Inverse Probleme im WS 2020/2021
Lernskript
Stand 3. Februar 2021, 10:41:58

Frank Wübbeling

3. Februar 2021

Contents

Chapter 1

Introduction and Examples for inverse problems

Chapter 2

Mathematical foundations

In this chapter, we rehearse mathematical basic knowledge and borrow a lot of stuff from other lectures in functional analysis.

2.1 About Linear Operators

Definition 2.1
Let X,Y normed vector spaces (usually Banach or Hilbert), and $K : X \mapsto Y$. Let $y \in Y$. The problem:

Find $x \in X$ such that $Kx = y$

is properly posed or well–posed, iff

1. $\forall\, y \in Y\ \exists\, x \in X : Kx = y$ *(solvability)*
2. $\forall\, y \in Y\ \exists!\, x \in X : Kx = y$ *(uniqueness)*
3. $K^{-1} : Y \mapsto X$ *is continuous w.r.t. the norms in X, Y (continuity).*

If one of the conditions is violated, the problem is improperly posed or ill–posed.

- Throughout the lecture, K will be linear (or affine linear).
- Usually, X and Y are ∞–dimensional function spaces, so our reference lecture is functional analysis.

Reminder: The operator norm of K is defined by

$$||K||_{X,Y} := \sup_{x \in X,\, x \neq 0} \frac{||Kx||_Y}{||x||_X} = \sup_{x \in X, ||x||_X = 1} ||Kx||_Y.$$

Reminder: K is continuous iff $||K|| < \infty$.

Remark: Let $x \in X$. Let $\widetilde{x} \in X$ an approximation to x. Then

$$||K\widetilde{x} - Kx|| \leq ||K||\,||\widetilde{x} - x||, \tag{2.1}$$

so the maximum error amplification factor when applying K is $||K||$.

Remark 2.2
The continuity of an operator depends on the norms in X and Y.

Example: Take $K : C^1[0,1] \mapsto C^0[0,1]$, $Kf := f'$.
Then K is discontinuous w.r.t. $(C^1([0,1]), ||\cdot||_\infty)$.
K is continuous w.r.t. $(C^1([0,1]), ||f||_\infty + ||f'||_\infty)$.

Example: Take any inverse problem for which an inverse operator exists. With the notation from 2.1: Define the norm in Y by

$$|||y||| := ||y||_Y + ||K^{-1}y||_X.$$

Then K is continuous. However, this does not help since in practice the norms are fixed and given by the application.

Remark: Let $\dim X = \infty$, and $Y \neq \{0\}$. Let $0 \neq y \in Y$ and (x_n) an infinite linearly independent system. Define $Kx_n := ny$ and extend it to a linear operator. Then K is discontinuous.

Remark: Let $\dim X < \infty$. Then all linear operators are continuous.

Definition 2.3
Let X, Y function spaces, $K : X \mapsto Y$. Assume that K is discontinuous. K is said to be ill–posed of the order m, if K becomes well–posed when adding norms of derivatives of order up to m to the norm in Y and ill-posed when adding norms of derivatives of order up to $m-1$.

Remark: The higher the order of ill–posedness, the harder it is to solve the inverse problem.

Remark: The inverse problem of integration (differentiation) is ill–posed of order one, differentiation of order m is ill–posed of order m.

Remark: The inverse problem for the diffusion equation never becomes continuous by adding derivatives, it has order ∞.

Remark: The inverse problem of the Radon Transform has order $\frac{1}{2}$, so it is easier to solve than differentiation of order one.

This is an off–hand definition. We will give a better definition in terms of the singular values of an operator later. And we will have to define differentiation of order $\frac{1}{2}$.

Definition 2.4
Let $\Sigma \subset \mathbb{R}^n$, $\Omega \subset \mathbb{R}^m$ measurable and open. Let $k : \Sigma \times \Omega \mapsto \mathbb{R}$ measurable. For $u : \Omega \mapsto \mathbb{R}$, $x \in \Sigma$ we define

$$(Ku)(x) := \int_\Omega k(x,y)u(y)dy.$$

K is an operator from a function space over Ω to a function space over Σ and is called integral operator.

Theorem 2.5
Let $\Sigma \subset \mathbb{R}^n$, $\Omega \subset \mathbb{R}^m$ bounded, $k \in L^2(\Sigma \times \Omega)$. Then $K : L^2(\Sigma) \mapsto L^2(\Omega)$ is continuous.

Reminder: Let X, Y Hilbertspaces, $K : X \mapsto Y$ linear and continuous. Then there exists a continuous linear adjoint operator $K^* : Y \mapsto X$ with the property that

$$(Kx, y)_Y = (x, K^*y)_X \ \forall x \in X,\ y \in Y.$$

Reminder: Let M a linear subspace of X. Denote by $\overline{M}$ the closure of M. Let

$$M^\perp = \{x \in X : (x,y) = 0 \, \forall y \in X\}.$$

Then $X = \overline{M} \oplus M^\perp$, so for all $x \in X$ we have $x = x_0 + x_1$ with $x_0 \in \overline{M}$, $x_1 \in M^\perp$. $||x - x_0|| \leq ||x - y|| \, \forall y \in \overline{M}$. The mapping P from x to x_0 is the orthogonal projection of x onto $\overline{M}$.

Definition 2.6
Let f, g functions on $\mathbb{R}^n$. The convolution of f and g is a function of $\mathbb{R}^n$ and defined as

$$(f * g)(x) = \int_{\mathbb{R}^n} f(x-y)g(y)\, dy$$

Lemma 2.7
For $f \in L^1$, $g \in L^1$ $f * g$ is well–defined and is in L^1. For fixed f or g the convolution is continuous from $L^1 \mapsto L^1$.

Remark: Not all inverse problems are ill–posed. Some examples:

Definition 2.8
Let f in the Schwartz–space $\mathcal{S}$. Then the Hilbert transform is defined as

$$(\mathcal{H}(f))(x) = \frac{1}{\pi} \not\int_{\mathbb{R}} \frac{f(y)}{x-y}\, dy.$$

The Hilbert transform is a continuous operator and its inverse is also continuous.

Remark: For the Fourier Transform, we have $||f||_{L^2} = ||\widehat{f}||_{L^2}$, which implies that the Fourier Transform and its Inverse are continuous w.r.t. L^2.

Remark: Let $K : X \mapsto Y$, $||K|| < 1$. Then $(I - K)$ is continuous, invertible and its inverse is continuous (idea of proof: Neumann–series).

2.2 Least squares and minimum norm solution

Definition 2.9
Let $K : X \mapsto Y$, $y \in Y$. Assume we are solving the inverse problem

$$Ku = y \Longleftrightarrow ||Ku - y|| = 0,$$

which might not be solvable. $u \in X$ is a Bestapproximation iff

$$||Ku - y||_Y \leq ||Kv - y|| \forall v \in X.$$

If X and Y are Hilbert spaces, u is called least squares solution.

Note: The least squares solution is not guaranteed to exist (later).

Definition 2.10
Let everything as in 2.9. The least squares solution might not be unique. $u \in X$ is called minimal norm solution (or Moore–Penrose–solution) iff

$$||u|| \leq ||v|| \forall \textit{ Bestapproximations } v.$$

Notation: We denote by

$$\text{ker}(K) = N(K) = \{u \in X : Ku = 0\}$$

the nullspace of K.

Notation: We denote by

$$\text{Range } (K) = R(K) = Im(K) = \{Ku \in Y : u \in X\}$$

the range of K.

Lemma 2.11
The kernel of a continuous linear operator K is always closed. The range is not necessarily closed. Example: Integral operator with kernel in $\mathcal{S}$ on L^2. Ku is in C^∞ which is not closed w.r.t. L^2.

For the following assume X, Y Hilbert spaces.

Lemma 2.12

$$R(K)^\perp = N(K^*)$$
$$\overline{R(K^*)} = N(K)^\perp$$

Theorem 2.13
The following is equivalent:

1. u *is a least squares solution of* $Ku = g$.
2. $K^*Ku = K^*g$ *(Gauss normal equations).*
3. $Ku = Pg$, P *orthogonal projection onto* $\overline{R(K)}$.

Corollary 2.14

1. $Ku = g$ *has a least squares solution iff*

$$g \in R(K) \oplus N(K^*) = R(K) \oplus R(K)^\perp.$$

2. $$\overline{R(K) \oplus R(K)^\perp} = \overline{R(K)} \oplus R(K) = Y$$

   The affine subspace of Y *for which a least squares solution exists is dense in* Y.

3. *If* u *and* v *are LSQ then* $u - v \in N(K)$.
4. *The set of all least squares solutions of* $Ku = g$ *is convex and closed. If it is nonempty, the minimum norm solution is unique.*
5. *Let* g *such that* $Ku = g$ *has at least one LSQ. Then* u *is a MNS iff* u *is an LSQ* $u \in \overline{R(K^*)}$.

2.3 Compact Operators

Motivation: In finite dimensional vector spaces, the MNS can easily be computed using the SVD. We want to do the same for infinite dimensional spaces. Here, the existence of the SVD is not guaranteed. We need an additional restriction for the operator, it needs to be compact. Luckily, most operators in inverse problems are compact. We collect their properties.

Definition 2.15
Let X, Y Banach spaces, $K : X \mapsto Y$ linear.

K is compact iff $\overline{K(B)}$ is compact for all bounded sets $B \subset X$

or equivalently

For all bounded sequences $(u_n) \subset X$ (Ku_n) has a convergent subseries.

Remark 2.16

1. *Let K_1, K_2 compact, $\alpha \in \mathbb{R}$. Then $K_1 + K_2$ and αK_1 are compact.*
2. *Let $\dim R(K) < \infty$ and K linear and continuous. Then K is compact.*
3. *Let K compact. Then K is continuous.*
4. *Let $K : X \mapsto Y$, $L : Y \mapsto Z$, K and L continuous. If K or L is compact, then LK is compact.*

Lemma 2.17
Let $K,\ K_n : X \mapsto Y$, and K_n compact.

$$||K - K_n||_{X,Y} \mapsto 0 \Longrightarrow K \text{ is compact.}$$

Remark: The convergence is with respect to the operator norm, a pointwise convergence for all elements in X is not sufficient! The convergence must be uniform.

Corollary 2.18
Let $k \in L^2(\Sigma \times \Omega)$, Σ and Ω bounded. Then $K : L^2(\Omega) \mapsto L^2(\Sigma)$,

$$(Ku)(x) := \int_\Omega k(x,y)u(y)dy$$

is well–defined (a.e.) and K is compact.

Idea of proof: Restrict to continuous k. Then the integral over k can be discretized (see homework), which defines operators with finite–dimensional range that converge towards K, so K must be compact. Now use that continuous functions are dense in L^2.

Lemma 2.19 *(Riesz' Lemma)*
Let X Banach, $U \subset X$ a closed subspace, $X \neq U$. Then there is an $x \in X$ with the property that $||x|| = 1$, $dist(x, U) \geq \frac{1}{2}$.

Corollary 2.20
If $\dim X = \infty$*, then there exists a sequence* $(x_n) \subset X$ *such that* $||x_n|| = 1$ *and* $dist(x_n, x_m) \geq \frac{1}{2}$ *for* $n \neq m$*.*

Remark: (x_n) has no convergent subsequence.

Corollary 2.21
Let $\dim X = \infty$*. Let* $K : X \mapsto Y$ *compact. Then* K *has no continuous inverse,* $Ku = g$ *is ill–posed.*

Idea: If the inverse of K exists, use the (x_n) from 2.20. Since K is continuous, $y_n = Kx_n$ is bounded, but $x_n = K^{-1}y_n$ has no convergent subsequence.

Now let's look at compact operators in Hilbert spaces.

Theorem 2.22 *(Bessel's inequality)*
Let (u_k) *ONS in* X*,* $u \in X$*. Then*

$$||u||_2^2 \geq \sum_k (u_k, u)^2.$$

Definition 2.23
Let (u_k) *ONS in* X*. Then* (u_k) *is complete iff*

$$X = \overline{span((u_k))}.$$

The span is the set of all finite linear combinations of the u_k*.*

Theorem 2.24 *(Parseval's identity):*
(u_k) *is complete iff*

$$||u||^2 = \sum_k (u, u_k)^2.$$

Remark: Let $U = \overline{span((u_k))}$ for an ONS (u_k). Then the Bestapproximation (orthogonal projection) for $x \in X$ in U is given by

$$u = \sum_k (x, u_k) u_k.$$

2.4 Representation theorem for s.a. Operators

Theorem 2.25
Let $K : X \mapsto X$ *compact,* X *Banach. Then*

$$\dim N(I - K) < \infty.$$

Idea of Proof: 2.20.

Corollary 2.26
If $\lambda \neq 0$ is an Eigenvalue of K, then the Eigenspace has finite dimension.

From now on let X, Y Hilbert (so scalar product and adjoint make sense).

Definition 2.27
$K : X \mapsto Y$ *is selfadjoint (s.a.) iff* $K = K^*$.

Lemma 2.28

1. *Eigenvalues of s.a. operators are real, even over complex vectorspaces.*
2. *Let K s.a. Then the eigenvectors for different eigenvalues are orthogonal.*

Theorem 2.29
Let K compact and s.a. Then it has an Eigenvalue λ with $|\lambda| = ||K||$.
$|\lambda| \geq |\mu|$ for all eigenvalues μ of K.

Example: Integral operator with $k(x, y) = \alpha(x)\alpha(y)$.

Remark: If $k(x, y) = k(y, x)$ then K is s.a. integral operator.

Remark: Every s.a. compact operator K has at least one eigenvalue. We use this to construct a series of nonzero eigenvalues (λ_k) and corresponding orthonormal eigenvectors (u_k) which is complete in the following sense: Every eigenvector of K corresponding to a nonzero eigenvalue is ín $span((u_k))$.

Theorem 2.30
Let $K : X \mapsto X$ compact and s.a. Let (u_k), (λ_k) a complete orthonormal set of eigenvectors and corresponding eigenvalues $\lambda_k \neq 0$.

Let $u \in U$. Then

$$u = \sum_k (u, u_k)u_k + u^{\perp},\ Ku^{\perp} = 0.$$

Remark: For finite dimensional operators this says: Self–adjoint operators (symmetric matrices) posess an ONB of eigenvectors (can be diagonalized by unitary matrices).

