<u>Fourier series</u>: we consider continuous 2π-periodic functions $f:\mathbb{R}\to\mathbb{C}$ with complex values, or equivalently functions on $S^1 \simeq \mathbb{R}/2\pi\mathbb{Z}$, with L^2 inner product $\langle f,g\rangle = \frac{1}{2\pi}\int_0^{2\pi} \overline{f(x)}\,g(x)\,dx$

The complex exponentials $e_n(x) = e^{inx}$, $n\in\mathbb{Z}$ satisfy $\langle e_i, e_j\rangle = \delta_{ij}$ - orthonormality.

<u>Def</u>: The Fourier coefficients of f are $c_n(f) = \langle e_n, f\rangle = \frac{1}{2\pi}\int_0^{2\pi} e^{-inx} f(x)\,dx$.

→ the <u>Fourier series</u> of f is $\sum_{n\in\mathbb{Z}} c_n e_n = \sum_{-\infty}^{\infty} c_n(f)\, e^{inx}$

<u>Q</u>: (Fourier, Dirichlet, Féjer, ...) does the Fourier series accurately represent f? (e.g. does it converge? to f?).

<u>Def</u>: <u>Trigonometric polynomials</u> = the vector space of <u>finite</u> linear combinations of e_n.

* Clearly this is an algebra, complex conj. invariant, and separates points of S^1, which is compact: hence by Stone-Weierstrass, trig. polynomials are <u>dense</u> in $(C^0(S^1), \|\cdot\|_\infty)$ ... hence also in L^2-norm $\left(\|f\|_{L^2} = \left(\frac{1}{2\pi}\int |f|^2\,dx\right)^{1/2} \le \sup|f|\right)$.

* The <u>n^{th} Fourier sum</u> $f_n = s_n(f) = \sum_{-n}^{n} c_k e^{ikx} = \sum_{-n}^{n} \langle e_k, f\rangle e_k$ is the orthogonal projection of f onto $V_n = \mathrm{span}(e_{-n},\dots,e_n)$ for $\langle\cdot,\cdot\rangle$.

Indeed: $\langle e_j, f_n\rangle = \sum_{k=-n}^{n} c_k \langle e_j, e_k\rangle = c_j = \langle e_j, f\rangle$, so $\langle e_j, f-f_n\rangle = 0$ $\forall -n\le j\le n$.

Thus: $\forall g\in V_n$, $\|f-f_n\|_{L^2} \le \|f-g\|_{L^2}$ - the point of V_n closest to f for $\|\cdot\|_{L^2}$

(This follows from $(f-f_n)\perp V_n$: $(f-g) = \underbrace{(f-f_n)}_{\perp V_n} + \underbrace{(f_n-g)}_{\in V_n} \Rightarrow \|f-g\|^2 = \|f-f_n\|^2 + \|f_n-g\|^2 \ge \|f-f_n\|^2$.)

⇒ <u>Theorem</u> (Parseval): Let $f\in C^0(S^1)$, $c_n = \langle e_n, f\rangle$ Fourier coeffts, $f_n = \sum_{-n}^{n} c_k e_k$ partial sums.
(1) $f_n \to f$ in L^2, ie. $\|f_n - f\|_{L^2}^2 = \frac{1}{2\pi}\int |f(x)-f_n(x)|^2\,dx \to 0$ as $n\to\infty$.
(2) $\sum_{n\in\mathbb{Z}} |c_n|^2 = \|f\|_{L^2}^2 = \frac{1}{2\pi}\int_0^{2\pi} |f(x)|^2\,dx$ (in particular $\sum |c_n|^2$ converges, so $c_n\to 0$ as $|n|\to\infty$)

<u>Pf</u>: (1) Since trig. polynomials $= \bigcup_n V_n$ are dense in $(C^0(S^1), \|\cdot\|_{L^2})$,
$\forall\varepsilon>0$ $\exists N$ st. $\exists g\in V_N$ with $\|f-g\|_{L^2} < \varepsilon$.
Now for $n\ge N$, $g\in V_N \subset V_n$ and f_n = closest point to f, so
$\|f-f_n\|_{L^2} \le \|f-g\|_{L^2} < \varepsilon$. This shows $f_n\to f$ in L^2.

(2) since $f_n\in V_n$ and $f-f_n \in V_n^\perp$, $\|f\|_{L^2}^2 = \|f_n\|_{L^2}^2 + \|f-f_n\|_{L^2}^2$ where $\|f_n\|_{L^2}^2 = \left\|\sum_{-n}^{n} c_k e_k\right\|^2 = \sum_{-n}^{n} |c_k|^2$ by orthonormality, and $\|f-f_n\|_{L^2}^2 \to 0$ by the first part. □

Corollary: if $f, g \in C^0(S^1)$ have same Fourier series then $\frac{1}{2\pi}\int |f-g|^2 dx = \sum |c_n(f) - c_n(g)|^2 = 0$, hence $f = g$.

* The fact that $f_n \to f$ in L^2 is the best approximation (in L^2 norm) of f by trig. polynomials, and that trig. polynomials are dense in $\|\cdot\|_\infty$ (so $\exists$ trig. polynomials $\to f$ uniformly) makes one hope that $f_n \to f$ uniformly or at least pointwise... alas not!

