Visualizing and Understanding the Performance of Randomized Optimization Methods

Rebecca A Lopez † ⋆

Mentored by Matt Menickelly⋆ and Stefan M. Wild ⋆

⋆ Mathematics and Computer Science Division, Argonne National Laboratory

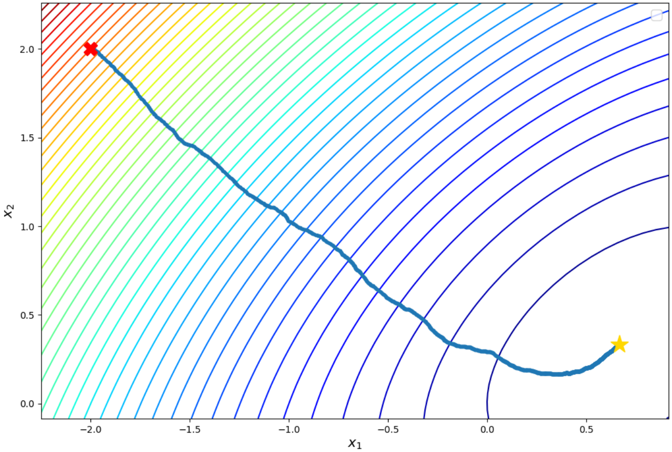
† Department of Applied Mathematics, University of Washington

August 6, 2021

Overview

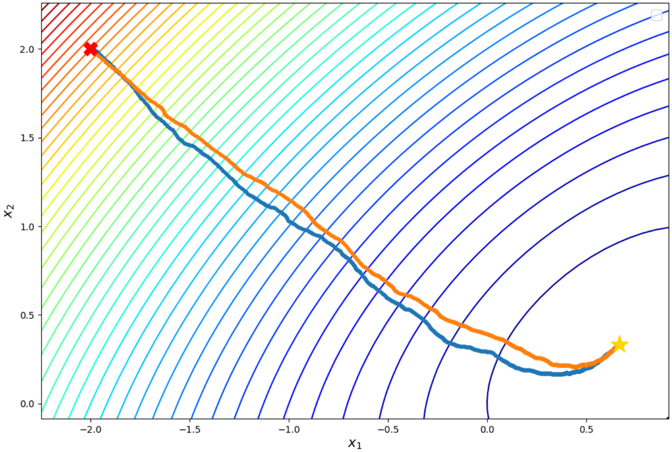
- Develop a better understanding of optimization algorithms that contain some form of randomization in their inner workings

- Produce different ways to visualize the mechanics of an optimization algorithms



Overview

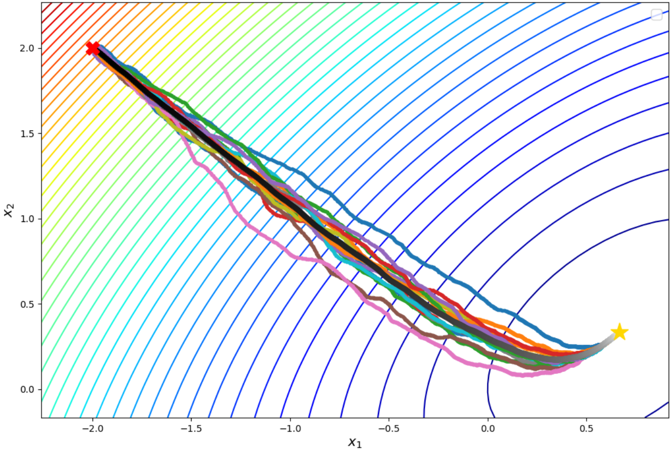
- Develop a better understanding of optimization algorithms that contain some form of randomization in their inner workings
- Produce different ways to visualize the mechanics of an optimization algorithms



Overview

- Develop a better understanding of optimization algorithms that contain some form of randomization in their inner workings

- Produce different ways to visualize the mechanics of an optimization algorithms



Randomized Algorithms for $\min_{\boldsymbol{x} \in \mathbb{R}^n} \; f(\boldsymbol{x})$

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \boldsymbol{d}_k, \qquad k = 0, 1, 2, \ldots,$$

where the direction $\boldsymbol{d}_k \in \mathbb{R}^n$ is *randomized* in some sense

