Functional Analysis II
Fall 2009
Prof. W. Hrusa*

Chris Almost†

Contents

*wh15@andrew.cmu.edu
†cdalmost@cmu.edu

1 Spectral Theory in Hilbert Spaces

1.1 Orthogonal Projections

Projection onto a convex set

This chapter is concerned with the geometric structure of linear transformations, i.e. spectral theory. Let X be a non-trivial Hilbert space over $\mathbb{K}$ with inner product $(\cdot,\cdot)$.

1.1.1 Theorem (Projection on a closed convex set). *Let X be a Hilbert space and $K \subseteq X$ be non-empty, closed, and convex. Let $x \in X$ be given. Then there is exactly one point $y_0 \in K$ such that $\|x - y_0\| = \inf\{\|x - y\| \mid y \in K\}$*

1.1.2 Theorem (Projection). *Let X be a Hilbert space and M a closed subspace of X. Let $x \in X$ be given. Then there is exactly one pair $(y,z) \in M \times M^{\perp}$ such that $x = y + z$.*

In fact, y is the point y_0 for $K = M$ in the previous theorem.

Orthogonal projections

Let M be a closed subspace. For each $x \in X$ let $P_M x$ be the unique element of M such that $\|x - P_M x\| \le \|x - y\|$ for all $y \in M$. Then P_M is linear, and $x - P_M x \in M^{\perp}$. Clearly $x = (x - P_M x) + P_M x$ and it follows that $\|P_M x\| \le \|x\|$, so P_M is continuous. It can be shown that $P_{M^{\perp}} = I - P_M$.

Orthonormal lists and bases

As list $(e_i \mid i \in I)$ is an *orthonormal list* provided that for all $i, j \in I$ we have $e_i \perp e_j$ if $i \neq j$ and $\|e_i\| = 1$. An orthonormal list is a *maximal orthonormal list* or an *orthonormal basis* if for all $x \in X$, if $x \perp e_i$ for all $i \in I$ then $x = 0$. These terms apply to sets of elements by considering it a self-indexed list.

Let $(e_i \mid i \in I)$ be an orthonormal basis and $x \in X$. Then $x = \sum_{i \in I}(x, e_i)e_i$, where there (generally uncountable) sum is interpreted in the usual way.

Duality

1.1.3 Theorem (Riesz Representation). *Let $x^* \in X^*$ be given. Then there exists exactly one $y \in X$ such that $\langle x^*, x\rangle = (x, y)$ for all $x \in X$, and moreover $\|x^*\|_* = \|y\|$.*

Last semester we introduced the *Riesz operator* $R : X \to X^*$, defined by $(R(y))(x) := (x, y)$. It is important to note that R is conjugate linear isometry.

Let $A \in \mathscr{L}(X;X)$ be given. The *Banach space adjoint* is $A^* : X^* \to X^*$, defined by $\langle A^* x^*, x\rangle = \langle x^*, Ax\rangle$. The *Hilbert space adjoint* is $A^*_H : X \to X$, defined by

$(Ax, y) = (x, A_H^* y)$. These adjoints are related by $A_H^* = R^{-1}A^*R$, and they are not equal in general.
Warning: Until further notice we will use A^* for the Hilbert space adjoint.

1.2 Self-adjoint Operators

Self-adjoint operators

We say that A is *self-adjoint* (or *Hermitian* in the complex case) provided $A = A^*$. We say that A is *normal* provided that it commutes with its adjoint, i.e. that $AA^* = A^*A$.

1.2.1 Proposition. *Assume that $\mathbb{K} = \mathbb{C}$ and let $A \in \mathscr{L}(X;X)$ be given. Then $A = A^*$ if and only if $(Ax, x) \in \mathbb{R}$.*

PROOF: If A is self-adjoint then $(Ax, x) = (x, Ax) = \overline{(Ax, x)}$ so $(Ax, x) \in \mathbb{R}$. Conversely, let $x, y \in X$ and $\alpha \in \mathbb{C}$ be given. Then $(A(x + \alpha y), x + \alpha y) \in \mathbb{R}$, so in particular $\alpha(Ay, x) + \overline{\alpha}(Ax, y)x \in \mathbb{R}$. It follows that

$$\begin{aligned} \alpha(Ay, x) + \overline{\alpha}(Ax, y)x &= \overline{\alpha}(x, Ay) + \alpha(y, Ax) \\ &= \overline{\alpha}(A^*x, y) + \alpha(A^*y, x) \end{aligned}$$

Take $\alpha = 1$ and $\alpha = i$ and reduce to see $(Ax, y) = (A^*x, y)$. Since this holds for arbitrary $x, y \in X$, it follows that $A = A^*$. □

1.2.2 Proposition. *If $A = A^*$ then $\|A\| = \sup\{|(Ax, x)| \mid x \in X, \|x\| = 1\}$.*

PROOF: Let $M = \sup\{|(Ax, x)| \mid x \in X, \|x\| = 1\}$, and let $x \in X$ with $\|x\| = 1$ be given. Then $|(Ax, x)| \le \|Ax\|\|x\| = \|Ax\| \le \|A\|$ so $M \le \|A\|$. Note that the self-adjointness of A was not used in this calculation.

Conversely, let $x, y \in X$ with $\|x\| = \|y\| = 1$ be given. Then

$$\begin{aligned} (A(x + y), x + y) &= (Ax, x) + (Ax, y) + (Ay, x) + (Ay, y) \\ &= (Ax, x) + (Ax, y) + (y, A^*x) + (Ay, y) \\ &= (Ax, x) + (Ax, y) + (y, Ax) + (Ay, y) \\ &= (Ax, x) + 2\Re(Ax, y) + (Ay, y) \\ \text{and } (A(x - y), x - y) &= (Ax, x) - 2\Re(Ax, y) + (Ay, y) \\ \text{so } 4\Re(Ax, y) &= (A(x + y), x + y) - (A(x - y), x - y) \end{aligned}$$

Whence

$$\begin{aligned} 4|\Re(Ax, y)| &\le |(A(x + y), x + y)| + |(A(x - y), x - y)| \\ &\le M(\|x + y\|^2 + \|x - y\|^2) \\ &\le M(2\|x\|^2 + 2\|x\|^2) = 4M \end{aligned}$$

Choose $\theta \in [0, 2\pi)$ such that $(Ax, y) = e^{i\theta}|(Ax, y)|$ (note that if $\mathbb{K} = \mathbb{R}$ then $\theta \in \{0, \pi\}$). Put $z = e^{-i\theta}x$. Then $\|z\| = 1$ and it follows that $|(Ax, y)| = \Re(Az, y) \le M$. Recall that $\|Ax\| = \sup\{|(Ax, y)| \mid y \in X, \|y\| = 1\}$, so $\|Ax\| \le M$. Since this holds for all $x \in X$ with $\|x\| = 1$, $\|A\| \le M$. $\square$

1.2.3 Corollary. *If $A = A^*$ and $(Ax, x) = 0$ for all $x \in X$ then $A = 0$.*

Note that the self-adjointness is required in the real case. Rotation R by $\frac{\pi}{2}$ in $\mathbb{R}^2$ is a non-zero linear operator but satisfies $(Rx, x) = 0$ for all $x \in \mathbb{R}^2$. Self-adjointness is *not* required in the complex case. Prove this as an exercise.

Normal operators

1.2.4 Proposition. *A is normal if and only if $\|Ax\| = \|A^*x\|$ for all $x \in X$.*

PROOF: Let $x \in X$ be given.

$$\begin{aligned}\|Ax\|^2 - \|A^*x\|^2 &= (Ax, Ax) - (A^*x, A^*x)\\ &= (A^*Ax, x) - (AA^*x, x)\\ &= ((A^*A - AA^*)x, x)\end{aligned}$$

Since $A^*A - AA^*$ is self-adjoint, the above expression is zero if and only if $A^*A = AA^*$. $\square$

1.2.5 Corollary. *If A is normal then* $\ker(A) = \ker(A^*)$.

Isometries

1.2.6 Definition. A is said to an *isometry* if $\|Ax\| = \|x\|$ for all $x \in X$. A surjective isometry is said to be *unitary*.

The right-shift operator on ℓ^2 is an isometry that is not surjective.

1.2.7 Proposition. *A is an isometry if and only if $(Ax, Ay) = (x, y)$ for all $x, y \in X$.*

PROOF: Assume that A is an isometry and let $x, y \in X$ and $\alpha \in \mathbb{K}$ be given. Then

$$\begin{aligned}\|A(x + \alpha y)\|^2 &= (A(x + \alpha y), A(x + \alpha y))\\ &= \|Ax\|^2 + 2\Re(\alpha(Ay, Ax)) + |\alpha|^2\|Ay\|^2\\ &= \|x\|^2 + 2\Re(\alpha(Ay, Ax)) + |\alpha|^2\|y\|^2\\ \|x + \alpha y\|^2 &= \|x\|^2 + 2\Re(\alpha(y, x)) + |\alpha|^2\|y\|^2\end{aligned}$$

Since the terms on the left are equal, $\Re(\alpha(y, x)) = \Re(\alpha(Ay, Ax))$. If $\mathbb{K} = \mathbb{R}$ then we are done. If $\mathbb{K} = \mathbb{C}$ then put $\alpha = 1, i$ to get the result. The other direction is clear. $\square$

1.2.8 Proposition. *A is isometric if and only if $A^*A = I$.*

PROOF: A is isometric if and only if $(Ax, Ay) = (x, y)$ for all $x, y \in X$, which is equivalent to $(A^*Ax, y) = (x, y)$ for all $x, y \in X$, which holds if and only if $x = A^*Ax$ for all $x \in X$. □

Warning: The theorem does not necessarily hold if we look at AA^* instead. Let R denote the right shift operator and L denote the left shift operator. R is an isometry, while L is not. Notice that $R^* = L$ and $L^* = R$, so $R^*R = LR = I$ and $RR^* = RL \neq I$. Keep in mind that R and L are not normal operators.

1.2.9 Proposition. *Assume that A is isometric. Then A is normal if and only if A is surjective.*

PROOF: If A is normal then $I = A^*A = AA^*$ so A is surjective since I is surjective. If A is surjective then A is invertible (since isometries are automatically injective) and A^{-1} is an isometry. Whence $(A^{-1})^*A^{-1} = I$, and since * commutes with $^{-1}$, we get $I = (A^*)^{-1}A^{-1} = (AA^*)^{-1}$, so $AA^* = I = A^*A$ since A is an isometry. □

1.3 Idempotent operators

1.3.1 Definition. $E \in \mathscr{L}(X; X)$ is said to be idempotent provided $E^2 = E$.

Note that every orthogonal projection is idempotent, but not every idempotent operator is an orthogonal projection. E.g. non-orthogonal projections are idempotent. More specifically, for $X = \mathbb{R}^2$, $E = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ is idempotent but not orthogonal. Note that E is not normal and $\|E\| = \sqrt{2} > 1$.

Let X be a Hilbert space and E an idempotent operator on X. Then $(I - E)^2 = I^2 - 2E + E^2 = I - E$, so $I - E$ is also idempotent. Now $x \in \ker(I - E)$ if and only if $Ex = x$, so $\ker(I - E) \subseteq \text{range}(E)$. Conversely, $y \in \text{range}(E)$ if there is $x \in X$ such that $y = Ex$, so $Ey = E^2x = Ex = y$ and so $y \in \ker(I - E)$. In particular $\ker(I - E) = \text{range}(E)$ and the range of E is closed.

Recall that $\ker(A) = \text{range}(A^*)^\perp$.

1.3.2 Proposition. *Assume that E is idempotent and put $M := \text{range}(E)$. Then $E = P_M$ if and only if $\ker(E) = \text{range}(E)^\perp$.*

PROOF: Exercise. □

If E is idempotent then $\|E\| = \|E^2\| \leq \|E\|^2$, so $\|E\| \geq 1$.

1.3.3 Proposition. *Assume that $E^2 = E$ and $E \neq 0$, and put $M := \text{range}(E)$. Then $E = P_M$ if and only $\|E\| = 1$.*

PROOF: If $E = P_M$ then for every $x \in X$, $x = P_M x + (I - P_M)x$, and these components are orthogonal. Therefore

$$\|x\|^2 = \|P_M x\|^2 + \|(I - P_M)x\|^2 \geq \|P_M x\|^2,$$

so $\|E\| \leq 1$ and it must be equal since E is a non-zero idempotent.

Assume $\|E\| = 1$. By 1.3.2, since range(E) is a closed subspace, it suffices to show that range$(E) = \ker(E)^{\perp}$. Let $x \in \ker(E)^{\perp}$ be given. Notice that range$(I - E) = \ker(E)$, so

$$0 = (x - Ex, x) = \|x\|^2 - (Ex, x) \geq \|x\|(\|x\| - \|Ex\|).$$

Since $\|E\| = 1$, it must be the case that $\|Ex\| = \|x\|$, and it follows that $\|Ex\| = \sqrt{(Ex, x)}$. Whence $\|x - Ex\|^2 = 0$ and $x \in$ range(E). Conversely, let $y \in$ range(E) be given. Write $y = x + z$ with $x \in \ker(E)$ and $z \in \ker(E)^{\perp} \subseteq$ range(E). Then $y = Ey = E(x + z) = Ez = z$, so $y \in \ker(E)^{\perp}$. □

1.3.4 Proposition. *Assume that $E^2 = E$ and let $M =$ range(E). The following are equivalent.*

(i) $E = P_M$*;*
(ii) $\|E\| = 1$*;*
(iii) $E = E^*$*;*
(iv) E *is normal.*

PROOF: (i) and (ii) are equivalent by 1.3.3. Assume $E = P_M$ and let $x, y \in X$. Write $x = x_1 + x_2$ and $y = y_1 + y_2$ where $x_1, y_1 \in M$ and $x_2, y_2 \in M^{\perp}$.

$$(Ex, y) = (Ex_1 + Ex_2, y_1 + y_2) = (x_1, y_1) = (x_1 + x_2, Ey_1 + Ey_2) = (x, Ey),$$

so (i) implies (iii). Clearly (iii) implies (iv). Assume that E is normal. Then for all $x \in X$, $\|Ex\| = \|E^* x\|$, and it follows that $\ker(E) = \ker(E^*) =$ range$(E)^{\perp}$. By 1.3.2, $E = P_M$ and (iv) implies (i). □

1.3.5 Proposition. *Assume $E^2 = E$ and put $M =$ range(E). Then $E = P_M$ if and only if $(Ex, x) \geq 0$ for all $x \in X$ (i.e. is real and non-negative).*

PROOF: Assume $E = P_M$, and note that, for any $x \in X$, we can write $x = x_1 + x_2$ with $x_1 \in M$ and $x_2 \in M^{\perp}$, so $(Ex, x) = (x_1, x_1) \geq 0$.

Conversely, in the complex case $(Ex, x) \in \mathbb{R}$ for all $x \in X$ implies that E is self-adjoint, so it is a projection by 1.3.4. In the real case, let $x \in X$ be given and write $x = Ex + (I - E)x =: y + z$.

$$0 \leq (Ey + Ez, y + z) = (y, y + z) = \|y\|^2 + (y, z),$$

Therefore $(y, z) = 0$ for all $y \in$ range(E) and $z \in$ range$(I - E) = \ker(E)$ and E is orthogonal. □

1.4 Spectral Theory

Invariant and reducing subspaces

Let X be a Hilbert space and $M \leq X$ a closed subspace. Note that $X = M \oplus M^{\perp}$. We can write an operator $A \in \mathscr{L}(X;X)$ as

$$\begin{pmatrix} P_M(Ax) \\ P_{M^\perp}(Ax) \end{pmatrix} = \begin{pmatrix} B & C \\ D & F \end{pmatrix} \begin{pmatrix} P_M x \\ P_{M^\perp} x \end{pmatrix}$$

where $B \in \mathscr{L}(M;M)$, $C \in \mathscr{L}(M^{\perp};M)$, $D \in \mathscr{L}(M;M^{\perp})$, and $F \in \mathscr{L}(M^{\perp};M^{\perp})$.

We say that M is *invariant under* A provided that $A[M] \subseteq M$, and M *reduces* A provided that $A[M] \subseteq M$ and $A[M^{\perp}] \subseteq M^{\perp}$. Notice that M reduces A if and only if $C = 0$ and $D = 0$. M ($M^{\perp}$) is invariant under A if and only if $D = 0$ ($C = 0$).

1.4.1 Proposition. *M is invariant under A if and only if $M^{\perp}$ is invariant under A^*.*

PROOF: Assume that M is invariant under A and let $x \in M$ and $y \in M^{\perp}$ be given. Then $(x, A^* y) = (Ax, y) = 0$. It follows that $A^* y \in M^{\perp}$, so $M^{\perp}$ is invariant under A^*. □

1.4.2 Proposition. *M is invariant under A if and only if $P_M A P_M = A P_M$.*

PROOF: Assume M is invariant under A and let $x \in X$ be given. Then $P_M x \in M$, so $AP_M x \in M$, so $P_M A P_M x = AP_M x$. Conversely, let $x \in M$ be given, and note that $Ax = AP_M x = P_M A P_M x \in M$, so M is invariant under A. □

1.4.3 Proposition. *The following are equivalent.*
(i) M reduces A;
(ii) $P_M A = A P_M$;
(iii) M is invariant under A and A^*.

PROOF: Exercise. □

The spectrum

Let $\alpha \in \mathbb{K}$ be given. We say that $\alpha \in \rho(A)$, the *resolvent set* of $A \in \mathscr{L}(X;X)$, provided that $\alpha I - A$ is bijective. Note that if $\alpha \in \rho(A)$ then $(\alpha I - A)^{-1} \in \mathscr{L}(X;X)$. The *spectrum* of A is defined to be $\sigma(A) = \mathbb{K} \setminus \rho(A)$. A number $\lambda \in \mathbb{K}$ is called an *eigenvalue* of A provided $\ker(\lambda I - A) \neq \{0\}$. Each non-zero member of $\ker(\lambda I - A)$ is called an *eigenvector*. The set of all eigenvalues of A is called the *point spectrum* of A and is denoted $\sigma_p(A)$. Clearly $\sigma_p(A) \subseteq \sigma(A)$. A number $\lambda \in \mathbb{K}$ is said be a *generalized eigenvalue* of A provided

$$\inf\{\|(\lambda I - A)x\| \mid x \in X, \|x\| = 1\} = 0.$$

In this case there is a sequence $\{x_n, n = 1, 2, \dots\}$ of unit vectors such that $(\lambda I - A)x_n \to 0$ as $n \to \infty$. It is clear that every eigenvalue is a generalized eigenvalue. Every generalized eigenvalue belongs to $\sigma(A)$.

1.4.4 Example. Let $X = \ell^2$ and $Ax = (x_1, \frac{1}{2}x_2, \frac{1}{3}x_3, \dots)$. Then A is injective, so $0 \notin \sigma_p(A)$, but $\|Ae^{(n)}\| = \frac{1}{n} \to 0$, so 0 is a generalized eigenvalue.

Spectral theory of compact operators

1.4.5 Proposition. *Assume that A is compact, $\lambda \in \sigma_p(A)$, $\lambda \neq 0$. Then $\ker(\lambda I - A)$ is finite dimensional.*

PROOF: Assume that $\ker(\lambda I - A)$ is infinite dimensional. Choose an orthonormal sequence $\{e_n, n \geq 1\}$, and choose a subsequence $\{e_{n_k}\}$ such that Ae_{n_k} converges strongly. For $j \neq k$,

$$\|Ae_{n_k} - Ae_{n_j}\|^2 = \|\lambda e_{n_k} - \lambda e_{n_j}\| = 2|\lambda|^2.$$

But this contradicts that Ae_{n_k} converges strongly. □

1.4.6 Proposition. *Assume that A is compact and $\lambda \in \mathbb{K} \setminus \{0\}$. If λ is a generalized eigenvalue of A then it is an eigenvalue of A.*

PROOF: Let $\{x_n, n \geq 1\}$ be a sequence of unit vectors such that $(\lambda I - A)x_n \to 0$. Choose a subsequence $\{x_{n_k}\}$ such that $Ax_{n_k} \to 0$. Since $\lambda \neq 0$ it follows that $x_{n_k} \to 0$, which is a contradiction unless $\lambda = 0$, which it isn't. □

1.4.7 Corollary. *Assume that A is compact and let $\lambda \in \mathbb{K} \setminus \{0\}$ be given. Assume that $\lambda \notin \sigma_p(A)$ and that $\overline{\lambda} \notin \sigma_p(A^*)$ (this second condition is redundant). Then $\lambda I - A$ is bijective (and $(\lambda I - A)^{-1}$ is bounded).*

PROOF: By 1.4.6, since $\lambda \notin \sigma_p(A)$, λ is not a generalized eigenvalue of A. Whence there is $c > 0$ such that $\|(\lambda I - A)x\| \geq c\|x\|$ for all $x \in X$. Therefore $\lambda I - A$ has closed range. It follows that

$$\operatorname{range}(\lambda I - A) = \ker(\overline{\lambda} I - A^*)^\perp = \{0\}^\perp = X,$$

so $\lambda I - A$ is surjective. It is injective because λ is not an eigenvalues, so by the corollary to the Open Mapping Theorem, $(\lambda I - A)^{-1}$ is bounded. □

Spectral theory of normal operators

1.4.8 Proposition. *Assume that A is normal and let $\lambda \in \mathbb{K}$ be given. Then $\ker(\lambda I - A) = \ker(\overline{\lambda} I - A^*)$ and $\ker(\lambda I - A)$ reduces A.*

PROOF: Clearly $\lambda I - A$ is normal, so

$$\ker(\lambda I - A) = \ker((\lambda I - A)^*) = \ker(\overline{\lambda} I - A^*).$$

To prove the second assertion, by 1.4.3, it suffices to show that $\ker(\lambda I - A)$ is invariant under both A and A^*. Let $x \in \ker(\lambda I - A)$ be given. Then $Ax = \lambda x \in \ker(\lambda I - A)$, and $A^* x = \overline{\lambda} x \in \ker(\lambda I - A)$. □

1.4.9 Proposition. *Assume A is normal and let $\lambda, \mu \in \sigma_p(A)$ be given, with $\lambda \neq \mu$. Then* $\ker(\lambda I - A) \perp \ker(\mu I - A)$.

Note that this is a generalization of the fact from basic linear algebra that the eigenspaces of a symmetric matrix are orthogonal.

