Modern methods of decision making 2021
Seminar · February 1st.

Exercise 1. Physical interpretation of gradients.

Let E be a Euclidean space with scalar product $\langle \cdot , \cdot \rangle$. Denote $\|.\|$ the inherited norm defined by $\|x\| = \sqrt{\langle x, x \rangle}$. Let $U \subset E$ be open and $f : U \to \mathbb{R}$ be differentiable. Select $x \in U$ such that $\nabla f(x) \neq 0$. For all $v \in E$, denote $\gamma_v(t) := x + tv$ (always defined for t small enough). Then, show that

$$\frac{\nabla f(x)}{\|\nabla f(x)\|} \in \underset{v \in E : \|v\| = 1}{\operatorname{argmax}} (f \circ \gamma_v)'(0),$$

and that

$$\|\nabla f(x)\| = \max_{v \in E : \|v\| = 1} (f \circ \gamma_v)'(0).$$

Solution.

Select $v \in E$ such that $\|v\| = 1$. Then

$$(f \circ \gamma_v)'(0) = \lim_{\substack{t \to 0 \\ t \neq 0}} \frac{f(x + tv) - f(x)}{t} = d_x f(v).$$

By definition of $\nabla f(x)$, we deduce that

$$(f \circ \gamma_v)'(0) = \langle v , \nabla f(x) \rangle.$$

Cauchy-Schwarz ineq. $\Rightarrow$

$$(f \circ \gamma_v)'(0) \leqslant \|v\| \cdot \|\nabla f(x)\| = \|\nabla f(x)\|.$$

It remains to observe that, letting $v^* = \dfrac{\nabla f(x)}{\|\nabla f(x)\|}$, we have

$$\begin{aligned}(f \circ \gamma_{v^*})'(0) &= \langle v^*, \nabla f(x)\rangle \\ &= \frac{\langle \nabla f(x), \nabla f(x)\rangle}{\|\nabla f(x)\|} \\ &= \|\nabla f(x)\|,\end{aligned}$$

which proves the claim.

Exercise 2. Chain Rule

Let E, F, G be normed vector spaces. Let $U \subset E$ and $V \subset F$ be open sets. Let finally $f : U \to F$ and $g : V \to G$ be differentiable maps such that $f(U) \subset V$. Show that $g \circ f : U \to G$ is differentiable and that $\forall x \in U, \forall h \in E$:

$$d_x(g \circ f)(h) = d_{f(x)} g\left(d_x f(h)\right).$$

Solution.

We need to prove that, $\forall x \in U$ and $\forall h \in E$ such that $x + h \in U$, we have

$$g \circ f(x+h) = g \circ f(x) + d_{f(x)} g\left(d_x f(h)\right) + \text{peanuts}(h),$$

where $\dfrac{\| \text{peanuts}(h) \|_G}{\| h \|_E} \xrightarrow[h \to 0]{} 0.$

To this aim, first notice that since f and g are differentiable, we have:

$$\begin{cases} f(x+h) = f(x) + d_x f(h) + \|h\|_E \, \varepsilon_f(h) \\ \text{and} \\ g(y+w) = g(y) + d_y g(w) + \|w\|_F \, \varepsilon_g(w) \end{cases}$$

As a result, we have

$$g \circ f(x+h) = g\Big(\overbrace{f(x)}^{y} + \overbrace{d_x f(h) + \|h\|_E \, \varepsilon_f(h)}^{w} \Big)$$

$$= g(f(x)) + d_{f(x)} g\big(d_x f(h) + \|h\|_E \, \varepsilon_f(h) \big)$$

$$+ \| d_x f(h) + \|h\|_E \, \varepsilon_f(h) \|_G \; \varepsilon_g\big(d_x f(h) + \|h\|_E \, \varepsilon_f(h) \big).$$

which, by linearity of $d_{f(x)} g$, can be written

$$= g \circ f(x) + d_{f(x)} g(d_x f(h)) + \|h\|_E \, d_{f(x)} g(\varepsilon_f(h))$$

$$+ \| d_x f(h) + \|h\|_E \, \varepsilon_f(h) \|_G \; \varepsilon_g\big(d_x f(h) + \|h\|_E \, \varepsilon_f(h) \big)$$

Call this (terrible) expression "peanuts(h)"

It is then rather straightforward to see that

$$\frac{\text{peanuts}(h)}{\|h\|_E} = d_{f(x)} g\left(\varepsilon_f(h)\right) + \left\| d_x f\left(\frac{h}{\|h\|_E}\right) + \varepsilon_f(h) \right\|_G \varepsilon_g\left(d_x f(h) + \|h\|_E \varepsilon_f(h)\right)$$

goes to 0 when $h \to 0$ since

$$\left.\begin{array}{l} \varepsilon_f(h) \underset{h \to 0}{\longrightarrow} 0 \\ \text{and} \\ d_{f(x)} g(v) \underset{v \to 0}{\longrightarrow} 0 \end{array}\right\} \Rightarrow d_{f(x)} g\left(\varepsilon_f(h)\right) \underset{h \to 0}{\longrightarrow} 0,$$

$$\begin{aligned} \left\| d_x f\left(\frac{h}{\|h\|_E}\right) + \varepsilon_f(h) \right\|_G &\leq \left\| d_x f\left(\frac{h}{\|h\|_E}\right) \right\|_G + \overbrace{\left\| \varepsilon_f(h) \right\|_G}^{\xrightarrow{h \to 0} 0} \\ &\leq \underbrace{\| d_x f \|_{op} + \| \varepsilon_f(h) \|_G}_{\text{Remains bounded}}, \end{aligned}$$

and

$$\varepsilon_g\left(d_x f(h) + \|h\|_E \varepsilon_f(h)\right) \underset{h \to 0}{\longrightarrow} 0.$$

