DISTRIBUTED NATURAL POLICY GRADIENTS WITH POISSONIZED DETERMINANTAL AVERAGING

Anonymous authors
Paper under double-blind review

ABSTRACT

Natural Policy Gradients (NPG) can exploit the low-rank structure of the empirical Fisher via the Woodbury identity, replacing a parameter-space solve with a sample-space system. While computationally attractive, this formulation incurs $\mathcal{O}(Bd)$ memory for B samples and d parameters. Data parallelism can reduce this cost, but naively averaging local empirical NPG directions does not recover the pooled-batch estimator due to finite-sample bias from Fisher inversion and Fisher–gradient coupling. We propose Poissonized Determinantal Averaging (PDA), a communication-efficient distributed NPG method based on Poissonized local sampling and determinant-weighted aggregation. PDA communicates only one parameter-sized direction and one scalar per worker. We show that, for any fixed number of workers N, PDA matches the full $1/(NB)$ leading-order bias term of pooled Poissonized NPG as the expected local batch size B grows, leaving an $\mathcal{O}_N(B^{-2})$ gap in expectation. For fixed B, PDA also converges to the population damped NPG direction as $N \to \infty$. Across eight MuJoCo and Procgen tasks, PDA more closely matches pooled NPG than uniform averaging while retaining favorable compute and memory scaling.

1 INTRODUCTION

Natural Policy Gradients (NPG) (Kakade, 2001) precondition the policy gradient with the inverse Fisher information matrix, yielding the direction of steepest local policy improvement under the KL-divergence geometry. Both the Fisher matrix and the policy gradient are population expectations under the policy-induced state-action distribution and are usually estimated from finite on-policy samples, giving rise to an empirical NPG update. Crucially, the empirical Fisher has a low-rank structure. Given B samples and d policy parameters, it is a $d \times d$ matrix formed from a sum of B rank-one score outer products, and therefore has rank at most B. This structure can be exploited through the Woodbury matrix identity. For a damped empirical Fisher, Woodbury transforms the $d \times d$ parameter-space linear system required by NPG into an equivalent $B \times B$ system in sample space, making the computation substantially more tractable when $B \ll d$. Such sample-space formulations have been widely studied for natural-gradient methods in supervised learning (Amari et al., 2000; Roux et al., 2007; Wu et al., 2024; Guzmán-Cordero et al., 2025) and have recently attracted increasing interest in reinforcement learning (Sun, 2026).

A key limitation of Woodbury-based methods for large-scale policy optimization is their memory footprint. In general, they require $\mathcal{O}(Bd)$ memory to store per-sample score vectors and $\mathcal{O}(B^2)$ memory for the corresponding sample-space Gram matrix. In the practically relevant regime $B \ll d$, the dominant memory cost is therefore typically $\mathcal{O}(Bd)$. A natural approach to reducing this cost is data-parallel distributed optimization, such as centralized SGD (CSGD), as commonly used in large-scale first-order training (Dekel et al., 2012; Goyal et al., 2017). Under synchronous data parallelism, N workers evaluate gradients at the same shared parameters using disjoint local mini-batches of size B, after which the local gradients are uniformly averaged to produce the global update. For an additive empirical objective, this procedure is exactly equivalent to computing a single gradient over the pooled batch of NB samples, while distributing its computation and memory across the N workers. Applying the same principle to Woodbury-based NPG is particularly appealing. Relative to processing the pooled NB-sample batch on a single worker, each worker only needs to store B score vectors, reducing the dominant per-worker memory cost from $\mathcal{O}(NBd)$ to $\mathcal{O}(Bd)$, i.e., by a factor of N. Moreover, operations involving the sample-space Gram matrix are performed on a $B \times B$

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

matrix rather than an $NB \times NB$ matrix, yielding up to a quadratic reduction in the corresponding per-worker Gram construction cost.

Unlike CSGD, however, uniformly averaging local empirical NPG directions does not recover the NPG update computed from the pooled batch. The key difficulty is that empirical NPG is a nonlinear function of its sampled quantities. In particular, the inverse of an empirical Fisher matrix is generally biased, and, in standard on-policy implementations, the empirical Fisher and policy gradient are typically estimated from the same samples, introducing additional dependence between curvature and gradient. Together, these effects induce a finite-sample bias in the empirical NPG direction. We characterize this bias and show that its leading-order term scales inversely with the sample size, and, unlike gradient averaging in CSGD, uniform averaging of local NPG directions leaves a systematic gap from the pooled estimator that does not vanish simply by increasing the number of workers. A natural candidate for correcting this discrepancy is determinantal averaging (DA) (Derezinski & Mahoney, 2019), which compensates for the bias introduced by inverting random local curvature estimates in distributed Newton methods. Extending DA to empirical NPG further requires handling both policy-induced random Fisher atoms and Fisher-gradient coupling. Separately, Derezinski et al. (2020b) establish useful determinant expectation identities for Poisson random designs, providing important algebraic machinery for extending determinant weighting to our setting.

Building on these ideas, we introduce *Poissonized Determinantal Averaging* (PDA). Each worker draws a Poisson-distributed number of on-policy samples, computes a local empirical NPG direction, and the resulting directions are aggregated using normalized determinant weights. Poisson factorial-moment identities recover the determinant and adjugate expectation identities required for policy-induced random Fisher atoms, and preserves the corresponding determinant-weighted identity when the Fisher and gradient share samples. Building on these identities, we show that PDA matches the pooled large-batch NPG estimator to leading order in expectation, while requiring only vector-level communication.

Our contributions are summarized as follows:

- We introduce Poissonized Determinantal Averaging (PDA), a data-parallel method for empirical NPG over continuous state-action spaces.
- We characterize the finite-sample bias of empirical NPG under Poisson sampling, and show that PDA recovers the full $\mathcal{O}((NB)^{-1})$ leading bias of pooled Poissonized NPG as $B \to \infty$ and converges to the population damped NPG direction as $N \to \infty$.
- Across eight MuJoCo and Procgen tasks, PDA more closely matches pooled NPG than uniform averaging and improves consistently with more workers. We further characterize its bias-variance tradeoff and validate its predicted compute and memory scaling.

2 Related Work

Sample-space empirical NPG. Woodbury-based methods exploit the low-rank structure of the empirical Fisher to compute the natural-gradient direction in sample space (Amari et al., 2000). This formulation has been studied extensively in supervised learning (Amari et al., 2000; Roux et al., 2007; Wu et al., 2024) and has recently gained interest in RL (Sun, 2026). Constructing and factorizing the resulting $B \times B$ Gram matrix requires $\mathcal{O}(B^2d + B^3)$ time and $\mathcal{O}(Bd + B^2)$ memory (Wu et al., 2024). Other approaches avoid explicit Fisher factorization altogether. Matrix-free methods, such as TRPO (Schulman et al., 2015), approximate the NPG direction using Fisher-vector products and iterative conjugate-gradient solves, while structured curvature methods such as K-FAC (Martens & Grosse, 2015) rely on block-diagonal and Kronecker-factored approximations. These approaches can also be distributed (Zhang & Lin, 2015; Jahani et al., 2020; Haider et al., 2021; Ba et al., 2017; Wu et al., 2017), but their dominant costs arise from iterative solver steps or layerwise curvature factors rather than explicit $B \times B$ sample-space systems. They therefore address a different computational regime from the sample-space bottleneck studied here.

Distributed natural gradient methods. Existing distributed natural-gradient methods in both supervised learning and RL rarely address the inverse Jensen gap induced by averaging local empirical Fisher inverses; instead, they typically avoid or ignore it. In supervised learning, SENG (Yang et al., 2022) assumes that the Woodbury sample-space Gram matrix is block diagonal across workers,

enabling direct uniform averaging of local directions. By contrast, Mu et al. (2022) avoids the inversion gap by aggregating curvature matrices before inversion, thereby recovering the pooled direction exactly but requiring matrix, rather than vector, communication. In RL, FedNPG (Yang et al., 2024) considers tabular softmax policies, for which the natural-gradient direction reduces to the corresponding Q-function and no explicit Fisher inversion is required. MDNPG (Chen et al., 2024) studies decentralized NPG for cooperative multi-agent RL using gradient tracking and averaging of local updates and parameters, but does not explicitly correct the bias arising from local Fisher inversion. The closest method to ours is FedNPG-ADMM (Lan et al., 2023), which targets the pooled undamped empirical NPG direction in a federated setting while communicating only vectors. For a fixed empirical system, multiple ADMM rounds can recover the pooled solution under suitable conditions. In policy optimization, however, the empirical system changes after each policy update, while its algorithm performs only one ADMM round per update. Therefore, it tracks, rather than directly reconstructs, the current pooled NPG direction.

Fisher/Hessian inversion bias. A closely related problem, Hessian inversion bias in distributed second-order optimization, has been studied more extensively. Surrogate sketching (Derezinski et al., 2020a) uses data-dependent non-i.i.d. sampling together with a population-dependent regularization correction. LESS (Derezinski et al., 2021b) requires sketch sizes exceeding the parameter dimension and leverage scores computed from the full dataset, while Newton-LESS (Derezinski et al., 2021a) relaxes this requirement to the effective dimension of the population Hessian. Optimal shrinkage (Zhang & Pilanci, 2023) similarly depends on population-level spectral quantities such as the effective rank. Moreover, these methods typically assume a shared population gradient rather than a stochastic gradient estimated jointly with the curvature. These assumptions are not available in our setting. Determinantal averaging (DA) (Derezinski & Mahoney, 2019) is more closely related, as it requires neither $B > d$, population-dependent correction factors, nor specialized data-dependent sampling. Nevertheless, its distributed Newton application considers local Hessians obtained by subsampling a fixed finite dataset and applies local inverse Hessians to a shared full-data gradient. In contrast, outside distributed second-order optimization, Derezinski et al. (2020b) develop determinant-preserving identities for Poisson random designs with jointly sampled feature–response pairs. Their Poisson determinant identity is closely related to the one used in PDA, but their focus is surrogate regression rather than distributed NPG and its bias.

3 Preliminaries

Notations. We consider a Markov decision process $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, r, \rho_0)$, where $\mathcal{S}$ and $\mathcal{A}$ are the state and action spaces, $\mathcal{P}(s' \mid s, a)$ is the transition kernel, $r : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is the immediate reward, and ρ_0 is the initial-state distribution. Parameterized by $\boldsymbol{\theta} \in \mathbb{R}^d$, the policy $\pi_{\boldsymbol{\theta}}(a \mid s)$ induces trajectories through $\mathrm{S}_0 \sim \rho_0$, $\mathrm{A}_t \sim \pi_{\boldsymbol{\theta}}(\cdot \mid \mathrm{S}_t)$, and $\mathrm{S}_{t+1} \sim \mathcal{P}(\cdot \mid \mathrm{S}_t, \mathrm{A}_t)$. Let $R_t := \sum_{k=0}^{\infty} \gamma^k r(\mathrm{S}_{t+k}, \mathrm{A}_{t+k})$ denote the discounted return, where $\gamma \in [0, 1)$. The action-value, state-value, and advantage functions are $Q^{\pi_{\boldsymbol{\theta}}}(s, a) := \mathbb{E}[R_t \mid \mathrm{S}_t = s, \mathrm{A}_t = a]$, $V^{\pi_{\boldsymbol{\theta}}}(s) := \mathbb{E}[R_t \mid \mathrm{S}_t = s]$, and $A^{\pi_{\boldsymbol{\theta}}}(s, a) := Q^{\pi_{\boldsymbol{\theta}}}(s, a) - V^{\pi_{\boldsymbol{\theta}}}(s)$, respectively, and the objective is $J(\boldsymbol{\theta}) := \mathbb{E}[R_0]$.

Natural Policy Gradients. We define NPG for a potentially continuous state–action space scenario. Let $(\mathcal{X}, \mathscr{X}, \nu_{\boldsymbol{\theta}})$ be the measurable state–action probability space induced by $\pi_{\boldsymbol{\theta}}$, where $\mathcal{X} = \mathcal{S} \times \mathcal{A}$ is the state–action space, $\mathscr{X}$ is the sigma algebra on $\mathcal{X}$, and $\nu_{\boldsymbol{\theta}}$ is the state–action probability measure under $\pi_{\boldsymbol{\theta}}$. Let the advantage function be $y(x) := A^{\pi_{\boldsymbol{\theta}}}(s, a)$. The per sample score function is $\boldsymbol{h}(x) := \nabla_{\boldsymbol{\theta}} \log \pi_{\boldsymbol{\theta}}(a \mid s) \in \mathbb{R}^d$. We define the population Fisher and gradient with an integral over $\mathcal{X}$ to support both discrete and continuous cases,

$$\boldsymbol{F} := \int_{\mathcal{X}} \boldsymbol{h}(x)\boldsymbol{h}(x)^\top \nu_{\boldsymbol{\theta}}(dx), \quad \boldsymbol{g} = \int_{\mathcal{X}} \boldsymbol{h}(x) y(x) \nu_{\boldsymbol{\theta}}(dx). \tag{1}$$

We treat the population damped NPG as the ground truth direction for the empirical estimators to compare against: $\boldsymbol{p} := (\boldsymbol{F} + \lambda \boldsymbol{I})^{-1}\boldsymbol{g}$, where $\lambda > 0$ is the damping parameter.

4 Poissonized Determinantal Averaging

We now introduce the distributed estimator studied in this work. Given a distributed system of N workers, instead of assigning each worker a deterministic local batch size, we Poissonize the number

of samples while preserving the same expected local sample budget B. Specifically, independently across workers,

$$\mathrm{K}_i \sim \mathrm{Poisson}(B), \tag{2}$$

and, conditional on K_i, sample

$$\mathrm{X}_{i1}, \ldots, \mathrm{X}_{i\mathrm{K}_i} \overset{\mathrm{iid}}{\sim} \nu_{\boldsymbol{\theta}}. \tag{3}$$

Thus $B = \mathbb{E}[\mathrm{K}_i]$, while K_i is the realized local batch size. The corresponding per worker Poissonized empirical Fisher and policy gradient are

$$\widehat{\boldsymbol{F}}_i := \frac{1}{B}\sum_{r=1}^{\mathrm{K}_i} \boldsymbol{h}(\mathrm{X}_{ir})\boldsymbol{h}(\mathrm{X}_{ir})^\top, \qquad \widehat{\boldsymbol{g}}_i := \frac{1}{B}\sum_{r=1}^{\mathrm{K}_i} \boldsymbol{h}(\mathrm{X}_{ir})y(\mathrm{X}_{ir}). \tag{4}$$

Importantly, both quantities are normalized by the Poisson intensity B, rather than by the realized count K_i. The two estimators in Equation (4) may be formed from the same samples. Each worker computes the damped empirical NPG direction

$$\widehat{\boldsymbol{p}}_i := (\widehat{\boldsymbol{F}}_i + \lambda \boldsymbol{I})^{-1}\widehat{\boldsymbol{g}}_i. \tag{5}$$

We associate worker i with the determinantal weight $D_i := \det(\widehat{\boldsymbol{F}}_i + \lambda \boldsymbol{I})$ and aggregate the local directions according to

$$\widehat{\boldsymbol{p}}_{\mathrm{PDA},N} := \frac{\sum_{i=1}^N D_i \widehat{\boldsymbol{p}}_i}{\sum_{i=1}^N D_i} = \sum_{i=1}^N \frac{D_i}{\sum_{i=1}^N D_i}\widehat{\boldsymbol{p}}_i \tag{6}$$

The role of Poissonization is not merely to randomize the local batch size. As discussed in the following section, its factorial-moment structure and Mecke identity establish the bias reduction property of determinantal averaging, even when the Fisher and gradient share the same samples.