PROOF: Let $x \in \ker(\lambda I - A)$ and $y \in \ker(\mu I - A)$ be given. We must show that $(x, y) = 0$. Since A is normal, $y \in \ker(\overline{\mu} I - A^*)$, and

$$\lambda(x, y) = (\lambda x, y) = (Ax, y) = (x, A^* y) = (x, \overline{\mu} y) = \mu(x, y),$$

so $(x, y) = 0$ since $\lambda \neq \mu$. □

1.4.10 Proposition. *If $A^* = A$ then $\sigma_p(A) \subseteq \mathbb{R}$.*

PROOF: Let $\lambda \in \sigma_p(A)$ and let $x \in \ker(\lambda I - A) = \ker(\overline{\lambda} I - A^*)$, with $x \neq 0$. Then $\lambda x = Ax = A^* x = \overline{\lambda} x$, so $\lambda = \overline{\lambda}$. □

Spectral theory of compact self-adjoint operators

1.4.11 Proposition. *Assume that A is compact and self-adjoint. Then one or both of $\pm \|A\|$ is an eigenvalue of A.*

PROOF: This is immediate if $A = 0$, so we may assume that $A \neq 0$. Since A is self-adjoint, by 1.2.2, $\|A\| = \sup\{|(Ax, x)| : \|x\| = 1\}$. Choose a sequence $\{x_n, n \geq 1\}$ of unit vectors such that $|(Ax_n, x_n)| \to \|A\|$. Actually, since this is a sequence of real numbers, we many choose $\{x_n\}$ so that $(Ax_n, x_n) \to \lambda$, where λ is $\|A\|$ or $-\|A\|$. Then,

$$\begin{aligned} 0 &\leq \|(A - \lambda I)x_n\|^2 \\ &= (Ax_n, Ax_n) - 2\lambda(Ax_n, x_n) + \lambda^2(x_n, x_n) \\ &\to \|A\|^2 - 2\|A\|^2 + \|A\|^2 = 0 \end{aligned}$$

Therefore λ is a generalized eigenvalue. Since A is compact and $\lambda \neq 0$, by 1.4.6, $\lambda \in \sigma_p(A)$. □

Let M and N be closed subspaces of X, with $M \perp N$. The *direct sum* of M and N is $M \oplus N = \{x + y \mid x \in M, y \in N\}$. Then $M \oplus N$ is a closed subspace of X and $(M \oplus N)^\perp = M^\perp \cap N^\perp$.

Let A be a compact self-adjoint operator on X. Let M be a closed subspace of X that reduces A. Define $\tilde{A} \in \mathscr{L}(M; M)$ by $\tilde{A} = A|_M$. Then $\tilde{A}$ is compact and self-adjoint, and $\|\tilde{A}\| \leq \|A\|$. Note that if $\lambda \in \sigma_p(A)$ then $|\lambda| \leq \|A\|$.

With all of this in mind, consider the following algorithm. Let $A_1 := A$.

- Choose $\lambda_1 \in \sigma_p(A_1)$ such that $|\lambda_1| = \|A_1\|$.
- Let $E_1 := \ker(\lambda_1 I - A)$. If $E_1 = X$ then we are done.

- Put $X_2 := E_1^{\perp}$. Note that X_2 reduces A_1 because E_1 reduces A_1.
- Put $A_2 := A|_{X_2}$, a compact self-adjoint operator on X_2.
- Choose $\lambda_2 \in \sigma_p(A_2)$ such that $|\lambda_2| = \|A_2\|$. Then $|\lambda_2| = \|A_2\| \le \|A_1\| = |\lambda_1|$ and $\lambda_2 \neq \lambda_1$.
- Put $E_2 := \ker(\lambda_2 I - A)$ and notice that $E_1 \perp E_2$. If $E_1 \oplus E_2 = X$ then stop.

In general, for $n \ge 2$,

- Put $X_{n+1} := (E_1 \oplus E_2 \oplus \cdots \oplus E_n)^{\perp}$. Note that X_{n+1} reduces A.
- Put $A_{n+1} := A|_{X_{n+1}} = A_n|_{X_{n+1}}$.
- Choose $\lambda_{n+1} \in \sigma_p(A_{n+1}) \subseteq \sigma_p(A)$ such that $|\lambda_{n+1}| = \|A_{n+1}\| \le |\lambda_n|$
- Put $E_{n+1} := \ker(\lambda_{n+1} I - A)$ and notice that $E_{n+1} \perp (E_1 \oplus \cdots \oplus E_n)$. If $E_1 \oplus \cdots \oplus E_{n+1} = X$ then stop.

If this process terminates in finitely many steps N, then $E_1 \oplus \cdots \oplus E_N = X$. Note that $\dim(E_n) < \infty$ for each $n < N$. If $\dim(E_N)$ is infinite then $\lambda_N = 0$, and otherwise $\dim(X) < \infty$. Assume that $\dim \operatorname{range}(A)$ is infinite. By induction, we get a sequence $\{\lambda_n\}_{n=1}^{\infty}$ of distinct eigenvalues of A with $|\lambda_1| \ge |\lambda_2| \ge \cdots$ and a sequence of orthogonal subspaces $\{E_n\}_{n=1}^{\infty}$ defined by $E_n = \ker(\lambda_n I - A)$. Moreover, $|\lambda_{n+1}| = \|A|_{(E_1 \oplus \cdots \oplus E_n)^{\perp}}\|$.

Let $\alpha := \lim_{n\to\infty} |\lambda_n|$. We claim that $\alpha = 0$. To see why, for each $n \ge 1$, choose $x_n \in E_n$ with $\|x_n\| = 1$. Note $(x_n, x_m) = \delta_{n,m}$. Since A is compact, $\{Ax_n\}_{n=1}^{\infty}$ has a convergent subsequence $\{Ax_{n_k}\}_{k=1}^{\infty}$. Now $Ax_{n_k} = \lambda_{n_k} x_{n_k}$, so

$$\|Ax_{n_k} - Ax_{n_j}\|^2 = \|\lambda_{n_k} x_{n_k} - \lambda_{n_j} x_{n_j}\|^2 = \lambda_{n_k}^2 + \lambda_{n_j}^2 \ge 2\alpha^2$$

Since convergent sequences have the Cauchy property, $\alpha = 0$.

Put $P_n := P_{E_n}$. We will show that $\|A - \sum_{j=1}^{\infty} \lambda_j P_j\| \to 0$ as $n \to \infty$. Indeed, let $n \in \mathbb{N}$ and $x \in X$, with $\|x\| = 1$, be given. Then $x = x_1 + \cdots + x_n + x_{\perp}$, where $x_i \in E_i$ and $x_{\perp} \in (E_1 \oplus \cdots \oplus E_n)^{\perp}$. Since $Ax_i = \lambda_i x_i$, it follows that

$$\left\|(A - \sum_{j=1}^{\infty} \lambda_j P_j)x\right\| = \|Ax_{\perp}\| = \|A|_{(E_1 \oplus \cdots \oplus E_n)^{\perp}} x_{\perp}\| \le |\lambda_{n+1}|.$$

Therefore $\|A - \sum_{j=1}^{\infty} \lambda_j P_j\| \to 0$ as $n \to \infty$.

Now we will show that $\{\lambda_i \mid i \in \mathbb{N}\} = \sigma_p(A) \setminus \{0\}$. Indeed, let $\mu \in \sigma_p(A) \setminus \{0\}$ and choose $x \in X$ such that $\|x\| = 1$ and $Ax = \lambda x$. If $\mu \notin \{\lambda_i \mid i \in \mathbb{N}\}$ then $(P_j x, x) = 0$ for all $j \in \mathbb{N}$. But this contradicts the note above, because $\mu x = Ax = \sum_{j=1}^{\infty} \lambda_j P_j x = 0$ but $x \neq 0$ and $\mu \neq 0$.

We have proven the following decomposition theorem.

1.4.12 Theorem. *Assume that A is a compact, self-adjoint operator on X.*

(i) $\sigma(A) \setminus \{0\} \subseteq \sigma_p(A)$.
(ii) *There is $\lambda \in \sigma_p(A)$ such that $|\lambda| = \|A\|$.*
(iii) *$\sigma_p(A)$ is countable and zero is the only possible accumulation point.*
(iv) *A has finite rank if and only if $\sigma_p(A)$ is a finite set.*
(v) *For each $\lambda \in \mathbb{K} \setminus \{0\}$, $\ker(\lambda I - A)$ is finite dimensional.*
(vi) *For all $\lambda, \mu \in \mathbb{K}$ with $\lambda \neq \mu$, $\ker(\lambda I - A) \perp \ker(\mu I - A)$.*
(vii) *There is an orthonormal basis $\{e_i \mid i \in I\}$ for X of eigenvectors of A. Moreover, for any such basis, $Ax = \sum_{i \in I} \lambda_i (x, e_i) e_i$, where $Ae_i = \lambda_i e_i$.*

Spectral theory of compact normal operators

Let A be a compact, normal operator on X. Assume for this section that $\mathbb{K} = \mathbb{C}$. (On $\mathbb{R}$ A need not have an eigenvalue.) Write

$$A = \tfrac{1}{2}(A + A^*) + i\tfrac{1}{2i}(A - A^*) =: B + iC,$$

and note that B and C are compact, self-adjoint, and $BC = CB$. We claim further that B and C are simultaneously diagonalizable.

1.4.13 Theorem. *Assume that A compact and normal and $\mathbb{K} = \mathbb{C}$. Then $\sigma_p(A)$ is non-empty, countable, and zero is the only possible accumulation point. Further, there exists an orthonormal basis $\{e_i \mid i \in I\}$ for X of eigenvectors of A, and for any such basis $Ax = \sum_{i\in I} \lambda_i(x, e_i)e_i$, where $Ae_i = \lambda_i e_i$.*

There are further spectral theories for unitary operators, bounded self-adjoint operators, unbounded self-adjoint operators, etc., but proper treatment of these theories would take the rest of the course.

2 Spectral Theory in Banach Spaces

2.1 The spectrum and resolvent set

Let X be a Banach space and $T \in \mathscr{L}(X;X)$. The *resolvent set* of T, denoted $\rho(T)$, is the set of all $\lambda \in \mathbb{K}$ such that $\lambda I - T$ is bijective. The *spectrum* of T, denoted $\sigma(T)$, is the complement of the resolvent set, $\sigma(T) := \mathbb{K} \setminus \rho(T)$. Eigenvectors and generalized eigenvectors are as before.

2.1.1 Proposition. *Let $\lambda_1, \dots, \lambda_n$ be distinct eigenvalues of T and $x_1, \dots, x_n$ be the associated eigenvectors. Then $\{x_1, \dots, x_n\}$ is linearly independent.*

2.1.2 Proposition. *Assume $\|T\| < 1$. Then $1 \in \rho(T)$ and the series $\sum_{k=0}^{\infty} T^k$ converges in the operator norm to $(I - T)^{-1}$.*

PROOF: Notice that $\|T^k\| \le \|T\|^k$ for all $k \in \mathbb{N}$, so $\sum_{k=0}^{\infty} \|T^k\|$ is convergent. Since X is complete, $\mathscr{L}(X;X)$ is complete, and $\sum_{k=0}^{\infty} T^k$ converges in operator norm.

Let $n \in \mathbb{N}$ be given and put $S_n = \sum_{k=0}^{n} T^k$. Notice that

$$(I - T)S_n = I - T^{n+1} = S_n(I - T).$$

As $n \to \infty$, $\|T^{n+1}\| \le \|T\|^n \to 0$, and

$$(I - T)\sum_{k=0}^{\infty} T^k = I = \left(\sum_{k=0}^{\infty} T^k\right)(I - T).$$

□

2.1.3 Corollary. *If $\lambda \in \mathbb{K}$ is such that $|\lambda| > \|T\|$ then $\lambda \in \rho(T)$.*

For all $\mu \in \rho(T)$, let $R(\mu; T) := (\mu I - T)^{-1}$, the *resolvent function* of T at μ. Let $\lambda_0 \in \rho(T)$ be given. Let $\lambda \in \mathbb{K}$ be such that $|\lambda - \lambda_0| \|R(\lambda_0; T)\| < 1$.

$$\begin{aligned} \lambda I - T &= (\lambda_0 I - T) + (\lambda - \lambda_0)I \\ &= (\lambda_0 I - T)(I + (\lambda - \lambda_0)(\lambda_0 I - T)^{-1}) \\ &= (\lambda_0 I - T)(I - (\lambda_0 - \lambda)R(\lambda_0; T)) \end{aligned}$$

$$\text{so } (\lambda I - T)\sum_{k=0}^{\infty}(\lambda_0 - \lambda)^k R(\lambda_0; T)^k = \lambda_0 I - T$$

$$\text{and } \sum_{k=0}^{\infty}(\lambda_0 - \lambda)^k R(\lambda_0; T)^k (\lambda I - T) = \lambda_0 I - T$$

Therefore $\lambda \in \rho(T)$ and $R(\lambda; T) = \sum_{k=0}^{\infty}(\lambda_0 - \lambda)^k R(\lambda_0; T)^{k+1}$.

2.1.4 Theorem. *$\rho(T)$ is open and $\sigma(T)$ is closed.*

2.1.5 Corollary. *When $|\lambda - \lambda_0| \|R(\lambda_0; T)\| < 1$,*

$$R(\lambda; T) = \sum_{k=0}^{\infty}(\lambda_0 - \lambda)^k R(\lambda_0; T)^{k+1}.$$

The mapping $\lambda \mapsto R(\lambda; T)$ is analytic on $\mathbb{K}$.

2.1.6 Theorem. *Let $\lambda, \mu \in \rho(T)$ and $s \in \mathscr{L}(X; X)$ be given.*

(i) $R(\lambda; T) - R(\mu; T) = (\mu - \lambda)R(\lambda; T)R(\mu; T)$.

(ii) *If $TS = ST$ then $SR(\lambda; T) = R(\lambda; T)S$.*

(iii) $R(\lambda; T)R(\mu; T) = R(\mu; T)R(\lambda; T)$.

PROOF:

(i) We employ a standard trick.

$$\begin{aligned} &R(\lambda; T) - R(\mu; T) \\ &= R(\lambda; T)(\mu I - T)R(\mu; T) - R(\lambda; T)(\lambda I - T)R(\mu; T) \\ &= R(\lambda; T)((\mu - \lambda)I)R(\mu; T) \\ &= (\mu - \lambda)R(\lambda; T)R(\mu; T) \end{aligned}$$

(ii) Note that $S(\lambda I - T) = (\lambda I - T)S$ since everything commutes with I. Multiply on the right and left by $R(\lambda; T)$.

(iii) Follows from either of the first two parts. □

2.1.7 Theorem. *If $\mathbb{K} = \mathbb{C}$ then $\sigma(T) \neq \varnothing$.*

Proof: The mapping $\lambda \mapsto R(\lambda; T)$ is analytic on $\rho(T)$. Suppose that $\sigma(T) = \varnothing$, so that $\rho(T) = \mathbb{C}$. Let $D := \{\lambda \in \mathbb{C} : |\lambda| \leq 2\|T\|\}$, which is non-empty and compact. Let $M := \max\{\|R(\lambda; T)\| : \lambda \in D\}$. For $\lambda \in \mathbb{C} \setminus D$,

$$R(\lambda; T) = (\lambda I - T)^{-1} = \tfrac{1}{\lambda}(I - \tfrac{T}{\lambda})^{-1},$$

and $\|\frac{T}{\lambda}\| \leq \frac{1}{2}$. Therefore

$$\|R(\lambda; T)\| = \left\| \frac{1}{\lambda} \sum_{n=0}^{\infty} \left(\frac{T}{|\lambda|} \right)^n \right\| \leq \frac{1}{\lambda} \sum_{n=0}^{\infty} \left\| \frac{T}{\lambda} \right\|^n \leq \|T\|.$$

Let $x \in X$ and $x^* \in X^*$ be given. Define $f : \mathbb{C} \to \mathbb{C}$ by $f(\lambda) := x^* R(\lambda; T)x$, so that $|f(\lambda)| \leq \|x^*\|\|x\| \max\{M, \|T\|\}$. But by assumption, f is entire. Since it is bounded, by Liouville's Theorem, f is constant. Since this holds for all x and x^*, this is a contradiction. □

Note that 2.1.7 might not hold if $\mathbb{K} = \mathbb{R}$, and might not hold if T is an unbounded operator.

2.1.8 Theorem (Spectral Mapping). *Assume that $\mathbb{K} = \mathbb{C}$. Let $p : \mathbb{C} \to \mathbb{C}$ be a non-constant polynomial. Then $\sigma(p(T)) = p[\sigma(T)]$.*

Proof: Later. □

The *spectral radius* of T, defined when $\sigma(T) \neq \varnothing$, is defined to be

$$r_\sigma(T) := \max\{|\lambda| : \lambda \in \sigma(T)\}.$$

Note that $0 \leq r_\sigma(T) \leq \|T\|$.

2.1.9 Theorem. *If $\mathbb{K} = \mathbb{C}$ then $r_\sigma(T) = \lim_{n\to\infty} \sqrt[n]{\|T^n\|}$ (and this limit exists).*

Proof: By 2.1.8, $r_\sigma(T^n) = (r_\sigma(T))^n$. Further, $r_\sigma(T^n) \leq \|T^n\|$, so

$$r_\sigma(T) = \sqrt[n]{r_\sigma(T^n)} \leq \sqrt[n]{\|T^n\|}.$$

Therefore $r_\sigma(T) \leq \liminf_n \sqrt[n]{\|T^n\|}$. For the other direction, consider the following.

$$R(\lambda; T) = \frac{1}{\lambda} \sum_{n=0}^{\infty} \left(\frac{T}{\lambda} \right)^n =: z \sum_{n=0}^{\infty} z^n T^n.$$

The function of z on the right is analytic on a disc centred at the origin. Put $a_n := \|T^n\|$. Then $\|\sum_{n=0}^{\infty} z^n T^n\| \leq \sum_{n=0}^{\infty} a_n |z|^n$. The radius of convergence r of this real power series satisfies $\frac{1}{r} = \limsup_n \sqrt[n]{a_n} = \limsup_n \sqrt[n]{\|T^n\|}$. Now, $\frac{1}{r}$ is the radius of the smallest disc centred at the origin whose exterior lies in $\rho(T)$. It follows that $r_\sigma(T) = \frac{1}{r} = \limsup_n \sqrt[n]{\|T^n\|}$. □

2.1.10 Corollary. *Assume that X is a complex Hilbert space and A is normal. Then $\|A\| = r_\sigma(A)$.*

Proof: It suffices to show that $\|A\|^2 = \|A^2\|$. Observe that

$$\begin{aligned}\|A\|^2 &= \sup\{|(Ax, Ax)| : \|x\| = 1\}\\ &= \sup\{|(A^*Ax, x)| : \|x\| = 1\}\\ &= \|A^*A\| && \text{since } A^*A \text{ is self-adjoint}\\ &= \|A^2\| && \text{since } A \text{ is normal}\end{aligned}$$

□

2.2 Spectral theory of compact operators

2.2.1 Theorem. *If T is a compact operator then $\sigma_p(T)$ is countable and 0 is the only possible accumulation point.*

Proof: For each $\varepsilon > 0$ put $\Lambda_\varepsilon := \{\lambda \in \sigma_p(T) : |\lambda| \geq \varepsilon\}$. We will show that Λ_ε is a finite set for every $\varepsilon > 0$.

Let $\varepsilon > 0$ be given and suppose that Λ_ε is infinite. Choose an injective sequence $\{\lambda_n\}_{n=1}^\infty$ in Λ_ε and choose a sequence $\{x_n\}_{n=1}^\infty$ of corresponding eigenvectors. For each n, put $M_n = \operatorname{span}\{x_1, \dots, x_n\}$. Notice that $\{M_n\}_{n=1}^\infty$ is increasing, and that $T[M_n] = M_n$ for each n.

Let $n \in \mathbb{N}$ and $x \in M_n$. Then $x = \alpha_1 x_1 + \cdots + \alpha_n x_n$ for some $\alpha_i \in \mathbb{K}$.

$$(\lambda_n I - T)x = \alpha_1(\lambda_n - \lambda_1)x_1 + \cdots + \alpha_{n-1}(\lambda_n - \lambda_{n-1})x_{n-1} + 0.$$

It follows that $(\lambda_n I - T)[M_n] = M_{n-1}$. Note finally that each M_n is a *closed* subspace of X. By the Riesz Lemma (from last semester) we can choose a sequence $\{y_n\}_{n=1}^\infty$ such that, for all $n \in \mathbb{N}$, $y_n \in M_n$, $\|y_n\| = 1$, and $\|y_n - x\| \geq \frac{1}{2}$ for all $x \in M_{n-1}$.

Let $m, n \in \mathbb{N}$ with $m < n$ be given.

$$Ty_n - Ty_m = \lambda_n y_n - (\lambda_n I - T)y_n - Ty_m =: \lambda_n(y_n - \tfrac{1}{\lambda_n}x),$$

and $x \in M_{n-1}$ since $Ty_m \in M_m \subseteq M_{n-1}$ and $(\lambda_n I - T)y_n \in M_{n-1}$. It follows from the properties of the sequence that $\|Ty_n - Ty_m\| \geq \frac{1}{2}|\lambda_n| \geq \frac{1}{2}\varepsilon$. Therefore the sequence $\{Ty_n\}_{n=1}^\infty$ has no convergent subsequences, which is a contradiction since T is compact. □

2.2.2 Proposition. *Assume that T is compact and let $\lambda \in \mathbb{K} \setminus \{0\}$ be given. Then $\ker(\lambda I - T)$ is finite dimensional.*

Proof: By 2.2.1 there are finitely many linearly independent eigenvectors of T associated with any eigenvalue of T. Therefore either $\ker(\lambda I - T) = \{0\}$ or $\ker(\lambda I - T)$ is finite dimensional. □

2.2.3 Theorem. *Assume that T is compact and let $\lambda \in \mathbb{K} \setminus \{0\}$ be given. Then* $\operatorname{range}(\lambda I - T)$ *is closed. It follows that* $\operatorname{range}(\lambda I - T) = {}^{\perp}\ker(\lambda I - T^*)$*, where T^* is the Banach space adjoint of T.*

PROOF: Suppose for contradiction that $\operatorname{range}(\lambda I - T)$ is not closed. Choose a sequence $\{x_n\}_{n=1}^{\infty}$ such that $y_n := (\lambda I - T)x_n \to y \notin \operatorname{range}(\lambda I - T)$.

Note that $y \neq 0$, so without loss of generality we may assume that $y_n \neq 0$ for all $n \in \mathbb{N}$. Hence $x_n \notin \ker(\lambda I - T)$ for all $n \in \mathbb{N}$. For each $n \in \mathbb{N}$, put

$$\delta_n = \inf\{\|x_n - z\| : z \in \ker(\lambda I - T)\} > 0,$$

and choose $z_n \in \ker(\lambda I - T)$ such that $a_n := \|x_n - z_n\| < 2\delta_n$.

We claim that $a_n \to \infty$ as $n \to \infty$. If not then $\{x_n - z_n\}_{n=1}^{\infty}$ would have a bounded subsequence $\{x_{n_k} - z_{n_k}\}_{k=1}^{\infty}$. But

$$x_{n_k} - z_{n_k} = \frac{1}{\lambda}((\lambda I - T) + T)(x_{n_k} - z_{n_k}) = \frac{1}{\lambda} y_{n_k} + \frac{1}{\lambda} T(x_{n_k} - z_{n_k})$$

so $\{x_{n_k} - z_{n_k}\}_{k=1}^{\infty}$ has a further subsequence that converges strongly. Let v denote the limit of that subsequence, and note that $(\lambda I - T)v = y$, which is a contradiction since $y \notin \operatorname{range}(\lambda I - T)$.

