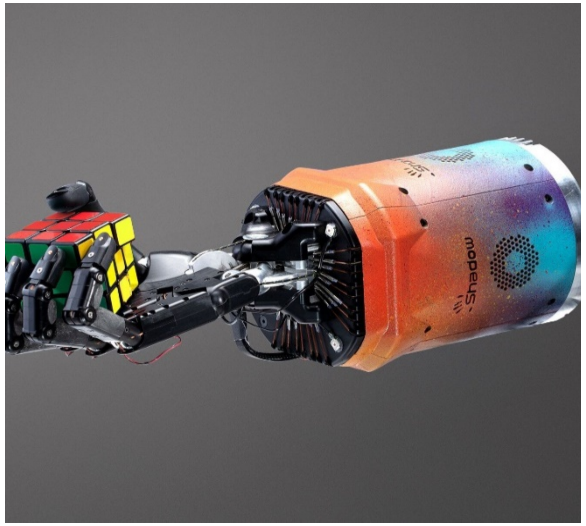
Do Differentiable Simulators Give Better Policy Gradients?

H.J. Terry Suh, Max Simchowitz, Kaiqing Zhang, Russ Tedrake

ICML 2022
Long Talk Presentation

Motivation. Policy Optimization for Physical Systems

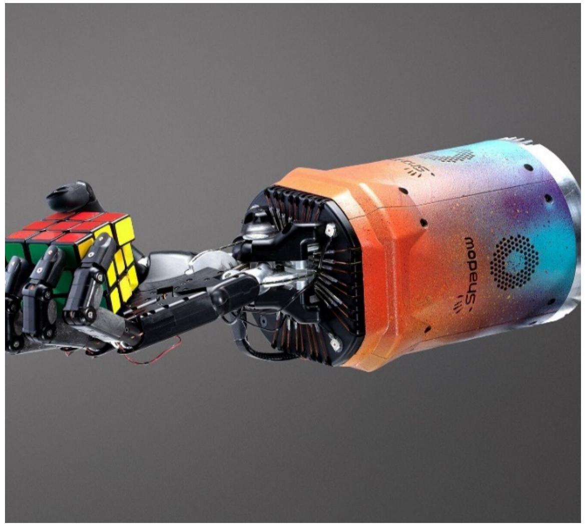


Success of Policy Optimization for Robotics

1. Collect data in simulation, run zero-order policy optimization (e.g. PPO)
2. Impressive results in manipulation, locomotion.

[1] OpenAI, “Solving Rubik’s Cube with a Robot Hand”
[2] T. Miki et al., “Learning robust perceptive locomotion for quadrupedal robots in the wild”.

Motivation. Policy Optimization for Physical Systems



Success of Policy Optimization for Robotics

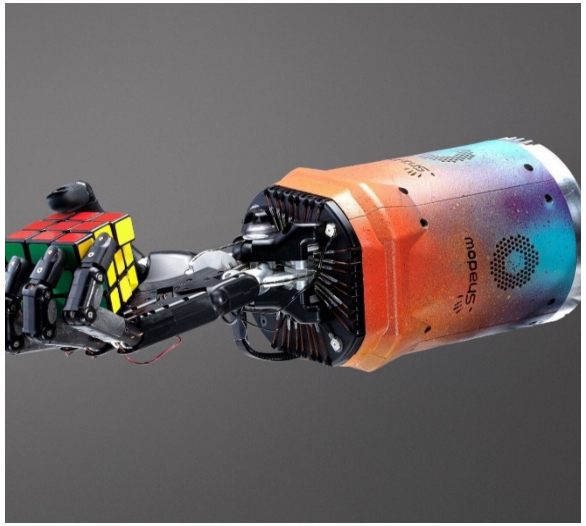
1. Collect data in simulation, run zero-order policy optimization (e.g. PPO)
2. Impressive results in manipulation, locomotion.

Still Leaves Much to be Desired.

1. Incredibly data hungry. Need many samples / time to train a policy.
2. We know structure for these systems. The fact that we cannot do better than blackbox optimization algorithms is both theoretically / practically unsatisfying.

[1] OpenAI, “Solving Rubik’s Cube with a Robot Hand”
[2] T. Miki et al., “Learning robust perceptive locomotion for quadrupedal robots in the wild”.

Motivation. Policy Optimization for Physical Systems



Success of Policy Optimization for Robotics

1. Collect data in simulation, run zero-order policy optimization (e.g. PPO)
2. Impressive results in manipulation, locomotion.

Still Leaves Much to be Desired.

1. Incredibly data hungry. Need many samples / time to train a policy.
2. We know structure for these systems. The fact that we cannot do better than blackbox optimization algorithms is both theoretically / practically unsatisfying.

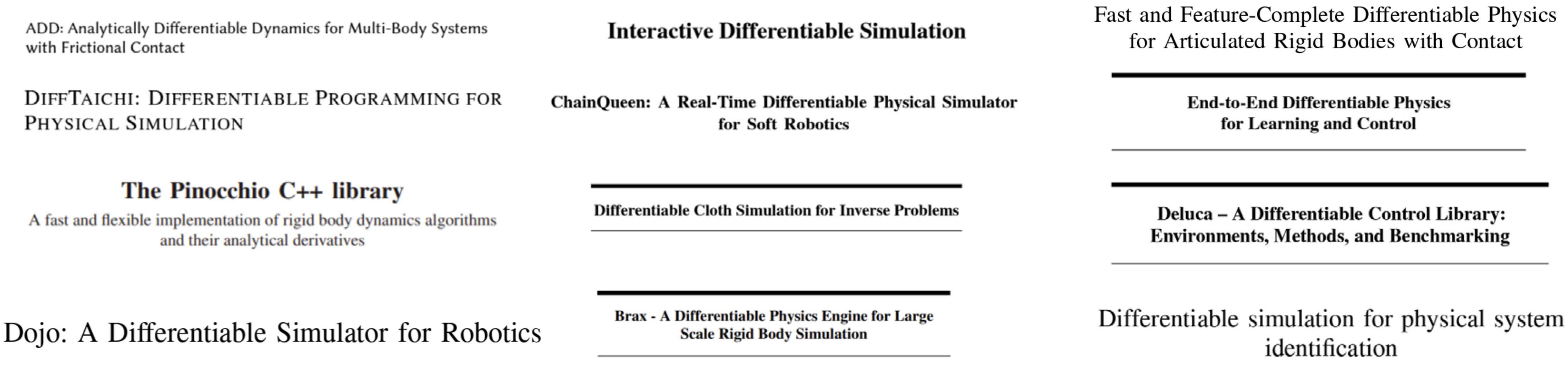
Can we do better by leveraging model structure for robotics?

[1] OpenAI, “Solving Rubik’s Cube with a Robot Hand”
[2] T. Miki et al., “Learning robust perceptive locomotion for quadrupedal robots in the wild”.

Motivation. Differentiable Simulation

Can we do better by leveraging model structure for robotics?

Yes, let’s make our models differentiable and use first-order gradients.



ADD: Analytically Differentiable Dynamics for Multi-Body Systems with Frictional Contact

DIFFTAICHI: DIFFERENTIABLE PROGRAMMING FOR PHYSICAL SIMULATION

The Pinocchio C++ library

A fast and flexible implementation of rigid body dynamics algorithms and their analytical derivatives

Dojo: A Differentiable Simulator for Robotics

Interactive Differentiable Simulation

ChainQueen: A Real-Time Differentiable Physical Simulator for Soft Robotics

Differentiable Cloth Simulation for Inverse Problems

Brax - A Differentiable Physics Engine for Large Scale Rigid Body Simulation

Fast and Feature-Complete Differentiable Physics for Articulated Rigid Bodies with Contact

End-to-End Differentiable Physics for Learning and Control

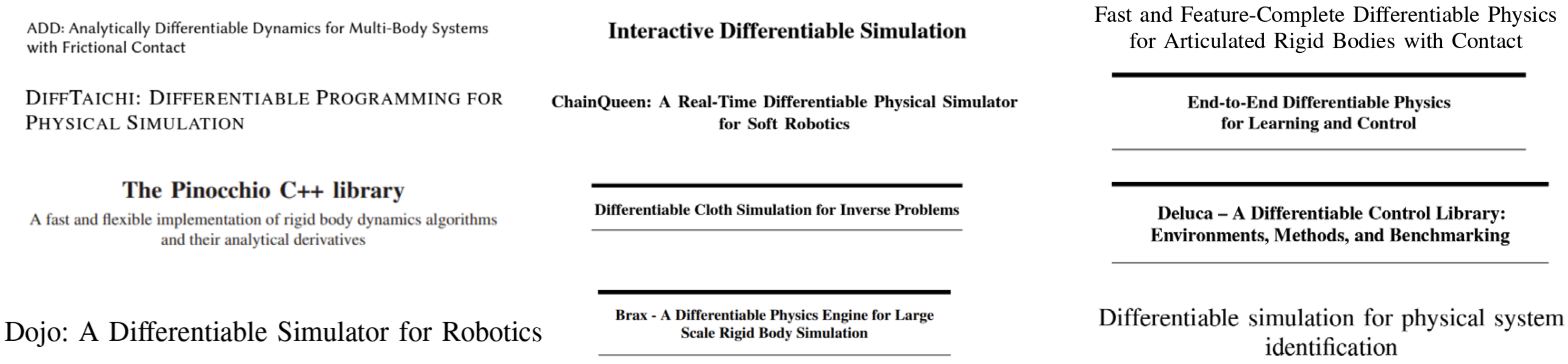
Deluca – A Differentiable Control Library: Environments, Methods, and Benchmarking

Differentiable simulation for physical system identification

Motivation. Differentiable Simulation

Can we do better by leveraging model structure for robotics?

