{k=1}^{N}\left(\gamma_k - 2L(n+4)\gamma_k^2\right)},$$

$$\mu \leq \frac{1}{\sqrt{(n+4)N}} \min\left\{\frac{D_f}{\sqrt{2(n+4)}}, D_X\right\}.$$

Then, under assumption A3, we have

$$\frac{1}{L}\mathbb{E}[\|\nabla f(x_R)\|^2] \leq C(N) := \frac{12(n+4)LD_f^2}{N} + 4\sqrt{n+4}\left[\bar{D} + \frac{D_f^2}{\bar{D}}\right]\frac{\sigma}{\sqrt{N}}.$$

A randomized stochastic gradient free (RSGF) method

Theorem: If the problem is convex with an optimal solution x^*, then,

$$\mathbb{E}[f(x_R) - f^*] \leq \frac{5L(n+4)D_X^2}{N} + 2\sqrt{n+4}\left[\bar{D} + \frac{D_X^2}{\bar{D}}\right]\frac{\sigma}{\sqrt{N}}$$

where $D_X = \|x_1 - x^*\|$.

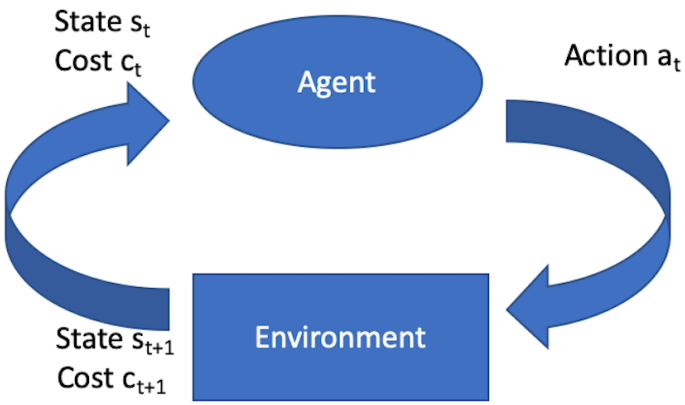
- Improving Nesterov's bound for general CP ($\mathcal{O}(n/\sqrt{N})$.

Further developments

- Stochastic nonconvex composite optimization
- Nonconvex accelerated gradient method
- Finite-sum and distributed optimization methods
  - Variance reduced gradient methods
  - Randomized primal-dual gradient methods
  - Gradient extrapolation method
  - Decentralized communication sliding
- Projection-free methods

See G. Lan, “First-order and Stochastic Optimization Methods for Machine Learning”, Springer, 2020.

Stochastic dynamic programming



- Interaction between agents and environment
- Improve agents' decisions through these interactions
- State: the status of environment
- Action: the behavior of agent
- Optimal policy: the best action at a state

Markov decision process (MDP)

Finite Markov decision process $M = (\mathcal{S}, \mathcal{A}, \mathcal{P}, c, \gamma)$.

- $\mathcal{S}$: a finite state space
- $\mathcal{A}$: a finite action space
- $\mathcal{P} : \mathcal{S} \times \mathcal{S} \times \mathcal{A} \to \mathbb{R}$: transition model
- $c : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$: cost function
- $\gamma \in (0, 1)$: discount factor
- policy $\pi : \mathcal{A} \times \mathcal{S} \to \mathbb{R}$:
  probability of selecting an action at a given state.

In Reinforcement Learning (RL): $\mathcal{P}$ unknown.

- Generative model: can query next state at any (s, a) pair.
- Single-trajectory model: does not permit system restarts at arbitrary state-action pairs; once initiated, the trajectory must be followed sequentially.

