OPTIMISATION NOTES following
OPTIMISATION BY VECTOR SPACE METHODS by David G. Leunberger

Nick Cao

1 Hilbert Spaces

Recall that an inner product $(\cdot|\cdot)$ satisfies:

1. $(x|y) = \overline{(y|x)}$
2. $(x+y|z) = (x|z) + (y|z)$ and $(\lambda x|y) = \lambda(x|y)$
3. $(x|x) \geq 0$ with equality iff $x = 0$

Some basic facts include:

- $\big|(x|y)\big| \leq \|x\|\|y\|$ with equality iff $x = \lambda y$ or $y = 0$ *(Cauchy Schwarz Inequality)*
- $\sqrt{(x|x)} = \|x\|$ defines a norm
- $(x|y) = 0$ for all y implies $x = 0$, the proof is to set $y = x$ and use property (3) in the definition
- $\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2$ *(Parallelogram Law)*, the proof requires expansion
- A *Hilbert Space* is a complete inner product space

Lemma 1 (Continuity of the Inner Product)**.**

Proof. Let $x_n \to x$ and $y_n \to y$. Then

$$\begin{aligned}\big|(x_n|y_n) - (x|y)\big| &= \big|(x_n|y_n) - (x_n|y) + (x_n|y) - (x|y)\big| \\ &\leq \big|(x_n|y_n - y)\big| + \big|(x_n - x|y)\big| \\ &\leq \|x_n\|\|y_n - y\| + \|x_n - x\|\|y\| \quad \text{(C-S)} \\ &\leq M\|y_n - y\| + \|x_n - x\|\|y\| \to 0.\end{aligned}$$

$\square$

1.1 Projection

If $(x|y) = 0$, then $x \perp y$. Further, $\|x+y\|^2 = \|x\|^2 + \|y\|^2$.

Theorem 1. *Let X be an inner product space, M a subspace of X, $x \in X$. If there is a vector $m_0 \in M$ such that $\|x - m_0\|^2 \leq \|x - m\|$ for all $m \in M$, then m_0 is unique. Further, $x - m_0 \in M^\perp$ necessarily; this is also sufficient.*

Such a projection m_0 is guaranteed to exist if X is complete (ie, a Hilbert space) and M is closed.

1.2 Orthogonal Complements

Define the orthogonal complement $S^\perp = \{x \in X : x \perp s, \ \forall s \in S\}$. $S^\perp$ is necessarily a closed subspace (closed by the continuity of the inner product). Further:

- $S \subset T$ implies $T^\perp \subset S^\perp$
- $S \subset \overline{[S]} = S^{\perp\perp}$

Define the *direct sum*: $X = M \oplus N$ if for all $x \in X$, there exist unique $m \in M$ and $n \in N$ such that $X = M + N$.

Theorem 2. *If M is a closed linear subspace of a Hilbert space H, then $H = M \oplus M^\perp$ and*

$M = M^{\perp\perp}$.

Proof. That $H = M + M^{\perp}$ follows from the projection theorem. For uniqueness, if $x = m_0 + n_0 = m_1 + n_1$, then $m_0 - m_1 + n_0 - n_1 = 0$. As $m_0 - m_1 \in M$ and $n_0 - n_1 \in M^{\perp}$, Pythagoras's Theorem states that $0 = m_0 - m_1 = n_0 - n_1$ so the representation is unique. □

Proposition 1. *An orthogonal set of nonzero vectors is a LI set.*

Proof. Let $\{x_1, \dots, x_n\}$ be a finite subset of the orthogonal set. Take scalars $\alpha_1, \dots, \alpha_n$ such that $\sum_{i=1}^n \alpha_i x_i = 0$. Then

$$\sum_{i=1}^n \alpha_i (x_i | x_k) = \left(\sum_{i=1}^n \alpha_i x_i \middle| x_k \right) = (0 | x_k) = 0.$$

Hence, as $x_k \neq 0$, $(x_k | x_k) > 0$ and $\alpha_k = 0$. Inducting (relying on AC if set is infinite), $\alpha_i = 0$ for all i. □

Theorem 3 (Gram-Schmidt)**.** *Let $\{x_i\}$ be a finite or countable sequence of LI vectors in inner product space X. Then there exists orthonormal sequence $\{e_i\}$ such that*

$$[e_1, \dots, e_n] = [x_1, \dots, x_n], \quad \forall n \in \mathbf{N}.$$

Proof. Set $e_1 = x_1 / \|x_1\|$, $z_n = x_n - \sum_{i=1}^{n-1} (x_n | e_i) e_i$, and $e_n = z_n / \|z_n\|$. □

1.3 Approximations

Say if we want to project $x \in H$ into $[y_1, \dots, y_n]$, a closed subspace of H. Denote the projection as $\alpha_1 y_1 + \cdots + \alpha_n y_n$. By the projection theorem,

$$\big(x - \alpha_1 y_1 - \cdots - \alpha_n y_n \big| y_i \big) = 0$$

for all $i = 1, \dots, n$. Arranging, we have:

$$\begin{aligned} (y_1|y_1)\alpha_1 + \cdots + (y_n|y_1)\alpha_n &= (x|y_1) \\ &\vdots \\ (y_1|y_n)\alpha_1 + \cdots + (y_n|y_n)\alpha_n &= (x|y_n) \end{aligned}$$

and thus

$$\underbrace{\begin{bmatrix} (y_1|y_1) & \cdots & (y_n|y_1) \\ \vdots & \ddots & \vdots \\ (y_1|y_n) & \cdots & (y_n|y_n) \end{bmatrix}}_{\text{Transpose of Gram matrix}} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} (x|y_1) \\ \vdots \\ (x|y_n) \end{bmatrix}$$

The determinant of the Gram matrix $\mathbf{G}$ is denoted $g(y_1, \dots, y_n)$.

Proposition 2. *$g(y_1, \dots, y_n) \neq 0$ iff $y_1, \dots, y_n$ are LI.*

Proof. We shall prove the contrapositive: $g(y_1, \dots, y_n) = 0$ iff $y_1, \dots, y_n$ are LD. Assume $y_1, \dots, y_n$ are LD. There exist $\alpha_1, \dots, \alpha_n$, not all zero, such that $\sum_{i=1}^n \alpha_i y_i = 0$. We must therefore have that $\big(\sum_{i=1}^n \alpha_i y_i \big| y_j \big) = 0$ for all $j = 1, \dots, n$. Then $(y_n|y_j) = \sum_{i=1}^n (-\alpha_i/\alpha_n)(y_i|y_j)$ for all j, so $\mathbf{G}$ must be rank deficient and $g = 0$. Conversely, if $\big(\sum_{i=1}^n \alpha_i y_i \big| y_j\big) = 0$ for all j and with some α_i not zero, then $0 = \sum_{j=1}^n \overline{\alpha}_j \big(\sum_{i=1}^n \alpha_i y_i \big| y_j\big) = \big(\sum_{i=1}^n \alpha_i y_i \big| \sum_{j=1}^n \alpha_j y_j\big) = \|\sum_{i=1}^n \alpha_i y_i\|^2$. □

Theorem 4. *Let $y_1, \dots, y_n$ be LI. Let δ be the minimum distasnce from x to the subspace M generated by $\{y_i\}$. Then*

$$\delta^2 = \frac{g(y_1, \dots, y_n, x)}{g(y_1, \dots, y_n)}$$

1.4 Abstract Fourier Series

Theorem 5. *Let $\{e_i\}$ be an orthonormal sequence in a Hilbert space H. A series of the form $\sum_{i=1}^{\infty} \xi_i e_i$ converges to some $x \in H$ iff $\sum_{i=1}^{\infty} |\xi_i|^2 < \infty$. In that case, $\xi_i = (x|e_i)$.*

Lemma 2 (Bessel's Inequality)**.** *Let x be an element in a Hilbert space H and suppose $\{e_i\}$ is an orthonormal sequence in H. Then $\sum_{i=1}^{\infty} \big|(x|e_i)\big|^2 \leq \|x\|^2$.*

Theorem 6. *Let x be an element in a Hilbert space H and suppose $\{e_i\}$ is an orthonormal sequence in H. Then the series $\sum_{i=1}^{\infty} (x|e_i)e_i$ converges to an element $\hat{x}$ in the closed subspace $M := \overline{[\{e_i\}_{i=1}^{\infty}]}$. The residual $x - \hat{x}$ is orthogonal to M.*

Lemma 3. *An orthonormal sequence $\{e_i\}$ in a Hilbert space H is complete (ie, $\overline{[\{e_i\}_{i=1}^{\infty}]} = H$) iff the only vector orthogonal to each e_i is the null vector.*

Proof. This follows from $H = M \oplus M^{\perp}$. □

A corollary of the previous lemma (but a long proof involving the Weierstrass Approximation Theorem) is that the countable space of polynomials is dense in $L_2[a, b]$.

The practical implication of this section is that we don't have to solve the linear system $\mathbf{G}^{\top} \boldsymbol{\alpha} = \big((x|y_1), \dots, (x|y_n)\big)^{\top}$ to find the projection. Instead, we can use Gram-Schmidt to turn $\{y_1, \dots, y_n\}$ into $\{e_1, \dots, e_n\}$ and then find $\hat{x} = \sum_{i=1}^{n} (x|e_i)e_i$.

1.5 Other Minimum Norm Problems

We move onto our first generalisation of the projection theorem.

