A unified view of likelihood ratio and reparameterization gradients

Paavo Parmas
Kyoto University[1]

Masashi Sugiyama
RIKEN and The University of Tokyo

Abstract

Reparameterization (RP) and likelihood ratio (LR) gradient estimators are used to estimate gradients of expectations throughout machine learning and reinforcement learning; however, they are usually explained as simple mathematical tricks, with no insight into their nature. We use a first principles approach to explain that LR and RP are alternative methods of keeping track of the movement of probability mass, and the two are connected via the divergence theorem. Moreover, we show that the space of all possible estimators combining LR and RP can be completely parameterized by a flow field $\boldsymbol{u}(\boldsymbol{x})$ and importance sampling distribution $q(\boldsymbol{x})$. We prove that there cannot exist a single-sample estimator of this type outside our characterized space, thus, clarifying where we should be searching for better Monte Carlo gradient estimators.

1 INTRODUCTION

Both likelihood ratio (LR) gradients (Glynn, 1990; Williams, 1992) and reparameterization (RP) gradients (Rezende et al., 2014; Kingma and Welling, 2013) give unbiased estimates of the gradient of an expectation w.r.t. the parameters of the distribution: $\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_{p(x;\theta)}\left[\phi(x)\right]$. This gradient estimation problem is fundamental in machine learning (Mohamed et al., 2019), where the gradients are used for optimization. LR is the basis of many reinforcement learning (RL) (Sutton and Barto, 1998; Schulman et al., 2015b, 2017; Sutton et al., 2000; Peters and Schaal, 2008) and evolutionary algorithms (Wierstra et al., 2008; Salimans et al., 2017; Ha and Schmidhuber, 2018; Conti et al., 2018). In RL, $\phi(x)$ represents the sum of rewards, and θ are the policy parameters. RP, on the other hand, is the backbone of stochastic variational inference (Hoffman et al., 2013), where $\phi(x)$ is the evidence lower bound, and θ are the variational parameters. For example, RP is used in autoencoders (Kingma and Welling, 2013). There is a vast body of research on both estimators (App. A), and there is no clear winner among RP and LR—both have advantages and disadvantages.

LR uses samples of the value of the function $\phi(x)$ to estimate the gradient, and is usually derived as

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_{p(x;\theta)}\left[\phi(x)\right] &= \int \frac{\mathrm{d}p\left(x;\theta\right)}{\mathrm{d}\theta}\phi(x)\mathrm{d}x\\ &= \int p\left(x;\theta\right)\frac{1}{p\left(x;\theta\right)}\frac{\mathrm{d}p\left(x;\theta\right)}{\mathrm{d}\theta}\phi(x)\mathrm{d}x\\ &= \int p\left(x;\theta\right)\frac{\mathrm{d}\log p\left(x;\theta\right)}{\mathrm{d}\theta}\phi(x)\mathrm{d}x\\ &= \mathbb{E}_{p(x;\theta)}\left[\frac{\mathrm{d}\log p\left(x;\theta\right)}{\mathrm{d}\theta}\phi(x)\right].\end{aligned} \tag{1}$$

On the other hand, RP uses samples of the gradient of the function $\frac{\mathrm{d}\phi(x)}{\mathrm{d}x}$, and it is derived by defining a mapping $g(\epsilon;\theta)=x$, where ϵ is sampled from a fixed simple distribution, $p(\epsilon)$, independent of θ, but x ends up being sampled from the desired distribution. For example, if x is Gaussian, $x \sim \mathcal{N}(\mu,\sigma)$, then the required mapping is $g(\epsilon;\theta)=\mu+\sigma\epsilon$, where $\epsilon \sim \mathcal{N}(0,1)$, and the RP gradient is derived as

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_{p(x;\theta)}\left[\phi(x)\right] &= \frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_{\epsilon\sim\mathcal{N}(0,1)}\left[\phi\left(g(\epsilon;\theta)\right)\right]\\ &= \mathbb{E}_{\epsilon\sim\mathcal{N}(0,1)}\left[\frac{\mathrm{d}\phi\left(g(\epsilon;\theta)\right)}{\mathrm{d}\theta}\right]\\ &= \mathbb{E}_{\epsilon\sim\mathcal{N}(0,1)}\left[\frac{\mathrm{d}\phi\left(g(\epsilon;\theta)\right)}{\mathrm{d}g}\frac{\mathrm{d}g(\epsilon;\theta)}{\mathrm{d}\theta}\right],\end{aligned} \tag{2}$$

where $\theta=[\mu,\sigma]$, $\frac{\mathrm{d}g}{\mathrm{d}\mu}=1$, $\frac{\mathrm{d}g}{\mathrm{d}\sigma}=\epsilon$ and $\frac{\mathrm{d}\phi(g(\epsilon;\theta))}{\mathrm{d}g}=\frac{\mathrm{d}\phi(x)}{\mathrm{d}x}$.

What do these derivations mean, and what is the relationship between the two methods? We give two possible answers to this question: (i) we give a first principles explanation that these are different methods

[1] Work mostly performed while affiliated with the Okinawa Institute of Science and Technology and partially performed while interning at RIKEN.

Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS) 2021, San Diego, California, USA. PMLR: Volume 130. Copyright 2021 by the author(s).

of keeping track of the movement of probability mass (Sec. 3), (ii) we show that RP and LR are *duals* under the divergence theorem when considering the integral of a probability mass flow (Sec. 4). Our theory gives a physical insight by analogy to fluid dynamics, and allows for intuitive visualizations. Our main technical result is in Thm. 3, where we formalize a generalized estimator that includes all previous LR and RP gradients as special cases, and we prove that there cannot exist an estimator of this type outside our characterized space. Finally, we advocate for a systematic approach in the search for novel gradient estimators (Sec. 5).

2 PRELIMINARIES

We introduce some preliminaries. In Eq. (4) we introduce a general form of all gradient estimators of the LR–RP type. Sec. 2.2 explains that the introduced equation indeed includes RP as a special case. Sec. 2.3 introduces the Monte Carlo (MC) integration principle, which provides a link between integral expressions and the corresponding gradient estimators. Sec. 2.4 introduces previous works on the relationship between LR and RP, and the limitations of these works. Sec. 2.5 gives basic knowledge about fluid dynamics and the divergence theorem necessary for understanding Sec. 4.

2.1 Setup

Problem statement. *Given one sample $\boldsymbol{x} \sim q(\boldsymbol{x})$, and while being allowed to evaluate $\phi(\boldsymbol{x})$ and $\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x})$, construct an estimator, E_{θ_i}, which may depend on $\boldsymbol{x}, \phi(\boldsymbol{x})$ and $\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x})$, s.t.*

$$\mathbb{E}_{q(\boldsymbol{x})}\left[E_{\theta_i}\right] = \frac{\mathrm{d}}{\mathrm{d}\theta_i}\mathbb{E}_{p(\boldsymbol{x})}\left[\phi(\boldsymbol{x})\right]. \qquad (3)$$

To obtain an estimator for the full gradient w.r.t. θ (as opposed to the derivative w.r.t. one element of the parameters θ_i), we can stack the estimators together: $E_\theta = [E_{\theta_1}, E_{\theta_2}, \ldots]$. Note that, while in the problem statement we consider one sample, $\boldsymbol{x} \sim q(\boldsymbol{x})$, the estimators can also be used with multiple samples in a batch by averaging the estimates together. The main reason we explicitly write out this problem statement is to emphasize the format based on having access to $\phi(\boldsymbol{x})$ and $\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x})$. We will further restrict our discussion to estimators having the following product form:

$$E_{\theta_i} = \boldsymbol{u}_{\theta_i}(\boldsymbol{x}) \cdot \nabla_{\boldsymbol{x}}\phi(\boldsymbol{x}) + \psi_{\theta_i}(\boldsymbol{x})\phi(\boldsymbol{x}), \qquad (4)$$

where $\boldsymbol{u}_{\theta_i}(\boldsymbol{x})$ is an arbitrary vector field, and $\psi_{\theta_i}(\boldsymbol{x})$ is an arbitrary scalar field (a function). Essentially, this equation is taking a weighted sum of the partial derivatives and value of $\phi(\boldsymbol{x})$, with a different weighting defined at each $\boldsymbol{x}$. Both LR and RP belong to this class of gradient estimators. Assuming that the sampling distribution is $q(\boldsymbol{x}) = p(\boldsymbol{x};\theta)$, we obtain LR by setting $\boldsymbol{u}_{\theta_i}(\boldsymbol{x}) = \boldsymbol{0}$, and $\psi_{\theta_i}(\boldsymbol{x}) = \frac{\mathrm{d}\log p(\boldsymbol{x};\theta)}{\mathrm{d}\theta_i}$. Note, that in Eq. (2), RP also has a $\frac{\mathrm{d}\phi(x)}{\mathrm{d}x}$ term, so it seems that it may also be described by the class of estimators in Eq. (4); however, there is a coordinate transformation to the ϵ-space, which may cause some confusion. In Sec. 2.2, we clear this confusion and show that, indeed, RP also belongs to the given class of estimators.

