Lecture 3: Finite element methods for Nonlinear PDEs: Semismooth Newton Method

Pr. Ismail Merabet

Univ. of K-M-Ouargla

March 11, 2026

Contents

Introduction

In this lecture, we will see how to extend the classical Newton method to a wider class of non differentiable functions.

This will be done in the context of constraint optimisation problems.

In the first part, we reformulate an optimisation problem into finding the roots of a non smooth functions.

In the second part, we will see how to find the root of this function

Newton's method for smooth function

Theorem (Unconstrained optimization)

Let $f : \mathbb{R} \to \mathbb{R}$ be differentiable. Then a solution $x^* \in \mathbb{R}$ to the optimization problem

$$\min_{x \in \mathbb{R}} f(x)$$

satisfies the first order optimality condition

$$f'(x^*) = 0.$$

Algorithm (Newton's method for unconstrained optimization)

Let $f : \mathbb{R} \to \mathbb{R}$ be twice differentiable. To compute a solution of $f'(x) = 0$ using Newton's method, proceed as follows.
Let $x_0 \in \mathbb{R}$ be given. Then compute iteratively

$$x_{n+1} := x_n - \frac{f'(x_n)}{f''(x_n)}.$$

What happens if we impose box-constraints on the optimization problem?

$$\min_{a \leq x \leq b} f(x)$$

Here $f'(x) = 0$ is not a necessary condition! Consider the following example:

$$f(x) = x, \qquad a = -1, \qquad b = 1$$

The obvious solution here is $x^* = -1$, but

$$f'(x^*) = 1 \neq 0!$$

Theorem (Constrained optimization)

Let $f : \mathbb{R} \to \mathbb{R}$ be differentiable. Then a solution $x^* \in \mathbb{R}$ to the optimization problem

$$\min_{a \le x \le b} f(x)$$

with $a, b \in \mathbb{R}$ and $a < b$ satisfies the first order optimality condition

$$f'(x^*)(x - x^*) \ge 0 \qquad \forall x \in [a, b].$$

Proof.

Assume there is a $p \in [a, b]$ such that

$$f'(x^*)(p - x^*) < 0$$

yielding $|f'(x^*)| > 0$. Define

$$h := p - x^*.$$

By continuity there exists $\hat{\delta} > 0$ such that for all $0 < \delta < \hat{\delta}$ we have

$$|f'(x^* + \delta h)| > 0, \qquad \operatorname{sign}(f'(x^*)) = \operatorname{sign}(f'(x^* + \delta h)).$$

Taylor expansion gives with a $0 < \bar{\delta} < \hat{\delta}$

$$f(p) = f(x^* + h) = f(x^*) + f'(x^* + \bar{\delta} h)h < f(x^*).$$

This is a contradiction to the optimality of x^*.

Projection

Let us define the following function (Projection) for $a, b \in \mathbb{R}$ and $a < b$:

$$P_{[a,b]} : \mathbb{R} \to \mathbb{R}, \qquad x \mapsto \begin{cases} a & \text{if } x \leq a \\ x & \text{if } a < x < b, \\ b & \text{if } x \geq b. \end{cases}$$

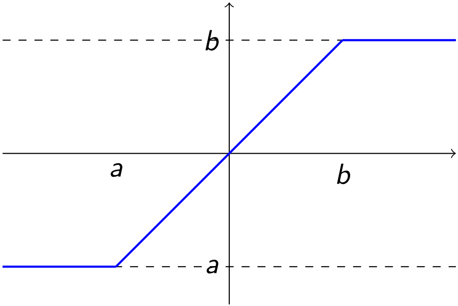

Figure: Plot of the projection function $P_{[a,b]}$.

Optimality conditions

Assume we want to minimize [1]

$$\min_{a \le x \le b} f(x) + \frac{1}{2}\alpha x^2 = g(x)$$

with some $\alpha > 0$.

The first order optimality conditions are

$$g'(x^*)(x - x^*) = \big(f'(x^*) + \alpha x^*\big)(x - x^*) \ge 0 \qquad \forall x \in [a, b].$$

[1] This is the Tikhonov regularization

Optimality conditions

Lemma (Nr. 1)

Assume that $x^* \in (a, b)$. Then we obtain

$$x^* = -\frac{1}{\alpha} f'(x^*).$$

Proof.

