



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	Tips and Tricks	8

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 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,

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

Le + rouge doit être un -?

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

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

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 n°3**

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

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

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

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

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

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

Tips and Tricks

- Approximation of the max:

$$\max\{z, 0\} = \frac{z + |z|}{2} = \frac{z + \sqrt{z^2}}{2} \approx \frac{z + \sqrt{z^2 + \delta}}{2} \quad (4.1)$$

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