Mathematical Concepts 1

Exercise 1: Gradient

Consider the bivariate function $f : \mathbb{R}^2 \to \mathbb{R}, (x_1, x_2) \mapsto x_1^2 + 0.5x_2^2 + x_1x_2$.

(a) Show that f is smooth (as defined in the lecture).

(b) Find the direction of greatest increase of f at $\mathbf{x} = (1, 1)$.

(c) Find the direction of greatest decrease of f at $\mathbf{x} = (1, 1)$.

(d) Find a direction in which f does not instantly change at $\mathbf{x} = (1, 1)$.

(e) Assume there exists a differentiable parametrization of a curve $\tilde{\mathbf{x}} : \mathbb{R} \to \mathbb{R}^2, t \mapsto \tilde{\mathbf{x}}(t)$ such that $\forall t \in \mathbb{R}$: $f(\tilde{\mathbf{x}}(t)) = f(1, 1)$. Show that at each point of the curve $\tilde{\mathbf{x}}$ the tangent line $\frac{\partial \tilde{\mathbf{x}}}{\partial t}$ is perpendicular to the gradient $\nabla f(\tilde{\mathbf{x}})$.

(f) Interpret (d), (e) geometrically

Exercise 2: Convexity

Consider two convex functions $f, g : \mathbb{R} \to \mathbb{R}$.

(a) Show that $f + g : \mathbb{R} \to \mathbb{R}, x \mapsto f(x) + g(x)$ is convex.

(b) Now, assume that g is additionally non-decreasing, i.e., $g(y) \geq g(x)\ \forall x \in \mathbb{R}, \forall y \in \mathbb{R}$ with $y > x$. Show that $g \circ f$ is convex.

Exercise 3: Taylor polynomials

Consider the bivariate function $f : \mathbb{R}^2 \to \mathbb{R}, (x_1, x_2) \mapsto \exp(\pi \cdot x_1) - \sin(\pi \cdot x_2) + \pi \cdot x_1 \cdot x_2$

(a) Compute the gradient of f for an arbitrary $\mathbf{x}$.

(b) Compute the Hessian of f for an arbitrary $\mathbf{x}$.

(c) State the first order taylor polynomial $T_{1,\mathbf{a}}(\mathbf{x})$ expanded around the point $\mathbf{a} = (0, 1)$.

(d) State the second order taylor polynomial $T_{2,\mathbf{a}}(\mathbf{x})$ expanded around the point $\mathbf{a} = (0, 1)$.

(e) Determine if $T_{2,\mathbf{a}}$ is a convex function.

Mathematical Concepts 1

Solution 1:
Gradient

(a) The gradient $\nabla f(\mathbf{x}) = (2x_1 + x_2, x_2 + x_1)^\top$ is continuous $\Rightarrow f \in \mathcal{C}^1$.

(b) The direction of greatest increase is given by the gradient, i.e., $\nabla f(1,1) = (3,2)^\top$.

(c) Let $\mathbf{v} \in \mathbb{R}^2$ be a direction with fixed length $\|\mathbf{v}\|_2 = r > 0$.
The directional derivative $D_\mathbf{v} f(\mathbf{x}) = \nabla f(\mathbf{x})^\top \mathbf{v} = \|\nabla f(\mathbf{x})\|_2 \|\mathbf{v}\|_2 \cos(\theta) = \|\nabla f(\mathbf{x})\|_2 \, r \cos(\theta)$. This becomes minimal if $\theta = \pi$, i.e., if $\mathbf{v}$ points in the opposite direction of $\nabla f \Rightarrow \mathbf{v} = -\nabla f(\mathbf{x})$ if $r = \|\nabla f(\mathbf{x})\|_2$. Here, the direction of greatest decrease is given by $-\nabla f(1,1) = (-3,-2)^\top$.

(d) $D_\mathbf{v} f(\mathbf{x}) = \nabla f(1,1)^\top \mathbf{v} \overset{!}{=} 0 \Rightarrow (3,2) \cdot \mathbf{v} = 0 \iff \mathbf{v} = \alpha \cdot (-2,3)^\top$ with $\alpha \in \mathbb{R}$ and $\alpha \neq 0$.

