1

Ch. 1 Introduction

- repetition of Multidim. Analysis
- examples

Ch 2 Line Search

- optimization along the line
- often : simple heuristics

We need to understand :

* Dichotomous search
* Golden section search
* Bisection search
* Newton's method
* Armijo's rule

self-study

2

Ch 3 Multidimensional search

The problem : $\min f(x)$, $x \in S \subset \mathbb{R}^n$

- Very hard! In general, only numerical methods.
- Common idea: to walk in S

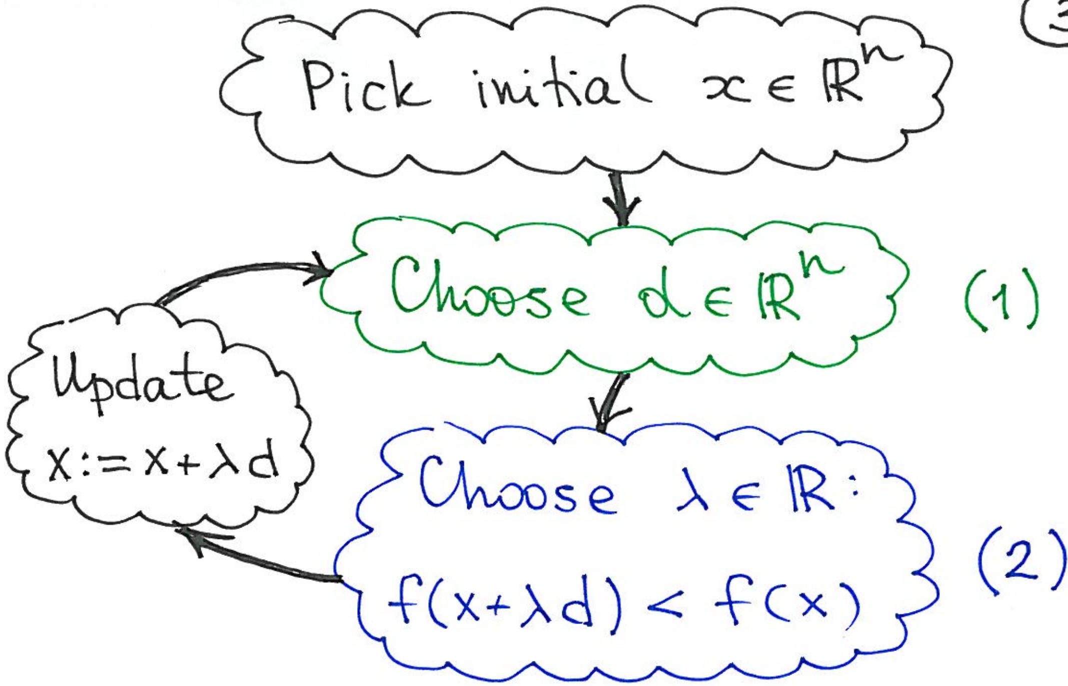
$x_1 \longmapsto x_2 \longmapsto x_3 \longmapsto x_4 \longmapsto \ldots$

trying to get $x_n \longrightarrow x_{min}$

- For now: let $S = \mathbb{R}^n$
- For a fixed $x \in \mathbb{R}^n$ and a fixed direction $d \in \mathbb{R}^n$ define

$$\varphi(\lambda) = f(x + \lambda d) \quad , \quad \lambda \in \mathbb{R}$$

$\varphi : \mathbb{R} \to \mathbb{R}$, we can use line search



(2): line search for $\varphi(\lambda) = f(x+\lambda d)$

- in teory: assume exact line search

  i.e. λ solves $\min\limits_{\lambda} f(x+\lambda d)$

- in practice: inexact line search,

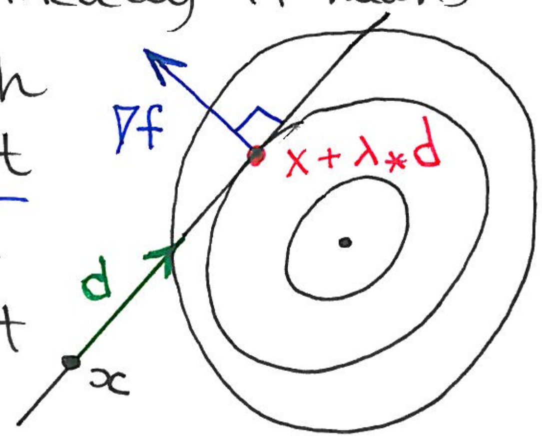
  e.g. Armijo's rule etc

Lemma 1: Assume $f \in C^1$, $x \in \mathrm{dom}(f)$, $d \neq 0$ are fixed.

Then λ_* solves $\min\limits_{\lambda} f(x+\lambda d) \Rightarrow$

$\Rightarrow \nabla f(x+\lambda_* d) \perp d$.

Remark: geometrically it means that the search line is tangent to the level set at the new point $x+\lambda_* d$.



