17. Unconstrained optimization

Gâteaux differential: $T$ is a possibly non-linear transformation from $X$ to $Y$. $\delta T(x; h)$ is called the Gâteaux differential of $T$ at $x$ in the direction $h$ if

$$\delta T(x; h) = \lim_{\alpha \to 0} \frac{T(x + \alpha h) - T(x)}{\alpha}$$

Note:

(i) This definition requires the concept of a norm in $Y$ (since the concept of a limit requires it), but $X$ can be just a vector space.

(ii) The Gâteaux differential generalizes the concept of a directional derivative of $f: \mathbb{R}^n \to \mathbb{R}$. 
(iii) If \( \delta T(x; h) \) exists \( \forall h \), then
\[ T \] is said to be Gateaux differentiable.

(iv) The definition makes sense only if \( x + \alpha h \in \text{Domain}(T) \) for sufficiently small \( \alpha \).

(v) For each fixed \( x \), \( \delta T(x; h) \)

is a possibly nonlinear transformation from \( x \) to \( y \).

(vi) Since the definition does not require a \( \| \cdot \| \) in \( X \), \( T \) need not be continuous for \( \delta T(x; h) \) to exist.

Frechet derivative: Suppose there exists a linear operator \( T(x) \in \mathcal{B}(x, Y) \) s.t.
\[
\lim_{\|h\| \to 0} \frac{\|T(x+h) - T(x) - T'(x)h\|}{\|h\|} = 0,
\]

then \( T'(x) \) is called the Frechet derivative of \( T \) at \( x \).

Note: (i) The Frechet derivative requires the concept of a norm in \( X \) as well.

(ii) If \( T'(x) \) exists, then the Gateaux differential \( \delta T(x; h) = T'(x)h \)

In this, \( \delta T(x; h) \) is also called the Frechet differential.

(iii) If \( T(x) \) exists, it is unique.

(iv) If \( T'(x) \) exists, then \( T \) is continuous at \( x \).

(v) We will write \( T'_x \) for \( T'(x) \)

sometimes

(vi) \( (\lambda_1 T_1 + \lambda_2 T_2)'(x) = \lambda_1 T_1'(x) + \lambda_2 T_2'(x) \)

(vii) Terminology: \( T'(x) : X \to Y \)

is a bounded linear operator.
For each fixed \( x \), it takes \( h \) as input and outputs \( T'(x)h \in Y \).

(Recall that in multivariable calculus \( \nabla f(x) \) is the gradient (Fréchet derivative of \( f \) at \( x \)) and \( \nabla f(x) h \) is the directional derivative (Gâteaux differential) in the direction \( h \).) \( T'(x) \) is a continuous linear operator.

We say that \( T'(x) \) is continuous at \( x_0 \) if \( \| T'(x) - T'(x_0) \| \leq \varepsilon \) whenever \( \| x - x_0 \| \leq \delta \). This concept is different from that of \( T'(x) \) being a continuous linear operator.

**Examples:**

(i) \( f(x, y) = \int \frac{x^2 y}{x^4 + y^2}, (x, y) \neq (0, 0) \)

0, else.
This is not a continuous function at 
$$(0,0)$$, so a Fréchet derivative cannot 
exist there. But a Gâteaux differential 
might as we will see.

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

$$= \lim_{\alpha \to 0} \left( \frac{h_1^2 h_2 \alpha^3}{(h_1^4 \alpha^4 + h_2^2 \alpha^2)} - 0 \right) \frac{1}{\alpha}$$

$$= \lim_{\alpha \to 0} \frac{h_1^2 h_2}{h_2^2 + h_4 \alpha^2} = \begin{cases} 0 & \text{if } h_2 = 0 \\ \frac{h_1}{h_2} & \text{if } h_2 \neq 0 \end{cases}$$

Note that $\delta f(x; h)$ is not 
even a linear operator on $h$.

