Inverse Problems

Benedikt Wirth

Contents

1 Introduction

Inverse Problem: Infer from an effect the cause, using a mathematical model.

- *genuine inverse problem:*

  effect is measured/observed, uniqueness of cause is desired

  Example 1 (Medical image reconstruction). *Construct from X-ray-CT measurement the X-ray attenuation (i. e. an image of the anatomy) within the patient.*

- *optimal control:*

  try to achieve desired effect by control of the cause (usually combined with optimization of further objective functionals), high regularity of cause is desired

  Example 2 (Cloaking). *Design the distribution of an optical material around a ball such that a light wave front behind the ball looks undisturbed (the ball is invisible).*

Associated forward problem: mathematical model which computes/produces the effect from the cause
A forward problem can typically be described by a (potentially nonlinear) map, the *forward operator*

$$A : X \to Y$$

for X the set or space of causes and Y of effects; the inverse problem thus is

$$\text{given } y \in Y, \text{ find } x \in X \text{ with } Ax = y. \tag{1}$$

Experimental design: For a sought, not directly measurable quantity of interest, choose/design a forward problem and associated measurements so that reconstructing the sought quantity via the associated inverse problem becomes as simple as possible.

Example 3 (X-ray directions). *Choose X-ray directions in CT such that they are few (reduction of radiation burden), but still allow a very good image reconstruction.*

Definition 4 (Well-posedness after Hadamard). *An inverse problem* (1) *is called* well-posed *if*

1. *it has a solution x*
2. *which is unique*
3. *and continuously depends on y.*

Remark 5 (Well-posedness). *Role of* (1), (2) *is clear.* (3) *is necessary, since the measurements y are never exact, but always contain small errors, so-called noise. These errors should not lead to completely different solutions x.*

Remark 6 (Typical inverse problems). *Inverse problems in applications are typically* ill-posed*, i. e. one of the conditions is violated (often all).*

Regularization: Method to produce a well-posed approximation for an inverse problem, i. e. $x = A^{-1}y$ is replaced by some $x = By$.

Example 7 (Tikhonov-regularization). *The inverse problem is replaced with*

$$x_\alpha = \arg\min_{x \in X} \|Ax - y\|_Y^2 + \alpha \|x\|_X^2.$$

Nomenclature: y = noise-free measurement, n^δ = noise (of some strength δ), $y^\delta = y + n^\delta$ = noisy measurement, $x = A^{-1}y$ = ground truth

Definition 8 (Space and convergence notions)**.**

1. *A* Banach space X *is a complete normed vector space (e. g.* $L^2((0,1))$*).*
2. *A* Hilbert space *is a Banach space, whose norm is induced by an inner product* $(\cdot,\cdot)$*.*
3. *The* dual space X^* *to a Banach space* X *is the space of all linear continuous maps* $\ell : X \to \mathbb{R}$ *with norm* $\|\ell\|_{X^*} = \sup_{\|x\|_X \le 1} |\ell(x)|$.
   One also writes $\ell(x) = \langle \ell, x\rangle_{X^*,X} = \langle \ell, x\rangle$.
4. $x_n \in X$ converges weakly *to* $x \in X$, $x_n \rightharpoonup x$, *if* $\ell(x_n) \to_{n\to\infty} \ell(x)$ $\forall \ell \in X^*$.
5. $\ell_n \in X^*$ converges weakly-$*$ *to* $\ell \in X^*$, $\ell_n \overset{*}{\rightharpoonup} \ell$, *if* $\ell_n(x) \to_{n\to\infty} \ell(x)$ $\forall x \in X$.
6. *Let* $A : X \to Y$ *be linear and continuous. The* adjoint operator $A^* : Y^* \to X^*$ *is defined by*
$$\langle A^*y', x\rangle = \langle y', Ax\rangle \quad \forall x \in X, y' \in Y^*.$$
7. *The* Hilbertian adjoint $A^H : Y \to X$ *of a linear and continuous operator* $A : X \to Y$ *is defined by*
$$(A^H y, x) = (y, Ax) \quad \forall x \in X, y \in Y.$$
8. *A functional* $f : X \to \mathbb{R}$ *is called* weakly lower semicontinuous *if* $\liminf_{n\to\infty} f(x_n) \ge f(x)$ *for all* $x_n \rightharpoonup x$. Weakly-* lower semicontinuous *is defined analogously.*

Theorem 9 (Banach space properties)**.**

1. *Riesz' representation theorem: A Hilbert space* X *is isometrically isomorphic to its dual space* X^* *via the Riesz isomorphism* $R_X : X^* \ni \ell \mapsto x_\ell \in X$ *with* $\langle \ell, \cdot\rangle = (x_\ell, \cdot)$*. Consequently,* $A^H = R_X A^* R_Y^{-1}$*.*
2. *Banach–Alaoglu theorem: Let* X *be a separable (i. e.* $\exists$ *countable dense subspace) or reflexive (i. e.* $(X^*)^*$ *is isometrically iomorphic to* X*) Banach space. The unit ball of* X^* *is weakly-* sequentially precompact.*
3. $\|\cdot\|_X$ *is weakly,* $\|\cdot\|_{X^*}$ *weakly-* lower semi-continuous.*

Theorem 10 (Well-posedness of Tikhonov regularization)**.** *Let* X, Y *be Hilbert spaces,* $A : X \to Y$ *be linear and continuous, then the Tikhonov regularization is well-posed.*

Proof. If x is a minimizer, then for $E_y(x) = \|Ax - y\|_Y^2 + \alpha\|x\|_X^2$ and any $\varphi \in X$ we have

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}t} E_y(x + t\varphi)|_{t=0} = 2(Ax - y, A\varphi) + 2\alpha(x, \varphi) \\ \Leftrightarrow \qquad & \underbrace{(Ax, A\varphi) + \alpha(x, \varphi)}_{=:B(x,\varphi)} = \underbrace{(y, A\varphi)}_{=:\ell(\varphi)}. \end{aligned}$$

Now

- $|B(x,\varphi)| \le (\alpha + \|A\|^2)\|x\|_X\|\varphi\|_X$,
- $B(x,x) \ge \alpha\|x\|_X^2$,
- $|\ell(\varphi)| \le \|y\|_Y\|A\|\|\varphi\|_X$,

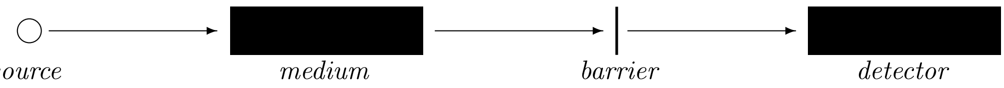
so by Lax–Milgram $\exists!$ solution x that continuously depends on y. It remains to show that this x really is a minimizer.

$$\begin{aligned} E_y(z) - E_y(x) &= \|Az - Ax\|_Y^2 + \alpha\|z - x\|_X^2 + 2(Ax - y, A(z - x)) + 2\alpha(x, (z - x)) \\ &= \|Az - Ax\|_Y^2 + \alpha\|z - x\|_X^2 > 0. \quad \square \end{aligned}$$

2 Illustration via integration/differentiation

Forward problem: $X = Y = L^2((0,1))$, $Ax = \big(s \mapsto \int_0^s x(t)\,\mathrm{d}t\big)$ (linear)

Example 11 (KATRIN experiment)**.** *Many electrons of a specific energy (normalized to 1) are sent into a medium (Xenon gas cloud). One wants to quantify the interaction, i. e. how many electrons are decelerated by how much energy, i. e. $x \in X$ is the probability density of the energy loss $\in [0,1]$. (The idea is that from x one can then read off the neutrino mass.) One can construct a barrier which blocks all electrons below a minimum energy s, and one can count the electrons behind, i. e. $y(s) = \int_0^{1-s} x(t)\,\mathrm{d}t$.*



The measurement y^δ contains errors, e. g. additive white Gaussian noise of standard deviation $\delta > 0$, i. e.

$$y^\delta = y + n^\delta$$

with n^δ the realization of a random variable such that $W_{a,b} = \int_a^b n^\delta(t)\,\mathrm{d}t$ has zero mean and variance $(b-a)\delta^2$ for all $0 \le a \le b \le 1$ and such that $W_{a,b}$ and $W_{c,d}$ have covariance $r\delta^2$ with r the length of $[a,b] \cap [c,d]$.

Theorem 12 (Non-well-posedness of differentiation)**.** *Consider the inverse problem $Ax = y$.*

1. *In general it has no solution.*
2. *If a solution exists, it is unique,*
3. *but not continuous in $y \in Y$.*

Proof. (2) Fundamental theorem of calculus $\Rightarrow \quad x(s) = y'(s)$, and the weak derivative is unique.
(1) If $y \notin W^{1,2}((0,1))$, it has no weak derivative in X.
(3) Set $\Delta y = \sin(nt)$, then $\|\Delta y\|_Y \le 1$, but

$$\|A^{-1}(y + \Delta y) - A^{-1}y\|_X = \|t \mapsto n\cos nt\|_X \ge Cn$$

gets arbitrarily big for $n \to \infty$. □

Tikhonov-regularization:

$$x_\alpha = \operatorname*{arg\,min}_{x \in X} \int_0^1 \left| \int_0^s x(t)\,\mathrm{d}t - y^\delta(s) \right|^2 \mathrm{d}s + \alpha \int_0^1 |x(t)|^2\,\mathrm{d}t.$$

Set $y_\alpha = Ax_\alpha$, then

$$y_\alpha = \operatorname*{arg\,min}_{y \in W^{1,2}((0,1)), y(0)=0} \int_0^1 \left| y - y^\delta \right|^2 \mathrm{d}s + \alpha \int_0^1 |y'|^2\,\mathrm{d}t$$

$$\Leftrightarrow \qquad y_\alpha \text{ solves } \begin{cases} -\alpha y_\alpha'' + y_\alpha - y^\delta = 0 & \text{on } (0,1) \\ y_\alpha(0) = 0,\ y_\alpha'(1) = 0, \end{cases}$$

thus y_α is the solution of an implicit Euler step with stepsize α of the heat equation with homogeneous Dirichlet-/Neumann-boundary conditions.
$\Rightarrow \qquad y^\delta$ first is smoothed to y_α!

Error estimate: Possible if noise-free measurement y or equivalently ground truth $x = A^{-1}y$ have additional regularity, e. g. $y \in C^2([0,1])$ & (for simplicity) $y'(1) = 0$.
Optimality condition: For all $\varphi \in W^{1,2}((0,1))$ with $\varphi(0) = 0$ we have

$$\begin{aligned} (y'_\alpha, \varphi') + \quad (y_\alpha, \varphi) &= (y^\delta, \varphi) \\ \alpha(y', \varphi') + \quad (y, \varphi) &= (y - \alpha y'', \varphi) \\ \hline \text{difference: } \alpha(y'_\alpha - y', \varphi') + (y_\alpha - y, \varphi) &= (y^\delta - y + \alpha y'', \varphi). \end{aligned}$$

With $\varphi = y_\alpha - y$ we get

$\Rightarrow$ $\alpha\|y'_\alpha - y'\|^2_{L^2} + \|y_\alpha - y\|^2_{L^2} = (y^\delta - y + \alpha y'', y_\alpha - y) \overset{\text{Young}}{\leq} \|y^\delta - y\|^2_{L^2} + \alpha^2\|y''\|^2_{L^2} + \frac{1}{2}\|y_\alpha - y\|^2_{L^2}$

$\Rightarrow$ $\|x_\alpha - x\|^2_{L^2} \leq \frac{1}{\alpha}\|y^\delta - y\|^2_{L^2} + \alpha\|y''\|^2_{L^2}$

$\Rightarrow$ for $\|y^\delta - y\|_{L^2} = \delta$, the optimal choice $\alpha = \frac{\delta}{\|y''\|_{L^2}}$ yields

$$\|x_\alpha - x\|_{L^2} \leq 2\sqrt{\|y''\|_{L^2}}\sqrt{\delta}$$

$\Rightarrow$ even with regularization we have reconstruction error $\gg$ measurement error

3 Some classical inverse/forward problems

1. Differentiation/Integration
2. X-ray transform

   Definition 13 (X-ray transform). *Let $\mathcal{C}' = \{(\theta, s) \in S^{d-1} \times \mathbb{R}^d \,|\, s \in \theta^\perp\}$. The* X-ray transform *on $B_1(0) \subset \mathbb{R}^d$ is the linear map*

   $$P : L^1(B_1(0)) \to L^1(\mathcal{C}'), \qquad Pu(\theta, s) = \int_{\{x \in B_1(0) \,|\, x = s + t\theta \text{ for some } t \in \mathbb{R}\}} u(x) \,\mathrm{d}\mathcal{L}^1(x).$$

   Remark 14 (Other function spaces). *The definition can also be extended to other function spaces such as Radon measures on $B_1(0)$ or on $\mathbb{R}^d$.*

   Example 15 (Computer tomography, CT). *An X-ray is taken from every direction $\theta \in S^2$. The attenuation of the X-ray at position $s + t\theta$ is proportional to the ray intensity I and the attenuation coefficient u at that position, thus*

   $$\tfrac{\mathrm{d}}{\mathrm{d}t} I(s + t\theta) = -u(s + t\theta) I(s + t\theta) \qquad \Rightarrow \qquad I(s + \theta) = I(s - \theta) \exp\left(-\int_{-1}^{1} u(s + t\theta) \,\mathrm{d}t\right).$$

   The measured intensity change is $\frac{I(s+\theta)}{I(s-\theta)} = \exp -Pu(\theta, s)$, thus $Pu(\theta, s) = \log \frac{I(s-\theta)}{I(s+\theta)}$.

   Theorem 16 (X-ray transform). *$P : L^1(B_1(0)) \to L^1(\mathcal{C}')$ is continuous.*

   Proof. $\int_{\mathcal{C}'} |Pu| \,\mathrm{d}(\theta, s) \leq \int_{S^{d-1}} \int_{\theta^\perp} \int_{\{x \in B_1(0) \,|\, x = s + t\theta \text{ for some } t \in \mathbb{R}\}} |u(x)| \,\mathrm{d}\mathcal{L}^1(x) \,\mathrm{d}s \,\mathrm{d}\theta \overset{\text{Fubini}}{=} \int_{S^{d-1}} \|u\|_{L^1} \,\mathrm{d}\theta = |S^{d-1}| \|u\|_{L^1}$. □

   Remark 17 (CT in other spaces). *Similarly in L^p.*

3. Parameter identification

   Determine coefficients of a PDE problem (e. g. inside PDE or IC or BC) from observation of the solution in a subdomain (e. g. on the boundary) for different right-hand sides (of the PDE or IC or BC).

   Example 18 (Oil production). *Let $\Omega \subset \mathbb{R}^3$ with smooth boundary represent rock and $x \in X = C(\Omega)$ with $x \geq c > 0$ represent the rock permeability. The measured liquid pressure within the rock is $y \in Y = W^{1,2}(\Omega)$, and $A : X \to Y$ maps x onto the solution y of*

   $$-\mathrm{div}(x(z)\nabla y(z)) = f(z) \text{ in } \Omega \text{ plus BC (Darcy flow)},$$

   where $f \in L^2(\Omega)$ is a controllable source.

Remark 19 (Nonlinearity of parameter identification)**.** *Parameter identification problems are typically nonlinear, e. g.* $A(2x) \neq 2A(x)$ *for the above example.*

Remark 20 (Solution formula in 1D)**.** *In 1D a single right-hand side* f *suffices, and we can derive a solution formula: Let* $\Omega = (0,1)$ *with homogeneous Neumann boundary conditions on the left boundary, then*

$$x(z)y'(z) = -\int_0^z f(s)\,\mathrm{d}s.$$

If $f > 0$ *or* $f < 0$*, we have* $\int_0^z f\,\mathrm{d}s \neq 0$*, thus* $y'(z) \neq 0$ *and*

$$x(z) = \frac{-\int_0^z f(s)\,\mathrm{d}s}{y'(z)}.$$

For $f = 0$*, however,* x *cannot be identified. The problem is ill-posed because of the differentiation; in addition there is error amplification for small* y'*.*

Definition 21 (Dirichlet-to-Neumann map)**.** *Let* $\Omega \subset \mathbb{R}^d$ *with smooth boundary,* $a \in C^0(\Omega)$ *with* $a \geq c > 0$*. The linear map*

$$\Lambda_a : H^{\frac{1}{2}}(\partial\Omega) \to H^{-\frac{1}{2}}(\partial\Omega), \quad f \mapsto a\frac{\partial u}{\partial n} \text{ for } u \text{ solution of } \begin{cases} -\mathrm{div}(a\nabla u) = 0 & \text{in } \Omega, \\ u = f & \text{on } \partial\Omega \end{cases}$$

is called Dirichlet-to-Neumann map*.*

($H^{\frac{1}{2}}(\partial\Omega)$ and $H^{-\frac{1}{2}}(\partial\Omega)$ are special Hilbert spaces, the traces of H^1 and L^2 functions.)

Example 22 (Electrical impedance tomography, EIT)**.** $\Omega \subset \mathbb{R}^d$ *patient body,* $x \in X = C^0(\Omega)$ *with* $x \geq c > 0$ *is electrical conductance,* $Y = L(H^{\frac{1}{2}}(\partial\Omega), H^{-\frac{1}{2}}(\partial\Omega))$*,* $Ax = \Lambda_x$*.* Λ_x *is measured by applying different voltages* f *and measuring the resulting currents.*

Remark 23 (1D EIT)**.** *For* $\Omega = (a,b)$ *we have* $Y = L(\mathbb{R}^2, \mathbb{R}^2) = \mathbb{R}^{2\times 2}$*. One measurement* $y \in \mathbb{R}^{2\times 2}$ *cannot suffice to reconstruct a function* $x \in C^0(\Omega)$*. For higher-dimensional domains reconstructions turn out to be possible.*

4. Inverse scattering

   Determine an object based on its scattering of (acoustic or electromagnetic) waves = special case of parameter identification for wave equation, Maxwell's equations, Schrödinger equation or other hyperbolic equations.

   Example 24 (Periodic wave field)**.** *The density or pressure* U *of an acoustic wave satisfies*

   $$\frac{\partial^2 U}{\partial t^2} = \frac{1}{n^2}\Delta U, \qquad \frac{1}{n(z)} = \text{speed of sound } (= 1 \text{ outside object}).$$

   For time-harmonic (i. e. periodic) waves $U(z,t) = e^{ikt}u(z)$ *this turns into Helmholtz' equation*

   $$\Delta u + k^2 n^2 u = 0$$

   for the observed wave u*. The incoming wave (which is sent) satisfies* $\Delta u^i + k^2 u^i = 0$*, the scattered wave is* $u^s = u - u^i$ *and satisfies*

   $$\Delta u^s + k^2 u^s = k^2(1 - n^2)(u^i + u^s).$$

   If $\mathcal{O} \subset B_1(0)$ *is the sought scattering object, one has* $1 - n^2 = c\chi_{\mathcal{O}}$ *for some fixed* $c > 0$*; the measurement typically is the far-field wave* $u|_{\partial B_R(0)}$ *with* $R \gg 1$*, i. e.*

   $$x \equiv \mathcal{O}, \quad y \equiv \{u|_{\partial B_R(0)} \text{ for a number of incoming waves } u^i \text{ of different frequencies}\}, \quad A : x \mapsto y.$$

   Variation: Sound is absorbed on $\partial\mathcal{O}$ *(*$u|_{\partial\mathcal{O}} = 0$*) or reflected (*$\frac{\partial u}{\partial n} = 0$*) or a mixture (*$\frac{\partial u}{\partial n} + \lambda u = 0$*).*

4 Linear integral operators

Many forward operators in inverse problems are linear integral operators, thus they form an interesting set of examples.

Definition 25 (Integral operator)**.** *Let $\Sigma \subset \mathbb{R}^n, \Omega \subset \mathbb{R}^d$ measurable and $k : \Sigma \times \Omega \to \mathbb{R}$ measurable. The linear* integral operator *with* integral kernel *k is defined for measurable functions $u : \Omega \to \mathbb{R}$ as*

$$Ku : \Sigma \to \mathbb{R}, \quad (Ku)(x) = \int_\Omega k(x,y)u(y)\,\mathrm{d}y.$$

Example 26 (Integration)**.** *Let $\Sigma = \Omega = (0,1)$, $k(x,y) = 1$ if $x \geq y$ and $k(x,y) = 0$ else. Then $(Ku)(x) = \int_0^1 k(x,y)u(y)\,\mathrm{d}y = \int_0^x u(y)\,\mathrm{d}y$.*

Example 27 (Convolution)**.** *Let $\Sigma = \Omega = \mathbb{R}^d$, $k(x,y) = G(x-y)$ for a measurable function $G : \mathbb{R}^d \to \mathbb{R}$. Then $Ku = G * u$.*

Remark 28 (X-ray transform)**.** *Generalizing k to measures (which we don't in this lecture), also the X-ray transform becomes a linear integral operator.*

Theorem 29 (Continuity of integral operator)**.** *Let Σ, Ω open and bounded and $k \in L^q(\Sigma \times \Omega)$ with $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$. Then $K : L^p(\Omega) \to L^r(\Sigma)$, $(Ku)(x) = \int_\Omega k(x,y)u(y)\,\mathrm{d}y$ is well-defined and continuous with $\|Ku\|_{L^r} \leq C\|k\|_{L^q}\|u\|_{L^p}$.*

Proof.

$$\begin{aligned}
\|Ku\|_{L^r}^r &= \int_\Sigma |\Omega|^r \left| \frac{1}{|\Omega|} \int_\Omega k(x,y)u(y)\,\mathrm{d}y \right|^r \mathrm{d}x \\
&\leq |\Omega|^{r-1} \int_\Sigma \int_\Omega |k(x,y)|^r |u(y)|^r \,\mathrm{d}y\,\mathrm{d}x \\
&\leq |\Omega|^{r-1} \left(\int_{\Sigma\times\Omega} |k(x,y)|^{rq/r} \,\mathrm{d}x\,\mathrm{d}y \right)^{r/q} \left(\int_{\Sigma\times\Omega} |u(y)|^{rp/r} \,\mathrm{d}x\,\mathrm{d}y \right)^{r/p} \\
&= |\Omega|^{r-1} |\Sigma|^{r/p} \|k\|_{L^q}^r \|u\|_{L^p}^r.
\end{aligned}$$

□

Theorem 30 (Young)**.** *Let $\Sigma = \Omega = \mathbb{R}^d$, $k(x,y) = G(x-y)$ for a $G \in L^q(\mathbb{R}^d)$ with $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$. Then $K : L^p(\Omega) \to L^r(\Sigma)$, $(Ku)(x) = \int_\Omega k(x,y)u(y)\,\mathrm{d}y$ is well-defined and continuous with $\|Ku\|_{L^r} \leq \|G\|_{L^q}\|u\|_{L^p}$.*

Proof.

$$\begin{aligned}
|u * G(x)| &\leq \int (|u(x-y)|^p |G(y)|^q)^{1/r} |u(x-y)|^{1-p/r} |G(y)|^{1-q/r} \,\mathrm{d}y \\
&\leq \underbrace{\|(|u(x-\cdot)|^p |G|^q)^{1/r}\|_{L^r}}_{=(\int |u(x-y)|^p |G(y)|^q \,\mathrm{d}y)^{\frac{1}{r}}} \underbrace{\||u|^{1-p/r}\|_{L^{\frac{1}{\frac{1}{p}-\frac{1}{r}}}}}_{=\|u\|_{L^p}^{1-p/r}} \underbrace{\||G|^{1-q/r}\|_{L^{\frac{1}{\frac{1}{q}-\frac{1}{r}}}}}_{=\|G\|_{L^q}^{1-q/r}}
\end{aligned}$$

$$\Rightarrow \quad \|u * G\|_{L^r}^r = \int |u * G|^r \,\mathrm{d}x \leq \|u\|_{L^p}^{r-p} \|G\|_{L^q}^{r-q} \underbrace{\int\int |u(x-y)|^p |G(y)|^q \,\mathrm{d}y\,\mathrm{d}x.}_{=\int\int |u(x-y)|^p \,\mathrm{d}x |G(y)|^q \,\mathrm{d}y = \|u\|_{L^p}^p \|G\|_{L^q}^q}$$

□

5 Compact operators

Most forward operators in inverse problems are compact operators.

Definition 31 (Compact operator)**.** *Let X, Y be Banach spaces. A linear operator $K : X \to Y$ is called* compact *if for any bounded set $B \subset X$ the image $K(B)$ is precompact in Y.*

Corollary 32 (Sequences under compact operator)**.** *K compact $\Leftrightarrow$ Kx_n contains a convergent subsequence for every bounded sequence $x_n \in X$.*

Proof. '⇒' Choose $B = \{x_1, x_2, \ldots\}$, then $\overline{\{Kx_1, Kx_2, \ldots\}}$ is compact; in a metric space compact = sequentially compact.

'⇐' Each sequence $x_n \in B$ is bounded ⇒ ∃ convergent subsequence of Kx_n ⇒ $\overline{K(B)}$ is sequentially compact and thus compact. □

Corollary 33 (Weak-strong convergence). *Let X be reflexive (i. e. $X^{**} = X$). K compact ⇔ $x_n \rightharpoonup x$ implies $Kx_n \to Kx$.*

Proof. '⇒' x_n is bounded ⇒ $Kx_n \to y \in Y$ (thus also $Kx_n \rightharpoonup y$) for a subsequence; furthermore $y = Kx$ because of $Kx_n \rightharpoonup Kx$. Now let x_k be a subsequence with $\|Kx_k - y\|_Y > c > 0$, then again Kx_k contains a convergent subsequence with $Kx_k \to Kx = y$ ↯

'⇐' Banach–Alaoglu: Every bounded sequence x_n is weakly precompact.
Eberlein–Šmulian: In a Banach space weakly compact = sequentially weakly compact.
⇒ x_n has weakly convergent subsequence ⇝ use previous corollary. □

Corollary 34 (Finite-dimensional image). *Any linear continuous operator K with finite-dimensional image is compact.*

Proof. B bounded implies KB bounded and finite-dimensional, thus precompact by Heine–Borel. □

Theorem 35 (Operations on compact operators). *Let X, Y, Z be Banach spaces, K, L linear operators.*

1. *$K, L : X \to Y$ compact ⇒ $K + L$ compact*
2. *$K : X \to Y$ compact, a real ⇒ aK compact*
3. *$K : X \to Y$ or $L : Y \to Z$ compact ⇒ $LK : X \to Z$ compact*

Proof. 1. Let x_n be bounded sequence ⇒ $Kx_{n_k} \to y \in Y$ for subsequence x_{n_k};
x_{n_k} bounded ⇒ $Lx_{n_{k_l}} \to \tilde{y} \in Y$ for subsequence $x_{n_{k_l}}$
⇒ $(K + L)x_{n_{k_l}} \to y + \tilde{y}$.

2. trivial

3. If K compact: Let x_n be bounded ⇒ $Kx_n \to y \in Y$ for subsequence ⇒ $LKx_n \to Ly$ for same subsequence
If L compact: Let x_n be bounded ⇒ Kx_n is bounded ⇒ LKx_n has convergent subsequence □

Theorem 36 (Schauder's theorem). *Let X, Y be Banach spaces, $K : X \to Y$ linear. K compact ⇔ K^* compact.*

Proof. '⇒' Let B_{Y^*} be the closed unit ball in Y^*, B_X the one in X.

- B_{Y^*} is equicontinuous, since $|\langle y', y\rangle - \langle y', \tilde{y}\rangle| \leq \|y'\|\|y - \tilde{y}\|_Y \leq \|y - \tilde{y}\|_Y \ \forall y, \tilde{y} \in Y, y' \in B_{Y^*}$
- let $E = \overline{KB_X}$ & note that E is compact
- let $y'_n \in B_{Y^*}$ be a sequence $\overset{\text{Arzelà–Ascoli}}{\Longrightarrow}$ ∃ uniformly convergent subsequence $y'_n|_E \to y'|_E$
⇒ $K^* y'_n$ is Cauchy (and thus K^* compact), since

$$\|K^* y'_n - K^* y'_m\| = \sup_{x \in B_X} |\langle K^* y'_n, x\rangle - \langle K^* y'_m, x\rangle| = \sup_{z \in E} |\langle y'_n, z\rangle - \langle y'_m, z\rangle| \xrightarrow[m,n\to\infty]{} 0$$

'⇐' Let $i : X \to X^{**}$, $j : Y \to Y^{**}$ be the inclusion.

- K^{**} is compact (by '⇒'), and $K^{**} \circ i = j \circ K \overset{\text{theorem 35}}{\Longrightarrow} j \circ K$ is compact
- KB_X is precompact in Y:
$x_n \in B_X$ sequence ⇒ subsequence jKx_n is Cauchy
⇒ $\|Kx_n - Kx_m\|_Y \overset{\text{Hahn–Banach}}{=} \sup_{y' \in B_{Y^*}} \langle y', Kx_n - Kx_m\rangle = \sup_{y' \in B_{Y^*}} \langle y', jKx_n - jKx_m\rangle \xrightarrow[m,n\to\infty]{} 0$

□

Theorem 37 (Operator norm of compact operators). *Let X, Y be Hilbert spaces, $K : X \to Y$ linear and compact. There exists $x \in X$ with $\|x\| = 1$ and $\|Kx\| = \|K\|$.*

Remark 38 (Norm of non-compact operators)**.** *In general false for non-compact operators (homework).*

Proof.
- let $y_n \in Y$ with $\|y_n\| = 1$ and $\|K^H y_n\| \to \|K^H\| = \|K\|$
- $K^H y_n \to z \in X$ along a subsequence (since K^H is compact), and $\|z\| = \lim_{n\to\infty} \|K^H y_n\| = \|K\|$
- $\|K^H y_n\|^2 = (y_n, KK^H y_n) \le \|KK^H y_n\| \le \|K\|^2$, thus $\|Kz\| = \lim_{n\to\infty} \|KK^H y_n\| = \|K\|^2$
- set $x = z/\|z\|$, then $\|Kx\| = \|Kz\|/\|K\| = \|K\|$ □

Compactness of an operator can be shown via approximation by compact operators (Fredholm considered compact operators as limits of operators with finite rank, 1900; the use and analysis of the below compactness condition originates from Frigyes Riesz, 1918).

Theorem 39 (Limit of compact operators)**.** *Let X, Y be Banach spaces and $K_n : X \to Y$ a sequence of compact operators with $K_n \to K$, then K is compact.*

Proof.
- Let $x_k \in X$ be bounded sequence, $\|x_k\|_X \le C < \infty \ \forall k$
- Let $I_1 \subset \{x_1, x_2, \dots\}$ be subsequence such that $\lim_{k\to\infty, x_k \in I_1} K_1 x_k$ exists,
  $I_2 \subset I_1$ one such that $\lim_{k\to\infty, x_k \in I_2} K_2 x_k$ exists,
  $I_n \subset I_{n-1}$ one such that $\lim_{k\to\infty, x_k \in I_n} K_n x_k$ exists.
- Let z_k be the kth element of I_k, then $\lim_{k\to\infty} K_n z_k$ exists $\forall n$
- Kz_k is Cauchy: Let $\epsilon > 0$, then choose n such that $\|K_n - K\| \le \frac{\epsilon}{3C}$, and choose N such that $\|K_n z_l - K_n z_m\|_Y \le \frac{\epsilon}{3} \ \forall m, l > N$.

$$\Rightarrow \quad \|Kz_l - Kz_m\|_Y \le \underbrace{\|Kz_l - K_n z_l\|_Y}_{\le \|K - K_n\| \|z_l\|_Y \le \frac{\epsilon}{3}} + \underbrace{\|K_n z_l - K_n z_m\|_Y}_{\le \frac{\epsilon}{3}} + \underbrace{\|K_n z_m - K z_m\|_Y}_{\le \frac{\epsilon}{3}} \le \epsilon \quad \forall m, l > N$$

□

Theorem 40 (Compactness of integral operators)**.** *Let $\Sigma \subset \mathbb{R}^n, \Omega \subset \mathbb{R}^d$ open and bounded and $k \in L^q(\Sigma \times \Omega)$ with $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$, $q < \infty$. Then $K : L^p(\Omega) \to L^r(\Sigma)$, $Ku(x) = \int_\Omega k(x,y)u(y)\,\mathrm{d}y$ is compact.*

Proof.
- Wlog we may assume that k is Lipschitz with $|k(x_1,y_1) - k(x_2,y_2)| \le L|(x_1,y_1) - (x_2,y_2)|$:
  - $C^{0,1}(\mathbb{R}^{n+d})$ is dense in $L^q(\Sigma \times \Omega) \subset L^q(\mathbb{R}^{n+d})$
  - let $k_n \in C^{0,1}(\mathbb{R}^{n+d})$ with $k_n \xrightarrow{L^q} k$, $K_n u(x) = \int_\Omega k_n(x,y)u(y)\,\mathrm{d}y$
  - $\|K_n - K\| \le C\|k_n - k\|_{L^q} \to_{n\to\infty} 0$
  - if K_n compact, then also K by previous result
- approximate K by K_n with finite-dimensional image:
  - for $n \in \mathbb{N}$ let $(\Omega_i^n)_i$ be finite partition of Ω with $\mathrm{diam}(\Omega_i^n) < \frac{1}{n}$
  - set $\phi_i^n(x) = \displaystyle\int_{\Omega_i^n} k(x,y)\,\mathrm{d}y/|\Omega_i^n|$ *(average)*

$$\psi_i^n(y) = \begin{cases} 1 & \text{if } y \in \Omega_i^n \\ 0 & \text{else} \end{cases}$$

$$k_n(x,y) = \sum_i \phi_i^n(x)\psi_i^n(y)$$

  - $|k_n(x,y) - k(x,y)| = |\phi_i^n(x) - k(x,y)| = |\int_{\Omega_i^n} k(x,z) - k(x,y)\,\mathrm{d}z|/|\Omega_i^n| \le \frac{L}{n}$ for $y \in \Omega_i^n$
    $\Rightarrow k_n \xrightarrow{L^q} k$, $K_n \to K$ with $K_n u(x) = \int_\Omega k_n(x,y)u(y)\,\mathrm{d}y$
- $K_n u = \sum_i \phi_i^n \int_{\Omega_i^n} u(y)\,\mathrm{d}y \in \mathrm{span}\{\phi_1^n, \phi_2^n, \dots\} \Rightarrow K_n$ compact $\Rightarrow K$ compact □

Remark 41 (Compactness of integral operators)**.** *An analogous proof partitions Σ instead of Ω (homework).*

Boundedness of the domain is important for compactness.

Theorem 42 (Convolution on unbounded domain is not compact)**.** *Let $\Sigma = \Omega = \mathbb{R}^d$, $k(x, y) = G(x-y)$ for some $G \in L^q(\mathbb{R}^d)$ with $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$. Then $K : L^p(\Omega) \to L^r(\Sigma)$, $Ku(x) = \int_\Omega k(x, y)u(y)\,\mathrm{d}y$ is not compact.*

Proof. Homework (construct a sequence $u_n \in L^p(\mathbb{R}^d)$ by translating a fixed function and show that Ku_n contains no convergent subsequence). □

We now show that compact operators do not possess a continuous inverse (in infinite dimensions).