2.2 Coordinate Transformations

In Eq. (2), $\frac{\mathrm{d}\phi(g(\epsilon;\theta))}{\mathrm{d}g} = \frac{\mathrm{d}\phi(x)}{\mathrm{d}x}$ requires no reference to ϵ, and could be computed by just knowing the x corresponding to the ϵ. Moreover, each ϵ is always in a one-to-one correspondence with a particular x.[2] Therefore, the whole estimator $\frac{\mathrm{d}\phi(g(\epsilon;\theta))}{\mathrm{d}g}\frac{\mathrm{d}g(\epsilon;\theta)}{\mathrm{d}\theta}$ could be computed by directly sampling $x \sim p(x;\theta)$, converting it to the corresponding ϵ, and computing the estimator. We denote the mapping from x to ϵ with $\epsilon = S(x;\theta)$, and call S the *standardization function*. This function is the inverse of the RP transformation, $x = g(\epsilon;\theta)$, defined in the introduction. The standardization function was used in implicit reparameterization gradients (Figurnov et al., 2018) to create an estimator without reference to ϵ that is applicable to a broader class of distributions than typical reparameterizations. The main point we wanted to emphasize here is that there is no need to refer to coordinate transformations at all to define RP gradients, and the ϵ-space is just a convenience that makes it easier to apply RP using automatic differentiation. In particular, given a sample $x \sim p(x;\theta)$, the RP estimator can be written as $\frac{\mathrm{d}\phi(x)}{\mathrm{d}x}\left.\frac{\partial g(\epsilon;\theta)}{\partial\theta}\right|_{\epsilon=S(x;\theta)}$, where $\left.\frac{\partial g(\epsilon;\theta)}{\partial\theta_i}\right|_{\epsilon=S(x;\theta)}$ corresponds to $\boldsymbol{u}_{\theta_i}(\boldsymbol{x})$.

2.3 Importance Sampling/MC Integration

Definition 1 (Integral expression)**.** *An integral expression is denoted by $\int_\Omega f(x)\mathrm{d}x$, and it comprises the domain of integration Ω, the function $f(x)$, and the measure of integration corresponding to $\mathrm{d}x$.*

The reason we make such a seemingly trivial definition is to distinguish between the *integral expression* and the *value* of the integral. For example, the integral expressions corresponding to the LR gradient (Eq. (1)) and the RP gradient (Eq. (2)) are different, but they *evaluate* to the same quantity. Thus, there is a duality between the integrals; however, it is not clear how this duality arises. In Sec. 4 we explain that the integrals are duals under the divergence theorem. Next, we

[2] Here, we assume that g is invertible as is usually the case; however, this assumption can be easily lifted by integrating across the pre-image of x (App. E.1).

explain how these integral expressions are related to the gradient estimators via importance sampling.

MC integration: Any integral $\int f(x)\ \mathrm{d}x$ can be estimated using Monte Carlo integration as follows:

$$\begin{aligned}\int f(x)\ \mathrm{d}x &= \int q(x)\frac{f(x)}{q(x)}\mathrm{d}x \\ &= \mathbb{E}_{q(x)}\left[\frac{f(x)}{q(x)}\right],\end{aligned} \tag{5}$$

where we are importance sampling from $x \sim q(x)$. From this method, we see that the LR gradient estimator $E_{\mathrm{LR}} = \frac{\mathrm{d}p(x;\theta)}{\mathrm{d}\theta}\phi(x)\Big/q(x)$ arises by applying the MC integration principle to the integral expression $I_{\mathrm{LR}} = \int \frac{\mathrm{d}p(x;\theta)}{\mathrm{d}\theta}\phi(x)\mathrm{d}x$, when $q(x) = p(x;\theta)$. Moreover, the integral expression corresponding to RP is given by $I_{\mathrm{RP}} = \int p(x;\theta)\frac{\mathrm{d}\phi(x)}{\mathrm{d}x}\left.\frac{\partial g(\epsilon;\theta)}{\partial\theta}\right|_{\epsilon=S(x;\theta)}\mathrm{d}x$, and the gradient estimator for general $q(x)$ is $E_{\mathrm{RP}} = \frac{p(x;\theta)}{q(x)}\frac{\mathrm{d}\phi(x)}{\mathrm{d}x}\left.\frac{\partial g(\epsilon;\theta)}{\partial\theta}\right|_{\epsilon=S(x;\theta)}$. In general, given an integral expression for a gradient estimator, one can always construct the estimator by directly applying MC integration. But the reverse is also true—given an estimator, E, one can always construct the integral expression as $I = \int E\mathrm{d}q$, where the $\mathrm{d}q$ indicates that we are integrating w.r.t. the measure corresponding to q ($\mathrm{d}q$ can be considered as being equivalent to $q(x)\mathrm{d}x$). *Thus, there is a one-to-one correspondence between the estimator and the integral expression, given the sampling distribution, $q(x)$.* Previously, in machine learning (ML), importance sampling was suggested as a principle for LR (Jie and Abbeel, 2010), but the link to RP has not been discussed in ML. We on the other hand, suggest importance sampling as a key component of any gradient estimator, including RP.

2.4 Prior Work on the Relationship between LR and RP Gradient Estimators

Measure theoretic view: LR and RP gradients are well-studied in operations research (L'Ecuyer, 1991), where their relationship has been described in terms of measure theory (L'Ecuyer, 1990). They defined the problem as finding the gradient of an expectation of a function $\phi(\omega;\theta)$ where the expectation is taken w.r.t. a probability measure P_θ. Here, ω represents a sample in this space, and, unlike our previous definition, ϕ may also depend on θ. Then, by sampling w.r.t. a different probability measure G, independent of θ, on the same space, the expectation can be written as

$$\int \phi(\omega;\theta)\mathrm{d}P_\theta = \int \phi(\omega;\theta)\frac{\mathrm{d}P_\theta}{\mathrm{d}G}\mathrm{d}G = \int \phi(\omega;\theta)L_\theta\mathrm{d}G, \tag{6}$$

where $L_\theta = \frac{\mathrm{d}P_\theta}{\mathrm{d}G}$ is the Radon-Nikodym derivative, a function f, s.t. $P_\theta(A) = \int_A f\mathrm{d}G$, where $P_\theta(A)$ denotes the measure of set A. If $p(\omega;\theta)$ and $q(\omega)$ are the pdf's of P_θ and G respectively, then we simply have $L_\theta = \frac{p(\omega;\theta)}{q(\omega)}$ is the likelihood ratio. Differentiating w.r.t. θ gives

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\int \phi(\omega;\theta)L_\theta\mathrm{d}G = \int \frac{\mathrm{d}\phi(\omega;\theta)}{\mathrm{d}\theta}L_\theta + \phi(\omega;\theta)\frac{\mathrm{d}L_\theta}{\mathrm{d}\theta}\mathrm{d}G. \tag{7}$$

Now, depending on how the probability space P_θ is defined, one obtains either the likelihood ratio gradient or the reparameterization gradient. If $\omega := x$, then ϕ is independent of θ, so $\frac{\mathrm{d}\phi(\omega;\theta)}{\mathrm{d}\theta} = 0$, and one obtains the likelihood ratio gradient term $\frac{\mathrm{d}L_\theta}{\mathrm{d}\theta} = \frac{\mathrm{d}p(x;\theta)}{\mathrm{d}\theta}/q(x)$, which is the same as in Eq. (1), except that importance sampling from q is used (set $q = p$ to get exactly the LR gradient). On the other hand, if $\omega := \epsilon$, and $\phi(\epsilon;\theta)$ is defined as $\phi\left(g(\epsilon;\theta)\right)$, then L is independent of θ, and one is left with only the reparameterization gradient term $\frac{\mathrm{d}\phi(\epsilon;\theta)}{\mathrm{d}\theta}$, as in Eq. (2).

A strength of this view is that it shows that LR and RP lie at opposite ends of a spectrum of estimators using both derivative and value information of ϕ. However, the additional intuition is still limited, as the theory does not explain how these opposite ends are related, and how to convert between the two—the theory only says that if one *can* choose probability spaces with specific properties, one obtains either RP or LR, but it does not explain *how* to achieve the desired properties.

Stein's identity/integration by parts: Another work on policy gradients (Liu et al., 2017) showed a connection between RP and LR via Stein's identity:

$$\int p\left(x;\theta\right)\left(\frac{\mathrm{d}\log p\left(x;\theta\right)}{\mathrm{d}x}\phi(x) + \frac{\mathrm{d}\phi(x)}{\mathrm{d}x}\right)\mathrm{d}x = 0; \tag{8}$$

however, note that the derivative here is w.r.t. x, not θ. They showed algebraically that it generalizes to derivatives w.r.t. θ, but to do so, they put infinitesimal Gaussian noise on x, and the additional intuition from their work is still limited.