(e) When we differentiate both sides of the equation $f(\tilde{\mathbf{x}}(t)) = f(1,1)$ w.r.t. t we arrive at $\frac{\partial f(\tilde{\mathbf{x}}(t))}{\partial t} = 0$. Via the chain rule it follows that $\underbrace{\frac{\partial f}{\partial \tilde{\mathbf{x}}}}_{=\nabla f(\tilde{\mathbf{x}})^\top} \frac{\partial \tilde{\mathbf{x}}}{\partial t} = 0$.

(f) The gradient is orthogonal to the tangent line of the level curves.

Solution 2:
Convexity

(a) Let $x, y \in \mathbb{R}$ and $t \in [0,1]$ then it holds that

$$\begin{aligned}
(f+g)(x + t(y-x)) &= f(x + t(y-x)) + g(x + t(y-x)) \\
&\leq f(x) + t(f(y) - f(x)) + g(x) + t(g(y) - g(x)) && (f, g \text{ are convex}) \\
&= f(x) + g(x) + t(f(y) + g(y) - (f(x) + g(x))) \\
&= (f+g)(x) + t((f+g)(y) - (f+g)(x)).
\end{aligned}$$

(b) Let $x, y \in \mathbb{R}$ and $t \in [0,1]$ then it holds that

$$\begin{aligned}
(g \circ f)(x + t(y-x)) &= g(f(x + t(y-x))) \\
&\leq g(f(x) + t(f(y) - f(x))) && (g \text{ is non-decreasing, } f \text{ is convex}) \\
&\leq g(f(x)) + t(g(f(y)) - g(f(x)))) && (g \text{ is convex}) \\
&= (g \circ f)(x) + t((g \circ f)(y) - (g \circ f)(x)).
\end{aligned}$$

Solution 3:
Convexity

Consider the bivariate function $f : \mathbb{R}^2 \to \mathbb{R}, (x_1, x_2) \mapsto \exp(\pi \cdot x_1) - \sin(\pi \cdot x_2) + \pi \cdot x_1 \cdot x_2$

(a) $\nabla f(\mathbf{x}) = \pi \cdot (\exp(\pi x_1) + x_2, -\cos(\pi x_2) + x_1)^\top$

(b) $\nabla^2 f(\mathbf{x}) = \pi \cdot \begin{pmatrix} \pi \exp(\pi x_1) & 1 \\ 1 & \pi \sin(\pi x_2) \end{pmatrix}$

(c) $T_{1,\mathbf{a}}(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a})^\top (\mathbf{x} - \mathbf{a}) = 1 + \pi \cdot (2,1) \cdot (x_1, x_2 - 1)^\top = 1 - \pi + 2\pi x_1 + \pi x_2$

Mathematical Concepts 2

Solution 1:
Matrix Calculus

(a) $\frac{\partial \|\mathbf{x}-\mathbf{c}\|_2^2}{\partial \mathbf{x}} = \frac{\partial \|\mathbf{u}\|_2^2}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}^\top \mathbf{u}}{\partial \mathbf{u}} \frac{\partial \mathbf{x}-\mathbf{c}}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}^\top \mathbf{I} \mathbf{u}}{\partial \mathbf{u}} (\mathbf{I} - \mathbf{0}) = \mathbf{u}^\top (\mathbf{I} + \mathbf{I}^\top) = 2(\mathbf{x} - \mathbf{c})^\top$

(b) $\frac{\partial \|\mathbf{x}-\mathbf{c}\|_2}{\partial \mathbf{x}} = \frac{\partial \sqrt{\|\mathbf{x}-\mathbf{c}\|_2^2}}{\partial \mathbf{x}} = \frac{0.5}{\sqrt{\|\mathbf{x}-\mathbf{c}\|_2^2}} \frac{\partial \|\mathbf{x}-\mathbf{c}\|_2^2}{\partial \mathbf{x}} \overset{(a)}{=} \frac{(\mathbf{x}-\mathbf{c})^\top}{\|\mathbf{x}-\mathbf{c}\|_2}$