Theorem 43 (Almost orthogonal element/Riesz lemma)**.** *Let X be a normed vector space and $U \subsetneq X$ a closed subspace. Then for every $\epsilon > 0$ there exists an $x \in X$ with $\|x\|_X = 1$ and $\mathrm{dist}(x, U) = \inf\{\|y - x\|_X \mid y \in U\} > 1 - \epsilon$.*

Proof.

- choose $v \in X \setminus U$ and $u \in U$ with $\|v - u\|_X < \frac{\mathrm{dist}(v,U)}{1-\epsilon}$
- set $x = \frac{v-u}{\|v-u\|_X}$, then $\|x\|_X = 1$ and

$$\begin{aligned}\mathrm{dist}(x, U) = \inf\{\|\tfrac{v-u}{\|v-u\|_X} - z\|_X \mid z \in U\} &= \tfrac{1}{\|v-u\|_X} \inf\{\|v - (u + \|v - u\|_X z)\|_X \mid z \in U\} \\ &= \tfrac{1}{\|v-u\|_X} \mathrm{dist}(v, U) > 1 - \epsilon\end{aligned}$$ □

Corollary 44 (Closed balls are noncompact in infinite dimensions)**.** *Let X be an infinite-dimensional Banach space. There exists a sequence $x_n \in X$ with $\|x_n\|_X = 1$ and $\|x_n - x_m\| \geq \frac{1}{2}$ for all $m \neq n$ (thus x_n contains no limit point).*
In particular, the closed unit ball in X is not compact, and the identity is not a compact operator on X. A closed ball on a Banach space is compact iff the space is finite-dimensional.

Proof.

- pick $x_1 \in X$ with $\|x_1\|_X = 1$
- pick $x_n \in X \setminus \mathrm{span}\{x_1, \ldots, x_{n-1}\}$ with $\|x_n\|_X = 1$ & $\mathrm{dist}(x_n, \mathrm{span}\{x_1, \ldots, x_{n-1}\}) > \frac{1}{2}$ □

Theorem 45 (Compact operators in infinite dimensions have no bounded inverse)**.** *Let X be an infinite-dimensional Banach space, Y a Banach space and $K : X \to Y$ compact. Then K has no bounded inverse. In particular, the inverse problem $Kx = y$ is ill-posed.*

Proof.

- let x_n the previous sequence & $y_n = Kx_n$
- K compact $\Rightarrow \exists$ convergent subsequence $y_n \to y \in Y$,
- but $x_n = K^{-1}y_n$ does not converge □

Theorem 46 (Bounded inverse theorem)**.** *Let X, Y be Banach spaces. A bijective linear continuous operator $L : X \to Y$ has a continuous inverse.*

Proof. L is surjective and thus by the open mapping theorem open. Thus preimages LU of open sets $U \subset X$ under L^{-1} are again open. □

Note: The previous result only holds on Banach spaces! Consequently, on Banach spaces, compact operators cannot be bijective, thus have no inverse!

Remark 47 (Conditional stability)**.** *Sometimes the inverse is continuous on certain subsets of the image (one speaks of* conditional stability*): E. g., let X, Y be Hilbert spaces, $K : X \to Y$ continuous, $x_i = K^H w_i$, $y_i = Kx_i$, $i = 1, 2$, then*

$$\|x_1-x_2\|_X^2 = (x_1-x_2, x_1-x_2) = (x_1-x_2, K^*(w_1-w_2)) = (y_1-y_2, w_1-w_2) \leq \|y_1-y_2\|_Y(\|w_1\|_Y+\|w_2\|_Y)$$

$\Rightarrow$ *For $W_C = \{x \in X \mid x = K^H w, \|w\|_Y < C\}$ there is a Hölder continuous inverse to $K : W_C \to KW_C$.*

6 The Riesz theorems

Our next aim is to understand the spectrum of compact operators $K : X \to X$ and their singular value decomposition (SVD). As for matrices, size differences of the singular values contain information about the stability of the inversion. The theory was developed by Frigyes Riesz. In this section we present the three main preparatory theorems that bear his name, which consider the operator $I - K$ (which should remind us of eigenvalues and eigenvectors). Since we only defined compact operators for Banach spaces we let X be a Banach space, but with obvious modifications everything would work for just normed vector spaces. We write $\|\cdot\|$ for $\|\cdot\|_X$.

Theorem 48 (Riesz' 1. theorem)**.** *Let X be a Banach space, $K : X \to X$ linear and compact. The kernel of $I - K$ is finite-dimensional.*

Proof.
- assume, $\ker(I - K)$ is infinite-dimensional
- let $x_n \in \ker(I - K)$ bounded sequence without converging subsequence *(Riesz lemma)*
- $x_n = Kx_n$, however, has a convergent subsequence ↯ □

Theorem 49 (Riesz' 2. theorem)**.** *Let X be a Banach space, $K : X \to X$ linear and compact. Then the range* $\operatorname{ran}(I - K)$ *is closed.*

Proof.
- let $\tilde{x}_n \in X$ and $y \in X$ with $(I - K)\tilde{x}_n \to y$
- set $x_n = \arg\min\{\|x\|^2 \mid x \in \tilde{x}_n + \ker(I - K)\}$
  - x_n is well-defined: minimizes coercive, cont. functional on finite-dim. space *(Riesz' 1. thm.)*
  - $\operatorname{dist}(x_n, \ker(I - K)) = \|x_n\|$
  - $(I - K)x_n \to y$ by definition
  - x_n is bounded: Otherwise,

$$\left.\begin{array}{l}(I-K)x_n/\|x_n\|_X \to 0\\ Kx_n/\|x_n\| \to z \text{ for subsequence}\end{array}\right\} \Rightarrow \frac{x_n}{\|x_n\|} \to z \ \&\ (I-K)z = 0$$

$$\Rightarrow\ 1 = \frac{\operatorname{dist}(x_n, \ker(I-K))}{\|x_n\|} \le \frac{\|x_n - \|x_n\|z\|}{\|x_n\|} = \left\|\frac{x_n}{\|x_n\|} - z\right\| \to 0 \ ↯$$

- along a subsequence, $Kx_n \to w \in X$, thus $\|x_n - y - w\| \le \|x_n - Kx_n - y\| + \|Kx_n - w\| \to 0$
- continuity of $(I - K) \Rightarrow y = \lim_{n\to\infty}(I - K)x_n = (I - K)(y + w)$ □

Theorem 50 (Riesz' 3. theorem)**.** *Let X be a Banach space, $K : X \to X$ linear and compact. Then there exists $r \in \mathbb{N}$ such that*

$$\begin{aligned}&\ker(I-K)^l \subsetneq \ker(I-K)^{l+1}, && \operatorname{ran}(I-K)^l \supsetneq \operatorname{ran}(I-K)^{l+1}, && \textit{if } l < r,\\ &\ker(I-K)^l = \ker(I-K)^{l+1}, && \operatorname{ran}(I-K)^l = \operatorname{ran}(I-K)^{l+1}, && \textit{if } l \ge r.\end{aligned}$$

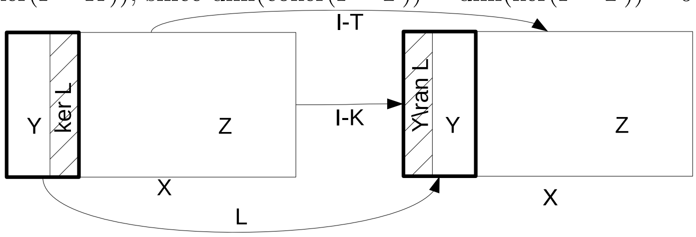
Furthermore, $X = \ker(I - K)^r \oplus \operatorname{ran}(I - K)^r$.

Proof.
- set $V_l = \operatorname{ran}(I - K)^l$ and show its properties:
  - $V_l \supset V_{l+1}$ by definition
  - $V_l = V_{l+1} \Rightarrow V_l = V_m$ for all $m > l$
  - assume $V_l \supsetneq V_{l+1}$ for all l
    * let $x_l \in V_l$ with $\|x_l\| = 1$ and $\operatorname{dist}(x_l, V_{l+1}) \ge \frac{1}{2}$ *(Riesz lemma & Riesz' 2. thm.)*
    * $\|Kx_l - Kx_m\| = \|x_l - x_m - (I-K)(x_l - x_m)\| = \|x_l - \underbrace{(x_m + (I-K)(x_l - x_m))}_{\in V_{l+1}}\| > \frac{1}{2}$

    for all $m > l$

    $\Rightarrow Kx_l$ has no convergent subsequence ↯

- set $W_l = \ker(I-K)^l$ and show its properties (potentially with different r): Homework (same way)
- $\dim(\ker(I-K)^l) = \dim(\operatorname{coker}(I-K)^l)$ for all l:
  - it suffices to show this for $l = 1$ (since $(I-K)^l = (I-L)$ for a compact operator L)
  - it suffices to show this for injective $I-K$:
    If $\dim(\ker(I-T)) = \dim(\operatorname{coker}(I-T))(=0)$ for every compact T with injective $I-T$, then:
    * "restrict $I-K$ (or equivalently K) to a subspace Z on which $I-K$ is injective by modding out the kernel, take $I-T = (I-K)|_Z$"
      · $Y = \ker(I-K)^r = \ker(I-K)^{r+1}$ satisfies $(I-K)Y \subset Y$, thus $KY \subset Y$
      · set $Z = X/Y$ with norm $\|z\|_Z = \inf\{\|x\| \,|\, x \in z+Y\}$
      · K induces a compact operator $T : Z \to Z$
      · $I-T$ is injective:
        otherwise there would be $x \notin Y$ with $(I-K)x \in Y$, thus $x \in \ker(I-K)^{r+1} = Y$
    * "replace $\ker(I-K)$ by the kernel of a finite-dimensional operator L for which thus $\dim\ker L = \dim\operatorname{coker} L$"
      · set $L : Y \to Y$, $L = (I-K)|_Y$, then $\ker(I-K) = \ker L$
      · $\dim(\ker(I-K)) = \dim(\ker L) = \dim(\operatorname{coker} L)$, since Y finite-dimensional *(Riesz 1. thm.)*
        $= \dim(\operatorname{coker}(I-K))$, since $\dim(\operatorname{coker}(I-T)) = \dim(\ker(I-T)) = 0$



  - still to show (for injective $I-K$): $\operatorname{coker}(I-K) = \{0\}$, i. e. $I-K$ surjective
    * assume $\operatorname{ran}(I-K) \subsetneq X$, then also $\operatorname{ran}(I-K)^l \supsetneq \operatorname{ran}(I-K)^{l+1}$ for all l ↯,
      since otherwise there would for every $x \in X$ be a $y \in X$ with $(I-K)^{l+1}y = (I-K)^l x$
      $\overset{I-K \text{ injective}}{\Rightarrow} (I-K)^l y = (I-K)^{l-1}x \Rightarrow \ldots \Rightarrow (I-K)y = x$ ↯
- either $\ker(I-K)^l = \ker(I-K)^{l+1}$ & $\operatorname{ran}(I-K)^l = \operatorname{ran}(I-K)^{l+1}$ or
  $\ker(I-K)^l \neq \ker(I-K)^{l+1}$ & $\operatorname{ran}(I-K)^l \neq \operatorname{ran}(I-K)^{l+1}$ (else contradiction to previous point)
  $\Rightarrow$ critical exponent r is the same for kernel and range
- let $0 = a + b$ with $a \in \ker(I-K)^r$, $b = (I-K)^r\beta$
  $\Rightarrow 0 = (I-K)^r a + (I-K)^r b = (I-K)^{2r}\beta$
  $\Rightarrow \beta \in \ker(I-K)^{2r} = \ker(I-K)^r \Rightarrow b = 0 \Rightarrow a = 0$
  $\overset{\dim(\ker(I-K)^r) = \dim(\operatorname{coker}(I-K)^r)}{\Rightarrow} X = \ker(I-K)^r \oplus \operatorname{ran}(I-K)^r$ □

7 The SVD of compact operators on Hilbert spaces

Definition 51 (Spectrum). *Let X be a normed $\mathbb{C}$-vector space, $K : X \to X$ linear and bounded.*

1. *The* spectrum *of K is the set $\sigma(K) = \{\lambda \in \mathbb{C} \,|\, \lambda I - K$ has no continuous inverse$\}$.*
2. *An* eigenvalue *of K is a $\lambda \in \mathbb{C}$ such that there exists a corresponding* eigenvector *$u \in X$ with $Ku = \lambda u$.*

Theorem 52 (Spectrum of compact operators). *Let X be an infinite-dimensional Banach space, $K : X \to X$ compact.*

1. $0 \in \sigma(K)$
2. $\lambda \in \sigma(K) \setminus \{0\} \Rightarrow \lambda$ *is eigenvalue of K with finite geometric multiplicity* $\dim(\ker(\lambda I - K))$

3. $\sigma(K)$ is countable with 0 as the only limit point

Proof. 1. already proven

2. Assume λ is no eigenvalue, i. e. $\ker(I - \frac{1}{\lambda}K) = \{0\}$
$\xRightarrow{\text{Riesz 3. thm.}} \operatorname{ran}(I - \frac{1}{\lambda}K) = X \xRightarrow{\text{bounded inv. thm.}} (I - \frac{1}{\lambda}K)^{-1}$ is continuous $\lightning$
$\dim(\ker(\lambda I - K)) = \dim(\ker(I - \frac{1}{\lambda}K)) < \infty$ already shown *(Riesz 1. thm.)*

3.
   - Let $\lambda_n \in \sigma(K)$ mutually different with $\lambda_n \to \lambda \neq 0$ and eigenvectors $x_n \in X$.
     Set $X_n = \operatorname{span}\{x_1, \ldots, x_n\}$ and choose $z_n \in X_n$ with $\|z_n\| = 1$, $\operatorname{dist}(z_n, X_{n-1}) \geq \frac{1}{2}$.
     Then Kz_n contains no convergent subsequence ($\Rightarrow \lightning$):
     Let $z_n = \sum_{i=1}^{n} \alpha_i x_i \Rightarrow Kz_n - \lambda_n z_n = \sum_{i=1}^{n-1} \alpha_i(\lambda_i - \lambda_n)x_i \in X_{n-1}$, thus
     $\|Kz_n - Kz_m\| = \|\lambda_n z_n - \underbrace{(Kz_m - (Kz_n - \lambda_n z_n))}_{\in X_{n-1}}\| \geq \operatorname{dist}(\lambda_n z_n, X_{n-1}) \geq \frac{\lambda_n}{2}\ \forall m < n.$
   - $\sigma(K) \subset \overline{B_{\|K\|}(0)} \subset \mathbb{C}$ & 0 is only limit point
     $\Rightarrow \sigma(K) \setminus B_{\frac{1}{n}}(0)$ is finite $\forall n \in \mathbb{N}$
     $\Rightarrow$ all elements can be numbered ☐

From now on let X, Y be real Hilbert spaces so that $X^* \equiv X$ and $Y^* \equiv Y$. Then we can form $K^H K$ and KK^H. Just as for matrices the singular values are going to be $\sigma_i = \sqrt{\lambda_i}$ for λ_i the eigenvalues of the positive semi-definite symmetric operator $K^H K : X \to X$. The largest singular value is $\|K\|$.

Definition 53 (Singular values & vectors)**.** *Let X, Y be Hilbert spaces, $K : X \to Y$ linear and compact. The* singular values *of K are*

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \ldots > 0,$$

where $(\sigma_n^2)_n$ are the nonzero eigenvalues of $K^H K$, counted with geometric multiplicity. The right singular vectors *of K are the corresponding normed eigenvectors $u_n \in X$, the* left singular vectors *are $v_n = Ku_n/\|Ku_n\|$ (for $\sigma_n > 0$).*

Theorem 54 (Singular value decomposition)**.** *Let X, Y be Hilbert spaces, $K : X \to Y$ linear and compact.*

1. v_n is eigenvector to eigenvalue σ_n^2 for KK^H

2. $\sigma(K^H K) = \sigma(KK^H)$, and the eigenspaces of σ_n^2 for $K^H K$ and KK^H have the same dimension

3. $Ku_n = \sigma_n v_n$ and $K^H v_n = \sigma_n u_n\ \forall n \in \mathbb{N}$

4. $\{u_n\}$ and $\{v_n\}$ are complete orthonormal systems in $\overline{\operatorname{ran} K^H K} = \overline{\operatorname{ran} K^H} = (\ker K)^\perp$ and $\overline{\operatorname{ran} KK^H} = \overline{\operatorname{ran} K} = (\ker K^H)^\perp$, respectively
(as long as eigenvectors of the same eigenvalue are chosen orthogonally).

Proof. 1. $KK^H v_n = K(K^H K u_n)/\|Ku_n\| = \sigma_n^2 Ku_n/\|Ku_n\| = \sigma_n v_n$

2. By 1, for every eigenvector of $K^H K$ there exists one of KK^H with same eigenvalue.
Analogously, for every eigenvector v of KK^H one obtains an eigenvector $K^H v/\|K^H v\|$ of $K^H K$ with same eigenvalue.
Furthermore, K and K^H are injective on the eigenspaces of $K^H K$ and KK^H, respectively.

3. $\|Ku_n\|^2 = (Ku_n, Ku_n) = (u_n, K^H K u_n) = \sigma_n^2(u_n, u_n) = \sigma_n^2$
$\Rightarrow v_n = Ku_n/\|Ku_n\| = Ku_n/\sigma_n$ & $K^H v_n = K^H K u_n/\|Ku_n\| = \sigma_n u_n$

4. u_n and v_n are normed by definition.
Orthogonality for $\sigma_n \neq \sigma_m$ follows from $\sigma_n^2(u_n, u_m) = (K^H K u_n, u_m) = (u_n, K^H K u_m) = \sigma_m^2(u_n, u_m)$ (analogous for v_n).
Let $U = \operatorname{span}\{u_1, u_2, \ldots\}$; need to show $U^\perp = \ker K$.
$K^H K(U^\perp) \subset U^\perp$: if $(u_n, K^H K u) \neq 0$ for some $u \in U^\perp$, then $0 \neq (K^H K u_n, u) = \sigma_n^2(u_n, u)\ \lightning$.
$\Rightarrow K^H : \overline{KU^\perp} \to U^\perp$
$\Rightarrow$ We can restrict K to a compact operator $L : U^\perp \to \overline{KU^\perp}$ with adjoint $L^H = K^H : \overline{KU^\perp} \to U^\perp$.

$\Rightarrow \exists u \in U^\perp$ with $\|u\| = 1$, $\|Lu\| = \|L\|$
$\Rightarrow \|L\|^2 = \|Lu\|^2 = (Lu, Lu) = (u, L^H Lu) \leq \|L^H Lu\| \leq \|L\|^2$ with equality only if $L^H Lu = \alpha u$
$\Rightarrow u$ is eigenvector to eigenvalue $\alpha = \|L\|^2$ of $L^H L$ and thus of $K^H K \Rightarrow L = 0 \Rightarrow U^\perp = \ker K$ □

Corollary 55 (SVD)**.** $Kx = K \sum_{n=1}^\infty (x, u_n) u_n = \sum_{n=1}^\infty (x, u_n) \sigma_n v_n$ *&* $K^H y = \sum_{n=1}^\infty (y, v_n) \sigma_n u_n$.

Corollary 56 (Picard criterion)**.** *Let X, Y be Hilbert spaces, $K : X \to Y$ linear and compact, $f \in \overline{\operatorname{ran} K}$. $Ku = f$ has a solution iff $\sum_{n=1}^\infty \frac{(f,v_n)^2}{\sigma_n^2} < \infty$.*

Proof. '⇒' Let $Ku = f$, then $(f, v_n) = (Ku, v_n) = (u, K^H v_n) = \sigma_n (u, u_n)$.
$\Rightarrow \sum_{n=1}^\infty \frac{(f,v_n)^2}{\sigma_n^2} = \sum_{n=1}^\infty (u, u_n)^2 \leq \|u\|^2$

'⇐' Set $u = \sum_{n=1}^\infty \frac{(f,v_n)}{\sigma_n} u_n$, then $Ku = \sum_{n=1}^\infty (f, v_n) v_n = f$ and $u \in X$ due to the condition. □

Definition 57 (Mildly and severely ill-posed)**.** *The inverse problem $Ku = f$ is called*

- severely ill-posed *if $\sigma_n = o(n^{-\alpha})$ for all $\alpha > 0$ (e.g. inverse heat flow),*
- mildly ill-posed *if $\sigma_n = O(n^{-\alpha})$ for some $\alpha > 0$ and it is not severly ill-posed (e.g. X-ray transform).*

Example 58 (Integration/differentiation mildly ill-posed)**.** $K : L^2((0,1)) \to L^2((0,1))$, $Ku(x) = \int_0^x u(s)\,\mathrm{d}s$
$\Rightarrow K^H v(y) = \int_y^1 v(s)\,\mathrm{d}s \Rightarrow K^H K u(y) = \int_y^1 \int_0^x u(s)\,\mathrm{d}s\,\mathrm{d}x$
Let σ^2 be eigenvalue of $K^H K$ with eigenvector u, i. e. $w \equiv K^H K u = \sigma^2 u$,
then $w''(s) = -u(s) = \frac{w(s)}{\sigma^2}$ with $w'(0) = 0$, $w(1) = 0$.
$\Rightarrow w(s) = \alpha \cos(\frac{s}{\sigma})$ *with* $w(1) = 0 \Rightarrow \sigma = \frac{2}{(2n-1)\pi}$, $u(s) = \frac{\alpha}{\sigma^2} \cos(\frac{s}{\sigma}) \Rightarrow \sigma_n = O(\frac{1}{n})$

The singular values help to better understand the effect of noise: Let $K : X \to Y$ compact, $Ku = f$, $Ku^\delta = f^\delta$ noisy measurement of f. We have

$$\|u^\delta - u\|_X^2 = \sum_{n=1}^\infty (u^\delta - u, u_n)^2 = \sum_{n=1}^\infty \frac{(f^\delta - f, v_n)^2}{\sigma_n^2}.$$

⇒ Noise at higher "frequencies" $\frac{1}{\sigma_n}$ (meaning noise components in $\operatorname{span}\{v_n\}$) is amplified more.

8 Generalized inverse

Even if an inverse problem has a solution it might not be unique. Likewise, the measurement may contain a component outside the range of the forward operator, which thus can actually be ignored. Both these situations refer to the first two conditions of well-definedness, injectivity and surjectivity. We now define how the solution of an inverse problem or its regularization should behave in these situations (on Hilbert spaces).

Definition 59 (Orthogonal projection)**.** *Let X be a Hilbert space, $M \subset X$ a closed subspace. The* orthogonal projection $\mathrm{P}_M : X \to M$ *is defined by*

$$(\mathrm{P}_M x, v) = (x, v) \quad \forall v \in M.$$

Theorem 60 (Orthogonal projection)**.** *The orthogonal projection is well-posed, linear, and continuous with $\|\mathrm{P}_M\| \leq 1$.*

Proof. Homework. □

Definition 61 (Least squares & minimum-norm solution)**.** *Let X, Y be Hilbert spaces, $A : X \to Y$ linear and continuous, $b \in Y$.*

1. *$x \in X$ is called* least squares solution *of $Ax = b$ if $\|Ax - b\|_Y \leq \|Az - b\|_Y$ $\forall z \in X$.*
2. *A least squares solution x is called* minimum norm solution *of $Ax = b$ if $\|x\|_X \leq \|z\|_X$ for all least squares solutions z.*

Theorem 62 (Least squares solution). *The following are equivalent:*

1. *x is least squares solution*
2. *$A^H Ax = A^H b$* $\quad (A^H$ applied to $Ax = b)$
3. *$Ax = \mathrm{P}_{\overline{\operatorname{ran} A}} b$*

Proof. 1.⇒2. opt. condition: $0 = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|A(x+tz)-b\|_Y^2|_{t=0} = (Ax-b, Az) = (A^H Ax - A^H b, z)\ \forall z \in X$

2.⇒3. 2. $\Rightarrow (Ax-b, Az) = 0$ for all $z \in X \Rightarrow (Ax-b, v) = 0$ for all $v \in \operatorname{ran} A \Rightarrow Ax = \mathrm{P}_{\overline{\operatorname{ran} A}} b$

3.⇒1. $\|Az-b\|^2 = \|(Ax-b)+A(z-x)\|^2 \overset{Ax-b\in(\operatorname{ran} A)^\perp \,\&\, A(z-x)\in\operatorname{ran} A}{=} \|(Ax-b)\|^2 + \|A(z-x)\|^2$ □

The last characterization shows that a least squares solution might not exist, e. g. if $\operatorname{ran} A$ is dense in Y.

Theorem 63 (Domain of minimum norm solution). *Let $b \in \operatorname{ran} A \oplus (\operatorname{ran} A)^\perp$.*

1. *A least squares solution exists.*
2. *The minimum norm solution is well-defined.*

Proof. 1. Let $b = Ax + w$ with $w \in (\operatorname{ran} A)^\perp = \ker A^H$,
then x is a least squares solution due to $A^H Ax = A^H(Ax+w) = A^H b$.

2. Homework (analogous to theorem 10) □

Definition 64 (Moore–Penrose inverse). *Let X, Y be Hilbert spaces, $A : X \to Y$ linear and bounded, $B : (\ker A)^\perp \to \operatorname{ran} A$ with $B = A|_{(\ker A)^\perp}$. The* Moore–Penrose (generalized) inverse *is the unique linear extension*

$$A^+ : \operatorname{ran} A \oplus (\operatorname{ran} A)^\perp \to (\ker A)^\perp$$

of B^{-1} with $\ker A^+ = (\operatorname{ran} A)^\perp$.

Note that the minimum norm solution and the Moore–Penrose inverse are not defined on all of Y, e. g. if $\operatorname{ran} A$ is dense in Y, then $(\operatorname{ran} A)^\perp = \{0\}$.

Theorem 65 (Moore–Penrose inverse). *The Moore–Penrose inverse is uniquely determined by*

1. *$AA^+A = A$*
2. *$A^+AA^+ = A^+$*
3. *$A^+A = I - \mathrm{P}_{\ker A}$*
4. *$AA^+ = \mathrm{P}_{\overline{\operatorname{ran} A}}|_{\operatorname{ran} A \oplus (\operatorname{ran} A)^\perp}$*

Proof. '⇒' (1) and (2) follow from (3) and (4), (3) and (4) follow from definition of A^+

'⇐'
- domain of A^+ follows from (4)
- range of A^+ follows from (2) & (3)
- $\ker A^+ \supset (\operatorname{ran} A)^\perp$ follows from (2) & (4)
- $A^+ = B^{-1}$ on $(\ker A)^\perp = \overline{\operatorname{ran} A}$ follows from (3) $\Rightarrow \ker A^+ = (\operatorname{ran} A)^\perp$ □

Theorem 66 (Minimum norm solution and Moore–Penrose inverse). *Let $b \in \operatorname{ran} A \oplus (\operatorname{ran} A)^\perp$. The minimum norm solution of $Ax = b$ is $x^* = A^+b$, and the least squares solutions are $x^* + \ker A$.*

Proof. Homework. □

Remark 67 (Moore–Penrose inverse of compact operator). *Let $K : X \to Y$ linear and compact with SVD σ_n, u_n, v_n and $b \in \operatorname{ran} K \oplus (\operatorname{ran} K)^\perp$, then*

$$\sum_{n=1}^\infty \sigma_n^2 (K^+b, u_n) u_n = K^H K K^+ b = K^H \mathrm{P}_{\overline{\operatorname{ran} K}} b = K^H b = \sum_{n=1}^\infty \sigma_n (b, v_n) u_n$$

$$\Rightarrow K^+ b = \sum_{n=1}^\infty (K^+b, u_n) u_n = \sum_{n=1}^\infty \frac{(b, v_n)}{\sigma_n} u_n.$$

9 Linear regularization

The Moore–Penrose inverse K^+ solves the inverse problem $Kx = y$ for $y \in \text{dom}(K^+) = \text{ran}\, K \oplus (\text{ran}\, K)^\perp$ and thus addresses the existence and uniqueness problems of an ill-posed inverse problem. However, K^+ is in general not continuous so that noise in y prevents the solution of $Kx = y$. This problem is solved by regularization.

Definition 68 (Regularization). *A family of continuous linear operators $R_\alpha : Y \to X$, $\alpha > 0$, and a map*

$$\alpha : (0, \infty) \times Y \to (0, \infty), \quad (\delta, y^\delta) \mapsto \alpha(\delta, y^\delta)$$

for the choice of the regularization parameter α *is called a* regularization of K^+ *if for every sequence y^δ with $\|y^\delta - y\| \le \delta$ one has*

$$R_{\alpha(\delta,y^\delta)} y^\delta \xrightarrow[\delta \to 0]{} K^+ y.$$

The parameter choice is called a priori *if $\alpha(\delta, y^\delta) = \alpha(\delta)$ (cf. corollary 72), else* a posteriori *(cf. remark 83). It is conventionally chosen such that $\alpha \to 0$ as $\delta \to 0$.*

For compact operators K on Hilbert spaces, one can obtain regularization operators R_α for K^+ via the SVD of K, by approximating

$$K^+ y = \sum_{n=1}^{\infty} \frac{(y, v_n)}{\sigma_n} u_n \qquad \text{with} \qquad R_\alpha y = \sum_{n=1}^{\infty} g_\alpha(\sigma_n)(y, v_n) u_n$$

for some $g_\alpha : (0, \infty) \to [0, \infty)$. For R_α to be an admissible regularization of K^+ one needs $g_\alpha(t) \to \frac{1}{t}$ for $\alpha \to 0$. In order to check the convergence, the error is estimated by

$$\|R_\alpha y^\delta - K^+ y\| \le \underbrace{\|R_\alpha y^\delta - R_\alpha y\|}_{\text{propagated measurement error}} + \underbrace{\|R_\alpha y - K^+ y\|}_{\text{approximation error}} .$$

Theorem 69 (Approximation error). *Let X, Y Hilbert spaces, $K : X \to Y$ linear and compact with SVD (σ_n, u_n, v_n), let $\Sigma = \{\sigma_1, \sigma_2, \ldots\} \subset (0, \|K\|]$. Let $R_\alpha : Y \to X$, $R_\alpha y = \sum_{n=1}^\infty g_\alpha(\sigma_n)(y, v_n) u_n$ with $g_\alpha : (0, \infty) \to [0, \infty)$.*

1. *R_α is continuous iff*
$$C_\alpha = \sup\{g_\alpha(t) \mid t \in \Sigma\} < \infty.$$
2. *$R_\alpha \to K^+$ pointwise on* $\text{dom} K^+$ *as $\alpha \to 0$ iff*
$$\gamma = \limsup_{\alpha \to 0} \sup\{\sigma g_\alpha(\sigma) \mid \sigma \in \Sigma\} < \infty,$$
$$g_\alpha(t) \to \frac{1}{t} \textit{ pointwise on } \Sigma \textit{ for } \alpha \to 0.$$

Proof. 1. "⇒": $\|R_\alpha y\|^2 = \sum_{n=1}^\infty |g_\alpha(\sigma_n)|^2 |(y, v_n)|^2 \le C_\alpha^2 \sum_{n=1}^\infty |(y, v_n)|^2 \le C_\alpha^2 \|y\|^2$
"⇐": Let $\epsilon > 0$ and $n \in \mathbb{N}$ with $g_\alpha(\sigma_n) > C_\alpha - \epsilon$, then $\|R_\alpha v_n\| > C_\alpha - \epsilon$.

2. "⇒":

- $\|R_\alpha y - K^+ y\|^2 = \sum_{n=1}^\infty |g_\alpha(\sigma_n) - \frac{1}{\sigma_n}|^2 |(y, v_n)|^2 = \sum_{n=1}^\infty |\sigma_n g_\alpha(\sigma_n) - 1|^2 |\frac{(y,v_n)}{\sigma_n}|^2$
- Picard criterion ⇒ $\sum_{n=1}^\infty |\frac{(y,v_n)}{\sigma_n}|^2 < \infty$
- for $\varepsilon > 0$ choose $N \in \mathbb{N}$ with $\sum_{n=N+1}^\infty |\frac{(y,v_n)}{\sigma_n}|^2 < \frac{\varepsilon}{2(1+\gamma)^2}$
and $\alpha_0 > 0$ with $\sum_{n=1}^N |\sigma_n g_\alpha(\sigma_n) - 1|^2 |\frac{(y,v_n)}{\sigma_n}|^2 < \frac{\varepsilon}{2}$ $\forall \alpha < \alpha_0$
- $\|R_\alpha y - K^+ y\|^2 = \underbrace{\sum_{n=1}^N |\sigma_n g_\alpha(\sigma_n) - 1|^2 \left|\frac{(y, v_n)}{\sigma_n}\right|^2}_{\le \frac{\varepsilon}{2}} + \underbrace{\sum_{n=N+1}^\infty \underbrace{|\sigma_n g_\alpha(\sigma_n) - 1|^2}_{\le (1+\gamma)^2} \left|\frac{(y, v_n)}{\sigma_n}\right|^2}_{\le \frac{\varepsilon}{2}} \le \varepsilon$

"$\Leftarrow$":

- if $g_\alpha(\sigma_n) \not\to \frac{1}{\sigma_n}$ as $\alpha \to 0$, then $R_\alpha v_n = g_\alpha(\sigma_n) u_n \not\to \frac{u_n}{\sigma_n} = K^+ v_n$
- Suppose $\alpha_j \to 0$, $n_j \to \infty$, and $S_j = \sigma_{n_j} g_{\alpha_j}(\sigma_{n_j}) \to \infty$ as $j \to \infty$, wlog $S_j > j+1$. Set $y = \sum_{j=1}^\infty \frac{\sigma_{n_j}}{j} v_{n_j}$, then $y \in \mathrm{dom} K^+$ due to $\sum_{n=1}^\infty |\frac{(y,v_n)}{\sigma_n}|^2 = \sum_{j=1}^\infty \frac{1}{j^2} < \infty$, but $\|R_{\alpha_j} y - K^+ y\| \geq |S_j - 1| \frac{|(y,v_{n_j})|}{\sigma_{n_j}} = |S_j - 1| \frac{1}{j} > 1$ for all j ↯ □

Remark 70 (Sufficient conditions)**.** *Replacing Σ with $(0, \|K\|]$ or even $(0, \infty)$ in the definitions of C_α or γ in theorem 69, one still gets sufficient conditions for boundedness and pointwise convergence of R_α.*

Remark 71 (Convergence speed)**.** *The convergence $R_\alpha y \to K^+ y$ can be arbitrarily slow. Indeed, given an arbitrarily small $\alpha_0 > 0$ and arbitrarily large $C > 0$, one can find n large enough such that $\|R_\alpha v_n - K^+ v_n\| = |g_\alpha(\sigma_n) - \frac{1}{\sigma_n}| > \frac{1}{\sigma_n} - \sup_{\alpha \geq \alpha_0} C_\alpha > C$ for all $\alpha \geq \alpha_0$.*

Corollary 72 (Convergence of regularization)**.** *Let K, g_α as above and $\alpha(\delta, y^\delta) \to_{\delta \to 0} 0$ such that $C_{\alpha(\delta, y^\delta)} \delta \to_{\delta \to 0} 0$, then*

$$R_{\alpha(\delta,y^\delta)} y^\delta \xrightarrow[\delta \to 0]{} K^+ y.$$

Proof. $\|R_\alpha y^\delta - K^+ y\| \leq \underbrace{\|R_\alpha\| \|y^\delta - y\|}_{\leq C_\alpha \delta \to 0} + \underbrace{\|R_\alpha y - K^+ y\|}_{\to 0}$ □

Example 73 (Different regularizations)**.** *1. Truncated singular value decomposition, TSVD*

$$g_\alpha(t) = \begin{cases} 0, & t \leq \alpha \\ \frac{1}{t}, & t > \alpha \end{cases} \qquad \Rightarrow \qquad C_\alpha = \frac{1}{\alpha},\ \gamma = 1$$

2. Lavrentiev regularization

$$g_\alpha(t) = \frac{1}{t + \alpha} \qquad \Rightarrow \qquad C_\alpha = \frac{1}{\alpha},\ \gamma = 1$$

3. Tikhonov regularization

$$g_\alpha(t) = \frac{t}{t^2 + \alpha} \qquad \Rightarrow \qquad C_\alpha = \frac{1}{2\sqrt{\alpha}},\ \gamma = 1$$

To implement the TSVD it suffices to determine the first singular values and vectors; for the Lavrentiev and Tikhonov regularization one can exploit the following.