Remark: Let $K : X \mapsto X$ compact and s.a. Let $\widetilde{u} \in X$ the (noisy) measurement of $u \in X$, $\widetilde{u} = u + n$, n noise. Let $||n|| \leq \epsilon$. Then by 2.1 we have

$$||Ku - K\widetilde{u}|| \leq ||K||\,\epsilon.$$

However, using the representation theorem, we can do better. Plugging in 2.30 we get

$$||Ku - K\widetilde{u}|| \leq \sum_k |\lambda_k| \, |(n, u_k)|.$$

So we see: The error in (n, u_1) gets amplified by $|\lambda_1| = ||K||$, but the error in $(n, u^\perp)$ has no effect at all, and errors in (n, u_k) for large k are small (since the λ_k converge to zero).

2.5 Singular Value Decomposition

Definition 2.31
Let $K : X \mapsto X$ compact and s.a.
K positive (semi–) definite iff

$$(x, Kx) > (\geq) \, 0 \, \forall x \in X, x \neq 0.$$

Remark 2.32

1. *K positive semidefinite iff K is s.a. and all Eigenvalues $\lambda_k \geq 0$.*
2. *Let $K : X \mapsto Y$. Then K^*K and KK^* are positive semidefinite.*

Theorem 2.33
Let $K : X \mapsto Y$ linear and compact, X, Y Hilbert spaces. Let (u_k), (λ_k) complete systems of Eigenvalues and Eigenvectors of K^*K as in 2.30. Let $\lambda_k = \sigma_k^2$, $\sigma_k > 0$.
Let $v_k = (Ku_k)/||Ku_k||$. Then v_k is a complete ONS of Eigenvectors of KK^* w.r.t. the Eigenvalues $\lambda_k = \sigma_k^2$.
Then we have

$$\forall u \in X : \; Ku = \sum_k \sigma_k (u, u_k) v_k$$

$$\forall v \in Y : \; K^*v = \sum_k \sigma_k (v, v_k) u_k$$

We denote the σ_k as singular values and the v_k, u_k as (left, right) singular vectors (or singular functions in function spaces) of K.

Remarks:

1. The series converges since

$$\sum_k (u, u_k)^2 \leq ||u||^2$$

by Bessel's inequality and the fact that σ_k converges to zero if the sum is infinite.

2. Analytically computing the SVD is usually difficult except in meaningless examples (see below).
3. It's usually a bad idea to use the SVD for computation. This is an analytical tool, in practice we can do better.
4. Most of the SVDs of operators in imaging share the property: u_n for large n (and that means σ_n small) is highly oscillating.

Example:

$$(Ku)(x) := \int_0^x u(t)\,dt,\ K : L^2(\Omega) \mapsto L^2(\Omega)$$

Leads to the ODE

$$u_k''(x) = -\frac{1}{\sigma_k^2} u_k(x),\ u_k(1) = 0,\ u_k'(0) = 0.$$

The singular values are

$$\sigma_k = \frac{2}{\pi}\frac{1}{(2n+1)} = O(\frac{1}{n}).$$

The singular functions are

$$u_k(t) = \sqrt{2}\cos\frac{t}{\sigma_n},\ v_k(t) = \sqrt{2}\sin\frac{t}{\sigma_n}$$

Corollary 2.34
Let $K : X \mapsto Y$ compact. Let σ_i, u_i, v_i the singular values and vectors of K. The set of all least squares solutions of $Ku = g$ is given by

$$\sum_k \frac{1}{\sigma_k}(g, v_k)u_k + u^\perp$$

where $u^\perp$ is the nullspace of K.
The Minimum Norm Solution of $Ku = g$ is given by

$$\sum_k \frac{1}{\sigma_k}(g, v_k)u_k.$$

If the sum does not converge, no least squares solution and no Minimum Norm Solution exist.

Definition 2.35 *Order (degree) of ill-posedness, see also 2.3*
An inverse problem $Ku = g$ is

- *mildly ill–posed of order α, iff*

$$\exists\, C : \sigma_n \leq C\frac{1}{n^\alpha}\, \forall\, n.$$

- *severely ill–posed, if no such C and α exist.*

Remark: This definition is compatible with the old one.

Definition 2.36 *Generalized Inverse, Pseudo Inverse, Moore–Penrose Inverse*
Let $K : X \mapsto Y$, X, Y Hilbert–spaces. The operator

$$K^+ :\ R(K) \oplus N(K^*) \mapsto X,\ K^+ g := u^+$$

where u^+ is the Minimum–Norm–Solution of $Ku = g$, is called Generalized (Pseudo, Moore–Penrose) Inverse.

Chapter 3

Regularization

3.1 Definition

Remark: Let (σ_k, u_k, v_k) a singular system for the operator K. Then for $\alpha > 0$ the family of operators

$$K_\alpha^+(g) := \sum_{\sigma_k \geq \alpha} \frac{1}{\sigma_k}(g, v_k)u_k$$

is called truncated singular value decomposition regularization.

Definition 3.1 *Regularization*
Let $K : X \mapsto Y$. A family of operators K_α^+, $\alpha \in \mathbb{R}^+$. is called a regularization of K^+ iff there is a function (parameter choice rule) $\alpha(\delta, g^\delta)$, such that

$$\forall \delta > 0,\ g \in R(K) \oplus N(K^*) : \lim_{\delta \mapsto 0} \sup_{||g - g_\delta|| \leq \delta} ||K_{\alpha(\delta, g_\delta)} g_\delta - K^+ g|| = 0.$$

If α does not depend on g_δ, α is called a priori.

Corollary 3.2
K_α^+ is a regularization iff for all (g_n) in Y with $||g_n - g|| \leq \delta_n$, $\delta_n \mapsto 0$

$$K^+_{\alpha(\delta_n, g_n)} g_n \mapsto K^+ g.$$

Remark: If α does not depend on δ, K^+ is continuous.

3.2 Tikhonov–Phillips Regularization

Definition 3.3 *Tikhonov–Phillips Regularizer*
f^α is the Tikhonov–Regularizer of $Kf = g$ iff

$$f^\alpha = \arg\min_f ||Kf - g||_Y^2 + \alpha^2 ||f||_X^2.$$

Lemma 3.4
Let $K : X \mapsto Y$ continuous, X, Y Hilbert spaces. Then Tikhonov–Regularization is well–defined and continuous. We have

$$f_\alpha^+ = K_\alpha^+ g = (K^*K + \alpha^2 I)^{-1} K^* g.$$

K_α^+ is continuous.

Remark: Let K compact, (σ_k, u_k, v_k) singular system of K. Then

$$f_\alpha^+ = \sum_k \frac{\sigma_k}{\alpha^2 + \sigma_k^2} (g, v_k)\, u_k.$$

3.3 Regularization and the SVD

Lemma 3.5 *Let $g_\alpha : \mathbb{R}^+ \mapsto \mathbb{R}^+$. Assume*

1. $$g_\alpha(\sigma) \to_{\alpha \to 0} \frac{1}{\sigma}.$$

2. $$\forall\, \alpha \exists\, C_\alpha : \sup_\sigma g_\alpha(\sigma) \leq C_\alpha.$$

3. $$\exists\, C : \sup_{\alpha,\sigma} \sigma g_\alpha(\sigma) \leq C.$$

Let $K : X \mapsto Y$ compact with singular system (σ_k, u_k, v_k). Then

$$K_\alpha^+ : Y \mapsto X,\ K_\alpha^+ g = \sum_k g_\alpha(\sigma_k)(g, v_k) u_k$$

is well–defined and continuous, $||K_\alpha^+|| \leq C_\alpha$. For $g \in R(K) \oplus N(K^*)$ we have

$$K_\alpha^+ g \mapsto_{\alpha \mapsto 0} K^+ g.$$

Remark: Tikhonov fits into this scheme setting

$$g_\alpha(\sigma) = \frac{\sigma}{\alpha^2 + \sigma^2},\ C_\alpha = \frac{1}{2\alpha},\ C = 1.$$

Remark: Truncated SVD fits into this scheme setting

$$g_\alpha(\sigma) = \begin{cases} 0, & \sigma \leq \alpha \\ \frac{1}{\sigma}, & \sigma > \alpha. \end{cases},\ C_\alpha = \frac{1}{\alpha},\ C = 1.$$

Remark: g is in the domain of K^+ iff

$$\sum_k \frac{1}{\sigma_k^2}(g, v_k)^2 < \infty$$

(Picard–criterion).

Corollary 3.6 *Let K_α^+ as above. Let the parameter choice rule α such that*

$$\alpha(\delta, g_\delta) \to_{\delta \to 0} 0,\ \delta\, ||K_{\alpha(\delta, g_\delta)}|| \to 0.$$

Then $(\alpha,\ K_\alpha^+)$ is a regularization of $Kf = g$.

Remark: If $g(\alpha) := ||K_\alpha^+||$ is strictly decreasing (as for Tikhonov, Truncated SVD), setting

$$\alpha(\delta) := g^{-1}(\frac{1}{\sqrt{\delta}})$$

satisfies the condition.

Example: The Lavrentiev Regularization is defined by

$$K_\alpha^+ g := \sum_k \frac{1}{\sigma_k + \alpha}(g, v_k)u_k.$$

If the range of K is positive definite and dense in Y, the Lavrentiev regularization can be computed via

$$(K + \alpha I)(K_\alpha^+ g) = g.$$

Lavrentiev regularization is a regularization with $C_\alpha = \frac{1}{\alpha}$ (for a proper choice of α).

Example: Iterative Regularization via iteratively solving the normal equation in the range of K^*. As an example we take Landweber (but conjugate gradient and others are also in use). Landweber iterations are defined according to

$$u^{(n+1)} = u^{(n)} - \omega K^*(Ku^{(n)} - g),\ u^{(0)} = 0,\ 0 < \omega < \frac{2}{\sigma_1}.$$

Note that all iterates are in the range of K^*.

Let $n = \lfloor \frac{1}{\alpha} \rfloor$. Then

$$K_\alpha^+ g = u^{(n)}$$

is a regularization (for a proper choice of α). Note that the idea here is early stopping: For a fixed δ, we choose an n where we stop the iteration, we do not let it converge towards a solution of the normal equation.

Example: Iterated Regularization

Here, we try to improve an existing approximation $u^{(n)}$ to $K^+ g$. Let

$$u^{(n+1)} = u^{(n)} + d^{(n)}.$$

We want $u^{(n+1)}$ to solve our equation, so we get the inverse problem

$$K(u^{(n)} + d^{(n)}) = g \Longrightarrow K d^{(n)} = g - K u^{(n)}.$$

Using any regularization scheme (e.g. Tikhonov), we determine an approximate solution $d_\alpha^{(n)}$ and set

$$u^{(n+1)} := u^{(n)} + d_\alpha^{(n)}.$$

Under conditions, for fixed n,

$$K_\alpha^+ g := u^{(n)}$$

is a regularization scheme. You will do this explicitly for Tikhonov in the homeworks.

Example: Discretizing by point evaluation

A simple regularization idea could be: Simply discretize the problem by point evaluations and use techniques from numerical analysis and numerical linear algebra to solve the inverse problem. This is very problematic. We look at the following example:

Let

$$Ku(t) := \int_0^\pi k(t,s) u(s) ds$$

an integral operator with continuous kernel function on the interval $[0, \pi]$ as before.

Fix a discretization parameter N and let (x_k), $k = 0 \dots N$, equidistant in $[0, \pi]$. Let $h = \frac{1}{N}$. Then we have

$$(Ku)(x_j) = \int_0^\pi k(t,s) u(s)\, dx \sim h \sum_l D_l k(x_j, x_l) u(x_l)$$

where the D_l are chosen according to the scheme we use for approximating the integral, we have e.g.

$$(D_k) = (\frac{1}{1}, 1, \ldots, 1, \frac{1}{2})$$

for the trapezoidal rule and

$$(D_k) = \frac{1}{3}(1, 4, 2, 4, 2, \ldots, 4, 2, 4, 1)$$

for the Simpson rule. At this point, both should deliver very similar results, and if one is preferrable, it should be Simpson due to its higher order.

Assume that $g(x_k) = (Ku)(x_k)$ were measured. Then setting

$$U := \begin{pmatrix} u(x_0) \\ \vdots \\ u(x_n) \end{pmatrix} \in \mathbb{R}^{n+1},\ G := \begin{pmatrix} g(x_0) \\ \vdots \\ g(x_n) \end{pmatrix} \in \mathbb{R}^{n+1},$$

$$\widetilde{K} := (k(x_j, x_l)) \in \mathbb{R}^{(n+1)\times(n+1)},\ D = diag((D_k)) \in \mathbb{R}^{(n+1)}$$

we have approximately

$$G \sim h\widetilde{K}Du.$$

Thus, the discretized inverse problem is

$$h\widetilde{K}D_{\text{Trapez}}U^{\text{Trapez}} = G$$

for a solution using the trapezoidal rule (D is chosen for trapezoidal rule) and

$$h\widetilde{K}D_{\text{Simpson}}U^{\text{Simpson}} = G.$$

However, assuming that $\widetilde{K}$ is invertible, this yields

$$U^{\text{Trapez}} = D^{-1}_{\text{Trapez}}D_{\text{Simpson}}U^{\text{Simpson}}.$$

This is valid independent of N (and thus h), and means that the two solutions cannot converge to the same limit function. In fact, since D_{Trapez} is almost the identity matrix, and D_{Simpson} is highly oscillating, at least one of the two vectors must be highly oscillating. We give a programming example that shows this in practice.

The result of this discussion is that point evaluations are an extremely bad choice for discretization.

Motivated by this discussion, and with truncated SVD in mind, we propose moment methods:

Example: Moment–Methods, Regularization by projection

Look for a vector u^+ in the span of $(u_1, \dots u_n)$ such that

$$\varphi_k(Ku^+ - g) = 0,\ \varphi_k(v) := (v_k, v).$$

If (σ_k, u_k, v_k) is a singular system, then this is equivalent to Truncated SVD.

For other choices, this discretization is a regularization under relaxed assumptions, mainly requiring that

$$\overline{\text{span } u_k} = X$$

and

$$\overline{\text{span } v_k} \supset \overline{R(K)}.$$

Let $Ku = g$, $||g - g_\delta|| \leq \delta$. If K^{-1} is continuous, we have

$$||K^{-1}g_\delta - u|| \leq ||K^{-1}||\,||g_\delta - g|| \leq ||K^{-1}||\delta = O(\delta).$$

If K^+ is discontinuous, we need to use regularization and have that $||K_\alpha^+||$ is unbounded (exercises). In this case we have

$$||K_\alpha^+ g_\delta - K^+ g|| \leq \underbrace{||K_\alpha^+ g_\delta - K_\alpha^+ g||}_{\leq ||K_\alpha^+||\,\delta} + \underbrace{||K_\alpha^+ g - K^+ g||}_{=:II}.$$

Corollary 3.7 *Let $Ku = g$ an inverse problem, $||g - g_\delta|| \leq \delta$, K^+ discontinuous. Then*

$$||K_\alpha^+ g_\delta - K^+ g|| \neq O(\delta).$$

So this is the price we have to pay for regularization: The error for the inverse problem does not go to zero with the same rate as for the continuous problem (1).

