To the question "why, if a wave operator simulates forward propagation, its adjoint simulates backward propagation?"

Herman Jaramillo

March 28, 2015

Contents

1 Introduction

The objective here is to answer the title question. However, to answer that question we need a good understanding of the concept of "adjoint".

There are three important reasons to pay attention to the concept of the "adjoint".

- Practical applications in signal processing as shown below in reference to Claerbout's work.
- Solution of equations. See section 4.
- Speeding of computations. The case of Waveform Inversion (WFI) [3] and references in there, also see section 5

Jon Claerbout in his Classroom website offers his free book [1]. His chapter 2 is called "Adjoint Operators" and it provides a great insight into the concept of "adjoint" from the seismic signal processing point of view. I can not help but copy from his book table (1), which is very illustrative:

After browsing over Clearbout's book I almost can say that "adjoint" is his preferred word. One of the reasons why Claerbout likes this concept is because it has the "kinematics" of the inverse and it is much easier to compute that the inverse. The inverse should account for dynamics and kinematics; however with just kinematics we go a long way with a few bucks. When the adjoint is the inverse, we say that the operator is *unitary.* Unitary operators are desirable because they are easy to invert and the inverse is closer to what we want that the adjoint. If the operator is unitary we kill two birds with one stone. Unfortunately there are not many unitary operators out there. A few examples of unitary operators are the right (or left) shift (3.2.1.2.1), the discrete Fourier transform (3.2.1.3.4), the continuous Fourier transform (3.130), etc. The operators for seismic imaging (Kirchhoff migration, modeling or finite difference, migration and modeling are not unitary). One way to convert a non–unitary (non–singular) operator into a unitary operator is through the

[1] http://sep.stanford.edu/sep/prof/bei11.2010.pdf

Forward Operator	**Adjoint Operator**
matrix multiply	conjugate-transpose matrix multiply
convolve	crosscorrelate
truncate	zero pad
replicate, scatter, spray	sum or stack
spray into neighborhood	sum in bins
derivative (slope)	negative derivative
causal integration	anticausal integration
add functions	do integrals
assignment statements	added terms
plane-wave superposition	slant stack / beam form
superpose on a curve	sum along a curve
stretch	squeeze
upward continue	downward continue
hyperbolic modeling	normal moveout and CDP stack
diffraction modeling	imaging by migration
ray tracing	tomography

Table 1: Claerbout's insight about the adjoint operator.

Grand–Schmidt orthogonalization process, but the process itself is as expensive as inversion.

The approach that I will take here is a bit more formal than Claerbout's approach, but is not totally formal. The concept of adjoint is embedded on the Hilbert spaces and this is a subject of Functional Analysis. Functional analysis is a blending of algebra, analysis and topology. From those three subjects, topology is the less known to a non–mathematician. Analysis deals with sequences, limits, distances, derivatives, integrals and so forth; algebra deals mainly with operations between points in a space (here a vector space), and topology deals with concepts such as open and closed sets, compact spaces, connected sets, convex sets, etc. I will try to stay away from topology to make the document more accessible to no mathematicians at the expense of sacrifying some rigor.

I define, inner product, Hilbert space and then adjoint. With those basic definitions I build examples that illustrate the concept of adjoint (perhaps more than the reader want). Then I show my own answer to the question in mind. Finally I explain the Waveform Inversion (WFI) of the acoustic constant density wave equation.

2 Inner product

I assume that we know the definition of a vector space. Let us start with vector space V over a field $\mathbb{F}$. The field $\mathbb{F}$ is the space of scalars (complex or real). From Wikipedia, an inner product space is a a vector space V over the field $\mathbb{F}$ together with an *inner product,* i.e., with a map

$$\langle\, .\, ,\, .\rangle : V \times V \to \mathbb{F} \tag{2.1}$$

that satisfies the following three axioms for all vectors $x, y, z \in V$ and all scalars $a \in \mathbb{F}$.

1. Conjugate symmetry: $\langle x\, , y\rangle = \overline{\langle y\, , x\rangle}$
2. Linearity in the first argument:
   $\langle a\, x\, , y\rangle = a\langle x\, , y\rangle$
   $\langle x + y\, , z\rangle = \langle x\, , z\rangle + \langle y\, , z\rangle$
3. Positive–definiteness $\langle x\, , x\rangle \geq 0$ with equality only for $x = 0$.

2.1 Examples of inner products

- In the n–dimensional vector space over the complex numbers $\mathbb{C}^n$. Assume

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{pmatrix} \quad \text{and} \quad \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix} \tag{2.2}$$

  The inner product of the two vectors is

$$\langle \boldsymbol{x}\,, \boldsymbol{y} \rangle = \boldsymbol{y}^H \boldsymbol{x} = \sum_{i=1}^{n} \overline{y_i}\, x_i \tag{2.3}$$

  where $\boldsymbol{x}^H$ is the conjugate transposed (Hermitian) of vector $\boldsymbol{x}$.

  If we have infinity sequences of x_i and y_i, the previous formula is extended to

$$\langle \boldsymbol{x}\,, \boldsymbol{y} \rangle = \boldsymbol{y}^H \boldsymbol{x} = \sum_{i=n}^{m} \overline{y_i}\, x_i \tag{2.4}$$

  where here n could be any integer number including $-\infty$ and $m \geq n$ is any integer number including ∞.

  Sometimes the inner product is defined on this space relative to a matrix A which is positive definite. The definition is

$$\langle \boldsymbol{x}\,, \boldsymbol{y} \rangle = \boldsymbol{y}^H A\, \boldsymbol{x} = \sum_{i=1}^{n} a_{ij}\, \overline{y_i}\, x_j \tag{2.5}$$

  It is common to see in statistics that A is the inverse covariance matrix. The matrix A can be interpreted as a weight or constraint.

- In the continuum. Assume the complex valued functions f and g are in the space of continuous functions defined in the closed real interval $[a, b]$. We define the inner product of f by g as

$$\langle u\,, v \rangle = \int_a^b u(x)\overline{v(x)} dx. \tag{2.6}$$

The interval $[a, b]$ could be extended to have $-\infty$ or ∞ in one of its ends or both ends.

We can generalize. The functions f and g do not have to be continuous but the integral (2.6) should exist. Now in general, if there is a positive integrable function (weight) w such that

$$\langle u\,, v\rangle = \int_a^b u(x)\overline{v(x)}w(x)dx. \tag{2.7}$$

exists for any u and v then equation (2.7) defines an inner product. We also assume that the real weight is $w(x) > 0$ as in the discrete case.

Now we have the elements to define the adjoint operator in a Hilbert space.

3 The adjoint operator

A Hilbert space is a vector space with an inner product and the condition that any sequence with elements in the space has a limit point which is in the space. We will not pay attention to this last condition in what follows but it is important for developing the calculus in Hilbert spaces. A Hilbert space is a generalization of the real numbers $\mathbb{R}$. Examples of Hilbert spaces are everywhere in mathematical physics. The set of real numbers $\mathbb{R}$, the set of complex numbers $\mathbb{C}$, a finite and infinite product of sets of real or complex numbers (with some constraints), the space of continuous functions in a close interval $[a, b]$ with any k number of derivatives $C^k[a, b]$, a product of sets of continuous functions with a given number of derivatives, etc. In all these spaces a proper definition of inner product should be addressed so that the convergence conditions will apply.

If L is linear operator from a Hilbert space $\mathcal{H}_1$ to a Hilbert space $\mathcal{H}_2$ ($L : \mathcal{H}_1 \to \mathcal{H}_2$), we want to find a linear operator L^* named the adjoint of L such that:

$$L^* : \mathcal{H}_2 \to \mathcal{H}_1. \tag{3.1}$$

satisfies

$$\langle Lx\,, y\rangle = \langle x\,, L^*y\rangle. \tag{3.2}$$

for all $x \in \mathcal{H}_1$, and $y \in \mathcal{H}_2$. Observe that the left inner product is defined in the space $\mathcal{H}_2$ and the right inner product is defined in $\mathcal{H}_1$.

If L is bounded, the existence is guaranteed. [2] We will study the conditions of existence case by case basis, since some of the operators that we will study here, which are necessary for the development of the topic, are not bounded.

3.1 Basic Properties of Adjoint operators

I list the basic properties of the adjoint operators without proof. For proofs of the following identities see [5].

Let as assume two linear operators A and B and any complex scalar λ.

- Symmetry:

$$(A^*)^* = A \tag{3.3}$$

  This property reduces my list of examples below to half. When I prove that the divergence is the adjoint of the gradient (3.2.2.7.1), then I do not have to prove that the gradient is the adjoint of the divergence.

- Linearity:

$$(A + B)^* = A^* + B^* \quad \text{and} \quad (\lambda A)^* = \overline{\lambda} A^*. \tag{3.4}$$

  Thanks to this property we only derive only adjoint operators for simple operators and then build more complicated operators as superposition of those simple operators. For example we derive the adjoint of a first order and second order differentiator and then use this property to show the adjoint of a general second order differential equation.

- Anticonmmutativity:

$$(A\,B)^* = B^*\,A^*. \tag{3.5}$$

  When we know the adjoint of two operators we also know the adjoint of its composition through this rule.

[2] I will not define bounded here or prove this existence theorem. The reader is referred to [5] for a good introduction in Functional Analysis where all these things are shown in great detail.

- Norm preserving:

$$\|A^*\| = \|A\|. \tag{3.6}$$

For the definition of a norm of an operator see [5]. I will not use this property in this document, however I believe it is interesting.

3.2 Examples of adjoint operators

3.2.1 The discrete case

3.2.1.1 The matrix operator: A linear operator that sends an n–dimensional complex vector into an m–dimensional complex vector can be represented by a complex matrix A of order $m \times n$. We write

$$\begin{aligned} L : \mathbb{R}^n &\rightarrow \mathbb{R}^m \\ \boldsymbol{x} &\mapsto \boldsymbol{y} = A\boldsymbol{x}. \end{aligned} \tag{3.7}$$

From the definition (3.2) we will find a matrix A^* such that

$$\langle A\boldsymbol{x}\,, \boldsymbol{y}\rangle = \langle \boldsymbol{x}\,, A^*\boldsymbol{y}\rangle \tag{3.8}$$

for all possible $\boldsymbol{x}$ and $\boldsymbol{y}$ vectors. Let us choose the two vectors: $\boldsymbol{x} = \hat{\boldsymbol{e}}_{\boldsymbol{j}}$ and $\boldsymbol{y} = \hat{\boldsymbol{e}}_{\boldsymbol{i}}$, the canonical vectors with all zero components except for a component 1 in the j and i entries respectively, so we find:

$$\langle A\boldsymbol{x}\,, \boldsymbol{y}\rangle = \hat{\boldsymbol{e}}_{\boldsymbol{i}}^H\, A\, \hat{\boldsymbol{e}}_{\boldsymbol{j}} = a_{ij} \tag{3.9}$$

and

$$\langle \boldsymbol{x}\,, A^*\boldsymbol{y}\rangle = \hat{\boldsymbol{e}}_{\boldsymbol{i}}^H\, (A^*)^H \hat{\boldsymbol{e}}_{\boldsymbol{j}} = \bar{a}^*_{ji} \tag{3.10}$$

so from the equality (3.8) and the previous two equations we find

$$a_{ij} = \bar{a_{ji}}^* \quad , \quad \text{that is} \quad a^*_{ij} = \bar{a}_{ji} \tag{3.11}$$

That is, the matrix A^* is the transposed conjugate (Hermitian) of A. So the adjoint of matrix is its Hermitian. We write:

$$\begin{aligned} L^* : \mathbb{R}^m &\rightarrow \mathbb{R}^n \\ \boldsymbol{y} &\mapsto \boldsymbol{x} = A^*\boldsymbol{y}. \end{aligned} \tag{3.12}$$

3.2.1.2 Examples in the Hilbert space l^2 of infinite sequences. The l^2 space is the space of infinity sequences $\{x_n, x_{n+1}, x_{n+2}...x_m\}$ (complex or real) where n could be any integer number including $-\infty$ and $m \geq n$ any integer number including ∞. Usually the set is defined with $n = 0$ and $m = \infty$ but we want to be more general for reasons that will be clear soon. The "2" in l^2 comes from the fact that we require finite energy. That is

$$E = \sum_{i=n}^{m} |x_i|^2 < \infty. \tag{3.13}$$

From all the spaces l^p [3], only the space l^2 is a Hilbert space [5] .

3.2.1.2.1 The right shift operator. Given a sequence $\{x_i\} \in l^2$ we define the right shift operator L as

$$\begin{aligned} L : l^2 &\rightarrow l^2 \\ \{x_i\} &\mapsto L(\{x_i\}) = \{x_{i+1}\}. \end{aligned} \tag{3.15}$$

It takes every sample into the next. Here we let the upper index m to go to ∞ so there is not a last sample. For this case, that is good news because otherwise the last sample would be lost. We will see that boundaries make things more complicated. From now on, I will avoid writing the braces around sequences (to simplify typing and avoid clustering of symbols). That is I write

$$\begin{aligned} L : l^2 &\rightarrow l^2 \\ x_i &\mapsto L(x_i) = x_{i+1}, \end{aligned} \tag{3.16}$$

instead of expression (3.15). And it is understood that x_i could mean a sequence on l^2 or the i–th element of that sequence. The context will determine the particular meaning of the symbol.

Let us find L^*. Given $u = \{x_i\}$ and $v = \{y_i\}$ in l^2 we have from the inner product definition (2.4) that

$$\langle Lu\,, v\rangle = \sum_{i=n}^{\infty} x_{i+1}\bar{y}_i = \sum_{i=n+1}^{\infty} x_i\bar{y}_{i-1}. \tag{3.17}$$

[3] For l^p it is required that

$$E = \sum_{i=n}^{m} |x_i|^p < \infty. \tag{3.14}$$

At this point we want to change $n+1$ to n, but this is only possible if $n = -\infty$. If this is the case we can write:

$$\langle Lu\,, v\rangle = \sum_{i=-\infty}^{\infty} x_i \bar{y}_{i-1} = \langle u\,, L^* v\rangle. \tag{3.18}$$

where L^* is the left shift operator defined by:

$$\begin{aligned} L^* : l^2 &\rightarrow l^2 \\ x_i &\mapsto L^*(x_i) = x_{i-1}. \end{aligned} \tag{3.19}$$

We can go further and say that L^* is the inverse of L. That is, $LL^* = L^*L = I$. This is obvious since shifting one place to the right, and then to the left, would leaves us at the starting point. We see then that L is a unitary operator.