Theorem 74 (Equivalent characterization of Lavrentiev and Tikhonov regularization)**.**

1. *Let R_α be the Tikhonov regularization operator, then $R_\alpha y = \arg\min_x \|Kx - y\|^2 + \alpha \|x\|^2$.*
2. *Let R_α be the Lavrentiev regularization operator and $U : X \to Y$, $U u_n = v_n$ $\forall n$ ($U|_{\overline{\mathrm{ran}\, K^H}}$ is an isometry; if $K = K^H : X \to Y = X$ with $\mathrm{ran}\, K$ dense in X, then $U = \mathrm{Id}$), then $R_\alpha = (K + \alpha U)^+$.*

Proof. Homework. □

Under additional regularity conditions on y or $K^+ y$ one can achieve convergence rates.

Definition 75 (Source condition)**.** *A* source condition *is a regularity condition on $x = K^+ y$ of the form*

$$x = (K^H K)^\mu w \text{ for some } \mu > 0 \text{ and } w \in X.$$

Remark 76 (Regularity in source condition)**.** *The source condition depends on K, i. e. the regularity is measured in terms of K. The higher μ the more regular are y and x.*

Theorem 77 (Error bound under source condition)**.** *Under the source condition $x = K^+ y = (K^H K)^\mu w$ we have*

$$\|R_\alpha y - K^+ y\| \leq \varphi_\mu(\alpha) \|w\| \qquad \text{for } \varphi_\mu(\alpha) = \max_{\sigma \in (0, \|K\|]} |g_\alpha(\sigma) \sigma^{2\mu+1} - \sigma^{2\mu}|.$$

Proof.
- $\sum_{n=1}^\infty \frac{(y,v_n)}{\sigma_n} u_n = K^+y = x = \sum_{n=1}^\infty \sigma_n^{2\mu}(w,u_n)u_n \Rightarrow (w,u_n) = \frac{(y,v_n)}{\sigma_n^{2\mu+1}}$
- $\|R_\alpha y - K^+y\|^2 = \sum_{n=1}^\infty \underbrace{\left[(g_\alpha(\sigma_n) - \frac{1}{\sigma_n})\sigma_n^{2\mu+1}\right]^2}_{\le \varphi_\mu(\alpha)^2} \underbrace{\left[\frac{(y,v_n)}{\sigma_n^{2\mu+1}}\right]^2}_{=(w,u_n)^2}$ □

Remark 78 (Convergence rates under source condition)**.** *Under the source condition we thus have* $\|R_\alpha y^\delta - K^+y\| \le C_\alpha \delta + \varphi_\mu(\alpha)\|w\|$*. The right-hand side can now be minimized for* α *to obtain an optimal convergence rate (and the associated choice of the regularization parameter* $\alpha(\delta)$*).*

Example 79 (Convergence rates under source condition)**.** *1. TSVD:*

$$\varphi_\mu(\alpha) = \sup_{\sigma>0}\left|\sigma^{2\mu+1}\cdot\begin{Bmatrix}0, & \sigma\le\alpha\\ \frac{1}{\sigma}, & \sigma>\alpha\end{Bmatrix} - \sigma^{2\mu}\right| = \sup_{\sigma>0}\sigma^{2\mu}\cdot\begin{Bmatrix}1, & \sigma\le\alpha\\ 0, & \sigma>\alpha\end{Bmatrix} = \alpha^{2\mu}, \quad C_\alpha = \frac{1}{\alpha}$$

$$C_\alpha\delta + \varphi_\mu(\alpha)\|w\| \to \min! \quad \Rightarrow \quad \alpha(\delta) = \left(\frac{\delta}{2\mu\|w\|}\right)^{\frac{1}{2\mu+1}}$$

$$\Rightarrow \quad \|R_{\alpha(\delta)}y^\delta - K^+y\| \le C\|w\|^{\frac{1}{2\mu+1}}\delta^{\frac{2\mu}{2\mu+1}}$$

2. Lavrentiev:

$$\varphi_\mu(\alpha) = \sup_{\sigma\le\|K\|}\left|\sigma^{2\mu+1}\frac{1}{\sigma+\alpha} - \sigma^{2\mu}\right| = \sup_{\sigma\le\|K\|}\frac{\sigma^{2\mu}\alpha}{\sigma+\alpha} = \begin{Bmatrix}\frac{\|K\|^{2\mu}\alpha}{\|K\|+\alpha}, & \mu\ge\frac{1}{2}\\ \frac{\alpha^{2\mu}}{2(1-\mu)(1-2\mu)^{2\mu-1}}, & \mu<\frac{1}{2}\end{Bmatrix}, \quad C_\alpha = \frac{1}{\alpha}$$

$$C_\alpha\delta + \varphi_\mu(\alpha)\|w\| \to \min! \quad \Rightarrow \quad \alpha(\delta) = C\begin{cases}\sqrt{\delta}, & \mu\ge\frac{1}{2}\\ \delta^{1/(2\mu+1)}, & \mu<\frac{1}{2}\end{cases}$$

$$\Rightarrow \quad \|R_{\alpha(\delta)}y^\delta - K^+y\| \le C\begin{cases}\sqrt{\delta}, & \mu>\frac{1}{2}\\ \delta^{2\mu/(2\mu+1)}, & \mu\le\frac{1}{2}\end{cases}$$

3. Tikhonov: Homework

Remark 80 (Maximum convergence rate)**.** *The smaller* μ*, the worse the convergence rate. Independent of the size of* μ*, the convergence rate is always strictly smaller than* δ^1 *due to the ill-posedness.*

Definition 81 (Qualification)**.** *The* qualification *of a regularization method is the largest* $\theta = 2\mu_0$ *so that the source condition for* $\mu < \mu_0$ *yields a slower convergence rate.*

Example 82 (Qualification)**.** *TSVD:* ∞*; Lavrentiev:* 1*; Tikhonov:* 2

Remark 83 (Mozorow's discrepancy principle)**.** *The discrepancy between the correct data* y *and the result* $KR_\alpha y$ *of the forward problem is*

$$\|y - KR_\alpha y\|^2 = \sum_{n=1}^\infty (1-\sigma_n g_\alpha(\sigma_n))^2(y,v_n)^2 = \sum_{n=1}^\infty (\sigma_n^{2\mu+1} - \sigma_n^{2\mu+2}g_\alpha(\sigma_n))^2\left(\frac{(y,v_n)}{\sigma_n^{2\mu+1}}\right)^2 \le \varphi_{\mu+\frac{1}{2}}(\alpha)^2\|w\|^2.$$

For the above examples, $\varphi_{\mu+\frac{1}{2}}(\alpha(\delta)) = \text{const.}\delta$*, thus* $\|y - KR_\alpha y\| \le \text{const.}\delta$*. This motivates* Mozorow's discrepancy principle*: Pick* α *such that* $\|KR_\alpha y^\delta - y^\delta\| \sim \delta$*.*

10 Tikhonov regularization for nonlinear inverse problems

For a nonlinear operator $F : X \to Y$ one cannot define a SVD or an adjoint. However, the formulation of Tikhonov regularization as minimization problem can be transferred onto the nonlinear inverse problem $F(x) = y$.

Definition 84 (Least squares and minimum norm solution, Tikhonov regularization)**.** *Let* X, Y *be Banach spaces,* $F : X \to Y$*,* $x^* \in X$*.*

- $x \in X$ *is called a* least squares solution *of* $F(x) = y$ *if* $\|F(x) - y\| \leq \|F(z) - y\| \; \forall z \in X$.
- *A least squares solution* x *of* $F(x) = y$ *is called* x^*-minimum norm solution *if* $\|x - x^*\| \leq \|z - x^*\|$ *for all least squares solutions* z.
- *The associated* Tikhonov regularization operator *is* $R_\alpha : Y \ni y \mapsto \arg\min_{x \in X} J_\alpha^y(x) \subset X$ *for* $J_\alpha^y(x) = \|F(x) - y\|^2 + \alpha\|x - x^*\|^2$.

Remark 85 (Consequences of nonlinearity)**.**

- *Uniqueness of the* x^**-minimum norm solution or the Tikhonov regularization cannot be expected for nonlinear* F*. Also, there can be local and global minimizers – we will only consider global ones.*
- *For linear inverse problems we picked* $x^* = 0$*, for nonlinear ones* 0 *plays no distinguished role.*

We now work through the standard program for nonlinear regularized inverse problems:

1. existence of minimizers
2. stability of minimizers
3. convergence of the regularization

Theorem 86 (Existence)**.** *Let* X, Y *be reflexive Banach spaces (i. e.* $X^{**} = X$, $Y^{**} = Y$*) and* $F : X \to Y$ *continuous and weakly sequentially continuous (i. e.* $F(x_n) \rightharpoonup F(x)$ *for* $x_n \rightharpoonup x$*).*

a. J_α^y *has a minimizer.*

b. If $F(x) = y$ *has a solution* $x \in X$*, then it has an* x^**-minimum norm solution.*

Proof. (a): "direct method of the calculus of variation"

1. $J_\alpha^y(x^*) = \|F(x^*) - y\|^2 < \infty$ & $J_\alpha^y \geq 0$
2. consider a "minimizing sequence" x_n with $J_\alpha^y(x_n) \to \inf J_\alpha^y$ monotonically
3. $\alpha\|x_n - x^*\|^2 \leq J_\alpha^y(x_n) \leq J_\alpha^y(x_0) < \infty \xRightarrow{\text{Banach–Alaoglu}}$ there exists a convergent subsequence $x_n \rightharpoonup x$
4. Due to weak lower semi-continuity of the norm and $F(x_n) \rightharpoonup F(x)$,
   $J_\alpha^y(x) = \|F(x) - y\|^2 + \alpha\|x - x^*\|^2 \leq \liminf_{n\to\infty} J_\alpha^y(x_n)$

(b): analogous, just restrict J_α^y to the weakly closed set of solutions to $F(x) = y$ ☐

In the linear case, $\|R_\alpha y^\delta - R_\alpha y\| < C_\alpha \|x^\delta - y\|$, i. e. the regularized solution converges strongly and linearly in the measurement error. In the nonlinear case we only obtain weak convergence of subsequences:

Theorem 87 (Stability/continuity of regularization)**.** *In addition to the conditions for existence let* $y_n \to y$ *in* Y *and* $x_n \in \arg\min_x J_\alpha^{y_n}(x)$*. Then* x_n *has a weakly convergent subsequence, and every weak limit point minimizes* $J_\alpha^y(x)$*.*

Proof. 1. $\alpha\|x_n - x^*\|^2 \leq J_\alpha^{y_n}(x_n) \leq J_\alpha^{y_n}(x^*) = \|F(x^*) - y_n\|^2 < C < \infty$
$\xRightarrow{\text{Banach–Alaoglu}}$ x_n has weakly convergent subsequence $x_n \rightharpoonup x$

2. Due to weak lower semi-continuity of the norm and $F(x_n) \rightharpoonup F(x)$,
   $J_\alpha^y(x) \leq \liminf_{n\to\infty} J_\alpha^{y_n}(x_n) \leq \liminf_{n\to\infty} J_\alpha^{y_n}(z) \overset{y_n \to y \text{ strongly}}{=} J_\alpha^y(z)$ for all $z \in X$ ☐

So far we showed existence and (weak subsequence-) continuity of the Tikhonov regularization (uniqueness is impossible in general), i. e. as much well-posedness as possible. Now we consider convergence.

Theorem 88 (Convergence)**.** *In addition to the above let* $F(x) = y$ *have a solution* $x \in X$ *and let* $y^\delta \in Y$ *with* $\|y^\delta - y\| \leq \delta$ *as well as* $x_\alpha^\delta \in \arg\min J_\alpha^{y^\delta}$*. If* $\alpha \to 0$ *and* $\delta/\sqrt{\alpha} \to 0$ *as* $\delta \to 0$ *for* $\alpha = \alpha(\delta, y^\delta)$*, then* x_α^δ *has a weakly convergent subsequence, and every weak limit point is an* x^**-minimum norm solution.*

Remark 89 (Condition on regularization parameter). *We require $\delta/\sqrt{\alpha} \to 0$, which is exactly the same as for Tikhonov regularization in the linear case ($C_\alpha \delta \to 0$).*

Proof. 1. There exists an x^*-minimum norm solution $x^\dagger$.

2. x_α^δ has weakly convergent subsequence:

   $\alpha\|x_\alpha^\delta - x^*\|^2 \le J_\alpha^{y^\delta}(x_\alpha^\delta) \le J_\alpha^{y^\delta}(x^\dagger) = \|F(x^\dagger) - y^\delta\|^2 + \alpha\|x^\dagger - x^*\|^2 \overset{F(x^\dagger)=y}{\le} \delta^2 + \alpha\|x^\dagger - x^*\|^2$
   $\Rightarrow x_\alpha^\delta$ bounded $\Rightarrow \exists$ weakly convergent subsequence $x_{\alpha_n}^{\delta_n} \rightharpoonup x$

3. x is x^*-minimum norm solution:

   - $\|x - x^*\|^2 \le \liminf_{n\to\infty} \|x_{\alpha_n}^{\delta_n} - x^*\|^2 \overset{F(x^\dagger)=y}{\le} \liminf_{n\to\infty} \frac{\delta_n^2}{\alpha_n} + \|x^\dagger - x^*\|^2 = \|x^\dagger - x^*\|^2$
   - $\|F(x_{\alpha_n}^{\delta_n}) - y\| \le \delta_n + \|F(x_{\alpha_n}^{\delta_n}) - y^\delta\| \le \delta_n + \sqrt{J_{\alpha_n}^{y^{\delta_n}}(x_{\alpha_n}^{\delta_n})} \le \delta_n + \sqrt{J_{\alpha_n}^{y^{\delta_n}}(x^\dagger)}$
     $\le \delta_n + \sqrt{\delta_n^2 + \alpha_n\|x^\dagger - x^*\|^2} \xrightarrow[n\to\infty]{} 0 \Rightarrow \|F(x)-y\| = \liminf_{n\to\infty} \|F(x_{\alpha_n}^{\delta_n}) - y\| = 0$ ☐

Corollary 90 (Strong stability and convergence in Hilbert space). *If X is a Hilbert space, the "weak" may be replaced with "strong" in the previous result.*

Remark 91 (Role of the Hilbert space). *Strong convergence is indeed specific to Hilbert spaces and cannot be expected in general Banach spaces: In Hilbert spaces, $x_n \rightharpoonup x$ & $\|x_n\| \to \|x\|$ imply $x_n \to x$, and we will just copy the corresponding proof.*

Proof. $\|x_{\alpha_n}^{\delta_n} - x\|^2 = \underbrace{\|x_{\alpha_n}^{\delta_n} - x^*\|^2}_{\substack{\le \frac{\delta_n^2}{\alpha_n} + \|x^\dagger - x^*\|^2 = \frac{\delta_n^2}{\alpha_n} + \|x - x^*\|^2 \\ (x^\dagger \text{ \& } x \text{ are } x^*\text{-min. nrm. sols.})}} - 2\underbrace{(x_{\alpha_n}^{\delta_n} - x^*, x - x^*)}_{\to \|x - x^*\|^2} + \|x - x^*\|^2 \to 0$ ☐

Still to consider: Convergence rates under additional smoothness conditions. To this end we restrict to Hilbert spaces in order to get rates in the norm (otherwise we would have to metrize the weak topology – which we will do later for measures).

Definition 92 (Fréchet differentiability). *A map $F: X \to Y$ between Banach spaces is called* Fréchet differentiable *in $x \in X$ with* Fréchet derivative *$F'(x)$, if $F'(x): X \to Y$ is linear and continuous with $\frac{\|F(y)-F(x)-F'(x)(y-x)\|}{\|x-y\|} \to 0$ as $\|y - x\| \to 0$. F is called* Fréchet differentiable*, if it is everywhere* Fréchet differentiable*.*

Theorem 93 (Differentiability of Tikhonov energy). *Let X, Y be Hilbert spaces. If $F: X \to Y$ is Fréchet differentiable, then so is J_α^y with $(J_\alpha^y)'(x) = (2(F'(x))^H(F(x) - y) + 2\alpha(x - x^*), \cdot)$.*

Proof. Homework ☐

For convergence rates we require a source condition. We consider the source condition for $\mu = \frac{1}{2}$. $x^\dagger = (K^H K)^{\frac{1}{2}} w$ for some $w \in X$ is equivalent to $x^\dagger = K^H p$ with $p = \sum_{n=1}^\infty (w, u_n) v_n$. This can be interpreted as the existence of a Lagrange multiplicator for the minimum norm solution problem $\min_x \frac{1}{2}\|x\|^2$ such that $Kx = y$, since

Lagrangian $\quad L(x,p) = \frac{1}{2}\|x\|^2 - (Kx - y, p)$

opt. cond. $\quad 0 = \frac{\partial L}{\partial p} = y - Kx \quad \& \quad 0 = \frac{\partial L}{\partial x} = x - K^H p.$

Analogously one proceeds for a nonlinear operator:

Lagrangian $\quad L(x,p) = \frac{1}{2}\|x - x^*\|^2 - (F(x) - y, p)$

opt. cond. $\quad 0 = \frac{\partial L}{\partial p} = y - F(x) \quad \& \quad 0 = \frac{\partial L}{\partial x} = x - x^* - (F'(x))^H p.$

Definition 94 (Source condition)**.** *The* source condition *with μ for an x^*-minimum norm solution $x^\dagger$ of the inverse problem $F(x)=y$ reads*

$$x^\dagger - x^* = [F'(x^\dagger)^H F'(x^\dagger)]^\mu w \qquad \text{for a } w \in X.$$

Theorem 95 (Convergence rate under source condition)**.** *In addition to the above let $x^\dagger - x^* = F'(x^\dagger)^H p$ for some $p \in Y$ and let F' have Lipschitz constant L with $L\|p\| < 1$. If we choose $\alpha(\delta, y^\delta) \sim \delta$, then there exists $D > 0$ with $\|x_\alpha^\delta - x^\dagger\| \leq \text{const.}\sqrt{\delta}\ \forall \delta < D$.*

Remark 96 (Relation to linear setting)**.** *In the linear setting, $L = 0$; also the choice of α and the resulting error estimate are the same as in the linear setting for $\mu = \frac{1}{2}$.*

Proof.

- $\|F(x_\alpha^\delta) - y^\delta\|^2 + \alpha\|x_\alpha^\delta - x^*\|^2 \leq \|F(x^\dagger) - y^\delta\|^2 + \alpha\|x^\dagger - x^*\|^2 \leq \delta^2 + \alpha\|x^\dagger - x^*\|^2$
  $\Leftrightarrow \|F(x_\alpha^\delta) - y^\delta\|^2 + \alpha\|x_\alpha^\delta - x^\dagger\|^2 \leq \delta^2 + 2\alpha(x^\dagger - x^*, x^\dagger - x_\alpha^\delta) = \delta^2 + 2\alpha(p, F'(x^\dagger)(x^\dagger - x_\alpha^\delta))$
- set $f(t) = (p, F(x^\dagger + t(x_\alpha^\delta - x^\dagger)))$
  $\Rightarrow f'(t) = (p, F'(x^\dagger + t(x_\alpha^\delta - x^\dagger))(x_\alpha^\delta - x^\dagger))$ has Lipschitz constant $L\|p\|\|x_\alpha^\delta - x^\dagger\|^2$
  $\Rightarrow \|(p, F(x_\alpha^\delta) - F(x^\dagger) - F'(x^\dagger)(x_\alpha^\delta - x^\dagger))\| = \|f(1) - f(0) - f'(0)\| = \|\int_0^1 f'(t) - f'(0)\,\mathrm{d}t\| \leq \frac{L\|p\|}{2}\|x_\alpha^\delta - x^\dagger\|^2$
- $\|F(x_\alpha^\delta) - y^\delta\|^2 + \alpha\|x_\alpha^\delta - x^\dagger\|^2 \leq \delta^2 + 2\alpha(p, F(x^\dagger) - F(x_\alpha^\delta)) + \alpha L\|p\|\|x_\alpha^\delta - x^\dagger\|^2$
  $\leq \delta^2 + 2\alpha(p, y^\delta - F(x_\alpha^\delta)) + 2\alpha(p, y - y^\delta) + \alpha L\|p\|\|x_\alpha^\delta - x^\dagger\|^2$
  $\Leftrightarrow \frac{1}{\alpha}\|F(x_\alpha^\delta) - y^\delta + \alpha p\|^2 + (1 - L\|p\|)\|x_\alpha^\delta - x^\dagger\|^2 \leq \frac{\delta^2}{\alpha} + \alpha\|p\|^2 + 2\delta\|p\| = (\underbrace{\tfrac{\delta}{\sqrt{\alpha}} + \sqrt{\alpha}\|p\|}_{\hat{=} C_\alpha \delta + \varphi_{\frac{1}{2}}(\alpha)\|w\|})^2$
  $\xRightarrow[c_1\delta \leq \alpha \leq c_2\delta]{} (1 - L\|p\|)\|x_\alpha^\delta - x^\dagger\|^2 \leq \delta(\frac{1}{\sqrt{c_1}} + \sqrt{c_2}\|p\|)^2$
  $\Rightarrow \|x_\alpha^\delta - x^\dagger\|^2 \leq \delta(\frac{1}{\sqrt{c_1}} + \sqrt{c_2}\|p\|)^2/(1 - L\|p\|)$ □

11 Short introduction to convex analysis

One often tries to choose convex regularizations since these are easier to minimize and come with simple error estimates.

Definition 97 (Convex functional)**.** *A subset C of a Banach space X is called* convex *if*

$$\theta x + (1-\theta)y \in C \quad \forall x, y \in C, \theta \in (0,1).$$

A functional $f : X \to (-\infty, \infty]$ is called proper and convex *if $f \not\equiv \infty$ and*

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y) \quad \forall x, y \in X, \theta \in (0,1).$$

The domain *of f is*

$$\operatorname{dom} f = \{x \in X \mid f(x) < \infty\}.$$

Example 98 (Convex functions)**.**

- *linear functionals*
- *$x \mapsto (x, Ax)$ for coercive linear operator A*
- *indicator function $\iota_C(x) = \begin{cases} 0 & \text{if } x \in C \\ \infty & \text{else} \end{cases}$ of a convex set C*
- *norms*
- *compositions of convex functionals with linear operators*

Convergence rates for regularizations of inverse problems can typically be obtained in the so-called Bregman distance, which we introduce next.

Definition 99 (Subdifferential)**.** *The* subdifferential *of a convex functional* $f : X \to (-\infty, \infty]$ *in* $x \in X$ *is*

$$\partial f(x) = \{s \in X^* \mid f(y) \geq f(x) + \langle s, y - x\rangle\}, \qquad \text{(“}f\text{ lies above its linearization”)}$$

its elements are called subgradients*.*

Example 100 (Subdifferentials)**.**

- $\ell \in X^* \Rightarrow \partial\ell(x) = \{\ell\}$
- f *Fréchet-differentiable in* $x \Rightarrow \partial f(x) = \{f'(x)\}$
- $\|\cdot\|$ *Hilbert space norm* $\Rightarrow \partial\|\cdot\|(x) = \begin{cases} \{(\frac{x}{\|x\|}, \cdot)\} & \text{if } x \neq 0 \\ \{(y, \cdot) \mid \|y\| \leq 1\} & \text{else} \end{cases}$

Definition 101 (Bregman distance)**.** *Let* f *be a proper convex functional on a Banach space* X *and* $w \in \partial f(x)$*. The* Bregman distance *of* $y \in X$ *to* $x \in X$ *is*

$$D_w^f(y, x) = f(y) - f(x) - \langle w, y - x\rangle \geq 0$$

(and it does not satisfy the axioms of a metric).

The (Bregman) distance to a minimizer of a convex functional as well as reconstruction errors in linear inverse problems and their convex regularizations can be estimated via duality methods.

Definition 102 (Legendre–Fenchel transform)**.** *The* Legendre–Fenchel conjugate *of a convex functional* f *on a Banach space* X *is*

$$f^* : X^* \to (-\infty, \infty], \qquad f^*(y) = \sup_{x \in X} \langle y, x\rangle - f(x).$$

The (predual) Legendre–Fenchel conjugate *of a convex functional* f *on a dual space* X^* *is*

$${}^*f : X \to (-\infty, \infty], \qquad {}^*f(x) = \sup_{y \in X^*} \langle y, x\rangle - f(y).$$

Example 103 (Legendre–Fenchel conjugate)**.**

- $(\|\cdot\|_X)^* = \iota_{\{y \in X^* \mid \|y\|_{X^*} \leq 1\}}$
- $f(x) = \frac{1}{2}(x, Ax)$ *for coercive* $A \Rightarrow f^*(y) = \frac{1}{2}(y, A^{-1}y)$

Theorem 104 (Fenchel–Moreau theorem)**.**

- *The Legendre–Fenchel conjugate is convex and lower semi-continuous.*
- *The Legendre–Fenchel biconjugate* ${}^*[f^*]$ *is the convex lower semi-continuous envelope of* f *(i. e. the largest lower semi-continuous convex function below* f*).*

Theorem 105 (Fenchel inequality)**.** *Let* f *be proper convex,* $x \in X$*,* $y \in X^*$*.*

- $\langle y, x\rangle \leq f(x) + f^*(y)$
- *equality* $\Leftrightarrow y \in \partial f(x) \Leftrightarrow x \in \partial f^*(y)$

Every convex optimization problem, when written as a sum of two terms, has an associated convex *dual optimization problem.* The relation between primal and dual problem allows to estimate the above-mentioned errors.

Theorem 106 (Fenchel–Rockafellar)**.** *Let* X, Y *be Banach spaces,* $F : Y \to (-\infty, \infty]$*,* $G : X \to (-\infty, \infty]$ *proper and convex and* $A : X \to Y$ *bounded linear.*

1. The primal optimization problem

$$p^* = \inf_{x\in X} F(Ax) + G(x)$$

and its dual problem

$$d^* = \sup_{y\in Y^*} -F^*(y) - G^*(-A^*y^*)$$

satisfy weak duality*, i. e.* $d^* \leq p^*$*.*

2. Let $\operatorname{relint} S$ *denote the* relative interior *of a set* S *(the interior relative to* $x + \operatorname{span}\{S - x\}$*). If*

(a) $\operatorname{relint}\operatorname{dom} F \cap A \operatorname{relint}\operatorname{dom} G \neq \emptyset$ *or*

(b) $(-A^*)\operatorname{relint}\operatorname{dom} F^* \cap \operatorname{relint}\operatorname{dom} G^* \neq \emptyset$ *and* F, G *are lower semi-continuous,*

then strong duality $p^* = d^*$ *holds. Under* (2a) *the supremum, under* (2b) *the infimum is attained.*

Proof of (1).

$$\begin{aligned} p^* &= \inf_{x\in X} F(Ax) + G(x) \\ &\geq \inf_{x\in X} {}^*[F^*](Ax) + G(x) \\ &= \inf_{x\in X} \sup_{y\in Y^*} \langle y, Ax\rangle - F^*(y) + G(x) \\ &\geq \sup_{y\in Y^*} \inf_{x\in X} \langle y, Ax\rangle - F^*(y) + G(x) \\ &= \sup_{y\in Y^*} \inf_{x\in X} \langle A^*y, x\rangle - F^*(y) + G(x) \\ &= \sup_{y\in Y^*} -F^*(y) - G^*(-A^*y) = d^* \end{aligned}$$

□

If $x \in X$, then with the help of the dual problem one can estimate how well x minimizes $F(Ax) + G(x)$. Indeed, for any $y \in Y^*$ we have

$$[F(Ax) + G(x)] - p^* \leq [F(Ax) + G(x)] - [-F^*(y) - G^*(-A^*y)] =: \varepsilon.$$

Then y is called a *dual certificate* for $F(Ax) + G(x) - p^* \leq \varepsilon$.

Corollary 107 (Primal-dual optimality conditions)**.** *Let strong duality hold.*

$$x \in X \textit{ solves the primal and } y \in Y^* \textit{ the dual problem} \quad \Leftrightarrow \quad \left\{ \begin{array}{ll} Ax & \in \partial F^*(y) \\ -A^*y & \in \partial G(x) \end{array} \right\}$$

Proof.

- (x, y) primal-dual optimal $\Leftrightarrow$ all inequalities must be equalities, i. e.

$$F(Ax) + G(x) = \langle y, Ax\rangle - F^*(y) + G(x) = \langle A^*y, x\rangle - F^*(y) + G(x) = -F^*(y) - G^*(-A^*y)$$

- by Fenchel's inequality, first equality $\Leftrightarrow$ $Ax \in \partial F^*(y)$, last equality $\Leftrightarrow$ $-A^*y \in \partial G(x)$

□

12 Tikhonov regularization in Banach spaces

So far we considered inverse problems and their regularizations on Hilbert spaces or at least reflexive Banach spaces. However, non-reflexive Banach spaces also important, standard, and more natural in many modern inverse problems (exemplarily, in the next chapters we will analyse inverse problems on the space of Radon measures).

A typical generalization of Tikhonov regularization for an inverse problem $Kx = y$ with an operator $K : X \to Y$ between Banach spaces would be

$$\operatorname*{arg\,min}_{x\in X} \underbrace{\|Kx - y\|_Y}_{\substack{\text{data/fidelity term} \\ \text{(ensures consistency} \\ \text{with measurement } y)}} + \alpha \underbrace{\|x\|_X}_{\text{regularization term}} .$$

Also more general data and regularization terms are often more appropriate, e. g.

- $\|Kx-y\|_Y^2$ (typically if Y is a Hilbert space),
- Kullback–Leibler divergence $\int d_{\mathrm{KL}}(Kx(s),y(s))\,\mathrm{d}s$ with $d_{\mathrm{KL}}(a,b)=b\log\frac{b}{a}-b+a$,
- entropy $\int e(x(s))\,\mathrm{d}s$ with $e(x)=x(\log x-1)$.

We consider the general setting

$$\operatorname*{arg\,min}_{x\in X} J_\alpha^y(x), \qquad J_\alpha^y(x)=\tfrac{1}{\alpha}F_y(Kx)+G(x)$$

with F_y,G proper convex, $F_y(y)=0$, $F_y>0$ else. For $\alpha=0$ we interpret this as constraint $F_y(Kx)=0$. We let $y^\dagger$ be the noise-free measurement and $x^\dagger$ the correct solution to the inverse problem $Kx^\dagger=y^\dagger$.

Theorem 108 (Vanishing Bregman distance for noiseless reconstruction)**.** *Let $x^\dagger\in X$ satisfy the source condition $-K^*w^\dagger\in\partial G(x^\dagger)$ for some $w^\dagger\in Y^*$. Then any minimizer x of $J_0^{y^\dagger}$ satisfies*

$$D^G_{-K^*w^\dagger}(x,x^\dagger)=0.$$

Proof. $D^G_{-K^*w^\dagger}(x,x^\dagger)=G(x)-G(x^\dagger)-\langle -K^*w^\dagger,x-x^\dagger\rangle$
$=G(x)-G(x^\dagger)+\langle w^\dagger,Kx-Kx^\dagger\rangle=G(x)-G(x^\dagger)=J_0^{y^\dagger}(x)-J_0^{y^\dagger}(x^\dagger)\leq 0$ □

Remark 109 (Interpretation of source condition)**.**

1. *Same source condition as in the setting with Hilbert space & nonlinear operator.*
2. *Source condition $-K^*w^\dagger\in\partial G(x^\dagger)$ is one of the two necessary and sufficient primal-dual optimality conditions for minimizing $J_0^{y^\dagger}$.*
3. *The other one is $Kx^\dagger\in\partial(\frac{1}{0}F_{y^\dagger})^*(w^\dagger)=\partial\iota^*_{\{y^\dagger\}}(w^\dagger)=\{y^\dagger\}$, thus automatically satisfied.*
4. *Thus, if strong duality holds, source condition $\Rightarrow$ $x^\dagger$ minimizes $J_0^{y^\dagger}$ & $w^\dagger$ certifies this.*

Now let y^δ be a noisy measurement with $F_{y^\delta}(y^\dagger)\leq\delta$ (this is how we now quantify the noise strength).

- If x^δ is an approximation of $x^\dagger$ and $J_0^{y^\dagger}$ smooth, the reconstruction error $x^\delta-x^\dagger$ can be estimated from the difference $J_0^{y^\dagger}(x^\delta)-J_0^{y^\dagger}(x^\dagger)$ and lower bounds on the Hessian of $J_0^{y^\dagger}$.
- For nonsmooth convex $J_0^{y^\dagger}$ the Hessian-based estimates are replaced by Bregman distances for $J_0^{y^\dagger}$.
- Since $J_0^{y^\dagger}(x^\delta)-J_0^{y^\dagger}(x^\dagger)=\infty$ if $Kx^\delta\neq y^\dagger$, plain Bregman distance would be ∞.
- Thus, fidelity term first needs to be dualized: for some fixed $w^\dagger\in Y^*$, instead of $J_0^{y^\dagger}$ consider

  $$G(\cdot)+\langle K^*w^\dagger,\cdot\rangle-(\tfrac{1}{0}F_{y^\dagger})^*(w^\dagger)$$

  (which by weak duality is never larger than $J_0^{y^\dagger}$).

