FuVar Notes

Naoki Pross – naoki.pross@ost.ch

Spring Semseter 2021

1 Preface

These are just my personal notes of the FuVar course, and definitively not a rigorously constructed mathematical text. The good looking LaTeX typesetting may trick you into thinking it is rigorous, but really, it is not.

2 Derivatives of vector valued scalar functions

Definition 1 (Partial derivative)**.** A vector values function $f : \mathbb{R}^m \to \mathbb{R}$, with $\mathbf{v} \in \mathbb{R}^m$, has a partial derivative with respect to v_i defined as

$$\partial_{v_i} f(\mathbf{v}) = f_{v_i}(\mathbf{v}) = \lim_{h \to 0} \frac{f(\mathbf{v} + h\mathbf{e}_j) - f(\mathbf{v})}{h}$$

Proposition 1. Under some generally satisfied conditions (continuity of n-th order partial derivatives) Schwarz's theorem states that it is possible to swap the order of differentiation.

$$\partial_x \partial_y f(x, y) = \partial_y \partial_x f(x, y)$$

Definition 2 (Linearization)**.** A function $f : \mathbb{R}^m \to \mathbb{R}$ has a linearization g at $\mathbf{x}_0$ given by

$$g(\mathbf{x}) = f(\mathbf{x}_0) + \sum_{i=1}^{m} \partial_{x_i} f(\mathbf{x}_0)(x_i - x_{i,0}),$$

if all partial derviatives are defined at $\mathbf{x}_0$.

Theorem 1 (Propagation of uncertanty)**.** Given a measurement of m values in a vector $\mathbf{x} \in \mathbb{R}^m$ with values given in the form $x_i = \bar{x}_i \pm \sigma_{x_i}$, a linear approximation the error of a dependent variable y is computed with

$$y = \bar{y} \pm \sigma_y \approx f(\bar{\mathbf{x}}) \pm \sqrt{\sum_{i=1}^{m} \left(\partial_{x_i} f(\bar{\mathbf{x}}) \sigma_{x_i}\right)^2}$$

Definition 3 (Gradient vector)**.** The *gradient* of a function $f(\mathbf{x}), \mathbf{x} \in \mathbb{R}^m$ is a vector containing the derivatives in each direction.

$$\nabla f(\mathbf{x}) = \sum_{i=1}^{m} \partial_{x_i} f(\mathbf{x}) \mathbf{e}_i = \begin{pmatrix} \partial_{x_1} f(\mathbf{x}) \\ \vdots \\ \partial_{x_m} f(\mathbf{x}) \end{pmatrix}$$

Definition 4 (Directional derivative)**.** A function $f(\mathbf{x})$ has a directional derivative in direction $\mathbf{r}$ (with $|\mathbf{r}| = 1$) given by

$$\frac{\partial f}{\partial \mathbf{r}} = \nabla_{\mathbf{r}} f = \mathbf{r} \cdot \nabla f$$

Theorem 2. The gradient vector always points towards *the direction of steepest ascent.*

2.1 Methods for maximization and minimization problems

Method 1 (Find stationary points)**.** Given a function $f : D \subseteq \mathbb{R}^m \to \mathbb{R}$, to find its maxima and minima we shall consider the points

- that are on the boundary of the domain ∂D,
- where the gradient ∇f is not defined,
- that are stationary, i.e. where $\nabla f = \mathbf{0}$.

Method 2 (Determine the type of stationary point for 2 dimensions)**.** Given a scalar function of two variables $f(x, y)$ and a stationary point $\mathbf{x}_s$ (where $\nabla f(\mathbf{x}_s) = \mathbf{0}$), we define the *discriminant*

$$\Delta = \partial_x^2 f \partial_y^2 f - \partial_y \partial_x f$$

- if $\Delta > 0$ then $\mathbf{x}_s$ is an extrema, if $\partial_x^2 f(\mathbf{x}_s) < 0$ it is a maximum, whereas if $\partial_x^2 f(\mathbf{x}_s) > 0$ it is a minimum;
- if $\Delta < 0$ then $\mathbf{x}_s$ is a saddle point;
- if $\Delta = 0$ we need to analyze further.

