Varitational Principles

Notes by Finley Cooper

11th October 2025

Contents

1 Calculus on $\mathbb{R}^n$

Consider $f : \mathbb{R}^n \to \mathbb{R}$. We'll write $(x_1, x_2, \ldots, x_n)$ as $\mathbf{x}$. Assume f is sufficiently differentiable (at least in C^2).

1.1 Stationary points

> **Definition.** (Stationary point) A point $\mathbf{a} \in \mathbb{R}^n$ is a *stationary point* of f if and only if $(\nabla f)(\mathbf{a}) = 0$

Taylor expand about such an $\mathbf{a}$:

$$f(\mathbf{x}) = f(\mathbf{a}) + \sum_{i=1}^{n}(x_i - a_i)(\nabla_i f)(\mathbf{a}) + \frac{1}{2}\sum_{i,j}(x_i - a_i)(x_j - a_j)H_{ij}(\mathbf{a}) + O(|\mathbf{x} - \mathbf{a}|^3).$$

Since $\mathbf{a}$ is stationary point, the linear term is zero, hence the behaviour of f around $\mathbf{a}$ is determined by the Hessian.

$$H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j} = H_{ji}$$

We can assume *wlog* that $\mathbf{a} = 0$ by a translation, so

$$f(\mathbf{x}) - f(\mathbf{a}) = \frac{1}{2}x_i H_{ij} x_j + O(|\mathbf{x}|^3).$$

Since H_{ij} is symmetric, we can diagonalise H by a rotation matrix so $x_i = R_{ij}x'_j$ we choose R so that $H' = R^H R = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Since H is symmetric, $\lambda_i \in \mathbb{R}$ for all i. So the series becomes

$$f(\mathbf{x}) - f(\mathbf{a}) = \frac{1}{2}\sum_{i=1}^{n}\lambda_i(x'_i)^2 + O(|\mathbf{x}|^3).$$

If $\lambda_i > 0$ for all i then $f(\mathbf{x}) > f(\mathbf{a})$ for small enough $\mathbf{x}$ so we get a local minimum.
If $\lambda_i < 0$ for all i then $f(\mathbf{x}) < f(\mathbf{a})$ for small enough $\mathbf{x}$ so we get a local maximum.
Otherwise if all non-zero we get a saddle point, and if some are zero then we get a degenerate stationary point, so we need to look at the next term in the Taylor expansion to describe the behaviour at the point.

If we're working in $\mathbb{R}^2$ we can work with determinates and traces since $\det H = \lambda_1\lambda_2$ and $\text{tr}(H) = \lambda_1 + \lambda_2$ so

(i) If $\det H > 0$, $\text{tr } H > 0 \implies$ local minimium.

(ii) If $\det H > 0$, $\text{tr } H < 0 \implies$ local maximium.

(iii) If $\det < 0$ we get a saddle point.

(iv) If $\det = 0$ we're in the degenerate case.

Remark. A local minimium is not a global minimium. There might be some other stationary point which is the global minimium, the minimium may also occur on the boundary of the domain or the function might be unbounded from below in the domain and not achieve a global minimium. We can make the same statements for maximiums.

3

Let's see an example with the function $f(x, y) = x^3 + y^3 - 3xy$.

$$\nabla f = (3x^2 - 3y, 3y^2 - 3x)$$

So $\nabla f = 0 \iff x^2 = y$ and $y^2 = x$. This gives the solutions $(0, 0), (1, 1)$ only.

$$H = \begin{pmatrix} 6x & -3 \\ -3 & 6y \end{pmatrix}$$

So

$$\det H = 9(4xy - 1) \text{ and } \operatorname{tr} H = 6(x + y).$$

Hence our point $(1, 1)$ is a local minimium and the point $(0, 0)$ is a saddle point.
We can work out the eigenvectors of the point $(0, 0)$ which are

$$\lambda_1 = -3 \quad e_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\lambda_2 = 3 \quad e_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

So it is a maximium along the line $y = x$ and a minimium along the line $y = -x$.
For this example there does not exist a global minimium or maximium since f increases/decreases
without bound on the domain.

1.2   Convex functions

Definition. (Convex set) A set S is *convex* if and only if $\forall x, y \in S$ and $\forall t \in (0, 1)$ we
have that $(1 - t)x + ty \in S$.

Now we'll define convexity for functions. Let f be a function domain with domain $D(f) \subseteq \mathbb{R}^n$.
The graph of f is the surface $z = f(\mathbf{x})$ in $\mathbb{R}^{n+1}$ with coordinates $(\mathbf{x}, z)$. A chord of f is a line
section joining two points of its graph.

Definition. (Convex function) A function $f : S \to \mathbb{R}$ is convex if
  (i) S is a convex set
  (ii) The graph of f lies below on or below all of its chords i.e we have that

$$f((1 - t)\mathbf{x} + t\mathbf{y}) \leq (1 - t)f(\mathbf{x}) + tf(\mathbf{y})$$

for all $t \in (0, 1)$ and for all $\mathbf{x}, \mathbf{y} \in S$.
We say that f is *strictly convex* if the condition in (ii) is strict when $\mathbf{x} \neq \mathbf{y}$.

Remark. See that we require the property (i) to exists for property (ii) to be well-defined.

Lemma. f is strictly concave if and only if f is strictly convex.

Proof. Trivial. □