Mathematics and Statistics Brush-up: Differentiation and Static Optimization

Albert Rodriguez Sala

UAB and Barcelona GSE

Differentiation

The derivative

The **derivative** of a function $f : \mathbb{R} \to \mathbb{R}$ at a point $x \in \mathbb{R}$ is a real number that describes the instantaneous change of the value f as x changes:

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

Local minimum/maximum

A point x^* is said to be **local maximum** or **maximizer** if there exists $\delta > 0$ such that $f(x^*) \geq f(x) \forall x \in N_\delta(x^*)$, that is $\forall x \in (x^* - \delta, x^* + \delta)$. A **local minimum** or **minimizer** is defined in an analogous way: $x^* := \{\exists \delta : f(x^*) \leq f(x) \forall x \in N_\delta(x^*)\}$

Proposition: Maximum/minimum in open sets

Proposition: Let O be an open subset of $\mathbb{R}$, and let $f : O \to \mathbb{R}$ a function with derivative at x, and let x^* be a local maximum (minimum). Then, $f'(x^*) = 0$

Proposition: Maximum/minimum in open sets

Proof.
Suppose x^* is a local maximum (for a local minimum is analogous). Then, $f(x^*+h) - f(x^*) \forall h > 0$, small enough, $(|h| < \delta$. Taking limits as h goes fo 0 from the right, it must be that

$$\lim_{h \to 0^+} \frac{f(x^*+h) - f(x^*)}{h} \leq 0 \quad \forall h \in (0, \delta)$$

Moreover, taking limits from the left

$$\lim_{h \to 0^-} \frac{f(x^*+h) - f(x^*)}{h} \geq 0 \quad \forall h \in (-\delta, 0)$$

Thus, since $f(x)$ has a derivative at x, then, it must be that:

$$\lim_{h \to 0^+} \frac{f(x^*+h) - f(x^*)}{h} = \lim_{h \to 0^-} \frac{f(x^*+h) - f(x^*)}{h} = f'(x^*) = 0$$

Properties of derivatives

- If $f(x) = c$ for all x. Then, $f'(x) = 0$
- If $f(x) = x$ then, $f'(x) = 1$
- $[f(x) \underset{(-)}{+} g(x)]' = f'(x) \underset{(-)}{+} g'(x)$ (Sum rule)
- $[kf(x)]' = kf'(x)$
- $[f(x) \cdot g(x)]' = f'(x) \cdot g(x) + g'(x) \cdot f(x)$ (Product rule)
- $\left[\frac{f(x)}{g(x)}\right]' = \frac{f'(x)g(x) - g'(x)f(x)}{[g(x)]^2}$ (quotient rule)
- $(f(g(x))' = f'(g(x)) \cdot g'(x)$ Equivalently, $\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$ (Chain rule)
- $\left[\frac{1}{f(x)}\right]' = -\frac{f'(x)}{[]f(x)]^2}$

Theorem
(Mean value theorem) *Let* $f : [a, b] \to \mathbb{R}$ *be a continuous function and differentiable on* (a, b)*. Then there exists a point* $c \in [a, b]$ *such that* $f'(c) = \frac{f(b)-f(a)}{b-a}$

Taylor Polynomials

When we approximate a function by its tangent line, we suffer an error. We can reduce such approximation error by approximating the function by a more sophisticated object than the tangent line: a polynomial of order higher than one.

If $f(a), f'(a), \ldots f^{(n)}(a)$ all exist, the **nth Taylor polynomial** of the function f at the point a is

$$T_n(x) = \sum_{k=0}^{n} \frac{f^k(a)}{k!}(x-a)^k$$
$$= f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \ldots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

Taylor series

If the function f has derivatives of all orders a, the **Taylor series** is a series expansion of a function about a point a given by

$$\begin{aligned} f(x) &= \sum_{n=0}^{\infty} T_n(x) \\ &= \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n \\ &= f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \ldots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \ldots \end{aligned}$$

Then, we say we do a **Taylor approximation of order nth** when we approximate $f(x)$ by computing a Taylor expansion till its *nth* polynomial.

Derivatives of real multivariable functions

Now suppose the case of of a real value multivariate function $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Then,

- The **directional derivative** of $f : \mathbb{R}^n \rightarrow \mathbb{R}$ at point x in the direction of $\mathbf{u}$ is defined by

$$D_u f(x) = \lim_{\alpha \rightarrow 0} \frac{f(x + \alpha \mathbf{u}) - f(x)}{\alpha}$$

  where $\mathbf{u}$ is a vector $\mathbf{u} = (u_1, u_2, \ldots, u_n)$, that gives the direction (set of coordinates) to where to move.

