Understanding Stuff for "Reparam in Policy" Paper 1

•) Differentiation under integral sign ☆

$$\frac{d}{dx}\int_a^b f(x,t)\,dt = \int_a^b \frac{\partial}{\partial x} f(x,t)\,dt$$

→ Partial

- Also called Leibniz integral rule
- used to solve complex integrals whose solution is not possible
- This simple example when 1) integration limits independent of π
  2) " discontinuities are independent of π

-) Indicator function

• denoted by $\mathbb{1}_A$

$\mathbb{1}_A(x) = 1$ if $x \in A$
or
$\mathbb{1}_A(x) = 0$ otherwise

If 'A' is a subset of some set 'X'

-) Importance Sampling in Monte Carlo Methods ☆

Goal, calculate $\int p(x) f(x)\,dx = E_p[f(x)]$

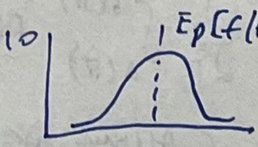
→ pdf of 'x'

if discrete then $\sum_x p(x) f(x)$

'X' too big to integrate ∴ we approximate using M.C.M

$$MCM \Rightarrow E_p[f(x)] \approx \underbrace{\frac{1}{N}\sum_{i=1}^{N} f(x_i)}_{S} \quad x_i \sim p(x)$$

since 'S' is random, it has its own distribution.



By Central Limit theorem

• $S \xrightarrow{d} \mathcal{N}(\mu, \sigma^2)$ ↓ Normal dist

$$\begin{cases} \mu = E_p[f(x)] \\ \sigma^2 = \frac{1}{N} V_p[f(x)] \end{cases}$$

$V_p[S] = \frac{1}{N} V_p[f(x)]$

↓ variance

-) Important Sampling : consider we have another dist $q(x)$

then: $E_p[f(x)] = \int p(x) f(x)\,dx$

$$\therefore E_p[f(x)] = \int p(x) f(x)\,dx \qquad \text{Multiply by } \frac{q(x)}{q(x)}$$

$$= \int q(x)\left[\frac{p(x)}{q(x)} f(x)\right] dx$$

$$= E_q\left[\frac{p(x)}{q(x)} f(x)\right] \approx \frac{1}{N}\sum_{i=1}^{N} \frac{p(x_i)}{q(x_i)} f(x_i) \quad x_i \sim q(x)$$

as long as $q(x) > 0$ wherever $p(x)f(x) \neq 0$

1) Identity Map

$f(X) = X$ for all elements X in M

Identity function 'f' defined on 'M' set is designed so.. M is its domain and co-domain

1) Control Variates

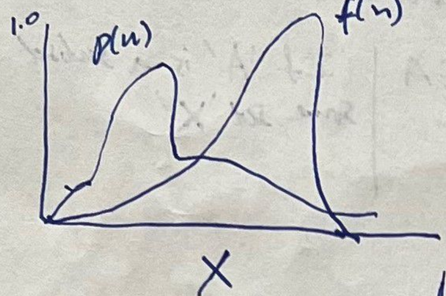
☆ why use 'r'? because $E_q[r] = E_p[f(n)]$ unbiased

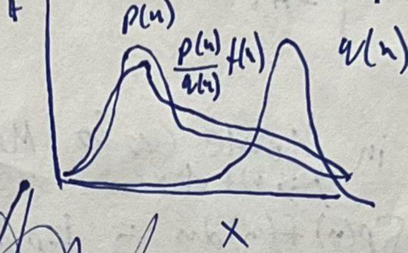
and

improved variance

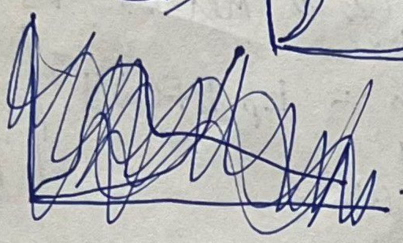
: $V_q[r] = \frac{1}{N} V_q\left[\frac{p(n)}{q(n)} f(n)\right]$

$V_q < V_p$





Importance Sampling



☆☆ Extension of Leibniz rule

$$\frac{d}{d\pi}\int_{a(\pi)}^{b(\pi)} f(n,\pi)\,dn = \underbrace{\int_{a(\pi)}^{b(\pi)} \frac{d}{d\pi} f(n,\pi)\,dn}_{\text{interior integral}}$$ [Move derivative inside integral]

Boundary terms

[Account for changes in integration limits] $+ f(b(\pi),\pi)\frac{d b(\pi)}{d\pi} - f(a(\pi),\pi)\frac{d a(\pi)}{d\pi}$