Ranganath et al. (2016) presented a derivation based on integration by parts, which can be seen as a generalized view compared to Stein's identity. We present this derivation and discuss it below.

Integration by parts is described by the identity:

$$\begin{aligned}&\int_a^b f(x)h(x)\mathrm{d}x \\ &= \left[\int_a^x f(z)\mathrm{d}z\ h(x)\right]_a^b - \int_a^b\int_a^x f(z)\mathrm{d}z\ \frac{\mathrm{d}h(x)}{\mathrm{d}x}\mathrm{d}x.\end{aligned} \tag{9}$$

We apply this identity on the integral for LR:

$$\int_{-\infty}^{+\infty} \frac{\mathrm{d}p(x;\theta)}{\mathrm{d}\theta}\phi(x)\mathrm{d}x = \left[\int_{-\infty}^{x} \frac{\mathrm{d}p(z;\theta)}{\mathrm{d}\theta}\mathrm{d}z\ \phi(x)\right]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty}\int_{-\infty}^{x} \frac{\mathrm{d}p(z;\theta)}{\mathrm{d}\theta}\mathrm{d}z\ \frac{\mathrm{d}\phi(x)}{\mathrm{d}x}\mathrm{d}x. \tag{10}$$

We can simplify as follows:

$$\int_{-\infty}^{x} \frac{\mathrm{d}p(z;\theta)}{\mathrm{d}\theta}\mathrm{d}z = \frac{\mathrm{d}}{\mathrm{d}\theta}\int_{-\infty}^{x} p(z;\theta)\mathrm{d}z = \frac{\mathrm{d}Q(x;\theta)}{\mathrm{d}\theta}, \tag{11}$$

where $Q(x;\theta)$ is the cumulative density function. The first term in Eq. (10) disappears because $\frac{\mathrm{d}Q(x;\theta)}{\mathrm{d}\theta} = 0$ at $x = \pm\infty$, and we end up with

$$\int p(x;\theta)\left(\frac{-1}{p(x;\theta)}\frac{\mathrm{d}Q(x;\theta)}{\mathrm{d}\theta}\right)\frac{\mathrm{d}\phi(x)}{\mathrm{d}x}\mathrm{d}x. \tag{12}$$

We can see that $\frac{-1}{p(x;\theta)}\frac{\mathrm{d}Q(x;\theta)}{\mathrm{d}\theta}$ is equivalent to $\boldsymbol{u}_{\theta_i}(\boldsymbol{x})$ in Eq. (4). [3] In the one-dimensional case, it turns out that this estimation method is the same as RP; however, the theory is still limited in several ways: (i) the derivation only considers one particular $\boldsymbol{u}_{\theta_i}(\boldsymbol{x})$ as opposed to an arbitrary one, (ii) in multiple dimensions, there are other RP gradients not conforming to this equation (Jankowiak and Obermeyer, 2018), (iii) the additional intuition from the derivation is limited—it appears to be just another "trick". It was suggested that the derivation is insightful, because the analytic computation of the zero term, $\left[\int_{-\infty}^{x} \frac{\mathrm{d}p(z;\theta)}{\mathrm{d}\theta}\mathrm{d}z\ \phi(x)\right]_{-\infty}^{+\infty} = 0$, is a reason for why RP has lower variance than LR (Ranganath et al., 2016; Cong et al., 2019). However, this argument is unsound because, on the contrary, adding a negatively correlated 0-mean *random* variable to the estimator, known as a control variate, is a common technique to reduce the variance (Greensmith et al., 2004a). Moreover, RP is not guaranteed to have lower variance than LR. For example, Parmas et al. (2018) showed a practical situation where LR is 10^6 more accurate than RP due to chaotic dynamics in the system. Other works showed toy problems where LR outperforms RP (Gal, 2016; Mohamed et al., 2019). Finally, it is unclear why analytically integrating a variable added to the integral expression should be related to the variance to begin with (as opposed to integrating a random variable in the *estimator*, a technique known as conditioning/Rao-Blackwellization (Owen, 2013)).

In conclusion, previous theories of the connection between RP and LR are still limited.

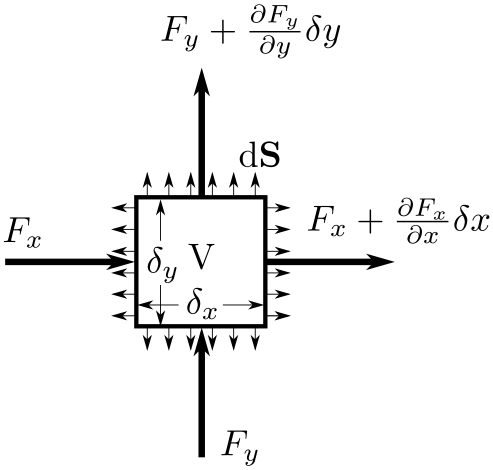

$F_y + \frac{\partial F_y}{\partial y}\delta y$
d**S**
F_x
$F_x + \frac{\partial F_x}{\partial x}\delta x$
δ_y
V
δ_x
F_y

Figure 1: Illustration of the divergence theorem.

2.5 Vector Calculus and Fluid Dynamics

Our unified theory in Sec. 4 relies on considering a "flow" of probability mass, so we give some background information. We illustrate the background in the 3-dimensional case, but it generalizes straightforwardly to higher dimensions.

Notation:
$\boldsymbol{F} = [F_x(x,y,z), F_y(x,y,z), F_z(x,y,z)]$ is a vector field.
$\phi(x,y,z)$ is a scalar field (a scalar function).
Divergence operator: $\nabla \cdot \boldsymbol{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}$.
Gradient operator: $\nabla\phi = \left[\frac{\partial \phi}{\partial x}, \frac{\partial \phi}{\partial y}, \frac{\partial \phi}{\partial z}\right]$.

The vector field $\boldsymbol{F}$ could be, for example, thought of as a local flow velocity of some fluid. If $\boldsymbol{F}$ is the density flow rate, then the divergence operator essentially measures how much the density is decreasing at a point. If the outflow is larger than the inflow, then the density decreases and vice versa. The divergence theorem, illustrated in Fig. 1, shows how this change in density can be measured in two equivalent ways: one could integrate the divergence across the volume, or one could integrate the inflow and outflow across the surface.

Theorem 1 (Divergence theorem)**.**

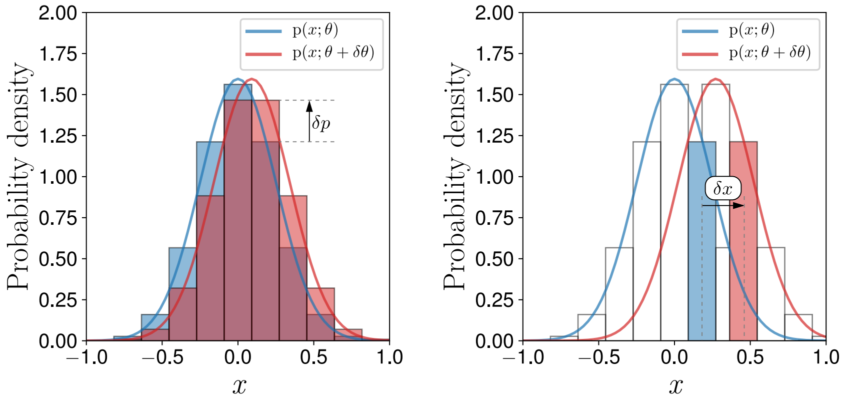
$$\int_V \nabla \cdot \boldsymbol{F}\mathrm{d}V = \int_S \boldsymbol{F}\cdot\mathrm{d}\boldsymbol{S}. \tag{13}$$

Proof. To prove the claim, consider the infinitesimal box in Fig. 1. The divergence can be calculated as

$$\int_V \nabla \cdot \boldsymbol{F}\mathrm{d}V = \delta x\delta y\left(\frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y}\right). \tag{14}$$

On the other hand, to take the integral across the surface, note that the surface normals point outwards,

[3] Note that substituting $\theta = x$ leads to $\frac{\mathrm{d}Q(x;\theta)}{\mathrm{d}x} = p(x;\theta)$, and the equation becomes Stein's identity in Eq. (8), showing that integration by parts generalizes it.



(a) LR probability "boxes" (b) RP probability "boxes"

Figure 2: LR keeps the boundaries of the "boxes" fixed, while RP keeps the probability mass fixed.

and the integral becomes

$$\begin{aligned}\int_S \boldsymbol{F}\cdot \mathrm{d}\boldsymbol{S} &= \delta_y\left(-F_x + F_x - \frac{\partial F_x}{\partial x}\delta x\right)\\ &\quad + \delta x\left(-F_y + F_y + \frac{\partial F_y}{\partial y}\delta y\right) \qquad (15)\\ &= \delta x \delta y\left(\frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y}\right),\end{aligned}$$

which is the same as the divergence. To generalize this to arbitrarily large volumes, notice that if one stacks the boxes next to each other, then the surface integral across the area where the boxes meet cancels out, and only the integral across the outer surface remains. □

For an incompressible flow, the density at any point does not change, and the divergence must be zero.

