Decomposition and Sampling Methods for Stochastic Optimization and Variational Problems

Uday V. Shanbhag (udaybag@illinois.edu), University of Illinois at Urbana-Champaign

DOE-DE-SC0003879

EXTENDING THE REALM OF OPTIMIZATION FOR COMPLEX SYSTEMS: UNCERTAINTY, COMPETITION, AND DYNAMICS (co-PIs: Tamer M. Başar, Prashant G. Mehta, and Sean P. Meyn)

Introduction

Consider a stochastic optimization problem given by

$$\boxed{\begin{array}{ll} \text{minimize} & \mathbb{E}[f(x;\omega)] \\ \text{subject to} & x \in K, \end{array}}$$

where $f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$, $f(x) \triangleq \mathbb{E}[f(x;\omega)]$ and $K \subseteq \mathbb{R}^n$ is a closed and convex set. Given $x_0 \in K$ and a sequence $\{\gamma_k\}$, a stochastic approximation algorithm is given by

$$x_{k+1} = \Pi_K\left(x_k - \gamma_k \nabla f(x_k;\omega_k)\right). \tag{1}$$

- Here, $x_0 \in X$ is a random initial point and we assume that $\mathbb{E}\left[\|x_0\|^2\right] < \infty$.
- Let $\mathcal{F}_k \triangleq \{x_0, \omega_0, \omega_1, \ldots, \omega_{k-1}\}$ for $k \geq 1$ and $\mathcal{F}_0 = \{x_0\}$
- Furthermore, $\mathbb{E}[w_k \mid \mathcal{F}_k] = 0$ for all $k \geq 0$, where $w_k = \nabla_x f(x_k;\omega_k) - \nabla f(x_k)$.

Assumption 1 (A1) *The function $f(\cdot;\omega)$ is convex on $\mathbb{R}^n$ for every $\omega \in \Omega$, $\mathbb{E}[f(x;\omega)]$ is finite for every $x \in \mathbb{R}^n$ and f is strongly convex with constant η and differentiable over K with Lipschitz gradients with constant $L > 0$.*

Assumption 2 (A2) *The stepsize is such that $\gamma_k > 0$ for all k. Furthermore, the following hold:*

(a) $\sum_{k=0}^{\infty}\gamma_k = \infty$ and $\sum_{k=0}^{\infty}\gamma_k^2 < \infty$.

(b) For some $\nu > 0$, the stochastic errors w_k satisfy $\mathbb{E}\left[\|w_k\|^2 \mid \mathcal{F}_k\right] < \nu^2$ a.s for $k \geq 0$.

Originally proposed by Robbins and Munro [5], also see recent monographs [3, 1]

Lemma 1 (Convergence of SA) *Suppose (A1) and (A2) hold. Let $\{x_k\}$ be generated by algorithm* (1). *Then*

$$\mathbb{E}\left[\|x_{k+1} - x^*\|^2 \mid \mathcal{F}_k\right] \leq (1 - \eta\gamma_k(2 - \gamma_k L))\|x_k - x^*\|^2 + \gamma_k^2\nu^2 \text{ holds a.s.}$$

- Choosing steplength sequence $\{\gamma_k\}$ satisfying $\sum_{k=0}^{\infty}\gamma_k = \infty$ and $\sum_{k=0}^{\infty}\gamma_k^2 < \infty$.
- A subset of choices given by $\gamma_k := \beta k^{-\alpha}$, where $\beta > 0$ and $\alpha \in (0.5, 1]$.
- Performance **very sensitive** to choices and problem parameters
- Goal: **Develop adaptive steplength rules that are robust to variation in problem parameters**

An adaptive recursive steplength scheme

1. Consider the following:

$$\mathbb{E}\left[\|x_{k+1} - x^*\|^2\right] \leq (1 - \eta\gamma_k(2 - \gamma_k L))\mathbb{E}\left[\|x_k - x^*\|^2\right] + \gamma_k^2\nu^2 \text{for all } k \geq 0. \tag{2}$$

2. When the stepsize is further restricted so that $0 < \gamma_k \leq \frac{1}{L}$, we have $1 - \eta\gamma_k(2 - \gamma_k L) \leq 1 - \eta\gamma_k$.
3. Thus, for $0 < \gamma_k \leq \frac{1}{L}$, inequality (2) yields

$$\mathbb{E}\left[\|x_{k+1} - x^*\|^2\right] \leq \underbrace{(1 - \eta\gamma_k)\mathbb{E}\left[\|x_k - x^*\|^2\right] + \gamma_k^2\nu^2}_{\triangleq e_{k+1}(\gamma_0,\ldots,\gamma_k)} \quad \text{for all } k \geq 0. \tag{3}$$