Assume that without loss of generality

$$f'(x^*) + \alpha x^* > 0.$$

Since $x^* \in (a, b)$ we find $\bar{x} \in (a, b)$ such that

$$\bar{x} - x^* < 0.$$

This gives

$$(f'(x^*) + \alpha x^*)(\bar{x} - x^*) < 0,$$

which is a contradiction to the first order optimality conditions.

The first order optimality conditions are

Lemma (Nr. 2)

Assume that $x^* = a$. Then we obtain

$$a = x^* \geq -\frac{1}{\alpha} f'(x^*).$$

Proof.

Since $a < b$ we find $a < \bar{x} < b$ and get

$$(f'(x^*) + \alpha x^*)(\bar{x} - a) \geq 0.$$

Since $\bar{x} - a > 0$, we obtain

$$f'(x^*) + \alpha x^* \geq 0.$$

Lemma (Nr. 3)

Assume that $x^* = b$. Then we obtain

$$b = x^* \leq -\frac{1}{\alpha} f'(x^*).$$

Proof.

Since $a < b$ we find $a < \bar{x} < b$ and get

$$(f'(x^*) + \alpha x^*)(\bar{x} - b) \geq 0.$$

Since $\bar{x} - b < 0$, we obtain

$$f'(x^*) + \alpha x^* \leq 0.$$

Lemma (Nr. 1)

Assume that $x^* \in (a, b)$. Then we obtain

$$x^* = -\frac{1}{\alpha} f'(x^*).$$

Lemma (Nr. 2)

Assume that $x^* = a$. Then we obtain

$$a = x^* \geq -\frac{1}{\alpha} f'(x^*).$$

Lemma (Nr. 3)

Assume that $x^* = b$. Then we obtain

$$b = x^* \leq -\frac{1}{\alpha} f'(x^*).$$

This shows that the first order optimality conditions can be rewritten to

$$x^* = P_{[a,b]}\left(-\frac{1}{\alpha} f'(x^*)\right).$$

Critical points of the constrained optimization problem

$$\min_{a \le x \le b} f(x) + \frac{1}{2}\alpha x^2$$

can be computed by solving the equation

$$x^* = P_{[a,b]}\left(-\frac{1}{\alpha} f'(x^*)\right).$$

Remark:

Note that this is not true if $\alpha = 0$!

Rewrite

$$x^* = P_{[a,b]}\left(-\frac{1}{\alpha} f'(x^*)\right)$$

to

$$F(x^*) = x^* - P_{[a,b]}\left(-\frac{1}{\alpha} f'(x^*)\right) = 0.$$

However, $P_{[a,b]}$ is not differentiable, hence Newton's method is not (directly) applicable to solve $F(x) = 0$!

Solution: Semismooth Newton's method!

Generalized gradient

Let consider the function

$$f(x) = \max(x, 0)$$

This function is non-differentiable, since at the point $x = 0$ we get two possible different values for the limit

$$\lim_{h\to 0^+} \frac{f(h) - f(0)}{h} = 1$$

$$\lim_{h\to 0^-} \frac{f(h) - f(0)}{h} = 0$$

The idea here is: Let's collect all possible limits in one set!

Generalized gradient

Define the set

$$D_f := \{x \in \mathbb{R} : f \text{ is differentiable at } x\}.$$

Definition

The set

$$\partial_B f(x) := \left\{ H \in \mathbb{R} \,\middle|\, \exists \{x^k\} \subset D_f : x^k \to x \text{ and } f'(x^k) \to H \right\}.$$

is called the Bouligand-subdifferential.
The convex hull[a]

$$\partial_C f(x) := \operatorname{conv}(\partial_B f(x))$$

is called the Clarke generalized gradient of f at x.

[a] The convex hull of set A is defined as the smallest convex set containing A.