3 A PROBABILITY "BOXES" VIEW OF LR AND RP GRADIENTS

Here we give our first explanation of the link between LR and RP gradients, illustrated in Fig. 2. In short, LR gradients estimate the change in expectation by measuring how the probability mass assigned to each $\phi(x)$ located at a fixed x changes, whereas RP gradients define "boxes" of fixed probability mass, keep track of where this "box" moves as the parameters θ change, and measure how the value ϕ corresponding to this "box" changes. For the ease of the explanation, consider a discrete space, where $x \sim P(x;\theta)$ can take N possible values. The continuous case can be recovered by letting $N \to \infty$. Fundamentally, the expectation, $\mathbb{E}_{P(x;\theta)}\left[\phi(x)\right]$, is a weighted average of function values $\sum_{i=1}^{N} P(x_i)\phi(x_i)$, where the weights sum to one: $\sum_{i=1}^{N} P(x_i) = 1$. Therefore, to determine the expectation, we must determine *how the probability mass is allocated to the different $\phi(x_i)$ values.* We can envision two different allocation procedures: (i) for each $\phi(x_i)$ at a *fixed* x_i, we determine how much probability mass P_i we assign to it; (ii) we predetermine the sizes of the "boxes" of *fixed* probability mass P_j, then, for each box with weight P_j we assign one of the available $\phi(x_i)$ values. Now, to measure the gradient of the expectation, $\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_{P(x;\theta)}\left[\phi(x)\right]$, *one must measure how the probability mass is reallocated* as the parameters θ are perturbed. We will see that allocation procedure (i) corresponds to LR gradients, whereas (ii) corresponds to RP gradients. A full formal derivation is given in App. C, but reading it should not be necessary to understand the concept, which we explain intuitively below. To perform the estimation, first note that the gradient is given by $\frac{\mathrm{d}}{\mathrm{d}\theta}\sum_{i=1}^{N} P(x_i)\phi(x_i) = \sum_{i=1}^{N}\frac{\mathrm{d}}{\mathrm{d}\theta}\big(P(x_i)\phi(x_i)\big)$.

LR estimator: In case (i): $\phi(x_i)$ is fixed, so $\frac{\mathrm{d}}{\mathrm{d}\theta}\big(P(x_i)\phi(x_i)\big) = \frac{\mathrm{d}P(x_i)}{\mathrm{d}\theta}\phi(x_i)$, and to estimate the gradient, *one must measure how the weight assigned to each particular $\phi(x_i)$ changes.* This corresponds precisely to what the LR gradient estimator does. To see this, first consider that any integral can be estimated by importance sampling from a distribution $q(x)$, and using MC integration, as shown in Eq. (5). Now, we set $q(x) = P(x;\theta)$, sample $x_i \sim P(x;\theta)$, and use the gradient estimator $E = \frac{1}{P(x_i;\theta)}\frac{\mathrm{d}P(x_i)}{\mathrm{d}\theta}\phi(x_i)$. Then this will satisfy $\mathbb{E}_{x_i\sim P(x;\theta)}\left[E\right] = \sum_{i=1}^{N}\frac{\mathrm{d}}{\mathrm{d}\theta}\big(P(x_i)\phi(x_i)\big)$. Note that $\frac{1}{P(x_i;\theta)}\frac{\mathrm{d}P(x_i;\theta)}{\mathrm{d}\theta} = \frac{\mathrm{d}}{\mathrm{d}\theta}\log P(x_i;\theta)$, and we see that it is the same as the LR gradient in Eq. (1). The transformation to the log term is known as the log-derivative trick, and it may appear to be the essence behind the LR gradient. However, actually the multiplication and division by $P(x;\theta)$ is just a special case of the more general MC integration principle. Rather than thinking of the LR gradient in terms of the log-derivative term, it may be better to think of it as simply estimating the integral of the probability gradient by applying the appropriate importance weights. Sometimes, the LR gradient is described as being "kind of like a finite difference gradient" (Salimans et al., 2017; Mania et al., 2018), but here we see that it is a different concept that does not rely on fitting a straight line between differences of ϕ (App. B), but estimates how probability mass is reallocated among different ϕ values.

RP estimator: In case (ii): $P(x_i)$ is fixed, but $\phi(x_i)$ may change—such a situation can occur when one has a fixed amount of probability mass P_i in the "box", but the location, x_i, changes. In this case, we have $\frac{\mathrm{d}}{\mathrm{d}\theta}\big(P(x_i)\phi(x_i)\big) = P(x_i)\frac{\mathrm{d}\phi(x_i)}{\mathrm{d}\theta} = P(x_i)\frac{\mathrm{d}\phi(x_i)}{\mathrm{d}x_i}\frac{\mathrm{d}x_i}{\mathrm{d}\theta}$, and to estimate the gradient, *one must measure how the function value ϕ in the "box" changes.* For example, consider shifting the mean location of a Gaussian distribution by $\delta\mu$, hence, also shifting the location of each of the "boxes" by the same quantity, as depicted in Fig. 2. The probability inside the box would stay fixed, but the function value ϕ would change. This situation corresponds to the RP gradient in Eq. (2). In this case, the position of the "box" is defined by $x_i := g(\epsilon_i;\theta)$, and the probability density assigned to ϵ_i stays fixed at $p(\epsilon_i)$. Finally, note that we can construct an estimator $E = \frac{\mathrm{d}\phi(x_i)}{\mathrm{d}x_i}\frac{\mathrm{d}x_i}{\mathrm{d}\theta}$ by sampling from $x_i \sim P(x_i)$, and this will be unbiased: $\mathbb{E}_{x_i\sim P(x;\theta)}[E] = \sum_{i=1}^{N}\frac{\mathrm{d}}{\mathrm{d}\theta}\big(P(x_i)\phi(x_i)\big)$.

We see that LR and RP are estimating the same quantity; the difference lies just in the way how one keeps track of the movement of the probability mass: LR measures how the probability mass assigned to a fixed location x_i changes, whereas RP measures how the function value ϕ corresponding to a moving "box" of probability mass changes.

4 A UNIFIED PROBABILITY FLOW VIEW OF LR AND RP GRADIENTS

Here we give another explanation of LR and RP. In this theory, both LR and RP come out of the same derivation, thus showing a link between the two. In particular, we define an incompressible flow of probability mass imposed by perturbing the parameters θ of $p(x;\theta)$, which can be used to express the derivative of the expectation as an integral over this flow. LR and RP estimators correspond to duals of this integral under the well-known divergence theorem (Thm. 1).

The main idea resembles RP, but in addition to sampling $\boldsymbol{x}$, we sample a height h for each point: $h = \epsilon_h p(\boldsymbol{x};\theta)$, where $\epsilon_h \sim \mathrm{unif}(0,1)$, i.e., the sampling space is extended with an additional dimension for the height $\tilde{\boldsymbol{x}} := [\boldsymbol{x}^T, h]^T$, and we are uniformly sampling in the volume under $p(\boldsymbol{x};\theta)$. The definition of g in the introduction is extended, s.t. $\tilde{g}(\epsilon_x,\epsilon_h) := \tilde{\boldsymbol{x}} = [g(\epsilon_x)^T, \epsilon_h p(\boldsymbol{x};\theta)]^T$. The expectation turns into

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}\theta}\int p(\boldsymbol{x};\theta)\,\phi(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\\ &=\frac{\mathrm{d}}{\mathrm{d}\theta}\int_{\epsilon_x}\int_{\epsilon_h} p(\epsilon_x)\,p(\epsilon_h)\,\phi(\tilde{g}(\epsilon_x,\epsilon_h))\,\mathrm{d}\epsilon_x\mathrm{d}\epsilon_h\\ &=\int_V \nabla_\theta\phi(\tilde{g}(\epsilon_x,\epsilon_h))\mathrm{d}V = \int_V \nabla_{\tilde{\boldsymbol{x}}}\phi(\tilde{\boldsymbol{x}})\nabla_\theta\tilde{g}(\epsilon_x,\epsilon_h)\mathrm{d}V.\end{aligned} \tag{16}$$

In Eq. (16), V is the volume under the curve of $p(\boldsymbol{x};\theta)$, and $\phi([\boldsymbol{x}^T,h]^T) := \phi(\boldsymbol{x})$ ignores the h-component. Each column i of $\nabla_\theta\tilde{g}(\epsilon_x,\epsilon_h)$ corresponds to a vector field induced by perturbing the i^{th} component of θ. The red lines in Fig. 3 show the flow fields for a Gaussian distribution as the mean and variance are perturbed. The other term, $\nabla_{\tilde{\boldsymbol{x}}}\phi(\tilde{\boldsymbol{x}})$, is the gradient of the scalar field $\phi(\tilde{\boldsymbol{x}})$. As ϕ is independent of h, the gradient is parallel to the $\boldsymbol{x}$ axes with magnitude $\frac{\mathrm{d}\phi}{\mathrm{d}\boldsymbol{x}}$.

