Proceedings of the 2024 Winter Simulation Conference
H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.

A SMOOTHED AUGMENTED LAGRANGIAN FRAMEWORK FOR CONVEX OPTIMIZATION WITH NONSMOOTH STOCHASTIC CONSTRAINTS

Peixuan Zhang[1]

[1]Dept. of Industrial and Manufacturing Eng., Pennsylvania State University, State College, PA, USA

ABSTRACT

Motivated by the need to develop simulation optimization methods for more general problem classes, we consider a convex stochastic optimization problem where both the objective and constraints are convex but possibly complicated by uncertainty and nonsmoothness. We present a smoothed sampling-enabled augmented Lagrangian framework that relies on inexact solutions to the AL subproblem. Under a constant penalty parameter, the dual suboptimality is shown to diminishes at a sublinear rate while primal infeasibility and suboptimality both diminish at a slower sublinear rate.

1 INTRODUCTION

Consider the convex optimization problem with possibly nonsmooth expectation-valued constraints.

$$\min_{\mathbf{x}\in\mathcal{X}} \left\{ f(\mathbf{x}) \triangleq \mathbb{E}[\tilde{f}(\mathbf{x},\boldsymbol{\xi})] \,|\, g_i(\mathbf{x}) \triangleq \mathbb{E}\left[\tilde{g}_i(\mathbf{x},\boldsymbol{\xi})\right] \le 0, i = 1,\cdots,m, \right\} \tag{NSCopt}$$

where $\mathcal{X} \subseteq \mathbb{R}^n$ is a closed and convex, $\boldsymbol{\xi} : \Omega \to \mathbb{R}^d$ is a d-dimensional random variable, $(\Omega, \mathcal{F}, \mathbb{P})$ denotes the probability space, $\Xi \triangleq \{\, \xi(\omega) \mid \omega \in \Omega \,\}$, and for any $\xi \in \Xi$, $\tilde{f}(\bullet, \xi)$ and $\tilde{g}_i(\bullet, \xi)$ are real-valued possibly nonsmooth (but smoothable (see Def. 1)) convex functions on $\mathcal{X}$ for $i = 1, \cdots, m$. Over the last fifteen years, there has been a pronounced effort in developing inexact Augmented Lagrangian Methods (AL) schemes with complexity guarantees for addressing deterministic convex optimization problems with possibly composite objectives and either conic or more general constraints. When f and g are expectation-valued, the only two available schemes are provided in (Zhang et al. 2023; Zhang et al. 2022) and both are equipped with a rate of $\mathcal{O}(\frac{1}{\sqrt{K}})$, but the first algorithm necessitates solving the AL problem exactly in finite time. Our approach aims to address the lack of efficient Augmented Lagrangian (AL) schemes for convex programs with nonsmooth, expectation-valued constraints.

2 A SMOOTHED STOCHASTIC AUGMENTED LAGRANGIAN FRAMEWORK

Definition 1 (Beck and Teboulle 2012) *Consider a closed, convex, proper function $h : \mathbb{R}^n \to \mathbb{R}$. A convex function is said to be (α, β)-smoothable if for any $\eta > 0$, there exists a convex C^1 function h_η such that*

$$\|\nabla_{\mathbf{z}} h_\eta(\mathbf{z}_1) - \nabla_{\mathbf{z}} h_\eta(\mathbf{z}_2)\| \le \tfrac{\alpha}{\eta}\|\mathbf{z}_1 - \mathbf{z}_2\|, h_\eta(\mathbf{z}) \le h(\mathbf{z}) \le h_\eta(\mathbf{z}) + \eta\beta, \forall \mathbf{z}_1, \mathbf{z}_2, \mathbf{z} \in \mathbb{R}^n.$$

