Simultaneous perturbation methods for stochastic non-convex optimization

Prashanth L.A.

Department of Computer Science & Engg
Indian Institute of Technology Madras
prashla@cse.iitm.ac.in

Outline

Simulation optimization: problem setting, practical motivation, challenges

First-order methods: gradient estimation, (near) unbiasednes, convergence

Second-order methods: why?, Hessian estimation, (near) unbiasednes, convergence

Applications: Service systems, transportation

Motivation


I DID
NOTHING
TODAY AND
STILL GOT PAID

Application I: Service System

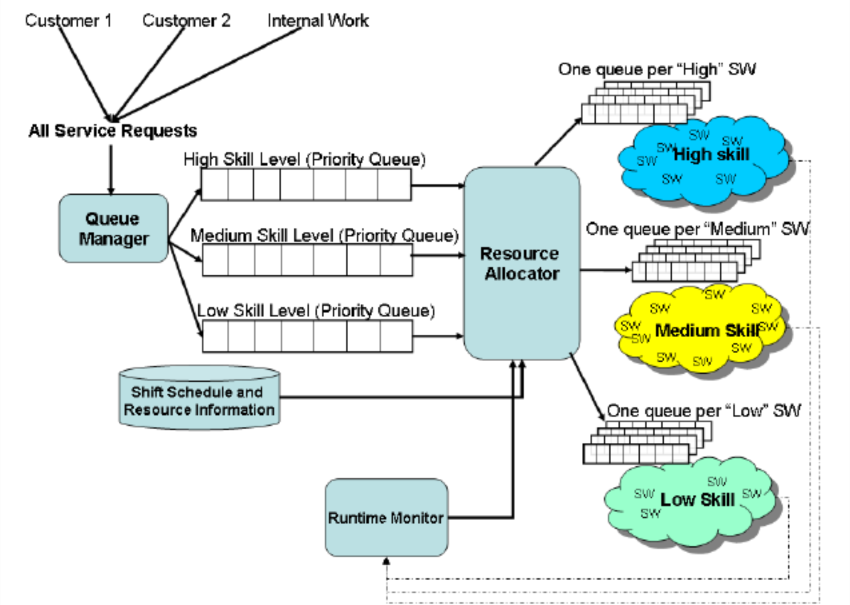


Table 1: Workers $W_{i,j}$

	Skill levels		
Shift	High	Med	Low
S1	1	3	7
S2	0	5	2
S3	3	1	2

Table 2: SLA targets $\gamma_{i,j}$

	Customers	
Priority	Bossy Corp	Cool Inc
P_1	4h	5h
P_2	8h	12h
P_3	24h	48h
P_4	18h	144h

Aim: Find the optimal number of workers for each shift and of each skill level

- that minimizes the labor cost and
- satisfies SLA requirements

On a good day, the traffic is . . .

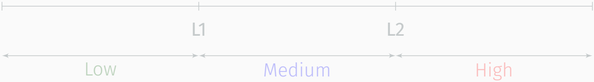
And on a bad day, it can be . . .

Aim: Maximize traffic flow

Input:
Coarse congestion estimates

Output:
Policy for switching traffic lights

Input: Coarse congestion estimates
Sensor loops at two points along the road



How to switch traffic lights given L1 and L2?

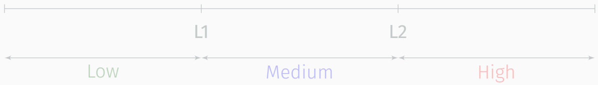
How to choose L1 and L2 for a given policy and road network?

Aim: Maximize traffic flow

Input:
Coarse congestion estimates

Output:
Policy for switching traffic lights

Input: Coarse congestion estimates
Sensor loops at two points along the road



How to switch traffic lights given L1 and L2?

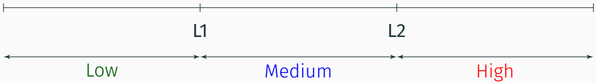
How to choose L1 and L2 for a given policy and road network?

Aim: Maximize traffic flow

Input:
Coarse congestion estimates

Output:
Policy for switching traffic lights

Input: Coarse congestion estimates
Sensor loops at two points along the road



How to switch traffic lights given L1 and L2?

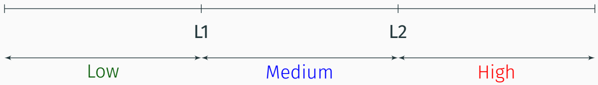
How to choose L1 and L2 for a given policy and road network?

Aim: Maximize traffic flow

Input:
Coarse congestion estimates

Output:
Policy for switching traffic lights

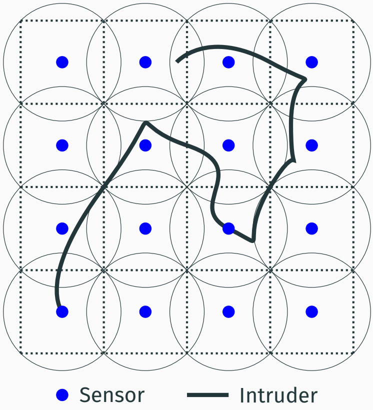
Input: Coarse congestion estimates
Sensor loops at two points along the road



How to switch traffic lights given L1 and L2?

How to choose L1 and L2 for a given policy and road network?

Application III: Intrusion detection using sensor networks



Aim:

- minimize the energy consumption of the sensors, while
- keeping tracking error to a minimum

Common application traits

Stochastic:
noisy observations

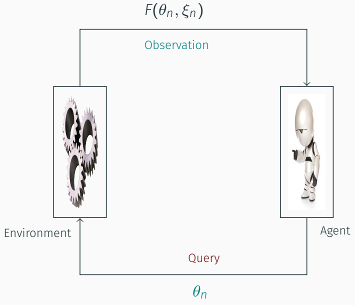
Model-free:
sample access to objective
* gradients unavailable

High-dimensional:
brute-force search infeasible

Solution:
Simultaneous perturbation methods

The framework

Basic optimization problem



Aim: $$\theta^* = \arg\min_{\theta \in \Theta} \left\{ f(\theta) \triangleq \mathbb{E}[F(\theta, \xi)] \right\},$$

- $f\colon \mathbb{R}^N \to \mathbb{R}$ is the performance measure
  - f *not* assumed to be convex
- $F(\theta, \xi)$ is the sample performance
- ξ is the noise factor that captures stochastic nature of the problem
- θ is the (vector) parameter of interest
- $\Theta \subseteq \mathbb{R}^N$ is the feasible region in which θ takes values.

Stochastic optimization via simulation

Stochastic optimization deals with highly nonlinear and high dimensional systems. The challenges with these problems are:

- Too complex to solve analytically.
- Many simplifying assumptions are required.

A good alternative of modeling and analysis is "Simulation"

Figure 1: Simulation optimization

Stochastic optimization via simulation

Stochastic optimization deals with highly nonlinear and high dimensional systems. The challenges with these problems are:

- Too complex to solve analytically.
- Many simplifying assumptions are required.

A good alternative of modeling and analysis is ”Simulation”

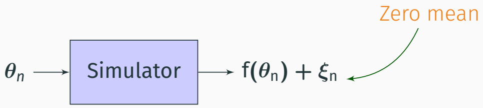


Figure 1: Simulation optimization

Noise controls

Recall: $f(\theta) = \mathbb{E}[F(\theta, \xi)]$.

Two settings for noise:

Controlled noise: ξ can be kept fixed between queries to obtain $F(\theta_1, \xi)$ and $F(\theta_2, \xi)$

Uncontrolled noise: $F(\theta, \xi)$ can be obtained at any point, but ξ is not controllable

Noise controls

Recall: $f(\theta) = \mathbb{E}[F(\theta, \xi)]$.

Two settings for noise:

Controlled noise ξ can be kept fixed between queries to obtain $F(\theta_1, \xi)$ and $F(\theta_2, \xi)$

Uncontrolled noise $F(\theta, \xi)$ can be obtained at any point, but ξ is not controllable

Challenges in simulation optimization

Deterministic optimization problem

- focus is on search for better solutions
- Complete information about objective function f, esp. gradients

Stochastic optimization problem

- f cannot be obtained directly, but we are given sample access, i.e.,
  $f(\theta) = E_\xi[F(\theta, \xi)]$
- Each sample $F(\theta, \xi)$ is obtained from an expensive simulation experiment or a (real) field test
- focus is on both search and evaluation
  - Tradeoff between evaluating better vs. finding more candidate solutions

Challenge: to find $\theta^* = \arg\min_{\theta \in \Theta} f(\theta)$, given only noisy function evaluations.

Challenges in simulation optimization

Deterministic optimization problem

- focus is on search for better solutions
- Complete information about objective function f, esp. gradients

Stochastic optimization problem

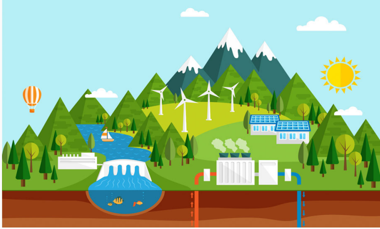
- f cannot be obtained directly, but we are given sample access, i.e., $f(\theta) \equiv E_{\xi}[F(\theta, \xi)]$
- Each sample $F(\theta, \xi)$ is obtained from an expensive simulation experiment or a (real) field test
- focus is on both search and evaluation
  - Tradeoff between evaluating better vs. finding more candidate solutions

Challenge: to find $\theta^* = \arg\min_{\theta \in \Theta} f(\theta)$, given only noisy function evaluations.

Some more applications

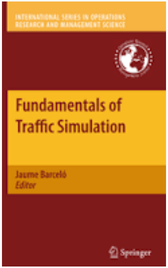
Energy Demand management

- Consumer demand, energy generation are uncertain.
- Objective is to minimize the difference.



Transportation

- Car-following model
- route choice
- traffic assignment model



Applications (contd)

- Service systems
  - banks, restaurants, call centers, amusement parks
- Transportation systems
  - airports: air space, runways, baggage, roads, queues
- Manufacturing systems
- Semiconductor fabrication facilities
- Supply chain management
- Communication networks
- Financial systems
  - risk management, retirement planning (portfolio opt)

Applications (contd)

- Service systems
  - banks, restaurants, call centers, amusement parks
- Transportation systems
  - airports: air space, runways, baggage, roads, queues
- Manufacturing systems
- Semiconductor fabrication facilities
- Supply chain management
- Communication networks
- Financial systems
  - risk management, retirement planning (portfolio opt)

Applications (contd)

- Service systems
  - banks, restaurants, call centers, amusement parks
- Transportation systems
  - airports: air space, runways, baggage, roads, queues
- Manufacturing systems
- Semiconductor fabrication facilities
- Supply chain management
- Communication networks
- Financial systems
  - risk management, retirement planning (portfolio opt)

Applications (contd)

- Service systems
  - banks, restaurants, call centers, amusement parks
- Transportation systems
  - airports: air space, runways, baggage, roads, queues
- Manufacturing systems
- Semiconductor fabrication facilities
- Supply chain management
- Communication networks
- Financial systems
  - risk management, retirement planning (portfolio opt)

Applications (contd)

- Service systems
  - banks, restaurants, call centers, amusement parks
- Transportation systems
  - airports: air space, runways, baggage, roads, queues
- Manufacturing systems
- Semiconductor fabrication facilities
- Supply chain management
- Communication networks
- Financial systems
  - risk management, retirement planning (portfolio opt)

Applications (contd)

- Service systems
  - banks, restaurants, call centers, amusement parks
- Transportation systems
  - airports: air space, runways, baggage, roads, queues
- Manufacturing systems
- Semiconductor fabrication facilities
- Supply chain management
- Communication networks
- Financial systems
  - risk management, retirement planning (portfolio opt)

Applications (contd)

- Service systems
    - banks, restaurants, call centers, amusement parks
- Transportation systems
    - airports: air space, runways, baggage, roads, queues
- Manufacturing systems
- Semiconductor fabrication facilities
- Supply chain management
- Communication networks
- Financial systems
    - risk management, retirement planning (portfolio opt)

Applications (contd)

- Service systems
  - banks, restaurants, call centers, amusement parks
- Transportation systems
  - airports: air space, runways, baggage, roads, queues
- Manufacturing systems
- Semiconductor fabrication facilities
- Supply chain management
- Communication networks
- Financial systems
  - risk management, retirement planning (portfolio opt)

Applications (contd)

- Service systems
    - banks, restaurants, call centers, amusement parks
- Transportation systems
    - airports: air space, runways, baggage, roads, queues
- Manufacturing systems
- Semiconductor fabrication facilities
- Supply chain management
- Communication networks
- Financial systems
    - risk management, retirement planning (portfolio opt)

Applications (contd)

- Service systems
  - banks, restaurants, call centers, amusement parks
- Transportation systems
  - airports: air space, runways, baggage, roads, queues
- Manufacturing systems
- Semiconductor fabrication facilities
- Supply chain management
- Communication networks
- Financial systems
  - risk management, retirement planning (portfolio opt)

Some real-world examples

- Kroger (Edelman 2013 finalist, gradient-based) Kroger Uses Simulation-Optimization to Improve Pharmacy Inventory Management
    - `www.youtube.com/watch?v=BNyDbBy-KYY` (start at 0:45)
    - `https://www.informs.org/About-INFORMS/News-Room/Press-Releases/Edelman-2013-Announcement`

      The Franz Edelman Award recognizes outstanding examples of innovative operations research and analytics that improves organizations and often change people's lives.

- Financial engineering
    - Monte Carlo simulation used widely on Wall Street.
    - Gradient estimates needed for hedging.
    - Hot research area: several research papers continue to be published

First-order methods

Stochastic analog of gradient descent

$$\theta_{n+1} = \theta_n - a_n G_n. \tag{1}$$

Suppose that

- G_n is an noisy estimate of the gradient $\nabla f(\theta_n)$, i.e., $\mathbb{E}(G_n) = \nabla f(\theta_n)$.
- $\{a_n\}$ are pre-determined step-sizes satisfying:
  $$\sum_{n=1}^{\infty} a_n = \infty, \quad \sum_{n=1}^{\infty} a_n^2 < \infty$$
- iterates are stable: $\sup_n \|\theta_n\| < \infty$.

Theorem (Variant of Robbins Monro stochastic approximation)
Letting $K := \{\theta \mid \nabla f(\theta) = 0\}$, we have

$$\theta_n \to K \text{ a.s. as } n \to \infty.$$

Stochastic analog of gradient descent

$$\theta_{n+1} = \theta_n - a_n G_n. \tag{1}$$

Suppose that

- G_n is an noisy estimate of the gradient $\nabla f(\theta_n)$, i.e., $\mathbb{E}(G_n) = \nabla f(\theta_n)$.
- $\{a_n\}$ are pre-determined step-sizes satisfying: $$\sum_{n=1}^{\infty} a_n = \infty, \quad \sum_{n=1}^{\infty} a_n^2 < \infty$$
- iterates are stable: $\sup_n \|\theta_n\| < \infty$.

Theorem (Variant of Robbins Monro stochastic approximation)
Letting $K := \{\theta \mid \nabla f(\theta) = 0\}$, we have

$$\theta_n \to K \text{ a.s. as } n \to \infty.$$

Stochastic analog of gradient descent

$$\theta_{n+1} = \theta_n - a_n G_n. \tag{1}$$

Suppose that

- G_n is an noisy estimate of the gradient $\nabla f(\theta_n)$, i.e.,
  $\mathbb{E}(G_n) = \nabla f(\theta_n)$.
- $\{a_n\}$ are pre-determined step-sizes satisfying:
  $$\sum_{n=1}^{\infty} a_n = \infty, \quad \sum_{n=1}^{\infty} a_n^2 < \infty$$
- iterates are stable: $\sup_n \|\theta_n\| < \infty$.

