Mittagsseminar

Variable Metric Random Pursuit

Sebastian U. Stich

joint work with
Christian L. Müller, Bernd Gärtner

ETH Zürich

December 4, 2012

Abstract

A continuous function f is to be minimized. Similar as in the discrete setting, people often use Randomized Search Heuristics for this task. While they often give good results in practice, most of them lack a thorough theoretical convergence analysis.

In this talk, we present and analyze Variable Metric Random Pursuit. This is an iterative algorithm where in every step a new approximation is calculated by searching along a randomly chosen direction. However, typically not every direction yields the same progress. The algorithm tries to "learn" successful search directions and samples them more often.

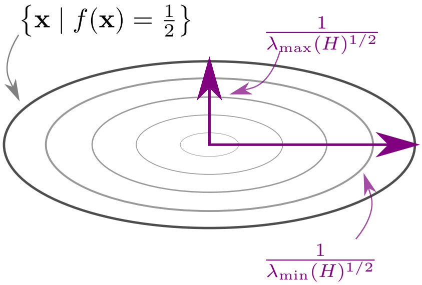
Table of contents

The recipe

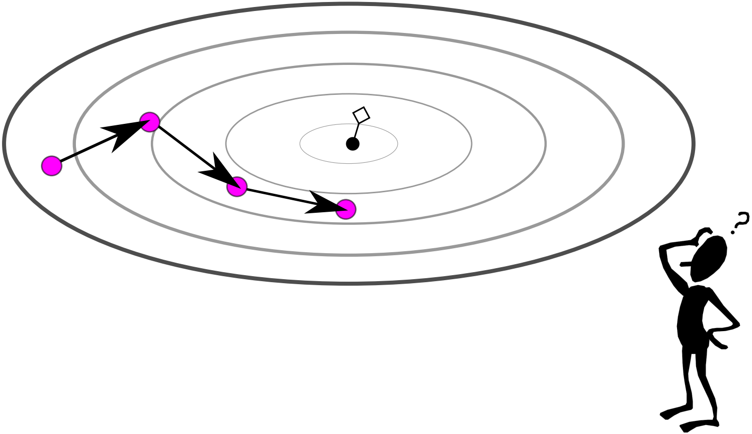
Problem: $\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x})$	$\rightarrow$	fancy heuristic	$\rightarrow$	apply		done ✓
$\downarrow$		$\downarrow$				
(general) problem class non-convex functions	$\rightarrow$	clean code, main features of state of the art algorithm	$\rightarrow$	analyze		PHD ✓
$\downarrow$		$\downarrow$				
convex functions this talk: quadratic	$\rightarrow$	2 features from 2 algorithms	$\rightarrow$	analyze		this works ✓

Quadratic functions

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T H \mathbf{x}$$



Where to go?

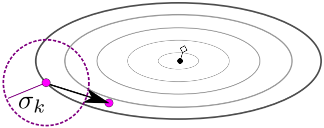


Where to go?

$$\mathbf{x}_{k+1} = \begin{cases} \mathbf{x}_k + \sigma_k \mathbf{u}_k & \text{if better} \\ \mathbf{x}_k & \text{otherwise} \end{cases}$$

(1+1)-Evolution Strategy (ES)

- $\mathbf{u}_k \sim \mathcal{N}(0, I_n)$
- σ_k "stepsize", empirically determined $\mathbf{Pr}\left[f(\mathbf{x}_k + \sigma_k \mathbf{u}_k) \leq f(\mathbf{x}_k)\right] = \frac{1}{5}$



$\downarrow$

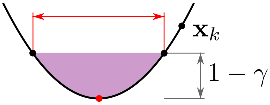
Random Pursuit

- $\mathbf{u}_k$ unifom random unit vector (direction)
- σ_k "best possible": line search in direction $\mathbf{u}_k$

Line Search

Sufficient decrease $0 < \gamma \leq 1$:

$$f(\mathbf{x}_{k+1}) \leq (1-\gamma) f(\mathbf{x}_k) + \gamma \min_{t \in \mathbb{R}} f(\mathbf{x}_k + t\mathbf{u}_k)$$



- We don't care about implementation: oracle
- count the number of oracle calls/line searches
- this talk: $\gamma = 1$
- (1+1)-ES with optimal scale σ_k:
  sample only one point $\Leftrightarrow$ γ is "large enough" *in expectation*

Convergence I

Let $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T H \mathbf{x}$.