Put $w_n = \frac{1}{a_n}(x_n - z_n)$. Notice that $\|w_n\| = 1$ and

$$(\lambda I - T)w_n = \frac{1}{a_n}(\lambda I - T)x_n \to 0$$

as $n \to \infty$. Choose a subsequence $\{w_{n_k}\}_{k=1}^{\infty}$ such that $\{Tw_{n_k}\}_{k=1}^{\infty}$ is strongly convergent. We concluded that $\{w_{n_k}\}_{k=1}^{\infty}$ is strongly convergent. Put $w := \lim_{k\to\infty} w_{n_k}$, and note that $(\lambda I - T)w = 0$. Put $u_n = z_n + a_n w \in \ker(\lambda I - T)$. By definition of δ_n,

$$\delta_n \leq \|x_n - u_n\| = \|x_n - z_n - a_n w\| = a_n \|w_n - w\| \leq 2\delta_n \|w_n - w\|,$$

so $\frac{1}{2} \leq \|w_n - w\|$ for all n. This is a contradiction. □

PROOF (ALTERNATE): Let $\{y_n\}_{n=1}^{\infty}$ be a convergent sequence in $\operatorname{range}(\lambda I - T)$. Put $y = \lim_{n\to\infty} y_n$. We hope to show that $y \in \operatorname{range}(\lambda I - T)$. Let x_n be such that $(\lambda I - T)x_n = y_n$. If $\{x_n\}_{n=1}^{\infty}$ were bounded then there would be a subsequence such that $\{Tx_{n_k}\}_{k=1}^{\infty}$ converges, and hence also that $\{x_{n_k}\}_{k=1}^{\infty}$ converges. We claim that there is a bounded sequence $\{z_n\}_{n=1}^{\infty}$ such that $(\lambda I - T)z_n = y_n$. □

2.2.4 Theorem. *Assume that T is compact and let $\lambda \in \mathbb{K} \setminus \{0\}$ be given. Then the following hold.*

(i) $\operatorname{range}(\lambda I - T) = {}^{\perp}\ker(\lambda I - T^*)$*; and*
(ii) $\operatorname{range}(\lambda I - T^*) = \ker(\lambda I - T)^{\perp}$*.*

2.2.5 Corollary. *Assume that T is compact and let $\lambda \in \mathbb{K} \setminus \{0\}$ be given. If $\lambda \in \sigma(T)$ then either $\lambda \in \sigma_p(T)$ or $\lambda \in \sigma_p(T^*)$ (or both).*

In fact, it is also true that $\dim\ker(\lambda I - T) = \dim\ker(\lambda I - T^*)$, so the "both" in the corollary always holds.

Assume that T is compact and let $\lambda \in \mathbb{K}\setminus\{0\}$ and $n \in \mathbb{N}$ be given.

$$(\lambda I - T)^n = \lambda^n I + \underbrace{\sum_{k=1}^{n} \binom{n}{k} (-1)^k \lambda^{n-k} T^k}_{\text{compact}} =: \mu I - L.$$

It follows that $\ker((\lambda I - T)^n)$ is finite dimensional and $\operatorname{range}((\lambda I - T)^n)$ is closed.

2.2.6 Lemma. *Assume that T is compact and let $\lambda \in \mathbb{K}\setminus\{0\}$ be given. There is a smallest non-negative integer r such that $\ker((\lambda I - T)^n) = \ker((\lambda I - T)^{n+1})$ for all $n \geq r$. Moreover, if $r > 0$ then the inclusions below are strict.*

$$\ker((\lambda I - T)^0) \subset \ker((\lambda I - T)^1) \subset \cdots \subset \ker((\lambda I - T)^r)$$

Proof: Put $K_m := \ker((\lambda I - T)^m)$. Suppose that $K_n \subsetneq K_{n+1}$ for all $n \geq 0$. By the Riesz Lemma there is a sequence $\{y_n\}_{n=1}^\infty$ such that $y_n \in K_{n+1} \setminus K_n$, $\|y_n\| = 1$, and $\operatorname{dist}(y_n, K_n) \geq \frac{1}{2}$ for all $n \geq 1$. Let $m < n$ be given and consider

$$(\lambda I - T)^n((\lambda I - T)y_n + Ty_m) = (\lambda I - T)^{n+1} y_n + T(\lambda I - T)^n y_m = 0,$$

so $(\lambda I - T)y_n + Ty_m \in K_n$. It follows that

$$\|Ty_n - Ty_m\| = \|\lambda y_n - ((\lambda I - T)y_n + Ty_m)\| \geq \frac{|\lambda|}{2}.$$

This is a contradiction because it implies $\{Ty_n\}_{n=1}^\infty$ has no convergent subsequences.

Let $n \geq 0$ be such that $K_n = K_{n+1}$. We will show $K_{n+1} = K_{n+2}$, proving the last statement of the lemma. Let $x \in K_{n+2}$ be given.

$$\begin{aligned} 0 = (\lambda I - T)^{n+2} x &= (\lambda I - T)^{n+1}(\lambda I - T)x \\ &= (\lambda I - T)^n(\lambda I - T)x && K_n = K_{n+1} \\ &= (\lambda I - T)^{n+1} x. \end{aligned}$$

□

2.2.7 Lemma. *Let $\lambda \in \mathbb{K}\setminus\{0\}$ be given and assume that T is compact. Then there is a smallest non-negative integer q such that $\operatorname{range}((\lambda I - T)^n) = \operatorname{range}((\lambda I - T)^q)$ for all $n \geq q$. Moreover, if $q > 0$ then the inclusions below are strict.*

$$\operatorname{range}((\lambda I - T)^0) \supset \operatorname{range}((\lambda I - T)^1) \supset \cdots \supset \operatorname{range}((\lambda I - T)^q).$$

2.2.8 Lemma. *Let $\lambda \in K \setminus \{0\}$ be given and assume T is compact. Then the number r from 2.2.6 is the same as q from 2.2.7.*

PROOF: We will show that $q \geq r$. The other inequality is easier. Let $K_n = \ker((\lambda I - T)^n)$ and $R_n = \text{range}((\lambda I - T)^n)$. Then $R_q = R_{q+1}$, so $(\lambda I - T)[R_q] = R_q$. Given $y \in R_q$ there is $x \in R_q$ such that $(\lambda I - T)x = y$.

We claim that $(\lambda I - T)|_{R_q}$ is injective. Suppose not. Choose $x_1 \in R_q \setminus \{0\}$ such that $(\lambda I - T)x_1 = 0$, $x_1 \in R_2 \setminus \{0\}$ such that $(\lambda I - T)x_2 = x_1$, etc. We get an infinite sequence $\{x_n\}_{n=1}^\infty$ such that $0 \neq x_1 = (\lambda I - T)^{n-1}x_n$ and $(\lambda I - T)^n x_n = (\lambda I - T)x_1 = 0$. So $x_n \in K_n \setminus K_{n-1}$, which contradicts 2.2.6.

Next we claim that $K_{q+1} = K_q$, proving the inequality. We already know $K_q \subseteq K_{q+1}$. Suppose the other inclusion does not hold, and choose $x_0 \in K_{q+1} \setminus K_q$. Put $y := (\lambda I - T)^q x \in R_q$, and note that $y \neq 0$. But $(\lambda I - T)y = 0$, contradicting the first claim. □

2.2.9 Theorem. *Let $\lambda \in \mathbb{K} \setminus \{0\}$ be given and assume that T is compact. If $\lambda \in \sigma(T)$ then $\lambda \in \sigma_p(T)$.*

PROOF: If $\lambda \notin \sigma_p(T)$ then $\ker(\lambda I - T) = \{0\}$, so $r = 0$. But then $q = 0$, so $\lambda I - T$ is surjective. Whence $\lambda I - T$ is bijective and $\lambda \notin \sigma(T)$. □

2.2.10 Theorem. *Let $\lambda \in \mathbb{K} \setminus \{0\}$ be given and assume that T is compact. Let r be as in 2.2.6. Then $X = \ker((\lambda I - T)^r) \oplus \text{range}((\lambda I - T^r)$.*

PROOF: Write K_r and R_r as before. Let $x \in X$ be given. Notice that $R_{2r} = R_r$. Choose $x_1 \in X$ such that $(\lambda I - T)^{2r}x_1 = (\lambda I - T)^r x$ and put $x_0 = (\lambda I - T)^r x_1 \in R_r$. Notice that $(\lambda I - T)^r x_0 = (\lambda I - T)^r x$. It follows that $x - x_0 \in K_r$, so $x = x_0 + (x - x_0)$.

Suppose that $x = \tilde{x}_0 + (x - \tilde{x}_0)$ is another decomposition. Put $v_0 = x_0 - \tilde{x}_0 \in R_r$. Choose $v \in X$ such that $v_0 = (\lambda I - T)^r v$. Note that $v_0 = (x - \tilde{x}_0) - (x - x_0) \in K_r$, so $(\lambda I - T)^{2r} v = (\lambda I - T)^r v_0 = 0$. But $v \in K_{2r} = K_r$, so $0 = (\lambda I - T)^r v = v_0$. □

2.2.11 Theorem. *Let $\lambda \in \mathbb{K} \setminus \{0\}$ be given and assume that T is compact. Then* $\dim \ker(\lambda I - T) = \dim \ker(\lambda I - T^*)$.

2.2.4 together with 2.2.11 is often known as the *Fredholm alternative*.

PROOF (IDEA): Let $\{x_1, \dots, x_n\}$ and $\{y_1^*, \dots, y_m^*\}$ be bases for $\ker(\lambda I - T)$ and $\ker(\lambda I - T^*)$, respectively. Choose a dual bases $\{x_1^*, \dots, x_n^*\}$ and $\{y_1, \dots, y_m\}$ in X^* and X.

If $n < m$ define $Sx = Tx + \sum_{i=1}^n x_i^*(x) y_i$. S is compact. It can be shown that $\ker(\lambda I - S) = \{0\}$, so $\lambda I - S$ is surjective, which is a contradiction. □

3 General Linear Operators

3.1 Introduction

Let X and Y be Banach spaces. Let $\mathscr{D}(A) \subseteq X$. We say that $A : \mathscr{D}(A) \to Y$ is *linear* if $\mathscr{D}(A)$ is a linear subspace of X (not necessarily closed) and A is linear map between these vector spaces. At this point there is no concept of norm or continuity. We say that A is *bounded* if there is $K \in \mathbb{R}$ such that $\|Ax\|_Y \leq K\|x\|_X$ for all $x \in \mathscr{D}(A)$. $\mathscr{L}(X;Y)$ is the set of all bounded linear function $X \to Y$ whose domain is all of X. If A is not bounded then we say that A is *unbounded*. We say that A is *closed* provided that $\mathrm{Gr}(A) = \{(x, Ax) \mid x \in \mathscr{D}(A)\}$ is closed in $X \times Y$. Recall the following theorem.

3.1.1 Theorem (Closed Graph). *Let X and Y be Banach spaces, $\mathscr{D}(A) \subseteq X$, and assume that $A : \mathscr{D}(A) \to Y$ is linear. If A is closed and $\mathscr{D}(A)$ is closed in X then A is bounded.*

3.1.2 Lemma. *With the notation of the closed graph theorem, A is closed if and only if for every $x \in X$ and $y \in Y$ and every sequence $\{x_n\}_{n=1}^\infty$ in $\mathscr{D}(A)$ such that $x_n \to x$ and $Ax_n \to y$ we have $x \in \mathscr{D}(A)$ and $Ax = y$.*

3.1.3 Examples.

(i) Let $X = Y = C[0,1]$, $\mathscr{D}(A) = C^1[0,1]$, and $Af = f'$ for all $f \in \mathscr{D}(A)$. Then A is closed and unbounded. Indeed, if $f_n \to f$ uniformly and $f_n' \to g$ uniformly then $f \in C^1[0,1]$ and $f' = g$, so we concluded with 3.1.2.

(ii) Let $X = Y = C[0,1]$, $\mathscr{D}(B) = C^2[0,1]$, and $Bf = f'$ for all $f \in \mathscr{D}(B)$. Then B is unbounded and not closed. Clearly B is not closed because of poor choice of domain.

3.1.4 Lemma. *Let $\mathscr{D}(A) \subseteq X$ and assume that $A : \mathscr{D}(A) \to Y$ is linear and bounded. If $\mathscr{D}(A)$ is closed then A is closed.*

PROOF: Let $\{x_n\}_{n=1}^\infty$ in $\mathscr{D}(A)$ be such that $x_n \to x \in X$ and $Ax_n \to y \in Y$. If $\mathscr{D}(A)$ is closed then $x \in \mathscr{D}(A)$. Since A is bounded,

$$\|Ax - Ax_n\|_Y \leq K\|x - x_n\|_X \to 0 \text{ as } n \to \infty. \qquad \square$$

3.1.5 Lemma. *Let $\mathscr{D}(A) \subseteq X$ and assume that $A : \mathscr{D}(A) \to Y$ is linear and closed. Then $\ker(A) := \{x \in \mathscr{D}(A) \mid Ax = 0\}$ is closed in X.*

PROOF: Let $\{x_n\}_{n=1}^\infty$ in $\ker(A)$ be such that $x_n \to x \in X$. Then $Ax_n = 0$ for each n, so in particular $Ax_n \to 0$ in Y. Since A is closed, $x \in \mathscr{D}(A)$ and $Ax = 0$, so $x \in \ker(A)$. $\square$

3.1.6 Lemma. *Let $\mathscr{D}(A) \subseteq X$ and assume that $A : \mathscr{D}(A) \to Y$ is linear, closed, and injective. Then $A^{-1} : \mathrm{range}(A) \to X$ is closed.*

PROOF: Let $\{y_n\}_{n=1}^\infty$ in $\mathrm{range}(A)$ be such that $y_n \to y \in Y$ and $A^{-1}y_n \to x \in X$. Then $A^{-1}y_n \in \mathscr{D}(A)$ and $AA^{-1}y_n \to y$, so since A is closed, $x \in \mathscr{D}(A)$ and $Ax = y$, i.e. $A^{-1}y = x$. $\square$

A is said to be *closable* provided there is a linear mapping $\tilde{A} : \mathscr{D}(\tilde{A}) \to Y$ such that $\mathrm{Gr}(A) \subseteq \mathrm{Gr}(\tilde{A})$, i.e. $\tilde{A}$ is a closed linear extension of A. In this case there is a minimal closed extension of A. The minimal closed extension is called the *closure* of A.

3.2 Adjoints

Let $A : \mathscr{D}(A) \to Y$ be linear. We want to find a linear $A^* : \mathscr{D}(A^*) \to X^*$ with $\mathscr{D}(A^*) \subseteq Y^*$ such that

$$\langle y^*, Ax \rangle = \langle A^* y^*, x \rangle$$

for all $x \in \mathscr{D}(A)$ and $y^* \in \mathscr{D}(A^*)$. For this to be reasonable, we require that $\mathscr{D}(A)$ is dense in A, for otherwise A^* would not be uniquely determined by the formula. In case that $\mathscr{D}(A)$ is dense in X, define

$$\mathscr{D}(A^*) := \{y^* \in A^* \mid \exists! z^* \in X^* \text{ such that } \langle y^*, Ax \rangle = \langle z^*, x \rangle \text{ for all } x \in \mathscr{D}(A)\},$$

and $A^* y^* = z^*$ for $y^* \in \mathscr{D}(A^*)$.

Warning: Even if $\mathscr{D}(A)$ is dense, it can happen that $\mathscr{D}(A^*) = \{0\}$.

Remark. If X is a Hilbert space and $A : \mathscr{D}(A) \subseteq X \to X$, the *Hilbert space adjoint* A_H^* is defined in the way one would expect. A is said to be *self-adjoint* provided $A_H^* = A$ and the domains are equal.

3.2.1 Example. Choose $p \in [1, \infty)$ and set $X = Y = \ell^p$. Then $X^* = Y^* = \ell^q$, where $\frac{1}{p} + \frac{1}{q} = 1$. Let $\mathscr{D}(A) := \mathbb{K}^{(\mathbb{N})}$ and define $Ax = (\sum_{n=1}^{\infty} n x_n, x_2, x_3, \dots)$.

Note that $\mathscr{D}(A)$ is dense in ℓ^p. We would like to find the adjoint. For $y \in \mathscr{D}(A^*)$ (to be determined) we want $z \in \ell^q$ such that $\langle y, Ax \rangle = \langle z, x \rangle$ for all $x \in \mathbb{K}^{(\mathbb{N})}$. The identity

$$y_1 \sum_{n=1}^{\infty} n x_n + \sum_{k=2}^{\infty} y_k x_k = z_1 x_1 + \sum_{k=2}^{\infty} z_k x_k$$

implies first that $y_1 = 0$. Once we have seen this, it follows that $z_k = y_k$ for all $k \geq 2$, so $\mathscr{D}(A^*) = \{y \in \ell^q \mid y_1 = 0\}$ and $A^* y = y$. Note that $\mathscr{D}(A^*)$ is not dense in ℓ^q.

3.2.2 Example. Let $\mathbb{K} = \mathbb{R}$ and take $X = Y = L^2[0,1]$. Define

$$\mathscr{D}(A) := \{f \in AC[0,1] : f' \in L^2[0,1], f(0) = f(1) = 0\},$$

and note that $\mathscr{D}(A)$ is dense in L^2. Define $Af = f'$ for $f \in \mathscr{D}(A)$. What is A^*?

For $g \in \mathscr{D}(A^*)$ (to be determined) we want $h \in L^2[0,1]$ such that for all $f \in \mathscr{D}(A)$, $\int_0^1 f' g \, dx = \int_0^1 f h \, dx$. If $g \in \mathscr{D}(A)$ then $\int_0^1 f' g \, dx = -\int_0^1 f g' \, dx$, integrating by parts. Therefore the required h is $-g'$ in this case. It follows that $\mathscr{D}(A) \subseteq \mathscr{D}(A^*)$

and $A^*g = -g'$ for all $g \in \mathscr{D}(A)$. Is $\mathscr{D}(A^*) = \mathscr{D}(A)$? Not quite: the elements of $\mathscr{D}(A^*)$ do not need to vanish at zero and one, i.e.

$$\mathscr{D}(A^*) = \{f \in AC[0,1] : f' \in L^2[0,1]\}.$$

It can be shown that $A^{**} = A$ (with equal domains).

We can extend the above example as follows. Define

$$\mathscr{D}(\Delta) := \{f \in AC[0,1] : f' \in AC[0,1], f'' \in L^2[0,1], f(0) = f(1) = 0\},$$

and $\Delta f = f''$. Then $\mathscr{D}(\Delta^*) = \mathscr{D}(\Delta)$ and $\Delta^* = \Delta$. We could also replace the conditions on the functions in $\mathscr{D}(\Delta)$ at the endpoints with conditions on the derivatives at the endpoints, etc.

3.2.3 Theorem. *Assume that $\mathscr{D}(A)$ is dense in X. Then A^* is closed.*

PROOF: Let $x^* \in X^*$, $y^* \in Y^*$, and let $\{y_n^*\}_{n=1}^{\infty}$ in $\mathscr{D}(A^*)$ be given. Assume $y_n^* \to y^*$ and that $A^*y_n^* \to x^*$. We need to show that $y^* \in \mathscr{D}(A^*)$ and that $A^*y^* = x^*$. For all $x \in \mathscr{D}(A)$ we have

$$\langle y_n^*, Ax\rangle \to \langle y^*, Ax\rangle \quad \text{and} \quad \langle y_n^*, Ax\rangle = \langle A^*y_n^*, x\rangle \to \langle x^*, x\rangle$$

as $n \to \infty$. It follows that $\langle y^*, Ax\rangle = \langle x^*, x\rangle$ for all $x \in \mathscr{D}(A)$, so $A^*y^* = x^*$ by the definition of the adjoint. □

3.2.4 Proposition. *Assume that $\mathscr{D}(A)$ is dense X, $A : \mathscr{D}(A) \to Y$ is linear, and Y is reflexive. Then A is closable if and only if $\mathscr{D}(A^*)$ is dense in Y^*.*

There are many identities like $\ker(A^*) = {}^{\perp}\mathrm{range}(A)$, etc. that can be proved about general linear operators. The book *Unbounded linear operators* by S. Goldberg answers many of these questions.

3.2.5 Theorem. *Assume $\mathscr{D}(A)$ is dense in X. Then $\mathscr{D}(A^*) = Y^*$ if and only if A is bounded (i.e. there is C such that $\|Ax\| \le C\|x\|$ for all $x \in \mathscr{D}(A)$). Moreover, if $\mathscr{D}(A^*) = Y^*$ then $A^* \in \mathscr{L}(Y^*; X^*)$ and*

$$\|A^*\| = \sup\{\|Ax\| : x \in \mathscr{D}(A), \|x\| \le 1\}.$$

PROOF: Assume that A is bounded. Let $y^* \in Y^*$ be given. Then $y^*A : \mathscr{D}(A) \to \mathbb{K}$ is a bounded linear functional. By the Hahn-Banach theorem we can choose $x^* \in X$ that extends y^*A. In particular, $\langle y^*, Ax\rangle = \langle x^*, x\rangle$ for all $x \in \mathscr{D}(A)$. It follows that $A^*y^* = x^*$, and $\mathscr{D}(A^*) = Y^*$. Conversely, by the closed graph theorem $A^* \in \mathscr{L}(Y^*; X^*)$. Put $B := \{x \in \mathscr{D}(A) : \|x\| \le 1\}$ and write $|A| := \sup\{\|Ax\| : x \in B\} \le \infty$. For any $y^* \in Y^* = \mathscr{D}(A^*)$,

$$\sup\{|\langle y^*, Ax\rangle| : x \in B\} = \sup\{|\langle A^*y^*, x\rangle| : x \in B\} \le \|A^*\|\|y^*\|.$$

We claim that $\sup\{\|Ax\| : x \in B\} < \infty$. Indeed, let J be the canonical injection of Y into Y^{**}. Put $S = A[B]$. For all $y^* \in Y^*$, $\sup\{|J(y)y^*| : y \in S\} < \infty$, so by the Principle of Uniform Boundedness, $\sup\{\|J(y)\| : y \in S\} < \infty$. Therefore $J[S]$ is bounded in Y^{**}, so it follows that S is bounded in Y. For all $y^* \in Y^*$,

$$\begin{aligned} \|A^* y^*\| &\le \sup\{|\langle A^* y^*, x\rangle| : x \in X, \|x\| \le 1\} \\ &= \sup\{|\langle A^* y^*, x\rangle| : x \in B\} && B \text{ dense} \\ &= \sup\{|\langle y^*, Ax\rangle| : x \in B\} \\ &\le \|y^*\||A| \end{aligned}$$

so $\|A^*\| \le |A|$. Finally, for all $x \in \mathscr{D}(A)$,

$$\begin{aligned} \|Ax\| &= \sup\{|\langle y^*, Ax\rangle| : y^* \in Y^*, \|y^*\| \le 1\} \\ &= \sup\{|\langle A^* y^*, x\rangle| : y^* \in Y^*, \|y^*\| \le 1\} \\ &\le \|A^*\|\|x\| \end{aligned}$$

so $|A| \le \|A^*\|$ and hence $|A| = \|A^*\|$. □

3.3 Spectral theory

Let X be a Banach space over $\mathbb{K}$, $\mathscr{D}(A) \subseteq X$, and $A : \mathscr{D}(A) \to X$ is linear.