Theorem 7. *Let M be a closed subspace of a Hilbert space H. Let $x \in H$ and $V := x + M$. Then there exists a unique $x_0 \in V$ of minimum norm. Furthermore, $x_0 \perp M$ (**not** $x_0 \perp V$).*

Proof. This is an x-shift of the problem of projecting $-x$ onto M. □

Theorem 8. *Let H be a Hilbert space and $y_1, \dots, y_n \in H$ be LI. Among all vectors $x \in H$ satisfying $(x|y_i) = c_i$ for $i = 1, \dots, n$, let x_0 have the minimum norm. Then $x_0 \in \overline{[y_1, \dots, y_n]}$ and may be written as*

$$x_0 = \sum_{i=1}^{n} \beta_i y_i$$

where satisfying $(x_0|y_i) = c_i$ for $i = 1, \dots, n$ necessitate

$$\begin{aligned}(y_1|y_1)\beta_1 + \cdots + (y_n|y_1)\beta_n &= c_1 \\ &\vdots \\ (y_1|y_n)\beta_1 + \cdots + (y_n|y_n)\beta_n &= c_n.\end{aligned}$$

This is equivalent to $\mathbf{G}^{\top} \boldsymbol{\beta} = \mathbf{c}$.

Proof. Let $M = \overline{[y_1, \dots, y_n]}$. The n constraints define a "linear variety" (shifted linear subspace) $x + M^\perp$, where $x \in H$ satisfies the n constraints. By the restated projection theorem, $x_0 \in M^{\perp\perp}$. As M is a closed subspace, $M = M^{\perp\perp}$. The condition $\mathbf{G}^\top \boldsymbol{\beta} = \mathbf{c}$ ensure $x_0 \in M$. $\square$

Do go back and read Example 1 on page 66, it's truly great.

Duality: Let M be a closed subspace of Hilbert space H and let $x \in H$. The two problems:

1. project x onto M
2. project x onto $M^\perp$

are complete symmetric because $M^{\perp\perp} = M$. If m_0 is the projection of x onto M, then $x - m_0 \in M^\perp$ is the projection of x onto $M^\perp$.

The second generalisation of the projection theorem is to convex sets:

Theorem 9. *Let x be a vector in a Hilbert space H and let K be a closed convex subset of H. Then there is a unique vector $k_0 \in K$ such that*

$$\|x - k_0\| \le \|x - k\|$$

for all $k \in K$. Furthermore, a necessary and sufficient condition that k_0 be the unique minimising vector is that $(x - k_0 | k - k_0) \le 0$ for all $k \in K$.

2 Dual Spaces

2.1 Linear Functionals

Let X be a vector space with the scalar field $\mathbf{K}$. A functional is a map $f : X \to \mathbf{K}$. A linear functional satisfies $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$ for any $x, y \in X$ and $\alpha, \beta \in \mathbf{K}$. Some basic facts about linear functionals on normed spaces:

- If a linear functional is continuous at a single point, it is continuous everywhere.
- A linear functional is bounded if there is some M such that $|f(x)| \le M\|x\|$ for all $x \in X$.
- A linear functional is bounded iff it is continuous.

The functional norm is thus defined:

$$\|f\| = \inf_M \{M : |f(x)| \le M\|x\|,\ \forall x \in X\} = \sup_{x \neq 0} \frac{|f(x)|}{\|x\|} = \sup_{\|x\| \le 1} |f(x)| = \sup_{\|x\| = 1} |f(x)|$$

The space of all bounded linear functionals over X is denoted X^* and is called the *(topological) dual* of X. It is a linear space when additional and scalar multiplication are defined in the usual way for functions.

Theorem 10. *If X is a Banach space, then X^* is a Banach space.*

Proof. Let $\{x_n^*\}$ be Cauchy in X^*. Then $\|x_n^* - x_m^*\| \to 0$ as $n, m \to 0$. Then $|x_n^*(x) - x_m^*(x)| \le \|x_n^* - x_m^*\|\|x\| \to 0$. Then $x_n^*(x) \to x^*(x) \in \mathbf{K}$ by the completeness of $\mathbf{K}$. By the linearity of limits, x^* is a linear functional. Now for all $\epsilon > 0$, there exists $N \in \mathbf{N}$ such that $|x_n^*(x) - x_m^*(x)| \le \epsilon\|x\|$ for all $n, m \ge N$. Take $m \to \infty$, we have $|x_n^*(x) - x^*(x)| \le \epsilon\|x\|$, so $\|x_n^* - x^*\| \le \epsilon$ and thus $x_n^* \to x^*$. Finally, we show that x^* is bounded: $|x^*(x)| \le |x^*(x) - x_n^*(x)| + |x_n^*(x)| \le (\epsilon + \|x_n^*\|)\|x\|$ for any $n \ge N$. $\square$

Proposition 3. *The dual of $\mathbf{R}^n$ (with the usual norm) is $\mathbf{R}^n$, i.e. a bounded linear functional f can be represented by $f(x_1, \ldots, x_n) = \sum_{i=1}^n y_i x_i$ for $y_i \in \mathbf{R}$ and every $\mathbf{y} \in \mathbf{R}^n$ defines a bounded linear functional f in the same way.*

Proof. It is obvious that the f defined by $\mathbf{y}$ is linear. Now

$$|f(\mathbf{x})| = \left|\sum_{i=1}^n y_i x_i\right| \le \left|\sum_{i=1}^n y_i^2\right|^{1/2} \left|\sum_{i=1}^n x_i^2\right|^{1/2} = \left|\sum_{i=1}^n y_i^2\right|^{1/2} \|\mathbf{x}\|$$

where we make use of the Cauchy-Schwarz inequality. Equality is achieved at $\mathbf{x} = \mathbf{y}$ and thus $\|f\| = \left|\sum_{i=1}^n y_i^2\right|^{1/2} = |\mathbf{y}|$. Thus, there is a one-to-one correspondence between $f \in (\mathbf{R}^n)^*$ and $\mathbf{y} \in \mathbf{R}^n$.

Let f be a bounded linear functional. Then

$$f(\mathbf{x}) = f\left(\sum_{i=1}^n x_i e_i\right) = \sum_{i=1}^n x_i f(e_i)$$

Setting $y_i = f(e_i)$, we obtain f in the desired form. Setting $\mathbf{x} = \mathbf{y}$, we obtain $f(\mathbf{y}) = |\mathbf{y}|$ and hence $\|f\| = \|\mathbf{y}\|$ (again $\|f\| \le \|\mathbf{y}\|$ because of Cauchy-Schwarz). ☐

Proposition 4. *The dual of ℓ_p is ℓ_q ($1/p + 1/q = 1$, $p < \infty$), i.e. a bounded linear functional f can be represented by $f(x) = \sum_{i=1}^\infty y_i x_i$ where $y \in \ell_q$, and every $y \in \ell_q$ defines a bounded linear functional f in the same way. Also, $\|f\| = \|y\|_q$.*

Proof. Let f be a bounded linear functional. Then

$$f(x) = f\left(\sum_{i=1}^\infty x_i e_i\right) = \sum_{i=1}^\infty x_i f(e_i) = \sum_{i=1}^\infty x_i y_i$$

where we set $y_i = f(e_i)$.

Suppose $1 < p < \infty$. Consider the sequence $x^{(N)} \in \ell_p$ such that $x_i^{(N)} = |y_i|^{q/p} \operatorname{sign}(y_i) \mathbf{1}_{[i \le N]}$. Now $\|x^{(N)}\|_p = \left(\sum_{i=1}^N |y_i|^q\right)^{1/p}$, and

$$f(x^{(N)}) = \sum_{i=1}^N |y_i|^{q/p+1} = \sum_{i=1}^N |y_i|^q = \left(\sum_{i=1}^N |y_i|^q\right)^{1/q} \|x^{(N)}\|.$$

By the definition of the functional norm, $\left(\sum_{i=1}^N |y_i|^q\right)^{1/q} \le \|f\|$. Taking $N \to \infty$, we see that $\|y\|_q \le \|f\|$ and thus $y \in \ell_q$.

It is obvious that the f defined by y is linear. Now

$$|f(x)| = \left|\sum_{i=1}^\infty y_i x_i\right| \le \left|\sum_{i=1}^\infty y_i^p\right|^{1/p} \left|\sum_{i=1}^\infty x_i^q\right|^{1/q} = \|y\| \|x\|$$

where we make use of the Hölder inequality. Hence $\|f\| \le \|y\|_q$ and hence y defines a bounded linear functional f.

Combining the two inequalities, we see that $\|f\| = \|y\|_q$.

