

A Function-Space Tour of Data Science: Banach Spaces and Sparse Methods

© 2026 Rahul Parhi. All rights reserved.

Disclaimer: *These notes carry no warranty of correctness, completeness, or proper citation of sources.*

1 Dual Spaces and Predual Spaces

Definition 1.1 (Dual space). Let \mathcal{F} be a normed space. The (*continuous*) *dual space* of \mathcal{F} is denoted by \mathcal{F}' and is the space of continuous linear functionals on \mathcal{F} . We write the canonical pairing as $\langle \cdot, \cdot \rangle : \mathcal{F} \times \mathcal{F}' \rightarrow \mathbb{R}$, i.e., if $f \in \mathcal{F}$ and $g \in \mathcal{F}'$, $\langle f, g \rangle = g[f]$.

Definition 1.2 (Bidual and canonical embedding). The *bidual* is $\mathcal{F}'' := (\mathcal{F}')'$. There is a canonical linear map $\iota : \mathcal{F} \rightarrow \mathcal{F}''$ defined by

$$(\iota(f))(g) := g[f] = \langle f, g \rangle, \quad g \in \mathcal{F}'. \quad (1)$$

Remark 1.3. We always have that $\|\iota(f)\|_{\mathcal{F}''} = \|f\|_{\mathcal{F}}$ (so ι is an isometry). In particular, we can always identify $\mathcal{F} \subset \mathcal{F}''$. If ι is onto, we say \mathcal{F} is *reflexive* and we can identify $\mathcal{F} = \mathcal{F}''$.

Definition 1.4 (Dual norm). For $g \in \mathcal{F}'$, we define the *dual norm* as

$$\|g\|_{\mathcal{F}'} := \sup_{\substack{f \in \mathcal{F} \\ \|f\|_{\mathcal{F}} \leq 1}} |\langle f, g \rangle|. \quad (2)$$

When endowed with this norm, the dual space \mathcal{F}' is a Banach space.

Definition 1.5 (Predual). A Banach space \mathcal{X} is a *predual* of a Banach space \mathcal{F} if $\mathcal{F} = \mathcal{X}'$ (isometrically). In this case we write the pairing as $\langle g, f \rangle$ for $g \in \mathcal{X}$ and $f \in \mathcal{F} = \mathcal{X}'$.

Remark 1.6. Not every Banach space admits a predual. For example, $L^1(\mathbb{R})$ has no predual.

Definition 1.7 (Strong and weak convergence). Let \mathcal{F} be a Banach space. A sequence $\{f_n\}_{n \in \mathbb{N}} \subset \mathcal{F}$ converges *strongly* (in norm) to $f \in \mathcal{F}$ if

$$\|f_n - f\| \rightarrow 0 \text{ as } n \rightarrow \infty.$$

A sequence $\{f_n\}_{n \in \mathbb{N}} \subset \mathcal{F}$ converges *weakly* to $f \in \mathcal{F}$ if

$$\langle f_n, g \rangle \rightarrow \langle f, g \rangle \text{ as } n \rightarrow \infty \text{ for all } g \in \mathcal{F}'.$$

Remark 1.8. The weak topology is still typically too fine for closed and bounded sets to be compact.

Definition 1.9 (Weak* convergence). Let \mathcal{F} be a Banach space such that $\mathcal{F} = \mathcal{X}'$ is a dual space with predual \mathcal{X} . A sequence $\{f_n\}_{n \in \mathbb{N}} \subset \mathcal{F}$ converges *weak** to $f \in \mathcal{F}$ if

$$\langle g, f_n \rangle \rightarrow \langle g, f \rangle \text{ as } n \rightarrow \infty \text{ for all } g \in \mathcal{X}. \quad (3)$$

Fact 1.10. *If \mathcal{H} is a Hilbert space, then the weak and weak* topologies coincide (after identifying $\mathcal{H} = \mathcal{H}'$ via the Riesz Representation Theorem). More generally, if \mathcal{F} is reflexive, then the weak and weak* coincide (by the identification $\mathcal{F} = \mathcal{F}''$).*

2 Banach–Alaoglu and Dixmier–Ng Theorems and Compactness

Theorem 2.1 (Banach–Alaoglu). *Let \mathcal{X} be a Banach space and let $\mathcal{F} = \mathcal{X}'$. Then the closed unit ball*

$$B_{\mathcal{F}} := \{f \in \mathcal{F} : \|f\|_{\mathcal{F}} \leq 1\} \quad (4)$$

is compact in the weak topology on \mathcal{F} (induced by \mathcal{X}).*

Remark 2.2. For optimization over dual Banach spaces, the weak* topology is often the “correct” topology to work with.

