# FINITE-TIME CONVERGENCE GUARANTEES OF CONTRACTIVE STOCHASTIC APPROXIMATION: MEAN-SQUARE AND TAIL BOUNDS

SIVA THEJA MAGULURI

We consider the problem of solving the fixed-point equation $\mathcal{H}(x) = x$, where $\mathcal{H} : \mathbb{R}^d \to \mathbb{R}^d$ is a contractive operator wrt a given norm $\|\cdot\|$ in $\mathbb{R}^d$. However, the operator $\mathcal{H}$ is unknown, and instead, one has access to a noisy oracle that at time $k$, given a vector $x_k$, return $H(x_k)+w_k$, where $w_k$ is the noise (that can depend on $\{x_0, x_1, \ldots x_k\}$). Robbins and Monro [1] introduced the Stochastic Approximation (SA) algorithm for this problem, where $x_{k+1} = x_k + \alpha_k (\mathcal{H}(x_k) - x_k + w_k)$, where $\{\alpha_k\}$ is the step-size sequence.

Such problems arise in various applications such as optimization and reinforcement learning. Of special interest in reinforcement learning is the case when the underlying norm is the $\ell_\infty$ norm which is much more challenging to study than the case of $\ell_2$ norm. We consider two cases. The first is when the magnitude of noise does not scale with the iterates, which we call the additive noise setting. The second is when the magnitude of noise is allowed to affinely grow with the iterates, which we call the multiplicative noise setting. An illustrative special case is when $\mathcal{H}$ is linear. Then the problem reduces to solving a linear equation of the form $Ax = b$, and the stochastic approximation algorithm reduces to $x_{k+1} = x_k + \alpha_k (A_k x_k - b_k)$, where $A_k$ and $b_k$ are noisy samples of $A$ and $b$ respectively. With appropriate choice of step sizes, it can be argued that the contractive property in this case reduces to $A$ being Hurwitz (i.e., all its eigenvalues have negative real parts). Additive noise here corresponds to the setting where $A$ is known (i.e., $A_k = A$), and $b$ alone is noisy. Multiplicative noise corresponds to when both $A$ and $b$ are noisy.

Prior work has established asymptotic convergence of the iterates $x_k$ to $x^*$, the unique fixed point of $\mathcal{H}$. The focus of our work is on establishing finite time error bounds, $\|x_k - x^*\|$. First, we present a bound on the mean-square error, $\mathbb{E}[\|x_k - x^*\|^2]$ under various choices of step-sizes. In particular, we show that when the step sizes are of the form $\frac{\alpha}{k+k_0}$ for appropriately chosen constants $\alpha$ and $k_0$, the mean square error is upper bounded by $\frac{C}{k}$ where $C$ is a problem-dependant constant. Such a result is well-known in the case when the underlying norm is the $\ell_2$ norm. However, obtaining this result for $\ell_\infty$ norm was open since the 90's, and our results is applicable for arbitrary norm. We use a Lyapunov (potential) function framework where the key is constructing a Lyapunov function. In the $\ell_2$ norm case, $V(x) = \|x - x^*\|^2$ works as a Lyapunov function. In the general setting, we construct a smooth (i.e., has Lipschitz gradients) approximation of $\|x-x^*\|^2$ using infimal convolution and the generalized Moreau envelop, and show that it serves as a Lyapunov function and gives the mean-square bounds.

Our next focus is on establishing tail bounds on the error $\|x - x^*\|^2$. Using Chebyschev inequality and the bound on the mean-square error, one immediately gets that $\mathbb{P}(\|x - x^*\|^2 \geq \frac{C}{k}z) \leq \frac{1}{z}$ for all $z > 0$. Our goal is to establish tail bounds that decay faster than any polynomial. In the additive noise setting, we establish exponentially decaying tail bound of the form $\mathbb{P}(\|x - x^*\|^2 \geq \frac{C}{k}z) \leq e^{-cz}$ where $c$ is a problem dependant constant. This implies the following sample complexity result: In order to obtain a solution $\hat{x}$ that is within $\epsilon$ of $x^*$ (i.e., $\|\hat{x} - x^*\| \leq \epsilon$) with probability $(1 - \delta)$, we need to run the stochastic approximation algorithm for $O\left(\frac{1}{\epsilon^2}\right) O\left(\log\left(\frac{1}{\delta}\right)\right)$ iterations. In the multiplicative noise setting, we establish Weibullian tails of the form $\mathbb{P}(\|x - x^*\|^2 \geq \frac{C}{k}z) \leq e^{-cz^{\frac{1}{R}}}$ for some $R \geq 1$, and this implies a sample complexity of $O\left(\frac{1}{\epsilon^2}\right) O\left(\left(\log\left(\frac{1}{\delta}\right)\right)^R\right)$. We also present a counter example to show that in the multiplicative noise setting, it is not possible to obtain exponential tails. In addition, we show that our results are applicable uniformly over the entire sample path.

In the additive noise setting, the iterates are bounded by a constant almost surely, and we construct an exponential supermartingale based on the Moreau envelop Lyapunov function. An application of the Ville's (aka Doob's) maximal inequality gives the result. The key challenge is in the multiplicative noise setting, where the iterates need not bounded by a constant. We adopt the following bootstrapping approach. First, we obtain an almost sure bound on the iterates that increases over time. Say the bound is of the form $O(k^R)$ for some $R > 0$. Starting from this upper bound, and going through the same steps as in the additive noise setting (i.e., using the exponential supermartingale), we obtain a tighter bound of the form $O(k^{R-1})$ with high probability. This process is inductively repeated $\lceil R+1 \rceil$ times to get the $O(1/k)$ bound.

This talk is based on joint work with Zaiwei Chen, Sanjay Shakkottai, Karthikeyan Shanmugam and Martin Zubeldia and presents results from [2, 3].

## References

[1] H. Robbins and S. Monro, "A stochastic approximation method," *The Annals of Mathematical Statistics*, pp. 400–407, 1951.

[2] Z. Chen, S. T. Maguluri, S. Shakkottai, and K. Shanmugam, "Finite-sample analysis of stochastic approximation using smooth convex envelopes," in *Advances in Neural Information Processing Systems*, 2020. [Online]. Available: https://proceedings.neurips.cc/paper/2020/file/5d44ee6f2c3f71b73125876103c8f6c4-Paper.pdf

[3] Z. Chen, S. T. Maguluri, and M. Zubeldia, "Concentration of contractive stochastic approximation: Additive and multiplicative noise," *Preprint arXiv:2303.15740*, 2023.