



LINMA2460 Nonlinear Programming

SIMON DESMIDT
ISSAMBRE L'HERMITE DUMONT

Academic year 2024-2025 - Q2



UCLouvain

Contents

1	Definitions, notations and random properties	2
1.1	Properties	3
1.2	Complexity table	3
2	TODO	4
3	Gradient descent without gradient	6
4	Local rates of convergence	8
4.1	Linear rate of GM	8
4.2	Local quadratic convergence of Newton's method	10
5	Tips and Tricks	11

Definitions, notations and random properties

- The Taylor expansion of order p of the function f around x_k and evaluated at y is:

$$T_p(y; x_k) = f(x_k) + \sum_{i=1}^p \frac{1}{i!} D^i f(x_k) (y - x_k)^i \quad (1.1)$$

- We can thus define the gradient w.r.t. y of the Taylor expansion of order p of f around x_k and evaluated at x_{k+1} :

$$\nabla_y T_p(x_{k+1}; x_k) = \nabla_y T_p(y; x_k) \big|_{y=x_{k+1}} \quad (1.2)$$

- An oracle is a "black box" that gives information about the derivatives based on x . The general form of an oracle is:

$$\text{p-order oracle: } x \mapsto \{D^i f(x)\}_{i=0}^p \quad (1.3)$$

And so we have the following simple oracles examples:

$$\begin{aligned} \text{Zero}^{th}\text{-order oracle: } x &\mapsto \{f(x)\} \\ \text{First-order oracle: } x &\mapsto \{f(x), \nabla f(x)\} \\ \text{Second-order oracle: } x &\mapsto \{f(x), \nabla f(x), \nabla^2 f(x)\} \end{aligned} \quad (1.4)$$

- $\mathcal{C}_L^p(\mathbb{R}^n)$: Class of functions p -times continuously differentiable with L -Lipschitz continuous p -order derivative, i.e. $\|D^p f(x) - D^p f(y)\| \leq L\|x - y\|, \forall x, y \in \mathbb{R}^n$. And so we have the following simple classes of problems:

- $\mathcal{C}_L^1(\mathbb{R}^n)$: Class of continuously differentiable functions with L -Lipschitz gradient;
- $\mathcal{C}_L^2(\mathbb{R}^n)$: Class of continuously differentiable functions with L -Lipschitz hessian.

- p th-order method (generalization of GM):

$$x_{k+1} = \arg \min_{y \in \mathbb{R}^n} \Omega_{x_k, y, p}(y) \equiv T_{x_k, p}(y) + \frac{M}{(p+1)!} \|y - x_k\|^{p+1} \quad (1.5)$$

1.1 Properties

- For a function $f \in \mathcal{C}^1(\Omega)$ and Ω is bounded, the following holds: $\|\nabla f(x)\| \leq L$ for all $x \in \Omega$ for some $L \geq 0$.
- By the mean value theorem, for a continuously differentiable function f , $\forall x, y \in \Omega$, $\exists z \in \Omega : f(y) - f(x) = \langle \nabla f(z), y - x \rangle$.
- For a matrix A and a scalar b , $\|A\| \leq b \implies |\lambda(A)| \leq b \implies |A| \preceq bI_n$, where the absolute value of the matrix is taken component wise.

1.2 Complexity table

Method	Lipschitz	∇f	$\nabla^2 f$...	$\nabla^p f$
Zero order		$O(n\varepsilon^{-2})$			
First order	$p = 1$	$O(\varepsilon^{-2})$			
Second order	$p = 2$	✗	$O(\varepsilon^{-3/2})$		
\vdots		✗	✗	\ddots	
p order		✗	✗	✗	$O(\varepsilon^{-\frac{p+1}{p}})$

TODO

We can generalise the property of a L-Lipschitz function to $f \in \mathcal{C}_L^p(\mathbb{R}^n)$. For $p = 1$, we had

$$f(y) \leq f(x_k) + \langle \nabla f(x_k), y - x_k \rangle + \frac{L}{2} \|y - x_k\|^2 \quad \forall y \in \mathbb{R}^n \quad (2.1)$$

For a general value of p , it becomes

$$f(y) \leq T_p(y; x_k) + \frac{L}{(p+1)!} \|y - x_k\|^{p+1} \quad \forall y \in \mathbb{R}^n \quad (2.2)$$

Using this, [we need a \$p\$ -th order oracle](#) for the method to work.

