LINMA2460 Nonlinear Programming

SIMON DESMIDT
ISSAMBRE L'HERMITE DUMONT

Academic year 2024-2025 - Q2

UCLouvain

Contents

Definitions, notations and random properties

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

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

- 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}:

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

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

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

And so we have the following simple oracles examples:

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

- $\mathcal{C}_L^p(\mathbb{R}^n)$: 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):

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

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, $\|A\| \leq b \implies |\lambda(A)| \leq b \implies |A| \preceq b I_n$, 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		$O(n\epsilon^{-2})$			
First order	$p = 1$	$O(\epsilon^{-2})$			
Second order	$p = 2$	X	$O(\epsilon^{-3/2})$		
$\vdots$		X	X	$\ddots$	
p order		X	X	X	$O(\epsilon^{-\frac{p+1}{p}})$

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

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

For a general value of p, it becomes

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

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

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

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

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

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

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

On the other hand, using the triangular inequality,

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

4

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

Combining equations (2.5) and (2.7),

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

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 $k = 0, \dots, T(\varepsilon) - 2$,

$$f(x_0) - f_{low} \geq f(x_0) - f(x_{T(\varepsilon)-1}) = \sum_{k=0}^{T(\varepsilon)-2} f(x_k) - f(x_{k+1})$$

$$\geq (T(\varepsilon) - 1)C(L)\varepsilon^{\frac{p+1}{p}} \tag{2.9}$$

$$\implies T(\varepsilon) \leq 1 + \frac{f(x_0) - f_{low}}{C(L)}\varepsilon^{-\frac{p+1}{p}} \equiv \mathcal{O}\left(\varepsilon^{-\frac{p+1}{p}}\right)$$

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 return $c(a) \in \mathbb{R}^m$, where $c_j(a) \in [0,1]$ is the probability of image a to be in class j. The classifier prediction is: $j(a) = \arg\max_{j \in [1,\dots,m]} c_j(a)$.