Fundamentals of MDP

- Let $\Delta_{\mathcal{A}} := \{p \in \mathbb{R}^{|\mathcal{A}|} : \sum_i p_i = 1, p_i \geq 0, i = 1, \ldots, |\mathcal{A}|\}$.
  Policy $\pi \in \Pi := \underbrace{\Delta_{\mathcal{A}} \times \ldots \Delta_{\mathcal{A}}}_{|\mathcal{S}| \text{ times}}$.
- State-value function:
  $V^{\pi}(s) = \mathbb{E}_{\pi}\left[\sum_{t=0}^{\infty} \underbrace{\gamma^t c_t}_{\text{discounting future costs}} | S_0 = s\right]$.
- Action-value function:
  $Q^{\pi}(s,a) := \mathbb{E}_{\pi}\left[\sum_{t=0}^{\infty} \gamma^t c(s_t, a_t) | s_0 = s, a_0 = a\right]$.
- Objectives: find $\pi^* \in \Pi$ such that
  $V^*(s) \equiv V^{\pi^*}(s) \leq V^{\pi}(s), \forall \pi \in \Pi$ for all $s \in \mathcal{S}$, and
  $Q^*(s,a) \equiv Q^{\pi^*}(s,a) \leq Q^{\pi}(s,a), \forall \pi \in \Pi$ for all $(s,a) \in \mathcal{S} \times \mathcal{A}$.
- Bellman's optimality condition guarantees the existence of π^* that achieves all these goals simultaneously.

Dynamic Optimization Methods

- Value iteration (Bellman 1957, Shapley 1953): iteratively updates the estimate on V^* according to $V_{k+1} = \mathcal{L}V_k$ with $\mathcal{L}V := \min_{\pi\in\Pi}\{c_\pi + \gamma P_\pi V\}$. Here $\mathcal{P}_\pi(s,s') := \sum_{a\in\mathcal{A}}\pi(a|s)\mathcal{P}(s'|s,a), \forall (s,s') \in \mathcal{S}\times\mathcal{S}$.
  - Linear convergence
  - Polynomial time algorithm (Tseng 1990).
  - Stochastic value iteration led by Sidford, Ye and co-authors, 2018.
  - Q-learning: similarly updates the estimate on Q^*.
- Policy iteration (Howard, 1960): iteratively calculates Q^{π_k} and sets $\pi_{k+1}(\cdot|s) \in \mathrm{Argmin}_{p\in\Delta_\mathcal{A}}\langle Q^{\pi_k}(s,\cdot), p\rangle, \forall s \in \mathcal{S}$.
  - Actor-critic algorithms, on-policy learning.
  - Linear convergence
  - Strongly polynomial time algorithm (Ye 2011).

Linear Optimization Methods

- Linear programming formulation (DÉpenoux 63, de Ghellinck 60):

$$\begin{aligned} \min \ & \textstyle\sum_s \sum_{a \in \mathcal{A}} c(s,a) x(s,a) \\ \text{s.t.} \ & \textstyle\sum_a x(s',a) - \sum_s \sum_{a \in \mathcal{A}} \gamma \mathcal{P}(s'|s,a) x(s,a) = 1, \forall s' \\ & x(s,a) \geq 0, \forall s, \forall a \end{aligned}$$

$$\begin{aligned} \max \ & \textstyle\sum_s V(s) \\ \text{s.t.} \ & V(s) \leq c(s,a) + \gamma \textstyle\sum_{s'} \mathcal{P}(s'|s,a) V(s'), \forall s, \forall a. \end{aligned}$$

- Dantzig's Simplex method (1947) is strongly polynomial (Ye, 2011).

Nonlinear Optimization Methods

Nonlinear programming formulation:

$$\begin{aligned} \min_{\pi} \quad & \left\{ f_\rho(\pi) := \sum_{s\in\mathcal{S}} \rho(s) V^\pi(s) \right\} \\ \text{s.t.} \quad & \pi(\cdot|s) \in \Delta_{\mathcal{A}}, s \in \mathcal{S}, \end{aligned}$$

where $\rho \in \mathbb{R}^{|\mathcal{S}|}$ is selected such that $\rho(s) > 0$, $s \in \mathcal{S}$, and $\sum_{s\in\mathcal{S}} \rho(s) = 1$.

- Objective function f_ρ is nonconvex with respect to π.
- The feasible set, represented by the Cartesian product of simplexes, is convex.
- Evaluating f_ρ typically involves sampling according to the transition probability $\mathcal{P}$.

