# Mathematics for Intelligent Systems Lecture 7 Homework Partial Solution

**(Gradients and Hessians)**

Andrea Baisero

## Gradient vs Direction of Steepest Descent

Recall that the derivative of a function (aka differential, aka directional derivative, aka linearization of a function) $f : \mathcal{V} \to \mathbb{R}$ at a point $x$ is defined as $df_x \in \mathcal{V}^*$,

$$df_x(v) = \lim_{t \to 0} \frac{f(x + tv) - f(x)}{t} \tag{1}$$

We have also seen that the gradient $\nabla f(x)$ is defined as the "column vector" of partial derivatives evaluated at $x$,

$$\nabla f(x) \equiv \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_d} \end{pmatrix} \tag{2}$$

Notice that the notion of derivative $df_x$ is a coordinate-less one; it exists and is the same whatever the choice of a basis is. On the other hand, the notion of gradient $\nabla f(x)$ is by definition dependent on an algebraeic expression (i.e. on a coordinate representation). These two notion are nonetheless related by $df_x(v) = \nabla f(x) \cdot v$, which holds in any basis.

We define the vector of steepest descent $\delta^* \in \mathcal{V}$ at the point $x$ as

$$\delta^* \equiv \operatorname*{arg\,min}_{\delta \in \mathcal{V}} df_x(\delta) \text{ s.t. } \delta^2 = 1 \tag{3}$$

The negative gradient $-\nabla f(x)$ is typically confused to be the vector of steepest descent. However, we will see that it is not necessarily the case.

(a) Find the vector of steepest descent $\delta^*$ in the case of Euclidean metric space, $\langle x, y \rangle = x^\intercal y$. Graphical argument is ok.

(b) Find the vector of steepest descent $\delta^*$ in the case of non-Euclidean metric space, with metric tensor $A\mathbb{R}^{d \times d}$, i.e. $\langle x, y \rangle = x^\intercal A y$. Graphical argument is allowed, although some algebraeic manipulation is required. Hint: Use the Cholesky decomposition of the metric tensor $A = B^\intercal B$.

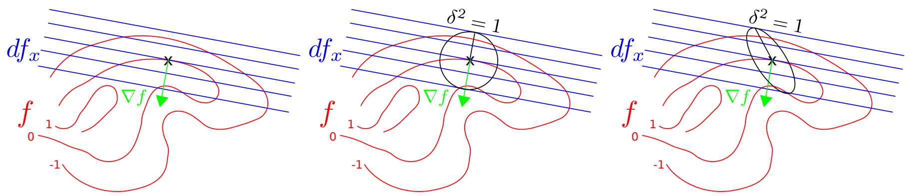


Figure 1: On the left, the relationship between function $f$, its differential $df_x$ as $x$, and its gradient $\nabla f(x)$ at $x$. Notice how the isolines of $df_x$ are straight, since it is a linear approximation, and that the gradient $\nabla f(x)$ is tangent to the isoline at $x$. In the middle we have the euclidean metric case, where the "unit sphere" around $x$ is an actual sphere. On the right we have the non-euclidean metric case, where the "unit sphere" around $x$ is actually an ellipsoid. The black straight line represents in each corresponding case the direction of steepest descent, i.e. the direction within the unit spheres which extends the most towards the lower-valued $df_x$ isolines.

## Solution

Figure 1 depicts the situation at hand.

(a) In the case of orthonormal basis, the "isoline" of the unit sphere $\delta^2 = \delta^\intercal \delta = 1$ really is spherical (see Fig. 1b). In such case, we notice that the direction of steepest descent coaligns itself with the gradient, and is $\delta^* = -\nabla f(x)$.

(b) In the case of non-orthonormal basis, the "isoline" of the unit sphere $\delta^2 = \delta^\intercal A \delta = 1$ is actually an ellipsoid (see Fig. 1c). We notice in the figure that the direction of steepest descent does not align with the gradient $\nabla f(x)$. To find exactly which direction we should take instead, we can apply the change of basis which is inherent to the metric matrix $A = B^\intercal B$, which is $z = B\delta$.

$$
\begin{aligned}
\delta^* &\equiv \arg\min_{\delta \in \mathcal{V}} df_x(\delta) \text{ s.t. } \delta^2 = 1 \\
&\equiv \arg\min_{\delta \in \mathcal{V}} \nabla f(x)^\intercal \delta \text{ s.t. } \delta^\intercal B^\intercal B \delta = 1 \\
z^* &\equiv \arg\min_{z \in \mathcal{V}} \nabla f(x)^\intercal B^{-1} z \text{ s.t. } z^\intercal z = 1
\end{aligned}
$$

We have now resorted to the euclidean case, so we know that the solution is $z^* = -B^{-\intercal} \nabla f(x)$. When we change back to the original coordinates, we obtain $\delta^* = B^{-1} z^* = -B^{-1} B^{-\intercal} \nabla f(x) = -A^{-1} \nabla f(x)$.