CGL Workshop 2013 – Athens

Optimization and Learning with Random Pursuit

Sebastian U. Stich

joint work with
Christian L. Müller, Bernd Gärtner

ETH Zürich

Oct 2, 2013

Abstract

We consider unconstrained randomized optimization of smooth convex functions in the gradient-free setting with Random Pursuit (RP). This algorithm only uses zeroth-order information of the objective function and computes an approximate solution by repeated optimization along a randomly chosen direction.

State-of-the-art (derivative-free) optimization algorithms often build and iteratively adapt a quadratic model of the objective function based only on the queried zeroth-order information. This allows to significantly increase the convergence rate of the optimization algorithm if the model describes the local geometry accurately enough.

We demonstrate how RP can not only be used for optimization, but the same algorithm can be used to build and iteratively update a quadratic model of a convex objective function. This model allows RP to achieve optimal convergence rates on quadratic functions. If time allows, we will also present some illustrative numerical experiments.

Table of contents

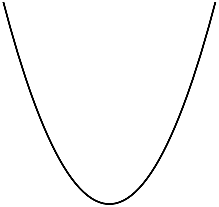
Derivative-Free Optimization

$\mathbf{x} \rightarrow$  $\rightarrow f(\mathbf{x})$

Given: $f\colon \mathbb{R}^n \rightarrow \mathbb{R}$

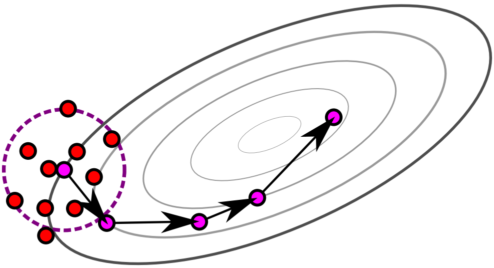
Goal: $\mathbf{x}^* \in \{\mathbf{x} \mid f(\mathbf{x}) - \min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \leq \epsilon\}$

- oracle access to $f(\mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^n$
- especially no access to gradients (may not even exist)
- no knowledge of the structure of f, but ...

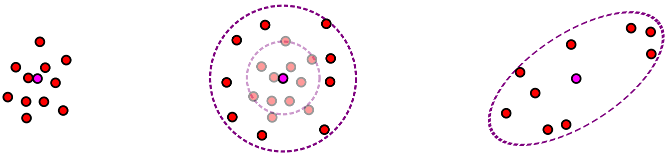


Assumption

- f convex



Internal Parameters



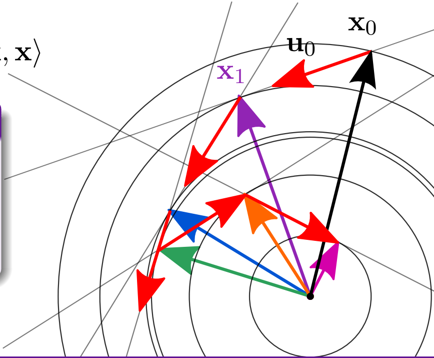
A Random Walk

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \langle \mathbf{x}_k, \mathbf{u}_k \rangle \, \mathbf{u}_k$$

- $\mathbf{x}_k \in E$
- $\mathbf{u}_k \in S \subset E$ uniform at random
- $\alpha > 0$ s.t. $\forall \mathbf{x}$: $\mathbb{E}_{\mathbf{u}}[\langle \mathbf{x}, \mathbf{u} \rangle^2] \geq \alpha \langle \mathbf{x}, \mathbf{x} \rangle$

Example

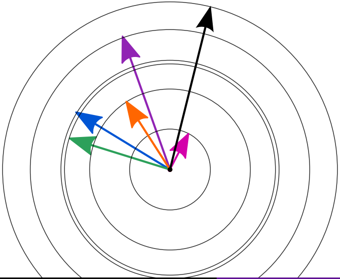
- $E = \mathbb{R}^n$
- $S = \{\mathbf{x} \mid \|\mathbf{x}\| = 1\}$
- $\alpha = \frac{1}{n}$



Two components

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \langle \mathbf{x}_k, \mathbf{u}_k \rangle \, \mathbf{u}_k$$

$\|\mathbf{x}_k\|$	$\mathbb{E}[\|\mathbf{x}_{k+1}\|^2] \leq (1-\alpha)\,\|\mathbf{x}_k\|^2 \quad (\to 0)$
$\frac{\mathbf{x}_k}{\|\mathbf{x}_k\|}$	Random Walk on the unit sphere.