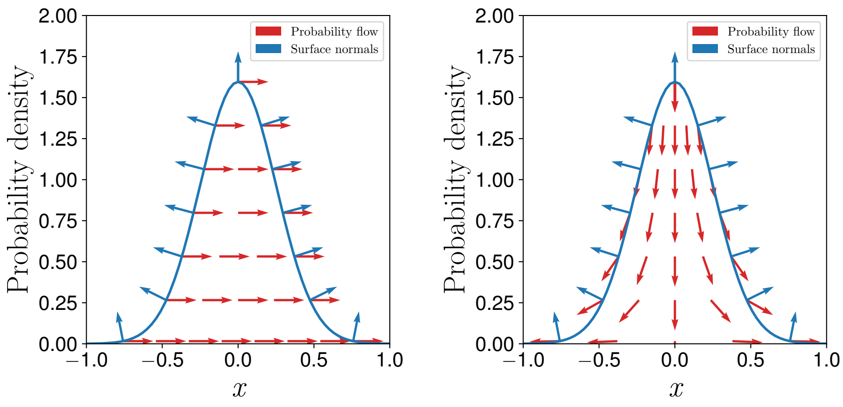
According to the divergence theorem in Eq. (13), the *volume integral* in Eq. (16) can be turned into a *surface integral* over the boundary $\boldsymbol{S}$ ($\mathrm{d}\boldsymbol{S}$ is a shorthand for $\hat{\boldsymbol{n}}\mathrm{d}S$, where $\hat{\boldsymbol{n}}$ is the surface normal vector), depicted by the blue lines in Fig. 3.

In Eq. (13), $\boldsymbol{F}$ is any vector field. A common corollary arises by picking $\boldsymbol{F} = \phi\boldsymbol{v}$, where ϕ is a scalar field, and $\boldsymbol{v}$ is a vector field. We choose $\boldsymbol{v} = \nabla_\theta\tilde{g}(\epsilon_x,\epsilon_h)\delta\theta$, where $\delta\theta$ is an arbitrary perturbation in θ, so that $\boldsymbol{F} = \phi(\tilde{\boldsymbol{x}})\nabla_\theta\tilde{g}(\epsilon_x,\epsilon_h)\delta\theta$, in which case $\nabla_{\tilde{\boldsymbol{x}}}\cdot\boldsymbol{F} = \nabla_{\tilde{\boldsymbol{x}}}\cdot(\phi(\tilde{\boldsymbol{x}})\nabla_\theta\tilde{g}(\epsilon_x,\epsilon_h)\delta\theta) = \nabla_{\tilde{\boldsymbol{x}}}\phi(\tilde{\boldsymbol{x}})\nabla_\theta\tilde{g}(\epsilon_x,\epsilon_h)\delta\theta + \phi(\tilde{\boldsymbol{x}})\nabla_{\tilde{\boldsymbol{x}}}\cdot\nabla_\theta\tilde{g}(\epsilon_x,\epsilon_h)\delta\theta$. Note that the term $\nabla_\theta\tilde{g}(\epsilon_x,\epsilon_h)\delta\theta$ corresponds to an incompressible flow (because the probability density does not change at any point in the augmented space). As the divergence of an incompressible flow is 0, $\nabla_{\tilde{\boldsymbol{x}}}\cdot\nabla_\theta\tilde{g}(\epsilon_x,\epsilon_h)\delta\theta = 0$, and the second term disappears. Noting that $\delta\theta$ can be canceled, because it is arbitrary, we are left with the equation:

$$\int_V \nabla_{\tilde{\boldsymbol{x}}}\phi(\tilde{\boldsymbol{x}})\nabla_\theta\tilde{g}(\epsilon_x,\epsilon_h)\mathrm{d}V = \int_S \phi(\tilde{\boldsymbol{x}})\nabla_\theta\tilde{g}(\epsilon_x,\epsilon_h)\cdot\mathrm{d}\boldsymbol{S}. \tag{17}$$

Now we explain how the left-hand side of Eq. (17) gives rise to the RP gradient estimator, while the right-hand side corresponds to the LR gradient estimator.

RP estimator: Consider the $\nabla_{\tilde{\boldsymbol{x}}}\phi(\tilde{\boldsymbol{x}})\nabla_\theta\tilde{g}(\epsilon_x,\epsilon_h)$ term. As the scalar field $\phi(\tilde{\boldsymbol{x}})$ is independent of the height location h, the component of the gradient in



(a) μ flow lines

(b) σ flow lines

Figure 3: Probability flow lines when μ and σ are perturbed.

that direction is 0, and $\nabla_{\tilde{\boldsymbol{x}}}\phi(\tilde{\boldsymbol{x}}) = [\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x}), 0]$. As the h-component is 0, the value of $\tilde{g}$ in the h-direction is multiplied by 0, and is irrelevant for the product, so $\nabla_{\tilde{\boldsymbol{x}}}\phi(\tilde{\boldsymbol{x}})\nabla_{\theta}\tilde{g}(\epsilon_x, \epsilon_h) = \nabla_{\boldsymbol{x}}\phi(\boldsymbol{x})\nabla_{\theta}g(\epsilon_x)$, which is just the term used in the RP estimator. Hence, the left-hand side of Eq. (17) corresponds to the RP gradient.

LR estimator: We will show that the LR estimator tries to integrate $\int_S \phi(\tilde{\boldsymbol{x}})\nabla_{\theta}\tilde{g}(\epsilon_x, \epsilon_h) \cdot \mathrm{d}\boldsymbol{S}$. To do so, note that $\mathrm{d}\boldsymbol{S} = \hat{\boldsymbol{n}}\mathrm{d}S$. It is necessary to express the normalized surface vector $\hat{\boldsymbol{n}}$, and then perform the integral over the surface. The derivation is in App. D.2, and the final result is

$$\int_S \phi(\tilde{\boldsymbol{x}})\nabla_{\theta}\tilde{g}(\epsilon_x, \epsilon_h) \cdot \mathrm{d}\boldsymbol{S} = \int_X \phi(\boldsymbol{x})\frac{\mathrm{d}p\left(\boldsymbol{x};\theta\right)}{\mathrm{d}\theta}\,\mathrm{d}\boldsymbol{x}. \tag{18}$$

We have already seen that MC integration of the right-hand side of Eq. (18) using samples from $p\left(\boldsymbol{x};\theta\right)$ yields the LR estimator. Thus, RP and LR are duals under the divergence theorem. To further strengthen this claim we prove that the LR gradient estimator is the unique estimator that takes weighted averages of the function values $\phi(\boldsymbol{x})$.

Theorem 2 (Uniqueness of LR estimator)**.** *$\psi(\boldsymbol{x}) = p\left(\boldsymbol{x};\theta\right)\frac{\mathrm{d}\log p(\boldsymbol{x};\theta)}{\mathrm{d}\theta}$ is the unique function ψ, s.t. $\int \psi(\boldsymbol{x})\phi(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \frac{\mathrm{d}}{\mathrm{d}\theta}\int p\left(\boldsymbol{x};\theta\right)\phi(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$ for all ϕ.*

Proof. Suppose that there exist $\psi \neq f$, s.t. $\int \phi(\boldsymbol{x})\psi(\boldsymbol{x})\ \mathrm{d}\boldsymbol{x} = \int \phi(\boldsymbol{x})f(\boldsymbol{x})\ \mathrm{d}\boldsymbol{x}$ for all ϕ. Rearrange the equation into $\int \phi(\boldsymbol{x})\left(\psi(\boldsymbol{x}) - f(\boldsymbol{x})\right)\ \mathrm{d}\boldsymbol{x} = 0$, then pick $\phi(\boldsymbol{x}) = \psi(\boldsymbol{x}) - f(\boldsymbol{x})$ from which we get $\int \left(\psi(\boldsymbol{x}) - f(\boldsymbol{x})\right)^2\ \mathrm{d}\boldsymbol{x} = 0$. This leads to $\psi = f$, which is a contradiction. Therefore, there cannot exist such $\psi \neq f$ that satisfy the condition for all ϕ. □

The result also follows from the Riesz representation theorem (Riesz, 1907). From the theorem, we see that Eq. (18) was immediately clear without having to go through the derivation in App. D.2. The same analysis does not work for RP (App. E.1). Indeed, there are infinitely many RP gradients (Jankowiak and Obermeyer, 2018).

