Faster Rates, Adaptive Algorithms, and Finite-Time Bounds for Linear Composition Optimization and Gradient TD Learning

Anant Raj [1] P. Joulani [2] A. György[2] C. Szepesvári [2,3]

[1]SIERRA, INRIA, Paris.

[2]DeepMind, London.

[3]University of Alberta, Edmonton, Canada.

AISTATS 2022

Outline

Stochastic (Linear) Composition Optimization

Consider minimizing the following objective with an iterative algorithm:

$$\theta^* = \arg\min_{\theta \in \Theta} \ell(\theta) := \frac{1}{2} \|b - A\theta\|^2_{M^{-1}},$$

where,

$$\Theta \subseteq \mathbb{R}^d, \qquad b \in \mathbb{R}^d, \qquad A \in \mathbb{R}^{d \times d}, \qquad M \in \mathbb{R}^{d \times d},$$

given sequential access to noisy estimates b_t, A_t, M_t of b, A, and M. Given symmetric, positive definite M,

$$\nabla \ell(\theta) = -A^\top M^{-1}(b - A\theta)$$

Composition of expectations: From b_t, A_t, M_t, it's not easy to estimate $\nabla \ell$ with "controlled" (e.g. i.i.d. / fast-mixing) noise.

Previous work

Finite-time bounds for iterates $\|\theta_T - \theta^*\|^2$ and/or objective $\ell(\bar{\theta}_T) - \ell(\theta^*)$:

- $\mathcal{O}(1/T^{1-\delta})$
- $\mathcal{O}(1/T)$
- $\mathcal{O}(1/\sqrt{T})$ with bounded domain, when $\sup_{\theta\in\Theta} \|\theta\| \leq r_\theta < \infty$

Especially studied under **Gradient TD Learning** (as we discuss later).

Great advances, with limitations:

- Tuning: Convergence guaranteed only if step-sizes are in a limited range and / or depend on knowledge of several problem parameters:
  - λ_A: upper-bound on max eigenvalue of A_t
  - λ_M: upper-bound on max eigenvalue of M_t
  - μ: lower-bound on min eigenvalue of $A^\top M^{-1} A$
  - β: lower-bound on min eigenvalue of M
  - upper-bound on $\|b_t\|$, lower-bound on min eigenvalue of A, ...
- Constants: The rates depend on the above parameters in obscure ways

Previous work: stochastic optimization

What if we had (independent, unbiased) direct estimates of $\nabla \ell(\theta_t)$?

- **Smooth objectives**: $\mathcal{O}(\frac{\lambda_A^2 r_\theta^2}{\beta T} + \frac{\sigma_* r_\theta}{\sqrt{T}})$ given $\sigma_*^2 \geq \mathbb{E}_t[\|b_t - A_t\theta^*\|^2]$
  - AdaGrad: with minimal knowledge of the problem parameters!
- **Strongly-convex objectives**: $\mathcal{O}(\frac{\sigma_*^2}{\mu T})$
- **Quadratic objectives**: $\mathcal{O}(\frac{1}{T})$, independent of μ! [Bach and Moulines, 2013]

But $\ell(\theta) = \frac{1}{2}\|b - A\theta\|^2_{M^{-1}}$ has all of these properties!

? Can we guarantee convergence with less prior info, e.g., as in AdaGrad?

? Can we expose the dependence of the bounds on λ_A, λ_M, β, μ, etc.?

? Can we get $\mathcal{O}(\frac{1}{T})$ rates independent of μ, as do Bach and Moulines [2013]?

Previous work: stochastic optimization

What if we had (independent, unbiased) direct estimates of $\nabla \ell(\theta_t)$?

