Differentiable functions and Lipschitz continuity

Clarify the multivariate meaning of $\|\nabla E(u)\|$ for Lipschitz continuous functions.

First assume $\|\nabla E(x)\|_{S^\infty} \le L$ for all x. Note that according to our notation $\nabla E(x)$ is the transpose of the Jacobian $JE(x)$. In particular, $\|\nabla E(x)\|_{S^\infty} = \|JE(x)\|_{S^\infty}$. Let

$$g(t) = \langle E(x) - E(y), E(tx + (1-t)y \rangle.$$

Using the mean value theorem and Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\|E(x) - E(y)\|_2^2 &= g(1) - g(0) = g'(\xi) && (1)\\
&= \langle E(x) - E(y), JE\left(\xi x + (1-\xi)y\right)(x-y)\rangle && (2)\\
&\le \|E(x) - E(y)\|_2 \|JE\left(\xi x + (1-\xi)y\right)(x-y)\|_2 && (3)\\
&\le \|E(x) - E(y)\|_2 \|JE\left(\xi x + (1-\xi)y\right)\|_{S^\infty} \|x-y\|_2 && (4)\\
&\le \|E(x) - E(y)\|_2 L \|x-y\|_2 . && (5)
\end{aligned}$$

Hence $E(x)$ has Lipschitz constant L.

Now assume that E has Lipschitz constant L. Then we have

$$\|JE(x)v\|_2 = \lim_{h\to 0} (1/h)\|E(x+hv) - E(x)\|_2 \le \lim_{h \to 0} (1/h) L \|hv\|_2 = L\|v\|_2 . \quad (6)$$

Taking the supremum on both sides yields the desired result:

$$\|\nabla E(x)\|_{S^\infty} = \sup_{\|v\|_2 = 1} \|\nabla E(x) v\|_2 \le \sup_{\|v\|_2=1} L\|v\|_2 = L. \quad (7)$$

Convergence of gradient descent for twice continuously differentiable L-smooth m-strongly convex functions

Define

$$G_\tau(u) = u - \tau \nabla E(u).$$

It holds that

$$\|G_\tau(u) - G_\tau(v)\|_2 = \|u - v\tau(\nabla E(u) - \nabla E(v))\|_2$$

Now by the mean value theorem (i.e. the same trick as above), it holds that there exists a z (which is some convex combination of u and v) such that

$$\nabla E(u) - \nabla E(v) = \nabla^2 E(z)(u-v).$$

We find

$$\begin{aligned}
\|G_\tau(u) - G_\tau(v)\|_2 &= \|u - v\tau\nabla^2 E(z)(u-v)\|_2\\
&= \|(I - \tau\nabla^2 E(z))(u-v)\|_2\\
&\le \|I - \tau\nabla^2 E(z)\|_{S^\infty} \|(u-v)\|_2.
\end{aligned}$$

Since E is L-smooth the eigenvalues of the Hessian are bounded from above by L. Since Since E is m-strongly convex the eigenvalues of the Hessian are bounded from below by m. Therefore it holds that

$$\|I - \tau\nabla^2 E(z)\|_{S^\infty} \leq \max\{|1 - \tau L|, |1 - \tau m|\},$$

and for $\tau \in]0, \frac{2}{L}[$ we find that G_τ is a contraction. Note that the contraction coefficient becomes minimal for $\tau = \frac{2}{L+m}$.