3.3.1 Definition. The *resolvent set* of A, denoted $\rho(A)$, is the set of $\lambda \in \mathbb{K}$ such that
(i) $\lambda I - A$ is injective;
(ii) range$(\lambda I - A)$ is dense in X; and
(iii) $(\lambda I - A)^{-1}$: range$(\lambda I - A) \to X$ is bounded.

The *spectrum* of A is defined to be $\sigma(A) := \mathbb{K} \setminus \rho(A)$. We divide the spectrum into three parts.
(i) The *point spectrum* of A, denoted $\sigma_p(A)$, is the set of $\lambda \in \mathbb{K}$ such that $\lambda I - A$ is not injective, i.e. for which (i) fails.
(ii) The *continuous spectrum* of A, denoted $\sigma_c(A)$, is the set of $\lambda \in \mathbb{K}$ such that (i) and (ii) hold but (iii) fails, i.e. $(\lambda I - A)^{-1}$ exists but is unbounded.
(iii) The *residual spectrum* of A, denoted $\sigma_r(A)$, is the set of $\lambda \in \mathbb{K}$ such that (i) holds but (ii) fails, i.e. range$(\lambda I - A)$ is not dense.

These pieces form a partition of the spectrum. The elements of $\sigma_p(A)$ are called eigenvalues and the elements of the respective kernels are called eigenvectors.

3.3.2 Example. Let $X = \ell^2$ and assume that $T \in \mathscr{L}(X;X)$ is compact. Then $0 \in \sigma(T)$. Zero can appear in any of the pieces.
(i) Let $Ax := (0, \frac{1}{2}x_2, \frac{1}{3}x_3, \frac{1}{4}x_4, \dots)$. Then $0 \in \sigma_p(A)$.
(ii) Let $Bx := (x_1, \frac{1}{2}x_2, \frac{1}{3}x_3, \frac{1}{4}x_4, \dots)$. Then $B^{-1}x = (x_1, 2x_2, 3x_3, 4x_4, \dots)$, which is unbounded, so $0 \in \sigma_c(B)$.
(iii) Let R denote the left shift operator and let $C = R \circ B$. Then C is compact and injective but range(C) is not dense, so $0 \in \sigma_r(A)$.

3.3.3 Proposition. *Assume that A is closed and let $\lambda \in \rho(A)$ be given. Then $(\lambda I - A)^{-1} \in \mathscr{L}(X;X)$.*

PROOF: If A is closed then $\lambda I - A$ is closed. By definition of $\rho(A)$, $\lambda I - A$ is injective. By 3.1.6, $(\lambda I - A)^{-1}$ is closed. Again by the definition $\rho(A)$, $(\lambda I - A)^{-1}$ is bounded. By 3.1.2, $\mathscr{D}((\lambda I - A)^{-1})$ is closed. Finally, again from the definition of $\rho(A)$, range$(\lambda I - A) = \mathscr{D}((\lambda I - A)^{-1})$ is dense, so $(\lambda I - A)^{-1}$ is defined on all of X. □

3.3.4 Example. Let $\mathbb{K} = \mathbb{C}$ and $X = L^2[0,1]$. We would like to find closed, densely defined A and B such that $\rho(A) = \varnothing$ and $\sigma(B) = \varnothing$.

(i) Let $\mathscr{D}(A) = \{f \in AC[0,1] : f' \in L^2[0,1]\}$ and $Af = f'$. Let $\lambda \in \mathbb{C}$ be given. Note that $(\lambda I - A)f = 0$ always has a solution, namely $f_\lambda(x) := e^{\lambda x}$. Therefore $\sigma_p(A) = \mathbb{C}$.

(ii) Let $\mathscr{D}(B) = \{f \in AC[0,1] : f' \in L^2[0,1], f(0) = 0\}$ and $Bf = f'$. (Note the boundary condition.) It can be shown that $\mathscr{D}(B)$ is dense in X. Let $\lambda \in \mathbb{C}$ and $g \in X$ be given. We would like to find $f \in \mathscr{D}(B)$ such that $(\lambda I - B)f = g$, which is equivalent to finding a solution to $f' - \lambda f = g$, $f(0) = 0$. Let

$$f_{\lambda,g} = -\int_0^x e^{\lambda(x-t)} g(t)\,dt,$$

and notice that $f_{\lambda,g}$ is the required solution, and it is unique. Finally, it can be shown that $(\lambda I - B)^{-1}$ is compact, so $\rho(B) = \mathbb{C}$.

3.3.5 Definition. Let $\lambda \in \rho(A)$ be given. Define the *resolvent function* of A at λ, $R(\lambda;A) \in \mathscr{L}(X;X)$, by $R(\lambda;A) = (\lambda I - A)^{-1}$.

Notice that $R(\lambda;A) : X \to \mathscr{D}(A)$. For all $x \in \mathscr{D}(A)$,

$$x = (\lambda I - A)R(\lambda;A)x = R(\lambda;A)(\lambda I - A)x,$$

so $AR(\lambda;A)x = R(\lambda;A)Ax = (\lambda R(\lambda;A) - I)x$.

3.3.6 Theorem. *Assume A is closed and let $\lambda_0 \in \rho(A)$ and $\lambda \in \mathbb{K}$ be given, with $|\lambda - \lambda_0| \|R(\lambda_0;A)\| < 1$ be given. Then $\lambda \in \rho(A)$ and*

$$R(\lambda;A) = \sum_{n=0}^{\infty} R(\lambda_0;A)^{n+1}(\lambda_0 - \lambda)^n$$

PROOF: The proof of 2.1.2 goes through unchanged, but one must take care regarding domains. □

3.3.7 Corollary. *$\rho(A)$ is open and $\sigma(A)$ is closed.*

Unlike the continuous case, $\sigma(A)$ is not necessarily bounded.

3.3.8 Theorem. *Assume that A is closed and let $\lambda, \mu \in \rho(A)$ be given. Then*

$$R(\lambda;A) - R(\mu;A) = (\mu - \lambda)R(\lambda;A)R(\mu;A) = (\mu - \lambda)R(\mu;A)R(\lambda;A).$$

PROOF: The proof of 2.1.6 goes through unchanged. □

As in the continuous case,

$$\frac{R(\lambda;A) - R(\mu;A)}{\lambda - \mu} = R(\lambda;A)R(\mu;A),$$

so $R'(\mu;A) = -R(\mu;A)^2$. Again, $R(\cdot;A)$ is an analytic function. Unlike in the continuous case, its composition with a linear functional may fail to be bounded, *cf.* 2.1.7.

4 Semigroups of linear operators

4.1 Introduction

Our goal is to define exponentials of linear operators. We will try to construct e^{tA} as a linear operator, where $A : \mathscr{D}(A) \to X$ is a general linear operator, not necessarily bounded. Notationally, it seems we are looking for a solution to $\dot{\mu}(t) = A\mu(t)$, $\mu(0) = \mu_0$, and we would like to write $\mu(t) = e^{tA}\mu_0$. It turns out that this will hold once we make sense of the terms.

How can we construct e^{tA} when A is a (finite) matrix? The most obvious way is to write down the power series: $\sum_{n=0}^{\infty} \frac{1}{n!}(tA)^n$. This series is absolutely convergent for every A and every $t \in \mathbb{R}$. In fact, this method works for $A \in \mathscr{L}(X;X)$, even if X is infinite dimensional.

A second method is to consider the connexion with the *explicit Euler scheme*. Consider the system of ODE $\dot{\mu}(t) = A\mu(t)$, $\mu(0) = \mu_0$. Partition $[0,t]$ into n parts and write

$$\dot{\mu}\left(\frac{kt}{n}\right) = \frac{n}{t}\left(\mu\left(\frac{(k+1)t}{n}\right) - \mu\left(\frac{kt}{n}\right)\right),$$

the *forward difference quotient* approximation. From the ODE we get

$$\begin{aligned} A\mu\left(\frac{kt}{n}\right) &= \frac{n}{t}\left(\mu\left(\frac{(k+1)t}{n}\right) - \mu\left(\frac{kt}{n}\right)\right) \\ \mu\left(\frac{(k+1)t}{n}\right) &= \left(I + \frac{t}{n}A\right)\mu\left(\frac{kt}{n}\right) \\ \mu(t) = \mu\left(\frac{nt}{n}\right) &\approx \left(I + \frac{t}{n}A\right)^n \mu_0. \end{aligned}$$

Thus $\mu(t) = \lim_{n\to\infty}(I + \frac{t}{n}A)^n\mu_0$ and we write $e^{tA} = \lim_{n\to\infty}(I + \frac{t}{n}A)^n$.

Both of these methods are doomed to failure if A is not bounded. When the explicit method fails, one would normally try the implicit method. The third method we consider is the connexion with the *implicit Euler scheme*. Partition $[0,t]$ into n parts and write

$$\dot{\mu}\left(\frac{(k+1)t}{n}\right)=\frac{n}{t}\left(\mu\left(\frac{(k+1)t}{n}\right)-\mu\left(\frac{kt}{n}\right)\right),$$

the *backward difference quotient* approximation. From the ODE we get

$$\begin{aligned}A\mu\left(\frac{(k+1)t}{n}\right)&=\frac{n}{t}\left(\mu\left(\frac{(k+1)t}{n}\right)-\mu\left(\frac{kt}{n}\right)\right)\\ \mu\left(\frac{(k+1)t}{n}\right)&=\left(I-\frac{t}{n}A\right)^{-1}\mu\left(\frac{kt}{n}\right)\\ \mu(t)=\mu\left(\frac{nt}{n}\right)&\approx\left(I-\frac{t}{n}A\right)^{-n}\mu_0.\end{aligned}$$

Thus $\mu(t)=\lim_{n\to\infty}(I-\frac{t}{n}A)^{-n}\mu_0$ and we write $e^{tA}=\lim_{n\to\infty}(I-\frac{t}{n}A)^{-n}$. This works for some unbounded A as well. The key point will be the behavior of $\|R(\lambda;A)^n\|$ for large n.

An engineer might consider the Laplace transform. If $f(t)=e^{tA}$ then it can be shown that $\hat{f}(\lambda)=(\lambda I-A)^{-1}=R(\lambda;A)$. There is an inversion formula, namely

$$e^{tA}=\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}e^{\lambda t}R(\lambda;A)\,d\lambda,$$

where γ is chosen so that the spectrum of A lies to the left of the line over which we are integrating. This formula can be interpreted and it works for many important unbounded linear operators.

A fifth method works for self-adjoint matrices. Let $\{e_k\}_{k=1}^N$ be an orthonormal basis of X of eigenvectors of A. For any $v\in X$, $v=\sum_{k=1}^N(v,e_k)e_k$ and $Av=\sum_{k=1}^N\lambda_k(v,e_k)e_k$. We take

$$e^{tA}v=\sum_{k=1}^N e^{\lambda_k t}(v,e_k)e_k.$$

In general, if X is a Hilbert space and $A:\mathscr{D}(A)\to X$ is self-adjoint then

$$A=\int_{-\infty}^{\infty}\lambda\,dP(\lambda),$$

where $\{P(\lambda)\mid\lambda\in\mathbb{R}\}$ is the *spectral family* associated with A. We know $\sigma(A)\subseteq\mathbb{R}$, so if $\sigma(A)$ is bounded above then we could define

$$e^{tA}=\int_{-\infty}^{\infty}e^{\lambda t}\,dP(\lambda).$$

Note that the matrix A can be recovered from its exponential via the formula

$$A=\lim_{t\downarrow 0}\tfrac{1}{t}(e^{tA}-I).$$

4.2 Linear C_0-semigroups

Let X be a Banach space over $\mathbb{K}$. By a *linear C_0-semigroup* (or a *strongly continuous semigroup*) we mean a mapping $T : [0,\infty) \to \mathscr{L}(X;X)$ such that

(i) $T(0) = I$;
(ii) $T(t+s) = T(t)T(s)$ for all $s, t \in [0,\infty)$; and
(iii) for all $x \in X$, $\lim_{t\downarrow 0} T(t)x = x$.

Remark.

(i) By the second condition, $T(t)T(s) = T(s)T(t)$ for all s, t.
(ii) We will sometimes use the notation $\{T(t)\}_{t\geq 0}$.
(iii) If we have some mapping $T : [0,\infty) \to \mathscr{L}(X;X)$ satisfying conditions (i) and (ii), (called a semigroup of bounded linear operators) then if the following condition holds then (iii) holds.
(iii') $\lim_{t\downarrow 0}\langle x^*, T(t)x\rangle = \langle x^*, x\rangle$ for all $x^* \in X^*$ and $x \in X$
(iv) The condition $\lim_{t\downarrow 0} \|T(t) - I\| = 0$ implies that $T(t) = \sum_{n=0}^{\infty} \frac{1}{n!}(tA)^n$ for all t, for some $A \in \mathscr{L}(X;X)$. This condition is too strong for practical purposes.
(v) The "C_0" in the name many come from "continuous at zero" or it may refer to the fact that these semigroups are (merely) continuous, as opposed to differentiable, etc.

Let T be a linear C_0-semigroup. The *infinitesimal generator* of T is the linear operator $A : \mathscr{D}(A) \to X$ defined as follows.

$$\mathscr{D}(A) = \left\{x \in X \mid \lim_{t\downarrow 0} \tfrac{1}{t}(T(t)x - x) \text{ exists}\right\}$$

and for all $x \in \mathscr{D}(A)$, $Ax = \lim_{t\downarrow 0} \frac{1}{t}(T(t)x - x)$. It is not immediately obvious that $\mathscr{D}(A) \neq \{0\}$. We will show that $\mathscr{D}(A)$ is dense and that A is a closed linear operator.

4.2.1 Example (Translation semigroup). Let $X = BUC(\mathbb{R})$ = bounded uniformly continuous functions $\mathbb{R} \to \mathbb{K}$. Define $(T(t)f)(x) := f(t+x)$ for all $t \in [0,\infty)$ and $x \in \mathbb{R}$. Clearly T satisfies (i) and (ii) of the definition. Uniform continuity is essential to get (iii). Indeed, if f is uniformly continuous then

$$\|T(t)f - f\|_\infty = \sup\{|f(t+x) - f(x)| : x \in \mathbb{R}\} \to 0 \text{ as } t \to 0.$$

The infinitesimal generator is

$$Af = \lim_{t\downarrow 0} \frac{f(t+x) - f(x)}{t} = f'(x),$$

i.e. differentiation. Note that the solution to the PDE $\mu_t(x,t) = \mu_x(x,t)$, $\mu(x,0) = \mu_0$ is $\mu(x,t) = \mu_0(x+t) = (T(t)\mu_0)(x)$.

4.2.2 Lemma. *Let T be a linear C_0-semigroup. Then there are $M, \omega \in \mathbb{R}$ such that $\|T(t)\| \leq Me^{\omega t}$ for all $t \in [0,\infty)$.*

PROOF: We claim first that there is $\eta > 0$ such that $\sup\{\|T(t)\| : t \in [0,\eta]\}$ is finite. Indeed, suppose there is no such η. Choose $\{t_n\}_{n=1}^{\infty}$ such that $t_n \downarrow 0$ and $\{\|T(t_n)\|, n \in \mathbb{N}\}$ is unbounded. However, for all $x \in X$, since $T(t_n)x \to x$, $\{T(t_n)x\}_{n=1}^{\infty}$ is a convergent sequence, so $\sup\{\|T(t_n)x\| : n \in \mathbb{N}\}$ is finite for each $x \in X$. By the Principle of Uniform Boundedness, $\sup\{\|T(t_n)\| : n \in \mathbb{N}\}$ is finite, a contradiction.

Choose $\eta > 0$ as claimed above. Set $M := \sup\{\|T(t)\| : t \in [0,\eta]\} \geq 1$. Let $t \in [0,\infty)$ be given. Choose $n \geq 0$ and $\alpha \in [0,\eta)$ such that $t = n\eta + \alpha$. Then $T(t) = T(n\eta+\alpha) = (T(\eta))^n T(\alpha)$ by the semigroup property. Whence

$$\|T(t)\| \leq \|T(\alpha)\|\|T(\eta)\|^n \leq MM^n.$$

Put $\omega = \frac{1}{\eta}\log M \geq 0$, so that $\omega t \geq n \log M$, and $\|T(t)\| \leq Me^{\omega t}$. □

4.2.3 Definition. Let T be a linear C_0-semigroup. We say that T is

(i) *uniformly bounded* if there is $M \in \mathbb{R}$ such that $\|T(t)\| \leq M$ for all $t \geq 0$.

(ii) *contractive* if $\|T(t)\| \leq 1$ for all $t \geq 0$.

(iii) *quasi-contractive* provided there is $\omega \in \mathbb{R}$ such that $\|T(t)\| \leq e^{\omega t}$ for all $t \geq 0$.

Contractive semigroups are much easier to study than general linear C_0-semigroups. If T is a linear C_0-semigroup satisfying $\|T(t)\| \leq Me^{\omega t}$ then $S(t) := e^{-\omega t}T(t)$ is a uniformly bounded linear C_0-semigroup. Note that the infinitesimal generator of S is related to that of T as follows.

$$\begin{aligned}
\lim_{t\downarrow 0} \frac{S(t)x - x}{t} &= \lim_{t\downarrow 0} \frac{e^{-\omega t}T(t)x - x}{t} \\
&= \lim_{t\downarrow 0} \frac{e^{-\omega t} - 1}{t} T(t)x + \lim_{t\downarrow 0} \frac{T(t)x - x}{t} \\
&= -\omega x + Ax = (A - \omega I)x
\end{aligned}$$

Further, there is an equivalent norm $|||\cdot|||$ on X such that S is contractive with respect to $|||\cdot|||$. In fact, we may take $|||x||| := \sup\{\|S(t)x\| : t \in [0,\infty)\}$. Indeed, for all $x \in X$,

$$|||S(t)x||| = \sup\{\|S(t+s)x\| : s \in [0,\infty)\} \leq |||x|||.$$

Warning: This norm $|||\cdot|||$ need not preserve all "nice" geometric properties of $\|\cdot\|$, such as the parallelogram law. See the book by Goldstein for an example.

4.2.4 Lemma. *Let T be a linear C_0-semigroup and let $x \in X$ be given. Then the mapping $t \mapsto T(t)x$ is continuous on $[0,\infty)$.*

PROOF: For continuity from the right, let $t \geq 0$ be given and notice that

$$T(t+h)x - T(t)x = (T(h) - I)(T(t)x) \to 0 \text{ as } h \to 0.$$

For continuity from the left, let $t > 0$ and $h \in (0, t)$ be given. Choose $M \geq 1$ and $\omega \geq 0$ such that $\|T(s)\| \leq Me^{\omega s}$ for all $s \in [0, \infty)$.

$$\begin{aligned}\|T(t-h)x - T(t)x\| &= \|T(t-h)(I - T(h))x\| \\ &\leq \|T(t-h)\|\|T(h)x - x\| \\ &\leq Me^{\omega(t-h)}\|T(h)x - x\| \to 0 \text{ as } h \to 0.\end{aligned}$$ □

4.2.5 Lemma. *Let T be a linear C_0-semigroup with infinitesimal generator A, and let $x \in X$ be given.*

(i) *For all $t \geq 0$, $\lim_{h\to 0} \frac{1}{h}\int_t^{t+h} T(s)x ds = T(t)x$ (where the limit is one sided if $t = 0$).*
(ii) *For all $t \geq 0$, $\int_0^t T(s)x ds \in \mathscr{D}(A)$ and $A\int_0^t T(s)x ds = T(t)x - x$.*

Proof: (i) This follows from 4.2.4 and basic calculus.
(ii) If $t = 0$ there is nothing to prove. Let $t > 0$ be given. For $h > 0$,

$$\begin{aligned}\frac{T(h)-I}{h}\int_0^t T(s)x ds &= \frac{1}{h}\int_0^t (T(s+h) - T(s))x ds \\ &= \frac{1}{h}\int_0^t T(s+h)x ds - \frac{1}{h}\int_0^t T(s)x ds \\ &= \frac{1}{h}\int_h^t T(u)x du + \frac{1}{h}\int_t^{t+h} T(u)x du \\ &\qquad - \frac{1}{h}\int_h^t T(s)x ds - \frac{1}{h}\int_0^h T(s)x ds \\ &= \frac{1}{h}\int_t^{t+h} T(u)x du - \frac{1}{h}\int_0^h T(s)x ds \\ &\to T(t)x - x \text{ as } h \to 0\end{aligned}$$

by part (a). The conclusions follows. □

4.2.6 Lemma. *Let T be a linear C_0-semigroup with infinitesimal generator A, and let $x \in \mathscr{D}(A)$ be given. Put $\mu(t) = T(t)x$ for all $t \geq 0$. Then $\mu(t) \in \mathscr{D}(A)$ for all $t \geq 0$, μ is differentiable on $[0, \infty)$, and for each $t \geq 0$,*

$$\dot{\mu}(t) = T(t)Ax = AT(t)x = A\mu(t).$$

Proof: Let $t \geq 0$ be given. For $h > 0$,

$$\frac{T(t+h)x - T(t)x}{h} = \left(\frac{T(h)-I}{h}\right)T(t)x = T(t)\left(\frac{T(h)-I}{h}\right)x \to T(t)Ax$$

as $h \downarrow 0$. In particular, $T(t)x \in \mathscr{D}(A)$ and $AT(t)x = T(t)Ax$. Furthermore, $D^+\mu(t)x = T(t)Ax$. Let $t > 0$ be given. For $h \in (0, t)$,

$$\frac{T(t-h)x - T(t)x}{h} = T(t-h)\left(\frac{x - T(h)x}{h}\right) \to -T(t)Ax \text{ as } h \to 0.$$

so $D^-\mu(t)x = T(t)Ax$. Since the left and right derivatives both exist and are equal, μ is differentiable and $\dot{\mu}(t) = A\mu(t)$. □

4.2.7 Lemma. *Let T be a linear C_0-semigroup with infinitesimal generator A, and let $x \in \mathscr{D}(A)$ be given. Then for all $s, t \in [0, \infty)$,*

$$T(t)x - T(s)x = \int_s^t AT(u)x\,du = \int_s^t T(u)Ax\,du.$$

PROOF: This follows from 4.2.6 and the Fundamental Theorem of Calculus. □

4.2.8 Theorem. *Let T be a linear C_0-semigroup with infinitesimal generator A. Then $\mathscr{D}(A)$ is dense in X and A is closed.*

PROOF: Let $x \in X$ be given. By 4.2.5, $x = \lim_{h\downarrow 0} \frac{1}{h}\int_0^h T(s)x\,ds$, and $\int_0^h T(s)x\,ds \in \mathscr{D}(A)$ for all $h \geq 0$, so $\mathscr{D}(A)$ is dense in X.