To solve $\min_{x \in \mathbb{R}^n} f(x)$, we can use the iteration

$$x_{k+1} = \arg \min_{y \in \mathbb{R}^n} T_p(y; x_k) + \frac{M}{(p+1)!} \|y - x_k\|^{p+1} \quad (2.3)$$

where the constant M is an approximation of the Lipschitz constant L . [Assuming \$f \in \mathcal{C}_L^p\(\mathbb{R}^n\)\$](#) , we have

$$\begin{aligned} f(x_{k+1}) &\leq T_p(x_{k+1}; x_k) + \frac{L}{(p+1)!} \|x_{k+1} - x_k\|^{p+1} \\ &= \underbrace{T_p(x_{k+1}; x_k) + \frac{M}{(p+1)!} \|x_{k+1} - x_k\|^{p+1}}_{\leq f(x_k)} + \frac{(L-M)}{(p+1)!} \|x_{k+1} - x_k\|^{p+1} \end{aligned} \quad (2.4)$$

where the inequality $\leq f(x_k)$ is due to the decrease of f and equation (2.3). [Suppose that \$M > 2L\$](#) . After some algebraic manipulations, we get

$$f(x_k) - f(x_{k+1}) \geq \frac{L}{(p+1)!} \|x_{k+1} - x_k\|^{p+1} \quad (2.5)$$

On the other hand, using the triangular inequality,

$$\begin{aligned} \|\nabla f(x_{k+1})\| &\leq \|\nabla f(x_{k+1}) - \nabla_y T_p(x_{k+1}; x_k)\| \\ &\quad + \underbrace{\left\| \nabla_y T_p(x_{k+1}; x_k) + \nabla \left(\frac{M}{(p+1)!} \|\cdot - x_k\|^{p+1} \right) \right\|_{y=x_{k+1}}}_{=0} \\ &\quad + \left\| \nabla \left(\frac{M}{(p+1)!} \|\cdot - x_k\|^{p+1} \right) \right\|_{y=x_{k+1}} \\ &\leq \frac{L}{p!} \|x_{k+1} - x_k\|^p + \frac{M}{p!} \|x_{k+1} - x_k\|^p \end{aligned} \quad (2.6)$$

$$\implies \|x_{k+1} - x_k\| \geq \left(\frac{p!}{L+M} \right)^{1/p} \|\nabla f(x_{k+1})\|^{1/p} \quad (2.7)$$

Combining equations (2.5) and (2.7),

$$f(x_k) - f(x_{k+1}) \geq \underbrace{\frac{L}{(p+1)!} \left(\frac{p!}{L+M} \right)^{\frac{p+1}{p}}}_{=:C(L)} \|\nabla f(x_{k+1})\|^{\frac{p+1}{p}} \quad (2.8)$$

Let $T(\varepsilon) = \inf\{k \in \mathbb{N} : \|\nabla f(x_k)\| \leq \varepsilon\}$. Assume that $T(\varepsilon) \geq 2$ and $f(x) \geq f_{low} \forall x \in \mathbb{R}^n$. Summing up (2.8) for $k = 0, \dots, T(\varepsilon) - 2$,

$$\begin{aligned} f(x_0) - f_{low} &\geq f(x_0) - f(x_{T(\varepsilon)-1}) = \sum_{k=0}^{T(\varepsilon)-2} f(x_k) - f(x_{k+1}) \\ &\geq (T(\varepsilon) - 1)C(L)\varepsilon^{\frac{p+1}{p}} \\ \implies T(\varepsilon) &\leq 1 + \frac{f(x_0) - f_{low}}{C(L)}\varepsilon^{-\frac{p+1}{p}} \equiv \mathcal{O}\left(\varepsilon^{-\frac{p+1}{p}}\right) \end{aligned} \quad (2.9)$$

Gradient descent without gradient

For this problem consider an adversarial attack on block-based image classifier. We have a machine learning model that given an image $a \in \mathbb{R}^p$ it returns $c(a) \in \mathbb{R}^m$, where $c_j(a) \in [0, 1]$ is the probability of image a to be in class j . The classifier prediction is: $j(a) = \arg \max_{j \in [1, \dots, m]} c_j(a)$.