For $p = 1$, repeat the proof with $x^{(N)} = \operatorname{sign}(y_N)\, e_N$. ☐

Proposition 5. *The dual of $L_p(\Omega, \mathcal{A}, \mu)$ is $L_q(\Omega, \mathcal{A}, \mu)$ ($1/p + 1/q = 1$, $p < \infty$), i.e. a bounded linear functional f can be represented by $f(x) = \int_\Omega x\, y\, d\mu$ where $y \in L_q$, and every $y \in L_q$ defines a bounded linear functional f in the same way. Also, $\|f\| = \|y\|_q$.*

Proof. Use lots of measure theory. □

Theorem 11 (Riesz Representation Theorem)**.** *If f is a bounded linear functional on a Hilbert space H, there exists a unique vector $y \in H$ such that for all $x \in H$, $f(x) = (x|y)$. Furthermore, we have $\|f\| = \|y\|$ and every y determines a unique bounded linear functional in this way. A Hilbert space is its own dual.*

Proof. Let $N = f^{-1}[\{0\}]$, which is a closed subspace of H. Now $H = N \oplus N^{\perp}$. If $N = H$, then $f = 0$ and $y = 0$. If $N \neq H$, then there exists $z \in N^{\perp}$. As $N^{\perp}$ is a subspace, we can take z such that $f(z) = 1$. Then for any $x \in H$, we have $x - f(x)z \in N$ as $f(x - f(x)z) = f(x) - f(x)f(z) = 0$. Then $z \perp N$ implies that $(x - f(x)z|z) = 0$ and that $(x|z) = f(x)(z|z)$ and hence we get the theorem by taking $y = z/\|z\|^2$. □

Let us define a function of bounded variation: $v : [a, b] \to \mathbf{R}$ is of bounded variation if

$$\sup\left\{\sum_{i=1}^{n} |v(x_i) - v(x_{i-1})| \,\middle|\, \{x_0, \ldots, x_n\} \text{ is a partition of } [a, b]\right\} < \infty$$

and the supremum is called the total variation of v.

Theorem 12 (Riesz-Kakutani-Markov Representation Theorem)**.** *Let f be a bounded linear functional on $X = C[a, b]$. Then there is a function v of bounded variation on $[a, b]$ such that*

$$f(x) = \int_a^b x(t)dv(t), \quad \forall x \in X$$

and $\|f\|$ is the total variation of v. Conversely, every function of bounded variation on $[a, b]$ defines a bounded linear functional on X in this way.

More generally, if X is a locally compact Hausdorff space, there is a unique regular Borel measure μ on X such that

$$f(x) = \int_X x(t)d\mu(t), \quad \forall x \in X.$$

The proof in the book uses the Stieltjes integral to prove the $[a, b]$ domain case. As usual, $BV[a, b]$ does not provide unique duals: we need to deal with annoying measure zero cases. We do this by defining the normalised space of bounded variation functions $NBV[a, b]$, which impose $v(a) = 0$ and right-continuity on (a, b).

2.2 Hahn-Banach Theorem

Let M, N be subspaces of vector space X with $M \subset N$, and $f : M \to \mathbf{K}$ be a linear functional. Then a linear functional $F : N \to \mathbf{K}$ is an extension of f from M to N if $f(m) = F(m)$ for all $m \in M$.

A sublinear functional p on a real vector space X satisfies:

1. $p(x_1 + x_2) \leq p(x_1) + p(x_2)$ for all $x_1, x_2 \in X$
2. $p(\alpha x) = \alpha p(x)$ for all $\alpha \geq 0$ and $x \in X$

Theorem 13 (Hahn-Banach, Extension form)**.** *Let X be a real linear normed space and p a continuous sublinear functional on X. Let f be a sublinear functional defined on a subspace M of X satisfying $f(m) \leq p(m)$ for all $m \in M$. Then there is an extension F of f from M to X such that $F(x) \leq p(x)$ on X.*

In particular, if f is a bounded linear functional, taking $p(x) = \|f\|_M\|x\|$, there is an extension F of f from M to X such that $F(x) \leq \|f\|_M\|x\|$. Thus, $\|F\| = \|f\|_M$.

A corollary is that for any $x \in X$, where X is a normed space, there is a nonzero bounded linear functional F on X such that $F(x) = \|F\|\|x\|$. The proof is to consider the one-dimensional subspace $[x]$ (or $[y]$ for $y \neq 0$ if $x = 0$) and note that $f(\alpha x) = \alpha\|x\|$ is a linear function for all $\alpha \in \mathbf{K}$ (and hence any $z \in [x]$) and is bounded by α (f is like a signed norm?). By Hahn-Banach, f can be extended to F on X. Set $\alpha = 1$ to obtain the result. The converse is not true (in non-reflexive spaces)—there are bounded linear functionals f such that $\|x\| < \|f\|$ for all $x \neq 0$.

2.3 Second Dual Space

Introduce the angle bracket notation: $\langle x, x^* \rangle := x^*(x)$. Note that by the Riesz Representation Theorem, inner products on Hilbert spaces are bounded linear functionals when one multiplier is held fixed, and hence this notation generalises inner products.

Fix $x \in X$. Then $f(x^*) = \langle x, x^* \rangle$ is a linear functional on X^*. Now $|f(x^*)| = |\langle x, x^* \rangle| = |x^*(x)| \leq \|x^*\|\|x\|$ and hence $\|f\| \leq \|x\|$. Conversely, by the corollary to the Hahn-Banach theorem, there exists $x^* \in X^*$ such that $f(x^*) = \langle x, x^* \rangle = \|x\|\|x^*\|$. Hence $\|f\| = \|x\|$. Hence f is a bounded linear functional. We can thus define the *natural mapping* $\phi : X \to X^{**}$ such that $\langle x, x^* \rangle = \langle x^*, \phi(x) \rangle$. Where ϕ is surjective, there is a bijection between X and X^{**} and we write $X = X^{**}$ and call X *reflexive*. This holds for the ℓ_p and L_p spaces where $1 < p < \infty$. However, $X \subset X^{**}$ holds strictly for non-reflexive spaces, such as ℓ_1 and L_1.

In a reflexive space, all bounded linear functionals f have some $x \in X$ such that $f(x) = \|f\|\|x\|$.

2.4 Alignment and Orthogonal Complements

$x^* \in X^*$ is aligned with $x \in X$ if $\langle x, x^* \rangle = \|x^*\|\|x\|$.

For example, $x \in L_p$ is aligned with $y \in L_q$ if and only if the condition for equality in the Hölder Inequality hold: $x(t) = K\,\mathrm{sign}(y(t))|y(t)|^{q/p}$ for some constant K. If $x \in C[a, b]$, then denoting $\Gamma^+ = x^{-1}[\|x\|]$ and $\Gamma^- = x^{-1}[-\|x\|]$, $v \in NBV[a, b]$ is aligned with x iff v is increasing in Γ^+, decreasing in Γ^-, and does not vary elsewhere. Under these conditions, $\int_a^b x(t)dv(t) = \|x\|\big(\int_{\Gamma^+} dv(t) - \int_{\Gamma^-} dv(t)\big) = \|x\|TV(v) = \|x\|\|v\|$.

$x^* \in X^*$ is orthogonal to $x \in X$ if $\langle x, x^* \rangle = 0$.

Let $S \subset X$. Define the orthogonal complement/*annihilator* $S^\perp = \{x^* \in X^* : \langle s, x^* \rangle = 0,\ \forall s \in S\} \subset X^*$. Similarly, if $U \subset X^*$, then $U^\perp \subset X^{**}$. More usefully, the *orthogonal complement of U in X* is ${}^\perp U = \{x \in X : \langle x, u \rangle = 0,\ \forall u \in U\} \subset X$. Note that ${}^\perp U = U^\perp \cap \phi[X]$ where ϕ is the natural mapping.

Theorem 14. *Let M be a closed subspace of normed space X. Then ${}^\perp[M^\perp] = M$.*

2.5 Minimum Norm Problems

Theorem 15. *Let X be a real normed vector space and M be a subspace of X. Let $x \in X$ and $d = \mathrm{dist}(x, M) = \inf_{m \in M}\|x - m\|$. Then $d = \max_{x^* \in M^\perp : \|x^*\| \leq 1}\langle x, x^* \rangle$, achieved at some $x_0^* \in M^\perp$. If the infimum is achieved at $m_0 \in M$, then x_0^* is aligned with $x - m_0$, that is to say, $\langle x - m_0, x_0^* \rangle = \|x - m_0\|\|x_0^*\|$. In more natural notation, if $\|x\|_M = \mathrm{dist}(x, M)$, then $\|x\|_M = \|x\|_{M^\perp}$ where the right-hand term is the usual functional norm applied to the naturally mapped $\phi(x)$.*

Proof. The proof is trivial if $x \in M$. Hence assume $x \notin M$.

Step 1: Prove $\langle x, x^* \rangle \leq d$. Let $\epsilon > 0$ be arbitrary, let $m_\epsilon \in M$ satisfy $\|x - m_\epsilon\| \leq d + \epsilon$, which exists because M is a subspace. Then for any $x^* \in M^\perp$ such that $\|x^*\| \leq 1$, we have $\langle m_\epsilon, x^* \rangle = 0$ and hence

$$\langle x, x^* \rangle = \langle x - m_\epsilon, x^* \rangle \leq \underbrace{\|x^*\|}_{\leq 1} \|x - m_\epsilon\| \leq d + \epsilon$$

As $\epsilon > 0$ was arbitrary, we have $\langle x, x^* \rangle \leq d$.

Step 2: Prove that the maximum is attained at some x_0^*. Let $N = [x + M]$. If $n \in N$, the representation $n = \alpha x + m$ is unique, where $\alpha \in \mathbf{R}$ and $m \in M$. Define the linear functional $f : N \to \mathbf{R}$ by $f(n) = \alpha d$. Note that if $m \in M$, then $\alpha = 0$ and $f(m) = 0$, and that $f(x) = d$. We have

$$\|f\| = \sup_N \frac{|f(n)|}{\|n\|} = \sup_N \frac{|\alpha| d}{\|\alpha x + m\|} = \sup_N \frac{|\alpha| d}{|\alpha| \|x + m/\alpha\|} = \frac{d}{\inf_N \|x + m/\alpha\|} = 1$$

The maximum x_0^* is attained at the Hahn-Banach extension of f from N to X. Note that $\|x_0^*\| = 1$ and $x_0^* \in M^\perp$. Further, $\langle x, x_0^* \rangle = d$.