- **Smooth objectives**: $\mathcal{O}(\frac{\lambda_A^2 r_\theta^2}{\beta T} + \frac{\sigma_* r_\theta}{\sqrt{T}})$ given $\sigma_*^2 \geq \mathbb{E}_t[\|b_t - A_t\theta^*\|^2]$
  - AdaGrad: with minimal knowledge of the problem parameters!
- **Strongly-convex objectives**: $\mathcal{O}(\frac{\sigma_*^2}{\mu T})$
- **Quadratic objectives**: $\mathcal{O}(\frac{1}{T})$, independent of μ! [Bach and Moulines, 2013]

But $\ell(\theta) = \frac{1}{2}\|b - A\theta\|^2_{M^{-1}}$ has all of these properties!

✓ We guarantee convergence with less prior info, with / without AdaGrad.

✓ We expose the dependence of the bounds on λ_A, λ_M, β, μ, etc.

✓ We get $\mathcal{O}(\frac{1}{T})$ rates independent of μ, as do Bach and Moulines [2013].

Previous work: using two updates

Learn a second estimate y_t that tracks $y_t^* = M^{-1}(b - A\theta_t)$:

$$\nabla \ell(\theta_t) = -A^\top \underbrace{M^{-1}(b - A\theta_t)}_{y_t^*} \approx -A^\top y_t$$

(Two-timescale) stochastic approximation, the basis of **GTD2, TDC,** etc.:

$$\theta_{t+1} = \theta_t + \eta_t^\theta A_t^\top y_t$$
$$y_{t+1} = y_t + \eta_t^y (b_t - A_t\theta_t - M_t y_t)$$

The two-update scheme is equivalent to stochastic gradient descent ascent (SGDA) for the **saddle-point** (θ^*, y^*): [Liu et al., 2018]

$$\min_{\theta\in\Theta} \ell(\theta) = \min_{\theta\in\Theta} \max_{y\in\mathbb{R}^d} L(\theta, y) := \langle b - A\theta, y\rangle - \frac{1}{2}\|y\|_M^2. \qquad (1)$$

Our approach: Online Linear Optimization

Study the optimization gap $\ell(\bar{\theta}_T) - \ell(\theta^*)$ instead of the saddle-point gap

- Expose the benefits of the curvature existing in the problem structure
- Fine-grained analysis with unbounded domains and updates

Reduce the problem to Online Linear Optimization (OLO)

- Basis of many optimization algorithms and fine-grained finite-time bounds (e.g., **AdaGrad** [Duchi et al., 2011, McMahan and Streeter, 2010])

Utilize distinct properties of Dual-Averaging **(DA)** vs Mirror-Descent **(MD)**

- DA: Convergence in unbounded domains
- MD: Faster adaptation

Error Decomposition to **Noise**, **Regret**, and **Curvature**

Let $\bar{\theta} = \theta_{1:T}/T$, and $\bar{y}^* = \arg\max_{y \in \mathcal{Y}} L\left(\bar{\theta}, y\right)$. We have:

$$\ell(\bar{\theta}) - \ell(\theta^\star) = \frac{\delta_{1:T}}{T} + \frac{R_T^\theta + R_T^y + R_T^\sigma}{T} - \frac{B_{1:T} + \bar{B}_{1:T}}{T},$$

where $B_t = \frac{1}{2}\|y_t\|_M^2$, $\bar{B}_t = \frac{1}{2}\|y_t - \bar{y}^*\|_M^2$, δ_t is a noise term arising from estimation of $\nabla_\theta L(\theta_t, y_t)$ and $\nabla_y L(\theta_t, y_t)$ by g_t^θ and g_t^y, and

$$R_T^\theta = \sum_{t=1}^{T} \langle g_t^\theta, \theta_t - \theta^* \rangle, \qquad R_T^y = \sum_{t=1}^{T} \langle -g_t^y, y_t - \bar{y}^* \rangle,$$

$$R_T^\sigma = \sum_{t=1}^{T} \langle \nabla_y L(\theta_t, y_t) - g_t^y, z_t - \bar{y}^* \rangle.$$