3.2.1.2.2 The forward difference operator. For simplicity assume that we sampled a continuous function with a sampling rate $\Delta t = 1$. The definition of a forward difference operator is given by:

$$\begin{aligned} L : l^2 &\rightarrow l^2 \\ x_i &\mapsto L(x_i) = x_{i+1} - x_i. \end{aligned} \tag{3.20}$$

That is

$$L\{x_i\} = R\{x_i\} - I\{x_i\} \tag{3.21}$$

where R is the right shift operator and I is the identity.

We use linearity (3.4) in (3.21) to find that

$$L^*(x_i) = x_{i-1} - x_i = -L(x_i) \tag{3.22}$$

That is, the adjoint of the forward operator is the backward operator, which happens to be the negative of the operator L. We will observe (see (3.2.2.1)) that this happens also in the continuum.

3.2.1.2.3 Convolution. The convolution (with a sequence $\{z_i\}$) operator is defined by

$$\begin{aligned} L_z : l^2 &\rightarrow l^2 \\ x_i &\mapsto L_z(x_i) = \sum_{j=-\infty}^{\infty} x_j z_{i-j}. \end{aligned} \tag{3.23}$$

Let us find L^*. Given $u = \{x_i\}$ and $v = \{y_i\}$ in l^2 we have from the inner product definition (2.4) and letting m, n be ∞ and $-\infty$ respectively that

$$\begin{aligned}
\langle Lu\,, v\rangle &= \sum_{i=-\infty}^{\infty}\left(\sum_{j=-\infty}^{\infty} x_j z_{i-j}\right)\bar{y}_i \\
&= \sum_{j=-\infty}^{\infty} x_j\left(\sum_{i=-\infty}^{\infty} z_{i-j}\bar{y}_i\right) \\
&= \langle u\,, L^*v\rangle
\end{aligned}$$

where

$$\begin{aligned}
L_z^* : l^2 &\rightarrow l^2 \\
y_i &\mapsto L_z(y_i) = \sum_{j=-\infty}^{\infty} y_i \bar{z}_{i-j}
\end{aligned} \tag{3.24}$$

is the cross–correlation of $\{y_i\}$ with $\{z_i\}$. So the adjoint of convolution is conjugate cross–correlation.

3.2.1.3 Examples in the Hilbert space $\mathbb{R}^n$ Things are much harder here, because of the boundaries. I will try to formulate the same problems as above but have to be more careful with the boundaries.

3.2.1.3.1 The first order forward difference differentiator. Assume we have 5 samples x_0, x_1, x_2, x_3, x_4 [4]. The forward difference operator for the first 4 samples would be something like

$$\begin{aligned}
\text{first} &= x_1 - x_0 \\
\text{second} &= x_2 - x_1 \\
\text{third} &= x_3 - x_2 \\
\text{fourth} &= x_4 - x_3
\end{aligned} \tag{3.25}$$

It is natural to think that the firth sample would be $x_5 - x_4$ but we do not have x_5 and this is the type of problems we will find all over the place

[4] Note that I start indexing in 0. Perhaps staring in 1 is more natural but being a programmer and programming in C for a few years, I can not help but start in 0.

when cutting the size of the data to fit finite models. We might want to say fifth $= -x_4$ as a boundary condition (assuming that if there is a x_5 this would be zero), but of course this is not the only option and we should understand that different boundary conditions produce different adjoint operators.

In general, assume we want to find the differentiation of a function sampled in an n–dimensional vector $\boldsymbol{x}$ with uniform sampling rate Δt. If we use, for example forward differences, we can apply the rule

$$\begin{aligned} L : \mathbb{R}^n &\rightarrow \mathbb{R}^n \\ x_i &\mapsto L(x_i) = \frac{1}{\Delta t} \begin{cases} x_{i+1} - x_i & i = 0..n-2 \\ -x_{n-1} & i = n-1 \end{cases} \end{aligned} \tag{3.26}$$

The L operator can be represented by the matrix vector multiplication $A\boldsymbol{x}$ where

$$A = \frac{1}{\Delta t} \begin{pmatrix} -1 & 1 & 0 & ... & & 0 \\ 0 & -1 & 1 & 0 & ... & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & ... & 0 & -1 & 1 \\ 0 & 0 & ... & 0 & 0 & -1 \end{pmatrix} \tag{3.27}$$

The last raw breaks the $\{-1, 1\}$ pattern.

The transpose of the matrix A is given by:

$$A^* = \frac{1}{\Delta t} \begin{pmatrix} -1 & 0 & 0 & ... & & 0 \\ 1 & -1 & 0 & 0 & ... & 0 \\ . & 1 & -1 & 0 & . & 0 \\ . & . & . & . & . & . \\ 0 & 0 & ... & 1 & -1 & 0 \\ 0 & 0 & ... & 0 & 1 & -1 \end{pmatrix} \tag{3.28}$$

Here it is the first raw of the matrix the one that breaks the pattern (now this pattern is $\{-1, 1\}$ which happens to be the negative of the pattern in the forward operator). So up to the boundaries the adjoint of the discrete differentiator is its negative. We will see in the next section that this is the case on the continuum but the boundaries should be set to zero.

We write:

$$\begin{aligned} L^* : \mathbb{R}^n &\rightarrow \mathbb{R}^n \\ x_i &\mapsto L(x_i) = \frac{1}{\Delta t} \begin{cases} -x_0 & i = 0 \\ x_i - x_{i+1} & i = 1..n-1 \end{cases} \end{aligned} \tag{3.29}$$

We say then that adjoint of the forward finite difference first order operator is the backward differences first order operator and the boundary condition that was at the upper end point in the forward operator was moved to the lower end point of the adjoint operator.

3.2.1.3.2 The second order central differences differentiator. As in the previous example, let us assume we have 5 samples x_0, x_1, x_2, x_3, x_4. The second order central differences, starting at the second point are defined as:

$$\begin{aligned} \text{second} &= x_2 - 2x_1 + x_0 \\ \text{third} &= x_3 - 2x_2 + x_1 \\ \text{fourth} &= x_4 - 2x_3 + x_2. \end{aligned} \tag{3.30}$$

We wish to extend this as follows:

$$\begin{aligned} \text{first} &= x_1 - 2x_0 + x_{-1} \\ \text{fifth} &= x_5 - 2x_4 + x_3. \end{aligned} \tag{3.31}$$

But, of course, there is no x_{-1} and x_5 samples. So the "best" we can do is

$$\begin{aligned} \text{first} &= x_1 - 2x_0 \\ \text{fifth} &= -2x_4 + x_3. \end{aligned} \tag{3.32}$$

that is we assume $x_5 = x_{-1} = 0$ (since they are not present).

In general, assume a set of values x_i sampled at a rate Δt. We can represent the second order central difference by the formula:

$$\begin{aligned} L : \mathbb{R}^n &\rightarrow \mathbb{R}^n \\ x_i &\mapsto L(x_i) = \frac{1}{\Delta t} \begin{cases} x_1 - 2x_0 & i = 0 \\ x_{i+1} - 2\,x_i + x_{i-1} & 0 < i < n-1 \\ x_{n-1} - 2x_{n-2} & i = n-1 \end{cases} \end{aligned}$$

The computation of the second derivative operator can be seen as a matrix vector multiplication where the operator L is associated with the matrix A,

which is

$$A = \frac{1}{(\Delta t)^2} \begin{pmatrix} -2 & 1 & 0 & ... & & 0 \\ 1 & -2 & 1 & 0 & ... & 0 \\ 0 & 1 & -2 & 1 & . & . \\ . & . & . & . & . & . \\ 0 & 0 & ... & 1 & -2 & 1 \\ 0 & 0 & ... & 0 & 1 & -2 \end{pmatrix} \tag{3.33}$$

It is clear that A is a symmetric matrix. That is $A = A^*$. When an operator is equal to its adjoint we say that the operator is *self–adjoint* . will see that in the continuum the second derivative agrees with this property.

3.2.1.3.3 An integrator. Think about a transform such as

$$Lu = \int_a^b k(s,t)u(t)dt. \tag{3.34}$$

This operator is known as the Fredholm integral operator. All integral (Fourier, Hankel, Laplace.....etc.) fall in this model (a and b could reach to $\pm\infty$ or even be complex numbers.) Here assume a and b are real and finite. We also assume that the Kernel $k(s,t)$ is sampled in the s, t domain spaces as well as the function u with a sampling rates Δs and Δt respectively. Then equation (3.34) can be represented as a matrix vector multiplication

$$\begin{aligned} L : \mathbb{R}^n &\rightarrow \mathbb{R}^m \\ \boldsymbol{y} &= A\,\boldsymbol{x} \end{aligned} \tag{3.35}$$

where $a_{ij} = k(s_0 + i\Delta s, t_0 + j\Delta t)$ and $x(i) = u(t_0 + i\Delta t)\Delta t$, with s_0 and t_0 are the initial (low boundary) points. So $a^*_{ij} = \bar{k}(s_0 + j\Delta s, t_0 + i\Delta t)$. If both s and t are in the same space, with the same origin t_0 and increment Δt (not usually the case) then $a^*_{ij} = \bar{k}(t_0 + j\Delta t, t_0 + i\Delta t) = (t_j, t_i)$, with $t_k = t_0 + k\Delta t$. This suggest that the adjoint L^* of the integral operator with kernel $k(s,t)$ is $\bar{k}(t,s)$.

3.2.1.3.4 The discrete Fourier transform. A sequence of n complex numbers $x_0, ..., x_{N-1}$ can be transformed to another sequence of N com-

plex numbers following the rule:

$$\begin{aligned} L : \mathbb{C}^N &\rightarrow \mathbb{C}^N \\ x_k &\mapsto X_k = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} x_i e^{-\mathrm{i}\ 2\pi ki/N} \end{aligned} \tag{3.36}$$

L is a linear operator. Let us find L^*. As usual, given $u = x_i$ and $v = y_i$ both in $\mathbb{C}^n$ we find

$$\begin{aligned} \langle Lu\,, v\rangle &= \sum_{j=0}^{N-1} X_j\, \overline{y}_j \\ &= \sum_{j=0}^{N-1} \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} x_j e^{-\mathrm{i}\ 2\pi kj/N}\ \bar{y}_j \\ &= \sum_{j=0}^{N-1} x_j \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} e^{-\mathrm{i}\ 2\pi kj/N}\ \bar{y}_j \\ &= \sum_{j=0}^{N-1} x_j \overline{Y}_j \end{aligned} \tag{3.37}$$

where

$$Y_j = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} y_i\, e^{\mathrm{i}\ 2\pi ki/N} \tag{3.38}$$

So

$$\begin{aligned} L^* : \mathbb{C}^N &\rightarrow \mathbb{C}^N \\ y_k &\mapsto Y_k = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} x_i e^{\mathrm{i}\ 2\pi ki/N} \end{aligned} \tag{3.39}$$

which actually is known as the inverse discrete Fourier transform. So the adjoint of the forward Fourier transform is the inverse discrete Fourier transform. The discrete Fourier transform is a unitary operator.

3.2.1.3.5 A causal integrator. Assume a vector of sample values x_i with sampling rate Δt. Causal integration is expressed as $y_i = y_{i-1} + x_i\, \Delta t$

and $y_{-1} = 0$. $i = 0..n-1$. Causal integration can be writen as a matrix vector product $A\boldsymbol{x}$ where the matrix A is given by:

$$A = \Delta t \begin{pmatrix} 1 & 0 & 0 & & ... & 0 \\ 1 & 1 & 0 & & ... & 0 \\ 1 & 1 & 1 & 0 & ... & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & 1 & ... & 1 & 1 & 1 \end{pmatrix} \tag{3.40}$$

which is a lower triangular matrix. The adjoint matrix A^* is given by the upper triangular matrix:

$$A^* = \Delta t \begin{pmatrix} 1 & 1 & 1 & ... & & 1 \\ 0 & 1 & 1 & ... & & 1 \\ 0 & 0 & 1 & ... & & 1 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & 0 & ... & 0 & 1 \end{pmatrix} \tag{3.41}$$

The recurrence relation here is $x_{i-1} = x_i + y_{i-1}\Delta t$. Here the y_j components are the input while the x_j components are the output; $i = n..0$ and we assume $x_n = 0$. We see that the adjoint of causal integration is anti–causal integration.

3.2.2 The continuum

3.2.2.1 The first order derivative The two "most" important operators in the continuum are the derivative and the integral.

Let us start with the one–dimensional differentiator of real functions of real values (so we do not have to deal with conjugates). To make my life easier I will assume that we live in the Hilbert space of $C^\infty[a,b]$. Why? [5]

$$Lu = \frac{du}{dx}. \tag{3.42}$$

[5]Here is why. If we pick the space of continuous differentiable functions $C^1[a,b]$ then the differential will send functions from the $C^1[a,b]$, space into the $C^0[a,b]$ space. The adjoint would have to be of the form $L^* : C^0[a,b] \to C^1[a,b]$. We will show that if the adjoint exists then it is also a differential and $C^0[a,b]$ has a bunch of non–differentiable functions (say for example $f(x) = |x|, atx = 0$).

Using definition (2.6) we have that

$$\begin{aligned}\langle Lu\,,v\rangle &= \int_a^b \frac{du}{dx}\,v dx \\ &= u\,v\,\Big|_a^b - \int_a^b u\frac{dv}{dx} \quad \text{(integration by parts)} \\ &= u(b)v(b) - u(a)v(a) + \left\langle u\,, -\frac{dv}{dx}\right\rangle.\end{aligned} \tag{3.43}$$

We are tempted to say that the adjoint of the differentiator operator is its negative. That is we wish to say

$$L^* = -\frac{d}{dx}; \tag{3.44}$$

however, unless

$$u(b)v(b) - u(a)v(a) = 0 \tag{3.45}$$

for all functions u and v, we can not comply with the definition of adjoint (3.2). Hence, the definition of adjoint of a derivative makes sense only for the space of all functions that vanish in the boundary of the interval $[a, b]$. Because in differential equations, not always we want solutions with zero boundary conditions, Keener [4] explains how to extend the operators to consider the case of non–zero boundary conditions. This extension is beyond the scope of this document.