The augmented Lagrangian (AL) function and the smoothed AL function are defined as

$$\mathcal{L}_\rho(\mathbf{x},\lambda) \triangleq \min_{\mathbf{v}\ge 0} \left\{ f(\mathbf{x}) + \lambda^\top (g(\mathbf{x}) + \mathbf{v}) + \tfrac{\rho}{2}\|g(\mathbf{x}) + \mathbf{v}\|^2 \right\}$$
$$\mathcal{L}_{\eta,\rho}(\mathbf{x},\lambda) \triangleq \min_{\mathbf{v}\ge 0} \left\{ f_\eta(\mathbf{x}) + \lambda^\top (g_\eta(\mathbf{x}) + \mathbf{v}) + \tfrac{\rho}{2}\|g_\eta(\mathbf{x}) + \mathbf{v}\|^2 \right\}$$

where $\mathbf{v} \in \mathbb{R}^m$ is a slack variable and $\lambda \in \mathbb{R}^m$ denotes the Lagrange multiplier.

Variance-reduced augmented Lagrangian scheme (VR-AL). Given $\mathbf{x}_0, \lambda_0$ and sequences $\{\rho_k, \epsilon_k, \eta_k, N_k, M_k\}$. For $k = 1, \cdots, K$,

[1] $\mathbf{x}_{k+1}$ satisfies $\mathbb{E}[\mathcal{L}_{\eta_k,\rho_k}(\mathbf{x}_{k+1}, \lambda_k) - \mathcal{D}_{\eta_k,\rho_k}(\lambda_k) \mid \mathcal{F}_k] \leq \epsilon_k \eta_k^b \quad a.s.$ with N_k evals of $\mathcal{SO}^{\text{fir}}$;

[2] $\lambda_{k+1} = \lambda_k + \rho_k \left(\nabla_\lambda \mathcal{L}_{\eta_k,\rho_k}(\mathbf{x}_{k+1}, \lambda_k) + w_k\right),$ where w_k requires M_k evals of $\mathcal{SO}^{\text{zer}}$,

where $w_k \triangleq \left(\Pi_+\left(\frac{\lambda_k}{\rho_k} + [\bar{g}_{\eta_k,M_k}(\mathbf{x}_{k+1})]\right) - \Pi_+\left(\frac{\lambda_k}{\rho_k} + g_{\eta_k}(\mathbf{x}_{k+1})\right)\right)$ and $\mathcal{F}_k$ is defined as $\mathcal{F}_0 \triangleq \{\mathbf{x}_0\}, \mathcal{F}_k \triangleq \mathcal{F}_{k-1} \cup \{\tilde{g}_{\eta_k}(\mathbf{x}_k, \xi_j)\}_{j=1}^{M_k} \cup \left\{\nabla_x \tilde{f}_{\eta_k}(\mathbf{x}, \xi_j) \cup \{\nabla_x \tilde{g}_{i,\eta_k}(\mathbf{x}, \xi_j)\}_{i=1}^m\right\}_{j=1}^{N_k}$, $\mathcal{SO}^{\text{zer}}, \mathcal{SO}^{\text{fir}}$ are zeroth and first-order oracles, $\mathcal{D}_{\eta,\rho}$ is the corresponding dual function.

Assumption 1 (i) $\tilde{f}(\bullet, \xi)$ and $g_i(\bullet, \xi)$ are smoothable real-valued functions; (ii) There exists a point $(\mathbf{x}^*, \lambda^*)$ satisfying the KKT conditions; (iii) $\mathcal{X} \in \mathbb{R}^n$ is convex and compact; (iv) There exists a vector $\bar{\mathbf{x}} \in \mathcal{X}$ such that $g_i(\bar{\mathbf{x}}) < 0$ for $i = 1, \ldots, m$. (v) For any $j \in [m]$ and $\mathbf{x} \in \mathcal{X}$, $\mathbb{E}[\|\tilde{g}_{\eta,j}(\mathbf{x}, \boldsymbol{\xi}) - g_{\eta,j}(\mathbf{x})\|^2] \leq \nu_j^2$; (vi) Suppose $\{\rho_k, \epsilon_k, \eta_k, M_k, N_k\}$ be such that $\left\{\sqrt{2\rho_k \epsilon_k \eta_k^b} + \frac{\nu_g \rho_k}{\sqrt{M_k}} + 2\sqrt{\eta_k \rho_k}\right\}$ be summable.