Theorem 2.3 (Dixmier–Ng). *Let \mathcal{F} be a normed space and let τ be a locally convex Hausdorff topology on \mathcal{F} . The following are equivalent:*

- (1) *The closed unit ball $B_{\mathcal{F}}$ is τ -compact.*
- (2) *There exists a Banach space \mathcal{X} such that $\mathcal{F} = \mathcal{X}'$.*

Remark 2.4. The Dixmier–Ng establishes a fundamental impossibility result: If a Banach space does not have a predual, there does not exist any “reasonable” topology such that its closed unit ball is compact.

3 Riesz–Markov–Kakutani Representation Theorem

Theorem 3.1 (Riesz–Markov–Kakutani). *Let Ξ be a compact (resp. locally compact) Hausdorff space. Let $C(\Xi)$ (resp. $C_0(\Xi)$) denote continuous real-valued functions on Ξ (continuous real-valued functions on Ξ vanishing at infinity) endowed with the L^∞ -norm. Then, $C(\Xi)' = \mathcal{M}(\Xi)$ (resp. $C_0(\Xi)' = \mathcal{M}(\Xi)$), where $\mathcal{M}(\Xi)$ is the space of finite (signed) Radon measures on Ξ .*

4 Fisher–Jerome Theorem

Let $\varphi : \Omega \times \Xi \rightarrow \mathbb{R}$ be such that for each x , $\varphi(x, \cdot)$ is in $C(\Xi)$ (resp. $C_0(\Xi)$) when Ξ is a compact (resp. locally compact) Hausdorff space. Define the notation $\varphi_\xi(x) := \varphi(x, \xi)$. Furthermore, define

$$f_\nu(x) := \int_{\Xi} \varphi_\xi(x) d\nu(\xi). \quad (5)$$

Theorem 4.1. *If, for every $y \in \mathbb{R}$, the loss function $\mathcal{L}(y, \cdot) : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous (l.s.c.), $\mathcal{L}(y, 0) < +\infty$,¹ and bounded from below, the solution set*

$$S := \arg \min_{\nu \in \mathcal{M}(\Xi)} \left\{ J(\nu) = \sum_{i=1}^n \mathcal{L}(y_i, f_\nu(x_i)) + \lambda \|\nu\|_{\mathcal{M}} \right\}, \quad \lambda > 0, \quad (6)$$

is nonempty. Furthermore, there always exists a solution $\nu^ \in S$ that yields the representation*

$$f_{\nu^*} = \sum_{k=1}^K v_k \varphi_{\xi_k}, \quad K \leq n. \quad (7)$$

Proof. Since $\mathcal{L}(y, \cdot)$ is l.s.c., J is weak* l.s.c. Indeed,

- the functional $\nu \mapsto f_\nu(x)$ is weak* continuous for all $x \in \Omega$ since

$$f_\nu(x) = \int_{\Xi} \varphi(x, \xi) d\nu(\xi) = \langle \varphi(x, \cdot), \nu \rangle \quad (8)$$

and $\varphi(x, \cdot) \in C(\Xi)$ (resp. $C_0(\Xi)$) when Ξ is compact (resp. locally compact).

¹This ensures that J is proper on $\mathcal{M}(\Xi)$.

- the norm $\|\cdot\|_{\mathcal{M}}$ is weak* l.s.c. on $\mathcal{M}(\Xi)$ since

$$\|\nu\|_{\mathcal{M}} = \sup_{\substack{\varphi \in C(\Xi) \text{ or } C_0(\Xi) \\ \|\varphi\|_{L^\infty} = 1}} |\langle \varphi, \nu \rangle|, \quad (9)$$

where we recall that the sup of weak* continuous functions is weak* l.s.c.

Since $\mathcal{L}(y, \cdot)$ is bounded from below, J is coercive. Indeed, J is the sum of something bounded from below and something coercive (all norms are obviously coercive). Since J is weak* l.s.c., it has weak* closed sublevel sets. Furthermore, since J is coercive, it has bounded sublevel sets. Thus, it has weak* compact sublevel sets and hence minimizers exist.

Let $\widehat{\nu} \in S$ be any minimizer and set $z_i := f_{\widehat{\nu}}(x_i)$ for $i = 1, \dots, n$. Consider the auxiliary problem

$$\min_{\nu \in \mathcal{M}(\Xi)} \|\nu\|_{\mathcal{M}} \quad \text{s.t.} \quad f_{\nu}(x_i) = z_i, i = 1, \dots, n. \quad (10)$$

Its feasible set is nonempty (it contains $\widehat{\nu}$). Let S_z be its solution set. Because the constraint set is weak*-closed and the objective is weak* l.s.c. with weak* compact sublevel sets, S_z is nonempty, weak* compact, and convex. Thus, S_z has extreme points by the Krein–Milman theorem. Finally, any solution to (10) is a solution to the original problem (6). Thus, it suffices to show that there exists a solution to (10) that is supported on at most n points.

