GRADIENT ESTIMATION VIEW OF ZEROTH-ORDER OPTIMIZERS

Anonymous authors
Paper under double-blind review

ABSTRACT

Zeroth-order (derivative-free) methods—including evolution strategies, finite differences, and sphere-sampling estimators—are often viewed as approximating gradients of smoothed objectives. But is this a universal property of ZO methods, or do only specific algorithms admit such a view? We answer this with an *update-first* reconstruction principle: given a sampler q_θ and update weights ψ, we reconstruct an implicit smoothing measure p_θ such that the expected update equals the derivative of $J(\theta) = \mathbb{E}_{x\sim p_\theta}[f(x)]$. The reconstruction equation $\partial_{\theta_j} p_\theta = q_\theta\, \psi_{\theta_j}$ cleanly separates *where* the algorithm samples (q_θ) from *which* objective it is smoothing (p_θ). It covers classical likelihood-ratio (REINFORCE) estimators, finite-difference and sphere rules whose weak derivatives live on boundaries, and previously-unexplained schemes such as random-denominator differences—for which we give the first explicit smoothed objective. Worked examples (Gaussian ES, forward and central FD, sphere sampling, randomized Kiefer–Wolfowitz, and an Epanechnikov-kernel uniform ES) yield closed-form kernels that make each method's induced smoothing visible, comparable, and interpretable.

1 INTRODUCTION

Zeroth-order (derivative-free) methods apply when gradients of $f : \mathbb{R}^d \to \mathbb{R}$ are unavailable, unreliable, or too expensive. They underpin classical derivative-free optimization (Conn et al., 2009; Larson et al., 2019), evolution strategies (Hansen & Ostermeier, 2001; Hansen, 2016; Wierstra et al., 2014; Salimans et al., 2017), and stochastic-approximation/bandit methods that infer gradient-like directions from one or a few function queries per iteration (Kiefer & Wolfowitz, 1952; Spall, 1992; Chen et al., 1999; Flaxman et al., 2005; Duchi et al., 2015; Nesterov & Spokoiny, 2017; Shamir, 2017; Ghadimi & Lan, 2013; Liu et al., 2020).

Despite their different presentations, many zeroth-order optimizers share a *sample, evaluate, weight, update* template. With parameters $\theta \in \mathbb{R}^m$ (e.g., a search mean and scale), a generic step is

$$\theta_{t+1} \;=\; \theta_t \;+\; \alpha_t\, A(\theta_t)\, \frac{1}{N} \sum_{i=1}^{N} \psi_{\theta_t}(x_i)\, f(x_i),\; x_i \sim q_{\theta_t}, \tag{1}$$

where q_{θ_t} is the sampler, $\psi_{\theta_t}(\cdot)$ the update weight vector, α_t the step size, and $A(\theta_t)$ an optional preconditioner (e.g., for Fisher/natural-gradient (Amari, 1998; Wierstra et al., 2014; Ollivier et al., 2017)). Often $f(x_i)$ is replaced by an *effective reward* r_i using baselines, monotone utilities, ranks, or elite indicators (CEM/CMA/IGO-style fitness shaping (Rubinstein, 1999; de Boer et al., 2005; Hansen, 2016; Ollivier et al., 2017)). Thus (1) specifies sampling (q_θ), how evaluations map to an update (ψ), optional geometry (A), and shaping ($f \mapsto r$).

A standard bridge to gradient-based optimization views many zeroth-order estimators as gradients of a *smoothed objective*

$$J(\theta) = \mathbb{E}_{x\sim p_\theta}[f(x)], \tag{2}$$

for some smoothing distribution p_θ. Score-function (likelihood ratio) estimators make this explicit through REINFORCE and Gaussian ES (Glynn, 1990; Williams, 1992; Wierstra et al., 2014; Hansen, 2016); sphere-sampling estimators give unbiased gradients of a *ball-smoothed* objective even when samples lie on the *sphere* (Flaxman et al., 2005; Duchi et al., 2015; Nesterov & Spokoiny, 2017;

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.

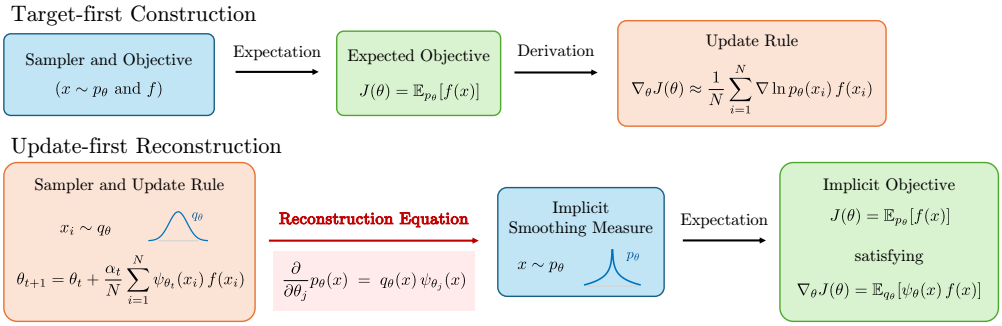

Figure 1: The conventional *target-first* and our *update-first* views. The target-first view constructs an update rule from a given sampler p_θ. In the update-first view, the algorithm specifies a sampler q_θ (where it queries) and weights $\psi_\theta(\cdot)$ (how it turns evaluations into an update). We then reconstruct an implicit smoothing measure p_θ such that the *expected* update equals $\nabla_\theta J(\theta)$ for $J(\theta) = \mathbb{E}_{x\sim p_\theta}[f(x)]$. In general $p_\theta \neq q_\theta$: score-function methods give $p_\theta = q_\theta$, while finite-difference and sphere estimators typically do not.

Shamir, 2017; Ghadimi & Lan, 2013). Such derivations are usually "target-first": choose a smoothing rule (Gaussian, ball, or interval) and derive an estimator.

Algorithms, however, are often specified update-first: implement q_θ, weights $\psi_\theta(\cdot)$, and optional shaping/preconditioning, then ask what objective is optimized, giving the inverse question:

> Given a sampler q_θ and weights ψ_θ in (1), what objective $J(\theta)$ in (2) is the method differentiating?

Our *update-first reconstruction* principle answers this question (Figure 1). Starting from q_θ and ψ_θ in (1), the population update is $\mathbb{E}_{x\sim q_\theta}[f(x)\psi_\theta(x)]$ (or the shaped-reward analogue). For a rich class of test functions, requiring this update to equal $\nabla_\theta J(\theta)$ is equivalent to a *reconstruction equation* for how an implicit smoothing measure p_θ changes with θ (Section 3); integrating it yields the objective $J(\theta) = \mathbb{E}_{x\sim p_\theta}[f(x)]$ differentiated by the method.

The reconstruction equation (15) separates *where we sample* (q_θ) from the smoothing measure (p_θ):

$$\underbrace{\partial_{\theta_j} p_\theta(x)}_{\text{smoothing measure}} = \underbrace{q_\theta(x)}_{\text{sampler}} \underbrace{\psi_{\theta_j}(x)}_{\text{weights}}.$$

In general $p_\theta \neq q_\theta$. For score-function weights $\psi_\theta(x) = \nabla_\theta \log q_\theta(x)$ (REINFORCE / Gaussian ES), the canonical reconstruction is $p_\theta = q_\theta$ after fixing the integration constant (Appendix I). The same calculus also covers finite-difference stencils, boundary/sphere estimators, and random-denominator schemes, yielding explicit kernels for inspection, visualization, and comparison (Appendix L, Appendix M, and Appendix K). Natural-gradient methods add a parameter-space preconditioner without changing the reconstructed smoothing measure (Appendix O.3).

For common location–scale parameterizations in evolution strategies, reconstruction reduces to transport and radial-flux identities, yielding practical integration templates and often closed-form kernels. More broadly, reconstruction completes the REINFORCE picture of zeroth-order optimization: Parmas & Sugiyama (2021) show that if an algorithm optimizes $J(\theta) = \mathbb{E}_{x\sim p_\theta}[f(x)]$ using an unbiased zeroth-order gradient estimator, then it must be a REINFORCE-style estimator. Our converse starts from the implemented sampler q_θ and weights $\psi_\theta(\cdot)$, then reconstructs the compatible smoothing measure p_θ whose induced objective the update differentiates. Thus, within the sampling-and-weighting template, essentially all centered zeroth-order updates fit the REINFORCE lens, with their smoothing behavior made explicit. For a newly proposed method in this template, reconstruction is a diagnostic: it identifies the objective differentiated by the implemented sampler and weights, rather than prescribing how q_θ or ψ_θ should be chosen.

Scope. "Sampling-and-weighting" here means any update of the form (1)—a sampler q_θ, pointwise weights $\psi_\theta(x)$, and a (possibly shaped) reward $r(x)$. This covers all the methods in Table 1 and

modern variants discussed in Appendix O (rank/elite shaping, natural-gradient preconditioning, MeZO-style finite differences). It does not cover DFO methods whose steps are chosen by geometric polling or local surrogate-model logic, such as direct search, MADS, and model-based trust-region methods (Kolda et al., 2003; Audet & Dennis, 2006; Conn et al., 2009; Larson et al., 2019). It also excludes step-size/self-adaptation rules that regulate scalar search statistics rather than estimate gradients of $\mathbb{E}_{x\sim p_\theta}[f(x)]$ (Rechenberg, 1973; Schwefel, 1995; Beyer & Schwefel, 2002; Hansen & Ostermeier, 2001); Appendix O.2 illustrates this boundary for the $1/5$ success rule.

For readability, we use "objective" and "component-wise objective" interchangeably when discussing one parameter component. More precisely, whenever the required integral exists, a fixed reference or boundary convention gives a unique full candidate $p_\theta^{(j)}$ and an objective $J_j(\theta) = \mathbb{E}_{x\sim p_\theta^{(j)}}[r(x)]$ satisfying $\partial_{\theta_j} J_j = g_j$. We usually suppress the superscript and write p_θ and J. A single joint measure exists when the component-wise candidates agree; otherwise, each component-wise derivative identity remains exact. Even when the candidates differ, the full expected update for a particular objective can still be a gradient field under the weaker objective-specific condition in Appendix E.

2 Background

This section fixes notation for sampling-based routes to gradient information. We distinguish (i) *small-step* estimators that approximate $\nabla f(\mu)$ as the step size vanishes from (ii) *exact* gradients of smoothed objectives $J(\theta) = \mathbb{E}_{x\sim p_\theta}[f(x)]$ at fixed scale. The key of our paper is that our reconstruction covers the breadth of all of these, and while some smoothing links are classical, others are new.

Quick taxonomy (names used below). We use *Gauss ES* for the Gaussian score estimator (9) and *Unif ES* for uniform sampling with the linear weight (100); *Sphere* for the uniform-ball boundary/sphere estimator (88) (Appendices H.2 and M); and *Gauss-denom/Unif-denom* for the random-denominator central difference (6) with ε drawn from a Gaussian/uniform law (Appendix K and Section 5.3).

2.1 Deterministic finite differences ($\Delta \to 0$)

In 1D, for step size $h > 0$, the forward and central finite differences for $f'(\mu)$ are

$$\widehat{g}_{\mathrm{fwd}}(\mu; h) := \frac{f(\mu + h) - f(\mu)}{h}, \qquad \widehat{g}_{\mathrm{cent}}(\mu; h) := \frac{f(\mu + h) - f(\mu - h)}{2h}, \tag{3}$$

and if f is differentiable, then both converge: $\lim_{h\to 0} \widehat{g}_{\mathrm{fwd}}(\mu; h) = \lim_{h\to 0} \widehat{g}_{\mathrm{cent}}(\mu; h) = f'(\mu)$, with truncation error $O(h)$ (forward) versus $O(h^2)$ (central) under additional smoothness.

In d dimensions, applying the central difference along coordinates e_i gives

$$\widehat{g}_i(\mu; h) := \frac{f(\mu + he_i) - f(\mu - he_i)}{2h}, \quad i = 1, \ldots, d, \tag{4}$$

at a cost of $2d$ function evaluations per iteration. With noisy evaluations (e.g., Kiefer–Wolfowitz), smaller h reduces bias but increases variance/noise amplification (Kiefer & Wolfowitz, 1952; Spall, 2003; Conn et al., 2009; Larson et al., 2019).

2.2 Randomized perturbations: fixed vs. random denominators

Coordinate-wise finite differences (4) cost $O(d)$ queries per iteration. Randomized schemes instead perturb along a random direction, using one or two evaluations per iteration independent of dimension (Chen et al., 1999; Spall, 1992).

Fixed-denominator directional difference (Gaussian or uniform directions). Let Z be a zero-mean isotropic direction with $\mathbb{E}[ZZ^\top] = I_d$ (e.g. $Z \sim \mathcal{N}(0, I_d)$, or a suitably scaled uniform-sphere direction). For $\sigma > 0$, define the two-point directional estimator

$$\widehat{g}_{\mathrm{dir}}(\mu; \sigma, Z) := \frac{f(\mu + \sigma Z) - f(\mu - \sigma Z)}{2\sigma} Z. \tag{5}$$

If f is sufficiently smooth and σ is small, then $\mathbb{E}[\widehat{g}_{\mathrm{dir}}(\mu; \sigma, Z)] = \nabla f(\mu) + O(\sigma^2)$.

Random-denominator central differences (Gauss-denom / Unif-denom, 1D). Using the realized perturbation as the step length yields a random-denominator estimator closely related to SPSA-style simultaneous-perturbation quotients (Chen et al., 1999; Spall, 1992; 2003). For a symmetric perturbation ε (e.g. Gaussian or uniform),

$$\widehat{g}_{\mathrm{den}}(\mu;\varepsilon) \;:=\; \frac{f(\mu+\varepsilon)-f(\mu-\varepsilon)}{2\varepsilon}. \tag{6}$$

These estimators can be viewed as $\sigma \to 0$ approximations to $f'(\mu)$, but at a *fixed* scale they correspond to smoothed objectives with qualitatively different kernels (logarithmic singularities); see Appendix K and the reconstructions in Section 5.3.

2.3 Likelihood-ratio gradients and Gaussian ES

The likelihood-ratio (score-function) identity provides a different viewpoint: rather than approximating $\nabla f(\mu)$ by $\sigma \to 0$, one defines an *explicitly smoothed* objective and considers its exact gradient.

Consider the Gaussian-smoothed objective

$$J(\mu,\sigma) \;:=\; \mathbb{E}\big[f(x)\big], \qquad x \sim \mathcal{N}(\mu, \sigma^2 I_d), \tag{7}$$

where $\mu \in \mathbb{R}^d$ is a location parameter and $\sigma > 0$ is a smoothing scale. A standard score-function calculation yields the exact gradient

$$\nabla_\mu J(\mu,\sigma) \;=\; \mathbb{E}\Big[f(x)\,\nabla_\mu \log \mathcal{N}(x;\mu,\sigma^2 I_d)\Big] \;=\; \mathbb{E}\left[f(x)\,\frac{x-\mu}{\sigma^2}\right]. \tag{8}$$

Equivalently, writing $x = \mu + \sigma Z$ with $Z \sim \mathcal{N}(0, I_d)$ gives the common reparameterized form

$$\nabla_\mu J(\mu,\sigma) \;=\; \frac{1}{\sigma}\,\mathbb{E}\big[f(\mu+\sigma Z)\,Z\big]. \tag{9}$$

Gaussian ES as an exact match. In its simplest form, Gaussian evolution strategies estimate (9) by Monte Carlo: draw $Z_1,\dots,Z_N \sim \mathcal{N}(0, I_d)$ and use

$$\widehat{\nabla_\mu J}(\mu,\sigma) \;:=\; \frac{1}{N\sigma}\sum_{i=1}^{N} f(\mu+\sigma Z_i)\,Z_i, \tag{10}$$

(or an antithetic/two-point variant for variance reduction). Thus Gaussian ES is an unbiased estimator of the *exact* gradient of the Gaussian-smoothed objective $J(\mu,\sigma) = \mathbb{E}[f(\mu+\sigma Z)]$ for the chosen smoothing scale σ (Wierstra et al., 2014; Glasmachers et al., 2010; Hansen, 2016). Appendix I verifies in the reconstruction framework that score weights force $p = q$.

Uniform interval and ball smoothing yield endpoint and boundary estimators, respectively; we review these identities in Appendix H.2.

Connection to the reconstruction principle. The estimators above are usually derived "target-first" from a chosen smoothing rule (Gaussian convolution, interval/ball averages). In the remainder of the paper we invert this logic: given a sampling-and-weighting rule, we reconstruct the implicit smoothing distribution and objective whose gradient it estimates (Section 3).

