ISMP 2012 – Berlin

# Convergence of Local Search


Sebastian U. Stich

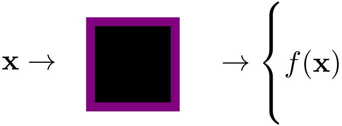
Institute of Theoretical Computer Science
Department of Computer Science
ETH Zürich


August 22, 2012

# Table of contents

# Black-box optimization

**Given:** $f : \mathbb{R}^n \mapsto \mathbb{R}$
**Goal:** $\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$



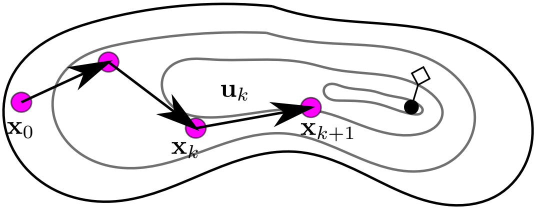

- **Problem class** $f$ convex
- **Oracle** access to $f(\mathbf{x})$, maybe $\nabla f(\mathbf{x})$
- **Complexity:** number of oracle calls sufficient to solve any problem of the class
- **Solution:** $\mathbf{y} : f(\mathbf{y}) - \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) \leq \epsilon$

# Motivation I - Local Search Algorithms

**Local Search Scheme:**

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

$\lambda_k$ step size $\mathbf{u}_k$ search direction



# Motivation II

**Local Search**

Derivative Free

Derivatives

RG

Randomized

Deterministic

GM NM

SPSA RP (1+1)-ES Variable Metric RH

Trust Region CMA-ES

**special applications/design principles**

fast

noise

inspired by nature

multi-funnel

non-convex

robust

invariance principles

scale invariant

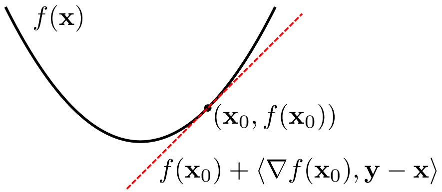
no derivatives

first step when analyzing new scheme:
**show consistency on convex functions**

# Convex functions I

**first-order condition:**

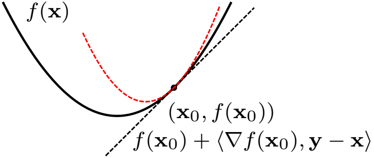
$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle\,, \quad \forall \mathbf{x}, \mathbf{y} \in E$$



# Convex functions II - Quadratic Model

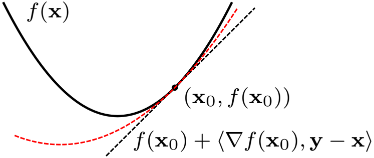
**Quadratic upper bound:**

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_{L_1}^2$$

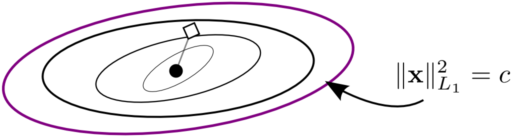



**Quadratic lower bound: (strongly convex)**

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_{M}^2$$

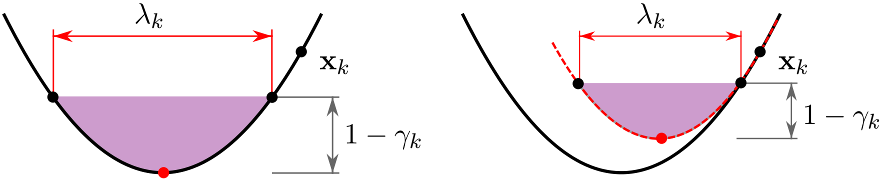



- Fully quadratic model: $L_1, M \succ 0$, $\|\mathbf{x}\|_M^2 := \mathbf{x}^T M \mathbf{x}$



# Sufficient Decrease

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




(D1) $\Rightarrow$ (D2)

