

Numerical Algorithms (MU4IN910)

Lecture 3: Introduction to optimization (1/2)

Stef Graillat

Sorbonne Université



Summary of the previous lecture

- Matrix decompositions and their uses
 - ① LU
 - ② QR
 - ③ Eigendecomposition (diagonalization, etc.)
 - ④ SVD (singular value decomposition)
- Software

- Existence and uniqueness of extrema
- Optimality conditions for extrema
- Algorithms for dimension 1

Optimization problems arise in all areas of science and engineering. We can cite for example:

- finance
- economy
- optimal control
- meteorology
- image processing
- power generation management
- molecular biology
- etc.

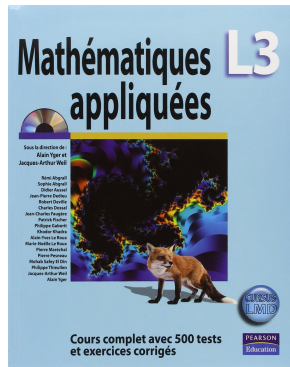
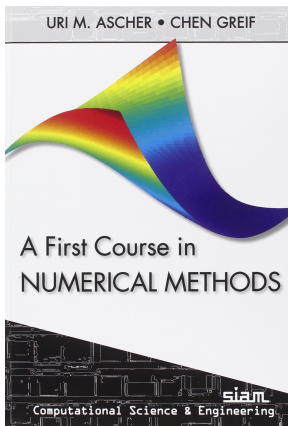
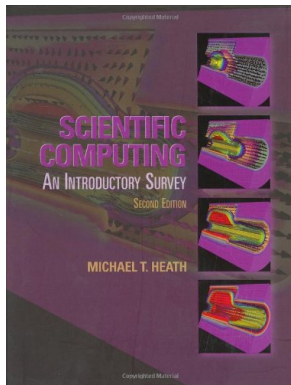
References

- **Scientific Computing, An Introductory Survey, Michael .T. Heath, Revised Second Edition, SIAM, 2018** (the lecture is mainly based on this book and on the associated slides)
- A First Course in Numerical Methods, Uri M. Ascher, Chen Greif, SIAM, 2011
- Mathématiques appliquées L3, sous la direction de Jacques-Arthur Weil et Alain Yger, Pearson, 2009
- Scientific Computing with Case Studies, Dianne P. O’Leary, SIAM, 2009
- Numerical Optimization, Jorgue Nocedal, Stephen Wright, Springer, 2006
- Introduction à l’Analyse Numérique Matricielle et à l’Optimisation, Philippe G. Ciarlet, Dunod, 1982
- Analyse numérique et optimisation, Grégoire Allaire, Éditions de l’École polytechnique, 2005
- L’optimisation, J.B. Hiriart-Urruty, PUF, 1996
- Convex Optimization, S. Boyd et L. Vandenberghe, Cambridge University Press, 2004

References (cont'd)

- Numerical optimization: Theoretical and practical aspects, F. Bonnans, J.C. Gilbert, C. Lemaréchal et C. Sagastizábal, Springer-Verlag, 2006
- Optimisation continue, F. Bonnans, Dunod, 2006
- Optimisation et contrôle des systèmes linéaires, M. Bergounioux, Dunod, 2001.
- Convex Analysis and Nonlinear Optimization, J.M. Borwein et A.S. Lewis, Springer, 2000
- Lectures on Modern Convex Optimization - Analysis, Algorithms and Engineering Applications, A. Ben-Tal, A. Nemirovski, SIAM, 2001
- Numerical Computing with MATLAB, Cleve Moler, SIAM, 2004
- Linear and Nonlinear Optimization, Igor Griva, Stephen G. Nash, Ariela Sofer, SIAM, 2009
- Practical Methods of Optimization, Roger Fletcher, Wiley, 1987
- A Mathematical View of Interior-Point Methods in Convex Optimization, James Renegar, SIAM, 2001
- Numerical Recipes. The Art of Scientific Computing, William Press, Saul Teukolsky, William Vetterling, Brian Flannery, 3rd Edition, Cambridge University Press, 2007

References (cont'd)



Outline of the lecture

- Basics of optimization
- Optimization in dimension 1
- Optimization in dimension $n \geq 2$

Basics of optimization

Optimization

- Given function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and set $S \subset \mathbb{R}^n$, find $x^* \in S$ such that $f(x^*) \leq f(x)$ for all $x \in S$
- x^* is called **minimizer** or **minimum** of f on S
- It suffices to consider only minimization, since maximum of f is minimum $-f$
- **Objective function** f is usually differentiable, and may be linear or nonlinear
- **Constraint** set S is defined by system of equations and inequalities, which may be linear or nonlinear
- Point $x \in S$ are called **feasible** points
- If $S = \mathbb{R}^n$, problem is **unconstrained**

