Fast Minimization of Expected Log-Loss

Chung-En Tsai, Hao-Chung Cheng, Yen-Huan Li (National Taiwan University)

Contributions

Theoretically and empirically fastest method for minimizing log-loss with complexity guarantees.

Problem Formulation

Quantum Setup: (maximum-likelihood quantum state tomography, PSD matrix permanent approximation)

$$f^\star := \min_{\rho \in \mathcal{D}} \left\{ f(\rho) := \frac{1}{n} \sum_{i=1}^{n} -\log \operatorname{tr}(A_i \rho) \right\},$$

where

- $\{A_i\}$ are $d \times d$ Hermitian PSD matrices;
- $\mathcal{D} := \{\rho \in \mathbb{C}^{d \times d} \mid \rho^* = \rho, \rho \geq 0, \operatorname{tr} \rho = 1\}$ is the set of quantum density matrices.

Classical Setup: (Poisson inverse problem, Kelly's portfolio)

$$f^\star := \min_{x \in \Delta} \left\{ f(x) := \frac{1}{n} \sum_{i=1}^{n} -\log \langle a_i, x \rangle \right\}.$$

Challenges

- High dimensions and large sample size.
- Lack of Lipschitz continuity and smoothness.

Existing Methods

Goal: $f(\hat{\rho}) - f^\star \leq \varepsilon$. Assume $n \gg d$.

Algorithms	Iteration complexity	Time complexity	
		Classical	Quantum
GD, OSEM, iMLE	may not converge		
Diluted iMLE, EMD with line search, PDHG, SPDHG	asymptotic convergence		
NoLips	d/ε	nd^2/ε	nd^3/ε
EM	$\log d/\varepsilon$	nd/ε	nd^2/ε
Frank-Wolfe	n/ε	n^2d/ε	n^2d^ω/ε
SSB, SLBOMD	d/ε^2	d^2/ε^2	$d^{\omega+1}/\varepsilon^2$
LB-SDA **($B=1$)**			
LB-SDA **($B=d$)**	$1/\varepsilon^2$		d^3/ε^2

Convergence Rate

LB-SDA with mini-batch size B satisfies

$$\mathbb{E}[f(\bar{\rho}_T) - f^\star] \leq \tilde{O}\left(\frac{d}{T} + \frac{\sqrt{d}}{\sqrt{BT}}\right).$$

Matching convergence rate of mini-batch SGD, although log-loss is not smooth.

Tools: A new local-norm-based online-to-batch conversion + "Smoothness" of log-loss

"Smoothness" of Log-Loss

Let $h(\rho) := -\log \det \rho$. Then,

$$\min_{\alpha \in \mathbb{R}} \|\nabla f(\rho) + \alpha I\|_{\rho,*}^2 \leq 4\left(f(\rho) - f^\star\right),$$

where $\|X\|_{\rho,*} := \left(\operatorname{tr}\left((\rho X)^2\right)\right)^{1/2}$ is the dual local norm of h.

Algorithm: LB-SDA

Let $\rho_1 = I/d$ and $f_i(\rho) := -\log \operatorname{tr}(A_i \rho)$.

For all $t \in \mathbb{N}$,

1. Sample $i_1, \ldots, i_B \in [n]$.
2. Compute $g_t = (1/B) \sum_{k=1}^{B} \nabla f_{i_k}(\rho_t)$.
3. Compute

$$\rho_{t+1} \in \operatorname*{arg\,min}_{\rho \in \mathcal{D}} \sum_{\tau=1}^{t} \operatorname{tr}(g_\tau \rho) + \frac{1}{\eta_t} h(\rho),$$

$$\eta_t = \frac{\sqrt{d}}{\sqrt{\sum_{\tau=1}^{t} \min_\alpha \|g_\tau + \alpha I\|_{\rho_\tau,*}^2 + 4d + 1}}.$$

Numerical Results

Quantum setup $(d, n) = (2^6, 4 \times 10^5)$