Theorem (Variant of Robbins Monro stochastic approximation)
Letting $K := \{\theta \mid \nabla f(\theta) = 0\}$, we have

$$\theta_n \to K \text{ a.s. as } n \to \infty.$$

Stochastic analog of gradient descent

$$\theta_{n+1} = \theta_n - a_n G_n. \tag{1}$$

Suppose that

- G_n is an noisy estimate of the gradient $\nabla f(\theta_n)$, i.e., $\mathbb{E}(G_n) = \nabla f(\theta_n)$.
- $\{a_n\}$ are pre-determined step-sizes satisfying:
  $$\sum_{n=1}^{\infty} a_n = \infty, \quad \sum_{n=1}^{\infty} a_n^2 < \infty$$
- iterates are stable: $\sup_n \|\theta_n\| < \infty$.

Theorem (Variant of Robbins Monro stochastic approximation)
Letting $K := \{\theta \mid \nabla f(\theta) = 0\}$, we have

$$\theta_n \to K \text{ a.s. as } n \to \infty.$$

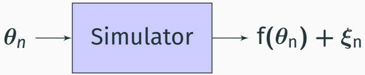
$$\theta_{n+1} = \theta_n - a_n G_n. \quad (2)$$

How to keep iterates stable?

Project θ_n onto a compact and convex set Θ $\leftarrow$ Projected stochastic approximation

$$\theta_{n+1} = \theta_n - a_n G_n. \tag{2}$$

How to estimate the gradient of f from samples?



Simultaneous perturbation methods.

Stochastic approximation (SA) alphabet soup

- **FDSA** Finite difference stochastic approximation
- **SPSA** Simultaneous perturbation stochastic approximation
- **SFSA** Smoothed functional stochastic approximation
- **RDSA** Random direction stochastic approximation

In the next few slides . . .

Q1) How to form G_n from function samples so that $G_n \approx \nabla f(\theta_n)$

Q2) Such a G_n - is it unbiased?

Q3) Does $\theta_{n+1} = \theta_n - a_n G_n$ converge to θ^* with such a G_n?

Q4) If answer is yes to above, what is the convergence rate?

Outline

Motivation

The framework

First-order methods

How are Gradients Estimated?

Analysis

Commercials

Second-order methods

Applications

Perfect measurements ⇔ No noise

Finite-difference stochastic approximation (FDSA) (Kiefer and Wolfowitz, 1952):

$$g^i = \frac{1}{\delta}\left(f(\theta + \delta e_i) - f(\theta)\right) , \quad i = 1, \dots, N .$$

Assume $f \in \mathcal{C}^3$

Taylor-series expansion:

$$f(\theta + \delta e_i) = f(\theta) + \delta \, \nabla f(\theta) e_i + \delta^2 \, e_i^\top \nabla^2 f(\theta) e_i + O(\delta^3).$$

Accuracy: $\|g - \nabla f(\theta)\|_2 = O(\delta)$.

Needs $N + 1$ queries.

Perfect measurements ⇔ No noise

Finite-difference stochastic approximation (FDSA) (Kiefer and Wolfowitz, 1952):

$$g^i = \frac{1}{\delta}\left(f(\theta + \delta e_i) - f(\theta)\right) , \quad i = 1, \ldots, N \, .$$

Assume $f \in \mathcal{C}^3$

Taylor-series expansion:

$$f(\theta + \delta e_i) = f(\theta) + \delta \, \nabla f(\theta) e_i + \delta^2 \, e_i^\top \nabla^2 f(\theta) e_i + O(\delta^3).$$

Accuracy: $\|g - \nabla f(\theta)\|_2 = O(\delta)$.

Needs $N + 1$ queries.

FDSA with two-sided Differences

Improved estimate:

$$g^i = \frac{1}{2\delta}\left(f(\theta + \delta e_i) - f(\theta - \delta e_i)\right), \quad i = 1, \ldots, N.$$

Taylor-series expansions:

$$f(\theta + \delta e_i) = f(\theta) + \delta\, \nabla f(\theta) e_i + \delta^2\, e_i^\top \nabla^2 f(\theta) e_i + O(\delta^3).$$

$$f(\theta - \delta e_i) = f(\theta) - \delta\, \nabla f(\theta) e_i + \delta^2\, e_i^\top \nabla^2 f(\theta) e_i + O(\delta^3).$$

Accuracy: $\|g - \nabla f(\theta)\|_2 = O(\delta^2)$.

Needs $2N$ queries.

FDSA with two-sided Differences

Improved estimate:

$$g^i = \frac{1}{2\delta}\left(f(\theta + \delta e_i) - f(\theta - \delta e_i)\right), \quad i = 1, \ldots, N.$$

Taylor-series expansions:

$$f(\theta + \delta e_i) = f(\theta) + \delta \nabla f(\theta) e_i + \delta^2 e_i^\top \nabla^2 f(\theta) e_i + O(\delta^3).$$

$$f(\theta - \delta e_i) = f(\theta) - \delta \nabla f(\theta) e_i + \delta^2 e_i^\top \nabla^2 f(\theta) e_i + O(\delta^3).$$

Accuracy: $\|g - \nabla f(\theta)\|_2 = O(\delta^2)$.

Needs $2N$ queries.

FDSA + Two-sided Differences + Noise

Improved estimate:

$$G^i = \frac{1}{2\delta}\left\{f(\theta+\delta e_i)+\xi_i^+ - (f(\theta-\delta e_i)+\xi_i^-)\right\}, \quad i = 1,\dots,N.$$

Taylor-series expansions:

$$f(\theta+\delta e_i) = f(\theta) + \delta\,\nabla f(\theta) e_i + \delta^2\, e_i^\top \nabla^2 f(\theta) e_i + O(\delta^3).$$
$$f(\theta-\delta e_i) = f(\theta) - \delta\,\nabla f(\theta) e_i + \delta^2\, e_i^\top \nabla^2 f(\theta) e_i + O(\delta^3).$$

Assumption: $\mathbb{E}\left[\xi^\pm\right] = 0$, $\mathbb{E}\left[(\xi^\pm)\right] \le \sigma^2 < +\infty$.

$\mathbb{E}\left[G^i\right] = g^i$. Hence

$$\left\|\mathbb{E}\left[G\right] - \nabla f(\theta)\right\|_2 = O(\delta^2)\,. \longleftarrow \text{bias}$$

Needs $2N$ queries.

FDSA + Two-sided Differences + Noise

Improved estimate:

$$G^i = \frac{1}{2\delta}\left\{f(\theta+\delta e_i) + \xi_i^+ - (f(\theta-\delta e_i)+\xi_i^-)\right\}, \quad i = 1,\dots,N.$$

Taylor-series expansions:

$$f(\theta+\delta e_i) = f(\theta) + \delta\,\nabla f(\theta)e_i + \delta^2\, e_i^\top \nabla^2 f(\theta) e_i + O(\delta^3).$$
$$f(\theta-\delta e_i) = f(\theta) - \delta\,\nabla f(\theta)e_i + \delta^2\, e_i^\top \nabla^2 f(\theta) e_i + O(\delta^3).$$

Assumption: $\mathbb{E}\left[\xi^\pm\right] = 0$, $\mathbb{E}\left[(\xi^\pm)\right] \le \sigma^2 < +\infty$.

$\mathbb{E}\left[G^i\right] = g^i$. Hence

$$\left\|\mathbb{E}\left[G\right] - \nabla f(\theta)\right\|_2 = O(\delta^2)\,. \longleftarrow \text{bias}$$

Needs $2N$ queries.

FDSA + Two-sided Differences + Noise

Improved estimate:

$$G^i = \frac{1}{2\delta} \left\{ f(\theta + \delta e_i) + \xi_i^+ - (f(\theta - \delta e_i) + \xi_i^-) \right\}, \quad i = 1, \ldots, N.$$

Taylor-series expansions:

$$f(\theta + \delta e_i) = f(\theta) + \delta \nabla f(\theta) e_i + \delta^2 e_i^\top \nabla^2 f(\theta) e_i + O(\delta^3).$$
$$f(\theta - \delta e_i) = f(\theta) - \delta \nabla f(\theta) e_i + \delta^2 e_i^\top \nabla^2 f(\theta) e_i + O(\delta^3).$$

Assumption: $\mathbb{E}\left[\xi^\pm\right] = 0$, $\mathbb{E}\left[(\xi^\pm)\right] \leq \sigma^2 < +\infty$.

$\mathbb{E}\left[G^i\right] = g^i$. Hence

$$\left\|\mathbb{E}\left[G\right] - \nabla f(\theta)\right\|_2 = O(\delta^2)\,. \longleftarrow \textbf{bias}$$

Needs $2N$ queries.

FDSA + Two-sided Differences + Noise

Improved estimate:

$$G^i = \frac{1}{2\delta}\left\{f(\theta+\delta e_i)+\xi_i^+ - (f(\theta-\delta e_i)+\xi_i^-)\right\}, \quad i=1,\dots,N.$$

Taylor-series expansions:

$$f(\theta+\delta e_i) = f(\theta) + \delta\,\nabla f(\theta) e_i + \delta^2\, e_i^\top \nabla^2 f(\theta) e_i + O(\delta^3).$$
$$f(\theta-\delta e_i) = f(\theta) - \delta\,\nabla f(\theta) e_i + \delta^2\, e_i^\top \nabla^2 f(\theta) e_i + O(\delta^3).$$

Assumption: $\mathbb{E}\left[\xi^\pm\right] = 0$, $\mathbb{E}\left[(\xi^\pm)\right] \le \sigma^2 < +\infty$.

$\mathbb{E}\left[G^i\right] = g^i$. Hence

$$\|\mathbb{E}[G] - \nabla f(\theta)\|_2 = O(\delta^2)\,. \longleftarrow \text{bias}$$

Needs $2N$ queries.

So far: with FDSA, we can get a gradient estimate

$$G^i = \frac{1}{2\delta}\left\{f(\theta+\delta e_i)+\xi_i^+ - (f(\theta-\delta e_i)+\xi_i^-)\right\}, \quad i = 1,\ldots,N.$$ with bias $O(\delta^2)$

what is second moment: $\mathbb{E}\left[\|G\|_2^2\right] =?$

$G_i = g_i + \frac{\xi_i^+ - \xi_i^-}{2\delta}$, hence $\mathbb{E}\left[G_i^2\right] = g_i^2 + \frac{2\sigma^2}{4\delta^2} = g_i^2 + \frac{\sigma^2}{2\delta^2}$ and

$$\mathbb{E}\left[\|G\|_2^2\right] = \|g\|_2^2 + O\left(\frac{N}{\delta^2}\right).$$

So far: with FDSA, we can get a gradient estimate

$$G^i = \frac{1}{2\delta}\left\{f(\theta+\delta e_i)+\xi_i^+ - (f(\theta-\delta e_i)+\xi_i^-)\right\}, \quad i=1,\ldots,N.$$ with bias $O(\delta^2)$

what is second moment: $\mathbb{E}\left[\|G\|_2^2\right] =?$

$G_i = g_i + \frac{\xi_i^+ - \xi_i^-}{2\delta}$, hence $\mathbb{E}\left[G_i^2\right] = g_i^2 + \frac{2\sigma^2}{4\delta^2} = g_i^2 + \frac{\sigma^2}{2\delta^2}$ and

$$\mathbb{E}\left[\|G\|_2^2\right] = \|g\|_2^2 + O\left(\frac{N}{\delta^2}\right).$$

So far: with FDSA, we can get a gradient estimate

$$G^i = \frac{1}{2\delta}\left\{f(\theta + \delta e_i) + \xi_i^+ - (f(\theta - \delta e_i) + \xi_i^-)\right\}, \quad i = 1, \ldots, N.$$

with bias $O(\delta^2)$

what is second moment: $\mathbb{E}\left[\|G\|_2^2\right] = ?$

$G_i = g_i + \frac{\xi_i^+ - \xi_i^-}{2\delta}$, hence $\mathbb{E}\left[G_i^2\right] = g_i^2 + \frac{2\sigma^2}{4\delta^2} = g_i^2 + \frac{\sigma^2}{2\delta^2}$ and

$$\mathbb{E}\left[\|G\|_2^2\right] = \|g\|_2^2 + O\left(\frac{N}{\delta^2}\right).$$

So far: with FDSA, we can get a gradient estimate

$$G^i = \frac{1}{2\delta}\left\{f(\theta + \delta e_i) + \xi_i^+ - (f(\theta - \delta e_i) + \xi_i^-)\right\}, \quad i = 1, \ldots, N.$$ with bias $O(\delta^2)$

what is second moment: $\mathbb{E}\left[\|G\|_2^2\right] =?$

$G_i = g_i + \frac{\xi_i^+ - \xi_i^-}{2\delta}$, hence $\mathbb{E}\left[G_i^2\right] = g_i^2 + \frac{2\sigma^2}{4\delta^2} = g_i^2 + \frac{\sigma^2}{2\delta^2}$ and

$$\mathbb{E}\left[\|G\|_2^2\right] = \|g\|_2^2 + O\left(\frac{N}{\delta^2}\right).$$

FDSA perturbed dimensions one-at-a-time, leading to $2N$ queries.
Can we reduce the number of queries?

Idea: Simultaneously randomly perturb all dimensions! (Spall, 1992)

Function measurements

$$y_n^+ = f(\theta_n + \delta_n d_n) + \xi_n^+, \quad y_n^- = f(\theta_n - \delta_n d_n) + \xi_n^-$$

Gradient estimate

$$G^i = \left[\frac{y_n^+ - y_n^-}{2\delta_n d_n^i}\right].$$

How to choose $d_n^i, i = 1, \dots, N$?

Only 2-queries, regardless of N!

$\mathbb{E}\left[G^i\right] = g^i$! Hence, $\|\mathbb{E}[G] - \nabla f(\theta)\|_2 = O(\delta^2)$.

FDSA perturbed dimensions one-at-a-time, leading to $2N$ queries.
Can we reduce the number of queries?

Idea: Simultaneously randomly perturb all dimensions! (Spall, 1992)

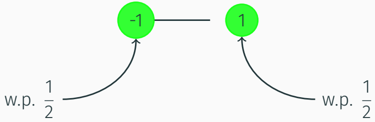
Function measurements

$$y_n^+ = f(\theta_n + \delta_n d_n) + \xi_n^+, \quad y_n^- = f(\theta_n - \delta_n d_n) + \xi_n^-$$

Gradient estimate

$$G^i = \left[\frac{y_n^+ - y_n^-}{2\delta_n d_n^i}\right].$$

How to choose $d_n^i, i = 1, \ldots, N$?



Only 2-queries, regardless of N!

$\mathbb{E}\left[G^i\right] = g^i$! Hence, $\|\mathbb{E}[G] - \nabla f(\theta)\|_2 = O(\delta^2)$.

FDSA perturbed dimensions one-at-a-time, leading to $2N$ queries.
Can we reduce the number of queries?