Yes, let's make our models differentiable and use first-order gradients.



ADD: Analytically Differentiable Dynamics for Multi-Body Systems with Frictional Contact

Interactive Differentiable Simulation

Fast and Feature-Complete Differentiable Physics for Articulated Rigid Bodies with Contact

DIFFTAICHI: DIFFERENTIABLE PROGRAMMING FOR PHYSICAL SIMULATION

ChainQueen: A Real-Time Differentiable Physical Simulator for Soft Robotics

End-to-End Differentiable Physics for Learning and Control

The Pinocchio C++ library

A fast and flexible implementation of rigid body dynamics algorithms and their analytical derivatives

Differentiable Cloth Simulation for Inverse Problems

Deluca – A Differentiable Control Library: Environments, Methods, and Benchmarking

Dojo: A Differentiable Simulator for Robotics

Brax - A Differentiable Physics Engine for Large Scale Rigid Body Simulation

Differentiable simulation for physical system identification

If we have access to autodiff gradients for physics simulation, is it better to use them for policy search?
Are there pathologies where using these gradients actually hurt?

Setup. Stochastic Optimization.

Stochastic First vs. Zeroth-Order Optimization (Additive Gaussian Noise)

$$\min_{\theta} F(\theta) = \min_{\theta} \mathbb{E}_{w \sim \mathcal{N}(w;0,\sigma^2 I)} f(\theta + w)$$

Setup. Stochastic Optimization.

Stochastic First vs. Zeroth-Order Optimization (Additive Gaussian Noise)

$$\min_\theta F(\theta) = \min_\theta \mathbb{E}_{w\sim\mathcal{N}(w;0,\sigma^2 I)} f(\theta + w)$$

ZoBG

Zero-Order Batch Gradient

$$\nabla_\theta \mathbb{E}_w f(\theta + w) = \frac{1}{\sigma^2}\mathbb{E}_w f(\theta + w)w$$

$$\approx \frac{1}{N}\sum_{i=1}^{N} f(\theta + w_i)w_i$$

REINFORCE
Likelihood Ratio
Score Function

Setup. Stochastic Optimization.

Stochastic First vs. Zeroth-Order Optimization (Additive Gaussian Noise)

$$\min_\theta F(\theta) = \min_\theta \mathbb{E}_{w \sim \mathcal{N}(w;0,\sigma^2 I)} f(\theta + w)$$

ZoBG

Zero-Order Batch Gradient

REINFORCE
Likelihood Ratio
Score Function

$$\nabla_\theta \mathbb{E}_w f(\theta + w) = \frac{1}{\sigma^2} \mathbb{E}_w f(\theta + w) w$$

$$\approx \frac{1}{N} \sum_{i=1}^{N} f(\theta + w_i) w_i$$

FoBG

First-Order Batch Gradient

Reparametrization
Pathwise Derivative
Backpropagation through Time

$$\nabla_\theta \mathbb{E}_w f(\theta + w) = \mathbb{E}_w \nabla_\theta f(\theta + w)$$

$$\approx \frac{1}{N} \sum_{i=1}^{N} \nabla_\theta f(\theta + w_i)$$

Setup. Stochastic Optimization.

Stochastic First vs. Zeroth-Order Optimization (Additive Gaussian Noise)

$$\min_\theta F(\theta) = \min_\theta \mathbb{E}_{w\sim\mathcal{N}(w;0,\sigma^2 I)} f(\theta + w)$$

ZoBG

Zero-Order Batch Gradient

REINFORCE
Likelihood Ratio
Score Function

$$\nabla_\theta \mathbb{E}_w f(\theta + w) = \frac{1}{\sigma^2}\mathbb{E}_w f(\theta + w)w$$

$$\approx \frac{1}{N}\sum_{i=1}^{N} f(\theta + w_i)w_i$$

FoBG

First-Order Batch Gradient

Reparametrization
Pathwise Derivative
Backpropagation through Time

$$\nabla_\theta \mathbb{E}_w f(\theta + w) = \mathbb{E}_w \nabla_\theta f(\theta + w)$$

$$\approx \frac{1}{N}\sum_{i=1}^{N} \nabla_\theta f(\theta + w_i)$$

Lessons from Stochastic Optimization

1. The two gradients converge to the same quantity under sufficient regularity conditions.
2. Convergence rate scales directly with variance of the estimators, ZoBG often has higher variance.

The Pathologies of FoBG

Lessons from Stochastic Optimization

1. The two gradients converge to the same quantity under sufficient regularity conditions.
   But the regularity conditions can be broken, leading to a biased FoBG.

2. Convergence rate scales directly with variance of the estimators, ZoBG often has higher variance.
   Often, but not always.

The Pathologies of FoBG

Lessons from Stochastic Optimization

1. The two gradients converge to the same quantity under sufficient regularity conditions.
   But the regularity conditions can be broken, leading to a biased FoBG.

2. Convergence rate scales directly with variance of the estimators, ZoBG often has higher variance.
   Often, but not always.

FoBG can be biased for discontinuous (not locally Lipschitz) landscapes!

The Pathologies of FoBG

Lessons from Stochastic Optimization

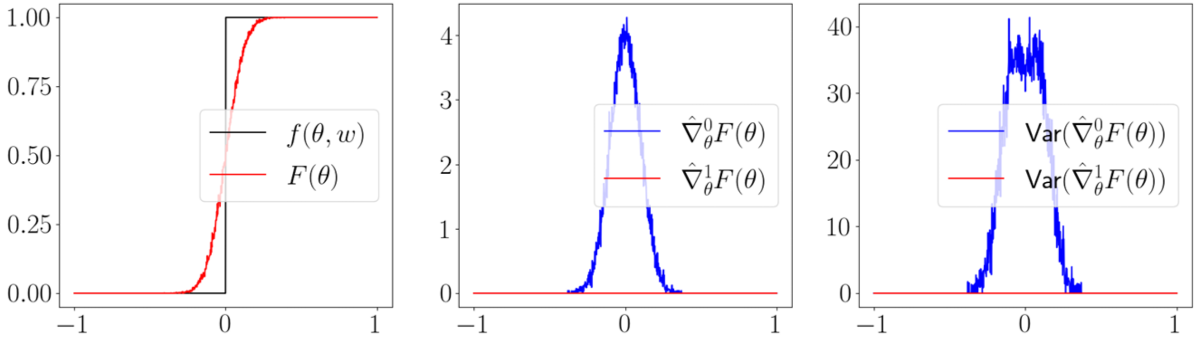
1. The two gradients converge to the same quantity under sufficient regularity conditions.
   But the regularity conditions can be broken, leading to a biased FoBG.
2. Convergence rate scales directly with variance of the estimators, ZoBG often has higher variance.
   Often, but not always.

FoBG can be biased for discontinuous (not locally Lipschitz) landscapes!

Illustrated best by sampling from a Heaviside.

$$f(\boldsymbol{\theta}, \mathbf{w}) = H(\boldsymbol{\theta} + \mathbf{w}), \quad H(t) = \begin{cases} 1 & t \geq 0 \\ 0 & t < 0 \end{cases},$$

$$F(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{w}}[H(\boldsymbol{\theta} + \mathbf{w})] = \text{erf}(-\boldsymbol{\theta}; \sigma^2),$$



The Pathologies of FoBG

Lessons from Stochastic Optimization

1. The two gradients converge to the same quantity under sufficient regularity conditions.
   But the regularity conditions can be broken, leading to a biased FoBG.

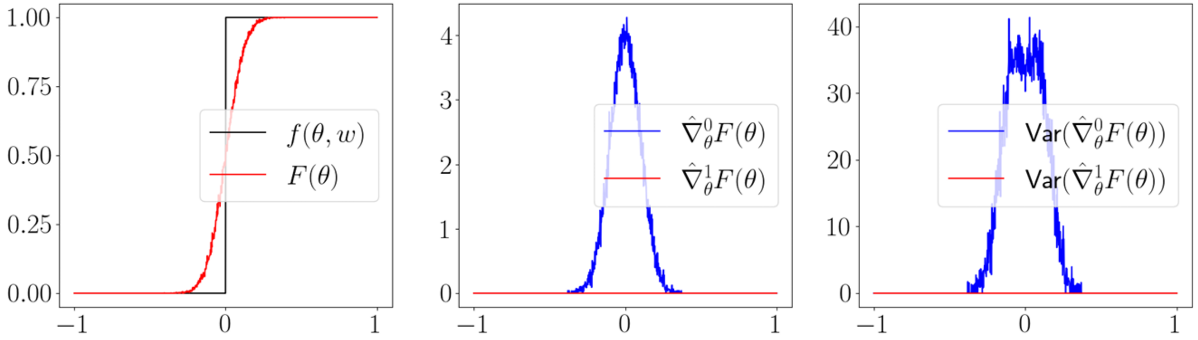
2. Convergence rate scales directly with variance of the estimators, ZoBG often has higher variance.
   Often, but not always.

FoBG can be biased for discontinuous (not locally Lipschitz) landscapes!

Illustrated best by sampling from a Heaviside.

$$f(\boldsymbol{\theta}, \mathbf{w}) = H(\boldsymbol{\theta} + \mathbf{w}), \quad H(t) = \begin{cases} 1 & t \geq 0 \\ 0 & t < 0 \end{cases},$$

$$F(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{w}}[H(\boldsymbol{\theta} + \mathbf{w})] = \mathrm{erf}(-\boldsymbol{\theta}; \sigma^2),$$



Note that since samples of first-order gradients are identically zero,

- The FoBG is zero, while the gradient of stochastic objective is non-zero everywhere.
- The empirical variance of the FoBG is also zero.