Let $\{x_n\}_{n=1}^\infty$ be a sequence in $\mathscr{D}(A)$ converging to $x \in X$ and suppose that $Ax_n \to y \in X$ as $n \to \infty$. We must show that $x \in \mathscr{D}(A)$ and that $Ax = y$. For $h > 0$, by 4.2.7,

$$T(h)x_n - x_n = \int_0^h T(s)Ax_n\,ds.$$

Let $n \to \infty$, noting that we may move the limit under the integral sign for the same reason we may do so in basic calculus, to see

$$T(h)x - x = \int_0^h T(s)y\,ds,$$

so by 4.2.5,

$$Ax = \lim_{h\downarrow 0}\frac{T(h)x - x}{h} = \lim_{h\downarrow 0}\frac{1}{h}\int_0^h T(s)y\,ds = y.$$

It follows that $x \in \mathscr{D}(A)$ and $Ax = y$. □

4.2.9 Lemma. *Let S, T be linear C_0-semigroups having the same infinitesimal generator A. Then $S(t) = T(t)$ for all $t \geq 0$.*

PROOF: Let $x \in \mathscr{D}(A)$ and $t > 0$ be given. Define the function $\mu : [0,t] \to X$ by $\mu(s) = T(t-s)S(s)x$ for all $x \in [0,t]$. We will show that μ is constant as follows. We claim that μ is differentiable on $[0,t]$ and

$$\dot{\mu}(s) = T(t-s)AS(s)x - T(t-s)AS(s)x = 0$$

for all $s \in [0,t]$. This will imply that μ is constant on $[0,t]$, so

$$T(t)x = \mu(0) = \mu(1) = S(t)x.$$

Since $\mathscr{D}(A)$ is dense in X, it will follow that $T(t) = S(t)$ on X for all $t \geq 0$. To prove the claim apply 4.2.6.

$$\begin{aligned}
&\frac{\mu(s+h)-\mu(s)}{h} \\
&= \frac{1}{h}(T(t-s-h)S(s+h)x - T(t-s)S(s)x) \\
&= \frac{1}{h}T(t-s-h)(S(s+h)-S(s))x + \frac{1}{h}(T(t-s-h)-T(t-s))S(s)x \\
&= T(t-s-h)\left(\frac{S(s+h)-S(s)}{h}\right)x + \left(\frac{T(t-s-h)-T(t-s)}{h}\right)S(s)x \\
&\to T(t-s)AS(s)x - T(t-s)AS(s)x = 0 \text{ as } h \to 0.
\end{aligned}$$

The mean value theorem holds as in basic calculus, so μ is constant. □

4.3 Infinitesimal Generators

Given a closed densely defined A, how do we tell if A generates a linear C_0-semigroup? Let $a \in \mathbb{R}$ and $n \in \mathbb{N}$ and put $f(t) = t^{n-1}e^{at}$ for all $t \geq 0$. Recall that the Laplace transform of f is

$$\hat{f}(\lambda) = \frac{(n-1)!}{(\lambda-a)^n}.$$

Let A be an $N \times N$ matrix and put $F(t) = e^{tA}$.

$$\begin{aligned}
\hat{F}(\lambda) = \int_0^\infty e^{-\lambda t}e^{tA}dt = \int_0^\infty e^{t(A-\lambda t)}dt \\
= (A-\lambda I)^{-1}e^{t(A-\lambda I)}\Big|_0^\infty = -(A-\lambda I)^{-1} = R(\lambda;A)
\end{aligned}$$

Recall that $e^{tA} = \lim_{n\to\infty}(I - \frac{t}{n}A)^{-n} = \lim_{n\to\infty}(\frac{n}{t})^n R(\frac{t}{n};A)^n$. To apply this to unbounded operators, the behavior of $R(\lambda;A)^n$ for large n will be key. We conjecture that

$$R(\lambda;A)^n = \frac{1}{(n-1)!}\int_0^\infty e^{-\lambda t}t^{n-1}e^{tA}dt.$$

4.3.1 Lemma. *Let $M, \omega \in \mathbb{R}$ and $\lambda \in \mathbb{K}$ with $\Re\lambda > \omega$ be given. Let T be a linear C_0-semigroup such that $\|T(t)\| \le Me^{\omega t}$ for all $t \ge 0$, and let A be the infinitesimal generator of T. Then $\lambda \in \rho(A)$ and, for all $x \in X$,*

$$R(\lambda;A)x = \int_0^\infty e^{-\lambda t} T(t)x\,dt.$$

PROOF: Put $I_1(\lambda)x = \int_0^\infty e^{-\lambda t} T(t)x\,dt$ for all $x \in X$. We need to show that $\lambda \in \rho(A)$ and $R(\lambda;A) = I_1(\lambda)$. Let $x \in \mathscr{D}(A)$ be given.

$$\begin{aligned}
I_1(\lambda)Ax &= \int_0^\infty e^{-\lambda t} T(t)Ax\,dt \\
&= \int_0^\infty e^{-\lambda t} \frac{d}{dt}(T(t)x)dt && \text{4.2.6} \\
&= -x + \lambda \int_0^\infty e^{-\lambda t} T(t)x\,dt && \text{integration-by-parts} \\
&= \lambda I_1(\lambda)x - x
\end{aligned}$$

Let $x \in X$ be given. We will show that $I_1(\lambda)x \in \mathscr{D}(A)$ and

$$AI_1(\lambda)x = \lambda I_1(\lambda)x - x.$$

Let $h > 0$ be given and compute the difference quotient.

$$\begin{aligned}
&\left(\frac{T(h)-I}{h}\right) I_1(\lambda)x \\
&= \frac{1}{h}\int_0^\infty e^{-\lambda t}(T(t+h)x - T(t)x)dt \\
&= \frac{1}{h}\int_0^\infty e^{-\lambda t} T(t+h)x\,dt - \frac{1}{h}\int_0^\infty e^{-\lambda t} T(t)x\,dt \\
&= \frac{1}{h}\int_h^\infty e^{-\lambda(s-h)} T(s)x\,ds - \frac{1}{h}\int_0^\infty e^{-\lambda t} T(t)x\,dt \\
&= \frac{1}{h}\int_0^\infty e^{-\lambda(t-h)} T(t)x\,dt - \frac{1}{h}\int_0^\infty e^{-\lambda t} T(t)x\,dt - \frac{1}{h}\int_0^h e^{-\lambda(t-h)} T(t)x\,dt \\
&= \int_0^\infty \frac{e^{-\lambda(t-h)} - e^{-\lambda t}}{h} T(t)x\,dt - e^{\lambda h}\frac{1}{h}\int_0^h e^{-\lambda(t-h)} T(t)x\,dt \\
&\to \lambda I_1(\lambda) - x \text{ as } h \to 0.
\end{aligned}$$

This proves the result. □

4.3.2 Lemma. *Let $M, \omega \in \mathbb{R}$ and $\lambda \in \mathbb{K}$ with $\Re\lambda > \omega$ be given. Let T be a linear C_0-semigroup such that $\|T(t)\| \leq Me^{\omega t}$ for all $t \geq 0$, and let A be the infinitesimal generator of T. Then $\lambda \in \rho(A)$ and, for all $n \in \mathbb{N}$ and all $x \in X$,*

$$R(\lambda;A)^n x = \frac{1}{(n-1)!}\int_0^\infty e^{-\lambda t} t^{n-1} T(t)x\,dt.$$

PROOF: We already know that $\rho(A) \supseteq \{\mu \in \mathbb{K} : \Re\mu > \omega\}$. We also know that $\mu \mapsto R(\mu;A)$ is analytic. We have seen that

$$R(\mu;A) = \sum_{n=0}^\infty (\lambda-\mu)^n R(\lambda;A)^{n+1} = \sum_{n=0}^\infty (-1)^n R(\lambda;A)^{n+1}(\mu-\lambda)^n$$

for $|\mu - \lambda|$ sufficiently small. Let $R^{(k)}(\lambda;A)$ denote the k^{th} derivative of $R(\mu;A)$ evaluated at $\mu = \lambda$. From the power series, for all $n \in \mathbb{N}$,

$$\frac{R^{(n-1)}(\lambda;A)}{(n-1)!} = (-1)^{n-1} R(\lambda;A)^n.$$

By 4.3.1, $R(\lambda;A)x = \int_0^\infty e^{-\lambda t} T(t)x\,dt$ for all $x \in X$. From this,

$$R^{(n-1)}(\lambda;A)x = (-1)^{n-1}\int_0^\infty e^{-\lambda t} t^{n-1} T(t)x\,dt.$$

This proves the result. □

4.3.3 Theorem (Hille-Yosida, 1948). *Let $M, \omega \in \mathbb{R}$ be given. Suppose that $A : \mathscr{D}(A) \to X$ is a linear operator with $\mathscr{D}(A) \subseteq X$. Then A is the infinitesimal generator of a linear C_0-semigroup T satisfying $\|T(t)\| \leq Me^{\omega t}$ for all $t \geq 0$ if and only if the following hold.*

(i) A is closed and $\mathscr{D}(A)$ is dense in X; and

(ii) $\rho(A) \supseteq \{\lambda \in \mathbb{R} : \lambda > \omega\}$ and $\|R(\lambda;A)^n\| \leq \frac{M}{(\lambda-\omega)^n}$ for all $\lambda \in \mathbb{R}$ with $\lambda > \omega$ and all $n \in \mathbb{N}$.

The inequality $\|R(\lambda;A)^n\| \leq \frac{M}{(\lambda-\omega)^n}$ might be tough to verify in practise. Notice that $\|R(\lambda;A)\| \leq \frac{M}{\lambda-\omega}$ implies that $\|R(\lambda;A)^n\| \leq \frac{M^n}{(\lambda-\omega)^n}$, so if $M = 1$, i.e. if the semigroup is quasi-contractive, then it is enough to verify the inequality for $n = 1$ only.

Hille-Yoside: proof of necessity

We have already seen that (i) holds, by 4.2.8, and that $\rho(A)$ contains $\{\lambda \in \mathbb{R} : \lambda > \omega\}$, by 4.3.1. By 4.3.2,

$$\begin{aligned} R(\lambda; A)^n x &= \frac{1}{(n-1)!} \int_0^\infty e^{-\lambda t} t^{n-1} T(t) x dt \\ \|R(\lambda; A)^n x\| &\le \frac{1}{(n-1)!} \int_0^\infty e^{-\lambda t} t^{n-1} \|T(t)x\| dt \\ &\le \frac{M}{(n-1)!} \|x\| \int_0^\infty e^{-\lambda t} t^{n-1} e^{\omega t} dt \\ &= \frac{M}{(n-1)!} \frac{(n-1)!}{(\lambda - \omega)^n} \|x\| = \frac{M}{(\lambda - \omega)^n} \|x\| \end{aligned}$$

(The evaluation of the integral can be found in any book explaining the Laplace transform.) This concludes the proof of necessity.

Hille-Yosida: proof of sufficiency

Should we try using the inverse Laplace transform? If we could write

$$T(t) = \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} e^{\lambda t} R(\lambda; A) d\lambda$$

then T would have higher order regularity in general. This method would work for so called "analytic" semigroups, but not for general C_0-semigroups.

How about the limit obtained from considering the implicit scheme? In general $T(t) = \lim_{n\to\infty}(I - \frac{t}{n}A)^{-n}$, and this method can be used, but we will not use it here. What we will do is approximate A with bounded operators $\{A_\lambda\}_{\lambda>\omega}$ and put $T_\lambda(t) = \sum_{n=0}^{\infty} \frac{1}{n!}(tA_\lambda)^n$. Then in theory $T_\lambda(t) \to T(t)$ as $\lambda \to \infty$.

4.3.4 Lemma. *Let $A : \mathscr{D}(A) \to X$ be a linear operator with $\mathscr{D}(A) \subseteq X$. Assume that (i) and (ii) of the Hille-Yosida theorem hold. Then, for all $x \in X$, $\lim_{\lambda\to\infty} \lambda R(\lambda; A)x = x$.*

Proof: Let $x \in \mathscr{D}(A)$ be given. For any $\lambda > \omega$,

$$\begin{aligned} (\lambda I - A) R(\lambda; A) x &= x \\ \lambda R(\lambda; A) x - x &= A R(\lambda; A) x \\ &= R(\lambda; A) A x \\ \|\lambda R(\lambda; A) x - x\| &= \|R(\lambda; A) A x\| \\ &\le \frac{M}{\lambda - \omega} \|Ax\| \\ &\to 0 \text{ as } \lambda \to \infty \end{aligned}$$

Since $\mathscr{D}(A)$ is dense in X, the result follows. □

The *Yosida approximation* A_λ of A, for $\lambda > \omega$, is defined by

$$A_\lambda x := \lambda AR(\lambda;A)x = (\lambda^2 R(\lambda;A) - \lambda I)x.$$

By 4.3.4, $A_\lambda x \to Ax$ as $\lambda \to \infty$ for all $x \in \mathscr{D}(A)$. We claim the following results, given as a homework exercises. Let $B \in \mathscr{L}(X;X)$ be given and define $e^{tB} = \sum_{n=0}^{\infty} \frac{1}{n!}(tB)^n$ for all $t \in \mathbb{R}$.

(i) $\{e^{tB}\}_{t\geq 0}$ is a linear C_0-semigroup with infinitesimal generator B.

(ii) $\lim_{t\to 0} \|e^{tB} - I\| = 0$

(iii) For all $\lambda \in \mathbb{K}$, $e^{t(B-\lambda I)} = e^{-\lambda t}e^{tB}$.

It can be shown that if T is a linear C_0-semigroup with the property that $\lim_{h\downarrow 0} \|T(h) - I\| = 0$ then $T(t) = e^{tB}$ for some $B \in \mathscr{L}(X;X)$.

Assume that conditions (i) and (ii) of the Hille-Yoside theorem hold. Put $A_\lambda = \lambda^2 R(\lambda;A) - \lambda I$ and notice that for any $\lambda > \omega$,

$$\begin{aligned}
e^{tA_\lambda} &= e^{-\lambda t}\sum_{n=0}^{\infty}\frac{\lambda^{2n}t^n R(\lambda;A)^n}{n!} \\
\|e^{tA_\lambda}\| &\leq Me^{-\lambda t}\sum_{n=0}^{\infty}\frac{\lambda^{2n}t^n}{(\lambda-\omega)^n n!} && \text{by (ii)} \\
&= Me^{-\lambda t}\exp\left(\frac{\lambda^2}{\lambda-\omega}t\right) && \lambda > \omega \\
&= M\exp\left(\frac{\lambda\omega}{\lambda-\omega}t\right).
\end{aligned}$$

It follows that $\|e^{tA_\lambda}\| \leq Me^{\omega_1 t}$ for any fixed $\omega_1 > \omega$, for all λ sufficiently large with respect to ω.

Put $T_\lambda(t) = e^{tA_\lambda}$ for all $t \geq 0$ and $\lambda > \omega$. Notice that $A_\lambda A_\mu = A_\mu A_\lambda$ and $A_\lambda T_\mu(t) = T_\mu(t)A_\lambda$ for all $\lambda, \mu > \omega$. Let $x \in \mathscr{D}(A)$ be given.

$$\begin{aligned}
T_\lambda(t)x - T_\mu(t)x &= \int_0^t \frac{d}{ds}(T_\mu(t-s)T_\lambda(s)x)ds \\
&= \int_0^t T_\mu(t-s)A_\lambda T_\lambda(s)x - T_\mu(t-s)A_\mu T_\lambda(s)x ds \\
&= \int_0^t (T_\mu(t-s)T_\lambda(s))(A_\lambda x - A_\mu x)ds \\
\|T_\lambda(t)x - T_\mu(t)x\| &\leq M^2 e^{\omega_1 t} t\|A_\lambda x - A_\mu x\|
\end{aligned}$$

Therefore $\{T_\lambda(t)x\}_{\lambda>\omega}$ has the Cauchy property, uniformly in t on bounded intervals. $\mathscr{D}(A)$ is dense in X and we have a bound on $\|T_\lambda(t)\|$ (in λ), so for all $x \in X$, $\lim_{\lambda\to\infty} T_\lambda(t)x$ exists.

For all $t \geq 0$ and $x \in X$, put $T(t)x = \lim_{\lambda\to\infty} T_\lambda(t)x$. Note that $\|T(t)\| \leq Me^{\omega_1 t}$, $T(t)T(s) = T(t+s)$ for all $s, t \geq 0$, and $T(0) = I$, since these relations hold for each T_λ. Continuity follows since the convergence is uniform for t in bounded intervals. Let B be the infinitesimal generator of T. We must show that $B = A$. First we will show that B extends A, and then we will use a resolvent argument to show that $\mathscr{D}(A) = \mathscr{D}(B)$. Let $x \in \mathscr{D}(A)$ be given.

$$\begin{aligned} \|T_\lambda(t)A_\lambda x - T(t)Ax\| &\le \|T_\lambda(t)(A_\lambda x - Ax)\| + \|(T_\lambda(t) - T(t))Ax\| \\ &\le Me^{\omega_1 t}\|A_\lambda x - Ax\| + \|(T_\lambda(t) - T(t))Ax\| \\ &\to 0 \text{ as } \lambda \to \infty \end{aligned}$$

Since the convergence is uniform in t on bounded intervals,

$$\begin{aligned} T(t)x - x &= \lim_{\lambda\to\infty} T_\lambda(t)x - x \\ &= \lim_{\lambda\to\infty} \int_0^t T_\lambda(s)A_\lambda x\,dx = \int_0^t T(s)Ax\,dx. \end{aligned}$$

Checking the definition of B, for any $h > 0$,

$$\frac{T(h)x - x}{h} = \frac{1}{h}\int_0^h T(s)Ax\,ds \to Ax \text{ as } h \downarrow 0.$$

Therefore $x \in \mathscr{D}(B)$ and $Bx = Ax$. B is closed since it is the infinitesimal generator of a linear C_0-semigroup, and A is closed by assumption. Since $\|T(t)\| \le Me^{\omega_1 t}$ for any $\omega_1 > \omega$, by 4.3.1 $\rho(B) \supseteq (\omega, \infty)$, so it follows that $\rho(B) \cap \rho(A) \ne \emptyset$. Choose $\lambda \in \rho(A) \cap \rho(B)$. By 3.3.3, since A and B are closed, $(\lambda I - A)[\mathscr{D}(A)] = X$ and $(\lambda I - B)[\mathscr{D}(B)] = X$. Further, since B extends A, $(\lambda I - B)[\mathscr{D}(A)] = (\lambda I - A)[\mathscr{D}(A)] = X$. To conclude the proof of the Hille-Yoside theorem, note that $\mathscr{D}(A) = R(\lambda; B)[X] = \mathscr{D}(B)$.

Remark. Let $A : \mathscr{D}(A) \to X$ be a linear operator with $\mathscr{D}(A) \subseteq X$. The following are equivalent.

(i) A is closed;
(ii) $(\lambda I - A) : \mathscr{D}(A) \to X$ is a bijection for some $\lambda \in \rho(A)$;
(iii) $(\lambda I - A) : \mathscr{D}(A) \to X$ is a bijection for all $\lambda \in \rho(A)$.

4.3.5 Corollary. *Assume that $A : \mathscr{D}(A) \to X$ is linear with $\mathscr{D}(A) \subseteq X$, and that $\mathscr{D}(A)$ is dense and A is closed. Then A generates a contractive linear C_0-semigroup if and only if $\rho(A) \supseteq (0, \infty)$ and $\|R(\lambda; A)\| \le \frac{1}{\lambda}$ for all $\lambda > 0$.*

4.4 Contractive semigroups

Let $T : [0, \infty) \to \mathscr{L}(X; X)$ be a contractive semigroup. For all $t, h \in [0, \infty)$,

$$\|T(t + h)\| = \|T(h)T(t)\| \le \|T(h)\|\|T(t)\| \le \|T(t)\|,$$

so $t \mapsto \|T(t)\|$ is a decreasing function. Assume for now that X is Hilbert space. Let $x \in \mathscr{D}(A)$ be given and put $\mu(t) = \|T(t)x\|^2 = (T(t)x, T(t)x)$. For all $t \ge 0$, since μ is decreasing,

$$0 \ge \dot{\mu}(t) = (T(t)x, T(t)Ax) + (T(t)Ax, T(t)x) = 2\Re(AT(t)x, x).$$

In particular, for $t=0$, $\Re(Ax,x)\le 0$ for all $x\in\mathscr{D}(A)$.

We will prove that if X is a Hilbert space and $A:\mathscr{D}(A)\to X$ is a linear operator then A generates a generates a contractive semigroup if and only if both of the following hold.

(i) $\Re(Ax,x)\le 0$ for all $x\in\mathscr{D}(A)$; and
(ii) there exists $\lambda_0>0$ such that $\lambda_0 I-A$ is surjective.

4.4.1 Definition. Let X be a Banach space over $\mathbb{K}$ with norm $\|\cdot\|$. By a *semi-inner product* on X, we mean a mapping $[\cdot,\cdot]:X\times X\to\mathbb{K}$ such that

(i) $[x+y,z]=[x,z]+[y,z]$ for all $x,y,z\in X$;
(ii) $[\alpha x,y]=\alpha[x,y]$ for all $x,y\in X$ and $\alpha\in\mathbb{K}$;
(iii) $[x,x]=\|x\|^2$ for all $x\in X$; and
(iv) $|[x,y]|\le\|x\|\|y\|$ for all $x,y\in X$.

Remark. The term "semi-inner product" is frequently used in a more general sense that is not linked to a pre-existing norm.

We must ask, do semi-inner products exist, and can there be more than one associated with any given norm? The answer to both is yes in general. However, if X^* is strictly convex then there cannot be more than one. We will see that if $\Re[Ax,x]\le 0$ with respect to one semi-inner product then it holds with respect to any semi-inner product.

4.4.2 Proposition. *There is at least one semi-inner product on a Banach space.*

PROOF: Let X be a Banach space. For every $x\in X$ put

$$\mathscr{F}(x):=\{x^*\in X^*\mid\langle x^*,x\rangle=\|x\|^2=\|x^*\|^2\}.$$

By the Hahn-Banach theorem $\mathscr{F}(x)$ is non-empty for every $x\in X$. For every $x\in X$, choose $F(x)\in\mathscr{F}(x)$. Define $[\cdot,\cdot]:X\times X\to\mathbb{K}$ by $[x,y]=\langle F(y),x\rangle$ for all $x,y\in X$. ☐

If X^* is strictly convex then there is exactly one semi-inner product, essentially because the set $\mathscr{F}(x)$ contains a single element.

4.4.3 Definition. Assume that $A:\mathscr{D}(A)\to X$ is linear with $\mathscr{D}(A)\subseteq X$. We say that A is *dissipative* provided that there is a semi-inner product on X such that $\Re[Ax,x]\le 0$ for all $x\in\mathscr{D}(A)$.

The notion of dissipativity depends on the particular norm used. Given a particular norm, it turns out that dissipativity will not depend on the semi-inner product used.

Aside: Consider $\mu_{tt}(x,t)=\Delta\mu(x,t)-\alpha(x)\mu_t(x,t)$ with $\mu|_{\partial\Omega}=0$, where α is non-negative, smooth, with compact support, and $\int_\Omega\alpha>0$. Then solutions μ tend to zero with t. Crazy!