TODO - Add mise en situation ou pas?

Given x_k let us decide:

$$x_{k+1} = x_k - \frac{1}{\sigma} g_{h_k}(x_k) \quad h_k > 0, \sigma > 0 \quad (3.1)$$

where $g_{h_k}(x_k) \in \mathbb{R}^n$ is given by:

$$[g_{h_k}(x_k)]_j = \frac{f(x_k + h e_j) - f(x_k)}{h_k} \quad \forall j \in [1, \dots, m] \quad (3.2)$$

Suppose that $f \in \mathcal{C}_L^1(\mathbb{R}^n)$. Then,

$$\|\nabla f(x_k) - g_{h_k}(x_k)\| \leq \frac{L\sqrt{n}}{2} h_k \quad (3.3)$$

Thus

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2} \|x_{k+1} - x_k\|^2 \\ &= f(x_k) + \langle g_{h_k}(x_k), x_{k+1} - x_k \rangle + \frac{\sigma}{2} \|x_{k+1} - x_k\|^2 \\ &\quad + \langle \nabla f(x_k) - g_{h_k}(x_k), x_{k+1} - x_k \rangle + \frac{(L - \sigma)}{2} \|x_{k+1} - x_k\|^2 \\ &\leq f(x_k) - \frac{1}{\sigma} \|g_{h_k}(x_k)\|^2 + \frac{1}{2\sigma} \|g_{h_k}(x_k)\|^2 \\ &\quad + \|\nabla f(x_k) - g_{h_k}(x_k)\| \frac{1}{\sigma} \|g_{h_k}(x_k)\| + \frac{(L - \sigma)}{2\sigma^2} \|g_{h_k}\|^2 \\ &\leq f(x_k) - \frac{1}{2\sigma} \|g_{h_k}(x_k)\|^2 + \frac{L\sqrt{n}}{2} h_k \frac{1}{\sigma} \|g_{h_k}\| + \frac{(L - \sigma)}{2\sigma^2} \|g_{h_k}\|^2 \\ &\leq f(x_k) - \frac{1}{2\sigma} \|g_{h_k}(x_k)\|^2 + \frac{L}{2} \left(\frac{nh_k^2}{2} + \frac{1}{2\sigma} \|g_{h_k}(x_k)\|^2 \right) + \frac{(L - \sigma)}{2\sigma^2} \|g_{h_k}\|^2 \\ &= f(x_k) - \left(\frac{2\sigma - L - 2(L - \sigma)}{4\sigma^2} \right) \|g_{h_k}(x_k)\|^2 + \frac{Ln}{4} h_k^2 \\ &= f(x_k) - \frac{(4\sigma - 3L)}{4\sigma} \|g_{h_k}(x_k)\|^2 + \frac{Ln}{4} h_k^2 \end{aligned} \quad (3.4)$$

$$\implies \frac{(4\sigma - 3L)}{4\sigma} \|g_{h_k}(x_k)\|^2 \leq f(x_k) - f(x_{k+1}) + \frac{Ln}{4} h_k^2 \quad (3.5)$$

If $\sigma \gg L$, then

$$\frac{1}{4\sigma} \|g_{h_k}(x_k)\|^2 \leq f(x_k) - f(x_{k+1}) + \frac{\sigma n}{4} h_k^2 \quad (3.6)$$

On the other hand, we have

$$\begin{aligned} \|\nabla f(x_k)\| &\leq \|\nabla f(x_k) - g_{h_k}(x_k)\| + \|g_{h_k}(x_k)\| \\ &\leq \frac{L\sqrt{n}}{2} h_k + \|g_{h_k}(x_k)\| \end{aligned} \quad (3.7)$$