Derivatives of real multivariable functions: the partial derivative

- The **Partial derivative** of f with respect to the $i-th$ argument, x_i, is defined by

$$\frac{\partial f(x)}{\partial x_i} = f_i(x) = D_{x_i} f(x) = Df(x, \mathbf{u_i}) \quad \text{with } \mathbf{u_i} = (\underset{1}{0}, \ldots, 0, \underset{i}{1}, 0, \ldots, \underset{n}{0})$$

Derivatives of real multivariable functions: The Jacobian

- The **Jacobian** of f is defined as the vector $(1 \times n)$ formed by all the partial derivatives of the function

$$\mathcal{J}_x(f) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} & \frac{\partial f(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix}$$

Derivatives of real multivariable functions: The Hessian

- the **Hessian** of f is defined as the matrix $(n \times n)$ formed by all the second-order partial derivatives of the function

$$\mathcal{H}_x(f) = \begin{bmatrix} \frac{\partial 2f}{\partial 2x_1} & \cdots & \frac{\partial f}{\partial x_1 \partial x_m} \\ \vdots & \ddots & \\ \frac{\partial 2f}{\partial x_m \partial x_1} & \cdots & \frac{\partial f}{\partial^2 x_m} \end{bmatrix}$$

The Hessian allows us to determine the convexity/concavity of the function f

1. A function $f : \mathbb{R}^n \to \mathbb{R}$ is concave if $\mathcal{H}_x(f)$ is negative semidefinite.
2. A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if $\mathcal{H}_x(f)$ is positive semidefinite.
3. A function $f : \mathbb{R}^n \to \mathbb{R}$ is strictly concave if $\mathcal{H}_x(f)$ is negative definite.
4. A function $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex if $\mathcal{H}_x(f)$ is positive definite.

Evaluating ccv/cvx with the Hessian

Then, the following equivalences are true

$$
\begin{aligned}
f \text{ concave} \iff \text{Hessian semidefinite negative} &\iff \text{its latent roots are } \lambda_i \leq 0 \\
&\longleftrightarrow \text{principal minors order } k\text{, have sign } -1^k \\
f \text{ convex} \iff \text{Hessian semidefinite positive} &\iff \text{its latent roots are } \lambda_i \geq 0 \\
&\longleftrightarrow \text{principal minors are } \geq 0
\end{aligned}
$$

Hessian matrix

A Hessian matrix will be symmetric if the second partial derivatives are continuous (Schwarz's theorem).

Differentiability

On the one hand, the derivative of a function $f : \mathbb{R} \longrightarrow \mathbb{R}$ at a point $x \in \mathbb{R}$ is defined as a real number that describes the instantaneous change of the value of f as x changes. On the other hand, the notion of derivative is tightly linked to the line that best approximates f near x.

Let f be a function from a vector space $V \subset \mathbb{R}^n$ to another vector space $W \subset \mathbb{R}$. We say that f is a linear map if for any two vectors $\mathbf{x}, \mathbf{y} \in V$ and any scalar $\alpha \in \mathbb{R}$, two conditions are satisfied:

1. $f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$ (additivity)
2. $f(\alpha \mathbf{x}) = \alpha f(\mathbf{x})$ (homogeneity of degree 1)

Therefore, we can only consider the map L s.t. $L(t) = f(x) - f'(x)(t - x)$

Differentiability

A function f is said to be **differentiable** at x is there exists such a linear map. For functions on the reals this is equivalent to the existence of a derivative: the linear map will be the one intersecting $f(x)$ with slope $f'(x)$. For functions defined on other spaces the definition of differentiability is more general.

A function $f : O \subseteq \mathbb{R}^n \to \mathbb{R}$ is said **differentiable** at $x \in O$ if there exists a matrix $\mathcal{J}_x(n \times 1)$, the Jacobian, such that

$$\lim_{\|h\| \to 0} \frac{\|f(x+h) - f(x) - J_x h\|}{\|h\|} = 0$$

Differentiability

Notice the following:

1. The derivative $Df : O \to \mathbb{R}^{\ltimes}$ assigns a Jacobian to each x.
2. the differential $df_x : \mathbb{R}^n \to \mathbb{R}$ is a mapping that assigns to each x the linear operator $df_x(h) = \mathcal{J}_x h$

Theorem
Consider the function $f : O \subseteq \mathbb{R}^n \to \mathbb{R}$. if f is differentiable at $x \in O$ then f is continuous at x. Suppose f has a well-defined partial derivatives and those are continuous. Then f is differentiable.