For term II we know that it converges to zero (Picard criterion). However, that convergence can be arbitrarily slow, and in fact, no general rate can be given for discontinuous K^+.

Theorem 3.8
Assume (in the usual setting)

$$||K_\alpha^+ g - K^+ g|| \leq f(\alpha)\,\forall\, g \in D(K^+),\ f(\alpha) \to_{\alpha \to 0} 0.$$

Then K^+ is continuous.

Assume that (σ_k, u_k, v_k) is a singular system for K, and g_α defines an SVD–based regularization scheme. Then

$$||K_\alpha^+ g - K^+ g||^2 = \sum_k \underbrace{(\sigma_k g_\alpha(\sigma_k) - 1)^2}\ \underbrace{\frac{1}{\sigma_k^2}(g, v_k)^2}_{\sum<\infty \text{ (Picard)}} .$$

The first term is bounded, but does not go to zero (fix α, let $\sigma \to 0$). The sum then converges, but arbitrarily slow. In order to get a rate, we have to make the assumption that the scalar products (g, v_k) drop fast. This is not an unusual condition, in Numerical Analysis you often proved error estimates for difference schemes for computing the derivative, based on the assumption that the derivative was one more time differentiable.

Definition 3.9 *(source conditions)*
u^+ satisfies a source condition of order μ iff u^+ is in the Range of $(K^*K)^\mu$.

We allow for μ not to be an integer (see below).

Theorem 3.10
Let $u^+ = (K^*K)^\mu v$, $v \in X$. Then

$$||K_\alpha^+ g - K^+ g||^2 \leq \underbrace{\max_{\sigma \in (0, \sigma_1]} |g_\alpha(\sigma)\sigma^{2\mu+1} - \sigma^{2\mu}|^2}_{=:\varphi_\mu(\alpha)^2} ||v||^2.$$

Now we have

$$||K_\alpha^+ g_\delta - K^+ g|| \leq C_\alpha \delta + \varphi_\mu(\alpha) \cdot ||v||.$$

Of course, we want to choose $\alpha(\delta)$ such that the right hand side (or at least the order with respect to δ) is minimal. Then, the parameter choice is called optimal (order optimal).

Example: Truncated SVD

The parameter choice

$$\alpha(\delta) := \left(\frac{\delta}{2\mu||v||}\right)^{\frac{1}{2\mu+1}}$$

is optimal. The rate achieved is $O(\delta^{2\mu/(2\mu+1)})$. The rate gets better with $\mu \to \infty$, approaches δ, but is always smaller than δ.

Example: Lavrentiev

For $\mu \geq \frac{1}{2}$, the rate is $O(\sqrt{\delta})$. For $\mu < \frac{1}{2}$, the rate is $O(\delta^{2\mu/(2\mu+1)})$. The best order is achieved for $\mu = \frac{1}{2}$ and does not improve for higher μ.

Example: Tikhonov

For $\mu \geq 1$, the rate achieved is $O(\delta^{2/3})$. For $\mu < 1$, the rate is $O(\delta^{2\mu/(2\mu+1)})$. The best order is achieved for $\mu = 1$.

Definition 3.11 *If the convergence rate increases up to a value μ_0 and then stays, we call $2\mu_0$ the qualification of a regularization.*

In a sense, this defines an optimal source space for the regularization scheme.

Example: TSVD has qualification ∞, Tikhonov 2, Lavrentiev 1.

Morozovs Discrepancy Principle:

Assume again that $||g_\delta - g|| \leq \delta$ and that the range of K is dense in Y. Since g_δ has an error bound of δ, it does not make sense to satisfy the data any better. So:

α is a good choice for a regularization parameter if

$$||(K(K_\alpha^+ g_\delta) - g_\delta|| \sim C\delta.$$

So to check a chosen regularization parameter, compute the norm of the defect. If it is on the order of δ, fine. If it is smaller, choose a higher value of α. If it is bigger, choose a smaller value of α.

There are many (theoretical) reasons why this yields a good choice for α. As an example, we show that for truncated SVD and an optimal choice of α the equation is correct.

Remark: All linear regularization methods can be analyzed using the SVD, and they all share the same property: For fixed α, the lower–order terms (small k) are almost the same in K_α^+ and K^+, the higher order terms are damped (their absolute value is smaller in the regularization, or vanishes as in the truncated SVD). Their only difference is the rate at which the terms get smaller.

Remark: Assuming that u_k is more and more oscillating (which is the case at least for the antiderivative, but can also be proven for many other operators like the Radon transform) this implies that we take away highly oscillating terms from the result, effectively smoothing the result.

Remark: The smoother a function is, the faster its Fourier coefficients vanish. Let e.g. $g : [-\pi, \pi] \mapsto \mathbb{R}$ n times differentiable, and a_k its Fourier coefficients.

Then

$$g^{(n)}(x) = \sum_k i^n k^n a_k e^{ikx}.$$

Since that function exists, we have

$$\sum_k (k^n a_k)^2 < \infty$$

and the a_k decay at least as $\frac{1}{n}$. This implies that for smooth functions, their share in highly oscillating functions vanishes rapidly. and is a hint that taking away the high–frequency parts is not too bad.

Remark: The SVD is often used to analyze image or signal denoising, where the operator is simply the identity. This is not strictly correct for the standard scalar product, since identity is not a compact operator, and in fact it posesses no SVD. However, choose the Fourier system

$$X = Y = L^2([-\pi, \pi]),\ u_k = v_k = \frac{1}{2\pi} e^{ikx},\ \sigma_k = 1.$$

Then obviously

$$Ku = \sum_k \sigma_k (u, u_k) v_k.$$

This system shares many properties of the SVD, including that the u_k are more and more oscillating with k. However, it is not a singular system (e.g. the σ_k do not go to zero).

Remark: For denoising ($K = I$) we have for Tikhonov

$$(K^*K + \alpha^2 I) u_\alpha^+ = K^* g \Longrightarrow u_\alpha^+ = \frac{1}{1+\alpha^2} g.$$

So classical Tikhonov simply multiplies the data with a constant. Choosing a different (semi-) norm for the penalty term helps. Assuming that u is n times differentiable and choosing

$$u_\alpha^+ \arg\min_u ||Ku - g||^2 + \alpha^2 ||u(n)||^2$$

we have

$$u_\alpha^+ = \sum \frac{1}{1 + \alpha^2 k^{2n}} (g, v_k) u_k.$$

Here, it makes sense to take n fractional. $n = \frac{1}{2}$ is a common choice.

Conclusion: Choosing an appropriate norm is important.

Conclusion: All linear regularizations share very similar properties.

In order to extend our capabilities, we need to turn to nonlinear regularizers K_α^+. A common choice is

$$u_\alpha^{TV} = \arg\min_u ||Ku - g||^2 + \alpha||\nabla u||_1$$

(Total variation). It penalizes the non–monotonicity of a function. We investigate this in a numerical experiment, but do not analyze it at this point, rather we turn to applications now.

Chapter 4

Fourier Transform and Distributions

In the preceding chapter, we saw that discrete and analytic Fourier Transform play an important role in image and signal analysis. We collect some basic definitions and properties, starting with the analytic transform. Most of this can be found in Forster, Analysis III, and is covered in the lecture Analysis III.

4.1 Fourier Transform

Definition 4.1 *Analytic Fourier Transform*
Let $f \in L^1(\mathbb{R}^n)$. Then the Fourier Transform $\widehat{f}$ of f is defined as

$$\widehat{f}(\xi) := (2\pi)^{-n/2} \int_{\mathbb{R}^n} f(x) e^{-ix\xi}\, dx.$$

Lemma 4.2
Always, let $f \in L^1(\mathbb{R}^n)$.

1. $$|\widehat{f}(\xi)| \leq (2\pi)^{-n/2} ||f||_1.$$

2. *$\widehat{f}$ is continuous.*

3. *Let f differentiable.*
   $$\widehat{f}'(\xi) = i\xi \widehat{f}(\xi).$$
   Let $xf(x)$ in L^1.
   $$\widehat{xf(x)}(\xi) = i \frac{d}{d\xi} \widehat{f}(\xi).$$

4. *Let f sufficiently differentiable, $\alpha \in \mathbb{N}_0^n$.*

$$\begin{aligned} D^\alpha f &:= \left(\frac{d}{dx_1}\right)^{\alpha_1} \cdots \left(\frac{d}{dx_n}\right)^{\alpha_n} f \\ x^\alpha &:= x_1^{\alpha_1} \cdot \ldots \cdot x_n^{\alpha_n} \\ |\alpha| &:= \alpha_1 + \ldots + \alpha_n \\ \widehat{D^\alpha f}(\xi) &= i^{|\alpha|} \xi^\alpha \widehat{f}(\xi) \\ \widehat{x^\alpha f(x)}(\xi) &= i^{|\alpha|} (D^\alpha \widehat{f})(\xi). \end{aligned}$$

There was a typo in the lecture for the last formula.

5. *Let $\lambda \neq 0$, $f_\lambda(x) := f(\lambda x)$.*

$$\widehat{f_\lambda}(\xi) = |\lambda|^{-n} \widehat{f}(\frac{\xi}{\lambda}).$$

6. *Let $a \in \mathbb{R}^n$, $f_a(x) := f(x + a)$.*

$$\widehat{f_a}(\xi) = e^{ia\xi} \widehat{f}(\xi).$$

7. *The Fourier Transform of $e^{-x^2/2}$ is $e^{-\xi^2/2}$.*
Idea of proof: Both solve the same initial value problem for an ODE.

8. *Let $f(x) = \chi_{[-1,1]}(x)$ and*

$$\operatorname{sinc}(x) := \begin{cases} 1, & \text{if } x = 1 \\ \frac{\sin x}{x}, & \text{otherwise.} \end{cases}$$

Then

$$\widehat{f}(x) = \sqrt{\frac{2}{\pi}} \operatorname{sinc}(x).$$

Let $f : \mathbb{R}^n \mapsto \mathbb{R}$, $f(x) := \chi_{[-1,1]^n}(x)$. Then

$$\widehat{f}(\xi) = \left(\frac{2}{\pi}\right)^{n/2} \operatorname{sinc}(\xi_1) \cdot \ldots \cdot \operatorname{sinc}(\xi_n).$$

Since $\widehat{f}$ is not in L^1, the Fourier transform is not an operator from L^1 to L^1.

Theorem 4.3 *Fourier Transform on L^2*
The Fourier Transform can be extended continuously as an operator from L^2 to L^2.

Theorem 4.4 *Inverse Fourier Transform*
Let $f \in L^2(\mathbb{R}^n)$ or $f, \widehat{f} \in L^1(\mathbb{R}^n)$. Then the inverse Fourier Transform of f is defined as

$$\widetilde{f}(\xi) := \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}^n} f(x) e^{ix\xi}\, dx$$

and

$$\widetilde{\widehat{f}} = \widehat{\widetilde{f}} = f.$$

Note: In the following, we will often use the Fourier transform of L^2-functions and use their integral representation, although the integral does not converge in the usual sense. E.g. for the sinc-function:

$$\widetilde{f}(x) = (2\pi)^{-\frac{1}{2}} \int_{\mathbb{R}} \sqrt{\frac{2}{\pi}} \frac{\sin x}{x} e^{ix\xi} d\xi = \chi_{[-1,1]}(x).$$

This is to be understood in the sense of Theorem 4.3: Represent sinc as the limit of a sequence of functions f_n in $L^2 \cap L^1$. Then $\widehat{f_n} \to F$, and we define F as the value of the integral above.

Theorem 4.5 *(Parseval)*

1. *Let $f, g \in L^1(\mathbb{R}^n)$. Then*

$$\int_{\mathbb{R}^n} f \cdot \widehat{g} = \int_{\mathbb{R}^n} \widehat{f} \cdot g.$$

2. *Let $f, g \in L^2(\mathbb{R}^n)$. Then*

$$\begin{aligned} (\widehat{f}, g)_{L^2} &= (f, \widetilde{g})_{L^2} \\ (\widehat{f}, \widehat{g})_{L^2} &= (f, g)_{L^2} \\ ||\widehat{f}||_2 &= ||f||_2. \end{aligned}$$

Parseval confirms a remark that we made at the beginning: The operator norm of the Fourier Transform and its inverse is 1. Both are continuous, thus the inverse problem of Fourier Transformation is well–posed.

We now turn to the connection between Fourier Transform and convolutions (see Definition 2.6).

Remark: $f * g = g * f$.

Remark: $supp(f * g) \subset (supp\, F) + (supp\, G)$.

Theorem 4.6 *(convolution theorem)*
Let $f, g \in L^1(\mathbb{R}^n)$. Then

$$\widehat{f * g}(\xi) = (2\pi)^{n/2} \widehat{f}(\xi)\widehat{g}(\xi).$$

Example: Let $f(x) = \chi_{[-1,1]}(x)$. Then

$$(f * f)(x) = \begin{cases} 0, & \text{if } |x| > 2 \\ 2 - |x|, & \text{otherwise.} \end{cases}.$$

$$\widehat{(f * f)}(\xi) = \sqrt{2\pi}\widehat{f}(\xi)^2 = \sqrt{2\pi}\frac{2}{\pi}\operatorname{sinc}(\xi)^2.$$

Remark: This gives us a way of defining the convolution of two L^2–functions. Let $f, g \in L^2$. Then $\widehat{f}$, $\widehat{g}$ are in L^2 and $\widehat{f}\widehat{g}$ is in L^1, so its inverse Fourier Transform exists. The only way of defining the convolution without breaking the convolution theorem is

$$f * g := \widetilde{(\widehat{f} \cdot \widehat{g})}.$$

Remark 4.7
Let $f, D^\alpha f \in L^1$. Then

$$|\widehat{D^\alpha f}(\xi)| = |\xi^\alpha|\,|\widehat{f}(\xi)|$$

which implies

$$|\widehat{f}(\xi)| \leq \frac{||D^\alpha f||_1}{|\xi^\alpha|}.$$

For f in C^∞, all derivatives in L^1, f decays faster than one over any polynomial.

4.2 Distributions or generalized functions

Always let $\Omega \subset \mathbb{R}^n$ open and $\neq \emptyset$. Usually we have $\Omega = \mathbb{R}^n$.

Let

$$D(\Omega) = C_0^\infty(\Omega) = \{f \in C^\infty(\Omega^n) : \exists K \subset \Omega \text{ compact} : supp\, f \subset K\}.$$

We call $D(\Omega)$ the space of test function and set $D = D(\mathbb{R}^n)$.