Proof: define $\varphi(\lambda) = f(x+\lambda d)$.

λ_* minimizes $\varphi \Rightarrow \varphi'(\lambda_*) = 0$.

But $\varphi'(\lambda) = \nabla f(x+\lambda d)^T d \Rightarrow$

$\Rightarrow \nabla f(x+\lambda_* d)^T d = 0$. ∎

(1): the most critical part.

Different methods ↔ different ways to choose the next d.

Cheap & Bad vs Good & expensive

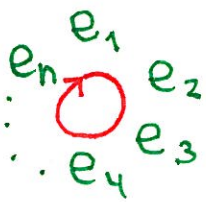
3.2. Cyclic coordinates

$$f : \mathbb{R}^n \to \mathbb{R}$$

Pick some basis of $\mathbb{R}^n$:

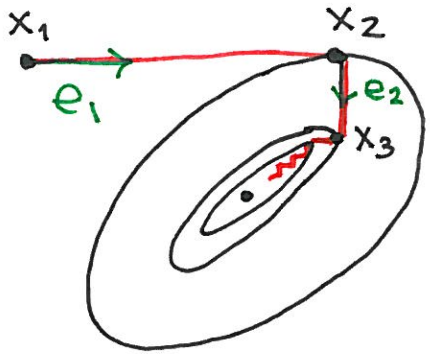
$\{e_1, e_2, e_3, \dots, e_n\}$, e.g. $e_i = \begin{array}{c}\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}\end{array}\begin{array}{l} \\ \\ \\ - i \\ \\ \\ \\ \end{array}$

and iterate



"$d = e_k$" in a loop.

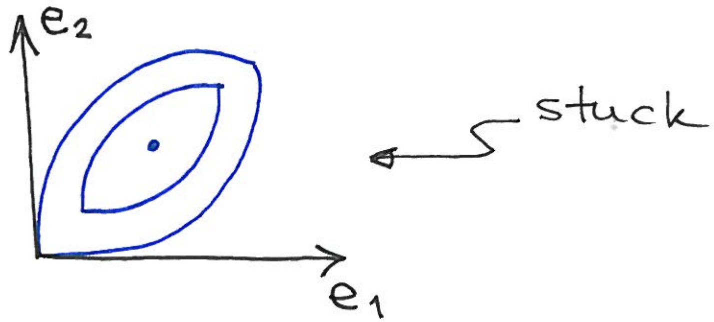
Ex. $\{e_1, e_2\}$



⊕ – easy to implement
– no need to know derivatives
– converges to a stationary point for $f \in C^1$.

⊖ – very slow (zigzagging)
– may stuck for $f \notin C^1$

Ex

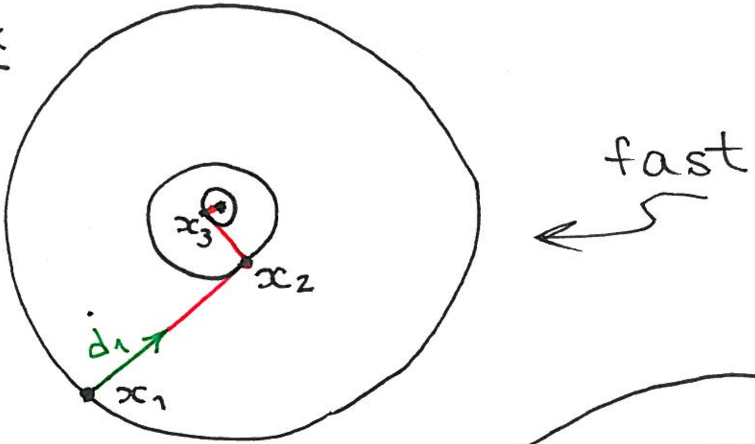


3.3. Steepest Descent

FlerDim: $\nabla f(x)$ is the direction of the steepest ascent.

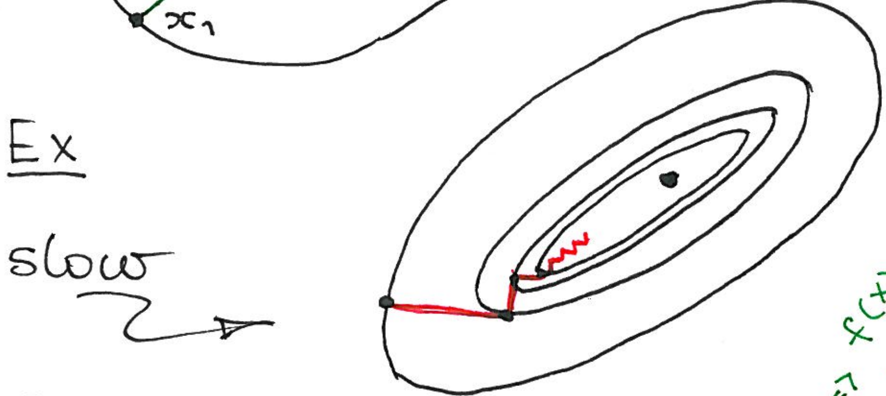
SD method: $d_k = -\nabla f(x_k)$

Ex



Ex

slow



$f(x) > 0 \quad \forall x \neq 0 \iff$

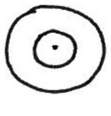
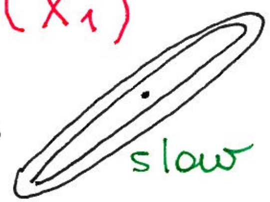
Th $f(x) = \frac{1}{2} x^T H x$, H - pos. def.