Analysis: use curvature to control regret

Adaptive **Dual-Averaging** Regret Bound

For an OLO algorithm of the DA family, we have

$$\sum_t \langle g_t, x_t - x^* \rangle \leq \frac{\eta_{T-1}}{2} \|x^* - x_1\|^2 + \sum_t \frac{\|g_t\|^2}{2\eta_{t-1}}.$$

Adaptive **Mirror-Descent** Regret Bound

For an OLO algorithm of the MD family, we have

$$\sum_t \langle g_t, x_t - x^* \rangle \leq \sum_t \frac{\eta_t - \eta_{t-1}}{2} \|x^* - x_t\|^2 + \sum_t \frac{\|g_t\|^2}{2\eta_t}.$$

Utilizing curvature: controlling iterate terms

Bound on $\|\bar{y}^*\|^2$ (Curvature of the **dual** objective)

We have $\|\bar{y}^*\|^2 \leq \frac{2}{\beta}\left(\ell(\bar{\theta}) - \ell(\theta^*)\right)$, where $\bar{\theta} = \theta_{1:T}/T$.

Bound on $\|\theta - \theta^*\|^2$ (Curvature of the **primal** objective)

For any $\theta \in \mathbb{R}^d$, we have $\|\theta - \theta^*\|^2 \leq \frac{2}{\mu}\left(\ell(\theta) - \ell(\theta^*)\right)$.

Utilizing curvature: controlling update terms

Bound on $\|g_t^\theta\|^2$ (**primal** update)

Recall $B_t = \frac{1}{2}\|y_t\|_M^2$. We have $\|g_t^\theta\|^2 \leq \frac{2\lambda_A^2}{\beta} B_t$.

Bound on $\|g_t^y\|^2$ (**dual** update)

Assume the noises are zero-mean and independent of each other. Then,

$$\mathbb{E}_t\left[\|g_t^y\|^2\right] \leq 3\left(\sigma_*^2 + \lambda_A^2\|\theta_t - \theta^*\|^2 + 2\lambda_M B_t\right) .$$

Adaptive $\mathcal{O}(\frac{1}{\sqrt{T}})$ rates: low prior info requirement

Update rule	Step-size	Bound on the Unnormalized Error: $T \cdot \mathbb{E}\left[\ell(\bar{\theta}_T) - \ell(\theta^*)\right]$
$\theta_{t+1} = \mathcal{P}_\Theta\left(\theta_t - g_t^\theta/\eta^\theta\right)$ $y_{t+1} = y_1 + g_{1:t}^y/\eta_t^y$	$\eta^\theta = 2\lambda_A^2/\beta$ $\eta_t^y = 12\lambda_M + \sqrt{t+1}$	$\frac{\lambda_A^2}{\beta} r_\theta^2 + (12\lambda_M + \sqrt{T})\frac{\lambda_A^2}{\beta^2} r_\theta^2 + 6\left(\sigma_*^2 + \lambda_A^2 r_\theta^2\right)\sqrt{T}$
$\theta_{t+1} = \mathcal{P}_\Theta\left(\theta_t - g_t^\theta/\eta_t^\theta\right)$ $y_{t+1} = y_1 + \frac{g_{1:t}^y}{12\lambda_M + \eta_t^y}$	$\eta_t^\theta = \eta^\theta\sqrt{\sum_{s=1}^{t-1}\|g_s^\theta\|^2}$ $\eta_t^y = \eta^y\sqrt{\sum_{s=1}^{t-1}\|g_s^y\|^2}$	$2\left(\frac{\eta^y\lambda_A^2}{2\beta^2} r_\theta^2 + \frac{1}{\eta^y}\right)\sqrt{3\sigma_*^2 T + 3\lambda_A^2 r_\theta^2 T} + \frac{2\lambda_A^2}{\beta}\left(\frac{\eta^\theta}{2} r_\theta^2 + \frac{1}{\eta^\theta}\right)^2$ $+24\lambda_M\left(\frac{\eta^y\lambda_A^2}{2\beta^2} r_\theta^2 + \frac{1}{\eta^y}\right)^2 + 12\lambda_M\frac{\lambda_A^2}{\beta^2} r_\theta^2 + \frac{\left(B^2 + \lambda_A^2 r_\theta^2\right)}{4\lambda_M}$
$\theta_{t+1} = \mathcal{P}_\Theta\left(\theta_t - g_t^\theta/\eta_t^\theta\right)$ $y_{t+1} = \mathcal{P}_\mathcal{Y}\left(y_t + g_t^y/\eta_t^y\right)$	$\eta_t^\theta = \eta^\theta\sqrt{\sum_{s=1}^{t-1}\|g_s^\theta\|^2}$ $\eta_t^y = \eta^y\sqrt{\sum_{s=1}^{t-1}\|g_s^y\|^2}$	$2\left(\frac{\eta^y c\lambda_A^2}{2\beta^2} r_\theta^2 + \frac{1}{\eta^y}\right)\sqrt{3\sigma_*^2 T + 3\lambda_A^2 r_\theta^2 T} + \frac{2\lambda_A^2}{\beta}\left(\frac{\eta^\theta}{2} r_\theta^2 + \frac{1}{\eta^\theta}\right)^2$ $+24\lambda_M\left(\frac{\eta^y c\lambda_A^2}{2\beta^2} r_\theta^2 + \frac{1}{\eta^y}\right)^2$