Characterizing the space of all LR and RP estimators: Now we can derive a *concrete* form of all estimators in the *abstract* class in Eqs. (4) and (7). The flow theory assumed that the flow is aligned with the change of the probability density. We can lift this restriction by subtracting the excess probability mass (App. E.2), giving a general gradient estimator combining both $\phi(\boldsymbol{x})$ and $\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x})$. This characterization is formalized in the theorem below.

Theorem 3 (The probability flow gradient estimator characterizes the space of all LR–RP gradient estimators)**.** *Given a sample, $\boldsymbol{x} \sim q(\boldsymbol{x})$, every unbiased gradient estimator, E_{θ_i}, s.t. $\mathbb{E}_{q(\boldsymbol{x})}\left[E_{\theta_i}\right] = \frac{\mathrm{d}}{\mathrm{d}\theta_i}\mathbb{E}_{p(\boldsymbol{x};\theta)}\left[\phi(\boldsymbol{x})\right]$, of the product form in Eq. (4),*

$$E_{\theta_i} = \boldsymbol{v}(\boldsymbol{x}) \cdot \nabla_{\boldsymbol{x}}\phi(\boldsymbol{x}) + \psi(\boldsymbol{x})\phi(\boldsymbol{x}),$$

where ϕ is an arbitrary function, and assuming[4] *$p(\boldsymbol{x};\theta)\boldsymbol{v}(\boldsymbol{x})\phi(\boldsymbol{x}) \to 0$ as $\|x\| \to \infty$, is a special case of*

[4] Note that the case where $p\left(\boldsymbol{x};\theta\right)\phi(\boldsymbol{x})\boldsymbol{v}(\boldsymbol{x}) \not\to 0$ does not correspond to any sensible estimator, because the value of $\phi(\boldsymbol{x})$ at $\|\boldsymbol{x}\| \to \infty$ will influence the gradient estimation. In that case, because $p\left(\boldsymbol{x};\theta\right) \to 0$ the probability of sampling at infinity will tend to 0, and the gradient variance will explode. This condition does however mean that if one wants to construct a sensible estimator, care must be taken to ensure that $\boldsymbol{u}_{\theta_i}(\boldsymbol{x})$ does not go to infinity too fast, e.g., as explained by Jankowiak and Karaletsos (2019).

the estimator characterized by

$$E_{\theta_i} = \frac{p(\boldsymbol{x};\theta)}{q(\boldsymbol{x})}\boldsymbol{u}_{\theta_i}(\boldsymbol{x}) \cdot \nabla_{\boldsymbol{x}}\phi(\boldsymbol{x}) + \frac{1}{q(\boldsymbol{x})}\left(\nabla_{\boldsymbol{x}} \cdot \big(p(\boldsymbol{x};\theta)\,\boldsymbol{u}_{\theta_i}(\boldsymbol{x})\big) + \frac{\mathrm{d}p(\boldsymbol{x};\theta)}{\mathrm{d}\theta_i}\right)\phi(\boldsymbol{x}), \tag{19}$$

where $\boldsymbol{u}_{\theta_i}$ is an arbitrary vector field. Note that, for simplicity, we also assumed continuity of p, ϕ and $\boldsymbol{u}_{\theta_i}$; however, this is unnecessary, and discontinuities are handled in App. E.4.

Proof. It is analogous to Thm 2. See App. E.2. □

The theorem says that given $\boldsymbol{v}$, one can derive the unique ψ necessary for unbiasedness. Thus, the probability flow gradient in Eq. (19) generalizes all previous LR–RP gradients in the literature, as well as all possible gradient estimators having the product form in Eq. (4).[5] By setting $\boldsymbol{u}_{\theta_i}(\boldsymbol{x}) = \mathbf{0}$ one recovers LR; by setting $\nabla_{\boldsymbol{x}} \cdot \big(p(\boldsymbol{x};\theta)\,\boldsymbol{u}_{\theta_i}(\boldsymbol{x})\big) + \frac{\mathrm{d}p(\boldsymbol{x};\theta)}{\mathrm{d}\theta_i} = 0$, one recovers the pathwise estimators described by Jankowiak and Obermeyer (2018).[6] Estimators combining both $\phi(\boldsymbol{x})$ and $\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x})$, such as the generalized RP gradient (Ruiz et al., 2016b), also conform to Eq. (19) (App. E.3).[7] Moreover, estimators in discontinuous situations, such as the GO gradient (Cong et al., 2019) or RP gradients for discontinuous models (Lee et al., 2018) also conform to this equation when taking into account for the discontinuities (App. E.4).

The terms in Eq. (19) can be readily interpreted. First note that $q(\boldsymbol{x})$ is just a factor due to MC integration by sampling from q (Sec. 2.3). The remaining terms can be made analogous to fluid motion. In the analogy, perturbing θ_i is equivalent to perturbing time measured in seconds (s), $p(\boldsymbol{x};\theta)$ is equivalent to the density measured in $\frac{\mathrm{kg}}{\mathrm{m}^3}$, and $\boldsymbol{u}_{\theta_i}$ is equivalent to the flow velocity measured in $\frac{\mathrm{m}}{\mathrm{s}}$. The term, $p(\boldsymbol{x};\theta)\boldsymbol{u}_{\theta_i}(\boldsymbol{x})$, is the probability mass flow rate per unit area, as is clear from multiplying the units: $\frac{\mathrm{kg}}{\mathrm{m}^3} \times \frac{\mathrm{m}}{\mathrm{s}} = \frac{\mathrm{kg/s}}{\mathrm{m}^2}$. The term $\nabla_{\boldsymbol{x}} \cdot \big(p(\boldsymbol{x};\theta)\,\boldsymbol{u}_{\theta_i}(\boldsymbol{x})\big)$ is the divergence of the probability mass flow rate, and it tells us the rate of change of density at $\boldsymbol{x}$ with a corresponding $\phi(\boldsymbol{x})$ caused by the probability flow, $p(\boldsymbol{x};\theta)\boldsymbol{u}_{\theta_i}(\boldsymbol{x})$ (see Sec. 2.5). The $p(\boldsymbol{x};\theta)\,\boldsymbol{u}_{\theta_i}(\boldsymbol{x}) \cdot \nabla_{\boldsymbol{x}}\phi(\boldsymbol{x})$ term, on the other hand, gives the rate of change of the $\phi(\boldsymbol{x})$ corresponding to a point moving on the probability flow. Integrated across the whole volume, the two terms involving $\boldsymbol{u}_{\theta_i}$ cancel out, leaving only the $\frac{\mathrm{d}p(\boldsymbol{x};\theta)}{\mathrm{d}\theta_i}$ term. The probability flow estimator could thus also be interpreted as adding a control variate to the standard LR gradient estimator.

Our explanation of the principle behind LR and RP gradients improves over previous explanations based on two main criteria: (i) greater generality, (ii) requiring fewer assumptions. In particular, Occam's razor states that given competing explanations, one should prefer the one with fewer assumptions. Our explanation does not require the reparameterization assumption used in many previous explanations of RP gradients. Instead, we argue that reparameterization is just a trick that allows implementing the estimator easily using automatic differentiation, but has little to do with the principle behind its operation. Moreover, the sufficient conditions for the probability flow gradient estimator are also necessary, so the explanation of the principle cannot be improved without expanding the class of estimators to go beyond Eq. (4).

Our characterization showed that the flow field $\boldsymbol{u}_{\theta_i}(\boldsymbol{x})$ and importance sampling distribution $q(\boldsymbol{x})$, together, fully describe the space of estimators; one remaining question is how to pick $\boldsymbol{u}_{\theta_i}(\boldsymbol{x})$ and $q(\boldsymbol{x})$, s.t. the variance of the estimator is low. Jankowiak and Obermeyer (2018) discussed how to pick $\boldsymbol{u}_{\theta_i}(\boldsymbol{x})$ when the $\phi(\boldsymbol{x})$ term disappears (i.e., for RP). In our concurrent work (Parmas and Sugiyama, 2019), we are discussing how to pick $q(\boldsymbol{x})$ when the $\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x})$ term disappears (i.e., for LR). The general question of the best combination of $\boldsymbol{u}_{\theta_i}(\boldsymbol{x})$ and $q(\boldsymbol{x})$ remains an open problem.

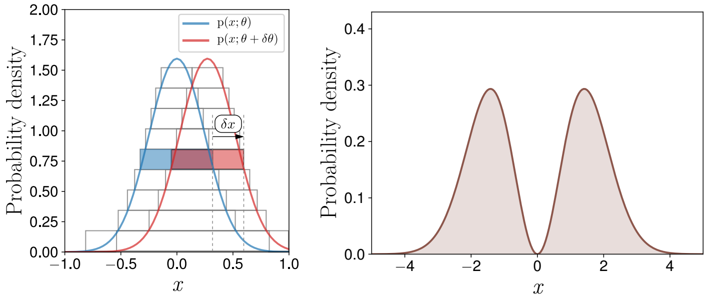
5 BENEFIT OF CHARACTERIZING THE SPACE OF ESTIMATORS

Characterizing the space of estimators via uniqueness claims is highly useful because it clarifies where we should be searching for new gradient estimators. In particular, often a new idea might appear promising at first sight, but uniqueness claims could immediately say that the idea will not lead to something novel, and will instead be a special case of the characterized space.

