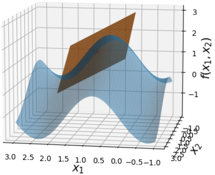
Optimization

Hessian Matrix & Taylor Series



Learning goals

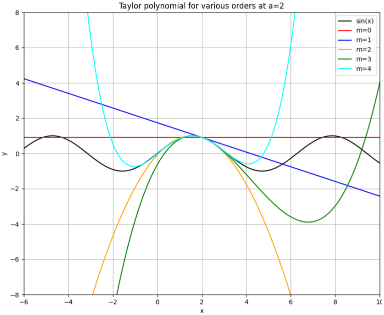
- Taylor series (Univariate)
- Hessian Matrix
- Taylor series (Multivariate)

DEFINITION TAYLOR'S THEOREM (UNIVARIATE)

Let $I \subseteq \mathbb{R}$ an open interval and $a, x \in I$ and $f \in \mathcal{C}^{m+1}(I, \mathbb{R})$. Then

$$f(x) = T_m(x, a) + R_m(x, a), \text{ with}$$

- m-th **Taylor polynomial**: $T_m(x, a) \overset{(*)}{=} \sum_{k=0}^{m} \frac{f^{(k)}(a)}{k!}(x-a)^k$
- **Remainder term**: $R_m(x, a)$ (we will cover this term later)



$^{(*)}$ $T_m(x, a) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \ldots + \frac{f^{(m)}(a)}{m!}(x-a)^m$

DEFINITION HESSIAN MATRIX

The 2nd derivative of a multivariate function $f \in \mathcal{C}^2(\mathcal{S}, \mathbb{R})$, $\mathcal{S} \subseteq \mathbb{R}^d$ (if it exists) is defined by the **Hessian** matrix

$$H(\mathbf{x}) = \nabla^2 f(\mathbf{x}) = \left(\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}\right)_{i,j=1\ldots d}$$

Example: Let $f(x_1, x_2) = sin(x_1) \cdot cos(x_2)$. Then:

$$H(\mathbf{x}) = \begin{pmatrix} \text{-}\cos(x_2) \cdot \sin(x_1) & \text{-}\cos(x_1) \cdot \cos(x_2) \\ \text{-}\cos(x_1) \cdot \sin(x_2) & \text{-}\cos(x_2) \cdot \sin(x_1) \end{pmatrix}$$

HESSIAN DESCRIBES LOCAL CURVATURE

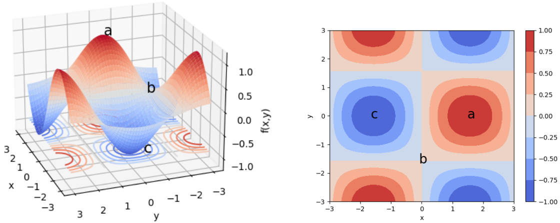
Let w.l.o.g. $A(\mathbf{x}) = \{\lambda_{1,\mathbf{x}}, ..., \lambda_{d,\mathbf{x}}\}$ be Eigenspectrum with $\lambda_{1,\mathbf{x}} \leq \lambda_{2,\mathbf{x}} \leq ... \leq \lambda_{d,\mathbf{x}}$ of $H(\mathbf{x})$; let $\boldsymbol{v}_{i,\mathbf{x}}$ define the respective Eigenvectors. We can read from it:

- $\boldsymbol{v}_d$ points into the direction of largest curvature
- $\boldsymbol{v}_1$ points into the direction of smallest curvature

Example (continued): $H(\mathbf{x}) = \begin{pmatrix} \text{-}\cos(x_2) \cdot \sin(x_1) & \text{-}\cos(x_1) \cdot \sin(x_2) \\ \text{-}\cos(x_1) \cdot \sin(x_2) & \text{-}\cos(x_2) \cdot \sin(x_1) \end{pmatrix}$.

- $H(a)$, $a = (\frac{-\pi}{2}, 0)$: $\lambda_1 = \lambda_2 = 1$; $v_1 = (1, 0)^\top$, $v_2 = (0, 1)^\top$
- $H(b)$, $b = (0, \frac{-\pi}{2})$: $\lambda_1 = -1, \lambda_2 = 1$; $v_1 = (-1, 1)^\top$, $v_2 = (1, 1)^\top$
- $H(c), c = (\frac{-\pi}{2}, 0), : \lambda_1 = \lambda_2 = 1; v_1 = (1, 0)^\top, v_2 = (0, 1)^\top$

HESSIAN DESCRIBES LOCAL CURVATURE



REMAINDER TERM

$$f(x) = T_m(x, a) + R_m(x, a)$$

How close is $T_m(x, a)$ to $f(x)$?