The function

$$f(x) := \begin{cases} e^{-\frac{1}{1-||x||^2}}, & \text{if } ||x|| \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

is in D. By scaling, shifting, convolving f with functions of compact support we can create a zoo of functions in D.

We view $D(\Omega)$ as a topological space, generated by the semi–norms

$$|f|_{\alpha,\infty} := \sup_{x\in K} |(D^\alpha f)(x)|.$$

For a sequence (φ_n) in $D(\Omega)$, $\varphi_n \to \varphi$ iff

1. $\exists\, K \subset \Omega$ compact : $supp\, \varphi_k \subset K\, \forall\, k$
2. $D^\alpha \varphi_k \mapsto D^\alpha \varphi$ uniformly

The topology is extremely strong.

Definition 4.8 *(Distributions)*
We define the space of distributions over the space of test functions $D(\Omega)$ by

$$D(\Omega)' := \{T : D(\Omega) \mapsto \mathbb{R},\, T \textit{ linear cotinuous}\}.$$

T is continuous iff

$$\varphi_n \to \varphi \Rightarrow T(\varphi_n) \to T(\varphi).$$

Lemma 4.9 *(Embedding of L^1_{loc} in $D(\Omega)'$)*
Let

$$L^1_{loc}(\Omega) = \{f : f \in L^1(K)\, \forall\, K \subset \Omega \textit{ compact}\}.$$

E.g. all continuous functions are in L^1_{loc}.

Let $f \in L^1_{loc}$. Then

$$T_f : D(\Omega) \mapsto \mathbb{R},\, T_f(\varphi) := \int_\Omega f(x)\varphi(x)\, dx$$

is in $D(\Omega)'$.

Important note: We will identify the function f with its corresponding distribution T_f. In this sense, $L^1_{loc} \subset D(\Omega)$.

Lemma 4.10 *(Delta–Distribution)*
Let

$$T : D(\Omega) \mapsto \mathbb{R},\, T\varphi := \varphi(0).$$

Then $T \in D(\Omega)'$. $\delta := T$ is called Dirac–Distribution or Delta–Distribution.

Remark: δ cannot be represented as a function f such that $T_f = \delta$. But defining

$$f_k := \begin{cases} \frac{1}{k}, & \text{if } |x| < \frac{1}{2k} \\ 0, & \text{otherwise.} \end{cases}$$

we obviously have

$$T_{f_k}(\varphi) \to \varphi(0) = \delta(\varphi),$$

so with pointwise convergence we have

$$f_k \equiv T_{f_k} \to \delta.$$

δ is the limit of functions that converge to 0 for $x \neq 0$, to ∞ for $x = 0$, and have integral 1, so δ is often viewed (in particular in the physics and engineering world) as generated by the "function"

$$\delta(x) := \begin{cases} \infty, & \text{if } x = 0 \\ 0, & \text{otherwise.} \end{cases}, \int_{\mathbb{R}^n} \delta(x)\, dx = 1.$$

Alternatively, the δ distribution can be approximated by smooth functions scaling the bump function from 4.8.

Corollary 4.11
Let α a Multiindex, $\overline{x} \in \Omega$ fixed.

$$T_{\alpha,\overline{x}}(\varphi) := (D^\alpha \varphi)(\overline{x})$$

is in $D(\Omega)'$.

Corollary 4.12
Let M a k–dimensional manifold and σ the k–dimensional induced measure. Then

$$T_M(\varphi) := \int_M \varphi(x) d\sigma(x)$$

is in $D(\Omega)'$.

Remark: Of course at this point we have the Radon transform in mind, it is a distribution.

Definition 4.13 *(Derivative of a distribution)*
Let $T \in D(\Omega)'$, α a multiindex. Then the derivative $D^\alpha T$ of T is defined as

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

Note: This definition is motivated by

$$(T_f)' \equiv f' \equiv T_{f'}.$$

Note: All distributions are differentiable. In particular, all functions in L^1_{loc} are differentiable (but of course, their derivatives might be distributions which cannot be represented as functions). We extend the definition of differentiability beyond weakness.

Example: For $f(x) = \operatorname{sgn}(x)$, we have $f'(x) = 2\delta(x)$.

Definition 4.14 *(Convolution of distribution and test function)*
Let $g \in D$, $g_x(y) := g(x - y)$ and $T \in D'$. Then

$$(T * g) : \mathbb{R}^n \mapsto \mathbb{R},\ (T * g)(x) := T(g_x).$$

Again, this is motivated by

$$T_f * g \equiv f * g \equiv T_{f*g}.$$

$T * g$ does not necessarily have compact support (take $T = 1$).

Example: $(\delta * g)(x) = g(x)$.

Note: Let $h_z = h(z + x)$. Then

$$\begin{aligned}(f * h_z)(y) &= \int_{\mathbb{R}} f(x)h(z + y - x)\,dx \\ &= (f * h)(z + y) = (f * h)_z(y).\end{aligned}$$

Thus: Viewing $f * g$ as an operator on g for fixed f, the result is shifted by z if the argument is shifted by z. We call this property shift–invariant. It is very common in imaging: shifting the object is shifting the image (neglecting boundary effects).

It is easily seen that any shift–invariant operator is a convolution (in the distributional sense). The kernel function can be recovered by applying the operator to the δ distribution.

A common application for this is microscopy. Typically, the outcoming images are blurred. Since they are shift–invariant (again discarding boundary effects), they are in fact convolutions. The kernel function f can be recovered by taking the image of

an approximate δ–distribution, a bright single spot. In order to recover the clear image g from a blurred image g_b we have to solve the inverse problem (deconvolution problem)

$$f * g = g_b.$$

Note: The Fourier Transform of a distribution in D is not easily defined. Using the same approach as above, we need to have

$$\widehat{T_f} = \widehat{f} = T_{\widehat{f}}.$$

So the only viable definition is

$$\widehat{T}(\varphi) := T(\widehat{\varphi}).$$

However, this makes no sense: A consequence of Paley–Wiener is that if φ and $\widehat{\varphi}$ have compact support, then $\varphi = 0$. So if $0 \neq \varphi \in D \Rightarrow \widehat{\varphi} \notin D$, and the definition makes no sense. We need to change our space of test functions such that the Fourier Transform of any test function is a test function.

4.3 Tempered Distributions

In order to be able to define Fourier Transforms, we change our space of test functions to the Schwartz space.

Definition 4.15 *(Schwartz space, tempered distributions)*
The functions space

$$\mathcal{S} = \{f \in C^\infty(\mathbb{R}^n) : \forall \alpha, \beta \in \mathbb{N}_0 \, \exists C_{\alpha,\beta} : |x^\alpha (D^\beta f)(x)| \leq C_{\alpha,\beta} \, \forall \, x \in \mathbb{R}^n\}$$

is the Schwartz space of rapidly decaying functions. It is the space of test functions for $\mathcal{S}'$, the space of tempered distributions, as in 4.8.

Using that D is dense in S, all remarks for D hold for $\mathcal{S}$.

Note: $D \subsetneq \mathcal{S}$: Obviously all functions from D are in $\mathcal{S}$. $\varphi(x) := \exp(-||x||^2/2)$ is in $\mathcal{S}$, but not in D.

Definition 4.16
Let $\varphi \in \mathcal{S}$, $T \in \mathcal{S}'$. Then the Fourier Transform $\widehat{T}$ and the Inverse Fourier Transform $\widetilde{T}$ of T are defined as

$$\widehat{T}(\varphi) := T(\widehat{\varphi})$$
$$\widetilde{T}(\varphi) := T(\widetilde{\varphi})$$

Due to 4.7 $\widehat{\varphi} \in \mathcal{S}$, so this is well–defined.

Of course, we have $\widetilde{\widehat{T}} = T$ (this is a typo in the video).

Example:

$$\widehat{\delta}(\varphi) = \delta(\widehat{\varphi}) = \widehat{\varphi}(0) = \int_{\mathbb{R}^n} (2\pi)^{-n/2} \varphi(x)\, dx = T_{(2\pi)^{-n/2}}.$$

So the Fourier Transform of δ is the constant function $(2\pi)^{-n/2}$:

$$\widehat{\delta} = (2\pi)^{-n/2} \cdot 1 = \widetilde{\delta}$$

$$\widehat{1} = (2\pi)^{n/2} \delta = \widetilde{1}$$

Neglecting the fact that δ has no function representation and inserting into the inverse Fourier Transform, we find for $n = 1$:

$$\delta(x) = (2\pi)^{-1/2} \int_{\mathbb{R}} \widehat{\delta}(\xi) e^{ix\xi} d\xi = \frac{1}{2\pi} \int_{\mathbb{R}} e^{ix\xi}\, d\xi.$$

The integral on the right hand side does not seem to make any sense at all, it does not converge when its limits go to infinity. However, in a distributional sense, the integral is fine. To give you the idea, let us approximate the integral by

$$\frac{1}{2\pi} \int_{-R}^{R} e^{ix\xi}\, d\xi = \frac{R}{\pi} \operatorname{sinc}(Rx) =: \delta_R(x).$$

This can easily be computed observing that

$$\sqrt{2\pi}\widehat{\delta_R} = \chi_{[-R,R]}.$$

We would expect that $\delta_R \to \delta$, or $\delta_R(\varphi) \to \delta(\varphi)$ (all for $R \to \infty$).

Using the convolution theorem, we have

$$(\varphi * \delta_R)(\xi) = \sqrt{2\pi}\widetilde{\widehat{\varphi} \cdot \widehat{\delta}_R}(\xi) = \int_{-R}^{R} \widehat{\varphi}(x) e^{ix\xi} dx \to \varphi(\xi)$$

and thus

$$T_{\delta_R}(\varphi) = \int_{\mathbb{R}} \delta_R(x)\varphi(x)\, dx = (\varphi * \delta_R)(0) \to \varphi(0)$$

and the integral converges in a distributional sense.

Theorem 4.17
Let $f \in S$. The Hilbert Transform from 2.8 has the property

$$\widehat{Hf}(\xi) = -i\, sign(\xi)\widehat{f}(\xi).$$

Corollary 4.18 ´
The Hilbert transform is invertible and continuous with operator norm 1. Its inverse is continuous with operator norm 1.

Idea of proof: Use Parseval and the inverse Fourier Transform.

Example: Fourier Transform of radially symmetric functions

Definition 4.19 *(Bessel Function of the first kind)*
Let $F : \mathbb{C} \mapsto \mathbb{C}$,

$$f(z) := e^{x(z-\frac{1}{z})\frac{1}{2}},\ x \in \mathbb{R}\ \textit{fixed}.$$

The coefficents $J_n(x)$ in the Laurent series

$$f(z) = \sum_{n=-\infty}^{\infty} J_n(x) z^n$$

are the Bessel functions of order n of the first kind.

Lemma 4.20 *(Integral representation of the Bessel function)*

$$J_n(x) = \frac{1}{2\pi}\int_0^{2\pi} e^{ix\sin\varphi - in\varphi} d\varphi = \frac{i^{-n}}{2\pi}\int_0^{2\pi} e^{ix\cos\varphi - in\varphi} d\varphi.$$

Lemma 4.21
J_n satisfies the Bessel differential equation

$$x^2 f''(x) + x f'(x) + (x^2 - n^2) f(x) = 0.$$

Remark: The Bessel differential equation has a second base solution with singularity at 0, Y_n, which we will use later for inverse scattering.

Lemma 4.22

$$J_n(x) = \sum_{m=0}^{\infty} (-1)^m \left(\frac{x}{2}\right)^{n+2m} \frac{1}{m!(n+m)!}$$

Lemma 4.23 *(Debye's relation)*

$$J_n(x) \sim 0 \textit{ for } |x| < n.$$

Remark: We will, throughout the script, in $\mathbb{R}^2$ use the notation

$$\theta(\varphi) = \begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix}, \theta^{\perp}(\varphi) = \begin{pmatrix} -\sin\varphi \\ \cos\varphi \end{pmatrix}$$

Lemma 4.24
For fixed ψ

$$\int_{S^1} e^{ir\theta\cdot\theta(\psi)} d\theta = \pi J_0(r).$$

Also, for $k \in \mathbb{Z}$*:*

$$\int_{S^1} e^{ir\theta\cdot\theta(\psi) - ik\varphi} d\theta = i^k e^{-ik\psi} 2\pi J_0(r)$$

where $\theta = \theta(\varphi)$*.*

Theorem 4.25 *(FT of radially symmetric functions)*
Let $f : \mathbb{R}^2 \mapsto \mathbb{R}$*,* f *radially symmetric (*$f(x) = f_0(||x||)$*).*
Then

$$\widehat{f}(\xi) = \int_{\mathbb{R}^+} r f_0(r) J_0(r||\xi||)\, dr.$$

$\widehat{f}$ *is also radially symmetric.*

Theorem 4.26
Let $f : \mathbb{R}^2 \mapsto \mathbb{R}$*,* $f(x) := \frac{1}{||x||}$*. Then* $\widehat{f} = f$*.*

Chapter 5

Radon–Transform

5.1 Definition and inversion theorems

Definition 5.1 *(Radon–Transform)*
Let $C := S^{n+1} \times \mathbb{R} \subset \mathbb{R}^{n+1}$. Let $X(\mathbb{R}^n)$ a function space on $\mathbb{R}^n$, $X(C)$ a function space on C.
Let σ the $(n-1)$–dimensional surface measure.

The Radon Transform $R : X(\mathbb{R}^n) \mapsto X(C)$ is defined as

$$\begin{aligned}(Rf)(\theta, s) &= \int_{x\cdot\theta=s} f(x)\, d\sigma(x) \\ &= \int_{x\cdot\theta=s} f(x)\, dx \\ &= \int_{y\cdot\theta=0} f(s\cdot\theta + y)\, dy \\ &= \int_{\mathbb{R}^n} f(x)\delta(x\cdot\theta - s)\, dx\end{aligned}$$

Definition 5.2 *(X–ray, Röntgen Transform)*
Let

$$C' \in \mathbb{R}^{2n} = \{(\theta, x) \in S^{n-1} \times \mathbb{R}^n : x\cdot\theta = 0\}.$$

Let $X(\mathbb{R}^n)$ a function space on $\mathbb{R}^n$, $X(C')$ a function space on C'.

The X–ray transform $P : X(\mathbb{R}^n) \mapsto X(C')$ is defined as

$$(Pf)(\theta, x) = \int_{\mathbb{R}} f(x + t\cdot\theta)\, dt.$$

Remark: In $\mathbb{R}^2$,

$$Rf(\theta, s) = (Pf)(\theta^\perp, s \cdot \theta).$$

Remark: $Rf(\theta, s) = Rf(-\theta, -s)$, So the Radon transform is measured only for $\varphi \in [0, \pi]$.