Using trick (5.3) in chapter 5,

$$\begin{aligned} \implies \|\nabla f(x_k)\|^2 &\leq 2 \left(\frac{L\sqrt{n}}{2} h_k \right)^2 + 2 \|g_{h_k}(x_k)\|^2 \\ &\leq \frac{L^2 n}{2} h_k^2 + 2 \|g_{h_k}(x_k)\|^2 \end{aligned} \quad (3.8)$$

$$\implies \frac{1}{8\sigma} \|\nabla f(x_k)\|^2 \leq \frac{L^2 n}{16\sigma} h_k^2 + \frac{1}{4\sigma} \|g_{h_k}(x_k)\|^2 \quad (3.9)$$

$$\implies \frac{1}{8\sigma} \|\nabla f(x_k)\|^2 \leq f(x_k) - f(x_{k+1}) + \frac{\sigma n}{4} h_k^2 + \frac{\sigma n}{16} h_k^2 \quad (3.10)$$

Let $T(\varepsilon) = \inf\{k \in \mathbb{N} : \|\nabla f(x_k)\| \leq \varepsilon\}$, with $f(x)$ bounded below by f_{low} , summing up (3.10) for $k = 0, \dots, T(\varepsilon) - 1$:

$$\frac{T(\varepsilon)}{8\sigma} \varepsilon^2 \leq f(x_0) - f_{low} + \frac{5\sigma n}{4} \sum_{k=0}^{T(\varepsilon)-1} h_k^2 \quad (3.11)$$

If $\{h_k^2\}$ is summable

$$\implies T(\varepsilon) \leq 8\sigma \left(f(x_0) - f_{low} + \frac{5\sigma n}{4} \sum_{k=0}^{T(\varepsilon)-1} h_k^2 \right) \varepsilon^2 = \mathcal{O}(\varepsilon^2) \quad (3.12)$$

In terms of call to the oracle, we have a complexity bound of $\mathcal{O}(n\varepsilon^2)$.

Local rates of convergence

4.1 Linear rate of GM

Let $f \in \mathcal{C}_M^{2,2}(\mathbb{R}^n)$. Assume f has a local minimizer x^* such that

$$\mu I_n \preceq \nabla^2 f(x^*) \preceq M I_n \quad (4.1)$$

Let $x_{k+1} = x_k - \frac{1}{L} \nabla f(x_k)$ for a given $x_0 \in \mathbb{R}^n$.

Notice that

$$\begin{aligned} \nabla f(x_k) &= \nabla f(x_k) - \nabla f(x^*) \\ &= \int_0^1 \nabla^2 f(x^* + \tau(x_k - x^*)) (x_k - x^*) d\tau \\ &= \int_0^1 \nabla^2 f(x^* + \tau(x_k - x^*)) d\tau (x_k - x^*) \\ &= G_k(x_k - x^*) \end{aligned} \quad (4.2)$$

Then,

$$\begin{aligned} \|x_{k+1} - x^*\| &= \|x_k - \frac{1}{L} \nabla f(x_k) - x^*\| \\ &= \|(I_n - \frac{1}{L} G_k)(x_k - x^*)\| \\ &\leq \|I_n - \frac{1}{L} G_k\| \|x_k - x^*\| \end{aligned} \quad (4.3)$$

Since $f \in \mathcal{C}_M^{2,2}(\mathbb{R}^n)$, we have $\|\nabla^2 f(x^* + \tau(x_k - x^*)) - \nabla^2 f(x^*)\| \leq \tau M \|x_k - x^*\|$ and using this we get:

$$|\langle \nabla^2 f(x^* + \tau(x_k - x^*)) - \nabla^2 f(x^*) v, v \rangle| \leq \tau M \|x_k - x^*\| \|v\|^2 \quad \forall v \in \mathbb{R}^n \quad (4.4)$$

Using the bound (4.1) and the previous inequality, we get:

$$\begin{aligned} \tau M \|x_k - x^*\| \|v\|^2 &\leq |\langle \nabla^2 f(x^* + \tau(x_k - x^*)) - \nabla^2 f(x^*) v, v \rangle| \leq \tau M \|x_k - x^*\| \|v\|^2 \\ \nabla^2 f(x^*) - \tau M \|x_k - x^*\| I_n &\preceq \nabla^2 f(x^* + \tau(x_k - x^*)) \preceq \nabla^2 f(x^*) + \tau M \|x_k - x^*\| I_n \\ (\mu - \tau M \|x_k - x^*\|) I_n &\preceq \nabla^2 f(x^* + \tau(x_k - x^*)) \preceq (L + \tau M \|x_k - x^*\|) I_n \end{aligned}$$