4.4.4 Lemma. *Assume that $A : \mathscr{D}(A) \to X$ is linear with $\mathscr{D}(A) \subseteq X$. Then A is dissipative if and only if $\|(\lambda I - A)x\| \geq \lambda\|x\|$ for all $x \in \mathscr{D}(A)$ and $\lambda > 0$.*

PROOF: Assume that A is dissipative. Choose a semi-inner product such that $\Re[Ax, x] \leq 0$ for all $x \in \mathscr{D}(A)$. Then for all $x \in \mathscr{D}(A)$ and $\lambda > 0$, we have

$$\Re[(A - \lambda I)x, x] = \lambda\|x\|^2 - \Re[Ax, x] \geq \lambda\|x\|^2.$$

Combining that with the fact that

$$\Re[(\lambda I - A)x, x] \leq |[(\lambda I - A)x, x]| \leq \|(\lambda I - A)x\|\|x\|$$

yields the result. Assume now that $\|(\lambda I - A)x\| \geq \lambda\|x\|$ for all $x \in \mathscr{D}(A)$ and $\lambda > 0$. As before, put

$$\mathscr{F}(x) := \{x^* \in X^* \mid \langle x^*, x\rangle = \|x\|^2 = \|x^*\|^2\}.$$

We identify three cases: $x = 0$, $x \in \mathscr{D}(A) \setminus \{0\}$, and $x \notin \mathscr{D}(A)$.

Let $x \in \mathscr{D}(A) \setminus \{0\}$ be given. Notice that $\|(nI - A)x\| \geq n\|x\|$ for all $n \in \mathbb{N}$. Choose $y_n^* \in \mathscr{F}(nx - Ax)$ and put $z_n^* = \frac{1}{\|y_n^*\|} y_n^*$ for all $n \in \mathbb{N}$.

$$\begin{aligned}
n\|x\| &\leq \|nx - Ax\| && \text{by assumption} \\
&= \frac{1}{\|y_n^*\|}\langle y_n^*, nx - Ax\rangle && \text{since } y_n^* \in \mathscr{F}(nx - Ax) \\
&= \langle z_n^*, nx - Ax\rangle && \text{(this is a real number)} \\
&= n\Re\langle z_n^*, x\rangle - \Re\langle z_n^*, Ax\rangle
\end{aligned}$$

Since $\|z_n^*\| = 1$ by construction,

$$n\|x\| \leq n\Re\langle z_n^*, x\rangle - \Re\langle z_n^*, Ax\rangle \leq n\|x\| - \Re\langle z_n^*, Ax\rangle$$

Therefore $\Re\langle z_n^*, Ax\rangle \leq 0$ and similarly $\Re\langle z_n^*, x\rangle \geq \|x\| - \frac{1}{n}\|Ax\|$. Assume with great loss of generality that X is reflexive or separable. Choose a subsequence $\{z_{n_k}^*\}_{k=1}^\infty$ of $\{z_n^*\}_{n=1}^\infty$ and $z^* \in X^*$ such that $z_{n_k}^* \xrightarrow{*} z^*$ as $k \to \infty$. (In general we would use nets.) Then $\|z^*\| \leq 1$, $\Re\langle z^*, Ax\rangle \leq 0$, and $\Re\langle z^*, x\rangle \geq \|x\|$. It follows that $\langle z^*, x\rangle = \|x\|$. Define a semi-inner product as before, but with

$$F(x) = \begin{cases} 0 & x = 0 \\ z^*\|x\| & x \in \mathscr{D}(A) \setminus \{0\} \\ \text{anything in } \mathscr{F}(x) & x \in X \setminus \mathscr{D}(A) \end{cases}$$

□

4.4.5 Lemma. *Assume that $A : \mathscr{D}(A) \to X$ is linear with $\mathscr{D}(A) \subseteq X$ and that A is dissipative. Let $\lambda_0 \in (0, \infty)$ be given and assume that $\lambda_0 I - A$ is surjective. Then A is closed, $\rho(A) \supseteq (0, \infty)$, and $\|R(\lambda; A)\| \leq \frac{1}{\lambda}$ for all $\lambda > 0$.*

PROOF: Notice that, by 4.4.4, $\|(\lambda I - A)x\| \geq \lambda\|x\|$ for all $x \in \mathscr{D}(A)$ and $\lambda > 0$. Whence immediately $\|R(\lambda; A)\| \leq \frac{1}{\lambda}$, provided the resolvent exists. The key points are to show that A is closed and that $\lambda I - A$ is surjective for all $\lambda > 0$.

Notice that $\lambda_0 I - A$ is bijective since it is surjective and bounded below, and further, $\|(\lambda_0 I - A)^{-1}x\| \leq \frac{1}{\lambda_0}\|x\|$. So $(\lambda_0 I - A)^{-1} \in \mathscr{L}(X; X)$, hence it is closed, so A is closed by 3.1.6.

To show that $\rho(A) \supseteq (0, \infty)$ it suffices to show that $(\lambda I - A)^{-1}$ is surjective for all $\lambda > 0$. Put $\Lambda = \{\lambda \in (0, \infty) : \lambda \in \rho(A)\}$, which is open (in the relative topology of $(0, \infty)$) and non-empty. We will show Λ is closed and conclude $\Lambda = (0, \infty)$. Let $\{\lambda_n\}_{n=1}^{\infty}$ be a sequence in Λ converging to $\lambda^* \in (0, \infty)$. We will show that $\lambda^* \in \Lambda$ by showing that $\lambda^* I - A$ is surjective. Let $y \in X$ be given. Produce $x \in X$ such that $(\lambda^* I - A)x = y$ as follows. For every $n \in \mathbb{N}$ put $x_n = R(\lambda_n; A)y$. Note that $\sup\{\frac{1}{\lambda_n} : n \in \mathbb{N}\} < \infty$.

$$\begin{aligned} \|x_n - x_m\| &= \|(R(\lambda_n; A) - R(\lambda_m; A))y\| \\ &= |\lambda_m - \lambda_n| \|R(\lambda_n; A)R(\lambda_m; A)y\| \\ &\leq |\lambda_m - \lambda_n| \frac{\|y\|}{\lambda_n \lambda_m} \\ &\to 0 \text{ as } n, m \to \infty \end{aligned}$$

Write $x_n \to x$. Finally, $\{x_n\}_{n=1}^{\infty} \subseteq \mathscr{D}(A)$, $x_n \to x$, and $Ax_n \to \lambda^* x - y$. Since A is closed $(\lambda^* I - A)x = y$. □

4.4.6 Theorem (Lumer-Phillips, 1961). *Assume $A : \mathscr{D}(A) \to X$ is linear with $\mathscr{D}(A)$ dense in X.*

(i) If A is dissipative and there is $\lambda_0 > 0$ such that $\lambda_0 I - A$ is surjective then A generates a contractive linear C_0-semigroup.

(ii) If A generates a contractive linear C_0-semigroup then $\lambda I - A$ is surjective for all $\lambda > 0$ and $\Re[Ax, x] \leq 0$ for all $x \in \mathscr{D}(A)$ and every semi-inner product on X (in particular, A is dissipative).

PROOF: The first part follows from 4.4.5 and the Hille-Yoside theorem, since $\|R(\lambda; A)\| \leq \frac{1}{\lambda}$ implies $\|R(\lambda; A)^n\| \leq \frac{1}{\lambda^n}$.

In the second part, the surjectivity conclusion follows from the Hille-Yosida theorem. Let $[\cdot, \cdot]$ be a semi-inner product on X. We need to show that $\Re[Ax, x] \leq 0$ for all $x \in \mathscr{D}(A)$. For all $h > 0$ and $x \in \mathscr{D}(A)$,

$$\begin{aligned} \Re[T(h)x - x, x] &= \Re[T(h)x, x] - \|x\|^2 \\ &\leq \|T(h)x\|\|x\| - \|x\|^2 \leq \|x\|^2 - \|x\|^2 \leq 0 \end{aligned}$$

Dividing by h and letting $h \downarrow 0$ we get $\Re[Ax, x] \leq 0$. □

4.4.7 Corollary. *Assume $B : \mathscr{D}(B) \to X$ is linear with $\mathscr{D}(B)$ dense in X. Let $\omega, \lambda_0 \in \mathbb{R}$ with $\lambda_0 > \omega$ be given. If $\lambda_0 I - B$ is surjective and there exists a semi-inner product on X such that $\Re[Bx, x] \leq \omega\|x\|^2$ for all $x \in \mathscr{D}(B)$, then B generates a linear C_0-semigroup T such that $\|T(t)\| \leq e^{\omega t}$.*

PROOF: Put $A = B - \omega I$ and apply the Lumer-Phillips theorem to A. □

4.4.8 Lemma. *Assume X that is reflexive and that $A : \mathscr{D}(A) \to X$ is linear with $\mathscr{D}(A) \subseteq X$. Let $\lambda_0 > 0$ be given and assume that A is dissipative and that $\lambda_0 I - A$ is surjective. Then $\mathscr{D}(A)$ is dense in X.*

This lemma shows that if X is reflexive then we do not need to assume that $\mathscr{D}(A)$ is dense in the Lumer-Phillips theorem. This is less helpful than it seems because in many applications it is trivial to check that the domain is dense.

Remark. Let M be a linear manifold in a Banach space X (not necessarily reflexive). Then M is dense in X if and only if for all $y \in X$ there is a sequence $\{x_n\}_{n=1}^{\infty} \subseteq M$ such that $x_n \rightharpoonup y$ as $n \to \infty$. Indeed, one direction is trivial. For the other, if y is not in the closure of M then $\text{dist}(M, y) > 0$. By Hahn-Banach there is $y^* \in X^*$ such that $y^*(x) = 0$ for all $x \in M$ and $y^*(y) \neq 0$.

PROOF: Let $y \in X$ be given. It suffices to prove that there is $\{x_n\}_{n=1}^{\infty} \mathscr{D}(A)$ such that $x_n \rightharpoonup y$ as $n \to \infty$. Put $x_n = (I - \frac{1}{n}A)^{-1} y = nR(n; A)y \in \mathscr{D}(A)$ for all $n \in \mathbb{N}$. Then

$$\|x_n\| \leq n\|R(n; A)\|\|y\| \leq n\frac{1}{n}\|y\| = \|y\|.$$

Choose a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ and $x \in X$ such that $x_{n_k} \rightharpoonup z$ as $k \to \infty$. We are done if we show $y = z$. But

$$A\left(\frac{x_{n_k}}{n_k}\right) = x_{n_k} - y \rightharpoonup z - y$$

and $x_{n_k} \rightharpoonup 0$ (in fact, $x_{n_k} \to 0$). $\text{Gr}(A)$ is closed and convex, so it is weakly closed. Since $(0, z - y) \in \text{Gr}(A)$, $z = y$. □

4.4.9 Theorem (Lumer-Phillips for Hilbert spaces).
Let X be a Hilbert space and assume that $B : \mathscr{D}(B) \to X$ is linear with $\mathscr{D}(B) \subseteq X$. Let $\omega \in \mathbb{R}$ and $\lambda_0 > \omega$ be given. Assume that $\Re(Bx, x) \leq \omega\|x\|^2$ for all $x \in \mathscr{D}(B)$ and that $\lambda_0 I - B$ is surjective. Then B generates a linear C_0-semigroup T such that $\|T(t)\| \leq e^{\omega t}$ for all $t \geq 0$.

4.4.10 Example. Let

$$\mathscr{D}(A) := \{u \in AC[0,1] : u' \in AC[0,1], u'' \in L^2[0,1], u(0) = u(1) = 0\},$$

and $Au := u''$. We have seen that A is closed and A is densely defined (in fact it is self-adjoint). For any $u \in \mathscr{D}(A)$,

$$(Au, u) = \int_0^1 u'' u \, dx = -\int_0^1 (u')^2 dx \leq 0$$

If we can solve the ODE $u - u'' = f$, $u(0) = u(1) = 0$ for any $f \in L^2(0,1)$, then A generates a contraction semigroup T by the Lumer-Phillips theorem. Thus the solutions to the heat equation

$$\begin{cases} u_t - u_{xx} = 0 & \text{on } (0,1) \\ u(t,0) = u(t,1) = 0 & \text{for all } t \geq 0 \\ u(0,x) = g(x) & \text{for all } x \in (0,1) \end{cases}$$

can be written $u(x,t) = (T(t)g)(x)$.

5 Fourier Transforms

5.1 Multi-index notation

Before introducing the Fourier transform we review the concept of *multi-indices*, and restate some well-known theorems in this notation. Let $n \in \mathbb{N}$ be given. By a *multi-index* of length n we mean a list $\alpha = (\alpha_1, \dots, \alpha_n)$ such that each α_i is a non-negative integer. The set of all multi-indices of length n is denoted by M_n. The notation of multi-indices was introduced by Whitney (reported by L. Tartar via personal communication with L. Schwartz). Write

- $|\alpha| := \sum_{i=1}^n \alpha_i$;
- $\alpha! := \alpha_1! \cdots \alpha_n!$; and
- $\alpha \leq \beta$ if and only if $\alpha_i \leq \beta_i$ for all $i = 1, \dots, n$.

For $\alpha \leq \beta$, write

$$\binom{\alpha}{\beta} := \frac{\beta!}{(\beta - \alpha)!\alpha!}.$$

For $f : \mathbb{R}^n \to \mathbb{K}$ of class C^∞, write

$$D^\alpha f(x) := \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}(x).$$

We have the Binomial Theorem,

$$(x+y)^\alpha = \sum_{\beta \leq \alpha} \binom{\alpha}{\beta} x^\beta y^{\alpha - \beta}$$

and Taylor's Theorem

$$f(x_0 + x) = \sum_{\alpha \in M_n} \frac{1}{\alpha!} D^\alpha f(x_0) x^\alpha.$$

Finally, we introduce one piece of non-standard notation,

$$P_\alpha(x) := x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}.$$

The standard notation is to simply write x^α, but it will be convenient to have a name for this oft-used function.

5.2 Fourier transforms

5.2.1 Definition. The *Fourier transform* of $f \in L^1(\mathbb{R}^n;\mathbb{K})$ is $\hat{f} : \mathbb{R}^n \to \mathbb{C}$ defined by

$$\hat{f}(\xi) := \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \exp(-ix\cdot\xi) f(x)dx.$$

There are several closely related, but slightly different, definitions of $\hat{f}$ appearing in the literature. Sometimes the normalizing constant is changed or left out, and sometimes the basis functions are modified by changing the sign or inserting a factor of 2π. Be very careful about any formulae related to Fourier series that you pull from an unfamiliar book.

5.2.2 Lemma. *For $f \in L^1(\mathbb{R}^n)$, $\hat{f}$ is continuous and* $\lim_{|\xi|\to\infty} \hat{f}(\xi) = 0$.

PROOF: Suppose that $\xi_m \to \xi$ in $\mathbb{R}^n$. For all $m \geq 1$, $|e^{-ix\cdot\xi_m} f(x)| = |f(x)|$ for all $x \in \mathbb{R}^n$. Therefore, by the Lebesgue dominated convergence theorem, $\hat{f}(\xi_m) \to \hat{f}(\xi)$. Since this holds for arbitrary convergent sequences, $\hat{f}$ is continuous.

The second assertion is the *Riemann-Lebesgue lemma* and its proof can be found in any text on Fourier transforms. □

Remark. Not every continuous function that vanishes at infinity is the Fourier transform of some L^1 function. Indeed, $\hat{}: L^1(\mathbb{R}^n) \to C_v(\mathbb{R}^n)$ is an injective linear mapping that is continuous, because $|\hat{f}(\xi)| \leq \|f\|_{L^1}$ for all $\xi \in \mathbb{R}^n$. But it can be shown that these spaces are not isomorphic, so the Fourier transform cannot be surjective. We will see that, for $1 \leq p \leq 2$ with $\frac{1}{p} + \frac{1}{q} = 1$, the Fourier transform can be defined on L^p and maps into L^q, but is surjective if and only if $p = q = 2$.

Suppose that f is really nice, by which we mean that it has sufficient differentiability and boundedness properties that all of the following computations are valid.

$$\begin{aligned}
(D^\alpha \hat{f})(\xi) &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} D_\xi^\alpha e^{-ix\cdot\xi} f(x)dx \\
&= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} (-ix)^\alpha e^{-ix\cdot\xi} f(x)dx \\
&= (-i)^{|\alpha|} \widehat{P_\alpha f}(\xi) \\
(D^\alpha f)\hat{}(\xi) &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{-ix\cdot\xi} D^\alpha f(x)dx \\
&= \frac{1}{(2\pi)^{n/2}} (-1)^{|\alpha|} \int_{\mathbb{R}^n} (-i\xi)^\alpha e^{-ix\cdot\xi} f(x)dx \\
&= i^{|\alpha|} P_\alpha(\xi)\hat{f}(\xi)
\end{aligned}$$

Therefore, $D^\alpha \hat{f} = (-i)^{|\alpha|}(P_\alpha f)\hat{}$ and $(D^\alpha f)\hat{} = i^{|\alpha|} P_\alpha \hat{f}$.

5.2.3 Definition (Schwartz space).

$$\mathscr{S}(\mathbb{R}^n) := \{\varphi \in C^\infty(\mathbb{R}^n) \mid \sup_{x\in\mathbb{R}} |x^\beta D^\alpha \varphi(x)| < \infty \text{ for all multi-indices } \alpha \text{ and } \beta\}$$

This is also known as the collection of *rapidly decreasing functions*.

5.2.4 Lemma. *Let $\varphi \in \mathscr{S}(\mathbb{R}^n)$ be given and let α be a multi-index. Then $\hat{\varphi} \in \mathscr{S}(\mathbb{R}^n)$, $D^\alpha \hat{\varphi} = (-i)^{|\alpha|}\widehat{P_\alpha \varphi}$, and $\widehat{D^\alpha \varphi} = i^{|\alpha|} P_\alpha \hat{\varphi}$.*

Proof: The computations above are valid for elements of $\mathscr{S}(\mathbb{R}^n)$. □

5.2.5 Lemma. *Let $\varphi, \psi \in \mathscr{S}(\mathbb{R}^n)$ and $x \in \mathbb{R}^n$ be given. Then*

$$\int_{\mathbb{R}^n} e^{ix\cdot\xi} \hat{\varphi}(\xi)\psi(\xi) d\xi = \int_{\mathbb{R}^n} \varphi(x+y)\hat{\psi}(y) dy$$

Proof: The following is valid for elements of $\mathscr{S}$.

$$\begin{aligned}\int_{\mathbb{R}^n} e^{ix\cdot\xi} \hat{\varphi}(\xi)\psi(\xi) d\xi &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} e^{ix\cdot\xi} e^{-iz\cdot\xi} \varphi(z)\psi(\xi) dz d\xi \\ &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} e^{-i(z-x)\cdot\xi} \psi(\xi)\varphi(z) d\xi dz \\ &= \int_{\mathbb{R}^n} \hat{\psi}(z-x)\varphi(z) dz = \int_{\mathbb{R}^n} \hat{\psi}(y)\varphi(x+y) dz\end{aligned}$$ □

5.2.6 Theorem (Fourier inversion). *Let $\varphi \in \mathscr{S}(\mathbb{R}^n)$ be given. Then*

$$\varphi(x) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \exp(ix\cdot\xi)\hat{\varphi}(\xi) d\xi.$$

Proof: Let $\psi \in \mathscr{S}(\mathbb{R}^n)$ be given. Define $\psi_\varepsilon : \mathbb{R}^n \to \mathbb{K}$ by $\psi_\varepsilon(z) := \psi(\varepsilon z)$.

$$\widehat{\psi_\varepsilon}(\xi) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{-ix\cdot\xi}\psi(\varepsilon x) dx = \frac{\varepsilon^{-n}}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{-i\frac{z}{\varepsilon}\cdot\xi}\psi(z) dz$$

Thus, by 5.2.5,

$$\begin{aligned}
\int_{\mathbb{R}^n} e^{ix\cdot\xi}\hat{\varphi}(\xi)\psi(\varepsilon\xi)d\xi &= \int_{\mathbb{R}^n} e^{ix\cdot\xi}\varphi(x+y)\widehat{\psi_\varepsilon}(y)dy \\
&= \frac{\varepsilon^{-n}}{(2\pi)^{n/2}}\int_{\mathbb{R}^n}\int_{\mathbb{R}^n}\varphi(x+y)e^{-i\frac{z}{\varepsilon}\cdot y}\psi(z)dzdy \\
&= \frac{\varepsilon^{-n}}{(2\pi)^{n/2}}\int_{\mathbb{R}^n}\int_{\mathbb{R}^n}\varphi(x+y)e^{-iz\cdot\frac{y}{\varepsilon}}\psi(z)dzdy \\
&= \varepsilon^{-n}\int_{\mathbb{R}^n}\varphi(x+y)\hat{\psi}\left(\frac{y}{\varepsilon}\right)dy \\
&= \int_{\mathbb{R}^n}\varphi(x+\varepsilon z)\hat{\psi}(z)dz \\
\int_{\mathbb{R}^n} e^{ix\cdot\xi}\hat{\varphi}(\xi)\psi(0)d\xi &= \int_{\mathbb{R}^n}\varphi(x)\hat{\psi}(z)dz \quad \text{letting } \varepsilon\downarrow 0.
\end{aligned}$$

Whence, for every $\psi \in \mathscr{S}(\mathbb{R}^n)$,

$$\psi(0)\int_{\mathbb{R}^n} e^{ix\cdot\xi}\hat{\varphi}(\xi)d\xi = \varphi(x)\int_{\mathbb{R}^n}\hat{\psi}(z)dz.$$

To prove the theorem we choose a convenient $\psi \in \mathscr{S}(\mathbb{R}^n)$. Namely, we pick the normal density function, $\psi(x) = \exp(-\frac{1}{2}|x|^2)$, so that

$$\psi(0) = (2\pi)^{n/2}\int_{\mathbb{R}^n}\hat{\psi}(z)dz. \tag{1}$$

The result follows. □

Remark. As a consequence of the proof, equation (1) holds for all $\psi \in \mathscr{S}(\mathbb{R}^n)$.

Given $\varphi \in \mathscr{S}(\mathbb{R}^n)$, define $\check{\varphi} \in \mathscr{S}(\mathbb{R}^n)$ by $\check{\varphi}(x) := \varphi(-x)$. Notice that $\check{\varphi} = \hat{\hat{\varphi}}$, and also

$$\hat{\check{\varphi}}(\xi) = \frac{1}{(2\pi)^{n/2}}\int_{\mathbb{R}^n} e^{ix\cdot\xi}\varphi(x)dx = \frac{1}{(2\pi)^{n/2}}\int_{\mathbb{R}^n} e^{-iz\cdot\xi}\check{\varphi}(z)dz = \hat{\check{\varphi}}(\xi).$$

It is a corollary of (5.2.6) that $\hat{}: \mathscr{S}(\mathbb{R}^n) \to \mathscr{S}(\mathbb{R}^n)$ is bijective. Namely, $\hat{\varphi} = 0$ implies $\varphi = 0$, so it is injective, and $\varphi = \hat{\check{\varphi}}$, so it is surjective.

