High probability and risk-averse guarantees for a stochastic accelerated primal-dual method[1]

Yassine Laguel [1]

[1] *Université Côte d'Azur, France, yassine.laguel@univ-cotedazur.fr*

Résumé. Les problèmes point-selle apparaissent dans de nombreuses applications, allant de l'apprentissage robuste à la théorie des jeux, en passant par les problèmes d'équité statistique. Dans ce projet, nous étudions les propriétés de convergence de `SAPD`, un algorithme du premier ordre pour les problèmes point-selle stochastiques fortement monotones. Nous démontrons la convergence à une vitesse accéléré de l'algorithme, en haute probabilités et pour différentes mesures de risque convexes. Pour les problèmes quadratiques sous perturbations gaussiennes, nous dérivons des formules analytiques sur la matrice de covariance limite des iterées ainsi que des bornes inférieures de complexité qui montrent que notre analyse générale est optimale. Nous illustrons nos résultats avec des expériences numériques sur des jeux à somme nulle et des problèmes d'apprentissage robustes.

Mots-clés. Problèmes point-selles, optimisation stochastique, preuve en grande probabilité, mesures de risque.

Abstract. Stochastic saddle point problems arise in many applications, ranging from distributionally robust learning to game theory and fairness in machine learning. We investigate the stochastic accelerated primal-dual algorithm for strongly-convex-strongly-concave (SCSC) saddle point problems. Our algorithm offers optimal complexity in several settings and we provide high probability guarantees for convergence to a neighbourhood of the saddle point. For quadratic problems under Gaussian perturbations, we derive analytical formulas for the limit covariance matrix together with lower bounds that show that our general analysis for SCSC problems is tight. Our risk-averse convergence analysis characterises the trade-offs between bias and risk in approximate solutions. We present numerical experiments on zero-sum games and robust learning problems..

Keywords. Saddle point problems, Stochastic optimization, High probability analyses, Risk measures.

Link to paper : `https://arxiv.org/pdf/2304.00444.pdf`

1 SAPD: A robust accelerated algorithm for stochastic min-max problems

1.1 Problem setting

We consider strongly convex/strongly concave (SCSC) saddle point problems of the form:

$$\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}} \mathcal{L}(x,y) \triangleq f(x) + \Phi(x,y) - g(y), \tag{1}$$

where $\mathcal{X}$ and $\mathcal{Y}$ are finite-dimensional Euclidean spaces, $f : \mathcal{X} \to \mathbb{R}\cup\{+\infty\}$ (resp. $g : \mathcal{Y}\times\mathbb{R}\cup\{+\infty\}$) is closed and μ_x-strongly convex (resp. μ_y-strongly convex), and $\Phi : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ is a smooth convex-concave function. Specifically, we assume that the partial gradients of $\nabla_x\Phi$ and $\nabla_y\Phi$ of Φ satisfy

$$\begin{aligned}
\|\nabla_x\Phi(x,y) - \nabla_x\Phi(\bar{x},\bar{y})\| \le L_{x,x}\|x-\bar{x}\| + L_{x,y}\|y-\bar{y}\|,\\
\|\nabla_y\Phi(x,y) - \nabla_y\Phi(\bar{x},\bar{y})\| \le L_{y,x}\|x-\bar{x}\| + L_{y,y}\|y-\bar{y}\|,
\end{aligned}$$

for all $(x,y),(\bar{x},\bar{y}) \in \mathbf{dom}\, f \times \mathbf{dom}\, g$. We further assume that these gradient are available only through light-tail stochastic estimates $\widetilde{\nabla}_x\Phi$ and $\widetilde{\nabla}_y\Phi$: the gradient noise terms $\Delta_k^x = \widetilde{\nabla}_x\Phi(x_k,y_{k+1}) - \nabla_x\Phi(x_k,y_{k+1})$ and $\Delta_k^y = \widetilde{\nabla}_y\Phi(x_k,y_k) - \nabla\Phi(x_k,y_k)$ are assumed norm-subGaussian [1] with respective proxies $\delta_x,\delta_y > 0$. Such setting arises frequently in large-scale optimization and machine learning applications where the gradients are estimated from either streaming data or from random samples of data (see e.g. [2, 3]).