Remark: We will usually assume

$$\text{supp } f \subset K_1(0),$$

so

$$(Rf)(\theta, s) = 0 \text{ for } |s| > 1.$$

Remark: We assume that the values of Rf are measured for $(\theta(\varphi_k), s_l)$, $\varphi_k = \frac{k\pi}{p}$, $k = 0 \dots p-1$, $s_l = \frac{l}{q}$, $l = -q \dots q$. This is called parallel scanning, in reality things are more difficult (fanbeam–scanning).

Theorem 5.3 *(Fourier Slice Theorem, Projection Theorem)*
Let $f \in \mathcal{S}(\mathbb{R}^n)$, $\theta \in \mathcal{S}^{n-1}$, $\sigma \in \mathbb{R}$. Then

$$\widehat{Rf}(\theta, \sigma) = (2\pi)^{(n-1)/2} \widehat{f}(\sigma\theta)$$

where $\widehat{Rf}$ is a 1d–Fourier Transform with respect to the second argument.

Theorem 5.4 *Fourier Slice for the X–ray transform*
Let $f \in \mathcal{S}$, $\theta \in S^{n-1}$, $\xi \cdot \theta = 0$.

$$\widehat{Pf}(\theta, \xi) = \sqrt{2\pi}\widehat{f}(\xi)$$

where $\widehat{Pf}$ is an $(n-1)$–dimensional Fourier Transform w.r.t. $\theta^\perp$ in the second argument.

Remark: When we want to use Fourier Slice for inversion directly, the inverse Fourier Transform has to be performed on a polar grid, which is not feasible.

Remark: Taking discrete data leads to the idea of backprojection: Add up the values for all measured lines that go through a point x in image space, thus

$$(R^* g)(x) = \sum_k g(\theta_k, x \cdot \theta_k).$$

This is not working, leads to very smooth images.

Geometrical considerations lead to the idea: Compute the backprojection, take its Fourier Transform, Multiply with $\frac{1}{||\xi||}$, take the inverse Fourier Transform. This is called ρ–filtered layergram.

Definition 5.5 *(Backprojection)*
Let $g \in \mathcal{S}(C)$, $x \in \mathbb{R}^n$.

$$R^* : \mathcal{S}(C) \mapsto \mathcal{S}(R^n),\ (R^*g)(x) := \int_{S^{n-1}} g(\theta, x \cdot \theta)\, d\theta$$

is the backprojection operator.

Theorem 5.6 L^2*–adjoint*
R^* *is the* L^2*–adjoint of* R.

Remark:

$$(R^*Rf)(x) = 2(\frac{1}{||x||} * f).$$

Theorem 5.7 *Convolution Theorem for the Radon Transform*
Let $h \in \mathcal{S}(C)$, $f \in \mathcal{S}(\mathbb{R}^n)$. *Then*

$$(R^*h) * f = R^*(h * Rf)$$

.

This implies that a convolution in image space can be performed in data space and vice versa.

Definition 5.8 *(Riesz–Potential)*
Let $\alpha < n$. *Then*

$$I^\alpha : \mathcal{S}(\mathbb{R}^n) \mapsto \mathcal{S}(\mathbb{R}^n),\ \widehat{(I^\alpha f)}(\xi) = ||\xi||^{-\alpha} \widehat{f}(\xi).$$

Let $\alpha < 1$. *Then*

$$I^\alpha : \mathcal{S}(C) \mapsto \mathcal{S}(C),\ \widehat{(I^\alpha g)}(\theta, \sigma) = |\sigma|^{-\alpha} \widehat{g}(\theta, \sigma).$$

Theorem 5.9 *(Riesz inversion formula for the Radon Transform)*
Let $0 \leq \alpha < n$. *Then for* $f \in \mathcal{S}(\mathbb{R}^n)$

$$f = \frac{1}{2}(2\pi)^{1-n} I^{-\alpha} R^* I^{\alpha-n+1} Rf.$$

For $\alpha = n - 1$, we get an algorithm that uses filter on the backprojected data, ρ–filtered layergram.

For $\alpha = 0$, we get an algorithm that backprojects filtered data, filtered backprojection.

Theorem 5.10 *(filtered backprojection (FBP), ρ–filtered layergram)*
Let $f \in \mathcal{S}(\mathbb{R}^n)$, $g = Rf$. Then

$$f = \frac{1}{2}(2\pi)^{1-n} I^{1-n} R^* g \ (\rho \textit{ filtered layergram})$$

and

$$f = \frac{1}{2}(2\pi)^{1-n} R^* I^{1-n} g \ \textit{ (filtered backprojection).}$$

Theorem 5.11
For $n = 3$, the algorithms reduce to

$$f(x) = -\frac{1}{8\pi^2}\Delta_x \int_{S^2} g(\theta, x \cdot \theta)\, d\theta$$

(Δ_x is the Laplace operator wrt x) and

$$f(x) = -\frac{1}{8\pi^2} \int_{S^2} g''(\theta, x \cdot \theta)\, d\theta.$$

Definition 5.12
For n odd, the inversion of the Radon transform is local in the sense that to reconstruct at a point $x \in \mathbb{R}^n$, only the values of integrals over hyperplanes that hit an arbitrarily small surrounding of the point is needed.

Remark: This means that to reconstruct just the heart in $\mathbb{R}^3$, we only need to measure on hyperplanes that hit the heart.

Theorem 5.13
For $n = 2$, we have

$$f(x) = \frac{1}{4\pi} \int_{S^1} \int_{\mathbb{R}} \frac{g'(\theta, s)}{x \cdot \theta - s}\, dx\, d\theta.$$

Theorem 5.14
For n even, the inversion of the Radon transform is not local.

Remark: This means that for line integrals in $\mathbb{R}^2$, to reconstruct the heart we need to measure on **all** lines that hit the support of f.

Remark: The original algorithm by Cormack, that was used in the first tomograph, is a direct solution of the inverse problem based on Fourier expansion of f and g.

Remark: Equivalent formulas exist for the Röntgen transform.

Theorem 5.15 *Characterization of the range of the Radon transform*
Let $g = Rf$. Then

$$p_m(\theta) := \int_R s^m g(\theta, s)\, ds$$

is a homogeneous polynomial in $\mathcal{P}_m$.
If this holds for a $g \in \mathcal{S}(C)$ and all m, then g is in the range of R.

Remark: The Radon transform is invertible on its range, since it is injective due to Fourier slice..

Definition 5.16 *(fractional Sobolev spaces)*
For $\alpha \geq 0$, we define the Sobolev space $H^\alpha(\mathbb{R}^n)$ by

$$H^\alpha(\mathbb{R}^n) = \{f : ||f||_{H^\alpha} < \infty\}$$

where

$$||f||^2_{H^\alpha} := \int_{\mathbb{R}^n} |\widehat{f}(\xi)|^2 (1 + ||\xi||^2)^\alpha d\xi.$$

We define the Sobolev space $H^\alpha(C)$ by

$$H^\alpha(C) = \{g : ||g||_{H^\alpha} < \infty\}$$

where

$$||g||^2_{H^\alpha} := \int_{S^{n-1}} \int_{\mathbb{R}} |\widehat{g}(\theta, \sigma)|^2 (1 + \sigma^2)^\alpha d\theta.$$

Remark: We allow α to be fractional.

Remark: $||\cdot||_{H^\alpha}$ is monotonous increasing with α.

Remark: $||\cdot||_{H^\alpha}$, $||\cdot||_{H^{\alpha'}}$ are not equivalent for $\alpha < \alpha'$.

Theorem 5.17 *(Sobolev estimate for the Radon transform)*
Let $f \in L^2(K_1(0))$ with compact support. Then

$$\exists C_1,\, C_2 :\, C_1||f||_{L^2} \leq ||Rf||_{H^{(n-1)/2}} \leq C_2||f||_{L^2}.$$

Remark: If f is just in L^2, then Rf is in $H^{(n-1)/2}$, so Rf is in a sense smoother than f, which makes inversion difficult.

Remark: For $g = Rf$, we have

$$\frac{1}{C_2}||g||_{H^{(n-1)/2}} \leq ||R^{-1}g||_{L^2} \leq \frac{1}{C_1}||g||_{H^{(n-1)/2}}.$$

This implies that the inversion of the Radon transform is ill-posed of the order $\frac{n-1}{2}$ or $\frac{1}{2}$ for $n = 2$. The inversion is **very** mildly ill–posed.

Chapter 6

Sampling and implementation

Theorem 6.1 *(Poisson's summation formula)*
Let $f \in \mathcal{S}$.

$$\sum_{l\in\mathbb{Z}^n} \widehat{f}(2\pi l) = \frac{1}{(2\pi)^{n/2}} \sum_{k\in\mathbb{Z}^n} f(k)$$

Corollary 6.2 *(Poisson's summation formula, second form)*
Let $f \in \mathcal{S}$.

$$\frac{1}{h^n} \sum_{l\in\mathbb{Z}^n} \widehat{f}(\frac{2\pi l}{h} + \xi) = \frac{1}{(2\pi)^{n/2}} \sum_{k\in\mathbb{Z}^n} f(k\,h) e^{-ihk\xi}$$

Definition 6.3 *(band–limit)*

1. *Let* $f \in \mathcal{S}$. f *is band–limited with bandlimit* Ω *iff* $\widehat{f}(\xi) = 0$ *for* $||\xi||_\infty > \Omega$.
2. *Let* f 2π*–periodic. Then* f *is band–limited with bandlimit* Ω *iff* $\widehat{f_k} = 0$ *for* $|k| > \Omega$, $\widehat{f_k}$ *the Fourier coefficients of* f.

Definition 6.4 *(band–limited, band–filtered version)*
Let $g \in \mathcal{S}$. *Then*

$$f(x) = \widetilde{(\widehat{f}(\xi)\chi_{[-\Omega,\Omega]}(\xi))}(x)$$

is the Ω*–band–limited or* Ω*–band–filtered version of* g.

Remark: Using

$$\widehat{\chi_{[-\Omega,\Omega]}}(\xi) = \sqrt{\frac{2}{\pi}}\Omega\,\mathrm{sinc}(\Omega\xi),$$

we have

$$f(x) = \Omega\sqrt{\frac{2}{\pi}}(g * \mathrm{sinc}(\Omega\cdot))(x).$$

Remark: Equivalently for periodic functions.

Remark: f is the best approximation to g in the space of band–limited functions.

Remark: If f is compactly supported, then f is not band–limited except for $f = 0$. Proof based on Paley–Wiener. See e.g. Hardy 1933, A theorem concerning Fourier Transform, Journal of the London Mathematical Society (`https://londmathsoc.onlinelibrary.wiley.com/doi/abs/10.1112/jlms/s1-8.3.227`). Therefore:

Definition 6.5 *(essentially band–limited)*
Let $f \in \mathcal{S}$. f is essentially band–limited with bandlimit Ω iff $\widehat{f}(\xi)$ is negligible for $||\xi||_\infty > \Omega$, or if f and its Ω–band–limited version are almost the same.

Example: The Ω–bandlimited version of $g(x) = \chi_{[-1,1]}$ is given by

$$f(x) = \frac{1}{\pi}(Si(\Omega(x+1)) - Si(\Omega(x-1)))$$

where Si is the antiderivative of the sinc function (integral sine).

Example: The Ω–bandlimited version of the δ–distribution is given by

$$f(x) = \Omega\sqrt{\frac{2}{\pi}}\,\mathrm{sinc}(\Omega x).$$

Interpretation: An infinitesimally small dot will appear as a sinc function in the signal or image.

Corollary 6.6 *to Poissons's summation formula:*

$$\widehat{f}(\xi) = (2\pi)^{-n/2}h^n \sum_{k\in\mathbb{Z}^n} f(kh)e^{-ikh\xi} - \sum_{l\neq 0} \widehat{f}(\frac{2\pi l}{h} + \xi).$$

Theorem 6.7 *(sampling Theorem by Kotelnikov, Shannon, Nyquist, Wiener...)*
Let $f \in \mathcal{S}$ Ω–band–limited.

1. *If $||\xi|| > \Omega$, then $\widehat{f}(\xi) = 0$.*

2. *Let* $||\xi|| \leq \Omega$, $h \leq \frac{\pi}{\Omega}$. *Then*

$$\widehat{f}(\xi) = (2\pi)^{-n/2} h^n \sum_{k \in \mathbb{Z}^n} f(kh) e^{-ikh\xi}.$$

So in this case $\widehat{f}$ *and thus* f *is uniquely determined by its values* $f(kh)$, $k \in \mathbb{Z}^n$.
The trapezoidal rule for the Fourier integral is exact.

3. *Let* $h \leq \frac{2\pi}{\Omega}$. *Then*

$$\int_{\mathbb{R}^n} f(x)\, dx = h^n \sum_{k \in \mathbb{Z}^n} f(kh).$$

The trapezoidal rule for the integral is exact.

4. *Let* $h \leq \frac{\pi}{\Omega}$. *then*

$$f(x) = \sum_{k \in \mathbb{Z}^n} f(h \cdot k) \operatorname{sinc} \frac{\pi}{h}(x - hk).$$

5. *Let* f, g Ω*–bandlimited,* $h \leq \frac{\pi}{\Omega}$. *Then* $f \cdot g$ *is* 2Ω*–bandlimited.*

$$\int_{\mathbb{R}^n} f(x)g(x)\, dx = h^n \sum_{k \in \mathbb{Z}^n} f(kh)g(kh).$$

6. *The sum converges arbitrarily slow, but can be accelerated by choosing* $h << \frac{\pi}{\Omega}$.

In the following, we will implement 5.7. We will implement it in such a way that the results are exact. We will always assume that the support of f is in the unit circle and f is essentially Ω–bandlimited.

Lemma 6.8
Let $V = \delta$ *or* $\widehat{V} = (2\pi)^{-n/2}$. *Then in 5.7* $V * f = f$.
Let f Ω*–bandlimited and* $\widehat{V}_\Omega(\xi) = (2\pi)^{-n/2}$ *for* $||\xi|| \leq \Omega$. *Then in 5.7* $V_\Omega * f = f$.

Lemma 6.9
Let $V = R^* v$, $v \in \mathcal{S}(C)$. *Assume that* v *is even and rotationally invariant, that is*

$$v(\theta, s) = v(s) = v(-s).$$

Then

$$\widehat{V}(\xi) = 2(2\pi)^{(n-1)/2} ||\xi||^{1-n} \widehat{v}(||\xi||).$$

Remark: For $\widehat{V}(\xi) = (2\pi)^{-n/2}$ 5.7 is 5.10, filtered backprojection.