By the properties of the semi-definite matrices, and the trick (5.4), we have:

$$\begin{aligned} \int_0^1 (\mu - \tau M \|x_k - x^*\|) \|v\|^2 d\tau &\leq \int_0^1 \langle \nabla^2 f(x^* + \tau(x_k - x^*)) v, v \rangle d\tau \\ &\leq \int_0^1 (L + \tau M \|x_k - x^*\|) \|v\|^2 d\tau \quad \forall v \in \mathbb{R}^n \end{aligned} \quad (4.5)$$

By using G_k and some constants, we get:

$$-\frac{1}{L}(L + \frac{M}{2}\|x_k - x^*\|)I_n \preceq -\frac{1}{L}G_k \preceq -\frac{1}{L}(\mu - \frac{M}{2}\|x_k - x^*\|)I_n \quad (4.6)$$

$$\left(1 - \frac{1}{L}(L + \frac{M}{2}\|x_k - x^*\|)\right) I_n \preceq I_n - \frac{1}{L}G_k \preceq \left(1 - \frac{1}{L}(\mu - \frac{M}{2}\|x_k - x^*\|)\right) I_n \quad (4.7)$$

And finally, we get:

$$\begin{aligned} \|I_n - \frac{1}{L}G_k\| &\leq \max \left\{ \left| 1 - \frac{1}{L}(L + \frac{M}{2}\|x_k - x^*\|) \right|, \left| 1 - \frac{1}{L}(\mu - \frac{M}{2}\|x_k - x^*\|) \right| \right\} \\ &= \max \left\{ \frac{M}{2L}\|x_k - x^*\|, 1 - \frac{\mu}{L} + \frac{M}{2L}\|x_k - x^*\| \right\} \\ &= 1 - \frac{\mu}{L} + \frac{M}{2L}\|x_k - x^*\| \end{aligned} \quad (4.8)$$

Suppose that $\frac{M}{2L}\|x_k - x^*\| \leq \frac{\mu}{2L} \iff \|x_k - x^*\| \leq \frac{\mu}{M}$

Then, in (4.8), we get:

$$\|I_n - \frac{1}{L}G_k\| \leq 1 - \frac{\mu}{2L} < 1 \quad (4.9)$$

And so, by (4.2)

$$\|x_{k+1} - x^*\| \leq \|I_n - \frac{1}{L}G_k\| \|x_k - x^*\| < \|x_k - x^*\| \quad (4.10)$$

If $\|x_0 - x^*\| < \frac{\mu}{M}$, it follows from the previous reasoning that:

$$\|x_2 - x^*\| \leq (1 - \frac{\mu}{2L})\|x_1 - x^*\| \leq (1 - \frac{\mu}{2L})^2\|x_0 - x^*\| \leq \frac{\mu}{M} \quad (4.11)$$

And so by induction, we can conclude that:

$$\|x_k - x^*\| \leq \left(1 - \frac{\mu}{2L}\right)^k \|x_0 - x^*\| \quad \forall k \geq 0 \quad (4.12)$$

\Rightarrow Linear rate of convergence

Given $\varepsilon > 0$, let $T(\varepsilon) = \inf\{k \in \mathbb{N} : \|x_k - x^*\| \leq \varepsilon\}$. Then, if $T(\varepsilon) \geq 1$ and using (4.12), we get:

$$\begin{aligned} \varepsilon &< \|x_{T(\varepsilon)-1} - x^*\| \leq \left(1 - \frac{\mu}{2L}\right)^{T(\varepsilon)-1} \|x_0 - x^*\| \\ \log \left(\frac{\varepsilon}{\|x_0 - x^*\|} \right) &\leq (T(\varepsilon) - 1) \log \left(1 - \frac{\mu}{2L} \right) \\ T(\varepsilon) - 1 &\leq \frac{\log \left(\frac{\varepsilon}{\|x_0 - x^*\|} \right)}{\log \left(1 - \frac{\mu}{2L} \right)} = \frac{\log (\|x_0 - x^*\| \varepsilon^{-1})}{|\log (1 - \frac{\mu}{2L})|} \end{aligned} \quad (4.13)$$

$$T(\varepsilon) \leq \mathcal{O}(\log(\varepsilon^{-1}))$$

\rightarrow Note: convexity was never assumed!