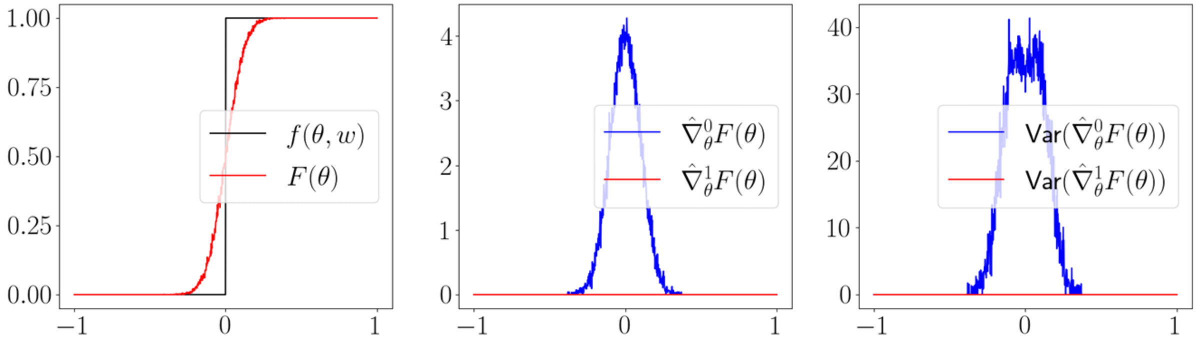
The Pathologies of FoBG: Discontinuities

FoBG can be biased for discontinuous (not locally Lipschitz) landscapes!

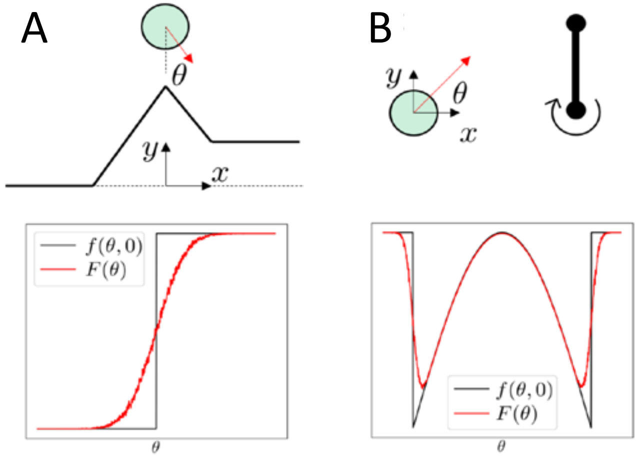
Illustrated best by sampling from a Heaviside.

$$f(\boldsymbol{\theta}, \mathbf{w}) = H(\boldsymbol{\theta} + \mathbf{w}), \quad H(t) = \begin{cases} 1 & t \geq 0 \\ 0 & t < 0 \end{cases},$$

$$F(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{w}}[H(\boldsymbol{\theta} + \mathbf{w})] = \mathrm{erf}(-\boldsymbol{\theta}; \sigma^2),$$



Not just a pathology, but common in physical systems involving contact.



1. Discontinuities caused by geometry (non-smooth surfaces, discontinuous normal).
2. Discontinuities caused by friction and tangential velocities.
3. Discontinuities caused by impacts.

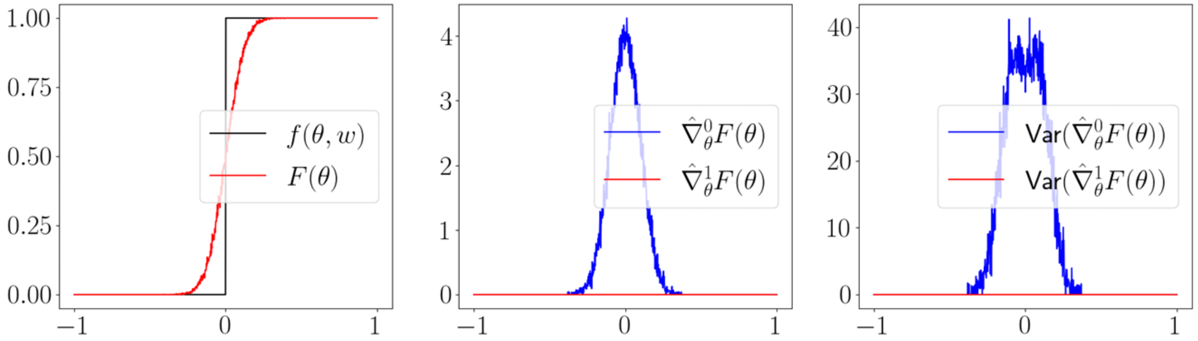
The Pathologies of FoBG: Discontinuities

FoBG can be biased for discontinuous (not locally Lipschitz) landscapes!

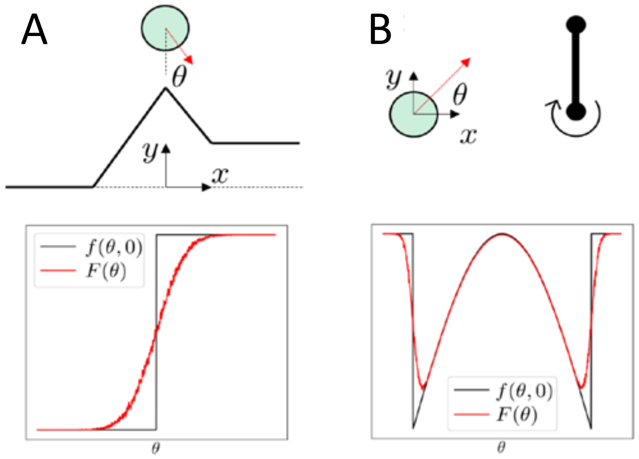
Illustrated best by sampling from a Heaviside.

$$f(\boldsymbol{\theta},\mathbf{w}) = H(\boldsymbol{\theta}+\mathbf{w}), \quad H(t) = \begin{cases}1 & t \geq 0 \\ 0 & t < 0\end{cases},$$

$$F(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{w}}[H(\boldsymbol{\theta}+\mathbf{w})] = \mathrm{erf}(-\boldsymbol{\theta};\sigma^2),$$



Not just a pathology, but common in physical systems involving contact.

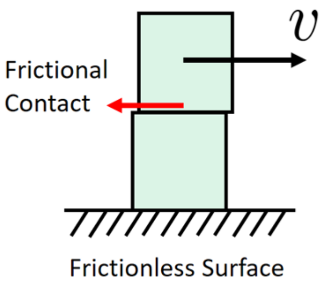


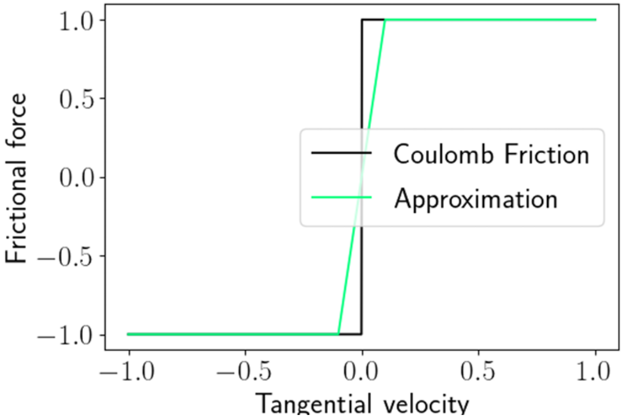
1. Discontinuities caused by geometry (non-smooth surfaces, discontinuous normal).
2. Discontinuities caused by friction and tangential velocities.
3. Discontinuities caused by impacts.

Perhaps the strictness of the discontinuity is a modeling decision, what if we soften it?

The Pathologies of FoBG: Empirical Bias

Empirical Bias: Continuous yet stiff Approximations of Discontinuities Look like Strict Discontinuities in the finite-sample regime.

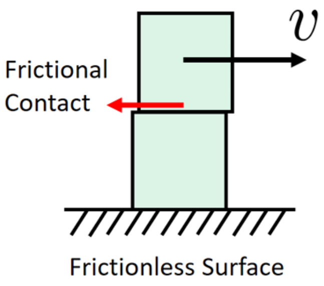


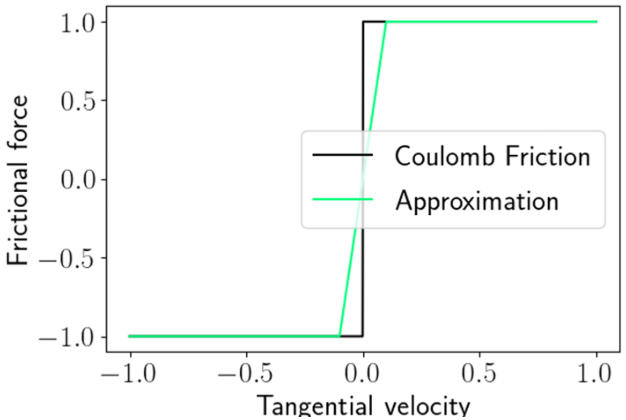


$$\bar{H}_\nu(t) = \begin{cases} 2t/\nu & \text{if } |t| \leq \nu/2 \\ H(t) & \text{else} \end{cases}$$

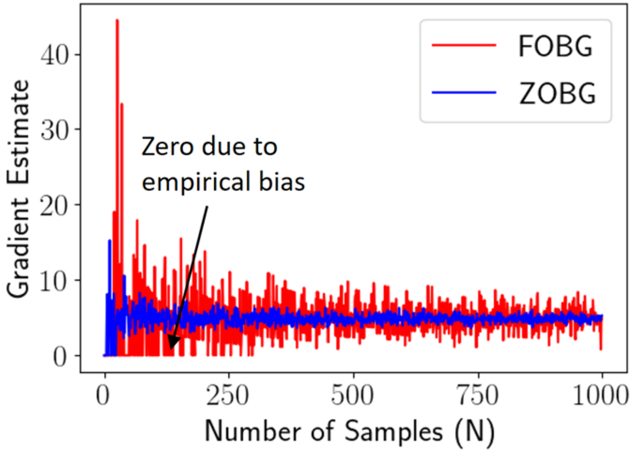
The Pathologies of FoBG: Empirical Bias

Empirical Bias: Continuous yet stiff Approximations of Discontinuities Look like Strict Discontinuities in the finite-sample regime.





$$\bar{H}_\nu(t) = \begin{cases} 2t/\nu & \text{if } |t| \leq \nu/2 \\ H(t) & \text{else} \end{cases}$$



1. Gradient of stiff approximations take high value with low probability
2. In finite-sample regime, there is no way to distinguish between strict discontinuity and its stiff continuous relaxations.