[5] Note, there still exist other gradient estimators that do not have the form $\boldsymbol{u}(\boldsymbol{x}) \cdot \nabla_{\boldsymbol{x}}\phi(\boldsymbol{x}) + \psi(\boldsymbol{x})\phi(\boldsymbol{x})$, e.g., gradient estimators with coupled samples (Walder et al., 2019; Mohamed et al., 2019), or gradient estimators using $\frac{\mathrm{d}^2\phi(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}^2}$.

[6] The work of Jankowiak and Obermeyer (2018) was concurrent to our initial derivations, and is the most similar publication to ours. They also used the divergence theorem, but they focused on deriving new pathwise estimators, and did not discuss the duality between LR and RP.

[7] Note that it is an open question whether the generalized RP and probability flow gradient estimator spaces are equal. To show that they are equal, one would have to find a generalized reparameterization corresponding to each arbitrary $\boldsymbol{u}_{\theta_i}(\boldsymbol{x})$. One main difficulty would arise if one requires a single reparameterization to simultaneously correspond to multiple different $\boldsymbol{u}_{\theta_i}(\boldsymbol{x})$ and $\boldsymbol{u}_{\theta_j}(\boldsymbol{x})$ for different dimensions i and j of the parameter vector θ. However, we believe that if finding such reparameterization is possible at all, the reparameterization corresponding to some complicated flow field may be quite bizarre, while in the flow framework, one just has to do a dot product between the flow and the gradient to compute the estimator.



(a) Probability slices

(b) Probability density

Figure 4: Motivation for slice integral sampling and the importance distribution for a Gaussian base distribution.

We illustrate this concept with two case studies below.

Case study 1 (Height reparameterization)**.** *Consider the example with a height reparameterization explained in Sec. 4. After some derivations, we reached Eq. (17), which we repeat below:*

$$\int_V \nabla_{\tilde{\boldsymbol{x}}}\phi(\tilde{\boldsymbol{x}})\nabla_\theta \tilde{g}(\epsilon_x, \epsilon_h)\mathrm{d}V = \int_S \phi(\tilde{\boldsymbol{x}})\nabla_\theta \tilde{g}(\epsilon_x, \epsilon_h) \cdot \mathrm{d}\boldsymbol{S}.$$

This equation looks promising because the left-hand side is known to correspond to RP, which is an unbiased gradient estimator. Therefore, we know that the right-hand side must also be unbiased, and it could potentially lead to some new interesting estimator using the $\phi(\boldsymbol{x})$ information. However, to compute the estimator, we have to perform the tedious error-prone derivations in App. D.2. Instead, based on the uniqueness claim in Thm. 2 we can immediately say that this approach cannot possibly lead to a new estimator, because it has the same form as LR (product with $\phi(\boldsymbol{x})$), saving us the trouble of going through the derivation.

Case study 2 (Horizontal slice integral sampling)**.** *Consider an algorithm that would sample horizontal slices of probability mass, illustrated in Fig. 4a, and motivated by flipping the vertical slices in Sec. 3. One could integrate the expectation over the slice, and try to estimate the gradient w.r.t. a parameter θ.*

Such an approach appears attractive, because if the location of the slice is moved by modifying the parameters of the distribution (e.g., by changing the mean, μ), then the derivative of the expected value of the integral over the slice will depend only on the value at the edges of the slice (because the probability density in the middle would not change). To clarify, consider a uniform distribution $p(x;\mu)$ between $\mu \pm \Delta$. The derivative is $\frac{\mathrm{d}}{\mathrm{d}\mu}\int_{\mu-\Delta}^{\mu+\Delta} p(x;\mu)\phi(x)\mathrm{d}x = C\left(\phi(\mu+\Delta) - \phi(\mu-\Delta)\right)$, where C is a constant. We could use importance sampling to sample on one of the two edges of the slice, to obtain an unbiased gradient estimator. However, at this point, it is clear that the estimator will belong to the same class as LR, as it will be averaging some value multiplied with $\phi(x)$, so we can already say that this idea is not promising. At best, it would lead to an LR gradient with a different sampling distribution $q(x)$—this is indeed the case, and the distribution is plotted in Fig. 4b. The full derivation is in App. F.

Systematic approach to deriving estimators: Rather than pursuing an ad hoc approach as in the two case studies, we propose that we should be using a more systematic approach in the search for new gradient estimators. It is not enough to find *one* novel estimator; one should find *all* estimators of a given novel class. Our proposed 3-step approach is below.

1. Find a new principle for a novel gradient estimator.
In the case of LR and RP, the principle is to measure the movement of probability mass, as described in Sec. 3.

2. Parameterize the class of estimators that encloses all estimators embodying the said principle.
In our case, this parameterization is in Eq. (4).

3. Find necessary and sufficient conditions for the estimator to be unbiased.
In our case, these conditions were given in Sec. 4.

6 CONCLUSIONS

We introduced a complete unified theory of LR and RP gradients, and characterized the space of all unbiased single sample gradient estimators taking a weighted sum of $\phi(\boldsymbol{x})$ and $\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x})$. Each estimator is defined by a vector field $\boldsymbol{u}_{\theta_i}(\boldsymbol{x})$ and an importance sampling distribution $q(\boldsymbol{x})$ that represent two "knobs" one can tune to improve gradient accuracy. We hope our work may lead to a systematic pursuit to characterizing all possible gradient estimators based on different principles of Monte Carlo gradient estimation.

Acknowledgements

PP was supported by KAKENHI Grants 16H06561, 16H06563 and 16K21738, by the research support of the Okinawa Institute of Science and Technology Graduate University for the Neural Computation Unit, by RIKEN during an internship, and by the Cyborg-AI project supported by NEDO, Japan. MS was supported by KAKENHI 17H00757.

References

Abel, N. H. (1895). *Untersuchungen über die Reihe: 1+(m/1) x+ m·(m-1)/(1· 2)· x2+ m·(m-1)·(m-2)/(1· 2· 3)· x3+...* Number 71. W. Engelmann. E.4

Asadi, K., Allen, C., Roderick, M., Mohamed, A.-r., Konidaris, G., and Littman, M. (2017). Mean actor critic. *stat*, 1050:1. A

Ciosek, K. and Whiteson, S. (2018). Expected policy gradients. In *Thirty-Second AAAI Conference on Artificial Intelligence*. A

Cong, Y., Zhao, M., Bai, K., and Carin, L. (2019). Go gradient for expectation-based objectives. *arXiv preprint arXiv:1901.06020*. 2.4, 4, E.4, E.4, E.4, E.4

Conti, E., Madhavan, V., Such, F. P., Lehman, J., Stanley, K., and Clune, J. (2018). Improving exploration in evolution strategies for deep reinforcement learning via a population of novelty-seeking agents. In *Advances in Neural Information Processing Systems*, pages 5027–5038. 1

Farquhar, G., Whiteson, S., and Foerster, J. (2019). Loaded dice: Trading off bias and variance in any-order score function estimators for reinforcement learning. *arXiv preprint arXiv:1909.10549*. A

Figurnov, M., Mohamed, S., and Mnih, A. (2018). Implicit reparameterization gradients. In *Advances in Neural Information Processing Systems*, pages 441–452. 2.2, A, E.3, E.3

Foerster, J., Farquhar, G., Al-Shedivat, M., Rocktäschel, T., Xing, E., and Whiteson, S. (2018). Dice: The infinitely differentiable Monte Carlo estimator. In *International Conference on Machine Learning*, pages 1529–1538. A

Gal, Y. (2016). *Uncertainty in deep learning*. PhD thesis, PhD thesis, University of Cambridge. 2.4, A

Geffner, T. and Domke, J. (2018). Using large ensembles of control variates for variational inference. In *Advances in Neural Information Processing Systems*, pages 9960–9970. A

Glynn, P. W. (1990). Likelihood ratio gradient estimation for stochastic systems. *Communications of the ACM*, 33(10):75–84. 1

Grathwohl, W., Choi, D., Wu, Y., Roeder, G., and Duvenaud, D. (2017). Backpropagation through the void: Optimizing control variates for black-box gradient estimation. *arXiv preprint arXiv:1711.00123*. A

Greensmith, E., Bartlett, P. L., and Baxter, J. (2004a). Variance reduction techniques for gradient estimates in reinforcement learning. *Journal of Machine Learning Research*, 5(Nov):1471–1530. 2.4

Greensmith, E., Bartlett, P. L., and Baxter, J. (2004b). Variance reduction techniques for gradient estimates in reinforcement learning. *Journal of Machine Learning Research*, 5(Nov):1471–1530. A

Grinfeld, P. (2013). *Introduction to tensor analysis and the calculus of moving surfaces*. Springer. E.4