Remark 1. The previous method is obtained by studying the second directional derivative $\nabla_{\mathbf{r}} \nabla_{\mathbf{r}} f$ at the stationary point in direction of a vector $\mathbf{r} = \mathbf{e}_1 \cos(\alpha) + \mathbf{e}_2 \sin(\alpha)$

Definition 5 (Hessian matrix)**.** Given a function $f : \mathbb{R}^m \to \mathbb{R}$, the square matrix whose entry at the i-th row and j-th column is the second derivative of f first with respect to x_j and then to x_i is know as the *Hessian* matrix. $(\mathrm{H}_f)_{i,j} = \partial_{x_i}\partial_{x_j} f$ or

$$\mathrm{H}_f = \begin{pmatrix} \partial_{x_1}\partial_{x_1} f & \cdots & \partial_{x_1}\partial_{x_m} f \\ \vdots & \ddots & \vdots \\ \partial_{x_m}\partial_{x_1} f & \cdots & \partial_{x_m}\partial_{x_m} f \end{pmatrix}$$

Because (almost always) the order of differentiation does not matter, it is a symmetric matrix.

Method 3 (Determine the type of stationary point in higher dimensions)**.** Given a scalar function of two variables $f(x, y)$ and a stationary point $\mathbf{x}_s$ (where $\nabla f(\mathbf{x}_s) = \mathbf{0}$), we compute the Hessian matrix $\mathrm{H}_f(\mathbf{x}_s)$. Then we compute its eigenvalues $\lambda_1, \ldots, \lambda_m$ and

- if all $\lambda_i > 0$, the point is a minimum;
- if all $\lambda_i < 0$, the point is a maximum;
- if there are both positive and negative eigenvalues, it is a saddle point.

In the other cases, when there are $\lambda_i \leq 0$ and/or $\lambda_i \geq 0$ further analysis is required.

Remark 2. Recall that to compute the eigenvalues of a matrix, one must solve the equation $(\mathrm{H} - \lambda \mathrm{I})\mathbf{x} = \mathbf{0}$. Which can be done by solving the characteristic polynomial $\det(\mathrm{H} - \lambda \mathrm{I}) = 0$ to obtain $\dim(\mathrm{H})$ λ_i, which when plugged back in result in a overdetermined system of equations.

Method 4 (Quickly find the eigenvalues of a 2×2 matrix)**.** Let

$$m = \frac{1}{2} \operatorname{tr} \mathrm{H} = \frac{a + d}{2} \text{ and } p = \det \mathrm{H} = ad - bc,$$

then

$$\lambda = m \pm \sqrt{m^2 - p}.$$

Method 5 (Search for a constrained extremum in 2 dimensions)**.** Let $n(x, y) = 0$ be a constraint in the search of the extrema of a function $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$. To find the extrema we look for points

- on the boundary $\mathbf{u} \in \partial D$ where $n(\mathbf{u}) = 0$;
- $\mathbf{u}$ where the gradient either does not exist or is $\mathbf{0}$, and satisfy $n(\mathbf{u}) = 0$;
- that solve the system of equations

$$\begin{cases} \partial_x f(\mathbf{u}) \cdot \partial_y n(\mathbf{u}) = \partial_y f(\mathbf{u}) \cdot \partial_x n(\mathbf{u}) \\ n(\mathbf{u}) = 0 \end{cases}$$

Method 6 (Search for a constrained extremum in higher dimensions, method of Lagrange multipliers)**.** We wish to find the extrema of $f : D \subseteq \mathbb{R}^m \to \mathbb{R}$ under $k < m$ constraints $n_1 = 0, \cdots, n_k = 0$. For that we consider the following points:

- Points on the boundary $\mathbf{u} \in \partial D$ that satisfy $n_i(\mathbf{u}) = 0$ for all $1 \leq i \leq k$,
- Points $\mathbf{u} \in D$ where either
  - any of $\nabla f, \nabla n_1, \ldots, \nabla n_k$ do not exist, or
  - $\nabla n_1, \ldots, \nabla n_k$ are linearly *dependent*,

  and that satisfy $0 = n_1(\mathbf{u}) = \ldots = n_k(\mathbf{u})$.
- Points that solve the system of $m + k$ equations

$$\begin{cases} \nabla f(\mathbf{u}) = \sum_{i=1}^{k} \lambda_k \nabla n_i(\mathbf{u}) & (m\text{-dimensional}) \\ n_i(\mathbf{u}) = 0 & \text{for } 1 \leq i \leq k \end{cases}$$

  The λ values are known as *Lagrange multipliers.*

License