A function $f : O \subseteq \mathbb{R}^n \to \mathbb{R}$ is said **continuously differentiable** if it is differentiable and Df is a continuous function. Therefore, a funcion is continuous differentiable if and only if the partial derivatives of the function f exists and are continuous.

Inverse function theorem

Theorem

Inverse Function Theorem Let $f : O \subseteq \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function and let $x^0 \in O$. If the determinant of the Jacobian of f is not equal to zero at x^0, then there exists an open neighborhood of x^0, U, such that:

1. *f is a one-to-one correspondance in U and f^{-1} is well defined.*
2. *$V = f(U)$ is an open set containing $f(x^0)$.*
3. *f^-1 is continuously differentiable with $D[f^{-1}(x^0)] = [Df(x^0)]^{-1}$*

This theorem is very useful when we cannot invert $f(x, y)$. We can finc the Jacobian of the inverse, by taking the jacobian and then invert it.

Static optimization

3 cases of static optimization

1. Convex constraint set (unconstraint optimization).
2. Lagrange problem.
3. Kuhn-Tucker problem.

Convex constraint set

Let S be a convex set in $\mathbb{R}^n$, $f : S \to \mathbb{R}$ and consider the problem

$$\underset{x}{optimize} f(x) : x \in S)$$

Suppose S is an open set. Then, a necessary condition for the optimum is to be a **stationary point**:

$$Df(x^*) = 0$$

That is all partial derivatives must be zero. In the more general case, this condition is not necessary,for example, if S was closed; nor sufficient, x^* could be a saddle point if f is not differenciable.

Convex constraint set: Sufficient conditions for global optimums

Sufficient conditions for global optimums: Suppose $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$ is a differentiable function defined on a convex set S and let x^* be a starionary point int the interior of S. Then:

1. If f is strictly concave in $S \longrightarrow x^*$ is a global maximum.
2. If f is strictly convex in $S \longrightarrow x^*$ is a global minimum.
3. The set of optimizers $\sigma(\theta) := \text{argmax/argmin}\{f(x) : x \in X\}$ is either empty or convex. That is, there cannot be multiple isolated points as maximizers (minimizers).

convexity and concavity of f with the Hessian

Thus, to determine whether a stationary point x^* is a maximum or a minimum we need to find the convexity and concavity of the function f.

$$f \text{ is concave} \iff \text{Hessian at } x^* \text{ semidefinite negative} \iff \text{its latent roots are } \lambda_i \leq 0$$
$$f \text{ is convex} \iff \text{Hessian at } x^* \text{ semidefinite positive} \iff \text{its latent roots are } \lambda_i \geq 0$$

Lagrange problem

Consider the problem of maximize or minimize a function $f : S \subseteq \mathbb{R}^n$ knowing that the variables must satisfy a system of J equations $g_j(x) = 0$. To do so, we use the **Lagrange method** defined by the **lagrange function**

$$L(x, \lambda) = f(x) - \sum_{j=1}^{n} \lambda_j g_j(x)$$

Where note that we impose a penalization λ_j from deviating from each constraint g_j.

FOC Lagrange problem

For a given λ_j we consider the following *first order conditions*:

$$\frac{\partial L}{\partial x} = \frac{\partial f}{\partial x} - \sum_{j=1}^{n} \lambda_j \frac{\partial g_j}{\partial x} = 0$$

$$\frac{\partial L}{\partial \lambda_j} = g_j(x) = 0 \;\; \forall j = 1, \dots, J$$

Any stationary point x^* of the Lagrangean function $L(x, \lambda)$ is a candidate to global maximum or minimum if the gradient vector of the functions g_i in the stationary point are linearly independent.