1. Standard Gaussian direction: Let $\boldsymbol{\xi}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$

$$\boldsymbol{d}_k(\boldsymbol{\xi}_k) = \left(\nabla_{\boldsymbol{x}} f(\boldsymbol{x}_k)^\top \boldsymbol{\xi}_k \right) \boldsymbol{\xi}_k.$$

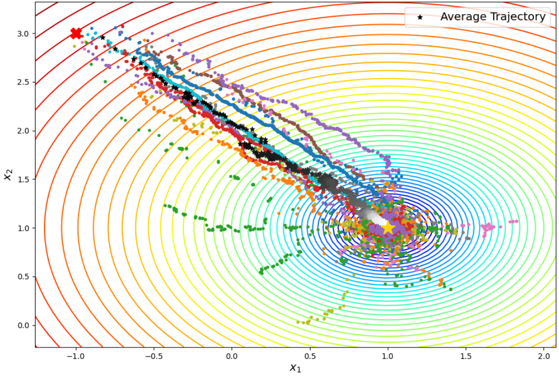
2. Randomized coordinate descent (RCD): Random coordinate $\xi_k \sim \mathcal{U}[1, \ldots, n]$

$$\boldsymbol{d}_k(\xi_k) = \left(\nabla_{\boldsymbol{x}} f(\boldsymbol{x}_k)^\top \boldsymbol{e}_{\xi_k} \right) \boldsymbol{e}_{\xi_k} = \frac{\partial}{\partial x_{\xi_k}} f(\boldsymbol{x}_k) \boldsymbol{e}_{\xi_k}.$$

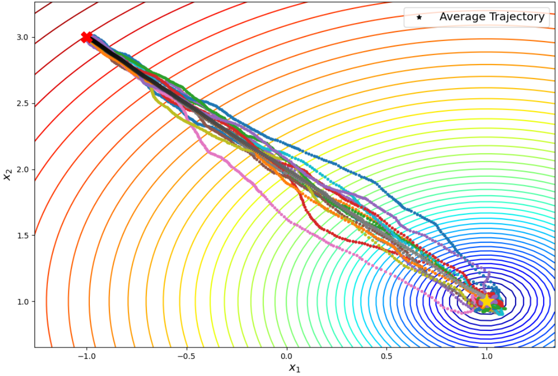
Step size α_k is typically chosen ahead of time or adaptively (e.g., Adam)

Visualizations of 2-D Convex Funnel

Comparison between the step size of $\alpha_k = \frac{M}{k+1}$ (Left) and the Adam method (Right):



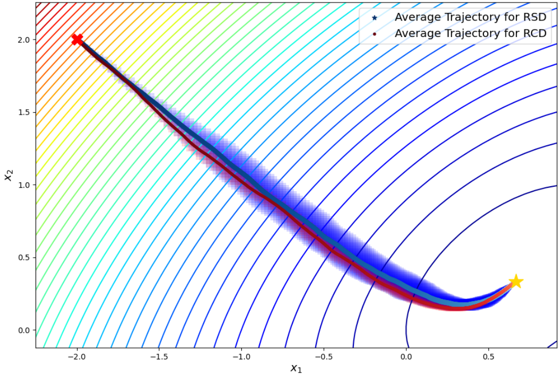
1,000 iterations

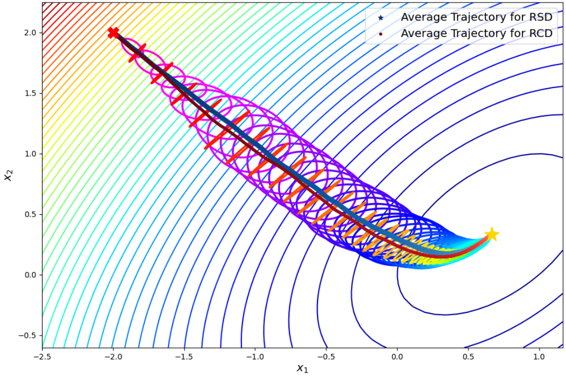


500 iterations

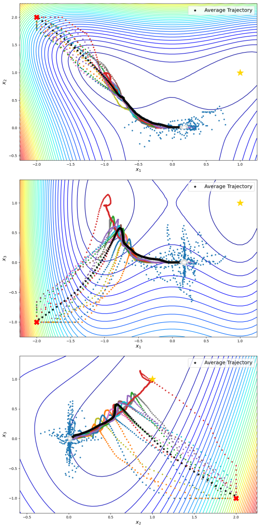
Visualizations of 2-D Convex Nesterov Quadratic