Theorem 110 (Bregman distance estimate for noisy reconstruction)**.** *Let $x^\dagger\in X$ satisfy the source condition $-K^*w^\dagger\in\partial G(x^\dagger)$ for some $w^\dagger\in Y^*$. Then a minimizer x_α^δ of $J_\alpha^{y^\delta}$ satisfies*

$$\begin{aligned}
D^G_{-K^*w^\dagger}(x_\alpha^\delta,x^\dagger)&\leq\Big(3\delta+F^*_{y^\delta}(2\alpha w^\dagger)+F^*_{y^\delta}(-2\alpha w^\dagger)\Big)/(2\alpha),\\
F_{y^\delta}(Kx_\alpha^\delta)&\leq\Big(3\delta+F^*_{y^\delta}(2\alpha w^\dagger)+F^*_{y^\delta}(-2\alpha w^\dagger)\Big),\\
\langle K^*w,x_\alpha^\delta-x^\dagger\rangle&\leq\Big(4\delta+F^*_{y^\delta}(2\alpha w^\dagger)+F^*_{y^\delta}(-2\alpha w^\dagger)+F^*_{y^\delta}(2\alpha w)+F^*_{y^\delta}(-2\alpha w)\Big)/(2\alpha)\quad\textit{for all } w\in Y^*.
\end{aligned}$$

Proof. 1. $\big[G(x_\alpha^\delta)+\langle K^*w^\dagger,x_\alpha^\delta\rangle-(\frac{1}{0}F_{y^\dagger})^*(w^\dagger)\big]-\big[G(x^\dagger)+\langle K^*w^\dagger,x^\dagger\rangle-(\frac{1}{0}F_{y^\dagger})^*(w^\dagger)\big]$
$=G(x_\alpha^\delta)-G(x^\dagger)-\langle -K^*w^\dagger,x_\alpha^\delta-x^\dagger\rangle=D^G_{-K^*w^\dagger}(x_\alpha^\delta,x^\dagger)$

2. optimality of x_α^δ:

$$G(x_\alpha^\delta)+\tfrac{1}{\alpha}F_{y^\delta}(Kx_\alpha^\delta)=J_\alpha^{y^\delta}(x_\alpha^\delta)\le J_\alpha^{y^\delta}(x^\dagger)=G(x^\dagger)+\tfrac{1}{\alpha}F_{y^\delta}(y^\dagger)\le G(x^\dagger)+\tfrac{\delta}{\alpha}.$$

3. Fenchel's inequality:

$$\langle K^*w,x_\alpha^\delta-x^\dagger\rangle=\frac{\langle 2\alpha w,Kx_\alpha^\delta\rangle+\langle -2\alpha w,Kx^\dagger\rangle}{2\alpha}\le\frac{F_{y^\delta}(Kx_\alpha^\delta)+F_{y^\delta}^*(2\alpha w)+F_{y^\delta}(Kx^\dagger)+F_{y^\delta}^*(-2\alpha w)}{2\alpha}$$

(already proves third statement in case second holds)

4. using both inequalities,

$$D^G_{-K^*w^\dagger}(x_\alpha^\delta,x^\dagger)=\left[G(x_\alpha^\delta)+\langle K^*w^\dagger,x_\alpha^\delta\rangle-(\tfrac{1}{0}F_{y^\dagger})^*(w^\dagger)\right]-\left[G(x^\dagger)+\langle K^*w^\dagger,x^\dagger\rangle-(\tfrac{1}{0}F_{y^\dagger})^*(w^\dagger)\right]$$
$$\le\frac{\delta}{\alpha}-\frac{1}{\alpha}F_{y^\delta}(Kx_\alpha^\delta)+\langle w^\dagger,K(x_\alpha^\delta-x^\dagger)\rangle\le\frac{1}{2\alpha}\Big(3\delta+F_{y^\delta}^*(2\alpha w^\dagger)+F_{y^\delta}^*(-2\alpha w^\dagger)\Big)-\frac{1}{2\alpha}F_{y^\delta}(Kx_\alpha^\delta).\quad\square$$

Remark 111 (Rates from estimates)**.** *$F_{y^\delta}(z)>0$ unless $z=y^\delta$*

$\Rightarrow$ *F_{y^δ} "strictly convex" in y^δ*

$\Rightarrow$ *$F_{y^\delta}^*(\pm 2\alpha w)$ is differentiable in $\alpha=0$ (homework)*

$\Rightarrow$ *$(F_{y^\delta}^*(2\alpha w)+F_{y^\delta}^*(-2\alpha w))/\alpha\to 0$ as $\alpha\to 0$ (symmetric finite difference)*

$\Rightarrow$ *choosing α as minimizer of $(\delta+F_{y^\delta}^*(2\alpha w^\dagger)+F_{y^\delta}^*(-2\alpha w^\dagger))/\alpha$ we get a convergence rate, e. g. $F_{y^\delta}(z)=\frac{1}{2}\|z-y^\delta\|^2$ in Hilbert space $\Rightarrow F_{y^\delta}^*(\pm 2\alpha w^\dagger)=2\alpha^2\|w^\dagger\|^2\pm\langle y^\delta,2\alpha w^\dagger\rangle$ $\Rightarrow(\delta+F_{y^\delta}^*(2\alpha w^\dagger)+F_{y^\delta}^*(-2\alpha w^\dagger))/\alpha=\frac{\delta}{\alpha}+4\|w^\dagger\|^2\alpha\Rightarrow\alpha=\frac{\sqrt{\delta}}{2\|w^\dagger\|}$, and rate $\sim\sqrt{\delta}$*

Remark 112 (Relation to Hilbert space setting)**.** *For X,Y Hilbert spaces and $G(x)=\|x\|^2$, $F_y(Kx)=\|Kx-y\|^2$, the above recovers the convergence rate of linear Tikhonov regularization with source condition for $\mu=\frac{1}{2}$ (and its proof reduces to the one we did for a nonlinear operator):*

- $D^G_{-K^Hw^\dagger}(x_\alpha^\delta,x^\dagger)=\|x_\alpha^\delta-x^\dagger\|^2$ (homework)
- $F_{y^\delta}^*(2\alpha w^\dagger)+F_{y^\delta}^*(-2\alpha w^\dagger)=2\|w^\dagger\|^2\alpha^2$ (homework)
- *$F_{y^\delta}(y^\dagger)=\|y^\delta-y^\dagger\|^2$, so δ here was called δ^2 before*

Remark 113 (Primal versus predual)**.** *Without any further changes one may also replace X^*,Y^* with predual spaces ${}^*X,{}^*Y$ (s. t. $({}^*Z)^*=Z$), Legendre–Fenchel conjugates with predual conjugates, and adjoints K^* with preadjoints ${}^*K$ (s. t. $({}^*K)^*=K$). This is the actual case of interest in the following.*

13 Short introduction to Radon measures

For superresolution microscopy or particle reconstruction applications, the natural space for inverse problems is the space of Radon measures. They also naturally occur (as derivatives) in inverse problems, whose reconstructions are piecewise constant/smooth, but we will only consider the former setting.

Definition 114 (Measure)**.**

1. *A set P of subsets of a set Ω is called* σ-algebra *if*
   (1) $\Omega\in P$ (2) $A\in P\Rightarrow\Omega\setminus A\in P$ (3) $A_i\in P\Rightarrow\bigcup_{i=1}^\infty A_i\in P$
   (due to $A\cap B=\Omega\setminus((\Omega\setminus A)\cup(\Omega\setminus B))$ it is also closed under countable intersections)
2. *The elements of P are called* measurable sets*, (Ω,P) is a* measurable space*.*
3. *The* Borel-algebra *of a topological space Ω is the smallest σ-algebra containing all open sets.*
4. *A* (positive/unsigned) measure *is a map $\mu:P\to[0,\infty]$ with*
   (1) $\mu(\emptyset)=0$ (2) $\mu(\bigcup_{n=1}^\infty A_i)=\sum_{i=1}^n\mu(A_i)$ for pairwise disjoint A_i ("countable additivity")

5. *A* signed measure *is a map* $\nu : P \to (-\infty, \infty]$ *with (1) and (2) absolutely convergent if finite.*

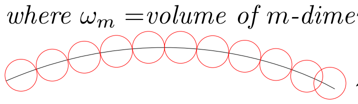
6. *The* support *of a measure* μ *on a Borel-measurable space is* $\operatorname{spt}\mu = \bigcap\{\overline{A} \in P \,|\, \mu(\Omega \setminus \overline{A}) = 0\}$.

Example 115 (Borel measures). *1.* Dirac measure $\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{else} \end{cases}$, $\operatorname{spt}\delta_x = \{x\}$

2. counting measure $\#(A) = \begin{cases} \text{number of elements of } A & \text{if } A \text{ finite,} \\ \infty & \text{else} \end{cases}$

3. Lebesgue measure $\mathcal{L}([a_1, b_1] \times \ldots \times [a_n, b_n]) = (b_1 - a_1) \cdots (b_n - a_n)$

4. Hausdorff measure $\mathcal{H}^m(A) = \lim_{\varepsilon \to 0} \inf \left\{ \sum_{i=1}^\infty \omega_m \left(\frac{\operatorname{diam} B_i}{2}\right)^m \,\middle|\, A \subset \bigcup_{i=1}^\infty B_i,\ \operatorname{diam} B_i < \varepsilon \right\}$, *where* ω_m *=volume of* m*-dimensional unit ball*



$\mathcal{H}^1(A)$ *=length of A*

5. Weighted measure $\nu = f\mu$ *for* μ *a measure,* f *measurable (see below);* $\nu(A) = \int_A f \,\mathrm{d}\mu$

Remark 116 (Point masses). *In some inverse problems one needs to reconstruct point sources, e.g. radioactive point sources in emission tomography, single fluorescent molecules in microscopy or iron particles in magnetic resonance tomography. A point source at position* $x \in R^n$ *with (radioactive/fluorescent/magnetic) intensity* $a > 0$ *can be described by* $a\delta_x$. *This motivates the use of a Banach space of Borel measures.*

Definition 117 (Transformations of measures). *Let* (Ω, P) *&* $(\tilde{\Omega}, \tilde{P})$ *be measurable spaces,* μ *a (signed) measure on* (Ω, P) *and* $B \in P$.

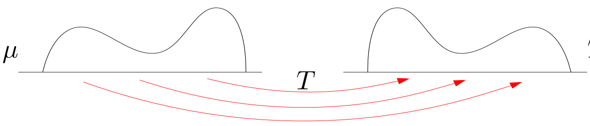
1. *The* restriction *of* μ *to* B *is* $\mu \llcorner B : P \to (-\infty, \infty]$, $\mu \llcorner B(A) = \mu(A \cap B)$.

2. $T : \Omega \to \tilde{\Omega}$ *is called* measurable *if* $T^{-1}(\tilde{A}) \in P \ \forall \tilde{A} \in \tilde{P}$.

3. *The* pushforward *of* μ *under* T *is* $T_\# \mu : \tilde{P} \to (-\infty, \infty]$, $T_\# \mu(\tilde{A}) = \mu(T^{-1}(\tilde{A}))$.

Remark 118 (Lebesgue integral). *For measurable functions* $f : \Omega \to \mathbb{R}$ *the Lebesgue integral* $\int_\Omega f \,\mathrm{d}\mu$ *can be defined.*

Example 119 (Pushforwards). *1.* μ



$T_\# \mu$

2. $\operatorname{proj}_i : \mathbb{R}^n \to \mathbb{R},\ x \mapsto x_i$; $\quad \operatorname{proj}_{i\#}\mu(A) = \mu(\mathbb{R}^{i-1} \times A \times \mathbb{R}^{n-i})$

   $\operatorname{proj}_i : \Omega_1 \times \ldots \times \Omega_n \to \Omega_i,\ (x_1, \ldots, x_n) \mapsto x_i$; $\quad \operatorname{proj}_{i\#}\mu(A) = \mu(\Omega_1 \times \ldots \times A \times \ldots \times \Omega_n)$

3. $\int_A f \circ T \,\mathrm{d}\mu = \int_{T(A)} f \,\mathrm{d}T_\#(\mu \llcorner A)$

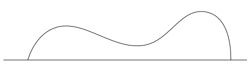
Definition 120 (Properties of measures). *1. A measure* μ *on* Ω *is called* σ-finite *if* $\Omega = \bigcup_{i=1}^\infty A_i$ *for a sequence* $A_i \subset \Omega$ *with* $|\mu(A_i)| < \infty$.

2. $\nu : P \to (-\infty, \infty]$ *is* absolutely continuous *wrt.* $\mu : P \to [0, \infty]$, $\nu \ll \mu$, *if* $\mu(A) = 0 \Rightarrow \nu(A) = 0$.

3. ν *&* μ *are called* singular, $\nu \perp \mu$, *if* $\exists A \in P : \mu(A) = 0, \nu(\Omega \setminus A) = 0$.

Example 121 (Properties of measures). • $\mathcal{L}$ *on* $\mathbb{R}^n$ *is* σ*-finite, but not finite.*

- $\delta_x \perp \mathcal{L}$

-  $\ll$ 

Theorem 122 (Hahn decomposition)**.** *For a signed measure $\mu : P \to (-\infty, \infty]$ on (Ω, P) there exists $N \in P$ such that $\begin{cases} \mu(A) \leq 0 & \text{if } A \subset N \\ \mu(A) \geq 0 & \text{if } A \subset \Omega \setminus N \end{cases}$ for all $A \in P$. We write $\mu^+ = \mu \llcorner (\Omega \setminus N)$, $\mu^- = \mu \llcorner N$ for the positive and negative part of μ.*

Theorem 123 (Radon–Nikodym)**.** *If μ is a σ-finite and ν a signed measure on (Ω, P) with $\nu \ll \mu$, then there exists a density function, called* Radon–Nikodym derivative*, i. e. a measurable $f : \Omega \to \mathbb{R}$ with $\nu(A) = \int_A f \, \mathrm{d}\mu \; \forall A \in P$. We write $f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$.*

Theorem 124 (Lebesgue decomposition)**.** *If μ is a σ-finite and ν a signed measure on (Ω, P), then there exists a unique decomposition $\nu = \tau + \pi$ with $\tau \ll \mu$, $\pi \perp \mu$.*

Certain measures form a Banach space; those are of particular interest to inverse problems.

Definition 125 (Variation and regularity)**.**

1. *Let μ be a signed measure on Ω with Hahn decomposition $\mu^{\pm}$. $|\mu| = \mu^+ - \mu^-$ is called* (total) variation (measure) *of μ.*
2. *$|\mu|(\Omega) = \sup\{\sum_{i=1}^{\infty} |\mu(A_i)| \,|\, A_i \subset \Omega$ measurable & pairwise disjoint$\}$ is called* total variation *of μ.*
3. *A measure μ on a topological space Ω is called* regular *if for all measurable $A \subset \Omega$ we have*

$$\mu(A) = \sup\{\mu(K) \,|\, K \subset A \text{ measurable \& compact}\} = \inf\{\mu(U) \,|\, U \supset A \text{ measurable \& open}\}.$$

A signed measure is regular if its variation measure is.

Theorem 126 (Regularity of Borel measures)**.** *A finite Borel measure on a compact metric space is regular.*

Theorem 127 (Riesz representation theorem)**.** *Let Ω be a compact metric space (e. g. $[0, 1]^n$).*

- *The space of* Radon measures *$\mathcal{M}(\Omega) = \{\mu$ regular signed Borel measure on $\Omega \,|\, |\mu|(\Omega) < \infty\}$ forms a Banach space with the norm $\|\mu\|_{\mathcal{M}} = \mathrm{TV}(\mu) = |\mu|(\Omega)$.*
- *$\mathcal{M}(\Omega) = (C(\Omega))^*$, $\langle f, \mu \rangle = \int_\Omega f \, \mathrm{d}\mu$*

Example 128 (Radon measures as dual objects)**.** *Homework:*

- $\mu = a\delta_x \Rightarrow \langle f, \mu \rangle = af(x)$
- $\mu = g\mathcal{L} \Rightarrow \langle f, \mu \rangle = \int_\Omega fg \, \mathrm{d}\mathcal{L}$
- $x_n \to x \in \Omega$, $a_n \to a \in \mathbb{R} \Rightarrow a_n\delta_{x_n} \overset{*}{\rightharpoonup} a\delta_x$
- $f_n \rightharpoonup f$ *in* $L^1(\Omega) \Rightarrow f_n\mathcal{L} \overset{*}{\rightharpoonup} f\mathcal{L}$
- $\| \sum_i a_i\delta_{x_i} \|_{\mathcal{M}} = \sum_i |a_i|$ *(x_i pairwise different)*
- $\|g\mathcal{L}\|_{\mathcal{M}} = \|g\|_{L^1}$

Regularization using the total variation typically leads to *sparse* results, i. e. measures that are zero almost everywhere, which fits to particle reconstruction.
As we have discussed for the Tikhonov regularization, in non-Hilbert spaces one cannot expect convergence rates for the norm of the reconstruction error in regularized inverse problems, but only gets weak(-*) convergence of the error to zero. Thus, if we want rates, we need to metrize weak-* convergence. The classical way to do so for measures is via optimal transport: One interprets two probability measures μ_1, μ_2 as initial and desired distribution of a material and calculates the cost for transporting the material from μ_1 to μ_2.

Definition 129 (Nonnegative measures)**.** *Let $\Omega \subset \mathbb{R}^n$ compact.*

- Nonnegative measures $\mathcal{M}_+(\Omega) = \{\mu \in \mathcal{M}(\Omega) \,|\, \mu = |\mu|\}$
- Probability measures $\mathcal{P}(\Omega) = \{\mu \in \mathcal{M}_+(\Omega) \,|\, \mu(\Omega) = 1\}$

Definition 130 (Monge formulation of optimal transport, Monge 1783)**.** *Let $\Omega \subset \mathbb{R}^n$ compact, $\mu_1, \mu_2 \in \mathcal{P}(\Omega)$ (representing the initial and final mass distribution), $c : \Omega \times \Omega \to \mathbb{R}$ ($c(x, y)$ is the cost for transporting unit mass from x to y). The* Monge formulation *of optimal transport reads*

$$\textit{find the } \text{transport map } T : \Omega \to \Omega \textit{ minimizing } C(T) = \int_\Omega c(x, T(x)) \,\mathrm{d}\mu_1 \textit{ among those with } T_\# \mu_1 = \mu_2.$$

Monge's formulation is not well-posed, e. g. for $\mu_1 = \delta_0$ and $\mu_2 = \frac{1}{2}\delta_x + \frac{1}{2}\delta_y$ half of the mass from 0 needs to be transported to x and half to y. Also, the minimization problem is very nonlinear and difficult to solve. In the 1950s Kantorovich found an alternative well-posed formulation as *convex* minimization problem, which won him the economics Nobel prize.

Definition 131 (Optimal transport cost, Kantorovich formulation)**.** *Let $\Omega \subset \mathbb{R}^n$ compact, $\mu_1, \mu_2 \in \mathcal{P}(\Omega)$, $c : \Omega \times \Omega \to \mathbb{R}$ continuous.*

- *A* transport plan *between μ_1, μ_2 is a $\pi \in \mathcal{P}(\Omega \times \Omega)$ with $\mathrm{proj}_{1\#}\pi = \mu_1$, $\mathrm{proj}_{2\#}\pi = \mu_2$. The set of transport plans is $\Pi(\mu_1, \mu_2)$. ($\pi(A \times B)$ is how much mass is transported from A to B.)*
- *The* tranport cost *of a transport plan π is $C(\pi) = \int_{\Omega\times\Omega} c(x, y) \,\mathrm{d}\pi(x, y)$.*
- *The* Kantorovich formulation *of optimal transport reads*

$$\textit{find the transport plan } \pi \in \Pi(\mu_1, \mu_2) \textit{ minimizing } C(\pi).$$

Theorem 132 (Existence of optimal transport plan)**.** *There exists an optimal transport plan.*

Proof. Homework (direct method of calculus of variations) □

Definition 133 (Wasserstein distance)**.** *Let $\Omega \subset \mathbb{R}^n$ compact, $\mu_1, \mu_2 \in \mathcal{P}(\Omega)$, $c(x, y) = |y - x|^p$, $p \geq 1$. The* Wasserstein-p distance *between μ_1 and μ_2 is*

$$W_p(\mu_1, \mu_2) = \left(\inf \left\{ \int_{\Omega\times\Omega} |x - y|^p \,\mathrm{d}\pi(x, y) \,\middle|\, \pi \in \Pi(\mu_1, \mu_2) \right\} \right)^{\frac{1}{p}}.$$

Theorem 134 (Wasserstein distance)**.** *The Wasserstein distance is a metric on $\mathcal{P}(\Omega)$.*

Proof. Homework □

Remark 135 (Wasserstein distance for non-probability measures)**.** *By rescaling μ_1, μ_2, π with the same factor, the Wasserstein distance is also defined for measures with non-unit mass.*

Example 136 (Wasserstein distances)**.**

- *$W_p(\delta_x, \delta_y) = |x - y|$ (homework)*
- *$W_p(\delta_x, \mu) = \left(\int_\Omega |x - y|^p \,\mathrm{d}\mu(y)\right)^{\frac{1}{p}}$ (homework)*
- *Let $f, g : [0, 1] \to [0, \infty)$ with $\int_0^1 f \,\mathrm{d}\mathcal{L} = \int_0^1 g \,\mathrm{d}\mathcal{L} = 1$, let F, G be antiderivatives of f, g with $F(0) = G(0) = 0$, set $T = G^{-1} \circ F$. Then $W_p(f\mathcal{L}, g\mathcal{L}) = \left(\int_0^1 |x - T(x)|^p f(x) \,\mathrm{d}\mathcal{L}(x)\right)^{\frac{1}{p}}$ (homework).*
- *Let A, B, C, D denote the corners of the unit square, $\mu_1 = \frac{1}{2}(\delta_A + \delta_B)$, $\mu_2 = \frac{1}{2}(\delta_C + \delta_D)$. $W_p(\mu_1, \mu_2) = 1$, $W_p(\delta_A, \mu_2) = (\frac{1}{2} + 2^{\frac{p}{2}-1})^{\frac{1}{p}}$ (homework).*

Theorem 137 (Metrization of weak-* convergence)**.** *The Wasserstein-p distance metrizes weak-* convergence on $\mathcal{P}(\Omega)$.*

Example 138 (Metrization of weak-* convergence)**.**

- *$\sum_{i=1}^N a_i^n \delta_{x_i^n} \overset{*}{\rightharpoonup} \sum_{i=1}^N a_i \delta_{x_i} \Leftrightarrow W_p(\sum_{i=1}^N a_i^n \delta_{x_i^n}, \sum_{i=1}^N a_i \delta_{x_i}) \to 0$ (homework)*
- *$\mu_n \overset{*}{\rightharpoonup} \delta_x \Leftrightarrow W_p(\mu_n, \delta_x) \to 0$ (homework)*

In some inverse problems one has to reconstruct a nonnegative Radon measure μ. The total variation or mass $\mu(\Omega)$ is usually not known beforehand. Thus it is not sufficient to restrict to probability measures. To define distances between measures of different mass one uses so-called *unbalanced optimal transport.* There are many variants, we just introduce the one most natural in our later setting.

Definition 139 (Unbalanced Wasserstein divergence)**.** *Let $\mu_1, \mu_2 \in \mathcal{M}_+(\Omega)$. For fixed $R > 0$ the* unbalanced Wasserstein-p divergence *between μ_1 and μ_2 is*

$$W_{p,R}^p(\mu_1, \mu_2) = \inf \left\{ W_p^p(\mu, \mu_2) + \tfrac{1}{2} R^p \|\mu_1 - \mu\|_{\mathcal{M}} \,\middle|\, \mu \in \mathcal{M}_+(\Omega) \right\}.$$

Remark 140 (Unbalanced Wasserstein divergence)**.** *It measures the cost for first changing the mass of μ_1 to some intermediate measure μ and then transporting that new mass distribution to μ_2. Up to distance R a mass transport is less costly than removing the mass in the initial position and regrowing it in the destination.*

Remark 141 (Metric properties)**.**

- $W_{p,R}^p(\mu_1, \mu_2) \geq 0$ *with equality iff* $\mu_1 = \mu_2$.
- $W_{p,R}^p(\mu_1, \mu_2) = W_{p,R}^p(\mu_2, \mu_1)$*:*
  - *Have* $\mu \ll \mu_1 + \mu_2$ *and* $\mu \leq \mu_1 + \mu_2$.
  - *Set* $\tilde{\mu} = \mu_2 + \mu_1 - \mu \in \mathcal{M}_+(\Omega)$.
  - $W_p^p(\mu, \mu_2) = W_p^p(\mu_1, \tilde{\mu})$ *and* $\|\mu_1 - \mu\|_{\mathcal{M}} = \|\mu_2 - \tilde{\mu}\|_{\mathcal{M}}$.
- *Triangle inequality is violated unless* $p = 1$.

Remark 142 (Unbalanced Wasserstein divergence for signed measures)**.** *We can extend the unbalanced Wasserstein-p divergence to signed measures $\mu_1, \mu_2 \in \mathcal{M}(\Omega)$ via*

$$W_{p,R}^p(\mu_1, \mu_2) = \inf \left\{ W_p^p(\mu, \mu_2^+) + W_p^p(\nu, |\mu_2^-|) + \tfrac{1}{2} R^p \|\mu_1 - \mu + \nu\|_{\mathcal{M}} \,\middle|\, \mu, \nu \in \mathcal{M}_+(\Omega), \mu, \nu \ll |\mu_1| + |\mu_2| \right\}.$$

14 Superresolution: Exact recovery

Superresolution = image reconstruction at a spatial resolution higher than the one of the measurement!
2014 chemistry Nobel prize: Betzig, Moerner, Hell for super-resolved fluorescence microscopy
$\Rightarrow$ resolution better than diffraction limit of light!
Examples: In *PALM* or *STORM*, most fluorescent molecules are switched off (e. g. by reversible photobleaching) so that at each time point only a few molecules emit light. In the camera, each molecule then appears as a diffuse blob, whose centre can be taken as the exact molecule position. Thus we try to reconstruct a linear combination of Dirac measures!
First rigorous analysis by de Castro & Gamboa 2012 and by Candès & Fernandez-Granda in 2013 & 2014; we do slightly different version.
Underlying theme: Infinite precision despite finite (!) measurements.

Remark 143 (Forward operator in superresolution)**.**

- *As forward operator Candès & Fernandez-Granda used a* truncated Fourier series

  $$K : \mathcal{M}(\Omega) \to \mathbb{C}^{(2k+1)^n}, \qquad K\mu = \left(\int_\Omega e^{-2\pi i \xi \cdot x} \, \mathrm{d}\mu(x) \right)_{\xi \in \{-k, -k+1, \ldots, k\}^n}$$

  on the domain $\Omega = [0,1]^n$ with periodic boundary (k = maximum frequency).
- *Other (finite-dimensional) forward operators are possible in principle, though they might potentially only allow slightly weaker results (e. g. because the source conditions have not as nice properties).*
- *We develop the theory independent of K and will a posteriori validate that it can be directly applied for K the truncated Fourier series.*

From now on we assume Ω to be a compact domain, $K : \mathcal{M}(\Omega) \to Y$ a linear forward operator with finite-dimensional Euclidean codomain Y, and the ground truth configuration to be

$$\mu^\dagger = \sum_{i=1}^{N} a_i \delta_{x_i} \in \mathcal{M}(\Omega).$$

Further, $y^\dagger = K\mu^\dagger$. We aim to reconstruct $\mu^\dagger$ from a (noisy or noiseless) measurement y by minimizing the Tikhonov functional

$$J_\alpha^y(\mu) = \frac{1}{\alpha}\|K\mu - y\|_Y^2 + \|\mu\|_{\mathcal{M}}.$$

Theorem 144 (Source condition). *The source condition for $\mu^\dagger$ in this setting reads*

$$\exists w^\dagger \in {}^*Y = Y \quad s.t. \qquad \|{}^*Kw^\dagger\|_{C^0} \leq 1, \qquad -{}^*Kw^\dagger(x_i) = \operatorname{sgn}(a_i), \quad i = 1, \ldots, N.$$

Proof. Homework. □

Theorem 145 (Support identification). *Let $\mu^\dagger$ satisfy a source condition with dual variable $w^\dagger$ and let $|{}^*Kw^\dagger| < 1$ on $\Omega \setminus \{x_1, \ldots, x_N\}$. Then any minimizer μ of $J_0^{y^\dagger}$ satisfies*

$$\operatorname{spt} \mu \subset \operatorname{spt} \mu^\dagger = \{x_1, \ldots, x_N\}.$$

Proof. Homework (use Bregman distance estimate). □

Theorem 146 (Exact recovery). *Assume in addition that for any $\nu \in \mathcal{M}(\Omega)$ with $\operatorname{spt} \nu \subset \operatorname{spt} \mu^\dagger$ a source condition holds. Then $\mu^\dagger$ is the unique minimizer of $J_0^{y^\dagger}$.*

Proof. Let μ be another minimizer, then $\operatorname{spt} \mu = \operatorname{spt} \mu^\dagger$.
Let $\mu - \mu^\dagger = \sum_{i=1}^N b_i \delta_{x_i}$ and w be the dual variable of the associated source condition, then

$$0 = -(w, K(\mu - \mu^\dagger)) = \langle -{}^*Kw, \mu - \mu^\dagger \rangle = \sum_{i=1}^{N} \operatorname{sgn}(b_i) b_i > 0$$

unless $b_1 = \ldots = b_N = 0$. □

Remark 147 (Relaxation of conditions). *Actually, asking a source condition with dual variable w_ν to hold for every ν is more than required. In fact one solely needs that $-{}^*Kw_\nu$ has the same sign as ν at every x_i.*

Remark 148 (Existence of dual variables). *Often, the conditions (of source conditions holding for any measure ν with $\operatorname{spt} \nu = \{x_1, \ldots, x_N\}$) are satisfied as long as the x_i have a minimum distance from each other (e. g. distance $\geq$ const./k if K is the truncated Fourier transform).*

15 Superresolution: Reconstruction from noisy data

Now let $y^\delta \in Y$ be a noisy measurement with $\|y^\delta - y^\dagger\|_Y^2 \leq \delta$. We will derive unbalanced Wasserstein divergence estimates for the reconstruction error.

- Let $\Delta > 0$ be the minimum distance between the x_i.
- For $R > 0$ let $B_R(S)$ denote the open R-neighbourhood of the set S.
- Abbreviate $B_i = B_R(\{x_i\})$, $B = \bigcup_{i=1}^N B_i$, $B^c = \Omega \setminus B$.

Theorem 149 (Unbalanced Wasserstein divergence from mass estimates)**.** *Let* $R \in (0, \Delta/2)$*. If* $\mu \in \mathcal{M}(\Omega)$ *satisfies*

$$|\mu|(B^c) \leq \alpha,$$

$$\sum_{i=1}^N |(\mu^\dagger - \mu)(B_i)| \leq \beta,$$

$$\sum_{i=1}^N \int_{B_i} \mathrm{dist}(x, x_i)^2 \,\mathrm{d}|\mu|(x) \leq \gamma,$$

then

$$W_{2,R}^2(\mu^\dagger, \mu) \leq \tfrac{1}{2} R^2(\alpha + \beta) + \gamma.$$

Proof. Homework. □

Theorem 150 (Mass distribution from Bregman distance)**.** *Let* $v^\dagger \in \partial \mathrm{TV}(\mu^\dagger) \cap C^0(\Omega)$*, i. e.* $|v^\dagger| \leq 1$ *and* $v^\dagger(x_i) = \mathrm{sgn}(a_i)$*. If* $v^\dagger$ *satisfies*

$$|v^\dagger| < 1 - \kappa R^2 \text{ on } B^c, \qquad \mathrm{sgn}(a_i) v^\dagger(x) \leq 1 - \kappa \mathrm{dist}(x, x_i)^2 \text{ on } B_i$$

for some $\kappa > 0, R \in (0, \Delta/2)$*, then for any* $\mu \in \mathcal{M}(\Omega)$ *we have*

$$|\mu|(B^c) \leq \frac{1}{\kappa R^2} D_{v^\dagger}^{\mathrm{TV}}(\mu, \mu^\dagger),$$

$$\sum_{i=1}^N \int_{B_i} \mathrm{dist}(x, x_i)^2 \,\mathrm{d}|\mu|(x) \leq \frac{1}{\kappa} D_{v^\dagger}^{\mathrm{TV}}(\mu, \mu^\dagger).$$

Furthermore, let $v \in \partial \mathrm{TV}(\nu) \cap C^0(\Omega)$ *for* $\nu = \sum_{i=1}^N (\mu - \mu^\dagger)(B_i)\delta_{x_i}$*, thus* $v(x_i) = \mathrm{sgn}(\nu(\{x_i\}))$ *for* $i = 1, \ldots, N$*. If for some* $\eta > 0$ *the function* v *satisfies*

$$v(x_i)v(x) \geq 1 - \eta \mathrm{dist}(x, x_i)^2 \quad \text{for all } x \in B_i, i = 1, \ldots, N,$$

then additionally

$$\sum_{i=1}^N |(\mu - \mu^\dagger)(B_i)| \leq \frac{1 + \eta R^2}{\kappa R^2} D_{v^\dagger}^{\mathrm{TV}}(\mu, \mu^\dagger) + \langle v, \mu - \mu^\dagger \rangle.$$

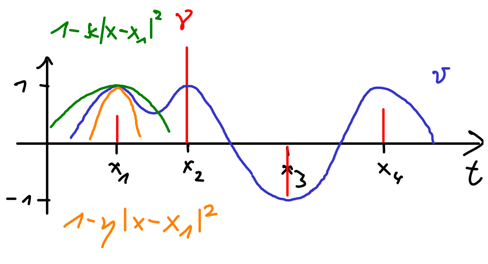
Proof.

- first two inequalities:

$$\kappa R^2 |\mu|(B^c) + \kappa \sum_{i=1}^N \int_{B_i} \mathrm{dist}(x, x_i)^2 \,\mathrm{d}|\mu|(x) \leq \int_{B^c} \frac{\mathrm{d}\mu}{\mathrm{d}|\mu|} - v^\dagger \,\mathrm{d}\mu(x) + \sum_{i=1}^N \int_{B_i} \frac{\mathrm{d}\mu}{\mathrm{d}|\mu|} - v^\dagger \,\mathrm{d}\mu(x)$$

$$= \int_\Omega \frac{\mathrm{d}\mu}{\mathrm{d}|\mu|} - v^\dagger \,\mathrm{d}\mu(x) = \|\mu\|_{\mathcal{M}} - \langle v^\dagger, \mu \rangle = \|\mu\|_{\mathcal{M}} - \|\mu^\dagger\|_{\mathcal{M}} - \langle v^\dagger, \mu - \mu^\dagger \rangle = D_{v^\dagger}^{\mathrm{TV}}(\mu, \mu^\dagger)$$

- third inequality:
  - ◦ $\sum_{i=1}^N |(\mu - \mu^\dagger)(B_i)| = \int_B v \,\mathrm{d}(\mu - \mu^\dagger) + \sum_{i=1}^N \int_{B_i} v(x_i) - v \,\mathrm{d}(\mu - \mu^\dagger)$
  - ◦ $\int_B v \,\mathrm{d}(\mu - \mu^\dagger) = \langle v, \mu - \mu^\dagger \rangle - \int_{B^c} v \,\mathrm{d}(\mu - \mu^\dagger)$
    $\leq \langle v, \mu - \mu^\dagger \rangle + |\mu - \mu^\dagger|(B^c) \leq \langle v, \mu - \mu^\dagger \rangle + \frac{1}{\kappa R^2} D_{v^\dagger}^{\mathrm{TV}}(\mu, \mu^\dagger)$
  - ◦ $\sum_{i=1}^N \int_{B_i} v(x_i) - v \,\mathrm{d}(\mu - \mu^\dagger) \leq \sum_{i=1}^N \int_{B_i} \eta \mathrm{dist}(x, x_i)^2 \,\mathrm{d}|\mu|(x) \leq \frac{\eta}{\kappa} D_{v^\dagger}^{\mathrm{TV}}(\mu, \mu^\dagger)$ □



Theorem 151 (Convergence rate of reconstruction)**.** *Assume that for any $\nu \in \mathcal{M}(\Omega)$ with $\operatorname{spt}\nu \subset \operatorname{spt}\mu^\dagger$ a source condition holds with dual variable w_ν and*

$$|{}^*Kw_\nu| < 1-\kappa R^2 \text{ on } B^c, \qquad 1-\eta\mathrm{dist}(x,x_i)^2 \leq -\mathrm{sgn}(\nu(\{x_i\}))\,{}^*Kw_\nu(x) \leq 1-\kappa\mathrm{dist}(x,x_i)^2 \text{ on } B_i$$

for some $\kappa,\eta>0$, $R\in(0,\Delta/2)$. Then any minimizer μ_α^δ of $J_\alpha^{y^\delta}$ satisfies

$$W_{2,R}^2(\mu^\dagger,\mu_\alpha^\delta) \leq C\tfrac{1+(\kappa+\eta)R^2}{\kappa}\left(\tfrac{\delta}{\alpha}+\alpha\right)$$

for some constant $C>0$ depending on $\mu^\dagger$. The choice $\alpha=\sqrt{\delta}$ thus yields $W_{2,R}^2(\mu^\dagger,\mu_\alpha^\delta)\leq \mathrm{const.}\sqrt{\delta}$.