5.2.7 Lemma. *Let $\varphi, \psi \in \mathscr{S}(\mathbb{R}^n)$ be given. Then*

$$\int_{\mathbb{R}^n}\varphi\hat{\psi}dx = \int_{\mathbb{R}^n}\hat{\varphi}\psi dx.$$

PROOF: By (5.2.5), $\int_{\mathbb{R}^n} e^{ix\cdot\xi}\hat{\varphi}(\xi)\psi(\xi)d\xi = \int_{\mathbb{R}^n}\varphi(x+y)\hat{\psi}(y)dy$. Put $x=0$. □

5.2.8 Theorem (Parseval's relation). *Let $\varphi,\psi \in \mathscr{S}(\mathbb{R}^n)$ be given. Then*

$$\int_{\mathbb{R}^n}\varphi\bar{\psi}dx = \int_{\mathbb{R}^n}\hat{\varphi}\bar{\hat{\psi}}dx,$$

i.e. the Fourier transform is an isometry with respect to the L^2-norm.

PROOF: We claim that $\bar{\psi} = \hat{\bar{\hat{\psi}}}$, so $\int \varphi\bar{\psi}dx = \int \varphi\hat{\bar{\hat{\psi}}}dx = \int \hat{\varphi}\bar{\hat{\psi}}dx$.

$$\bar{\hat{\psi}}(\xi) = \frac{1}{(2\pi)^{n/2}}\overline{\int_{\mathbb{R}^n} e^{-ix\cdot\xi}\psi(x)dx} = \frac{1}{(2\pi)^{n/2}}\int_{\mathbb{R}^n} e^{ix\cdot\xi}\bar{\psi}(x)dx = \check{\bar{\psi}}(\xi)$$

Taking the Fourier transform of both sides proves the claim. □

5.2.9 Example. Given $f \in \mathscr{S}(\mathbb{R}^n)$, find $u \in \mathscr{S}(\mathbb{R}^n)$ such that $-\Delta u + u = f$, where Δ is the Laplacian operator. Put $P(\xi) = 1 + |\xi|^2$. Apply the Fourier transform to the equation to get

$$\begin{gathered} -\widehat{\Delta u} + \hat{u} = \hat{f} \\ (1+|\xi|^2)\hat{u}(\xi) = \hat{f}(\xi) \\ \text{i.e. } P\hat{u} = \hat{f} \end{gathered}$$

Now, $\hat{u} \in \mathscr{S}(\mathbb{R}^n)$ if and only if $\hat{f}/P \in \mathscr{S}(\mathbb{R}^n)$. But P is never zero so we can divide by it with no problem. Therefore there is a unique $u \in \mathscr{S}(\mathbb{R}^n)$ solving the equation, namely $u = (\hat{f}/P)\check{}$. The equation $-\Delta u = f$ cannot be solved in this manner because the resulting polynomial is zero at $\xi = 0$.

5.3 Tempered distributions

Topologize $\mathscr{S}(\mathbb{R}^n)$ as follows. For every $N \in \mathbb{N}\cup\{0\}$, define

$$|||\varphi|||_N := \sum_{|\alpha|,|\beta|\le N}\|P_\alpha D^\beta\varphi\|_\infty.$$

Define the metric $\rho : \mathscr{S}(\mathbb{R}^n)\times\mathscr{S}(\mathbb{R}^n) \to [0,\infty)$ by

$$\rho(\varphi,\psi) := \sum_{k=0}^{\infty}\frac{|||\varphi-\psi|||_k 2^{-k}}{1+|||\varphi-\psi|||_k}.$$

It can be shown that $(\mathscr{S}(\mathbb{R}^n),\rho)$ is complete and ρ is translation invariant.

5.3.1 Definition. A continuous linear mapping $\mu : \mathscr{S}(\mathbb{R}^n) \to \mathbb{K}$ is called a *tempered distribution*. We write $\langle\mu,\varphi\rangle$ for $\mu(\varphi)$ with $\varphi \in \mathscr{S}(\mathbb{R}^n)$, and let $\mathscr{S}'(\mathbb{R}^n)$ denote the set of all tempered distributions.

Remark.

(i) Notice that $\psi_k \to \psi$ in $\mathscr{S}(\mathbb{R}^n)$ if and only if $P_\beta D^\alpha \varphi_k \to P_\beta D^\alpha \psi$ uniformly on $\mathbb{R}^n$ for all multi-indices α and β.

(ii) The equation $-u''+u=0$ has solution $u(x)=c_1e^x+c_2e^{-x}$ for any $c_1,c_2\in\mathbb{R}$. But $c_1e^x+c_2e^{-x}\in\mathscr{S}'(\mathbb{R}^n)$ if and only if $c_1=c_2=0$.

Many linear operations that are defined on $\mathscr{S}$ can be extended to $\mathscr{S}'$ in a natural way, by thinking of the elements of $\mathscr{S}'$ as "integrating" elements of $\mathscr{S}$. Let $\mu\in\mathscr{S}'(\mathbb{R}^n)$ be given.

- Define $\hat{\mu}\in\mathscr{S}'(\mathbb{R}^n)$ by $\langle\hat{\mu},\varphi\rangle=\langle\mu,\hat{\varphi}\rangle$ for all $\varphi\in\mathscr{S}(\mathbb{R}^n)$.
- Define $\check{\mu}\in\mathscr{S}'(\mathbb{R}^n)$ by $\langle\check{\mu},\varphi\rangle=\langle\mu,\check{\varphi}\rangle$ for all $\varphi\in\mathscr{S}(\mathbb{R}^n)$.
- Define $P_\alpha\mu\in\mathscr{S}'(\mathbb{R}^n)$ by $\langle P_\alpha\mu,\varphi\rangle=\langle\mu,P_\alpha\varphi\rangle$ for all $\varphi\in\mathscr{S}(\mathbb{R}^n)$.
- Define $D^\alpha\mu\in\mathscr{S}'(\mathbb{R}^n)$ by $\langle D^\alpha\mu,\varphi\rangle=(-1)^{|\alpha|}\langle\mu,D^\alpha\varphi\rangle$ for all $\varphi\in\mathscr{S}(\mathbb{R}^n)$.

5.3.2 Example. For $\varphi\in\mathscr{S}(\mathbb{R}^n)$ define $L_\varphi\in\mathscr{S}'(\mathbb{R}^n)$ by $\langle L_\varphi,\psi\rangle:=\int_{\mathbb{R}^n}\varphi\psi\,dx$. Let $p\in[1,\infty)$ and $\mu\in L^p(\mathbb{R}^n)$ be given. Define $L_\mu\in\mathscr{S}'(\mathbb{R}^n)$ by

$$\langle L_\mu,\varphi\rangle:=\int_{\mathbb{R}^n}\mu(x)\varphi(x)dx.$$

For $\mu\in L^p(\mathbb{R}^n)$, we generally identify μ and L_μ. Notice that if $\langle L_\mu,\varphi\rangle=0$ for all $\varphi\in\mathscr{S}(\mathbb{R}^n)$ then $\mu=0$ a.e., so $L:L^p\to\mathscr{S}'(\mathbb{R}^n)$ is an injection.

Let $\mu\in L^1(\mathbb{R}^n)$ be given. Then

$$\langle L_\mu,\hat{\varphi}\rangle=\frac{1}{(2\pi)^{n/2}}\int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n}e^{-x\cdot\xi}\varphi(\xi)d\xi\right)\mu(x)dx=\langle L_{\hat{\mu}},\varphi\rangle.$$

Therefore the definition of Fourier transform for tempered distributions agrees with the definition given for L^1 functions at the beginning of the chapter. It can be checked that $\mu\in L^2(\mathbb{R}^n)$ if and only if $\hat{\mu}\in L^2(\mathbb{R}^n)$ for $L_\mu\in\mathscr{S}'(\mathbb{R}^n)$.

5.3.3 Theorem. *Let $\mu\in\mathscr{S}'(\mathbb{R}^n)$ and a multi-index α be given. Then*

(i) $\hat{\hat{\mu}}=\check{\mu}$;

(ii) $\check{\hat{\mu}}=\hat{\check{\mu}}$;

(iii) $D^\alpha\hat{\mu}=(-i)^{|\alpha|}\widehat{P_\alpha\mu}$; *and*

(iv) $\widehat{D^\alpha\mu}=i^{|\alpha|}P_\alpha\mu$.

PROOF: Exercise. □

5.3.4 Example (Delta function). Define $\delta_0\in\mathscr{S}'(\mathbb{R}^n)$ by $\delta_0(\varphi)=\varphi(0)$.

$$\langle\hat{\delta}_0,\varphi\rangle=\langle\delta_0,\hat{\varphi}\rangle=\hat{\varphi}(0)=\frac{1}{(2\pi)^{n/2}}\int_{\mathbb{R}^n}\varphi(x)dx$$

This final expression looks like "$(2\pi)^{-n/2}L_1$," but the constant function 1 is not an element of $\mathscr{S}(\mathbb{R}^n)$. It is, however, an element of $\mathscr{S}'(\mathbb{R}^n)$, and we write $\hat{\delta}_0=(2\pi)^{-n/2}\mathbf{1}$. Notice

$$\langle\check{\delta}_0,\varphi\rangle=\langle\delta_0,\check{\varphi}\rangle=\check{\varphi}(0)=\varphi(0)=\langle\delta_0,\varphi\rangle,$$

so by 5.3.3, $\hat{\mathbf{1}}=(2\pi)^{n/2}\hat{\hat{\delta}}_0=(2\pi)^{n/2}\delta_0$.

5.4 Convolution

5.4.1 Definition. Given $\varphi, \psi \in \mathscr{S}(\mathbb{R}^n)$, define the *convolution* of φ and ψ to be $\varphi * \psi : \mathbb{R}^n \to \mathbb{K}$, where

$$\varphi * \psi(x) = \int_{\mathbb{R}^n} \varphi(x-y)\psi(y)dy = \int_{\mathbb{R}^n} \varphi(y)\psi(x-y)dy.$$

Note that $\varphi * \psi \in \mathscr{S}(\mathbb{R}^n)$ and convolution is associative. For any multi-index α, $D^\alpha(\varphi * \psi) = (D^\alpha \varphi) * \psi = \varphi * (D^\alpha \psi)$, so the convloution is as smooth as its smoothest argument.

5.4.2 Lemma. *Let $\varphi, \psi \in \mathscr{S}(\mathbb{R}^n)$ be given. Then*
(i) $\widehat{\varphi * \psi} = (2\pi)^{n/2}\hat{\varphi}\hat{\psi}$*; and*
(ii) $\widehat{\varphi\psi} = (2\pi)^{-n/2}\hat{\varphi} * \hat{\psi}$.

Proof: By the inversion theorem we only need to get down to the nitty-gritty for one of the parts.

$$\begin{aligned}\widehat{\varphi * \psi}(\xi) &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{ix\cdot\xi} \left(\int_{\mathbb{R}^n} \varphi(y)\psi(x-y)dy \right) dx \\ &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} e^{i(x-y)\cdot\xi} e^{-iy\cdot\xi} \varphi(y)\psi(x-y)dydx \\ &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} e^{iz\cdot\xi} e^{-iy\cdot\xi} \varphi(y)\psi(z)dzdx \\ &= (2\pi)^{n/2}\hat{\varphi}(\xi)\hat{\psi}(\xi)\end{aligned}$$

For part (ii),

$$\widehat{\varphi\psi} = (2\pi)^{-n/2}\widehat{\widehat{\check{\hat{\varphi}} * \check{\hat{\psi}}}} = (2\pi)^{-n/2}(\check{\hat{\varphi}} * \check{\hat{\psi}})^\vee = (2\pi)^{-n/2}\hat{\varphi} * \hat{\psi} \qquad \square$$

To define the convolution on distributions, it will be convenient to introduce the *translation operator*. Given $h \in \mathbb{R}^n$ and $\varphi \in \mathscr{S}(\mathbb{R}^n)$, define $\tau_h\varphi \in \mathscr{S}(\mathbb{R}^n)$ by $(\tau_h\varphi)(x) := \varphi(x-h)$ for all $x \in \mathbb{R}^n$. Beware that some sources define the translation operator with a "+". Now,

$$\begin{aligned}\langle L_{\tau_h\mu}, \varphi\rangle &= \int_{\mathbb{R}^n} \tau_h\mu(x)\varphi(x)dx = \int_{\mathbb{R}^n} \mu(x-h)\varphi(x)dx \\ &= \int_{\mathbb{R}^n} \mu(x)\varphi(x+h)dx = \int_{\mathbb{R}^n} \mu(x)\tau_{-h}\varphi(x)dx = \langle L_\mu, \tau_{-h}\varphi\rangle,\end{aligned}$$

so for $\mu \in \mathscr{S}'(\mathbb{R}^n)$ we should define $\tau_h\mu \in \mathscr{S}'(\mathbb{R}^n)$ by $\langle \tau_h\mu, \varphi\rangle = \langle \mu, \tau_{-h}\varphi\rangle$ for all $\varphi \in \mathscr{S}(\mathbb{R}^n)$. For and $\varphi, \psi \in \mathscr{S}(\mathbb{R}^n)$,

$$(\varphi * \psi)(x) = \int_{\mathbb{R}^n} \varphi(y)\psi(x-y)dy = \int_{\mathbb{R}^n} \varphi(y)\check{\psi}(y-x)dy = \int_{\mathbb{R}^n} \varphi(y)\tau_x\check{\psi}(y)dy.$$

Therefore, for $\mu \in \mathscr{S}'(\mathbb{R}^n)$ and $\varphi \in \mathscr{S}(\mathbb{R}^n)$, we should define $\mu * \varphi : \mathbb{R}^n \to \mathbb{K}$ by $(\mu * \varphi)(x) = \langle \mu, \tau_x \check{\varphi} \rangle$ for all $x \in \mathbb{R}^n$. It can be shown that $\mu * \varphi \in C^\infty(\mathbb{R}^n)$, but there is no reason to expect that it is rapidly decreasing. It does however have at most polynomial growth, so it lives in $\mathscr{S}'$.

5.4.3 Lemma. *Let $\mu \in \mathscr{S}'(\mathbb{R}^n)$, $\varphi, \psi \in \mathscr{S}(\mathbb{R}^n)$, and a multi-index α be given.*

(i) $(\mu * \varphi) * \psi = \mu * (\varphi * \psi)$;
(ii) $D^\alpha(\mu * \varphi) = \mu * (D^\alpha \varphi) = (D^\alpha \mu) * \varphi$;
(iii) $(\mu * \psi)\hat{} = (2\pi)^{n/2} \hat{\psi}\hat{\mu}$; *and*
(iv) $\widehat{\psi\mu} = (2\pi)^{-n/2} \hat{\mu} * \hat{\psi}$.

Proof: We prove (iii) and leave the rest as exercises. Let $\varphi \in \mathscr{S}(\mathbb{R}^n)$ be given.

$$
\begin{aligned}
\langle (\mu * \psi)\hat{}, \varphi \rangle &= \langle \mu * \psi, \hat{\varphi} \rangle && \text{by definition of } \hat{} \\
&= \langle \mu * \psi, \tau_0 \check{\check{\hat{\varphi}}} \rangle && \text{trivially} \\
&= ((\mu * \psi) * \check{\hat{\varphi}})(0) && \text{by definition of } * \\
&= (\mu * (\psi * \check{\hat{\varphi}}))(0) && \text{by part (i)} \\
&= (\mu * (\check{\psi} * \hat{\varphi})\check{})(0) && \text{it can be “checked”} \\
&= \langle \mu, \check{\psi} * \hat{\varphi} \rangle && \text{as above} \\
&= \langle \mu, (2\pi)^{n/2} \widehat{\hat{\psi}\varphi} \rangle && \text{5.4.2} \\
&= (2\pi)^{n/2} \langle \hat{\mu}, \hat{\psi}\varphi \rangle \\
&= (2\pi)^{n/2} \langle \hat{\psi}\hat{\mu}, \varphi \rangle
\end{aligned}
$$

□

5.4.4 Example. The *fundamental solution* of a PDE is defined to be the solution with δ_0 on the right hand side, i.e. the solution to $Lu = \delta_0$, if there is a solution.

5.5 Sobolev spaces

Let $p \in [1, \infty]$ and a non-negative integer m be given. Define

$$W^{m,p}(\mathbb{R}^n) := \{u \in \mathscr{S}'(\mathbb{R}^n) : D^\alpha u \in L^p(\mathbb{R}^n), |\alpha| \le m\},$$

where $W^{0,p}(\mathbb{R}^n) = L^p$, and the norm

$$\|u\|_{m,p} = \begin{cases} \left(\sum_{|\alpha| \le m} \|D^\alpha u\|_{L^p}^p\right)^{1/p} & 1 \le p < \infty \\ \sum_{|\alpha| \le m} \|D^\alpha u\|_{L^\infty} & p = \infty \end{cases}$$

For $1 < p < \infty$, $(W^{m,p}, \|\cdot\|_{m,p})$ is uniformly convex and reflexive. Neither is true for $p = 1$ or $p = \infty$.

The case $p = 2$ is special. Note that $u \in W^{m,2}(\mathbb{R}^n)$ if and only if $Q_{m/2}\hat{u} \in L^2(\mathbb{R}^n)$, where $Q_s = (1 + |x|^2)^s$. Indeed, the Fourier transform is a bijection from

L^2 to L^2 and $\widehat{D^\alpha u} = i^{|\alpha|} P_\alpha \hat{u}$ for all multi-indices α. An equivalent inner product on $W^{m,2}(\mathbb{R}^n)$ is given by

$$(u,v)_m := \int_{\mathbb{R}^n} Q_m(\xi)\hat{u}(\xi)\bar{\hat{v}}(\xi)d\xi.$$

For $s \in \mathbb{R}$, define $H^s(\mathbb{R}^n) := \{u \in \mathscr{S}'(\mathbb{R}^n) : Q_{s/2}\hat{u} \in L^2(\mathbb{R}^n)\}$, with inner product $(\cdot,\cdot)_s$. Notice that $H^m(\mathbb{R}^n) = W^{m,2}(\mathbb{R}^n)$ for non-negative integers m.

5.5.1 Theorem (Sobolev Embedding, special case).
Let C_v denote the collection of all continuous functions $f : \mathbb{R}^n \to \mathbb{K}$ such that $\sup|f(x)| \to 0$ as $|x| \to \infty$, equipped with the norm $\|\cdot\|_\infty$. Then for all real $s > n/2$, $H^s(\mathbb{R}^n) \hookrightarrow C_v(\mathbb{R}^n)$.

Proof: Let $u \in H^s(\mathbb{R}^n)$, so that $Q_{s/2}\hat{u} \in L^2(\mathbb{R}^n)$. Write $\hat{u} = (Q_{s/2}\hat{u})/Q_{s/2}$.

$$\check{u} = \hat{\hat{u}} = \left(\frac{Q_{s/2}\hat{u}}{Q_{s/2}}\right)^{\wedge}$$

so it suffices to show that $1/Q_{s/2} \in L^2$, for then $\hat{u} \in L^1(\mathbb{R}^n)$, which implies that $\check{u} \in C_v$ by 5.2.2. Changing to polar coördinates,

$$\int_{\mathbb{R}^n} \left(\frac{1}{Q_{s/2}}\right)^2 dx = \int_{\mathbb{R}^n} \frac{1}{(1+|x|)^s} dx = \int_0^\infty \frac{1}{(1+r^2)^s} r^{n-1} dr < \infty$$

if and only if $s > n/2$. □

5.5.2 Example (Heat equation). The *heat equation* is $u_t(x,t) = \Delta u(x,t)$, for $x \in \mathbb{R}^n$ and $t \geq 0$, with initial condition $u(x,0) = u_0(x)$ for $x \in \mathbb{R}^n$. Our basic space is $X = L^2(\mathbb{R}^n)$. Formally, we want an operator $A : \mathscr{D}(A) \to X$ such that $Av = \Delta v$ for all $v \in \mathscr{D}(A)$. If $I - A$ is surjective and A is dissipative then A generates a contractive linear C_0-semigroup (X is reflexive so we do not need to worry about the domain being dense to apply the Lumer-Phillips theorem.)

Given $f \in X = L^2(\mathbb{R}^n)$, we have to find $u \in \mathscr{D}(A)$ such that $u - \Delta u = f$. Put $P(\xi) = 1 + |\xi|^2 = Q_1$. Then, taking the Fourier transform, $P\hat{u} = \hat{f}$, so $\hat{u} = \hat{f}/P$. Therefore there is such a u and it lives in $H^2(\mathbb{R}^n)$. We take $\mathscr{D}(A) = H^2(\mathbb{R}^n)$, and $I - A$ is a bijection between $\mathscr{D}(A)$ and $L^2(\mathbb{R}^n)$.

Looking to the inner product as the obvious semi-inner product, we need to know that $(Av, v) \leq 0$ for all $v \in \mathscr{D}(A)$. We could integrate by parts

$$(Av, v) = \int_{\mathbb{R}^n} \Delta u(x)\bar{u}(x)dx = -\int_{\mathbb{R}}^n u^2(x)dx \leq 0,$$

or we could also use Parseval's relation,

$$(\widehat{Av}, \hat{v}) = \int_{\mathbb{R}^n} \widehat{\Delta u}\bar{\hat{u}} d\xi = -\int_{\mathbb{R}^n} |\xi|^2 \hat{u}\bar{\hat{u}} dx \leq 0.$$

Thus, by the Lumer-Phillips theorem, A generates a contractive linear C_0-semigroup, and for all $u_0 \in \mathscr{D}(A)$, the mapping $t \mapsto T(t)u_0$ is differentiable on $[0,\infty)$ and $\frac{d}{dt}(T(t)u_0) = AT(t)u_0$.

The results here are not optimal. In fact, given $u_0 \in L^2(\mathbb{R}^n)$, it can be shown $T(t)u_0 \in \bigcap_{n=1}^{\infty} \mathscr{D}(A^n)$ for all $t > 0$ and $t \mapsto T(t)$ is analytic on $[0,\infty)$ in the uniform operator topology. To get these results one would have to develop the theory of analytic contractive semigroups.

5.5.3 Example (Wave equation). The *wave equation* is $u_{tt}(x,t) = \Delta u(x,t)$, for $x \in \mathbb{R}^n$ and $t \geq 0$, with initial conditions $u(x,0) = u_0(x)$ and $u_t(x,0) = v_0(x)$ for $x \in \mathbb{R}^n$. Walter Litmann 1967 showed that if $n > 2$ then the wave operator generates a C_0-semigroup in L^p if and only if $p = 2$, so the following results are optimal.

For now we take $\mathbb{K} = \mathbb{R}$. Formally multiply the equation by u_t and integrate over $\mathbb{R}^n$ to get

$$\begin{aligned}\int_{\mathbb{R}^n} u_t u_{tt}\,dx &= \int_{\mathbb{R}^n} (\Delta u) u_t\,dx \\ \frac{1}{2}\frac{d}{dt}\int_{\mathbb{R}^n}(u_t)^2 dx &= -\int_{\mathbb{R}^n} \nabla u \cdot \nabla u_t\,dx = -\frac{1}{2}\frac{d}{dt}\int_{\mathbb{R}^n} |\nabla u|^2 dx \\ 0 &= \frac{1}{2}\frac{d}{dt}\int_{\mathbb{R}^n}(u_t^2 + |\nabla u|^2)dx.\end{aligned}$$

This is referred to as *conservation of energy*. We have, for all $t \geq 0$,

$$\frac{1}{2}\int_{\mathbb{R}^n}(u_t^2 + |\nabla u|^2)(x,t)dx = \frac{1}{2}\int_{\mathbb{R}^n}((v_0(x))^2 + |\nabla u_0(x)|^2)dx =: \frac{\varepsilon_0}{2}.$$

Note that $\int_{\mathbb{R}^n} \frac{1}{2}(u(x))^2 dx$ is conspicuously missing. When n is odd, it can be shown that $\int_{\mathbb{R}^n} \frac{1}{2} u_t^2 dx \to \frac{1}{2}\varepsilon_0$ and $\int_{\mathbb{R}^n} \frac{1}{2}|\nabla u(x)|^2 dx \to \frac{1}{2}\varepsilon_0$ as $t \to \infty$. This is referred to as *equipartition of energy*.