Fact: $\exists f \in C^0(S^1)$ st. the Fourier series of f does not converge ($s_n(f)(0)$ unbounded!) (but the example is hard to construct).

Thm (Dirichlet) ‖ if f is C^1 then $f_n = s_n(f) \longrightarrow f$ uniformly.

The proof uses convolution - redefine, for periodic functions, $(f*g)(x) = \frac{1}{2\pi}\int_0^{2\pi} f(t)\, g(x-t)\, dt$.

& note $c_n e_n(x) = \frac{1}{2\pi}\left(\int f(t) e^{-int} dt\right) e^{inx} = (f * e_n)(x)$.

So: $s_n(f) = \sum_{-n}^{n} c_k e_k = f * \left(\sum_{-n}^{n} e_k\right) = f * D_n$ where

$$D_n(x) = \sum_{-n}^{n} e^{ikx} = \frac{e^{i(n+\frac{1}{2})x} - e^{-i(n+\frac{1}{2})x}}{e^{ix/2} - e^{-ix/2}} = \frac{\sin(n+\frac{1}{2})x}{\sin(\frac{x}{2})} \qquad \underline{\text{Dirichlet kernel}}$$

Dirichlet's proof studies this convolution for $f \in C^1$ to prove unif. convergence.

The fact that convergence can sometimes fail makes it remarkable that $\forall f \in C^0$, f can be recovered from the partial sums $s_n(f) = f_n = \sum_{-n}^{n} c_k e^{ikx}$ ...

Thm (Féjer): ‖ If $f \in C^0(S^1)$ then $\dfrac{s_0(f) + \dots + s_{n-1}(f)}{n}$ converges uniformly to f.

The reason is that this process amounts to convolution with the Féjer kernel $F_n = \frac{D_0 + \dots + D_{n-1}}{n}$, which actually approximates identity (in the sense seen last time) unlike D_n.

Differentiation in several variables

Def: ‖ $U \subset \mathbb{R}^n$ open, $f : U \to \mathbb{R}^m$ is <u>differentiable</u> at $x \in U$ if $\exists$ linear map $L : \mathbb{R}^n \to \mathbb{R}^m$ st. $\lim_{v \to 0} \frac{|f(x+v) - f(x) - Lv|}{|v|} = 0$ (also write: $f(x+v) = f(x) + Lv + o(|v|)$)

$o(|v|)$ means: $\ll |v|$, ie. $(\dots / |v|) \to 0$

‖ The <u>differential</u> of f at x is then $Df(x) = L \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$
↳ aka $f'(x)$ or $df(x)$

- Conceptually, the input of $Df(x)$ is a <u>tangent vector</u> to U at x, and output $Df(x)v$ is a tangent vector at $f(x)$.
- Natural norm on $\mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$: the operator norm $\|L\| = \sup_{v \neq 0} \frac{|Lv|}{|v|}$ $(= \sup\{|Lv| / |v| \leq 1\})$
- Say $f \in C^1(U, \mathbb{R}^m)$ is f is differentiable $\forall x \in U$ and $Df : U \to \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$ is continuous.

- As a matrix, the entries of $Df(x)$ are the <u>partial derivatives</u> $\frac{\partial f_i}{\partial x_j}$ = derivative of f_i as function of x_j (keeping other x_k = const.) then $(Df(x)v)_i = \sum_j \frac{\partial f_i}{\partial x_j} v_j$

(Pf: take $v \parallel e_j$ in def.^{n} of the differential).

Thm: $f \in C^1(U, \mathbb{R}^m)$ iff $\forall i,j$ $\frac{\partial f_i}{\partial x_j}$ exists and is continuous.

$\Rightarrow$ is clear, but $\Leftarrow$ is more subtle: the existence of $\frac{\partial f_i}{\partial x_j}$ does not imply the differentiability or even the continuity of f!

Ex: $f(x,y) = \frac{x^3}{x^2+y^2}$, $f(0,0)=0$ $\Rightarrow$ $f(x,0) = x$ $\frac{\partial f}{\partial x}(0,0) = 1$ so if $Df(0)$ exists,
$f(0,y) = 0$ $\frac{\partial f}{\partial y}(0,0) = 0$ it maps $(v_1,v_2) \mapsto v_1$.

However $f(t,t) = \frac{t}{2} \neq t + o(|t|)$!

Pf $\Leftarrow$: enough to consider $f = f_i : U \to \mathbb{R}$ one component at a time.

