A short proof of the Newton-Kantarovich theorem

Domingos M Cardoso[1] and Luís A Vieira [2]

October 22, 2008

1 Basic notions

Consider a nonempty open convex subset $C \subseteq \mathbb{R}^n$. Then a function $f : C \to \mathbb{R}$ is m-strongly convex in C if it has second derivatives on C and $\exists m > 0$ such that

$$(1) \qquad \forall x \in C \qquad m\|h\|^2 \leq h^T \nabla^2 f(x) h \;\; \forall h \in \mathbb{R}^n,$$

where $\nabla^2 f(x)$ denotes the Hessian matriz of f in x. It is immediate that every strongly convex function is strictly convex, however the converse is not true. For instance, the function $f : \mathbb{R} \to \mathbb{R}$, such that $f(x) \mapsto x^{2k}$, with $k \in \mathbb{N} \setminus \{1\}$, is strictly convex but it is not strongly convex (since its second derivative in $x = 0$ is null and then (1) is not fulfilled). It is also immediate that the Hessian matrix of each strongly convex function is positive definite and then it is invertible.

Lemma 1.1 *If $C \subseteq \mathbb{R}^n$ is a nonempty open convex subset and the function $f : C \to \mathbb{R}$ is m-strongly convex in C, then*

$$\|(\nabla^2 f(x))^{-1}\| \leq \frac{1}{m},$$

where $\|A\| = \sup_{h \neq 0} \frac{\|Ah\|_2}{\|h\|_2}$.

Proof: $\|(\nabla^2 f(x))^{-1}\| = \sup_{y \neq 0} \frac{\|(\nabla^2 f(x))^{-1} y\|_2}{\|y\|_2} = \sup_{\nabla^2 f(x) h \neq 0} \frac{\|h\|_2}{\|\nabla^2 f(x) h\|_2} = \frac{1}{\inf_{h \neq 0} \frac{\|\nabla^2 f(x) h\|_2}{\|h\|_2}}$ (since $h \neq 0 \Leftrightarrow \nabla^2 f(x) h \neq 0$). Taking into account that $\forall h \in \mathbb{R}^n \setminus \{0\}$

$$m\|h\|_2^2 \leq h^T \nabla^2 f(x) h \leq \|h\|_2 \|\nabla^2 f(x) h\|_2 \Rightarrow m \leq \frac{\|\nabla^2 f(x) h\|\|_2}{\|h\|_2},$$

it follows $m \leq \inf_{h \neq 0} \frac{\|\nabla^2 f(x) h\|_2}{\|h\|_2} \Leftrightarrow \frac{1}{\inf_{h \neq 0} \frac{\|\nabla^2 f(x) h\|_2}{\|h\|_2}} \leq \frac{1}{m}$ and therefore, $\|(\nabla^2 f(x))^{-1}\| \leq \frac{1}{m}$. ■

[1]University of Aveiro, Portugal (dcardoso@ua.pt)
[2]University of Porto, Portugal (lvieira@fe.up.pt).

Consider the nonempty subset $D \subseteq \mathbb{R}^n$. Then a function $F : D \to \mathbb{R}^m$ is Lipschitzian in D, with constant of Lipschitz L, if $\forall x, y \in D \ \ \|f(x) - f(y)\| \leq L\|x - y\|$.

Lemma 1.2 *If $C \subseteq \mathbb{R}^n$ is a nonempty open convex subset, the function $f : C \to \mathbb{R}$ has second derivatives on C and $\nabla^2 f(x)$ is Lipschitzian with constant of Lipschitz L, then $\forall h \in \mathbb{R}^n$, such that $x + h \in C$,*

$$\|\nabla f(x+h) - \nabla f_x(h)\| \leq \frac{L}{2}\|h\|^2, \tag{2}$$

where $\nabla f(y)$ denotes the gradient of f in y and $f_x(h) = f(x) + (\nabla f(x))^T h + \frac{1}{2} h^T \nabla^2 f(x) h$ is the quadratic approximation of the function f in a neighborhood of x in C.

Proof: First, it should be noted that $\nabla f_x(h) = \nabla f(x) + \nabla^2 f(x) h$, and then $\nabla f(x+h) - \nabla f_x(h) = \nabla f(x+h) - \nabla f(x) - \nabla^2 f(x) h$. Therefore,

$$\begin{aligned}
\|\nabla f(x+h) - \nabla f_x(h)\| &= \|\nabla f(x+h) - \nabla f(x) - \nabla^2 f(x) h\| \\
&= \| \int_0^1 \nabla^2 f(x + th) h dt - \nabla^2 f(x) h\| \\
&= \| \int_0^1 (\nabla^2 f(x + th) - \nabla^2 f(x)) h dt\| \\
&\leq \int_0^1 \|\nabla^2 f(x + th) - \nabla^2 f(x)\| \|h\| dt \\
&\leq \int_0^1 Lt\|h\|^2 dt. \\
&= \frac{L}{2}\|h\|^2.
\end{aligned}$$