Proof. Homework (combine theorems 110, 149 and 150). □

16 Source conditions: Trigonometric polynomials

For the previous two sections we need the existence of dual variables w_ν such that $-{}^*Kw_\nu$ satisfies certain growth conditions. As an exemplary case we show when this is possible for K the truncated Fourier series on $[0,1]$. Then $g\equiv -{}^*Kw_\nu$ is of the form

$$g(x)=\sum_{j=-k}^{k} c_j e^{2\pi ijx},$$

i. e. a trigonometric polynomial with maximum frequency k and coefficients $c_j\in\mathbb{C}$ (homework). Hence, we need to show that for any $\nu=\sum_{i=1}^N b_i\delta_{x_i}$ with minimum distance Δ between the x_i there exists a trigonometric polynomial g with maximum frequency k such that

$$|g|\leq 1-\kappa R^2 \text{ on } B^c, \qquad 1-\eta\mathrm{dist}(x,x_i)^2 \leq \mathrm{sgn}(b_i)g(x) \leq 1-\kappa\mathrm{dist}(x,x_i)^2 \text{ on } B_i.$$

To this end one uses a special basis of trigonometric polynomials.

Definition 152 (Dirichlet kernel, Fejér kernel)**.**

1. *The* Dirichlet kernel *of frequency k is $D_k(x)=\sum_{j=-k}^k e^{2\pi ijx}$.*
2. *The* Fejér kernel *of frequency k is $F_k(x)=\frac{1}{k+1}\sum_{j=0}^k D_j(x)=\sum_{j=-k}^k(1-\frac{|j|}{k+1})e^{2\pi ijx}$.*

Theorem 153 (Dirichlet kernel, Fejér kernel)**.**

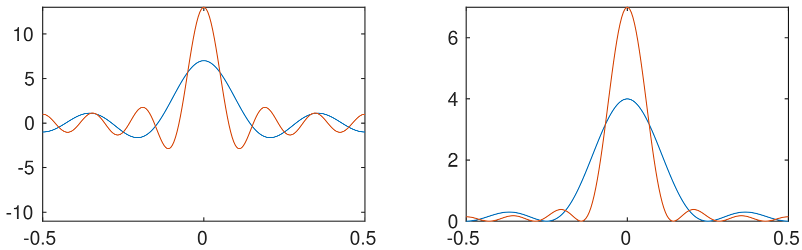
1. $D_k(x)=\frac{\sin((2k+1)\pi x)}{\sin(\pi x)}$
2. $F_k(x)=\frac{1}{k+1}\left(\frac{\sin((k+1)\pi x)}{\sin(\pi x)}\right)^2$

Proof. 1. geometric series: $\sum_{j=-k}^k s^j = s^{-k}\sum_{j=0}^{2k}s^j = s^{-k}\frac{1-s^{2k+1}}{1-s} = \frac{s^{-k-1/2}-s^{k+1/2}}{s^{-1/2}-s^{1/2}}$
$\Rightarrow \sum_{j=-k}^k e^{2\pi ixj} = \frac{e^{-(2k+1)\pi ix}-e^{(2k+1)\pi ix}}{e^{-\pi ix}-e^{\pi ix}} = \frac{-2i\sin((2k+1)\pi x)}{-2i\sin(\pi x)}$

2. $\sum_{j=0}^k \sin((2j+1)\pi x) = \frac{\sin^2((k+1)\pi x)}{\sin(\pi x)}$
   - induction basis: $k=0$
   - induction step: $\sum_{j=0}^{k+1}\sin((2j+1)\pi x) = \frac{\sin^2((k+1)\pi x)}{\sin(\pi x)}+\sin((2k+3)\pi x) = \frac{\sin^2((k+1)\pi x)+\sin(\pi x)\sin((2k+3)\pi x)}{\sin(\pi x)}$ by addition theorems, $\sin^2(t)=(1-\cos(2t))/2$, thus

$$\sin^2((k+1)\pi x)-\sin^2((k+2)\pi x) = \tfrac{\cos(2(k+2)\pi x)-\cos(2(k+1)\pi x)}{2}$$
$$= \tfrac{\cos((2k+3)\pi x)\cos(\pi x)-\sin((2k+3)\pi x)\sin(\pi x)-[\cos((2k+3)\pi x)\cos(-\pi x)-\sin((2k+3)\pi x)\sin(-\pi x)]}{2}$$
$$= -\sin((2k+3)\pi x)\sin(\pi x)$$

$F_k(x) = \frac{1}{k+1}\sum_{j=0}^{k} D_j(x) = \frac{1}{k+1}\sum_{j=0}^{k} \frac{\sin((2j+1)\pi x)}{\sin(\pi x)} = \frac{1}{k+1}\left(\frac{\sin((k+1)\pi x)}{\sin(\pi x)}\right)^2$ □



D_3 D_6 $\qquad$ F_3 F_6

Theorem 154 (Estimates of Fejér kernel)**.** *There exists $C > 0$ such that on $[-\frac{1}{2}, \frac{1}{2}]$ the Fejér kernel satisfies*

1. $0 \leq F_k(x) \leq \frac{1}{(k+1)\sin^2(\pi x)}$,
2. $|F_k'(x)| \leq \frac{3\pi}{\sin^2(\pi x)}$,
3. $|F_k^{(j)}(x)| \leq \frac{C(k+1)^{j-1}}{\sin^2(\pi x)}$, $j = 2, 3, 4$.

Proof. 1. obvious

2.
$$\begin{aligned}|F_k'(x)| &= \frac{1}{k+1}\left|\frac{2(k+1)\pi\sin((k+1)\pi x)\cos((k+1)\pi x)}{\sin^2(\pi x)} - 2\frac{\pi\cos(\pi x)\sin^2((k+1)\pi x)}{\sin^3(\pi x)}\right| \\ &= \frac{\pi}{(k+1)\sin^2(\pi x)}\left|k\sin(2\pi(k+1)x) - 2\tfrac{\cos(\pi x)\sin^2((k+1)\pi x) - \sin(\pi x)\sin((k+1)\pi x)\cos((k+1)\pi x)}{\sin(\pi x)}\right| \\ &= \frac{k\pi}{(k+1)\sin^2(\pi x)}\left|\sin(2\pi(k+1)x) - 2\tfrac{\sin(k\pi x)}{k\sin(\pi x)}\sin((k+1)\pi x)\right|\end{aligned}$$

and $\left|\frac{\sin(k\pi x)}{k\sin(\pi x)}\right| \leq 1$

3. homework □

The Fejér kernel F_k has a pronounced maximum at 0 and quickly decays to zero away from 0 (for $k \to \infty$ it approximates δ_0). We now construct a trigonometric polynomial g with

$$g(x_i) = \mathrm{sgn}(b_i), \qquad g'(x_i) = 0, \qquad i = 1, \ldots, N$$

(and hopefully the x_i being global extrema, since we want $|g| \leq 1$).

- idea: take g as linear combination of the shifted kernels $F_k(x - x_i)$
- while $F_k(x - x_i)$ has a pronounced maximum at x_i, the other summands (though small) may shift the extremum slightly away from x_i
- as a remedy, could perturb x_i to $\tilde{x}_i$; however, finding the correct $\tilde{x}_i$ is highly nonlinear problem
- instead: exploit $F_k(x - \tilde{x}_i) \approx F_k(x - x_i) + (\tilde{x}_i - x_i)F_k'(x - x_i) \Rightarrow$ take ansatz

$$g(x) = \sum_{j=1}^{N} \alpha_j F_k(x - x_j) + \beta_j F_k'(x - x_j).$$

Theorem 155 (Fejér coefficients). *The coefficients $\alpha = (\alpha_1, \ldots, \alpha_N)^T$ and $\beta = (\beta_1, \ldots, \beta_N)^T$ satisfy*

$$\underbrace{\begin{pmatrix} \frac{D_0}{k+1} & \frac{D_1}{\sqrt{2/3}\pi(k+1)^2} \\ \frac{-D_1}{\sqrt{2/3}\pi(k+1)^2} & \frac{-D_2}{\frac{2}{3}\pi^2(k+1)^3} \end{pmatrix}}_{M} \underbrace{\begin{pmatrix} (k+1)\alpha \\ \sqrt{2/3}\pi(k+1)^2\beta \end{pmatrix}}_{V} = \underbrace{\begin{pmatrix} \mathrm{sgn}(b_1) \\ \vdots \\ \mathrm{sgn}(b_N) \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{W}$$

for $D_0 = (F_k(x_l - x_j))_{l,j}$, $D_1 = (F_k'(x_l - x_j))_{l,j}$, $D_2 = (F_k''(x_l - x_j))_{l,j}$.
Moreover, there exists $\bar{C} > 0$ such that $\Delta \geq \frac{\bar{C}}{k+1}$ implies that the equation is solvable, and

$$\|V - W\|_\infty \leq \frac{\bar{C}}{\Delta^2(k+1)^2}.$$

Proof.

- $g(x_i) = \mathrm{sgn}(b_i)$ & $g'(x_i) = 0$ for $i = 1, \ldots, N$
  $\Leftrightarrow \left(\begin{smallmatrix} D_0 & D_1 \\ D_1 & D_2 \end{smallmatrix}\right)\left(\begin{smallmatrix} \alpha \\ \beta \end{smallmatrix}\right) = (\mathrm{sgn}(b_1), \cdots, \mathrm{sgn}(b_N), 0, \cdots, 0)^T$
  $\Leftrightarrow$ given equation system with M
- M is symmetric and diagonally dominant:
  - let M_d be the diagonal and $\tilde{M}$ the rest of M, i.e.

$$M = M_d + \tilde{M} \qquad \text{with} \qquad M_d = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 1-\frac{1}{(k+1)^2} & \\ & & & & \ddots & \\ & & & & & 1-\frac{1}{(k+1)^2} \end{pmatrix}.$$

  - compute $\|\tilde{M}\|_\infty$
    * for $l \leq N$ we have

$$\begin{aligned} \sum_{j \neq l} |M_{lj}| &= \sum_{j \neq l, j \leq N} \left|\frac{(D_0)_{lj}}{k+1}\right| + \sum_{j=1}^{N} \left|\frac{(D_1)_{lj}}{\sqrt{2/3}\pi(k+1)^2}\right| \\ &\leq \sum_{j \neq l, j \leq N} \frac{1}{4(k+1)^2 \mathrm{dist}^2(x_j, x_l)} + \sum_{j \neq l, j \leq N} \frac{3\sqrt{3}}{4\sqrt{2}(k+1)^2 \mathrm{dist}^2(x_j, x_l)} \\ &\leq \frac{1}{2(k+1)^2\Delta^2} \sum_{j=1}^{1/2\Delta} \frac{1}{j^2} + \frac{3\sqrt{3}}{2\sqrt{2}(k+1)^2\Delta^2} \sum_{j=1}^{1/2\Delta} \frac{1}{j^2} \leq \frac{4}{(k+1)^2\Delta^2} \end{aligned}$$

    * again for $l \leq N$ we have

$$\begin{aligned} \sum_{j \neq N+l} |M_{N+l,j}| &= \sum_{j=1}^{N} \left|\frac{(D_1)_{lj}}{\sqrt{2/3}\pi(k+1)^2}\right| + \sum_{j \neq l, j \leq N} \left|\frac{(D_2)_{lj}}{2\pi^2(k+1)^3/3}\right| \\ &\leq \sum_{j \neq l, j \leq N} \frac{3\sqrt{3}}{4\sqrt{2}(k+1)^2 \mathrm{dist}^2(x_j, x_l)} + \sum_{j \neq l, j \leq N} \frac{C}{(k+1)^2 \mathrm{dist}^2(x_j, x_l)} \\ &\leq \frac{3\sqrt{3}}{2\sqrt{2}(k+1)^2\Delta^2} \sum_{j=1}^{1/2\Delta} \frac{1}{j^2} + \frac{2C}{(k+1)^2\Delta^2} \sum_{j=1}^{1/2\Delta} \frac{1}{j^2} \leq \frac{4C+4}{(k+1)^2\Delta^2} \end{aligned}$$

    * $\|\tilde{M}\|_\infty \leq \frac{4C+2}{(k+1)^2\Delta^2}$
  - $\Delta > \frac{2\sqrt{4C+2}}{k+1} \Rightarrow \|\tilde{M}\|_\infty \leq \frac{1}{4} \Rightarrow \|M^{-1}\|_\infty \leq \left(1 - \frac{1}{(k+1)^2} - \|\tilde{M}\|_\infty\right)^{-1} \leq 2$

- $\|V\|_\infty = \|M^{-1}W\|_\infty \leq \|M^{-1}\|_\infty \|W\|_\infty \leq 2 \cdot 1$
- $M_d V - W = -\tilde{M}V \Rightarrow V - W = V - M_d^{-1}W = -M_d^{-1}\tilde{M}V$
  $\Rightarrow \|V - W\|_\infty \leq \|M_d^{-1}\|_\infty \|\tilde{M}\|_\infty \|V\|_\infty \leq \frac{\text{const.}}{(k+1)^2\Delta^2}$ □

Theorem 156 (Existence of trigonometric polynomial)**.** *There exist constants $c_1, c_2, c_3, c_4 > 0$ such that if $\Delta \geq \frac{c_1}{k+1}$, then g satisfies the desired conditions with $R = \frac{c_2}{k+1}$, $\eta = c_3(k+1)^2$, $\kappa = c_4(k+1)^2$.*

Proof.

- let wlog. x_1 be closest to x, then
  (using $|F_k''| \leq |F_k''(0)| = \frac{2}{3}\pi^2 k(k+1)(k+2)$ & $|F_k'''| \leq 30(k+1)^4$)

$$\begin{aligned}
|g''(x)| &\leq \sum_{i=1}^N |\alpha_i||F_k''(x-x_i)| + |\beta_i||F_k'''(x-x_i)| \\
&\leq \text{const.} k(k+2) + \text{const.}(k+1)^2 + \sum_{i=2}^N \frac{\text{const.}}{\sin^2(\pi(x-x_i))} + \frac{\text{const.}}{\sin^2(\pi(x-x_i))} \\
&\leq \tilde{C}\left((k+1)^2 + \sum_{i=1}^{1/2\Delta} \frac{1}{i^2\Delta^2}\right) \leq \hat{C}\left((k+1)^2 + \frac{1}{\Delta^2}\right) \leq \hat{C}(1+1/c_1^2)(k+1)^2
\end{aligned}$$

  thus can choose $\eta = \hat{C}(1+1/c_1^2)(k+1)^2/2 = c_3(k+1)^2$
- analogously $|g'''(x)| \leq \text{const.}(1+1/c_1^2)(k+1)^3$
- wlog. consider second derivative at x_1

$$\begin{aligned}
-\text{sgn}(b_1)g''(x_1) &\geq -\text{sgn}(b_1)\alpha_1 F_k''(0) - \sum_{i=2}^N (|\alpha_i||F_k''(x_1-x_i)| + |\beta_i||F_k'''(x_1-x_i)|) \\
&\geq \left(1 - \frac{\bar{C}}{c_1^2}\right)\frac{2}{3}\pi^2 k(k+2) - \sum_{i=2}^N \left(\frac{\text{const.}}{\sin^2(\pi(x-x_i))} + \frac{\text{const.}}{\sin^2(\pi(x-x_i))}\right) \\
&\geq \left(1 - \frac{\bar{C}}{c_1^2}\right)\frac{2}{3}\pi^2 k(k+2) - \tilde{C}\sum_{i=1}^{1/2\Delta}\frac{1}{i^2\Delta^2} \\
&\geq \check{C}\left(1 - \tfrac{\hat{C}}{c_1^2}\right)(k+1)^2
\end{aligned}$$

  $\Rightarrow$ if c_1 is chosen large and c_2 small enough, $-\text{sgn}(b_1)g''(x) > c_4(k+1)^2$ whenever $\text{dist}(x, x_1) \leq \frac{c_2}{k+1}$
- $\Rightarrow$ choose $R = \frac{c_2}{k+1}$, $\kappa = c_4(k+1)^2/2$
- Assume $R < \Delta/2$ and $F_k(R) \leq F_k(0) + F_k''(0)R^2$ (else decrease c_2) and let $x \in B^c$, x_1 closest to x.

$$\begin{aligned}
|g(x)| &\leq \sum_{i=1}^N |\alpha_i| F_k(x-x_i) + |\beta_i||F_k'(x-x_i)| \\
&\leq |\alpha_1|F_k(R) + |\beta_1||F_k'(x-x_1)| + \sum_{i=2}^N \frac{2}{(k+1)^2\sin^2(\pi(x-x_i))} + \sum_{i=2}^N \frac{\text{const.}}{c_1^2(k+1)^2}\frac{3\pi}{\sin^2(\pi(x-x_i))} \\
&\leq (1 + \tfrac{\text{const.}}{c_1^2})\frac{F_k(R)}{k+1} + \frac{\text{const.}}{c_1^2} + \frac{\text{const.}}{(k+1)^2}\sum_{i=2}^N \frac{1}{\text{dist}^2(x,x_i)} \\
&\leq (1 + \tfrac{\text{const.}}{c_1^2})\frac{F_k(0) + F_k''(0)R^2}{k+1} + \frac{\text{const.}}{c_1^2} + \frac{\text{const.}}{(k+1)^2}\sum_{i=1}^{1/2\Delta}\frac{1}{i^2\Delta^2} \\
&\leq (1 + \tfrac{\text{const.}}{c_1^2})(1 - \tfrac{2}{3}\pi^2 k(k+2)R^2) + \frac{\text{const.}}{c_1^2} < 1 - \kappa R^2 \text{ if } c_4 \text{ small, } c_1 \text{ big enough}
\end{aligned}$$

□

Remark 157 (Higher dimensions)**.** *In higher dimensions one just builds the dual variables as linear combinations of tensor products $F_k(x)F_k(y)F_k(z)\cdots$ of Fejér kernels and their derivatives.*

17 Fourier transform

The Fourier transform is the forward operator in magnetic resonance tomography. However, it also helps to express other forward operators (such as convolution or the Radon or X-ray transform) in a basis that simplifies their understanding. Below all functions will be complex-valued without explicit mention.

Definition 158 (Fourier transform)**.** *The* Fourier transform *is the linear map* $\mathcal{F} : L^1(\mathbb{R}^d) \to L^\infty(\mathbb{R}^d)$*,*

$$\mathcal{F}(f)(\xi) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(x) e^{-ix\cdot\xi} \, \mathrm{d}x.$$

We write $\hat{f} = \mathcal{F}(f)$*. If* f *is vector-valued,* $\mathcal{F}$ *is applied to each component. The* inverse Fourier transform *is defined as the linear map* $\mathcal{F}^{-1} : L^1(\mathbb{R}^d) \to L^\infty(\mathbb{R}^d)$*,*

$$\mathcal{F}^{-1}(g)(x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} g(\xi) e^{ix\cdot\xi} \, \mathrm{d}\xi.$$

We write $\check{f} = \mathcal{F}^{-1}(f)$*.*

Remark 159 (Fourier transform on Radon measures)**.** *One can even extend* $\mathcal{F}$ *(and analogously* $\mathcal{F}^{-1}$*) to Radon measures* ν *by*

$$\mathcal{F}(\nu)(\xi) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{-ix\cdot\xi} \, \mathrm{d}\nu(x).$$

Definition 160 (Multiindex)**.** *A* multiindex *in* $\mathbb{R}^d$ *is a vector* $\alpha \in \mathbb{N}_0^d$*. One writes*

$$D^\alpha = (\tfrac{\partial}{\partial x_1})^{\alpha_1} \cdots (\tfrac{\partial}{\partial x_d})^{\alpha_d}, \qquad x^\alpha = x_1^{\alpha_1} \cdots x_d^{\alpha_d}, \qquad |\alpha| = \alpha_1 + \ldots + \alpha_d.$$

Theorem 161 (Growth properties of Fourier transform)**.**

1. $\|\hat{f}\|_{L^\infty}, \|\check{f}\|_{L^\infty} \leq (2\pi)^{-d/2} \|f\|_{L^1}$
2. $f \in L^1(\mathbb{R}^d) \Rightarrow \hat{f}, \check{f} \in C^0(\mathbb{R}^d)$
3. $f, \nabla f \in L^1(\mathbb{R}^d) \Rightarrow \widehat{\nabla f}(\xi) = i\hat{f}(\xi)\xi$
4. $g \in L^1(\mathbb{R}^d)$ *for* $g(x) = xf(x) \Rightarrow \hat{g}(\xi) = i\nabla\hat{f}(\xi)$
5. *Let* $\alpha \in \mathbb{N}_0^d$*. If* f *is sufficiently differentiable and* f*,* $D^\alpha f$*,* $f_\alpha(x) = x^\alpha f(x)$ *are integrable,*

$$\widehat{D^\alpha f}(\xi) = i^{|\alpha|} \xi^\alpha \hat{f}(\xi),$$
$$\widehat{f_\alpha}(\xi) = i^{|\alpha|} D^\alpha \hat{f}(\xi).$$

Thus, a differentiability order of f *implies a decay order of* $\hat{f}$*; a decay order of* f *implies a differentiability order of* $\hat{f}$*. Analogously for inverse Fourier transform.*

Proof.

1. $|\hat{f}(\xi)|, |\check{f}(\xi)| \leq (2\pi)^{-d/2} \int_{\mathbb{R}^d} |f(x)| \, \mathrm{d}x$
2. $\lim_{\xi\to\xi_0} \hat{f}(\xi) = (2\pi)^{-d/2} \lim_{\xi\to\xi_0} \int_{\mathbb{R}^d} f(x)e^{-ix\cdot\xi} \, \mathrm{d}x \overset{\text{dom. conv. thm.}}{=} (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(x) e^{-ix\cdot\xi_0} \, \mathrm{d}x = \hat{f}(\xi_0)$
3. $\widehat{\partial f/\partial x_i}(\xi) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \frac{\partial f}{\partial x_i} e^{-ix\cdot\xi} \, \mathrm{d}x = -(2\pi)^{-d/2} \int_{\mathbb{R}^d} f e^{-ix\cdot\xi}(-i\xi_i) \, \mathrm{d}x = i\xi_i \hat{f}(\xi)$
4. $i\frac{\partial \hat{f}}{\partial \xi_i}(\xi) = (2\pi)^{-d/2} \int_{R^d} f(x) e^{-ix\cdot\xi} x_i \, \mathrm{d}x$
5. induction using previous two points □

Proposition 162 (Algebraic identities for the Fourier transform)**.**

1. $f, g \in L^1(\mathbb{R}^d) \Rightarrow \int_{\mathbb{R}^d} f(x)\hat{g}(x) \, \mathrm{d}x = \int_{\mathbb{R}^d} \hat{f}(x) g(x) \, \mathrm{d}x$
2. $g(x) = \bar{f}(-x) \Rightarrow \hat{g} = \bar{\hat{f}}$

3. $f \in L^1(\mathbb{R}^d) \Rightarrow \bar{\hat{f}} = \check{\bar{f}}$

4. $\lambda \neq 0,\ f_\lambda(x) = f(\lambda x) \Rightarrow \widehat{f_\lambda}(\xi) = |\lambda|^{-d}\hat{f}(\frac{\xi}{\lambda})$

5. $a \in \mathbb{R}^d,\ f_a(x) = f(x+a) \Rightarrow \widehat{f_a}(\xi) = e^{ia\cdot\xi}\hat{f}(\xi)$

6. let $\mathcal{F}_i(f)(x) = \mathcal{F}(f(x_1, \dots, x_{i-1}, \cdot, x_{i+1}, \dots, x_d))(x_i)$*, then* $\mathcal{F} = \mathcal{F}_1 \cdots \mathcal{F}_d$

7. $f(x) = f_1(x_1)\cdots f_d(x_d) \Rightarrow \hat{f}(\xi) = \widehat{f_1}(\xi_1)\cdots\widehat{f_d}(\xi_d)$

Proof. 1. $\hat{f}, \hat{g} \in L^\infty(\mathbb{R}^d)$, and both expressions equal $(2\pi)^{-d/2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} e^{-ix\cdot\xi}f(x)g(\xi)\,\mathrm{d}x\,\mathrm{d}\xi$

2. $\hat{g}(\xi) = (2\pi)^{-d/2}\int_{\mathbb{R}^d} e^{-ix\cdot\xi}\bar{f}(-x)\,\mathrm{d}x = (2\pi)^{-d/2}\int_{\mathbb{R}^d}\overline{e^{-iy\cdot\xi}f(y)}\,\mathrm{d}y = \bar{\hat{f}}$

3. trivial

4. $\widehat{f_\lambda}(\xi) = (2\pi)^{-d/2}\int_{\mathbb{R}^d} f(\lambda x)e^{-i\xi\cdot x}\,\mathrm{d}x = (2\pi)^{-d/2}|\lambda|^{-d}\int_{\mathbb{R}^d} f(y)e^{-i\frac{\xi}{\lambda}\cdot y}\,\mathrm{d}y$

5. trivial

6. Fubini: $\hat{f}(\xi) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\cdots\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f(x_1,\dots,x_d)e^{-ix_d\xi_d}\,\mathrm{d}x_d\dots e^{-ix_2\xi_2}\,\mathrm{d}x_2\,e^{-ix_1\xi_1}\,\mathrm{d}x_1$

7. follows from previous point and linearity ☐

Example 163 (Fourier transforms)**.**

- $f(x) = e^{-|x|^2/2} \Rightarrow \hat{f}(\xi) = e^{-|\xi|^2/2}$
  - $f(x) = f_1(x_1)\cdots f_d(x_d)$ *with* $f_i(x_i) = e^{-x_i^2/2}$*, thus sufficient to consider* $d = 1$*.*
  - $f'(x) + xf(x) = 0$
  - *taking Fourier transform,* $i\xi\hat{f}(\xi) + i\hat{f}'(\xi) = 0$
  - $f(0) = 1$ *and* $\hat{f}(0) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-x^2/2}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}}\left(\int_{\mathbb{R}}\int_{\mathbb{R}} e^{-(x^2+y^2)/2}\,\mathrm{d}x\,\mathrm{d}y\right)^{\frac{1}{2}}$
  $$= \frac{1}{\sqrt{2\pi}}\left(\int_0^{2\pi}\int_0^\infty e^{-r^2/2}r\,\mathrm{d}r\,\mathrm{d}\varphi\right)^{\frac{1}{2}} = \left([-e^{-r^2/2}]_0^\infty\right)^{\frac{1}{2}} = 1$$
  - $\Rightarrow f$ *and* $\hat{f}$ *solve same ODE with same initial condition*
- $f(x) = \chi_{[-1,1]}(x) \Rightarrow \hat{f}(\xi) = \sqrt{2/\pi}\,\mathrm{sinc}(\xi)$ *for* $\mathrm{sinc}(x) = \sin(x)/x$ *if* $x \neq 0$, $\mathrm{sinc}(0) = 1$
  - $\hat{f}(\xi) = \frac{1}{\sqrt{2\pi}}\int_{-1}^1 e^{-ix\xi}\,\mathrm{d}x = [\frac{1}{-i\xi\sqrt{2\pi}}e^{-ix\xi}]_{-1}^1 = \frac{e^{-i\xi}-e^{i\xi}}{-i\xi\sqrt{2\pi}} = \sqrt{\frac{2}{\pi}}\mathrm{sinc}\xi$
- $\nu = \delta_x \Rightarrow \hat{\nu}(\xi) = (2\pi)^{-d/2}e^{-ix\cdot\xi}$

Proposition 164 (Fourier transform of radially symmetric functions)**.** *If* $f \in L^1(\mathbb{R}^d)$ *is radially symmetric, i. e.* $f(x) = F(|x|)$*, then so is* $\hat{f}$ *with*

$$\hat{f}(\xi) = \int_0^\infty r^{d-1}F(r)J(r|\xi|)\,\mathrm{d}r \qquad \text{for } J(s) = (2\pi)^{-d/2}\int_{S^{d-1}} e^{-is(1\ 0\ \dots)^T\cdot\theta}\,\mathrm{d}\mathcal{H}^{d-1}(\theta).$$

For $d = 2$ *we have* $J = J_0$ *with* J_n *the nth order Bessel function of the first kind, the bounded solution to*

$$s^2J_n''(s) + sJ_n'(s) + (s^2 - n^2)J_n(s) = 0 \qquad \text{with} \qquad \int_0^\infty J_n(s)\,\mathrm{d}s = 1.$$

For $d = 3$ *we have* $J(s) = \frac{2}{\sqrt{2\pi}}\mathrm{sinc}(s)$*.*

Proof. Using polar coordinates $(r,\theta)\in[0,\infty)\times S^{d-1}$,

$$\hat f(\xi)=(2\pi)^{-\frac d2}\int_{S^{d-1}}\int_0^\infty f(r\theta)e^{-ir\theta\cdot\xi}r^{d-1}\,\mathrm dr\,\mathrm d\mathcal H^{d-1}(\theta)=\int_0^\infty r^{d-1}F(r)(2\pi)^{-\frac d2}\int_{S^{d-1}}e^{-ir|\xi|\theta\cdot\frac{\xi}{|\xi|}}\,\mathrm d\mathcal H^{d-1}(\theta)\,\mathrm dr.$$

Properties for $d=2$ follow by direct calculation. For $d=3$,

$$\begin{aligned}J(s)=(2\pi)^{-3/2}\int_{S^2}e^{-is(1\ 0\ 0)^T\cdot\theta}\,\mathrm d\mathcal H^2(\theta)&=(2\pi)^{-3/2}\mathcal H^1(S^1)\int_0^\pi\sin\varphi\, e^{-is\cos\varphi}\,\mathrm d\varphi\\&=(2\pi)^{-3/2}\mathcal H^1(S^1)\left[\tfrac{e^{-is\cos\varphi}}{is}\right]_{\varphi=0}^\pi=2(2\pi)^{-3/2}\mathcal H^1(S^1)\mathrm{sinc}(s).\end{aligned}$$ □

Theorem 165 (Convolution theorem)**.** *If $f,g\in L^1(\mathbb R^d)$, then $\widehat{f*g}=(2\pi)^{d/2}\hat f\hat g$.*

Proof. By Young's convolution theorem, $f*g\in L^1(\mathbb R^d)$.

$$\begin{aligned}\widehat{f*g}(\xi)&=(2\pi)^{-d/2}\int_{\mathbb R^d}e^{-ix\cdot\xi}\int_{\mathbb R^d}f(z)g(x-z)\,\mathrm dz\,\mathrm dx\\&=(2\pi)^{-d/2}\int_{\mathbb R^d}e^{-iz\cdot\xi}f(z)\int_{\mathbb R^d}e^{-i(x-z)\cdot\xi}g(x-z)\,\mathrm dx\,\mathrm dz\\&=\hat g(\xi)\int_{\mathbb R^d}e^{-iz\cdot\xi}f(z)\,\mathrm dz\\&=(2\pi)^{d/2}\hat f(\xi)\hat g(\xi)\end{aligned}$$ □

Theorem 166 (Plancherel's theorem)**.** *If $f\in L^1(\mathbb R^d)\cap L^2(\mathbb R^d)$, then $\|\hat f\|_{L^2}=\|f\|_{L^2}$.*

Proof.

- $\|f\|_{L^2}^2=(f*g)(0)$ for $g(x)=\bar f(-x)$
- $\|\hat f\|_{L^2}^2=\int_{\mathbb R^d}|\hat f|^2\,\mathrm d\xi=(2\pi)^{-d/2}\int_{\mathbb R^d}\widehat{f*g}\,\mathrm d\xi$
- set $v_\epsilon(x)=e^{-\epsilon|x|^2/2}$, then $\hat v_\epsilon(\xi)=\epsilon^{-d/2}e^{-|\xi|^2/2\epsilon}$ and
  $\hat v_\epsilon(\xi)\ge0,\qquad \int_{\mathbb R^d}\hat v_\epsilon(\xi)\,\mathrm d\xi=\int_{\mathbb R^d}e^{-|\xi|^2/2}\,\mathrm d\xi=(2\pi)^{d/2},\qquad \hat v_\epsilon\xrightarrow[\epsilon\to0]{}(2\pi)^{d/2}\delta_0$
- set $w=f*g$, then $\hat w=(2\pi)^{d/2}|\hat f|^2\ge0$ and $w\in C^0(\mathbb R^d)$ since
  $$\lim_{h\to0}w(x+h)=\lim_{h\to0}(f*g(\cdot+h))(x)=f*g(x)\quad\text{due to}\quad g(\cdot+h)\xrightarrow[h\to0]{}g\text{ in }L^2(\mathbb R^d)$$
- $$(2\pi)^{d/2}w(0)=\lim_{\epsilon\to0}\int_{\mathbb R^d}w(x)\hat v_\epsilon(x)\,\mathrm dx=\lim_{\epsilon\to0}\int_{\mathbb R^d}\hat w(x)v_\epsilon(x)\,\mathrm dx\overset{\text{mon. conv. thm.}}{=}\int_{\mathbb R^d}\hat w\,\mathrm d\xi$$ □

Theorem 167 (Parseval's theorem)**.** *If $f,g\in L^1(\mathbb R^d)\cap L^2(\mathbb R^d)$, then $(f,g)_{L^2}=(\hat f,\hat g)_{L^2}$.*

Proof. $(f,g)_{L^2}+(g,f)_{L^2}=\|f+g\|_{L^2}^2-\|f\|_{L^2}^2-\|g\|_{L^2}^2=\|\hat f+\hat g\|_{L^2}^2-\|\hat f\|_{L^2}^2-\|\hat g\|_{L^2}^2=(\hat f,\hat g)_{L^2}+(\hat g,\hat f)_{L^2}$
$i(f,g)_{L^2}-i(g,f)_{L^2}=\|if+g\|_{L^2}^2-\|f\|_{L^2}^2-\|g\|_{L^2}^2=\|i\hat f+\hat g\|_{L^2}^2-\|\hat f\|_{L^2}^2-\|\hat g\|_{L^2}^2=i(\hat f,\hat g)_{L^2}-i(\hat g,\hat f)_{L^2}$
take first equation minus i times second □

Theorem 168 (Inverse Fourier transform)**.** *If f and $\hat f$ (thus by proposition 162 also $\check f$) lie in $L^1(\mathbb R^d)\cap L^2(\mathbb R^d)$, then $\mathcal F^{-1}\mathcal Ff=\mathcal F\mathcal F^{-1}f=f$.*

Proof. $(g,\bar f)_{L^2}=(\hat g,\hat{\bar f})_{L^2}=(\hat g,\bar{\check f})_{L^2}=\int_{\mathbb R^d}\hat g\check f\,\mathrm dx=\int_{\mathbb R^d}g\hat{\check f}\,\mathrm dx=(g,\bar{\hat{\check f}})_{L^2}\ \forall g\in L^1(\mathbb R^d)\cap L^2(\mathbb R^d)$
analogously $\check{\hat f}=f$ □

$\hat f\in L^1(\mathbb R^d)\cap L^2(\mathbb R^d)$ is for instance implied by $f\in W^{d+4,1}(\mathbb R^d)$. Indeed, let $\tilde d$ be the largest even number below $d+4$, then by theorem 161(5) $|\xi|^{\tilde d}\hat f(\xi)=\widehat{(-\Delta)^{\tilde d/2}f}\in L^\infty(\mathbb R^d)$ so that $\hat f\in L^1(\mathbb R^d)\cap L^2(\mathbb R^d)$.