**Sufficient decrease $0 < \gamma_k \leq 1$:**

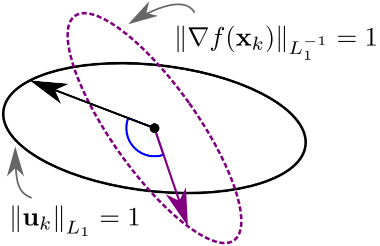
$$f(\mathbf{x}_k) - f(\mathbf{x}_{k+1}) \geq \gamma_k \left( f(\mathbf{x}_k) - \min_{t \in \mathbb{R}} f(\mathbf{x}_k + t\mathbf{u}_k) \right) \qquad \text{(D1)}$$

$$f(\mathbf{x}_k) - f(\mathbf{x}_{k+1}) \geq \gamma_k \frac{\langle \nabla f(\mathbf{x}_k), \mathbf{u}_k \rangle^2}{2 \|\mathbf{u}_k\|_{L_1}^2} \qquad \text{(D2)}$$

# Search Directions

**Quality of search direction $\mathbf{u}_k$ :**

$$\beta_k := \left\langle \frac{\nabla f(\mathbf{x}_k)}{\|\nabla f(\mathbf{x}_k)\|_{L_1^{-1}}}, \frac{\mathbf{u}_k}{\|\mathbf{u}_k\|_{L_1}} \right\rangle$$




- $\gamma_k$ : quality of the stepsize
- $\beta_k^2$ : quality of the search direction

# Single step progress

**Quadratic upper bound:**

$$f(\mathbf{x}_k) - f(\mathbf{x}_{k+1}) \geq \frac{\beta_k^2 \gamma_k}{2} \|\nabla f(\mathbf{x}_k)\|^2_{L_1^{-1}}$$

$+$ **Quadratic lower bound:**

$$f(\mathbf{x}_k) - f(\mathbf{x}_{k+1}) \geq \underbrace{\lambda_{\min}(M L_1^{-1})}_{:=\kappa} \beta_k^2 \gamma_k \left(f(\mathbf{x}_k) - f(\mathbf{x}^*)\right)$$

$=$ **Progress:**

$$\underbrace{f(\mathbf{x}_{k+1}) - f(\mathbf{x}^*)}_{:=f_{k+1}} \leq \left(1 - \kappa \beta_k^2 \gamma_k\right) \cdot \underbrace{\left(f(\mathbf{x}_k) - f(\mathbf{x}^*)\right)}_{:=f_k}$$

# Global convergence

After $N$ iterations:

$$\ln f_N - \ln f_0 \leq \sum_{k=0}^{N-1} \ln f_{k+1} - \ln f_k$$
$$\leq -\kappa \sum_{k=0}^{N-1} \beta_k^2 \gamma_k$$

$$f_N \leq f_0 \cdot \exp\left[ -\kappa \underbrace{\sum_{k=0}^{N-1} \beta_k^2 \gamma_k}_{:= S_N} \right]$$

# A typical example

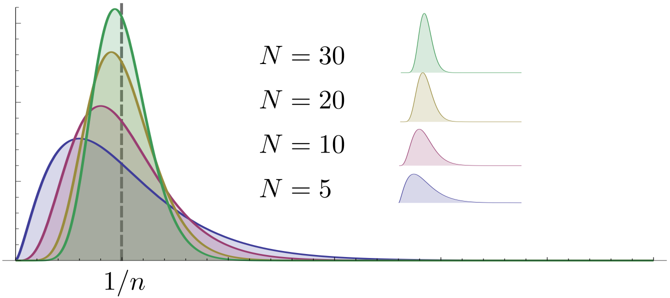
$$\mathbf{u}_k \sim \mathcal{N}(\mathbf{0}, L_1^{-1})$$

We calculate:

$$\mathbb{E}\left[\beta_k^2\right] = \frac{1}{n} \qquad \text{and} \qquad \mathbb{V}\left[\beta_k^2\right] = \frac{2(n-1)}{n^2(n+2)}$$

- independent of $\nabla f(\mathbf{x}_k)$
- $\beta_k^2$ follows a beta distribution $\sim \beta\left(\frac{1}{2}, \frac{n-1}{2}\right)$
- the sum $\sum \beta_k^2$ is concentrated around its expectation $\frac{N}{n}$ (measure concentration on the sphere)