[Account for discontinuities of integrand that depend on π] $+ \sum_i \left(f(c_i(\pi)^-,\pi) - f(c_i(\pi)^+,\pi)\right)\frac{d c_i(\pi)}{d\pi}$

$$f(n,\pi) = \begin{cases} 0 & n < 2\pi \\ 1 & n > 2\pi \end{cases}$$

$$\frac{d}{d\pi}\int_0^{4\pi} f(n,\pi)\,dn = \int_0^{2\pi}\frac{d}{d\pi} 0\,dn + \int_{\pi}^{4\pi}\frac{d}{d\pi} 1\,dn$$

$$+ 1\frac{d(4\pi)}{d\pi} - 0\frac{d0}{d\pi}$$

$$+ (0-1)\frac{d(2\pi)}{d\pi}$$

Reparameterising Discontinuous Integrands for Differentiable Rendering

-) This approach does not rely on sampling silhouette edges [As in PyTorch3D]

- Differentiable path tracer to reconstruct 3D geometry and material of real world objects from real photos.

- Basically Paper presents new technique in <u>differentiable path tracing</u>
  1) to find edges with sufficient density
  2) produce gradients with low variance
  ] shortcoming of Li et al 2018

- Main Contrib:
  1) New tech for differential path tracing that solves this
  2) Avoid explicit sampling of discontinuities by applying careful chosen changes of variables that remove dependence of discontinuities on scene parameters.
  3) of course resulting integrands contain discontinuities but they are "static" in the re-parameterised integral and allows differentiation under integral sign.
  4) Therefore this technique produces high quality gradients in scenes with very high geometric complexity.

1) Preliminaries:

$$I = \int_X f(x, \Theta)\, dx$$

, f defined on domain X (unit sphere S^2)

Θ = Scene parameters like
i) vertex position, normals, texture etc

↓

pixel and shading integral

↓

Physically based rendering relies on Monte Carlo estimate of this

After this we assume single parameter $\Theta \in \mathbb{R}$

Now we will see how to compute gradients of integrands and how change in para helps in case of <u>non-diff integral</u>.

i) <u>3.1</u> smooth case for diff Monte Carlo estimators :

$$\frac{\partial I}{\partial \theta} = \frac{\partial}{\partial \theta} \int f(x, \theta)\, dx$$

grad-based optim requires access to partial deriv of integral I wrt θ :

- Using Leibniz integral rule giving continuity of both 'f' and its partial derivative θ

$$\therefore \frac{\partial}{\partial \theta} \int f(x,\theta)\, dx = \int \frac{\partial}{\partial \theta} f(x,\theta)\, dx$$

using Monte Carlo integration to estimate 'I' with a probability density function $p(x)$ that doesn't depend on 'θ'

$$I \approx E = \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i, \theta)}{p(x_i)}$$

here $P_X(x_i)$ where $\sum_i P_X(x_i) = 1$

- So an estimator of the derivative is simply derivative of the estimator

$$\therefore \frac{\partial I}{\partial \theta} \approx \frac{1}{N} \sum \frac{\partial}{\partial \theta} \frac{f(x_i, \theta)}{p(x_i)} = \frac{\partial E}{\partial \theta}$$

- Sampling depends on θ.

$$\therefore I \approx E = \frac{1}{N} \sum \frac{f(x_i(\theta), \theta)}{p(x_i(\theta), \theta)}$$

- Since 'p' is differentiable wrt θ, and when sampling implements a differentiable mapping from random to samples $x_i(\theta)$ then we differentiate Monte Carlo in the same way

$$\therefore \frac{\partial I}{\partial \theta} \approx \frac{\partial E}{\partial \theta} = \frac{1}{N} \sum \frac{\partial}{\partial \theta} \frac{f(x_i(\theta), \theta)}{p(x_i(\theta), \theta)}$$

3.2 Non diff integrands

1) Even if mappings are differentiable the function $f(x, \theta)$ is non diff in θ due to visibility changes.

2) These manifest as discontinuities ~~...~~, whose position in 'X' are a function of θ.

∴ diff not valid in such conditions

1) We overcome this by using "a change of variables" that removes the discontinuity of the integrand in θ.

if $T: y \to X$ transformation exists

then the re-parameterised integral

$$\int_X f(x, \theta)\, dx = \int_y f(T(y, \theta), \theta) \,|\det J_T|\, dy$$

can be handle using previous estimators $\hat{I}$

- Now considering an example.
  1) Continous integration kernel multiplied by 1D indicator func which acts as a visibility function

$$f(x) = \mathbb{1}_{x > \theta} K(x), \quad \text{where} \int_X K(x)\, dx = 1$$

θ = parameter that determines position of discontinuity that prevents diff under integral sign