3.2.2.2 A scaled first order derivative Let us consider a bit more general first order derivative operator

$$Lu = c(x)\frac{du}{dx}. \tag{3.46}$$

Using definition (2.6) we have that

$$\begin{aligned}\langle Lu\,,v\rangle &= \int_a^b c(x)\,\frac{du}{dx}\,v\,dx, \\ &= \quad c\,u\,v\,\Big|_a^b - \int_a^b u(x)\frac{d(cv)}{dx}\,dx, \\ &= \quad c(b)\,u(b)\,v(b) - c(a)\,u(a)\,v(a) + \left\langle u\,, -\frac{dcv}{dx}\right\rangle.\end{aligned}$$

where again, we used integration by parts. So the adjoint operator exists provided that

$$c(b)\,u(b)\,v(b) - c(a)\,u(a)\,v(a) = 0 \tag{3.47}$$

and

$$L^*u(x) = -\frac{d}{dx}\left(c(x)u(x)\right). \tag{3.48}$$

3.2.2.3 The second order derivative: Let us assume that the operator L is a second order differentiator. That is,

$$Lu = \frac{d^2u}{dx^2} \tag{3.49}$$

Using definition (2.6) we find that

$$\begin{aligned}
\langle Lu\,,v\rangle &= \int_a^b \frac{d^2u}{dx^2}\,v dx \\
&= \frac{du}{dx}\,v\,|_a^b - \int_a^b \frac{du}{dx}\frac{dv}{dx} \quad \text{(integration by parts)} \\
&= \frac{du(b)}{dx}v(b) - \frac{du(a)}{dx}v(a) - \int_a^b \frac{du}{dx}\frac{dv}{dx}dx \\
&= \frac{du(b)}{dx}v(b) - \frac{du(a)}{dx}v(a) - u\frac{dv}{dx}\Big|_a^b + \int_a^b u\frac{d^2v}{dx^2} \\
&= \frac{du(b)}{dx}v(b) + u(a)\frac{dv(a)}{dx} - \frac{du(a)}{dx}v(a) - u(b)\frac{dv(b)}{dx} + \left\langle u\,,\frac{d^2v}{dx^2}\right\rangle.
\end{aligned} \tag{3.50}$$

Integration by parts was used twice. In second and fourth steps. As in the first derivative case we need the function

$$\frac{du(b)}{dx}v(b) + u(a)\frac{dv(a)}{dx} - \frac{du(a)}{dx}v(a) - u(b)\frac{dv(b)}{dx}a \tag{3.51}$$

to vanish in the boundary and the first derivatives on the boundary to be finite. If this is the case, then the adjoint of the second derivative operator is itself. This agrees with the discrete case where we found out that the second order central difference operator was self–adjoint.

Equation (3.51) can be rewritten as:

$$J(u,v) = \frac{du}{dx}v - u\frac{dv}{dx}\bigg|_a^b = W(v,u)|_a^b \tag{3.52}$$

where $W(v,u)$ is the Wronskian of u and v. The notation $J(u,v)$ is taken from Keener's book. I though at first it was a Jacobian but it is not. Let us consider a more general second order differential operator

3.2.2.4 A scaled second order derivative Following the pattern we now find the adjoint of a scaled second order derivative operator. That is, define

$$Lu = c(x)\frac{d^2u}{dx^2}. \tag{3.53}$$

Using definition (2.6) we find that

$$\begin{aligned}
\langle Lu\,,v\rangle &= \int_a^b c(x)\,\frac{d^2u}{dx^2}\;v\,dx \\
&= \frac{du}{dx}\,c\,v\bigg|_a^b - \int_a^b \frac{du}{dx}\,\frac{d}{dx}(cv)\,dx \\
&= \frac{du}{dx}\,c\,v - u\frac{d}{dx}(cv)\;\bigg|_a^b + \int_a^b u\,\frac{d^2}{dx^2}(cv)\,dx \\
&= \frac{du}{dx}\,c\,v - u\frac{d}{dx}(cv)\;\bigg|_a^b + \langle u\,,\frac{d^2}{dx^2}(cv)\rangle.
\end{aligned}$$

So provided that

$$J(u,v) = W(cv,u)|_a^b = 0 \tag{3.54}$$

we can say that

$$L^*[v(x)] = \frac{d^2}{dx^2}\left(c(x)v(x)\right). \tag{3.55}$$

From equation (3.4) and the fact that the adjoint of $Lu = a(x)u$ for a scalar real $a(x)$ is the same

$$Lv = a(x)v, \tag{3.56}$$

combined with the equations (3.48) and (3.55) we find the more general second order differential operator as we see next.

3.2.2.5 The second order differential operator: A general second order ordinary differential operator is given by

$$Lu = a_2(x)\,\frac{d^2u}{dx^2} + a_1(x)\frac{du}{dx} + a_0(x)u. \tag{3.57}$$

We use equations (3.48), (3.55), and (3.56) together with the addition property (3.4) to find that

$$L^*v = \frac{d^2}{dx^2}(a_2(x)v) - \frac{d}{dx}(a_1(x)v) + a_0(x)\,v, \tag{3.58}$$

provided that the boundary condition (the superposition of both boundary conditions for first and second derivative examples above):

$$J(u,v) = w[u(x), a_2(x)v(x)] + a_1(x)\,u(x)\,v(x)|_a^b = 0 \tag{3.59}$$

are satisfied.

3.2.2.6 The Sturm–Liouville operator: This operator is one of the most important in mathematical physics, and by the end of this section the reader should understand why.

The Sturm–Liouville operator is defined by the equation

$$Lu = \frac{1}{\omega(x)}\,\frac{d}{dx}\left(p(x)\frac{du}{dx}\right) + q(x)u, \quad \text{where} \quad \omega(x) > 0. \tag{3.60}$$

Actually, the classical Sturm–Liouville operator is defined with $\omega(x) = 1$ and can be naturally obtained from the general second order differential equation (3.57) after assuming $a_2' = a_1$, but equation (3.60) provides a useful generalization. We show that this operator is formally, self–adjoint with respect to the inner product defined in equation (2.7), provided certain boundary conditions that we will derive.

Let us see:

$$\begin{aligned}
\langle Lu\,,v\rangle &= \int_a^b \left[\frac{v(x)}{\omega(x)}\frac{d}{dx}\left(p(x)\frac{du}{dx}\right)+q(x)u(x)v(x)\right]\omega(x)dx \\
&= \int_a^b \left[v(x)\frac{d}{dx}\left(p(x)\frac{du}{dx}\right)+q(x)u(x)v(x)\omega(x)\right]dx \qquad (3.61)\\
&= v(x)p(x)\frac{du}{dx}\bigg|_a^b - \int_a^b \left[p(x)\frac{du}{dx}\frac{dv}{dx}-q(x)u(x)v(x)\omega(x)\right]dx \\
&= \left[v(x)p(x)\frac{du}{dx}-u(x)p(x)\frac{dv}{dx}\right]\bigg|_a^b + \\
&\qquad \int_a^b \left[u(x)\frac{d}{dx}\left(p(x)\frac{dv}{dx}\right)+q(x)u(x)v(x)\omega(x)\right]dx. \\
&= \left[v(x)p(x)\frac{du}{dx}-u(x)p(x)\frac{dv}{dx}\right]\bigg|_a^b + \\
&\qquad \int_a^b \left[\frac{1}{\omega(x)}\frac{d}{dx}\left(p(x)\frac{dv}{dx}\right)+q(x)v(x)\right]u(x)\,\omega(x)dx \\
&= \langle u\,,Lv\rangle,
\end{aligned}$$

provided that

$$J(u,v) = p(x)\left[v(x)\frac{du}{dx}-u(x)\frac{dv}{dx}\right]\bigg|_a^b = p(x)\,W\left[v(x),u(x)\right]\big|_a^b = 0 \qquad (3.62)$$

Observe how convenient was to cancel $\omega(x)$ in the second step and preserve it in the second term of this second step. The third and fourth steps (as usual) are integration by parts.

If a vector ϕ satisfies the equation

$$L\,\phi = \lambda\phi, \qquad (3.63)$$

for some complex number λ we say that ϕ is an *eigenvector* and λ is an *eigenvalue* for the operator L. The set of all eigenvalues is known as the spectrum of L.

One interesting thing about self–adjoint operators is that the eigenfunctions u_1 and u_2 corresponding to two different eigenvalues λ_1 and λ_2 are orthogonal. That is their inner product is zero or $\langle u_1\,u_2\rangle = 0$. The proof is

Name associated with function	$p(x)$	$q(x)$	$r(x)$	interval	Boundary conditions	eigenvalues n=1,2...,	Normalized eigenvectors
Fourier	1	0	1	$-[\pi, \pi]$	$y(\pi) = y(-\pi)$ $y'(\pi) = y'(-\pi)$	$0, n^2$	$1, \sin nx, \cos nx$
Bessel	x	$k^2 x^{-1}$	x	$[0, 1]$	$y(0) =$ finite $y(1) = 0$	$\mu_n, J_k(\mu_n) = 0$	$J_k(\mu_n(x)$
Legendre polynomials	$1 - x^2$	0	1	[-1,1]	$y(-1)$ and $y(1)$ finite	$n(n+1)$	$P_n(x)$
Tchebycheff polynomials	$(1-x^2)^{1/2}$	0	$(1-x^2)^{-1/2}$	[-1,1]	$y(-1)$ and $y(1)$ finite	n^2	$T_n(x)$
Hermite polynomials	$\exp(-x^2)$	0	$\exp(-x^2)$	$(-\infty, \infty)$	As $\|x\| \to \infty$, $y = \mathcal{O}(x^k)$ for some $k > 0$	$0, 2n$	$H_0(x), H_n(x)$
Laguerre	$x \exp(-x)$	0	$x \exp(-x)$	$[0, \infty)$	$y(0)$ finite; as $x \to \infty$, $y = \mathcal{O}(x^k)$ for some $k > 0$	n	$L_n(x)$

Table 2: Properties of the Sturm–Liouville operator.

easy and we can include it here. Let us assume that λ_1 and λ_2 are eigenvalues of L with eigenvectors u_1 and u_2. That is,

$$Lu_1 = \lambda_1 u_1 \qquad \text{and} \qquad Lu_2 = \lambda_2 u_2. \tag{3.64}$$

So

$$\langle Lu_1 \, , Lu_2 \rangle = \langle \lambda_1 u_1 \, , \lambda_2 u_2 \rangle = \lambda_1 \overline{\lambda_2} \langle u_1 \, , u_w \rangle. \tag{3.65}$$

On the other hand

$$\begin{aligned} \langle Lu_1 \, , Lu_2 \rangle &= \langle u_1 \, , L(Lu_2) \rangle \\ &= \langle u_1 \, , L(\lambda_2 u_2) \rangle \\ &= \langle u_1 \, , \lambda_2^2 u_2) \rangle \\ &= \overline{\lambda_2^2} \langle u_1 \, , u_2 \rangle \end{aligned} \tag{3.66}$$

so because $\lambda_1 \neq \lambda_2$, then $\lambda_1 \overline{\lambda_2} \neq \overline{\lambda_2^2}$ and so, the only way that equations (3.65) and (3.66) are equal is if $\langle u_1 \, , u_2 \rangle = 0$. So that the eigenfunctions u_1 and u_2 are orthogonal. We will have much more to say about the spectrum of an operator in the section 4.3.3. One of the nice properties of the Sturm–Liouville operator is that it generates a great deal of orthogonal polynomials used in the mathematical physics. Table (2) (taken from [6]) illustrates this.

3.2.2.7 Partial differential equations: I follow the same methodology used for ordinary differential equations. There, I found the adjoint for the following sequence: the first derivative, the first derivative with a factor, a second derivative, a second derivative with a factor, and then jumped to a more general second order ordinary differential equation.

However the multi–dimensional differential operators come in many flavors. For example, while in one dimension we only have a first derivative operator, in several dimensions we find for scalar fields gradient (∇) and for vector fields divergence ($\nabla.$) and curl ($\nabla\times$). All of them use first order partial derivatives. In the case of second order I only explore the Laplacian operator (∇^2). The trick to find adjoints is always integration by parts. Integration by parts comes from the derivative of the product of two functions. In symbols, if

$$d(uv) = udv + vdu, \tag{3.67}$$

then

$$uv = \int (udv + vdu) \tag{3.68}$$

or

$$\int udv = uv - \int vdu. \tag{3.69}$$

I will use bold capital letters for vector fields and non–bold non–capital letters for scalar fields. I assume vector fields from $\mathbb{R}^n$ into $\mathbb{R}^m$ (m could be equal to n) and scalar fields from $\mathbb{R}^n$ into $\mathbb{R}$.

Therefore the following set of equations will be of great use in the following sections:

$$\nabla \cdot (f\,\boldsymbol{G}) = \boldsymbol{G} \cdot \nabla f + f\,\nabla \cdot \boldsymbol{G} \tag{3.70}$$

$$\nabla \cdot (\boldsymbol{F} \times \boldsymbol{G}) = \boldsymbol{G} \cdot (\nabla \times \boldsymbol{F}) - \boldsymbol{F} \cdot (\nabla \times \boldsymbol{G}) \tag{3.71}$$

This list could be made larger but what we are shown here is what is most commonly used. The reader could google for "multidimensional integration by parts" to find many sources on this topic, and proofs of these are found in texts in multi–dimensional calculus or multi–dimensional analysis. We add that

$$\nabla.\nabla f = \nabla^2 f. \tag{3.72}$$

The Laplacian $\nabla^2 f$ is sometimes noted as Δf. If in equation 3.70 we change $\boldsymbol{G}$ by ∇g, we find:

$$\nabla \cdot (f\, \nabla g) \;\; = \;\; f\, \nabla \cdot \nabla g + \nabla f \cdot \nabla g. \tag{3.73}$$

That is

$$\nabla \cdot (f\, \nabla g) \;\; = \;\; f\, \nabla^2 g + \nabla f \cdot \nabla g. \tag{3.74}$$

The identities in this section are commonly used to reduce dimensionality of problems. For example to reduce an integral over a volume to the integral over its surface (the boundary). That is the case of the Green's and Stoke's theorems. With Green identities, also built from the identities in this section, Green's functions for wave operators are built. Here we are somehow going the other way around. We want data that vanishes in the boundaries in order to find the adjoint operator in acting on the whole volumes.

Before we start applying the tools at hand, we should understand inner products in functions of several variables. Assume that f and g

$$f, g : V \subseteq \mathbb{R}^n \to \mathbb{R} \tag{3.75}$$

are scalar differentiable fields, in some domain V. We define the inner product $\langle\, . \rangle$ of f with g as

$$\langle f\, , g \rangle = \int_V f\, g\, d\boldsymbol{x}, \tag{3.76}$$

where $\boldsymbol{x}$ rolls over all points of the set V, given that this multi–dimensional integral exists.

Now let us define the corresponding inner product for vector fields. Assume that $\boldsymbol{F} = \{f_i\}$ and $\boldsymbol{G} = g_i$, $i = 1, 2, .., n$.

$$\boldsymbol{F}, \boldsymbol{G} : V \subseteq \mathbb{R}^n \to \mathbb{R}^m \tag{3.77}$$

are vector differentiable fields, in some domain V. We define the inner product $\langle\, . \rangle$ of f with g as

$$\langle \boldsymbol{F}\, , \boldsymbol{G} \rangle = \int_V \sum_{i=1}^{m} f_i\, g_i\, d\boldsymbol{x}, = \int_V \boldsymbol{F} \cdot \boldsymbol{G}\; d\boldsymbol{x}. \tag{3.78}$$

where $\boldsymbol{x}$ rolls over all points of the set V, given that this multi–dimensional integrals exists.