From now on let

$$\widehat{V_\Omega}(\xi) = (2\pi)^{-n/2}\widehat{\phi}\left(\frac{||\xi||}{\Omega}\right),\ \phi(\sigma) \sim \begin{cases} 1, & \text{if } |\sigma \leq 1 \\ 0, & \text{otherwise} \end{cases},\ V_\Omega = R^* v_\Omega.$$

Remark: We have $f \sim R^*(v_\Omega * g)$ with equality for "=" in the definition of ϕ.

Remark: $R^*(v_\Omega * g)$ is always Ω–bandlimited (even if f is not). Think of it as a projection onto the set of bandlimited functions. Note that the projection is not orthogonal (due to aliasing, see below).

Remark: $R^*(v_\Omega * g)$ is continuous since $\widehat{v_\Omega}$ is bounded.

Lemma 6.10

$$v_\Omega(s) = \frac{1}{2}(2\pi)^{-n}\int_{-\Omega}^{\Omega} |\sigma|^{n-1}\widehat{\phi}\left(\frac{\sigma}{\Omega}\right) e^{is\sigma}\, d\sigma.$$

Corollary 6.11

1. *Let*

$$\widehat{\phi}(\sigma) = \begin{cases} 1, & \textit{if } |\sigma| \leq 1 \\ 0, & \textit{otherwise}. \end{cases}$$

   and $n = 2$. Then

$$v_\Omega(s) = \frac{\Omega^2}{4\pi^2}\left(\operatorname{sinc}(s\,\Omega) + \frac{1}{\Omega^2 s^2}(\cos(s\,\Omega) - 1)\right)$$

   and $v_\Omega(0) = \frac{\Omega^2}{8\pi}$. This formula is due to Ramachandran and Lakshminarayanan (1971), v_Ω is called the Ram-Lak filter.

2. *Alternative choices for $\widehat{\phi}(\sigma)$ include $\operatorname{sinc}(\sigma\frac{\pi}{2})$ with a cutoff at 1. This was proposed by Shepp and Logan (and again v_Ω can be computed from 6.10). Another popular choice is $\cos(\sigma\frac{\pi}{2})$ with a cutoff at 1, the cosine filter.*

We will now turn to implementation of parallel scanning. That was already defined before 5.3. We measure $g(\theta(\varphi_k), s_l)$ where

$$\varphi_k = \frac{k\pi}{p},\ k = 0 \ldots p-1; s_l = \frac{l}{q},\ l = -q \ldots q.$$

We sum up the steps needed:

1. Select Ω and the parameters p and q for parallel scanning.
2. Select a filter function ϕ and compute $\widehat{v_\Omega}$.
3. Measure data g.
4. Compute $g * v_\Omega$ **exactly**.
5. Compute $F = R^*(g * v_\Omega)$ **exactly**.

We want to implement and select everything in such a way that $F = f$ for Ω–bandlimited f.

Note: We have discrete measurements, and all implementations are discrete, but the result is analytically correct (neglecting measurement errors, of course). Additionally, the procedure is continuous.

Assume now that v_Ω is Ram-Lak, $\operatorname{supp} f \subset K_1(0)$, and f is essentially Ω–bandlimited.

Lemma 6.12
g and v_Ω are Ω–bandlimited. The trapezoidal rule for

$$(v_\Omega * g)(\theta, s) = \int_{-1}^{1} v_\Omega(s - s')\, g(\theta, s')\, ds'$$

is exact when $q \geq \frac{\Omega}{\pi}$ (6.7).

Lemma 6.13
Let

$$h(\varphi) = (v_\Omega * g)(\theta(\varphi), x \cdot \theta(\varphi)).$$

Then we have for its Fourier coefficients $\widehat{h_k} \sim 0$ for $|k| > 2\Omega$.

In the Exercises you prove:

Corollary 6.14
Let $p > \Omega$. Then

$$R^*(v_\Omega * g)(x) = \int_{-\pi}^{\pi} h(\varphi) d\varphi = \frac{\pi}{p} \sum_{j=0}^{2p-1} h(\varphi_j).$$

Note: There is an error in the video at this point, for some reason I come up with the wrong condition. You will prove that this is the right one.

Corollary 6.15
Let $f \in \mathcal{S}(R^2)$ Ω–bandlimited, $p \geq \Omega$, $q \geq \frac{\Omega}{\pi}$. Then

$$f(x) = (V_\Omega * f)(x) = 2\frac{\pi}{p}\frac{1}{q}\sum_{j=0}^{p-1}\sum_{l=-q}^{q} v_\Omega(x\,\theta(\varphi_j) - s_l)g(\theta_j, s_l)$$

is an exact, continuous reconstruction formula and only uses discrete measurement values from parallel scanning.

Remark: Assume that p and q are fixed. Then $\Omega = \min(p, \pi\, q)$. This implies that improving the number of measurement angles or sensors alone does not improve resolution, we have to improve both. In the optimal case, we should have $p = \pi\, q$ (which we assume from now on).

Remark: Regarding the implementation of 6.14: Since f is essentially Ω–bandlimited, we should sample it with a grid spacing of $h = \frac{\pi}{\Omega}$. Since f has compact support, we must compute 6.14 for $O(\Omega^2)$ values of x. For each value we must compute the double sum, resulting in a complexity of $O(\Omega^4)$.

This can be reduced with the following idea: Let

$$h_j(s) = \sum_{l=-q}^{q} v_\Omega(s - s_l)g(\theta_j, s_l).$$

Then

$$f(x) = (V_\Omega * f)(x) = 2\frac{\pi}{p}\frac{1}{q}\sum_{j=0}^{p-1} h_j(x\,\theta(\varphi_j)).$$

h_j is Ω–bandlimited, so it should be sampled with grid spacing $h = \frac{\pi}{\Omega}$. So beforehand, we compute

$$h_j(s_k) = \sum_{l=-q}^{q} v_\Omega(s_{k-l})g(\theta_j, s_l).$$

This can be done using the convolution theorem and fast Fourier transform for each j, resulting in a complexity of $O(\Omega^2 \log \Omega)$.

$f(x)$ is now computed by interpolating h_j, using the computed values. To be exact, we would have to use 6.7 for the interpolation. However, in practice, it suffices to use simply linear interpolation. Usually, in implementations of the inverse Radon transform (Matlab, python, ...) the choice of interpolation function is left to the user. For linear interpolation, the complexity of computing $f(x)$ is $O(\Omega)$, resulting in an overall complexity of $O(\Omega^3)$ for the backprojection step.

All in all, with this simplification the complexity of filtered backprojection is $O(\Omega^3)$.

Remark: For Ram–Lak and $q = \frac{\Omega}{\pi}$, we have

$$v_\Omega(s_k) = \frac{\Omega^2}{2\pi^2} \begin{cases} \frac{1}{4}, & \text{if } l = 0 \\ 0, & \text{if } l \neq 0 \text{ even} \\ \frac{1}{\pi^2 l^2}, & \text{otherwise.} \end{cases}$$

Theorem 6.16 *(sampling theorem by Petersen and Middleton)*
Let $f \in \mathcal{S}(\mathbb{R}^n)$. Let K the support of $\widehat{f}$. Let $W \in \mathbb{R}^{n\times n}$ such that

$$K \cap K + 2\pi W^{-t} k = \emptyset \ \forall\, k \in \mathbb{Z}^n,\ k \neq 0, .$$

Then f is uniquely determined by the values of $g(Wk)$, $k \in \mathbb{Z}^n$.

Theorem 6.17 *(Natterer, Theorem 3.1)*
Let $g = Rf$, $f \in \mathcal{S}(\mathbb{R}^2)$ Ω–bandlimited. Then the essential support K of the 2D–Fouriertransform of g is given by

$$K := \{(k, \sigma) :\ |\sigma| < \Omega,\ |k| < |\sigma|\}.$$

Note: g is 2π–periodic in the first variable, so the Fourier transform has to be understood in the distributional sense, according to the exercises. Note that we will nevertheless make use of 6.16, although we only proved that for functions in $\mathcal{S}$.

Note: This is a simplification, the essential support is in fact a little bit bigger.

Corollary 6.18 *(2D–sampling)*

1. *Let*
$$W = \begin{pmatrix} \frac{\pi}{p} & 0 \\ 0 & \frac{1}{q} \end{pmatrix},\ p = \Omega,\ q = \frac{\Omega}{\pi}.$$
Then 6.16 is satisfied. This results in parallel scanning.

2. *Let*
$$W = \begin{pmatrix} \frac{2\pi}{p} & -\frac{1}{q} \\ 0 & \frac{1}{q} \end{pmatrix},\ p = \Omega,\ q = \frac{\Omega}{\pi}.$$
Then 6.16 is satisfied. Wk is a grid that has only half the points of the grid of parallel scanning, thus uses less measurements, but still g is uniquely determined. This is called interlaced or efficient scanning.

Chapter 7

3D CT

We only have a very small glimpse at true 3D scanning.

Definition 7.1 *(Cone–Beam–Transform)*
Let $f \in \mathcal{S}(R^n)$. Let $a, x \in \mathbb{R}^n$. Then

$$(Df)(a,x) := \int_0^\infty f(a+tx)\,dt$$

is the Cone–Beam–Transform.

Theorem 7.2
In distributional sense, we have

$$\widehat{Df}(a,\xi) = \int_0^\infty \rho^{n-2} e^{i\rho\, a\cdot\xi} \widehat{f}(\rho\xi) d\rho$$

where the Fourier Transform is with respect to the second variable.

Definition 7.3 *(Tuy's condition)*
Let $f \in \mathcal{S}(\mathbb{R}^3)$. Let $a(\lambda)$, $\lambda \in [0,1]$, a differentiable curve in $\mathbb{R}^3$. a satisfies Tuy's condition, if every plane that hits the support of f contains at least one point of the curve, and the curve is not tangential to the plane in this point.

This boils down to: Let $x \in$ supp f. Let $\theta \in S^2$. Then there exists $\lambda(\theta, x)$ such that

$$a(\lambda(\theta,x)) \cdot \theta = x \cdot \theta,\ a'(\lambda(\theta,x)) \cdot \theta \neq 0.$$

Theorem 7.4 *(Tuy's reconstruction formula)*
If Tuy's condition is satisfied, then f can be reconstructed from the values of

$$(Df)(a(\lambda), x),\ \lambda \in [0,1],\ x \in \mathbb{R}^3.$$

In particular for $x \in$ supp f:

$$f(x) = (2\pi)^{-3/2} \frac{1}{i} \int_{S^2} \left(\frac{1}{a'(\lambda) \cdot \theta} \frac{d}{d\lambda} \widehat{Df}(a(\lambda), \theta) \right) \Big|_{\lambda = \lambda(\theta, x)} d\theta.$$

Note: Tuy's condition is necessary for stable reconstruction. The simple geometry of a circle around the object does not satisfy Tuy's condition.

Remark: Definitely, FDK should be introduced here as an example (as in the old lecture). Do this next time.

Chapter 8

Stability of reconstruction formulas

Assume dimension 2.

Theorem 8.1 *Assume that $\widetilde{g} = Rf + n$, noise n is an uncorrelated random variable in measurement space with mean 0 and variance σ^2. Let f Ω–bandlimited and v an exact reconstruction filter (Ram–Lak).*

Let

$$\widetilde{f}(x) = \int_{S^1} \int_{\mathbb{R}} \widetilde{g}(\theta, s) v(x \cdot \theta - s)\, dx\, d\theta.$$

Then $\widetilde{f}$ is a random variable with mean f and variance less or equal to $\frac{\Omega^3}{24\pi^2}\sigma^2$.

Theorem 8.2 *Let $f \in \mathcal{S}(\mathbb{R}^2)$, $f = f_\Omega + f_\Omega^*$, f_Ω the Ω–bandlimited version of f. Let $\widetilde{g} = Rf + n$ with noise n, $||n||_2 \leq \epsilon$. Let v_Ω an exact reconstruction filter for Ω–bandlimited functions (Ram–Lak). Then*

$$||v_\Omega * \widetilde{g}||_2^2 \leq \underbrace{\int_{S^1} \int_{-\Omega}^{\Omega} \sigma^2 |\widehat{Rf_\Omega^*}(\theta, \sigma)|^2 d\sigma\, d\theta}_{\text{Aliasing}} + \underbrace{||f_\Omega^*||_2^2}_{\text{model error}} + \underbrace{2\Omega^2 ||n||_2^2}_{\text{noise}}.$$

Filtered backprojection is a regularization of projection type with regularization parameter $1/\Omega$.

Chapter 9

Discretization, ART, Kaczmarz

A completely different reconstruction method is derived by simply discretizing. We differentiate:

- $Rf(\theta, s) = g(\theta, s)$ (continuous problem)
- $Rf(\theta_l, s) = g(\theta_l, s) = g_l(s)$ (semidiscrete form)
- $Rf(\theta_l, s_l) = g(\theta_l, s_l) = g_l$ (discrete form)

In image space, f is discretized as

$$f(x) \sim \sum_k f_k \chi_k(x),\ f_k \in \mathbb{R}$$

with Ansatzfunctions χ_k. Typically, χ_k is the characteristic function of a pixel or voxel or

$$\chi_k(x) = e^{-\lambda ||x - x_k||^2/2}.$$

We restrict these remarks to pixels in $\mathbb{R}^2$.

In the fully discrete Radon problem, we end up with an equation

$$Rf = g,\ R = (a_{lk}),\ f = (f_k),\ g = (g_l)$$

where

$$a_{lk} = \int_{L_l} \chi_k(x)\, dx,\ L_l = L(\theta_l, s_l).$$

For pixel ansatz functions, the system matrix R can be computed very efficiently, so it is not stored but computed on the fly (Siddon/Bresenham).

Since dimensions are large and the system matrix is sparse, we need iterative methods. Since Landweber (page 18) is too expensive, we go for the Kaczmarz method.

Definition 9.1 *(Kaczmarz)*
Let $R_k : X \mapsto Y_k$, $k = 0 \dots p-1$, X, Y_k Hilbert spaces. Solve

$$R_k f = g_k,\ k = 0 \dots p-1.$$

Let $f^{(0)} \in X$. Choose ω fixed. Then the Kaczmarz method is defined by

$$f^{(l+1)} = f^{(l)} + \omega R_k^*(R_k R_k^*)^{-1}(g_k - R_k f^{(l)}),\ k = k(l).$$

A typical choice for k is $k = l \mod p$. In this case we call $f^{(lp)}$ the sweeps. The effort for computing one sweep is roughly equivalent to computing one step in the Landweber method.

