ThRaSH Workshop 2012

Convergence of Local Search

Sebastian U. Stich[a,b]

joint work with
Christian L. Müller[a,b], Bernd Gärtner[a]

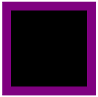
Institute of Theoretical Computer Science
Department of Computer Science[a], Swiss Institute of Bioinformatics[b]
ETH Zürich

Mai 3, 2012

Table of contents

Black-box optimization

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

$\mathbf{x} \rightarrow$  $\rightarrow \left\{ f(\mathbf{x}) \right.$

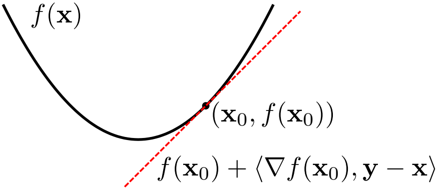
- **Problem class** f convex

- **Oracle** access to $(f(\mathbf{x}))$
- **Complexity:** number of oracle calls sufficient to solve any problem of the class
- **Solution:** $\mathbf{y} : f(\mathbf{y}) - \min_{\mathbf{x} \in E} f(\mathbf{x}) \leq \epsilon$

Convex functions

first-order condition:

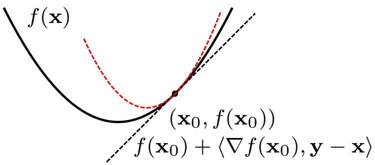
$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle , \quad \forall \mathbf{x}, \mathbf{y} \in E$$



Convex functions II

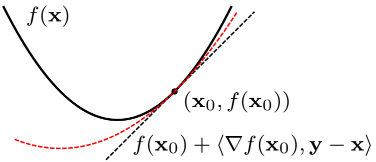
Quadratic upper bound:

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



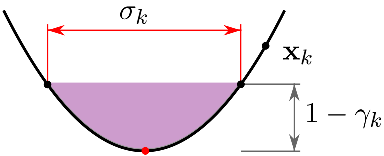
Quadratic lower bound: (strongly convex)

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



- We call $\kappa := L/\mu$ condition number; $\mu \cdot I_n \preceq \nabla^2 f(\mathbf{x}) \preceq L \cdot I_n$
- Only (!) for strongly convex: $\|\mathbf{x} - \mathbf{x}^*\|^2 \leq \frac{2}{\mu}(f(\mathbf{x}) - f(\mathbf{x}^*))$

Local search



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

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

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

Sufficient decrease

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

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

$$\Rightarrow$$

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

Set $t = -\underbrace{\left\langle \frac{\nabla f(\mathbf{x}_k)}{\|\nabla f(\mathbf{x}_k)\|_2}, \mathbf{u}_k \right\rangle}_{:=\beta_k} \cdot \frac{\|\nabla f(\mathbf{x}_k)\|_2}{L}$

Single step progress

We use this t together with our assumptions.

Quadratic upper bound:

$$f(\mathbf{x}_k) - f(\mathbf{x}_{k+1}) \geq \frac{\gamma}{2}\beta_k^2 \left\|\nabla f(\mathbf{x}_k)\right\|_2^2$$

Quadratic lower bound:

$$f(\mathbf{x}_k) - f(\mathbf{x}_{k+1}) \geq \frac{\gamma\mu}{L}\beta_k^2 \left(f(\mathbf{x}_k) - f(\mathbf{x}^*)\right)$$

Progress:

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

Global convergence

After N steps:

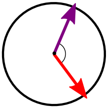
$$\begin{aligned}\ln f_N - \ln f_0 &\leq \sum_{k=0}^{N-1} \ln f_{k+1} - \ln f_k \\ &\leq \sum_{k=0}^{N-1} \ln\left(1 - \beta_k^2 \frac{\gamma}{\kappa}\right) \\ &\leq -\sum_{k=0}^{N-1} \beta_k^2 \frac{\gamma}{\kappa}\end{aligned}$$

$$f_N \leq f_0 \cdot \exp\left(-\frac{\gamma}{\kappa} \sum_{k=0}^{N-1} \beta_k^2\right)$$