∴ we define a new integration variable $y = x - \theta$ and change variables accordingly, here $|\det J_T| = 1$

$$I = \int_X f(x)\, dx = \int_y \mathbb{1}_{y>0} K(y+\theta)\, dy$$

Now is differentiable wrt θ at every point y

so we get

$$\frac{\partial I}{\partial \theta} \approx \frac{1}{N} \sum \frac{\partial}{\partial \theta} \frac{\mathbb{1}_{y_i > 0} K(y_i + \theta)}{p(y_i)}$$

2

$\therefore$) The above transformation can have 2 effects

1) Instead of integrating a func with a discontinity who depends on θ, we integrate in a space where discontinity does not move when θ changes

2) This is equivalent to importance sampling the integral $\int f(x)\,dx$ using samples $x_i(\theta) = y_i + \theta$

$\therefore$ If change of variables T is available, we draw samples y_i from density 'p' and evaluate $f(T(y_i))$ and $p(y_i)$ and the Jacobian then the estimate is differentiated using A.D.

Imp:
1) Transformation does not affect primal computation of 'I'
2) transformation 'T' should be designed to yield identity map when $\theta = \theta_0$, θ_0 = concrete parameter value for which grad to be evaluated. in 1D example it is $T(y,\theta) = y + \theta - \theta_0$
3) density 'p' from which 'y_i' are sampled must not depend on θ, otherwise discontinuities would be reintroduced in integral.

Method:

1) Assuming all relevant integrals have been reparameterized over 'spherical domain'
2) Integrants vanish outside of a small neighbourhood around a given direction.

4.1 Removing discontinuities using rotations

·) when integrands have small angular support, the discontinuities typically consist of the silhouette of a single object as in fig 2.
" Main Assumption of Method "

·) displacement of this silhouette on S^2 for infinitesimal perturbations of θ is well approximated using spherical rotation. As in fig 4a

·) Main Assumption

$$I = \int_{S^2} f(\omega, \theta) d\omega = \int_{S^2} f(R(\omega, \theta), \theta) d\omega$$

·) That there exists a rotation so that change of variables makes $f(R(\omega,\theta),\theta)$ continuous wrt θ for each direction 'ω'

·) $|\det J_T| = 1$ for this change of variables + Rotation must depend on para value θ.

·) Using rotation 'R' to reparametrize the integral leads to Monte Carlo estimate

$$E = \frac{1}{N} \sum \frac{f(R(\omega_i, \theta), \theta)}{p(\omega_i; \theta_0)} \approx I$$

ω_i sampled from distribution 'p' - normally, θ_0 used instead of θ to remove dependencies of samples ω_i on θ

·) 4.2 Generalizing to func with large support

·) when complex arrangement of ~~sig~~ singularities violate assumptions in 4.1 it produces significant bias in grad est.

∴ Solution → $\int_{S^2} f(\omega) d\omega = \int_{S^2} \int_{S^2} f(\mu) K(\mu, \omega) d\mu d\omega$,

where K is a spherical convolution kernel satisfying

$$\int_{S^2} K(\mu, \omega) d\mu = 1, \ \forall \omega \in S^2$$

Integral of 'f' is equal to the integral of a convolution of 'f'

1) We choose 'K' making it viable for ~~the~~ assumption in 4.1

1) Therefore to evaluate and estimate this we sample directions ω_i for the outer integral and then sample offset directions $\mu_i \sim K(\cdot, \omega_i)$ which gives us:

$$I \approx E = \frac{1}{N} \sum \frac{f(x_i(\mu_i,\theta),\theta)\, K(x_i(\mu_i,\theta),\omega_i(\theta),\theta)}{p(\omega_i(\theta),\theta)\, p_K(\mu_i)}$$

Size of kernel 'K' provides trade offs in variance and bias

1) Determining suitable rotations using ray tracing
   1) Benefit of our approach is that we dont even want to know if the integrand ~~has~~ has a silhouette.
   1) Displacement of points on silhouette edges will closely approximate displacement of other positions on the object.
   2) We intersect 4 rays within the "support" of the integrand against scene geometry.
   3) Using the resulting info about their distance and surface normals we apply a heuristic to select a point that belongs to the occluder whose motion wrt scene para matches that of silhouette.
   1) All this for efficient computation of rotations to compensate for discontinuities.
   1) See fig 5, estimation of changes in variables is done independently for each integral at render time, and some rays created for this task can be ~~used~~ reused for sampling the integrals.
1) See fig 13, projection direction in S^2 of integration, of the selected ~~point~~ point $\omega_p(\theta)$ and set $\omega_{p_0} = \omega_p(\theta_0)$

$\therefore$ we build a differentiable rot-Mat $R(\theta)$ that satisfies: $R(\theta_0)\omega = \omega, \forall \omega \in S^2$ and

$$\frac{\partial}{\partial\theta} R(\theta)\,\omega_{p_0} = \frac{\partial}{\partial\theta}\,\omega_p(\theta)$$

1) direction of current para config