Idea: Simultaneously randomly perturb all dimensions! (Spall, 1992)

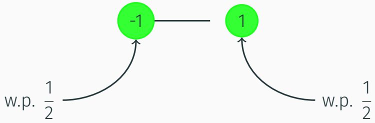
Function measurements

$$y_n^+ = f(\theta_n + \delta_n d_n) + \xi_n^+, \quad y_n^- = f(\theta_n - \delta_n d_n) + \xi_n^-$$

Gradient estimate

$$G^i = \left[\frac{y_n^+ - y_n^-}{2\delta_n d_n^i}\right].$$

How to choose $d_n^i, i = 1, \ldots, N$?



Only 2-queries, regardless of N!

$\mathbb{E}\left[G^i\right] = g^i$! Hence, $\|\mathbb{E}\left[G\right] - \nabla f(\theta)\|_2 = O(\delta^2)$.

FDSA perturbed dimensions one-at-a-time, leading to $2N$ queries.
Can we reduce the number of queries?

Idea: Simultaneously randomly perturb all dimensions! (Spall, 1992)

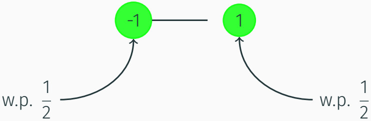
Function measurements

$$y_n^+ = f(\theta_n + \delta_n d_n) + \xi_n^+, \quad y_n^- = f(\theta_n - \delta_n d_n) + \xi_n^-$$

Gradient estimate

$$G^i = \left[\frac{y_n^+ - y_n^-}{2\delta_n d_n^i} \right].$$

How to choose $d_n^i, i = 1, \ldots, N$?



Only 2-queries, regardless of N!

$\mathbb{E}\left[G^i\right] = g^i$! Hence, $\|\mathbb{E}[G] - \nabla f(\theta)\|_2 = O(\delta^2)$.

FDSA perturbed dimensions one-at-a-time, leading to $2N$ queries.
Can we reduce the number of queries?

Idea: Simultaneously randomly perturb all dimensions! (Spall, 1992)

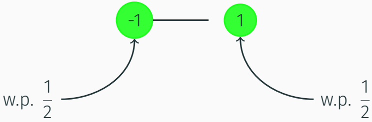
Function measurements

$$y_n^+ = f(\theta_n + \delta_n d_n) + \xi_n^+, \quad y_n^- = f(\theta_n - \delta_n d_n) + \xi_n^-$$

Gradient estimate

$$G^i = \left[\frac{y_n^+ - y_n^-}{2\delta_n d_n^i}\right].$$

How to choose $d_n^i, i = 1, \ldots, N$?



Only 2-queries, regardless of N!

$\mathbb{E}\left[G^i\right] = g^i$! Hence, $\|\mathbb{E}[G] - \nabla f(\theta)\|_2 = O(\delta^2)$.

"Mother" of all two-point sim-pert estimates

$$G = \frac{(f(\theta + U) + \xi^+) - (f(\theta - U) + \xi^-)}{2\delta} V \,.$$

Choose U, V such that $\mathbb{E}\left[VU^\top\right] = I$, $\mathbb{E}[V] = 0$.

One-point estimate!

$$G = \frac{(f(\theta + U) + \xi^+)}{\delta} V \,.$$

Choose U, V such that $\mathbb{E}\left[VU^\top\right] = I$, $\mathbb{E}[V] = 0$. Works??

$$\mathbb{E}[G] = \mathbb{E}\left[G - \frac{f(\theta)}{\delta} V\right] = \mathbb{E}\left[\frac{(f(\theta + U) + \xi^+) - f(\theta)}{\delta} V\right].$$

“Mother” of all two-point sim-pert estimates

$$G = \frac{(f(\theta + U) + \xi^+) - (f(\theta - U) + \xi^-)}{2\delta} V \,.$$

Choose U, V such that $\mathbb{E}\left[VU^\top\right] = I$, $\mathbb{E}[V] = 0$.

One-point estimate!

$$G = \frac{(f(\theta + U) + \xi^+)}{\delta} V \,.$$

Choose U, V such that $\mathbb{E}\left[VU^\top\right] = I$, $\mathbb{E}[V] = 0$. Works??

$$\mathbb{E}[G] = \mathbb{E}\left[G - \frac{f(\theta)}{\delta} V\right] = \mathbb{E}\left[\frac{(f(\theta + U) + \xi^+) - f(\theta)}{\delta} V\right].$$

“Mother” of all two-point sim-pert estimates

$$G = \frac{(f(\theta + U) + \xi^+) - (f(\theta - U) + \xi^-)}{2\delta} V\,.$$

Choose U, V such that $\mathbb{E}\left[VU^\top\right] = I$, $\mathbb{E}\left[V\right] = 0$.

One-point estimate!

$$G = \frac{(f(\theta + U) + \xi^+)}{\delta} V\,.$$

Choose U, V such that $\mathbb{E}\left[VU^\top\right] = I$, $\mathbb{E}\left[V\right] = 0$. Works??

$$\mathbb{E}\left[G\right] = \mathbb{E}\left[G - \frac{f(\theta)}{\delta} V\right] = \mathbb{E}\left[\frac{(f(\theta + U) + \xi^+) - f(\theta)}{\delta} V\right].$$

“Mother” of all two-point sim-pert estimates

$$G = \frac{(f(\theta + U) + \xi^{+}) - (f(\theta - U) + \xi^{-})}{2\delta} V \,.$$

Choose U, V such that $\mathbb{E}\left[VU^\top\right] = I$, $\mathbb{E}\left[V\right] = 0$.

One-point estimate!

$$G = \frac{(f(\theta + U) + \xi^{+})}{\delta} V \,.$$

Choose U, V such that $\mathbb{E}\left[VU^\top\right] = I$, $\mathbb{E}\left[V\right] = 0$. Works??

$$\mathbb{E}\left[G\right] = \mathbb{E}\left[G - \frac{f(\theta)}{\delta} V\right] = \mathbb{E}\left[\frac{(f(\theta + U) + \xi^{+}) - f(\theta)}{\delta} V\right].$$

Family of "Sim-pert" Gradient Estimates

- $U \sim \delta\,\mathcal{N}(0, I)$, $V = \delta^{-1}\,U$
  - Smoothed functional by Katkovnik and Kulchitsky (1972);
  - Refined by Polyak and Tsybakov (1990); also studied by Dippon (2003); Nesterov and Spokoiny (2011).
- $U \sim \delta\,\mathrm{Unif}(\mathbb{S}_N)$, $V = N\delta^{-1}\,U$
  - RDSA by Kushner and Clark (1978); Enhanced by Prashanth et al. (2017)
  - Rediscovered by Flaxman et al. (2005)
- $U_i \sim \delta\,\mathrm{Rademacher}(\pm 1)$, $V = \delta^{-1}\,U$
  - SPSA by Spall (1992).
- Deterministic perturbations by Bhatnagar et al. (2003)
- ...

Does it matter which of these we select? Not really:
Bias is always $O(\delta^2)$, while variance is $O(1)$ or $O(\delta^{-2})$ (noise controlled or not)

Family of "Sim-pert" Gradient Estimates

- $U \sim \delta\,\mathcal{N}(0, I)$, $V = \delta^{-1}\,U$
  - Smoothed functional by Katkovnik and Kulchitsky (1972);
  - Refined by Polyak and Tsybakov (1990); also studied by Dippon (2003); Nesterov and Spokoiny (2011).
- $U \sim \delta\,\mathrm{Unif}(\mathbb{S}_N)$, $V = N\delta^{-1}\,U$
  - RDSA by Kushner and Clark (1978); Enhanced by Prashanth et al. (2017)
  - Rediscovered by Flaxman et al. (2005)
- $U_i \sim \delta\,\mathrm{Rademacher}(\pm 1)$, $V = \delta^{-1}\,U$
  - SPSA by Spall (1992).
- Deterministic perturbations by Bhatnagar et al. (2003)
- $\ldots$

Does it matter which of these we select? Not really:
Bias is always $O(\delta^2)$, while variance is $O(1)$ or $O(\delta^{-2})$ (noise controlled or not)

Family of "Sim-pert" Gradient Estimates

- $U \sim \delta\,\mathcal{N}(0, I)$, $V = \delta^{-1}\,U$
  - Smoothed functional by Katkovnik and Kulchitsky (1972);
  - Refined by Polyak and Tsybakov (1990); also studied by Dippon (2003); Nesterov and Spokoiny (2011).
- $U \sim \delta\,\mathrm{Unif}(\mathbb{S}_N)$, $V = N\delta^{-1}\,U$
  - RDSA by Kushner and Clark (1978); Enhanced by Prashanth et al. (2017)
  - Rediscovered by Flaxman et al. (2005)
- $U_i \sim \delta\,\mathrm{Rademacher}(\pm 1)$, $V = \delta^{-1}\,U$
  - SPSA by Spall (1992).
- Deterministic perturbations by Bhatnagar et al. (2003)
- ...

Does it matter which of these we select? Not really:
Bias is always $O(\delta^2)$, while variance is $O(1)$ or $O(\delta^{-2})$ (noise controlled or not)

Family of "Sim-pert" Gradient Estimates

- $U \sim \delta\,\mathcal{N}(0, I)$, $V = \delta^{-1}\,U$
  - Smoothed functional by Katkovnik and Kulchitsky (1972);
  - Refined by Polyak and Tsybakov (1990); also studied by Dippon (2003); Nesterov and Spokoiny (2011).
- $U \sim \delta\,\mathrm{Unif}(\mathbb{S}_N)$, $V = N\delta^{-1}\,U$
  - RDSA by Kushner and Clark (1978); Enhanced by Prashanth et al. (2017)
  - Rediscovered by Flaxman et al. (2005)
- $U_i \sim \delta\,\mathrm{Rademacher}(\pm 1)$, $V = \delta^{-1}\,U$
  - SPSA by Spall (1992).
- Deterministic perturbations by Bhatnagar et al. (2003)
- ...

Does it matter which of these we select? Not really:
Bias is always $O(\delta^2)$, while variance is $O(1)$ or $O(\delta^{-2})$ (noise controlled or not)

Family of "Sim-pert" Gradient Estimates

- $U \sim \delta\,\mathcal{N}(0, I)$, $V = \delta^{-1}\,U$
  - Smoothed functional by Katkovnik and Kulchitsky (1972);
  - Refined by Polyak and Tsybakov (1990); also studied by Dippon (2003); Nesterov and Spokoiny (2011).
- $U \sim \delta\,\mathrm{Unif}(\mathbb{S}_N)$, $V = N\delta^{-1}\,U$
  - RDSA by Kushner and Clark (1978); Enhanced by Prashanth et al. (2017)
  - Rediscovered by Flaxman et al. (2005)
- $U_i \sim \delta\,\mathrm{Rademacher}(\pm 1)$, $V = \delta^{-1}\,U$
  - SPSA by Spall (1992).
- Deterministic perturbations by Bhatnagar et al. (2003)
- ...

Does it matter which of these we select? Not really:
Bias is always $O(\delta^2)$, while variance is $O(1)$ or $O(\delta^{-2})$ (noise controlled or not)

Family of "Sim-pert" Gradient Estimates

- $U \sim \delta\,\mathcal{N}(0, I)$, $V = \delta^{-1}\,U$
  - Smoothed functional by Katkovnik and Kulchitsky (1972);
  - Refined by Polyak and Tsybakov (1990); also studied by Dippon (2003); Nesterov and Spokoiny (2011).
- $U \sim \delta\,\mathrm{Unif}(\mathbb{S}_N)$, $V = N\delta^{-1}\,U$
  - RDSA by Kushner and Clark (1978); Enhanced by Prashanth et al. (2017)
  - Rediscovered by Flaxman et al. (2005)
- $U_i \sim \delta\,\mathrm{Rademacher}(\pm 1)$, $V = \delta^{-1}\,U$
  - SPSA by Spall (1992).
- Deterministic perturbations by Bhatnagar et al. (2003)
- ...

Does it matter which of these we select? Not really:
Bias is always $O(\delta^2)$, while variance is $O(1)$ or $O(\delta^{-2})$ (noise controlled or not)

Family of "Sim-pert" Gradient Estimates

- $U \sim \delta\,\mathcal{N}(0, I)$, $V = \delta^{-1}\,U$
  - Smoothed functional by Katkovnik and Kulchitsky (1972);
  - Refined by Polyak and Tsybakov (1990); also studied by Dippon (2003); Nesterov and Spokoiny (2011).
- $U \sim \delta\,\mathrm{Unif}(\mathbb{S}_N)$, $V = N\delta^{-1}\,U$
  - RDSA by Kushner and Clark (1978); Enhanced by Prashanth et al. (2017)
  - Rediscovered by Flaxman et al. (2005)
- $U_i \sim \delta\,\mathrm{Rademacher}(\pm 1)$, $V = \delta^{-1}\,U$
  - SPSA by Spall (1992).
- Deterministic perturbations by Bhatnagar et al. (2003)
- ...

Does it matter which of these we select? Not really:
Bias is always $O(\delta^2)$, while variance is $O(1)$ or $O(\delta^{-2})$ (noise controlled or not)

Family of “Sim-pert” Gradient Estimates

- $U \sim \delta\,\mathcal{N}(0, I)$, $V = \delta^{-1}\,U$
  - Smoothed functional by Katkovnik and Kulchitsky (1972);
  - Refined by Polyak and Tsybakov (1990); also studied by Dippon (2003); Nesterov and Spokoiny (2011).
- $U \sim \delta\,\mathrm{Unif}(\mathbb{S}_N)$, $V = N\delta^{-1}\,U$
  - RDSA by Kushner and Clark (1978); Enhanced by Prashanth et al. (2017)
  - Rediscovered by Flaxman et al. (2005)
- $U_i \sim \delta\,\mathrm{Rademacher}(\pm 1)$, $V = \delta^{-1}\,U$
  - SPSA by Spall (1992).
- Deterministic perturbations by Bhatnagar et al. (2003)
- ...

Does it matter which of these we select? Not really:
Bias is always $O(\delta^2)$, while variance is $O(1)$ or $O(\delta^{-2})$ (noise controlled or not)

Family of “Sim-pert” Gradient Estimates

- $U \sim \delta\,\mathcal{N}(0, I)$, $V = \delta^{-1}\,U$
  - Smoothed functional by Katkovnik and Kulchitsky (1972);
  - Refined by Polyak and Tsybakov (1990); also studied by Dippon (2003); Nesterov and Spokoiny (2011).
- $U \sim \delta\,\mathrm{Unif}(\mathbb{S}_N)$, $V = N\delta^{-1}\,U$
  - RDSA by Kushner and Clark (1978); Enhanced by Prashanth et al. (2017)
  - Rediscovered by Flaxman et al. (2005)
- $U_i \sim \delta\,\mathrm{Rademacher}(\pm 1)$, $V = \delta^{-1}\,U$
  - SPSA by Spall (1992).
- Deterministic perturbations by Bhatnagar et al. (2003)
- ...

Does it matter which of these we select? Not really:
Bias is always $O(\delta^2)$, while variance is $O(1)$ or $O(\delta^{-2})$ (noise controlled or not)

Family of “Sim-pert” Gradient Estimates

- $U \sim \delta\,\mathcal{N}(0, I)$, $V = \delta^{-1}\,U$
  - Smoothed functional by Katkovnik and Kulchitsky (1972);
  - Refined by Polyak and Tsybakov (1990); also studied by Dippon (2003); Nesterov and Spokoiny (2011).
- $U \sim \delta\,\mathrm{Unif}(\mathbb{S}_N)$, $V = N\delta^{-1}\,U$
  - RDSA by Kushner and Clark (1978); Enhanced by Prashanth et al. (2017)
  - Rediscovered by Flaxman et al. (2005)
- $U_i \sim \delta\,\mathrm{Rademacher}(\pm 1)$, $V = \delta^{-1}\,U$
  - SPSA by Spall (1992).
- Deterministic perturbations by Bhatnagar et al. (2003)
- ...