Proposition 2 (Under constant ρ_k) Consider the sequence $\{(\mathbf{x}_k, \lambda_k)\}$ generated by (**VR-AL**). Suppose $\rho_k = \rho$ for every $k \geq 0$. Then the following holds for $\bar{\mathbf{x}}_K = \frac{\sum_{i=0}^{K-1} \mathbf{x}_i}{K}$and for any $K > 0$.

$$\mathbb{E}\left[f^* - \mathcal{D}_\rho(\bar{\lambda}_K)\right] \leq \frac{1}{K}\mathbb{E}[\|\lambda_0 - \lambda^*\|^2 + \frac{1}{K}\sum_{k=0}^{K-1}\left(\left(\frac{\nu_G}{\sqrt{M_k}} + \frac{\sqrt{2\epsilon_k \eta_k^b}}{\sqrt{\rho}} + \eta_k m\beta\right) B_\lambda + 2\eta_k(b_\lambda m + 1)\beta\right)$$

$$\mathbb{E}\left[d_-(g(\bar{\mathbf{x}}_K))\right] \leq \sqrt{\frac{m}{K}\sum_{i=0}^{K-1}\left(\frac{6\epsilon_i \eta_i^b}{\rho} + 3m^2\eta_i^2\beta^2\right) + \frac{6mC_D}{\rho K} + \frac{C_B}{\rho K}\sum_{i=0}^{K-1}\eta_i} \triangleq \sqrt{\frac{\tilde{C}_d}{K}}$$

$$f(\mathbf{x}^*) - \mathbb{E}[f(\bar{\mathbf{x}}_K)] \leq \eta_K \beta + \frac{\rho \tilde{C}_d}{2K} + \frac{b_{\lambda,\eta}\sqrt{\tilde{C}_d}}{\sqrt{K}} \leq \frac{C_f}{\sqrt{K}}$$

$$f(\mathbf{x}^*) - \mathbb{E}[f(\bar{\mathbf{x}}_K)] \geq -\frac{\|\lambda_0\|^2}{2\rho K} - \frac{1}{K}\sum_{k=0}^{K-1}\left((B_\lambda + b_\lambda)\frac{\nu_G}{\sqrt{M_k}} + \frac{(\rho+1)\nu_G^2}{M_k} + \epsilon_k \eta_k^b + \eta_k \beta\right) \geq -\frac{\tilde{C}_f}{K}$$

where $\tilde{C}_d, C_f, \tilde{C}_f$ are non-negative constants, $d\_(\bullet)$ is the distance function to non-positive orthant.

When $\{\rho_k\}$ is an increasing sequence and $\eta_k = \mathcal{O}(1/\rho_k)$ and $M_k = \mathcal{O}(1/\rho_k^2)$, then it can be shown that the expected suboptimality and infeasibility diminish at the rate of $\mathcal{O}(1/\rho_k)$.

3 CONCLUDING REMARKS

In this paper, we present an efficient inexact sampling-enabled AL framework for contending with convex optimization problems with possibly nonsmooth and convex expectation-valued constraints with rates guarantees. Future work will consider developing an overall complexity analysis as well as extensions to compositional constraints.

REFERENCES

Beck, A. and M. Teboulle. 2012. "Smoothing and first order methods: A unified framework". *SIAM Journal on Optimization* 22(2):557–580.

Zhang, L., Y. Zhang, J. Wu, and X. Xiao. 2022. "Solving stochastic optimization with expectation constraints efficiently by a stochastic augmented lagrangian-type algorithm". *INFORMS Journal on Computing* 34(6):2989–3006.

Zhang, L., Y. Zhang, X. Xiao, and J. Wu. 2023. "Stochastic approximation proximal method of multipliers for convex stochastic programming". *Mathematics of Operations Research* 48(1):177–193.