Note that we quietly assumed that R_k is surjective. This is not needed in the generalized Kaczmarz method.

Example: Let

$$R = \begin{pmatrix} R_0^t \\ \vdots \\ R_{p-1}^t \end{pmatrix}, R_k \in \mathbb{R}^m,\ g = \begin{pmatrix} g_0 \\ \vdots \\ g_{p-1} \end{pmatrix}.$$

The Kaczmarz method for the solution of $Rf = g$ is then with $k = k(l)$:

$$f^{(l+1)} = f^{(l)} + \omega \frac{g_k - R_k^t f^{(l)}}{||R_k||^2} R_k.$$

Example: (semidiscrete tomography problem, parallel geometry):
Let $(R_k f)(s) = Rf(\theta_k, s)$. Find an image f such that $(R_k f)(s) = g_k(s)$, $k = 0 \dots p-1\ \forall s \in \mathbb{R}$.

We restrict the Radon transform to functions with support in the unit circle, so

$$R_k : L^2(K_1(0)) \mapsto L^2([-1,1]).$$

Then the semidiscrete Kaczmarz method (without discretization in image space) is given by

$$f^{(l+1)}(x) = f^{(l)}(x) + \omega \frac{1}{2\sqrt{1-(x\cdot\theta_k)^2}}(g_k(x\cdot\theta_k) - Rf(\theta_k, x\cdot\theta_k))$$

Theorem 9.2 *(Convergence theorem for the general Kaczmarz method, Tanabe 1971)*
Let $R_k : X \mapsto Y_k$, $k = 0 \dots p-1$, X, Y_k Hilbert spaces. Let $g_k \in Y_k$. Let

$$R = \begin{pmatrix} R_0 \\ \vdots \\ R_{p-1} \end{pmatrix}, \; g = \begin{pmatrix} g_0 \\ \vdots \\ g_{p-1} \end{pmatrix}.$$

Assume $\exists f \in X : \; Rf = g$ (consistency).
Let $C_k : X \mapsto X$ continuous and positive definite, and $C_k \geq R_k^* R_k$ in the sense that

$$(Cx, x) \geq (R_k^* R_k x, x).$$

Let $0 < \omega < 2$ and $f^{(0)} \in X$.
The general Kaczmarz method is then defined by

$$f^{(l+1)} = f^{(l)} + \omega R_k^* C_k^{-1} (g_k - R_k f^{(l)}),$$

where $k = k(l)$, we assume $k = l \mod p$.
Let $f^{(0)} \in Ker(R)^\perp$, e.g. $f^{(0)} = 0$.
Then $f^{(lp)} \to_{l \to \infty} R^+ g$, i.e. the sweeps converge towards the minimum norm solution of $Rf = g$.

Note: The numerical experiment showed that the order of the operators is important for the convergence speed. For the Radon transform, we should choose the directions θ_j in such an order that subsequent iterations are as orthogonal as possible to the ones before. To this end, clever ordering schemes were designed. However, a simple random order will behave similarly efficient in practice.

Chapter 10

Statistical Inverse Problems

Our main example for statistical problems is emission tomography. While this is an inherently 3D problem (since emitted photons cannot be restricted to a plane like in the CT case), for simplicity, we view it in 2D only. However, since we will be considering discrete algorithms only, everything can simply be transferred to 3D, apart from the numerical complexity.

10.1 Emission Tomography

In emission tomography, a (radioactive) tracer is injected into the body. The external radioactivity is measured. The goal is to reconstruct the density $f(x)$ of the tracer inside the body. Since radiation is very low, single photons (rather than a constant flow of photons like in CT) are measured. The photons are sent out by the tracer at random, so by nature this is a random process which has to be modeled using statistics.

Definition 10.1 *(attenuated Radon transform, one-sided X-Ray transform)*
Let $f,\ \mu \in \mathcal{S}(\mathbb{R}^2)$. *Then*

$$D : \mathcal{S}(\mathbb{R}^2) \mapsto \mathcal{S}(\mathbb{R}^n \times S^{n-1}), (D\mu)(x,\theta) := \int_0^\infty \mu(x + t\,\theta)\,dt$$

is the one–sided X-Ray transform.

$$R_\mu : \mathcal{S}(\mathbb{R}^2) \mapsto \mathcal{S}(C),\ (R_\mu f)(\theta, s) := \int_{x\cdot\theta=s} e^{-(D\mu)(x,\theta^\perp)} f(x)\,dx$$

is the attenuated Radon transform. Here, $\theta^\perp$ *is the rotation of* θ *by* $\frac{\pi}{2}$ *such that* $\det(\theta, \theta^\perp) = 1$.

Note: Other than for the Radon transform, in general we have

$$(R_\mu f)(\theta, s) \neq (R_\mu f)(-\theta, -s).$$

SPECT (Single Photon Emission Computerized Tomography): In the analytical model for SPECT, the attenuated Radon transform of the tracer density f is measured in a parallel geometry. μ is usually assumed to be known. However, even then analytical reconstruction apart from simple cases where the attenuation is constant (see exercises) is difficult.

PET (Positron Emission Tomography): In the analytical model for PET, the Radon transform of the tracer density f is measured. The geometry consists of all lines connecting the sensors which are equally distributed on a circle around the object. The attenuation μ is assumed to be known and can be corrected for beforehand. So PET boils down to a CT measurement in an unusual geometry. However, the main difference is that the measurements are statistical by nature.

For both cases, we will consider only discrete ART–like reconstruction formulas. This is consistent with what is used in modern devices (traditionally, filtered backprojection was used for reconstruction ignoring the differences in the model).

The process of radioactive decay is modeled by the Poisson distribution.

Definition 10.2 *(Poisson–distribution)*
Let X the random variable for the number of (observed) decays of a radioactive probe in a time interval of length T_0. Let c the total radioactivity in the probe such that

$$E(X) = cT_0 =: \lambda.$$

Then for $\mu \in N_0$

$$p(X = \mu) = e^{-\lambda}\frac{\lambda^\mu}{\mu!}.$$

X is Poisson–distributed.

Note: Since the radioactive strength of the probe is reduced over time, c is in fact time dependent. We ignore this for our model, in practice this is easily corrected for with the half time known.

Theorem 10.3
Let X Poisson–distributed with parameter λ. Then

$$E(X) = \lambda$$

and

$$\mathit{Var}(X) = E((X - E(X))^2) = \lambda.$$

This leads to a statistical model for PET. Let X_l the random variable for the number of decays on the line L_l in a time interval of length T_0, $l = 0 \dots L-1$. Assume that the reconstruction area is divided into pixels and that the tracer density is given by

$$f(x) = \sum_k f_k \chi_k(x)$$

where χ_k is the characteristic function of pixel k. Further let $a_{l,k}$ the length of the intersection of Pixel k and line L_l. Then the total radioactivity on the line is given by

$$c_l = \sum_k a_{lk} f_k$$

(everything as in the ART case). So the expected value of X_l is $T_0 c_l$. Since X_l is Poisson–distributed, its parameter must be

$$\lambda_l = T_0 c_l.$$

In the following, we set $T_0 = 1$ so X_l is Poisson–distributed with parameter

$$\lambda_l = \sum_k a_{l,k} f_k.$$

This boils down to:

Lemma 10.4
The probability of having exactly g_l decays on line L_l, under the assumption that for the density $f(x)$ of a radioactive tracer we have $f(x) = \sum_k f_k \chi_k(x)$, is given by

$$p\left(X_l = g_l \middle| f(x) = \sum_k f_k \chi_k(x)\right) = e^{-\lambda_l} \frac{\lambda_l^{g_l}}{g_l!}$$

where

$$\lambda_l = \sum_k a_{l,k} f_k.$$

Now let $A = (a_{l,k})$, $X = (X_l)$, $G = (g_l)$, $F = (f_k)$. Then $\lambda_l = (AF)_l$. Under the assumption that the X_l are independent, we finally have

$$p\left(X = G \middle| f(x) = \sum_k f_k \chi_k(x)\right) = \Pi_l \left(e^{-(AF)_l} \frac{(AF)_l^{g_l}}{g_l!}\right) =: L_G(F)$$

(where χ_k, $a_{l,k}$ etc. are as above).

Now assume that a fixed vector G has been measured. Then for every vector F, $L_G(F)$ is the probability of measuring G if the true radioactivity function is $f(x) = \sum_k f_k \chi_k(x)$.

We reconstruct by determining a vector F where the probability of actually measuring G is high. This gives rise to two reconstruction ideas.

Definition 10.5 *(Maximum Likelihood)*
View F as a random variable. Then

$$f_{ML} = \arg\max_F p(F|\textit{G is measured})$$

is the maximum likelihood approximation. For the PET case this means

$$f_{ML} = \arg\max_F L_G(F).$$

Definition 10.6 *(a posteriori) Bayes estimate*
View F as a random variable. Then

$$E(F|\textit{G is measured})$$

is the Bayes estimate.

Note: In the literature, varying definitions are used.

10.2 Statistics Basics

We remind of some definitions from stochastics about multidmensional random variables. In the following, X is a random variable in $\mathbb{R}^n$ with probability distribution $p(X = x)$.

Definition 10.7 *(mean, expected value)*

$$E(X) = \int_{\mathbb{R}^n} x\, p(X = x)\, dx$$

is the mean of X. E is linear, that is

$$E(AX + \mu) = AE(X) + \mu.$$

Definition 10.8 *(covariance matrix)*

$$\textit{cov}\,(X) = E((X - E(X))(X - E(X))^t)$$

is the covariance matrix. The covariance matrix is positive semidefinite (we will assume it is positive definite). When the off-diagonals are 0, the X_i are uncorrelated. If the X_i are independent, they are uncorrelated. We have

$$\textit{cov}\,(AX+\mu) = A\,\textit{cov}\,(X)A^t.$$

The diagonal entries of the matrix are the variance (in a noise model, the size of the noise), the off–diagonals stand for the connection between random variables (in a noise model, this should be zero).

Definition 10.9 *(normal distribution)*
For a positive definite Matrix $\Sigma \in \mathbb{R}^{n\times n}$ and a vector $\mu \in \mathbb{R}^n$, the normal distribution is defined by

$$p(X=x) := (2\pi)^{-n/2}\frac{1}{\sqrt{\det\Sigma}}e^{-\frac{1}{2}||\Sigma^{-1/2}(y-\mu)||^2} = (2\pi)^{-n/2}\frac{1}{\sqrt{\det\Sigma}}e^{-\frac{1}{2}(y-\mu)^t\Sigma^{-1}(y-\mu)}.$$

We have $\textit{cov}\,X = \Sigma$ and $E(X) = \mu$. The standard normal distribution is given by $\Sigma = I$ and $\mu = 0 \in \mathbb{R}^n$.
Note (I forgot to mention this in the video): Obviously, if X is normally distributed, then $Bx + c$ is normally distributed.

Definition 10.10 *(conditional probability)*
Let $X = (X_1, X_2)$. Then the probability that $X_1 = f$ provided $X_2 = g$ is given by

$$p(X_1 = f|X_2 = g) := \frac{p(X=(f,g))}{\int p(X_1 = x, X_2 = g)\,dx}.$$

Theorem 10.11 *(conditional probability for normal distributions)*
Let X normally distributed with $\textit{cov}\,(X) = K$, $E(X) = \mu$.
Let

$$X = (F,G),\ \mu = (\widetilde{f},\widetilde{g}),\ K = \begin{pmatrix} K_{11} & K_{12} \\ K_{12}^t & K_{22}\end{pmatrix},\ F \in \mathbb{R}^n,\ G \in \mathbb{R}^m$$

where $K_{11} \in \mathbb{R}^{n\times n}$ etc.
Then $p(F = f|G = g)$ is normally distributed with respect to f for fixed g with mean

$$\widetilde{f} + K_{12}K_{22}^{-1}(g - \widetilde{g})$$

and covariance matrix

$$K_{11} - K_{12}K_{22}^{-1}K_{12}^t.$$

10.3 Bayes– and ML–estimates for inverse problems

Theorem 10.12 *(Bayes–estimate for normally distributed random variables with noise)*
Assume $f \in \mathbb{R}^N$ is a random variable, normally distributed with covariance matrix F and mean $\overline{f}$ (image).
Let $A \in \mathbb{R}^{M\times N}$ (observation/measurement operator).
Assume $n \in \mathbb{R}^M$ is a random variable, normally distributed with covariance $\sigma^2\, I$ and mean zero (noise).
Assume that n and F are uncorrelated. Then (f, n) is normally distributed with mean $(\overline{f}, 0)$ and covariance

$$\begin{pmatrix} F & 0 \\ 0 & \sigma^2 I \end{pmatrix}.$$

Let $g = Af + n$ (measurement). Then

$$X = \begin{pmatrix} f \\ g \end{pmatrix} = \begin{pmatrix} I & 0 \\ A & I \end{pmatrix} \begin{pmatrix} f \\ n \end{pmatrix}$$

is a random variable in $\mathbb{R}^{N+M}$. X is normally distributed,

$$E(X) = \begin{pmatrix} \overline{f} \\ A\overline{f} \end{pmatrix}, \ \textit{cov}\,(X) = \begin{pmatrix} F & FA^t \\ AF & AFA^t + \sigma^2 I \end{pmatrix} =: \begin{pmatrix} K_{11} & K_{12} \\ K_{12}^t & K_{22} \end{pmatrix}.$$

Assume that $\widetilde{g}$ was measured, a realization of g. Then

$$f_{Bayes} = E(f|g = \widetilde{g}) = \overline{f} + FA^t(AFA^t + \sigma^2 I)^{-1}(\widetilde{g} - A\overline{f}).$$

Assuming f is uncorrelated, $F = I$:

$$\begin{aligned} f_{Bayes} &= \overline{f} + A^t(AA^t + \sigma^2\, I)^{-1}(\widetilde{g} - A\overline{f}) \\ &= \overline{f} + (A^tA + \sigma^2 I)^{-1}A^t(\widetilde{g} - A\overline{f}). \end{aligned}$$

Note that, since $f|g = \widetilde{g}$ is normally distributed, from the definition of normal distribution we have $f_{ML} = f_{Bayes}$.
The Bayes estimate is a generalization of Tikhonov–regularization in this very special case.

Now we do this for emission tomography.