Optimization problems

- General continuous optimization problem:

$$\min f(x) \quad \text{subject to} \quad g(x) = 0 \text{ and } h(x) \leq 0.$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$

- **Linear programming:** f , g and h are all linear
- **Nonlinear programming :** at least one of f , g et h is nonlinear

Examples of optimization problems

- Minimize weight of structure subject to constraint on its strength, or maximize its strength subject to constraint on its weight
- Minimize surface area of cylinder subject to constraint on its volume:

$$\begin{aligned} \min_{x_1, x_2} f(x_1, x_2) &= 2\pi x_1(x_1 + x_2) \\ \text{subject to } g(x_1, x_2) &= \pi x_1^2 x_2 - V = 0 \end{aligned}$$

where x_1 and x_2 are radius and height of cylinder, and V is required volume

Local vs global optimization

- $x^* \in S$ is a **global minimum** if $f(x^*) \leq f(x)$ for all $x \in S$
- $x^* \in S$ is a **local minimum** if $f(x^*) \leq f(x)$ for all feasible x in some neighborhood of x^*



- Finding, or even verifying, global minimum is difficult, in general
- Most optimization methods are designed to find local minimum, which may or may not be global minimum
- If global minimum is desired, one can try several widely separated starting points and see if all produce same result
- For some problems, such as linear programming, global optimization is more tractable

Existence of minimum

- If f is continuous on closed and bounded set $S \subset \mathbb{R}^n$ then f has a global minimum on S
- If S is not closed or is unbounded, then f may have no local or global minimum on S
- Continuous function f on an unbounded set S is **coercive** if

$$\lim_{\|x\| \rightarrow +\infty} f(x) = +\infty,$$

i.e. $f(x)$ must be large whenever $\|x\|$ is large

- If f is coercive on closed, unbounded set $S \subset \mathbb{R}^n$ then f has global minimum on S

- A **level set** for a function $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$ is the set of all points in S for which f has some given constant value
- Given $\gamma \in \mathbb{R}$, the **sublevel set** is

$$L_\gamma = \{x \in S : f(x) \leq \gamma\}$$

- If a continuous function f on $S \subset \mathbb{R}^n$ has nonempty sublevel set that is closed and bounded, then f has global minimum on S
- If S is unbounded, then f is coercive on S iff all of its sublevel sets are bounded

Uniqueness of minimum

- A set S is **convex** if it contains line segment between any two of its points

$$\forall x, y \in S, \forall \alpha \in [0;1], \quad \alpha x + (1 - \alpha)y \in S$$

- A function $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$ is **convex** on a convex set S if

$$\forall x, y \in S, \forall \alpha \in [0;1], \quad f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$$

- A function $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$ is **strictly convex** on a convex set S if

$$\forall x, y \in S, \forall \alpha \in]0;1[, \quad f(\alpha x + (1 - \alpha)y) < \alpha f(x) + (1 - \alpha)f(y)$$

- Any local minimum of convex function f on convex set $S \subset \mathbb{R}^n$ is global minimum f on S
- Any local minimum of a **strictly convex** function f on convex set $S \subset \mathbb{R}^n$ is **unique global minimum** of f on S

First-order optimality condition

- For function of one variable, we can find extremum by differentiating function and setting derivative to zero
- Generalization to function of n variables is to find **critical point**, *i.e.*, solution of nonlinear system

$$\nabla f(x) = 0$$

where $\nabla f(x)$ is gradient vector of f whose i th component is $\partial f(x)/\partial x_i$

- For continuously differentiable $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$, any interior point x^* of S at which f has local minimum must be a critical point of f
- But not all critical points are minima: they can also be maxima or saddle points

Second-order optimality condition

- For twice continuously differentiable $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$ we can distinguish among critical points by considering **Hessian matrix** $H_f(x)$ defined by

$$(H_f(x))_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

which is symmetric. We also note $H_f(x) = \nabla^2 f(x)$.