Adaptive $\mathcal{O}(\frac{1}{T})$ rates: low prior info given knowledge of β

Update rule	Step-size	Bound on the Unnormalized Error: $T \cdot \mathbb{E}\left[\ell(\bar{\theta}_T) - \ell(\theta^*)\right]$
$\theta_{t+1} = \mathcal{P}_\Theta\left(\theta_t - g_t^\theta/\eta^\theta\right)$ $y_{t+1} = y_t + g_t^y/\eta_t^y$	$\eta^\theta = 2\lambda_A^2/\beta$ $\eta_t^y = 12\lambda_M + \frac{\beta}{2}t$	$\frac{\lambda_A^2}{\beta} r_\theta^2 + 12\lambda_M \frac{\lambda_A^2}{\beta^2} r_\theta^2 + \frac{6\left(\sigma_*^2 + \lambda_A^2 r_\theta^2\right)}{\beta} \log(T+1)$
$\theta_{t+1} = \mathcal{P}_\Theta\left(\theta_t - g_t^\theta/\eta_t^\theta\right)$ $y_{t+1} = y_t + g_t^y/\eta_t^y$	$\eta_t^\theta = \eta^\theta \sqrt{\sum_{s=1}^{t-1} \|g_s^\theta\|^2}$ $\eta_t^y = 12\lambda_M + \frac{\beta}{2}t$	$\frac{\lambda_A^2}{\beta}\left(\frac{\eta^\theta}{2} r_\theta^2 + \frac{1}{\eta^\theta}\right)^2 + 12\lambda_M \frac{\lambda_A^2}{\beta^2} r_\theta^2 + \frac{6\left(\sigma_*^2 + \lambda_A^2 r_\theta^2\right)}{\beta} \log(T+1)$
$\theta_{t+1} = \mathcal{P}_\Theta\left(\theta_t - g_t^\theta/\eta_t^\theta\right)$ $y_{t+1} = \mathcal{P}_\mathcal{Y}\left(y_t + g_t^y/\eta_t^y\right)$	$\eta_t^\theta = \eta^\theta \sqrt{\sum_{s=1}^{t-1} \|g_s^\theta\|^2}$ $\eta_t^y = \frac{\beta}{2}t$	$\frac{\lambda_A^2}{\beta}\left(\frac{\eta^\theta}{2} r_\theta^2 + \frac{1}{\eta^\theta}\right)^2 + \frac{6\left(\sigma_*^2 + \lambda_A^2 r_\theta^2\right)}{\beta} \log(T+1)$ $+ \frac{6\lambda_M^2 c \lambda_A^2 r_\theta^2}{\beta^3} \log\left(\min\{T+1, 24\lambda_M\}\right)$

Restarting Algorithm

two_stage_algo.png

Convergence with independent zero-mean noise

Convergence of the Restarting Algorithm (without Projections)

If T_1, the length of the first epoch, is large enough and the noises at time steps $t = 1, 2, \ldots, T$, are zero-mean and independent of each other, then, after $S > 0$ epochs, we have $\ell(\bar{\theta}_S) - \ell(\theta^*) = \mathcal{O}(\frac{1}{\beta T_{1:S}})$.