(2) $f: \mathbb{E}^n \to \mathbb{R}$ is a function 
with continuous partial derivatives

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

Gâteaux
\[
\frac{\partial f}{\partial x_i} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}
\]

\[f'(x) = \left( \frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right)\]

Frechet

also denoted by \(\nabla f\), the gradient.

These follow from multivariable calculus.

(3) \(X = C[0, 1]\) and

\[f(x) = \int_0^1 g(x(t), t) \, dt,\]

where \(g_x\) is continuous in both \(x\) and \(t\),

\[
\delta f(x, h) = \lim_{\alpha \to 0} \left( \int_0^1 g \left( x(t) + \alpha h(t), t \right) dt \right) - \int_0^1 g \left( x(t), t \right) dt
\]

\[
= \int_0^1 g_x \left( x(t), t \right) h(t) \, dt.
\]
To prove that
\[ f'(x) h = \int_0^1 g_x(x(t), t) h(t) \, dt, \]
we have to show that
\[ \lim_{\| h \| \to 0} \frac{\| f(x + h) - f(x) - f'(x) h \|}{\| h \|} = 0. \]

See book for details.

**Theorem:**
\[ T: Y \to Z, \quad S: X \to Y \]
\[ (TS)(x) = T'(S(x)) S'(x). \]

(\text{Note: this is an extension of)
\[ \frac{d}{dx} f(g(x)) = f'(g(x)) g'(x). \]

**Theorem:**
\[ \| T(x + h) - T(x) \| \leq \| h \| \sup_{0 < \beta < 1} \| T'(x + \beta h) \|. \]
\[ \text{for } h \text{ if } x + \beta h \in \text{domain of } T, \quad \beta \in [0, 1]. \]
Proof: By H–B theorem, \( \exists y^* \ s.t. \)
\[
y^* \left[ T(x+h) - Tx \right] = \| T(x+h) - T(x) \|
\]
and \( \|y^*\| = 1 \). Consider the function
\[
\phi(x) = y^* \left[ T(x + xh) - T(x) \right]
\]

Note that \( y^* \) is a bounded, linear functional (with linear being the important point for what we are going to say next). Therefore, we claim \( y^*(y)h = y^*(h) \).

To see this, note that
\[
\frac{\| y^*(y+h) - y^*(y) - y^*(h) \|}{\| h \|} = 0, \text{ hence the limit as } \| h \| \to 0 \\
\text{is also zero.}
\]

Thus, by the chain rule,
\( \varphi'(x) = y^* \left[ T'(x + \alpha h) \right] \)

by the mean-value theorem,

\( \varphi(x) = \varphi(0) + \varphi'(x_0) \) for some

\( 0 < x_0 < 1 \)

\( \implies y^* \left[ T(x + h) - T(x) \right] \]

\( = y^* \left[ T'(x + \alpha_0 h) \right] \)

LHS = \( \| T(x + h) - T(x) \| \) by our choice of \( y^* \)

RHS \leq \( \| y^* \| \| T'(x + \alpha_0 h) \| \)

\( \frac{1}{2} \| h \|^2 \sup_{0 < \alpha < 1} \| T''(x + \alpha h) \| \)

**Theorem:**

\( \| T(x + h) - T(x) - T'(x) h \| \)

\( \leq \frac{1}{2} \| h \|^2 \sup_{0 < \alpha < 1} \| T''(x + \alpha h) \| \)

\( \forall x + \beta h \in \text{Dom}(T) \forall \beta \in [0,1]. \)
Here $T''$ is the derivative of $T'.$

Optimization:

$$\min_{x \in \Omega} f(x).$$

Global min $x_0 \in \Omega$: $f(x_0) \leq f(x) \quad \forall x \in \Omega.$

Local min $x_0 \in \Omega$ (aka relative min.):