Does it matter which of these we select? Not really:
Bias is always $O(\delta^2)$, while variance is $O(1)$ or $O(\delta^{-2})$ (noise controlled or not)

Family of "Sim-pert" Gradient Estimates

- $U \sim \delta\,\mathcal{N}(0, I)$, $V = \delta^{-1}\,U$
  - Smoothed functional by Katkovnik and Kulchitsky (1972);
  - Refined by Polyak and Tsybakov (1990); also studied by Dippon (2003); Nesterov and Spokoiny (2011).
- $U \sim \delta\,\mathrm{Unif}(\mathbb{S}_N)$, $V = N\delta^{-1}\,U$
  - RDSA by Kushner and Clark (1978); Enhanced by Prashanth et al. (2017)
  - Rediscovered by Flaxman et al. (2005)
- $U_i \sim \delta\,\mathrm{Rademacher}(\pm 1)$, $V = \delta^{-1}\,U$
  - SPSA by Spall (1992).
- Deterministic perturbations by Bhatnagar et al. (2003)
- ...

Does it matter which of these we select? Not really:
Bias is always $O(\delta^2)$, while variance is $O(1)$ or $O(\delta^{-2})$ (noise controlled or not)

What we have learned so far?

For performing gradient descent:

$$\theta_{n+1} = \theta_n - a_n G_n,$$

we can construct nearly unbiased gradient estimate G_n using simultaneous perturbation trick

Noise $\rightarrow$ Gradient estimate $\downarrow$	Controlled	Uncontrolled
Bias	$C_1\delta^2$	$C_1\delta^2$
Variance	C_2	$\frac{C_2}{\delta^2}$

This assumed $f \in \mathcal{C}^3$. Holds also for f convex, smooth.

A few answers so far...

Q1) How to form G_n from function samples so that $G_n \approx \nabla f(\theta_n)$

Use simultaneous perturbation trick

Q2) Such a G_n - is it unbiased?

Almost ... what we get is an asymptotically unbiased estimate?

Q3) Does $\theta_{n+1} = \theta_n - a_n G_n$ converge to θ^* with such a G_n?

??

Q4) If answer is yes to above, what is the convergence rate?

??

A few answers so far...

Q1) How to form G_n from function samples so that $G_n \approx \nabla f(\theta_n)$

Use simultaneous perturbation trick

Q2) Such a G_n - is it unbiased?

Almost ... what we get is an asymptotically unbiased estimate?

Q3) Does $\theta_{n+1} = \theta_n - a_n G_n$ converge to θ^* with such a G_n?

??

Q4) If answer is yes to above, what is the convergence rate?

??

A few answers so far...

Q1) How to form G_n from function samples so that $G_n \approx \nabla f(\theta_n)$

Use simultaneous perturbation trick

Q2) Such a G_n - is it unbiased?

Almost ... what we get is an asymptotically unbiased estimate?

Q3) Does $\theta_{n+1} = \theta_n - a_n G_n$ converge to θ^* with such a G_n?

??

Q4) If answer is yes to above, what is the convergence rate?

??

Outline

xkcd on real analysis

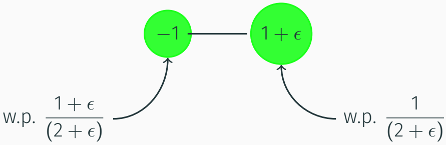
Function measurements

$$y_n^+ = f(\theta_n + \delta_n d_n) + \xi_n^+, \quad y_n^- = f(\theta_n - \delta_n d_n) + \xi_n^-$$

RDSA Gradient estimate

$$G_n = \frac{1}{1+\epsilon} d_n \left[\frac{y_n^+ - y_n^-}{2\delta_n} \right].$$

Asymmetric Bernoulli distribution for $d_n^i, i = 1, \ldots, N$:



Assumptions

Smoothness $f \in \mathcal{C}^3$, i.e., f is three times continuously differentiable

Zero-mean noise $\mathbb{E}\left[\xi_n^+ - \xi_n^- \,\middle|\, d_n, \mathcal{F}_n\right] = 0$, where $\mathcal{F}_n = \sigma(\theta_m, m < n)$.

Need these to establish (asymptotic) unbiasedness of gradient estimate

---

Second moment bound $\mathbb{E}\left|\xi_n^{\pm}\right|^2 \leq \alpha_1$, $\mathbb{E}\left|f(x_n \pm \delta_n d_n)\right|^2 \leq \alpha_2$

Step-sizes $a_n, \delta_n \to 0$ as $n \to \infty$, $\sum_n a_n = \infty$ and $\sum_n \left(\frac{a_n}{\delta_n}\right)^2 < \infty$.

Stable iterates $\sup_n \|\theta_n\| < \infty$ w.p. 1.

Assumptions

Smoothness $f \in \mathcal{C}^3$, i.e., f is three times continuously differentiable

Zero-mean noise $\mathbb{E}\left[\xi_n^+ - \xi_n^- \,\middle|\, d_n, \mathcal{F}_n\right] = 0$, where $\mathcal{F}_n = \sigma(\theta_m, m < n)$.

Second moment bound $\mathbb{E}\left|\xi_n^{\pm}\right|^2 \leq \alpha_1$, $\mathbb{E}\left|f(x_n \pm \delta_n d_n)\right|^2 \leq \alpha_2$

Step-sizes $a_n, \delta_n \to 0$ as $n \to \infty$, $\sum_n a_n = \infty$ and $\sum_n \left(\frac{a_n}{\delta_n}\right)^2 < \infty$.

So that the noise effects vanish asymptotically

Stable iterates $\sup_n \|\theta_n\| < \infty$ w.p. 1.

Assumptions

Smoothness $f \in \mathcal{C}^3$, i.e., f is three times continuously differentiable

Zero-mean noise $\mathbb{E}\left[\xi_n^+ - \xi_n^- \,\middle|\, d_n, \mathcal{F}_n\right] = 0$, where $\mathcal{F}_n = \sigma(\theta_m, m < n)$.

Second moment bound $\mathbb{E}\left|\xi_n^\pm\right|^2 \leq \alpha_1$, $\mathbb{E}\left|f(x_n \pm \delta_n d_n)\right|^2 \leq \alpha_2$

Step-sizes $a_n, \delta_n \to 0$ as $n \to \infty, \sum_n a_n = \infty$ and $\sum_n \left(\frac{a_n}{\delta_n}\right)^2 < \infty$.

Stable iterates $\sup_n \|\theta_n\| < \infty$ w.p. 1.

Needed to establish convergence of gradient-descent scheme. Trick: use projection

Ordinary differential equations (ODE) approach for stochastic approximation

$\theta_{n+1} = \theta_n - a_n G_n$ is equivalent to $\theta_{n+1} = \theta_n - a_n \left(\nabla f(x_n) + \eta_n + \beta_n \right)$

$\eta_n = G_n - \mathbb{E}(G_n \mid \mathcal{F}_n) \leftarrow$ martingale difference,
$\beta_n = \mathbb{E}(G_n \mid \mathcal{F}_n) - \nabla f(x_n) \leftarrow$ gradient estimation bias $= O(\delta_n^2)$

Mean ODE $\dot{\theta}_t = -\nabla f(\theta_t)$ with limit set $K = \left\{\theta : \nabla f(\theta)\right) = 0\}$

"If" there is no bias and no noise, then it is straightforward(?) to see that θ_n converges a.s. to K.

Can we conclude the same with bias and noise elements?

$$\theta_{n+1} = \theta_n - a_n\left(\nabla f(x_n) + \eta_n + \beta_n\right)$$

$\eta_n = G_n - \mathbb{E}(G_n \mid \mathcal{F}_n) \leftarrow$ martingale difference $\beta_n = \mathbb{E}(G_n \mid \mathcal{F}_n) - \nabla f(x_n) \leftarrow$ gradient estimation bias $= O(\delta_n^2)$

To apply Kushner-Clark lemma we verify a few conditions:

1) "$\beta_n \to 0$ almost surely" $\leftarrow$ holds since we assume $\delta_n \to 0$ and $\beta_n = O(\delta_n^2)$

2) "$\forall \epsilon > 0, \lim_{n\to\infty} \underbrace{P\left(\sup_{m\geq n}\left\|\sum_{i=n}^{m} a_i\eta_i\right\| \geq \epsilon\right)}_{(*)} = 0.$"

$$(*) \leq \frac{1}{\epsilon^2}\mathbb{E}\left\|\sum_{i=n}^{\infty} a_i\eta_i\right\|^2 = \frac{1}{\epsilon^2}\sum_{i=n}^{\infty} a_i^2\mathbb{E}\left\|\eta_i\right\|^2 \leq \frac{C}{\epsilon^2}\lim_{n\to\infty}\sum_{i=n}^{\infty}\frac{a_i^2}{\delta_i^2} \to 0$$

Thus,

$$\theta_n \to K \text{ a.s. as } n \to \infty$$

A few answers so far...

Q1) How to form G_n from function samples so that $G_n \approx \nabla f(\theta_n)$

Use simultaneous perturbation trick

Q2) Such a G_n - is it unbiased?

Almost ... what we get is an asymptotically unbiased estimate

Q3) Does $\theta_{n+1} = \theta_n - a_n G_n$ converge to θ^* with such a G_n?

Yes!

Q4) If answer is yes to above, what is the convergence rate?

Asymptotic normality

A few answers so far. . .

Q1) How to form G_n from function samples so that $G_n \approx \nabla f(\theta_n)$

Use simultaneous perturbation trick

Q2) Such a G_n - is it unbiased?

Almost . . . what we get is an asymptotically unbiased estimate

Q3) Does $\theta_{n+1} = \theta_n - a_n G_n$ converge to θ^* with such a G_n?

Yes!

Q4) If answer is yes to above, what is the convergence rate?

Asymptotic normality

Asymptotic normality

$$n^{\beta/2}(x_n - x^*) \xrightarrow{dist} \mathcal{N}(\mu, PMP^\intercal)$$

where $\beta = 2/3$ and μ, M depend on a_n, d_n and f at θ^*.

Under some conditions, this implies

$$n^{\beta}\mathbb{E}\left\|x_n - x^*\right\|^2 \to \mu^\intercal\mu + \text{trace}(PMP^\intercal) \text{ as } n \to \infty$$

asymptotic mean square error (AMSE) is the limit above

To achieve a given accuracy, the number of samples needed by 1SPSA (n_{1SPSA}) to that of 1FDSA (n_{1FDSA}) is $\frac{n_{1\mathcal{SPSA}}}{n_{1\mathcal{FDSA}}} = \frac{1}{N}$

Bottomline: Simultaneously randomly perturbing all dimensions is equivalent to perturbing dimensions one-at-a-time!

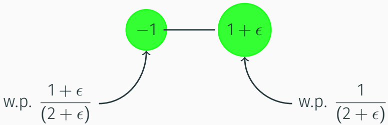
Function measurements

$$y_n^+ = f(\theta_n + \delta_n d_n) + \xi_n^+, \quad y_n^- = f(\theta_n - \delta_n d_n) + \xi_n^-$$

RDSA Gradient estimate

$$G_n = \frac{1}{1+\epsilon} d_n \left[\frac{y_n^+ - y_n^-}{2\delta_n}\right].$$

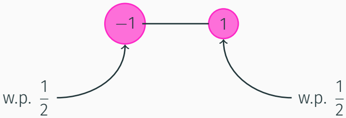
Asymmetric Bernoulli distribution for $d_n^i, i = 1, \ldots, N$:



SPSA Gradient estimate

$$G_n = d_n^{-1} \left[\frac{y_n^+ - y_n^-}{2\delta_n}\right].$$

Symmetric Bernoulli distribution for $d_n^i, i = 1, \ldots, N$:



So, which perturbation choice works best?

The competitors

Samples $y_n^{\pm}$ at $x_n \pm \delta_n d_n$

Algorithm	d_n	G_n
1SPSA	Rademacher	$d_n^{-1}\left[\frac{y_n^+ - y_n^-}{\delta_n}\right]$
1RDSA-Gaussian	Standard Gaussian	$d_n\left[\frac{y_n^+ - y_n^-}{\delta_n}\right]$
1RDSA-Unif	$U[-1,1]$	$3d_n\left[\frac{y_n^+ - y_n^-}{2\delta_n}\right]$
1RDSA-AsymBer	Asymmetric Bernoulli	$\frac{1}{1+\epsilon}d_n\left[\frac{y_n^+ - y_n^-}{2\delta_n}\right]$

So, which perturbation choice works best?

Letting (A) and (B) denote problem-dependent quantities, we have

Fact 1: $$\frac{AMSE_{1RDSA-Gaussian}}{AMSE_{1SPSA}} = \frac{9(A)+(B)}{(A)+(B)}$$

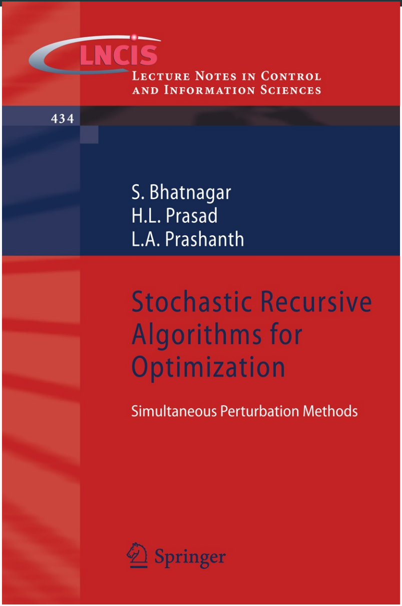
Fact 2: $$\frac{AMSE_{1RDSA-Unif}}{AMSE_{1SPSA}} = \frac{3.24(A)+(B)}{(A)+(B)}$$

Fact 3: With $\epsilon = 0.01$,

$$\frac{AMSE_{1RDSA-AsymBer}}{AMSE_{1SPSA}} = \frac{1.00019(A)+(B)}{(A)+(B)}$$

Commercials

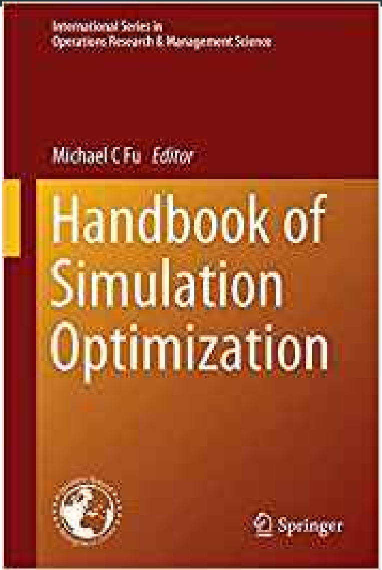
For deep-dive into simultaneous perturbation methods



Rigorous treatment of SPSA and friends
includes both first as well as second-order schemes

Prerequisities: probability theory, stochastic approximation (short appendices cover the main results)