Step 3: Alignment. Assume there exists $m_0 \in M$ such that $\|x - m_0\| = d$. Let $x_0^* \in M^\perp$, $\|x_0^*\| = 1$ obtain the maximum. Then

$$\langle x - m_0, \underbrace{x_0^*}_{\in M^\perp} \rangle = \langle x, x_0^* \rangle = d = \underbrace{\|x_0^*\|}_{=1} \|x - m_0\|$$ □

A corollary is thus: let x be an element of real normed vector space X and M be a subspace of X. Then $m_0 \in M$ satisfies $\|x - m_0\| \leq \|x - m\|$ for all $m \in M$ iff there is a nonzero $x^* \in M^\perp$ aligned with $x - m_0$. In this sense, $x - m_0$ is orthogonal to M, like the Hilbert projection theorem.

Theorem 16. *Let X be a real normed vector space and M be a subspace of X. Let $x^* \in X^*$ and denote $d = \operatorname{dist}(x^*, M^\perp)$. $d = \min_{m^* \in M^\perp} \|x^* - m^*\|$ is achieved at some $m_0^* \in M^\perp$, and $d = \sup_{x \in M : \|x\| \leq 1} \langle x, x^* \rangle = \|x^*\|_M$. If the supremum is achieved at $x_0 \in M$, then $x^* - m_0^*$ is aligned with x_0. In more natural notation, if $\|x^*\|_{M^\perp} = \operatorname{dist}(x^*, M^\perp)$, then $\|x^*\|_{M^\perp} = \|x^*\|_M$.*

Proof. Step 1: Prove $\|x^* - m^*\| \geq \|x^*\|_M$. For any $m^* \in M^\perp$, we have

$$\|x^* - m^*\| = \sup_{\|x\| \leq 1} (\langle x, x^* \rangle - \langle x, m^* \rangle) \geq \sup_{x \in M : \|x\| \leq 1} (\langle x, x^* \rangle - \langle x, m^* \rangle)$$

Noting that $\langle x, m^* \rangle = 0$ for all $x \in M$, we have $\|x^* - m^*\| \geq \|x^*\|_M$.

Step 2: Prove that the minimum is attained at some m_0^*. Let y^* be the Hahn-Banach extension of $x^*|_M$ from M to X. Then $x^* - y^* = 0$ on M. Set $m_0^* = x^* - y^*$. Then $m_0^* \in M^\perp$ and $\|x^* - m_0^*\| = \|x^* - x^* + y^*\| = \|y^*\| = \|x^*\|_M$.

Step 3: Alignment. Assume there exists $x_0 \in M$ such that $\langle x_0, x^* \rangle = d$, which from Step 1 implies that $x_0 = \arg\max_{x \in M : \|x\| \leq 1} \langle x, x^* \rangle$. Then $\|x_0\| = 1$ (else $\langle x_0 / \|x_0\|, x^* \rangle > d$, contradicting steps 1–2). Also,

$$\|x_0\| \|x^* - m_0^*\| = d = \langle x_0, x^* \rangle = \langle x_0, x^* - m_0^* \rangle$$

where the last equality follows from $m_0^* \in M^\perp$, implying that $\langle x_0, m_0^* \rangle = 0$. □

As an application, we will prove one of Tonelli's theorems.

Proposition 6. *If f is continuous on $[a,b]$ and p_0 is a polynomial of up to degree n minimising $\|f-p\|_\infty$, then $|f(t)-p_0(t)|$ achieves its maximum on at least $n+2$ points on $[a,b]$.*

Proof. We formulate this as projecting f in space $X = C[a,b]$ onto the $(n+1)$-dimensional subspace M of n^{th} degree polynomials. p_0 exists by the finite-dimensionality of M. Let $d = \|f-p_0\|_\infty > 0$ and $\Gamma = |f-p_0|^{-1}[\{d\}]$ denote the arg-maxima of $|f-p_0|$. By the first minimum norm theorem, $f-p_0$ must be aligned with some $v^* \in M^\perp \subset C[a,b]^* = NBV[a,b]$. As $d>0$, $v^* \neq 0$. From the alignment section, we know that v^* varies only on Γ.

Assume by way of contradiction that $|\Gamma| < n+2$. Let $t_k \in \Gamma$ such that v^* varies at t_k. Then the polynomial $q(t) = \prod_{s\in\Gamma\setminus\{t_k\}}(t-s)$ has at most an order of $n+1$ so $q \in M$, but $\langle q, v^*\rangle = \int_a^b q\,dv^* = \prod_{s\in\Gamma\setminus\{t_k\}}(t_k - s) \times (v^*(t_k) - \lim_{t\to t_k^-} v^*(t)) \neq 0$, so $v^* \notin M^\perp$. Contradiction! □

As another application, we will solve minimum norm problems with linear constraints, just like the Hilbert Space section.

Proposition 7. *Let $y_i \in X$ for $i = 1,\dots,n$ and suppose $D = \{x^* \in X^* : \langle y_i, x^*\rangle = c_i,\ i = 1,\dots,n\}$ is nonempty (so the constraints are consistent). Then*

$$\min_{x^*\in D}\|x^*\| = \max_{\|Ya\|\le 1} c^\top a.$$

Proof. Let $M = \operatorname{span}\{y_1,\dots,y_n\}$ and $\bar{x}^* \in D$. The n constraints define a linear variety $\bar{x}^* + M^\perp$. Hence, the problem is an $\bar{x}^*$-shift of the problem of projecting $-\bar{x}^*$ onto $M^\perp$. Thus, if we let $d = \min_{\langle y_i,x^*\rangle=c_i}\|x^*\|$, then $d = \min_{m^*\in M^\perp}\|\bar{x}^* - m^*\|$. By the second minimum norm theorem, $d = \sup_{x\in M:\|x\|\le 1}\langle x, \bar{x}^*\rangle = \sup_{\|\sum_i a_i y_i\|\le 1}\langle \sum_i a_i y_i, \bar{x}^*\rangle$ where the second equality follows from the definition of M. Now $\langle \sum_i a_i y_i, \bar{x}^*\rangle = \sum_{i=1}^n a_i\langle y_i, \bar{x}^*\rangle = \sum_{i=1}^n a_i c_i = c^\top a$ where the second equality follows from $\bar{x}^* \in D$. □

2.6 Weak convergence

The following convergence notions appear in decreasing order of strength. Consider $\{x_n\}$ in normed vector space X.

- $x_n \to x$ (strongly) if $\|x_n - x\| \to 0$ as $n \to \infty$
- $x_n \rightharpoonup x$ (weakly) if $\langle x_n, x^*\rangle \to \langle x, x^*\rangle$ as $n\to\infty$ for all $x^* \in X^*$
- $x_n^* \rightharpoonup^* x^*$ (weak*) if $\langle x, x_n^*\rangle \to \langle x, x^*\rangle$ as $n\to\infty$ for all $x \in X$

Strong implies weak as $|\langle x_n, x^*\rangle - \langle x, x^*\rangle| \le \|x^*\|\|x_n - x\|$. Weak implies weak* as X^{**} contains $\phi(X)$, where ϕ is the natural mapping.

More definitions:

- A set $K \subset X^*$ is weak* compact if every infinite sequence from K contains a weak* convergent subsequence (converging to a point in K).
- A functional $f : X \to \mathbf{K}$ is weakly continuous at x if given $\epsilon > 0$, there is a $\delta > 0$ and finite $\{x_1^*,\dots,x_m^*\} \subset X^*$ such that $|f(y) - f(x)| < \epsilon$, for all $y \in X$ such that $|\langle y, x_i^*\rangle - \langle x, x_i^*\rangle| < \delta$ for all $i = 1,\dots,m$.
- A functional $f : X^* \to \mathbf{K}$ is weak* continuous at x^* if given $\epsilon > 0$, there is a $\delta > 0$ and finite $\{x_1,\dots,x_m\} \subset X$ such that $|f(y^*) - f(x^*)| < \epsilon$, for all $y^* \in X^*$ such that $|\langle x_i, y^*\rangle - \langle x_i, x^*\rangle| < \delta$ for all $i = 1,\dots,m$.

Theorem 17 (Alaoglu)**.** *Let X be a real normed vector space. The closed unit sphere in X^* is weak* compact.*

Proposition 8. *If $x_n \rightharpoonup x$ and f is weakly continuous, then $f(x_n) \to f(x)$. Similarly, if $x_n^* \rightharpoonup^* x^*$ and f is weak* continuous, then $f(x_n^*) \to f(x^*)$.*

Proof. Fix $\epsilon > 0$. As f is weakly continuous, there exists $\delta > 0$ and $x_i^* \in X^*$ such that $|f(y) - f(x)| < \epsilon$ for all $y \in X$ satisfying $|\langle y, x_i^* \rangle - \langle x, x_i^* \rangle| < \delta$ for all $i = 1, \ldots, m$. By definition of weak convergence, there exists $N \in \mathbf{N}$ such that $|\langle x_n, x_i^* \rangle - \langle x, x_i^* \rangle| < \delta$ for all $n \geq N$ and all i. Hence $|f(x_n) - f(x)| < \epsilon$ for all $n \geq N$. For weak* convergence, simply change the variables in the proof. $\square$

Proposition 9 (Extreme Value Theorem)**.** *Let $S \subset X^*$ be weak* compact and $f : S \to \mathbf{R}$ be weak* continuous. Then f is bounded on S and achieves its maximum on S.*

Proof. Let $\{y_n\} \subset f(S) \subset \mathbf{R}$ be an arbitrary sequence. Then there exists $\{x_n^*\} \subset S$ such that $f(x_n^*) = y_n$ for all $n \in \mathbf{N}$. By weak* compactness, there exists subsequence $\{x_{n_k}^*\}$ such that $x_{n_k}^* \rightharpoonup^* x^* \in S$. As f is weak* continuous, by the previous proposition, $f(x_{n_k}^*) \to f(x^*)$. Let $y_{n_k} = f(x_{n_k}^*)$ for all $k \in \mathbf{N}$ and $y = f(x^*)$. Then $y_{n_k} \to y$ as $k \to \infty$, so $\{y_n\}$ has a convergent subsequence and $f(S)$ is compact. By the Heine-Borel theorem, f must be bounded (in the traditional sense, *not* in the linear functional sense).