(c) $\frac{\partial \mathbf{u}^\top \mathbf{v}}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}^\top \mathbf{I} \mathbf{v}}{\partial \mathbf{x}} = \mathbf{u}^\top \mathbf{I} \frac{\partial \mathbf{v}}{\partial \mathbf{x}} + \mathbf{v}^\top \mathbf{I}^\top \frac{\partial \mathbf{u}}{\partial \mathbf{x}} = \mathbf{u}^\top \frac{\partial \mathbf{v}}{\partial \mathbf{x}} + \mathbf{v}^\top \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$

(d) $\frac{\partial \mathbf{Y}^\top \mathbf{u}}{\partial \mathbf{x}} = \frac{\partial \begin{pmatrix} \mathbf{y}_1^\top \mathbf{u} \\ \vdots \\ \mathbf{y}_d^\top \mathbf{u} \end{pmatrix}}{\partial \mathbf{x}} \overset{(c)}{=} \begin{pmatrix} \mathbf{y}_1^\top \frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \mathbf{u}^\top \frac{\partial \mathbf{y}_1}{\partial \mathbf{x}} \\ \vdots \\ \mathbf{y}_d^\top \frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \mathbf{u}^\top \frac{\partial \mathbf{y}_d}{\partial \mathbf{x}} \end{pmatrix}$

(e) Note for $\mathbf{y} : \mathbb{R}^d \to \mathbb{R}^d, \mathbf{x} \mapsto \mathbf{y}(\mathbf{x})$ the $i-$th column of $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$ is $\frac{\partial \mathbf{y}}{\partial x_i}$. With this it follows that

$$
\begin{aligned}
\frac{\partial^2 \mathbf{u}^\top \mathbf{v}}{\partial \mathbf{x} \partial \mathbf{x}^\top} &= \frac{\partial}{\partial \mathbf{x}} \left(\frac{\partial \mathbf{u}^\top \mathbf{v}}{\partial \mathbf{x}^\top} \right) \\
&= \frac{\partial}{\partial \mathbf{x}} \left[\left(\frac{\partial \mathbf{u}^\top \mathbf{v}}{\partial \mathbf{x}} \right)^\top \right] \\
&\overset{(c)}{=} \frac{\partial (\mathbf{u}^\top \frac{\partial \mathbf{v}}{\partial \mathbf{x}} + \mathbf{v}^\top \frac{\partial \mathbf{u}}{\partial \mathbf{x}})^\top}{\partial \mathbf{x}} \\
&= \frac{\partial \left(\left(\frac{\partial \mathbf{v}}{\partial \mathbf{x}} \right)^\top \mathbf{u} + \left(\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \right)^\top \mathbf{v} \right)}{\partial \mathbf{x}} \\
&\overset{(d)}{=} \begin{pmatrix} \mathbf{u}^\top \frac{\partial^2 \mathbf{v}}{\partial x_1 \partial \mathbf{x}} + \frac{\partial \mathbf{v}}{\partial x_1}^\top \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \\ \vdots \\ \mathbf{u}^\top \frac{\partial^2 \mathbf{v}}{\partial x_d \partial \mathbf{x}} + \frac{\partial \mathbf{v}}{\partial x_d}^\top \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \end{pmatrix}^\top + \begin{pmatrix} \mathbf{v}^\top \frac{\partial^2 \mathbf{u}}{\partial x_1 \partial \mathbf{x}} + \frac{\partial \mathbf{u}}{\partial x_1}^\top \frac{\partial \mathbf{v}}{\partial \mathbf{x}} \\ \vdots \\ \mathbf{v}^\top \frac{\partial^2 \mathbf{u}}{\partial x_d \partial \mathbf{x}} + \frac{\partial \mathbf{u}}{\partial x_d}^\top \frac{\partial \mathbf{v}}{\partial \mathbf{x}} \end{pmatrix}^\top \\
&= \begin{pmatrix} \mathbf{u}^\top \frac{\partial^2 \mathbf{v}}{\partial x_1 \partial \mathbf{x}} \\ \vdots \\ \mathbf{u}^\top \frac{\partial^2 \mathbf{v}}{\partial x_d \partial \mathbf{x}} \end{pmatrix}^\top + \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \left(\frac{\partial \mathbf{v}}{\partial \mathbf{x}} \right)^\top + \frac{\partial \mathbf{v}}{\partial \mathbf{x}} \left(\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \right)^\top + \begin{pmatrix} \mathbf{v}^\top \frac{\partial^2 \mathbf{u}}{\partial x_1 \partial \mathbf{x}} \\ \vdots \\ \mathbf{v}^\top \frac{\partial^2 \mathbf{u}}{\partial x_d \partial \mathbf{x}} \end{pmatrix}^\top
\end{aligned}
$$