For a broader view



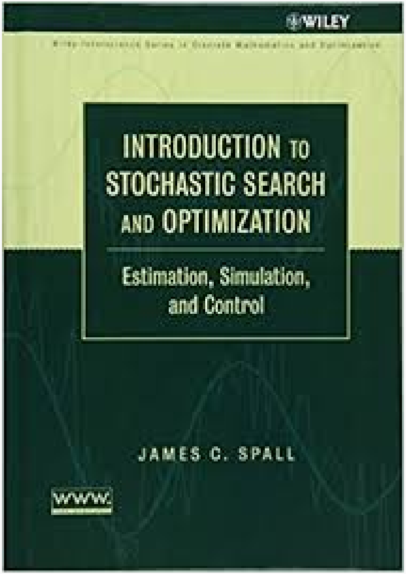
Chapter 3: Ranking & Selection aka Best-arm identification in multi-armed bandits

Chapter 5: Stochastic Gradient Estimation

Chapter 6: An Overview of Stochastic Approximation

Chapter 10: Solving Markov Decision Processes via Simulation

For a even more broader view



1) Random search
2) Machine (reinforcement) learning
3) Recursive linear estimation
4) Model selection
5) Stochastic approximation
6) Simulation-based optimization
7) Simulated annealing
8) Markov chain Monte Carlo
9) Genetic and evolutionary algorithms
10) Optimal experimental design

Some more books and other references

1. Spall, J. C. (1998), An Overview of the Simultaneous Perturbation Method for Efficient Optimization, Johns Hopkins APL Technical Digest, vol. 19(4), pp. 482–492.
2. Michael Fu (2002) Optimization for Simulation: Theory vs. Practice (Feature Article), INFORMS Journal on Computing, Vol.14, No.3, 192-215.
3. Henderson/Nelson (editors) (2006) Handbook of Operations Research and Management Science: Simulation Vol.13
    - Chapters 17-21: Selecting the Best System, Metamodel-Based Simulation Optimization, Gradient Estimation, Random Search, Metaheuristics
4. SPSA web site `www.jhuapl.edu/SPSA`
5. Vivek Borkar (2008), Stochastic approximation: a dynamical systems viewpoint, Cambridge university press

Software

1. OptQuest (Arena, Crystal Ball, et al.)
    - standalone module, most widely implemented – scatter search, tabu search, neural networks
2. Simulation Optimization Testbed: `http://simopt.org`
3. AutoStat (AutoMod from Autosimulations, Inc.)
    - part of a complete statistical output analysis package – dominates semiconductor industry
    - evolutionary (variation of genetic algorithms)
4. SimRunner (ProModel): evolutionary
5. Optimizer (WITNESS): simulated annealing, tabu search
6. Risk Solver (Excel): `www.solver.com/simulation-optimization`

Second-order methods

Why second-order methods?

Gradient-descent (GD)

$$\theta_{n+1} = \theta_n - a_n \nabla f(\theta_n)$$

- optimum convergence speed requires knowledge of curvature of f
- declines fast initially, but slows down towards the end (when near θ^*)
- *not* scale invariant: change $\theta \to B\theta$, GD update would depend on B
- Efficient update $\Leftrightarrow$ low per-iteration cost

Newton methods

$$\theta_{n+1} = \theta_n - a_n (\nabla^2 f(\theta_n))^{-1} \nabla f(\theta_n)$$

- optimum speed of convergence without knowledge of $\lambda_{\min}(\nabla^2 f(\theta^*))$.
- faster convergence in final phase; equivalent to minimizing a quadratic model of f
- scale invariant: auto-adjusts to the scale of θ
- high per-iteration cost $\leftarrow$ matrix inversion, more samples for estimation

Why second-order methods?

Gradient-descent (GD)

$$\theta_{n+1} = \theta_n - a_n \nabla f(\theta_n)$$

- optimum convergence speed requires knowledge of curvature of f
- declines fast initially, but slows down towards the end (when near θ^*)
- *not* scale invariant: change $\theta \to B\theta$, GD update would depend on B
- Efficient update $\Leftrightarrow$ low per-iteration cost

Newton methods

$$\theta_{n+1} = \theta_n - a_n (\nabla^2 f(\theta_n))^{-1} \nabla f(\theta_n)$$

- optimum speed of convergence without knowledge of $\lambda_{\min}(\nabla^2 f(\theta^*))$.
- faster convergence in final phase; equivalent to minimizing a quadratic model of f
- scale invariant: auto-adjusts to the scale of θ
- high per-iteration cost $\leftarrow$ matrix inversion, more samples for estimation

Why second-order methods?

Gradient-descent (GD)

$$\theta_{n+1} = \theta_n - a_n \nabla f(\theta_n)$$

- optimum convergence speed requires knowledge of curvature of f
- declines fast initially, but slows down towards the end (when near θ^*)
- *not* scale invariant: change $\theta \to B\theta$, GD update would depend on B
- Efficient update $\Leftrightarrow$ low per-iteration cost

Newton methods

$$\theta_{n+1} = \theta_n - a_n (\nabla^2 f(\theta_n))^{-1} \nabla f(\theta_n)$$

- optimum speed of convergence without knowledge of $\lambda_{\min}(\nabla^2 f(\theta^*))$.
- faster convergence in final phase; equivalent to minimizing a quadratic model of f
- scale invariant: auto-adjusts to the scale of θ
- high per-iteration cost $\leftarrow$ matrix inversion, more samples for estimation

Why second-order methods?

Gradient-descent (GD)

$$\theta_{n+1} = \theta_n - a_n \nabla f(\theta_n)$$

- optimum convergence speed requires knowledge of curvature of f
- declines fast initially, but slows down towards the end (when near θ^*)
- *not* scale invariant: change $\theta \to B\theta$, GD update would depend on B
- Efficient update $\Leftrightarrow$ low per-iteration cost

Newton methods

$$\theta_{n+1} = \theta_n - a_n (\nabla^2 f(\theta_n))^{-1} \nabla f(\theta_n)$$

- optimum speed of convergence without knowledge of $\lambda_{\min}(\nabla^2 f(\theta^*))$.
- faster convergence in final phase; equivalent to minimizing a quadratic model of f
- scale invariant: auto-adjusts to the scale of θ
- high per-iteration cost $\leftarrow$ matrix inversion, more samples for estimation

Stochastic analog of Newton-Raphson method

- Matrix projection
- Gradient estimate

$$\theta_{n+1} = \theta_n - a_n \Upsilon(\overline{H}_n)^{-1} G_n , \tag{3}$$

$$\overline{H}_n = \left(1 - \frac{1}{n+1}\right) \overline{H}_{n-1} + \frac{1}{n+1} \widehat{H}_n , \tag{4}$$

- Averaging
- Hessian estimate

Stochastic analog of Newton-Raphson method

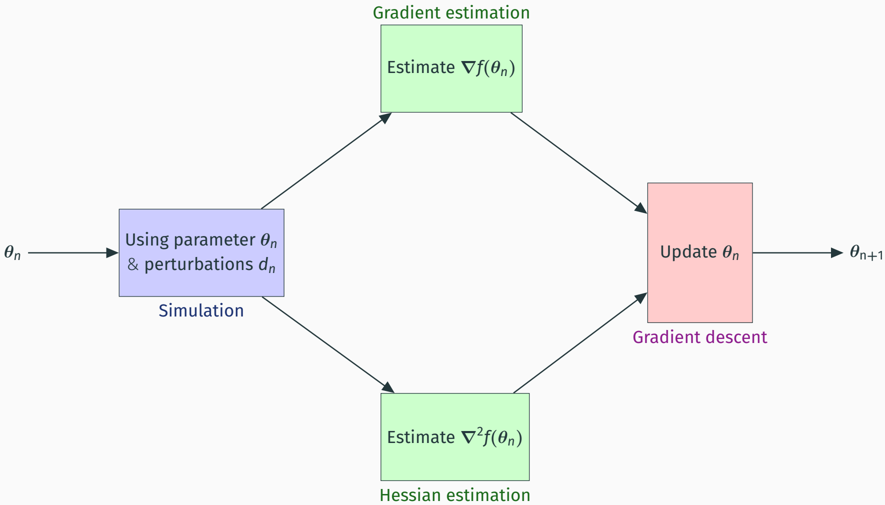
- Matrix projection
- Gradient estimate

$$\theta_{n+1} = \theta_n - a_n \Upsilon(\overline{H}_n)^{-1} G_n, \qquad (3)$$

$$\overline{H}_n = \left(1 - \frac{1}{n+1}\right) \overline{H}_{n-1} + \frac{1}{n+1} \widehat{H}_n, \qquad (4)$$

- Averaging
- Hessian estimate

Overall flow



What's up next

Second-order FDSA Fabian (1971) requires $O(N^2)$ samples to estimate Hessian

Simultaneous perturbation in action:

(Spall 2000)[1]	Second-order SPSA (2SPSA)	4 simulations/iteration
(Prashanth L.A. et al 2016)[2]	Second-order RDSA (2RDSA)	3 simulations/iteration

[1] J. C. Spall (2000), "Adaptive stochastic approximation by the simultaneous perturbation method," *IEEE TAC.*

[2] Prashanth L. A. et al. (2016) "Adaptive system optimization using random directions stochastic approximation," *IEEE TAC.*

RDSA gradient estimate

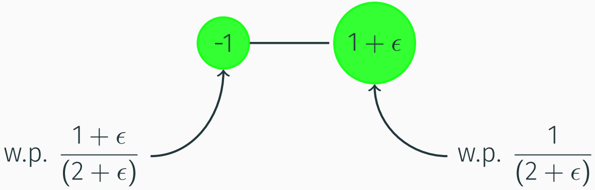
Function measurements

$$y_n^+ = f(\theta_n + \delta_n d_n) + \xi_n^+, \quad y_n^- = f(\theta_n - \delta_n d_n) + \xi_n^-$$

Gradient estimate

$$G_n = \frac{1}{1+\epsilon} d_n \left[\frac{y_n^+ - y_n^-}{2\delta_n} \right]. \tag{5}$$

Asymmetric Bernoulli distribution for $d_n^i, i = 1, \ldots, N$:



2RDSA Hessian estimate

Function measurements

$$y_n^+ = f(\theta_n + \delta_n d_n) + \xi_n^+,\ \ y_n^- = f(\theta_n - \delta_n d_n) + \xi_n^-,\ \ y_n = f(\theta_n) + \xi_n$$

Hessian estimate $\widehat{H}_n$

$$\begin{aligned}
\widehat{H}_n &= M_n\left(\frac{y_n^+ + y_n^- - 2y_n}{\delta_n^2}\right)\\
&= M_n\left[\left(\frac{f(\theta_n+\delta_n d_n) + f(\theta_n-\delta_n d_n) - 2f(\theta_n)}{\delta_n^2}\right)\right.\\
&\qquad\qquad \left. + \left(\frac{\xi_n^+ + \xi_n^- - 2\xi_n}{\delta_n^2}\right)\right]\\
&= M_n\left(d_n^\top \nabla^2 f(\theta_n) d_n + O(\delta_n^2) + \left(\frac{\xi_n^+ + \xi_n^- - 2\xi_n}{\delta_n^2}\right)\right). \qquad (6)
\end{aligned}$$

Want to recover $\nabla^2 f(\theta_n)$ from this

Zero-mean

2RDSA Hessian estimate

Function measurements

$$y_n^+ = f(\theta_n + \delta_n d_n) + \xi_n^+, \;\; y_n^- = f(\theta_n - \delta_n d_n) + \xi_n^-, \;\; y_n = f(\theta_n) + \xi_n$$

Hessian estimate $\widehat{H}_n$

$$\begin{aligned}
\widehat{H}_n &= M_n \left(\frac{y_n^+ + y_n^- - 2y_n}{\delta_n^2} \right) \\
&= M_n \left[\left(\frac{f(\theta_n + \delta_n d_n) + f(\theta_n - \delta_n d_n) - 2f(\theta_n)}{\delta_n^2} \right) \right. \\
&\qquad \left. + \left(\frac{\xi_n^+ + \xi_n^- - 2\xi_n}{\delta_n^2} \right) \right] \\
&= M_n \left(d_n^\top \nabla^2 f(\theta_n) d_n + O(\delta_n^2) + \left(\frac{\xi_n^+ + \xi_n^- - 2\xi_n}{\delta_n^2} \right) \right). \qquad (6)
\end{aligned}$$

Want to recover $\nabla^2 f(\theta_n)$ from this

Zero-mean

How to choose M_n?

Asymmetric Bernoulli Perturbation

$$M_n = \begin{bmatrix} \frac{1}{\kappa}\left((d_n^1)^2 - (1+\epsilon)\right) & \cdots & \frac{1}{2(1+\epsilon)^2} d_n^1 d_n^N \\ \frac{1}{2(1+\epsilon)^2} d_n^2 d_n^1 & \cdots & \frac{1}{2(1+\epsilon)^2} d_n^2 d_n^N \\ \cdots & \cdots & \cdots \\ \frac{1}{2(1+\epsilon)^2} d_n^N d_n^1 & \cdots & \frac{1}{\kappa}\left((d_n^N)^2 - (1+\epsilon)\right) \end{bmatrix}, \tag{7}$$

where $\kappa = \tau \left(1 - \frac{(1+\epsilon)^2}{\tau}\right)$ and $\tau = E(d_n^i)^4 = \frac{(1+\epsilon)(1+(1+\epsilon)^3)}{(2+\epsilon)}$, for any $i = 1, \ldots, N$.

2SPSA - Hessian estimation - main idea

Suppose $G_n(\theta_n \pm \delta_n d_n)$ are approximations to the gradient of f at $\theta_n \pm \delta_n d_n$. Let $\Delta G_n = G_n(\theta_n + \delta_n d_n) - G_n(\theta_n - \delta_n d_n)$.

Simultaneous perturbation trick suggests

$$\widehat{H}_n = \frac{\Delta G_n}{4\delta_n d_n}$$

What remains to be specified: G_n

Use Simultaneous perturbation trick again!

$$G_n(\theta_n \pm \delta_n d_n) = d_n^{-1} \frac{y(\theta_n \pm \delta_n d_n + \delta_n \hat{d}_n) - y(\theta_n \pm \delta_n d_n)}{\delta_n}$$

where $\hat{d}_n$ are another independent set of perturbations having same distribution as d_n.

Convergence analysis

Under regularity conditions that aren't too far from those for 1SPSA/1RDSA, we have

Bias in Hessian estimate For $i, j = 1, \ldots, N$,

$$\left| \mathbb{E}\left[\widehat{H}_n(i,j) \middle| \mathcal{F}_n \right] - \nabla^2_{ij} f(\theta_n) \right| = O(\delta_n^2). \tag{8}$$

Strong Convergence of Hessian

$$\theta_n \to \theta^*, \overline{H}_n \to \nabla^2 f(\theta^*) \text{ a.s. as } n \to \infty.$$

[1] Here $\widehat{H}_n(i,j)$ and $\nabla^2_{ij} f(\cdot)$ denote the (i,j)th entry in the Hessian estimate $\widehat{H}_n$ and the true Hessian $\nabla^2 f(\cdot)$, respectively.

2SPSA vs. 2RDSA: An asymptotic mean-square error (AMSE) comparison

Letting (A) and (B) denote problem-dependent quantities and with $\epsilon = 0.01$ for 2RDSA-AsymBer, we have

$$\frac{\mathcal{AMSE}_{2\mathcal{RDSA}-\text{AsymBer}}}{\mathcal{AMSE}_{2\mathcal{SPSA}}} = \frac{1.00019(A) + (B)}{(A) + (B)}$$

However, 2SPSA uses 4 samples/iteration, while 2RDSA-AB uses 3. So,

$$\frac{\hat{n}_{2\mathcal{RDSA}-\text{AsymBer}}}{\hat{n}_{2\mathcal{SPSA}}} = \frac{3}{4} \times \frac{\mathcal{AMSE}_{2\mathcal{RDSA}-\text{AsymBer}}}{\mathcal{AMSE}_{2\mathcal{SPSA}}} = \frac{3.00057(A) + 3(B)}{4(A) + 4(B)} < 1$$

Bottomline: 2RDSA with asymmetric Bernoulli perturbations is better than 2SPSA on all problem instances!