Let $y_n \to \sup f(S)$. Then there exists $\{x_n^*\} \subset S$ such that $f(x_n^*) = y_n$ for all $n \in \mathbf{N}$. By weak* compactness, there exists subsequence $\{x_{n_k}^*\}$ such that $x_{n_k}^* \rightharpoonup^* x^* \in S$. As f is weak* continuous, by the previous proposition, $f(x_{n_k}^*) \to f(x^*)$. In addition, $f(x^*)$ must be $\sup f(S)$, so f achieves its maximum on S. $\square$

Combining Alaoglu's Theorem and the Extreme Value Theorem, set S to be the unit sphere in X^* and $f(x^*) = \langle x, x^* \rangle$ for some $x \in X$, which is weak* continuous. Then the EVT tells us that $f(x^*)$ achieves its maximum on the unit sphere. By the first alignment theorem, taking $M = \{0\}$, we have alignment with the optimal x_0^*, so $\langle x, x_0^* \rangle = \|x_0^*\| \|x\| = \|x\|$, which is equivalent to the Corollary to the Hahn-Banach Theorem.

2.7 Hyperplanes

A hyperplane H in vector space X is a maximal proper linear variety. It satisfies:

- $H \neq X$
- If V is a linear variety containing X, then $V = H$ or $V = X$

Since $\bar{H}$ is also a linear variety, then $\bar{H} \in \{H, X\}$ so H is either closed or dense.

Proposition 10. *Hyperplanes are pre-images under linear functionals of singleton sets. Conversely, if f is a non-zero linear functional on X, $f^{-1}[\{c\}]$ is a hyperplane in X where $c \in \mathbf{K}$.*

Proof. If H contains the origin, then there is a linear functional f such that $H = f^{-1}[\{0\}]$. We can construct f by taking $x_1 \notin H$ and defining $f(x) = \alpha$ where $x = \alpha x_1 + h$ for some $h \in H$ (recall by linear independence that α and h are unique).

If H does not contain the origin, then $H = x_0 + M$ for some linear subspace M and $x_0 \notin M$. Then construct $f(x) = \alpha$ where $x = \alpha x_0 + m$ for some $m \in M$ and $H = f^{-1}[\{1\}]$.

Conversely, let $M = f^{-1}[\{0\}]$, a linear subspace of X. Let $x_0 \in X \setminus M$ with $f(x_0) = 1$. Then for any $x \in X$, we have $f(x - f(x)x_0) = f(x) - f(x_0)f(x) = 0$, and thus $x - f(x)x_0 \in M$. Thus, $x = m + f(x)x_0$ for some $m \in M$, and hence $X = [x_0 + M]$. Hence M is a proper maximal subspace. Letting $x_1 \in M$ such that $f(x_1) = c$, we have $f^{-1}[\{c\}] = x_1 + M$ is a hyperplane. □

In addition, if H is not a subspace (i.e. it does not contain the origin), the representation $f^{-1}[\{1\}]$ is unique. To prove this, assume $H = f^{-1}[\{1\}] = g^{-1}[\{1\}]$. Then $H \subset (f - g)^{-1}[\{0\}]$, which is a linear subspace. But by the definition of H as a maximal linear variety, $(f - g)^{-1}[\{0\}] = X$ so $f = g$.

Proposition 11. *Let f be a nonzero linear functional on a normed vector space X. Then the hyperplane $H = f^{-1}[\{c\}]$ is closed for every $c \in \mathbf{K}$ iff f is continuous/bounded.*

Let f be a nonzero linear functional on vector space X. The hyperplane $H = f^{-1}[c]$ determines four half-spaces (open and closed only make sense if f is continuous):

	Negative	Positive
Open	$f^{-1}[(-\infty, c)] = \{x : f(x) < c\}$	$f^{-1}[(c, \infty)] = \{x : f(x) > c\}$
Closed	$f^{-1}[(-\infty, c]] = \{x : f(x) \leq c\}$	$f^{-1}[[c, \infty)] = \{x : f(x) \geq c\}$

The proof of the hyperplane separation theorem requires the introduction of a specific sublinear functional, the *Minkowski functional*:

$$p_K(x) = \inf \left\{ r : \frac{x}{r} \in K,\ r > 0 \right\}$$

where $0 \in \operatorname{int} K$. Beyond sublinearity, we have $p > 0$, p is continuous, $\operatorname{int} K = p^{-1}[[0, 1)]$, and $\overline{K} = p^{-1}[[0, 1]]$.

Theorem 18 (Hyperplane Separation Theorem)**.** *Let K be a convex set in real normed vector space X, $\operatorname{int} K \neq 0$, V is a linear variety in X such that $V \cap \operatorname{int} K = 0$. Then there is a closed hyperplane in X containing V but no interior points of K. Hence there is $x^* \in X^*$ and $c \in \mathbf{R}$ such that $\langle v, x^* \rangle = c$ for all $v \in V$ and $\langle k, x^* \rangle < c$ for all $k \in \operatorname{int} K$.*

There are various other definitions and restatements (X, K are defined as in the theorem):

- A closed hyperplane H in X supports convex set K if K is contained in one closed half-space determined by H and $H \cap \overline{K} \neq \emptyset$.
- If $x \notin \operatorname{int} K \neq \emptyset$, then there is a closed hyperplane H containing x such that K lies on one side of H.
- Let K_1 and K_2 be convex sets in X such that $\operatorname{int} K_1 \neq \emptyset$ and $K_2 \cap \operatorname{int} K_1 = \emptyset$. Then there is a closed hyperplane H separating K_1 and K_2. That is to say, there is an $x^* \in X^*$ such that $\sup_{x \in K_1} \langle x, x^* \rangle \leq \inf_{x \in K_2} \langle x, x^* \rangle$.
- Let K be closed and convex and $x \notin K$. Then there is a closed half-space containing K but not x.
- Let K be closed and convex. Then K is the intersection of all closed half-spaces containing it.
- The *support functional* of K is $h(x^*) = \sup_{x \in K} \langle x, x^* \rangle$.

Finally, there is another dual theorem:

Theorem 19. *Let X be a real normed vector space and K be a convex subset with support functional h. Let $x_1 \in X$ and $d = \operatorname{dist}(x_1, K) = \inf_{x \in K} \|x - x_1\|$. Then $d = \max_{\|x^*\| \leq 1} \big(\langle x_1, x^* \rangle - h(x^*) \big)$, achieved at some $x_0^* \in X^*$. If the infimum is achieved at $x_0 \in K$, then $-x_0^*$ is aligned with $x_0 - x_1$.*

3 Hilbert Space exercises

Q3.1 *Show that $|(x|y)| = \|x\|\|y\|$ iff $\alpha x + \beta y = 0$ for some scalars α, β.*

Take the usual proof of the inequality and notice that $0 = (x - \lambda y|x - \lambda y)$ if and only if $x - \lambda y = 0$ by the property of the inner product.

Q3.2 *Consider the set X of real functions x defined in $\mathbf{R}$ for which*

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |x(t)|^2 dt < \infty.$$

Let M be the subspace where the limit is zero. Part (a) is trivial: show that the space $H = X/M$ becomes a pre-Hilbert space when the inner product is defined as

$$([x]|[y]) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)y(t)dt.$$

Part (b): show that H is not separable.

Q3.3 *Let H consist of all $m \times n$ real matrices with addition and scalar multiplication defined as the usual corresponding operations with matrices, and with the inner product of two matrices A, B defined as*

$$(A|B) = \operatorname{tr}(A^\top QB)$$

where Q is a symmetric, positive-definite $m \times m$ matrix. Prove that H is a Hilbert space.

Symmetry and linearity are trivial to show. As Q is positive-definite, if $A \neq 0$, then letting $A = \begin{bmatrix} a_1 & \dots & a_n \end{bmatrix}$ with some $a_i \neq 0$, we have that

$$A^\top QA = \begin{bmatrix} a_1^\top \\ \vdots \\ a_n^\top \end{bmatrix} Q \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} = \begin{bmatrix} a_1^\top Qa_1 & \dots & a_1^\top Qa_n \\ \vdots & \ddots & \vdots \\ a_n^\top Qa_1 & \dots & a_n^\top Qa_n \end{bmatrix}$$

By the positive-definiteness of Q, $a_i^\top Qa_i \geq 0$ and strictly if $a_i \neq 0$. Then $(A|A) = \operatorname{tr}(A^\top QA) = \sum_{i=1}^n a_i^\top Qa_i > 0$. On the other hand $(0|0) = \operatorname{tr}(0) = 0$.

