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

Ch. 3 Multidimensional search

\[
\min f(x) \quad x \in S \subseteq \mathbb{R}^n
\]

- Very hard! In general, only numerical methods.
- Common idea: to walk in \( S \)
  \( x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow x_4 \rightarrow \ldots \)
  trying to get \( x_n \rightarrow x_{\text{min}} \)
- 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
  \[
  \psi(\lambda) = f(x + \lambda d), \quad \lambda \in \mathbb{R}
  \]
  \( \psi : \mathbb{R} \rightarrow \mathbb{R} \), we can use line search
Pick initial $x \in \mathbb{R}^n$.

Choose $d \in \mathbb{R}^n$. (1)

$\text{Update } x := x + \lambda d$

Choose $\lambda \in \mathbb{R}$: $f(x + \lambda d) < f(x)$. (2)

Lemma 1: Assume $f \in C^1$, $x \in \text{dom}(f)$, $d \neq 0$ are fixed. Then $\lambda_x$ solves $\min_{\lambda} f(x + \lambda d) \Rightarrow \nabla f(x + \lambda_x d) \perp d$.

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

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

$\lambda_x$ minimizes $\varphi \Rightarrow \varphi'(\lambda_x) = 0$.

But $\varphi'(\lambda) = \nabla f(x + \lambda d)^T d \Rightarrow \nabla f(x + \lambda_x d)^T d = 0$. 

- (2): Line search for $\varphi(\lambda) = f(x + \lambda d)$
  - In theory: assume exact line search
    - i.e. $\lambda_x$ solves $\min_{\lambda} f(x + \lambda d)$
  - In practice: inexact line search, e.g. Armijo's rule etc
(1): the most critical part.
Different methods ↔ different ways to choose the next $d$.
Cheap & bad vs Good & expensive

3.2. Cyclic coordinates

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

Pick some basis of $\mathbb{R}^n$: $\{e_1, e_2, e_3, \ldots, e_n\}$, e.g. $e_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \cdots 0 \end{bmatrix}$ and iterate

"$d = e_k$" in a loop.

Ex. $\{e_1, e_2\}$

3.3. Steepest Descent

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

SD method: $d_k = -\nabla f(x_k)$
3.4. Newton's method

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

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

\[
f'(x) = 2ax + b = 0 \iff x = -\frac{b}{2a}.
\]

If \( f(x) \) is not quadratic?

Let's approximate it by a quadratic.

We hope that \( x_k \to x_{\text{min}} \).
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:

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

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

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

Call $x_2 = x$ and repeat.

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

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

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

$H_{x_k}$ invertible

$d_k$ - descent direction.