2SPSA vs. 2RDSA: An asymptotic mean-square error (AMSE) comparison

Letting (A) and (B) denote problem-dependent quantities and with $\epsilon = 0.01$ for 2RDSA-AsymBer, we have

$$\frac{\mathcal{AMSE}_{2\mathcal{RDSA}-\text{AsymBer}}}{\mathcal{AMSE}_{2\mathcal{SPSA}}} = \frac{1.00019(A) + (B)}{(A) + (B)}$$

However, 2SPSA uses 4 samples/iteration, while 2RDSA-AB uses 3. So,

$$\frac{\hat{n}_{2\mathcal{RDSA}-\text{AsymBer}}}{\hat{n}_{2\mathcal{SPSA}}} = \frac{3}{4} \times \frac{\mathcal{AMSE}_{2\mathcal{RDSA}-\text{AsymBer}}}{\mathcal{AMSE}_{2\mathcal{SPSA}}} = \frac{3.00057(A) + 3(B)}{4(A) + 4(B)} < 1$$

Bottomline: 2RDSA with asymmetric Bernoulli perturbations is better than 2SPSA on all problem instances!

Some questions before diving into applications...

Q1) Can I solve constrainted optimization problems using simultaneous perturbation methods?

Yes! See service systems application next

Q2) So far, the focus has been on continuous optimization problems. Can SPSA/its friends be used for discrete parameter optimization?

Yes! See (again) service systems application next

Q3) Analysis showed convergence to local optima. Is global convergence achievable?

Yes. See (Maryak and Chin, 2008)

Q4) Instead of full inverted Hessian, can we subsample/use a sparse representation and still approximate Hessian inverse well?

??

Some questions before diving into applications...

Q1) Can I solve constrainted optimization problems using simultaneous perturbation methods?

Yes! See service systems application next

Q2) So far, the focus has been on continuous optimization problems. Can SPSA/its friends be used for discrete parameter optimization?

Yes! See (again) service systems application next

Q3) Analysis showed convergence to local optima. Is global convergence achievable?

Yes. See (Maryak and Chin, 2008)

Q4) Instead of full inverted Hessian, can we subsample/use a sparse representation and still approximate Hessian inverse well?

??

Some questions before diving into applications. . .

Q1) Can I solve constrainted optimization problems using simultaneous perturbation methods?

Yes! See service systems application next

Q2) So far, the focus has been on continuous optimization problems. Can SPSA/its friends be used for discrete parameter optimization?

Yes! See (again) service systems application next

Q3) Analysis showed convergence to local optima. Is global convergence achievable?

Yes. See (Maryak and Chin, 2008)

Q4) Instead of full inverted Hessian, can we subsample/use a sparse representation and still approximate Hessian inverse well?

??

Some questions before diving into applications. . .

Q1) Can I solve constrainted optimization problems using simultaneous perturbation methods?

Yes! See service systems application next

Q2) So far, the focus has been on continuous optimization problems. Can SPSA/its friends be used for discrete parameter optimization?

Yes! See (again) service systems application next

Q3) Analysis showed convergence to local optima. Is global convergence achievable?

Yes. See (Maryak and Chin, 2008)

Q4) Instead of full inverted Hessian, can we subsample/use a sparse representation and still approximate Hessian inverse well?

??

Applications

Outline


WE NEED TO CUT OUR EXPENSES.
I RECOM— MEND ELI— MINATING TED'S JOB.
@ScottAdamsSays
Dilbert.com
WHAT? ? ? I RECOMMEND ELIMINATING DILBERT'S JOB!
10-03-16 © 2016 Scott Adams, Inc. /Dist. by Universal Uclick
JUST BECAUSE HE SAID IT FIRST? ? ?
LET'S NOT OVER— ANALYZE IT.

Service Systems

An organization composed of the resources that support, and the processes that drive service interactions so that the outcomes meet customer expectations Examples: call centers, BPOs, data-center management

Challenges:

- Each customer has unique environments, expectations (SLAs)
- Randomness in service times, arrivals of service requests
- Not all service workers can support many customers / types of work
- Continuous change in scope of work, number/skills of workers

How do we staff such SS?

Service Systems

Examples: call centers, BPOs, data-center management

Challenges:

- Each customer has unique environments, expectations (SLAs)
- Randomness in service times, arrivals of service requests
- Not all service workers can support many customers / types of work
- Continuous change in scope of work, number/skills of workers

How do we staff such SS?

Service Systems

Examples: call centers, BPOs, data-center management

Challenges:

- Each customer has unique environments, expectations (SLAs)
- Randomness in service times, arrivals of service requests
- Not all service workers can support many customers / types of work
- Continuous change in scope of work, number/skills of workers

How do we staff such SS?

Application I: Service System

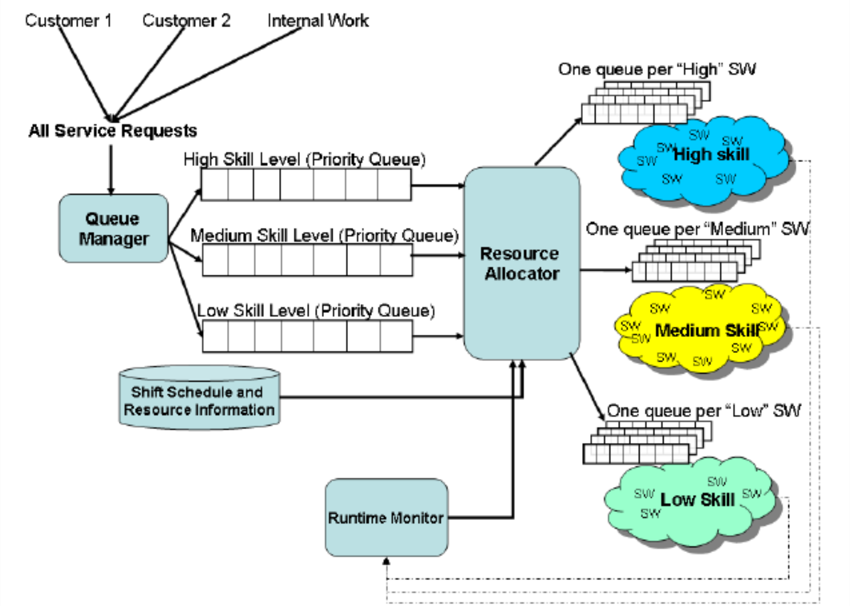


Table 3: Workers $W_{i,j}$

	Skill levels		
Shift	High	Med	Low
S1	1	3	7
S2	0	5	2
S3	3	1	2

Table 4: SLA targets $\gamma_{i,j}$

	Customers	
Priority	Bossy Corp	Cool Inc
P_1	4h	5h
P_2	8h	12h
P_3	24h	48h
P_4	18h	144h

Aim: Find the optimal number of workers for each shift and of each skill level

- that minimizes the labor cost
- subject to SLA constraints

Labor Cost Optimization

The problem we are looking at

Find the optimal number of workers for each shift and of each skill level

- that minimizes the average labor cost; and
- satisfies service level agreement (SLA) constraints

how do we solve it?

Simulation optimization!

Challenges

- discrete worker parameter
- SLA constraints

Labor Cost Optimization

The problem we are looking at

Find the optimal number of workers for each shift and of each skill level

- that minimizes the average labor cost; and
- satisfies service level agreement (SLA) constraints

how do we solve it?

Simulation optimization!

Challenges

- discrete worker parameter
- SLA constraints

Labor Cost Optimization

The problem we are looking at

Find the optimal number of workers for each shift and of each skill level

- that minimizes the average labor cost; and
- satisfies service level agreement (SLA) constraints

how do we solve it?

Simulation optimization!

Challenges

- discrete worker parameter
- SLA constraints

Notation: Shifts A, Skills B, Customers C, Priorities P

State:

$$X_n = (\underbrace{\mathcal{N}_1(n), \ldots, \mathcal{N}_{|B|}(n)}_{\textit{complexity queue lengths}}, \underbrace{u_{1,1}(n), \ldots\ldots, u_{|A|,|B|}(n)}_{\textit{worker utilizations}}, \underbrace{\gamma'_{1,1}(n), \ldots\ldots, \gamma'_{|C|,|P|}(n)}_{\textit{SLAs attained}}, q(n)),$$

Single-stage cost:

$$c(X_n) = \left(1 - \sum_{i=1}^{|A|}\sum_{j=1}^{|B|} \alpha_{i,j} \times u_{i,j}(n)\right) + \left(\sum_{i=1}^{|C|}\sum_{j=1}^{|P|} \left|\gamma'_{i,j}(n) - \gamma_{i,j}\right|\right)$$

under-utilization of workers

over/under-achievement of SLAs

Constraints:

$$g_{i,j}(X_n) = \gamma_{i,j} - \gamma'_{i,j}(n) \leq 0, \forall i, j \qquad \textbf{(SLA attainments)}$$

$$h(X_n) = 1 - q(n) \leq 0, \qquad \textbf{(Queue Stability)}$$

Notation: Shifts A, Skills B, Customers C, Priorities P

State:

$$X_n = (\underbrace{\mathcal{N}_1(n), \ldots, \mathcal{N}_{|B|}(n)}_{\text{complexity queue lengths}}, \underbrace{u_{1,1}(n), \ldots\ldots, u_{|A|,|B|}(n)}_{\text{worker utilizations}}, \underbrace{\gamma'_{1,1}(n), \ldots\ldots, \gamma'_{|C|,|P|}(n)}_{\text{SLAs attained}}, q(n)),$$

Single-stage cost:

$$c(X_n) = \boxed{\left(1 - \sum_{i=1}^{|A|}\sum_{j=1}^{|B|} \alpha_{i,j} \times u_{i,j}(n)\right)} + \boxed{\left(\sum_{i=1}^{|C|}\sum_{j=1}^{|P|} \left|\gamma'_{i,j}(n) - \gamma_{i,j}\right|\right)}$$

under-utilization of workers

over/under-achievement of SLAs

Constraints:

$$g_{i,j}(X_n) = \gamma_{i,j} - \gamma'_{i,j}(n) \le 0, \forall i, j \qquad \text{(SLA attainments)}$$

$$h(X_n) = 1 - q(n) \le 0, \qquad \text{(Queue Stability)}$$

Notation: Shifts A, Skills B, Customers C, Priorities P

State:

$$X_n = (\underbrace{\mathcal{N}_1(n), \ldots, \mathcal{N}_{|B|}(n)}_{\text{complexity queue lengths}}, \underbrace{u_{1,1}(n), \ldots\ldots, u_{|A|,|B|}(n)}_{\text{worker utilizations}}, \underbrace{\gamma'_{1,1}(n), \ldots\ldots, \gamma'_{|C|,|P|}(n)}_{\text{SLAs attained}}, q(n)),$$

Single-stage cost:

$$c(X_n) = \boxed{\left(1 - \sum_{i=1}^{|A|}\sum_{j=1}^{|B|} \alpha_{i,j} \times u_{i,j}(n)\right)} + \boxed{\left(\sum_{i=1}^{|C|}\sum_{j=1}^{|P|} \left|\gamma'_{i,j}(n) - \gamma_{i,j}\right|\right)}$$

under-utilization of workers

over/under-achievement of SLAs

Constraints:

$$g_{i,j}(X_n) = \gamma_{i,j} - \gamma'_{i,j}(n) \leq 0, \forall i, j \qquad \text{(SLA attainments)}$$

$$h(X_n) = 1 - q(n) \leq 0, \qquad \text{(Queue Stability)}$$

Constrained Optimization Problem

Parameter $\theta = (\underbrace{W_{1,1}, \ldots\ldots, W_{|A|,|B|}}_{number\ of\ workers})^T$

Average Cost $J(\theta) \triangleq \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} E[c(X_m)]$

subject to

SLA constraints $G_{i,j}(\theta) \triangleq \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} E[g_{i,j}(X_m)] \leq 0,$

Queue Stability $H(\theta) \triangleq \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} E[h(X_m)] \leq 0$

θ^* cannot be found by traditional methods - not a closed form formula!

$$\max_{\lambda} \min_{\theta} L(\theta, \lambda) \triangleq J(\theta) + \sum_{i=1}^{|C|} \sum_{j=1}^{|P|} \lambda_{i,j} G_{i,j}(\theta) + \lambda_f H(\theta)$$

Three-Stage Solution:

inner-most stage simulate the SS for several time steps

intermediate stage estimate $\nabla_\theta L(\theta, \lambda)$ using simulation results and then update θ along descent direction

outer-most stage update the Lagrange multipliers λ in the ascent direction using the constraint sample

SASOC Algorithm

Multi-timescale stochastic approximation SASOC runs all three loops simultaneously with varying step-sizes

SPSA for estimating $\nabla L(\theta, \lambda)$ using simulation results

Lagrange theory SASOC does gradient descent on the primal using SPSA and dual-ascent on the Lagrange multipliers

Generalized projection All SASOC algorithms involve a certain generalized smooth projection operator that helps imitate a continuous parameter system

Update rule

$$W_i(n+1) = \bar{\Gamma}_i \left[W_i(n) + b(n) \left(\frac{\bar{L}(nK) - \bar{L}'(nK)}{\delta \Delta_i(n)} \right) \right], \forall i = 1, 2, \ldots, N$$

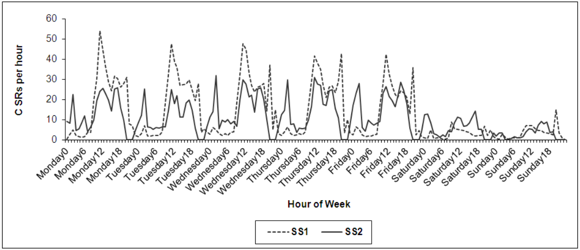
where for $m = 0, 1, \ldots, K-1$,

$$\begin{aligned}
\bar{L}(nK+m+1) &= \bar{L}(nK+m) + d(n)(l(X_{nK+m}, \lambda(nK)) - \bar{L}(nK+m)), \\
\bar{L}'(nK+m+1) &= \bar{L}'(nK+m) + d(n)(l(\hat{X}_{nK+m}, \lambda(nK)) - \bar{L}'(nK+m)), \\
\lambda_{i,j}(n+1) &= \left(\lambda_{i,j}(n) + a(n) g_{i,j}(X_n) \right)^+, \forall i = 1, 2, \ldots, |C|, j = 1, 2, \ldots, |P|, \\
\lambda_f(n+1) &= \left(\lambda_f(n) + a(n) h(X_n) \right)^+.
\end{aligned}$$

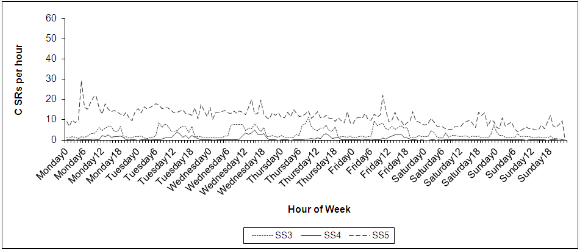
In the above, $l(X, \lambda) = c(X) + \sum_{i=1}^{|C|} \sum_{j=1}^{|P|} \lambda_{i,j} g_{i,j}(X) + \lambda_f h(X)$.