Let $\{A_j\}$ be a Cauchy sequence in H. Then $\|A_j - A_k\|^2 \to 0$ as $j, k \to \infty$. Hence $(A_j - A_k|A_j - A_k) \to 0$ as $j, k \to \infty$. Hence

$$\sum_{i=1}^{n} (a_i^{(j)} - a_i^{(k)})^\top Q(a_i^{(j)} - a_i^{(k)}) \to 0 \quad \text{as } j, k \to \infty$$

Now $(a_i^{(j)} - a_i^{(k)})^\top Q(a_i^{(j)} - a_i^{(k)}) \geq 0$ for all i by the positive definiteness of Q. Hence we must have $(a_i^{(j)} - a_i^{(k)})^\top Q(a_i^{(j)} - a_i^{(k)}) \to 0$ as $j, k \to \infty$ for all $i = 1, \dots, n$. By the positive definiteness of Q, we must have $a_i^{(j)} - a_i^{(k)} \to 0$ as $j, k \to \infty$. Therefore, we have that $\{a_{i,l}^{(j)}\}$ is a Cauchy sequence in $\mathbf{R}$ for all $i = 1, \dots, m$ and $l = 1, \dots, n$. By the completeness of $\mathbf{R}$, we have that $\{a_{i,l}^{(j)}\} \to a_{i,l} \in \mathbf{R}$ and therefore $A^j \to [a_{i,l}]_{i=1,\dots,m,\ l=1,\dots,n}$ as $j \to \infty$. Hence H is complete and therefore a Hilbert space.

Q3.4 *Show that if $g(y_1, \dots, y_n) = 0$, the normal equations possess a solution but it is not unique.*

Say if $\operatorname{rank} G = m$. Then take $y_1, \dots, y_m$ to be a LI set of vectors (rearranging if necessary). From the proposition, we know that $\mathbf{G}(y_1, \dots, y_m)(\alpha_1, \dots, \alpha_m)^\top = ((x|y_1), \dots, (x|y_m))^\top$

has a unique solution. Set $k = 1, \dots, n-m$. Then $y_{m+k} = \sum_{j=1}^{n} \beta_j y_j$ where some β_j is nonzero. Then:

$$\begin{aligned}
\sum_{i=1}^{n} (y_i | y_{m+k}) \alpha_i &= \sum_{i=1}^{n} (y_i | \textstyle\sum_j \beta_j y_j) \alpha_i \\
&= \sum_j \overline{\beta}_j \sum_{i=1}^{n} \alpha_i (y_i | y_j) \\
&= \sum_j \overline{\beta}_j (x | y_j) \quad \text{from solution} \\
&= (x | \textstyle\sum_j \beta_j y_j) = (x | y_{m+k})
\end{aligned}$$

Hence $(\alpha_1, \dots, \alpha_m, 0, \dots, 0)^\top$ is a solution to the normal equations. However, as there is more than one way to choose a set of LI vectors out of $y_1, \dots, y_m$, this solution cannot be unique.

Q3.5 *Find the linear function $x(t) = a + bt$ minimising $\int_{-1}^{1} [t^2 - x(t)]^2 dt$.*

Let $p_k(t) = t^k$. The problem is to minimise $\|x - p_2\|_{L_2[-1,1]}$, where $x \in \operatorname{span}\{1, p_1\}$. From the projection theorem, $(x - p_2) \perp \operatorname{span}\{1, p_1\}$. Hence, $(x - p_2) \perp 1$:

$$0 = \int_{-1}^{1} [t^2 - bt - a] dt = \frac{1}{3}[t^3]_{-1}^{1} - \frac{1}{2} b [t^2]_{-1}^{1} - a[t]_{-1}^{1} = \frac{2}{3} - 2a$$

and hence $a = 1/3$. In addition, $(x - p_2) \perp p_1$:

$$0 = \int_{-1}^{1} [t^2 - bt - a] t \, dt = \frac{1}{4}[t^4]_{-1}^{1} - \frac{1}{3} b [t^3]_{-1}^{1} - \frac{1}{2} a [t^2]_{-1}^{1} = -\frac{2}{3} b$$

and hence $b = 0$. The projection is $x(t) = 1/3$.

Q3.6 *Given a function $x \in L_2[0,1]$, we seek a polynomial p of degree n or less which minimises $\int_0^1 |x(t) - p(t)|^2 dt$ while satisfying $\int_0^1 p(t) dt = 0$.*

Part (a): Show that this problem has a unique solution. The space of polynomials of degree n or less satisfying $\int_0^1 p(t)dt = 0$ form a closed linear subspace. To prove that it is closed, consider a convergent sequence p_k. If $\int_0^1 p_k(t)dt = 0$ for all k and $p_k \to p$ (which is indeed a degree $\leq n$ polynomial by the finite-dimensionality and hence closedness of that space), then by the Cauchy-Schwarz inequality,

$$\int_0^1 |p_k(t) - p(t)| dt \leq \left(\int_0^1 |p_k(t) - p(t)|^2 dt \right)^{1/2} \left(\int_0^1 dt \right)^{1/2} \to 0 \quad \text{as } k \to \infty.$$

By the triangle inequality for integrals:

$$\left| \int_0^1 p_k(t) dt - \int_0^1 p(t) dt \right| = \left| \int_0^1 p_k(t) - p(t) dt \right| \leq \int_0^1 |p_k(t) - p(t)| dt$$

and by the reverse triangle inequality,

$$\left| \int_0^1 p_k(t) dt \right| - \left| \int_0^1 p(t) dt \right| \leq \left| \int_0^1 p_k(t) dt - \int_0^1 p(t) dt \right| \to 0 \quad \text{as } k \to \infty.$$

Hence

$$\int_0^1 p(t) dt = \lim_{k \to \infty} \int_0^1 p_k(t) dt = 0.$$

Then the existence of a unique solution is guaranteed by the projection theorem.

Part (b): Show that this problem can be solved by firs finding the polynomial q of degree n or less which minimises $\int_0^1 |x(t)-q(t)|^2 dt$ and then finding p of degree n or less which minimises $\int_0^1 |q(t)-p(t)|^2 dt$ while satisfying the requirement $\int_0^1 p(t)dt = 0$.

Denote the degree $\leq n$ polynomials by P and the subset of polynomials satisfying the integral requirement as $P_0 \subset P$. By the projection theorem, $(x-q)\perp P$ and $(q-p)\perp P_0$. Then Let $\tilde{p} \in P_0 \subset P$ be arbitrary. We have that $(x-p|\tilde{p}) = (x-q+q-p|\tilde{p}) = (x-q|\tilde{p})+(q-p|\tilde{p}) = 0$ and hence $x-p\perp P_0$. Hence by the projection theorem, p is indeed the minimiser.

Q3.7 *Let M and N be orthogonal closed subspaces of a Hilbert space H and let x be an arbitrary vector in H. Show that the subspace $M \oplus N$ is closed and that the orthogonal projection of x onto $M \oplus N$ is equal to $P_M(x) + P_N(x)$.*

Let $y_k \in M \oplus N$ and let $y_k \to y$. Now $y_k = m_k + n_k$ admits a unique representation for $m_k \in M$ and $n_k \in N$. We have that for all $\epsilon > 0$, there exists $K \in \mathbf{N}$ such that $\|y_k - y\| < \epsilon$, for all $k \geq K$. Thus, $\|m_k + n_k - P_M(y) + P_M(y) - y\| < \epsilon$, for all $k \geq K$, where $P_M(y)$ exists due to the projection theorem. Now $m_k - P_M(y) \in M$ and $y - P_M(y) \in M^\perp$ by the projection theorem. As $n_k \in N \subset M^\perp$, we deduce that $n_k + P_M(y) - y \in M^\perp$ (as $M^\perp$ is a linear subspace of H). By Pythagoras's theorem, we may deduce that

$$\|m_k - P_M(y)\|^2 + \|n_k + P_M(y) - y\|^2 < \epsilon^2, \quad \forall k \geq K$$

and hence $m_k \to P_M(y)$ as $k \to \infty$. Similarly, $n_k \to P_N(y)$ as $k \to \infty$. Hence y has a unique representation $P_M(y) + P_N(y) \in M \oplus N$, so we may deduce that $M \oplus N$ is closed.

Let $x \in H$, $m \in M$, and $n \in N$ be arbitrary. We have that

$$\big(x - P_M(x) - P_N(x)|m+n\big) = \big(x - P_M(x)|m\big) + \big(x - P_N(x)|n\big) = 0$$

as $(x - P_M(x))\perp m$ and $(x - P_N(x))\perp n$ by the projection theorem. Further, as $P_M(x) \in M$, $P_N(x) \in N$, and $M\perp N$, then $P_M(x)\perp n$ and $P_N(x)\perp m$. Hence, by the projection theorem, $P_{M\oplus N}(x) = P_M(x) + P_N(x)$.