A practical way of estimating $\widehat{\boldsymbol{p}}_i$ is leveraging the Woodbury matrix identity to formulate NPG from a potentially intractable parameter space system to a tractable sample space system (see Appendix G for details). Let $\boldsymbol{H}_i := [\boldsymbol{h}(\mathrm{X}_{i1}) \quad \cdots \quad \boldsymbol{h}(\mathrm{X}_{i\mathrm{K}_i})] \in \mathbb{R}^{d\times \mathrm{K}_i}$, the Woodbury matrix identity yields

$$\widehat{\boldsymbol{p}}_i = \frac{1}{\lambda}\left[\widehat{\boldsymbol{g}}_i - \frac{1}{B}\boldsymbol{H}_i\left(\lambda \boldsymbol{I}_{\mathrm{K}_i} + \frac{1}{B}\boldsymbol{H}_i^\top \boldsymbol{H}_i\right)^{-1}\boldsymbol{H}_i^\top \widehat{\boldsymbol{g}}_i\right]. \tag{7}$$

The determinantal weight can be evaluated entirely in the sample space.

$$D_i = \det\left(\lambda \boldsymbol{I}_d + \frac{1}{B}\boldsymbol{H}_i\boldsymbol{H}_i^\top\right) = \lambda^{d-\mathrm{K}_i}\det\left(\lambda \boldsymbol{I}_{\mathrm{K}_i} + \frac{1}{B}\boldsymbol{H}_i^\top \boldsymbol{H}_i\right). \tag{8}$$

Taking logarithms gives $\log D_i = d\log\lambda + \log\det\left(\lambda \boldsymbol{I}_{\mathrm{K}_i} + \frac{1}{B}\boldsymbol{H}_i^\top \boldsymbol{H}_i\right) - \mathrm{K}_i \log\lambda$. Since the term $d\log\lambda$ is common to all workers, it cancels when the determinantal weights are normalized. Let

$$s_i := \log\det\left(\lambda \boldsymbol{I}_{\mathrm{K}_i} + \frac{1}{B}\boldsymbol{H}_i^\top \boldsymbol{H}_i\right) - \mathrm{K}_i\log\lambda. \tag{9}$$

The determinantal scores are normalized using a softmax, and the resulting weights are used to aggregate the local directions:

$$w_i := \frac{\exp(s_i)}{\sum_{j=1}^N \exp(s_j)}, \qquad \widehat{\boldsymbol{p}}_{\mathrm{PDA},N} = \sum_{i=1}^N w_i \widehat{\boldsymbol{p}}_i. \tag{10}$$

5 Theoretical Analysis of Poissonized Determinantal Averaging

In this section, we show how closely PDA reproduces the expected behavior of pooled empirical NPG while retaining distributed local computation. Bias provides a natural basis for this comparison: relative to the population damped NPG direction $\boldsymbol{p}$, it captures the systematic error of each estimator, and differences in bias directly characterize the gap between their expected updates. We therefore compare local estimator $\widehat{\boldsymbol{p}}_i$, pooled estimator $\widehat{\boldsymbol{p}}_{\mathrm{pool}}$, uniform averaging $\widehat{\boldsymbol{p}}_{\mathrm{avg},N}$, and PDA $\widehat{\boldsymbol{p}}_{\mathrm{PDA},N}$ through their respective biases under the Poisson sampling model introduced in Section 4.

We begin by characterizing the finite-sample bias of a single-worker empirical NPG estimator under the Poisson sampling model and then use this result to analyze its distributed counterparts.

Proposition 5.1 (Finite-sample bias of empirical NPG)**.** *Suppose the empirical Fisher and gradient are constructed from the same Poisson samples as in Section 4. If*

$$\int_{\mathcal{X}} \left(\|\boldsymbol{h}(x)\|^8 + \|\boldsymbol{h}(x)\|^4 |y(x)|^4 \right) \nu_{\boldsymbol{\theta}}(dx) < \infty, \tag{11}$$

then, as $B \to \infty$*,*

$$\mathrm{Bias}(\widehat{\boldsymbol{p}}_i) = \mathbb{E}[\widehat{\boldsymbol{p}}_i] - \boldsymbol{p} = \frac{1}{B}\boldsymbol{c} + \mathcal{O}(B^{-2}), \tag{12}$$

where

$$\boldsymbol{c} := \boldsymbol{G} \int_{\mathcal{X}} \boldsymbol{h}(x)\boldsymbol{h}(x)^\top \boldsymbol{G} \left[\boldsymbol{h}(x)\boldsymbol{h}(x)^\top \boldsymbol{p} - \boldsymbol{h}(x)y(x) \right] \nu_{\boldsymbol{\theta}}(dx). \tag{13}$$

defines the leading bias coefficient, and $\boldsymbol{G} := (\boldsymbol{F} + \lambda \boldsymbol{I})^{-1}$.

The proof is provided in Appendix A.2. Proposition 5.1 is an asymptotic expansion of the bias with respect to the expected Poisson batch size B. Thus, when $\boldsymbol{c} \neq \boldsymbol{0}$, the term $\boldsymbol{c}/B$ dominates the $\mathcal{O}(B^{-2})$ remainder, and the bias has asymptotic order B^{-1}.

To better understand the source of this bias, the leading term can equivalently be written as

$$\frac{\boldsymbol{c}}{B} = \boldsymbol{G}\mathbb{E}[\boldsymbol{\Delta}_i \boldsymbol{G} \boldsymbol{\Delta}_i]\boldsymbol{p} - \boldsymbol{G}\mathbb{E}[\boldsymbol{\Delta}_i \boldsymbol{G} \boldsymbol{\delta}_i], \tag{14}$$

where $\boldsymbol{\Delta}_i := \widehat{\boldsymbol{F}}_i - \boldsymbol{F}$ and $\boldsymbol{\delta}_i := \widehat{\boldsymbol{g}}_i - \boldsymbol{g}$. The first term is the inversion–curvature contribution. It arises from the nonlinear dependence of the NPG direction on the empirical Fisher and reflects the inverse Jensen gap, $\mathbb{E}[(\widehat{\boldsymbol{F}}_i + \lambda \boldsymbol{I})^{-1}] \neq (\boldsymbol{F} + \lambda \boldsymbol{I})^{-1}$. The second term arises because the empirical Fisher and policy gradient are estimated from the same samples. Classical determinantal averaging (Derezinski & Mahoney, 2019) considers a random inverse applied to a fixed vector so there is no analogous coupling term. In contrast, we later show that PDA accounts for both contributions in the same-sample empirical NPG setting.

Recall the distributed setting introduced in Section 4, in which each of the N workers computes a local empirical NPG direction using an expected batch size of B. With Proposition 5.1, each local estimator has bias $\mathrm{Bias}(\widehat{\boldsymbol{p}}_i)$. Pooling from all nodes, we define the pooled empirical NPG as follows

$$\widehat{\boldsymbol{p}}_{\mathrm{pool}} := \left(\lambda \boldsymbol{I} + \frac{1}{N} \sum_{i=1}^{N} \widehat{\boldsymbol{F}}_i \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \widehat{\boldsymbol{g}}_i \right),$$

and it yields the following bias.

Corollary 5.2 (Bias of pooled empirical NPG)**.** *Suppose the conditions of Proposition 5.1 hold. The superposition of* N *independent Poisson processes of intensity* $B\nu_{\boldsymbol{\theta}}$ *is a Poisson process of intensity* $NB\nu_{\boldsymbol{\theta}}$*. Therefore, the pooled estimator is distributionally equivalent to a single Poissonized empirical NPG estimator with expected sample size* NB*. Consequently, as* $NB \to \infty$*,*

$$\mathrm{Bias}(\widehat{\boldsymbol{p}}_{\mathrm{pool}}) = \mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{pool}}] - \boldsymbol{p} = \frac{1}{NB}\boldsymbol{c} + \mathcal{O}((NB)^{-2}). \tag{15}$$

The proof is provided in Appendix A.3. The pooled estimator therefore provides a natural statistical target: by using the combined expected sample budget, it reduces the leading bias from $\mathcal{O}(B^{-1})$ to $\mathcal{O}((NB)^{-1})$. Our goal is therefore to match the pooled estimator leading bias while computing each local NPG direction independently.

The following corollary explains why uniform averaging, $\widehat{\boldsymbol{p}}_{\mathrm{avg},N} := \frac{1}{N}\sum_{i=1}^{N} \widehat{\boldsymbol{p}}_i$, despite being standard in first-order distributed optimization, does not achieve this objective.

Corollary 5.3 (Uniform averaging retains the local bias)**.** *Since the workers are identically distributed,* $\mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{avg},N}] = \mathbb{E}[\widehat{\boldsymbol{p}}_i]$*. Consequently, as* $B \to \infty$

$$\mathrm{Bias}(\widehat{\boldsymbol{p}}_{\mathrm{avg},N}) = \mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{avg},N}] - \boldsymbol{p} = \frac{1}{B}\boldsymbol{c} + \mathcal{O}(B^{-2}), \tag{16}$$

regardless of the number of workers N.

See proof in Appendix A.3. Increasing the number of workers through uniform averaging does not change the leading bias, and hence not recovering the leading bias behavior of the pooled estimator.

In contrast, with Poissonization and determinant weighting, PDA can recover the leading-order pooled bias. Recall from Equation (6) that PDA is a ratio of empirical averages. For any fixed expected local batch size B, the law of large numbers implies that, as $N \to \infty$, this ratio converges to $\mathbb{E}[D_i \widehat{\boldsymbol{p}}_i]/\mathbb{E}[D_i]$. Let $D := \det(\boldsymbol{F} + \lambda \boldsymbol{I})$. Thus, the exact identities $\mathbb{E}[D_i] = D$ and $\mathbb{E}[D_i \widehat{\boldsymbol{p}}_i] = D\boldsymbol{p}$ make this limit equal to $\boldsymbol{p}$.

Poissonization makes these identities exact. The first identity is established by Derezinski et al. (2020b). In the determinant expansion, each order-k term contains $B^{-k}\mathbb{E}[(\mathrm{K}_i)_k]$, while for $\mathrm{K}_i \sim \mathrm{Poisson}(B)$, $\mathbb{E}[(\mathrm{K}_i)_k] = B^k$. The cancellation of these factors serves as the basis for the first identity. Second, we further show that for shared Fisher-gradient samples, the Poisson-Mecke formula (Last & Penrose, 2017, Chapter 4) isolates a single sample from the Poissonized batch, while adjugate rank-one invariance ensures that its Fisher contribution has no effect on the corresponding adjugate-weighted score term, yielding the second identity. With deterministic $\mathrm{K}_i = B$, however, $\mathbb{E}[(\mathrm{K}_i)_k] = (B)_k \neq B^k$ for $k \geq 2$, so the cancellation underlying the first identity is lost. Isolating one sample also changes the remaining batch size from B to $B-1$, so the argument underlying the second identity no longer applies. Consequently, the exact identities and resulting many-worker guarantee do not extend directly to deterministic fixed-size sampling. Detailed derivations are in Appendix B and Appendix F. Together, these identities result in $\mathbb{E}[D_i(\widehat{\boldsymbol{p}}_i - \boldsymbol{p})] = 0$, and hence PDA exactly centers the weighted local NPG error, which forms the basis of its bias analysis. We first characterize PDA's bias as $B \to \infty$.

Theorem 5.4 (PDA's leading-order bias matching with pooled NPG)**.** *Under the same-sample Poisson construction of Section 4 and the moment condition in Equation (11), for every fixed $N \geq 1$, as $B \to \infty$,*

$$\mathrm{Bias}(\widehat{\boldsymbol{p}}_{\mathrm{PDA},N}) = \mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{PDA},N}] - \boldsymbol{p} = \frac{\boldsymbol{c}}{NB} + \mathcal{O}_N\left(B^{-2}\right). \tag{17}$$

Combining this expansion with Corollary 5.2 further gives

$$\mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{PDA},N}] - \mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{pool}}] = \mathcal{O}_N\left(B^{-2}\right). \tag{18}$$

The proof is provided in Appendix D. Under stronger moment conditions, Appendix E further sharpens the remainder to $\mathcal{O}(1/(NB^2))$ uniformly over $N \geq 1$. We next fix B and let $N \to \infty$.

Corollary 5.5 (Many-worker consistency)**.** *For any fixed $B \geq 1$, suppose the worker numerator–denominator pairs are i.i.d. and integrable. Then $\widehat{\boldsymbol{p}}_{\mathrm{PDA},N} \xrightarrow[N\to\infty]{\text{a.s.}} \boldsymbol{p}$. Moreover, if for some $\epsilon > 0$, $\mathbb{E}[D_i|\widehat{\boldsymbol{g}}_i|^{1+\epsilon}] < \infty$, then $\mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{PDA},N}] \longrightarrow \boldsymbol{p}$.*

See proof in Appendix C.1. In Appendix C.2, we further show that under the stronger moment conditions, the fixed-B bias admits the expansion $\mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{PDA},N}] - \boldsymbol{p} = \boldsymbol{\kappa}(B)/N + \mathcal{O}_B(N^{-2})$ as $N \to \infty$, and is therefore $\mathcal{O}_B(N^{-1})$. Although $\boldsymbol{\kappa}(B)$ need not equal the pooled coefficient $\boldsymbol{c}/B$ at finite B, $\boldsymbol{\kappa}(B) = \boldsymbol{c}/B + \mathcal{O}(B^{-2})$ as $B \to \infty$.

6 Experimental results

To evaluate the empirical effectiveness, we compare PDA against four baselines. *Uniform averaging* is included to empirically test our theoretical finding that averaging local empirical NPG directions does not eliminate their bias. *FedNPG-ADMM*, discussed in Section 2, is a closely related data-parallel distributed NPG method. Because it is agnostic to the particular NPG solver, it can be combined directly with a Woodbury-based implementation. We also consider the *Single NPG* and *Pooled NPG*, where the former computes an empirical NPG direction using the same batch size as one distributed worker and the latter computes a single direction using all samples collected across the workers. These two baselines provide practical reference points: an effective distributed method should improve upon the local estimator and ideally approach the pooled estimator. To isolate the effect of the distributed aggregation rule, all methods are implemented on top of RAT (Sun, 2026), a Woodbury based NPG algorithm.