Corollary 169 (Fourier transform on L^2)**.** *The Fourier transform and inverse Fourier transform can be uniquely extended to isometric isomorphisms of $L^2(\mathbb R^d)$ which are inverses of each other and satisfy proposition 162 with L^1 replaced by L^2.*

Proof.

- $\mathcal{F}$, $\mathcal{F}^{-1}$ are isometries on the dense subset $L^1 \cap L^2$ of L^2
- unique norm-preserving extension onto L^2 by Hahn–Banach
- $\mathcal{F}\mathcal{F}^{-1}f = \mathcal{F}^{-1}\mathcal{F}f = f$ for the dense subset $L^1 \cap L^2 \cap W^{d+4,1}(\mathbb{R}^d)$ of L^2
- all properties extend by continuity □

Remark 170 (Integral representation of Fourier transform)**.** *For an L^2-function f we might sometimes write $\hat{f}(\xi) = \int_{\mathbb{R}^d} f(x)e^{-ix\cdot\xi}\,\mathrm{d}x$ even though the integral is actually not well-defined; it then has to be interpreted as the limit of $\int_{\mathbb{R}^d} f_n(x)e^{-ix\cdot\xi}\,\mathrm{d}x$ for a sequence $f_n \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ with $f_n \to f$ in $L^2(\mathbb{R}^d)$.*

Summarizing, the great strengths of the Fourier transform are that it is an isometric isomorphism (i. e. an orthonormal basis change) on $L^2(\mathbb{R}^d)$ that turns convolution into pointwise multiplication and differentiation into multiplication with the frequency.

18 Tempered distributions

The Fourier transform and related transforms can actually be extended to much larger spaces. As an example we have already seen the extension to measures. We will extend it to the dual space of the so-called Schwartz space, the so-called tempered distributions.

Definition 171 (Schwartz space)**.** *The* Schwartz space of rapidly decaying functions *is the function space*

$$\mathcal{S}(\mathbb{R}^d) = \{f \in C^\infty(\mathbb{R}^d) \,|\, \|f\|_n < \infty \,\forall n \in \mathbb{N}\}$$

with the (semi-)norms

$$\|f\|_n = \sup_{x\in\mathbb{R}^d} \max_{|\alpha|,|\beta|\leq n} |x^\alpha D^\beta f(x)|$$

(with respect to which it is not complete). It becomes a Fréchet space (a complete metric space) with the metric

$$d_{\mathcal{S}}(f,g) = \sum_{n=1}^\infty \frac{1}{2^n}\frac{\|f-g\|_n}{1+\|f-g\|_n}.$$

Example 172 (Schwartz functions)**.**

- *infinitely smooth functions with compact support, e. g.*

$$f(x) = \begin{cases} \exp(-\frac{1}{1-|x|^2}) & \text{if } |x| < 1 \\ 0 & \text{else} \end{cases}$$

  or its translations, scalings, sums, convolutions with compactly supported functions

- *normal distribution $f(x) = \exp(-|x|^2/2)$*

It can readily be checked that $f_n \to f$ in $\mathcal{S}(\mathbb{R}^d)$ if and only if $\|f_n - f\|_k \to 0$ for all k and that differentiation D^α and translation $f \mapsto f(\cdot + x)$ are continuous from $\mathcal{S}(\mathbb{R}^d)$ into itself. Likeweise, pointwise multiplication $\mathcal{S}(\mathbb{R}^d) \times \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d)$ is continuous. Note further that for $B \subset \mathbb{R}^d$ the unit ball with complement B^c and for $p(x) = |x|^{-d-1}$ and any $f : \mathbb{R}^d \to \mathbb{C}$ we have

$$\|f\|_{L^1} = \int_B |f|\,\mathrm{d}x + \int_{B^c} |f|\,\mathrm{d}x \leq \mathcal{L}^d(B)\|f\|_0 + \|p\|_{L^1(B^c)}\|f/p\|_{L^\infty(B^c)} \leq C(d)\|f\|_{d+2}$$

with a constant $C(d)$ depending on the dimension.

Theorem 173 (Fourier transform on Schwartz space)**.** *$\mathcal{F}$ is a continuous automorphism on $\mathcal{S}(\mathbb{R}^d)$.*

Proof.

- $\xi^\alpha D^\beta \hat{f}(\xi) = i^{|\beta|}\xi^\alpha \widehat{f_\beta}(\xi) = i^{|\alpha|+|\beta|}\widehat{D^\alpha f_\beta}(\xi)$ for $f_\beta(x) = x^\beta f(x)$

- thus $\|\hat f\|_n = \max_{|\alpha|,|\beta|\leq n} \|\widehat{D^\alpha f_\beta}\|_{L^\infty} \leq \max_{|\alpha|,|\beta|\leq n} \|D^\alpha f_\beta\|_{L^1} \leq \text{const.}\|f\|_{n+d+2}$
- thus $f \in \mathcal{S}(\mathbb{R}^d) \Rightarrow \hat f \in \mathcal{S}(\mathbb{R}^d)$, and $\mathcal{F}$ is continuous in $0 \in \mathcal{S}(\mathbb{R}^d)$ and thus on $\mathcal{S}(\mathbb{R}^d)$
- analogously, $\mathcal{F}^{-1}$ is continuous from $\mathcal{S}(\mathbb{R}^d)$ into itself □

Definition 174 (Tempered distributions)**.** *The space $\mathcal{S}'(\mathbb{R}^d)$ of* tempered distributions *on $\mathbb{R}^d$ is the space of continuous linear functionals $\mathcal{S}(\mathbb{R}^d) \to \mathbb{C}$.*

Example 175 (Tempered distributions)**.**

- *Any $g \in L^p(\mathbb{R}^d)$ induces a tempered distribution T_g via $T_g(f) = \int_{\mathbb{R}^d} fg\,\mathrm{d}x$.*
- *Any $\nu \in \mathcal{M}(\mathbb{R}^d)$ induces a tempered distribution T_ν via $T_\nu(f) = \int_{\mathbb{R}^d} f\,\mathrm{d}\nu$.*
- *Let $\alpha \in \mathbb{N}_0^d$, $x \in \mathbb{R}^d$, then $T_{\alpha,x} \in \mathcal{S}'(\mathbb{R}^d)$ for $T_{\alpha,x}(f) = D^\alpha f(x)$.*
- *Any polynomial g on $\mathbb{R}^d$ induces a tempered distribution T_g via $T_g(f) = \int_{\mathbb{R}^d} fg\,\mathrm{d}x$.*
- *Special cases: $\delta_x, \int_S \cdot\,\mathrm{d}\mathcal{H}^k \in \mathcal{S}(\mathbb{R}^d)$ for $x \in \mathbb{R}^d$, $S \subset \mathbb{R}^d$ k-dimensional and smooth (think of X-ray/Radon transform)*

We will identify L^p-functions or Radon measures with distributions, thus $L^p(\mathbb{R}^d), \mathcal{M}(\mathbb{R}^d) \subset \mathcal{S}'(\mathbb{R}^d)$. Below, by a tilde we will denote the map

$$f \mapsto \tilde f, \quad \tilde f(x) = f(-x).$$

Definition 176 (Operations on distributions)**.** *Let $T \in \mathcal{S}'(\mathbb{R}^d)$, $g \in \mathcal{S}(\mathbb{R}^d)$, $\alpha \in \mathbb{N}_0^d$.*

1. *The* (distributional) derivative $D^\alpha T \in \mathcal{S}'(\mathbb{R}^d)$ *of T is defined by*

$$(D^\alpha T)(f) = (-1)^{|\alpha|} T(D^\alpha f).$$

2. *The* product $gT \in \mathcal{S}'(\mathbb{R}^d)$ *of T with g is defined by*

$$(gT)(f) = T(gf).$$

3. *The* convolution $T * g \in \mathcal{S}'(\mathbb{R}^d)$ *of T with g is defined by*

$$(T * g)(f) = T(\tilde g * f).$$

4. *The* Fourier transform *and* inverse Fourier transform $\hat T, \check T \in \mathcal{S}'(\mathbb{R}^d)$ *of T are defined by*

$$\hat T(f) = T(\hat f), \qquad \check T(f) = T(\check f).$$

Remark 177 (Motivation for formulas)**.** *The above formulas are chosen for consistency with the case when T equals a Schwartz function $\phi \in \mathcal{S}(\mathbb{R}^d)$, in which*

$$\begin{aligned} \int_{\mathbb{R}^d} D^\alpha \phi\, f\,\mathrm{d}x &= (-1)^{|\alpha|} \int_{\mathbb{R}^d} \phi\, D^\alpha f\,\mathrm{d}x,\\ \int_{\mathbb{R}^d} (\phi g)\, f\,\mathrm{d}x &= \int_{\mathbb{R}^d} \phi\,(gf)\,\mathrm{d}x,\\ \int_{\mathbb{R}^d} (\phi * g)\, f\,\mathrm{d}x &= \int_{\mathbb{R}^d} \phi\,(\tilde g * f)\,\mathrm{d}x,\\ \int_{\mathbb{R}^d} \hat\phi\, f\,\mathrm{d}x &= \int_{\mathbb{R}^d} \phi\,\hat f\,\mathrm{d}x. \end{aligned}$$

Example 178 (Fourier transform of Dirac & $1/|x|$)**.**

- *$\hat\delta_0(f) = \delta_0(\hat f) = \hat f(0) = (2\pi)^{-\frac d2} \int_{\mathbb{R}^d} f\,\mathrm{d}x$, thus $\hat\delta_0 = (2\pi)^{-\frac d2}$*
- *$f(x) = \frac{1}{|x|}$ in $\mathbb{R}^2 \Rightarrow \hat f(\xi) = \int_0^\infty J_0(r|\xi|)\,\mathrm{d}r = \frac{1}{|\xi|}$*

Obviously $\hat{\check{T}} = \check{\hat{T}} = T$ and $\widehat{T * g} = (2\pi)^{d/2}\hat{g}\hat{T}$.
Since differentiation, multiplication and the Fourier transform are continuous on $\mathcal{S}(\mathbb{R}^d)$, the above definitions of distributional derivative and (inverse) Fourier transform are well-defined (they indeed yield tempered distributions). The well-definedness of the convolution follows from the following.

Theorem 179 (Convolution of tempered distributions)**.** *Let $T \in \mathcal{S}'(\mathbb{R}^d)$ and $g \in \mathcal{S}(\mathbb{R}^d)$, then $T * g \in \mathcal{S}'(\mathbb{R}^d)$ is well-defined. Moreover we have $T * g \in C^\infty(\mathbb{R}^d)$ and $(T * g)(x) = T(g(x - \cdot))$.*

Proof. 1. Let $f \in \mathcal{S}(\mathbb{R}^d)$, then $T * g(f)$ is well-defined.

- $f, g \in \mathcal{S}(\mathbb{R}^d) \Rightarrow f, \tilde{g} \in L^1(\mathbb{R}^d) \Rightarrow \tilde{g} * f \in L^1(\mathbb{R}^d)$
- $\hat{\tilde{g}}, \hat{f} \in \mathcal{S}(\mathbb{R}^d) \Rightarrow \hat{\tilde{g}}\hat{f} \in \mathcal{S}(\mathbb{R}^d) \Rightarrow \tilde{g} * f = (2\pi)^{d/2}\widecheck{(\hat{\tilde{g}}\hat{f})} \in \mathcal{S}(\mathbb{R}^d)$

2. $T * g$ is linear (trivial) and continuous on $\mathcal{S}(\mathbb{R}^d)$, thus $T * g \in \mathcal{S}'(\mathbb{R}^d)$:

- due to linearity suffices to show continuity in 0, so let $f_n \to 0$ in $\mathcal{S}(\mathbb{R}^d)$
- $\Rightarrow \hat{f}_n \to 0$ in $\mathcal{S}(\mathbb{R}^d) \Rightarrow \hat{\tilde{g}}\hat{f}_n \to 0$ in $\mathcal{S}(\mathbb{R}^d) \Rightarrow \tilde{g} * f_n \to 0$ in $\mathcal{S}(\mathbb{R}^d) \Rightarrow T(\tilde{g} * f_n) \to 0$

3. $T * g(x) = T(g(x - \cdot))$:

- let $f_n \stackrel{*}{\rightharpoonup} \delta_x$ in $\mathcal{M}(\mathbb{R}^d)$, then $\tilde{g} * f_n \to g(x - \cdot)$ in $\mathcal{S}(\mathbb{R}^d)$
- thus $\lim_{n\to\infty}(T * g)(f_n) = \lim_{n\to\infty} T(\tilde{g} * f_n) = T(g(x - \cdot))$

4. Abbreviate $h(x) = T(g(x - \cdot))$, then $h \in C^\infty(\mathbb{R}^d)$, since $D^\alpha h(x) = T((D^\alpha g)(x - \cdot))$ for all $\alpha \in \mathbb{N}_0^d$:

- suffices to consider $\alpha = (1\ 0\ \dots\ 0)$
  (other first derivatives follow analogously and higher ones by induction)
- note $\frac{g(x+(\epsilon\ 0\dots 0)-\cdot)-g(x-\cdot)}{\epsilon} \to_{\epsilon\to 0} \partial_{x_1} g(x - \cdot)$ in $\mathcal{S}(\mathbb{R}^d)$
- $\partial_{x_1} h(x) = \lim_{\epsilon\to 0} T(\frac{g(x+(\epsilon\ 0\dots 0)-\cdot)-g(x-\cdot)}{\epsilon}) = T(\lim_{\epsilon\to 0} \frac{g(x+(\epsilon\ 0\dots 0)-\cdot)-g(x-\cdot)}{\epsilon}) = T(\partial_{x_1} g(x - \cdot))$ □

Remark 180 (Convolution theorem)**.** *Under additional conditions on two tempered distributions R, T (e. g. when their singular supports are disjoint) one can even define their product and sometimes even their convolution or the product of their Fourier transforms. In those cases the convolution theorem $\widehat{T * R} = (2\pi)^{d/2}\hat{T}\hat{R}$ still holds.*

Definition 181 (Shift-invariance)**.** *A bounded linear operator $A : L^p(\mathbb{R}^d) \to L^q(\mathbb{R}^d)$, $1 \le p, q \le \infty$, is called* shift-invariant *if it commutes with translation by $z \in \mathbb{R}^d$, that is, $A(f(\cdot + z)) = (Af)(\cdot + z)$.*

Convolutions $f \mapsto T * f$ are shift-invariant. In fact, they are the only such operators.

Theorem 182 (Shift-invariant operators and convolutions)**.** *Let $A : L^p(\mathbb{R}^d) \to L^q(\mathbb{R}^d)$ be a bounded linear shift-invariant operator, then there exists $T \in \mathcal{S}'(\mathbb{R}^d)$ with $Af = T * f$ for all $f \in \mathcal{S}(\mathbb{R}^d)$.*

Proof.

- $D^\alpha(Af) = A(D^\alpha f)$ (for f sufficiently differentiable):
  - suffices to consider $\alpha = (1\ 0\ \dots\ 0)$
    (other first derivatives follow analogously and higher ones by induction)
  - set $f_h(x) = f(x_1 + h, x_2, \dots, x_d)$, then
    $\|A(\frac{f_h - f}{h}) - A(\partial_{x_1} f)\|_{L^q} = \|A(\frac{f_h - f}{h} - \partial_{x_1} f)\|_{L^q} \le \|A\|\|\frac{f_h - f}{h} - \partial_{x_1} f\|_{L^p} \to_{h\to 0} 0$
  - thus, pointwise limit as $h \to 0$ of $\frac{(Af)(x_1+h,x_2,\dots,x_d)-Af(x)}{h} = \frac{Af_h(x)-Af(x)}{h} = A(\frac{f_h - f}{h})(x)$ exists a. e. and equals $A(\partial_{x_1} f)$
- if $Af = T * f$ $\forall f \in \mathcal{S}(\mathbb{R}^d)$, then necessarily $T(f) = T(\tilde{f}(0 - \cdot)) = T * \tilde{f}(0) = A\tilde{f}(0)$ $\forall f \in \mathcal{S}(\mathbb{R}^d)$
- $T \in \mathcal{S}'(\mathbb{R}^d)$:
  - T is linear

– $|T(f)| = |A\tilde f(0)| \le \|A\tilde f\|_{C^0} \lesssim \|A\tilde f\|_{W^{d+1,q}} \lesssim \max_{|\alpha|\le d+1} \|D^\alpha(A\tilde f)\|_{L^q}$ and

$$\begin{aligned}\|D^\alpha(A\tilde f)\|_{L^q}^p = \|A(D^\alpha \tilde f)\|_{L^q}^p \lesssim \|D^\alpha \tilde f\|_{L^p}^p \\ = \|D^\alpha f\|_{L^p}^p \le \|(1+|x|^2)^{dp}|D^\alpha f|^p\|_{L^\infty}\|(1+|x|^2)^{-dp}\|_{L^1} \lesssim \|f\|_{|\alpha|+2d}^p,\end{aligned}$$

thus $|T(f)| \le \|f\|_{3d+1}$

- let $f \in \mathcal{S}(\mathbb{R}^d)$, then $(Af)(x) = A(f(x+\cdot))(0) = T(f(x-\cdot)) = (T * f)(x)$ □

Remark 183 (Forward operator of microscopy)**.** *Ignoring boundary effects due to a microscope's finite field of view, the forward operator of any microscopy is shift-invariant: Shifting the sample results in the same shift of the recorded image. Thus the forward operator is a convolution, whose kernel, called* point spread function*, is found by imaging a Dirac measure.*

Remark 184 (Space of test functions and distributions)**.** *The tempered distributions $\mathcal{S}'(\mathbb{R}^d)$ actually form a subspace of the space $\mathcal{D}'(\mathbb{R}^d)$ of distributions, the continuous linear functionals on the space $\mathcal{D}(\mathbb{R}^d) = C_c^\infty(\mathbb{R}^d)$ of test functions (infinitely smooth functions with compact support). For these distributions, differentiation, multiplication and convolution can be defined in the same way as for tempered distributions, but the Fourier transform cannot: In the defining equality $\hat T(f) = T(\hat f)$, both f and $\hat f$ would have to have compact support, which is impossible by the Schwartz–Paley–Wiener theorem.*

19 Radon and X-ray transform

For $\theta \in S^{d-1}$ let us abbreviate $\theta^\perp = \{x \in \mathbb{R}^d \,|\, x\cdot\theta = 0\}$.

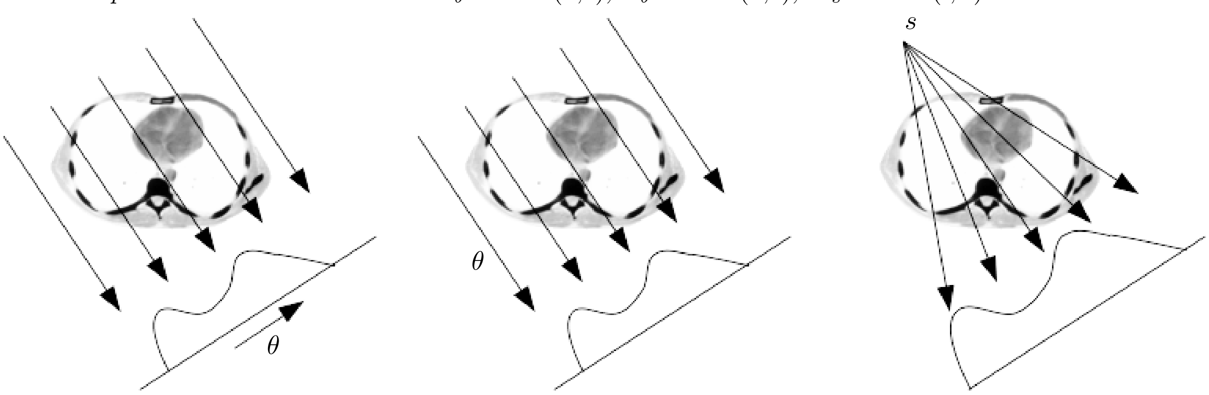
Definition 185 (Radon, X-ray, and divergent beam transform)**.** *Let*

$$\begin{aligned}\mathcal{C} &= S^{d-1}\times\mathbb{R} \subset \mathbb{R}^{d+1},\\ \mathcal{C}' &= \{(\theta,s)\in S^{d-1}\times\mathbb{R}^d \,|\, s\in\theta^\perp\} \subset \mathbb{R}^{2d},\\ \mathcal{C}'' &= S^{d-1}\times\mathbb{R}^d \subset \mathbb{R}^{2d}.\end{aligned}$$

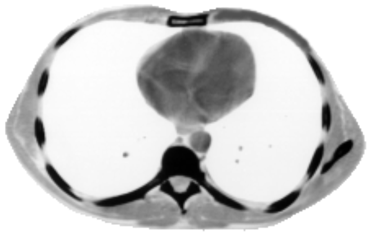
The Radon*,* X-ray*, and* divergent beam transform *are defined as the linear maps*

$$\begin{aligned}&R : \mathcal{S}(\mathbb{R}^d)\to\mathcal{S}(\mathcal{C}), && Ru(\theta,s) = \int_{\{x\in\mathbb{R}^d \,|\, x\cdot\theta=s\}} u(x)\,\mathrm{d}\mathcal{H}^{d-1}(x),\\ &P : \mathcal{S}(\mathbb{R}^d)\to\mathcal{S}(\mathcal{C}'), && Pu(\theta,s) = \int_{-\infty}^{\infty} u(s+t\theta)\,\mathrm{d}t,\\ &D : \mathcal{S}(\mathbb{R}^d)\to C^\infty(\mathcal{C}''), && Du(\theta,s) = \int_0^{\infty} u(s+t\theta)\,\mathrm{d}t.\end{aligned}$$

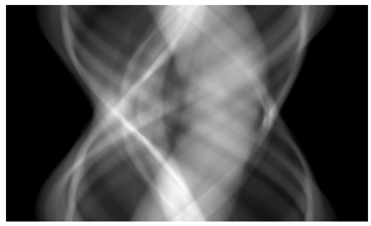
The divergent beam transform is also known as fanbeam transform *in two and as* conebeam transform *in three space dimensions. We write $R_\theta u = Ru(\theta,\cdot)$, $P_\theta u = Pu(\theta,\cdot)$, $D_s u = Du(\cdot,s)$.*



Radon transform X-ray transform divergent beam transform



original u



“sinogram” Ru ($\downarrow S^1$; $\to \mathbb{R}$)

Remark 186 (Extension to Radon measures). *The transforms can be extended to act on Radon measures $\nu \in \mathcal{M}(\mathbb{R}^d)$ via*

$$\begin{aligned} R_\theta \nu &= [x \mapsto x \cdot \theta]_\# \nu \in \mathcal{M}(\mathbb{R}), \\ P_\theta \nu &= [x \mapsto x - (x \cdot \theta)\theta]_\# \nu \in \mathcal{M}(\theta^\perp), \\ D_s \nu &= [x \mapsto \tfrac{x-s}{|x-s|}]_\# \nu \in \mathcal{M}(S^{d-1}). \end{aligned}$$

Remark 187 (Point symmetry). *The Radon and X-ray transform satisfy the point symmetry*

$$Ru(\theta, s) = Ru(-\theta, -s), \qquad Pu(\theta, s) = Pu(-\theta, s).$$

Remark 188 (Relation between the transforms in two and higher dimensions). *In $d = 2$ dimensions,*

$$Ru(\theta, s) = Pu(\theta', s\theta) \qquad \text{or equivalently} \qquad Pu(\theta, s) = Ru(\theta', \theta' \cdot s),$$

where $\theta' = (-\theta_2\ \theta_1)^T$ denotes the counterclockwise rotation by $\frac{\pi}{2}$. In higher dimensions one can express the Radon transform R_θ as an integral of the X-ray transform P_φ with any $\varphi \in \theta^\perp \cap S^{d-1}$ via

$$Ru(\theta, s) = \int_{\{x \in \varphi^\perp \mid x \cdot \theta = s\}} Pu(\varphi, x) \,\mathrm{d}\mathcal{H}^{d-2}(x).$$

Similarly, the X-ray transform P_θ can be reduced to a family of two-dimensional Radon transforms R_φ with any $\varphi \in \theta^\perp \cap S^{d-1}$ via

$$Pu(\theta, s) = R\tilde{u}((0\ 1)^T, 0) \qquad \text{with} \quad \tilde{u}(x) = u(s + (\theta|\varphi)x).$$

Finally, in any dimension,

$$Pu(\theta, s) = Du(\theta, s) + Du(-\theta, s).$$

Remark 189 (Forward operator in X-ray and emission tomography). *The X-ray transform is the forward operator of emission tomography: A (radioactive) mass at a point leads to photon emissions along all lines through that point; thus one measures total mass along every line in space.*
The divergent beam transform on a subset of $\mathcal{C}''$ (typically on $S^{d-1} \times C$ for a one-dimensional curve around the imaged object) is the forward operator of computed tomography: An X-ray point source is moved around the imaged object, and the arriving X-ray intensity is measured in a grid of detectors on the opposite side.
For sufficiently large distances between the X-ray source and the imaged object one can approximate the divergent beam transform by the X-ray transform.
In reality the situation is slightly more complicated: The photons in emission tomography may be absorbed (or even scattered) so that the X-ray transform actually has to be replaced by the so-called attenuated X-ray transform. *Likewise, the X-ray intensity in computed tomography actually is proportional to the negative exponential of the divergent beam transform so that first the logarithm of the measurements has to be taken. However, if the X-ray source is not monoenergetic and the imaged materials show different absorption behaviour for X-rays of different energies, one cannot remove the exponential nonlinearity from the forward operator.*

20 Inverse formulas for Radon and X-ray transform

Computing the inverse operator is usually based on a particular relation with the Fourier transform. To this end it is helpful to define the Fourier transform on a k-dimensional subspace M of $\mathbb{R}^d$: Given a complex-valued function or Radon measure u on M and an orthonormal basis $\theta_1,\ldots,\theta_k$ of M, we set

$$\mathcal{F}_M u : M \to \mathbb{C}, \qquad \mathcal{F}_M u(\xi_1\theta_1+\ldots+\xi_k\theta_k) = \mathcal{F}\tilde{u}(\xi_1,\ldots,\xi_k), \qquad \text{where } \tilde{u}(x_1,\ldots,x_k) = u(x_1\theta_1+\ldots+x_k\theta_k).$$

$\mathcal{F}_M$ is independent of the chosen orthonormal basis. We also write $\hat{u} = \mathcal{F}_M u$.

Theorem 190 (Projection-slice theorem)**.** *Let $M \subset \mathbb{R}^d$ be a k-dimensional subspace and denote by $\pi_M : \mathcal{M}(\mathbb{R}^d) \to \mathcal{M}(M)$ the projection onto and by $\sigma_M : C^0(\mathbb{R}^d) \to C^0(M)$ the restriction to M,*

$$\pi_M \nu = \mathrm{P}_{M\#}\nu, \qquad \sigma_M u = u|_M.$$

Then $\mathcal{F}_M \pi_M = (2\pi)^{(d-k)/2}\sigma_M \mathcal{F}$ on $\mathcal{M}(\mathbb{R}^d)$.

Proof. Let $\nu \in \mathcal{M}(\mathbb{R}^d)$, fix orthonormal basis $\theta_1,\ldots,\theta_k$ of M, set $\widetilde{\pi_M\nu}(s) = \pi_M\nu(s_1\theta_1+\ldots+s_k\theta_k)$.

$$\begin{aligned}
(2\pi)^{(d-k)/2}\sigma_M\mathcal{F}\nu(\xi_1\theta_1+\ldots+\xi_k\theta_k) &= (2\pi)^{-k/2}\int_{\mathbb{R}^d} e^{-i(\xi_1\theta_1+\ldots+\xi_k\theta_k)\cdot x}\,\mathrm{d}\nu(x)\\
&= (2\pi)^{-k/2}\int_{\mathbb{R}^k} e^{-i\xi\cdot s}\,\mathrm{d}[x\mapsto(\theta_1\cdot x,\ldots,\theta_k\cdot x)]_{\#}\nu(s)\\
&= (2\pi)^{-k/2}\int_{\mathbb{R}^k} e^{-i\xi\cdot s}\,\mathrm{d}\widetilde{\pi_M\nu}(s)\\
&= \mathcal{F}_M\pi_M\nu(\xi_1\theta_1+\ldots+\xi_k\theta_k).
\end{aligned}$$

□

Corollary 191 (Fourier slice theorem)**.** *Let $u \in \mathcal{S}(\mathbb{R}^d)$, $\theta \in S^{d-1}$, then*

$$\begin{aligned}
\widehat{Ru}(\theta,\xi) &= (2\pi)^{(d-1)/2}\hat{u}(\xi\theta) && \forall\xi\in\mathbb{R},\\
\widehat{Pu}(\theta,\xi) &= (2\pi)^{1/2}\hat{u}(\xi) && \forall\xi\in\theta^\perp,
\end{aligned}$$

where the Fourier transform on the left-hand side is with respect to the second argument.

Proof. $Ru(\theta,\cdot) = \pi_M u$ for $M = \mathrm{span}\{\theta\}$ (identifying on the left-hand side $\mathbb{R}$ with M);
$Pu(\theta,\cdot) = \pi_M u$ for $M = \theta^\perp$. □

Remark 192 (Fourier slice theorem for divergent beam transform)**.** *A Fourier slice theorem for the divergent beam transform does not exist in a simple form. People seem to have generalized it to this setting, though (see Zhao, Halling: A new Fourier method for fan beam reconstruction, 1995).*

Corollary 193 (Transforms and convolution/differentiation)**.** *Let $u, v \in \mathcal{S}(\mathbb{R}^d)$.*

1. *$R_\theta(u*v) = R_\theta u * R_\theta v$ and $P_\theta(u*v) = P_\theta u * P_\theta v$*
2. *$R_\theta D^\alpha u = \theta^\alpha D^{|\alpha|}R_\theta u$ and $P_\theta D^\alpha u = D^\alpha((P_\theta u)\circ \mathrm{P}_{\theta^\perp})$*

Proof. Homework □

In addition, formulas for the inverse transforms involve the backprojection.

Definition 194 (Backprojection)**.** *The* backprojections *of the Radon, X-ray, and divergent beam transform are defined as*

$$\begin{aligned}
R^\# &: \mathcal{S}(\mathcal{C}) \to C_0^\infty(\mathbb{R}^d), & R^\# v(x) &= \int_{S^{d-1}} v(\theta, x\cdot\theta)\,\mathrm{d}\mathcal{H}^{d-1}(\theta),\\
P^\# &: \mathcal{S}(\mathcal{C}') \to C_0^\infty(\mathbb{R}^d), & P^\# v(x) &= \int_{S^{d-1}} v(\theta, \mathrm{P}_{\theta^\perp}x)\,\mathrm{d}\mathcal{H}^{d-1}(\theta),\\
D^\# &: \mathcal{S}(\mathcal{C}'') \to C_0^\infty(\mathbb{R}^d), & D^\# v(x) &= \int_{S^{d-1}}\int_0^\infty v(\theta, x-t\theta)\,\mathrm{d}t\,\mathrm{d}\mathcal{H}^{d-1}(\theta).
\end{aligned}$$

Theorem 195 (Backprojection)**.** *The backprojections are well-defined, i. e. indeed map into $C_0^\infty(\mathbb{R}^d)$.*

Proof.
- differentiation D^α can be pulled into the integral
  $\Rightarrow$ yields integral of same type of a Schwartz function (e. g. of $\theta^\alpha\partial_s^{|\alpha|}v(\theta, x\cdot\theta)$ in case of $R^\#$)
  $\Rightarrow$ suffices to show that such integrals decay to zero
- let $v \in \mathcal{S}(\mathcal{C})$, then $\lim_{|x|\to\infty} R^\# v(x) = 0$: Let $\varepsilon > 0$ arbitrary.
  - for $x \in \mathbb{R}^d$, $n > 0$ set $S(x,n) = \{\theta \in S^{d-1} \,|\, |x\cdot\theta| < n\} = \{\theta \in S^{d-1} \,|\, |\frac{x}{|x|}\cdot\theta| < \frac{n}{|x|}\}$
  - pick $n > 0$ such that $|v(\theta, s)| < \varepsilon$ for $|s| > n$
  - pick $m > 0$ such that $\mathcal{H}^{d-1}(S(x,n)) < \varepsilon$ for $|x| > m$
  - if $|x| > m$,

$$|R^\# v(x)| \le \left|\int_{S(x,n)} v(\theta, x\cdot\theta)\,\mathrm{d}\mathcal{H}^{d-1}(\theta)\right| + \left|\int_{S^{d-1}\setminus S(x,n)} v(\theta, x\cdot\theta)\,\mathrm{d}\mathcal{H}^{d-1}(\theta)\right|$$
$$\le \mathcal{H}^{d-1}(S(x,n)) \sup_{(\theta,s)\in\mathcal{C}} |v(\theta,s)| + \mathcal{H}^{d-1}(S^{d-1})\varepsilon \le \varepsilon(\sup_{(\theta,s)\in\mathcal{C}} |v(\theta,s)| + \mathcal{H}^{d-1}(S^{d-1}))$$

- analogous for other backprojections □

The integrals of the backprojection in fact also make sense for less regular functions v than Schwartz functions. The backprojection applied to $v = Au$ with A being R, P, or D moves all measurements to where they potentially stem from (hence the name; its result is sometimes called a *layergram*). Using microlocal analysis one can show that this way on can identify the singularities of the imaged object, however, the singularities will be of a slightly different type so that to the human eye the backprojection looks very different from the imaged object.