3 Reconstruction Principle: From (q,ψ) to an Implicit Objective

Much of the zeroth-order literature clarifies unbiased gradient estimators once the target expectation $\mathbb{E}_{p_\theta}[\cdot]$ is specified. However, there also exist zeroth-order optimizers that have not been characterized in terms of unbiased gradient estimators. In practice, algorithms specify (i) how to *sample* query points and (ii) how to *weight* evaluations to form an update. We therefore ask: Given an algorithm's sampler q_θ and its weight function $\psi_\theta(\cdot)$, what smoothed objective is it (exactly) differentiating?

Our answer is an *update-first reconstruction* principle: we reconstruct a smoothing distribution (kernel) p_θ whose parameter derivative matches the algorithm's expected weighted update. Crucially, the sampler q_θ and induced smoothing distribution p_θ need not coincide: score-function methods often have $p_\theta = q_\theta$, whereas finite-difference and boundary estimators generally do not.

Table 1 summarizes this section: reconstruction recovers known cases (e.g., Gaussian ES) and adds the new Unif/Gauss-denom and Unif ES with its Epanechnikov kernel (Section 5).

Table 1: Canonical reconstructions. For each method, the sampling rule q and weight ψ determine a (possibly generalized) reconstruction of $\partial_\mu p$, which integrates to an explicit smoothing density p. The induced objective is $J(\mu) = \int f(x)p(x \mid \mu)\,\mathrm{d}x$ (or its (μ, σ) analogue).

Method	**Sampler** q	**Weight** ψ_μ **(location)**	**Reconstructed** p	**Induced objective** J
Forward FD	$x = \mu, \mu + h$	$(\mu) = -\frac{2}{h}, (\mu+h) = +\frac{2}{h}$	$\frac{1}{h}\mathbb{1}\{\mu \le x \le \mu + h\}$	$\frac{1}{h}\int_{\mu}^{\mu+h} f(x)\,\mathrm{d}x$
Central FD	$x = \mu \pm h$	$(\mu \pm h) = \pm\frac{1}{h}$	$\frac{1}{2h}\mathbb{1}\{\mu-h \le x \le \mu+h\}$	$\frac{1}{2h}\int_{\mu-h}^{\mu+h} f(x)\,\mathrm{d}x$
Unif-denom	$\frac{1}{2\sigma}\mathbb{1}_{\mathcal{I}_\sigma(\mu)}$	$\frac{1}{x-\mu}$	$\frac{1}{2\sigma}\log\!\big(\frac{\sigma}{\lvert x-\mu\rvert}\big)\mathbb{1}_{\mathcal{I}_\sigma(\mu)}$	$\int f(x)p(x \mid \mu, \sigma)\,\mathrm{d}x$
Gauss-denom	$\mathcal{N}(\mu, \sigma^2)$	$\frac{1}{x-\mu}$	$-\frac{1}{2\sigma\sqrt{2\pi}}\,\mathrm{Ei}\big(-\frac{(x-\mu)^2}{2\sigma^2}\big)$	$\int f(x)p(x \mid \mu, \sigma)\,\mathrm{d}x$
Unif ES	$\frac{1}{2\sigma}\mathbb{1}_{\mathcal{I}_\sigma(\mu)}$	$\frac{3(x-\mu)}{\sigma^2}$	$\frac{3}{4\sigma}\Big(1-\frac{(x-\mu)^2}{\sigma^2}\Big)\mathbb{1}_{\mathcal{I}_\sigma(\mu)}$	$\int_{\mathbb{R}} f(x)\,p(x \mid \mu)\,\mathrm{d}x$
Gaussian ES	$\mathcal{N}(\mu, \sigma^2 I_d)$	$\nabla_\mu \log q = \frac{x-\mu}{\sigma^2}$	$p = q$	$\mathbb{E}_{x\sim\mathcal{N}(\mu,\sigma^2 I_d)}[f(x)]$
Sphere estimator	$\propto \delta(\lVert x-\mu\rVert - \sigma)$	$\frac{d}{\sigma^2}(x-\mu)$	$\frac{1}{V_d\sigma^d}\mathbb{1}_{B(\mu,\sigma)}$	$\frac{1}{V_d\sigma^d}\int_{B(\mu,\sigma)} f(x)\,\mathrm{d}x$

3.1 Setup: sample, evaluate, weight

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a black-box objective and let $\theta \in \mathbb{R}^m$ denote algorithm parameters (e.g., a location μ and scale σ, treated explicitly in Section 4). At each iteration, the algorithm samples $x \sim q_\theta(\cdot)$, evaluates a scalar *reward* $r(x)$ derived from $f(x)$ (possibly after a pointwise utility or other transform), and forms an update direction

$$\widehat{g}_j(\theta) \;=\; r(x)\,\psi_{\theta_j}(x), \qquad j = 1, \dots, m, \tag{11}$$

(in practice averaged over multiple samples). Here $\psi_{\theta_j}(\cdot)$ is the jth component of the weight function. Taking expectations under the sampler gives the population update direction

$$g_j(\theta) \;:=\; \mathbb{E}_{x\sim q_\theta}\big[r(x)\,\psi_{\theta_j}(x)\big] \;=\; \int r(x)\,q_\theta(x)\,\psi_{\theta_j}(x)\;\mathrm{d}x, \tag{12}$$

where we write densities for simplicity.

3.2 Target form: gradients of smoothed objectives

We say the update is a gradient estimator of a smoothed objective if there exists a family of smoothing densities p_θ such that, for each j,

$$\frac{\partial}{\partial\theta_j}J(\theta) \;=\; g_j(\theta) \quad \text{with} \quad J(\theta) \;:=\; \mathbb{E}_{x\sim p_\theta}\big[r(x)\big] \;=\; \int r(x)\,p_\theta(x)\;\mathrm{d}x. \tag{13}$$

If p_θ is classically differentiable in θ, then differentiating under the integral gives

$$\frac{\partial}{\partial\theta_j}J(\theta) \;=\; \int r(x)\,\frac{\partial}{\partial\theta_j}p_\theta(x)\;\mathrm{d}x. \tag{14}$$

When p_θ has sharp boundaries (interval/ball smoothing), the identity still holds with weak (distributional) parameter derivatives, yielding endpoint/boundary terms (Appendix H.2).

3.3 The reconstruction equation

Comparing (12) and (14) gives the *reconstruction equation*

$$\frac{\partial}{\partial\theta_j}p_\theta(x) \;=\; q_\theta(x)\,\psi_{\theta_j}(x). \tag{15}$$

For a fixed reward r, (15) is a sufficient way to make (13) hold. For a rich class of rewards, for example all smooth compactly supported test functions, it is also necessary in the weak sense: equality of the pairings in (12) and (14) for all such r forces the distributions $\partial_{\theta_j}p_\theta$ and $q_\theta\psi_{\theta_j}$ to agree. Thus the algorithmic weight specifies the weak parameter derivative that the implicit smoothing measure must have.

The reconstruction equation is an "inverse" counterpart of the score-function identity: rather than starting with a known p_θ, we start from the algorithmic weights ψ and integrate (15) to recover p_θ.

Recovering p_θ by integration. For a fixed component j, Equation (15) is a first-order differential equation in θ_j. Holding the remaining parameters fixed, integration determines p_θ up to a term that is independent of θ_j. Whenever the equation admits a solution in the stated function or measure class, a specified reference family or boundary convention fixes this term and makes the component-wise reconstruction unique; normalization and nonnegativity are separate checks. More generally, any two of p_θ, q_θ, and ψ_{θ_j} determine the third through (15), provided the required derivatives exist and division is used only where the denominator is nonzero. For score-function weights $\psi_j = \partial_{\theta_j} \log q_\theta$, (15) implies $\partial_{\theta_j} p_\theta = \partial_{\theta_j} q_\theta$, so $p_\theta = q_\theta + C$ for a θ_j-independent offset; the canonical Gaussian ES reconstruction fixes $C = 0$.

3.4 Centering and the mass-drift decomposition

The smoothing density must keep unit total mass, $\int p_\theta(x)\, \mathrm{d}x = 1$ for all θ. Under $\partial_{\theta_j} p_\theta(x) = q_\theta(x)\psi_{\theta_j}(x)$, this requires $\mathbb{E}_{x\sim q_\theta}[\psi_{\theta_j}(x)] = 0$. When it fails, the expected update $g_j(\theta) = \mathbb{E}_{q_\theta}[r(x)\psi_{\theta_j}(x)]$ decomposes into a covariance term plus a deterministic *mass drift*:

$$g_j(\theta) = \mathrm{Cov}_{q_\theta}\big(r(x), \psi_{\theta_j}(x)\big) + \mathbb{E}_{q_\theta}[r(x)]\, \mathbb{E}_{q_\theta}[\psi_{\theta_j}(x)].$$

The drift changes total mass and cannot arise from the gradient of a normalized expectation.

Centering removes this obstruction and isolates the covariance term: either center the weights, $\psi^{\mathrm{ctr}}_{\theta_j}(x) = \psi_{\theta_j}(x) - \mathbb{E}_{q_\theta}[\psi_{\theta_j}(x)]$, or subtract a mean baseline $b(\theta) = \mathbb{E}_{q_\theta}[r(x)]$ from the reward. The two are equivalent in expectation and algebraically identical with the same batch means (Appendix C). Appendix D gives a leave-one-out baseline for unbiased covariance estimates. Thus the limitation is benign: centering or a mean baseline removes the drift.

4 Location–Scale Families: Explicit Reconstruction Templates

In general, (15) cannot be integrated analytically without additional assumptions. Here we impose the common location–scale structure used by many evolutionary-computation methods: A location–scale family has the form

$$p(x \mid \mu, \sigma) = \frac{1}{\sigma^d}\, p_0\left(\frac{x-\mu}{\sigma}\right) \tag{16}$$

with a *location* (search center μ) and a *scale* (step size σ). The same identity holds for q when q is itself a location–scale family. Under this structure, reconstruction yields formulas that turn (q, ψ) into an explicit candidate smoothing density p.

Intuitively, changing the mean μ translates probability mass, while changing the scale σ induces radial in/out "flux"; these pictures yield simple PDEs and 1D integration rules.

4.1 Setup: sampler, weights, and reconstruction equations

Let $\mu \in \mathbb{R}^d$ and $\sigma > 0$ be parameters, and let $q(x \mid \mu, \sigma)$ be the algorithm's sampling density. Assume the algorithm uses weights $\psi_\mu(x) \in \mathbb{R}^d$ and $\psi_\sigma(x) \in \mathbb{R}$ to form expected update directions:

$$g_\mu(\mu, \sigma) = \mathbb{E}_{x\sim q}[r(x)\, \psi_\mu(x)], \qquad g_\sigma(\mu, \sigma) = \mathbb{E}_{x\sim q}[r(x)\, \psi_\sigma(x)].$$

As in Section 3, reconstruction seeks $p(x \mid \mu, \sigma)$ s.t. for $J(\mu, \sigma) = \mathbb{E}_{x\sim p(\cdot\mid\mu,\sigma)}[r(x)]$, $\nabla_\mu J = g_\mu$ and $\partial_\sigma J = g_\sigma$. For a rich class of rewards, this is guaranteed by the reconstruction equations

$$\frac{\partial}{\partial \mu_k} p(x \mid \mu, \sigma) = q(x \mid \mu, \sigma)\, \psi_{\mu_k}(x),\ k = 1, \ldots, d, \tag{17}$$

$$\frac{\partial}{\partial \sigma} p(x \mid \mu, \sigma) = q(x \mid \mu, \sigma)\, \psi_\sigma(x). \tag{18}$$

Centering assumption (mass conservation). To avoid the drift terms explained in Section 3.4, we assume either (i) centered weights, $\psi \mapsto \psi^{\mathrm{ctr}}$, or (ii) an equivalent mean reward baseline.

4.2 Location derivatives become transport equations

A basic calculus fact is that differentiating with respect to μ is the same as differentiating with respect to x with a minus sign:

$$\frac{\partial}{\partial \mu_k} p(x \mid \mu, \sigma) = -\frac{\partial}{\partial x_k} p(x \mid \mu, \sigma). \tag{19}$$

This is just translation invariance: increasing μ_k shifts the whole density to the right in coordinate k.

Plugging (19) into (17) yields a first-order PDE in x:

$$\frac{\partial}{\partial x_k} p(x \mid \mu, \sigma) = -\, q(x \mid \mu, \sigma)\, \psi_{\mu_k}(x). \tag{20}$$

This is a *transport equation*: the right-hand side specifies how the density must slope in x in order for its μ-derivative to match the algorithmic weighting rule.

Explicit integration. When $d = 1$, (20) is an ODE. If we impose the natural boundary condition $p(x \mid \mu, \sigma) \to 0$ as $x \to +\infty$, then integrating from x to $+\infty$ gives the explicit reconstruction

$$p(x \mid \mu, \sigma) = \int_x^\infty q(t \mid \mu, \sigma)\, \psi_\mu(t)\, \mathrm{d}t. \tag{21}$$

This formula is the mechanism behind our 1D examples: given q and ψ_μ, you can literally integrate once to recover p. In higher dimensions, the same one-dimensional integral, with the other coordinates held fixed, gives a unique full candidate $p^{(j)}(x \mid \mu, \sigma)$ for each location coordinate j (Appendix E).

Consistency. A single joint p exists exactly when these full candidates agree. Under standard differentiability on a simply connected domain, an equivalent test is that the Jacobian of $q\psi_\mu$ is symmetric, the usual "curl-free" condition for a vector field to be the gradient of a scalar potential. Even when the candidates differ, the expected update for a particular f may still be a gradient field under the weaker objective-specific condition. Appendix E explains both conditions and their assumptions. Reconstruction also gives a compatible scale update: after reconstructing p from the location rule, the unique scale weight for the same objective is $\psi_\sigma^{\mathrm{comp}} = \partial_\sigma p / q$ wherever $q > 0$. Score weights give the familiar jointly compatible case $p = q$.

4.3 Scale derivatives become radial-flux equations

The scale (σ) case can be derived by rewriting $\partial_\sigma p$ as a radial-flux/divergence expression and then reducing the resulting PDE along rays (Appendix G). The final integration template is simple: write $x = \mu + ru$ with $r = \|x - \mu\| \geq 0$ and $u \in \mathbb{S}^{d-1}$, then

$$p(\mu + ru \mid \mu, \sigma) = -\frac{\sigma}{r^d} \int_0^r t^{d-1}\, q(\mu + tu \mid \mu, \sigma) \psi_\sigma(\mu + tu)\, \mathrm{d}t. \tag{22}$$

4.4 What the templates give you in practice

Equations (21) and (22) turn the abstract reconstruction equation into concrete integration rules. Mean updates solve the transport equation $\nabla_x p = -q\, \psi_\mu$, while scale updates solve a radial-flux equation. After integration, two checks remain: normalization, $\int p(x \mid \mu, \sigma)\, \mathrm{d}x = 1$, and nonnegativity, $p \geq 0$. Score-function methods satisfy both automatically with the canonical choice $p = q$; boundary, and finite-difference estimators often yield valid interval/ball kernels. If either check fails, reconstruction still identifies a signed or unnormalized smoothing functional plus any mass drift (Section 3.4).

5 Worked Reconstructions via Location–Scale Templates

Section 4 derived practical templates: transport equations for location parameters and radial-flux equations for scales, with closed-form 1D integrations. We apply them across several zeroth-order estimators to show the breadth of settings covered and make explicit the mapping

$$(q, \psi) \quad \longmapsto \quad p \quad \longmapsto \quad J(\mu, \sigma) = \mathbb{E}_{x \sim p}[f(x)]$$

The breadth of reconstructed kernels is visualized in Figure 2.

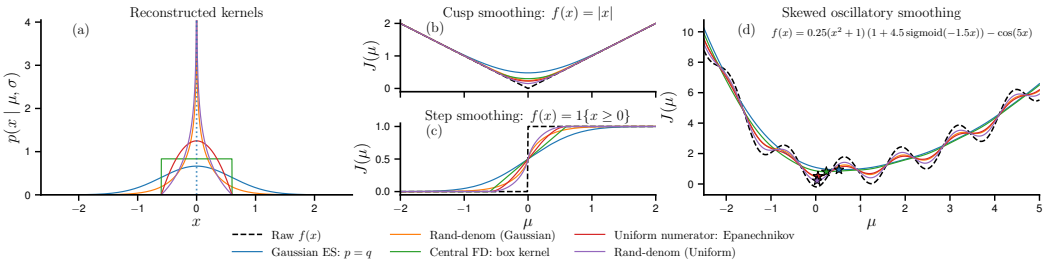

Figure 2: Panel (a) shows reconstructed kernels $p(x \mid \mu, \sigma)$; panels (b)–(d) show the induced smoothed objectives for a cusp, a step, and a skewed oscillatory objective. The key observation is that the derived densities p_θ allow observing qualitative differences in the resulting objective function.