1.2 The SAPD Algorithm

We consider the Stochastic Accelerated Primal Dual algorithm (SAPD), introduced in [4], which interleaves stochastic proximal gradient ascent steps with respect to y and stochastic proximal gradient descent steps with respect to x. Precisely, for $k \ge 0$, it maintains the iterates, x_k, y_k, and s_k as

$$\begin{aligned}
&\tilde{\mathrm{s}}_k \leftarrow \tilde{\nabla}_{\mathrm{y}}\Phi\left(x_k, y_k, \omega_k^y\right) + \theta\left(\tilde{\nabla}_{\mathrm{y}}\Phi\left(x_{k+1}, y_{k+1}, \omega_{k+1}^y\right) - \tilde{\nabla}_{\mathrm{y}}\Phi\left(x_k, y_k\right)\right)\\
&y_{k+1} \leftarrow \mathrm{prox}_{\sigma g}\left(y_k + \sigma\tilde{\mathrm{s}}_k\right)\\
&x_{k+1} \leftarrow \mathrm{prox}_{\tau f}\left(x_k - \tau\tilde{\nabla}_{\mathrm{x}}\Phi\left(x_k, y_{k+1}\right)\right).
\end{aligned} \tag{2}$$

where τ and σ denote respectively stepsize parameters with respect to x and y, and $\theta \in (0,1)$ a momentum parameter. The use of momentum on the dual variable, as displayed in the previous equation, allows for provably accelerated convergence of SAPD toward an approximate solution of (1).

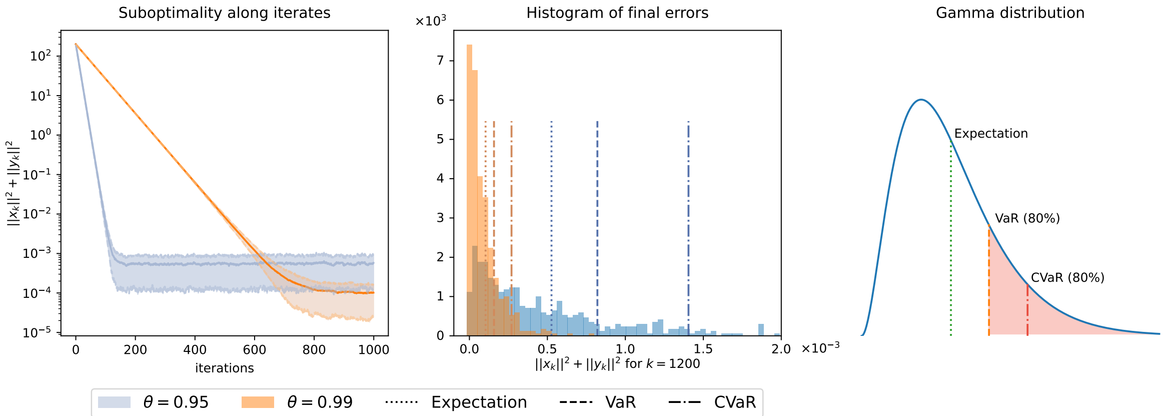

Figure 1: (Left) Convergence of SAPD on the saddle point problem $\min_{x\in\mathbb{R}}\max_{y\in\mathbb{R}} x^2/2 + xy + y^2/2$, initialized at $x_0 = y_0 = 10$ with momentum parameters $\theta = 0.95$ and $\theta = 0.99$. **(Middle)** Histogram of the distribution of the SAPD iterates (x_n, y_n) after $n = 1000$ iterations for 500 runs, with corresponding momentum parameters $\theta = 0.95$ and $\theta = 0.99$. **(Right)** Illustration of the expectation $\mathbb{E}(X)$, p-th quantile $(\mathrm{VaR}_{1-p}(X))$ and $\mathrm{CVaR}_p(X)$ for $p = 80\%$ and X a gamma-distributed random variable with shape parameter 3 and scale parameter 5.