Let $X = H^1(\mathbb{R}^n) \times L^2(\mathbb{R}^n)$. We write the wave equation as the system

$$\begin{aligned}\dot{u}_1 &= u_2 \\ \dot{u}_2 &= \Delta u_2.\end{aligned}$$

Formally, we define the operator $A : \mathscr{D}(A) \to X$ by $A(\varphi_1, \varphi_2) = (\varphi_2, \Delta\varphi_1)$ (the parentheses represent ordered pairs, not inner products). Then the system may be written $(\dot{u}_1, \dot{u}_2) = A(u_1, u_2)$. We would like to see that A generates a linear C_0-semigroup. Equip X with the inner product

$$\langle(\varphi_1,\varphi_2),(\psi_1,\psi_2)\rangle = \int_{\mathbb{R}^n} \varphi_1\psi_1 + \nabla\varphi_1 \cdot \nabla\psi_1 + \varphi_2\psi_2\,dx.$$

Based on the formal computation, we would expect the term first term in the inner product to "cause issues" in what follows. The domain of A must be $\mathscr{D}(A) = H^2(\mathbb{R}^n) \times H^1(\mathbb{R}^n)$ since we need A to map into X.

We must show that A is dissipative. Take $(\varphi_1, \varphi_2) \in \mathscr{D}(A)$ and consider

$$\begin{aligned}\langle A(\varphi_1, \varphi_2), (\varphi_1, \varphi_2)\rangle &= \langle (\varphi_2, \Delta\varphi_1), (\varphi_1, \varphi_2)\rangle \\ &= \int_{\mathbb{R}^n} \varphi_2\varphi_1 + \nabla\varphi_2 \cdot \nabla\varphi_1 + (\Delta\varphi_1)\varphi_2 dx = \int_{\mathbb{R}^n} \varphi_2\varphi_1 dx\end{aligned}$$

by parts or by Parseval's relation. (Note that if the domain were bounded then we would have Poincaré's inequality and be done.) Finally,

$$\int_{\mathbb{R}^n} \varphi_2\varphi_1 dx \le \frac{1}{2}\int_{\mathbb{R}^n} \varphi_1^2 + \varphi_2^2 dx \le \frac{1}{2}\|(\varphi_1, \varphi_2)\|^2,$$

so A is quasi-dissipative with $\omega = \frac{1}{2}$ (this is not optimal). Now we must show there is $\lambda_0 > \frac{1}{2}$ such that $\lambda_0 I - A$ is surjective. We may as well take $\lambda_0 = 1$ and consider the system $(\varphi_1, \varphi_2) - A(\varphi_1, \varphi_2) = (f_1, f_2)$, i.e.

$$\begin{aligned}\varphi_1 - \varphi_2 &= f_1 \\ \varphi_2 - \Delta\varphi_1 &= f_2.\end{aligned}$$

Solving, we need $\varphi_1 - \Delta\varphi_1 = f_2 + f_1$, and we've already seen that there is exactly one $\varphi_1 \in H^2(\mathbb{R}^n)$ that satisfies this equation. Therefore A generates a quasi-contractive linear C_0-semigroup $\{T(t)\}_{t\ge 0}$.

From the homework, for all $m \in \mathbb{N}$, $\mathscr{D}(A^m)$ is invariant under $T(t)$ for all $t \ge 0$. Therefore if $u_0 \in H^{m+1}(\mathbb{R}^n)$ and $v_0 \in H^m(\mathbb{R}^n)$ then the solution $(u(\cdot, t), u_t(\cdot, t))$ lives in $H^{m+1}(\mathbb{R}^n) \times H^m(\mathbb{R}^n)$ for all $t \ge 0$.

6 Non-linear Operators

6.1 Monotone operators

6.1.1 Definition. Let X be a real Banach space and X^* be its dual space. We say that a (not-necessarily linear) mapping $F : X \to X^*$ is

(i) a *monotone mapping* provided $\langle F(u) - F(v), u - v\rangle \ge 0$ for all $u, v \in X$;

(ii) a *strictly monotone mapping* provided $\langle F(u) - F(v), u - v\rangle > 0$ for all $u, v \in X$ with $u \ne v$;

(iii) a *bounded mapping* provided $F[B]$ is bounded in X^* for all bounded subsets B of X;

(iv) *hemi-continuous* provided the mapping $\lambda \mapsto \langle F(u + \lambda v), w\rangle$ is continuous as a function from $\mathbb{R}$ to $\mathbb{R}$ for all $u, v, w \in X$;

(v) *coercive* provided $\langle F(u), u\rangle / \|u\| \to \infty$ as $\|u\| \to \infty$.

This definition of coercivity is a slight generalization of the definition of coercivity from the Lax-Milgram theorem.

6.1.2 Lemma. *Let $m \in \mathbb{N}$ be given and assume that $P : \mathbb{R}^m \to \mathbb{R}^m$ is continuous and satisfies $\xi \cdot P(\xi) \geq 0$ for all $\xi \in \mathbb{R}^m$ with $|\xi| = \rho$. Then there is $\eta \in \mathbb{R}^M$ with $|\eta| \leq \rho$ such that $P(\eta) = 0$.*

PROOF: Suppose there is no such η. Define $\Psi(\xi) = -\rho P(\xi)/|P(\xi)|$, so that Ψ is continuous and maps B_ρ, the closed ball of radius ρ, to itself. By Brouwer's Fixed Point theorem Ψ has a fixed point $\xi_0 \in B_\rho$. Then $|\xi_0| = |\Psi(\xi_0)| = \rho$ and

$$0 \leq \xi_0 \cdot P(\xi_0) = -\rho |P(\xi_0)| < 0. \qquad \square$$

6.1.3 Theorem. *Let X be a separable, reflexive Banach space and assume that $F : X \to X^*$ is monotone, bounded, hemi-continuous, and coercive. Then F is surjective.*

PROOF: Let $f \in X^*$ be given. We must find $u \in X$ such that $F(u) = f$. Assume that X is finite dimensional. Since X is separable, we may choose a linearly independent sequence $\{v_m\}_{m=1}^\infty$ such that the linear span of $\{v_m\}_{m=1}^\infty$ is dense in X. For $m \in \mathbb{N}$ put $V_m := \operatorname{span}\{v_1, \ldots, v_m\}$.

Fix $m \in \mathbb{N}$. We look for $u_m \in V_m$ such that $\langle F(u_m), v_i \rangle = \langle f, v_i \rangle$ for all $i = 1, \ldots, m$. Write $u_m = \sum_{j=1}^m \xi_j v_j$ and solve for the required $\xi \in \mathbb{R}^m$. Define a mapping $P : \mathbb{R}^m \to \mathbb{R}^m$ by $P_i(\xi) = \langle F(u_m) - f, v_i \rangle$. P is continuous because F is hemi-continuous. Then

$$\begin{aligned} \xi \cdot P(\xi) &= \sum_{i=1}^m \xi_i P_i(\xi) = \langle F(u_m) - f, u_m \rangle \\ &= \langle F(u_m), u_m \rangle - \langle f, u_m \rangle \\ &\geq \|u_m\| \left(\frac{\langle F(u_m), u_m \rangle}{\|u_m\|} - \|f\|_{X^*} \right). \end{aligned}$$

Choose $\rho > 0$ (independent of m) such that $\xi \cdot P(\xi) \geq 0$ for all $\xi \in \mathbb{R}^m$ with $|\xi| = \rho$. This can be done because F is coercive. By the lemma, P has a zero in B_ρ, and this zero is the ξ giving the desired u_m.

Generate a sequence $\{u_m\}_{m=1}^\infty$ in this way with $\|u_m\| \leq \rho$ for all $m \in \mathbb{N}$. Then $\{F(u_m)\}_{m=1}^\infty$ is bounded because F is bounded. Bounded sequences have weakly convergence subsequences because X is reflexive. Without loss of generality we may assume that $\{u_m\}_{m=1}^\infty$ is weakly convergent in X and $\{F(u_m)\}_{m=1}^\infty$ is weak* (weakly) convergent in X^*. Let u be the weak limit of $\{u_m\}_{m=1}^\infty$ in X and let φ be the weak limit of $\{F(u_m)\}_{m=1}^\infty$ in X^*. We will show that $\varphi = f$ and $F(u) = \varphi$.

$\langle F(u_m), v \rangle = \langle f, v \rangle$ for all $v \in V_m$, so $\langle \varphi, v \rangle = \langle f, v \rangle$ for all $v \in \bigcup_{m=1}^\infty V_m$. This set is dense, so $\varphi = f$. By monotonicity, for all $m \in \mathbb{N}$ and $v \in X$,

$$\begin{aligned} 0 &\le \langle F(v) - F(u_m), v - u_m \rangle \\ &= \langle F(v), v\rangle - \langle F(v), u_m\rangle - \langle F(u_m), v\rangle + \langle F(u_m), u_m\rangle \\ &= \langle F(v), v\rangle - \langle F(v), u_m\rangle - \langle F(u_m), v\rangle + \langle f, u_m\rangle \\ &\to \langle F(v), v\rangle - \langle F(v), u\rangle - \langle f, v\rangle + \langle f, u\rangle \\ \text{so } 0 &\le \langle F(v) - f, v - u\rangle. \end{aligned}$$

Let $w \in X$ be given and put $v = u + \lambda w$ in the inequality. For all $\lambda > 0$, $\langle F(u + \lambda w) - f, w\rangle \ge 0$, so letting $\lambda \downarrow 0$ we see that $\langle F(u) - f, w\rangle \ge 0$ by hemi-continuity, and this holds for all $w \in X$. Therefore $F(u) = f$. □

6.2 Differentiability

6.2.1 Definition. Let X and Y be Banach spaces. $F : X \to Y$ is said to be *Fréchet differentiable* at $x_0 \in X$ provided there is $L \in \mathscr{L}(X; Y)$ such that

$$\lim_{\substack{h \in X \\ \|h\| \downarrow 0}} \frac{F(x_0 + h) - F(x_0) - Lh}{\|h\|} = 0.$$

We write $F'(x_0) := \nabla F(x_0) := DF(x_0) := L$. We say that F *Fréchet differentiable* if it is Fréchet differentiable at each $x_0 \in X$.

Remark. Fréchet differentiability implies continuity in norm.

6.2.2 Proposition. *Let Z be a third Banach space and let $F : X \to Y$ and $G : Y \to Z$ and $x_0 \in X$ be given. Assume that F is Fréchet differentiable at x_0 and G is Fréchet differentiable at $F(x_0)$. Then $G \circ F$ is Fréchet differentiable at x_0 and $(G \circ F)'(x_0) = G'(F(x_0))F'(x_0)$.*

PROOF: Using "little-o" notation,

$$\begin{aligned} F(x_0 + h) &= F(x_0) + F'(x_0)h + o(\|h\|) \\ G(F(x_0 + h)) &= G(F(x_0) + F'(x_0)h + o(\|h\|)) \\ &= G(F(x_0)) + G'(F(x_0))(F'(x_0)h + o(\|h\|)) + o(\|h\|) \\ &= G(F(x_0)) + G'(F(x_0))F'(x_0)h + o(\|h\|)) \end{aligned}$$

which proves the result. □

6.2.3 Definition. Let $F : X \to Y$ and $x_0, v \in X$ be given. We say that F has a *Gâteaux variation* in the direction v provided the following limit exists.

$$\delta F(x_0; v) := \lim_{t \to 0} \frac{F(x_0 + tv) - F(x_0)}{t}$$

We say that F is *Gâteaux differentiable* at x_0 provided it has a Gâteaux variation in each direction $v \in X$.

Remark. $\delta F(x_0; v)$ need not be linear in v, i.e. it need not be the case that $\delta F(x_0; v) = L(x_0)v$ for some $L \in \mathscr{L}(X; Y)$. Linearity in v might be part of the definition in some books.

6.2.4 Example. Gâteaux differentiability does not imply continuity, even in finite dimensions. Let $S = \{x \in \mathbb{R}^2 \mid 2x_1^2 > x_2 > x_1^2\}$ and consider the characteristic function of S. $\mathbf{1}_S(0,0) = 0$, and given any line through the origin there is a neighbourhood of the origin such that, in that neighbourhood, that line does not intersect S. Therefore $\mathbf{1}_S$ is Gâteaux differentiable at $(0,0)$ in every direction with derivative 0, but $\mathbf{1}_S$ is not continuous at $(0,0)$.

6.2.5 Proposition. *If F is Fréchet differentiable at x_0 then F is Gâteaux differentiable at x_0 and $\delta F(x_0; v) = F'(x_0)v$ for all $v \in X$.*

6.2.6 Theorem. *Let $F : X \to Y$ and $x_0 \in X$ be given. Assume that there is an open set U containing x_0 such that*

(i) F is Gâteaux differentiable at each $x \in U$;

(ii) for all $x \in U$ there is $L(x) \in \mathscr{L}(X; Y)$ such that $\delta F(x; v) = L(x)v$ for all $v \in X$; and

(iii) $x \mapsto L(x)$ is continuous from U to $\mathscr{L}(X; Y)$.

Then F is Fréchet differentiable at x_0 and $F'(x_0)h = \delta F(x_0; h)$ for all $h \in X$.

PROOF: Choose $\delta > 0$ such that $B_\delta(x_0) \subseteq U$ and let $h \in B_\delta(0)$ be given. Define $\psi : [0,1] \to X$ by $\psi(t) = F(x_0 + th)$, so that $\psi(0) = F(x_0)$ and $\psi(1) = F(x_0 + h)$. We need to show that $\psi(1) - \psi(0) - L(x_0)h$ is $o(\|h\|)$. By (i) and (ii), $\psi'(t) = \delta F(x_0 + th; h) = L(x_0 + th)h$, which is continuous by (iii). Therefore, by the fundamental theorem of calculus,

$$F(x_0 + h) - F(x_0) - L(x_0)h = \int_0^1 (\psi'(t) - L(x_0)h)dt$$

$$\|F(x_0 + h) - F(x_0) - L(x_0)h\| \le \left(\int_0^1 \|L(x_0 + th) - L(x_0)\| dt \right) \|h\|$$

This goes to zero as $\|h\| \to 0$ since $x \mapsto L(x)$ is continuous. □

6.2.7 Theorem (Inverse function theorem).
Assume that $F : X \to Y$ is Fréchet differentiable on X and that $x \mapsto F'(x)$ is continuous. Let $x_0 \in X$ be given and assume that $F'(x_0)$ is bijective. Then there are $U \subseteq X$ and $V \subseteq Y$, both open, such that $x_0 \in U$, $F(x_0) \in V$, and $F|_U^V$ is bijective. In fact, the inverse is continuously Fréchet differentiable.

Remark. Moreover, there is $\eta > 0$ such that, for all $y \in B_\eta(F(x_0))$, the sequence $\{x_n\}_{n=1}^\infty$ generated by $x_{n+1} := x_n + [F'(x_0)]^{-1}(y - F(x_n))$ converges to the unique solution of $y = F(x)$ in U.

PROOF: Without loss of generality we may assume $x_0 = 0$ and $F(x_0) = 0$. Let $y \in Y$ be given. We will find conditions on y that will allow us to solve the equation $F(x) = y$. Note that $F(x) = F'(0)x + R(x)$, where

$$R(x) = F(x) - F'(0)x = o(\|x\|).$$

Hence, if x solves $F(x) = y$ then

$$\begin{aligned} F'(0)x &= y - R(x) \\ x &= (F'(0))^{-1}(y - (F(x) - F'(0)x)). \end{aligned}$$

Define $T_y(x) := (F'(0))^{-1}(y - (F(x) - F'(0)x))$. We will show that if $\|y\|$ is sufficiently small then T_y maps a closed ball into itself and T_y is a strict contraction. Let $M := \|(F'(0))^{-1}\|$. Since $\|F(x) - F'(0)x\|$ decreases super-linearly in $\|x\|$, there is $\varepsilon > 0$ such that $\|F(x) - F'(0)x\| \leq \frac{\varepsilon}{2M}$ when $\|x\| < \varepsilon$. Assume that $\|y\| < \frac{\varepsilon}{2M}$, so that

$$\|T_y(x)\| \leq M\|y\| + M\|F(x) - F'(0)x\| < \varepsilon$$

Therefore there is $\varepsilon > 0$ such that $T_y : \overline{B}_\varepsilon(0) \to \overline{B}_\varepsilon(0)$. Further, T_y is Fréchet differentiable and $T_y'(x) = -(F'(0))^{-1}(F'(x) - F'(0))$. This goes to zero as $\|x\| \to 0$ since we have assumed that $x \mapsto F'(x)$ is continuous. Further restrict ε so that $\|T_y'(x)\| \leq \frac{1}{2}$ for all $x \in \overline{B}_\varepsilon(0)$. Notice that

$$\begin{aligned} T_y(x) - T_y(z) &= \int_0^1 T_y'(z + t(x - z))(x - z)dt \\ \|T_y(x) - T_y(z)\| &= \left(\int_0^1 \|T_y'(z + t(x - z))\| dt \right) \|x - z\| \leq \frac{1}{2}\|x - z\|. \end{aligned}$$

Therefore T_y is a strict contraction. By the contractive mapping theorem T_y has a fixed point, which by construction solves $F(x) = y$. □

6.3 Convexity

6.3.1 Definition. A function $f : X \to \mathbb{R} \cup \{+\infty\}$ is said to be a *proper function* if there is $x_0 \in X$ such that $f(x_0) < \infty$. If f is convex and proper then the *subdifferential* or *subgradient* of f at $x_0 \in X$ is defined to be

$$\partial f(x_0) := \{x^* \in X^* \mid f(x) \geq f(x_0) + x^*(x - x_0) \text{ for all } x \in X\}.$$

6.3.2 Example. Take $X = \mathbb{R}$ and $f(x) = |x|$. Then

$$\partial f(x) = \begin{cases} \{-1\} & x < 0 \\ [-1, 1] & x = 0 \\ \{1\} & x > 0. \end{cases}$$

Remark.
(i) It can happen that $\partial f(x_0) = \varnothing$. Example?
(ii) If f is convex and proper then f attains a minimum at x_0 if and only if $0 \in \partial f(x_0)$.

From here on assume that $\mathbb{K} = \mathbb{R}$.

6.3.3 Theorem. *Assume that $f : X \to \mathbb{R}$ is continuous and convex and let $x_0 \in X$ be given. Then $\partial f(x_0) \neq \varnothing$.*

PROOF: Use the Hahn-Banach theorem. Put $E := \{(x, \lambda) \mid x \in X, f(x) < \lambda\}$, i.e. the epigraph of f. Notice that E is nonempty, open, and convex. Separate E from $\{(x_0, f(x_0))\}$ by applying a strong form of the separating hyperplane theorem. Choose $(x^*, \alpha) \in X^* \times \mathbb{R}$ and $\beta \in \mathbb{R}$ such that
(i) $\|x^*\| + |\alpha| > 0$, i.e. $(x^*, \alpha) \neq (0, 0)$;
(ii) $x^*(x) + \alpha\lambda \geq \beta$ whenever $\lambda > f(x)$; and
(iii) $x^*(x_0) + \alpha f(x_0) \leq \beta$.
It cannot be the case that $\alpha = 0$ because then we would have $x^*(x) \geq \beta$ for all $x \in X$, which would imply that $x^* = 0$ as well, contradicting (i). Further, $\alpha > 0$ because otherwise we could take $\lambda \to \infty$ and contradict (ii). Put $y^* = -\frac{1}{\alpha}x^*$. Then $f(x) \geq \frac{\beta}{\alpha} + y^*(x)$ for all $x \in X$ by (ii) and so $f(x_0) = \frac{\beta}{\alpha} + y^*(x_0)$ by (iii). Subtracting, $f(x) \geq f(x_0) + y^*(x - x_0)$, so $y^* \in \partial f(x_0)$. □

6.3.4 Theorem. *Assume that $f : X \to \mathbb{R}$ is convex and continuous, and let $x_0 \in X$ be given. The following are equivalent.*
(i) $\partial f(x_0)$ is a singleton.
(ii) f is Gâteaux differentiable at x_0 and there is $L(x_0) \in X^*$ such that $\delta f(x_0; v) = L(x_0)v$ for all $v \in X$.

PROOF: Assume (ii) holds. Let $v \in X$ be given and put $\psi(t) = f(x_0 + tv)$. Then ψ is continuous, convex, and differentiable at 0, with

$$\psi'(0) = \delta f(x_0; v) = L(x_0)v.$$

Since ψ is convex,

$$\begin{aligned} \psi(1) &\geq \psi(0) + \psi'(0)1 \\ \text{so } f(x_0 + v) &\geq f(x_0) + L(x_0)v \end{aligned}$$

Therefore $L(x_0) \in \partial f(x_0)$. Let $z^* \in \partial f(x_0)$ be given and $y \in X$.

$$\begin{aligned} f(x_0 + ty) &\geq f(x_0) + z^*(ty) = f(x_0) + tz^*(y) \\ \text{and } f(x_0 + ty) &= f(x_0) + tL(x_0)y + o(t) \end{aligned}$$

Combining these,

$$tL(x_0)y + o(t) \geq tz^*(y)$$
$$L(x_0)y + \frac{o(t)}{t} \begin{cases} \geq z^*(y) & \text{if } t > 0 \\ \leq z^*(y) & \text{if } t < 0 \end{cases}$$

Letting $t \to 0$ we conclude that $L(x_0)y = z^*(y)$. □

6.3.5 Proposition. *Assume that $f : X \to \mathbb{R}$ is Fréchet differentiable, and notice that $f' : X \to X^*$. The following are equivalent.*
(i) f is convex.
(ii) $f(u) \geq f(v) + \langle f'(v), u - v\rangle$ for all $u, v \in X$.
(iii) $\langle f'(u) - f'(v), u - v\rangle \geq 0$ for all $u, v \in X$, i.e. f' is monotone.

Proof: Assume that f is convex. Let $u, v \in X$ be given. Then for all $t \in (0, 1]$,

$$f(tu + (1-t)v) \leq tf(u) + (1-t)f(v)$$
$$f(v + t(u-v)) \leq f(v) + t(f(u) - f(v))$$
$$\frac{f(v + t(u-v)) - f(v)}{t} \leq f(u) - f(v)$$

so, taking $t \downarrow 0$, $\langle f'(v), u - v\rangle \leq f(u) - f(v)$. The rest of the implications are left as exercises. □

Index