Applying mean value theorem successively, for $x \in U$ and $v \in \mathbb{R}^n$ st. $B_{|v|}(x) \subset U$:

$$f(x_1+v_1,\dots,x_n+v_n) = f(x_1+v_1,\dots,x_{n-1}+v_{n-1},x_n) + \frac{\partial f}{\partial x_n}(x_1+v_1,\dots,x_{n-1}+v_{n-1},y_n)\, v_n$$

for some $y_n \in (x_n, x_n+v_n)$, by mean val. thm for $\partial f/\partial x_n$.

$$= \dots \quad \left(\text{apply mean val. thm. to } \partial f/\partial x_{n-1}, \dots, \frac{\partial f}{\partial x_1} \text{ successively}\right)$$

$$= f(x_1,\dots,x_n) + \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x_1+v_1,\dots,x_{j-1}+v_{j-1},\underset{\substack{\cap \\ (x_j,\, x_j+v_j)}}{y_j},x_{j+1},\dots,x_n) \cdot v_j$$

All these points are within distance $|v|$ of x, so using continuity of $\partial f/\partial x_j$ we get that for $|v| \to 0$ this is well approximated (within $o(|v|)$) by $f(x) + \sum_j \frac{\partial f}{\partial x_j}(x) v_j$.

Hence f is differentiable and $Df(x) = \left(\frac{\partial f}{\partial x_1} \cdots \frac{\partial f}{\partial x_n}\right)$, which depends continuously on x. □

- Usual rules of differentiation hold; in particular

Thm (<u>chain rule</u>): if g is differentiable at $x \in \mathbb{R}^n$ and f is differentiable at $g(x) \in \mathbb{R}^m$, then $f \circ g$ is differentiable at x and $D(f\circ g)(x) = Df(g(x)) \circ Dg(x)$

Pf: $g(x+v) = g(x) + \underbrace{Dg(x)\, v + r(v)}_{=w}$ where $r(v) = o(|v|)$ (i.e. $\lim \frac{|r(v)|}{|v|} = 0$).

so $$f\circ g(x+v) = f(g(x)+w) = f(g(x)) + Df(g(x))\, w + o(|w|)$$
$$= f(g(x)) + Df(g(x)) . Dg(x)\, v + o(|v|).$$ □

- Mean value thm doesn't hold, eg. $f : \mathbb{R} \to \mathbb{R}^2$, $t \mapsto (\cos t, \sin t)$ $\quad f(2\pi) = f(0) \neq f(0) + 2\pi f'(t)$ $\forall t \in [0,2\pi]$.

However we have the <u>mean value inequality</u>:

Thm: $f : U \to \mathbb{R}^m$ differentiable at every point of the line segment $[a,b] = \{tb + (1-t)a \,/\, t \in [0,1]\} \Rightarrow |f(b) - f(a)| \le |b-a| \cdot \sup_{x\in[a,b]} \|Df(x)\|$.

Pf: u = unit vector in direction of $f(b)-f(a)$, let $g(t) = \langle u, f(a+tv)\rangle$
v = —"— $b-a$

then $g'(t) = \langle u, Df(a+tv)\,v\rangle$ so $|g'(t)| \le \|Df(a+tv)\|$. The result then follows from the single-variable mean value ineq. for g on $[0, |b-a|]$. $\square$

- Higher order derivatives: f is C^2 if $Df : U \to \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m) \simeq \mathbb{R}^{n\times m}$ is C^1, etc.

The main important fact about higher partial derivatives is:

Prop: if $\left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)$ exist and are continuous then $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$

Pf: enough to consider the case of $f(x,y)$. For h and k small $\neq 0$, consider
$$\frac{1}{hk}\Big(f(x+h,y+k) - f(x+h,y) - f(x,y+k) + f(x,y)\Big)$$
writing this in terms of $g(x,y) = \frac{f(x,y+k) - f(x,y)}{k}$, this is $\frac{1}{h}\big(g(x+h,y) - g(x,y)\big)$

so by mean value thm for $\frac{\partial g}{\partial x}$, $\exists h_1 \in (0,h)$ st. this equals
$$\frac{\partial g}{\partial x}(x+h_1, y) = \frac{1}{k}\left(\frac{\partial f}{\partial x}(x+h_1, y+k) - \frac{\partial f}{\partial x}(x+h_1, y)\right).$$
In turn, by mean value thm for $\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)$, $\exists k_1 \in (0,k)$ st. this equals $\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)(x+h_1, y+k_1)$.

Doing the same calculation in opposite order shows $= \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)(x+h_2, y+k_2)$ for some $h_2 \in (0,h)$, $k_2 \in (0,k)$.

Since these 2nd derivatives are continuous by assumption, taking limits as $h, k \to 0$ gives the result. $\square$

- Hence: the Hessian matrix $H = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)$ is symmetric. and should be interpreted as a symmetric bilinear form on tangent vectors. If $f \in C^2$ then
$$f(x+v) = f(x) + Df(x)\cdot v + \frac{1}{2} H(x)(v,v) + o(|v|^2) \qquad \text{(\& so on, Taylor !)}.$$

- Because of the local approximation $f(x+v) = f(x) + Df(x)v + r(v)$, the behavior of $Df(x)$ governs that of f near x. In particular:

  → if $Df(x)$ is injective then f is injective on a (suff. small) neighborhood of x.

  → if $Df(x)$ is surjective then f maps a neighborhood of x surjectively onto a nbd of $f(x)$.

When both hold, f is a local diffeomorphism, by the inverse function theorem.

Def: a map $f : U \to V$ between open subsets of $\mathbb{R}^n$ is a diffeomorphism if it is a homeomorphism and both f and f^{-1} are C^1.