4. Thus, in the worst case, the error satisfies the following recursive relation:

$$e_{k+1}(\gamma_0,\ldots,\gamma_k) = (1 - \eta\gamma_k)e_k(\gamma_0,\ldots,\gamma_{k-1}) + \gamma_k^2\nu^2.$$

Idea: Why not minimize the upper bound of the error?
If $\mathbb{G}_k \triangleq \left\{z \in \mathbb{R}^k : z_j \in (0, 1/L), j = 0,\ldots,k-1\right\}$, then the minimization problem is given by

$$\min_{(\gamma_j)_{j=0}^k \in \mathbb{G}_{k+1}} \; e_{k+1}(\gamma_0,\ldots,\gamma_k)$$

Proposition 1 (Optimality of sequence within a range) *Let $e_0 > 0$ be such that $\frac{\eta}{2\nu^2}e_0 \leq \frac{1}{L}$ and consider the following recursive rule:*

$$\gamma_0^* = \frac{\eta}{2\nu^2}e_0, \qquad \gamma_k^* = \gamma_{k-1}^*\left(1 - \frac{\eta}{2}\gamma_{k-1}^*\right) \quad \text{for all } k \geq 1. \tag{4}$$

Then, the following hold:

(a) The error e_k satisfies $e_k(\gamma_0^*,\ldots,\gamma_{k-1}^*) = \frac{2\nu^2}{\eta}\gamma_k^*$ for all $k \geq 0$.

(b) For each $k \geq 1$, the vector $(\gamma_0^*, \gamma_1^*, \ldots, \gamma_{k-1}^*)$ is the minimizer of the function $e_k(\gamma_0,\ldots,\gamma_{k-1})$ over $\mathbb{G}_k$ and $e_k(\gamma_0,\ldots,\gamma_{k-1}) - e_k(\gamma_0^*,\ldots,\gamma_{k-1}^*) \geq \nu^2(\gamma_{k-1} - \gamma_{k-1}^*)^2$.

Proposition 2 (Global convergence of RSA scheme) *Suppose (A3) and (A4) hold and $\{\gamma_k\}$ is generated by the recursive scheme. Then, the sequence $\{x_k\}$ generated by algorithm (1) converges almost surely to x^*.*

Proof idea: Suffices to show that $\sum_{k=0}^{\infty}\gamma_k = \infty$ and $\sum_{k=0}^{\infty}\gamma_k^2 < \infty$.

Addressing nonsmoothness

Given a convex function $f(x)$, then a smooth approximation [4, 2] is defined as $\hat{f}(x) \triangleq \mathbb{E}_Z f(x+Z)$.

Lemma 2 (Approximation quality) *Let $z \in \mathbb{R}^n$ be a random vector with a support given by an n-dim. ball centered at the origin with radius ϵ and $\mathbb{E}[z] = 0$. Assume that $f(x)$ is a convex function and there exists a $C > 0$ such that $\|g\| \leq C$ for all $g \in \partial f(x)$ and $x \in \mathbb{R}^n$. Then we have:*

(a) $\hat{f}$ is convex and differentiable over X, with $\nabla \hat{f}(x) = \mathbb{E}[g(x+z)]\ \forall x \in X$, and $g(x+z) \in \partial f(x+z)$ a.s. Furthermore, $\|\nabla \hat{f}(x)\| \leq C$ for all $x \in X$.

(b) $f(x) \leq \hat{f}(x) \leq f(x) + \epsilon C$ for all $x \in X$.

Suppose $z \in \mathbb{R}^n$ has uniform distribution over the n-dimensional ball centered at the origin with radius ϵ and density†

$$p_u(z) = \begin{cases} \frac{1}{c_n \epsilon^n} & \text{for } \|z\| \leq \epsilon, \\ 0 & \text{otherwise.} \end{cases} \quad \text{and } c_n = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}, \Gamma\left(\frac{n}{2}+1\right) = \begin{cases} \left(\frac{n}{2}\right)! & \text{if } n \text{ is even,} \\ \sqrt{\pi}\frac{n!!}{2^{(n+1)/2}} & \text{if } n \text{ is odd.} \end{cases} \tag{5}$$

- **Need a Lipschitz constant to employ RSA**

Lemma 3 (Lipschitz bounds on smooth approximation) *Under the stated assumptions, we have*

$$\|\nabla \hat{f}(x) - \nabla \hat{f}(y)\| \leq \kappa \frac{n!!}{(n-1)!!}\frac{C}{\epsilon}\|x - y\| \qquad \text{for all } x, y \in X,$$

where $\kappa = \frac{2}{\pi}$ if n is even, and $\kappa = 1$ otherwise.