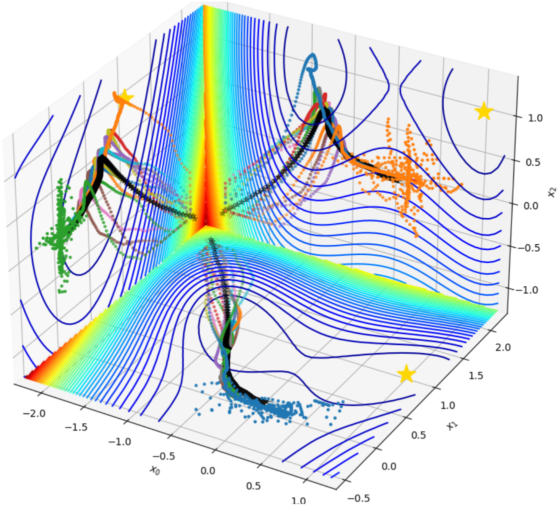
To understand variability, we calculate standard deviation for each of our $\boldsymbol{x}$ values, creating a two-dimensional confidence intervals (Left) as well as look at covariance confidence ellipses (Right).



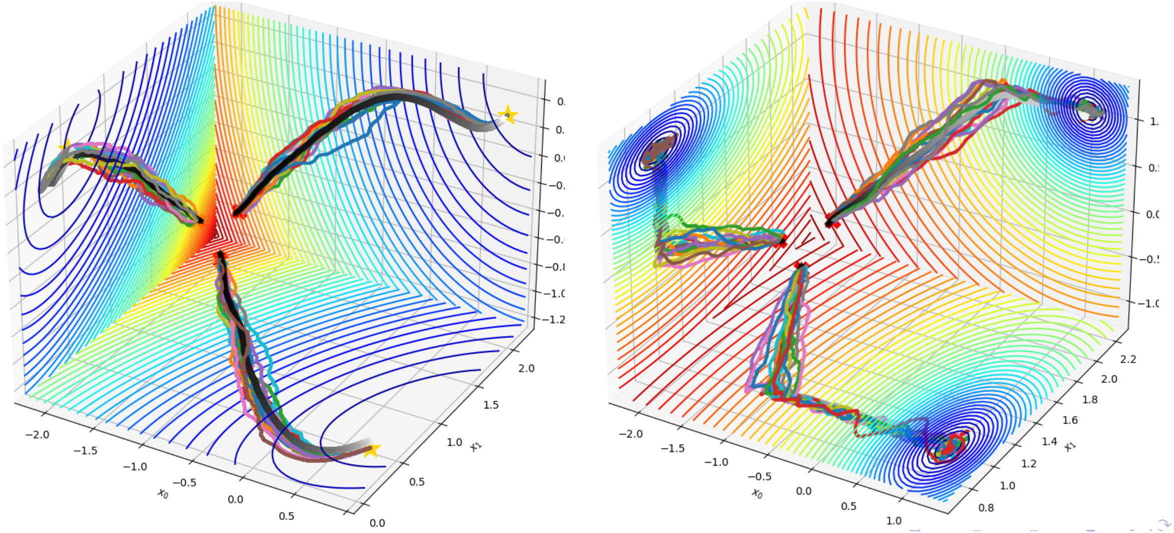


Visualizations of N-D Chained Alternative Rosenbrock





Visualizations of N-D Convex Nesterov (L) & Convex Funnel (R)



Future Work & Acknowledgments

- Conduct visualizations on additional nonconvex functions with higher dimensionality
- Apply principal component analysis and produce visualizations for the understanding of such principal components

Future Work & Acknowledgments

- Conduct visualizations on additional nonconvex functions with higher dimensionality
- Apply principal component analysis and produce visualizations for the understanding of such principal components

I would like to acknowledge and thank the Argonne National Laboratory as well as the National GEM Consortium.

Thank you for listening!

Any Questions?

Feel free to contact me through `rlopez@uw.edu`.