$\alpha = \dfrac{\lambda_{max}(H)}{\lambda_{min}(H)}$ (condition number)

$\alpha \geq 1$

$C_\alpha = \left(\dfrac{\alpha - 1}{\alpha + 1}\right)^2$ $\quad (0 \leq C_\alpha < 1)$

Then $f(x_{k+1}) \leq C_\alpha^k \cdot f(x_1)$

Remark: $\alpha \approx 1$  , $\alpha \approx \infty$ 
fast

3.4. Newton's method

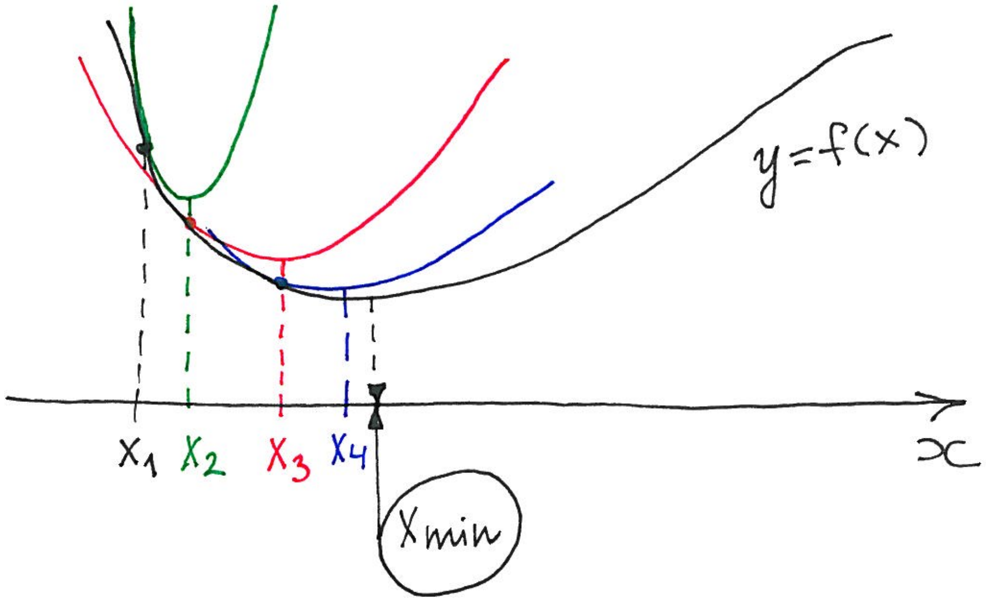
Ex $f(x) = ax^2 + bx + c$, $a > 0$

To minimize $f(x)$: $f'(x) = 0$.

$f'(x) = 2ax + b = 0 \iff x = -\dfrac{b}{2a}$.

If $f(x)$ is not quadratic?

Let's approximate it by a quadratic.



We hope that $x_k \to x_{min}$.

FlerDim: Taylor formula at x_1

$$f(x) = f(x_1) + \nabla f(x_1)^T (x - x_1) + \frac{1}{2}(x-x_1)^T H(x_1)(x-x_1) + \ldots = p_2(x) + \ldots$$

We minimize p_2 instead:

$\nabla p_2(x) = 0 \iff$ [use Ex. 1.2e, 1.5a]

$$\iff \nabla f(x_1) + H(x_1)(x - x_1) = 0 \iff$$

$$\iff x = x_1 - H(x_1)^{-1} \nabla f(x_1).$$

Call $x_2 = x$ and repeat.

N $$x_{k+1} = x_k - H(x_k)^{-1} \nabla f(x_k)$$

Remark:
- $d_k = -H(x_k)^{-1} \nabla f(x_k)$
- $\lambda_k = 1$ (unit step)

⊕ • if converges then the convergence is very fast
- converges locally for $f \in C^2$ under some mild assumptions.

⊖ • $H(x_k)$ may be singular.
- may diverge
- may converge to a maximum

A possible remedy:

1) modify d_k.
2) add line search.

Idea: replace the "old" d_k with

$$d_k = -H_{\varepsilon_k}^{-1} \nabla f(x_k)$$

where $H_{\varepsilon_k} \approx H(x_k)$,

H_{ε_k} invertible

d_k – descent direction.