■

2 The Newton-Kantorovich theorem

Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ with continuous second derivatives and its quadratic approximation in a neighborhood of $x^k \in \mathbb{R}^n$,

$$f_{x^k}(h) = f(x^k) + (\nabla f(x^k))^T h + \frac{1}{2} h^T \nabla^2 f(x^k) h.$$

If $\nabla^2 f(x)$ is invertible, then the well known Newton's iterates

$$x^{k+1} = x^k - (\nabla^2 f(x^k))^{-1} \nabla f(x^k), \ \ k = 0, 1, \ldots, \tag{3}$$

follows since the critical point that solves the system of equations $\nabla f_{x^k}(h) = 0 \Leftrightarrow \nabla f(x^k) + \nabla^2 f(x^k)h = 0 \Leftrightarrow \nabla^2 f(x^k)h = -\nabla f(x^k)$ is $h = -(\nabla^2 f(x^k))^{-1}\nabla f(x^k)$. The Newton-Kantorovich theorem states some conditions that assure the convergence of the Newton's iterates (3), $x^{k+1} = x^k + h$, to a critical point of f.

Theorem 2.1 *Consider a nonempty open set $C \subseteq \mathbb{R}^n$ and a function $f : \mathbb{R}^n \to \mathbb{R}$ with second derivatives in C. Assuming that $\nabla^2 f(x)$ is a Lipschitzian function with Lipschitz constant L, if the function f is m-strongly convex in a open ball with center in a critical point x^* and radius $\frac{m}{L}$, $B(x^*, \frac{m}{L}) \subset C$, then $\forall x^0 \in B(x^*, \frac{m}{L})$ the sequence* (3) *has quadratic convergence to x^*.*

Proof: If x^{k+1} is determined from x^k, according to (3), then

$$\begin{aligned} x^{k+1} - x^* &= x^k - (\nabla^2 f(x^k))^{-1}\nabla f(x^k) - x^* \\ &= x^k - x^* - (\nabla^2 f(x^k))^{-1}\nabla f(x^k) + (\nabla^2 f(x^k))^{-1}\nabla f(x^*) \\ &\qquad\qquad\qquad\qquad \text{(notice that } \nabla f(x^*) = 0) \\ &= (\nabla^2 f(x^k))^{-1}\nabla^2 f(x^k)(x^k - x^*) - (\nabla^2 f(x^k))^{-1}\nabla f(x^k) \\ &\qquad + (\nabla^2 f(x^k))^{-1}\nabla f(x^*) \\ &= (\nabla^2 f(x^k))^{-1}[\nabla^2 f(x^k)(x^k - x^*) - (\nabla f(x^k) - \nabla f(x^*))]. \end{aligned}$$

Therefore, assuming $x^k \in B(x^*, \frac{m}{2^s L})$, with $s \geq 0$, it follows

$$\begin{aligned} \|x^{k+1} - x^*\| &= \|(\nabla^2 f(x^k))^{-1}[\nabla f(x^*) - \nabla f(x^k) - \nabla^2 f(x^k)(x^* - x^k)]\| \\ &\leq \|(\nabla^2 f(x^k))^{-1}\|\|\nabla f(x^*) - \nabla f(x^k) - \nabla^2 f(x^k)(x^* - x^k)\| \\ &\leq \frac{1}{m}\|\nabla f(x^*) - \nabla f(x^k) - \nabla^2 f(x^k)(x^* - x^k)\| \text{ (by Lemma 1.1)} \\ &\leq \frac{L}{2m}\|x^* - x^k\|^2 \text{ (by Lemma 1.2)} \\ &< \frac{L}{2m}\frac{m^2}{2^{2s}L^2} = \frac{m}{2^{2s+1}L} \text{ (since } \|x^* - x^k\| < \frac{m}{2^s L}). \end{aligned}$$

Thus $x^k \in B(x^*, \frac{m}{2^s L}) \Rightarrow x^{k+1} \in B(x^*, \frac{m}{2^{2s+1}L})$ and since, by hypothesis, $x^0 \in B(x^*, \frac{m}{L})$, then

$$x^0 \in B(x^*, \frac{m}{L}) \Rightarrow x^1 \in B(x^*, \frac{m}{2L}) \Rightarrow x^2 \in B(x^*, \frac{m}{2^3 L}) \Rightarrow x^3 \in B(x^*, \frac{m}{2^7 L}) \Rightarrow \cdots .$$

Furthermore, since $\|x^{k+1} - x^*\| \leq \frac{L}{2m}\|x^k - x^*\|^2 \Leftrightarrow \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|^2} \leq \frac{L}{2m}$, the sequence of Newton iterates (3) has quadratic convergence. ■