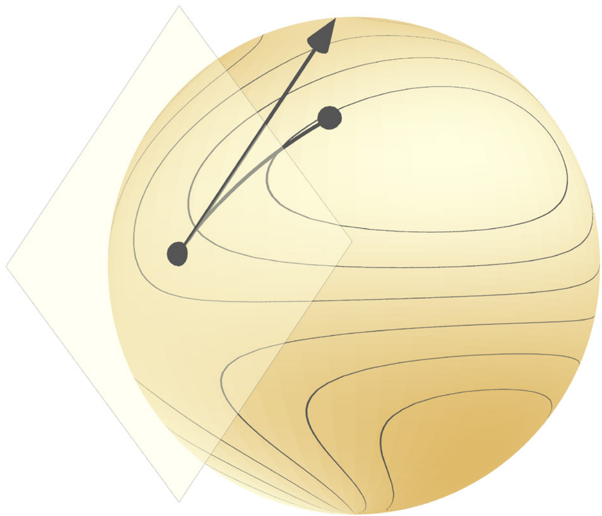
204

Choosing a step size



Spring 2023

Optimization on manifolds, MATH 512 @ EPFL

Instructor: Nicolas Boumal

Decrease without knowing L?

We make a Lipschitz-like assumption about f and R together:

A2 $\quad f\big(R_x(s)\big) \leq f(x) + \langle \mathrm{grad} f(x), s \rangle_x + \frac{L}{2} \|s\|_x^2$ for all $(x, s) \in \mathrm{T}\mathcal{M}$.

With $x_{k+1} = R_{x_k}\big(-\alpha \, \mathrm{grad} f(x_k)\big)$, this implies:

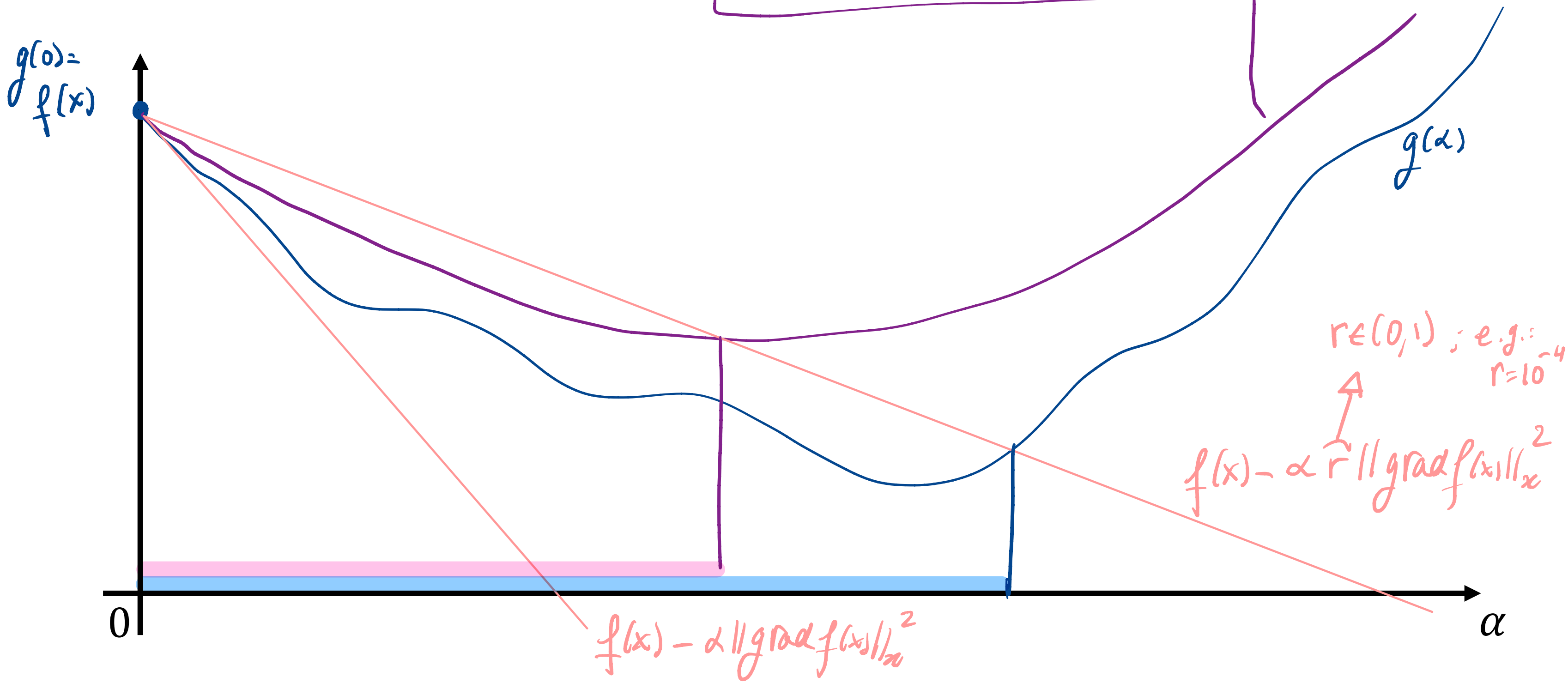
A3 $\quad$ Sufficient decrease: $f(x_k) - f(x_{k+1}) \geq c \|\mathrm{grad} f(x_k)\|_{x_k}^2 \quad \forall k$

where $c = \alpha\left(1 - \frac{\alpha L}{2}\right)$. $\quad$ Positive if $\alpha \in (0, 2/L)$. But rarely know L...

$g'(0) = Df(x)[-\mathrm{grad}f(x)] = -\|\mathrm{grad}f(x)\|_x^2$

Backtracking line-search

$$g(\alpha) \stackrel{\text{def}}{=} f\left(R_x\left(-\alpha\,\mathrm{grad}f(x)\right)\right) \overset{\text{A2}}{\leq} \boxed{f(x) - \alpha\|\mathrm{grad}f(x)\|_x^2 + \frac{\alpha^2 L}{2}\|\mathrm{grad}f(x)\|_x^2}$$



I accept α if $g(\alpha) \leq f(x) - \alpha r \|\text{grad} f(x)\|_x^2 \equiv f(x) - f(R_x(-\alpha \text{grad} f(x)))$
$\geq \alpha r \|\text{grad} f(x)\|_x^2$

Armijo backtracking for gradient descent

Parameters: $\tau \in (0,1)$, $r \in (0,1)$ (for example, $\tau = \frac{1}{2}$ and $r = 10^{-4}$)

Input: $x \in \mathcal{M}, \bar{\alpha} > 0$

Algorithm: Let $\alpha \leftarrow \bar{\alpha}$

While $f(x) - f\left(R_x(-\alpha \text{grad} f(x))\right) < r\alpha \|\text{grad} f(x)\|_x^2$

$\alpha \leftarrow \tau\alpha$

Output α

Theorem: If **A2** holds, then **A3** holds with $c = r \min\left(\bar{\alpha}, \frac{2\tau(1-r)}{L}\right)$.

"While" loops at most $\max\left(1, 2 + \log_{\tau^{-1}}\left(\frac{\bar{\alpha}L}{2(1-r)}\right)\right)$ times.