Generalized gradient

Lets compute the generalized gradient for $f(x) = \max(x, 0)$. If $x > 0$ we directly get

$$\partial f(x) = \{1\}.$$

Similar for $x < 0$ we get

$$\partial f(x) = \{0\}.$$

For $x = 0$ we get

$$\partial_B f(0) = \{0, 1\}.$$

Computing the convex hull gives

$$\partial f(0) = \mathrm{conv}(\{0, 1\}) = [0, 1].$$

Generalized gradient

Finally we get for $f(x) = \max(x, 0)$:

$$\partial f(x) = \begin{cases} \{0\} & \text{if } x < 0, \\ [0,1] & \text{if } x = 0, \\ \{1\} & \text{if } x > 0. \end{cases}$$

Similar we can show:

$$\partial P_{[a,b]}(x) = \begin{cases} \{0\} & \text{if } x < a \ \vee \ x > b, \\ [0,1] & \text{if } x = a \ \vee \ x = b, \\ \{1\} & \text{if } a < x < b. \end{cases}$$

Generalized gradient

Remark:

- For the generalized gradient results resembling the chain- and product rule can be proved. (Note that the generalized gradient is a set!)
- Furthermore if f is locally Lipschitz the generalized gradient is a nonempty (convex and compact) set.

Semismooth functions

Definition

A locally Lipschitz continuous function f is called semismooth at x if

$$\sup_{H \in \partial f(x+d)} \|f(x+d) - f(x) - Hd\| = o(\|d\|) \qquad \text{for } d \to 0.$$

Note that the generalized gradient is evaluated at the point $x + d$ not at the point x, as in the classical definition (e.g. for Fréchet-differentiability).

It can be shown that $\max(\cdot, 0)$ and therefore $P_{[a,b]}$ are semismooth functions!

Semismooth Newton method

Algorithm (Semismooth Newton method)

1. Let $x_0 \in \mathbb{R}$ and set $k := 0$.
2. If a suitable stopping criterion is satisfied, stop the algorithm.
3. Choose an element $H_k \in \partial f(x_k)$ and compute
$$x_{k+1} := x_k - \frac{f(x_k)}{H_k}.$$
4. Set $k = k + 1$ and go to Step 2.

Question: For this formulation it is not needed that f is semismooth? Why is it called semismooth Newton method?

Semismooth Newton method

Theorem

Let $f : \mathbb{R} \to \mathbb{R}$ be semismooth and let x^* be a solution of $f(x) = 0$. Then there exists $\varepsilon > 0$ such that for all starting points x_0 with $|x^* - x_0| < \varepsilon$ we have the following statements:

- The sequence defined by the semismooth Newton method is well defined and x_k converges to x^*.
- The convergence rate is superlinear if all elements from $\partial_B f(x^*)$ are non-zero.

Details on the implementation

Recall: We want to minimize

$$\min_{a \le x \le b} \; f(x) + \frac{1}{2}\alpha x^2$$

resulting in finding the zeros of the semismooth function

$$0 = x - P_{[a,b]}\left(-\frac{1}{\alpha} f'(x)\right).$$

Details on the implementation

For the function

$$F(x) = x - P_{[a,b]}\left(-\frac{1}{\alpha} f'(x)\right)$$

we get a suitable derivative $F'(x) \in \partial F(x)$ with

$$F'(x) = 1 + \begin{cases} 0, & \text{if } x \leq a \ \vee \ x \geq b, \\ \frac{1}{\alpha} f''(x), & \text{if } a < x < b. \end{cases}$$

which allows us to easily implement the semismooth Newton method.

Details on the implementation

For the function

$$F(x) = x - P_{[a,b]}\left(-\frac{1}{\alpha} f'(x)\right)$$

we obtain a suitable generalized derivative $F'(x) \in \partial F(x)$ given by

$$F'(x) = 1 + \begin{cases} 0, & \text{if } x \leq a \;\vee\; x \geq b, \\ \dfrac{1}{\alpha} f''(x), & \text{if } a < x < b. \end{cases}$$

This expression allows us to easily implement the semismooth Newton method.

Recap

- We extended the classical definition of differentiability to a wider class of (non-differentiable) functions.
- This definition can be used to define the semismooth Newton method.
- Convergence and convergence speed can be shown.
- This method can be used in constrained optimization.
- Semismoothness can be extended to higher dimensions and even function spaces.

Theorem (Kantorovich (1948))

Then

1. The Newton sequence $(x_k)_{k=0}^{\infty}$ defined by

$$x_{k+1} = x_k - H'(x_k)^{-1} H(x_k)$$

converges to a root $x^\star \in \overline{B(x_0, r)}$ of H.

2. For each $k \geq 0$,

$$\|x^\star - x_k\|_X \leq \frac{r}{2^k}$$

3. The root $x^\star$ is the locally unique, i.e. $x^\star$ is the only root of H in the ball $\overline{B(x_0, r)}$.