Remark 196 (Backprojection does not map into Schwartz space)**.** *Note that the backprojection does not map into Schwartz space. For instance, let $v(\theta, s) = e^{-s^2/2}$, then*

$$R^\# v(x) = \int_{S^{d-1}} e^{-(x\cdot\theta)^2/2}\,\mathrm{d}\mathcal{H}^{d-1}(\theta) > \int_{S(x,n)} e^{-n^2/2}\,\mathrm{d}\mathcal{H}^{d-1} = e^{-n^2/2}\mathcal{H}^{d-1}(S(x,n)) \gtrsim e^{-n^2/2}(\tfrac{n}{|x|})^{d-1}.$$

Theorem 197 (Adjoint transform)**.** *The backprojection $A^\#$ with A being R, P, or D is the adjoint A^* restricted to Schwartz space, i. e. for all $u \in \mathcal{S}(\mathbb{R}^d)$, $v \in \mathcal{S}(C)$ with C being $\mathcal{C}$, $\mathcal{C}'$, or $\mathcal{C}''$ we have*

$$\langle Ru, v\rangle_{\mathcal{S},\mathcal{S}'} = \langle u, R^\# v\rangle_{\mathcal{S},\mathcal{S}'} \qquad \langle Pu, v\rangle_{\mathcal{S},\mathcal{S}'} = \langle u, P^\# v\rangle_{\mathcal{S},\mathcal{S}'} \qquad \langle Du, v\rangle_{\mathcal{S},\mathcal{S}'} = \langle u, D^\# v\rangle_{\mathcal{S},\mathcal{S}'}.$$

Proof. Homework □

Remark 198 (Generalization of transforms)**.** *The Radon, X-ray, and divergent beam transform can be extended to the space $\mathcal{E}'(\mathbb{R}^d)$ of distributions of compact support (the dual space to $\mathcal{E}(\mathbb{R}^d) = C^\infty(\mathbb{R}^d)$, where $\mathcal{E}'(\mathbb{R}^d) \subset \mathcal{S}'(\mathbb{R}^d) \subset \mathcal{D}'(\mathbb{R}^d)$) by simply defining them as the adjoint of their backprojection. For compactly supported Radon measures, for instance, this will yield the same extension as in remark 186. Note that non-compactly supported Radon measures do not lie in $\mathcal{E}'(\mathbb{R}^d)$, but only in the dual space to $C_0^\infty(\mathbb{R}^d)$. Hence an extension to all Radon measures needs to exploit that the backprojection of a Schwartz function decays to zero at infinity. For this it is essential to have measurements along all angles $\theta \in S^{d-1}$ (or at least angles from a relatively open subset): If we for instance only measure $v = R_\theta u$ for a single angle θ, then the corresponding backprojection would be $R_\theta^\# v(x) = v(x\cdot\theta)$, which is constant on $\theta^\perp$. Similarly, the backprojections can be extended to tempered distributions (distributions of compact support in case of $D^\#$) by interpreting them as the adjoint of R, P, and D, respectively.*

Theorem 199 (Convolution theorem for projection transforms)**.** *For $u \in \mathcal{S}(\mathbb{R}^d)$ and $v \in \mathcal{S}(C)$ with $C = \mathcal{C}$ or $C = \mathcal{C}'$, respectively, we have*

$$(R^\# v) * u = R^\#(v * Ru),$$
$$(P^\# v) * u = P^\#(v * Pu),$$

where the convolution on the right-hand side is with respect to the second argument.

Proof. Only for R (analogous for P).

$$\begin{aligned}
R^\# v * u(x) &= \int_{\mathbb{R}^d} R^\# v(x-y)u(y)\,\mathrm{d}y \\
&= \int_{\mathbb{R}^d}\int_{S^{d-1}} v(\theta,(x-y)\cdot\theta)\,\mathrm{d}\mathcal{H}^{d-1}(\theta)\,u(y)\,\mathrm{d}y \\
&= \int_{S^{d-1}}\int_{\mathbb{R}^d} v(\theta,(x-y)\cdot\theta)\,u(y)\,\mathrm{d}y\,\mathrm{d}\mathcal{H}^{d-1}(\theta) \\
&\overset{y=s\theta+z}{=} \int_{S^{d-1}}\int_{\mathbb{R}}\int_{\theta^\perp} v(\theta,x\cdot\theta-s)\,u(s\theta+z)\,\mathrm{d}\mathcal{H}^{d-1}(z)\,\mathrm{d}s\,\mathrm{d}\mathcal{H}^{d-1}(\theta) \\
&= \int_{S^{d-1}}\int_{\mathbb{R}} v(\theta,x\cdot\theta-s)\,Ru(\theta,s)\,\mathrm{d}s\,\mathrm{d}\mathcal{H}^{d-1}(\theta) \\
&= \int_{S^{d-1}} (v*Ru)(\theta,x\cdot\theta)\,\mathrm{d}\mathcal{H}^{d-1}(\theta) \\
&= R^\#(v*Ru)(x)
\end{aligned}$$
□

The final ingredient for the inverse operator is the Riesz potential.

Definition 200 (Riesz potential)**.** *For $\alpha < d$ the linear operator $I^\alpha : \mathcal{S}(\mathbb{R}^d) \to L^\infty(\mathbb{R}^d)$ is defined by*

$$\widehat{I^\alpha u}(\xi) = |\xi|^{-\alpha}\hat{u}(\xi).$$

$I^\alpha u$ is called the Riesz potential *of u. If applied to functions on $\mathcal{C}$ or $\mathcal{C}'$, it shall act on the second variable. I^α is injective; its inverse on its range is denoted $I^{-\alpha}$, since for $\alpha > -d$, $(I^\alpha)^{-1}|_{\mathcal{S}(\mathbb{R}^d)}$ obviously coincides with the Riesz potential of exponent $-\alpha$.*

The Riesz potential lies in $L^\infty(\mathbb{R}^d)$, because $\widehat{I^\alpha u} \in L^1(\mathbb{R}^d)$ for $u \in \mathcal{S}(\mathbb{R}^d)$. It can be thought of as the inversion of the fractional Laplacian $(-\Delta)^{\alpha/2}$, which in Fourier space becomes multiplication with $|\xi|^\alpha$. For $\alpha > 0$ it is thus a smoothing operator, and for α nonpositive and even it maps into Schwartz space.

Lemma 201 (Integral formula)**.** *For $f \in L^1(\mathbb{R}^d)$ we have*

$$\int_{\mathbb{R}^d} f(x)\,\mathrm{d}x = \frac{1}{\mathcal{H}^{d-2}(S^{d-2})}\int_{S^{d-1}}\int_{\theta^\perp} |s|f(s)\,\mathrm{d}\mathcal{H}^{d-1}(s)\,\mathrm{d}\mathcal{H}^{d-1}(\theta).$$

Proof.

- Let $S = \{(\theta,\varphi) \in S^{d-1}\times S^{d-1} \,|\, \theta \perp \varphi\}$ and $\mathrm{proj}_2 : S \to S^{d-1}$, $\mathrm{proj}_2(\theta,\varphi) = \varphi$, then

$$\mathrm{proj}_{2\#}(\mathcal{H}^{d-1}\otimes\mathcal{H}^{d-2})\llcorner S = \mathrm{proj}_{2\#}(\mathcal{H}^{d-2}\otimes\mathcal{H}^{d-1})\llcorner S = \mathcal{H}^{d-2}(S^{d-2})\mathcal{H}^{d-1}\llcorner S^{d-1}.$$

- In each subspace $\theta^\perp$ use polar coordinates (r,φ):

$$\begin{aligned}
\int_{S^{d-1}}\int_{\theta^\perp} |s|f(s)\,\mathrm{d}\mathcal{H}^{d-1}(s)\,\mathrm{d}\mathcal{H}^{d-1}(\theta) &= \int_{S^{d-1}}\int_{S^{d-1}\cap\theta^\perp}\int_0^\infty r^{d-1}f(r\mathrm{proj}_2(\theta,\varphi))\,\mathrm{d}r\,\mathrm{d}\mathcal{H}^{d-2}(\varphi)\,\mathrm{d}\mathcal{H}^{d-1}(\theta) \\
&= \mathcal{H}^{d-2}(S^{d-2})\int_{S^{d-1}}\int_0^\infty r^{d-1}f(r\varphi)\,\mathrm{d}r\,\mathrm{d}\mathcal{H}^{d-1}(\varphi) = \mathcal{H}^{d-2}(S^{d-2})\int_{\mathbb{R}^d} f(x)\,\mathrm{d}x.
\end{aligned}$$
□

Theorem 202 (Riesz inversion formula)**.** *Let $u \in \mathcal{S}(\mathbb{R}^d)$, then for any $\alpha < d$ we have*

$$\begin{aligned}
u &= \tfrac{1}{2}(2\pi)^{1-d}I^{-\alpha}R^\# I^{\alpha-d+1}(Ru), \\
u &= \tfrac{1}{\mathcal{H}^{d-2}(S^{d-2})}(2\pi)^{-1}I^{-\alpha}P^\# I^{\alpha-1}(Pu).
\end{aligned}$$

Proof.

$$
\begin{aligned}
I^\alpha u(x) &= (2\pi)^{-d/2}\int_{\mathbb{R}^d} e^{ix\cdot\xi}|\xi|^{-\alpha}\hat u(\xi)\,\mathrm{d}\xi\\
&= (2\pi)^{-d/2}\int_{S^{d-1}}\int_0^\infty e^{isx\cdot\theta}s^{d-1-\alpha}\hat u(s\theta)\,\mathrm{d}s\,\mathrm{d}\mathcal{H}^{d-1}(\theta)\\
&= (2\pi)^{\frac12-d}\int_{S^{d-1}}\int_0^\infty e^{isx\cdot\theta}s^{d-1-\alpha}\widehat{Ru}(\theta,s)\,\mathrm{d}s\,\mathrm{d}\mathcal{H}^{d-1}(\theta)\\
&= (2\pi)^{\frac12-d}\frac12\int_{S^{d-1}}\int_{\mathbb{R}} e^{isx\cdot\theta}|s|^{d-1-\alpha}\widehat{Ru}(\theta,s)\,\mathrm{d}s\,\mathrm{d}\mathcal{H}^{d-1}(\theta)\\
&= (2\pi)^{1-d}\frac12\int_{S^{d-1}} I^{\alpha+1-d}Ru(\theta,x\cdot\theta)\,\mathrm{d}\mathcal{H}^{d-1}(\theta)\\
&= \tfrac12(2\pi)^{1-d}R^\# I^{\alpha+1-d}(Ru)(x)
\end{aligned}
$$

$$
\begin{aligned}
I^\alpha u(x) &= (2\pi)^{-d/2}\int_{\mathbb{R}^d} e^{ix\cdot\xi}|\xi|^{-\alpha}\hat u(\xi)\,\mathrm{d}\xi\\
&= (2\pi)^{-d/2}\frac{1}{\mathcal{H}^{d-2}(S^{d-2})}\int_{S^{d-1}}\int_{\theta^\perp} e^{ix\cdot s}|s|^{1-\alpha}\hat u(s)\,\mathrm{d}\mathcal{H}^{d-1}(s)\,\mathrm{d}\mathcal{H}^{d-1}(\theta)\\
&= (2\pi)^{-(d+1)/2}\frac{1}{\mathcal{H}^{d-2}(S^{d-2})}\int_{S^{d-1}}\int_{\theta^\perp} e^{i\mathrm{P}_{\theta^\perp}x\cdot s}|s|^{1-\alpha}\widehat{Pu}(\theta,s)\,\mathrm{d}\mathcal{H}^{d-1}(s)\,\mathrm{d}\mathcal{H}^{d-1}(\theta)\\
&= (2\pi)^{-1}\frac{1}{\mathcal{H}^{d-2}(S^{d-2})}\int_{S^{d-1}} I^{\alpha-1}Pu(\theta,\mathrm{P}_{\theta^\perp}x)\,\mathrm{d}\mathcal{H}^{d-1}(\theta)\\
&= \tfrac{1}{\mathcal{H}^{d-2}(S^{d-2})}(2\pi)^{-1}P^\# I^{\alpha-1}(Pu)(x)
\end{aligned}
$$

□

Remark 203 (Riesz inversion formula for Radon transform)**.** *For the Radon transform, the case $\alpha = d-1$ is known as ρ-*filtered layergram *(one takes the layergram and applies a filter which in Fourier space is ρ^{1-n} for ρ the radial variable), the case $\alpha = 0$ as* filtered backprojection *(one first filters or sharpens the measurement via I^{1-d} before applying the backprojection).*

Remark 204 (Locality of Radon transform)**.** *For odd dimension d (in particular $d = 3$) the ρ-filtered layergram and the filtered backprojection obviously read*

$$
\begin{aligned}
u(x) &= \tfrac12(2\pi)^{1-d}(-\Delta_x)^{\frac{d-1}{2}}\int_{S^{d-1}} Ru(\theta,x\cdot\theta)\,\mathrm{d}\mathcal{H}^{d-1}(\theta),\\
u(x) &= \tfrac12(2\pi)^{1-d}\int_{S^{d-1}}(-1)^{\frac{d-1}{2}}\partial_2^{d-1}Ru(\theta,x\cdot\theta)\,\mathrm{d}\mathcal{H}^{d-1}(\theta)
\end{aligned}
$$

with Δ_x the Laplace operator in the x-variable. As these formulas tell, the inversion of the Radon transform is local in the sense that to reconstruct u at a point x from Ru, one only requires the values of Ru belonging to hyperplanes arbitrarily close to x.
This is not true for the Radon transform in even dimensions (or for the X-ray transform, which, as we know, is related to families of 2D Radon transforms). In particular there exist Schwartz functions u that are nonzero on the unit ball, but satisfy $Ru = 0$ on $S^{d-1}\times[-1,1]$.

Corollary 205 (Representation of $A^\# A$)**.** *Let $u \in \mathcal{S}(\mathbb{R}^d)$, then*

$$
\begin{aligned}
R^\# Ru &= \mathcal{H}^{d-2}(S^{d-2})g * u && \text{for } g(x) = |x|^{-1},\\
P^\# Pu &= 2h * u && \text{for } h(x) = |x|^{1-d}.
\end{aligned}
$$

Proof.

$$
\begin{aligned}
\hat{g}(\xi) &= \tfrac{(2)^{d/2-1}\Gamma((d-1)/2)}{\Gamma(1/2)}|\xi|^{1-d}\\
\hat{h}(\xi) &= \tfrac{2^{1-d/2}\Gamma(1/2)}{\Gamma((d-1)/2)}|\xi|^{-1}\\
\mathcal{H}^{d-2}(S^{d-2}) &= \tfrac{2\pi^{(d-1)/2}}{\Gamma((d-1)/2)}\\
\Gamma(1/2) &= \sqrt{\pi}\\
\widehat{R^{\#}Ru}(\xi) &= 2(2\pi)^{d-1}\widehat{I^{d-1}u}(\xi) = 2(2\pi)^{d-1}|\xi|^{1-d}\hat{u}(\xi)\\
&= \tfrac{\pi^{d-1}2^{d/2+1}\Gamma(1/2)}{\Gamma((d-1)/2)}\hat{g}(\xi)\hat{u}(\xi) = (2\pi)^{d/2}\mathcal{H}^{d-2}(S^{d-2})\hat{g}(\xi)\hat{u}(\xi) = \mathcal{H}^{d-2}(S^{d-2})\widehat{g*u}(\xi)\\
\widehat{P^{\#}Pu}(\xi) &= 2\pi\mathcal{H}^{d-2}(S^{d-2})\widehat{I^{1}u}(\xi) = \tfrac{4\pi^{(d+1)/2}}{\Gamma((d-1)/2)}|\xi|^{-1}\hat{u}(\xi) = 2(2\pi)^{d/2}\hat{h}(\xi)\hat{u}(\xi) = 2\widehat{h*u}(\xi)
\end{aligned}
$$
□

Remark 206 (Extension onto L^2)**.** *Similarly to the comment on Radon measures in remark 198 one can extend $A \equiv R$ or $A \equiv P$ to a bounded linear operator $A : L^2(\mathbb{R}^d) \to \mathcal{S}'(C)$ with C being $\mathcal{C}$ or $\mathcal{C}'$ (simply by setting ${}^*A = A^{\#}$). From the above formulas we see that this extension cannot be continuous into $L^2(C)$: This would imply ${}^*A = A^H = A^{\#} : L^2(C) \to L^2(\mathbb{R}^d)$ to be bounded so that also $A^{\#}A : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is bounded, which is false. Exemplarily, consider $d = 2$, in which case $R^{\#}R = P^{\#}P$. Let g_1 be the restriction of g to the unit ball and g_2 to the complement, then $g_1 \in L^1(\mathbb{R}^2)$ and $g_2 \in L^p(\mathbb{R}^2) \setminus L^2(\mathbb{R}^2)$ for any $p > 2$. By Young's convolution theorem there exists a $u \in L^2(\mathbb{R}^2)$ with $g_2 * u \notin L^2(\mathbb{R}^2)$ (while $g_1 * u \in L^2(\mathbb{R}^2)$), thus $g * u = g_1 * u + g_2 * u \notin L^2(\mathbb{R}^2)$.*

Remark 207 (Ill-posedness)**.** *Since we defined them on an unbounded domain, the Radon or X-ray transform are not compact on shift-invariant function spaces (cf. theorem 42). Still their inversion is ill-posed, e. g. with respect to the L^2-metric on domain and codomain. For instance, in $d = 2$ dimensions we have $\widehat{P^{\#}Pu}(\xi) = \widehat{R^{\#}Ru}(\xi) = const.\hat{u}(\xi)/|\xi|$ so that a small perturbation of $P^{\#}Pu = R^{\#}Ru$ leads to an arbitrarily large change of u, if the perturbation happened at low frequencies (recall that the Fourier transform is an isometry on $L^2(\mathbb{R}^d)$).*

21 The range of Radon and X-ray transform

Theorem 208 (Helgason–Ludwig consistency/moment conditions)**.** *If $u \in \mathcal{S}(\mathbb{R}^d)$, then for any $m \in \mathbb{N}_0$ there exist polynomials p_m, q_m homogeneous of degree m with*

$$
\int_{\mathbb{R}} s^m R_\theta u(s)\,\mathrm{d}s = p_m(\theta), \qquad \int_{\theta^\perp} (x\cdot y)^m P_\theta u(x)\,\mathrm{d}\mathcal{H}^{d-1}(x) = q_m(y)\ \forall\theta\in S^{d-1}, y \in \theta^\perp.
$$

Proof. $\int_{\mathbb{R}} s^m R_\theta u(s)\,\mathrm{d}s = \int_{\mathbb{R}} s^m \int_{\theta^\perp} u(s\theta+y)\,\mathrm{d}\mathcal{H}^{d-1}(y)\,\mathrm{d}s = \int_{\mathbb{R}^d}(x\cdot\theta)^m u(x)\,\mathrm{d}x$
is homogeneous polynomial in θ
$\int_{\theta^\perp}(x\cdot y)^m P_\theta u(x)\,\mathrm{d}\mathcal{H}^{d-1}(x) = \int_{\theta^\perp}(x\cdot y)^m \int_{\mathbb{R}} u(x+t\theta)\,\mathrm{d}t\,\mathrm{d}\mathcal{H}^{d-1}(x) = \int_{\mathbb{R}^d}(z\cdot y)^m u(z)\,\mathrm{d}z$
is homogeneous polynomial in y, independent of θ □

Note that all p_m might be zero even if $u \neq 0$ (e. g. $R_\theta u(s) = f(\theta)h(s)$ for f, h from remark 212 later).

Theorem 209 (Range of Radon transform)**.** *Let $v \in \mathcal{S}(\mathcal{C})$ with $v(\theta, s) = v(-\theta, -s)$ and the Helgason–Ludwig condition*

$$
\int_{\mathbb{R}} s^m v(\theta, s)\,\mathrm{d}s = p_m(\theta)
$$

for m-homogeneous polynomials p_m, $m \in \mathbb{N}_0$. Then there exists $u \in \mathcal{S}(\mathbb{R}^d)$ with $v = Ru$.

Proof. Define u via $\hat{u}(\xi) = (2\pi)^{(1-d)/2}\hat{v}(\frac{\xi}{|\xi|}, |\xi|)$.
Suffices to show $\hat{u} \in \mathcal{S}(\mathbb{R}^d)$, then $Ru = v$ by Fourier slice theorem.
$\hat{u}$ has derivatives up to an arbitrary order q (already clear for $\xi \neq 0$):

- $e^{it} = \sum_{m=0}^{q} \frac{(it)^m}{m!} + e_q(t)$ with $e_q(t) = \sum_{m=q+1}^{\infty} \frac{(it)^m}{m!}$

- $\hat{u}(\xi) = (2\pi)^{-d/2} \int_{\mathbb{R}} e^{-is|\xi|} v(\tfrac{\xi}{|\xi|}, s)\,\mathrm{d}s$
$$= (2\pi)^{-d/2} \left(\sum_{m=0}^{q} \frac{(-i|\xi|)^m}{m!} \int_{\mathbb{R}} s^m v(\tfrac{\xi}{|\xi|}, s)\,\mathrm{d}s + \int_{\mathbb{R}} e_q(-|\xi|s) v(\tfrac{\xi}{|\xi|}, s)\,\mathrm{d}s \right)$$
$$= (2\pi)^{-d/2} \left(\sum_{m=0}^{q} \frac{(-i)^m p_m(\xi)}{m!} + \int_{\mathbb{R}} e_q(-|\xi|s) v(\tfrac{\xi}{|\xi|}, s)\,\mathrm{d}s \right)$$

- $D^\alpha_\xi \Big(e_q(-|\xi|s) v(\tfrac{\xi}{|\xi|}, s)\Big)$ is Schwartz in s and continuous & unif. bdd. in $\xi \in B_1(0)\setminus\{0\}$ for any $|\alpha| \le q+1$:
  - sufficient to show: $D^\alpha_\xi \Big(e_q(-|\xi|s) v(\tfrac{\xi}{|\xi|}, s)\Big)$ is finite linear combination of terms
  $$|\xi|^{q+1-|\alpha|} a(|\xi|s) h(\tfrac{\xi}{|\xi|}, s)$$
  with $h \in \mathcal{S}(\mathcal{C})$ and $a \in C^\infty(\mathbb{R})$ s. t. $a^{(n)}$ grows at most polynomially for any $n \in \mathbb{N}_0$
  - induction basis ($|\alpha| = 0$): take $h(\theta, s) = s^{q+1} v(\theta, s)$ and $a(t) = e_q(-t)/t^{q+1}$
  - induction step: for $|\alpha| \le q$ assume claim holds; differentiate one of the terms,
  $$\nabla_\xi (|\xi|^{q+1-|\alpha|} a(|\xi|s) h(\tfrac{\xi}{|\xi|}, s)) = |\xi|^{q-|\alpha|} a(|\xi|s) \left[\tfrac{\xi}{|\xi|} h(\tfrac{\xi}{|\xi|}, s)\right] + |\xi|^{q-|\alpha|} [(|\xi|s) a'(|\xi|s)] \left[\tfrac{\xi}{|\xi|} h(\tfrac{\xi}{|\xi|}, s)\right] + |\xi|^{q-|\alpha|} a(|\xi|s) \left[\left(I - \tfrac{\xi}{|\xi|} \otimes \tfrac{\xi}{|\xi|}\right) \nabla_\theta h(\tfrac{\xi}{|\xi|}, s)\right]$$
  $\Rightarrow$ also next higher derivatives are linear combinations of terms with required properties
- For any $|\alpha| = q$, $D^\alpha_\xi \hat{u}(\xi)$ can be continuously extended to $\xi = 0$:
  - $\nabla_\xi D^\alpha_\xi \hat{u}$ is bounded and continuous on $B_1(0) \setminus \{0\}$ by previous point

$D^\alpha_\xi \hat{u}(\xi)$ decays faster than any power of $|\xi|$:

- $\sup_{|\xi|>1} |\xi^\beta D^\alpha_\xi \hat{u}(\xi)| = (2\pi)^{(1-d)/2} \sup_{|\xi|>1} |\xi^\beta D^\alpha_\xi \hat{v}(\tfrac{\xi}{|\xi|}, |\xi|)| < \infty$ since v is Schwartz ☐

Remark 210 (Helgason–Ludwig conditions)**.** *If one considers non-Schwartz functions $v : \mathcal{C} \to \mathbb{R}$, it is known that the Helgason–Ludwig conditions on v determine the decay of $R^{-1}v$ (see Madych & Solmon: A range theorem for the Radon transform, 1988). Roughly, if the conditions hold for $m = 0, \ldots, k$, then (under additional smoothness conditions) $v = Ru$ for some function u which decays at least like $|x|^{-d-k-1}$. This is quite natural: The necessity of the Helgason–Ludwig conditions followed from the identity $\int_{\mathbb{R}} s^m R_\theta u(s)\,\mathrm{d}s = \int_{\mathbb{R}^d} (x \cdot \theta)^m u(x)\,\mathrm{d}x$, whose right-hand side is only well-defined if u decays fast enough (faster than $|x|^{-d-k}$).*

Theorem 211 (Range of X-ray transform)**.** *Let $w \in \mathcal{S}(\mathcal{C}')$ with $w(\theta, s) = 0$ for $|s| \ge R$ and the Helgason–Ludwig condition*
$$\int_{\theta^\perp} (x \cdot y)^m w(\theta, x)\,\mathrm{d}\mathcal{H}^{d-1}(x) = q_m(y)\ \forall \theta \in S^{d-1}, y \in \theta^\perp$$
for m-homogeneous polynomials q_m, $m \in \mathbb{N}_0$. Then there exists $u \in \mathcal{S}(\mathbb{R}^d)$ with $w = Pu$.

Proof. With $\varphi \in S^{d-1} \cap \theta^\perp$ we can compute the Radon transform v from the X-ray transform w via
$$v(\theta, s) = \int_{\{x \in \varphi^\perp \,|\, x \cdot \theta = s\}} w(\varphi, x)\,\mathrm{d}\mathcal{H}^{d-2}(x).$$

- $v = Ru$ for some $u \in \mathcal{S}(\mathbb{R}^d)$ with support in the R-ball:
  - $\int_{\mathbb{R}} s^m v(\theta, s)\,\mathrm{d}s = \int_{\mathbb{R}} s^m \int_{\{x \in \varphi^\perp \,|\, x \cdot \theta = s\}} w(\varphi, x)\,\mathrm{d}\mathcal{H}^{d-2}(x)\,\mathrm{d}s = \int_{\varphi^\perp} (x \cdot \theta)^m w(\varphi, x)\,\mathrm{d}x = q_m(\theta)$

- $v(\theta,s)=0$ for $|s|\geq R \overset{\text{theorem 224 later}}{\Rightarrow} u$ has support in R-ball

- v does not depend on φ:
  - polynomials are dense on $L^2((-R,R))$
  - $\Rightarrow v(\theta,\cdot)$ uniquely specified by $q_0,q_1,\dots$
- $w=Pu$, since integrals of w and Pu over arbitrary hyperplanes in $\theta^\perp$ coincide:
  - pick hyperplane $H=\{x\in\theta^\perp\,|\,x\cdot\varphi=s\}$ for arbitrary $s\in\mathbb{R}$, $\varphi\in S^{d-1}\cap\theta^\perp$, then

$$\begin{aligned}\int_H w(\theta,x)\,\mathrm{d}\mathcal{H}^{d-2}(x)&=v(\varphi,s)=Ru(\varphi,s)\\ \int_H Pu(\theta,x)\,\mathrm{d}\mathcal{H}^{d-2}(x)&=\int_H\int_{\mathbb{R}}u(x+t\theta)\,\mathrm{d}t\,\mathrm{d}\mathcal{H}^{d-2}(x)\\ &=\int_{\theta^\perp\cap\varphi^\perp}\int_{\mathbb{R}}u(s\varphi+y+t\theta)\,\mathrm{d}t\,\mathrm{d}\mathcal{H}^{d-2}(y)\\ &=\int_{\varphi^\perp}u(s\varphi+x)\,\mathrm{d}\mathcal{H}^{d-1}(x)\\ &=Ru(\varphi,s)\end{aligned}$$

  - Thus Radon transforms of w and Ru within $\theta^\perp$ coincide, but Radon transform is injective. □

Remark 212 (Noncompact support). *If $d>2$, the support condition $w(\theta,s)=0$ for $|s|\geq R$ cannot be dropped (for $d=2$ it can since R and P and their moment conditions are equivalent). Indeed, there exists a nonzero even $h\in\mathcal{S}(\mathbb{R})$ with*

$$\int_0^\infty s^m h(s)\,\mathrm{d}s=0\qquad\text{for all } m\in\mathbb{N}_0.$$

If $w(\theta,s)=f(\theta)h(|s|)$ for some $f\in C^\infty(S^{d-1})$ with $f(-\theta)=f(\theta)$, the Helgason–Ludwig condition is satisfied due to

$$\begin{aligned}\int_{\theta^\perp}(s\cdot y)^m w(\theta,s)\,\mathrm{d}\mathcal{H}^{d-1}(s)&=f(\theta)\int_{\theta^\perp}(s\cdot y)^m h(|s|)\,\mathrm{d}\mathcal{H}^{d-1}(s)\\ &=f(\theta)\int_{S^{d-1}\cap\theta^\perp}(\varphi\cdot y)^m\,\mathrm{d}\mathcal{H}^{d-2}(\varphi)\int_0^\infty r^{d-2+m}h(r)\,\mathrm{d}r=0.\end{aligned}$$

If $w=Pu$ for some $u\in\mathcal{S}(\mathbb{R}^d)$, then for $\varphi\in S^{d-1}\cap\theta^\perp$

$$Ru(\varphi,s)=\int_{\{x\in\theta^\perp\,|\,x\cdot\varphi=s\}}w(\theta,x)\,\mathrm{d}\mathcal{H}^{d-2}(x)=f(\theta)\underbrace{\int_{\{x\in\theta^\perp\,|\,x\cdot\varphi=s\}}h(|x|)\,\mathrm{d}\mathcal{H}^{d-2}(x)}_{\text{function solely of } s\text{, since integrand only depends on } |x|},$$

a contradiction unless $f=$const.

Remark 213 (Continuous inverse on Schwartz space). *The range* $\operatorname{ran}R\subset\mathcal{S}(\mathcal{C})$ *of $R:\mathcal{S}(\mathbb{R}^d)\to\mathcal{S}(\mathcal{C})$ is closed: If $v_n\to v\in\mathcal{S}(\mathcal{C})$ with v_n satisfying point symmetry and the Helgason–Ludwig conditions with polynomials p_m^n, then also v satisfies point symmetry and the Helgason–Ludwig conditions with p_m being the uniform limit of the p_m^n (which again must be an m-homogeneous polynomial). Thus by the open mapping theorem (which also holds on Fréchet spaces), R is an open map from $\mathcal{S}(\mathbb{R}^d)$ onto its range. Since by the Fourier slice theorem R is also injective on $\mathcal{S}(\mathbb{R}^d)$, it follows that R has a continuous inverse $R^{-1}:\operatorname{ran}R\to\mathcal{S}(\mathbb{R}^d)$ – one does not see ill-posedness on the level of Schwartz functions!*
Similarly, the range of $P:\mathcal{S}(\Omega)\to\mathcal{S}(\mathcal{C}')$ is closed (where $\mathcal{S}(\Omega)$ are the Schwartz functions with support in the compact $\Omega\subset\mathbb{R}^d$) and P has a continuous inverse.

22 Fractional Hilbert spaces

Definition 214 (Fractional Hilbert space). *For $\gamma \in \mathbb{R}$ the* (fractional) Hilbert space $H^\gamma(\mathbb{R}^d)$ *is defined as*

$$H^\gamma(\mathbb{R}^d) = \{u \in \mathcal{S}'(\mathbb{R}^d) \,|\, \|u\|_{H^\gamma(\mathbb{R}^d)} < \infty\} \qquad \text{with norm } \|u\|_{H^\gamma(\mathbb{R}^d)} = \left(\int_{\mathbb{R}^d} |\hat{u}(\xi)|^2 (1+|\xi|^2)^\gamma \,\mathrm{d}\xi\right)^{1/2}.$$

Similarly we define $H^\gamma(S^{d-1}\times\mathbb{R}^n) = \{u \in \mathcal{S}'(S^{d-1}\times\mathbb{R}^n) \,|\, \|u\|_{H^\gamma(S^{d-1}\times\mathbb{R}^n)} < \infty\}$ with norm

$$\|u\|_{H^\gamma(S^{d-1}\times\mathbb{R}^n)} = \left(\int_{S^{d-1}}\int_{\mathbb{R}^n} |\hat{u}(\theta,\xi)|^2 (1+|\xi|^2)^\gamma \,\mathrm{d}\xi \,\mathrm{d}\mathcal{H}^{d-1}(\theta)\right)^{1/2},$$

where the Fourier transform is with respect to the second argument of u.
For $\Omega \subset \mathbb{R}^d$ open and bounded, the completion of $\{u \in \mathcal{S}(\mathbb{R}^d) \,|\, u = 0$ outside $\Omega\}$ with respect to $\|\cdot\|_{H^\gamma(\mathbb{R}^d)}$ is

$$H_0^\gamma(\Omega) = \{u \in H^\gamma(\mathbb{R}^d) \,|\, \operatorname{spt} u \subset \overline{\Omega}\}.$$

$H_0^\gamma(S^{d-1}\times\Omega)$ for $\Omega \subset \mathbb{R}^n$ open and bounded is defined analogously.

Remark 215 (Identification with periodic functions). *If $\Omega = (-\pi,\pi)^d$ and $\gamma \geq 0$, then any $u \in H_0^\gamma(\Omega) \subset H_0^0(\Omega) = L^2(\Omega)$ can be interpreted as (a periodic) L^2-function on Ω. An orthonormal basis of $L^2(\Omega)$ is given by $(b_k)_{k\in\mathbb{Z}^d}$ with $b_k(x) = (2\pi)^{-d/2}e^{ik\cdot x}$. Therefore we can decompose u into its Fourier series*

$$u(x) = \sum_{k\in\mathbb{Z}^d} \hat{u}_k b_k(x) = (2\pi)^{-d/2} \sum_{k\in\mathbb{Z}^d} \hat{u}_k e^{ik\cdot x} \qquad \text{with } \hat{u}_k = (u, b_k)_{L^2(\Omega)} = (2\pi)^{-d/2}\int_\Omega u(x)e^{-ik\cdot x}\,\mathrm{d}x$$

(it is always clear from the context whether $\hat{u}$ refers to the Fourier transform or the Fourier series coefficients). It turns out that the norms $\|u\|_{H_0^\gamma(\Omega)}$ and

$$\|u\|_{H^\gamma_{\mathrm{per}}(\Omega)} = \left(\sum_{k\in\mathbb{Z}^d} (1+|k|^2)^\gamma |\hat{u}_k|^2\right)^{1/2}$$

are equivalent on $H_0^\gamma(\Omega)$, where $H^\gamma_{\mathrm{per}}(\Omega)$ is the completion of infinitely smooth periodic functions on Ω with respect to $\|\cdot\|_{H^\gamma_{\mathrm{per}}(\Omega)}$. This is usually proved by interpreting $H_0^\gamma(\Omega)$ as interpolation between $H_0^{\lfloor\gamma\rfloor}(\Omega)$ and $H_0^{\lceil\gamma\rceil}(\Omega)$ and the analogous for $H^\gamma_{\mathrm{per}}(\Omega)$ (see Lemma VII.4.4 and references in Natterer, Mathematical Methods of Computerized Tomography, 2001). For other periodic domains an analogous statement holds true.

