Differentiable functions and Lipschitz continuity

First assume $\|\nabla E(x)\| \leq L$ for all x. 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

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

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

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

$$\|\nabla E(x)v\| = \lim_{h\to 0} (1/h)\|E(x+hv) - E(x)\| \leq \lim_{h\to 0} (1/h)L\|hv\| = L\|v\|. \qquad (6)$$

Taking the supremum on both sides yields the desired result:

$$\|\nabla E(x)\| = \sup_{\|v\|=1} \|\nabla E(x)v\| \leq \sup_{\|v\|=1} L\|v\| = L. \qquad (7)$$

Characterization of L-smooth functions

We will prove $1 \Rightarrow 2 \Rightarrow 4 \Rightarrow 1$ and $2 \Leftrightarrow 5$, $2 \Leftrightarrow 3$ to show equivalence of all statements.

- $2 \Leftrightarrow 5$: See exercise 1.4 for equivalence of positivity of Hessian and convexity.
- $2 \Leftrightarrow 3$: See hint in exercise 1.4, first order definition of convexity is equivalent to convexity. Applying the first order definition of convexity to the function $\frac{L}{2}\|u\|^2 - E(u)$ we have

$$\begin{aligned}
&\frac{L}{2}\|v\|^2 - E(v) \geq \frac{L}{2}\|u\|^2 - E(u) + \langle Lu - \nabla E(u), v-u\rangle && (8)\\
\Leftrightarrow\ & E(u) + \langle \nabla E(u), v-u\rangle - \frac{L}{2}\|u\|^2 + \frac{L}{2}\|v\|^2 + L\|u\|^2 - L\langle u, v\rangle \geq E(v) && (9)\\
\Leftrightarrow\ & E(v) \leq E(u) + \langle \nabla E(u), v-u\rangle + \frac{L}{2}\|v-u\|^2 && (10)
\end{aligned}$$

  Note that this is a quadratic upper bound on E. It is minimized

$$\min_v E(u) + \langle \nabla E(u), v-u\rangle + \frac{L}{2}\|v-u\|^2$$

  at the point $v^* = u - \frac{1}{L}\nabla E(u)$ with

$$E^* \leq E(u) - \frac{1}{2L}\|\nabla E(u)\|^2,$$

  where E^* denotes the global optimum.

- $\underline{1 \Rightarrow 2}$: To show convexity, we prove monotonicity of the gradient of $\frac{L}{2}\|u\|^2 - E(u)$:

$$\begin{aligned}
&\langle L(u-v) - (\nabla E(u) - \nabla E(v)), u - v\rangle = L\|u-v\|^2 - \langle \nabla E(u) - \nabla E(v), u - v\rangle \\
&\geq L\|u-v\|^2 - \|\nabla E(u) - \nabla E(v)\|\,\|u-v\| \\
&\geq L\|u-v\|^2 - L\|u-v\|\,\|u-v\| = 0.
\end{aligned}$$

- $\underline{2 \Rightarrow 4}$: Define

$$E_v(w) = E(w) - \langle \nabla E(v), w\rangle$$

$E_v(w)$ is L-smooth and convex since only a linear term is added. The gradient is given as:

$$\nabla E_v(w) = \nabla E(w) - \nabla E(v)$$

Clearly, v minimizes $\nabla E_v(w)$ since the gradient is zero. Since v minimizes E_v we have from the observation in $2 \Leftrightarrow 3$ that

$$E_v(w) - E_v(v) \geq \frac{1}{2L}\|\nabla E(w)\|^2.$$

Now we have

$$\begin{aligned}
E(v) - E(u) - \langle \nabla E(u), v - u\rangle &= E(v) - \langle \nabla E(u), v\rangle - E(u) + \langle \nabla E(u), u\rangle && (11)\\
&= E_u(v) - E_u(u) && (12)\\
&\geq \frac{1}{2L}\|\nabla E_u(v)\|^2 && (13)\\
&= \frac{1}{2L}\|\nabla E(v) - \nabla E(u)\|^2 && (14)
\end{aligned}$$

And also

$$\begin{aligned}
E(u) - E(v) - \langle \nabla E(v), u - v\rangle &= E(u) - \langle \nabla E(v), u\rangle - E(v) + \langle \nabla E(v), v\rangle && (15)\\
&= E_v(u) - E_v(v) && (16)\\
&\geq \frac{1}{2L}\|\nabla E_v(u)\|^2 && (17)\\
&= \frac{1}{2L}\|\nabla E(u) - \nabla E(v)\|^2 && (18)
\end{aligned}$$

Combining these two estimates gives:

$$\begin{aligned}
&-\langle \nabla E(v), u - v\rangle - \langle \nabla E(u), v - u\rangle \geq \frac{1}{L}\|\nabla E(u) - \nabla E(v)\|^2 && (19)\\
\Leftrightarrow\ &\langle \nabla E(v) - \nabla E(u), v - u\rangle \geq \frac{1}{L}\|\nabla E(u) - \nabla E(v)\|^2 && (20)
\end{aligned}$$

- $\underline{4 \Rightarrow 1}$: Follows directly from Cauchy-Schwarz inequality:

$$\frac{1}{L}\|\nabla E(u) - \nabla E(v)\|^2 \leq \langle \nabla E(u) - \nabla E(v), u - v\rangle \leq \|\nabla E(u) - \nabla E(v)\|\,\|u - v\|$$

Multiplying the above by $L/\|\nabla E(u) - \nabla E(v)\|$ yields the result.