5.1 Score weights as a sanity check: Gaussian ES gives $p = q$

For likelihood-ratio/score-function estimators, the reconstruction often collapses: if $\psi_\theta(x) = \nabla_\theta \log q(x \mid \theta)$, then $q(x \mid \theta)\psi_\theta(x) = \nabla_\theta q(x \mid \theta)$, so the reconstruction equation implies $\nabla_\theta p = \nabla_\theta q$. Fixing the offset in (94) to zero gives $p = q$. This is the familiar case of Gaussian ES and related NES/IGO updates. Because this serves mainly as a consistency check, we place the details in Appendix I.

5.2 Uniform perturbations with ES-style weights

A commonly used alternative to Gaussian perturbations is to draw random directions with i.i.d. uniform components and form a two-point estimate along the direction (see, e.g., Prashanth et al. (2017)). In one dimension (to keep notation simple), the resulting two-point update takes the form

$$\widehat{g}(\mu) \;=\; \frac{3}{\eta^2}\,\xi\,\frac{f(\mu+\delta\xi) - f(\mu-\delta\xi)}{2\delta}, \qquad \xi \sim \mathrm{Unif}([-\eta, \eta]). \tag{23}$$

The analysis in Prashanth et al. (2017) treats this estimator as a randomized finite-difference rule and studies its bias via Taylor expansion as $\delta \to 0$. From the reconstruction viewpoint, we can instead identify an *exact* smoothed objective whose gradient is estimated by (23) for any fixed δ.

Applying the reconstruction equation for location families yields the corresponding smoothing density with $\psi_\mu(x) = 3(x-\mu)/\sigma^2$ (see Appendix J) as

$$p(x \mid \mu) \;=\; \frac{3}{4\sigma}\Big(1 - \frac{(x-\mu)^2}{\sigma^2}\Big)\,\mathbb{1}_{\mathcal{I}_\sigma(\mu)}, \tag{24}$$

i.e., a compactly supported quadratic (Epanechnikov) kernel. Therefore, the two-point update (23) estimates $\nabla_\mu J(\mu)$ for the smoothed objective $J(\mu) = \mathbb{E}_{x\sim p(\cdot|\mu)}[f(x)]$ with p given by (24).

5.3 Random-denominator central differences: explicit p for non-score weights

We now apply (21) to a method that *does not* use a score weight. Consider the 1D antithetic estimator with a *random denominator* in (6). This is closely related to randomized-difference quotients and SPSA-type constructions (Spall, 1992; 2003; Liu et al., 2020).

The reconstructed densities for uniform perturbations and gaussian perturbations are derived as

$$p_{\mathrm{rd,unif}}(x \mid \mu, \sigma) = \frac{1}{2\sigma}\log\Big(\frac{\sigma}{|x-\mu|}\Big)\;\mathbb{1}_{\mathcal{I}_\sigma(\mu)} \tag{25}$$

$$p_{\mathrm{rd,gauss}}(x \mid \mu, \sigma) = -\frac{1}{2\sigma\sqrt{2\pi}}\;\mathrm{Ei}\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big), \tag{26}$$

respectively. The induced objective is $J_{\mathrm{rd}}(\mu, \sigma) = \int f(x) p_{\mathrm{rd}}(x \mid \mu, \sigma)\,\mathrm{d}x$ whenever $f \in L^1(p_{\mathrm{rd}})$, i.e., whenever its expectation under the reconstructed distribution exists. Bounded measurable f is a simple sufficient case; continuity is not required. We retain the antithetic quotient as the primary update, and Appendix K uses the one-point weighted form only as symmetry-based shorthand.

Why this example matters. The random-denominator estimator (6) has been studied since (Spall, 1992; Chen et al., 1999), and understood as *approximating* the gradient of $f(x)$. To our knowledge, no prior work identifies an objective whose *exact* gradient this estimator computes. Reconstruction supplies one—and shows that uniform vs. Gaussian perturbations correspond to *qualitatively different* smoothing kernels (both with logarithmic singularities, but with different envelopes). This is a concrete instance of the framework producing a previously-unstated result.

6 Discussion and Limitation

Reconstruction provides an update-first, REINFORCE-style interpretation of zeroth-order optimization: given an implemented sampler q_θ and weights ψ_θ, we construct an implicit smoothing measure p_θ such that the expected update equals the weak gradient of $J(\theta) = \mathbb{E}_{x\sim p_\theta}[r(x)]$. This makes the question "what objective is this code optimizing?" concrete: it is encoded by the kernel p_θ.

A key consequence is that sampling and smoothing are distinct: q_θ can live on a boundary or stencil, while p_θ describes the interior averaging that defines the objective (Table 1). Making p_θ explicit lets us compare estimators by smoothing behavior rather than sampling rule. Figure 2 illustrates how updates regularize cusps and steps. In the rightmost panel, for example, Rand-denom and uniform numerator variants preserve local minima, whereas Gaussian ES and Central FD shift the global minimum differently. A fair comparison would tune the smoothing scale per algorithm, but nevertheless reconstruction exposes the update's smoothing properties.

We envision many benefits of this framework. By unifying zeroth-order optimizers under the REINFORCE lens, it becomes possible to create a unified convergence analysis that depends on the properties of the derived density p_θ and gradient variance. Moreover, methods that are classically viewed as approximations to the derivative such as the randomized denominator methods, are shown by our analysis to actually also be just REINFORCE estimators of a different objective.

As one concrete example, Appendix F connects reconstruction to biased-gradient analysis. Our analysis with fixed-scale location reconstructions shows that, for an L-smooth objective, the bias $\|g_\mu(\mu) - \nabla f(\mu)\|$ is bounded by ζ_{rec}, the expected perturbation norms of the component-wise kernels. Thus the relative-plus-additive bias model of Ajalloeian & Stich (2020) holds with $m = 0$ and $\zeta = \zeta_{\mathrm{rec}}$. Their theory turns this bias into a stationarity or optimization floor. If f is λ-strongly convex, every expected equilibrium $g_\mu(\widehat{\mu}) = 0$ also satisfies $\|\widehat{\mu} - \mu^\star\| \le \zeta_{\mathrm{rec}}/\lambda$. These results remain valid when the component-wise kernels differ. In compatible common-kernel cases, the same representation reduces to standard smoothing-bias analyses (Nesterov & Spokoiny, 2017; Berahas et al., 2022). A separate conditional-variance bound, calculated from the original estimator (q, ψ, f), would complete the estimator-specific inputs to existing stochastic convergence results.

Two limitations are worth separating. First, if $\mathbb{E}_{q_\theta}[\psi_j] \neq 0$, the update contains deterministic *mass drift* that cannot arise from a normalized expectation. Centering/baselines remove this term (restoring "mass conservation"), so this limitation is usually benign; if a method reduces to drift after centering, it is more naturally interpreted as a control law (e.g., step-size rules). Second, reconstruction integrates one parameter direction at a time, yielding one candidate smoothing distribution per gradient component; these candidates need not be compatible with a single joint p_θ. Even then, the construction identifies the smoothed objective differentiated by each separate gradient component.

Accounting for these small adjustments brings a broad range of practical methods into the framework. Further discussion of batch-dependent ranks and elite selection, preconditioning, and LLM-scale ZO methods such as MeZO (Malladi et al., 2023) appears in Appendices O.1, O.3, and B, respectively.

7 Conclusion

We presented a unification of many zeroth-order optimizers under the likelihood ratio gradient lens. Starting from an algorithm's sampling distribution q_θ and update weights ψ, we reconstruct implicit smoothing measures such that each expected update component is the corresponding partial derivative of its induced smoothed objective. When these reconstructions agree, the full update is the gradient of a common objective. The reconstruction unifies classical likelihood ratio estimators with finite-difference estimators, clarifies how reward shaping changes the effective objective, and enables unified theoretical analyses.

AI USE STATEMENT

Generative AI was used to assist with drafting and editing the manuscript, algebraic calculations, the literature review, and presentation, including LaTeX and bibliography formatting, compilation troubleshooting, and layout checks. The key research ideas were conceptualized by the authors.

REFERENCES

Ahmad Ajalloeian and Sebastian U. Stich. On the convergence of SGD with biased gradients. *arXiv preprint arXiv:2008.00051*, 2020. doi: 10.48550/arXiv.2008.00051. URL https://arxiv.org/abs/2008.00051.

Shun-ichi Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10(2):251–276, 1998. doi: 10.1162/089976698300017746. URL https://direct.mit.edu/neco/article/10/2/251/6143/Natural-Gradient-Works-Efficiently-in-Learning.

Charles Audet and J. E. Dennis. Mesh adaptive direct search algorithms for constrained optimization. *SIAM Journal on Optimization*, 17(1):188–217, 2006. doi: 10.1137/040603371. URL https://doi.org/10.1137/040603371.

Albert S. Berahas, Liyuan Cao, Krzysztof Choromanski, and Katya Scheinberg. A theoretical and empirical comparison of gradient approximations in derivative-free optimization. *Foundations of Computational Mathematics*, 22(2):507–560, 2022. doi: 10.1007/s10208-021-09513-z.

Hans-Georg Beyer and Hans-Paul Schwefel. Evolution strategies: A comprehensive introduction. *Natural Computing*, 1(1):3–52, 2002. doi: 10.1023/A:1015059928466. URL https://doi.org/10.1023/A:1015059928466.

Han-Fu Chen, Tyrone E. Duncan, and Bozenna Pasik-Duncan. A Kiefer–Wolfowitz algorithm with randomized differences. *IEEE Transactions on Automatic Control*, 44(3):442–453, 1999. doi: 10.1109/9.751340. URL https://doi.org/10.1109/9.751340.

Andrew R. Conn, Katya Scheinberg, and Luis N. Vicente. *Introduction to Derivative-Free Optimization*. MPS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, USA, 2009. doi: 10.1137/1.9780898718768. URL https://epubs.siam.org/doi/book/10.1137/1.9780898718768.

Pieter-Tjerk de Boer, Dirk P. Kroese, Shie Mannor, and Reuven Y. Rubinstein. A tutorial on the cross-entropy method. *Annals of Operations Research*, 134(1):19–67, 2005. doi: 10.1007/s10479-005-5724-z. URL https://people.smp.uq.edu.au/DirkKroese/ps/aortut.pdf.

Michel C. Delfour and Jean-Paul Zolesio. *Shapes and Geometries: Metrics, Analysis, Differential Calculus, and Optimization*, volume 22 of *Advances in Design and Control*. Society for Industrial and Applied Mathematics (SIAM), 2 edition, 2011. ISBN 9780898719369. doi: 10.1137/1.9780898719826. URL https://doi.org/10.1137/1.9780898719826.

John C. Duchi, Michael I. Jordan, Martin J. Wainwright, and Andre Wibisono. Optimal rates for zero-order convex optimization: The power of two function evaluations. *IEEE Transactions on Information Theory*, 61(5):2788–2806, 2015. doi: 10.1109/TIT.2015.2409256. URL https://www.cs.yale.edu/homes/wibisono/DucJorWaiWib15.pdf.

Abraham D. Flaxman, Adam Tauman Kalai, and H. Brendan McMahan. Online convex optimization in the bandit setting: Gradient descent without a gradient. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 385–394, 2005. URL https://www.microsoft.com/en-us/research/wp-content/uploads/2016/11/2005-Online_Convex_Optimization_in_the_Bandit_Setting.pdf.

Gerald B. Folland. *Real Analysis: Modern Techniques and Their Applications*. John Wiley & Sons, 2 edition, 1999.

Michael C. Fu. Gradient estimation. In Shane G. Henderson and Barry L. Nelson (eds.), *Simulation*, volume 13 of *Handbooks in Operations Research and Management Science*, chapter 19, pp. 575–616. Elsevier, 2006. doi: 10.1016/S0927-0507(06)13019-4. URL https://doi.org/10.1016/S0927-0507(06)13019-4.

Tanmay Gautam, Youngsuk Park, Hao Zhou, Parameswaran Raman, and Wooseok Ha. Variance-reduced zeroth-order methods for fine-tuning language models. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

Saeed Ghadimi and Guanghui Lan. Stochastic first- and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, 23(4):2341–2368, 2013. doi: 10.1137/120880811. URL https://optimization-online.org/wp-content/uploads/2012/06/3499.pdf.

Tobias Glasmachers, Tom Schaul, Sun Yi, Daan Wierstra, and J"urgen Schmidhuber. Exponential natural evolution strategies. In *Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation (GECCO)*, pp. 393–400, 2010. doi: 10.1145/1830483.1830557. URL https://people.idsia.ch/~juergen/xNES2010gecco.pdf.

Peter W. Glynn. Likelihood ratio gradient estimation for stochastic systems. *Communications of the ACM*, 33(10):75–84, 1990. doi: 10.1145/84537.84552. URL https://dl.acm.org/doi/10.1145/84537.84552.

Pierre Grinfeld. Hadamard's formula inside and out. *Journal of Optimization Theory and Applications*, 146(3):654–690, 2010. doi: 10.1007/s10957-010-9681-6. URL https://doi.org/10.1007/s10957-010-9681-6.

Nikolaus Hansen. The CMA evolution strategy: A tutorial. *arXiv preprint arXiv:1604.00772*, 2016. URL https://arxiv.org/abs/1604.00772.

Nikolaus Hansen and Andreas Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001. doi: 10.1162/106365601750190398. URL https://doi.org/10.1162/106365601750190398.

J. Kiefer and J. Wolfowitz. Stochastic estimation of the maximum of a regression function. *The Annals of Mathematical Statistics*, 23(3):462–466, 1952. doi: 10.1214/aoms/1177729392. URL https://doi.org/10.1214/aoms/1177729392.

Tamara G. Kolda, Robert Michael Lewis, and Virginia Torczon. Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Review*, 45(3):385–482, 2003. doi: 10.1137/S003614450242889. URL https://doi.org/10.1137/S003614450242889.

Jeffrey Larson, Matt Menickelly, and Stefan M. Wild. Derivative-free optimization methods. *Acta Numerica*, 28:287–404, 2019. doi: 10.1017/S0962492919000060. URL https://www.math.ucdavis.edu/~saito/data/ZerothOrderOptim/larson-etal_derivative-free-optimization-methods.pdf.

Pierre L'Ecuyer. A unified view of the IPA, SF, and LR gradient estimation techniques. *Management Science*, 36(11):1364–1383, 1990. doi: 10.1287/mnsc.36.11.1364. URL https://doi.org/10.1287/mnsc.36.11.1364.

Pierre L'Ecuyer. An overview of derivative estimation. In *Proceedings of the 1991 Winter Simulation Conference*, pp. 207–217, 1991. doi: 10.1109/WSC.1991.185617. URL https://doi.org/10.1109/WSC.1991.185617.

Zhenhua Li, Xi Lin, Qingfu Zhang, and Hailin Liu. Evolution strategies for continuous optimization: A survey of the state-of-the-art. *Swarm and Evolutionary Computation*, 56:100694, 2020. doi: 10.1016/j.swevo.2020.100694. URL https://doi.org/10.1016/j.swevo.2020.100694.

Sijia Liu, Pin-Yu Chen, Bhavya Kailkhura, Gaoyuan Zhang, Alfred Hero, and Pramod K. Varshney. A primer on zeroth-order optimization in signal processing and machine learning: Principals, recent advances and applications. *arXiv preprint arXiv:2006.06224*, 2020. URL https://arxiv.org/abs/2006.06224.

Sadhika Malladi, Tianyu Gao, Eshaan Nichani, Alex Damian, Jason D Lee, Danqi Chen, and Sanjeev Arora. Fine-tuning language models with just forward passes. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 53038–53075. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/a627810151be4d13f907ac898ff7e948-Paper-Conference.pdf.

Shakir Mohamed, Mihaela Rosca, Michael Figurnov, and Andriy Mnih. Monte carlo gradient estimation in machine learning. *Journal of Machine Learning Research*, 21(132):1–62, 2020. URL https://www.jmlr.org/papers/v21/19-346.html.