3.2.2.7.1 The gradient: Assume that f is a scalar differentiable field in $V \subseteq \mathbb{R}^n$. Let us call the Hilbert space $\mathcal{H}_1$ the space of scalar field functions

$$f : V \subseteq \mathbb{R}^n \to \mathbb{R} \tag{3.79}$$

which, where we can take as many we need partial derivatives (C^∞ if needed) and $\mathcal{H}_2$ the space of vector field functions

$$\boldsymbol{G} : V \subseteq \mathbb{R}^n \to \mathbb{R}^n \tag{3.80}$$

where also we can take as many derivatives as needed. Call L the gradient operator

$$\begin{array}{rcl} L : \mathcal{H}_1 & \to & \mathcal{H}_2 \\ & & \\ f & \mapsto & L(f) = \begin{pmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ . \\ . \\ . \\ \partial f/\partial x_n \end{pmatrix} \end{array}$$

we want to find L^*. Let us take a vector field $\boldsymbol{G}$ in $\mathcal{H}_2$. Then, from the inner product definition 3.78 we have:

$$\begin{aligned} \langle L(f)\,, \boldsymbol{G} \rangle &= \int_V \nabla f \cdot \boldsymbol{G}\, d\boldsymbol{x} \\ &= \int_V [\nabla \cdot (f\,\boldsymbol{G}) - f\,\nabla \cdot \boldsymbol{G}\,]\, d\boldsymbol{x} \\ &= \int_\Sigma f\,\boldsymbol{G} \cdot d\Sigma - \int_V f\,\nabla \cdot \boldsymbol{G}\, d\boldsymbol{x}. \end{aligned} \tag{3.81}$$

The second equation comes from identity 3.70 and the third equation comes from the divergence theorem. Here Σ is the boundary of the volume V. Assuming that the product $f\,\boldsymbol{G}$ vanishes in the boundary Σ we find

$$\langle L(f)\,, \boldsymbol{G} \rangle = -\int_V f\,\nabla \cdot \boldsymbol{G}\, d\boldsymbol{x} = \langle f\,, -\nabla \cdot \boldsymbol{G} \rangle.$$

where the second inner product is in the space of scalar functions (see 3.76). That is, provided that the product $f\,\boldsymbol{G}$ vanishes in the boundary we found that

$$\langle L(f)\,, \boldsymbol{G} \rangle = \langle f\,, L^*(\boldsymbol{G}) \rangle \tag{3.82}$$

with

$$\begin{aligned} L^* : \mathcal{H}_2 &\rightarrow \mathcal{H}_1 \\ \boldsymbol{G} &\mapsto L^*(\boldsymbol{G}) = -\nabla \cdot \boldsymbol{G}. \end{aligned} \tag{3.83}$$

We say that the adjoint of the gradient is the negative divergence. By the symmetry relation 3.3 we can say that the adjoint of the divergence is the negative of the gradient.

3.2.2.7.2 Product of a scalar function and a gradient: Making the previous example a bit more general, let us assume that

$$\begin{aligned} L : \mathcal{H}_1 &\rightarrow \mathcal{H}_2 \\ f &\mapsto a(\boldsymbol{x})\nabla f, \end{aligned} \tag{3.84}$$

then, given a vector field $\boldsymbol{G}$ in $\mathcal{H}_2$

$$\begin{aligned} \langle L(f)\,, \boldsymbol{G} \rangle &= \int_V a(\boldsymbol{x})\, \nabla f \cdot \boldsymbol{G}\, d\boldsymbol{x} \\ &= \int_V \nabla f \cdot a(\boldsymbol{x})\, \boldsymbol{G}\, d\boldsymbol{x} \\ &= \int_V [\nabla \cdot (f a(\boldsymbol{x})\, \boldsymbol{G}) - f\, \nabla \cdot a(\boldsymbol{x})\, \boldsymbol{G}\,]\, d\boldsymbol{x} \\ &= \int_\sum (f\, a(\boldsymbol{x})\, \boldsymbol{G}) \cdot d\,\Sigma - \int_V f\, \nabla \cdot (a(\boldsymbol{x})\, \boldsymbol{G})\, d\boldsymbol{x} \end{aligned} \tag{3.85}$$

The second equation is the commutation of the scalar under the dot product, the third equation comes from identity 3.70 and the fourth equation comes from the divergence theorem. Here Σ is the boundary of the volume V. Assuming that the product $f\, a(\boldsymbol{x})\, \boldsymbol{G}$ vanishes in Σ we find

$$\langle L(a(\boldsymbol{x})\, f)\,, \boldsymbol{G} \rangle = -\int_V f\, \nabla \cdot (a(\boldsymbol{x})\, \boldsymbol{G})\, d\boldsymbol{x} = \langle f\,, -\nabla \cdot (a(\boldsymbol{x})\, \boldsymbol{G})\, \rangle.$$

where the second inner product is in the space of scalar functions (see 3.76). That is, provided that the product $f\, \boldsymbol{G}$ vanishes in the boundary we found that

$$\langle L(f)\,, \boldsymbol{G} \rangle = \langle f\,, L^*(a(\boldsymbol{x})\, \boldsymbol{G}) \rangle \tag{3.86}$$

with

$$\begin{aligned} L^* : \mathcal{H}_2 &\rightarrow \mathcal{H}_1 \\ \boldsymbol{G} &\mapsto L^*(\boldsymbol{G}) = -\nabla \cdot [a(\boldsymbol{x})\, \boldsymbol{G}]. \end{aligned} \tag{3.87}$$

In words, we say that the adjoint maps gradient scaled into scaled divergence. That is, a factor outside the gradient will be sent inside the divergence symbol under the adjoint operation.

3.2.2.7.3 The curl: The curl can be defined in n–dimensional spaces but here we will only use the three–dimensional space version, in order to simplify the development and because it is the most commonly used. The definition of curl is a linear operator L from $\mathcal{H}_2$ to $\mathcal{H}_2$ such that:

$$\begin{aligned} L : \mathcal{H}_2 &\rightarrow \mathcal{H}_2 \\ \boldsymbol{G} &\mapsto L(\boldsymbol{G}) = \nabla \times \boldsymbol{G} = \begin{pmatrix} \partial g_2/\partial x_3 - \partial g_3/\partial x_2 \\ \partial g_3/\partial x_1 - \partial g_1/\partial x_3 \\ \partial g_1/\partial x_2 - \partial g_2/\partial x_1 \end{pmatrix} \end{aligned}$$

provided that all partial derivatives exist.

We now find the adjoint of the curl. Using the inner product definition 3.78 we have that for any vector field function $\boldsymbol{F}$ in $\mathcal{H}_2$

$$\begin{aligned} \langle\, L\,\boldsymbol{G}\,, \boldsymbol{F} \rangle &= \int_V (\nabla \times \boldsymbol{G}) \cdot \boldsymbol{F}\, d\boldsymbol{x} \qquad (3.88) \\ &= \int_V \nabla \cdot (\boldsymbol{F} \times \boldsymbol{G}) + G \cdot (\nabla \times \boldsymbol{F}) \\ &= \int_\Sigma (\boldsymbol{F} \times \boldsymbol{G})\, d\,\Sigma + \int_V \boldsymbol{G} \cdot (\nabla \times \boldsymbol{F}) \\ &= \int_\Sigma (\boldsymbol{F} \times \boldsymbol{G})\, d\,\Sigma + \langle\, \boldsymbol{G}\,, L(\boldsymbol{F})\,\rangle. \qquad (3.89) \end{aligned}$$

Here in the second step we used the identity 3.71 and in the third step the divergence theorem. So, provided the boundary condition $\boldsymbol{F} \times \boldsymbol{G} = 0$ on Σ, we have that the curl operator is self–adjoint.

3.2.2.7.4 Product of a scalar function and a curl: In the previous problem let us multiply the curl by scalar field a; that is

$$\begin{aligned} L : \mathcal{H}_2 &\rightarrow \mathcal{H}_2 \\ \boldsymbol{G} &\mapsto L(\boldsymbol{G}) = a(\boldsymbol{x})\, \nabla \times \boldsymbol{G}. \end{aligned} \tag{3.90}$$

Give any vector field $\boldsymbol{F}$ in $\mathcal{H}_2$ we find:

$$\begin{aligned}
\langle\, L\,\boldsymbol{G}\,,\boldsymbol{F}\rangle &= \int_V (a(\boldsymbol{x})\,\nabla\times\boldsymbol{G})\cdot\boldsymbol{F}\,d\boldsymbol{x} && (3.91)\\
&= \int_V (\nabla\times\boldsymbol{G})\cdot(a(\boldsymbol{x})\,\boldsymbol{F})\,d\boldsymbol{x}\\
&= \int_V \nabla\cdot(a(\boldsymbol{x})\,\boldsymbol{F}\times\boldsymbol{G}) + \boldsymbol{G}\cdot(\nabla\times(a(\boldsymbol{x})\,\boldsymbol{F}))\\
&= \int_{\sum} (a(\boldsymbol{x})\,\boldsymbol{F}\times\boldsymbol{G})\,d\,\Sigma + \int_V \boldsymbol{G}\cdot(\nabla\times(a(\boldsymbol{x})\,\boldsymbol{F}))\\
&= \int_\Sigma ((a(\boldsymbol{x})\,\boldsymbol{F})\times\boldsymbol{G})\,d\,\Sigma + \langle\,\boldsymbol{G}\,,L^*(\boldsymbol{F})\,\rangle. && (3.92)
\end{aligned}$$

with

$$\begin{aligned}
L^*:\mathcal{H}_2 &\rightarrow \mathcal{H}_2\\
\boldsymbol{F} &\mapsto L(\boldsymbol{F}) = \nabla\times(a(\boldsymbol{x})\boldsymbol{F}). && (3.93)
\end{aligned}$$

So, provided the boundary condition $(a(\boldsymbol{x})\,\boldsymbol{F})\times\boldsymbol{G} = 0$ in Σ we have that the adjoint of the product of a scalar field and a curl is the curl of the product of the scalar field and its vector field argument.

3.2.2.7.5 The Laplacian: is a linear operator L from $\mathcal{H}_1$ to $\mathcal{H}_1$ such that:

$$\begin{aligned}
L:\mathcal{H}_1 &\rightarrow \mathcal{H}_1\\
f &\mapsto L(f) = \nabla^2 f = \sum_{i=1}^{n}\frac{\partial^2 f}{\partial x^2}. && (3.94)
\end{aligned}$$

provided that all partial derivatives exist.

We want to find the adjoint of the Laplacian. We use the results found above. That is, since, from 3.72 $L(f) = \nabla^2 f = \nabla\cdot\nabla f$, we can write Laplacian as the cascade of two operators:

$$S(f) = \nabla f \qquad \text{and} \qquad T(\boldsymbol{G}) = \nabla\cdot\boldsymbol{G}, \tag{3.95}$$

with $\boldsymbol{G} = \nabla f$, so

$$\begin{aligned}
\langle L(f)\,, g\rangle &= \langle TSf\,, g\rangle \\
&= \langle Sf\,, T^* g\rangle \qquad \text{definition of adjoint} \\
&= \langle Sf\,, -\nabla g\rangle \qquad \text{see comment after 3.83} \\
&= \langle f\,, -S^*\,\nabla g\rangle \quad \text{definition of adjoint} \\
&= \langle f\,, \nabla\cdot\,\nabla g\rangle \quad \text{property 3.83} \\
&= \langle f\,, \nabla^2 g\rangle \\
&= \langle f\,, Lg\rangle. \qquad (3.96)
\end{aligned}$$

This is an interesting way to prove that the Laplacian is a self–adjoint operator, however it does not explicitly shows the boundary conditions for this to happen. Therefore I will try this again using the (already familiar) methodology (integration by parts) used on this document.

$$\begin{aligned}
\langle L(g)\,, f\rangle &= \int_V f\,\nabla^2 g\; d\boldsymbol{x} \\
&= \int_V \nabla\cdot(f\nabla g) - \nabla f\cdot\nabla g\; d\boldsymbol{x} \qquad \text{see equation 3.73} \\
&= \int_\Sigma f\;\nabla g\cdot d\Sigma - \int_V \nabla f\cdot\nabla g\; d\boldsymbol{x} \quad \text{divergence theorem} \\
&= \int_\Sigma f\;\nabla g\cdot d\Sigma - \int_V \nabla\cdot(g\nabla f) - g\nabla^2 f\; d\boldsymbol{x} \quad \text{3.73 again but} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{change } f \text{ by } g \\
&= \int_\Sigma (f\;\nabla g - g\;\nabla f)\cdot d\Sigma + \int_V g\nabla^2 f\; d\boldsymbol{x} \quad \text{divergence theorem} \\
&= \int_\Sigma (f\;\nabla g - g\;\nabla f)\cdot d\Sigma + \langle g\,, L(f)\rangle. \qquad (3.97)
\end{aligned}$$

So we showed again that the adjoint of the Laplacian operator is itself (the Laplacian operator is self–adjoint) and exposed the boundary condition. We require

$$f\nabla g - g\;\nabla f = 0 \qquad (3.98)$$

in the boundary Σ. This proof could have been written in a couple of lines by using the second Green's identity. Actually we proved the second green

identity

$$\int_V (f\,\nabla^2 g - g\,\nabla^2 f)\,d\,\boldsymbol{x} = \int_\Sigma (f\nabla g - g\,\nabla f)\cdot d\Sigma. \tag{3.99}$$

3.2.2.7.6 Product of a scalar function and a Laplacian: Let assume that the operator L is such that

$$\begin{aligned} L : \mathcal{H}_1 &\rightarrow \mathcal{H}_1 \\ f &\mapsto L(f) = a(\boldsymbol{x})\nabla^2 f. \end{aligned} \tag{3.100}$$

As usual we write, for any given scalar field g in $\mathcal{H}_1$

$$\begin{aligned} \langle L(f)\,, g\rangle &= \int_V a\,\nabla^2 f\; g\; d\,\boldsymbol{x} \\ &= \int_V a\; g\;(\nabla^2 f)\, d\,\boldsymbol{x} \\ &= \int_V f\,[\nabla^2\,(a\,g)] + \int_\Sigma [f\,\nabla(a\,g) - (a\,g)\,\nabla f]\cdot d\Sigma. \\ &= \langle f, L(g)\rangle. + \int_\Sigma [f\,\nabla(a\,g) - (a\,g)\,\nabla f]\cdot d\Sigma. \end{aligned} \tag{3.101}$$

In the third line we used the second Green's identity 3.99. So provided the boundary condition:

$$f\,\nabla(a\,g) - (a\,g)\,\nabla f = 0 \tag{3.102}$$

in Σ, we find that the adjoint of L is:

$$\begin{aligned} L^* : \mathcal{H}_1 &\rightarrow \mathcal{H}_1 \\ g &\mapsto L^*(g) = \nabla^2(a\;f). \end{aligned} \tag{3.103}$$

3.2.2.7.7 The dot product of a vector function and a gradient. The general first order linear differential operator: We can write a general first order linear differential operator acting on a scalar field $f \in \mathcal{H}_1$ as

$$Lf = \sum_i b_i(\boldsymbol{x})\frac{\partial f}{\partial x_i} + (c\,f)(\boldsymbol{x}) = \boldsymbol{B}\cdot\nabla f + (c\,f)(\boldsymbol{x}), \tag{3.104}$$

for scalar fields $b_i(\boldsymbol{x})$ and $c(\boldsymbol{x})$ all in $\mathcal{H}_1$. Here $\boldsymbol{B}$ is a vector with components $b_i(\boldsymbol{x})$.