Convergence with high probability

$$\beta^2 = \langle \mathbf{v}, \mathbf{u} \rangle^2$$

$$\mathbf{v} \in S^{n-1}, \qquad \mathbf{u} \sim S^{n-1}$$



$$\mathbb{E}\left[\beta^2\right] = \mathbb{E}\left[\langle \mathbf{v}, \mathbf{u} \rangle^2\right] = \frac{1}{n} \qquad \mathrm{Var}\left[\beta^2\right] = \mathrm{Var}\left[\langle \mathbf{v}, \mathbf{u} \rangle^2\right] \leq \frac{2}{n^2}$$

$$\mathbb{P}\left[\sum_{k=1}^{N-1} \langle \mathbf{v}, \mathbf{u}_k \rangle^2 < (1-\epsilon)\frac{N}{n}\right] \leq \frac{\mathrm{Var}\left[\beta^2\right]}{N} \cdot \frac{1}{\epsilon^2\, \mathbb{E}\left[\beta^2\right]^2} = \frac{2}{\epsilon^2 N}$$

Convergence with high probability

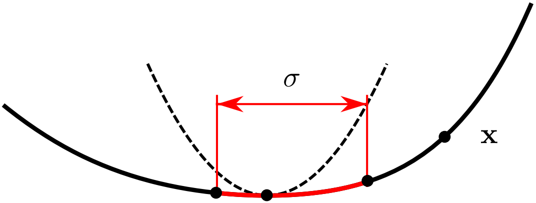
For $N = \Omega(n)$:

$$f_N \leq f_0 \cdot \exp\left(-\frac{\gamma}{\kappa}\sum_{k=0}^{N-1}\beta_k^2\right) \leq f_0 \cdot \exp\left(-\frac{\gamma}{\kappa}\cdot\frac{(1-\epsilon)N}{n}\right) \qquad \text{w.h.p.}$$

Example I - Random Pursuit

$\mathbf{u} \sim S^{n-1}$ + approximate line search:

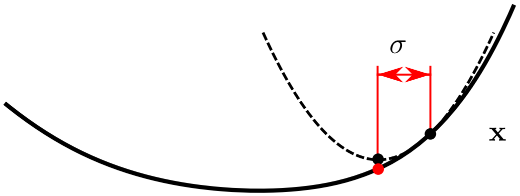
$$\sigma_k \in \left[0.5 \cdot \arg\min_{h \in \mathbb{R}} f(\mathbf{x}_k + h\mathbf{u}) \, , \arg\min_{h \in \mathbb{R}} f(\mathbf{x}_k + h\mathbf{u}) + \delta \right]$$



Example II - Random Gradient Method [Nesterov 2011]

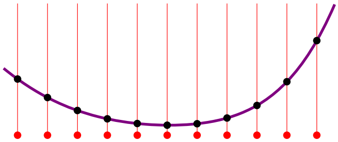
$\mathbf{u} \sim S^{n-1}$ $+$ estimated stepsize:

$$\sigma_k \approx -\frac{1}{L} f'(\mathbf{u}, \mathbf{x}_k)$$



Example III & IV Different spaces

Discrete space



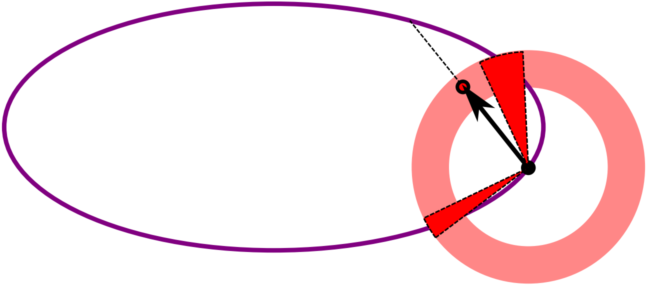
Matrices

$$f\colon \mathbb{R}^{n\times n} \to \mathbb{R}$$

Example V (?) - Optimal 1/5 rule

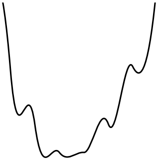
$\mathbf{u} \sim \mathcal{N}(0, I_n)$ + '1/5'- rule:

$$\mathbb{P}\left[f(\mathbf{x}_k + \sigma_k \mathbf{u}) \leq f(\mathbf{x}_k)\right] = \text{const}$$



Outlook and Open Problems

- Possible:
  - Concentration for $N = \Omega(\log n)$
  - Different search directions (not only S^{n-1})
  - Non-isotropic sampling
  - Interesting spaces (e.g. matrices)
  - Smooth convex functions
- Would be very nice:
  - Constraint handling
  - Apply to 1/5-rule stepsize rule [like (1+1)-ES]
- Open:
  - Extension of the model to *"almost convex functions"*



Thank you

References

- S.U. Stich, C.L. Müller, B. Gärtner. Optimization of convex functions with Random Pursuit 2011.
- S.U. Stich. Convergence of Local Search, *Manuscript* 2012.