Yurii Nesterov and Vladimir Spokoiny. Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics*, 17(2):527–566, 2017. doi: 10.1007/s10208-015-9296-2. URL https://doi.org/10.1007/s10208-015-9296-2.

Yann Ollivier, Ludovic Arnold, Anne Auger, and Nikolaus Hansen. Information-geometric optimization algorithms: A unifying picture via invariance principles. *Journal of Machine Learning Research*, 18(18):1–65, 2017. URL https://jmlr.org/papers/v18/14-467.html.

Paavo Parmas and Masashi Sugiyama. A unified view of likelihood ratio and reparameterization gradients. In Arindam Banerjee and Kenji Fukumizu (eds.), *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 4078–4086. PMLR, 2021. URL https://proceedings.mlr.press/v130/parmas21a.html.

L. A. Prashanth, Shalabh Bhatnagar, Michael C. Fu, and Steven I. Marcus. Adaptive system optimization using random directions stochastic approximation. *IEEE Transactions on Automatic Control*, 62(5):2223–2238, 2017. doi: 10.1109/TAC.2016.2600643. URL https://doi.org/10.1109/TAC.2016.2600643.

Ingo Rechenberg. Evolutionsstrategie: Optimierung technischer systeme nach prinzipien der biologischen evolution, 1973. URL https://gwern.net/doc/reinforcement-learning/exploration/1973-rechenberg.pdf. Frommann-Holzboog, Stuttgart.

Luis Miguel Rios and Nikolaos V. Sahinidis. Derivative-free optimization: A review of algorithms and comparison of software implementations. *Journal of Global Optimization*, 56:1247–1293, 2013. doi: 10.1007/s10898-012-9951-y. URL https://doi.org/10.1007/s10898-012-9951-y.

Reuven Y. Rubinstein. The cross-entropy method for combinatorial and continuous optimization. *Methodology and Computing in Applied Probability*, 1(2):127–190, 1999. doi: 10.1023/A:1010091220143. URL https://link.springer.com/article/10.1023/A:1010091220143.

Tim Salimans, Jonathan Ho, Xi Chen, Szymon Sidor, and Ilya Sutskever. Evolution strategies as a scalable alternative to reinforcement learning. *arXiv preprint arXiv:1703.03864*, 2017. URL https://arxiv.org/abs/1703.03864.

Hans-Paul Schwefel. *Evolution and Optimum Seeking*. John Wiley & Sons, New York, NY, USA, 1995. ISBN 9780471571483. URL https://www.researchgate.net/profile/Hans-Paul-Schwefel/publication/220690578_Evolution_and_Optimum_Seeking/links/544bd6c70cf24b5d6c409084/Evolution-and-Optimum-Seeking.pdf.

Ohad Shamir. On the complexity of bandit and derivative-free stochastic convex optimization. In Shai Shalev-Shwartz and Ingo Steinwart (eds.), *Proceedings of the 26th Annual Conference on Learning Theory*, volume 30 of *Proceedings of Machine Learning Research*, pp. 3–24. PMLR, 2013. URL https://proceedings.mlr.press/v30/Shamir13.html.

Ohad Shamir. An optimal algorithm for bandit and zero-order convex optimization with two-point feedback. *Journal of Machine Learning Research*, 18(52):1–11, 2017. URL https://jmlr.csail.mit.edu/papers/volume18/16-632/16-632.pdf.

Jan Sokolowski and Jean-Paul Zolesio. *Introduction to Shape Optimization: Shape Sensitivity Analysis*, volume 16 of *Springer Series in Computational Mathematics*. Springer, 1992. doi: 10.1007/978-3-642-58106-9. URL https://doi.org/10.1007/978-3-642-58106-9.

James C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control*, 37(3):332–341, 1992. doi: 10.1109/9.119632. URL https://ieeexplore.ieee.org/document/119632.

James C. Spall. *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. John Wiley & Sons, Hoboken, NJ, USA, 2003. ISBN 9780471330523. doi: 10.1002/0471722138. URL https://doi.org/10.1002/0471722138.

Michael Spivak. *Calculus on Manifolds: A Modern Approach to Classical Theorems of Advanced Calculus*. W. A. Benjamin, 1965.

Charles Stein. Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, 9(6):1135–1151, 1981. doi: 10.1214/aos/1176345632. URL https://doi.org/10.1214/aos/1176345632.

Daan Wierstra, Tom Schaul, Tobias Glasmachers, Yi Sun, Jan Peters, and Jürgen Schmidhuber. Natural evolution strategies. *Journal of Machine Learning Research*, 15(27):949–980, 2014. URL https://jmlr.org/papers/v15/wierstra14a.html.

Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3):229–256, 1992. doi: 10.1007/BF00992696. URL https://link.springer.com/article/10.1007/BF00992696.

Yihua Zhang, Pingzhi Li, Junyuan Hong, Jiaxiang Li, Yimeng Zhang, Wenqing Zheng, Pin-Yu Chen, Jason D. Lee, Wotao Yin, Mingyi Hong, Zhangyang Wang, Sijia Liu, and Tianlong Chen. Revisiting zeroth-order optimization for memory-efficient LLM fine-tuning: A benchmark. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

Appendix

A Reconstruction as a practical recipe

We summarize the reconstruction viewpoint as a concrete workflow.

(1) Identify its sampler q_θ and its weight function $\psi_\theta(\cdot)$ (and the reward r used in the update).

(2) If $\mathbb{E}_{q_\theta}[\psi_j] \neq 0$, apply centering (either ψ-centering or a mean baseline for r).

(3) Integrate the reconstruction equation $\partial_{\theta_j} p_\theta = q_\theta \psi^{\mathrm{ctr}}_{\theta_j}$ (with appropriate boundary/normalization conditions) to obtain the implied smoothing density p_θ.

(4) Read off the implied objective $J(\theta) = \mathbb{E}_{x\sim p_\theta}[r(x)]$.

B Extended related work

This section briefly situates the paper with respect to several closely related literatures. The main text focuses on an "update-first" reconstruction view; many of the references below provide complementary, "target-first" perspectives or cover broader algorithmic contexts.

B.1 Derivative-free optimization and stochastic approximation

Derivative-free optimization (DFO) studies algorithms that optimize a function using only function evaluations. Modern treatments cover both direct-search methods and model-based trust-region schemes, as well as complexity results and practical implementations (Conn et al., 2009; Larson et al., 2019; Rios & Sahinidis, 2013; Kolda et al., 2003; Audet & Dennis, 2006). In parallel, stochastic approximation originates with Kiefer–Wolfowitz-type finite-difference schemes for optimizing an unknown regression function (Kiefer & Wolfowitz, 1952) and includes randomized-difference and SPSA-style constructions (Chen et al., 1999; Spall, 1992; 2003). These traditions emphasize bias/variance tradeoffs (e.g., step-size selection) and robustness to noise, which are orthogonal to our focus on identifying the *implicit* smoothed objective induced by an implemented update.

B.2 Gradient estimation identities in simulation and Monte Carlo

A long line of work in simulation optimization studies general gradient estimators based on infinitesimal perturbation analysis (IPA), score-function / likelihood-ratio (LR) methods, and related identities (L'Ecuyer, 1990; 1991; Fu, 2006; Glynn, 1990). In machine learning, broad overviews of Monte Carlo gradient estimators and variance-reduction techniques are given by Mohamed et al. (2020). The reconstruction equation in this paper addresses an inverse question relative to these works: instead of starting from a *target* expectation and deriving an estimator, we start from the estimator/update and identify (when possible) a compatible expectation.

B.3 Bandit and zeroth-order convex optimization

In online/bandit convex optimization and related zeroth-order settings, smoothing by averaging over a ball (or other sets) is a standard device that turns function evaluations into unbiased gradient estimates of the smoothed objective. Canonical results and estimators include the one-point and two-point sphere estimators and their optimization rates (Flaxman et al., 2005; Shamir, 2013; Duchi et al., 2015; Nesterov & Spokoiny, 2017; Shamir, 2017; Ghadimi & Lan, 2013). Our reconstructions (Table 1) recover these ball-smoothed objectives directly from the sampling-and-weighting rule, highlighting the sampler/smoother distinction (q versus p).

B.4 Evolution strategies, information geometry, and natural gradients

Evolution strategies (ES) originate in classical work on stochastic search and step-size adaptation (Rechenberg, 1973; Schwefel, 1995; Beyer & Schwefel, 2002) and have many modern variants and analyses (Hansen & Ostermeier, 2001; Hansen, 2016; Wierstra et al., 2014; Ollivier et al., 2017; Li

et al., 2020; Salimans et al., 2017). A recurring theme is invariance and geometry: NES/IGO-style methods interpret score-function updates as gradients in the space of distributions, often paired with natural-gradient (Fisher) preconditioning (Amari, 1998; Ollivier et al., 2017; Wierstra et al., 2014). Section O.3 connects this preconditioning view to our reconstruction calculus.

B.5 Weak derivatives, boundary terms, and shape calculus

Finally, the "boundary/endpoint" estimators emphasized in Appendix H.2 are instances of a general phenomenon: differentiating an integral over a moving domain produces boundary terms. This is central in shape calculus and geometric sensitivity analysis (Grinfeld, 2010; Delfour & Zolesio, 2011; Sokolowski & Zolesio, 1992). In our context, weak derivatives provide a convenient language for viewing finite differences and sphere estimators as exact gradients of non-smooth (indicator-kernel) smoothing measures.

B.6 Zeroth-order optimization for Large Language Model Training

Recent advances in large language model (LLM) training and fine-tuning have renewed interest in memory-efficient zeroth-order optimization methods (Malladi et al., 2023; Gautam et al., 2024; Zhang et al., 2024). These methods aim to replace backpropagation with gradient-free updates, enabling optimization under strict memory constraints where storing activations or full gradients is prohibitive. A representative approach is MeZO (Malladi et al., 2023), which approximates directional derivatives using two-point finite differences along randomly sampled directions while reusing random seeds to reconstruct perturbations on-the-fly. This design eliminates the need to store perturbation vectors or intermediate activations, substantially reducing memory usage compared to first-order methods. Our framework also provides a unified view that clarifies their intrinsic objectives and their relationship to classical zeroth-order estimators.

C Centering and the mass-drift decomposition

This appendix collects derivations underlying the centering/mass-drift discussion in Section 3.4.

The normalization condition for reconstruction,

$$\int p_\theta(x)\ \mathrm{d}x = 1 \quad \forall\theta, \tag{27}$$

implies the necessary constraint

$$0 = \frac{\partial}{\partial\theta_j}\int p_\theta(x)\ \mathrm{d}x = \int \frac{\partial}{\partial\theta_j} p_\theta(x)\ \mathrm{d}x. \tag{28}$$

Under the reconstruction equation $\partial_{\theta_j} p_\theta(x) = q_\theta(x)\psi_{\theta_j}(x)$, this becomes

$$\mathbb{E}_{x\sim q_\theta}\big[\psi_{\theta_j}(x)\big] = 0. \tag{29}$$

Score-function weights satisfy (29) automatically, but general algorithmic weights need not. When (29) fails, it is helpful to make explicit what part of the update is incompatible with a *normalized* smoothing density.

The drift term. Consider the expected update direction for parameter component θ_j:

$$g_j(\theta) := \mathbb{E}_{x\sim q_\theta}\big[r(x)\,\psi_{\theta_j}(x)\big]. \tag{30}$$

Add and subtract the mean of ψ_{θ_j} under q_θ:

$$g_j(\theta) = \mathbb{E}_{q_\theta}\Big[r(x)\big(\psi_{\theta_j}(x) - \mathbb{E}_{q_\theta}[\psi_{\theta_j}(x)]\big)\Big] \;+\; \mathbb{E}_{q_\theta}[r(x)]\,\mathbb{E}_{q_\theta}[\psi_{\theta_j}(x)]. \tag{31}$$

The second term,

$$d_j(\theta) := \mathbb{E}_{q_\theta}[r(x)]\,\mathbb{E}_{q_\theta}[\psi_{\theta_j}(x)], \tag{32}$$

is a deterministic *drift* term (given θ): it does not depend on any particular sample realization beyond these means. Importantly, $d_j(\theta)$ is exactly the part that prevents $\partial_{\theta_j} p_\theta = q_\theta\psi_j$ from holding with a normalized density p_θ, because it corresponds to a nonzero derivative of total mass.

The first term in (31) has a mean-zero weight and is therefore compatible with mass conservation. It is precisely a covariance:

$$\mathbb{E}_{q_\theta}\Big[r(x)\big(\psi_{\theta_j}(x) - \mathbb{E}_{q_\theta}[\psi_{\theta_j}(x)]\big)\Big] = \mathrm{Cov}_{q_\theta}\big(r(x), \psi_{\theta_j}(x)\big). \tag{33}$$

Interpretation without centering: "gradient of an unnormalized integral" + drift. Even if (29) fails, one can still integrate

$$\frac{\partial}{\partial\theta_j}\widetilde{p}_\theta(x) = q_\theta(x)\,\psi_{\theta_j}(x) \tag{34}$$

to obtain an *unnormalized* kernel $\widetilde{p}_\theta$ whose total mass $\widetilde{Z}(\theta) := \int \widetilde{p}_\theta(x)\,dx$ is not constant. Defining the corresponding unnormalized objective

$$\widetilde{J}(\theta) := \int r(x)\,\widetilde{p}_\theta(x)\;\mathrm{d}x, \tag{35}$$

we obtain

$$\frac{\partial}{\partial\theta_j}\widetilde{J}(\theta) = \mathbb{E}_{q_\theta}\big[r(x)\psi_{\theta_j}(x)\big] = g_j(\theta). \tag{36}$$

Thus, *without centering* the update can still be seen as a true gradient, but of an integral whose normalizing mass drifts with θ.

If we additionally normalize $\widetilde{p}_\theta$ into a density $\overline{p}_\theta := \widetilde{p}_\theta/\widetilde{Z}(\theta)$ and define the normalized expectation $J(\theta) := \mathbb{E}_{x\sim\overline{p}_\theta}[r(x)]$, then $\widetilde{J}(\theta) = \widetilde{Z}(\theta)\,J(\theta)$ implies the decomposition

$$g_j(\theta) = \frac{\partial}{\partial\theta_j}\widetilde{J}(\theta) = \widetilde{Z}(\theta)\,\frac{\partial}{\partial\theta_j}J(\theta) \;+\; J(\theta)\,\frac{\partial}{\partial\theta_j}\widetilde{Z}(\theta). \tag{37}$$

The second term is exactly a "mass drift" contribution. Centering (below) enforces $\partial_{\theta_j}\widetilde{Z}(\theta) = 0$, eliminating this drift and recovering a gradient of a proper expectation.

Fix A (center the weight). Define the centered weight

$$\psi^{\mathrm{ctr}}_{\theta_j}(x) := \psi_{\theta_j}(x) - \mathbb{E}_{x\sim q_\theta}\big[\psi_{\theta_j}(x)\big], \tag{38}$$

so that $\mathbb{E}_{q_\theta}[\psi^{\mathrm{ctr}}_{\theta_j}] = 0$ holds automatically. Then the centered reconstruction equation

$$\frac{\partial}{\partial\theta_j}p_\theta(x) = q_\theta(x)\,\psi^{\mathrm{ctr}}_{\theta_j}(x) \tag{39}$$

is compatible with mass conservation. A valid smoothing density additionally requires nonnegativity and unit mass; a joint reconstruction requires compatibility (Appendix E). When these conditions hold, the corresponding objective $J(\theta) = \mathbb{E}_{p_\theta}[r(x)]$ satisfies

$$\begin{aligned}\frac{\partial}{\partial\theta_j}J(\theta) &= \mathbb{E}_{q_\theta}\big[r(x)\psi^{\mathrm{ctr}}_{\theta_j}(x)\big] \\ &= \mathrm{Cov}_{q_\theta}\big(r(x), \psi_{\theta_j}(x)\big).\end{aligned} \tag{40}$$

Fix B (subtract a mean baseline from the reward). Alternatively, keep ψ_j unchanged and define a mean baseline

$$b(\theta) := \mathbb{E}_{x\sim q_\theta}[r(x)], \qquad \widetilde{r}(x;\theta) := r(x) - b(\theta). \tag{41}$$

Then

$$\mathbb{E}_{q_\theta}\big[\widetilde{r}(x;\theta)\,\psi_{\theta_j}(x)\big] = \mathbb{E}_{q_\theta}\big[r(x)\psi_{\theta_j}(x)\big] - \mathbb{E}_{q_\theta}[r(x)]\,\mathbb{E}_{q_\theta}[\psi_{\theta_j}(x)], \tag{42}$$

which subtracts exactly the drift term (32). Thus, centering the reward or centering the weight yields the same expected direction:

$$\mathbb{E}_{q_\theta}\big[(r(x) - \mathbb{E}_{q_\theta}[r(x)])\,\psi_{\theta_j}(x)\big] = \mathbb{E}_{q_\theta}\big[r(x)\,(\psi_{\theta_j}(x) - \mathbb{E}_{q_\theta}[\psi_{\theta_j}(x)])\big]. \tag{43}$$

Practical note (finite batches). In implementations, expectations are replaced by batch averages from samples $x_{1:N} \sim q_\theta$. Let $r_i := r(x_i)$ and $\psi_i := \psi_{\theta_j}(x_i)$ for the parameter component θ_j, and define

$$\bar{r} := \frac{1}{N}\sum_{i=1}^{N} r_i, \qquad \bar{\psi} := \frac{1}{N}\sum_{i=1}^{N}\psi_i.$$

Then the two centering approaches are not only equivalent in expectation, but *algebraically identical* when the same batch means are used:

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}(r_i-\bar{r})\,\psi_i &= \frac{1}{N}\sum_{i=1}^{N} r_i\,(\psi_i-\bar{\psi}) \\ &= \frac{1}{N}\sum_{i=1}^{N} r_i\psi_i \;-\; \bar{r}\,\bar{\psi}.\end{aligned} \tag{44}$$

Thus "subtracting a mean baseline from r" and "centering ψ" produce the same estimator (for a given batch), just written in two different ways.