Theorem 216 (Properties of fractional Hilbert spaces). *Let $\gamma, \beta \in \mathbb{R}$, $\Omega \subset \mathbb{R}^d$ open and bounded.*

1. *For $\gamma \in \mathbb{N}_0$ the fractional Hilbert spaces $H^\gamma(\mathbb{R}^d)$ and $H_0^\gamma(\Omega)$ coincide with their usual notion, and the corresponding norms are equivalent.*
2. *$u \mapsto u^{\beta-\gamma}$ with $\widehat{u^{\beta-\gamma}}(\xi) = \hat{u}(\xi)(1+|\xi|^2)^{(\beta-\gamma)/2}$ is an isometric isomorphism $H^\beta(\mathbb{R}^d) \to H^\gamma(\mathbb{R}^d)$.*
3. *$H^\gamma(\mathbb{R}^d)$ is a Hilbert space.*
4. *Its norm is shift-invariant.*
5. *$(H^\gamma(\mathbb{R}^d))^* = H^{-\gamma}(\mathbb{R}^d)$ with dual pairing $\langle u, v\rangle = \int_{\mathbb{R}^d} \hat{u}\hat{v}\,\mathrm{d}\xi$;*
   furthermore $(H_0^\gamma(\Omega))^* \subset H_0^{-\gamma}(\Omega)$ if $\gamma \geq 0$ (opposite inclusion for $\gamma \leq 0$ by reflexivity).
6. *$\|\cdot\|_{H^\gamma} > \|\cdot\|_{H^\beta}$ for $\gamma > \beta$ and $H^\gamma(\mathbb{R}^d) \subsetneq H^\beta(\mathbb{R}^d)$, $H_0^\gamma(\Omega) \subsetneq H_0^\beta(\Omega)$.*
7. *$H_0^\gamma(\Omega)$ embeds compactly into $H_0^\beta(\Omega)$ for $\gamma > \beta$.*

Proof. 1.
$$\sum_{|\alpha|\leq n}\int_{\mathbb{R}^d} |D^\alpha u|^2\,\mathrm{d}x = \sum_{|\alpha|\leq n}\int_{\mathbb{R}^d} |\xi^\alpha \hat{u}(\xi)|^2\,\mathrm{d}\xi \begin{cases} \leq \int_{\mathbb{R}^d} (1+|\xi|^2)^n |\hat{u}(\xi)|^2\,\mathrm{d}\xi \\ \geq \text{const.} \int_{\mathbb{R}^d} (1+|\xi|^2)^n |\hat{u}(\xi)|^2\,\mathrm{d}\xi \end{cases}$$

2. straightforward
3. inner product $(u,v)_{H^\gamma} = \int_{\mathbb{R}^d} \hat{u}(\xi)\overline{\hat{v}(\xi)}(1+|\xi|^2)^\gamma \,\mathrm{d}\xi$
   Cauchy sequence $u_n \in H^\gamma$ induces Cauchy sequence u_n^γ in L^2;
   convergence of the latter to u in L^2 implies convergence of the former to $u^{-\gamma}$ in H^γ
4. straightforward
5. By 2., there is a one-to-one correspondence between $\ell \in (H^\gamma(\mathbb{R}^d))^*$ and $l \in (L^2(\mathbb{R}^d))^* = L^2(\mathbb{R}^d)$ via $\ell(u) = l(u^\gamma) = \int_{\mathbb{R}^d} \hat{l}\overline{\widehat{u^\gamma}} \,\mathrm{d}\xi = \int_{\mathbb{R}^d} \widehat{l^\gamma}\overline{\hat{u}} \,\mathrm{d}\xi = \langle l^\gamma, u\rangle$, where $l^\gamma \in H^{-\gamma}(\mathbb{R}^d)$.
6. straightforward
7. Let $u_n \rightharpoonup u$ in $H_0^\gamma(\Omega)$; need to show $u_n \to u$ in $H_0^\beta(\Omega)$.

   Wlog. $\beta \geq 0$.

   - if $\beta < 0$ set $m = 2\lceil -\beta/2\rceil \in \mathbb{N}_0$
   - let $\chi \in C_c^\infty(\mathbb{R}^d)$ with $\chi = 1$ on a neighbourhood of Ω, $\chi = 0$ outside $\tilde{\Omega} \supset \Omega$
   - $u_n \rightharpoonup u$ in $H_0^\gamma(\Omega) \Rightarrow u_n^{-m} \rightharpoonup u^{-m}$ in $H^{\gamma+m}(\mathbb{R}^d) \Rightarrow \chi u_n^{-m} \rightharpoonup \chi u^{-m}$ in $H_0^{\gamma+m}(\tilde{\Omega}) \subset H^{\gamma+m}(\mathbb{R}^d)$
   - $(\chi(u_n^{-m} - u^{-m}))^m = (\mathrm{id} - \Delta)^{m/2}(\chi(u_n^{-m} - u^{-m})) = (\mathrm{id} - \Delta)^{m/2}(u_n^{-m} - u^{-m}) = u_n - u$ on Ω.
   - if $\chi u_n^{-m} \to \chi u^{-m}$ in $H_0^{\beta+m}(\tilde{\Omega})$, then

   $$\begin{aligned} \|u_n - u\|_{H_0^\beta(\Omega)} &= \sup_{\|v\|_{H^{-\beta}} \leq 1, \operatorname{spt} v \subset \overline{\Omega}} \langle v, (\chi(u_n^{-m} - u^{-m}))^m\rangle \\ &\leq \sup_{\|v\|_{H^{-\beta}} \leq 1} \langle v, (\chi(u_n^{-m} - u^{-m}))^m\rangle \\ &= \sup_{\|v\|_{H^{-\beta-m}} \leq 1} \langle v, \chi(u_n^{-m} - u^{-m})\rangle \to 0 \end{aligned}$$

   Wlog. $\Omega = (-\pi, \pi)^d$.

   - by shifting and rescaling coordinates we achieve $\Omega \subset (-\pi,\pi)^d$ without changing convergences
   - $(H_0^\gamma((-\pi,\pi)^d))^* \subset (H_0^\gamma(\Omega))^*$; thus $u_n \rightharpoonup u$ in $H_0^\gamma((-\pi,\pi)^d)$
   - $H_0^\beta(\Omega)$ is a closed subset of $H_0^\beta((-\pi,\pi)^d)$;
     hence, $u_n \to u$ in $H_0^\beta((-\pi,\pi)^d)$ implies $u_n \to u$ in $H_0^\beta(\Omega)$

   $u_n \rightharpoonup u$ in $H_0^\gamma(\Omega) \Rightarrow (\widehat{u_n})_k \to \hat{u}_k$ for all $k \in \mathbb{Z}^d$.

   - $\gamma \geq 0 \Rightarrow u_n \rightharpoonup u$ in $L^2(\Omega)$
   - $(u_n - u, b_k)_{L^2(\Omega)} = ((\widehat{u_n})_k - \hat{u}_k)$

   $u_n \to u$ in $H_0^\beta(\Omega)$.

   - abbreviate $M = \max\{\|u\|_{H^\gamma_{\mathrm{per}}(\Omega)}, \sup_n \|u_n\|_{H^\gamma_{\mathrm{per}}(\Omega)}\}$ and fix an arbitrary $\varepsilon > 0$
   - let $R^2 > (8M^2/\varepsilon)^{\frac{1}{\gamma-\beta}} - 1$
   - let $N > 0$ large enough such that $\sum_{k\in\mathbb{Z}^d, |k|\leq R}(1+|k|^2)^\beta |(\widehat{u_n})_k - \hat{u}_k|^2 \leq \frac{\varepsilon}{2}$ for all $n > N$
   - $$\begin{aligned} \|u_n - u\|^2_{H^\beta_{\mathrm{per}}(\Omega)} &= \sum_{k\in\mathbb{Z}^d, |k|\leq R}(1+|k|^2)^\beta |(\widehat{u_n})_k - \hat{u}_k|^2 + \sum_{k\in\mathbb{Z}^d, |k|> R}(1+|k|^2)^\beta |(\widehat{u_n})_k - \hat{u}_k|^2 \\ &\leq \frac{\varepsilon}{2} + 2\sum_{k\in\mathbb{Z}^d, |k|> R}(1+|k|^2)^\beta (|(\widehat{u_n})_k|^2 + |\hat{u}_k|^2) \\ &\leq \frac{\varepsilon}{2} + 2(1+R^2)^{\beta-\gamma} 2M^2 \\ &\leq \varepsilon \end{aligned}$$

□

Remark 217 (Dirac measure)**.** *We have $\delta_x \in H^{-\gamma}(\mathbb{R}^d)$ if and only if $\gamma > \frac{d}{2}$ (homework). Similarly, $\gamma > \frac{d}{2}$ is equivalent to point evaluation at x being a continuous linear operator on $H^\gamma(\mathbb{R}^d)$ (homework). Now let $\gamma > \frac{d}{2}$ and $x \in \partial\Omega$ for a bounded open $\Omega \subset \mathbb{R}^d$, then $\delta_x \in H_0^{-\gamma}(\Omega)$, but $u(x) = 0$ for any $u \in H_0^\gamma(\Omega)$ so that $(H_0^{-\gamma}(\Omega))^* \not\subset H_0^\gamma(\Omega)$.*

Remark 218 (Compact embedding versus shift-invariance)**.** *The embedding $H^\gamma(\mathbb{R}^d) \hookrightarrow H^\beta(\mathbb{R}^d)$ is never compact due to the shift-invariance of $H^\gamma(\mathbb{R}^d)$ and $H^\beta(\mathbb{R}^d)$ (cf. theorem 42).*

To understand later the degree of ill-posedness of the Radon and X-ray transform we now aim to derive how the singular values of the compact embedding $H_0^\gamma(\Omega) \hookrightarrow H_0^\beta(\Omega)$ decay.

Theorem 219 (Courant–Fisher–Weyl min-max principle)**.** *Let X, Y be Hilbert spaces, $K : X \to Y$ linear and compact. The kth singular value ρ_k of K equals the numbers*

$$C_k := \max\{\inf\{\|Kx\|_Y \,|\, x \in S,\ \|x\|_X \geq 1\} \,|\, S \subset X \textit{ is } k\textit{-dimensional subspace}\},$$
$$D_k := \min\left\{\sup\{\|Kx\|_Y \,|\, x \in S^\perp,\ \|x\|_X \leq 1\} \,\middle|\, S \subset X \textit{ is } (k-1)\textit{-dimensional subspace}\right\}.$$

Proof.

- $Kx = \sum_{n=1}^\infty \rho_n (x, u_n)_X v_n$ for orthonormal left & right singular vectors $u_n \in X$, $v_n \in Y$
- take $S = \operatorname{span}\{u_1, \ldots, u_k\}$, then
$$C_k \geq \inf\{\|Kx\|_Y \,|\, x \in \operatorname{span}\{u_1, \ldots, u_k\},\ \|x\|_X \geq 1\}$$
$$= \inf\left\{\left(\textstyle\sum_{n=1}^k \rho_n^2 (x, u_n)_X^2\right)^{1/2} \,\middle|\, x \in \operatorname{span}\{u_1, \ldots, u_k\},\ \|x\|_X \geq 1\right\} = \rho_k$$
- consider arbitrary k-dimensional subspace $S \subset X$;
there exists some $x \in \operatorname{span}\{u_1, \ldots, u_{k-1}\}^\perp \cap S$ (k unknowns, $k-1$ equations); set $v = x/\|x\|_X$;
$\Rightarrow \|Kv\|_Y^2 = \sum_{n=k}^\infty \rho_n^2 (v, u_n)_X^2 \leq \rho_k^2 \sum_{n=k}^\infty (v, u_n)_X^2 = \rho_k^2$
$\Rightarrow C_k \leq \max\{\rho_k \,|\, S \subset X \text{ is } k\text{-dimensional subspace}\} = \rho_k$
- analogous argument for D_k □

Theorem 220 (Singular values of composition)**.** *Let W, X, Y, Z be Hilbert spaces, $K : X \to Y$ linear and compact with singular values $(\sigma_k)_{k\in\mathbb{N}_0}$, and $J : W \to X$ as well as $L : Y \to Z$ linear and bounded. Then the singular values $(\lambda_k)_{k\in\mathbb{N}_0}$ of LKJ satisfy*

$$\lambda_k \leq \|L\|\|J\|\sigma_k.$$

If L and J are bijective, then also

$$\lambda_k \geq \tfrac{1}{\|L^{-1}\|\|J^{-1}\|}\sigma_k.$$

Proof. For any $w \in W$ and $S \subset W$ we have

$$\|LKJw\|_Z \leq \|L\|\|KJw\|_Y,$$
$$\{x \in X \,|\, x \in JS,\ \|x\|_X \geq \|J\|\} = \{Jw \in X \,|\, w \in S,\ \|Jw\|_X \geq \|J\|\} \subset \{Jw \in X \,|\, w \in S,\ \|w\|_W \geq 1\}.$$

Therefore

$$\begin{aligned}
\lambda_k &= \max\{\inf\{\|LKJw\|_Z \,|\, w \in S,\ \|w\|_W \geq 1\} \,|\, S \subset W \text{ is } k\text{-dimensional subspace}\} \\
&\leq \|L\| \max\{\inf\{\|KJw\|_Y \,|\, w \in S,\ \|w\|_W \geq 1\} \,|\, S \subset W \text{ is } k\text{-dimensional subspace}\} \\
&= \|L\| \max\{\inf\{\|KJw\|_Y \,|\, w \in S,\ \|w\|_W \geq 1\} \,|\, S \subset W \text{ is } k\text{-dimensional subspace, } J \text{ injective on } S\} \\
&= \|L\| \max\{\inf\{\|Kx\|_Y \,|\, x \in JS,\ \|x\|_X \geq \|J\|\} \,|\, S \subset W \text{ is } k\text{-dimensional subspace, } J \text{ injective on } S\} \\
&= \|L\| \max\{\inf\{\|Kx\|_Y \,|\, x \in JS,\ \|x\|_X \geq \|J\|\} \,|\, JS \subset X \text{ is } k\text{-dimensional subspace}\} \\
&\leq \|L\|\|J\| \max\left\{\inf\{\|Kx\|_Y \,|\, x \in \tilde{S},\ \|x\|_X \geq 1\} \,\middle|\, \tilde{S} \subset X \text{ is } k\text{-dimensional subspace}\right\} \\
&= \|L\|\|J\|\sigma_k.
\end{aligned}$$

Now let L, J both have bounded inverse, then

$$\|KJw\|_Y = \|L^{-1}LKJw\|_Y \leq \|L^{-1}\|\|LKJw\|_Z,$$
$$\{Jw \,|\, w \in S,\ \|w\|_W \geq 1\} = \{x \,|\, x \in JS,\ \|J^{-1}x\|_W \geq 1\} \subset \{x \,|\, x \in JS,\ \|x\|_X \geq \tfrac{1}{\|J^{-1}\|}\}.$$

Therefore

$$\begin{aligned}
\lambda_k &= \max\{\inf\{\|LKJw\|_Z \,|\, w \in S,\ \|w\|_W \geq 1\} \,|\, S \subset W \text{ is } k\text{-dimensional subspace}\} \\
&\geq \tfrac{1}{\|L^{-1}\|} \max\{\inf\{\|KJw\|_Y \,|\, w \in S,\ \|w\|_W \geq 1\} \,|\, S \subset W \text{ is } k\text{-dimensional subspace}\} \\
&\geq \tfrac{1}{\|L^{-1}\|} \max\left\{\inf\{\|Kx\|_Y \,|\, x \in JS,\ \|x\|_X \geq \tfrac{1}{\|J^{-1}\|}\} \,\middle|\, S \subset W \text{ is } k\text{-dimensional subspace}\right\} \\
&= \tfrac{1}{\|L^{-1}\|\|J^{-1}\|} \max\left\{\inf\{\|Kx\|_Y \,|\, x \in \tilde{S},\ \|x\|_X \geq 1\} \,\middle|\, \tilde{S} \subset X \text{ is } k\text{-dimensional subspace}\right\} \\
&= \tfrac{1}{\|L^{-1}\|\|J^{-1}\|}\sigma_k.
\end{aligned}$$

□

Theorem 221 (SVD of H^γ-embedding)**.** *Let $\gamma > \beta \geq 0$ and $\Omega \subset \mathbb{R}^d$ open and bounded, then the singular values σ_k of the embedding $\iota : H_0^\gamma(\Omega) \hookrightarrow H_0^\beta(\Omega)$ decay like $\sigma_k \sim k^{\frac{\beta-\gamma}{d}}$.*

Proof. First consider $\Omega = (-\pi, \pi)^d$.

1. The singular values λ_k of $H^\gamma_{\mathrm{per}}(\Omega) \hookrightarrow H^\beta_{\mathrm{per}}(\Omega)$ decay like $\lambda_k \sim k^{\frac{\beta-\gamma}{d}}$:
   - $(v_n^\alpha)_{n\in\mathbb{Z}^d}$ with $(\widehat{v_n^\alpha})_n = (1+|n|^2)^{-\alpha/2}$ and $(\widehat{v_n^\alpha})_k = 0$ else forms orthonormal basis of $H^\alpha_{\mathrm{per}}(\Omega)$
   - $v_n^\gamma = (1+|n|^2)^{\frac{\beta-\gamma}{2}} v_n^\beta$, hence singular values are $(1+|n|^2)^{\frac{\beta-\gamma}{2}}$
   - kth singular value corresponds to $n_k \in \mathbb{Z}^d$ with kth smallest norm, thus $k \sim |n_k|^d$
   - $\lambda_k = (1+|n_k|^2)^{\frac{\beta-\gamma}{2}} \sim k^{\frac{\beta-\gamma}{d}}$
2. The $H^\alpha_{\mathrm{per}}(\Omega)$-orthogonal projection $\mathrm{P}_\alpha : H^\alpha_{\mathrm{per}}(\Omega) \to H^\alpha_0(\Omega)$ for $\alpha \geq 0$ is well-posed:
   - $H^\alpha_0(\Omega) \subset H^\alpha_{\mathrm{per}}(\Omega)$
   - $H^\alpha_0(\Omega)$=completion of Schwartz functions with support in Ω wrt. $\|\cdot\|_{H^\alpha_0(\Omega)}$, thus wrt. $\|\cdot\|_{H^\alpha_{\mathrm{per}}(\Omega)}$
   - $\Rightarrow H^\alpha_0(\Omega)$ is closed subset of $H^\alpha_{\mathrm{per}}(\Omega)$
3. σ_k decay at least like $\sigma_k \lesssim k^{\frac{\beta-\gamma}{d}}$:
   - let $\iota^\alpha_{0,\mathrm{per}} : H^\alpha_0(\Omega) \hookrightarrow H^\alpha_{\mathrm{per}}(\Omega)$ and $\iota_{\mathrm{per}} : H^\gamma_{\mathrm{per}}(\Omega) \hookrightarrow H^\beta_{\mathrm{per}}(\Omega)$
   - $\iota = \mathrm{P}_\beta \circ \iota_{\mathrm{per}} \circ \iota^\gamma_{0,\mathrm{per}}$
   - $\iota^\gamma_{0,\mathrm{per}}$ and P_β are bounded; now use theorem 220 to obtain $\sigma_k \lesssim \lambda_k$
4. σ_k decay at most like $\sigma_k \gtrsim k^{\frac{\beta-\gamma}{d}}$:
   - let $E^\alpha : H^\alpha_{\mathrm{per}}(\Omega) \to H^\alpha_{\mathrm{per}}(\Omega)$ be given by $(\widehat{E^\alpha u})_{2k} = \hat{u}_k$ and $(\widehat{E^\alpha u})_k = 0$ else;
     then $\|E^\alpha\| \leq 2^\alpha$ and $E^\alpha u(x) = u(2x)$ (assuming u on rhs is periodically extended)
   - let $\chi \in C^\infty_c(\Omega)$ with $\chi = 1$ on a neighbourhood of $\Omega/2$ and $F^\alpha : H^\alpha_{\mathrm{per}}(\Omega) \to H^\alpha_0(\Omega)$, $F^\alpha u = \chi u$;
     F^α is bounded (which can be seen using the norm $\|\cdot\|_{H^\alpha_{\mathrm{per}}(\Omega)}$ on domain and codomain, since $\widehat{F^\alpha u} = \hat{\chi} * \hat{u}$, where $\hat{\chi}_k$ decays faster than any power of $|k|$)
   - $X = \mathrm{ran}\,(F^\beta E^\beta)$ is closed in $H^\beta_{\mathrm{per}}(\Omega)$:
     let $u_n \in X$ with $u_n \to u$ in $H^\beta_{\mathrm{per}}(\Omega)$,
     then also $u_n \to u$ in $L^2(\Omega)$ and thus pointwise a. e. along a subsequence
     $v \in X \Leftrightarrow v \in H^\beta_{\mathrm{per}}(\Omega)$ & pointwise condition $v(x) = \chi(x)v(x/2)$ if $x \in \Omega \setminus \frac{\Omega}{2}$
   - $X \subset H^\beta_{\mathrm{per}}(\Omega)$ closed
     $\Rightarrow$ orthogonal projection $P : H^\beta_{\mathrm{per}}(\Omega) \to X$ is well-defined,
     $\Rightarrow (F^\beta E^\beta)$ has bounded inverse $T : X \to H^\beta_{\mathrm{per}}(\Omega)$ by injectivity and bounded inverse theorem
     $\Rightarrow$ by Hahn–Banach, T can be extended to a bounded linear operator $T : H^\beta_{\mathrm{per}}(\Omega) \to H^\beta_{\mathrm{per}}(\Omega)$

- $\iota_{\text{per}} = T \circ P \circ \iota^{\beta}_{0,\text{per}} \circ \iota \circ F^{\gamma} \circ E^{\gamma}$; now use theorem 220 to obtain $\lambda_k \lesssim \sigma_k$

For an arbitrary domain Ω let $\bar{\Omega} = (-\pi, \pi)^d$ and G, J be domain translations and rescalings such that

$$G\Omega \subset\subset \bar{\Omega} \text{ and } J\bar{\Omega} \subset\subset \Omega,$$

and let $\bar{\iota}$ be the embedding $H^{\gamma}_0(\bar{\Omega}) \hookrightarrow H^{\beta}_0(\bar{\Omega})$, then

$$\iota : H^{\gamma}_0(\Omega) \xrightarrow{\text{bounded } \circ G} H^{\gamma}_0(G\Omega) \hookrightarrow H^{\gamma}_0(\bar{\Omega}) \overset{\bar{\iota}}{\hookrightarrow} H^{\beta}_0(\bar{\Omega}) \xrightarrow{\text{orth. proj.}} H^{\beta}_0(G\Omega) \xrightarrow{\text{bounded } \circ G^{-1}} H^{\beta}_0(\Omega)$$
$$\bar{\iota} : H^{\gamma}_0(\bar{\Omega}) \xrightarrow{\text{bounded } \circ J^{-1}} H^{\gamma}_0(J^{-1}\bar{\Omega}) \hookrightarrow H^{\gamma}_0(\Omega) \overset{\iota}{\hookrightarrow} H^{\beta}_0(\Omega) \xrightarrow{\text{orth. proj.}} H^{\beta}_0(J^{-1}\bar{\Omega}) \xrightarrow{\text{bounded } \circ J} H^{\beta}_0(\bar{\Omega})$$

so that the singular values of ι and $\bar{\iota}$ decay at the same rate. □

Remark 222 (Embedding for negative β)**.** *Let $0 \leq \beta \leq \gamma$. The adjoint of the embedding $H^{\gamma}_0(\Omega) \hookrightarrow H^{\beta}_0(\Omega)$ is the embedding $(H^{\beta}_0(\Omega))^* \hookrightarrow (H^{\gamma}_0(\Omega))^*$, so its singular values also decay at the same rate. Exploiting the embeddings $H^{-\alpha}_0(\underline{\Omega}) \subset (H^{\alpha}_0(\Omega))^* \subset H^{-\alpha}_0(\overline{\Omega})$ for $\underline{\Omega} \subset\subset \Omega \subset\subset \overline{\Omega}$, as in the previous proof one obtains the decay rate $k^{\frac{\beta-\gamma}{d}}$ even for $H^{\gamma}_0(\Omega) \hookrightarrow H^{\beta}_0(\Omega)$ with arbitrary $\gamma > \beta$.*

23 Radon and X-ray transform on bounded domains

Remarks 206 and 207 depended on the unbounded domain. On a bounded domain things get simpler. Below we identify a pair $(\theta, s) \in \mathcal{C}$ with the hyperplane $s\theta + \theta^{\perp}$ and a pair $(\theta, s) \in \mathcal{C}'$ with the line $s + \theta\mathbb{R}$.

Lemma 223 (Support of integral transforms)**.** *If u is compactly supported, then so are Ru and Pu (they only have support on hyperplanes or lines (θ, s) that intersect the support of u).*

Proof. trivial □

The reverse holds true as well (which in case of nonnegative u is trivial). It can be shown using Cormack's original inversion formula for the Radon transform (which we will not derive).

Theorem 224 (Support of inverse integral transforms)**.** *Let $u \in \mathcal{S}(\mathbb{R}^d)$ and $\Omega \subset \mathbb{R}^d$ be convex and compact.*

1. *If $Ru(\theta, s) = 0$ for every hyperplane (θ, s) not intersecting Ω, then $u = 0$ on $\mathbb{R}^d \setminus \Omega$.*
2. *If $Pu(\theta, s) = 0$ for every line (θ, s) not intersecting Ω, then $u = 0$ on $\mathbb{R}^d \setminus \Omega$.*

Proof. 1. Suffices to consider balls Ω, since for any $x \notin \Omega$ there is a ball containing Ω, but not x.
By coordinate transform suffices to consider Ω to be the unit ball.
Cormack's inversion formula shows $Ru = 0$ on $S^{d-1} \times (\mathbb{R} \setminus [-1, 1]) \Rightarrow u = 0$ outside unit ball.
2. Follows from 1. since each hyperplane not intersecting Ω is spanned by lines not intersecting Ω. □

Theorem 225 (Sobolev estimates for Radon and X-ray transform)**.** *Let $\Omega \subset \mathbb{R}^d$ be bounded and open and $\gamma \in \mathbb{R}$. There exist constants $c, C > 0$, depending only on γ, d, and Ω such that*

$$c\|u\|_{H^{\gamma}_0(\Omega)} \leq \|Ru\|_{H^{\gamma+\frac{d-1}{2}}(\mathcal{C})}, \|Pu\|_{H^{\gamma+\frac{1}{2}}(\mathcal{C}')} \leq C\|u\|_{H^{\gamma}_0(\Omega)} \qquad \text{for all } u \in C^{\infty}_0(\Omega).$$

Proof. 1.
$$\begin{aligned}
\|Ru\|^2_{H^{\gamma+\frac{d-1}{2}}(\mathcal{C})} &= \int_{S^{d-1}} \int_{\mathbb{R}} (1 + \sigma^2)^{\gamma+\frac{d-1}{2}} |\widehat{R_{\theta}u}(\sigma)|^2 \,\mathrm{d}\sigma \,\mathrm{d}\mathcal{H}^{d-1}(\theta) \\
&= (2\pi)^{d-1} \int_{S^{d-1}} \int_{\mathbb{R}} (1 + \sigma^2)^{\gamma+\frac{d-1}{2}} |\hat{u}(\sigma\theta)|^2 \,\mathrm{d}\sigma \,\mathrm{d}\mathcal{H}^{d-1}(\theta) \\
&= 2(2\pi)^{d-1} \int_{S^{d-1}} \int_0^{\infty} (1 + \sigma^2)^{\gamma+\frac{d-1}{2}} |\hat{u}(\sigma\theta)|^2 \,\mathrm{d}\sigma \,\mathrm{d}\mathcal{H}^{d-1}(\theta) \\
&= 2(2\pi)^{d-1} \int_{\mathbb{R}^d} (1 + |\xi|^2)^{\gamma+\frac{d-1}{2}} |\hat{u}(\xi)|^2 |\xi|^{1-d} \,\mathrm{d}\xi
\end{aligned}$$

- $\|Ru\|^2_{H^{\gamma+\frac{d-1}{2}}(\mathcal{C})} \geq 2(2\pi)^{d-1}\int_{\mathbb{R}^d}(1+|\xi|^2)^\gamma|\hat u(\xi)|^2\,\mathrm{d}\xi = 2(2\pi)^{d-1}\|u\|^2_{H^\gamma(\mathbb{R}^d)}$
- $\frac{\|Ru\|^2_{H^{\gamma+\frac{d-1}{2}}(\mathcal{C})}}{2(2\pi)^{d-1}} = \underbrace{\int_{\{|\xi|\leq 1\}}(1+|\xi|^2)^{\gamma+\frac{d-1}{2}}\underbrace{|\hat u(\xi)|^2}_{\leq\|\hat u\|^2_{L^\infty(\{|\xi|\leq1\})}}|\xi|^{1-d}\,\mathrm{d}\xi}_{\leq\text{const.}\|\hat u\|^2_{L^\infty(\{|\xi|\leq1\})}} + \underbrace{\int_{\{|\xi|>1\}}(1+|\xi|^2)^{\gamma+\frac{d-1}{2}}|\hat u(\xi)|^2\underbrace{|\xi|^{1-d}}_{\leq(\frac{1+|\xi|^2}{2})^{\frac{1-d}{2}}}\,\mathrm{d}\xi}_{\leq 2^{\frac{d-1}{2}}\|u\|^2_{H^\gamma(\mathbb{R}^d)}}$
- let $\chi\in C_0^\infty(\mathbb{R}^d)$ be one on Ω, and set $\chi_\xi(x)=e^{-ix\cdot\xi}\chi(x)$, then

$$
\begin{aligned}
(2\pi)^{d/2}|\hat u(\xi)| &= \left|\int_{\mathbb{R}^d}\chi_\xi(x)u(x)\,\mathrm{d}x\right| = \left|\int_{\mathbb{R}^d}\check\chi_\xi(\eta)\hat u(\eta)\,\mathrm{d}\eta\right| \\
&\leq \left(\int_{\mathbb{R}^d}\frac{1}{(1+|\eta|^2)^\gamma}|\check\chi_\xi(\eta)|^2\,\mathrm{d}\eta\right)^{\frac12}\left(\int_{\mathbb{R}^d}(1+|\eta|^2)^\gamma|\hat u(\eta)|^2\,\mathrm{d}\eta\right)^{\frac12} = \|\chi_\xi\|_{H^{-\gamma}(\mathbb{R}^d)}\|u\|_{H^\gamma(\mathbb{R}^d)}
\end{aligned}
$$

- $\|\chi_\xi\|_{H^{-\gamma}(\mathbb{R}^d)}$ depends continuously on ξ, thus the supremum over $|\xi|\leq 1$ is bounded

2.
$$
\begin{aligned}
\|Pu\|^2_{H^{\gamma+\frac12}(\mathcal{C}')} &= \int_{S^{d-1}}\int_{\theta^\perp}(1+|\xi|^2)^{\gamma+\frac12}|\widehat{P_\theta u}(\xi)|^2\,\mathrm{d}\mathcal{H}^{d-1}(\xi)\,\mathrm{d}\mathcal{H}^{d-1}(\theta) \\
&= 2\pi\int_{S^{d-1}}\int_{\theta^\perp}(1+|\xi|^2)^{\gamma+\frac12}|\hat u(\xi)|^2\,\mathrm{d}\mathcal{H}^{d-1}(\xi)\,\mathrm{d}\mathcal{H}^{d-1}(\theta) \\
&= 2\pi\mathcal{H}^{d-2}(S^{d-2})\int_{\mathbb{R}^d}(1+|\eta|^2)^{\gamma+\frac12}|\hat u(\eta)|^2|\eta|^{-1}\,\mathrm{d}\eta
\end{aligned}
$$

(using lemma 201 in last step); rest analogous to Radon transform □

Remark 226 (Compactness of transforms). *We see that on a bounded domain Ω, not only is R (analogously P) bounded from $L^2(\Omega)=H_0^0(\Omega)$ into $L^2(\mathcal{C})$, but even compact: It is the composition of the compact embedding $H_0^0(\Omega)\hookrightarrow H_0^{-\frac{d-1}{2}}(\Omega)$ with the bounded $R: H_0^{-\frac{d-1}{2}}(\Omega)\to L^2(\Omega)$. (Note that $H_0^{\frac{d-1}{2}}(S^{d-1}\times[a,b])$ does not embed compactly into $L^2(\mathcal{C})$ since it has no additional regularity along S^{d-1}, but one can show that the subspace satisfying the Helgason–Ludwig conditions does; in other words, R is continuous from $H_0^\gamma(\Omega)$ into an even more regular space than $H^{\gamma+\frac{d-1}{2}}(\mathcal{C})$ – one with additional regularity along S^{d-1}.)*

Remark 227 (Ill-posedness). *Obviously, inversion of R (analogously for P) is well-posed if R is interpreted as an operator from $L^2(\Omega)$ to $H^{\frac{d-1}{2}}(\mathcal{C})$. However, typically the measurement error lies in $L^2(\mathcal{C})$ or is even less regular (for instance, Gaussian white noise is in $H^{-\gamma}(\mathbb{R}^d)$ if and only if $\gamma>\frac12$), and we often require the reconstruction to have small errors in $L^2(\Omega)$ or even $H_0^1(\Omega)$. Thus we need to interpret R as an operator from $L^2(\Omega)$ or $H_0^1(\Omega)$ into $L^2(\mathcal{C})$ or $H^{-\gamma}(\mathcal{C})$, which is compact.*

Corollary 228 (Mild ill-posedness). *Let $\beta\leq\gamma$. The singular values of $R: H_0^\gamma(\Omega)\to H^\beta(\mathcal{C})$ decay like $\sigma_k\sim k^{\frac{\beta-\gamma}{d}-\frac{d-1}{2d}}$, the singular values of $P: H_0^\gamma(\Omega)\to H^\beta(\mathcal{C}')$ like $\sigma_k\sim k^{\frac{\beta-\gamma}{d}-\frac{1}{2d}}$, so inversion of both is (very) mildly ill-posed.*

Proof. Interpret R as composition

$$H_0^\gamma(\Omega)\xrightarrow{\iota\text{ (compact embedding)}}H_0^{\beta+\frac{1-d}{2}}(\Omega)\xrightarrow{R\text{ (boundedly invertible on its range)}}H^\beta(\mathcal{C}),$$

and apply theorem 220 and remark 222. Analogous argument for P. □

In fact, explicit singular value decompositions for the Radon transform between multiple different spaces are known, for instance between weighted L^2-spaces on bounded domains.