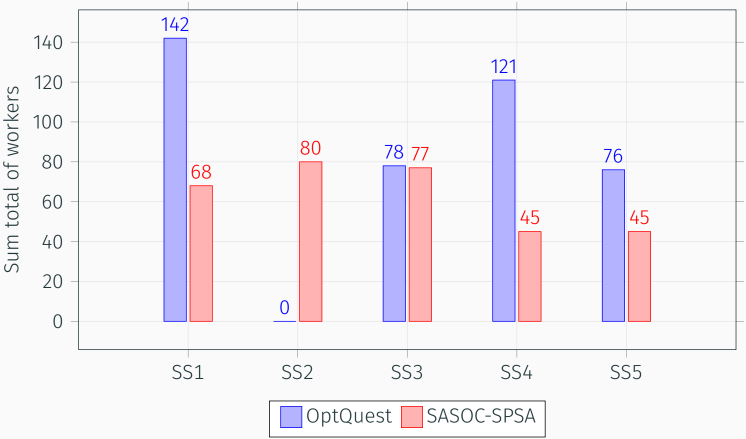
Work arrival patterns over a week for five real-life SS supporting IBM’s customers

SS1 and SS2



SS3, SS4 and SS5

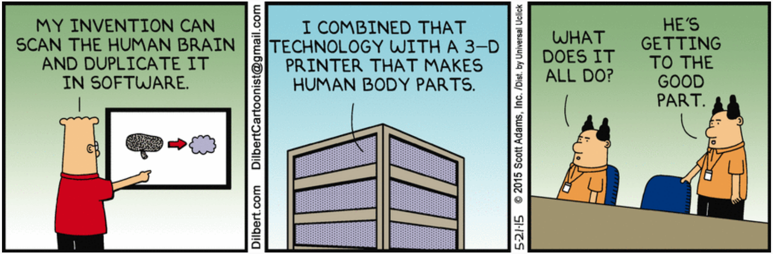




- **SASOC** is compared against **OptQuest** – a state-of-the-art optimization package – on five real-life SS via AnyLogic Simulation Toolkit
- **SASOC** is an order of magnitude faster than **OptQuest** and finds better solutions

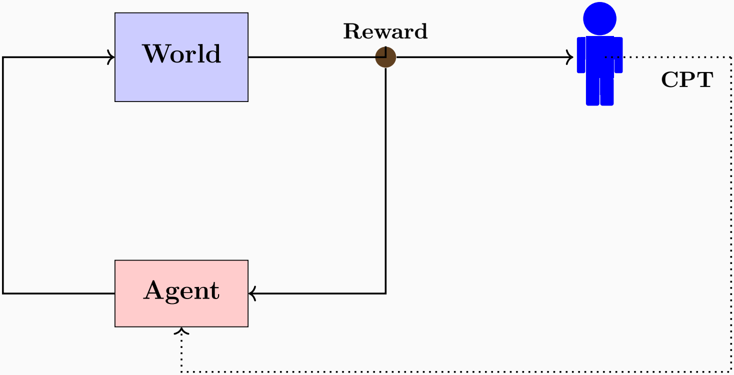
Outline

Dilbert on AI



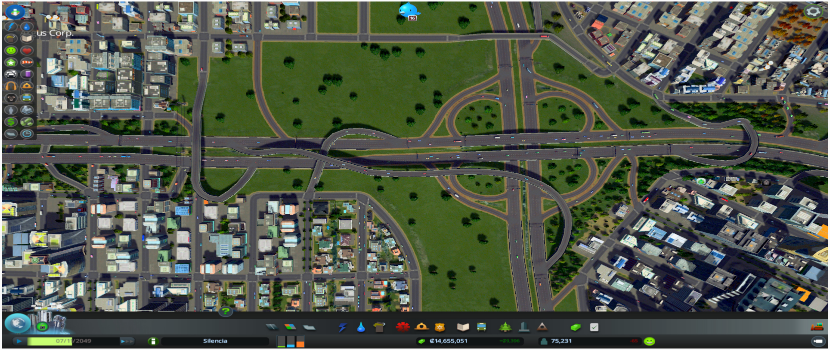
AI that benefits humans

Sequential decision making (RL/bandits) setting with rewards evaluated by **humans**



Cumulative prospect theory (CPT) *captures* ***human preferences***

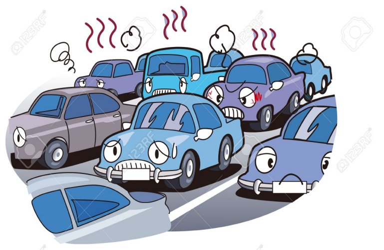
Going to office



On every day

1. Pick a route to office
2. Reach office and record (suffered) delay

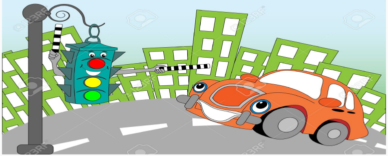
Why not distort?



Delays are stochastic

In choosing between routes, humans ***need not*** minimize **expected delay**

Plans based on average assumptions are wrong on average. – Sam L. Savage



Two-route scenario: Average delay(Route 2) slightly below that of Route 1

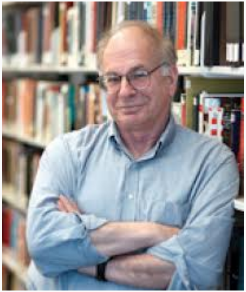
Route 2 has a *small* chance of *very* high delay, e.g. jammed traffic

I might prefer Route 1

In choosing between routes,
humans ***need not**** *minimize* ***expected delay***

Prospect Theory and its refinement (CPT)

Amos Tversky



Daniel Kahneman

Kahneman & Tversky (1979) "*Prospect Theory: An analysis of decision under risk*" is the second most cited paper in economics during the period, 1975-2000

Cumulative prospect theory - Tversky & Kahneman (1992)
   Rank-dependent expected utility - Quiggin (1982)

CPT-value

For a given r.v. X, CPT-value $\mathbb{C}(X)$ is

$$\mathbb{C}(X) := \underbrace{\int_0^{\infty} w^+\left(\mathbb{P}\left(u^+(X) > z\right)\right) dz}_{\text{Gains}} - \underbrace{\int_0^{\infty} w^-\left(\mathbb{P}\left(u^-(X) > z\right)\right) dz}_{\text{Losses}}$$

Utility functions $u^+, u^- : \mathbb{R} \to \mathbb{R}_+$, $u^+(x) = 0$ when $x \leq 0$, $u^-(x) = 0$ when $x \geq 0$

Weight functions $w^+, w^- : [0,1] \to [0,1]$ with $w(0) = 0$, $w(1) = 1$

Connection to expected value:

$$\mathbb{C}(X) = \int_0^{\infty} \mathbb{P}\left(X > z\right) dz - \int_0^{\infty} \mathbb{P}\left(-X > z\right) dz$$
$$= \mathbb{E}\left[(X)^+\right] - \mathbb{E}\left[(X)^-\right]$$

$(a)^+ = \max(a, 0)$, $(a)^- = \max(-a, 0)$

CPT-value

For a given r.v. X, CPT-value $\mathbb{C}(X)$ is

$$\mathbb{C}(X) := \underbrace{\int_0^{\infty} w^+ \left(\mathbb{P}\left(u^+(X) > z\right)\right) dz}_{\text{Gains}} - \underbrace{\int_0^{\infty} w^- \left(\mathbb{P}\left(u^-(X) > z\right)\right) dz}_{\text{Losses}}$$

Utility functions $u^+, u^- : \mathbb{R} \to \mathbb{R}_+$, $u^+(x) = 0$ when $x \leq 0$, $u^-(x) = 0$ when $x \geq 0$

Weight functions $w^+, w^- : [0, 1] \to [0, 1]$ with $w(0) = 0$, $w(1) = 1$

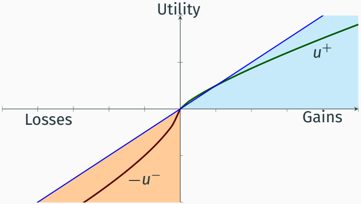
Connection to expected value:

$$\begin{aligned}\mathbb{C}(X) &= \int_0^{\infty} \mathbb{P}\left(X > z\right) dz - \int_0^{\infty} \mathbb{P}\left(-X > z\right) dz \\ &= \mathbb{E}\left[(X)^+\right] - \mathbb{E}\left[(X)^-\right]\end{aligned}$$

$(a)^+ = \max(a, 0)$, $(a)^- = \max(-a, 0)$

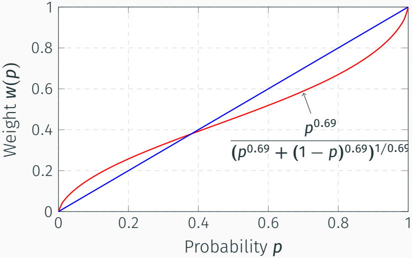
Utility and weight functions

Utility functions



For losses, the disutility $-u^-$ is **convex**, for gains, the utility u^+ is **concave**

Weight function



Overweight low probabilities, **underweight** high probabilities

CPT-value estimation

Problem: Given samples $X_1, \ldots, X_n$ of X, estimate

$$\mathbb{C}(X) := \int_0^{\infty} w^+ \left(\mathbb{P}\left(u^+(X) > z\right)\right) dz - \int_0^{\infty} w^- \left(\mathbb{P}\left(u^-(X) > z\right)\right) dz$$

Nice to have: Sample complexity $O\left(1/\epsilon^2\right)$ for accuracy ϵ

Empirical distribution function (EDF): Given samples $X_1, \ldots, X_n$ of X,

$$\hat{F}_n^+(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{(u^+(X_i)\leq x)}, \quad \text{and} \quad \hat{F}_n^-(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{(u^-(X_i)\leq x)}$$

Using EDFs, the CPT-value $\mathbb{C}(X)$ is estimated by

$$\overline{\mathbb{C}}_n = \underbrace{\int_0^\infty w^+(1-\hat{F}_n^+(x))dx}_{\text{Part (I)}} - \underbrace{\int_0^\infty w^-(1-\hat{F}_n^-(x))dx}_{\text{Part (II)}}$$

Computing Part (I): Let $X_{[1]}, X_{[2]}, \ldots, X_{[n]}$ denote the order-statistics

$$\text{Part (I)} = \sum_{i=1}^{n} u^+(X_{[i]})\left(w^+\left(\frac{n+1-i}{n}\right) - w^+\left(\frac{n-i}{n}\right)\right),$$

Empirical distribution function (EDF): Given samples $X_1, \dots, X_n$ of X,

$$\hat{F}_n^+(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{(u^+(X_i)\leq x)}, \quad \text{and} \quad \hat{F}_n^-(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{(u^-(X_i)\leq x)}$$

Using EDFs, the CPT-value $\mathbb{C}(X)$ is estimated by

$$\overline{\mathbb{C}}_n = \underbrace{\int_0^{\infty} w^+(1-\hat{F}_n^+(x))dx}_{\text{Part (I)}} - \underbrace{\int_0^{\infty} w^-(1-\hat{F}_n^-(x))dx}_{\text{Part (II)}}$$

Computing Part (I): Let $X_{[1]}, X_{[2]}, \dots, X_{[n]}$ denote the order-statistics

$$\textbf{Part (I)} = \sum_{i=1}^{n} u^+(X_{[i]}) \left(w^+\left(\frac{n+1-i}{n}\right) - w^+\left(\frac{n-i}{n}\right)\right),$$

(A1). Weights w^+, w^- are Hölder continuous, i.e.,
$|w^+(x) - w^+(y)| \leq H|x-y|^\alpha, \forall x, y \in [0,1]$

(A2). Utilities $u^+(X)$ and $u^-(X)$ are bounded above by $M < \infty$

Sample Complexity:

Under (A1) and (A2), for any $\epsilon, \delta > 0$, we have

$$\mathbb{P}\left(\left|\overline{\mathbb{C}}_n - \mathbb{C}(X)\right| \leq \epsilon\right) > 1 - \delta, \forall n \geq \ln\left(\frac{1}{\delta}\right) \cdot \frac{4H^2M^2}{\epsilon^{2/\alpha}}$$

Special Case: Lipschitz weights ($\alpha = 1$)

Sample complexity $O\left(1/\epsilon^2\right)$ for accuracy ϵ

(A1). Weights w^+, w^- are Hölder continuous, i.e.,
$|w^+(x) - w^+(y)| \leq H|x-y|^\alpha, \forall x, y \in [0,1]$

(A2). Utilities $u^+(X)$ and $u^-(X)$ are bounded above by $M < \infty$

Sample Complexity:

Under (A1) and (A2), for any $\epsilon, \delta > 0$, we have

$$\mathbb{P}\left(\left|\overline{\mathbb{C}}_n - \mathbb{C}(X)\right| \leq \epsilon\right) > 1 - \delta \,, \forall n \geq \ln\left(\frac{1}{\delta}\right) \cdot \frac{4H^2M^2}{\epsilon^{2/\alpha}}$$

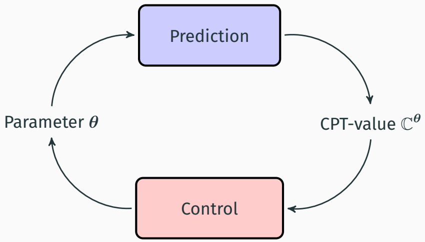
Special Case: Lipschitz weights ($\alpha = 1$)

Sample complexity $O\left(1/\epsilon^2\right)$ for accuracy $\boldsymbol{\epsilon}$

CPT-value optimization

$$\text{Find} \quad \theta^* = \arg\max_{\theta \in \Theta} \mathbb{C}(X^\theta)$$

RL application: θ = policy parameter, X^θ = return



Two-Stage Solution:

inner stage Obtain samples of X^θ and estimate $\mathbb{C}(X^\theta)$;

outer stage Update θ using gradient ascent

$\nabla_i \mathbb{C}(X^\theta)$ is not given

Update rule: $\theta^i_{n+1} = \Gamma_i\left(\theta^i_n + \gamma_n \widehat{\nabla}_i \mathbb{C}(X^{\theta_n})\right), \quad i = 1, \ldots, d.$

Projection operator (Γ_i) · Step-sizes (γ_n) · Gradient estimate ($\widehat{\nabla}_i \mathbb{C}(X^{\theta_n})$)

Challenge: estimating $\nabla_i \mathbb{C}(X^\theta)$ given only biased estimates of $\mathbb{C}(X^\theta)$

Solution: use SPSA [Spall'92]

$$\widehat{\nabla}_i \mathbb{C}(X^\theta) = \frac{\overline{\mathbb{C}}_n^{\theta_n + \delta_n \Delta_n} - \overline{\mathbb{C}}_n^{\theta_n - \delta_n \Delta_n}}{2\delta_n \Delta^i_n}$$

$\mathbb{C}_n^{\theta_n \pm \delta_n \Delta_n}$ are estimates of CPT-value for policies $\theta_n \pm \delta_n \Delta_n$.

Δ_n is a vector of independent Rademacher r.v.s and $\delta_n > 0$ vanishes asymptotically.