For i.i.d. samples, this common estimator is proportional to the sample covariance, and its expectation differs from the population covariance by the familiar factor $(N-1)/N$. If exact unbiasedness of the covariance estimate is important (without tracking correction factors), a leave-one-out baseline is a simple remedy; see Appendix D.

D LEAVE-ONE-OUT BASELINES FOR UNBIASED COVARIANCE ESTIMATES

This appendix complements Section 3.4. In finite batches, the usual "baseline"/"centering" trick is (up to scaling) a sample covariance; with the common $1/N$ normalization, it therefore carries the familiar $(N-1)/N$ factor in expectation. Here we record a simple leave-one-out (LOO) baseline that removes this factor and is convenient to reference from the main text.

Let $(r_i,\psi_i)_{i=1}^{N}$ be i.i.d. draws of a reward $r(x)$ and a scalar weight component $\psi_{\theta_j}(x)$ under $x\sim q_\theta$ (the statements below apply component-wise when ψ is vector-valued). Define the batch means

$$\bar{r} := \frac{1}{N}\sum_{i=1}^{N} r_i, \qquad \bar{\psi} := \frac{1}{N}\sum_{i=1}^{N}\psi_i.$$

Centered estimator (baseline/centering) and its bias. Using Eq. (44), the standard centered estimator admits the decomposition

$$\begin{aligned}\widehat{g}_{\mathrm{ctr}} := \frac{1}{N}\sum_{i=1}^{N}(r_i-\bar{r})\,\psi_i &= \frac{1}{N}\sum_{i=1}^{N} r_i\,(\psi_i-\bar{\psi}) \\ &= \underbrace{\frac{1}{N}\sum_{i=1}^{N} r_i\psi_i}_{\text{raw term}} - \underbrace{\bar{r}\,\bar{\psi}}_{\text{(batch) baseline term}}.\end{aligned} \tag{45}$$

A direct expectation calculation yields

$$\mathbb{E}[\widehat{g}_{\mathrm{ctr}}] = \frac{N-1}{N}\,\underbrace{\mathrm{Cov}\big(r(x),\psi_{\theta_j}(x)\big)}_{\text{population covariance target}}. \tag{46}$$

Leave-one-out (LOO) baseline and unbiasedness. Define the leave-one-out mean

$$\bar{r}_{-i} := \frac{1}{N-1}\sum_{k\neq i} r_k,$$

and the estimator

$$\widehat{g}_{\mathrm{LOO}} := \frac{1}{N}\sum_{i=1}^{N}\big(r_i - \underbrace{\bar{r}_{-i}}_{\text{independent of }(r_i,\,\psi_i)}\big)\,\psi_i. \tag{47}$$

Since $\bar{r}_{-i}$ is independent of (r_i,ψ_i), we have

$$\mathbb{E}\big[(r_i-\bar{r}_{-i})\psi_i\big] = \mathbb{E}[r(x)\psi_{\theta_j}(x)] - \mathbb{E}[r(x)]\,\mathbb{E}[\psi_{\theta_j}(x)] = \mathrm{Cov}\big(r(x),\psi_{\theta_j}(x)\big),$$

and therefore

$$\mathbb{E}[\widehat{g}_{\mathrm{LOO}}] = \mathrm{Cov}\big(r(x), \psi_{\theta_j}(x)\big). \tag{48}$$

When $\mathbb{E}[\psi_{\theta_j}(x)] = 0$ (after weight-centering), this reduces to $\mathbb{E}[\widehat{g}_{\mathrm{LOO}}] = \mathbb{E}[r(x)\psi_{\theta_j}(x)]$.

Useful identity (implementation shortcut). Using $\bar{r}_{-i} = (N\bar{r} - r_i)/(N-1)$, we obtain

$$r_i - \bar{r}_{-i} = \frac{N}{N-1}\,(r_i - \bar{r}), \qquad \Rightarrow \qquad \widehat{g}_{\mathrm{LOO}} = \frac{N}{N-1}\,\widehat{g}_{\mathrm{ctr}}. \tag{49}$$

Thus LOO can be implemented either explicitly via $\bar{r}_{-i}$ or equivalently by rescaling the centered estimator; (49) removes the $(N-1)/N$ attenuation in (46) without changing the algebraic structure of the update.

E Explicit integration of location–scale families for higher dimensions

The one-dimensional formula (21) already contains the main idea. In several dimensions, we hold the other coordinates fixed and apply the same integral. This produces one full candidate over x from each location coordinate. We first derive these candidates and compare them directly; afterwards, we give the familiar curl-free test as a convenient smooth equivalent.

Reconstructing a full candidate from one coordinate. Fix a coordinate j, write $x = (x_j, x_{-j})$, and define

$$F_j(x \mid \mu, \sigma) := q(x \mid \mu, \sigma)\, \psi_{\mu_j}(x). \tag{50}$$

The joint transport equation (20) gives

$$\frac{\partial}{\partial x_j} p(x \mid \mu, \sigma) = -\,F_j(x \mid \mu, \sigma). \tag{51}$$

For fixed x_{-j}, this is the same one-dimensional equation as in (21). Using the boundary convention $p((t, x_{-j}) \mid \mu, \sigma) \to 0$ as $t \to +\infty$, its solution is

$$p^{(j)}(x \mid \mu, \sigma) = \int_{x_j}^{\infty} F_j((t, x_{-j}) \mid \mu, \sigma)\ \mathrm{d}t. \tag{52}$$

Although only t is integrated out, the right-hand side still depends on x_{-j}. Thus $p^{(j)}$ is a full function on $\mathbb{R}^d$, not a marginal. It is unique under the chosen boundary convention: the difference of two solutions has zero x_j-derivative and tends to zero at the same boundary, so the difference is zero.

Compatibility by comparing the candidates. Repeating (52) for every coordinate gives uniquely determined full candidates $p^{(1)}, \ldots, p^{(d)}$. The direct compatibility criterion is simply that these candidates agree.

Proposition E.1 (Compatibility of the location reconstructions)**.** *Assume that every candidate (52) is well defined under the same boundary convention. A common function p satisfying all location-coordinate reconstruction equations and that boundary convention exists if and only if*

$$p^{(1)}(x \mid \mu, \sigma) = \cdots = p^{(d)}(x \mid \mu, \sigma) \qquad \textit{for almost every } x. \tag{53}$$

If, additionally, $F = (F_1, \ldots, F_d)$ is continuously differentiable on an open simply connected domain and the common boundary condition is satisfied, then this is equivalent to the symmetric-Jacobian condition

$$\frac{\partial F_j}{\partial x_k} = \frac{\partial F_k}{\partial x_j} \qquad \textit{for all } j, k. \tag{54}$$

Proof. If (53) holds, the common value satisfies every coordinate equation. Conversely, any common solution must equal each $p^{(j)}$ by the uniqueness argument above.

For the smooth statement, if $\nabla_x p = -F$, equality of mixed derivatives implies (54). Conversely, the standard conservative-vector-field theorem states that a continuously differentiable field with symmetric Jacobian on a simply connected domain has a scalar potential, unique up to an additive constant (Spivak, 1965). Applying this to $-F$ gives a common solution, and the boundary convention fixes the additive constant. □

The smooth sufficiency can also be checked directly from the explicit candidate. Assume differentiation can be passed under the integral in (52) and $F_k((t, x_{-j}) \mid \mu, \sigma) \to 0$ as $t \to +\infty$. For $k \neq j$,

$$\begin{aligned}\frac{\partial}{\partial x_k} p^{(j)}(x \mid \mu, \sigma) &= \int_{x_j}^{\infty} \frac{\partial F_j}{\partial x_k}((t, x_{-j}) \mid \mu, \sigma)\ \mathrm{d}t \\ &= \int_{x_j}^{\infty} \frac{\partial F_k}{\partial t}((t, x_{-j}) \mid \mu, \sigma)\ \mathrm{d}t \\ &= -F_k(x \mid \mu, \sigma). \end{aligned} \tag{55}$$

The first equality differentiates beneath the integral, the second uses (54), and the last uses the boundary condition and the fundamental theorem of calculus. Together with $\partial_{x_j} p^{(j)} = -F_j$, this proves that $p^{(j)}$ satisfies every location equation.

When is the integral well defined? The basic reconstruction (52) uses only an ordinary one-dimensional integral. For almost every fixed x_{-j}, it is enough that $t \mapsto F_j((t, x_{-j}) \mid \mu, \sigma)$ is measurable, locally integrable, and has a convergent tail integral from x_j to $+\infty$. Absolute tail integrability is a simple sufficient condition. Under these assumptions, $p^{(j)}$ is absolutely continuous in x_j and satisfies (51) almost everywhere (Folland, 1999, Chapter 3). No continuity or classical smoothness is required for this basic construction. Additional regularity is needed only for the symmetric-Jacobian test, differentiation with respect to another coordinate, or interchange of differentiation and integration. Boundary-supported cases can instead be interpreted weakly, as in Appendix M. Finally, a candidate is a probability density only after checking nonnegativity and unit total mass.

General parameter compatibility and a compatible scale update. For a general parameter vector, let $G_j(\theta, x) := q_\theta(x)\psi_{\theta_j}(x)$. If one common family p_θ satisfies $\partial_{\theta_j} p_\theta = G_j$ for every j, then commuting mixed parameter derivatives requires

$$\partial_{\theta_k} G_j(\theta, x) = \partial_{\theta_j} G_k(\theta, x) \qquad \text{for all } j, k. \tag{56}$$

Under the same continuously differentiable and simply connected-domain assumptions, this condition is also sufficient after one common reference measure is fixed, by the same conservative-field argument.

Reconstruction can also construct a compatible update. Suppose a location rule and its boundary convention have reconstructed a unique $p(x \mid \mu, \sigma)$. A scale update differentiates the same objective exactly when

$$q(x \mid \mu, \sigma)\, \psi_\sigma(x) = \partial_\sigma p(x \mid \mu, \sigma). \tag{57}$$

Therefore, wherever $q(x \mid \mu, \sigma) > 0$, the compatible scale weight is

$$\psi_\sigma^{\mathrm{comp}}(x) = \frac{\partial_\sigma p(x \mid \mu, \sigma)}{q(x \mid \mu, \sigma)}. \tag{58}$$

This weight is unique on the support of q: if two weights satisfy (57), their difference multiplied by q is zero. If $\partial_\sigma p$ is nonzero where $q = 0$, then no pointwise weighting rule under that sampler can represent the same scale derivative. Thus reconstruction gives the unique compatible scale update whenever such an update exists under q.

A common objective without a common measure. A common smoothing measure is stronger than necessary for the expected update on one fixed objective f to be a gradient field. Define

$$g_j(\theta) := \int f(x) G_j(\theta, x)\ \mathrm{d}x. \tag{59}$$

If g is continuously differentiable on a simply connected parameter domain, it is the gradient of a scalar objective if and only if its Jacobian is symmetric:

$$\partial_{\theta_k} g_j(\theta) = \partial_{\theta_j} g_k(\theta) \qquad \text{for all } j, k. \tag{60}$$

This condition can hold even when (56) fails, because integration against the particular f can cancel the incompatible part. For fixed-scale location reconstructions $p_\mu^{(j)}(x) = \kappa_j(x - \mu)$ with

$U^{(j)} \sim \kappa_j$, differentiation under the expectation gives $g_j(\mu) = \mathbb{E}[\partial_j f(\mu + U^{(j)})]$. Thus (60) becomes $\mathbb{E}[\partial_{jk} f(\mu + U^{(j)})] = \mathbb{E}[\partial_{jk} f(\mu + U^{(k)})]$. It holds automatically for quadratic objectives, whose Hessian is constant, and for coordinate-separable objectives, whose mixed derivatives vanish.

Examples. For Gaussian score weights, $q\psi_{\theta_j} = \partial_{\theta_j} q$ for both location and scale, so all components reconstruct the same canonical measure $p = q$ (Appendix I). The sphere estimator likewise reconstructs the same uniform-ball measure for every location coordinate in the weak sense (Appendix M). The non-score finite-difference, uniform-ES, and random-denominator examples in Section 5 are presented for one optimized parameter. Arbitrary scale weights are not expected to be compatible with a location reconstruction, because (58) gives exactly one compatible choice. For example, choosing $\psi_\sigma = 0$ is incompatible whenever the p reconstructed from the location update varies with σ, since the scale equation would require $\partial_\sigma p = 0$.

F Reconstruction-based gradient-bias and equilibrium bounds

This section gives one concrete use of the reconstructed kernels. Standard biased-gradient analyses begin with a bound on the difference between the population update and the true gradient. Ajalloeian & Stich (2020, Definition 1 and Assumption 4) write

$$\widehat{g}_\mu(\mu, \xi) = \nabla f(\mu) + b(\mu) + n(\mu, \xi), \qquad \mathbb{E}_\xi[n(\mu, \xi)] = 0, \tag{61}$$

and assume

$$\|b(\mu)\|^2 \le m\|\nabla f(\mu)\|^2 + \zeta^2, \qquad 0 \le m < 1. \tag{62}$$

Here m controls relative bias, while ζ is an additive bias that creates an irreducible accuracy floor. Their Lemma 3 gives a nonconvex stationarity bound, and their Theorem 6 gives a function-value floor under the Polyak–Łojasiewicz condition, with bias terms proportional to $\zeta^2/(1-m)$. In their Gaussian zeroth-order example, they obtain $m = 0$ and an explicit ζ from the Gaussian-specific bound of Nesterov & Spokoiny (2017).

Reconstruction supplies the same population-bias parameter for general component-wise location kernels. After additionally bounding the conditional second moment of the stochastic error directly from the original estimator (q, ψ, f), these quantities can be inserted into existing biased-gradient convergence theorems. We leave that estimator-specific variance analysis outside the present scope.

F.1 Setting and the component-wise smoothed derivatives

Let $\widehat{g}_\mu(\mu, \xi) \in \mathbb{R}^d$ denote a possibly stochastic location update, and define its population update and bias relative to the original objective by

$$g_\mu(\mu) := \mathbb{E}_\xi[\widehat{g}_\mu(\mu, \xi)], \qquad b(\mu) := g_\mu(\mu) - \nabla f(\mu). \tag{63}$$

For each component j, assume that the boundary-fixed reconstruction is a fixed-scale location kernel

$$p_\mu^{(j)}(x) = \kappa_j(x - \mu), \qquad U^{(j)} \sim \kappa_j, \tag{64}$$

where $U^{(j)} \in \mathbb{R}^d$ and κ_j does not depend on μ. Define the component-wise smoothed objective

$$J_j(\mu) := \mathbb{E}[f(\mu + U^{(j)})]. \tag{65}$$

Reconstruction gives $g_{\mu,j} = \partial_{\mu_j} J_j$. Assume that differentiation can be passed under the expectation. A standard sufficient condition is that $\partial_j f(\mu + u)\kappa_j(u)$ is dominated, for μ in the region of interest, by an integrable function of u. The usual location-family reparameterization identity then gives

$$g_{\mu,j}(\mu) = \partial_{\mu_j} \mathbb{E}[f(\mu + U^{(j)})] = \mathbb{E}[\partial_j f(\mu + U^{(j)})]. \tag{66}$$

For Gaussian kernels, this is equivalent to the usual score-function identity through Stein's lemma (Stein, 1981). No common kernel and no scalar potential for the full vector field g_μ are assumed.