- Exact representation of $R_m(x, a)$:

$$R_m(x, a) := \int_a^x \frac{f^{(k+1)}(t)}{k!}(x - t)^k \, \mathrm{d}t$$

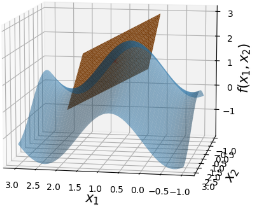
(integral form of remainder; alternative formulas exist, but are not covered here.)

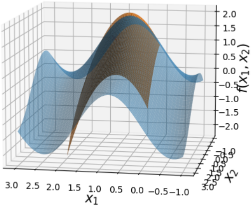
- In order of magnitude:

$$R_m(\boldsymbol{a}) \in \mathcal{O}(\|\mathbf{x} - \boldsymbol{a}\|^m) \text{ for } \mathbf{x} \to \boldsymbol{a}$$

REMAINDER TERM

- Higher m gives a better approximation
- The m^{th} order taylor series is the best m^{th} order approximation to $f(\mathbf{x})$ near $\boldsymbol{a}$





Consider $T_2(\mathbf{x}, \boldsymbol{a}) = f(\boldsymbol{a}) + \nabla f(\boldsymbol{a})^\top(\mathbf{x} - \boldsymbol{a}) + \frac{1}{2}(\mathbf{x} - \boldsymbol{a})^\top \boldsymbol{H}(\boldsymbol{a})(\mathbf{x} - \boldsymbol{a})$. The first term ensures the **value** of T_2 and f match at $\boldsymbol{a}$. The second term ensures the **slopes** of T_2 and f match at $\boldsymbol{a}$. The third term ensures the **curvature** of T_2 and f match at $\boldsymbol{a}$.

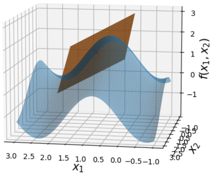
MULTIVARIATE TAYLOR SERIES

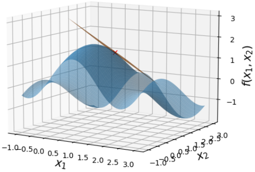
Taylor's theorem (1st order):

$$f(\mathbf{x}) = \underbrace{f(\boldsymbol{a}) + \nabla f(\boldsymbol{a})^\top(\mathbf{x} - \boldsymbol{a})}_{T_1(\mathbf{x},\boldsymbol{a})} + R_1(\mathbf{x}, a)$$

Example: $f(\mathbf{x}) = \sin(2x_1) + \cos(x_2)$, $\boldsymbol{a} = (1, 1)^\top$. Since $\nabla f(\mathbf{x}) = \begin{pmatrix} 2 \cdot \cos(2x_1) \\ -\sin(x_2) \end{pmatrix}$

$$\begin{aligned} f(\mathbf{x}) &= T_1(\mathbf{x}) + R_1(\mathbf{x}, \boldsymbol{a}) = f(\boldsymbol{a}) + \nabla f(\boldsymbol{a})^\top(\mathbf{x} - \boldsymbol{a}) + R_1(\mathbf{x}, \boldsymbol{a}) \\ &= \sin(2) + \cos(2) + (2 \cdot \cos(2), -\sin(1)) \begin{pmatrix} x_1 - 1 \\ x_2 - 1 \end{pmatrix} + R_1(\mathbf{x}, \boldsymbol{a}) \end{aligned}$$





MULTIVARIATE TAYLOR SERIES

Taylor's theorem (2nd order):

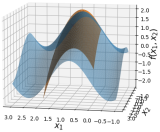
$$f(\mathbf{x}) = \underbrace{f(\boldsymbol{a}) + \nabla f(\boldsymbol{a})^\top(\mathbf{x} - \boldsymbol{a}) + \frac{1}{2}(\mathbf{x} - \boldsymbol{a})^\top \boldsymbol{H}(\boldsymbol{a})(\mathbf{x} - \boldsymbol{a})}_{T_2(\mathbf{x},\boldsymbol{a})} + R_2(\mathbf{x}, a)$$

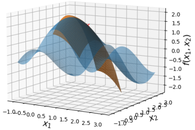
Example (continued): $f(\mathbf{x}) = \sin(2x_1) + \cos(x_2)$, $\boldsymbol{a} = (1, 1)^\top$. Since