Gu, S., Levine, S., Sutskever, I., and Mnih, A. (2015). MuProp: Unbiased backpropagation for stochastic neural networks. *arXiv preprint arXiv:1511.05176*. A

Gu, S., Lillicrap, T., Ghahramani, Z., Turner, R. E., and Levine, S. (2016). Q-prop: Sample-efficient policy gradient with an off-policy critic. *arXiv preprint arXiv:1611.02247*. A

Gu, S. S., Lillicrap, T., Turner, R. E., Ghahramani, Z., Schölkopf, B., and Levine, S. (2017). Interpolated policy gradient: Merging on-policy and off-policy gradient estimation for deep reinforcement learning. In *Advances in Neural Information Processing Systems*, pages 3846–3855. A

Ha, D. and Schmidhuber, J. (2018). Recurrent world models facilitate policy evolution. In *Advances in Neural Information Processing Systems*, pages 2450–2462. 1

Hoffman, M. D., Blei, D. M., Wang, C., and Paisley, J. (2013). Stochastic variational inference. *The Journal of Machine Learning Research*, 14(1):1303–1347. 1

Jang, E., Gu, S., and Poole, B. (2016). Categorical reparameterization with Gumbel-Softmax. *arXiv preprint arXiv:1611.01144*. A

Jankowiak, M. and Karaletsos, T. (2019). Pathwise derivatives for multivariate distributions. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 333–342. 3, 8

Jankowiak, M. and Obermeyer, F. (2018). Pathwise derivatives beyond the reparameterization trick. In *International Conference on Machine Learning*, pages 2240–2249. 2.4, 4, 4, 5, A, E.1, E.2, E.2

Jiang, N. and Li, L. (2016). Doubly robust off-policy value evaluation for reinforcement learning. In *International Conference on Machine Learning*, pages 652–661. A

Jie, T. and Abbeel, P. (2010). On a connection between importance sampling and the likelihood ratio policy gradient. In *Advances in Neural Information Processing Systems*, pages 1000–1008. 2.3, A, C

Kingma, D. P. and Welling, M. (2013). Auto-encoding variational Bayes. *arXiv preprint arXiv:1312.6114*. 1

L'Ecuyer, P. (1990). A unified view of the IPA, SF, and LR gradient estimation techniques. *Management Science*, 36(11):1364–1383. 2.4

L'Ecuyer, P. (1991). An overview of derivative estimation. In *1991 Winter Simulation Conference Proceedings.*, pages 207–217. IEEE. 2.4

Lee, W., Yu, H., and Yang, H. (2018). Reparameterization gradient for non-differentiable models. In *Advances in Neural Information Processing Systems*, pages 5553–5563. 4, E.4, E.4, E.4, E.4, E.4

Liu, H., Feng, Y., Mao, Y., Zhou, D., Peng, J., and Liu, Q. (2017). Action-depedent control variates for policy optimization via stein's identity. *arXiv preprint arXiv:1710.11198*. 2.4

Maddison, C. J., Mnih, A., and Teh, Y. W. (2016). The concrete distribution: A continuous relaxation of discrete random variables. *arXiv preprint arXiv:1611.00712*. A

Mania, H., Guy, A., and Recht, B. (2018). Simple random search of static linear policies is competitive for reinforcement learning. In *Advances in Neural Information Processing Systems*, pages 1800–1809. 3, B, C

Mao, J., Foerster, J., Rocktäschel, T., Al-Shedivat, M., Farquhar, G., and Whiteson, S. (2019). A baseline for any order gradient estimation in stochastic computation graphs. In *International Conference on Machine Learning*, pages 4343–4351. A

Metz, L., Maheswaranathan, N., Nixon, J., Freeman, C. D., and Sohl-Dickstein, J. (2019). Understanding and correcting pathologies in the training of learned optimizers. In *International Conference on Machine Learning*. A

Mohamed, S., Rosca, M., Figurnov, M., and Mnih, A. (2019). Monte Carlo gradient estimation in machine learning. *arXiv preprint arXiv:1906.10652*. 1, 2.4, 4, A

Munos, R., Stepleton, T., Harutyunyan, A., and Bellemare, M. (2016). Safe and efficient off-policy reinforcement learning. In *Advances in Neural Information Processing Systems*, pages 1054–1062. A

Neal, R. M. (2003). Slice sampling. *The annals of statistics*, 31(3):705–767. F

Nesterov, Y. and Spokoiny, V. (2017). Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics*, 17(2):527–566. A

Owen, A. B. (2013). *Monte Carlo theory, methods and examples*. 2.4

Parmas, P. (2018). Total stochastic gradient algorithms and applications in reinforcement learning. In *Advances in Neural Information Processing Systems*, pages 10204–10214. A

Parmas, P., Rasmussen, C. E., Peters, J., and Doya, K. (2018). PIPPS: Flexible model-based policy search robust to the curse of chaos. In *International Conference on Machine Learning*, pages 4062–4071. 2.4, A, A

Peters, J. and Schaal, S. (2008). Reinforcement learning of motor skills with policy gradients. *Neural networks*, 21(4):682–697. 1

Petersen, K. B. and Pedersen, M. S. (2012). The matrix cookbook (version: November 15, 2012). E.3, E.3

Ranganath, R., Tran, D., and Blei, D. (2016). Hierarchical variational models. In *International Conference on Machine Learning*, pages 324–333. 2.4, 2.4

Rezende, D. J., Mohamed, S., and Wierstra, D. (2014). Stochastic backpropagation and approximate inference in deep generative models. In *International Conference on Machine Learning*, pages 1278–1286. 1, A

Riesz, F. (1907). *Sur une espèce de géométrie analytique des systèmes de fonctions sommables*. Gauthier-Villars. 4

Ruiz, F., Titsias, M., and Blei, D. (2016a). Overdispersed black-box variational inference. In *32nd Conference on Uncertainty in Artificial Intelligence 2016, UAI 2016*, pages 647–656. A

Ruiz, F. J., Titsias, M. K., and Blei, D. (2016b). The generalized reparameterization gradient. In *Advances in Neural Information Processing Systems*, pages 460–468. 4, E.3, E.3

Salimans, T., Ho, J., Chen, X., Sidor, S., and Sutskever, I. (2017). Evolution strategies as a scalable alternative to reinforcement learning. *arXiv preprint arXiv:1703.03864*. 1, 3, B, C

Schulman, J., Heess, N., Weber, T., and Abbeel, P. (2015a). Gradient estimation using stochastic computation graphs. In *Advances in Neural Information Processing Systems*, pages 3528–3536. A

Schulman, J., Levine, S., Abbeel, P., Jordan, M., and Moritz, P. (2015b). Trust region policy optimization. In *International Conference on Machine Learning*, pages 1889–1897. 1

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*. 1

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge. 1

Sutton, R. S., McAllester, D. A., Singh, S. P., and Mansour, Y. (2000). Policy gradient methods for reinforcement learning with function approximation. In *Advances in neural information processing systems*, pages 1057–1063. 1

Thomas, P. and Brunskill, E. (2016). Data-efficient off-policy policy evaluation for reinforcement learning. In *International Conference on Machine Learning*, pages 2139–2148. A

Titsias, M. K. and Lázaro-Gredilla, M. (2015). Local expectation gradients for black box variational inference. In *Advances in neural information processing systems*, pages 2638–2646. A

Tucker, G., Bhupatiraju, S., Gu, S., Turner, R., Ghahramani, Z., and Levine, S. (2018). The mirage of action-dependent baselines in reinforcement learning. In *International Conference on Machine Learning*, pages 5022–5031. A

Tucker, G., Mnih, A., Maddison, C. J., Lawson, J., and Sohl-Dickstein, J. (2017). REBAR: Low-variance, unbiased gradient estimates for discrete latent variable models. In *Advances in Neural Information Processing Systems*, pages 2627–2636. A

Walder, C. J., Nock, R., Ong, C. S., and Sugiyama, M. (2019). New tricks for estimating gradients of expectations. *arXiv preprint arXiv:1901.11311*. 4

Weaver, L. and Tao, N. (2001). The optimal reward baseline for gradient-based reinforcement learning. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pages 538–545. Morgan Kaufmann Publishers Inc. B

Weber, T., Heess, N., Buesing, L., and Silver, D. (2019). Credit assignment techniques in stochastic computation graphs. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2650–2660. A

Wierstra, D., Schaul, T., Peters, J., and Schmidhuber, J. (2008). Natural evolution strategies. In *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*, pages 3381–3387. IEEE. 1

Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3-4):229–256. 1

Wu, A. (2019). Generalized transformation-based gradient. *arXiv preprint arXiv:1911.02681*. A

Xu, M., Quiroz, M., Kohn, R., and Sisson, S. A. (2019). Variance reduction properties of the reparameterization trick. In *International Conference on Artificial Intelligence and Statistics*. A