The last term (a factor of a scalar field by f) is obviously self–adjoint and non–interesting. We are after the adjoint of the first term, which is the dot product of the vectors $\boldsymbol{B}$ and ∇f and we all L_B.

As usual, let us pick an arbitrary scalar function $g \in \mathcal{H}_1$. Then,

$$\begin{aligned}
\langle L_B(f)\,, g\rangle &= \int_V (\boldsymbol{B}\cdot\nabla f)\; g\, d\boldsymbol{x} \\
&= \int_V (g\,\boldsymbol{B})\cdot\nabla f\, d\boldsymbol{x} \\
&= \int_V \nabla\cdot(g\, f\cdot\boldsymbol{A}) - f\nabla\cdot g\,\boldsymbol{B}\, d\boldsymbol{x} \\
&= \int_\Sigma f\, g\,\boldsymbol{B}\;\cdot d\Sigma - \int_V f\nabla\cdot g\boldsymbol{B}\, d\boldsymbol{x} \\
&= \int_\Sigma f\, g\,\boldsymbol{B}\;\cdot d\,\Sigma - \langle f\,, L^* g\rangle. \qquad (3.105)
\end{aligned}$$

Second line is commutation, third line is due to identity 3.70 , and fourth line is due to the divergence theory. Given the boundary condition

$$f\, g\,\boldsymbol{B} = 0, \qquad (3.106)$$

in Σ, the adjoint operator L^* is given by:

$$\begin{aligned}
L_B^* : \mathcal{H}_1 &\to \mathcal{H}_1 \\
g &\mapsto L^* g = -\nabla\cdot g\;\boldsymbol{B}. \qquad (3.107)
\end{aligned}$$

So the adjoint of the second order operator L is given by

$$\begin{aligned}
L^*(g) &= L_B^*(g) + b\, g(\boldsymbol{x}) \\
&= -\nabla\cdot g\boldsymbol{B} + c\, v(\boldsymbol{x}) \\
&= -\sum_{i=1}^{n} \frac{\partial\; bg(\boldsymbol{x})}{\partial x_i} + (c\, g)(\boldsymbol{x}). \qquad (3.108)
\end{aligned}$$

with the boundary condition:

$$(f\, g\, b_i)(\boldsymbol{x}) = 0, \qquad (3.109)$$

for any $\boldsymbol{x}$ in Σ.

3.2.2.7.8 A general linear second order differential operator: A general linear second order differentiator operator in $\mathcal{H}_1$ can be written as

$$\begin{aligned} L : \mathcal{H}_1 &\rightarrow \mathcal{H}_1 \\ f &\mapsto L(f) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i \frac{\partial f}{\partial x_i} + c\, f. \end{aligned} \tag{3.110}$$

The purpose of this section is to find the adjoint L^*. The first order part was solved in the previous section. We focus now in the first double sum of second order derivatives. That is, let us define

$$\begin{aligned} L_A : \mathcal{H}_1 &\rightarrow \mathcal{H}_1 \\ f &\mapsto L_A(f) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} + \end{aligned} \tag{3.111}$$

We are tempted to start applying the definition of inner product and integration by parts to get L_A^*. However I do not know about any integration by parts that will take care of the whole matrix (of a_{ij}) to matrix (of second order partial derivatives) inner product. This looks like a second order weighted divergence operator but I do not know the theory for this higher dimensional "divergence". Instead, based on previous experience for ordinary differential equations and the operators shown in the sections above I will make an intelligent guess on what L^* should look like and then verify that it is the adjoint we are looking for.

I claim that

$$\begin{aligned} L_A^* : \mathcal{H}_1 &\rightarrow \mathcal{H}_1 \\ g &\mapsto L_A^*(g) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 (a_{ij}\, g)}{\partial x_i \partial x_j} + \end{aligned} \tag{3.112}$$

Then I compute

$$\langle L_A f\, , g \rangle - \langle L_A^* g\, , f \rangle \tag{3.113}$$

and show that it is an exact differential, which after applying the divergence theorem produces the condition in the boundary Σ. We see that:

$$\langle L_A f\, , g \rangle - \langle L_A^* g\, , f \rangle = \int_V \sum_{i=1}^{n} \sum_{j=1}^{n} \left(a_{ij}\, g \frac{\partial^2 f}{\partial x_i \partial x_j} - f\, \frac{\partial^2 (a_{ij}\, g)}{\partial x_i \partial x_j} \right) d\boldsymbol{x} \tag{3.114}$$

The expression in the parenthesis can be written as

$$\begin{aligned} a_{ij}\, g\frac{\partial^2 f}{\partial x_i \partial x_j} - f\,\frac{\partial^2 (a_{ij}\, g)}{\partial x_i \partial x_j} &= a_{ij}\, g\frac{\partial^2 f}{\partial x_i \partial x_j} + \frac{\partial (a_{ij}\, g)}{\partial x_i \partial x_j}\frac{\partial f}{\partial x_j} \\ &\quad -\frac{\partial (a_{ij}\, g)}{\partial x_i \partial x_j}\frac{\partial f}{\partial x_j} - f\,\frac{\partial^2 (a_{ij}\, g)}{\partial x_i \partial x_j} \\ &= \frac{\partial}{\partial x_i}\left(a_{ij} g \frac{\partial f}{\partial x_j}\right) - \frac{\partial}{\partial x_j}\left(f\frac{\partial (a_{ij} g)}{\partial x_i}\right). \end{aligned}$$

Now since

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial}{\partial x_j}\left(f\frac{\partial (a_{ij} g)}{\partial x_i}\right). = \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial}{\partial x_i}\left(f\frac{\partial (a_{ji} g)}{\partial x_j}\right).$$

we find that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\, g\frac{\partial^2 f}{\partial x_i \partial x_j} - f\,\frac{\partial^2 (a_{ij}\, g)}{\partial x_i \partial x_j} = \sum_{i=1}^{n}\frac{\partial}{\partial x_i}\left[\sum_{j=1}^{n} a_{ij} g\frac{\partial f}{\partial x_j} - f\frac{\partial (a_{ji} g)}{\partial x_j}\right].$$

and so

$$\begin{aligned} \langle L_A f\,, g\rangle - \langle L_A^* g\,, f\rangle &= \int_V \sum_{i=1}^{n}\frac{\partial}{\partial x_i}\left[\sum_{j=1}^{n} a_{ij} g\frac{\partial f}{\partial x_j} - f\frac{\partial (a_{ji} g)}{\partial x_j}\right]. \\ &= \int_V \nabla \cdot p_i d\boldsymbol{x} \end{aligned}$$

Is an exact differential with

$$p_i = \sum_{j=1}^{n} a_{ij} g\frac{\partial f}{\partial x_j} - f\frac{\partial (a_{ji} g)}{\partial x_j},$$

and we can apply the divergence theorem to finally get

$$\langle L_A f\,, g\rangle - \langle L_A^* g\,, f\rangle = \int_\Sigma p_i \cdot d\Sigma. \tag{3.115}$$

So provided that $p_i = 0$ in the boundary Σ, the adjoint of L_A exists and it is given in equation 3.112. Combining this result with the adjoint of the general linear first order differential operator in equation 3.108 we find that the

adjoint of the general linear second order partial differential operator 3.110 is

$$\begin{aligned} L^* : \mathcal{H}_1 &\rightarrow \mathcal{H}_1 \\ g &\mapsto L(g) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 (a_{ij} g)}{\partial x_i \partial x_j} - \sum_{i=1}^{n} \frac{\partial \left(b_i \, g \right)}{\partial x_i} + c \, g. \end{aligned} \quad (3.116)$$

with the boundary condition

$$p_i = \sum_{j=1}^{n} a_{ij} g \frac{\partial f}{\partial x_j} - f \frac{\partial (a_{ji} g)}{\partial x_j} + b_i f \, g = 0$$

for all $\boldsymbol{x}$ in Σ.

All wave equations are contained here, including any type of anisotropy. In particular the acoustic wave equation for variable density can be written in terms of the linear operator in the frequency domain as:

$$\begin{aligned} L : \mathcal{H}_1 &\rightarrow \mathcal{H}_1 \\ f &\mapsto L(f) = \rho(\boldsymbol{x}) \, \nabla \cdot \left[\frac{1}{\rho(\boldsymbol{x})} \nabla f \right] + \frac{\omega^2}{c^2(\boldsymbol{x})} f \end{aligned} \quad (3.117)$$

Let us expand the operator.

$$\begin{aligned} L(f) &= \nabla^2 f + \rho(\boldsymbol{x}) \, \nabla \left[\frac{1}{\rho(\boldsymbol{x})} \right] \cdot \nabla f + \frac{\omega^2}{c^2(\boldsymbol{x})} f \\ &= \sum_{i=1}^{n} \left[\frac{\partial^2 f}{\partial x^2} - \frac{1}{\rho(\boldsymbol{x})} \frac{\partial \rho(\boldsymbol{x})}{\partial x_i} \frac{\partial f}{\partial x_i} \right] + \frac{\omega^2}{c^2(\boldsymbol{x})} f, \end{aligned} \quad (3.118)$$

and applying general formula 3.116 we find

$$\begin{aligned} L^* g &= \sum_{i=1}^{n} \left[\frac{\partial^2 g}{\partial x^2} + \frac{\partial}{\partial x_i} \left(\frac{1}{\rho(\boldsymbol{x})} \frac{\partial \rho(\boldsymbol{x})}{\partial x_i} \; g \right) \right] + \frac{\omega^2}{c^2(\boldsymbol{x})} \\ &= \nabla^2 g + \nabla \cdot \left[\frac{1}{\rho(\boldsymbol{x})} \nabla \rho(\boldsymbol{x}) \, g \right] + \frac{\omega^2}{c^2(\boldsymbol{x})} \\ &= \nabla \cdot \left[\nabla g + \frac{1}{\rho(\boldsymbol{x})} \nabla \rho(\boldsymbol{x}) \, g \right] + \frac{\omega^2}{c^2(\boldsymbol{x})} \\ &= \nabla \cdot \left[\frac{1}{\rho(\boldsymbol{x})} \nabla \left(\rho(\boldsymbol{x}) \, g \right) \right] + \frac{\omega^2}{c^2(\boldsymbol{x})} \end{aligned}$$

3.2.2.8 Integral operators: As indicated in equation 3.34, the operator

$$Lu = \int_a^b k(s,t)u(t)dt. \tag{3.119}$$

represents a Fredholm integral operator, and it is a key concept for the theory of transforms. Let us generalize it.

Define a Hilbert space $\mathcal{H}$ of the functions $u(\boldsymbol{x})$

$$u : \mathbb{C}^n \to \mathbb{C}. \tag{3.120}$$

[6]

We write a general Fredholm operator L as:

$$\begin{aligned} L : \mathcal{H} &\to \mathcal{H} \\ u(\boldsymbol{x}) &\mapsto L\,u(\boldsymbol{x}) = \int_{V_y} k(\boldsymbol{x},\boldsymbol{y})u(\boldsymbol{y})d\boldsymbol{y}. \end{aligned} \tag{3.121}$$

Here instead of the limits a and b we consider the domain as the volume of integration V_y and $k(\boldsymbol{x},\boldsymbol{y})$ is a complex number.

Given $v(\boldsymbol{x}) \in \mathcal{H}$

$$\begin{aligned} \langle L\,u\,, v\rangle &= \int_{V_x}\left[\int_{V_y} k(\boldsymbol{x},\boldsymbol{y})\,u(\boldsymbol{y})d\boldsymbol{y}\right]\overline{v}(\boldsymbol{x})\;d\boldsymbol{x} \\ &= \int_{V_y}\left[\int_{V_x} k(\boldsymbol{x},\boldsymbol{y})\,\overline{v}(\boldsymbol{x})d\boldsymbol{x}\right]u(\boldsymbol{y})\;d\boldsymbol{x} \\ &= \int_{V_y} u(\boldsymbol{y})\,\overline{L^*}\;vd\boldsymbol{x} \end{aligned}$$

with

$$L^*v(\boldsymbol{y}) = \int_{V_x} \overline{k}(\boldsymbol{x},\boldsymbol{y})\;v(\boldsymbol{x})\;d\boldsymbol{x}. \tag{3.122}$$

We assume that we can change the order of integration. This is not always possible. Note that in the definition of L the dummy variable is the second

[6] I assume that any operation on these functions is well defined, that is the function has all the continuity and differentiability attributes needed.

($\boldsymbol{x}$) variable, while here the dummy variable is ($\boldsymbol{y}$). Rewriting this equation in terms of $\boldsymbol{x}$ instead of $\boldsymbol{y}$ we find:

$$L^* v(\boldsymbol{x}) = \int_{V_y} \overline{k}(\boldsymbol{y}, \boldsymbol{x})\, v(\boldsymbol{y})\, d\boldsymbol{y}. \tag{3.123}$$

Therefore the adjoint of the Fredholm integral operator represented by the kernel $k(\boldsymbol{x}, \boldsymbol{y})$ is given by another Fredholm integral operator with kernel $\overline{k}(\boldsymbol{y}, \boldsymbol{x})$. This resembles the case of the finite dimensional spaces of matrices where the adjoint is given by the Hermitian, that is the conjugate of the transposed (here instead of transposed is the switch of the $\boldsymbol{x}$, $\boldsymbol{y}$ variables).

The generality of this integral representation opens the door to a wealth of examples. I only show the application of this to the integral operator as a function of the Volterra operator, the Fourier transform, and to the integrals of seismic migration and demigration, because they are very relevant in our geophysical applications.