1. Lipshitz constant given by $\kappa\frac{n!!}{(n-1)!!}\frac{C}{\epsilon}$, grows at $\sqrt{n}$ with n
2. We consider a smoothed approximation $\tilde{f}(x) \triangleq \mathbb{E}\left[\hat{f}(x;\omega)\right]$ where $\hat{f}(x;\omega) \triangleq \mathbb{E}_Z[f(x+z;\omega)]$.
3. A modified SA scheme where we sample in the product space of z and ω:

$$x_{k+1} = \Pi_K[x_k - \gamma_k \nabla f(x_k + z_k;\omega_k)] \qquad \text{for } k \geq 0, \tag{6}$$

4. The proposed RSA scheme is employed in this regime

DOE PI Meeting 2011

Numerical results

Consider the following stochastic utility problem, $\min_{x \in X} f(x)$, where $f(x) \triangleq \mathbb{E}[f_i(x;\xi)]$ where $f_i(x;\xi) \triangleq \phi\left(\sum_{i=1}^n \left(\frac{i}{n} + \xi_i\right)x_i\right)$, $X \triangleq \{x \in R^n | x \geq 0, \sum_{i=1}^n x_i = 1\}$, ξ_i are iid normals with mean zero and variance one, The function $\phi(\cdot)$ is a piecewise linear convex function given by $\phi(t) \triangleq \max_{1 \leq i \leq m}\{v_i + s_i t\}$, where $v_i, s_i \in [0,1]$. We choose $n = 20, N = 4000$ with smoothing parameter $\epsilon = 0.5$ and regularization $\eta = 0.5$

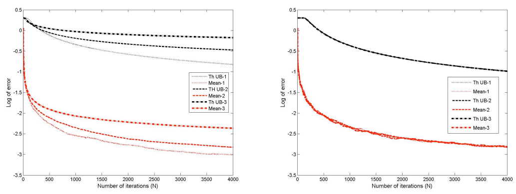

Figure 1: *Sensitivity of HSA ($\gamma_k = 1/k$) (L) and RSA (R) for three problem instances*

Extensions

- Strongly monotone stochastic VIs: $x_{k+1} := \Pi_K\left(x_k - \gamma_k F(x_k;\omega_k)\right)$.
- Extensions to merely monotone SVIs: $x_{k+1} := \Pi_K\left(x_k - \gamma_k(\nabla f(x_k;\omega_k) + \eta_k x_k)\right)$.
  Key: γ_k, η_k updated after every iteration
- Adaptive smoothing generalizations: where γ_k, ϵ_k updated after every iteration

Ref: F. Yousefian, A. Nedich, and U. V. Shanbhag, On stochastic gradient and subgradient methods with adaptive steplength sequences, To appear in Automatica, 2011

Cartesian monotone stochastic variational inequalities

Motivation: Two-period stochastic Nash games

- Consider an $N-$player deterministic Nash game in which the jth agent solves

$$\boxed{\text{A}(z_{-j}) \qquad \begin{array}{ll} \text{maximize} & \pi_j(z_j; z_{-j}) \\ \text{subject to} & \begin{array}{l} d(z) \geq 0 \\ z_j \in \mathbf{Z}_j, \end{array} \end{array}}$$

  where $\pi_j(z_j; z_{-j})$ is a convex differentiable function of z_j for all z_{-j}, $d(z)$ is a concave differentiable function of z and $\mathbf{Z}_j$ is a closed and convex set.
- Then (z^*, λ^*) is an equilibrium if and only if (z^*, λ^*) is a solution of fixed-point problems:

$$z_j = \Pi_{\mathbf{Z}_j}(z_j - \gamma \mathbf{F}_j(z, \lambda)) \tag{7}$$
$$\lambda = \Pi_{\mathbb{R}^+_m}(\lambda - \gamma \mathbf{F}_\lambda(z, \lambda)), \tag{8}$$

  where $\mathbf{F}_i(z, \lambda) = -\nabla_{z_i}\pi_i - \nabla_{z_i}d(z)^T\lambda$ and $\mathbf{F}_\lambda(z, \lambda) = d(z)$.