2 Main contributions

2.1 High Probability Bounds

Our main contribution lies in providing the first analysis of an accelerated algorithm for SCSC problems with high probability guarantees, where our bounds reflect the accelerated decay of the initialization bias scaling linearly with the condition number of the problem. More specifically, our high-probability bounds imply that given target accuracy $\varepsilon > 0$, `SAPD`, with a proper choice of parameters that we explicit, can generate a solution (x_n, y_n) that satisfies $\mu_x\|x_n - x^\star\|^2 + \mu_y\|y_n - y^\star\|^2 \le \varepsilon$ with probability $p \in (0,1)$ after

$$n = \mathcal{O}\left(\left[\frac{L_{x,x}}{\mu_x} + \frac{L_{y,x}}{\sqrt{\mu_x\mu_y}} + \frac{L_{y,y}}{\mu_y} + \left(1 + \frac{L_{x,y}}{L_{y,x}} + \frac{L_{x,y}^2}{L_{y,x}^2}\right)\max\left(\frac{\delta_x^2}{\mu_x}, \frac{\delta_y^2}{\mu_y}\right)\frac{1+\log\left(\frac{1}{1-p}\right)}{\varepsilon}\right]\log\left(\frac{(1+\frac{L_{x,y}^2}{L_{y,x}^2})\mathcal{W}_0}{\varepsilon}\right)\right) \quad (3)$$

iterations where $\mathcal{W}_0 \triangleq \mu_x\|x_0 - x^\star\|_2^2 + \mu_y\|y_0 - y^\star\|_2^2$.

2.2 Refined analysis for quadratics

We provide an in-depth analysis of the behavior of `SAPD` on a class of quadratic problems subject to i.i.d. isotropic Gaussian noise where we can characterize the behavior of the distribution of the iterates explicitly. In particular, we derive an analytical formula for the limiting covariance matrix of `SAPD`'s iterates. This is achieved through the solving of an intricate Lyapunov equation parameterized by the stepsize and momentum parameters of the algorithm. We leverage this formula to demonstrate the tightness of our high probability bounds for general SCSC problems.

2.3 Further risk-averse bounds

We finally provide finite-time risk guarantees for the convergence of `SAPD`, where we measure the risk in terms of the Conditional Value at Risk [5], the Entropic Value at Risk [6] and χ^2-divergence based risk measure of the distance to the saddle point. To our knowledge, these are the first risk-averse guarantees that quantify the risk associated with an *approximate* solution generated by a primal-dual algorithm for saddle point problems.

References

[1] Chi Jin, Praneeth Netrapalli, Rong Ge, Sham M Kakade, and Michael I Jordan. A short note on concentration inequalities for random vectors with subGaussian norm. *arXiv preprint arXiv:1902.03736*, 2019.

[2] Mert Gürbüzbalaban, Andrzej Ruszczyński, and Landi Zhu. A stochastic subgradient method for distributionally robust non-convex and non-smooth learning. *Journal of Optimization Theory and Applications*, 194(3):1014–1041, 2022.

[3] Léon Bottou, Frank E Curtis, and Jorge Nocedal. Optimization methods for large-scale machine learning. *SIAM review*, 60(2):223–311, 2018.

[4] Xuan Zhang, Necdet Serhat Aybat, and Mert Gürbüzbalaban. Robust accelerated primal-dual methods for computing saddle points, 2023. Available at https://arxiv.org/pdf/2111.12743.

[5] R Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of banking & finance*, 26(7):1443–1471, 2002.

[6] Amir Ahmadi-Javid. Entropic value-at-risk: A new coherent risk measure. *Journal of Optimization Theory and Applications*, 155:1105–1123, 2012.