For the following, assume that $f \in \mathbb{R}^N$ is an image, $A \in \mathbb{R}^{M\times N}$ is a measurement operator. $g \in \mathbb{R}^M$ is a Poisson distributed variable with $E(g_i) = (Af)_i$, $\widetilde{g}$ is a realization of g (a measurement). According to 10.4, the $\log$–likelihood function is given by

$$l(f) = \log L(f) = \sum_i \widetilde{g}_i \log(Af)_i \cdot (Af)_i - \sum_i \log(\widetilde{g}_i!).$$

Theorem 10.13 *(EM algorithm, Kuhn–Tucker–conditions)*
Let

$$f_{ML} := \arg\max_{f \geq 0} l(f).$$

Then

$$0 = f_{ML} \cdot (\nabla l)(f_{ML}) = f_{ML} \cdot A^t \frac{g}{A f_{ML}} - f_{ML} \cdot A^t \mathbf{1}$$

and

$$f_{ML} = f_{ML} \cdot \frac{\mathbf{1}}{A^t \mathbf{1}} \cdot A^t \frac{g}{A f_{ML}}.$$

Here, the vector division and multiplication is componentwise, and $\mathbf{1}$ *is a vector of ones in* $\mathbb{R}^m$.
The fixpoint algorithm for this equation

$$f^{(k+1)} = f^{(k)} \cdot \left(\frac{\mathbf{1}}{A^t \mathbf{1}} \cdot A^t \frac{g}{A f^{(k)}} \right)^{\omega}$$

converges towards f_{ML} *provided* ω *is small enough. This is called the EM (Expectation Maximization) algorithm.*

Replacing multiplication by addition and division by subtraction, this looks very much like Landweber. So it makes sense to consider a Kaczmarz–like algorithm to speed up EM (which is notoriously slow), where in each iteration, only a subset of the data and the equations is used. This is called OSEM (Ordered Subset Expectation Maximization). In a way, it is a multiplicative version of the ART–algorithm, also called MART (multiplicative ART), in a slightly different notation.

Since a couple of years, OSEM/MART is the standard algorithm for solving the emission tomography problem in clinical devices. Up to that point, filtered backprojection–based algorithms were used.

Chapter 11

Applications

11.1 MRI

In Magnetic Resonance Imaging, the Fourier transform of the solution is measured directly (where in the simplest case, the function is the proton density of the body).

The measured function is

$$g(t) = \int_{\mathbb{R}^2} f(x, y, z_0)\mu(x, y)e^{i\omega_0 t + x\, G_x\, t + y \cdot G_y T}\, dx\, dy$$

where f is the proton density, μ is the attenuation dependent on the distance between coil and (x, y), ω_0 is the Larmor frequency without gradient field in the (x, y, z_0)–plane, G_y is the strength of the gradient field in y direction, T is the length of the pulse in y direction, G_x is the strength of the field in x direction. (In the video, I have set $\mu = 1$ for simplicity.)

So we get

$$\widehat{f\mu}(G_x\, t, G_y\, T, z_0) = \frac{e^{-i\omega_0 t}}{2\pi} g(t).$$

Now varying G_y, the Fourier Transform of f can be measured everywhere. Typically, in each single measurement the Fourier Transform on a line (or path, by changing G_y during the measurement) can be computed.

Theorem 11.1
The inverse problem of MRI is well–posed (due to Parseval), assuming full measurements.

Theorem 11.2 *(Radial sampling)*
Assume that $\widehat{f}(\sigma_k, \theta_l)$ has been measured with MRI for equidistant σ_k and θ_l. Then, due to Fourier Slice, we have that

$$\widehat{f}(\sigma_k\, \theta_l) = \widehat{Rf}(\theta_l, \sigma_k).$$

So parallel geometry Radon data $(Rf)(\theta_l, \sigma_k)$ can be computed by inverse Fourier Transform and the image can be computed using filtered backprojection.

11.2 Ultrasound: Inverse Scattering

In inverse scattering, an object is irradiated by an incoming wave, generating a scattered wave. The scattered wave is measured outside of the object. The goal is to reconstruct the sound speed from the scattered wave.

We view the time–harmonic problem. Let $U(x,t)$ the sound pressure of the total wave, $U = U^i + U^s$, U^i, U^s are the incoming and scattered waves, respectively. Assume that the time–dependence is given by $e^{i\omega t}$, so

$$U(x,t) = e^{i\omega t} u(x)$$

etc. Since $U(x,t)$ satisfies the wave equation

$$c(x)^2 \Delta U(x,t) = \frac{d^2 U}{dt^2}(x,t)$$

we have

$$\Delta u(x) + \underbrace{\frac{\omega^2}{c(x)^2}}_{k(x)^2} u(x) = 0$$

(and a boundary condition (radiation condition) at infinity). This is the Helmholtz equation.

Assume that $c(x) = c_0$ is constant. Then there is no scattered wave, thus $u = u^i$ which means that u^i must satisfy the equation

$$\Delta u(x) + k^2 u(x) = 0$$

where $k = \omega / c_0$. Specifically, the generated incoming waves in devices are modelled as plane waves in direction $\theta \in S^1$ given by

$$u^i_\theta(x) = e^{ikx\cdot\theta}.$$

In general, we assume that the sound speed is constant (=c_0) away from the body and varies slightly inside the body, so we have

$$k(x)^2 = k^2(1 + q(x))$$

where $q(x)$ is a function with support inside the body which represents the difference between sound speeds inside the body and the outside.

In the inverse scattering problem, we try to reconstruct q given measurements of the scattered wave u^s outside the body for various (all) directions θ of incoming waves.

The measurement operator is given by

$$M : L^2(K_1(0)) \mapsto L^2(S^1 \times S^1),\ (Mq)(\theta, \psi) = u^s_\theta(\psi)$$

and u^s_θ is the scattered wave generated by the incoming wave u^i_θ.

u satisfies the Helmholtz equation, which is (affine) linear for fixed q (or fixed u). However, since q and u are unknown in the interior in our case, the dependency of u^s on q is non–linear.

Therefore, in the simplest case the Born approximation is used instead.

Theorem 11.3 *(Lippmann–Schwinger–Equation)*
Let $u = u^i + u^s$ a solution to the nonhomogeneous Helmholtz equation with contrast function $q(x)$ and support in the unit circle. Then

$$u^s(x) = k^2 \int_{K_1(0)} q(x)u(x)G(x, y)\, dy$$

where $G(x, y)$ is Green's function for the Helmholtz equation with radiation condition.
In $\mathbb{R}^2$, $G(x, y) = H_0(k||x - y||)$ where H_0 is the Hankel function of order zero.

Definition 11.4 *(Born–approximation)*
Let everything as in 11.3. Then

$$u^s(x) = k^2 \int_{K_1(0)} q(x)u(x)G(x, y)\, dy$$

is the Born approximation for the time harmonic inverse scattering problem. It approximates the true solution for small q.

11.3 Analytical inversion for diffraction tomography (Born approximation)

Definition 11.5
The term inverse crime refers to the habit of testing an algorithm with artificially generated data that was produced using the (possibly unrealistic) same assumptions as for deriving the algorithm.

Examples include solving the inverse scattering problem with data produced in the Born approximation and then applying an algorithm that only works for the Born approximation, or simply using exactly the same program for producing the data and computing the inverse. If at all possible, analytical solutions to the base problem should be used for testing.

Theorem 11.6 *(Green's function for the Helmholtz equation with radiation condition in 2D)*
Green's function for the radiation condition in 2D is given by

$$G(x, y) = H_0(k||x - y||)$$

where $H_0 = J_0 + iY_0$ is the Hankel function of zero order, J_0 and Y_0 are Bessel functions of the first and second kind, the base solutions for the Bessel equation.

Theorem 11.7 *(Wolf's projection theorem)*
Let $u^i_\theta(x) = e^{ikx\theta}$ with θ fixed (for the moment) the incident plane wave in direction theta. Let u^B the Born approsximation of the scattered wave for this incident field and a contrast function q. Let

$$g(\theta, s) := (Rq)(\theta, s) := u^B(L\,\theta + s\,\theta^\perp)$$

where

$$u^B(L\,\theta + s\,\theta^\perp) = \frac{k^2 i}{4} \int_{K_1(0)} H_0(k\,||L\,\theta + s\,\theta^\perp - y||)\, q(y)\, e^{ik\theta\cdot y}\, dy.$$

and $|L| > 1$ fixed. This is the measurement of the incident field behind ($L > 0$, transmission measurement) or before ($L < 0$, reflection measurement) the object. Let

$$a(\sigma) = \begin{cases} \pm\sqrt{k^2 - \sigma^2}, & \text{if } |\sigma| \le k \\ \pm i\sqrt{k^2 - \sigma^2}, & \text{otherwise} \end{cases}$$

where the $\pm$ is chosen as the sign of L. Note that a is real in the first and purely imaginary in the second case. Then

$$\widehat{Rq}(\theta, \sigma) = i\sqrt{\frac{\pi}{2}}\frac{k^2}{a(\sigma)} e^{i|L|a(\sigma)}\, \widehat{q}((a(\sigma) - k)\,\theta + \sigma\,\theta^\perp).$$

Korollar 11.8
Assume that $\widehat{g}(\theta, \sigma)$ is measured for all $\theta \in S^1$ and $|\sigma| \leq k$.
For fixed θ, $\widehat{q}$ can be computed on a sphere of radius k around $-k\theta$. For $L > 0$, the part of the semicircle containing the origin can be recovered, the rest from $L < 0$ (Ewald–sphere).
Let $L > 0$ (transmission measurement). Then using data for all $\theta \in S^1$, $\widehat{q}(\xi)$ can be computed from this data for $||\xi|| \leq \sqrt{2}k$.
Let $L < 0$ (reflection measurement). Then using data for all $\theta \in S^1$, $\widehat{q}(\xi)$ can be computed from this data for $2k \geq ||\xi|| \geq \sqrt{2}k$.
If measurements for both L and $-L$ are available, then $\widehat{q}(\xi)$ can be computed from this data for $||\xi|| \leq 2k$.

So the obtainable resolution is limited in the frequency domain by $2k$. This is known as the diffraction limit and corresponds to the Nyquist rate.

Korollar 11.9
In a sense, the projection theorem is a generalization of Fourier slice for smaller wavelength. For $\omega \to \infty$, the part of the sphere that is computed from transmission turns into a line through the origin in direction θ (as in Fourier slice). u has to be scaled by the incoming wave for the factors to vanish (Take $v = \frac{1}{k}\frac{u}{u^i}$, the projection theorem is often found in the literature in this form).

11.4 Solving nonlinear inverse problems with the Kaczmarz method

In this section, we look at a numerical method for solving inverse problems that involve a variant of Kaczmarz and a Newton–like algorithm. Note that other than in the Born approximation, we are solving the full nonlinear problem here. No convergence results are given, and we do this on an example only.

We rewrite the inverse scattering problem as an initial value problem. This can be done using the analytic solution of the Helmholtz equation in the exterior of S^1 in the previous chapter. Again, we look at everything in 2D. Our setting is:

For $\theta \in S^1$, u_θ is a solution to the perturbed Helmholtz initial value problem

$$\Delta u_\theta + k^2(1 + q(x))u_\theta = 0,\ u_\theta|_{L_1} = f_0,\ \frac{\partial}{\partial \nu} u_\theta|_{L_1} = f_1$$

where L_1 is a line perpendicular to θ before the object, i.e.

$$L_1 = \{-L\theta + s\theta^\perp,\ s \in \mathbb{R}\},\ L > 1.$$

As usual, we assume that supp $q \subset K_1(0)$.
For the data, we measure u_θ on a line L_2 behind the object (transmission measurement), i.e.

$$(Rq)(\theta, s) := (R_\theta q)(s) := g_\theta(s) = u_\theta|_{L_2}(s) = u_\theta(L\theta + s\theta^\perp).$$

As before, Rq is nonlinear (since it contains the products of the unknowns q and u_θ).
In the inverse problem, we search for a common solution q of the equations for a given measurement,

$$(R_\theta q) = g_\theta,\ \theta \in S^1.$$

This is exactly the situation of 9.1. Again, our iterative scheme is: Given a current guess $q^{(l)}$, select a fixed θ_k and update $q^{(l+1)} = q^{(l)} + dq$ using only the equation for θ_k.

Unfortunately, 9.1 only makes sense for linear operators. As in Newton's method, we linearize R_θ *at the current guess*, that is

$$R_\theta(q^{(l)} + dq) \sim R_\theta q^{(l)} + (R'_\theta(q^{(l)}))dq,\ dq \text{ small}.$$

where $R'_\theta(q^{(k)})$ is a linear operator. Note that the difference to the Born approximation is that we take the linearization not for q small (linearization around zero), but for dq small (linearization around $q^{(l)}$).

Then the nonlinear Kaczmarz method, for discrete θ_k and $R_k = R_{\theta_k}$, reads (very informal definition):

Definition 11.10 *(nonlinear Kaczmarz)*
Let $R_k : X \mapsto Y_k$, $k = 0 \dots p-1$, X, Y_k Hilbert spaces, R_k (possibly) nonlinear. Solve

$$R_k f = g_k,\ k = 0 \dots p-1.$$

Let $f^{(0)} \in X$. Choose ω fixed. Then the nonlinear Kaczmarz method is defined by

$$q^{(l+1)} = q^{(l)} + \omega R'_k(q^{(k)})^*(R'_k(q^{(k)})R'_k(q^{(k)})^*)^{-1}(g_k - R_k q^{(l)}),\ k = k(l).$$

Here, R'_k is a linearization of R_k, e.g. the Frechet–derivative.

The nonlinear method shares many properties of the linear method. Again, it is very efficient, and the inner inversion is usually replaced by an approximation (a constant times the identity matrix).

To apply it, it remains to compute the derivative and its adjoint. In the following, we fix θ_k and write $R := R_{\theta_k}$ for clarity.

Theorem 11.11
For the model problem,

$$R'(q)dq = \widetilde{du}|_{L_2}$$

where $\widetilde{du}$ is the solution to the initial value problem

$$\Delta\widetilde{du} + k^2(1+q(x))\widetilde{du} = -k^2 u\, dq,\ \widetilde{du}|_{L_1} = 0,\ \frac{\partial}{\partial\nu}\widetilde{du}|_{L_1} = 0$$

and u is the solution to the initial value problem

$$\Delta u + k^2(1+q(x))u = 0,\ u|_{L_1} = f_1,\ \frac{\partial}{\partial\nu}u|_{L_1} = f_2.$$

Observe that obviously this is linear in dq.

Remark: The notation is justified, in appropriate spaces, R' is the Frechet derivative of R.

Theorem 11.12
The adjoint of the derivative is given by

$$R'(q)^*\, g = -k^2 u\, v$$

where u is as above and v is a solution to the initial value problem

$$\Delta v + k^2(1+q(x))v = 0,\ v|_{L_2} = 0,\ \frac{\partial}{\partial\nu}v|_{L_2} = -g$$

(with respect to L^2).

Now we have everything set up, and Kaczmarz can be applied.