Assume that f has an L-Lipschitz gradient,

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\| \qquad \text{for all } x, y \in \mathbb{R}^d, \tag{67}$$

and that

$$a_j := \mathbb{E}[\|U^{(j)}\|] < \infty \qquad \text{for } j = 1, \dots, d. \tag{68}$$

F.2 A PLUG-IN BOUND FOR THE $m = 0$ CASE

Proposition F.1 (Bias from reconstructed kernels). *Under (66)–(68),*

$$|b_j(\mu)| \le L a_j \qquad \text{for every } j, \tag{69}$$

and therefore

$$\|b(\mu)\| \le \zeta_{\mathrm{rec}} := L \left(\sum_{j=1}^{d} a_j^2 \right)^{1/2}. \tag{70}$$

Thus (62) holds with $m = 0$ and $\zeta = \zeta_{\mathrm{rec}}$.

Proof. For component j, (66) gives

$$\begin{aligned} |b_j(\mu)| &= \left| \mathbb{E}[\partial_j f(\mu + U^{(j)})] - \partial_j f(\mu) \right| \\ &\le \mathbb{E}\left[\left| \partial_j f(\mu + U^{(j)}) - \partial_j f(\mu) \right| \right] \\ &\le \mathbb{E}\left[\|\nabla f(\mu + U^{(j)}) - \nabla f(\mu)\| \right] \\ &\le L\mathbb{E}[\|U^{(j)}\|] = L a_j. \end{aligned} \tag{71}$$

The first inequality is $|\mathbb{E}Z| \le \mathbb{E}|Z|$, the second uses $|v_j| \le \|v\|$, and the third uses the L-Lipschitz-gradient assumption. Finally, $\|b(\mu)\|^2 = \sum_j b_j(\mu)^2 \le L^2 \sum_j a_j^2$, which proves (70). □

F.3 RELATION TO GAUSSIAN ZEROTH-ORDER BOUNDS

The preceding result recovers the type of bias calculation used in prior zeroth-order analyses through a single plug-in rule. Ajalloeian and Stich consider

$$g_{\mathrm{GS}}(x, u) = \frac{f(x + \tau u) - f(x)}{\tau} u, \qquad u \sim \mathcal{N}(0, I_d), \tag{72}$$

and use the bound

$$\|\mathbb{E}[g_{\mathrm{GS}}(x, u)] - \nabla f(x)\| \le \frac{L\tau}{2}(d + 3)^{3/2} \tag{73}$$

from Nesterov & Spokoiny (2017). For Gaussian reconstruction, $U^{(j)} \sim \mathcal{N}(0, \tau^2 I_d)$ and $a_j = \tau\mathbb{E}\|u\| \le \tau\sqrt{d}$. Proposition F.1 therefore gives

$$\zeta_{\mathrm{rec}} \le L\tau d. \tag{74}$$

Thus, under the same L-smoothness assumption, the general component-wise result has d rather than $d^{3/2}$ dependence relative to the particular bound used by Ajalloeian and Stich, while also applying to non-Gaussian and mutually incompatible kernels.

When every component shares one common kernel, as in Gaussian ES, the vector can be bounded before separating its coordinates:

$$\|g_\mu(\mu) - \nabla f(\mu)\| = \left\| \mathbb{E}[\nabla f(\mu + U) - \nabla f(\mu)] \right\| \le L\mathbb{E}\|U\| \le L\tau\sqrt{d}. \tag{75}$$

A bound of this form for exact Gaussian smoothing also appears in Berahas et al. (2022). Such target-first analyses begin with a known smoothing kernel; reconstruction supplies that kernel from (q, ψ). The common-kernel estimate is a compatible special case, whereas (70) remains valid without compatibility. The comparison here concerns the population-bias term only; the cited works also analyze estimator-specific second moments and convergence.

F.4 DISTANCE BETWEEN AN EXPECTED EQUILIBRIUM AND THE OPTIMUM

Corollary F.2 (Equilibrium shift). *In addition to the assumptions of Proposition F.1, suppose f is λ-strongly convex:*

$$f(y) \ge f(x) + \langle \nabla f(x), y - x \rangle + \frac{\lambda}{2}\|y - x\|^2 \qquad \text{for all } x, y \in \mathbb{R}^d. \tag{76}$$

Let $\mu^\star$ be its unique minimizer. Any equilibrium $\widehat{\mu}$ of the population update, $g_\mu(\widehat{\mu}) = 0$, satisfies

$$\|\widehat{\mu} - \mu^\star\| \le \frac{\zeta_{\mathrm{rec}}}{\lambda} = \frac{L}{\lambda}\left(\sum_{j=1}^{d}\left(\mathbb{E}\|U^{(j)}\|\right)^2\right)^{1/2}. \tag{77}$$

This conclusion does not require the component-wise kernels to be compatible or g_μ to be a gradient field.

Proof. Strong convexity implies strong monotonicity of the gradient:

$$\langle \nabla f(x) - \nabla f(y), x - y\rangle \ge \lambda \|x - y\|^2. \tag{78}$$

Taking $y = \mu^\star$, using $\nabla f(\mu^\star) = 0$, and applying Cauchy–Schwarz gives

$$\|\nabla f(x)\| \ge \lambda \|x - \mu^\star\|. \tag{79}$$

At an expected equilibrium,

$$\nabla f(\widehat{\mu}) = -b(\widehat{\mu}), \qquad \|\nabla f(\widehat{\mu})\| \le \zeta_{\mathrm{rec}}, \tag{80}$$

where the inequality is Proposition F.1. Combining the last two displays proves (77). □

The corollary locates an expected equilibrium but does not by itself prove that a stochastic iteration converges to it. To apply the full results of Ajalloeian & Stich (2020), one additionally bounds the conditional second moment of the zero-mean stochastic error $\widehat{g}_\mu(\mu, \xi) - g_\mu(\mu)$ directly from the original estimator (q, ψ, f), and verifies the corresponding step-size and objective assumptions. Together with ζ_{rec}, these quantities can then be inserted into existing biased-gradient convergence results.

G Proofs of location–scale identities

This appendix proves the calculus identities used in the location–scale templates of Section 4, and records the full scale-parameter derivation (summarized in the main text) leading to the ray-integration template (22). Let

$$p(x \mid \mu, \sigma) = \sigma^{-d} p_0\left(\frac{x - \mu}{\sigma}\right), \qquad z := \frac{x - \mu}{\sigma}.$$

We will show the location derivative (19) and the scale derivative

$$\sigma \frac{\partial}{\partial \sigma} p(x \mid \mu, \sigma) = -\nabla_x \cdot ((x - \mu) p(x \mid \mu, \sigma)). \tag{81}$$

Identity (19) (location derivative as transport). For each coordinate k, the chain rule gives

$$\begin{aligned}\partial_{\mu_k} p(x \mid \mu, \sigma) &= \underbrace{\sigma^{-d}}_{\text{Jacobian}} \underbrace{\nabla p_0(z)^\top}_{\text{base-density gradient}} \underbrace{\partial_{\mu_k} z}_{=-e_k/\sigma} \\ &= -\sigma^{-d-1}\, \partial_{z_k} p_0(z),\end{aligned}$$

where e_k is the kth standard basis vector. Similarly,

$$\begin{aligned}\partial_{x_k} p(x \mid \mu, \sigma) &= \underbrace{\sigma^{-d}}_{\text{Jacobian}}\, \nabla p_0(z)^\top\, \underbrace{\partial_{x_k} z}_{=e_k/\sigma} \\ &= \sigma^{-d-1}\, \partial_{z_k} p_0(z).\end{aligned}$$

Thus $\partial_{\mu_k} p = -\partial_{x_k} p$, proving (19).

Identity (81) (scale derivative as divergence). Differentiate $p(x \mid \mu, \sigma) = \sigma^{-d} p_0(z)$ with respect to σ:

$$\partial_\sigma p = \underbrace{\partial_\sigma(\sigma^{-d})}_{=-d\sigma^{-d-1}}\, p_0(z) + \sigma^{-d}\, \nabla p_0(z)^\top \partial_\sigma z, \qquad \partial_\sigma z = -\frac{x - \mu}{\sigma^2}.$$

Multiplying by σ and using $p = \sigma^{-d} p_0(z)$ and $\nabla_x p = \sigma^{-d-1} \nabla p_0(z)$ yields

$$\sigma\, \partial_\sigma p = -d\, p - \underbrace{(x-\mu)^\top \nabla_x p}_{\text{radial derivative term}}.$$

On the other hand, the product rule gives the divergence identity

$$\nabla_x \cdot \big((x-\mu)p\big) = \underbrace{\nabla_x \cdot (x-\mu)}_{=d}\, p + (x-\mu)^\top \nabla_x p = d\, p + (x-\mu)^\top \nabla_x p. \tag{82}$$

Combining the two displays gives $\sigma\, \partial_\sigma p = -\nabla_x \cdot ((x-\mu)p)$, proving (81).

From the reconstruction equation to the scale flux PDE. Multiplying the scale reconstruction equation (18) by σ and substituting (81) yields the divergence-form PDE

$$\nabla_x \cdot \big((x-\mu)\, p(x \mid \mu, \sigma)\big) = -\,\sigma\, q(x \mid \mu, \sigma)\, \cdot\, \psi_\sigma(x). \tag{83}$$

Ray reduction (85) from the scale PDE (83). Fix a unit direction $u \in \mathbb{S}^{d-1}$ and write $x = \mu + ru$ with $r = \|x - \mu\| \geq 0$. Define the ray-wise mass function

$$g_u(r) := r^d\, p(\mu + ru \mid \mu, \sigma). \tag{84}$$

A standard polar-coordinate divergence calculation gives

$$\frac{\mathrm{d}}{\mathrm{d}r} g_u(r) = -\,\sigma\, r^{d-1}\, q(\mu + ru \mid \mu, \sigma)\, \psi_\sigma(\mu + ru). \tag{85}$$

Integrating from 0 to r yields

$$p(\mu + ru \mid \mu, \sigma) = \frac{g_u(0)}{r^d} - \frac{\sigma}{r^d} \int_0^r t^{d-1}\, q(\mu + tu \mid \mu, \sigma)\, \cdot\, \psi_\sigma(\mu + tu)\, \mathrm{d}t. \tag{86}$$

For most smooth densities, finiteness at the center implies $g_u(0) = 0$, which recovers the simplified template stated in the main paper as (22).

H Background supplement: method taxonomy and weak-derivative details

Section 2 is intentionally compact. This appendix collects three short add-ons that are useful when reading the worked reconstructions: (i) a taxonomy of common estimators in "numerator" vs. "denominator" form, (ii) the weak-derivative intuition behind endpoint/boundary gradients for uniform kernels, and (iii) a closed-form reconstruction for the 1D uniform random-denominator estimator.

H.1 Taxonomy: perturbations in the numerator vs. the denominator

Many sampling-based updates can be written (in expectation) as

$$\mathbb{E}[\widehat{g}(\theta)] = \mathbb{E}_{X \sim \underbrace{q_\theta}_{\text{sampler}}}\Big[\, \underbrace{f(X)}_{\text{evaluation}}\, \underbrace{\psi_\theta(X)}_{\text{weight}}\Big],$$

and the reconstruction equation (15) asks for a (possibly nonsmooth) smoothing density p_θ such that

$$\mathbb{E}[\widehat{g}(\theta)] = \nabla_\theta \underbrace{\mathbb{E}_{X \sim p_\theta}[f(X)]}_{J(\theta)}.$$

A convenient informal split is whether the random perturbation appears in the *numerator* (as a direction/score) or in the *denominator* (as a random step length):

- **Gauss ES (score / numerator).** Sampling $X = \mu + \sigma Z$ with $Z \sim \mathcal{N}(0, I_d)$ and using score weights yields (9). Reconstruction gives $p = q$ (Appendix I).

- **Sphere (boundary / numerator).** Uniform-in-ball smoothing produces a gradient supported on the boundary sphere, leading to the one-point estimator (88). Appendix M verifies the corresponding transport equation in the distributional sense.
- **Fixed-denominator directional differences (numerator).** The two-point directional estimator (5) also multiplies by a direction Z (numerator), but is most naturally motivated as a small-σ approximation to $\nabla f(\mu)$ unless one explicitly identifies an induced kernel via reconstruction.
- **Random-denominator differences (denominator).** The estimator (6) divides by the realized perturbation ε. Reconstruction yields kernels with logarithmic singularities at the center (Section 5.3); the uniform case admits a simple closed form given below.

H.2 Endpoint/boundary estimators: uniform kernels

Uniform (box/ball) smoothing leads to gradients supported on endpoints or boundaries. Forward and central finite differences (3) can be seen as exact derivatives of interval averages (Appendix L).

In d dimensions, let $B(\mu,\sigma) = \{x \in \mathbb{R}^d : \|x-\mu\| \le \sigma\}$ be the radius-σ ball centered at μ and let V_d denote the volume of the unit ball in $\mathbb{R}^d$. Define the ball-smoothed objective

$$J_{\text{ball}}(\mu;\sigma) \;:=\; \frac{1}{V_d\,\sigma^d}\int_{B(\mu,\sigma)} f(x)\;\mathrm{d}x. \tag{87}$$

Differentiating with respect to μ yields a boundary-gradient formula in terms of values of f on the sphere $\|x-\mu\| = \sigma$. Let U be uniform on the unit sphere $\mathbb{S}^{d-1}$. Then

$$\nabla_\mu J_{\text{ball}}(\mu;\sigma) \;=\; \mathbb{E}_{U\sim\text{Unif}(\mathbb{S}^{d-1})}\left[\frac{d}{\sigma}\,f(\mu+\sigma U)\,U\right]. \tag{88}$$

This is the one-point sphere estimator used in bandit and zeroth-order optimization (Flaxman et al., 2005; Duchi et al., 2015; Shamir, 2017; Nesterov & Spokoiny, 2017; Ghadimi & Lan, 2013).

Two-point (antithetic) sphere estimator. A widely used variance-reduced alternative evaluates both $\mu+\sigma U$ and $\mu-\sigma U$:

$$\widehat{g}_{\text{sph},2}(\mu;\sigma,U) \;:=\; \frac{d}{2\sigma}\big(f(\mu+\sigma U)-f(\mu-\sigma U)\big)\,U. \tag{89}$$

By symmetry of U, the two-point form is unbiased for the same gradient as (88) and typically has reduced variance (Duchi et al., 2015; Shamir, 2017).

H.3 Weak-derivative intuition: why endpoints and boundaries appear

The endpoint/boundary estimators can be understood by viewing smoothing as convolution with a kernel that may have *sharp boundaries*.

1D interval kernels. For the symmetric interval average, the kernel is

$$k_h(t) = \frac{1}{2h}\mathbf{1}\{|t|\le h\}, \qquad J_{\text{cent}}(\mu;h) = (f * k_h)(\mu).$$

Although k_h is not classically differentiable at $\pm h$, its derivative in the distributional sense is supported on the endpoints:

$$k_h'(t) = \frac{1}{2h}\big(\delta(t+h)-\delta(t-h)\big).$$

Placing this into $(f*k_h)'(\mu)$ yields the endpoint difference formula (3). Appendix L gives the same conclusion from the reconstruction transport equation.

Higher dimensional ball kernels. The uniform-in-ball kernel

$$k_\sigma(u) = \frac{1}{V_d\sigma^d}\mathbf{1}\{\|u\|\le\sigma\}$$

has a weak gradient supported on the boundary sphere $\|u\| = \sigma$; heuristically,

$$\nabla k_\sigma(u) \text{ “behaves like” } -\frac{1}{V_d\sigma^d}\,n(u)\,\delta(\|u\|-\sigma), \qquad n(u) = \frac{u}{\|u\|}.$$

Placing this into $\nabla(f*k_\sigma)$ produces the boundary/sphere identity (88). Appendix M checks the corresponding transport equation $-\partial_x p = q\psi$ rigorously in the sense of distributions.