3.2.2.8.1 The Volterra integral operator Define the function

$$\begin{aligned} L : \mathcal{H} &\rightarrow \mathcal{H} \\ u(\boldsymbol{x}) &= L\, u(\boldsymbol{x}) \int_{-\boldsymbol{a}}^{\boldsymbol{x}} k(\boldsymbol{x}, \boldsymbol{y})\, u(\boldsymbol{y}) d\boldsymbol{y}. \end{aligned} \tag{3.124}$$

First, note the $\boldsymbol{x}$ as an upper index. This distinguishes this integral operator from the Fredholm integral operator. Let us further assume that $-\boldsymbol{a}$ is the boundary at $-\infty$ for all dimensions. We can then rewrite the Volterra integral as a Fredholm type integral operator with the help of the multi–dimensional Heaviside function. That is, we can write 3.124 as follows:

$$\begin{aligned} L : \mathcal{H} &\rightarrow \mathcal{H} \\ u(\boldsymbol{x}) &= L\, u(\boldsymbol{x}) \int_{-\boldsymbol{a}}^{\boldsymbol{a}} S(\boldsymbol{x} - \boldsymbol{y}) k(\boldsymbol{x}, \boldsymbol{y})\, u(\boldsymbol{y}) d\boldsymbol{y}. \end{aligned} \tag{3.125}$$

where

$$S(\boldsymbol{x} - \boldsymbol{y}) = \prod_i^n s(x_i - y_i), \tag{3.126}$$

and

$$s(x_i - y_i) = \begin{cases} 1 & \text{if} \quad x_i \geq y_i \\ 0 & \text{if} \quad x_i < y_i \end{cases} \tag{3.127}$$

So we can absorb the Heaviside function into the kernel k, and rewriting it as

$$K(\boldsymbol{x},\boldsymbol{y}) = S(\boldsymbol{x}-\boldsymbol{y})k(\boldsymbol{x},\boldsymbol{y}) \tag{3.128}$$

and write the adjoint of the Volterra integral operator, following equation 3.123. That is,

$$\begin{aligned} L^* v(\boldsymbol{x}) &= \int_{-\boldsymbol{a}}^{\boldsymbol{a}} \overline{K}(\boldsymbol{y},\boldsymbol{x})\, v(\boldsymbol{y})\, d\boldsymbol{y} \\ &= \int_{-\boldsymbol{a}}^{\boldsymbol{a}} S(\boldsymbol{y}-\boldsymbol{x})\, \overline{k}(\boldsymbol{y},\boldsymbol{x})\, v(\boldsymbol{y})\, d\boldsymbol{y} \\ &= \int_{-\boldsymbol{x}}^{\boldsymbol{a}} \overline{k}(\boldsymbol{y},\boldsymbol{x})\, v(\boldsymbol{y}) d\boldsymbol{y}. \end{aligned} \tag{3.129}$$

Interestingly, the adjoint of a "causal" Volterra integral equation is an "anti–causal" Volterra integral equation. The variable $\boldsymbol{x}$ was moved from the upper to the lower index. We will use this result in the last section of the document.

3.2.2.8.2 The Continuous Fourier Transform: The (forward) Fourier transform is usually defined as

$$\begin{aligned} L : \mathcal{H} &\rightarrow \mathcal{H} \\ u(\boldsymbol{k}) &\mapsto L\, u(\boldsymbol{k}) = \frac{1}{(2\pi)^{n/2}} \int_{V_x} \mathrm{e}^{\mathrm{i}\, \boldsymbol{k}\cdot\boldsymbol{x}} u(\boldsymbol{x}) d\boldsymbol{x}. \end{aligned} \tag{3.130}$$

The set V_x is identified with $\mathbb{R}^n$. The vector $\boldsymbol{k}$ is the wavenumber vector. We identify the kernel here as

$$k(\boldsymbol{x},\boldsymbol{k}) = \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}, \tag{3.131}$$

and so the adjoint of the Fourier transform is associated with the kernel

$$\overline{k}(\boldsymbol{k},\boldsymbol{x}) = \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}, \tag{3.132}$$

and the adjoint of the Fourier transform is given by:

$$\begin{aligned} L^* : \mathcal{H} &\rightarrow \mathcal{H} \\ v(\boldsymbol{x}) &\mapsto L^*\, v(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}} \int_{V_k} \mathrm{e}^{-\mathrm{i}\, \boldsymbol{x}\cdot\boldsymbol{k}} v(\boldsymbol{k}) d\boldsymbol{k}. \end{aligned} \tag{3.133}$$

This happens to be the inverse Fourier transform. Therefore we say that the Fourier transform is a unitary operator.

In mathematical physics we are interested in both spatial and temporal coordinates. The kernel for the Fourier transform is usually defined as:

$$k(\boldsymbol{x}, \boldsymbol{k}, t, \omega) = \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x} - \omega\, t}, \tag{3.134}$$

with $\omega = 2\,\pi f$ being the circular frequency. The minus "-" sign in front of the t comes with causality. Anti–causal Fourier transform are defined with a "+" sign instead but they are uncommon. We can embed this definition into the space of functions in $\mathbb{C}^{n+1}$. In this space the dot product could be defined as

$$\boldsymbol{x} \cdot \boldsymbol{y} = \sum_{i=1}^{n-1} x_i\, y_i - x_n y_n. \tag{3.135}$$

Note that I did not use the conjugate for the second ($\boldsymbol{y}$) variable. In practice those variables are real.

3.2.2.8.3 The adjoint of Kirchhoff migration is Kirchhoff demigration Kirchoff migration, in the frequency domain, can be written as the following integral operator.

$$L[D(\boldsymbol{\xi}, \omega)] = I(\boldsymbol{x}) = \int_{\Sigma}\int_{\omega} A(\boldsymbol{x}, \boldsymbol{\xi}, \omega) D(\boldsymbol{\xi}, \omega)\, \mathrm{e}^{\mathrm{i}\phi(\boldsymbol{x}, \boldsymbol{\xi}, \omega)}\, d\boldsymbol{\xi}\, d\omega. \tag{3.136}$$

The reader is referred to [1] for the description of the meaning of each symbol here. We identify the kernel of the operator as

$$k(\boldsymbol{x}, \boldsymbol{y}) = A(\boldsymbol{x}, \boldsymbol{y})\, \mathrm{e}^{\mathrm{i}\phi(\boldsymbol{x}, \boldsymbol{y})} \tag{3.137}$$

and the input function

$$u(\boldsymbol{x}) = D(\boldsymbol{x}) \tag{3.138}$$

where $\boldsymbol{y} = (\boldsymbol{\xi}, \omega)$. Then, from equation 3.122, the adjoint of Kirchhoff migration is given by

$$\begin{aligned} L^*[v(\boldsymbol{y})] &= \int_{\Gamma} \overline{A}(\boldsymbol{y}, \boldsymbol{x})\, v(\boldsymbol{x})\ \mathrm{e}^{-\mathrm{i}\,\phi(\boldsymbol{y}, \boldsymbol{x})} d\boldsymbol{x} \\ &= \int_{\Gamma} \overline{A}(\boldsymbol{\xi}, \omega, \boldsymbol{x}) v(\boldsymbol{x})\, \mathrm{e}^{-\mathrm{i}\phi(\boldsymbol{\xi}, \omega, \boldsymbol{x})} d\boldsymbol{x}. \end{aligned} \tag{3.139}$$

which is Kirchhoff demigration. Here $I(\boldsymbol{y}) = v(\boldsymbol{y})$ is the image produced by the migration operator 3.136. While this operator is not "true" amplitude, it works well at the kinematic level. Asymptotically, Kirchhoff migration is a stack (spread) over/along diffraction (isochrone) surfaces (again see Bleistein for further explanations) while demigration is a spread (stack) along/over diffraction (isochrone) surfaces.

3.2.3 Why do I think that the continuum is "easier" than the discrete

This section might not belong to this manuscript, but I always have had the need to justify this statement. I showed a set of examples of adjoint operators in the discrete and in the continuum. We need to acknowledge the work of Newton, Leibniz, Gauss, Euler, Lagrange, Cauchy, and all the great mathematicians that invented and developed the integral and differential calculus, analysis, differential equations and all that. The theory of the continuum (used for example in continuum mechanics) is a super–highway used to speed up the mathematical physics development that has sent us this far in so many fields (including geophysics, of course). I just want to show two simple examples that justify my thinking.

- **The sum of integer powers:** Let us assume that we want to add the first m powers of integer numbers. That is,

$$\sum_{i=1}^{m} i^n. \tag{3.140}$$

If m is big, this is not a simple matter, without a computer. There are actually closed formulas for this expression and the Bernoulli brothers worked hard on this problem. I will not even try to write the solution for this problem here, but the reader can find good sources in Wikipedia. The only thing that I can say for sure is that the leading order term is $m^{n+1}/n+1$. That is

$$\sum_{i=1}^{m} i^n = \frac{m^{n+1}}{n+1} + \mathcal{O}(m^n). \tag{3.141}$$

On the other hand, it is very easy to find that

$$\int_0^m x^n dx = \frac{m^{n+1}}{n+1}, \tag{3.142}$$

and this is exact. The difference between equations 3.140 and 3.142 is the total area chopped when cutting rectangles to fit the smooth polynomial curve x^n. This difference is complicated and shows that the continuum is a blessing compared to the discrete.

- **The finite difference approximation to the wave equation:** Think about the three dimensional acoustic wave equation and its computer implementation using finite differences. The equation operator (density $\rho = 1$) can be written as

$$Lu(\boldsymbol{x}, t) = \left[\nabla^2 - \frac{\partial^2}{\partial t^2} \right] u(\boldsymbol{x}, t). \tag{3.143}$$

This looks simple enough. However, let us assume that we are going to use 9 point differentiators in space and 4 points in time, in a finite difference approximation. If we are going to write this finite difference operators as a matrix vector multiplication, we need a huge Toeplitz banded matrix. For each spatial derivative there would be 9 diagonals on this matrix, and they are not contiguous. The time derivative will add terms to those 25 (since 9 x 3-2=25, the center was counted 3 times) diagonal lines of matrix components. We have not even added here the source term and the boundary conditions that complicate much more the problem. The writing of all these matters could take at least one page and the computer implementation is a big task. For example, if instead of thinking about matrices we think about fourth rank tensors (objects with four indices) just the Laplacian would look like:

$$\begin{aligned} \nabla^2 u \approx a_{xyz0}\ u^i_{j,k,l} + \sum_{n=1}^{n=4} \ [\ & a_{xn}(u^i_{j+n,k,l} + u^i_{j-n,k,l}) \\ & + a_{yn}(u^i_{j,k+n,l} + u^i_{j,k-n,l}) \\ & + a_{zn}(u^i_{j,k,l+n} + u^i_{j,k,l-n})\ ] \end{aligned} \tag{3.144}$$

Where the coefficients a_{ij}, correspond to the stencil of the finite difference approximation for each partial differential operator. The indices i, j, k, l are for t, x, y, z coordinates respectively.

True: Real life is discrete, but the mathematical physics has come such a long way with the continuum tools: Electromagnetic theory

(Maxwell equations), thermodynamics (heat equation) , wave mechanics (wave equations), fluid dynamics (Navier–Stokes equations) quantum mechanics (Schrodringer equation), etc.

Back to reality: Our job as software developers is to take those continuum equations and make them discrete with the understanding that aliasing, dispersion, diffusion, stability, boundary conditions, wraparound, and other numerical artifacts are waiting anxiously to surprise us, together with all details of input/output and memory management.

I like the matrix formulation because it is compact. I like the expression A, but not the expression a_{ij}. Keeping track of indices could be hard and I ignore them whenever I can. In the same way I like to refer, for example, to the operator L. A lot of what I have written here applies for abstract operators L. The devil is in the details, but having a look to the forest is very important before landing on a particular tree far from our final target. I have been able to write a lot of details in certain derivations here because I have a clear landing spot, and have found the details with a good compass navigator (the understanding of the basic principles of adjoint operators) and a good editor (VI plus LaTeX [7]).

4 The use of the adjoint in the solution of linear equations

The solution of the general linear equation $Lu = f$ is at the center of the applications of mathematical physics for all developing technologies. Modeling and inversion, if linear, fit this equation (integral equations, differential equations, algebraic equations etc.). If the modeling problem is non–linear (and unfortunately this is almost always the case in real life) it is approximated by a linear operator (for example the Born approximation, one iteration of the gradient method, straight ray tomography, etc) or a sequence of linear

[7] Although it could be unbelievable, I do not understand my own hand–writing. When equations are large, since my hand–writing uses big characters, equations do not fit in one line, and sometimes in one page. I hate pencils and erasers. I end up using pages and pages of paper with the same things written on them that go directly to the trash basket. LaTeX makes things much simpler for me. It is like "programming" the derivation of the equation.

problems (for example, the Newton method, steepest descents, conjugate gradients, etc).

It is therefore very important to build a road map on the solution of that equation. The logic steps for this are:

- Is there a solution? (existence, the Fredholm alternative theorem)
- Is it unique? (uniqueness)
- If it exists how do we find it? (Green's functions)
- If it does not exist, can we approximate it? (the pseudo–inverse, the least square method).

In this section I address these points from the abstract point of view. The problem of existence can be complicated. For example, there is a wealth of literature about the Dirichlet problem that happens when L is a partial differential operator (the classical one is the Laplacian). I will not even try any level of detail in those directions. The idea here is to factor out what is common in all methods and maximize the cost–benefit relationship between time spent versus learning on modeling and inversion.

I show how the adjoint plays a central roll in the solution of the linear equation problem.

4.1 Basic Definitions

4.1.1 Kernel, Cokernel, Range

Let us assume a linear operator L from a Hilbert space $\mathcal{H}_1$ into a Hilbert space $\mathcal{H}_2$.

The kernel of L, also known as the null space is the set of all vectors (functions) $u \in \mathcal{H}_1$ such that

$$Lu = 0. \tag{4.1}$$

We note the null space of L as $\mathcal{N}(L)$. The null space is important to proof uniqueness.

The cokernel is the kernel of the adjoint. That is, the set of all vectors $v \in \mathcal{H}_2$ such that

$$L^* v = 0. \tag{4.2}$$

The range of L is the set of functions $f \in \mathcal{H}_2$ such that

$$Lu = f, \tag{4.3}$$

for all $u \in \mathcal{H}_1$. The notation for range of L is $\mathcal{R}(L)$. Saying that f is in the range of L is saying that the equation 4.3 has a solution. However proving that $f \in \mathcal{R}(A)$ is as hard as finding the solution of the equation. In the following section I will show how the cokernel of L can be used as an indicator for the existence of solutions of equation 4.3.