We evaluate the methods on four MuJoCo tasks (Todorov et al., 2012), involving continuous control of simulated physical systems from vectorized observations, and four Procgen tasks (Cobbe et al.,

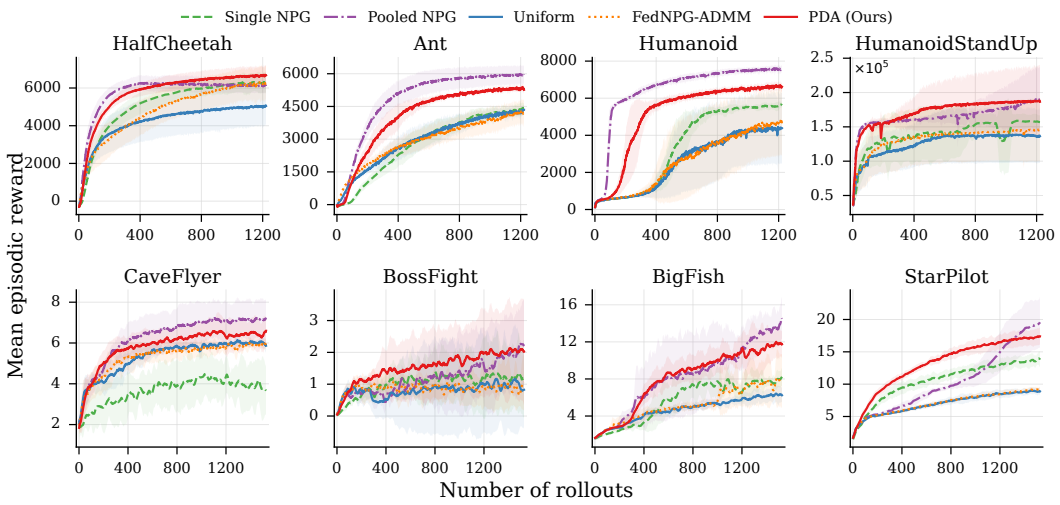


Figure 1: Training performance for $N = 8$ workers and $B = 512$. Curves show the mean and one standard deviation over 5 seeds. Procgen curves are smoothed with a 25-update running average.

2019), involving discrete control from pixel observations. For all cases, actor and critic share the same neural network architecture but hold separate parameters, and we train the critic centrally.

Unless otherwise stated, we use Fisher damping coefficient 0.1 for MuJoCo and 0.5 for Procgen, $N = 8$, and $B = 512$. PDA and the single and pooled NPG estimators follow the Poissonized sampling model to match the setting in our theoretical analysis. Thus, $B = 512$ is the expected local batch size, and the pooled estimator has expected batch size $NB = 4096$. Uniform averaging and FedNPG-ADMM instead use fixed local batches of $B = 512$ samples, reflecting their conventional implementations. Nevertheless, we separately provide a four-way estimator study in Figure 3, isolating the effects of Poissonization and determinant weighting. We also present a training ablation of Poissonization in Appendix I.1, showing that Poissonization generally yields higher training performance than fixed-size sampling under determinantal averaging. Full implementation details and hyperparameters are provided in Appendix H.

Training performance. Figure 1 reports the training mean episodic rewards across the eight tasks. PDA is the strongest distributed method across all eight environments and most consistently approaches the performance of Pooled NPG. It substantially outperforms uniform averaging on most tasks. On certain tasks, the mean return achieved by PDA even exceeds that of Pooled NPG. One possible explanation is that the additional stochasticity of determinant-weighted aggregation provides an implicit regularization or exploration effect during training. Uniform averaging, by contrast, provides limited and inconsistent improvement over Single NPG. It outperforms Single NPG on only a small subset of the tasks and performs similarly or worse on the remainder. This behavior is consistent with our theoretical analysis: averaging retains their systematic local-estimator bias, preventing the additional workers from being fully translated into improved policy updates, demonstrating the practical limitation motivating our bias-aware aggregation method. FedNPG-ADMM generally performs similarly to uniform averaging and remains below PDA across the evaluated tasks. Following its original formulation, FedNPG-ADMM performs a single ADMM round per policy update to maintain communication efficiency. Consequently, the consensus problem is not solved to convergence at each update, which may limit how closely its direction approximates Pooled NPG and helps explain its similarity to uniform averaging.

Scaling the number of workers. Figure 2 evaluates PDA on Ant and StarPilot with damping $\lambda = 0.1$, expected local batch size $B = 512$, and varying N. Performance improves consistently through $N = 16$ on Ant and $N = 32$ on StarPilot, while the next evaluated worker count achieves similar performance rather than further gains. This behavior is consistent with diminishing returns: for fixed B, PDA's finite-worker ratio bias decays as $\mathcal{O}(1/N)$, so each doubling of N produces a smaller absolute correction. Once aggregation error is no longer the dominant source of error, further

improvements in direction estimation may not translate into measurable return gains. Nevertheless, PDA converts additional workers into performance improvements over a substantial range while keeping each sample-space problem at the local size B, which is useful when per-device memory or computation is limited.

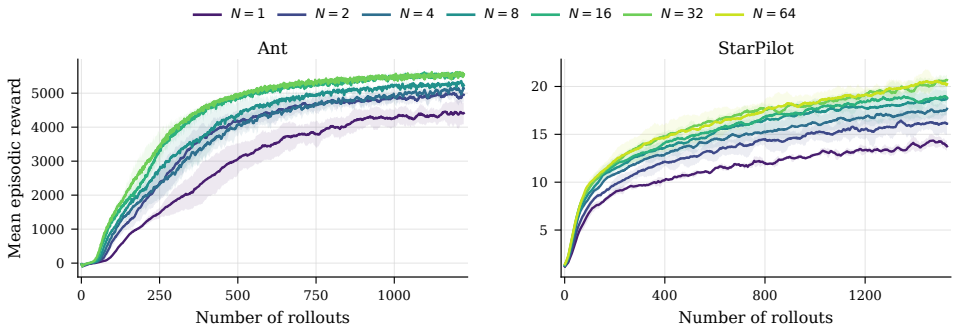
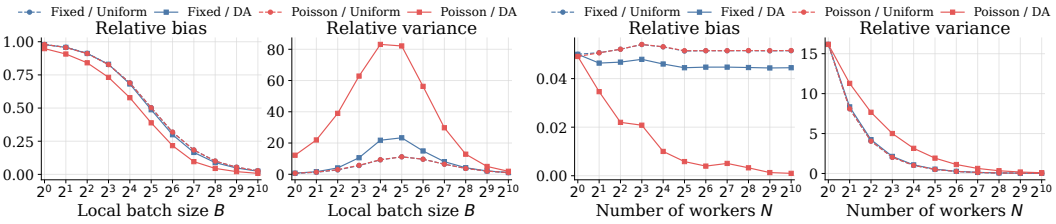


Figure 2: Effect of the number of workers N on PDA with $B = 512$, $\lambda = 0.1$ on Ant and StarPilot.

Bias–variance analysis. We isolate the estimator behavior at a fixed checkpoint of a trained StarPilot policy (Figure 3). A high-accuracy reference NPG direction is constructed from 12,800 transitions using an FP64 Cholesky solve. We evaluate the four combinations of fixed or Poisson sampling and uniform or determinantal averaging in two complementary regimes: fixed $N = 8$ with $B \in \{2^0, \ldots, 2^{10}\}$, and fixed $B = 512$ with $N \in \{2^0, \ldots, 2^{10}\}$, using $\lambda = 0.5$. Each configuration is repeated 8000 times. Let $\widehat{\mathbf{u}}_r$ be the aggregated NPG direction in repetition r, $\overline{\mathbf{u}} = R^{-1} \sum_{r=1}^{R} \widehat{\mathbf{u}}_r$, and $\mathbf{u}_\star$ the reference direction. We report the relative variance $\widehat{V}_{\mathrm{rel}} = \left[(R-1)\|\mathbf{u}_\star\|_2^2\right]^{-1} \sum_{r=1}^{R} \|\widehat{\mathbf{u}}_r - \overline{\mathbf{u}}\|_2^2$ and relative bias $\widehat{b}_{\mathrm{rel}} = \sqrt{\|\overline{\mathbf{u}} - \mathbf{u}_\star\|_2^2 / \|\mathbf{u}_\star\|_2^2 - \widehat{V}_{\mathrm{rel}}/R}$.



(a) $B \in \{2^0, \ldots, 2^{10}\}$, $N = 8$ (b) $B = 512$, $N \in \{2^0, \ldots, 2^{10}\}$

Figure 3: Relative bias and variance of aggregated NPG directions on StarPilot.

The bias trends qualitatively agree with the two theoretical regimes. For fixed N, PDA's bias decreases as B grows, consistent with its large-batch bias expansion. Uniform averaging also improves as each local estimator is less biased. More notably, for fixed B, its bias continues to decrease as workers are added, consistent with many-worker consistency, whereas the other schemes approach higher plateaus. This regime is particularly relevant in practice because scaling N is computationally cheaper than scaling B. Across both sweeps, PDA generally attains the lowest bias, confirming the joint role of its two components. Specifically, determinant weighting with fixed-size batches provides a modest reduction, while Poisson sampling with uniform averaging behaves similarly to fixed-batch uniform averaging. The strongest reduction arises when Poissonization and determinant weighting are combined.

Variance is reported as a secondary diagnostic. PDA generally exhibits the largest variance. Compared to uniform averaging, uneven normalized determinant weights introduce an additional source of fluctuation that can be amplified by variation in the realized Poisson batch size, potentially leading to PDA's higher variance. Relative variance also dominates squared relative bias in the corresponding mean squared error. Nevertheless, the earlier training experiments show that PDA most closely

matches the pooled NPG baseline, suggesting that correcting systematic directional bias appears more important for training performance than reducing one-step estimator variance in the regimes studied.

Computational scaling. Figure 4 compares the expected FLOPs and nominal peak memory per update for four implementations. PDA (parallel) reports the per-worker cost under concurrent execution, whereas PDA (serial) executes the N local computations sequentially on one device, representing a single-machine simulation of the distributed procedure. Each worker's local sample count is modeled as $\mathrm{K}_i \sim \mathrm{Poisson}(B)$ with $B = 512$. Specifically, expected FLOPs are computed from the exact Poisson moments without sampling realized counts, while memory is evaluated at $\mathrm{K}_i = B$. The results are derived analytically from operation counts and tensor dimensions rather than wall-clock measurements, avoiding hardware- and implementation-dependent effects.

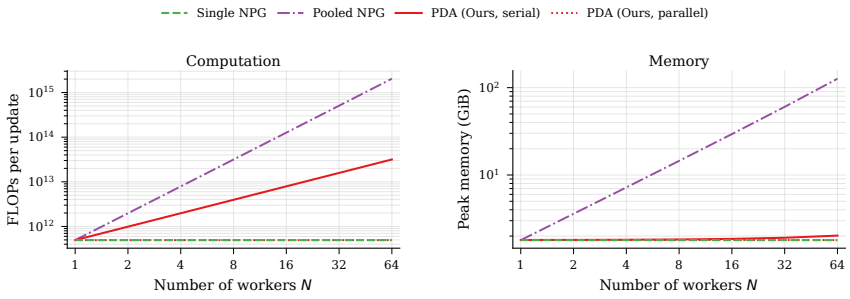


Figure 4: Analytical computation and memory cost per update as a function of the number of workers N, with $B = 512$ and $P = 938{,}719$ policy parameters.

Consistent with the sample-space complexity discussed in Section 2, the computation of Pooled NPG grows approximately quadratically with N, while its memory grows approximately linearly. Parallel PDA remains close to the constant cost of Single NPG because each worker continues to process only B samples. Serial PDA requires computation that grows linearly with N, but keeps the local solver memory nearly constant. Together with the findings in Figure 2, this provides a practical route for scaling Woodbury-based empirical NPG: parallel execution maintains approximately constant per-worker cost, subject to communication overhead; when only one device is available, the same local computations can instead be executed serially, trading additional computation time for improved performance while retaining approximately single-worker solver memory (see Appendix I.2 for measured single-GPU results).

7 Limitations and Conclusion

Limitations. Our theory compares PDA and pooled empirical NPG under a common damping parameter λ and targets the population damped NPG direction. In practice, however, the smaller local Fisher estimates used by PDA may require stronger damping for numerical stability than the pooled Fisher. Moreover, under Poissonization, small damping can amplify variation in determinant weights across workers, limiting how aggressively λ can be reduced. Extending the theory to adaptive damping is left for future work.

Conclusion. We studied data-parallel empirical NPG in the sample-space (Woodbury) regime and showed that finite-sample bias is a fundamental obstacle to naive distribution. Uniform averaging of local directions keeps the $\mathcal{O}(B^{-1})$ local bias no matter how many workers are added. We proposed Poissonized Determinantal Averaging (PDA), which draws Poisson local sample counts and combines local NPG directions with normalized determinant weights. PDA keeps sample-space solvers fully local and communicates only one direction and one scalar per worker. Poissonization yields exact determinant and adjugate identities even when the Fisher and policy gradient share samples. As a result, PDA reduces the leading bias to $\mathcal{O}((NB)^{-1})$ and recovers the full leading bias coefficient of pooled empirical NPG. On MuJoCo and Procgen, PDA consistently outperforms uniform averaging and FedNPG-ADMM and most closely approaches pooled NPG. It does so at roughly single-worker per-device cost, and can run serially at near-constant solver memory.

AI USE STATEMENT

In this work, generative AI assisted us in reviewing and proofreading theoretical models, checking mathematical formulations, refining hypotheses, implementing methods, polishing writing, and writing code. Generative AI was not used to provide critical ingredients for proving mathematical claims, generate synthetic datasets, design or provide feedback on research methodology or experiments, assist with translation, clean or reformat datasets, support qualitative or thematic data analysis, or interpret results. We manually reviewed and verified all AI-assisted theoretical content, code implementations, and writing. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

PDA is described in Section 4, and the main theoretical results are presented in Section 5. Detailed proofs and derivations are provided in Appendices A to G. Implementation and experimental details are documented in Appendix H. Code is provided in the supplementary materials.

REFERENCES

Shun-ichi Amari, Hyeyoung Park, and Kenji Fukumizu. Adaptive method of realizing natural gradient learning for multilayer perceptrons. *Neural Comput.*, 12(6):1399–1409, 2000. doi: 10.1162/089976600300015420. URL https://doi.org/10.1162/089976600300015420.

Jimmy Ba, Roger B. Grosse, and James Martens. Distributed second-order optimization using kronecker-factored approximations. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. URL https://openreview.net/forum?id=SkkTMpjex.

Jinchi Chen, Jie Feng, Weiguo Gao, and Ke Wei. Decentralized natural policy gradient with variance reduction for collaborative multi-agent reinforcement learning. *J. Mach. Learn. Res.*, 25:172:1–172:49, 2024. URL https://jmlr.org/papers/v25/22-1036.html.

Karl Cobbe, Christopher Hesse, Jacob Hilton, and John Schulman. Leveraging procedural generation to benchmark reinforcement learning. *arXiv preprint arXiv:1912.01588*, 2019.

Ofer Dekel, Ran Gilad-Bachrach, Ohad Shamir, and Lin Xiao. Optimal distributed online prediction using mini-batches. *J. Mach. Learn. Res.*, 13:165–202, 2012. doi: 10.5555/2503308.2188391. URL https://dl.acm.org/doi/10.5555/2503308.2188391.

Michal Derezinski and Michael W. Mahoney. Distributed estimation of the inverse hessian by determinantal averaging. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pp. 11401–11411, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/b1b20d09041289e6c3fbb81850c5da54-Abstract.html.

Michal Derezinski, Burak Bartan, Mert Pilanci, and Michael W. Mahoney. Debiasing distributed second order optimization with surrogate sketching and scaled regularization. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020a. URL https://proceedings.neurips.cc/paper/2020/hash/4a46fbfca3f1465a27b210f4bdfe6ab3-Abstract.html.

Michal Derezinski, Feynman T. Liang, and Michael W. Mahoney. Exact expressions for double descent and implicit regularization via surrogate random design. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020,*

virtual, 2020b. URL https://proceedings.neurips.cc/paper/2020/hash/37740d59bb0eb7b4493725b2e0e5289b-Abstract.html.