Solution 2:
Optimality in 1d

Let $f : [-1, 2] \to \mathbb{R}, x \mapsto \exp(x^3 - 2x^2)$

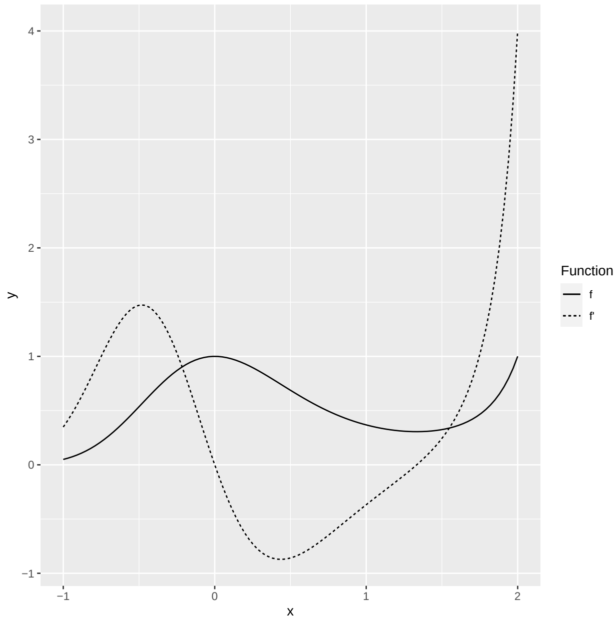
(a) $f'(x) = \exp(x^3 - 2x^2) \cdot (3x^2 - 4x)$

(b)
```
library(ggplot2)

f <- function(x) exp(x^3 - 2*x^2)
df <- function(x) f(x) * (3*x^2 - 4*x)

ggplot(data.frame(x = seq(-1, 2, by=0.005)), aes(x)) +
```

```
  geom_function(fun = f, aes(linetype = "f")) +
  geom_function(fun = df, aes(linetype = "f'"))    +
  scale_linetype_discrete(name = "Function")
```



(c) f is continuously differentiable $\Rightarrow$ candidates can only be stationary points and boundary points. Find stationary points, i.e., points where
$f'(x) = 0 \iff \underbrace{\exp(x^3 - 2x^2)}_{>0} \cdot (3x^2 - 4x) = 0 \iff 3x^2 - 4x = 0 \iff x(3x - 4) = 0.$
$\Rightarrow x_1 = 0, x_2 = 4/3$. The other candidates are boundary points, i.e., $x_3 = -1, x_4 = 2$.

(d) $f''(x) = \exp(x^3 - 2x^2) \cdot (3x^2 - 4x)^2 + \exp(x^3 - 2x^2) \cdot (6x - 4)$

(e) $f''(x_1) = \exp(0) \cdot (-4) < 0$
$\Rightarrow x_1$ is a local maximum

$f''(x_2) = \exp((4/3)^3 - 2(4/3)^2) \cdot (4) > 0$
$\Rightarrow x_2$ is a local minimum.

The boundary points x_3 and x_4 are not considered as *local* optima.

(f) $f(x_1) = \exp(0) = 1$
$f(x_2) = \exp((4/3)^3 - 2(4/3)^2) \approx 0.3057$
$f(x_3) = \exp(-3) \approx 0.05$
$f(x_4) = \exp(0) = 1$
$\Rightarrow x_1, x_4$ are global maxima. x_3 is global minimum.