The Pathologies of FoBG: High Variance

Lessons from Stochastic Optimization

1. The two gradients converge to the same quantity under sufficient regularity conditions.
   But the regularity conditions can be broken, leading to a biased FoBG.

2. Convergence rate scales directly with variance of the estimators, ZoBG often has higher variance.
   Often, but not always.

The Pathologies of FoBG: High Variance

Lessons from Stochastic Optimization

1. The two gradients converge to the same quantity under sufficient regularity conditions.
   But the regularity conditions can be broken, leading to a biased FoBG.

2. Convergence rate scales directly with variance of the estimators, ZoBG often has higher variance.
   Often, but not always.

ZoBG

Zero-Order Batch Gradient

REINFORCE
Likelihood Ratio
Score Function

$$\mathbf{Var}(\hat{\nabla}^{[0]}F(\theta)) \leq \frac{\overset{\dim w}{\swarrow}\,n}{N\sigma^2}\max\nolimits_w \|f(\theta+w)\|_2^2$$

1. Scaling with dimension of injected noise.
2. Scaling with function value.

The Pathologies of FoBG: High Variance

Lessons from Stochastic Optimization

1. The two gradients converge to the same quantity under sufficient regularity conditions.
   But the regularity conditions can be broken, leading to a biased FoBG.

2. Convergence rate scales directly with variance of the estimators, ZoBG often has higher variance.
   Often, but not always.

ZoBG

Zero-Order Batch Gradient

REINFORCE
Likelihood Ratio
Score Function

$$\mathbf{Var}(\hat{\nabla}^{[0]}F(\theta)) \leq \frac{\overset{\overset{\dim w}{\swarrow}}{n}}{N\sigma^2}\max_w \|f(\theta+w)\|_2^2$$

1. Scaling with dimension of injected noise.
2. Scaling with function value.

FoBG

First-Order Batch Gradient

Reparametrization
Pathwise Derivative

$$\mathbf{Var}(\hat{\nabla}^{[1]}F(\theta)) \leq \frac{1}{N}\max_w \|\nabla_\theta f(\theta+w)\|_2^2$$

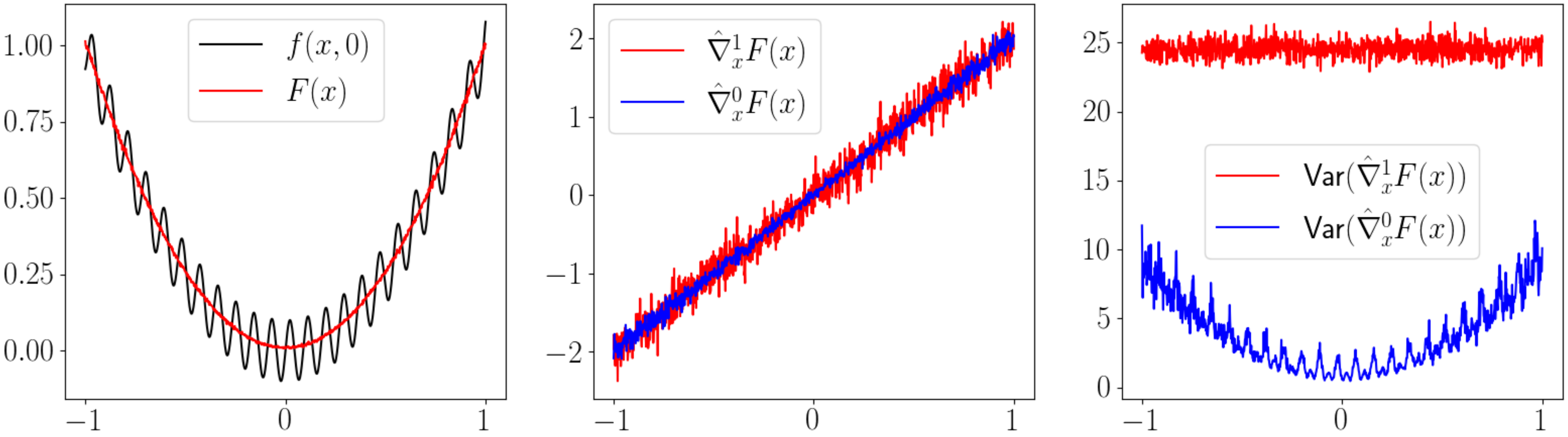
1. No scaling with dimension.
2. Scaling with value of gradient.

The Pathologies of FoBG: High Variance

Lessons from Stochastic Optimization

1. The two gradients converge to the same quantity under sufficient regularity conditions.
   But the regularity conditions can be broken, leading to a biased FoBG.

2. Convergence rate scales directly with variance of the estimators, ZoBG often has higher variance.
   Often, but not always.



The Pathologies of FoBG: High Variance

ZoBG

Zero-Order Batch Gradient

$$\mathbf{Var}(\hat{\nabla}^{[0]}F(\theta)) \leq \frac{n}{N\sigma^2}\max_w \|f(\theta+w)\|_2^2$$

FoBG

First-Order Batch Gradient

$$\mathbf{Var}(\hat{\nabla}^{[1]}F(\theta)) \leq \frac{1}{N}\max_w \|\nabla_\theta f(\theta+w)\|_2^2$$

The Pathologies of FoBG: High Variance

ZoBG
Zero-Order Batch Gradient

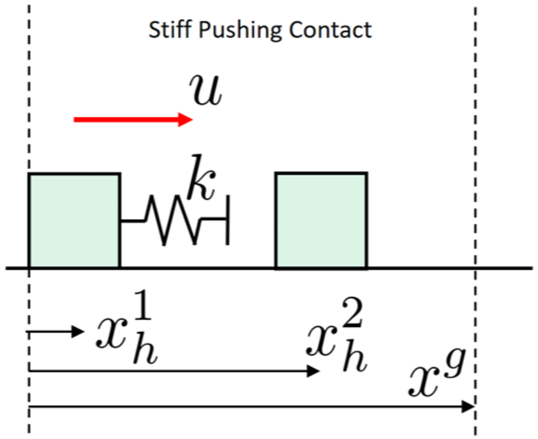
$$\mathbf{Var}(\hat{\nabla}^{[0]} F(\theta)) \leq \frac{n}{N\sigma^2} \max_w \|f(\theta + w)\|_2^2$$

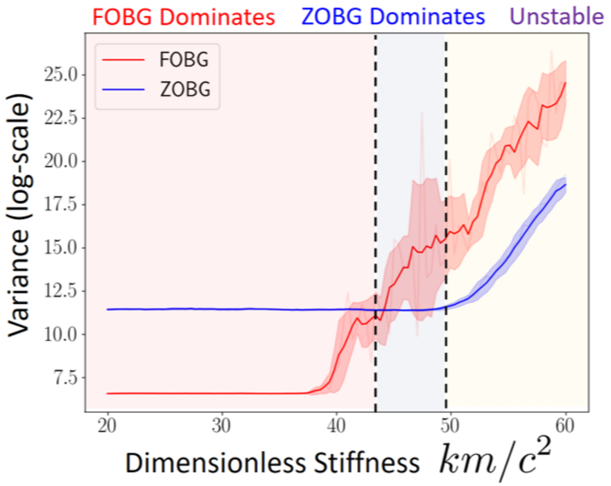
FoBG
First-Order Batch Gradient

$$\mathbf{Var}(\hat{\nabla}^{[1]} F(\theta)) \leq \frac{1}{N} \max_w \|\nabla_\theta f(\theta + w)\|_2^2$$

The FoBG CAN have more variance if function values are bounded, but gradients are high.