Exercise 3. Classical computations

Let E be Euclidean with scalar product $\langle \cdot, \cdot \rangle$, associated norm $\|.\|$.

1) Let $L : E \to E$ be a symmetric linear map, i.e., such that $\langle L(x), y \rangle = \langle x, L(y) \rangle$

for all $x, y \in E$. Fix $b \in E$ and $c \in \mathbb{R}$. Consider the map

$$f(x) = \langle x, L(x) \rangle + \langle x, b \rangle + c.$$

Show that f is differentiable on E and compute its differential and gradient.

2) Fix an integer $K \geq 1$, vectors $\beta_1, \dots, \beta_K \in E$, constants $\beta_{1,0}, \dots, \beta_{K,0} \in \mathbb{R}$ as well as $\alpha_0, \alpha_1, \dots, \alpha_K \in \mathbb{R}$. Consider finally a differentiable function $\sigma : \mathbb{R} \to \mathbb{R}$. Define $f : E \to \mathbb{R}$ by

$$f(x) = \alpha_0 + \sum_{k=1}^{K} \alpha_k \, \sigma\left(\langle x, \beta_k \rangle + \beta_{k,0} \right).$$

Compute the differential and gradient of f.

Solution.

1) Consider $x, h \in E$. Then, by bilinearity of the scalar product and linearity of L:

$$f(x+h) = f(x) + \underbrace{\left(\langle h, L(x) \rangle + \langle x, L(h) \rangle + \langle h, b \rangle \right)}_{=: a(h)} + \underbrace{\langle h, L(h) \rangle}_{=: b(h)}$$

By symmetry of L, we clearly have that

$$a(h) = \langle h, 2L(x) + b \rangle$$

which is clearly linear in h.

Then, we observe that

$$\frac{b(h)}{\|h\|_E} = \frac{\langle h, L(h)\rangle}{\|h\|_E}$$

$$\leqslant \|L(h)\|_E \qquad \text{(Cauchy - Schwarz)}$$

$$\leqslant \|L\|_{op} \, \|h\|_E$$

where $\|L\|_{op} := \sup_{h \neq 0} \frac{\|L(h)\|_E}{\|h\|_E} < +\infty$ is the operator norm of L. In particular, we see that

$\frac{b(h)}{\|h\|_E} \xrightarrow[h \to 0]{} 0$. Hence, by definition (and uniqueness) of the differential, we deduce that

$$d_x f(h) = \langle h, 2L(x) + b\rangle$$

and therefore

$$\nabla f(x) = 2L(x) + b.$$

2) Introduce $g_k : E \to \mathbb{R}$ defined by $g_k(x) = \langle x, \beta_k\rangle + \beta_{k,0}$. Then

$$f(x) = \alpha_0 + \sum_{k=1}^{K} \alpha_k \, \sigma(g_k(x)).$$

We deduce that

$$d_x f(h) = \sum_{k=1}^{K} \alpha_k \, d_x(\sigma \circ g_k)(h).$$

By the chain rule,

$$d_x(\sigma \circ g_k)(h) = d_{g_k(x)}\sigma\left(d_x g_k(h)\right).$$

But

$$d_x g_k(h) = \langle h, \beta_k \rangle$$

and

$$\forall u, v \in \mathbb{R} : \quad d_u \sigma(v) = \lim_{\substack{t \to 0 \\ t \neq 0}} \frac{\sigma(u+tv) - \sigma(u)}{t}$$
$v \neq 0$

$$= v . \lim_{\substack{t \to 0 \\ t \neq 0}} \frac{\sigma(u+tv) - \sigma(u)}{tv}$$

$$= v . \lim_{\substack{\varepsilon \to 0 \\ \varepsilon \neq 0}} \frac{\sigma(u+\varepsilon) - \sigma(u)}{\varepsilon}$$

$$= v \, \sigma'(u).$$

Combining all this, we obtain that:

$$d_x f(h) = \sum_{k=1}^{K} \alpha_k . \langle h, \beta_k \rangle \, \sigma'(\langle x, \beta_k \rangle + \beta_{k,0})$$

$$= \langle h, \sum_{k=1}^{K} \alpha_k \sigma'(\langle x, \beta_k \rangle + \beta_{k,0}) \beta_k \rangle$$

which brings

$$\nabla f(x) = \sum_{k=1}^{K} \alpha_k \sigma'(\langle x, \beta_k \rangle + \beta_{k,0}) \beta_k$$