Pros and Cons

- Dynamic optimization
  - Value iteration: simple schemes, rich theory for deterministic and stochastic cases, but unfriendly to function approximation to handle large state/action spaces, requiring generative model.
  - Policy iteration: simple schemes, rich theory for deterministic case only.
- Linear optimization
  - Rich theory for deterministic and stochastic cases.
  - Unfriendly to function approximation and possible nonlinear components, requiring generative model.
- Nonlinear optimization
  - Friendly to function approximation, nonlinear components, applicable to both generative and single-trajectory models.
  - Theoretical studies are lacking behind.

Research objectives

- **Goal:** Novel stochastic first-order algorithms for RL
- **Approach:** study NLP algorithms for both policy optimization and policy evaluation and their interactions.
  - (i) Policy optimization methods (e.g., policy mirror descent) to update policy
  - (ii) Policy evaluation method (e.g., stochastic variational inequality methods) to compute first-order information for a given policy
- **Focus:** performance guarantees (i.e., complexity bounds).

Existing Policy Gradient Methods

- Widely used methods utilizing first-order information of f.
- Performance guarantees are NOT competitive with other MDP/RL methods (Agarwal, Kakade, Lee, Mahajan '19) .
  - Iteration complexity in the deterministic case: $\mathcal{O}(1/\epsilon)$
  - Sample complexity for RL: $\mathcal{O}(1/\epsilon^4)$

Research questions

Can policy gradient type methods achieve comparable or even stronger convergence properties than other RL methods?

Policy Mirror Descent

Algorithm 4 The policy mirror descent (PMD) method (Lan 2021)

Input: initial points π_0 and stepsizes $\eta_k \geq 0$.

for $k = 0, 1, \ldots,$ **do**

$$\pi_{k+1}(\cdot|s) = \operatorname{argmin}_{p(\cdot|s)\in\Delta_{|\mathcal{A}|}} \left\{ \langle Q^{\pi_k}(s,\cdot), p(\cdot|s)\rangle + \frac{1}{\eta_k} D^p_{\pi_k}(s) \right\}.$$

end for

Note:

- Reduces to Policy Iteration if $\eta_k = \infty$.
- Proximal mapping subproblem may have an explicit solution, e.g., when $D^p_{\pi_k}$ is the KL divergence.
- Linear convergence (Lan 2021; Li, Lan and Zhao 2022; Xiao 2022)

Stochastic Policy Mirror Descent (SPMD)

- Given stochastic estimator $\mathcal{Q}^{\pi_k,\xi_k}$, SPMD update:
  $\pi_{k+1}(\cdot|s) = \operatorname{argmin}_{p(\cdot|s)\in\Delta_{|\mathcal{A}|}} \left\{\eta_k \langle \mathcal{Q}^{\pi_k,\xi_k}(s,\cdot), p(\cdot|s)\rangle + D^p_{\pi_k}(s)\right\}.$
- Convergence analysis exploits the separation of bias and variance.
- Sample complexity (Lan 2021): under both generative and single-trajectory models
  - $\mathcal{O}(1/\epsilon^2)$ for general problems.
  - $\mathcal{O}(1/\epsilon)$ for regularized problems.
  - The latter has not been achieved by any other methods.

Other Outcomes

- Surprising properties of PMD
  - Implicitly explores state and action spaces (Li and Lan 23).
  - Distribution-free convergence, strongly polynomial time and accuracy certificates (Ju and Lan 2024)
- Policy optimization over general state and action space (Ju and Lan 2022)
- Accelerated policy evaluation (Kostalis, Lan and Li 2019, Li, Lan and Panajady 21)
- Safe RL (Xu, Liang and Lan, 2020), Robust RL (Li, Lan and Zhao 2021), Average-cost RL (Li, Wu, Lan 2022)

Summary

- Gradient/mirror descent and its acceleration
- Stochastic gradient/mirror descent (SGD)
- Accelerated SGD (SGD with momentum)
- Nonconvex stochastic gradient descent
- Policy mirror descent for reinforcement learning

Active research areas

- Parameter-free methods
- Risk-averse/distributional robust optimization
- Reinforcement learning through optimization
- Integrated modeling and algorithm design