Challenges:

- Need for scalable distributed algorithms and decomposition schemes
- Projection problem **costly** since $\mathbf{Z}_i$ is given by a set of constraints of cardinality $|\Omega|$ (as arising from two-period stochastic programs)

Two-timescale bounded complexity dual scheme

Two timescale dual scheme: A dual method requires that for every update in the dual space, an *exact* primal solution is required. In particular, for $k \geq 0$, this leads to a set of iterations given by

$$z_j^k = \Pi_{\mathbf{Z}_j}(z_j^k - \gamma_d(\mathbf{F}_z(z_j^k; z_{-j}^k, \lambda^k) + \epsilon^\ell z_j^k)), \text{ for all } j \tag{9}$$
$$\lambda^{k+1} = \Pi_{\mathbb{R}^+_m}(\lambda^k - \gamma_p(\mathbf{F}_\lambda(z^k, \lambda^k) + \epsilon^\ell \lambda^k)), \tag{10}$$

where γ_p and γ_d are the primal and dual steplengths,

- Shortcoming: Need for exact primal solutions for every dual solution.
- Our intent: bounded complexity variant requiring K iterations of the primal scheme be made for a given value of the dual iterates:

$$z_j^{t+1} = \Pi_{\mathbf{Z}_j}(z_j^t - \gamma_d(\mathbf{F}_z(z_j^t; z_{-j}^t, \lambda^k) + \epsilon^\ell z_j^t)), \text{ for all } j, t = 0,\ldots,K-1. \tag{11}$$

- We present results for a networked stochastic Nash game with N_f firms, N_g generating nodes and $n = |\Omega|$

Proposition 3 (Error bounds for inexact-dual scheme) *Consider the inexact dual scheme given by* (11) *and* (10). *If $d(z(\lambda))$ is co-coercive with constant $\epsilon/\|B\|^2$, $\|B\| \leq \sqrt{N_f N_g n}$, $\|z\| \leq M_z$ and γ^l satisfies $\gamma_d < \frac{2\epsilon}{2\epsilon^2 + N_f N_g n}$, then we have*

$$\|\lambda^k - \lambda_\epsilon^*\| \leq q_d^k\|\lambda^0 - \lambda_\epsilon^*\|^k + \left(\frac{1 - q_d^k}{1 - q_d}\right)\left(\left(\frac{2}{\epsilon^2} + 4\right)(N_f N_g n)^{1/2} q_p^{K/2} M_z^2 (1 + (N_f N_g n)^{1/2} q_p^{K/2})\right).$$

Lemma 4 *Consider the inexact dual scheme given by* (11) *and* (10). *If $d(z(\lambda))$ is co-coercive with constant $\epsilon/\|B\|^2$, $\|z\| \leq M_z$ and γ^l satisfies $\gamma_d < \frac{2\epsilon}{2\epsilon^2 + N_f N_g n}$. Then for any nonnegative integers $k, K \geq 0$, we have*

$$\|z_K^k - z_\epsilon^*\| \leq q_p^{K/2} M_z + \frac{\sqrt{N_f N_g n}}{\epsilon}\|\lambda^k - \lambda_\epsilon^*\|, \quad \max(0, -d(z_K^k)) \leq \sqrt{N_f N_g n}\left(q_p^{K/2} M_z + \frac{\sqrt{N_f N_g n}}{\epsilon}\|\lambda^k - \lambda_\epsilon^*\|\right).$$

Scalability Results

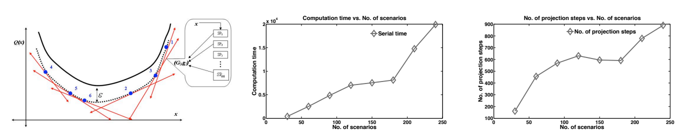

Figure 2: *Cutting plane methods for solving projection problem (L), Scalability (C,R)*

Ref: A. Kannan, U.V. Shanbhag, and H. M. Kim, Addressing supply-side Risk in uncertain power markets: stochastic generalized Nash models, scalable algorithms and error analysis, under second revision in Optimization Methods and Software.

Acknowledgements: We would like to acknowledge the support of DOE Award DOE-DE-SC0003879 and NSF EAGER: 09848905