Q3.9 *Prove that if $S \subset H$, then $S^{\perp\perp} = \overline{[S]}$.*

We know that $H = S^\perp \oplus S^{\perp\perp}$ as $S^\perp$ is a closed linear subspace. We also know that $H = \overline{[S]} \oplus \overline{[S]}^\perp$. Now $S \subset \overline{[S]}$ implies that $\overline{[S]}^\perp \subset S^\perp$. We shall prove the reverse inclusion: let $x \in S^\perp$ and $y \in \overline{[S]}$. Then there exists a sequence $y_n \in [S]$ such that $y_n \to y$. Now for all $n \in \mathbf{N}$, $y_n = \sum_k \beta_{nk} s_{nk}$ is a finite sum with scalars β_{nk} and $s_{nk} \in S$ for all n, k. Hence

$$(x|y_n) = (x|\textstyle\sum_k \beta_{nk} s_{nk}) = \sum_k \overline{\beta}_{nk}(x|s_{nk}) = 0.$$

Using the continuity of the inner product,

$$(x|y) = \big(x|\lim_{n\to\infty} y_n\big) = \lim_{n\to\infty}(x|y_n) = 0$$

and hence $S^\perp \subset \overline{[S]}^\perp$. Hence $S^\perp = \overline{[S]}^\perp$ and hence $S^{\perp\perp} = \overline{[S]}$.

Q3.10 A Hilbert space H of functions on a set S is said to be a *reproducing kernel Hilbert space* if there is a function K defined on $S \times S$ having the properties:

1. $K(\cdot, t) \in H$ for each $t \in S$
2. $x(t) = (x|K(\cdot, t))$ for each $x \in H$, $t \in S$.

Such a function K is called a reproducing kernel.

Prove that a reproducing kernel, if it exists, is unique.

Assume by way of contradiction that K_1 and K_2 are reproducing kernels of H, and $K_1 \neq K_2$. Then $K_1(\cdot,t) - K_2(\cdot,t) \neq 0$. From part 1 of the definition and because H is a linear space, $K_1(\cdot,t) - K_2(\cdot,t) \in H$. By the positive definiteness of the inner product,

$$\begin{aligned} &\big(K_1(\cdot,t) - K_2(\cdot,t) | K_1(\cdot,t) - K_2(\cdot,t)\big) > 0 \\ \implies &\big(K_1(\cdot,t) - K_2(\cdot,t) | K_1(\cdot,t)\big) > \big(K_1(\cdot,t) - K_2(\cdot,t) | K_2(\cdot,t)\big) \end{aligned}$$

But this contradicts part 2 of the definition.

Q3.13 *Show that the Gram determinant $g(x_1, \dots, x_n)$ is never negative.*

Note: For $n = 2$, this is the Cauchy-Schwarz inequality. In fact, we are generalising it. Further, we will prove something stronger, that G is positive semi-definite.

Let $y \in \mathbf{R}^n$. Then

$$y^\top G y = \sum_{i=1}^n \sum_{j=1}^n y_i y_j (x_i | x_j) = \sum_{i=1}^n \sum_{j=1}^n (y_i x_i | y_j x_j) = \left(\sum_{i=1}^n y_i x_i \middle| \sum_{j=1}^n y_j x_j \right)$$

where we use the linearity of the inner product. Then changing the index, we have

$$y^\top G y = \left(\sum_{i=1}^n y_i x_i \middle| \sum_{i=1}^n y_i x_i \right) \geq 0$$

by the positive definiteness of the inner product.

Q3.14 *Let $\{y_1, \dots, y_n\}$ be LI vectors in pre-Hilbert space X and $x \in X$. Show that the best approximation to x in the subspace generated by $\{y_i\}$ has the explicit representation*

$$\hat{x} = \frac{\begin{vmatrix} (y_1|y_1) & \dots & (y_n|y_1) & (x|y_1) \\ \vdots & \ddots & \vdots & \vdots \\ (y_1|y_n) & \dots & (y_n|y_n) & (x|y_n) \\ y_1 & \dots & y_n & 0 \end{vmatrix}}{-g(y_1, \dots, y_n)}.$$

Show that the minimum error $\hat{x} - x$ is given by

$$\hat{x} - x = \frac{\begin{vmatrix} (y_1|y_1) & \dots & (y_n|y_1) & (x|y_1) \\ \vdots & \ddots & \vdots & \vdots \\ (y_1|y_n) & \dots & (y_n|y_n) & (x|y_n) \\ y_1 & \dots & y_n & x \end{vmatrix}}{-g(y_1, \dots, y_n)}.$$

Denote $\hat{x} = \alpha_1 y_1 + \dots + \alpha_n y_n$. Then using the Gram matrix and Cramer's rule, we deduce that

$$\alpha_i = \frac{\begin{vmatrix} (y_1|y_1) & \dots & (y_{i-1}|y_1) & (x|y_1) & (y_{i+1}|y_1) & \dots & (y_n|y_1) \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ (y_1|y_n) & \dots & (y_{i-1}|y_n) & (x|y_n) & (y_{i+1}|y_n) & \dots & (y_n|y_n) \end{vmatrix}}{g(y_1, \dots, y_n)}$$

Then we can move the $((x|y_1), \dots, (x|y_n))^\top$ term to the end, which involves $n - i$ pairwise interchanges of rows. As each pairwise interchange reverses the sign of the determinant, we have:

$$\alpha_i = (-1)^{n-i} \frac{\begin{vmatrix} (y_1|y_1) & \dots & (y_{i-1}|y_1) & (y_{i+1}|y_1) & \dots & (y_n|y_1) & (x|y_1) \\ \vdots & & \vdots & \vdots & & \vdots & \vdots \\ (y_1|y_n) & \dots & (y_{i-1}|y_n) & (y_{i+1}|y_n) & \dots & (y_n|y_n) & (x|y_n) \end{vmatrix}}{g(y_1, \dots, y_n)}$$

Hence

$$\hat{x} = \sum_{i=1}^{n} (-1)^{n-i} \frac{\begin{vmatrix} (y_1|y_1) & \dots & (y_{i-1}|y_1) & (y_{i+1}|y_1) & \dots & (y_n|y_1) & (x|y_1) \\ \vdots & & \vdots & \vdots & & \vdots & \vdots \\ (y_1|y_n) & \dots & (y_{i-1}|y_n) & (y_{i+1}|y_n) & \dots & (y_n|y_n) & (x|y_n) \end{vmatrix}}{g(y_1, \dots, y_n)} y_i$$

$$= \frac{\begin{vmatrix} (y_1|y_1) & \dots & (y_n|y_1) & (x|y_1) \\ \vdots & \ddots & \vdots & \vdots \\ (y_1|y_n) & \dots & (y_n|y_n) & (x|y_n) \\ y_1 & \dots & y_n & 0 \end{vmatrix}}{-g(y_1, \dots, y_n)}.$$

Note that we must introduce the negative sign (in the denominator for notational clarity) because when $n - i$ is odd, the checkerboard pattern of the determinant treats y_i positively and when $n - i$ is even, the checkerboard pattern of the determinant treats y_i as negative.

For $\hat{x} - x$, note that the checkerboard pattern keeps all terms on the main diagonal positive. Therefore, we have that the right hand side is

$$\hat{x} + x \frac{\begin{vmatrix} (y_1|y_1) & \cdots & (y_n|y_1) \\ \vdots & \ddots & \vdots \\ (y_1|y_n) & \cdots & (y_n|y_n) \end{vmatrix}}{-g(y_1, \dots, y_n)} = \hat{x} - x \frac{g(y_1, \dots, y_n)}{g(y_1, \dots, y_n)} = \hat{x} - x.$$

Q3.16

Proposition 12 (Parseval's Equality)**.** *An orthonormal sequence $\{e_i\}$ is complete in a Hilbert space H iff for all $x, y \in H$,*

$$(x|y) = \sum_{i=1}^{\infty} (x|e_i)(e_i|y).$$

Prove Parseval's equality.

Assume $\{e_i\}$ is a complete orthonormal sequence. Then $x = \sum_i (x|e_i)e_i$ and $y = \sum_i (y|e_i)e_i$. Then we have

$$\begin{aligned} (x|y) &= \left(\sum_{i=1}^{\infty} (x|e_i)e_i \middle| \sum_{j=1}^{\infty} (y|e_j)e_j \right) \\ &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} (x|e_i)(e_j|y) \underbrace{(e_i|e_j)}_{=\delta_{ij}} \\ &= \sum_{i=1}^{\infty} (x|e_i)(e_i|y) \end{aligned}$$

where δ_{ij} denotes the Dirac delta function.

Assume Parseval's equality holds. Then say if $y \in H$ is orthogonal to the subspace generated by $\{e_i\}$. This implies that $(e_i|y) = 0$ for all i. Then $(x|y) = 0$ for all $x \in H$ by the equality. In particular, $(y|y) = 0$, and therefore $y = 0$ by the positive definiteness of the inner product. Hence the only element of $\overline{[\{e_i\}]}^{\perp}$ is 0, and thus $\{e_i\}$ is a complete orthonormal sequence.

Q3.17 *Let $\{y_1, \dots, y_n\}$ be LI and suppose $\{e_1, \dots, e_n\}$ are obtained from the y_is by the Gram-Schmidt procedure. Let*

$$\hat{x} = \sum_{i=1}^{n} (x|e_i)e_i = \sum_{i=1}^{n} \alpha_i y_i.$$

Show that the coefficients α_i can be easily obtained from the Fourier coefficients $(x|e_i)$.