Convergence with Markov noise

Convergence of the Restarting Algorithm (without Projections)

Suppose that the noises at time step $t = 1, 2, \ldots, T$, have a finite mixing time τ for a given tolerance Δ. Then, after $S > 0$ epochs, for small enough Δ and large enough T_1, we have

$$\ell(\bar{\theta}_S) - \ell(\theta^*) = \mathcal{O}\left(\frac{\tau}{\beta T_{1:S}} + \Delta\right).$$

Special Case: Policy Evaluation in Reinforcement Learning

- Markov Decision Process with finite state- and action-space.
- Stream of transitions $\{(s_t, a_t, r_t = r(s_t, a_t), s'_t, \rho_t)\}_{t=1}^{\infty}$, where r_t is the reward after choosing action a_t in state s_t and transitioning to s'_t.
- $\phi(s)$: feature vectors for state s, with $\phi_t = \phi(s_t)$ and $\phi'_t = \phi(s'_t)$.
- **Goal**: assuming $a_t \sim \pi_b(s_t)$ where π_b is the behaviour policy, find $\theta \in \mathbb{R}^d$ s.t. $v_\theta(s) = \phi(s)^\top \theta$ approximates the cumulative γ-discounted reward under target policy π, starting from any state s.
- Surrogate loss: Projected Bellman Error $\ell(\theta) = \frac{1}{2}\|b - A\theta\|^2_{M^{-1}}$, where

$$b = \lim_{t\to\infty} \mathbb{E}[b_t], \qquad A = \lim_{t\to\infty} \mathbb{E}[A_t], \qquad M = \lim_{t\to\infty} \mathbb{E}[M_t],$$

with $\rho_t = \frac{\pi(a_t|s_t)}{\pi_b(a_t|s_t)}$, $A_t = \rho_t \phi_t (\phi_t - \gamma \phi'_t)^\top$, $b_t = \rho_t \phi_t r_t$, and $M_t = \phi_t \phi_t^\top$.

Summary and future work

We studied a special linear composition optimization problem which arises in off-policy policy evaluation, and showed

- Step-size tuning strategies that depend on much less prior knowledge than previous work;
- Presented a fine-grained analysis framework that exposed the dependence of our finite-time bounds not only on the number of updates T, but also on other problem parameters;
- Showed a $\mathcal{O}(1/T)$ rate under i.i.d. and Markov noise settings which is independent of μ, a result that had so far been available only for stochastic optimization of quadratic objectives [Bach and Moulines, 2013].

Future work includes simplifying this framework, generalizing it beyond the quadratic objective, and extending it to other algorithmic techniques.

References I

Francis Bach and Eric Moulines. Non-strongly-convex smooth stochastic approximation with convergence rate o (1/n). In *Proceedings of the 26th International Conference on Neural Information Processing Systems-Volume 1*, pages 773–781, 2013.

John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(Jul):2121–2159, 2011.

Bo Liu, Ian Gemp, Mohammad Ghavamzadeh, Ji Liu, Sridhar Mahadevan, and Marek Petrik. Proximal gradient temporal difference learning: Stable reinforcement learning with polynomial sample complexity. *Journal of Artificial Intelligence Research*, 63:461–494, 2018.

H Brendan McMahan and Matthew Streeter. Adaptive bound optimization for online convex optimization. *arXiv preprint arXiv:1002.4908*, 2010.