Case 1. Stiff Contact Models





The Pathologies of FoBG: High Variance

ZoBG

Zero-Order Batch Gradient

$$\mathbf{Var}(\hat{\nabla}^{[0]}F(\theta)) \leq \frac{n}{N\sigma^2} \max_w \|f(\theta+w)\|_2^2$$

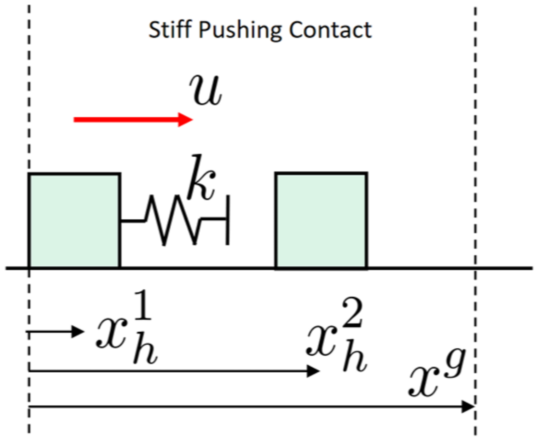
FoBG

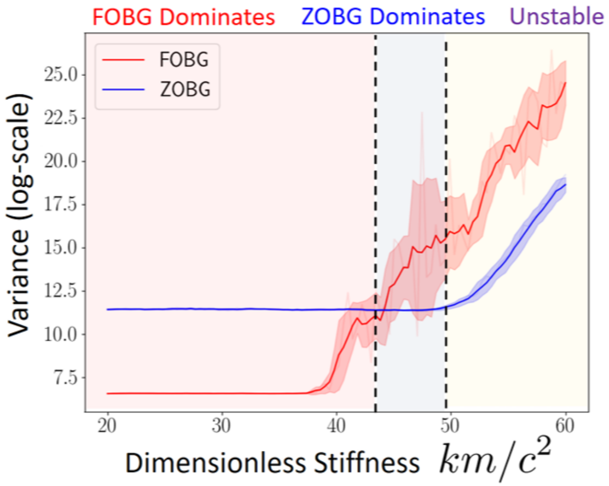
First-Order Batch Gradient

$$\mathbf{Var}(\hat{\nabla}^{[1]}F(\theta)) \leq \frac{1}{N} \max_w \|\nabla_\theta f(\theta+w)\|_2^2$$

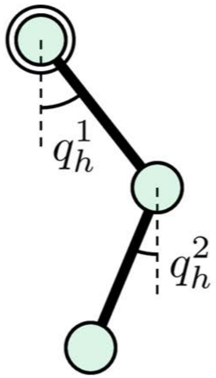
The FoBG CAN have more variance if function values are bounded, but gradients are high.

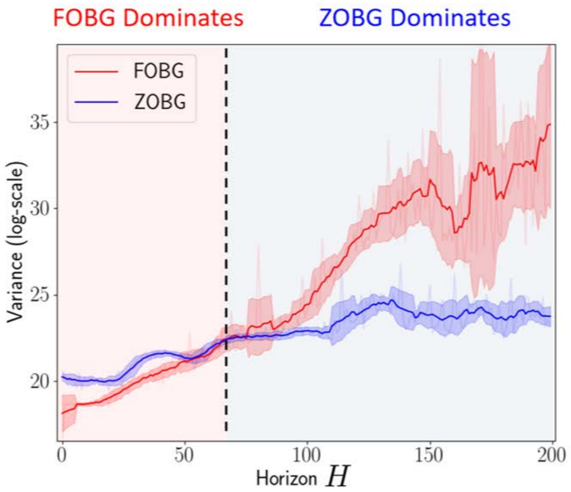
Case 1. Stiff Contact Models





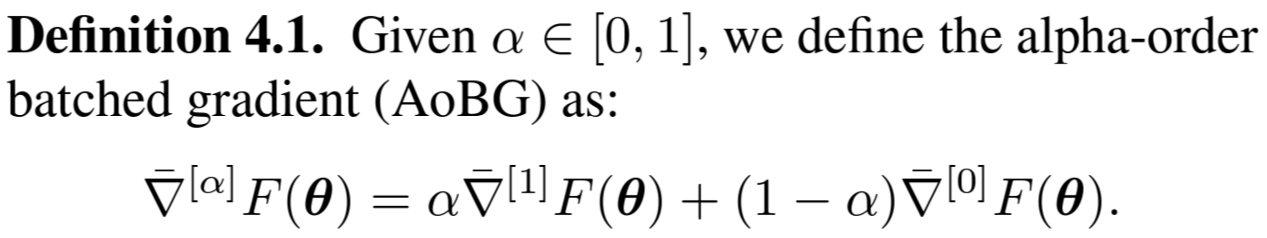
Case 2. Chaos





Interpolating the First and Zero-Order Gradients

Consider an interpolated gradient of the two objectives. How should we choose alpha?

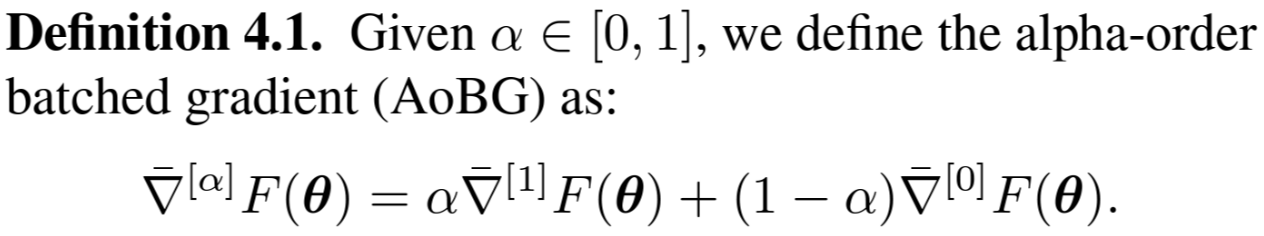


Definition 4.1. Given $\alpha \in [0,1]$, we define the alpha-order batched gradient (AoBG) as:

$$\bar{\nabla}^{[\alpha]} F(\boldsymbol{\theta}) = \alpha \bar{\nabla}^{[1]} F(\boldsymbol{\theta}) + (1-\alpha) \bar{\nabla}^{[0]} F(\boldsymbol{\theta}).$$

Interpolating the First and Zero-Order Gradients

Consider an interpolated gradient of the two objectives. How should we choose alpha?



Definition 4.1. Given $\alpha \in [0, 1]$, we define the alpha-order batched gradient (AoBG) as:

$$\bar{\nabla}^{[\alpha]} F(\boldsymbol{\theta}) = \alpha \bar{\nabla}^{[1]} F(\boldsymbol{\theta}) + (1 - \alpha) \bar{\nabla}^{[0]} F(\boldsymbol{\theta}).$$

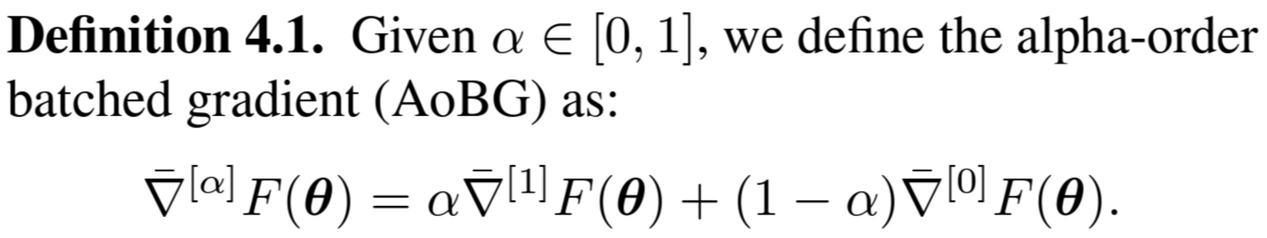
Original Motivation behind some of these approaches:

1. The FoBG may be subject to high variance because of chaos.
2. But the empirical variance can be queried online which can inform us which gradient to use more.
3. Assuming the samples used to obtain both estimates are uncorrelated, we can minimize expected variance of the interpolated gradient:

$$\min_{\alpha \in [0,1]} \quad \alpha^2 \hat{\sigma}_1^2 + (1 - \alpha)^2 \hat{\sigma}_0^2 \qquad\qquad \alpha^* = \frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2 + \hat{\sigma}_0^2}$$

Interpolating the First and Zero-Order Gradients

Consider an interpolated gradient of the two objectives. How should we choose alpha?



Definition 4.1. Given $\alpha \in [0, 1]$, we define the alpha-order batched gradient (AoBG) as:

$$\bar{\nabla}^{[\alpha]} F(\boldsymbol{\theta}) = \alpha \bar{\nabla}^{[1]} F(\boldsymbol{\theta}) + (1 - \alpha) \bar{\nabla}^{[0]} F(\boldsymbol{\theta}).$$

Original Motivation behind some of these approaches:

1. The FoBG may be subject to high variance because of chaos.
2. But the empirical variance can be queried online which can inform us which gradient to use more.
3. Assuming the samples used to obtain both estimates are uncorrelated, we can minimize expected variance of the interpolated gradient:

$$\min_{\alpha \in [0,1]} \quad \alpha^2 \hat{\sigma}_1^2 + (1 - \alpha)^2 \hat{\sigma}_0^2 \qquad\qquad \alpha^* = \frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2 + \hat{\sigma}_0^2}$$

But recall the empirical bias phenomenon….empirical variance can be misleading!

Robust Interpolation

Consider an interpolated gradient of the two objectives. How should we choose alpha?

Definition 4.1. Given $\alpha \in [0, 1]$, we define the alpha-order batched gradient (AoBG) as:

$$\bar{\nabla}^{[\alpha]} F(\boldsymbol{\theta}) = \alpha \bar{\nabla}^{[1]} F(\boldsymbol{\theta}) + (1 - \alpha) \bar{\nabla}^{[0]} F(\boldsymbol{\theta}).$$

How do we achieve robust interpolation to potential bias of the FoBG?

1. We know the ZoBG is always unbiased.
2. We can unit-test the FoBG against the ZoBG based on some confidence statistics of the ZoBG.