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2 \cdot \cos(2x_1) \\ -\sin(x_2) \end{pmatrix} \text{ and } H(\mathbf{x}) = \begin{pmatrix} -4\sin(2x_1) & 0 \\ 0 & -\cos(x_2) \end{pmatrix}$$

we get

$$f(\mathbf{x}) \quad = \quad T_1(\mathbf{x}, \boldsymbol{a}) + \frac{1}{2}\begin{pmatrix} x_1 - 1 \\ x_2 - 1 \end{pmatrix}^\top \begin{pmatrix} -4\sin(2) & 0 \\ 0 & -\cos(1) \end{pmatrix} \begin{pmatrix} x_1 - 1 \\ x_2 - 1 \end{pmatrix} + R_2(\mathbf{x}, \boldsymbol{a})$$





MULTIVARIATE TAYLOR SERIES

What can be written down nicely for first and second order Taylor Series is (notationally) a bit more cumbersome for general k.

Let $f : \mathbb{R}^d \to \mathbb{R}$, $f \in \mathcal{C}^k$ at $\boldsymbol{a} \in \mathbb{R}^d$. Then

$$f(x) = T_m(\mathbf{x}, \boldsymbol{a}) + R_m(\mathbf{x}, \boldsymbol{a}), \text{ with}$$

$$T_m(\mathbf{x}, \boldsymbol{a}) = \sum_{|\boldsymbol{\alpha}| \leq k} \frac{D^{\boldsymbol{\alpha}} f(\boldsymbol{a})}{\boldsymbol{\alpha}!} (\mathbf{x} - \boldsymbol{a})^{\boldsymbol{\alpha}} \text{ and } \lim_{\mathbf{x} \to \boldsymbol{a}} R_m(\mathbf{x}, \boldsymbol{a}) = 0$$

with $\boldsymbol{\alpha} \in \mathbb{N}^d$ and the multi-index notation

- $|\boldsymbol{\alpha}| = \alpha_1 + \cdots + \alpha_d$
- $\boldsymbol{\alpha}! = \alpha_1! \cdots \alpha_d!$
- $\mathbf{x}^{\boldsymbol{\alpha}} = x_1^{\alpha_1} \cdots x_d^{\alpha_d}$
- $D^{\boldsymbol{\alpha}} f = \frac{\partial^{|\boldsymbol{\alpha}|} f}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}$

MULTIVARIATE TAYLOR SERIES

Let's check for $f : \mathbb{R}^2 \to \mathbb{R}$ and $k = 1$. We have for $|\alpha| \leq 1$:

- $\alpha_1 = 0, \alpha_2 = 0$: $|\boldsymbol{\alpha}| = 0, \boldsymbol{\alpha}! = 1, \mathbf{x}^{\boldsymbol{\alpha}} = 1, D^{\boldsymbol{\alpha}} f = 1$
- $\alpha_1 = 1, \alpha_2 = 0$: $|\boldsymbol{\alpha}| = 1, \boldsymbol{\alpha}! = 1, \mathbf{x}^{\boldsymbol{\alpha}} = x_1, D^{\boldsymbol{\alpha}} f = \frac{\partial f}{\partial x_1}$
- $\alpha_1 = 0, \alpha_2 = 1$: $|\boldsymbol{\alpha}| = 1, \boldsymbol{\alpha}! = 1, \mathbf{x}^{\boldsymbol{\alpha}} = x_2, D^{\boldsymbol{\alpha}} f = \frac{\partial f}{\partial x_2}$

and therefore:

$$
\begin{aligned}
T_m(\mathbf{x}, \boldsymbol{a}) &= \sum_{|\boldsymbol{\alpha}| \leq k} \frac{D^{\boldsymbol{\alpha}} f(\boldsymbol{a})}{\boldsymbol{\alpha}!} (\mathbf{x} - \boldsymbol{a})^{\boldsymbol{\alpha}} \\
&= \frac{1 \cdot f(\boldsymbol{a})}{1} \cdot 1 + \frac{\partial f}{\partial x_1}(\boldsymbol{a})(x_1 - a_1) + \frac{\partial f}{\partial x_2}(\boldsymbol{a})(x_2 - a_2) \\
&= f(a) + \begin{pmatrix} \frac{\partial f}{\partial x_1}(\boldsymbol{a}) \\ \frac{\partial f}{\partial x_2}(\boldsymbol{a}) \end{pmatrix}^\top \begin{pmatrix} x_1 - a_1 \\ x_2 - a_2 \end{pmatrix} = f(\boldsymbol{a}) + \nabla f(\boldsymbol{a})^\top (\mathbf{x} - \boldsymbol{a}).
\end{aligned}
$$