Direct Acceleration of Stochastic Root-Finding Without Variance Reduction and Regularization

TaeHo Yoon Nicolas Loizou

Department of Applied Mathematics & Statistics
Mathematical Institute for Data Science
Johns Hopkins University

NeurIPS 2026

Root-finding/Fixed-point problems

We consider the root-finding problem

$$\underset{x\in\mathbb{R}^d}{\text{find}} \quad \mathbb{F}(x) = 0$$

or the equivalent fixed-point problem

$$\underset{x\in\mathbb{R}^d}{\text{find}} \quad x = \mathbb{T}(x)$$

where $\mathbb{F}\colon \mathbb{R}^d \to \mathbb{R}^d$ and $\mathbb{T} = \mathbb{I} - \alpha\mathbb{F}$ with $\alpha > 0$. We consider the problems where

$$\mathbb{F} \text{ is } \tfrac{1}{L}\text{-cocoercive} \iff \mathbb{T} \text{ is nonexpansive for } 0 < \alpha \leq \tfrac{2}{L}.$$

Deterministic acceleration

For nonexpansive fixed-point problems, we have optimal[1] accelerated algorithms w.r.t. squared fixed-point residual $\|x_{N-1} - \mathbb{T}(x_{N-1})\|^2$:

	Anchoring (Halpern[2])	**Dual anchoring**
Terminal iteration bound[3,4,5]	$\frac{4\|x_0 - x_\star\|^2}{N^2}$	$\frac{4\|x_0 - x_\star\|^2}{N^2}$
Is the method anytime?	✓	✗

Anchoring is **anytime** acceleration which has guarantee at all iterations. However, Dual-OHM (dual anchoring) must choose N in advance.

Why consider dual anchoring then?

[1]Park and Ryu (2022); [2]Halpern (1967); [3]Lieder (2021); [4]Kim (2021); [5]Yoon et al. (2024).

Stochastic extension of deterministic acceleration

Anchoring: OHM

$$x_{k+1} = \frac{1}{k+2}x_0 + \frac{k+1}{k+2}\mathbb{T}(x_k).$$

Dual anchoring: Dual-OHM

$$x_{k+1} = x_k + \frac{N-k-1}{N-k}(\mathbb{T}(x_k) - \mathbb{T}(x_{k-1})).$$

OHM stands for *Optimal Halpern Method*. For Dual-OHM, $\mathbb{T}(x_{-1}) = x_0$ and $k = 0, \ldots, N-2$.

Direct stochastic extensions replace $\mathbb{T}(x_k)$ by a mini-batch estimate.

However, stochastic anchoring is known to accumulate errors.[6,7]
Is Dual-OHM better?

[6] Lee and Kim (2021); [7] Cai et al. (2022).

Yes, dual anchoring admits a better stochastic extension

Assume unbiased oracle with variance $\leq \sigma^2$ and $1/L$-cocoercivity in expectation.

With $\alpha = 1/L$ and constant batch-size B, stochastic Dual-OHM satisfies

$$\mathbb{E}\|\mathbb{F}(x_{N-1})\|^2 \leq \underbrace{\frac{4L^2\|x_0 - x_\star\|^2}{N^2}}_{\text{deterministic rate}} + \underbrace{\frac{6\sigma^2}{B}}_{\substack{\text{stochastic term} \\ \textbf{does not grow with } N}} .$$

To achieve $\mathbb{E}\|\mathbb{F}(x_{N-1})\| \leq \epsilon$, it suffices to use

$$\underbrace{N = \mathcal{O}(\epsilon^{-1})}_{\text{iterations}} \times \underbrace{B = \mathcal{O}(\epsilon^{-2})}_{\text{samples per iteration}} = \boxed{\mathcal{O}(\epsilon^{-3}) \text{ oracle calls}} .$$

No variance reduction. No recursive regularization.

Noise separates the two equally optimal deterministic methods

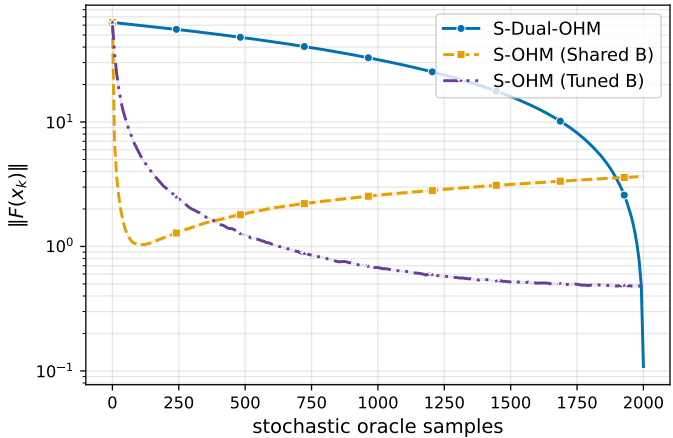
Known guarantees for direct mini-batch stochastic extensions

	Anchoring	**Dual anchoring**
Batch size for accuracy ϵ	$B_k = \Theta(k\epsilon^{-2})$	$B = \Theta(\epsilon^{-2})$
Total oracle complexity	$\mathcal{O}(\epsilon^{-4})$	$\mathcal{O}(\epsilon^{-3})$

The anchoring bound controls noise accumulation with **growing batches**.
Dual anchoring permits **constant batches**.

Halpern can also achieve $\mathcal{O}(\epsilon^{-3})$ via variance reduction techniques such as PAGE (Cai et al., 2022). However, dual-anchor achives the same complexity **directly and much more easily**.

Experiments



Noisy worst-case fixed-point problem

All algorithms use same total sample budget.

S-OHM may even diverge, with same constant B used as in S-Dual-OHM.

Even with separately tuned B, S-OHM is slower than S-Dual-OHM.

Takeaway

Deterministic setting

- Same exact optimal last-iterate rate is achieved by different algorithms.
- Anchoring is anytime, while dual-anchoring requires a fixed horizon.

Stochastic setting

- Anchoring accumulates noise, and requires additional techniques such as variance reduction to achieve $\mathcal{O}(\epsilon^{-3})$ complexity.

• Dual-anchoring is much more robust to noise, and achieves the $\mathcal{O}(\epsilon^{-3})$ complexity easily and directly.