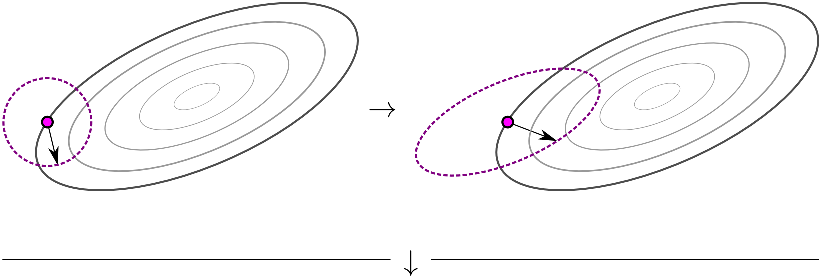
After k steps:

$$\mathbb{E}\left[f(\mathbf{x}_k)\right] \leq \left(1 - \frac{\lambda_{\min}(H)}{\mathrm{Tr}[H]}\right)^k f(\mathbf{x}_0)$$

- for $H = I_n$ the factor equals $\left(1 - \frac{1}{n}\right)$, this is optimal
- what if $H \neq I_n$?

Some directions are better than others. . .

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



Fixed Metric Random Pursuit

- $\mathbf{u}_k := \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|_2}$ random unit vector (direction) $\mathbf{v}_k \sim \mathcal{N}(0, \Sigma)$
- σ_k "best possible": line search in direction $\mathbf{u}_k$

Convergence II

Let $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T H \mathbf{x}$.

After k steps:

$$\mathbb{E}\left[f(\mathbf{x}_k)\right] \leq \left(1 - \frac{\lambda_{\min}(H\Sigma)}{\mathrm{Tr}[H\Sigma] + 2\lambda_{\max}(H\Sigma)}\right)^k f(\mathbf{x}_0)$$

- for $\Sigma = H^{-1}$ the factor equals $\left(1 - \frac{1}{n+2}\right)$
- how can we find Σ?

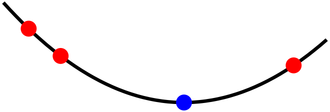
How to find Σ?

Goal: find $\Sigma^{-1} = H$

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

X_k is a $n \times n$ matrix:

- search direction $U_k = \mathbf{u}_k \mathbf{u}_k^T$, with $\mathbf{u}_k \sim \mathcal{N}(\mathbf{0}, I_n)$
- stepsize: exact line search, use the fact that g is a quadratic function: 3 function evaluations are sufficient



$$g(X_k + \alpha U_k) = \underbrace{\frac{1}{2} \|X_k - H\|_F^2 + \frac{1}{2} \|U_k\|_F^2}_{\text{constant}} + \alpha \langle X_k - H, U_k \rangle$$

Observation: $\langle H, U_k \rangle = f(\mathbf{x} + \mathbf{u}_k) - 2f(\mathbf{x}) + f(\mathbf{x} - \mathbf{u}_k)$

Randomized Hessian Update

Randomized Hessian Update [Leventhal, Levis, 2011]

- $U_k = \mathbf{u}_k \mathbf{u}_k^T$, with $\mathbf{u}_k \sim \mathcal{N}(\mathbf{0}, I_n)$ search direction
- $X_{k+1} = X_k - \langle X_k - H_k, U_k \rangle U_k$ (exact line search)

$$\mathbb{E}\left[\|X_k - H\|_F^2\right] \leq \left(1 - \frac{2}{n(n+2)}\right)^k \|X_0 - H\|_F^2$$

$\downarrow$

Variable Metric Random Pursuit

- Random Pursuit with $\mathbf{u}_k \sim \mathcal{N}(0, \Sigma_k)$
- update $\Sigma_k^{-1} \to \Sigma_{k+1}^{-1}$ with Randomized Hessian Update

Does this help?

- If the error $\|X_k - H\|_F$ is relatively large we don't know.
- If the error is relatively small $\|X_k - H\|_F \leq c\lambda_{\min}(H)$, $c < 1$:

$$1 - \frac{\lambda_{\min}(H\Sigma_k)}{\mathrm{Tr}[H\Sigma_k] + 2\lambda_{\max}(H\Sigma_k)} \leq 1 - \frac{(1-c)^2}{n+2}$$

- the progress is independent of H!

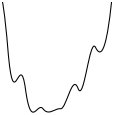
Outlook and Open problems

1. Dependence on $\lambda(H)$
2. When can we stop (with the learning)?

(general) problem class non-convex functions	$\rightarrow$	clean code, main features of state of the art algorithm	$\rightarrow$	analyze
$\downarrow\uparrow$		$\downarrow\uparrow$		
convex functions	$\rightarrow$	2 features from 2 algorithms	$\rightarrow$	analyze

3. Metric learning scheme from state of the art algorithm (update directly on Σ_k instead of Σ_k^{-1})
4. *some* non-convex functions

Thank you



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, C.L. Müller, On spectral invariance of Randomized Hessian and Covariance Matrix Adaptation schemes. In: PPSN, 2012.
- S.U. Stich, Convergence of Local Search, *Manuscript*, 2012.
- S.U. Stich, C.L. Müller and B. Gärtner, Matrix-valued Iterative Random Projections, *Manuscript*, 2012.

Acknowledgement

Project CG Learning, FET-Open grant number: 255827