4.2 The Fredholm Alternative

The question of uniqueness should come, logically, after the question of existence. However, existence is in general much more difficult to prove than uniqueness and it is common to see that uniqueness is shown first. That is the case here.

4.2.0.1 The uniqueness criterium: If the null space of L is the single element 0, then the solutions to the equation

$$Lu = f, \tag{4.4}$$

if they exist, are unique. The proof of this theorem is simple. Assume there are at least two vectors u_1 and u_2 for which equation 4.4 is satisfied. Since L is linear the difference $u_1 - u_2$ satisfies $L(u_1 - u_2) = 0$, but since the null space only contains the 0 vector, this means that $u_1 = u_2$. So there can not be more than one vector satisfying $Lu = f$.

4.2.0.2 The existence criterium: a necessary condition. This criterium is known as the Fredholm alternative. If there is a solution u for the equation 4.4 then f has to be orthogonal to the cokernel of L. That is, for each vector v in satisfying $L^*v = 0$, we should have $\langle f\,, v\rangle = 0$.

The proof of this is also short. Let us see. Assume $v \in \mathcal{H}_2$ such that $L^*v = 0$, so

$$\langle f\,, v\rangle = \langle Lu, v\rangle = \langle u, L^*v\rangle = \langle u, 0\rangle = 0. \tag{4.5}$$

The interesting and more complicated problem is the implication in the other direction. [8] That is, we ask, if given that the "source" term f is orthogonal to the cokernel of L, then this implies that the solution of equation 4.4 exists? I show a proof here for finite dimensional spaces (matrices) due to Keener [4]. The problem in general is more complicated and it belongs to the field of functional analysis. Differential operators are in general hard to deal with (the Dirichlet problem, for example), I try to stay away from these difficulties (due to lack of time).

A different way to rephrase what we found here is that the cokernel of L is orthogonal to the range of L, that is $\mathcal{N}(L^*) \perp \mathcal{R}(L)$. In a finite dimensional space the cokernel of L is the orthogonal complement of the range of L. That is, we can write any vector b as $(b = b_0 + b_r)$ the sum of two components. One component along the $\mathcal{N}(L^*)$ (component h_0) and the other (component h_r) along $\mathcal{R}(L)$. The cokernel of L and the range of L expand the whole space. Unfortunately this is not always true and therefore the following theorem is not valid in all cases.

4.2.0.3 The existence criterium: a sufficient condition for matrices. Assume a matrix equation

$$A\,x = b. \tag{4.6}$$

If $\langle b\,, v\rangle = 0$ for each v in the cokernel of A, then there exist a solution x for the system 4.6. Let us prove the theorem by contradiction. Assume that there are no solutions for 4.6. That is, there is no any b in the range of A. From our observation before starting this section, we can write b as the sum $b = b_0 + b_r$ where $b_0 \in \mathcal{N}(A^*)$ and $b_r \in \mathcal{R}(A)$, with $b_0 \neq 0$. So,

$$0 = \langle b\,, b_0\rangle = \langle b_0 + b_r\,, b_0\rangle = \langle b_0\,, b_0\rangle + \langle b_r\,, b_0\rangle = \langle b_0\,, b_0\rangle. \tag{4.7}$$

In the first equality we used the assumption that the cokernel is orthogonal to the vector b, and in the last, the fact that the cokernel of A and the range of A are orthogonal. The contradiction is that we said that $b_0 \neq 0$, and found the contrary here. So $b_0 = 0$ and $b = b_r$ is in the range of A. So the equation 4.6 has a solution.

This proof looks as general as we want, and in a way it is. It works even for some infinite dimensional Hilbert spaces. The type of operators for which

[8] Most of the equivalences in mathematics are easy to prove in one direction but not in the other, and here we have a typical situation.

this proof work is called compact operators [9] The bad news is that operators as common as any differential operator are not bounded (and therefore no compact) and the proof does not apply to these particular cases.

Showing that there exists a solution is sometimes enough for a mathematician but never practical for a software developer. The next few sessions will show how to find solutions of linear equations $Lu = v$.

4.3 The Green's function method for the solution of linear equations

4.3.1 motivation

Let us motivate the introduction of the Green's function in the finite dimensional space $\mathbb{C}^n$ which is more familiar to us.

We define the linear operator

$$\begin{aligned} L : \mathbb{C}^n &\rightarrow \mathbb{C}^n \\ \boldsymbol{x} &\mapsto L(\boldsymbol{x}) = A\boldsymbol{x}. \end{aligned} \tag{4.8}$$

Let us assume we want to solve the following sequence of equations

$$A\boldsymbol{x}_i = \boldsymbol{f}_i \quad , \quad i = 1..m. \tag{4.9}$$

Where A is a non–singular matrix and $\boldsymbol{f}_i$ is a sequence of vectors. If we know A^{-1} then the solutions are

$$\boldsymbol{x}_i = A^{-1}\boldsymbol{f}_i, \tag{4.10}$$

that is, we have to compute the inverse only ONCE and apply a matrix vector multiplication to each of the source terms $\boldsymbol{f}_i$. But, what is the inverse of A ?

In order to make an easy transition to the general case, let us make clear the following notation facts. For a matrix B in $\mathbb{C}^n \times \mathbb{C}^n$, and a vector v in $\mathbb{C}^n$ we write

$$\langle B\,, \boldsymbol{v} \rangle_j \tag{4.11}$$

[9] Compact operators are the "natural" extension of matrices to infinite dimensional spaces. For the definition of compact operators and their main properties, the reader can check Kenner's book [4].

to indicate the inner product of the j–th row of the matrix B with the vector $\boldsymbol{v}$. Let us assume that there is a matrix G such that

$$L^* G = I \tag{4.12}$$

where I is the identify matrix. From the adjoint definition we have that

$$\langle \boldsymbol{x}\,, L^* G \rangle_j = \langle L\,\boldsymbol{x}\,, G \rangle_j. \tag{4.13}$$

So,

$$\langle \boldsymbol{x}\,, I \rangle_j = \langle \boldsymbol{f}\,, G \rangle_j, \tag{4.14}$$

that is, the solution $\boldsymbol{x}$ is given by

$$x_j = \langle \boldsymbol{f}\,, G \rangle_j \tag{4.15}$$

which in matrix notation is

$$\boldsymbol{x} = G^H \boldsymbol{f} \tag{4.16}$$

Note that $G = [L^*]^{-1} = [A^{-1}]^H$ as expected. We use this same idea for the general case, as shown next.

4.3.2 the roll of the adjoint in the Green's function method

We want to solve the equation

$$L\,u(\boldsymbol{x}) = f(\boldsymbol{x}), \tag{4.17}$$

where u and b are vectors (functions) in a Hilbert space $\mathcal{H}$.

The problem about finding the Green's function is set up as

$$L^*\,G(\boldsymbol{x}, \boldsymbol{y}) = \delta(\boldsymbol{x} - \boldsymbol{y}). \tag{4.18}$$

But, what is $\delta(\boldsymbol{x} - \boldsymbol{y})$? I want to write a problem that applies equally to many type of operators, either in the continuum or the discrete; either differential or integral. What would be this δ in the discrete?

A physicist could see the Dirac delta as

$$\delta(\boldsymbol{x} - \boldsymbol{x}) = \begin{cases} 0 & \text{if } \boldsymbol{x} \neq \boldsymbol{y} \\ \lim_{\boldsymbol{x}\to\boldsymbol{y}} \frac{d\boldsymbol{x}}{\|\boldsymbol{x}-\boldsymbol{y}\|} \approx 1 & \text{if } \boldsymbol{x} = \boldsymbol{y}. \end{cases} \tag{4.19}$$

Something clearly 0 for $\boldsymbol{x} \neq \boldsymbol{y}$, but that integrates to 1 as we approach $\boldsymbol{x} = \boldsymbol{y}$. With this interpretation it is clear that the corresponding discrete version of the Dirac Delta is the Kronecker delta

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases} \tag{4.20}$$

Since the Kronecker delta δ_{ij} represents the identity matrix, then we see how the equation 4.18 matches the discrete case in equation 4.12.

However, the definition 4.19 lacks rigor and should be formalized. The definition of a Dirac delta rests in the world of "Distribution Theory". Some authors [2] prefer to use the name "Generalized Functions" but I will keep the name distribution theory, which was the original name (translated to English) proposed by Laurent Schwartz, the inventor. I only touch over the surface of distribution theory to collect the minima elements to attack the question in the title.

First, I define a functional. A *functional* is a linear function from the Hilbert space $\mathcal{H}$ into the field $\mathbb{f}$ (here the complex numbers $\mathbb{C}$). A way to create a functional is to take an element of the Hilbert space $\mathcal{H}$ and find the inner product of it with each element of $\mathcal{H}$. In fact, it can be shown (see Riesz theorem, [5]) that all continuous linear functionals in a Hilbert space are inner products. The set of all functionals in a Hilbert space makes another Hilbert space known as the *dual* .

Let us then pick a vector ϕ in a Hilbert space of functions over the complex numbers $\mathbb{C}$, and create a functional L_ϕ by taking the inner product of ϕ with each vector of the Hilbert space. That is,

$$\begin{aligned} L : \mathcal{H} &\to \mathbb{C} \\ \alpha &\mapsto L_\phi(\alpha) = \langle \alpha \, , \phi \rangle \end{aligned} \tag{4.21}$$

If we use definition 2.6, for example, then

$$L_\phi(\alpha) = \int_a^b \alpha(\boldsymbol{x}) \, \phi(\boldsymbol{x}) d\boldsymbol{x}. \tag{4.22}$$

The integral smooths out the meaning of α. It would be good if we could unveil what α is. We show how this is possible thanks to the introduction of the Dirac delta.

The Dirac delta is a functional that maps a vector (function, $\phi(\boldsymbol{x})$) in $\mathcal{H}$ into its evaluation at $\boldsymbol{x} = 0$. That is

$$\begin{aligned} \delta : \mathcal{H} &\rightarrow \mathbb{C} \\ \phi(\boldsymbol{x}) &\mapsto \phi(0). \end{aligned} \tag{4.23}$$

If we write the functional δ as an inner product (using definition 2.6) then we have that

$$\int_a^b \phi(\boldsymbol{x})\delta(\boldsymbol{x})d\boldsymbol{x} = \phi(0). \tag{4.24}$$

[10]

So the Dirac delta extracts the zero coordinate of the vector ϕ. We first see that the Dirac delta works the same for $\phi(\boldsymbol{x})$ that for $\phi(-\boldsymbol{x})$. We say that the Dirac delta is even. From now on, we consider the limits a and b expanded to infinity. In distribution theory the functions are localized in finite intervals with the help of neutralizers (test functions).

To extract any value from the function $\phi(\boldsymbol{x})$ we can use the shifting property of the integrals as follows

$$\int_a^b \phi(\boldsymbol{y})\delta(\boldsymbol{x}-\boldsymbol{y})d\boldsymbol{y} = \int_{-b}^{-a} \phi(\boldsymbol{x}-\boldsymbol{w})\delta(\boldsymbol{w})d\boldsymbol{w} = \phi(\boldsymbol{x}). \tag{4.25}$$

where we used the change of variables $\boldsymbol{w} = \boldsymbol{x}-\boldsymbol{y}$. This is the sifting property of the Dirac delta and we can write it in bracket form as

$$\langle\, \phi\,, \delta(\boldsymbol{x}-\boldsymbol{y})\rangle = \phi(x) * \delta(\boldsymbol{x}) = \phi(\boldsymbol{x}). \tag{4.26}$$

So the Dirac delta is the identity under the convolution operation. Similarly we observe that the Kronecker delta δ_{ij} satisfies:

$$\langle \boldsymbol{x}\,, \delta_{ij}\rangle = \sum_{j=1}^{n} \delta_{ij}x_j = x_i. \tag{4.27}$$

It is clear the similarity between equations 4.27 for the discrete and 4.26 for the continuum.

[10] If instead of using the definition 2.6, we would have used the definition 2.7, then the delta symbol $\delta(\boldsymbol{x})$, should carry the inverse of the weight $w(\boldsymbol{x})$. That is, we still need to end up with $\phi(0)$ regardless the definition of inner product.

Let us now see how a Green function satisfying equation 4.18 can help us solve the general equation $L\,u = f$. From the definition of adjoint we have

$$\langle Lu\,, G(\boldsymbol{x}, \boldsymbol{y})\rangle = \langle u, L^* G(\boldsymbol{x}, \boldsymbol{y})\rangle \tag{4.28}$$

That is,

$$\langle f\,, G(\boldsymbol{x}, \boldsymbol{y})\rangle = \langle u, \delta(\boldsymbol{x} - \boldsymbol{y})\rangle = u(\boldsymbol{x}). \tag{4.29}$$

So the solution of $Lu = f$ is given by

$$u(\boldsymbol{x}) = \langle f\,, G(\boldsymbol{x}, \boldsymbol{y})\rangle, \tag{4.30}$$

with $G(\boldsymbol{x}, \boldsymbol{y})$, the Green's function of the adjoint operator L^*. With the inner product definition 2.6 we can write equation 4.30 turns out to be

$$u(\boldsymbol{x}) = \int_V \overline{G}(\boldsymbol{x}, \boldsymbol{y})\, f(\boldsymbol{y}) d\boldsymbol{y}, \tag{4.31}$$

which is an integral operator of the Fredholm type. Therefore the adjoint of the kernel $\overline{G}(\boldsymbol{x}, \boldsymbol{y})$ is the kernel $G(\boldsymbol{y}, \boldsymbol{x})$ (see equation 3.122). So by, following the same reasoning, this time for the equation

$$L^* v(\boldsymbol{x}) = f(\boldsymbol{x}) \tag{4.32}$$

we find the solution

$$v(\boldsymbol{x}) = \int_V G(\boldsymbol{y}, \boldsymbol{x})\, f(\boldsymbol{y}) d\boldsymbol{y}, \tag{4.33}$$

or in bracket notation

$$v(\boldsymbol{x}) = \langle f\,, \overline{G}(\boldsymbol{y}, \boldsymbol{x})\rangle. \tag{4.34}$$

If the operator is self–adjoint, then $L = L^*$, and so $v(\boldsymbol{x}) = u(\boldsymbol{x})$, that is

$$\langle f\,, G(\boldsymbol{x}, \boldsymbol{y})\rangle = \langle f\,, \overline{G}(\boldsymbol{y}, \boldsymbol{x})\rangle, \tag{4.35}$$

and since this should be true for any arbitrary source (test function) f, then we conclude that

$$G(\boldsymbol{x}, \boldsymbol{y}) = \overline{G}(\boldsymbol{y}, \boldsymbol{x}). \tag{4.36}$$

This is known as the reciprocity relation for Green's functions of self–adjoint operators.