- At critical point x^* , if $H_f(x^*)$ is
 - positive definite, then x^* is a minimum of f
 - negative definite, then x^* is a maximum of f
 - indefinite, then x^* is saddle point of f
 - singular, then various pathological situations are possible

Constrained optimality

- If problem is constrained, only **feasible** directions are relevant
- For equality-constrained problem

$$\min f(x) \text{ subject to } g(x) = 0$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$, with $m \leq n$, necessary condition for feasible point x^* to be solution is that negative gradient of f lie in space spanned by constraint normals

$$-\nabla f(x^*) = J_g(x^*)^T \lambda$$

where J_g is the Jacobian matrix of g and λ is the vector of **Lagrange multipliers**

- This condition says we cannot reduce objective function without violating constraints

Constrained optimality (cont'd)

- The **Lagrangian function** $\mathcal{L} : \mathbb{R}^{n+m} \rightarrow \mathbb{R}$ is defined by

$$\mathcal{L}(x, \lambda) = f(x) + \lambda^T g(x)$$

- Its gradient is given by

$$\nabla \mathcal{L}(x, \lambda) = \begin{pmatrix} \nabla f(x) + J_g^T(x) \lambda \\ g(x) \end{pmatrix}$$

- Its Hessian is given by

$$H_{\mathcal{L}}(x, \lambda) = \begin{pmatrix} B(x, \lambda) & J_g^T(x) \\ J_g(x) & 0 \end{pmatrix}$$

where

$$B(x, \lambda) = H_f(x) + \sum_{i=1}^m \lambda_i H_{g_i}(x)$$

Constrained optimality (cont'd)

- Together, necessary condition and feasibility imply critical point of Lagrangian function

$$\nabla \mathcal{L}(x, \lambda) = \begin{pmatrix} \nabla f(x) + J_g^T(x)\lambda \\ g(x) \end{pmatrix} = 0$$

- Hessian of Lagrangian is symmetric, but not positive definite, so critical point of \mathcal{L} is saddle point rather than minimum or maximum
- A critical point (x^*, λ^*) of \mathcal{L} is a constrained minimum of f if $B(x^*, \lambda^*)$ is positive definite on null space $J_g(x^*)$
- If columns of Z form basis for null space of $J_g(x^*)$, then test projected Hessian $Z^T B Z$ for positive definiteness

- Taylor series expansion of f in the neighborhood of x^*

$$f(\widehat{x}) = f(x^* + h) = f(x^*) + f'(x^*)h + \frac{1}{2}f''(x^*)h^2 + \mathcal{O}(h^3)$$

Since $f'(x^*) = 0$, if $|f(\widehat{x}) - f(x^*)| \leq \varepsilon$ then $|\widehat{x} - x^*|$ is of the order of $\sqrt{2\varepsilon/|f''(x^*)|}$

- Consequently, based only on function evaluation, a minimum can only be calculated about half precision

One-dimensional optimization

- For minimizing function of one variable, we need “bracket” for solution analogous to sign change for nonlinear equation
- Real-valued function f is **unimodal** on interval $[a, b]$ if there is unique $x^* \in [a, b]$ such that $f(x^*)$ is a minimum of f on $[a, b]$, f is strictly decreasing for $x \leq x^*$ and is strictly increasing for $x^* \leq x$
- Unimodality enables discarding portions of interval based on sample function values, analogous to interval bisection

Golden section search

- Suppose f is unimodal on $[a, b]$, and let x_1 and x_2 be two points within $[a, b]$, with $x_1 < x_2$
- Evaluating and comparing $f(x_1)$ and $f(x_2)$, we can discard either $[a, x_1)$ or $(x_2, b]$, with minimum known to lie in remaining subinterval
- To repeat process, we need compute only one new function evaluation
- To reduce length of interval by fixed fraction at each iteration, each new pair of points must have same relationship with respect to new interval that previous pair had with respect to previous interval

Golden section search (cont'd)

- To accomplish this, we choose relative positions of two points as τ and $1 - \tau$ where $\tau^2 = 1 - \tau$. So $\tau = (\sqrt{5} - 1)/2 \approx 0.618$ and $1 - \tau \approx 0.382$
- Whichever subinterval is retained, its length will be τ relative to previous interval, and interior point retained will be at position either τ or $1 - \tau$ relative to new interval
- To continue iteration, we need to compute only one new function value, at complementary point
- Golden section search is safe but convergence rate is only linear

Golden section search (cont'd)

$$\tau = (\sqrt{5} - 1)/2$$

$$x_1 = a + (1 - \tau)(b - a); \quad f_1 = f(x_1)$$

$$x_2 = a + \tau(b - a); \quad f_2 = f(x_2)$$

while $((b - a) > \text{tol})$ do

 if $(f_1 > f_2)$ then

$$a = x_1$$

$$x_1 = x_2$$

$$f_1 = f_2$$

$$x_2 = a + \tau(b - a)$$

$$f_2 = f(x_2)$$

else

$$b = x_2$$

$$x_2 = x_1$$

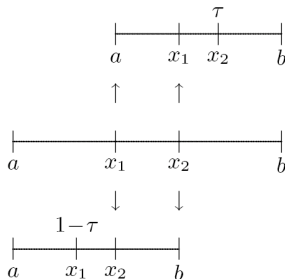
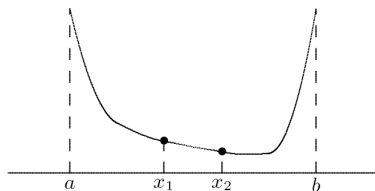
$$f_2 = f_1$$