H.4 Uniform random-denominator central differences (1D): closed-form kernel

Consider the Unif-denom estimator (6) with $\varepsilon \sim \mathrm{Unif}([-\sigma, \sigma])$. We retain the antithetic quotient as the primary update. To explain the one-point shorthand, fix $\delta > 0$. Symmetry of ε gives

$$\mathbb{E}\left[\mathbf{1}\{|\varepsilon| > \delta\}\frac{f(\mu+\varepsilon) - f(\mu-\varepsilon)}{2\varepsilon}\right] = \mathbb{E}\left[\mathbf{1}\{|\varepsilon| > \delta\}\frac{f(\mu+\varepsilon)}{\varepsilon}\right]. \tag{90}$$

Letting $\delta \downarrow 0$ recovers the antithetic expectation whenever it exists. Thus, below we use the one-point form only as shorthand for this symmetric pairing. Let $X := \mu + \varepsilon$. Then $X \sim q(\cdot \mid \mu)$ with

$$q(x \mid \mu) = \frac{1}{2\sigma}\mathbf{1}\{|x-\mu| \le \sigma\}, \qquad \psi_\mu(x) = \frac{1}{x-\mu},$$

so the symmetry-paired one-point form is represented by $q(x \mid \mu)\psi_\mu(x)$.

Reconstruction. For a location family, $\partial_\mu p(x \mid \mu) = -\partial_x p(x \mid \mu)$, so the reconstruction equation $\partial_\mu p = q\psi$ takes the transport form

$$-\partial_x p(x \mid \mu) = q(x \mid \mu)\,\psi_\mu(x) = \frac{1}{2\sigma}\,\frac{1}{x-\mu}\,\mathbf{1}\{|x-\mu| \le \sigma\}. \tag{91}$$

Integrating (91) over x on $[\mu-\sigma, \mu+\sigma]$ and imposing $p(\mu \pm \sigma \mid \mu) = 0$ yields the normalized kernel

$$p(x \mid \mu) = \frac{1}{2\sigma}\log\left(\frac{\sigma}{|x-\mu|}\right)\mathbf{1}\{|x-\mu| \le \sigma\}. \tag{92}$$

The kernel (92) is integrable but unbounded at $x = \mu$ (a logarithmic singularity). Consequently, $J(\mu) := \mathbb{E}_{X\sim p(\cdot|\mu)}[f(X)]$ satisfies $J'(\mu) = \mathbb{E}[\widehat{g}_{\mathrm{den}}(\mu;\varepsilon)]$ for a rich class of test functions (in the weak/distributional sense).

Remark (Gaussian case). For Gauss-denom ($\varepsilon \sim \mathcal{N}(0,\sigma^2)$) the same steps give $q(x \mid \mu) = \mathcal{N}(x;\mu,\sigma^2)$ and $\psi_\mu(x) = 1/(x-\mu)$, leading to an even kernel expressible as an integral $p(x \mid \mu) = \int_{|x-\mu|}^{\infty} q_0(u)\,u^{-1}\,du$ with q_0 the Gaussian density. See Section K.2 for the explicit form.

I Gaussian ES: score weights reconstruct $p = q$

This appendix spells out the "sanity check" mentioned in Section 5.1. It shows that when evolution strategies uses *score weights*

$$\psi_{\theta_j}(x) = \partial_{\theta_j}\log q(x \mid \theta),$$

the reconstruction equation (15) integrates to $p_\theta = q_\theta$. Consequently, the update is the gradient of the q_θ-smoothed objective in the sense of Section 3.2.

I.1 A general lemma: score weights force $p = q$

Let $q(x \mid \theta)$ be a differentiable density and set $\psi_{\theta_j}(x) = \partial_{\theta_j}\log q(x \mid \theta)$. Then

$$\underbrace{q(x \mid \theta)\,\psi_{\theta_j}(x)}_{\text{weighting in (15)}} = q(x \mid \theta)\,\partial_{\theta_j}\log q(x \mid \theta) = \underbrace{\partial_{\theta_j} q(x \mid \theta)}_{\text{exact parameter derivative}}. \tag{93}$$

Therefore the reconstruction equation (15),

$$\partial_{\theta_j} p(x \mid \theta) = q(x \mid \theta)\psi_{\theta_j}(x),$$

becomes $\partial_{\theta_j} p = \partial_{\theta_j} q$, implying

$$p(x \mid \theta) = q(x \mid \theta) + C(x) \tag{94}$$

for some C independent of θ. Thus $p = q + C$ for a θ-independent signed offset. Normalization alone gives only $\int C(x)\,\mathrm{d}x = 0$; the canonical reconstruction fixes $C = 0$ by a reference condition (e.g., $p(\cdot \mid \theta_0) = q(\cdot \mid \theta_0)$) or by the boundary/regularity conditions of the chosen smoothing family. With this convention,

$$p(x \mid \theta) = q(x \mid \theta). \tag{95}$$

I.2 Apply to isotropic Gaussian ES

Consider d dimensions with the isotropic Gaussian sampler

$$q(x \mid \mu, \sigma) = \mathcal{N}(\mu, \sigma^2 I_d) = \frac{1}{(2\pi\sigma^2)^{d/2}} \exp\left(-\frac{\|x-\mu\|^2}{2\sigma^2}\right). \tag{96}$$

The score weights are

$$\begin{aligned} \psi_\mu(x) &= \nabla_\mu \log q(x \mid \mu, \sigma) = \underbrace{\frac{x-\mu}{\sigma^2}}_{\text{location score}}, \\ \psi_\sigma(x) &= \partial_\sigma \log q(x \mid \mu, \sigma) = \underbrace{-\frac{d}{\sigma}}_{\text{Jacobian term}} + \underbrace{\frac{\|x-\mu\|^2}{\sigma^3}}_{\text{quadratic term}}. \end{aligned} \tag{97}$$

Using (93) with $\theta = (\mu, \sigma)$ gives

$$\partial_{\mu_k} p(x \mid \mu, \sigma) = \partial_{\mu_k} q(x \mid \mu, \sigma), \qquad \partial_\sigma p(x \mid \mu, \sigma) = \partial_\sigma q(x \mid \mu, \sigma),$$

so the lemma (95) implies

$$p(x \mid \mu, \sigma) = q(x \mid \mu, \sigma) = \mathcal{N}(\mu, \sigma^2 I). \tag{98}$$

I.3 Induced objective

With (98), the generic smoothed objective becomes the Gaussian-smoothed objective

$$J(\mu, \sigma) = \int_{\mathbb{R}^d} f(x) \underbrace{p(x \mid \mu, \sigma)}_{=q(x|\mu,\sigma)} \, \mathrm{d}x = \mathbb{E}_{x \sim \mathcal{N}(\mu, \sigma^2 I_d)}[f(x)]. \tag{99}$$

J Uniform perturbations with ES-style weights

Let $\sigma := \delta\eta$, and define the interval $\mathcal{I}_\sigma(\mu) := \{x \in \mathbb{R} : |x-\mu| \le \sigma\}$. The sampling distribution is $q(x \mid \mu) = \frac{1}{2\sigma} \mathbb{1}_{\mathcal{I}_\sigma(\mu)}$. Using symmetry, the expected update induced by (23) can be written in the LR form $\mathbb{E}_{x \sim q(\cdot|\mu)}[\, f(x)\, \psi_\mu(x)\,]$ with

$$\psi_\mu(x) \;=\; \frac{3(x-\mu)}{\sigma^2}. \tag{100}$$

Applying the reconstruction equation for location families yields the corresponding smoothing density (24).

K Random-denominator central differences: explicit p for non-score weights

We derive the explicit reconstructed density p_θ and the induced objective for random-denominator central differences using Gauss-denom and Unif-denom, respectively.

Rewrite as a weighted expectation. Let $x := \mu + \varepsilon$ and let $q(\cdot \mid \mu, \sigma)$ denote the density of x. We retain

$$\widehat{g}_{\mathrm{rd}}(\mu; \varepsilon) = \frac{f(\mu+\varepsilon) - f(\mu-\varepsilon)}{2\varepsilon}$$

as the primary estimator. By the symmetric-pairing argument in Appendix H, its expectation can be represented in the one-point weighted form

$$\mathbb{E}[\widehat{g}_{\mathrm{rd}}(\mu; \varepsilon)] = \mathbb{E}_{x \sim q(\cdot|\mu,\sigma)}\left[f(x)\, \psi_\mu(x)\right], \qquad \psi_\mu(x) = \frac{1}{x-\mu}, \tag{101}$$

where the right-hand side is understood through that symmetric pairing rather than as a separate absolute-integrability claim. Thus, in our reconstruction notation, (q, ψ_μ) is known explicitly. Plugging $\psi_\mu(x) = 1/(x-\mu)$ into (20) yields

$$\partial_x p(x \mid \mu, \sigma) = -\frac{q(x \mid \mu, \sigma)}{x - \mu}. \tag{102}$$

We now integrate (102) for two common perturbation laws.

K.1 Uniform perturbations: a logarithmic kernel

Assume $\varepsilon \sim \mathrm{Unif}[-\sigma, \sigma]$. Then $x = \mu + \varepsilon$ has density

$$q(x \mid \mu, \sigma) = \frac{1}{2\sigma}\, \mathbb{1}_{\mathcal{I}_\sigma(\mu)}. \tag{103}$$

Substituting (103) into (102) gives

$$\partial_x p(x \mid \mu, \sigma) = -\frac{1}{2\sigma}\, \frac{\mathbb{1}_{\mathcal{I}_\sigma(\mu)}}{x - \mu}.$$

Outside $[\mu - \sigma, \mu + \sigma]$, the right-hand side is zero, so p is constant there; the boundary condition $p(x) \to 0$ as $|x| \to \infty$ implies $p(x) = 0$ for $|x - \mu| > \sigma$.

For $x \in (\mu, \mu + \sigma]$, integrate from x to $\mu + \sigma$ (where $p = 0$):

$$p(x \mid \mu, \sigma) = \int_x^{\mu+\sigma} \frac{1}{2\sigma} \frac{1}{t - \mu}\, \mathrm{d}t = \frac{1}{2\sigma} \log\left(\frac{\sigma}{x - \mu}\right).$$

For $x \in [\mu - \sigma, \mu)$, integrate from $\mu - \sigma$ to x (again using $p(\mu - \sigma) = 0$):

$$p(x \mid \mu, \sigma) = -\int_{\mu-\sigma}^{x} \frac{1}{2\sigma} \frac{1}{t - \mu}\, \mathrm{d}t = \frac{1}{2\sigma} \log\left(\frac{\sigma}{\mu - x}\right).$$

Combining both sides yields the explicit reconstructed density

$$p_{\mathrm{rd,unif}}(x \mid \mu, \sigma) = \frac{1}{2\sigma} \log\left(\frac{\sigma}{|x - \mu|}\right) \mathbb{1}_{\mathcal{I}_\sigma(\mu)}. \tag{104}$$

This kernel has a logarithmic (integrable) singularity at $x = \mu$ and integrates to one (a direct check).

Induced objective. The method is therefore an unbiased estimator of $\partial_\mu J_{\mathrm{rd,unif}}(\mu, \sigma)$ where

$$J_{\mathrm{rd,unif}}(\mu, \sigma) = \int f(x)\, p_{\mathrm{rd,unif}}(x \mid \mu, \sigma)\, \mathrm{d}x. \tag{105}$$

K.2 Gaussian perturbations: an exponential-integral kernel

Assume $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. Then

$$q(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right). \tag{106}$$

For $x > \mu$, substituting (106) into (21) and setting $u = t - \mu$ gives the reconstruction. Extending by symmetry to $x < \mu$ yields the ordinary tail integral

$$p(x \mid \mu, \sigma) = \int_{|x-\mu|}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{u^2}{2\sigma^2}\right) \frac{1}{u}\, \mathrm{d}u, \qquad x \neq \mu. \tag{107}$$

This integral can be evaluated in closed form using the exponential integral $\mathrm{Ei}(\cdot)$, yielding

$$p_{\mathrm{rd,gauss}}(x \mid \mu, \sigma) = -\frac{1}{2\sigma\sqrt{2\pi}}\, \mathrm{Ei}\left(-\frac{(x - \mu)^2}{2\sigma^2}\right). \tag{108}$$

As in the uniform case, the singularity at $x = \mu$ is logarithmic and integrable, so $p_{\mathrm{rd,gauss}}$ is a valid density.

Induced objective. The estimator (6) is an unbiased estimator of $\partial_\mu J_{\mathrm{rd,gauss}}(\mu, \sigma)$ where

$$J_{\mathrm{rd,gauss}}(\mu, \sigma) = \int f(x)\, p_{\mathrm{rd,gauss}}(x \mid \mu, \sigma)\, \mathrm{d}x. \tag{109}$$

L FINITE-DIFFERENCE RECONSTRUCTIONS

This appendix spells out the reconstruction recipe (17)–(20) for the classical one-dimensional finite-difference rules. The key point is that each finite difference is *exactly* the gradient of a simple *box-kernel* average of f.

L.1 FORWARD DIFFERENCE: ONE-SIDED (ENDPOINT) BOX SMOOTHING

The forward estimator is

$$\widehat{g}_{\mathrm{fwd}}(\mu) = \frac{f(\mu+h) - f(\mu)}{h}.$$

Introduce the two-point sampling measure

$$q_{\mathrm{fwd}}(x \mid \mu) = \frac{1}{2}\delta(x-\mu) + \frac{1}{2}\delta(x-(\mu+h)), \qquad \psi_\mu(\mu) = -\frac{2}{h}, \;\; \psi_\mu(\mu+h) = +\frac{2}{h},$$

so that

$$\mathbb{E}_{X\sim q_{\mathrm{fwd}}(\cdot|\mu)}\big[f(X)\psi_\mu(X)\big] = \frac{f(\mu+h)-f(\mu)}{h} = \widehat{g}_{\mathrm{fwd}}(\mu).$$

Reconstructing p via the transport equation. For a location parameter, $\partial_\mu p(x \mid \mu) = -\partial_x p(x \mid \mu)$, so the 1D instance of (20) becomes

$$\underbrace{-\partial_x p(x \mid \mu)}_{\text{transport ODE}} = \underbrace{q_{\mathrm{fwd}}(x \mid \mu)}_{\text{sampling}}\underbrace{\psi_\mu(x)}_{\text{weights}} = -\frac{1}{h}\delta(x-\mu) + \frac{1}{h}\delta(x-(\mu+h)). \tag{110}$$

Integrating (110) using $\partial_x \mathbb{1}\{x \geq a\} = \delta(x-a)$ and imposing that p has no mass outside the interval gives the one-sided box kernel

$$p_{\mathrm{fwd}}(x \mid \mu) = \frac{1}{h}\,\mathbb{1}\{\mu \leq x \leq \mu+h\}. \tag{111}$$

Induced objective. Defining $J_{\mathrm{fwd}}(\mu) \coloneqq \mathbb{E}_{X\sim p_{\mathrm{fwd}}(\cdot|\mu)}[f(X)]$ yields

$$J_{\mathrm{fwd}}(\mu) = \frac{1}{h}\int_\mu^{\mu+h} f(x)\,\mathrm{d}x, \qquad J'_{\mathrm{fwd}}(\mu) = \frac{f(\mu+h)-f(\mu)}{h} = \widehat{g}_{\mathrm{fwd}}(\mu). \tag{112}$$

Thus the forward difference is the exact gradient of a one-sided, endpoint-smoothed objective.

L.2 CENTRAL DIFFERENCE: SYMMETRIC BOX SMOOTHING

The central estimator is

$$\widehat{g}_{\mathrm{cent}}(\mu) = \frac{f(\mu+h) - f(\mu-h)}{2h}.$$

Use the symmetric two-point measure

$$q_{\mathrm{cent}}(x \mid \mu) = \frac{1}{2}\delta(x-(\mu+h)) + \frac{1}{2}\delta(x-(\mu-h)), \qquad \psi_\mu(\mu+h) = +\frac{1}{h}, \;\; \psi_\mu(\mu-h) = -\frac{1}{h}.$$

Then $\mathbb{E}_{q_{\mathrm{cent}}}[f(X)\psi_\mu(X)] = \widehat{g}_{\mathrm{cent}}(\mu)$, and the transport equation gives

$$-\partial_x p(x \mid \mu) = q_{\mathrm{cent}}(x \mid \mu)\,\psi_\mu(x) = \frac{1}{2h}\delta(x-(\mu+h)) - \frac{1}{2h}\delta(x-(\mu-h)). \tag{113}$$

Integrating (113) yields the symmetric box kernel

$$p_{\mathrm{cent}}(x \mid \mu) = \frac{1}{2h}\,\mathbb{1}\{\mu-h \leq x \leq \mu+h\}. \tag{114}$$

Consequently,

$$J_{\mathrm{cent}}(\mu) \coloneqq \mathbb{E}_{X\sim p_{\mathrm{cent}}(\cdot|\mu)}[f(X)] = \frac{1}{2h}\int_{\mu-h}^{\mu+h} f(x)\,\mathrm{d}x, \qquad J'_{\mathrm{cent}}(\mu) = \widehat{g}_{\mathrm{cent}}(\mu). \tag{115}$$

M SPHERE-SAMPLING: DISTRIBUTIONAL CHECK

This section checks that the uniform-in-a-ball kernel underlying J_{ball} in (87) satisfies the transport equation (20) in the sense of distributions.