The next logical step is the finding of the Green's function. There is not a unique method to find Green's functions. For example, one way to find the free space Green's function for the homogeneous acoustic wave equation

$$\nabla^2 G(\boldsymbol{x}, \boldsymbol{x}_s, t, t_s) - \frac{1}{c} \frac{\partial^2 G(\boldsymbol{x}, \boldsymbol{x}_s, t, t_s)}{\partial t^2} = \delta(\boldsymbol{x} - \boldsymbol{x}_s)\delta(t - t_s), \tag{4.37}$$

could be carried out by taking the multi–dimensional Fourier transform

$$\left[\boldsymbol{k}^2 - \frac{\omega^2}{c}\right] \hat{G}(\boldsymbol{k}, \boldsymbol{x}_s, \boldsymbol{\omega}, t_s) \mathrm{e}^{(\omega\, t_s - \boldsymbol{k}\cdot\boldsymbol{x}_s)} = 1, \tag{4.38}$$

and then perform contour integration on the wavenumber $\boldsymbol{k}$ and frequency ω domain. That is,

$$G(\boldsymbol{x}, \boldsymbol{x}_s, t, t_s) = \left(\frac{1}{2\pi}\right)^{(n+1)/2} \int \frac{\mathrm{e}^{\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{x}_s)-\omega(t-t_s)}}{\boldsymbol{k}^2 - \omega^2/c^2} d\boldsymbol{k}\, d\omega. \tag{4.39}$$

Here $\boldsymbol{k}^2 = \boldsymbol{k}\cdot\boldsymbol{k}$ and the integration is carried out over all $n+1$th–dimensional space of wavenumbers and frequencies. I prefer to use the factor outside of the integral as a square root of a power of $1/2\pi$ for symmetry reasons. I will not evaluate this integral (it is solved in many textbooks). The purpose here is only to show one way in which a Green's function is computed.

An interesting Green's function is that needed to solve integral equations. The Fredholm integral equation of the second kind, it given by [11].

$$u(\boldsymbol{x}) = f(\boldsymbol{x}) + \frac{1}{\lambda} \int k(\boldsymbol{x}, \boldsymbol{y}) u(\boldsymbol{x}) d\boldsymbol{x}. \tag{4.40}$$

This integral equation is fundamental for the solution of multiple scattering problems. It can be written in operator form as

$$(K - \lambda I)u = -f. \tag{4.41}$$

The Green's function is written as

$$G = (K - \lambda I)^{-1}. \tag{4.42}$$

[11] This is the Fredholm operator defined in 3.121, but with a source term f and a scaling factor λ

and it is also known as the resolvent of K. This operator is usually written in terms of its power series expansion:

$$G = \sum_{i=0}^{\infty} (1/\lambda)^i K^i, \tag{4.43}$$

and is also known as the Neumann series, from which its linear term

$$G \approx I + \frac{1}{\lambda} K \tag{4.44}$$

is the Born approximation. Of course we have to know when the series expansion for G makes sense. If norm of the $\lambda^{-1} K$ operator is less than 1, then there is convergence.

If λ is an eigenvalue of the Fredholm operator K, then G is singular. We will have more to say about the Fredholm operator in the next section.

I want to treat the solution of linear operators in a more general framework, therefore in the next section, I indicate a method for the expansion of a Green's function in terms of its eigenfunctions.

4.3.3 the spectrum of an operator and its use to find a Green function

A few bullets for this section:

- Representation of the Green's function an terms of eigenfunctions.
- Representation of the resolvent in terms of the Green's function.
- Representation of a Dirac delta in terms of a Green's function.
- Representation of a transform pair from the Green's function.

Let me motivate this section with the linear operators defined in $\mathbb{R}^n$. That is the space of $n \times n$ matrices. The easiest system to solve is that represented by the diagonal matrix. Here the diagonal elements are the eigenvalues of the matrix and the solution of the equation

$$A\,\boldsymbol{x} = \boldsymbol{y}, \tag{4.45}$$

is given by

$$x_i = \frac{y_i}{\lambda_i} \tag{4.46}$$

with λ_i the i–th diagonal entry (eigenvalue) of the matrix A, assuming, of course, that each $\lambda_i \neq 0$. Unfortunately, is almost never the case, that our system is given as a diagonal system.

The next level of difficulty would be that of a system $A\,\boldsymbol{x} = \boldsymbol{y}$, that is not diagonal but it is diagonalizable. That is, assume that we can write the matrix A as

$$A = P^{-1}\,DP, \tag{4.47}$$

for P a non–singular matrix and D a diagonal matrix. Then the system 4.45 can be writen as

$$P^{-1}\,D\,P\,\boldsymbol{x} = \boldsymbol{y}. \tag{4.48}$$

If we multiply both sides of the previous equation by P we find

$$DP^{-1}\boldsymbol{x} = P^{-1}\,\boldsymbol{y}. \tag{4.49}$$

So by calling $\boldsymbol{z} = P^{-1}\boldsymbol{x}$, and $\boldsymbol{w} = P^{-1}\,\boldsymbol{y}$ we have the system

$$D\boldsymbol{z} = \boldsymbol{w}, \tag{4.50}$$

where now the system is diagonal, and we can go back to the solution

$$z_i = \frac{w_i}{\lambda_i} \Rightarrow \boldsymbol{x} = P^{-1}D^{-1}P\boldsymbol{y} \tag{4.51}$$

λ_i the diagonal value. We will see that the λ_i are in this case also eigenvalues of the matrix A. The transformation matrix P is a change of basis matrix that will be seen later as a transform (such as a Fourier transform or any other transform). We take the "data" $\boldsymbol{y}$ and transform it to a new coordinate system with friendlier coordinates. We solve the easy, diagonal system, and finally we have to transform back by pre–muiltiplying the solution in the new basis with the inverse (transform) P^{-1}. In this way we return to the original basis. But this is not always easy.

From equation 4.48, by pre–multiplying by P both sides we find

$$A\,P = P\,D \tag{4.52}$$

That is, we get n equations

$$A\,\boldsymbol{p}_i = \lambda_i\,\boldsymbol{p}_i \tag{4.53}$$

where $\boldsymbol{p}_i$ is the i–the column of P. We recognize that the columns of P are the eigenvectors of A, and the λ_i are the eigenvalues of A. The eigenvalues of A are known as the spectrum of the operator, the concept of spectrum is quite different if the space is not finite dimensional, as will be shown below. Equation 4.53 is actually the starting point of the diagonalization process and that is how we will solve the problem for a general operator L in the Hilbert space $\mathcal{H}$. If the matrix A is self–adjoint, then we showed in section 3.2.2.6 that their eigenvectors are orthogonal. This is the case for all the eigenfunctions of the Sturm–Liouville operator shown in table 2.

Let us now move to a more general problem of an operator L in a Hilbert space $\mathcal{H}$ which spectrum is discrete, no necessarily finite.

We already showed in section 3.2.2.6 that if an operator L is self–adjoint the eigenfunctions corresponding to different eigenvalues are orthogonal. We will soon see how convenient is to have orthogonality of those eigenfunctions.

But, can we guarantee that, even if the operator is self–adjoint, the eigenvalues exits? This is certainly true for finite dimensional spaces but it is not for infinite–dimensional dimensional spaces and I provide an example here.

Let us define $\mathcal{H}$ as the space of square integrable functions in the interval $[0, 1]$. That is, given $\mathcal{H} = L^2[0, 1]$, we define

$$\begin{aligned} L : \mathcal{H} &\rightarrow \mathcal{H} \\ f(x) &\mapsto x\, f(x). \end{aligned} \tag{4.54}$$

If λ is an eigenvalue of L then

$$L[f(x)] = \lambda f(x) = x f(x). \tag{4.55}$$

So

$$f(x)(\lambda - x) = 0, \tag{4.56}$$

and since λ is independent of x then $f(x) = 0$. Even, if $f(x) \neq 0$ at the single value λ, as a function of $L^2[0, 1]$ we say that $f = 0$. Therefore this operator does not have eigenvalues [12].

Yet another example is the following. The Volterra operator

$$V(f(x)) = \int_0^1 f(t)k(x, t)dt,$$

[12] for λ to be an eigenvalue the corresponding eigenvector should not be zero

is linear in (no self–adjoint) f. We will think that f and its range is in the space $L^2[0,1]$. If the kernetl $k(x,t) = 1_{[0,x]}(t)$, that is the characteristic function of the set $[0,x]$ defined as

$$1_{[0,x]}(t) = \begin{cases} 1 & 0 \leq t \leq x, \\ 0 & \text{otherwise} \end{cases}$$

then

V, as the integration in the $[0,x]$ interval. That is,

$$V(f) = \int_0^x f(t)dt$$

We show that V does not have eigvenvalues/eigenvectors . Let us assume that $\lambda \neq 0$ is an eigenvalue of V that is $Vf = \lambda f$, then

$$f(x) = \frac{1}{\lambda}\int_0^x f(t)dt. \tag{4.57}$$

Now using the fact that f is in L^2 we have that for $x < y$

$$\begin{aligned} |f(y) - f(x)| &= \frac{1}{\lambda}\left|\int_x^y f(t)dt\right| \leq \frac{1}{\lambda}\int_x^y |f(t)|dt = \frac{1}{\lambda}\int_0^1 |f(t)|1_{[x,y]}(t)dt \\ &\leq \frac{\|f\|_2}{\lambda}\left(\int_0^1 (1_{[x,y]})^2 dt\right)^{1/2} = \frac{\|f\|_2}{\lambda}\sqrt{y-x}, \end{aligned}$$

so f is continouos. In additon f is differentiable and taking derivatives in 4.57 we find $f = \lambda f'$. The solution of this elementary differential equation is $f(t) = ce^{t/\lambda}$, and since $f(0) = 0$, then $f(t) = 0$. So there can not be non-trivial eigenfunctions.

4.3.3.1 The point spectrum: If the eigenfunctions of L expand the whole space $\mathcal{H}$, [13] we can express any vector f as a linear combination of the eigenfunctions. That is , for

$$Lu = f, \tag{4.58}$$

[13] This is true in many cases. Particularly in all finite dimensional non–singular operators and in the differential operators that we are interested (wave equation operators.) If the set of eigenfunctions expand the whole space, we say that the set is complete. I will ignore those cases of non–complete sets of eigenfunctions.

we can expand

$$f = \sum_i c_i \, \phi_i \tag{4.59}$$

for some complex coefficients c_i. If the eigenfunctions ϕ_i are orthogonal, the c_i are easy to find, otherwise the eigenfunctions could be made orthonormal by a Grand–Schmidt orthogonalization method. I assume, to avoid introducing complexity to the problem, that the eigenfunctions are orthonormal (that is, they are orthogonal and scaled by the inverse of their norm). If the operator L is self–adjoint this is not a big issue. Recall table 2, which shows the sets of orthogonal eigenfunctions generated by the Sturm–Liouville operator under different boundary conditions and definitions of inner products. For the purpose of answering the question in this document we have enough with these assumptions. We now take the inner of product 4.59 with an arbitrary function ϕ_j and then

$$\langle f \, , \phi_j \rangle = \sum_i c_i \, \langle \phi_i \, , \phi_j \rangle = \sum_i c_i \, \delta_{ij} = c_j. \tag{4.60}$$

where we used the orthonormality condition $\langle \phi_i \, , \phi_j \rangle = \delta_{ij}$. The c_j are known as the Fourier coefficients, because in the particular case that the eigenfunctions are trigonometrical polynomials the c_i are the weights of the Fourier series approximation to the function f. We can write equations 4.59 and 4.60 as the pair:

$$\begin{aligned} f &= \langle c_i \, , \overline{\phi_i} \rangle \\ c_j &= \langle f \, , \phi_j \rangle \end{aligned} \tag{4.61}$$

This is a transform pair. The great symmetry here is due to the orthonormality of the eigenfunctions ϕ_j. Otherwise, some scaling factors would be in one and the other side to balance things.

If u is a solution of the equation $Lu = f$, then since the set of eigenfunctions ϕ_i expand the whole space, there are coefficients α_i such that

$$u = \sum_i \alpha_i \, \phi_i, \tag{4.62}$$

and

$$L \, u = L \left[\sum_i \alpha_i \, \phi_i \right] = \sum_i \alpha_i \, L[\phi_i] = \sum_i \alpha_i \lambda_i \phi_i \tag{4.63}$$

and matching coefficients in equations 4.63 and 4.59 we find that

$$\alpha_i = \frac{c_i}{\lambda_i} = \frac{\langle f\,, \phi_i \rangle}{\lambda_i}. \tag{4.64}$$

and so

$$u = \sum_i \frac{\langle f\,, \phi_i \rangle}{\lambda_i} \phi_i, \tag{4.65}$$

provided that $\lambda_i \neq 0$ for each i. This is a very interesting result. It says that the solution of $Lu = f$ is representation of f in the eigenfunction basis, where each projection coefficient (Fourier coefficient) is inverse weighted by the spectrum (frequency in the case of Fourier) value. Of course this representation fails when an eigenvalue is 0, and we also assume that the set of eigenvalues exist, is discrete and the sum 4.65 converges. Since the Green's function satisfies equation 4.29, then from equation 4.65 we find that [14]

$$G(\boldsymbol{x}, \boldsymbol{y}) = \sum_i \frac{1}{\lambda_i} \phi_i(\boldsymbol{x})\, \phi_i(\boldsymbol{y}). \tag{4.67}$$

4.3.3.2 When things do not work that well

The Jordan canonical form.....

4.4 The Least square problem: pseudoinverse

We have now the elements to solve the question that motivated this document.

5 The response to the question.

As

[14] Note that we have been ignoring the weight w used in the inner product, definition 2.7. If we would have included it in our derivations we would have found that

$$G(\boldsymbol{x}, \boldsymbol{y}) = w(\boldsymbol{x}) \sum_i \frac{1}{\lambda_i} \phi_i(\boldsymbol{x})\, \phi_i(\boldsymbol{y}). \tag{4.66}$$

References

[1] N. Bleistein, J.K. Cohen, and J. Stockwell. *Mathematics of multidimensional seismic Migration, Imaging and Inversion.* Springer–Verlag, 2000.

[2] I. M. Gel'fand and G. E. Shilov. *Generalized functions, Volume 1:Properties and Operations.* Academic Press, 1964.

[3] H Jaramillo. An informal argument about the principle of gradient computation in wave form inversion (wfi). *my own notes*, 2012.

[4] J. P. Keener. *Principles of Applied Mathematics*, page 156. 1988.

[5] Erwin Kreyszig. *Introductory Functional Analysis with Application.* John Wiley & Sons Inc., 1978.

[6] L. A. Segel. *Mathematics Applied to Continuum Mechanics*, page 557. 1987.