Michal Derezinski, Jonathan Lacotte, Mert Pilanci, and Michael W. Mahoney. Newton-less: Sparsification without trade-offs for the sketched newton update. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pp. 2835–2847, 2021a. URL https://proceedings.neurips.cc/paper/2021/hash/16837163fee34175358a47e0b51485ff-Abstract.html.

Michal Derezinski, Zhenyu Liao, Edgar Dobriban, and Michael W. Mahoney. Sparse sketches with small inversion bias. In Mikhail Belkin and Samory Kpotufe (eds.), *Conference on Learning Theory, COLT 2021, 15-19 August 2021, Boulder, Colorado, USA*, volume 134 of *Proceedings of Machine Learning Research*, pp. 1467–1510. PMLR, 2021b. URL http://proceedings.mlr.press/v134/derezinski21a.html.

Priya Goyal, Piotr Dollár, Ross B. Girshick, Pieter Noordhuis, Lukasz Wesolowski, Aapo Kyrola, Andrew Tulloch, Yangqing Jia, and Kaiming He. Accurate, large minibatch SGD: training imagenet in 1 hour. *CoRR*, abs/1706.02677, 2017. URL http://arxiv.org/abs/1706.02677.

Andrés Guzmán-Cordero, Felix Dangel, Gil Goldshlager, and Marius Zeinhofer. Improving energy natural gradient descent through woodbury, momentum, and randomization. In Danielle Belgrave, Cheng Zhang, Laura N. Montoya, Hsuan-Tien Lin, Razvan Pascanu, Piotr Koniusz, Marzyeh Ghassemi, Nancy Chen, Iván Vladimir Meza Ruíz, and Arturo Loaiza-Bonilla (eds.), *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2025, NeurIPS 2025, San Diego, CA, USA, December 2-7, 2025 / Mexico City, Mexico, November 30 - December 5, 2025*, 2025. URL http://papers.nips.cc/paper_files/paper/2025/hash/a5809a25c65040fd43e146126b45b150-Abstract-Conference.html.

Adnan Haider, Chao Zhang, Florian L. Kreyssig, and Philip C. Woodland. A distributed optimisation framework combining natural gradient with hessian-free for discriminative sequence training. *Neural Networks*, 143:537–549, 2021. doi: 10.1016/J.NEUNET.2021.05.011. URL https://doi.org/10.1016/j.neunet.2021.05.011.

Majid Jahani, Xi He, Chenxin Ma, Aryan Mokhtari, Dheevatsa Mudigere, Alejandro Ribeiro, and Martin Takác. Efficient distributed hessian free algorithm for large-scale empirical risk minimization via accumulating sample strategy. In Silvia Chiappa and Roberto Calandra (eds.), *The 23rd International Conference on Artificial Intelligence and Statistics, AISTATS 2020, 26-28 August 2020, Online [Palermo, Sicily, Italy]*, volume 108 of *Proceedings of Machine Learning Research*, pp. 2634–2644. PMLR, 2020. URL http://proceedings.mlr.press/v108/jahani20a.html.

Sham M. Kakade. A natural policy gradient. In Thomas G. Dietterich, Suzanna Becker, and Zoubin Ghahramani (eds.), *Advances in Neural Information Processing Systems 14 [Neural Information Processing Systems: Natural and Synthetic, NIPS 2001, December 3-8, 2001, Vancouver, British Columbia, Canada]*, pp. 1531–1538. MIT Press, 2001. URL https://proceedings.neurips.cc/paper/2001/hash/4b86abe48d358ecf194c56c69108433e-Abstract.html.

Guangchen Lan, Han Wang, James Anderson, Christopher G. Brinton, and Vaneet Aggarwal. Improved communication efficiency in federated natural policy gradient via admm-based gradient updates. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. URL http://papers.nips.cc/paper_files/paper/2023/hash/bc6a1f968f8b1dae3e880f3f723d7d46-Abstract-Conference.html.

Günter Last and Mathew Penrose. *Lectures on the Poisson Process*, volume 7 of *Institute of Mathematical Statistics Textbooks*. Cambridge University Press, 2017. doi: 10.1017/9781316104477. URL https://doi.org/10.1017/9781316104477.

James Martens and Roger B. Grosse. Optimizing neural networks with kronecker-factored approximate curvature. In Francis R. Bach and David M. Blei (eds.), *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, volume 37 of *JMLR Workshop and Conference Proceedings*, pp. 2408–2417. JMLR.org, 2015. URL http://proceedings.mlr.press/v37/martens15.html.

Baorun Mu, Saeed Soori, Bugra Can, Mert Gürbüzbalaban, and Maryam Mehri Dehnavi. Hylo: A hybrid low-rank natural gradient descent method. In Felix Wolf, Sameer Shende, Candace Culhane, Sadaf R. Alam, and Heike Jagode (eds.), *SC22: International Conference for High Performance Computing, Networking, Storage and Analysis, Dallas, TX, USA, November 13-18, 2022*, pp. 47:1–47:16. IEEE, 2022. doi: 10.1109/SC41404.2022.00052. URL https://doi.org/10.1109/SC41404.2022.00052.

Nicolas Le Roux, Pierre-Antoine Manzagol, and Yoshua Bengio. Topmoumoute online natural gradient algorithm. In John C. Platt, Daphne Koller, Yoram Singer, and Sam T. Roweis (eds.), *Advances in Neural Information Processing Systems 20, Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 3-6, 2007*, pp. 849–856. Curran Associates, Inc., 2007. URL https://proceedings.neurips.cc/paper/2007/hash/9f61408e3afb633e50cdf1b20de6f466-Abstract.html.

John Schulman, Sergey Levine, Pieter Abbeel, Michael I. Jordan, and Philipp Moritz. Trust region policy optimization. In Francis R. Bach and David M. Blei (eds.), *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, volume 37 of *JMLR Workshop and Conference Proceedings*, pp. 1889–1897. JMLR.org, 2015. URL http://proceedings.mlr.press/v37/schulman15.html.

Mingfei Sun. Randomized advantage transformation (RAT): computing natural policy gradients via direct backpropagation. *CoRR*, abs/2605.18591, 2026. doi: 10.48550/ARXIV.2605.18591. URL https://doi.org/10.48550/arXiv.2605.18591.

Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5026–5033. IEEE, 2012. doi: 10.1109/IROS.2012.6386109.

Xiaodong Wu, Wenyi Yu, Chao Zhang, and Philip C. Woodland. An improved empirical fisher approximation for natural gradient descent. In Amir Globersons, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang (eds.), *Advances in Neural Information Processing Systems 37: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024. URL http://papers.nips.cc/paper_files/paper/2024/hash/f23098fa0cfcdef0e743b134d380eeb9-Abstract-Conference.html.

Yuhuai Wu, Elman Mansimov, Roger B. Grosse, Shun Liao, and Jimmy Ba. Scalable trust-region method for deep reinforcement learning using kronecker-factored approximation. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pp. 5279–5288, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/361440528766bbaaaa1901845cf4152b-Abstract.html.

Minghan Yang, Dong Xu, Zaiwen Wen, Mengyun Chen, and Pengxiang Xu. Sketch-based empirical natural gradient methods for deep learning. *J. Sci. Comput.*, 92(3):94, 2022. doi: 10.1007/S10915-022-01911-X. URL https://doi.org/10.1007/s10915-022-01911-x.

Tong Yang, Shicong Cen, Yuting Wei, Yuxin Chen, and Yuejie Chi. Federated natural policy gradient and actor critic methods for multi-task reinforcement learning. In Amir Globersons, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang (eds.), *Advances in Neural Information Processing Systems 37: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024. URL http://papers.nips.cc/paper_files/paper/2024/hash/dbdea7859f1d2fc10f2c9e79b8f5ae54-Abstract-Conference.html.

Fangzhao Zhang and Mert Pilanci. Optimal shrinkage for distributed second-order optimization. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pp. 41523–41549. PMLR, 2023. URL https://proceedings.mlr.press/v202/zhang23ah.html.

Yuchen Zhang and Xiao Lin. Disco: Distributed optimization for self-concordant empirical loss. In Francis R. Bach and David M. Blei (eds.), *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, volume 37 of *JMLR Workshop and Conference Proceedings*, pp. 362–370. JMLR.org, 2015. URL http://proceedings.mlr.press/v37/zhangb15.html.

APPENDIX CONTENTS

A Local, uniform-average, and pooled bias **15**
A.1 Moment identities used in the proof . . . 15
A.2 Proof of Proposition 5.1 . . . 15
A.3 Proofs of Corollary 5.3 and Corollary 5.2 . . . 15
A.4 Proof of Lemma A.1 . . . 16

B Exact determinant, adjugate, and shared-sample identities **16**
B.1 A determinant identity with a deterministic base . . . 16
B.2 Poisson adjugate identity . . . 17
B.3 Shared-sample determinant-weighted identity . . . 17

C Finite-worker ratio bias and many-worker consistency **18**
C.1 Proof of Corollary 5.5 . . . 19
C.2 An explicit expansion when the Poisson intensity is fixed . . . 19

D Proof of Theorem 5.4 **20**

E Uniform joint worker–batch bounds and leading coefficients **21**
E.1 Nonasymptotic joint worker–batch bias bound . . . 21
E.2 A concrete sufficient condition for the uniform expansion . . . 21
E.3 Uniform-in-N refinement . . . 22
E.4 Proof of Lemma E.2 . . . 23

F Fixed-size batches and conditionally unbiased gradients **24**
F.1 Deterministic batch sizes . . . 24
F.2 Conditionally unbiased or independently sampled gradients . . . 24

G Sample-space identities and empty-batch conventions **25**

H Experimental Details **26**

I Additional Experimental Results **27**
I.1 Ablation of Poissonization . . . 27
I.2 Measured wall-clock time and memory on a single GPU . . . 27

A LOCAL, UNIFORM-AVERAGE, AND POOLED BIAS

Throughout the appendices, the policy parameter is fixed and all population quantities and estimators retain their definitions in the main text. In particular, $\boldsymbol{G} = (\boldsymbol{F} + \lambda \boldsymbol{I})^{-1}$, $\boldsymbol{p} = \boldsymbol{G}\boldsymbol{g}$, $\boldsymbol{\Delta}_i = \widehat{\boldsymbol{F}}_i - \boldsymbol{F}$, and $\boldsymbol{\delta}_i = \widehat{\boldsymbol{g}}_i - \boldsymbol{g}$. We use Euclidean vector norms and operator matrix norms. Constants implicit in $\mathcal{O}(\cdot)$ may depend on the fixed population, d, and λ. Additional dependence on fixed N or fixed B is indicated by a subscript.

A.1 MOMENT IDENTITIES USED IN THE PROOF

Lemma A.1 (Centered Poisson moments)**.** *Under the sampling scheme of Section 4 and Equation (11),*

$$\mathbb{E}[\boldsymbol{\Delta}_i \boldsymbol{G} \boldsymbol{\Delta}_i] = \frac{1}{B} \int_{\mathcal{X}} \boldsymbol{h}(x)\boldsymbol{h}(x)^\top \boldsymbol{G}\boldsymbol{h}(x)\boldsymbol{h}(x)^\top \, \nu_{\boldsymbol{\theta}}(dx), \tag{19}$$

$$\mathbb{E}[\boldsymbol{\Delta}_i \boldsymbol{G} \boldsymbol{\delta}_i] = \frac{1}{B} \int_{\mathcal{X}} \boldsymbol{h}(x)\boldsymbol{h}(x)^\top \boldsymbol{G}\boldsymbol{h}(x) y(x) \, \nu_{\boldsymbol{\theta}}(dx), \tag{20}$$

$$\mathbb{E}[\boldsymbol{\Delta}_i \boldsymbol{G} \boldsymbol{\Delta}_i \boldsymbol{G} (\boldsymbol{\delta}_i - \boldsymbol{\Delta}_i \boldsymbol{p})] = \frac{1}{B^2} \int_{\mathcal{X}} \boldsymbol{h}(x)\boldsymbol{h}(x)^\top \boldsymbol{G}\boldsymbol{h}(x)\boldsymbol{h}(x)^\top \boldsymbol{G}\big[\boldsymbol{h}(x)y(x) - \boldsymbol{h}(x)\boldsymbol{h}(x)^\top \boldsymbol{p}\big] \, \nu_{\boldsymbol{\theta}}(dx). \tag{21}$$

Moreover, as $B \to \infty$,

$$\mathbb{E}\|\boldsymbol{\Delta}_i\|^4 + \mathbb{E}\|\boldsymbol{\delta}_i\|^4 = \mathcal{O}(B^{-2}). \tag{22}$$

The same second-moment rule applies to any bilinear expression in entries of $\boldsymbol{\Delta}_i$ and $\boldsymbol{\delta}_i$: its expectation is B^{-1} times the corresponding raw single-mark product. For three centered entries the factor is B^{-2}.

A.2 PROOF OF PROPOSITION 5.1

Proof. The local NPG error admits the following representation

$$\widehat{\boldsymbol{p}}_i - \boldsymbol{p} = (\widehat{\boldsymbol{F}}_i + \lambda \boldsymbol{I})^{-1} (\boldsymbol{\delta}_i - \boldsymbol{\Delta}_i \boldsymbol{p}) \,. \tag{23}$$

Iterating the resolvent identity three times gives the finite, exact identity

$$\begin{aligned}(\widehat{\boldsymbol{F}}_i + \lambda \boldsymbol{I})^{-1} = \boldsymbol{G} - \boldsymbol{G}\boldsymbol{\Delta}_i\boldsymbol{G} + \boldsymbol{G}\boldsymbol{\Delta}_i\boldsymbol{G}\boldsymbol{\Delta}_i\boldsymbol{G} \\ - \boldsymbol{G}\boldsymbol{\Delta}_i\boldsymbol{G}\boldsymbol{\Delta}_i\boldsymbol{G}\boldsymbol{\Delta}_i(\widehat{\boldsymbol{F}}_i + \lambda \boldsymbol{I})^{-1}.\end{aligned} \tag{24}$$

No small-norm or convergent Neumann-series assumption is needed. Multiply by $\boldsymbol{\delta}_i - \boldsymbol{\Delta}_i\boldsymbol{p}$ and use Equation (23). The first term has zero expectation. By Lemma A.1, the next term has expectation

$$\begin{aligned}-\boldsymbol{G}\mathbb{E}[\boldsymbol{\Delta}_i \boldsymbol{G}(\boldsymbol{\delta}_i - \boldsymbol{\Delta}_i \boldsymbol{p})] &= \frac{1}{B}\boldsymbol{G} \int_{\mathcal{X}} \boldsymbol{h}(x)\boldsymbol{h}(x)^\top \boldsymbol{G}\big[\boldsymbol{h}(x)\boldsymbol{h}(x)^\top \boldsymbol{p} - \boldsymbol{h}(x)y(x)\big] \, \nu_{\boldsymbol{\theta}}(dx) \\ &= \frac{\boldsymbol{c}}{B}.\end{aligned}$$

The cubic term has expectation $\mathcal{O}(B^{-2})$ by Equation (21). Both inverse norms are at most λ^{-1}, so the expected norm of the remaining term is at most

$$\lambda^{-4}\left(\mathbb{E}[\|\boldsymbol{\Delta}_i\|^3\|\boldsymbol{\delta}_i\|] + \|p\|\mathbb{E}\|\boldsymbol{\Delta}_i\|^4\right) = \mathcal{O}(B^{-2}).$$

Here Hölder's inequality bounds the mixed expectation by $(\mathbb{E}\|\boldsymbol{\Delta}_i\|^4)^{3/4}(\mathbb{E}\|\boldsymbol{\delta}_i\|^4)^{1/4}$. This proves Equation (12), including its stated remainder. □

A.3 PROOFS OF COROLLARY 5.3 AND COROLLARY 5.2

Proof. Linearity of expectation and identical worker distributions give

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} \widehat{\boldsymbol{p}}_i\right] - \boldsymbol{p} = \mathbb{E}[\widehat{\boldsymbol{p}}_1] - \boldsymbol{p}.$$

Proposition 5.1 proves the uniform-average claim. For pooling, the estimator referred to in Corollary 5.2 is explicitly

$$\widehat{\boldsymbol{p}}_{\text{pool}} = \left(\lambda \boldsymbol{I} + \frac{1}{N}\sum_{i=1}^{N} \widehat{\boldsymbol{F}}_i\right)^{-1} \left(\frac{1}{N}\sum_{i=1}^{N} \widehat{\boldsymbol{g}}_i\right).$$

The total count $\sum_i \mathrm{K}_i$ is Poisson with mean NB; conditional on this count, the pooled marks are i.i.d. with law $\nu_{\boldsymbol{\theta}}$. Both pooled sums are normalized by NB. Thus the pooled pair has exactly the one-worker distribution with B replaced by NB. Applying the proposition with this intensity proves Equation (15). □

A.4 Proof of Lemma A.1

Proof. Conditioning on the count gives $\mathbb{E}\mathrm{K}_i = B$ and $\mathbb{E}[\mathrm{K}_i(\mathrm{K}_i - 1)] = B^2$. In the product of two empirical sums, equal sample indices contribute the raw single-mark product with coefficient B^{-1}; unequal indices contribute the product of the population means. Centering removes the latter, proving Equation (19) and Equation (20). For three centered sums, all terms with an index appearing only once cancel. Only the common-index contribution remains, with coefficient $B/B^3 = B^{-2}$, proving Equation (21).