Update rule: $\theta^i_{n+1} = \Gamma_i \left(\theta^i_n + \gamma_n \widehat{\nabla}_i \mathbb{C}(X^{\theta_n}) \right), \quad i = 1, \ldots, d.$

Projection operator

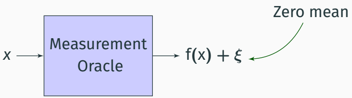
Step-sizes

Gradient estimate

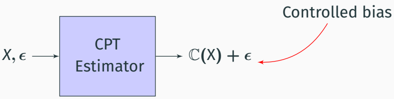
$$\widehat{\nabla}_i \mathbb{C}(X^{\theta}) = \frac{\overline{\mathbb{C}}_n^{\theta_n + \delta_n \Delta_n} - \overline{\mathbb{C}}_n^{\theta_n - \delta_n \Delta_n}}{2\delta_n \Delta^i_n}$$

$\mathbb{C}_n^{\theta_n \pm \delta_n \Delta_n}$ are estimates of CPT-value for policies $\theta_n \pm \delta_n \Delta_n$.

Δ_n is a vector of independent Rademacher r.v.s and $\delta_n > 0$ vanishes asymptotically.



Simulation optimization



CPT-value optimization

Figure 2: Overall flow of CPT-SPSA

How to choose m_n to ignore estimation bias? Ensure $\dfrac{1}{m_n^{\alpha/2}\delta_n} \to 0$

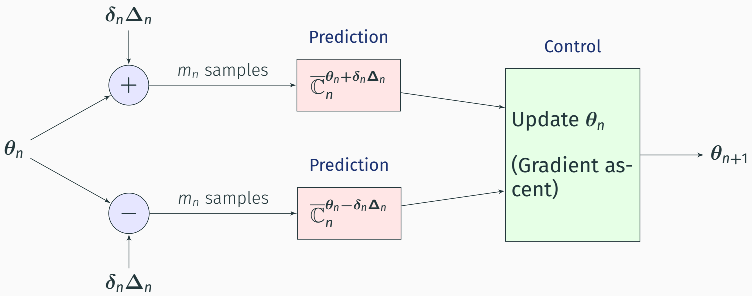
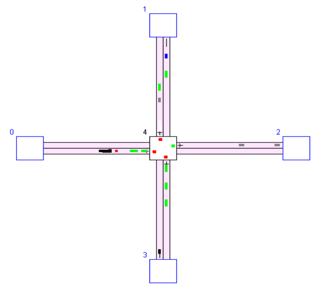


Figure 2: Overall flow of CPT-SPSA

How to choose m_n to ignore estimation bias? Ensure $\frac{1}{m_n^{\alpha/2}\delta_n} \rightarrow 0$

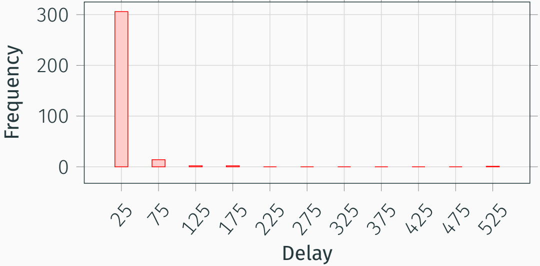
Application: Traffic signal control



- For any path $i = 1, \ldots, \mathcal{M}$ and policy θ, let
    - X_i^θ be the delay r.v.
    - B_i be the reference delay, calculated with a pre-timed traffic light controller
    - μ^i be the proportion of traffic on path i
- CPT captures the road users' evaluation of the delay

Goal: $$\max_{\theta \in \Theta} \mathrm{CPT}(X_1^\theta, \ldots, X_{\mathcal{M}}^\theta) = \sum_{i=1}^{\mathcal{M}} \mu_i^\theta \mathbb{C}(B_i - X_i^\theta)$$

AVG-SPSA



CPT-SPSA

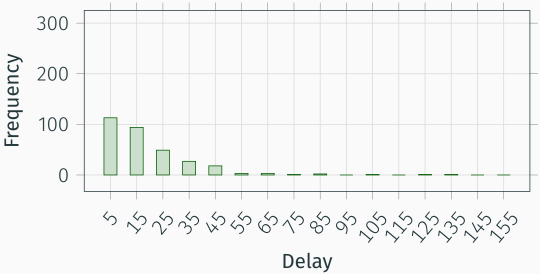


Figure 3: Histogram of the sample delays for the path from node 0 to 1 for AVG-SPSA that minimizes overall expected delay and CPT-SPSA that maximizes CPT-value of differential delay.

Conclusions

Simultaneous perturbation methods can make a difference!

- **Simulation**: problem cannot be solved via closed-form expressions. System too complex.
- **Optimization**: hand-tuning too difficult, classic gradient-based approaches are *not* directly applicable
- **Simultaneous perturbation methods**: widely applicable, easy to implement, handles noisy samples, efficient in high-dimensions!
- **Gradient/Hessian Estimation** via simultaneous perturbation trick
- **Theoretical guarantees**: nearly unbiased gradient/Hessian estimates, proven convergence to local optima
- **Applications**: from queueing networks to transportation to finance.

Acknowledgments

Shalabh Bhatnagar

(IISc)

Michael Fu

(UMD)

Steve Marcus

(UMD)

Csaba Szepesvari

(U Alberta)

H L Prasad
(Astrome Tech)

Nirmit Desai
(IBM Research)

Thank you

Selected publications

Prashanth L.A., S. Bhatnagar, Michael Fu and Steve Marcus (2015), *Adaptive system optimization using random directions stochastic approximation, IEEE transactions on Automatic Control*.

Prashanth L.A. and M. Ghavamzadeh (2013), *Actor-Critic Algorithms for Risk-Sensitive MDPs, NIPS (Full oral) (Longer version in MLJ)*.

S. Bhatnagar and **Prashanth L.A.** (2015), *Simultaneous Perturbation Newton Algorithms for Simulation Optimization, Journal of Optimization Theory and Applications*.

Prashanth L.A., A. Chatterjee and S. Bhatnagar (2014), *Two Timescale Convergent Q-learning for Sleep-Scheduling in Wireless Sensor Networks*, Wireless Networks.

Prashanth L.A., H.L. Prasad, N. Desai, S. Bhatnagar and G.Dasgupta (2015), *Simultaneous Perturbation Methods for Adaptive Labor Staffing in Service Systems*, Simulation.

H.L. Prasad, **Prashanth L.A.**, S. Bhatnagar, and N. Desai (2013), *Adaptive smoothed functional algorithms for optimal staffing levels in service systems*, Service Science (INFORMS).

Xiaowei Hu, **Prashanth L.A.**, András György and Csaba Szepesvári (2016), *(Bandit)*

Convex Optimization with Biased Noisy Gradient Oracles, AISTATS.

References

Bhatnagar, S., Fu, M. C., Marcus, S. I., Wang, I., et al. (2003). Two-timescale simultaneous perturbation stochastic approximation using deterministic perturbation sequences. *ACM TOMACS*, 13(2):180–209.

Dippon, J. (2003). Accelerated randomized stochastic optimization. *The Annals of Statistics*, 31(4):1260–1281.

Fabian, V. (1971). Stochastic approximation. In *Optimizing Methods in Statistics (ed. J.J.Rustagi)*, pages 439–470, New York. Academic Press.

Flaxman, A. D., Kalai, A. T., and McMahan, H. B. (2005). Online convex optimization in the bandit setting: gradient descent without a gradient. In *SODA*, pages 385–394.

Katkovnik, V. Y. and Kulchitsky, Y. (1972). Convergence of a class of random search algorithms. *Automation Remote Control*, 8:1321–1326.

Kiefer, J. and Wolfowitz, J. (1952). Stochastic estimation of the maximum of a regression function. *The Annals of Mathematical Statistics*, 23(3):462–466.

Kushner, H. J. and Clark, D. S. (1978). *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Springer Verlag, New York.

Maryak, J. L. and Chin, D. C. (2008). Global random optimization by simultaneous perturbation stochastic approximation. *IEEE Transactions on Automatic Control*, 53:780–783.

Nesterov, Y. and Spokoiny, V. (2011). Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics*, pages 1–40.

Polyak, B. and Tsybakov, A. (1990). Optimal orders of accuracy for search algorithms of stochastic optimization. *Problems in Information Transmission*, pages 126–133.

Prashanth, L. A., Bhatnagar, S., Fu, M., and Marcus, S. (2017). Adaptive system optimization using random directions stochastic approximation. *IEEE Transactions on Automatic Control*.

Spall, J. C. (1992). Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control*, 37(3):332–341.

Bonus Application:
Risk-Sensitive Reinforcement Learning

Risk-Sensitive Sequential Decision-Making

Optimal policy

$$\pi^* = \arg\max_{\pi} \left\{ V^{\pi}(x^0) = E\left[\sum_{n=0}^{\infty} \gamma^n \; r(x_n, \pi(x_n)) \;\middle|\; x_0 = x^0, \;\; \pi \right] \right\}$$

Value function   Reward   Policy

- a criterion that penalizes the ***variability*** induced by a given policy
- minimize some measure of ***risk*** as well as maximizing a usual optimization criterion

Objective: to optimize a risk-sensitive criterion such as

- expected exponential utility *(Howard & Matheson 1972)*
- variance-related measures *(Sobel 1982; Filar et al. 1989)*
- percentile performance *(Filar et al. 1995)*

Open Question ???

construct conceptually meaningful and computationally tractable criteria

mainly negative results:

(e.g., Sobel 1982; Filar et al., 1989; Mannor & Tsitsiklis, 2011)

Discounted Reward MDPs

A class of parameterized stochastic policies $\{\pi(\cdot|x;\theta),\ x \in \mathcal{X},\ \theta \in \Theta \subseteq \Re^{\kappa_1}\}$

Return: $D^\theta(x) = \sum_{n=0}^{\infty} \gamma^n r(x_n, a_n) \mid x_0 = x,\ \theta$

Mean of Return: $V^\theta(x) = \mathbb{E}\big[D^\theta(x)\big]$

Variance of Return: $\Lambda^\theta(x) = \mathbb{E}\big[D^\theta(x)^2\big] - V^\theta(x)^2 = U^\theta(x) - V^\theta(x)^2$

Optimization Problem

$$\max_\theta\ V^\theta(x^0) \quad \text{s.t.} \quad \Lambda^\theta(x^0) \leq \alpha$$

$$\Updownarrow$$

$$\max_\lambda\ \min_\theta\ L(\theta,\lambda)\ \triangleq\ -V^\theta(x^0) + \lambda\big(\Lambda^\theta(x^0) - \alpha\big)$$

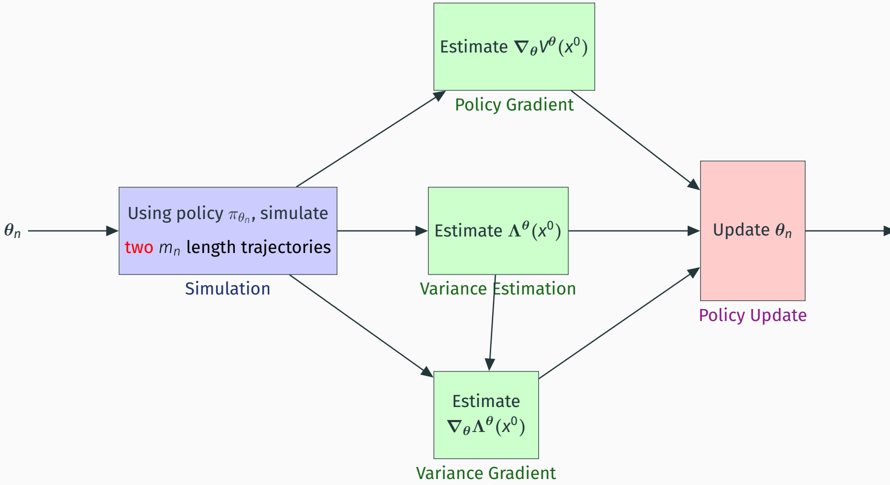


Figure 4: Solving the risk-sensitive MDP

Why Estimating the Gradient is Challenging?

The Gradient $\nabla_\theta L(\theta, \lambda)$

$$(1-\gamma)\nabla_\theta V^\theta(x^0) = \sum_{x,a} d_\gamma^\theta(x, a|x^0) \; \nabla_\theta \log \pi(a|x;\theta) \; Q^\theta(x, a)$$

$$(1-\gamma^2)\nabla_\theta U^\theta(x^0) = \sum_{x,a} \widetilde{d}_\gamma^\theta(x, a|x^0) \; \nabla_\theta \log \pi(a|x;\theta) \; W^\theta(x, a)$$

$$+ 2\gamma \sum_{x,a,x'} \widetilde{d}_\gamma^\theta(x, a|x^0) \; P(x'|x, a) \; r(x, a) \; \nabla_\theta V^\theta(x')$$

$d_\gamma^\theta(x, a|x^0)$ and $\widetilde{d}_\gamma^\theta(x, a|x^0)$ are γ and γ^2 discounted visiting state distributions of the Markov chain under policy θ

Why Simultaneous Perturbation?

Challenge: estimating $\nabla_\theta L(\theta, \lambda)$

- two different sampling distributions (d_γ^θ and $\widetilde{d}_\gamma^\theta$) used for $\nabla V^\theta(x^0)$ and $\nabla U^\theta(x^0)$
- $\nabla V^\theta(x')$ appears in the second sum of $\nabla U^\theta(x^0)$ equation

Solution: use SPSA *(Spall 1992)*

$$\nabla_i V^{\theta_n}(x^0) \quad \approx \quad \frac{V^{\theta_n + \beta_n \Delta_n}(x^0) - V^{\theta_n}(x^0)}{\beta_n \Delta_n^{(i)}}, \qquad i = 1, \ldots, \kappa_1$$

$\Delta_n = (\Delta_n^{(1)}, \ldots, \Delta_n^{(\kappa_1)})^\top$ is a vector of independent Rademacher random variables and

β_n are perturbation constants that vanish asymptotically

Traffic Signal Control MDP:

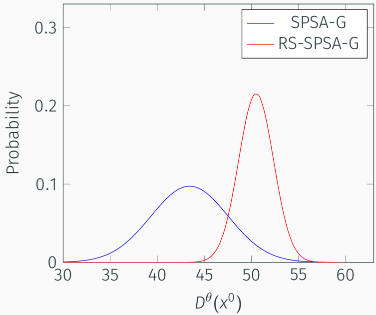
State. $x_n = (\underbrace{q_1(n), \cdots, q_N(n)}_{\text{queue lengths}}, \underbrace{t_1(n), \cdots, t_N(n)}_{\text{elapsed times}})$

Actions. $a_n = \{\text{feasible sign configurations in state } s_n\}$

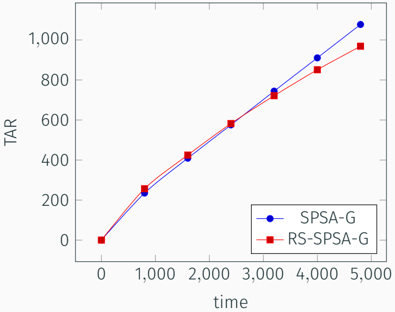
Cost.
$$r(x_n, a_n) = -\left[\xi_1 \times \left(\sum_{i \in I_p} (q_i(n) + t_i(n))\right) + \xi_2 \times \left(\sum_{i \notin I_p} (q_i(n) + t_i(n))\right)\right]$$

Aim: find a risk-sensitive control strategy that minimizes the total delay experienced by road users, while also reducing the variations

Simulation Results



(a) Distribution of $D^\theta(x^0)$



(b) Total Arrived Road Users