$\exists \delta > 0$ s.t. $B_\delta(x) \subseteq \Omega$ and

$$f(x_0) \leq f(x) \quad \forall x \in B_\delta(x).$$

Strict local min if

$$f(x_0) < f(x) \quad \forall x \in B_\delta(x).$$

Theorem: let $\delta f(x; h)$ be the

Gâteaux differential of a functional $f$ over a normed space $X$, $x_0$ is a local min of $f$. Then,
\[ (i) \quad \delta f(x; h) = 0 \quad \forall \ h \in X \]

\[ (ii) \quad \text{If } f \text{ is defined only over a convex set } \Omega, \text{ then } \delta f(x; x - x_0) = 0 \quad \forall \ x \in \Omega. \]

**Proof:**

\[ (i) \quad \frac{f(x_0 + \alpha h) - f(x_0)}{\alpha} \geq 0 \]

\[ \Rightarrow \quad \delta f(x_0; h) \geq 0 \]

Consider the direction \(-h\):

\[ \frac{f(x_0 - \alpha h) - f(x_0)}{\alpha} \geq 0 \]

Define \( \beta = -\alpha \):

\[ \frac{f(x_0 + \beta h) - f(x_0)}{-\beta} \geq 0 \]

\[ \Rightarrow \quad \frac{f(x_0 + \beta h) - f(x_0)}{\beta} \leq 0 \]
Letting \( \beta \to 0 \), we get \( \delta f(x_0; h) \leq 0 \)

(iii) If \( \Omega \) is a convex set, then
\[
\alpha x + (1-\alpha) x_0 \in \Omega \quad \forall x \in \Omega
\]

\[
f(x_0 + \alpha (x - x_0)) \geq f(x_0)
\]

\[
\frac{f(x_0 + \alpha (x - x_0)) - f(x_0)}{\alpha} \geq 0
\]

As \( \alpha \to 0 \), \( \delta f(x_0; x - x_0) \geq 0 \)

Example: \( \min \quad J = \int_{t_1}^{t_2} f(x(t), \dot{x}(t), t) \, dt \)

\( x(t_1), x(t_2) \) are given.

\[
\delta J(x; h) = \int_{t_1}^{t_2} \left( \frac{d}{dx} f(x + \alpha h, \dot{x} + \alpha \dot{h}, t) \right) \, dt
\]

\[
= \int_{t_1}^{t_2} \left( \frac{\partial f}{\partial x}(x, \dot{x}, t) h(t) + \frac{\partial f}{\partial \dot{x}}(x, \dot{x} + \alpha \dot{h}, t) \dot{h}(t) \right) \, dt
\]
\[
= \int_{t_1}^{t_2} \frac{\partial f(x, \dot{x}, t)}{\partial x} h(t) \, dt \\
+ \left. \frac{\partial f(x, \dot{x}, t)}{\partial x} h(t) \, dt \right|_{t_1}^{t_2} \\
- \int_{t_1}^{t_2} \dot{h}(t) \frac{d}{dt} \frac{\partial f(x, \dot{x}, t)}{\partial x} \, dt
\]

**Second term:** \( h(t_1) = h(t_2) = 0 \) since \( x(t_1), x(t_2) \) are fixed, \( \Rightarrow \) second term = 0.

\[\Rightarrow \delta f(x; h) = \int_{t_1}^{t_2} \left( \frac{\partial f(x, \dot{x}, t)}{\partial x} - \frac{d}{dt} \frac{\partial f(x, \dot{x}, t)}{\partial x} \right) h(t) \, dt\]

Setting this equal to zero gives us:

\[\frac{\partial f(x, \dot{x}, t)}{\partial x} = \frac{d}{dt} \frac{\partial f(x, \dot{x}, t)}{\partial \dot{x}}\]
This is called the Euler-Lagrange equation.

Note: Here we assumed that $X = C_1 [t_1, t_2]$. An additional assumption that $f_x$ was continuous in $t$ was also made in doing integration by path. But this can be removed (see book).