For completeness, for any coordinate function f of $\boldsymbol{h}\boldsymbol{h}^\top$ or $\boldsymbol{h}y$, direct expansion of the fourth centered moment gives

$$\mathbb{E}\left[\frac{1}{B}\sum_{r=1}^{\mathrm{K}_i} f(\mathrm{X}_{ir}) - \int_{\mathcal{X}} f \, d\nu_{\boldsymbol{\theta}}\right]^4 = \frac{1}{B^3}\int_{\mathcal{X}} f^4 \, d\nu_{\boldsymbol{\theta}} + \frac{3}{B^2}\left(\int_{\mathcal{X}} f^2 \, d\nu_{\boldsymbol{\theta}}\right)^2.$$

The two terms correspond respectively to one block of four indices and two blocks of two. The assumed fourth moments of $\boldsymbol{h}\boldsymbol{h}^\top$ and $\boldsymbol{h}y$ are precisely controlled by Equation (11). Finite-dimensional norm equivalence now proves Equation (22). □

B Exact determinant, adjugate, and shared-sample identities

B.1 A determinant identity with a deterministic base

The following identities are a specialization of the determinant-preserving Poisson framework of Derezinski et al. (2020b, Lemma 5). We give a direct proof in our setting to express the result in Fisher notation and make the conditions explicit.

Lemma B.1 (Deterministic positive-definite base)**.** *Suppose that*

$$\int_{\mathcal{X}} \|\boldsymbol{h}(x)\|^2 \, \nu_{\boldsymbol{\theta}}(dx) < \infty. \tag{25}$$

Under the Poisson sampling scheme, for every deterministic $\boldsymbol{R} \succ 0$,

$$\mathbb{E}[\det(\boldsymbol{R} + \widehat{\boldsymbol{F}}_i)] = \det(\boldsymbol{R} + \boldsymbol{F}). \tag{26}$$

Here $\boldsymbol{R}$ is the arbitrary base already used in the main-text proof sketch; it does not change the population Fisher or damping.

Proof of Lemma B.1. Conditional on $\mathrm{K}_i = m$, Gram–Cauchy–Binet gives

$$\det(\boldsymbol{R} + \widehat{\boldsymbol{F}}_i) = \det(\boldsymbol{R}) \sum_{k=0}^{d} B^{-k} \sum_{1 \le r_1 < \cdots < r_k \le m} \det\left[\boldsymbol{h}(\mathrm{X}_{ir_a})^\top \boldsymbol{R}^{-1} \boldsymbol{h}(\mathrm{X}_{ir_b})\right]_{a,b=1}^{k}. \tag{27}$$

The empty determinant is one and a sum over unavailable indices is zero. For a symmetric nonnegative kernel f of k marks, conditioning on K_i and using its factorial moment yields

$$\mathbb{E} \sum_{1\le r_1<\cdots<r_k\le \mathrm{K}_i} f(\mathrm{X}_{ir_1},\dots,\mathrm{X}_{ir_k}) = \frac{\mathbb{E}[(\mathrm{K}_i)_k]}{k!}\int_{\mathcal{X}^k} f(x_1,\dots,x_k)\prod_{a=1}^{k}\nu_{\boldsymbol{\theta}}(dx_a) = \frac{B^k}{k!}\int_{\mathcal{X}^k} f\, d\nu_{\boldsymbol{\theta}}^{\otimes k}. \tag{28}$$

Consequently,

$$\mathbb{E}\det(\boldsymbol{R}+\widehat{\boldsymbol{F}}_i) = \det(\boldsymbol{R})\sum_{k=0}^{d}\frac{1}{k!}\int_{\mathcal{X}^k}\det\big[\boldsymbol{h}(x_a)^\top\boldsymbol{R}^{-1}\boldsymbol{h}(x_b)\big]_{a,b=1}^{k}\prod_{a=1}^{k}\nu_{\boldsymbol{\theta}}(dx_a). \tag{29}$$

To identify this last expression, expand each Gram determinant by Cauchy–Binet into squared coordinate minors of $\boldsymbol{R}^{-1/2}\boldsymbol{h}(x_a)$. Expanding the two determinants in each square into permutations and integrating each independent column gives $k!$ times the corresponding principal minor of $\boldsymbol{R}^{-1/2}\boldsymbol{F}\boldsymbol{R}^{-1/2}$. Summing the principal minors over k is exactly $\det(\boldsymbol{I}+\boldsymbol{R}^{-1/2}\boldsymbol{F}\boldsymbol{R}^{-1/2})$. Multiplication by $\det(\boldsymbol{R})$ proves Equation (26). All integrals are finite because Hadamard's inequality gives

$$0 \le \det\big[\boldsymbol{h}(x_a)^\top\boldsymbol{R}^{-1}\boldsymbol{h}(x_b)\big]_{a,b=1}^{k} \le \|\boldsymbol{R}^{-1}\|^k\prod_{a=1}^{k}\|\boldsymbol{h}(x_a)\|^2.$$

Thus only the second score moment is needed, irrespective of d. □

Lemma B.2 (Poisson determinant identity)**.** *Under Equation (25) and the Poisson construction of Section 4,*

$$\mathbb{E}[D_i] = \mathbb{E}\Big[\det(\widehat{\boldsymbol{F}}_i+\lambda\boldsymbol{I})\Big] = \det(\boldsymbol{F}+\lambda\boldsymbol{I}) = D. \tag{30}$$

Proof of Lemma B.2. Apply Lemma B.1 with $\boldsymbol{R}=\lambda\boldsymbol{I}$. □

B.2 Poisson adjugate identity

Corollary B.3 (Poisson adjugate identity)**.** *Under the assumptions of Lemma B.2,*

$$\mathbb{E}\Big[\operatorname{adj}(\widehat{\boldsymbol{F}}_i+\lambda\boldsymbol{I})\Big] = \operatorname{adj}(\boldsymbol{F}+\lambda\boldsymbol{I}). \tag{31}$$

Equivalently,

$$\mathbb{E}\Big[D_i(\widehat{\boldsymbol{F}}_i+\lambda\boldsymbol{I})^{-1}\Big] = D(\boldsymbol{F}+\lambda\boldsymbol{I})^{-1}. \tag{32}$$

Proof. For any deterministic vector $\boldsymbol{u}$ and $t\ge 0$, apply Lemma B.1 to $\boldsymbol{R}=\lambda\boldsymbol{I}+t\boldsymbol{u}\boldsymbol{u}^\top$. The determinant lemma on both sides and subtraction of the $t=0$ identity give

$$\mathbb{E}\Big[\boldsymbol{u}^\top\operatorname{adj}(\widehat{\boldsymbol{F}}_i+\lambda\boldsymbol{I})\boldsymbol{u}\Big] = \boldsymbol{u}^\top\operatorname{adj}(F+\lambda\boldsymbol{I})\boldsymbol{u}.$$

The adjugates are positive semidefinite. Their diagonal entries have finite expectation by taking $\boldsymbol{u}=\boldsymbol{e}_j$; absolute off-diagonal entries are bounded by half the sum of the corresponding diagonal entries. Thus all entries are integrable, and polarization proves Equation (31). The equality $\operatorname{adj}(\widehat{\boldsymbol{F}}_i+\lambda\boldsymbol{I}) = D_i(\widehat{\boldsymbol{F}}_i+\lambda\boldsymbol{I})^{-1}$ proves Equation (32). The argument also proves the adjugate identity for every deterministic positive-definite base $\boldsymbol{R}$ by starting with $\boldsymbol{R}+t\boldsymbol{u}\boldsymbol{u}^\top$. □

B.3 Shared-sample determinant-weighted identity

Proposition B.4 (Shared-sample determinant-weighted identity)**.** *Suppose that $\widehat{\boldsymbol{F}}_i$ and $\widehat{\boldsymbol{g}}_i$ are constructed from the same Poisson samples as in Section 4, and assume*

$$\int_{\mathcal{X}}\big(1+\|\boldsymbol{h}(x)\|^2\big)\|\boldsymbol{h}(x)y(x)\|\,\nu_{\boldsymbol{\theta}}(dx) < \infty. \tag{33}$$

Then

$$\mathbb{E}[D_i\widehat{\boldsymbol{p}}_i] = D\boldsymbol{p}. \tag{34}$$

Proof. The Poisson–Mecke equation (Last & Penrose, 2017, Chapter 4) applied to the sample sum in $\widehat{\boldsymbol{g}}_i$ gives

$$\mathbb{E}[D_i\widehat{\boldsymbol{p}}_i] = \int_{\mathcal{X}} \mathbb{E}\left[\operatorname{adj}\left(\widehat{\boldsymbol{F}}_i + \lambda\boldsymbol{I} + \frac{1}{B}\boldsymbol{h}(x)\boldsymbol{h}(x)^\top\right)\boldsymbol{h}(x)y(x)\right]\nu_{\boldsymbol{\theta}}(dx). \tag{35}$$

In this finite-intensity setting, the formula follows directly by conditioning on $\mathrm{K}_i = m$: symmetry replaces the sum by m times its first term, and $m\mathbb{P}(\mathrm{K}_i = m)/B = \mathbb{P}(\mathrm{K}_i = m-1)$ leaves an independent Poisson batch after the selected mark is removed.

By the adjugate identity with base $\lambda\boldsymbol{I} + B^{-1}\boldsymbol{h}(x)\boldsymbol{h}(x)^\top$, the expectation inside Equation (35) is $\operatorname{adj}(\boldsymbol{F} + \lambda\boldsymbol{I} + B^{-1}\boldsymbol{h}(x)\boldsymbol{h}(x)^\top)\boldsymbol{h}(x)y(x)$. Sherman–Morrison and the determinant lemma give

$$\begin{aligned}&\operatorname{adj}\left(\boldsymbol{F} + \lambda\boldsymbol{I} + \frac{1}{B}\boldsymbol{h}(x)\boldsymbol{h}(x)^\top\right)\\ &\qquad = D\boldsymbol{G} + \frac{D}{B}\left[\big(\boldsymbol{h}(x)^\top\boldsymbol{G}\boldsymbol{h}(x)\big)\boldsymbol{G} - \boldsymbol{G}\boldsymbol{h}(x)\boldsymbol{h}(x)^\top\boldsymbol{G}\right].\end{aligned} \tag{36}$$

The term in brackets annihilates $\boldsymbol{h}(x)y(x)$, since

$$\big(\boldsymbol{h}(x)^\top\boldsymbol{G}\boldsymbol{h}(x)\big)\boldsymbol{G}\boldsymbol{h}(x)y(x) - \boldsymbol{G}\boldsymbol{h}(x)\big(\boldsymbol{h}(x)^\top\boldsymbol{G}\boldsymbol{h}(x)\big)y(x) = 0.$$

Integration therefore yields $\mathbb{E}[D_i\widehat{\boldsymbol{p}}_i] = D\boldsymbol{G}\boldsymbol{g} = D\boldsymbol{p}$, as required.

Here is an explicit integrability justification. The added-point adjugate is positive semidefinite, so its operator norm is bounded by its trace. Its expected trace equals the trace of the right-hand side of Equation (36), which is bounded by

$$D\operatorname{tr}(\boldsymbol{G}) + \frac{D}{B}\|\boldsymbol{G}\|\operatorname{tr}(\boldsymbol{G})\|\boldsymbol{h}(x)\|^2.$$

Multiplication by $\|\boldsymbol{h}(x)y(x)\|$ is integrable under Equation (33). Mecke applied first to the nonnegative norm bound justifies all signed/vector expectation exchanges above. □

The cancellation uses the scalar-multiple structure of each gradient atom $\boldsymbol{h}(x)y(x)$. It does not follow merely from Fisher–gradient dependence or unbiasedness of the two empirical sums separately.

C Finite-worker ratio bias and many-worker consistency

Lemma B.2 gives

$$\mathbb{E}[D_i] = D,$$

while Proposition B.4 gives

$$\mathbb{E}[D_i\widehat{\boldsymbol{p}}_i] = D\boldsymbol{p}.$$

Together they imply the exact centering property

$$\mathbb{E}[D_i(\widehat{\boldsymbol{p}}_i - \boldsymbol{p})] = 0. \tag{37}$$

Define

$$\mathbf{z}_i := D_i(\widehat{\boldsymbol{p}}_i - \boldsymbol{p}), \qquad \mathrm{u}_i := D_i - D.$$

Then PDA can be written exactly as the centered random ratio

$$\widehat{\boldsymbol{p}}_{\mathrm{PDA},N} - \boldsymbol{p} = \frac{N^{-1}\sum_i \mathbf{z}_i}{D + N^{-1}\sum_i \mathrm{u}_i}. \tag{38}$$

The exact centering Equation (37) and the representation Equation (38) are used throughout this section. In particular,

$$\frac{1}{N}\sum_{i=1}^{N} D_i \geq \lambda^d > 0.$$

C.1 Proof of Corollary 5.5

Proof. The strong law applied separately to the integrable i.i.d. numerator and denominator gives

$$\frac{1}{N}\sum_{i=1}^{N} D_i\widehat{\boldsymbol{p}}_i \longrightarrow D\boldsymbol{p}, \qquad \frac{1}{N}\sum_{i=1}^{N} D_i \longrightarrow D \quad \text{almost surely.}$$

Their ratio converges to $\boldsymbol{p}$. To pass to expectations under $\mathbb{E}\big[D_i\|\widehat{g}_i\|^{1+\epsilon}\big] < \infty$, convexity and $\|\widehat{\boldsymbol{p}}_i\| \le \lambda^{-1}\|\widehat{\boldsymbol{g}}_i\|$ imply

$$\begin{aligned}\|\widehat{\boldsymbol{p}}_{\mathrm{PDA},N}\|^{1+\epsilon} &\le \lambda^{-(1+\epsilon)}\sum_{i=1}^{N} w_i\|\widehat{\boldsymbol{g}}_i\|^{1+\epsilon} \\ &\le \frac{1}{N\lambda^{d+1+\epsilon}}\sum_{i=1}^{N} D_i\|\widehat{\boldsymbol{g}}_i\|^{1+\epsilon}.\end{aligned}$$

The expectation is bounded uniformly in N, proving uniform integrability and hence convergence in expectation. □

C.2 An explicit expansion when the Poisson intensity is fixed

Proposition C.1 (Fixed-B ratio coefficient)**.** *Under the exact identities, let $B > 0$ be fixed, and suppose*

$$\mathbb{E}|D_i - D|^4 + \mathbb{E}\|D_i(\widehat{\boldsymbol{p}}_i - \boldsymbol{p})\|^4 < \infty. \tag{39}$$

Then, as $N \to \infty$,

$$\begin{aligned}\mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{PDA},N}] - \boldsymbol{p} &= -\frac{\mathbb{E}[D_1(\widehat{\boldsymbol{p}}_1 - \boldsymbol{p})(D_1 - D)]}{ND^2} + \mathcal{O}_B(N^{-2}) \\ &= \frac{\boldsymbol{p}\,\mathrm{Var}(D_1) - \mathrm{Cov}(D_1\widehat{\boldsymbol{p}}_1, D_1)}{ND^2} + \mathcal{O}_B(N^{-2}).\end{aligned} \tag{40}$$

The covariance is taken componentwise in its vector argument.