Example

Kuhn-Tucker problem

Consider the problem of maximize (minimize) a function $f : S \subseteq \mathbb{R}^n$ subject to J constraints $g_j(x) \leq 0$, where both f and g are continuously differentiable functions from $\mathbb{R}^n$ to $\mathbb{R}$.

$$max(min)_x f(x) \quad s.t. g(x) \leq 0$$

The constraints can be

1. **binding** or **active** at a feasible point x^0 if they hold with equality, in which case we are in the Lagrangian case. That is there exists $\lambda_j \neq 0$
2. **Inactive** otherwise, in which case they will not have effects on the local properties of the solution. That is $\forall \lambda_j, \lambda_j = 0$

Kuhn-Tucker conditions

To solve this problem, we define the function

$$L(x, \lambda) = f(x) - \sum_{j=1}^{n} \lambda_j g_j(x)$$

Then, the **Kuhn-Tucker conditions** for the problem are

1. $D_x L(x, \lambda) = Df(x) - \sum_{j=1}^{n} \lambda_j Dg_j(x) = 0$
2. $\lambda_j \underset{(\leq)}{\geq} 0, g_j(x) \leq 0$ and $\lambda_j[g_j(x)] = 0$ for $j = 1, \ldots, J$

   Where

   - If x^* is a maximum $\Longrightarrow$ all $\lambda_i \geq 0$
   - If x^* is a minimum $\Longrightarrow$ all $\lambda_i \leq 0$

Example (Optimization in a non-closed region). Find the maximum and minimum of the function

$$f(x,y)=\frac{x}{2}-y \quad s.t. \begin{cases} x+e^{-x}\leq y \\ x\geq 0 \end{cases}$$

Solution: First we define the Lagrangian function

$$L(x,y,\lambda)=x/2-y-\lambda_1(x+e^{-x}-y)-\lambda_2(-x)$$

the first order conditions are

$$\begin{aligned} D_xL(x,y,\lambda)=1/2-\lambda_1(1-e^{-x})+\lambda_2=0 \\ D_yL(x,y\lambda)=-1+\lambda_1=0 \end{aligned}$$

Example: Kuhn-Tucker problem

From which we get $\lambda_1 = 1, e^{-x} + 0 = 1/2$. Then, we have the following 4 cases to study:

- 1. $[\lambda_1 = 0 \text{ and } \lambda_2 \neq 0]$
  we do not have any candidate.
- 2. $[\lambda_1 \neq 0 \text{ and } \lambda_2 = 0]$
  We have a candidate given by the conditions $e^{-x} + \lambda_2 = 1/2$ and $x + e^{-x} = y$. So that,

$$e^{-x} = \frac{1}{2} \qquad x = ln\frac{1}{2}^{-1} \qquad x = ln2$$

and

$$y = ln2 + e^{-ln2} = ln2 + 1/2$$

so that we find the point $(ln2, 1/2 + ln2)$ which is a candidate to a maximum $(\lambda = 1 \geq 0)$.

Example: Kuhn-Tucker problem

- 3. $[\lambda_1 = 0 \text{ and } \lambda_2 = 0]$
  We do not have any candidate since $\lambda_1 = 1$
- 4. $[\lambda \neq 0 \text{ and } \lambda \neq= 0]$
  In this case we have that $x = 0$ and $y = x + e^{-x} = y$ so that we find the point $(0, 1)$ which is neither a maximum or a minimum because $\lambda_1 = 1 \geq 0$ and $\lambda_2 = -1/2$

Example: Kuhn-Tucker problem

We found $(ln2, 1/2 + ln2)$ as a candidate to maximum with $\lambda_1 = 1, \lambda_2 = 0$, so that the Lagrangian function becomes $L(x, y, \lambda) = x/2 - y - (x + e^{-x} - y)$ and its Hessian is:

$$\mathcal{H}(f) = \begin{bmatrix} -e^{-x} & 0 \\ 0 & 0 \end{bmatrix}$$

And $det(\mathcal{H}(f)|_{(ln2, 1/2+ln2)} = 0$ so that we cannot determine the convexity/concavity with the Hessian. However, we can do so computing the eigenvalues of $\mathcal{H}(f)$. Solving the characteristic equation we find $\lambda_1 = -1/2$ and $\lambda_2 = 0$. Because $\lambda \leq 0$, the Hessian is semidefinite negative and thus f is concave. The point **(ln2, 1/2 +ln2) is a maximum**. Also note that in this exercise there is no minimum (the region is not bounded!) :

$$\lim_{x=0, y\to\infty} f(x, y) = \lim_{x=0, y\to\infty} \frac{0}{2} - y = -\infty$$