We know that $y_i = \sum_{j=1}^{n} (y_i|e_j)e_j$ as the y_is lie in the space spanned by $\{e_i\}$. Hence

$$\sum_{j=1}^{n} (x|e_j)e_j = \hat{x} = \sum_{i=1}^{n} \alpha_i y_i = \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{n} (y_i|e_j)e_j = \sum_{j=1}^{n} \sum_{i=1}^{n} \alpha_i (y_i|e_j)e_j$$

Hence we have $(x|e_j) = \sum_{i=1}^{n} \alpha_i (y_i|e_j)$ for all $j = 1, \dots, n$. Thus the various α_i be be obtained by solving the linear system

$$\begin{bmatrix} (y_1|e_1) & (y_2|e_1) & \dots & (y_n|e_1) \\ (y_1|e_2) & (y_2|e_2) & \dots & (y_n|e_2) \\ \vdots & \vdots & \ddots & \vdots \\ (y_1|e_n) & (y_2|e_n) & \dots & (y_n|e_n) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} (x|e_1) \\ (x|e_2) \\ \vdots \\ (x|e_n) \end{bmatrix}$$

Q3.21 *Using the projection theorem, solve the finite-dimensional problem:*

$$\begin{aligned} &\min x^\top Q x \\ &\textit{subject to } Ax = b \end{aligned}$$

where x is an n-vector, Q a positive-definite symmetric $n \times n$ matrix, A an $m \times n$ $(m < n)$, and b an m-vector.

Define the inner product $(x|y) = x^\top Q y$. This is linear and the positive-definiteness of Q guarantees positive definiteness of the inner product. The problem can be formulated as a minimum norm problem: minimise $\|x\|^2$ in $\mathbf{R}^n$ such that $Ax = b$. Write

$$A = \begin{bmatrix} a_1^\top \\ \vdots \\ a_m^\top \end{bmatrix}$$

where we must be careful to note that a_i is the i^{th} **row** of A, and not the i^{th} column. The constraint may be written as $a_i^\top x = b_i$ for $i = 1, \dots, m$. We may rewrite this as $a_i^\top Q^{-1} Q x = b_i$, and hence $(Q^{-1}a_i|x) = b_i$ (note that Q being symmetric implies that Q^{-1} is also symmetric). By the theorem, the (unique) solution takes the form

$$x_0 = \sum_{i=1}^{m} \beta_i Q^{-1} a_i = Q^{-1} \begin{bmatrix} a_1 & \dots & a_m \end{bmatrix} \beta = Q^{-1} A^\top \beta$$

where the β_i are chosen to satisfy $Ax_0 = b$. Expanding, we have

$$Ax_0 = AQ^{-1}A^\top \beta$$

Notice that if A is full rank, then by the positive definiteness of Q, we have $AQ^{-1}A^\top$ is full rank and therefore $\beta = (AQ^{-1}A^\top)^{-1}b$. Therefore, $x_0 = Q^{-1}A^\top(AQ^{-1}A^\top)^{-1}b$, although this seems wrong.

Q3.24 The following theorem is valid in a Hilbert space H. If K is a closed convex set in H and $x \in H$, $x \notin K$; there is a unique vector $k_0 \in K$ such that $\|x - k_0\| \leq \|x - k\|$ for all $k \in K$.

Show that this theorem does not apply in arbitrary Banach space.

Consider the space $\mathbf{R}^2$ endowed with the supremum norm. The set $K = \mathbf{R}_+ \times \mathbf{R}$ is closed (by the equivalence of the p-norms) and convex. Consider $x = (-1, 0)$. Then $\min_{k \in \mathbf{R}_+ \times \mathbf{R}} \|x - k\| = 1$, but candidates for the argmin k_0 include the entire set $\{0\} \times [-1, 1]$.

4 Dual Space Exercises

Q5.1 *Define linear functional f on $L_2[0,1]$ by*

$$f(x) = \int_0^1 a(t) \int_0^t b(s)x(s)\ ds\ dt$$

where $a, b \in L_2[0,1]$. Show that f is a bounded linear functional and find $y \in L_2$ such that $f(x) = (x|y)$.

Let $c(t) = \int_0^t b(s)x(s)\ ds$. Then $f(x) = (a|c)$. By the Cauchy-Schwarz inequality, $|f(x)| \leq \|a\|\|c\|$. Now

$$\|c\| = \int_0^1 \left| \int_0^t b(s)x(s)ds \right|^2 dt \leq \int_0^1 \left(\int_0^t |b(s)x(s)|\, ds \right)^2 dt \leq \int_0^1 \int_0^t |b(s)|^2 ds \int_0^t |x(s)|^2 ds\ dt$$

where the first inequality is the triangular inequality for integrals and the second inequality is the Cauchy-Schwarz inequality. Hence

$$\|c\| \leq \int_0^1 \int_0^1 |b(s)|^2 ds \int_0^1 |x(s)|^2 ds\ dt = \|b\|\|x\| \int_0^1 dt = \|b\|\|x\|$$

Hence $|f(x)| \leq \|a\|\|b\|\|x\|$ and thus f is a bounded linear functional. Note that

$$f(x) = \int_0^1 \int_0^t a(t)b(s)x(s)\ ds\ dt = \int_0^1 \int_s^1 a(t)b(s)x(s)\ dt\ ds$$

by considering the region in $\mathbf{R}^2$ over which integration is taken. Hence

$$f(x) = \int_0^1 b(s) \int_s^1 a(t)\ dt\ x(s)\ ds$$

so $y(s) = b(s) \int_s^1 a(t)\ dt$.

Q5.2 *Characterise the dual space of c, the space of convergent sequences.*

Let $x \in c$. Then

$$|f(x)| = |f(\textstyle\sum_{i=1}^\infty x_i e_i)| = \left| \displaystyle\sum_{i=1}^\infty x_i f(e_i) \right| \leq \|x\|_\infty \left| \sum_{i=1}^\infty y_i \right| = \|x\|_\infty \|y\|_1$$

where the second equality follows from the linearity of f. As $x \in c$ arbitrary, $\|f\| \leq \|y\|_1$.

Conversely, consider $x^{(N)}$ such that $x_i^{(N)} = \text{sign}(y_i)\mathbf{1}_{[i \leq N]}$. As $x_i^{(N)} \to 0$, we have $x^{(N)} \in c$, and $\|x^{(N)}\| = 1$. We have

$$f(x^{(N)}) = f(\textstyle\sum_{i=1}^\infty x_i^{(N)} e_i) = \displaystyle\sum_{i=1}^N |y_i| = \sum_{i=1}^N |y_i| \|x^{(N)}\|_\infty$$

recalling that $\|x^{(N)}\|_\infty = 1$. By the definition of the operator norm, $\sum_{i=1}^N |y_i| \leq \|f\|$. As $N \in \mathbf{N}$ arbitrary, we can take $N \to \infty$ and deduce that $\|y\|_1 = \sum_{i=1}^\infty |y_i| \leq \|f\|$.

Combining the two inequalities, we have $\|y\|_1 = \|f\|$. Hence the dual of c is ℓ_1, where we associate each $y \in \ell_1$ with the bounded linear functional $f(x) = \sum_{i=1}^{\infty} x_i y_i$.

Q5.3 *Let X^* be the dual of normed space X. Show that if X^* is separable, then X is separable.*

Q5.4 *Show that the normed space $C[a,b]$ is not reflexive.*

$C[a,b]$ is separable by the Stone-Weierstrass Theorem. Consider δ_x, δ_y, the Dirac delta function for $x, y \in [a,b]$ and $x \neq y$. Then $\|\delta_x - \delta_y\| = 2$ in the total variation norm. Hence there are uncountably infinite disjoint open balls of the form $B(\delta_x, 1)$ with $x \in [a,b]$, all of which are subsets of $NBV[a,b]$. Hence $NBV[a,b]$ is not separable, and its dual cannot be the separable space $C[a,b]$ by Question 5.3.

Q5.5 *Prove that $x \in L_p$ is aligned with $y \in L_q$ iff $x(t) = K \operatorname{sign}(y(t))|y(t)|^{q/p}$.*

If:

$$\langle y, x\rangle = \int_a^b x(t)y(t)\ dt = K\int_a^b |y(t)|^{1+q/p}\ dt = K\int_a^b |y(t)|^q\ dt$$

and we have

$$\|x\| = \left(\int_a^b |x(t)|^p\ dt\right)^{1/p} = K\left(\int_a^b |y(t)|^q\ dt\right)^{1/p}$$

and

$$\|y\| = \left(\int_a^b |y(t)|^q\ dt\right)^{1/q}.$$

Hence

$$\|x\|\|y\| = K\left(\int_a^b |y(t)|^q\ dt\right)^{1/p+1/q} = K\int_a^b |y(t)|^q\ dt = \langle y, x\rangle$$

Only if: the proof of Hölder's Inequality proceeds by Young's Inequality, which shows that for any positive s, t that $st \leq s^p/p + t^q/q$. As this proof uses the concavity of the logarithm, we have inequality if $s^p = t^q$, and Hölder uses $s = |x|/\|x\|$ and $t = |y|/\|y\|$. Thus, $\langle |y|, |x|\rangle = \|x\|\|y\|$ if and only if $(|x|/\|x\|)^p = (|y|/\|y\|)^q$. Thus, $|x|^p = |y|^q \times \|x\|^p/\|y\|^q$. Setting $K^p = \|x\|^p/\|y\|^q$, we have $|x|^p = K^p|y|^q$ and $|x| = K|y|^{q/p}$. Finally, for alignment, we need to ensure that $\langle y, x\rangle = \langle |y|, |x|\rangle$, which is guaranteed if $\operatorname{sign} x = \operatorname{sign} y$.

Q5.6