Proof. Use the following exact reciprocal identity, with all averages left in their original notation:

$$\begin{aligned}\frac{1}{N^{-1}\sum_i D_i} = \frac{1}{D} - \frac{N^{-1}\sum_i(D_i - D)}{D^2} + \frac{(N^{-1}\sum_i(D_i - D))^2}{D^3} \\ - \frac{(N^{-1}\sum_i(D_i - D))^3}{D^3(N^{-1}\sum_i D_i)}.\end{aligned} \tag{41}$$

Multiplying by $N^{-1}\sum_i D_i(\widehat{\boldsymbol{p}}_i - \boldsymbol{p})$, the constant term has zero expectation. Independence and centering give

$$\mathbb{E}\left[\left(\frac{1}{N}\sum_i D_i(\widehat{\boldsymbol{p}}_i - \boldsymbol{p})\right)\left(\frac{1}{N}\sum_j (D_j - D)\right)\right] = \frac{1}{N}\mathbb{E}[D_1(\widehat{\boldsymbol{p}}_1 - \boldsymbol{p})(D_1 - D)], \tag{42}$$

$$\mathbb{E}\left[\left(\frac{1}{N}\sum_i D_i(\widehat{\boldsymbol{p}}_i - \boldsymbol{p})\right)\left(\frac{1}{N}\sum_j (D_j - D)\right)^2\right] = \frac{1}{N^2}\mathbb{E}[D_1(\widehat{\boldsymbol{p}}_1 - \boldsymbol{p})(D_1 - D)^2]. \tag{43}$$

In each expansion, an index that occurs only once has zero contribution. Each centered worker average has fourth moment $\mathcal{O}_B(N^{-2})$: the nonzero fourth-order index patterns are one repeated index or two paired indices. Therefore the expected norm of the last reciprocal term, after multiplication by the numerator, is at most

$$\begin{aligned}\frac{1}{D^3\lambda^d}\left(\mathbb{E}\left\|\frac{1}{N}\sum_i D_i(\widehat{\boldsymbol{p}}_i - \boldsymbol{p})\right\|^4\right)^{1/4} \\ \times\left(\mathbb{E}\left|\frac{1}{N}\sum_i (D_i - D)\right|^4\right)^{3/4} = \mathcal{O}_B(N^{-2}).\end{aligned}$$

This proves the first equality. Expanding its covariance and using $\mathbb{E}[D_1\widehat{\boldsymbol{p}}_1] = D\boldsymbol{p}$ proves the second. □

The coefficient in Equation (40) need not equal $\boldsymbol{c}/B$ for a fixed finite B. Matching the pooled coefficient requires the additional large-B expansion proved below. Holding the Poisson intensity fixed still leaves the realized count K_i random; it is distinct from using a deterministic batch size.

D Proof of Theorem 5.4

Proof. Fix any $N \geq 1$ and assume the moment condition in Equation (11). The first-order determinant expansion and normalization give

$$D_i = D\,[1 + \operatorname{tr}(\boldsymbol{G}\boldsymbol{\Delta}_i)] + \text{terms of degree at least two in } \boldsymbol{\Delta}_i,$$
$$w_i = \frac{1}{N}\left[1 + \operatorname{tr}(\boldsymbol{G}\boldsymbol{\Delta}_i) - \frac{1}{N}\sum_{j=1}^{N}\operatorname{tr}(\boldsymbol{G}\boldsymbol{\Delta}_j)\right] + \text{terms of degree at least two.} \tag{44}$$

The first two degrees of the local direction error are

$$\boldsymbol{G}(\boldsymbol{\delta}_i - \boldsymbol{\Delta}_i\boldsymbol{p}) - \boldsymbol{G}\boldsymbol{\Delta}_i\boldsymbol{G}(\boldsymbol{\delta}_i - \boldsymbol{\Delta}_i\boldsymbol{p}).$$

Since $\sum_i w_i = 1$, it suffices to expand $\sum_i w_i(\widehat{\boldsymbol{p}}_i - \boldsymbol{p})$; no expansion of a weight multiplying a nonzero constant error is needed. The expectation of the local quadratic term, with constant weight $1/N$, sums to $\boldsymbol{c}/B$. For the remaining quadratic contribution, independence and centering give

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\left(\operatorname{tr}(\boldsymbol{G}\boldsymbol{\Delta}_i) - \frac{1}{N}\sum_{j}\operatorname{tr}(\boldsymbol{G}\boldsymbol{\Delta}_j)\right)\boldsymbol{G}(\boldsymbol{\delta}_i - \boldsymbol{\Delta}_i\boldsymbol{p})\right]$$
$$= \left(1 - \frac{1}{N}\right)\mathbb{E}[\operatorname{tr}(\boldsymbol{G}\boldsymbol{\Delta}_1)\boldsymbol{G}(\boldsymbol{\delta}_1 - \boldsymbol{\Delta}_1\boldsymbol{p})]. \tag{45}$$

Lemma A.1 evaluates the last expectation as

$$\frac{1}{B}G\int_{\mathcal{X}}\operatorname{tr}(\boldsymbol{G}\boldsymbol{h}(x)\boldsymbol{h}(x)^\top)\big[\boldsymbol{h}(x)y(x) - \boldsymbol{h}(x)\boldsymbol{h}(x)^\top p\big]\,\nu_{\boldsymbol{\theta}}(dx)$$
$$= \frac{1}{B}G\int_{\mathcal{X}}\boldsymbol{h}(x)\boldsymbol{h}(x)^\top G\big[\boldsymbol{h}(x)y(x) - \boldsymbol{h}(x)\boldsymbol{h}(x)^\top p\big]\,\nu_{\boldsymbol{\theta}}(dx) = -\frac{\boldsymbol{c}}{B}. \tag{46}$$

Indeed, the vector in brackets is a scalar multiple of $\boldsymbol{h}(x)$, and $\operatorname{tr}(\boldsymbol{G}\boldsymbol{h}\boldsymbol{h}^\top) = \boldsymbol{h}^\top\boldsymbol{G}\boldsymbol{h}$. Combining the two quadratic contributions yields

$$\frac{\boldsymbol{c}}{B} - \left(1 - \frac{1}{N}\right)\frac{\boldsymbol{c}}{B} = \frac{\boldsymbol{c}}{NB}.$$

The preceding degree calculation alone is not a remainder estimate. To justify it, regard the entire estimator as a function of all Fisher matrices and gradients and use the segment

$$\lambda\boldsymbol{I} + \boldsymbol{F} + t\boldsymbol{\Delta}_i = \lambda\boldsymbol{I} + (1-t)\boldsymbol{F} + t\widehat{\boldsymbol{F}}_i \succeq \lambda\boldsymbol{I}, \qquad 0 \leq t \leq 1.$$

Every inverse along this segment has norm at most λ^{-1}. Derivatives of a log determinant are traces of products of these inverses and the differentiation directions. Derivatives of normalized weights are finite sums of products of these traces and weights between zero and one. Consequently, for fixed N, d, λ, the Fisher derivatives through order four of a normalized weight times an inverse are uniformly bounded.

First split the estimator into the contribution with every gradient equal to g and the contribution linear in $\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_N$. Expand the former through Fisher degree three and the latter through Fisher degree two. This is the expansion through total degree three. The remainder is bounded by a constant, independent of the realizations and of B, times

$$\|\boldsymbol{g}\|\left(\sum_i\|\boldsymbol{\Delta}_i\|\right)^4 + \left(\sum_i\|\boldsymbol{\Delta}_i\|\right)^3\sum_i\|\boldsymbol{\delta}_i\|. \tag{47}$$

This splitting is needed because derivatives of a direction can depend linearly on its gradient argument; no uniform bound on sampled gradients is being assumed.

Every degree-one expectation vanishes. Every degree-three expectation is $\mathcal{O}_N(B^{-2})$: terms involving independent workers factor, any singleton factor has zero mean, and three entries on one worker are third centered Poisson moments. Finally, Equation (22), Hölder's inequality, and fixed N give expectation $\mathcal{O}_N(B^{-2})$ for Equation (47). Thus

$$\mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{PDA},N}] - \boldsymbol{p} = \frac{\boldsymbol{c}}{NB} + \mathcal{O}_N(B^{-2}).$$

By Corollary 5.2,

$$\mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{pool}}] - \boldsymbol{p} = \frac{\boldsymbol{c}}{NB} + \mathcal{O}((NB)^{-2}).$$

Since N is fixed, $\mathcal{O}((NB)^{-2}) = \mathcal{O}_N(B^{-2})$. Subtracting this expansion from the one above therefore gives

$$\mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{PDA},N}] - \mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{pool}}] = \mathcal{O}_N(B^{-2}),$$

which proves the claimed expectation gap.

□

E Uniform joint worker–batch bounds and leading coefficients

E.1 Nonasymptotic joint worker–batch bias bound

Proposition E.1 (Nonasymptotic joint worker–batch bias bound)**.** *Suppose that, for all $B \geq B_0 > 0$, there exist constants C_1, C_2 independent of N and B such that*

$$\mathbb{E}\Big[\|D_i(\widehat{\boldsymbol{p}}_i - \boldsymbol{p})\|^2\Big] \leq \frac{C_1}{B}, \quad \mathrm{Var}(D_i) \leq \frac{C_2}{B}. \tag{48}$$

Then, for every $N \geq 1$ and $B \geq B_0$,

$$\|\mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{PDA},N}] - \boldsymbol{p}\| \leq \frac{\sqrt{C_1 C_2}}{NB\,\det(\boldsymbol{F} + \lambda \boldsymbol{I})\lambda^d}. \tag{49}$$

Consequently, the finite-worker bias is uniformly $\mathcal{O}((NB)^{-1})$.

Proof. Exact centering gives

$$\begin{aligned}\mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{PDA},N}] - \boldsymbol{p} &= \mathbb{E}\left[\left(\frac{1}{N}\sum_i D_i(\widehat{\boldsymbol{p}}_i - \boldsymbol{p})\right)\left(\frac{1}{N^{-1}\sum_i D_i} - \frac{1}{D}\right)\right] \\ &= -\frac{1}{D}\mathbb{E}\left[\frac{(N^{-1}\sum_i D_i(\widehat{\boldsymbol{p}}_i - \boldsymbol{p}))(N^{-1}\sum_i (D_i - D))}{N^{-1}\sum_i D_i}\right].\end{aligned}$$

Since the denominator is at least λ^d, Cauchy–Schwarz and independence yield

$$\begin{aligned}\|\mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{PDA},N}] - \boldsymbol{p}\| &\leq \frac{1}{D\lambda^d}\left(\frac{\mathbb{E}\|D_1(\widehat{\boldsymbol{p}}_1 - \boldsymbol{p})\|^2}{N}\right)^{1/2}\left(\frac{\mathrm{Var}(D_1)}{N}\right)^{1/2} \\ &\leq \frac{\sqrt{C_1 C_2}}{NBD\lambda^d}.\end{aligned}$$

This proves Equation (49) under the stated second-moment bounds alone. □

E.2 A concrete sufficient condition for the uniform expansion

For the sharper statement, a sufficient condition, expressed entirely in the existing marks, is

$$\int_{\mathcal{X}} \left(\|\boldsymbol{h}(x)\|^2 + \|\boldsymbol{h}(x)y(x)\|\right)^{4d} \nu_{\boldsymbol{\theta}}(dx) < \infty. \tag{50}$$

Bounded $\boldsymbol{h}$ and $\boldsymbol{h}y$ are a special case. This is a sufficient, not a minimal, condition. The dimension-dependent requirement is used only for the uniform expansion and its determinant moments; it does not replace the weaker assumptions of the exact identities or the fixed-worker expansion.

Lemma E.2 (Uniform moments and the one-worker covariance)**.** *Under Equation (50), for every fixed $B_0 > 0$, uniformly over $B \geq B_0$,*

$$\mathbb{E}\|D_i(\widehat{\boldsymbol{p}}_i - \boldsymbol{p})\|^2 + \mathbb{E}|D_i - D|^2 = \mathcal{O}(B^{-1}), \tag{51}$$

$$\mathbb{E}\|D_i(\widehat{\boldsymbol{p}}_i - \boldsymbol{p})\|^4 + \mathbb{E}|D_i - D|^4 = \mathcal{O}(B^{-2}), \tag{52}$$

$$\mathbb{E}[D_i(\widehat{\boldsymbol{p}}_i - \boldsymbol{p})(D_i - D)] = -\frac{D^2\boldsymbol{c}}{B} + \mathcal{O}(B^{-2}), \tag{53}$$

$$\mathbb{E}[D_i(\widehat{\boldsymbol{p}}_i - \boldsymbol{p})(D_i - D)^2] = \mathcal{O}(B^{-2}). \tag{54}$$

E.3 Uniform-in-N refinement

Proposition E.3 (Uniform-in-N PDA bias expansion)**.** *Under the same-sample Poisson construction of Section 4 and the higher moment condition in Equation (50), as $B \to \infty$,*

$$\mathrm{Bias}(\widehat{\boldsymbol{p}}_{\mathrm{PDA},N}) = \mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{PDA},N}] - \boldsymbol{p} = \frac{\boldsymbol{c}}{NB} + \mathcal{O}\left(\frac{1}{NB^2}\right), \tag{55}$$

uniformly over $N \geq 1$. Combining this expansion with Corollary 5.2 further gives

$$\mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{PDA},N}] - \mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{pool}}] = \mathcal{O}\left(\frac{1}{NB^2}\right), \tag{56}$$

uniformly over $N \geq 1$.