# Convergence with high probability



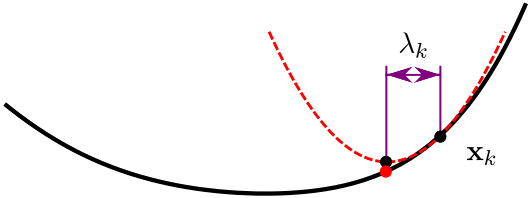

Let $0 < \epsilon \leq 1$,

$$\mathbb{P}\left[\sum_{k=0}^{N-1} \beta_k^2 \leq (1-\epsilon)\frac{N}{n}\right] \leq \exp\left[-\frac{\epsilon^2-\epsilon^3}{8}N\right]$$

# Example I - Random Gradient Method [Nesterov 2011]

$\mathbf{u}_k \sim \mathcal{N}(\mathbf{0}, L_1^{-1})$ + estimate stepsize with gradient oracles:

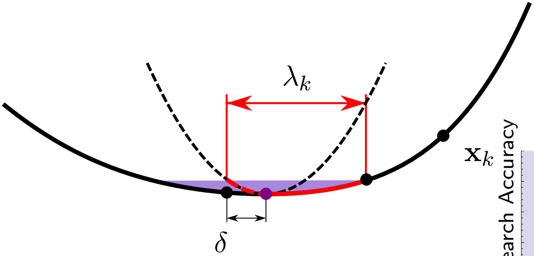
$$\lambda_k \propto -f'(\mathbf{x}_k, \mathbf{u}_k)$$

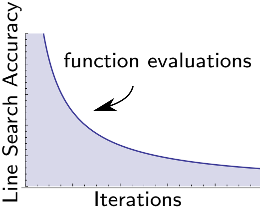


# Example II - Random Pursuit

$\mathbf{u}_k \sim \mathcal{N}(\mathbf{0}, L_1^{-1})$ + (approximate) line search:

$$\lambda_k \in \left[\mu \cdot \arg\min_{t \in \mathbb{R}} f(\mathbf{x}_k + t\mathbf{u}_k)\,, \arg\min_{t \in \mathbb{R}} f(\mathbf{x}_k + t\mathbf{u}_k) + \delta\right]$$

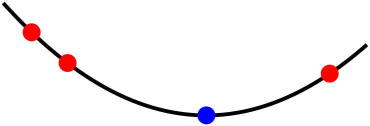



- find the optimal accuracy $\mu$

# Example III

**quadratic function:** 3 function evaluations are sufficient



Where to find quadratic functions?

# Example III - Random Hessian [Lewis, Leventhal 2011]

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

- To minimize $f$ efficiently, estimate the metric $H$!

**minimize** $g(X) := b^T\mathbf{x} + \frac{1}{2}\|X - H\|_F^2$ **with Random Pursuit**

$$\mathbf{u}_k \sim \mathcal{N}(\mathbf{0}, I_n) \qquad \text{search direction : } U_k = \mathbf{u}_k\mathbf{u}_k^T$$

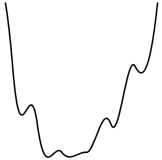
- Evaluate $g$ on 3 points on a line given by $U_k$:

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

**observe:** $\langle H, U_k\rangle \approx \frac{f(\mathbf{x} + \epsilon\mathbf{u}_k) - 2f(\mathbf{x}) + f(\mathbf{x} - \epsilon\mathbf{u}_k)}{\epsilon^2}$

# Outlook and Open Problems

- many more examples . . .
- extension to accelerated (random) schemes
- "nice" non-convex functions ?
- schemes with adaptive step size control
  - sum of correlated random variables
- constraints



Thank you

# References


- S.U. Stich. Convergence of Local Search, *Manuscript*, 2012.
- V.G. Karmanov. On convergence of a random search method in convex minimization problems, 1974.
- D. Leventhal and A.S. Lewis, Randomized Hessian estimation and directional search, 2011.
- Y. Nesterov, Random Gradient-Free Minimization of Convex Functions, 2011.
- S.U. Stich, C.L. Müller and B. Gärtner, Optimization of convex functions with Random Pursuit, 2012.
- S.U. Stich, C.L. Müller and B. Gärtner, Variable Metric Random Pursuit, *in preparation*, 2012.


**Acknowledgements** Christian Müller and Bernd Gärtner, ETH Zürich

Project CG Learning, FET-Open grant number: 255827