Let ν^* be an extreme point of S_z . We claim that ν^* is supported on at most n points. Assume for contradiction that ν^* is supported on more than n points. Let $\nu^* = \nu_+^* - \nu_-^*$ be its Jordan decomposition and let $\Xi = P \sqcup N$ be a Hahn decomposition for ν^* (so $\nu^* \geq 0$ on P and $\nu^* \leq 0$ on N). Since $|\nu^*|(\Xi) > 0$ and $\text{supp}(|\nu^*|)$ has more than n points, we can find pairwise disjoint Borel sets $A_1, \dots, A_{n+1} \subset \Xi$ such that $|\nu^*|(A_j) > 0$ for each j , and each A_j lies entirely in P or entirely in N . Finally, define $A_0 = \Xi \setminus \bigcup_{j=1}^{n+1} A_j$. Define the restricted signed measures

$$\nu_j := \nu^*|_{A_j}, \quad j = 0, \dots, n+1. \quad (11)$$

Then, $\nu^* = \sum_{j=0}^{n+1} \nu_j$, and the signed measures $\nu_0, \nu_1, \dots, \nu_{n+1}$ are pairwise mutually singular.

For each j , define $b_j \in \mathbb{R}^n$ by

$$(b_j)_i := f_{\nu_j}(x_i), \quad i = 1, \dots, n. \quad (12)$$

Since we have $n+1$ vectors in \mathbb{R}^n , they are linearly dependent. Thus, there exist scalars a_1, \dots, a_{n+1} , not all zero, such that $\sum_{j=1}^{n+1} a_j b_j = 0$. Let

$$\mu := \sum_{j=1}^{n+1} a_j \nu_j. \quad (13)$$

Then for each i ,

$$f_{\mu}(x_i) = \sum_{j=1}^{n+1} a_j f_{\nu_j}(x_i) = \sum_{j=1}^{n+1} a_j (b_j)_i = 0, \quad (14)$$

so $\nu^* + t\mu$ satisfies the constraints in (10) for all $t \in \mathbb{R}$.

Because each ν_j has constant sign, there exists $\varepsilon > 0$ such that $1 \pm \varepsilon a_j \geq 0$ for all j (e.g., take $\varepsilon \leq \min_{a_j \neq 0} 1/(2|a_j|)$). For such an ε ,

$$\nu^* \pm \varepsilon \mu = \nu_0 + \sum_{j=1}^{n+1} (1 \pm \varepsilon a_j) \nu_j, \quad (15)$$

and no sign flips occur within each A_j . Consequently, for $t \in [-\varepsilon, \varepsilon]$ the map $t \mapsto \|\nu^* + t\mu\|_{\mathcal{M}}$ is affine:

$$\|\nu^* + t\mu\|_{\mathcal{M}} = \|\nu_0\|_{\mathcal{M}} + \sum_{j=1}^{n+1} \|(1 + ta_j)\nu_j\|_{\mathcal{M}} = \|\nu_0\|_{\mathcal{M}} + \sum_{j=1}^{n+1} (1 + ta_j) \|\nu_j\|_{\mathcal{M}}. \quad (16)$$

where we used the fact that the A_j are disjoint. But ν^* minimizes $\nu \mapsto \|\nu\|_{\mathcal{M}}$ over the feasible affine set, so $t = 0$ is a minimizer of this affine function on $[-\varepsilon, \varepsilon]$. Therefore its slope must be zero, i.e.,

$$\sum_{j=1}^{n+1} a_j \|\nu_j\|_{\mathcal{M}} = 0, \quad (17)$$

and hence

$$\|\nu^* \pm \varepsilon \mu\|_{\mathcal{M}} = \|\nu^*\|_{\mathcal{M}}. \quad (18)$$

Thus, $\nu^* \pm \varepsilon \mu \in S_z$ and they are distinct (since $\mu \neq 0$). Thus,

$$\nu^* = \frac{1}{2}(\nu^* + \varepsilon \mu) + \frac{1}{2}(\nu^* - \varepsilon \mu), \quad (19)$$

contradicting that ν^* is an extreme point of S_z .

We conclude that ν^* is supported on at most n points, i.e.,

$$\nu^* = \sum_{k=1}^K v_k \delta_{\xi_k}, \quad K \leq n, \quad (20)$$

and therefore,

$$f_{\nu^*} = \int_{\Xi} \varphi_{\xi} d\nu^*(\xi) = \sum_{k=1}^K v_k \varphi_{\xi_k}, \quad (21)$$

which completes the proof. \square

Remark 4.2. The \mathcal{M} -norm is not strictly convex, so uniqueness of minimizers for (6) is hard to guarantee, just like finite-dimensional ℓ^1 -minimization problems. In finite dimensions, compressed sensing gives conditions implying uniqueness. In infinite dimensions, various “off-the-grid” analogues exist which provide conditions for uniqueness using tools from convex duality.