Proof. Use the exact reciprocal expansion Equation (41) in Equation (38). The constant term has zero expectation by exact centering. For the first covariance,

$$\mathbb{E}\left[\left(\frac{1}{N}\sum_i D_i(\widehat{\boldsymbol{p}}_i - \boldsymbol{p})\right)\left(\frac{1}{N}\sum_j (D_j - D)\right)\right] = \frac{1}{N^2}\sum_i\sum_j \mathbb{E}[D_i(\widehat{\boldsymbol{p}}_i - \boldsymbol{p})(D_j - D)].$$

For $i \neq j$, independence across workers and centering give

$$\mathbb{E}[D_i(\widehat{\boldsymbol{p}}_i - \boldsymbol{p})(D_j - D)] = \mathbb{E}[D_i(\widehat{\boldsymbol{p}}_i - \boldsymbol{p})]\mathbb{E}[D_j - D] = 0.$$

Hence only the N diagonal terms remain, and since the workers are identically distributed,

$$\mathbb{E}\left[\left(\frac{1}{N}\sum_i D_i(\widehat{\boldsymbol{p}}_i - \boldsymbol{p})\right)\left(\frac{1}{N}\sum_j (D_j - D)\right)\right] = \frac{1}{N}\mathbb{E}[D_1(\widehat{\boldsymbol{p}}_1 - \boldsymbol{p})(D_1 - D)].$$

Thus its contribution is

$$-\frac{1}{ND^2}\mathbb{E}[D_1(\widehat{\boldsymbol{p}}_1 - \boldsymbol{p})(D_1 - D)] = \frac{\boldsymbol{c}}{NB} + \mathcal{O}\left(\frac{1}{NB^2}\right).$$

By Equation (43) and Equation (54), the next term is $\mathcal{O}(N^{-2}B^{-2})$. Independence, centering, and the second and fourth moments in Lemma E.2 give

$$\mathbb{E}\left\|\frac{1}{N}\sum_i D_i(\widehat{\boldsymbol{p}}_i - \boldsymbol{p})\right\|^4 = \mathcal{O}(N^{-2}B^{-2}),$$

$$\mathbb{E}\left|\frac{1}{N}\sum_i (D_i - D)\right|^4 = \mathcal{O}(N^{-2}B^{-2}). \tag{57}$$

Indeed, the same-worker terms are $N^{-3}\mathcal{O}(B^{-2})$ and the paired terms are $N^{-2}\mathcal{O}(B^{-2})$; this also covers $N = 1$. The final reciprocal term is bounded by Hölder and the denominator lower bound

exactly as in Appendix C, now giving $\mathcal{O}(N^{-2}B^{-2})$ uniformly. Combining the three contributions proves

$$\mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{PDA},N}] - \boldsymbol{p} = \frac{\boldsymbol{c}}{NB} + \mathcal{O}\left(\frac{1}{NB^2}\right) \quad (N \geq 1,\ B \geq B_0).$$

All constants are independent of N and B in this range.

By Corollary 5.2,

$$\mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{pool}}] - \boldsymbol{p} = \frac{\boldsymbol{c}}{NB} + \mathcal{O}\left(\frac{1}{N^2B^2}\right).$$

Since $N \geq 1$,

$$\frac{1}{N^2B^2} \leq \frac{1}{NB^2}.$$

Subtracting the two expansions therefore gives

$$\mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{PDA},N}] - \mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{pool}}] = \mathcal{O}\left(\frac{1}{NB^2}\right),$$

uniformly over $N \geq 1$. □

E.4 Proof of Lemma E.2

Proof. The exact polynomial identities

$$D_i - D = \det(F + \lambda \boldsymbol{I} + \boldsymbol{\Delta}_i) - D,$$
$$D_i(\widehat{\boldsymbol{p}}_i - \boldsymbol{p}) = \mathrm{adj}(F + \lambda \boldsymbol{I} + \boldsymbol{\Delta}_i)(\boldsymbol{\delta}_i - \boldsymbol{\Delta}_i \boldsymbol{p})$$

show that both quantities are polynomials of total degree at most d in the centered entries of $\boldsymbol{\Delta}_i, \boldsymbol{\delta}_i$, with no constant term. This removes any inverse remainder from the present moment calculation.

We spell out the moment accounting. For coordinate functions $f_1, \ldots, f_k$ selected from the entries of $\boldsymbol{h}\boldsymbol{h}^\top$ and $\boldsymbol{h}y$, the joint centered Poisson moment is

$$\mathbb{E} \prod_{a=1}^{k} \left(\frac{1}{B} \sum_{r=1}^{\mathrm{K}_i} f_a(\mathrm{X}_{ir}) - \int_{\mathcal{X}} f_a \, d\nu_{\boldsymbol{\theta}} \right)$$
$$= \sum_{\substack{\pi \text{ a partition of } \{1,\ldots,k\} \\ \text{every block has size at least 2}}} B^{|\pi|-k} \prod_{A \in \pi} \int_{\mathcal{X}} \prod_{a \in A} f_a(x) \, \nu_{\boldsymbol{\theta}}(dx). \tag{58}$$

Here π and A are only summation indices for partitions and their blocks. To verify the formula, expand the empirical sums by equality patterns among sample indices and use $\mathbb{E}[(\mathrm{K}_i)_j] = B^j$; centering cancels exactly the singleton blocks. Equivalently each block of size $j \geq 2$ is a joint cumulant with factor B^{1-j}.

For $k \geq 2$, a partition without singletons has at most $\lfloor k/2 \rfloor$ blocks. A degree-k centered product therefore has expectation $\mathcal{O}(B^{-\lceil k/2 \rceil})$, uniformly for $B \geq B_0$. Condition Equation (50) supplies every mixed raw moment through order $4d$. Expanding the squared and fourth powers of the polynomials above uses degrees from 2 to $2d$ and from 4 to $4d$, respectively. Applying Equation (58) entrywise and finite-dimensional norm inequalities proves Equation (51) and Equation (52).

The product in Equation (54) has minimum degree three and maximum degree $3d$, so the same rule proves its $\mathcal{O}(B^{-2})$ bound. Finally, the linear terms of the two polynomials are respectively $D\,\mathrm{tr}(\boldsymbol{G}\boldsymbol{\Delta}_i)$ and $D\boldsymbol{G}(\boldsymbol{\delta}_i - \boldsymbol{\Delta}_i\boldsymbol{p})$. Their product has expectation

$$D^2\mathbb{E}[\mathrm{tr}(\boldsymbol{G}\boldsymbol{\Delta}_i)G(\boldsymbol{\delta}_i - \boldsymbol{\Delta}_i\boldsymbol{p})] = -\frac{D^2\boldsymbol{c}}{B}$$

by Equation (46). All remaining terms have total degree at least three and at most $2d$, hence expectation $\mathcal{O}(B^{-2})$. This proves Equation (53) and completes the lemma. □

The pooled estimator has remainder $\mathcal{O}((NB)^{-2})$, whereas the uniform PDA remainder proved here is $\mathcal{O}((NB^2)^{-1})$. The complete leading coefficients agree, but equality of estimators, exact finite-worker unbiasedness, and matching higher-order terms are not asserted.

F FIXED-SIZE BATCHES AND CONDITIONALLY UNBIASED GRADIENTS

F.1 DETERMINISTIC BATCH SIZES

Only in this subsection, replace the Poisson count by $\mathrm{K}_i \equiv B$ with B a positive integer, retaining the empirical sums and the normalization $1/B$ in Equation (4). All population symbols remain unchanged. This alternative sampling scheme is different from holding the Poisson intensity B fixed while N increases.

The same Gram–Cauchy–Binet expansion as in Appendix B now yields

$$\mathbb{E}[D_i] = \sum_{k=0}^{d} \frac{\lambda^{d-k}}{k!} \frac{(B)_k}{B^k} \int_{\mathcal{X}^k} \det\big[\boldsymbol{h}(x_a)^\top \boldsymbol{h}(x_b)\big]_{a,b=1}^{k} \prod_{a=1}^{k} \nu_{\boldsymbol{\theta}}(dx_a), \tag{59}$$

$$D = \sum_{k=0}^{d} \frac{\lambda^{d-k}}{k!} \int_{\mathcal{X}^k} \det\big[\boldsymbol{h}(x_a)^\top \boldsymbol{h}(x_b)\big]_{a,b=1}^{k} \prod_{a=1}^{k} \nu_{\boldsymbol{\theta}}(dx_a). \tag{60}$$

Here $(B)_0 = 1$, $(B)_k = B(B-1)\cdots(B-k+1)$ for $k \le B$, and $(B)_k = 0$ for $k > B$. Conditional on the count, there are exactly $\binom{B}{k} = (B)_k/k!$ subsets of size k, which proves the first formula. The second is the integral determinant expansion proved in Appendix B. Thus the correction factors are generally different from one. In particular, the Poisson expectation identity cannot be transferred to a fixed-size batch in general.

For the coupled gradient, independence of the B marks gives

$$\begin{aligned}&\mathbb{E}[\boldsymbol{\Delta}_i \boldsymbol{G}(\boldsymbol{\Delta}_i \boldsymbol{p} - \boldsymbol{\delta}_i)] \\ &\quad = \frac{1}{B} \int_{\mathcal{X}} \big[\boldsymbol{h}(x)\boldsymbol{h}(x)^\top - \boldsymbol{F}\big] \boldsymbol{G} \left[\big(\boldsymbol{h}(x)\boldsymbol{h}(x)^\top - \boldsymbol{F}\big)\boldsymbol{p} - \big(\boldsymbol{h}(x)y(x) - \boldsymbol{g}\big)\right] \nu_{\boldsymbol{\theta}}(dx).\end{aligned} \tag{61}$$

Unlike the Poisson formula, the atoms in this covariance are centered. The finite resolvent proof of Proposition 5.1 continues to apply: third centered sample-mean moments are $\mathcal{O}(B^{-2})$ and fourth moments are $\mathcal{O}(B^{-2})$ under Equation (11). Therefore

$$\begin{aligned}\mathbb{E}[\widehat{\boldsymbol{p}}_i] - \boldsymbol{p} &= \frac{1}{B} \boldsymbol{G} \int_{\mathcal{X}} \big[\boldsymbol{h}(x)\boldsymbol{h}(x)^\top - \boldsymbol{F}\big] \boldsymbol{G} \\ &\qquad \times \left[\big(\boldsymbol{h}(x)\boldsymbol{h}(x)^\top - \boldsymbol{F}\big)\boldsymbol{p} - \big(\boldsymbol{h}(x)y(x) - \boldsymbol{g}\big)\right] \nu_{\boldsymbol{\theta}}(dx) + \mathcal{O}(B^{-2}) \\ &= \frac{1}{B} \left[\boldsymbol{c} + \lambda \boldsymbol{G}\boldsymbol{F}\boldsymbol{G}\boldsymbol{p}\right] + \mathcal{O}(B^{-2}).\end{aligned} \tag{62}$$

To verify the last equality, use $\boldsymbol{F}\boldsymbol{p} - \boldsymbol{g} = -\lambda \boldsymbol{p}$ and expand the centered product:

$$\begin{aligned}&\boldsymbol{G} \int_{\mathcal{X}} (\boldsymbol{h}\boldsymbol{h}^\top - \boldsymbol{F}) G \big[(\boldsymbol{h}\boldsymbol{h}^\top - \boldsymbol{F})\boldsymbol{p} - (\boldsymbol{h}y - \boldsymbol{g})\big] \, d\nu_{\boldsymbol{\theta}} \\ &\qquad = \boldsymbol{G} \int_{\mathcal{X}} \boldsymbol{h}\boldsymbol{h}^\top \boldsymbol{G}(\boldsymbol{h}\boldsymbol{h}^\top \boldsymbol{p} - \boldsymbol{h}y) \, d\nu_{\boldsymbol{\theta}} - \boldsymbol{G}\boldsymbol{F}\boldsymbol{G}(\boldsymbol{F}\boldsymbol{p} - \boldsymbol{g}) = \boldsymbol{c} + \lambda \boldsymbol{G}\boldsymbol{F}\boldsymbol{G}\boldsymbol{p}.\end{aligned}$$

Uniform averaging retains this local coefficient, whereas pooling N independent fixed-size batches replaces B by NB. The PDA theorem does not give an exact fixed-size determinantal correction. If $\boldsymbol{F}$ remains uniformly nonsingular, the displayed coupled-coefficient difference tends to zero as $\lambda \downarrow 0$; the results here themselves assume $\lambda > 0$.

If a Poisson count is used but the empirical sums are instead divided by the realized K_i, conditioning on K_i reintroduces the factorial corrections in Equation (59). The event $\mathrm{K}_i = 0$ also needs a separate convention. This explains why the deterministic normalization $1/B$ in the proposed Poisson estimator is essential.

F.2 CONDITIONALLY UNBIASED OR INDEPENDENTLY SAMPLED GRADIENTS

Return now to the Poisson count model. As an alternative to the shared sample construction, suppose

$$\mathbb{E}[\widehat{\boldsymbol{g}}_i \mid \widehat{\boldsymbol{F}}_1, \ldots, \widehat{\boldsymbol{F}}_N] = \boldsymbol{g}, \qquad \mathbb{E}[\|\operatorname{adj}(\widehat{\boldsymbol{F}}_i + \lambda \boldsymbol{I})\| \|\widehat{\boldsymbol{g}}_i\|] < \infty. \tag{63}$$

An independent unbiased gradient batch is sufficient. This condition is an alternative sampling assumption, not an assumption used in the main shared-sample theorems. Conditioning first on the Fisher matrices and then using Corollary B.3 gives

$$\mathbb{E}[D_i\widehat{\boldsymbol{p}}_i] = \mathbb{E}[D_i(\widehat{\boldsymbol{F}}_i + \lambda \boldsymbol{I})^{-1}\boldsymbol{g}] = D\boldsymbol{G}\boldsymbol{g} = D\boldsymbol{p}.$$

Consequently the exact numerator and ratio-of-expectations identities also hold in this setting.

For the local bias, integrate out the gradient noise first. Expanding the inverse through cubic Fisher degree with its fourth-order resolvent remainder, as in Appendix A, gives

$$\mathbb{E}[\widehat{\boldsymbol{p}}_i] - \boldsymbol{p} = \frac{1}{B}\boldsymbol{G}\int_{\mathcal{X}} \boldsymbol{h}(x)\boldsymbol{h}(x)^\top \boldsymbol{G}\boldsymbol{h}(x)\boldsymbol{h}(x)^\top \boldsymbol{p}\, \nu_{\boldsymbol{\theta}}(dx) + \mathcal{O}(B^{-2}) \tag{64}$$

under $\int \|\boldsymbol{h}(x)\|^8 \nu_{\boldsymbol{\theta}}(dx) < \infty$ and the displayed integrability conditions. The second centered Fisher moment supplies the leading term; the third has expectation $\mathcal{O}(B^{-2})$, and the fourth resolvent remainder is bounded by $\lambda^{-5}\|\boldsymbol{g}\|\mathbb{E}\|\boldsymbol{\Delta}_i\|^4$.