Kernel and induced (q, ψ) pair. Let

$$p(x \mid \mu, \sigma) = \frac{1}{V_d \sigma^d} \mathbb{1}\{\|x - \mu\| \le \sigma\}.$$

The one-point sphere estimator (88) corresponds to sampling on the boundary sphere, i.e.

$$\underbrace{q(x \mid \mu, \sigma)}_{\text{uniform on } \|x-\mu\|=\sigma} = \frac{1}{S_{d-1}\sigma^{d-1}} \delta(\|x - \mu\| - \sigma), \qquad \underbrace{\psi_\mu(x)}_{\text{score weight}} = \frac{d}{\sigma^2}(x - \mu). \tag{116}$$

Using $S_{d-1} = dV_d$ (surface area vs. volume of the unit ball), we obtain

$$q(x \mid \mu, \sigma)\, \psi_{\mu_k}(x) = \frac{x_k - \mu_k}{V_d \sigma^{d+1}} \delta(\|x - \mu\| - \sigma). \tag{117}$$

Weak derivative of the ball kernel. The distributional derivative of the indicator is supported on the boundary:

$$\underbrace{-\partial_{x_k} \mathbb{1}\{\|x - \mu\| \le \sigma\}}_{\text{outward flux}} = \underbrace{\frac{x_k - \mu_k}{\|x - \mu\|}}_{\text{unit normal } n_k(x)} \delta(\|x - \mu\| - \sigma).$$

Therefore,

$$-\partial_{x_k} p(x \mid \mu, \sigma) = \frac{1}{V_d \sigma^d} \frac{x_k - \mu_k}{\|x - \mu\|} \delta(\|x - \mu\| - \sigma) = \frac{x_k - \mu_k}{V_d \sigma^{d+1}} \delta(\|x - \mu\| - \sigma),$$

where the last equality uses that $\|x - \mu\| = \sigma$ on the support of the delta. Comparing with (117) shows that $-\partial_{x_k} p = q\, \psi_{\mu_k}$, i.e. (20) (equivalently (17)) holds in the distributional sense.

N DERIVATION FOR UNIFORM ES-STYLE PERTURBATIONS (1RDSA-UNIF)

This appendix derives the explicit smoothing density $p(\cdot \mid \mu)$ induced by the uniform-perturbation estimator of Prashanth et al. (2017) (*1RDSA-Unif*). The steps mirror the general reconstruction recipe used throughout the paper: (i) rewrite the *expected* update in the (q, ψ) form, and (ii) solve the resulting transport equation (the 1D case of (20)) to recover p and the smoothed objective $J(\mu) = \mathbb{E}_{x \sim p(\cdot|\mu)}[f(x)]$.

For transparency, we present the one-dimensional case. The estimator (from (23)) is

$$\widehat{g}(\mu) \;=\; \frac{3}{\eta^2}\, d\, \frac{f(\mu + \delta d) - f(\mu - \delta d)}{2\delta}, \qquad d \sim \text{Unif}([-\eta, \eta]). \tag{23}$$

Step 1: reduce the two-point form using symmetry. Taking expectations and using that the law of d is symmetric ($d \overset{d}{=} -d$), we obtain

$$\begin{aligned}\mathbb{E}[\widehat{g}(\mu)] &= \frac{3}{\eta^2}\, \mathbb{E}\left[d\, \frac{f(\mu + \delta d) - f(\mu - \delta d)}{2\delta}\right] \\ &= \frac{3}{\eta^2} \cdot \frac{1}{2\delta} \left(\mathbb{E}[df(\mu + \delta d)] - \mathbb{E}[df(\mu - \delta d)]\right).\end{aligned} \tag{118}$$

For the second term, substitute $d \mapsto -d$:

$$\mathbb{E}[df(\mu - \delta d)] = \mathbb{E}[-df(\mu + \delta d)] = -\mathbb{E}[df(\mu + \delta d)].$$

Hence,

$$\mathbb{E}[\widehat{g}(\mu)] = \frac{3}{\eta^2 \delta}\, \mathbb{E}\big[d\, f(\mu + \delta d)\big]. \tag{119}$$

Step 2: express in the (q, ψ) form. Let $\sigma := \delta\eta$ and define the random variable $x = \mu + \delta d$. Since $d \sim \mathrm{Unif}([-\eta, \eta])$, we have $x \sim \mathrm{Unif}([\mu - \sigma, \mu + \sigma])$, i.e.

$$\underbrace{q(x \mid \mu)}_{\text{sampling law}} = \frac{1}{2\sigma}\, \mathbb{1}_{\mathcal{I}_\sigma(\mu)}.$$

Moreover, $d = (x - \mu)/\delta$, and $\eta = \sigma/\delta$, so

$$\frac{3}{\eta^2\delta}\, d = \frac{3}{(\sigma/\delta)^2\delta} \cdot \frac{x - \mu}{\delta} = \underbrace{\frac{3(x - \mu)}{\sigma^2}}_{\psi_\mu(x)}.$$

Substituting into (119) yields the likelihood-ratio form

$$\mathbb{E}[\widehat{g}(\mu)] = \mathbb{E}_{x \sim q(\cdot \mid \mu)}\big[f(x)\, \psi_\mu(x)\big], \qquad \psi_\mu(x) = \frac{3(x - \mu)}{\sigma^2}. \tag{120}$$

Note that $\mathbb{E}_{x \sim q}[\, \psi_\mu(x)\,] = \frac{3}{\sigma^2}\mathbb{E}[x - \mu] = 0$, so no additional drift term is required.

Step 3: reconstruct the smoothing density p. For a location family $p(x \mid \mu) = p_0(x - \mu)$, we have $\partial_\mu p(x \mid \mu) = -\partial_x p(x \mid \mu)$. Combining this identity with the reconstruction equation $\partial_\mu p = q\, \psi_\mu$ and (120) gives

$$\underbrace{-\partial_x p(x \mid \mu)}_{\text{transport ODE}} = \underbrace{q(x \mid \mu)}_{\frac{1}{2\sigma}\mathbb{1}_{\mathcal{I}_\sigma(\mu)}}\ \underbrace{\psi_\mu(x)}_{\frac{3(x-\mu)}{\sigma^2}} = \frac{3(x - \mu)}{2\sigma^3}\, \mathbb{1}_{\mathcal{I}_\sigma(\mu)}. \tag{121}$$

Integrating (121) over x on $[\mu - \sigma, \mu + \sigma]$ yields

$$p(x \mid \mu) = C - \frac{3}{4\sigma^3}(x - \mu)^2, \qquad |x - \mu| \le \sigma.$$

Imposing the boundary condition $p(\mu \pm \sigma \mid \mu) = 0$ (no jump mass at the endpoints) gives $C = \frac{3}{4\sigma}$, and therefore the (Epanechnikov) quadratic kernel

$$p(x \mid \mu) = \frac{3}{4\sigma}\left(1 - \frac{(x - \mu)^2}{\sigma^2}\right) \mathbb{1}_{\mathcal{I}_\sigma(\mu)}. \tag{122}$$

Consequently, defining the smoothed objective

$$J(\mu) := \mathbb{E}_{x \sim p(\cdot \mid \mu)}[f(x)]$$

gives

$$\nabla_\mu J(\mu) = \mathbb{E}[\widehat{g}(\mu)],$$

i.e. 1RDSA-Unif performs gradient ascent/descent on the quadratic-kernel-smoothed objective J.

O Beyond mean updates: reward transformations and step-size control

So far, our worked examples have focused on *mean* (location) updates, where the reconstruction equations yield an explicit smoothing density $p(\cdot \mid \mu, \sigma)$ and an induced smoothed objective $J(\mu, \sigma) = \int f(x) p(x \mid \mu, \sigma)\, \mathrm{d}x$. In practice, many evolution strategies and derivative-free methods include two additional ingredients: (i) *reward transformations* (ranks, elites, fitness shaping) and (ii) *step-size* (and sometimes covariance) adaptation. This section briefly explains how both fit into the reconstruction viewpoint, and highlights a representative σ-update rule—the classical $1/5$ success rule—as an example where the update is naturally interpreted as a *drift* term rather than a gradient of a normalized smoothed objective.

O.1 Reward transformations: ranks and elites change the objective, not the reconstruction

A large fraction of ES algorithms do not use the raw evaluations $f(x_i)$ directly, but first apply a deterministic *reward transformation* to obtain utilities r_i (or weights) and then update parameters using the generic form

$$\Delta\theta \;=\; \sum_{i=1}^{N} r_i\, \psi_\theta(x_i), \qquad x_i \sim q(\cdot \mid \theta).$$

Two canonical examples are: (i) *elite selection* in the cross-entropy method (CEM), which assigns r_i as an elite indicator or a truncated weight (Rubinstein, 1999; de Boer et al., 2005), and (ii) *rank-based* utilities in CMA-ES and related algorithms (Hansen, 2016; Ollivier et al., 2017; Wierstra et al., 2014), which assign weights based only on the ordering of fitness values (making the method invariant to monotone transformations of f).

In the reconstruction viewpoint, it is useful to separate two roles: *(a) the smoothing distribution* $p(\cdot \mid \theta)$ implied by (q, ψ) via the reconstruction equations, and *(b) the integrand* being averaged under p. Reward transformations primarily affect (b): when utilities are pointwise, $r_i = u(f(x_i))$, they replace the objective by a utility-smoothed objective $\mathbb{E}_{x\sim p}[u(f(x))]$; when utilities depend on the batch (e.g. ranks, elites), they correspond more naturally to *population* objectives or quantile-type criteria. In many ES/IGO-style updates, one chooses the score weight $\psi_\theta(x) = \nabla_\theta \log q(x \mid \theta)$, in which case the reconstruction yields $p = q$ (Appendix I), while the reward transformation determines which functional of f is being optimized (Ollivier et al., 2017; Wierstra et al., 2014).

O.2 A σ-update example: the $1/5$ success rule as a drift term

Step-size control rules are often presented as part of the "engineering" of evolution strategies. A historically important example is Rechenberg's $1/5$ success rule for the $(1+1)$-ES (Rechenberg, 1973; Schwefel, 1995). In minimization form, the algorithm samples an offspring

$$x = \mu + \sigma Z, \qquad Z \sim \mathcal{N}(0, I),$$

and declares *success* if the offspring improves on the parent:

$$s(x; \mu) := \mathbf{1}\{f(x) \le f(\mu)\}. \tag{123}$$

The $1/5$ rule adapts σ to keep the success probability near $1/5$.

A convenient parameterization. A common "textbook" form of the $1/5$ success rule updates σ multiplicatively: if the iteration is successful, $\sigma \leftarrow \sigma\alpha$, and otherwise $\sigma \leftarrow \sigma\alpha^{-1/4}$, where $\alpha > 1$ is fixed.

Write the step-size parameter as $\theta := \log\sigma$ (so $\sigma = e^\theta$). The multiplicative rule gives the log-step increment

$$\Delta\theta = \begin{cases} \log\alpha, & s(x;\mu) = 1, \\ -\frac{1}{4}\log\alpha, & s(x;\mu) = 0. \end{cases}$$

Writing $p_s := \Pr(f(x) \le f(\mu))$,

$$\mathbb{E}[\Delta\theta] = p_s \log\alpha - (1-p_s)\frac{1}{4}\log\alpha = \frac{5p_s - 1}{4}\log\alpha,$$

so the fixed point is $p_s = 1/5$. A smooth additive proxy with the same root is therefore

$$\Delta\theta \;=\; \eta\big(s(x;\mu) - \tfrac{1}{5}\big), \tag{124}$$

where η absorbs the constant factor.

Attempting reconstruction: identify (q, ψ_θ, r). Under the Gaussian sampler,

$$x \sim q(\cdot \mid \mu, \sigma) = \mathcal{N}(\mu, \sigma^2 I),$$

the additive update (124) can be written in our "sample-then-weight" form as

$$\Delta\theta \;=\; r(x)\,\psi_\theta(x), \qquad r(x) = s(x;\mu) - \tfrac{1}{5}, \qquad \psi_\theta(x) \equiv \eta.$$

Crucially, the weight ψ_θ is *constant* (it does not depend on x).

Centering decomposition and the drift term. Section 3.4 showed that when $\mathbb{E}_q[\psi] \neq 0$, an update decomposes into a gradient-compatible part plus a constant drift term:

$$\mathbb{E}_q[r(x)\psi(x)] = \mathbb{E}_q\big[r(x)\big(\psi(x) - \mathbb{E}_q[\psi(x)]\big)\big] + \mathbb{E}_q[\psi(x)]\,\mathbb{E}_q[r(x)]. \tag{125}$$

Applying (125) with $\psi_\theta(x) \equiv \eta$ yields

$$\psi_\theta(x) - \mathbb{E}_q[\psi_\theta(x)] \equiv \eta - \eta \equiv 0,$$

so the gradient-compatible term vanishes identically. The expected update is therefore *pure drift*:

$$\mathbb{E}[\Delta\theta] = \mathbb{E}_q[r(x)\psi_\theta(x)] = \eta\,\mathbb{E}_q[r(x)] = \eta\big(\Pr_{x\sim q}(f(x) \leq f(\mu)) - \tfrac{1}{5}\big). \tag{126}$$

Interpretation in the reconstruction framework. If we enforce normalization by centering the weight (Section 3.4), then for the $1/5$ rule the centered weight is $\psi_\theta^{\mathrm{ctr}} \equiv 0$. The corresponding reconstruction equation becomes

$$\partial_\theta p(x \mid \theta) = q(x \mid \theta)\,\psi_\theta^{\mathrm{ctr}}(x) \equiv 0,$$

so there is no nontrivial smoothing density $p(\cdot \mid \theta)$ whose *normalized* gradient matches (124). Instead, the update (124) is naturally interpreted as a stochastic control / root-finding mechanism that pushes the success probability $\Pr(f(\mu + \sigma Z) \leq f(\mu))$ toward the target value $1/5$.

Takeaway. The $1/5$ success rule illustrates a boundary of the reconstruction principle: mean updates based on likelihood-ratio or generalized-gradient weights admit a direct $(q, \psi) \mapsto p$ reconstruction of a smoothed objective gradient, whereas some step-size control rules are most naturally understood as *drift* updates that regulate a scalar statistic (here, success probability) rather than estimate $\partial_\theta \mathbb{E}_{x\sim p_\theta}[f(x)]$ for a normalized p_θ.

O.3 Natural gradients as preconditioning in the reconstruction framework

Many ES variants described as "natural-gradient" methods can be viewed in our framework as taking the same basic stochastic direction and then applying a *preconditioner* in parameter space. Concretely, starting from an update of the form

$$\Delta\theta \;=\; \sum_{i=1}^{N} r_i\,\psi_\theta(x_i), \qquad x_i \sim q(\cdot \mid \theta),$$

a natural-gradient-like method replaces it by

$$\Delta\theta_{\mathrm{NG}} \;=\; A(\theta)\,\Delta\theta, \qquad \text{typically } A(\theta) = F(\theta)^{-1}, \tag{127}$$

where $F(\theta)$ is the Fisher information matrix of the search distribution $q(\cdot \mid \theta)$. This preconditioning changes the *geometry* of the parameter update (rescaling and mixing coordinates), but it does not change the reconstruction logic: the implied smoothing distribution p is determined by the pair (q, ψ) (and by any centering used to enforce normalization), while $A(\theta)$ simply post-multiplies the resulting direction. In particular, when one uses score weights $\psi_\theta(x) = \nabla_\theta \log q(x \mid \theta)$ (as in NES/IGO-style updates), the reconstruction yields $p = q$ (Appendix I), and the Fisher preconditioner provides a principled way to map the score-gradient into the familiar moment-style update directions used in practice.