Previous Interpolation

$$\min_{\alpha \in [0,1]} \quad \alpha^2 \hat{\sigma}_1^2 + (1 - \alpha)^2 \hat{\sigma}_0^2$$

Robust Interpolation

Consider an interpolated gradient of the two objectives. How should we choose alpha?

Definition 4.1. Given $\alpha \in [0, 1]$, we define the alpha-order batched gradient (AoBG) as:

$$\bar{\nabla}^{[\alpha]} F(\boldsymbol{\theta}) = \alpha \bar{\nabla}^{[1]} F(\boldsymbol{\theta}) + (1 - \alpha) \bar{\nabla}^{[0]} F(\boldsymbol{\theta}).$$

How do we achieve robust interpolation to potential bias of the FoBG?

1. We know the ZoBG is always unbiased.
2. We can unit-test the FoBG against the ZoBG based on some confidence statistics of the ZoBG.

Previous Interpolation

$$\min_{\alpha \in [0,1]} \quad \alpha^2 \hat{\sigma}_1^2 + (1 - \alpha)^2 \hat{\sigma}_0^2$$

Robust Interpolation

$$\min_{\alpha \in [0,1]} \quad \alpha^2 \hat{\sigma}_1^2 + (1 - \alpha)^2 \hat{\sigma}_0^2$$

$$\text{s.t.} \quad \epsilon + \alpha \underbrace{\left\| \bar{\nabla}^{[1]} F - \bar{\nabla}^{[0]} F \right\|}_{B} \leq \gamma.$$

Confidence interval on the ZoBG estimate.

User-defined threshold on allowable bias of FoBG.

Robust Interpolation

How do we achieve robust interpolation to potential bias of the FoBG?

1. We know the ZoBG is always unbiased.
2. We can unit-test the FoBG against the ZoBG based on some confidence statistics of the ZoBG.

Robust Interpolation

$$\begin{aligned} \min_{\alpha\in[0,1]} \quad & \alpha^2\hat{\sigma}_1^2 + (1-\alpha)^2\hat{\sigma}_0^2 \\ \text{s.t.} \quad & \epsilon + \alpha \underbrace{\left\|\bar{\nabla}^{[1]}F - \bar{\nabla}^{[0]}F\right\|}_{B} \leq \gamma. \end{aligned}$$

Confidence interval on the ZoBG estimate.

User-defined threshold on allowable bias of FoBG.

This constraint enforces a chance constraint on the allowable bias of the FoBG.

Robust Interpolation

How do we achieve robust interpolation to potential bias of the FoBG?

1. We know the ZoBG is always unbiased.
2. We can unit-test the FoBG against the ZoBG based on some confidence statistics of the ZoBG.

Robust Interpolation

$$\begin{aligned}\min_{\alpha\in[0,1]} \quad & \alpha^2\hat{\sigma}_1^2 + (1-\alpha)^2\hat{\sigma}_0^2 \\ \text{s.t.} \quad & \epsilon + \alpha\underbrace{\|\bar{\nabla}^{[1]}F - \bar{\nabla}^{[0]}F\|}_{B} \leq \gamma.\end{aligned}$$

Confidence interval on the ZoBG estimate.

User-defined threshold on allowable bias of FoBG.

This constraint enforces a chance constraint on the allowable bias of the FoBG.

Definition 4.2 (Accuracy). α is (γ, δ)-accurate if the bound on the *error* of AoBG is satisfied with probability δ:

$$\|\bar{\nabla}^{[\alpha]}F(\boldsymbol{\theta}) - \nabla F(\boldsymbol{\theta})\| \leq \gamma. \tag{3}$$

Lemma 4.3 (Robustness). *Suppose that $\epsilon + \alpha B \leq \gamma$ with probability δ. Then, α is (γ, δ)-accurate.*

Robust Interpolation

How do we achieve robust interpolation to potential bias of the FoBG?

1. We know the ZoBG is always unbiased.
2. We can unit-test the FoBG against the ZoBG based on some confidence statistics of the ZoBG.

Robust Interpolation

$$\min_{\alpha\in[0,1]} \quad \alpha^2\hat{\sigma}_1^2 + (1-\alpha)^2\hat{\sigma}_0^2$$

$$\text{s.t.} \quad \epsilon + \alpha\underbrace{\|\bar{\nabla}^{[1]}F - \bar{\nabla}^{[0]}F\|}_{B} \leq \gamma.$$

Confidence interval on the ZoBG estimate.

User-defined threshold on allowable bias of FoBG.

This constraint enforces a chance constraint on the allowable bias of the FoBG.

Definition 4.2 (Accuracy). α is (γ,δ)-accurate if the bound on the *error* of AoBG is satisfied with probability δ:

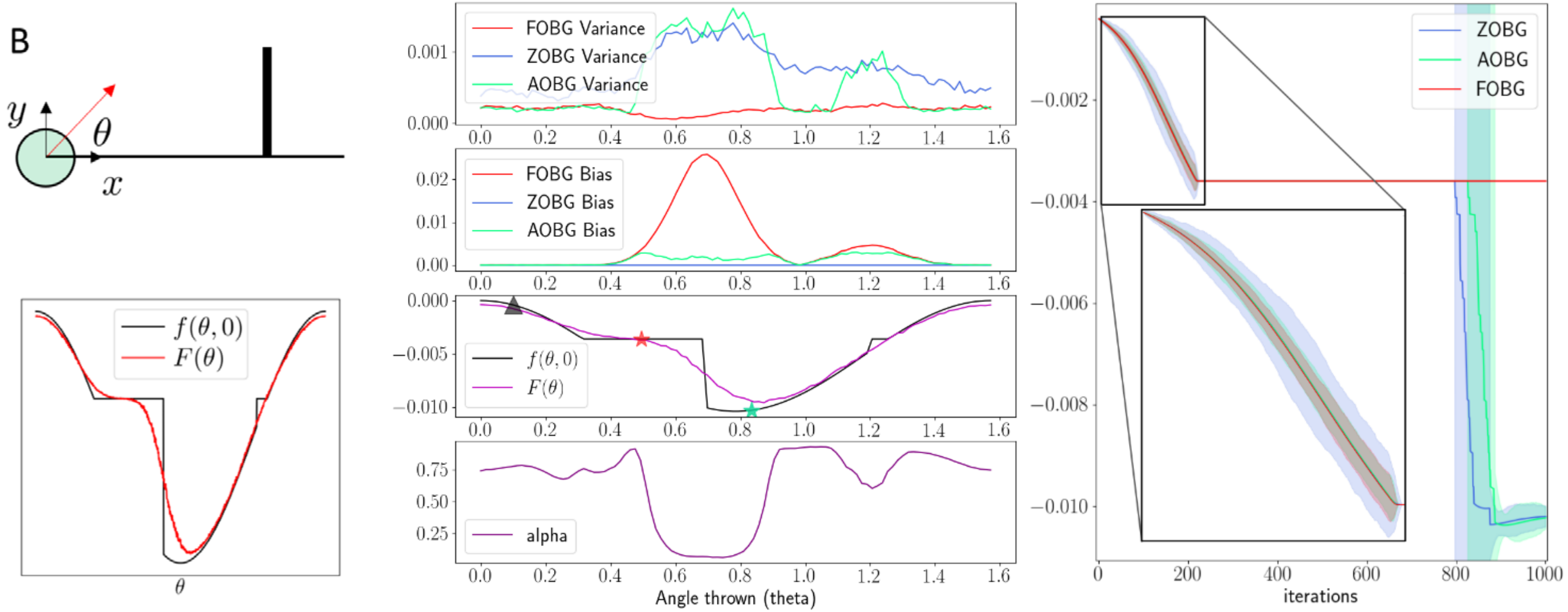
$$\|\bar{\nabla}^{[\alpha]}F(\boldsymbol{\theta}) - \nabla F(\boldsymbol{\theta})\| \leq \gamma. \tag{3}$$

Lemma 4.3 (Robustness). *Suppose that $\epsilon + \alpha B \leq \gamma$ with probability δ. Then, α is (γ,δ)-accurate.*

Lemma 4.4. *With $\gamma=\infty$, the optimal α is $\alpha_\infty := \frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2+\hat{\sigma}_0^2}$. For finite $\gamma \geq \epsilon$, Eq* (4) *is*