Uniform averaging retains Equation (64); pooling replaces B by NB. For fixed N, conditioning reduces PDA to the same map with every gradient equal to $\boldsymbol{g}$. The weight calculation of Appendix D now uses

$$\mathbb{E}[\mathrm{tr}(\boldsymbol{G}\boldsymbol{\Delta}_i)(-\boldsymbol{G}\boldsymbol{\Delta}_i\boldsymbol{p})] = -\frac{1}{B}\boldsymbol{G}\int_{\mathcal{X}} \boldsymbol{h}\boldsymbol{h}^\top \boldsymbol{G}\boldsymbol{h}\boldsymbol{h}^\top \boldsymbol{p}\, d\nu_{\boldsymbol{\theta}}.$$

Thus

$$\mathbb{E}[\widehat{\boldsymbol{p}}_{\mathrm{PDA},N}] - \boldsymbol{p} = \frac{1}{NB}G\int_{\mathcal{X}} \boldsymbol{h}\boldsymbol{h}^\top \boldsymbol{G}\boldsymbol{h}\boldsymbol{h}^\top \boldsymbol{p}\, d\nu_{\boldsymbol{\theta}} + \mathcal{O}_N(B^{-2}). \tag{65}$$

For fixed B, the ratio expansion in Appendix C instead has coefficient

$$\frac{\boldsymbol{p}\,\mathrm{Var}(D_1) - \mathrm{Cov}(D_1(\widehat{\boldsymbol{F}}_1 + \lambda\boldsymbol{I})^{-1}\boldsymbol{g}, D_1)}{ND^2},$$

with remainder $\mathcal{O}_B(N^{-2})$ under the corresponding fourth moments. Its conditional mean is bounded in norm by $\|g\|/\lambda$. Applying the strong law to the determinant and inverse-weighted quantities and then dominated convergence proves convergence of expectations to p. Almost-sure consistency of the actual noisy ratio additionally uses i.i.d. integrable worker pairs, exactly as in the main-text corollary.

The uniform proof in Appendix E applies to the conditional mean by replacing $\mathrm{adj}(\boldsymbol{F} + \lambda\boldsymbol{I} + \boldsymbol{\Delta}_i)(\boldsymbol{\delta}_i - \boldsymbol{\Delta}_i\boldsymbol{p})$ with $-\mathrm{adj}(\boldsymbol{F} + \lambda\boldsymbol{I} + \boldsymbol{\Delta}_i)\boldsymbol{\Delta}_i\boldsymbol{p}$. For example, $\int \|\boldsymbol{h}(x)\|^{8d}\nu_{\boldsymbol{\theta}}(dx) < \infty$ suffices. It sharpens Equation (65) to a uniform remainder $\mathcal{O}((NB^2)^{-1})$.

Finally, under deterministic batch size $\mathrm{K}_i \equiv B$ and conditional unbiasedness, the local expansion is instead

$$\begin{aligned}\mathbb{E}[\widehat{\boldsymbol{p}}_i] - \boldsymbol{p} &= \frac{1}{B}\boldsymbol{G}\int_{\mathcal{X}} (\boldsymbol{h}\boldsymbol{h}^\top - \boldsymbol{F})G(\boldsymbol{h}\boldsymbol{h}^\top - \boldsymbol{F})\boldsymbol{p}\, d\nu_{\boldsymbol{\theta}} + \mathcal{O}(B^{-2}) \\ &= \frac{1}{B}\left[\boldsymbol{G}\int_{\mathcal{X}} \boldsymbol{h}\boldsymbol{h}^\top \boldsymbol{G}\boldsymbol{h}\boldsymbol{h}^\top \boldsymbol{p}\, d\nu_{\boldsymbol{\theta}} - \boldsymbol{G}\boldsymbol{F}\boldsymbol{G}\boldsymbol{F}\boldsymbol{p}\right] + \mathcal{O}(B^{-2}).\end{aligned} \tag{66}$$

The difference from the Poisson coefficient is $\boldsymbol{G}\boldsymbol{F}\boldsymbol{G}\boldsymbol{F}\boldsymbol{p}$; unlike the coupled difference in Equation (62), it need not vanish as $\lambda \downarrow 0$ even when F is nonsingular.

G Sample-space identities and empty-batch conventions

We verify the implementation identities using the score matrix $\boldsymbol{H}_i$ already defined in Section 4. Since $\widehat{\boldsymbol{F}}_i = B^{-1}\boldsymbol{H}_i\boldsymbol{H}_i^\top$, direct multiplication by $\lambda\boldsymbol{I} + B^{-1}\boldsymbol{H}_i\boldsymbol{H}_i^\top$ verifies

$$(\lambda\boldsymbol{I} + B^{-1}\boldsymbol{H}_i\boldsymbol{H}_i^\top)^{-1} = \lambda^{-1}\boldsymbol{I} - \frac{1}{\lambda B}\boldsymbol{H}_i(\lambda\boldsymbol{I}_{\mathrm{K}_i} + B^{-1}\boldsymbol{H}_i^\top\boldsymbol{H}_i)^{-1}\boldsymbol{H}_i^\top. \tag{67}$$

Multiplying by $\widehat{\boldsymbol{g}}_i$ proves Equation (7). Sylvester's determinant identity similarly yields

$$\begin{aligned}D_i &= \lambda^d \det(\boldsymbol{I}_d + (\lambda B)^{-1}\boldsymbol{H}_i\boldsymbol{H}_i^\top) \\ &= \lambda^d \det(\boldsymbol{I}_{\mathrm{K}_i} + (\lambda B)^{-1}\boldsymbol{H}_i^\top\boldsymbol{H}_i) \\ &= \lambda^{d-\mathrm{K}_i}\det(\lambda\boldsymbol{I}_{\mathrm{K}_i} + B^{-1}\boldsymbol{H}_i^\top\boldsymbol{H}_i).\end{aligned}$$

This proves Equation (8) for every realized K_i, including $\mathrm{K}_i > d$. Factoring out the common λ^d gives precisely s_i and w_i in Equation (9) and Equation (10); the worker-dependent term $-\mathrm{K}_i \log \lambda$ must be retained.

When $\mathrm{K}_i = 0$, the shared-sample sums are zero, $\widehat{\boldsymbol{p}}_i = 0$, $D_i = \lambda^d$, and $s_i = 0$. With the empty determinant convention $\det(I_0) = 1$, the sample-space formulas remain valid. Dropping zero-count workers would change the stated estimator and its sampling law.

An exact sample-space implementation constructs a $\mathrm{K}_i \times \mathrm{K}_i$ Gram matrix at cost $\mathcal{O}(d\mathrm{K}_i^2)$ and solves it at cost $\mathcal{O}(\mathrm{K}_i^3)$; storing scores and the Gram costs $\mathcal{O}(d\mathrm{K}_i + \mathrm{K}_i^2)$. These are algebraic costs excluding score evaluation, and depend on the realized count. With $\mathrm{K}_i \sim \mathrm{Poisson}(B)$,

$$\mathbb{E}\mathrm{K}_i^2 = B^2 + B, \qquad \mathbb{E}\mathrm{K}_i^3 = B^3 + 3B^2 + B.$$

Thus the expected dense linear-algebra work per worker is $\mathcal{O}(d(B^2 + B) + B^3 + 3B^2 + B)$; communication remains one d-vector and one scalar per worker. These identities establish estimator equivalence for exact solves. They do not by themselves establish equivalence after local clipping, inexact solves, or additional nonlinear update rules.

H Experimental Details

Actor training. We use Randomized Advantage Transformation (RAT) (Sun, 2026) as the backbone Woodbury-based NPG algorithm. For the distributed methods involved, i.e., uniform averaging, FedNPG-ADMM, and PDA, each worker computes a local RAT direction, after which the directions are aggregated with the respective method. Following the original RAT design, the aggregated update norm is clipped to ensure training stability.

Critic training. The critic is optimized using Adam, with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and weight decay$= 0$. Because our study focuses on the distributed computation and aggregation of the actor's NPG direction, implementation-wise, we train the critic centrally using the pooled samples. Since this is equivalent to aggregating sample-weighted local critic gradients before applying a single shared Adam update, training behavior remains unaffected.

Network architectures. For MuJoCo, we use a two-layer MLP with 256 hidden units per layer, ReLU activations, and no dropout, with around 7×10^4 parameters. Observations are normalized using running mean and variance statistics and clipped to $[-5, 5]$. The Gaussian policy uses a linear mean head and a learned state-independent log-standard-deviation vector, while the value head is linear with PopArt normalization.

For Procgen, we use a compact IMPALA-style ResNet with around 9×10^5 parameters. RGB observations are scaled to $[-1, 1]$ and processed by two convolutional stages with 8 and 16 channels. Each stage consists of a 3×3 convolution, 3×3 max-pooling with stride 2, and two residual blocks containing two 3×3 convolutions each. Batch normalization is disabled. The resulting features are flattened and projected to a 256-dimensional ReLU embedding, followed by linear policy and PopArt value heads.

In both cases, the actor and critic use the same backbone architecture but have separate networks.

Hyperparameters. The hyperparameters are presented in Table 1. Each worker maintains a dedicated set of 8 environments. For fixed-batch regime, At each rollout, 512 transition steps are collected from each environment. For Poissonized-batch regime, at the beginning of each rollout, independently across N workers and M minibatches, we draw $\mathrm{K}_{i,j} \sim \mathrm{Poisson}(B)$, $j = 1, \ldots, M$. Worker i then collects $T_i := \sum_{j=1}^{M} \mathrm{K}_{i,j}$ transitions from its environments. The collected data are then used to perform PPO-like multi-epoch minibatch updates, following the implementation of RAT (Sun, 2026). The actor learning rate is scheduled according to the KL divergence induced by each rollout update: if the KL divergence exceeds the upper threshold, the learning rate is divided by 1.5; if it falls below the lower threshold, the learning rate is multiplied by 1.5.

Differences between theory and implementation. Our analysis assumes i.i.d. samples and sample-local gradient signals. In practice, rollout transitions are temporally dependent, and batch-wise advantage normalization makes each gradient signal depend on other samples in the batch. We

Table 1: Hyperparameters used in the experiments.

Hyperparameter	**MuJoCo**	**Procgen**
Discount factor γ	0.999	0.999
GAE parameter λ_{GAE}	0.95	0.95
Environments per worker	8	8
(Expected) Rollout steps per environment	512	512
Batch size / Poinsson Intensity	512	512
Epochs per rollout	10	4
Number of minibatches	8	8
NPG damping	0.1	0.5
Maximum NPG norm	0.25	0.25
Initial actor learning rate	0.05	0.005
Critic learning rate	0.001	0.0002
KL upper threshold	0.016	0.0025
KL lower threshold	0.004	0.0005
Number of rollouts	1220	1525
Number of Procgen levels	N/A	10
FedNPG-ADMM-specific hyperparameter		
ADMM penalty ρ	0.1	

retain these departures because rollout sampling and advantage normalization are standard components of practical deep-RL training. The experiments therefore evaluate PDA under realistic training conditions rather than exactly reproducing the assumptions of the theoretical analysis.

I Additional Experimental Results

I.1 Ablation of Poissonization

Figure 5 compares determinantal averaging with fixed-size and Poissonized sample sizes while holding the remaining training settings fixed.

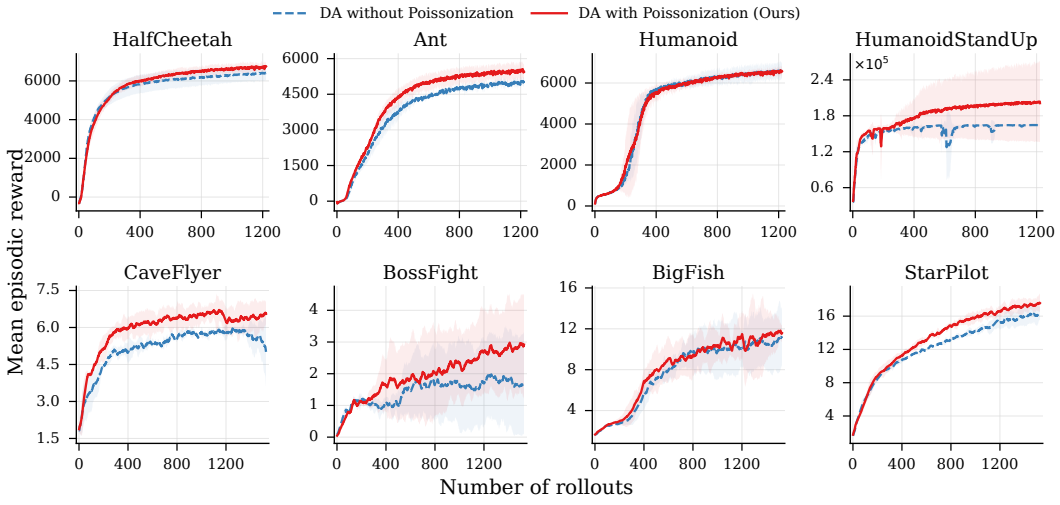


Figure 5: Training returns of determinantal averaging with fixed-size versus Poissonized sample size across eight tasks, over 3 seeds.

I.2 Measured wall-clock time and memory on a single GPU

To complement the analytical estimates in Figure 4, Figure 6 reports the measured time and peak memory per update for Single NPG, Pooled NPG, and serial PDA on a single Nvidia H200 with expected $B = 512$, for StarPilot. The time of serial PDA grows linearly in N, as it solves N local

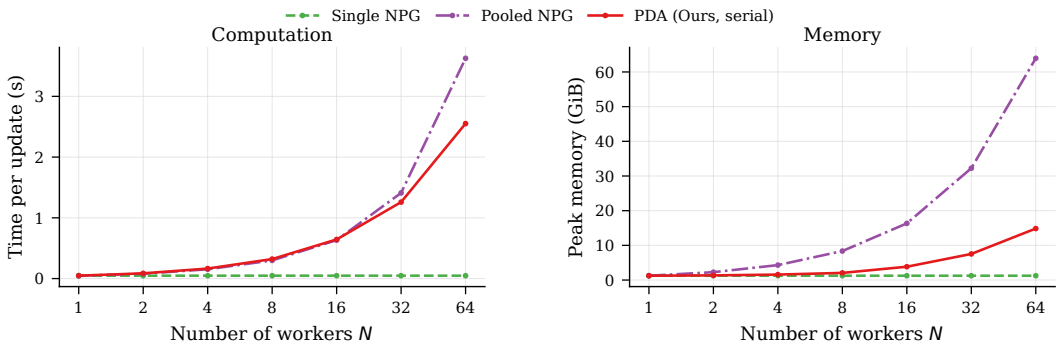


Figure 6: Measured wall-clock time and peak GPU memory per update on a single GPU for Single NPG, Pooled NPG, and serial PDA on StarPilot with expected local batch size $B = 512$ and $P = 938{,}719$ policy parameters.

problems in sequence. Pooled NPG closely tracks serial PDA for small N, where costs linear in the number of samples dominate and the larger Gram system is absorbed by GPU parallelism, but grows faster at large N as its quadratic cost takes over. Memory separates the methods more clearly. Pooled NPG grows linearly in N because it stores all NB score vectors at once, whereas serial PDA remains far below it across the full range. The residual growth of serial PDA does not come from the local sample-space solve, whose workspace is independent of N. It comes from the surrounding single-device training pipeline, which keeps the rollout observations of all N simulated workers resident on the GPU and evaluates the post-update KL divergence over the full batch. In a genuinely distributed deployment, each worker would hold only its own B samples, and this storage would not accumulate on any single device.

With a single GPU, we do not measure parallel PDA directly, but its per-worker cost follows from the $N = 1$ case. Each worker solves one local problem of expected size B, so its compute and memory match Single NPG, up to waiting for the worker with the largest realized Poisson batch in each synchronous step. Communication has the same structure as synchronous data-parallel SGD: each worker contributes one parameter-sized vector and one scalar to an all-reduce, so PDA can reuse standard collective implementations and should inherit their communication behaviour.