4.2 Local quadratic convergence of Newton's method

Let $f \in \mathcal{C}_M^{2,2}(\mathbb{R}^n)$. Assume f has a local minimizer x^* such that

$$\mu I_n \preceq \nabla^2 f(x^*) \quad \mu > 0 \quad (4.14)$$

Given $x_0 \in \mathbb{R}^n$, let:

$$x_{k+1} = x_k - \nabla^{-2} f(x_k) \nabla f(x_k) \quad (4.15)$$

We have, by the previous equation and the definition of G_k (4.2):

$$\begin{aligned} \|x_{k+1} - x^*\| &= \|x_k - \nabla^{-2} f(x_k) \nabla f(x_k) - x^*\| \\ &= \|(x_k - x^*) - \nabla^{-2} f(x_k) G_k(x_k - x^*)\| \\ &= \|\nabla^{-2} f(x_k) \left(\nabla^2 f(x_k) - \int_0^1 \nabla^2 f(x^* + \tau(x_k - x^*)) d\tau \right) (x_k - x^*)\| \\ &= \|\nabla^{-2} f(x_k) \left(\int_0^1 \nabla^2 f(x_k) - \nabla^2 f(x^* + \tau(x_k - x^*)) d\tau \right) (x_k - x^*)\| \\ &\leq \|\nabla^{-2} f(x_k)\| \left(\int_0^1 \|\nabla^2 f(x_k) - \nabla^2 f(x^* + \tau(x_k - x^*))\| d\tau \right) \|x_k - x^*\| \\ &\leq \|\nabla^{-2} f(x_k)\| \left(\int_0^1 M(1 - \tau) \|x_k - x^*\| d\tau \right) \|x_k - x^*\| \\ &\leq \|\nabla^{-2} f(x_k)\| \|x_k - x^*\|^2 \frac{M}{2} \end{aligned} \quad (4.16)$$

Since $f \in \mathcal{C}_M^{2,2}(\mathbb{R}^n)$, we have

$$\nabla^2 f(x^* + \tau(x_k - x^*)) - \nabla^2 f(x^*) \succeq \tau M \|x_k - x^*\| I_n \quad (4.17)$$

$$\begin{aligned} \nabla^2 f(x_k) &\succeq \nabla^2 f(x^*) - M \|x_k - x^*\| I_n \\ &\succeq (\mu - M \|x_k - x^*\|) I_n \\ \lambda_{\min}(\nabla^2 f(x_k)) &\geq \mu - M \|x_k - x^*\| \end{aligned} \quad (4.18)$$

Suppose that $-M \|x_k - x^*\| \geq -\frac{\mu}{2} \Leftrightarrow \|x_k - x^*\| \leq \frac{\mu}{2M}$
Then,

$$\begin{aligned} \lambda_{\min}(\nabla^2 f(x_k)) &\geq \frac{\mu}{2} \\ \lambda_{\max}(\nabla^{-2} f(x_k)) &\leq \frac{2}{\mu} \\ \Rightarrow \|\nabla^{-2} f(x_k)\| &\leq \frac{2}{\mu} \end{aligned} \quad (4.19)$$

Therefore, by (4.16), we conclude that:

$$\begin{aligned} \|x_{k+1} - x^*\| &\leq \frac{M}{2} \|\nabla^{-2} f(x_k)\| \|x_k - x^*\| \\ &\leq \frac{M}{\mu} \|x_k - x^*\|^2 \end{aligned} \quad (4.20)$$

Tips and Tricks

1. Approximation of the max:

$$\max\{z, 0\} = \frac{z + |z|}{2} = \frac{z + \sqrt{z^2}}{2} \approx \frac{z + \sqrt{z^2 + \delta}}{2} \quad (5.1)$$

2.

$$ab \leq \frac{a^2 + b^2}{2} \quad (5.2)$$

3.

$$(a + b)^2 \leq 2a^2 + 2b^2 \quad (5.3)$$

4. V-trick:

$$\langle xv, v \rangle \leq \|x\| \|v\|^2 \quad (5.4)$$