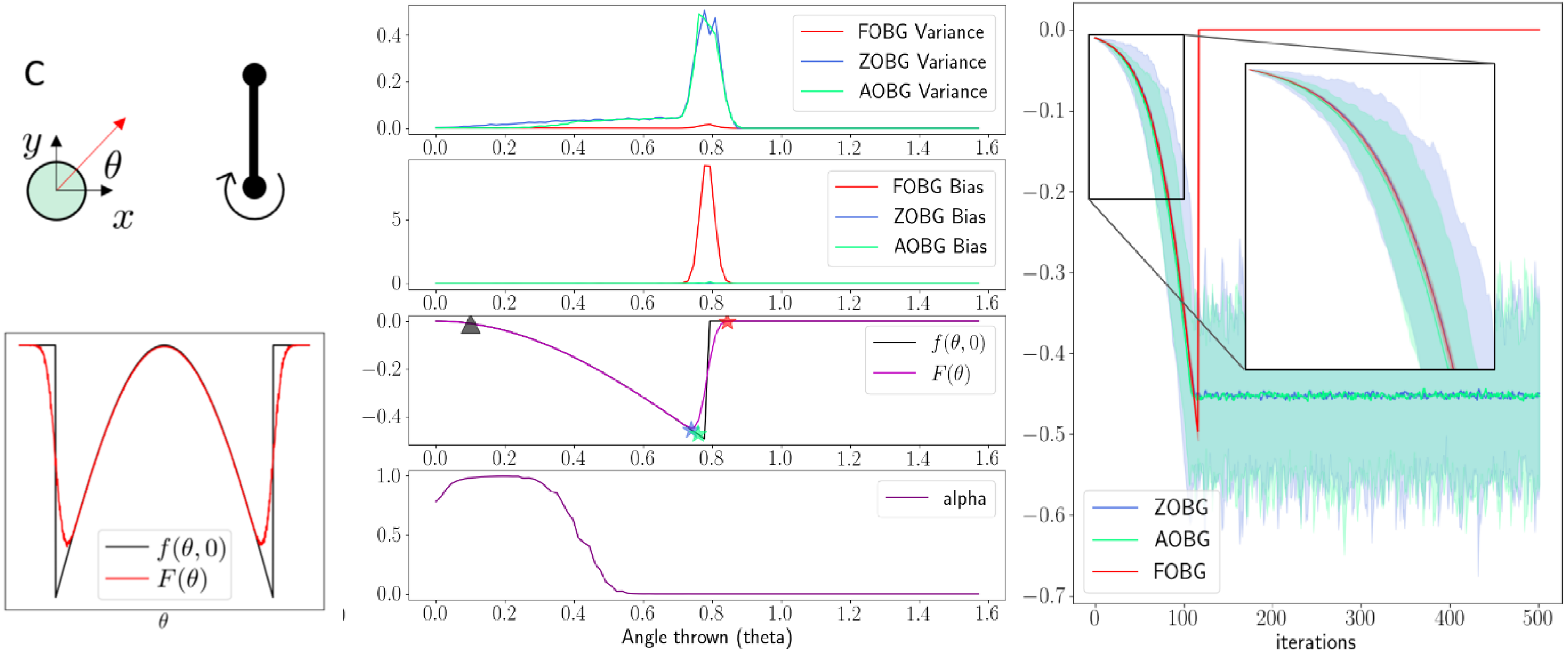
$$\alpha_\gamma := \begin{cases} \alpha_\infty & \textit{if} \quad \alpha_\infty B \leq \gamma - \varepsilon \\ \frac{\gamma-\varepsilon}{B} & \textit{otherwise} \; . \end{cases} \tag{5}$$

Results on Robust Interpolation Gradient



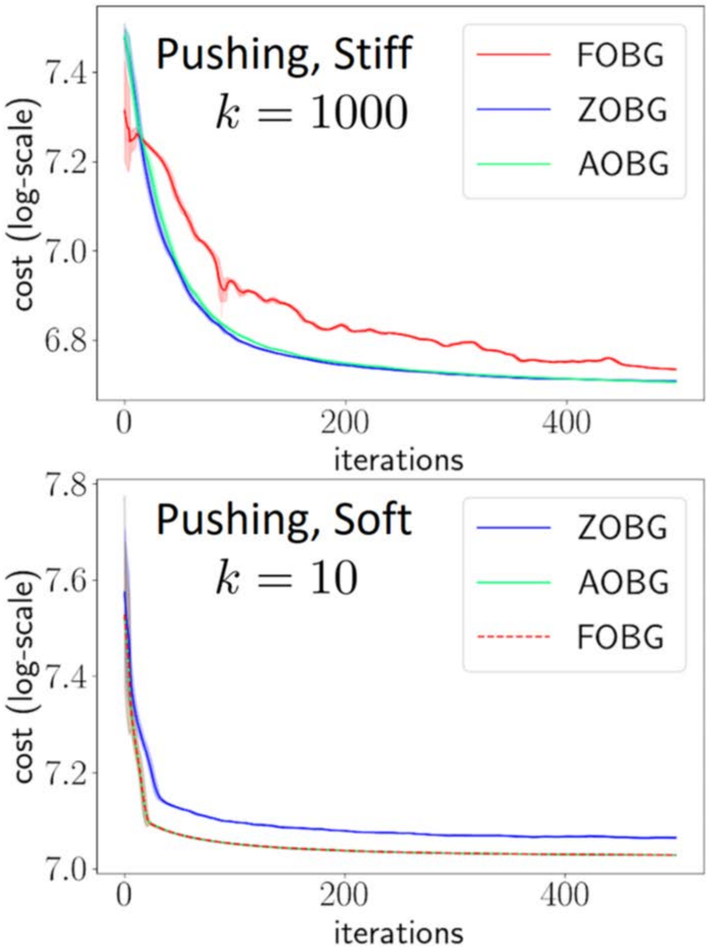
1. The AoBG obeys some threshold on bias.
2. The variance of AoBG is between FoBG and ZoBG – on some coordinates results in lower bias.
3. NOTE: bias-variance characteristics not only result in convergence-rate arguments, but result in different local minima.

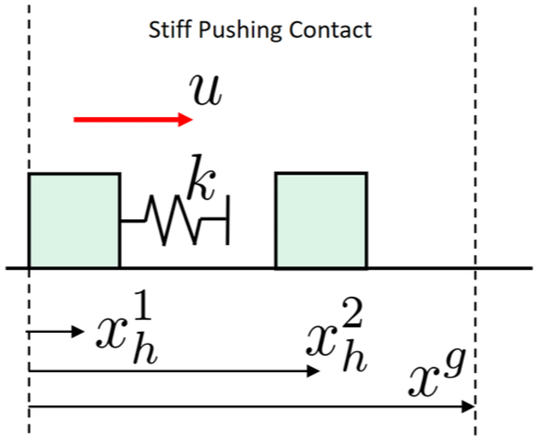
Results on Robust Interpolation Gradient



1. The AoBG obeys some threshold on bias.
2. The variance of AoBG is between FoBG and ZoBG – on some coordinates results in lower bias.
3. NOTE: bias-variance characteristics not only result in convergence-rate arguments, but result in different local minima.

Scaling Results on Practical Examples

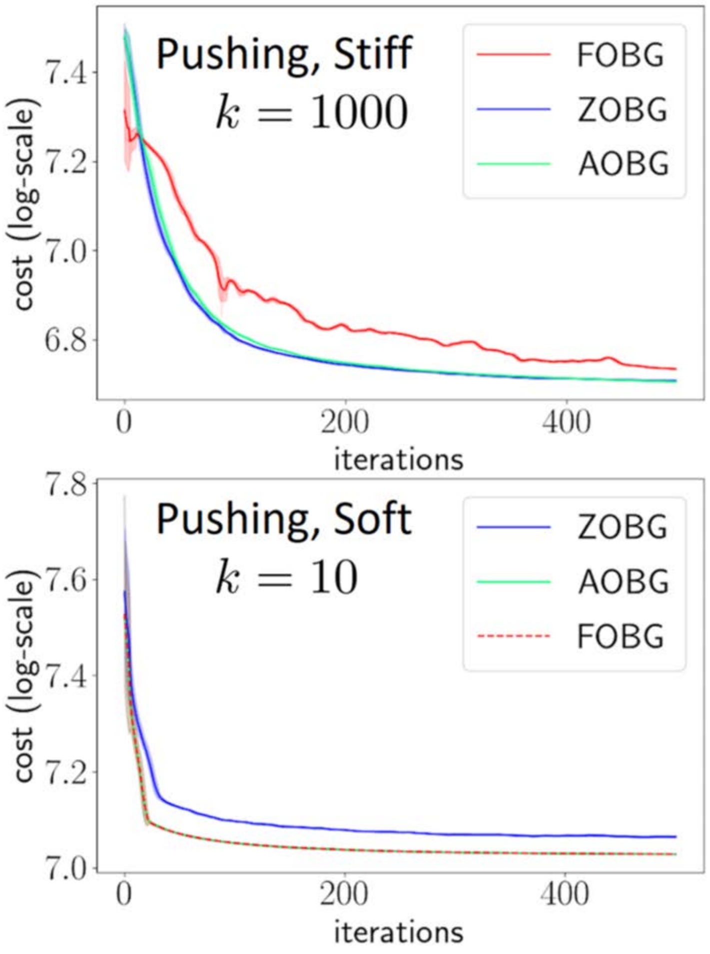


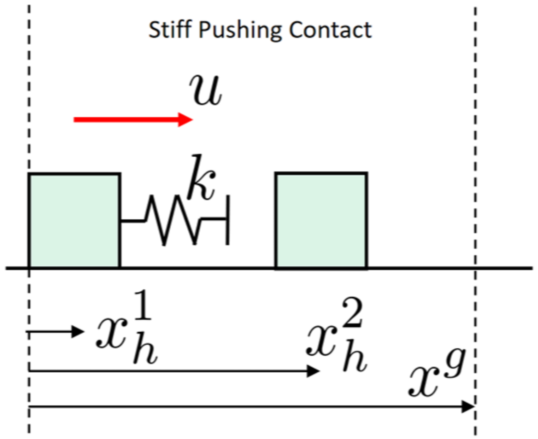


Trajectory Optimization Example

How does performance of gradient descent with different estimators perform as we increase the stiffness of contact?