$$x_1 = a + (1 - \tau)(b - a)$$

$$f_1 = f(x_1)$$

end

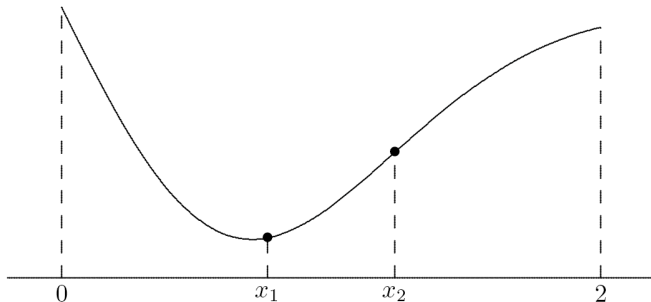
end



Golden section search: example

Use golden section search to minimize

$$f(x) = 0.5 - xe^{-x^2}$$

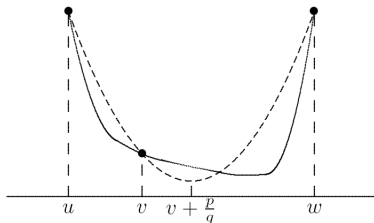


Golden section search: example (cont'd)

x_1	f_1	x_2	f_2
0.764	0.074	1.236	0.232
0.472	0.122	0.764	0.074
0.764	0.074	0.944	0.113
0.656	0.074	0.764	0.074
0.584	0.085	0.652	0.074
0.652	0.075	0.695	0.071
0.695	0.071	0.721	0.071
0.679	0.072	0.695	0.071
0.695	0.071	0.705	0.071
0.705	0.071	0.711	0.071

Quadratic interpolation method

- The function is approximated by a quadratic function (polynomial of degree 2) in 3 values
- The quadratic function is minimized to obtain an approximation of the minimum

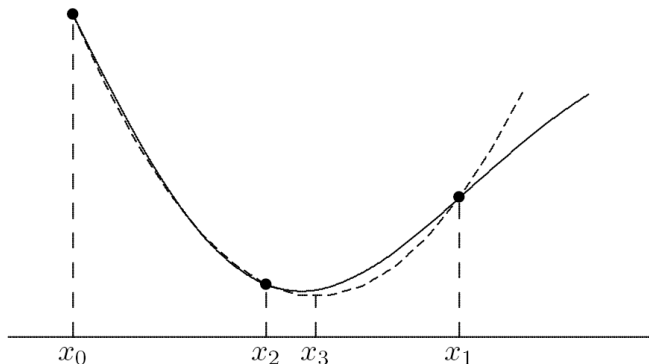


- We replace one of the three points by the new point and we iterate the process until convergence
- The convergence rate is superlinear

Quadratic interpolation method

Using the quadratic interpolation method to minimize

$$f(x) = 0.5 - xe^{-x^2}$$



Quadratic interpolation method

x_k	$f(x_k)$
0.000	0.500
0.600	0.081
1.200	0.216
0.754	0.073
0.721	0.071
0.692	0.071
0.707	0.071

- A local quadratic approximation is truncated Taylor series

$$f(x+h) \approx f(x) + f'(x)h + \frac{1}{2}f''(x)h^2$$

- By differentiation, minimum of this quadratic function of h is given by $h = -f'(x)/f''(x)$

- Iteration scheme

$$x_{k+1} = x_k - f'(x_k)/f''(x_k)$$

is **Newton's method** for solving nonlinear equation $f'(x) = 0$

- Newton's method for finding minimum normally has quadratic convergence rate, but must be started close enough to solution to converge

Newton's Method: example

- Use Newton's method to minimize $f(x) = 0.5 - xe^{-x^2}$
- First and second derivatives of f are given by

$$f'(x) = (2x^2 - 1)e^{-x^2} \quad \text{and} \quad f''(x) = 2x(3 - 2x^2)e^{-x^2}$$

- Newton iteration is given by

$$x_{k+1} = x_k - (2x_k^2 - 1)/(2x_k(3 - 2x_k^2))$$

- Using starting guess $x_0 = 1$, we obtain

x_k	$f(x_k)$
1.000	0.132
0.500	0.111
0.700	0.071
0.707	0.071

Conclusion

We presented

- generalities on the problems of optimization (existence, uniqueness, etc.)
- algorithms to solve optimization problems in dimension 1

In the next lecture, we will present

- algorithms to solve optimization problems in dimension $n \geq 2$