Random Pursuit

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \sigma_k \mathbf{u}_k$$

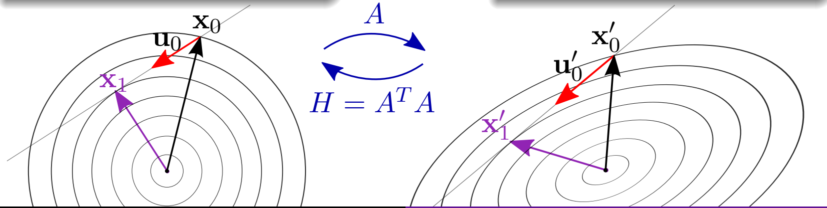
$$f(\mathbf{x}) = \mathbf{x}^T I_n \mathbf{x} \qquad f(\mathbf{x}) = \mathbf{x}^T H \mathbf{x}$$

$$\sigma_k = -\mathbf{x}_k^T \mathbf{u}_k \qquad \sigma_k = -\frac{\mathbf{x}_k^T H \mathbf{u}_k}{\mathbf{u}_k H \mathbf{u}_k}$$

$$\mathbb{E}[f(\mathbf{x}_{k+1})] \leq \left(1 - \frac{\lambda_{\min}(H)}{\mathrm{Tr}[H]}\right) f(\mathbf{x}_k)$$

✓ optimal $\left(1 - \frac{1}{n}\right)$

↯ → learn H



Metric Learning

minimize $g(X) = \frac{1}{2} \|X - H\|_F^2$ **with Random Pursuit**

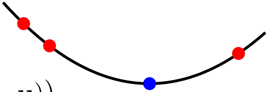
$$E_{k+1} = E_k - \langle E_k, U_k \rangle U_k$$

where $E_k := X_k - H$ is the error at step k.

$U_k \in S$ — we take $S = \left\{ U = \mathbf{u}^T \mathbf{u} \mid \|\mathbf{u}\| = 1 \right\}$

- $\|U\|_F = 1$
- $\mathbb{E}_U[\langle E, U \rangle] = \mathbb{E}_{\mathbf{u}}[(\mathbf{u}^T E \mathbf{u})^2] \geq \frac{2}{n(n+2)} \|E\|_F^2$

$\mathbf{u}^T E \mathbf{u}$ — can be estimated by knowing 3 function values on the line $\{\mathbf{y} \mid \mathbf{x} + t\mathbf{u}\}$



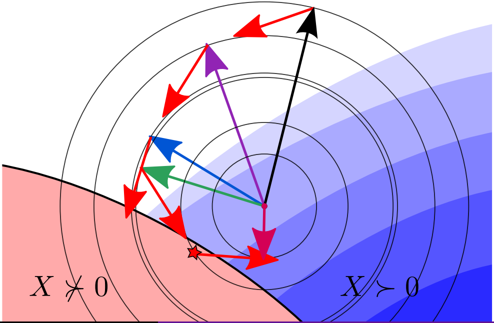
$$\langle X_k - H, U \rangle = \mathbf{u}^T X_k \mathbf{u} + \left(f(\mathbf{x} + \mathbf{u}) - 2f(\mathbf{x}) + f(\mathbf{x} - \mathbf{u}) \right)$$

Variable Metric Random Pursuit

Combine optimization ($\mathbf{x}_k$) with learning (X_k)

$$\mathbb{E}[f(\mathbf{x}_{k+1})] \leq \left(1 - \frac{\lambda_{\min}(HX_k^{-1})}{\mathrm{Tr}[HX_k^{-1}]}\right) f(\mathbf{x}_k)$$

- Balance optimization and learning
  - initialization, special cases: f "almost" quadratic, etc.
- X_k must be positive definite



References

- S.U. Stich, C.L. Müller and B. Gärtner, Variable Metric Random Pursuit, *submitted*, 2012.
- D. Leventhal and A.S. Lewis, Randomized Hessian estimation and directional search, 2011.
- S.U. Stich, C.L. Müller and B. Gärtner, Optimization of convex functions with Random Pursuit, 2012.
- S.U. Stich, Convergence of Local Search, *Manuscript*, 2013+.
- S.U. Stich, C.L. Müller and B. Gärtner, Matrix-valued Iterative Random Projections, *Manuscript*, 2013+.

Acknowledgement

Project CG Learning, FET-Open grant number: 255827