Scaling Results on Practical Examples



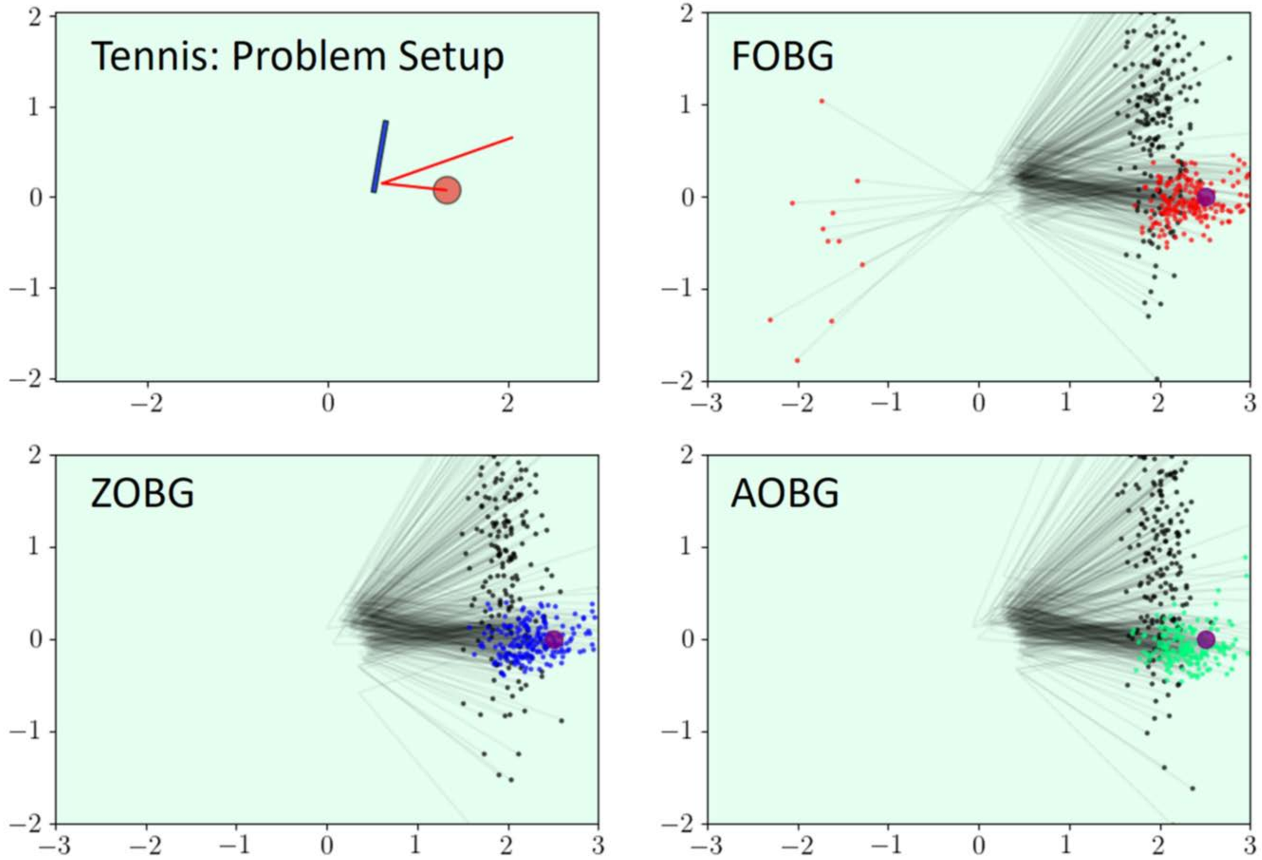


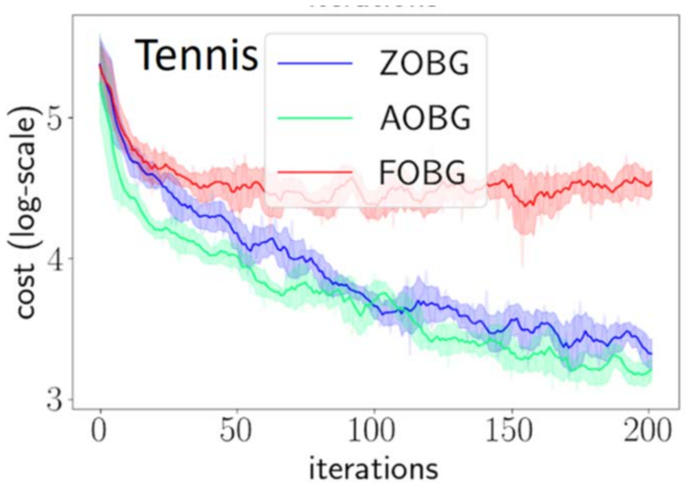
Trajectory Optimization Example

How does performance of gradient descent with different estimators perform as we increase the stiffness of contact?

1. FoBG results in worse performance as we increase stiffness
2. ZoBG results in worse performance for softer systems
3. AoBG automates the procedure of selecting between the two.

Scaling Results on Practical Examples

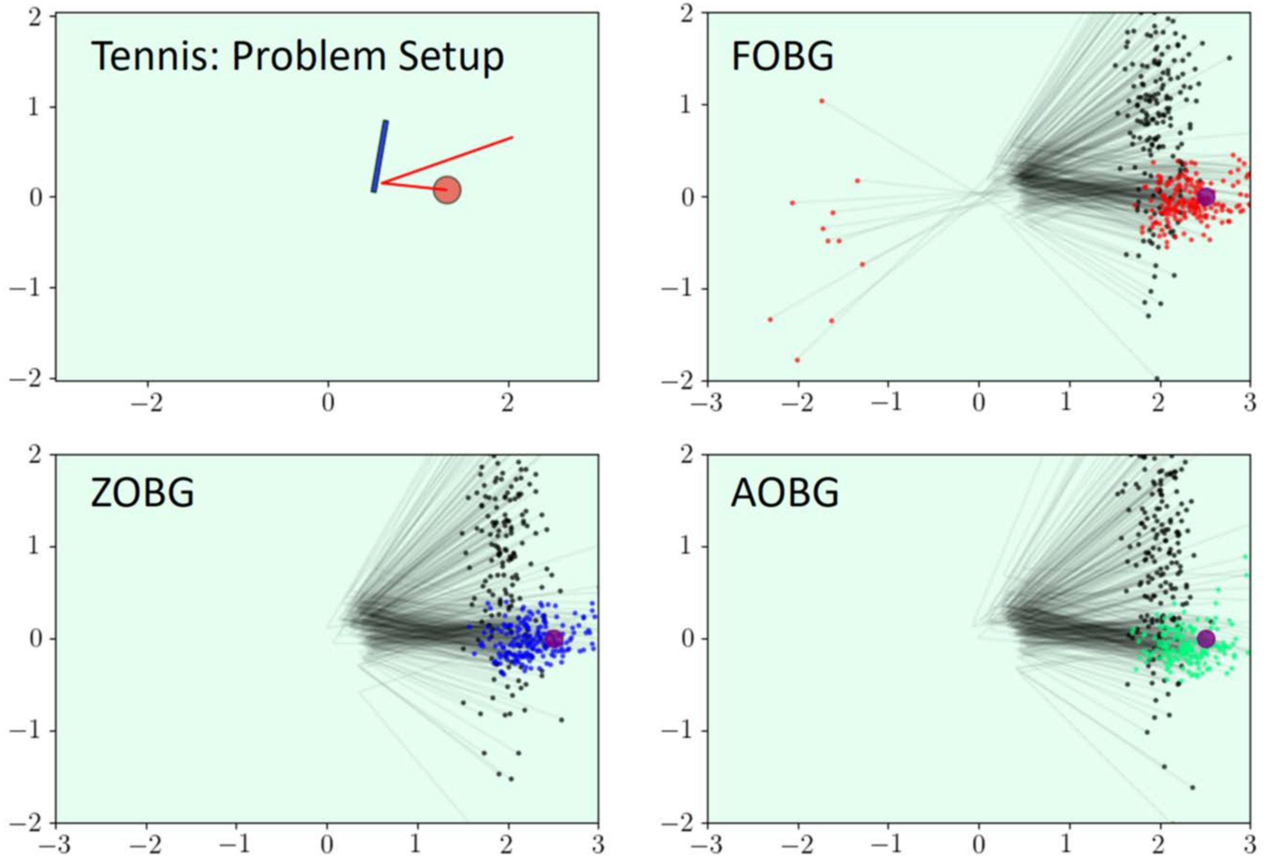


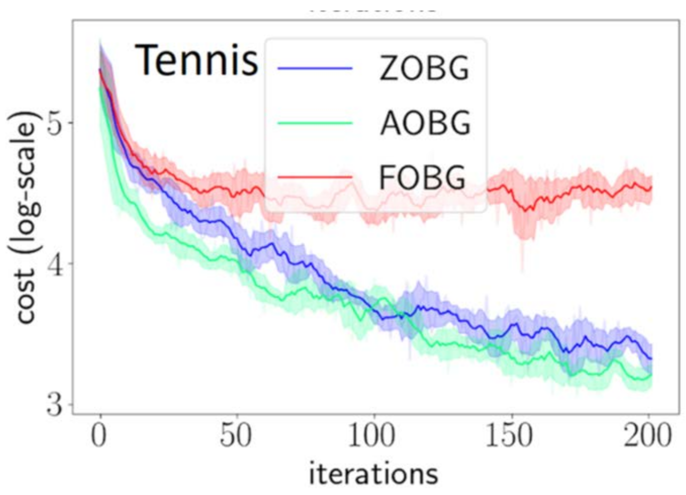


Policy Optimization Example

How do different policy gradients perform on policy optimization?

Scaling Results on Practical Examples





Policy Optimization Example

How do different policy gradients perform on policy optimization?

1. FoBG does worse the ZoBG asymptotically.
2. AoBG descends down faster than ZoBG
3. However, a wide enough distribution will contain discontinuities, and AoBG will tend to utilize ZoBG more. Limitation of the method.