Duality in Mathematics

Zachary McNulty

Duality is a somewhat ill-defined principle in mathematics, but it generally describes the idea that problems in one setting can often be translated to problems in another setting in a one-to-one fashion, often in the form of an *involution*. This becomes particularly useful when the dual space has some structure that the primal space lacks, allowing us to apply a different set of tools to study the same underlying problem, or when the problem somehow simplifies dramatically in the dual setting. In this note we will explore several realizations of this phenomenon. These examples do not all arise from a single formal construction; rather, they share the recurring themes of translation, reversal, and the recovery of one object from another.

1 Motivating Examples of Duality

We begin with some simple examples and move toward more sophisticated applications. Some of the more complicated examples can be skipped on a first reading.

1.1 Set Complements

Given a subset $A \subseteq S$ of some set S, properties of A can often be framed in terms of its complement A^c. For example, in topology a set A is closed if and only if A^c is open. This allows us to state many topological properties in terms of either open sets or closed sets, for example:

- *Continuity*: the preimage of every open set is open if and only if the preimage of every closed set is closed.
- *Compactness*: an open cover $\{U_\alpha\}$ of K has a finite subcover if and only if K^c contains a finite intersection of the U_α^c.
- *Closure and interior*: $\mathrm{Int}(S)^c = \overline{S^c}$.
- *Denseness*: every open set $U \subseteq X$ intersects S if and only if $\overline{S} = X$.
- *The T_1 separation property*: every singleton is closed if and only if $\{x\} = \bigcap_{F \in C_x} F$, where $C_x = \{F \text{ closed} : x \in F\}$.

The map $A \mapsto A^c$ is an involution, and $A \subseteq B$ implies $A^c \supseteq B^c$, so it reverses the natural inclusion order. Here, duality is the idea that in the framework of topology, we can often translate statements about open sets to statements about closed sets, and vice versa.

1.2 Dual Vector Spaces

Given a normed vector space V, define its *continuous dual space* by

$$V^* := \{\phi : V \to \mathbb{R} : \phi \text{ is a bounded linear functional}\}.$$

If V is a real Hilbert space, the Riesz representation theorem says that every $\phi \in V^*$ has a unique representation

$$\phi(x) = \langle x, v_\phi \rangle$$

for some $v_\phi \in V$. The correspondence $\phi \mapsto v_\phi$ is a canonical isometric isomorphism $V^* \cong V$. This duality is useful because V^* is a space of scalar-valued functions, allowing analytic statements about functionals to be translated into geometric statements about vectors.

For example, the derivative of a differentiable functional $F : V \to \mathbb{R}$ is naturally an element $DF(x) \in V^*$. Riesz representation produces the gradient $\nabla F(x) \in V$ characterized by

$$DF(x)[h] = \langle \nabla F(x), h \rangle \qquad \text{for every } h \in V.$$

Thus the familiar gradient is the vector representative of the derivative, an observation underlying gradient-based optimization in Hilbert spaces.

A bounded linear map $T : V \to W$ induces a dual map

$$T' : W^* \to V^*, \qquad T'(\phi) := \phi \circ T.$$

Notice that the direction of the arrow is reversed. When V and W are Hilbert spaces, the Riesz identifications turn T' into the Hilbert-space adjoint $T^* : W \to V$, characterized by

$$\langle Tx, y \rangle_W = \langle x, T^* y \rangle_V.$$

Finally, the evaluation map $V \to V^{**}$ is an isometric isomorphism for a Hilbert space, so Hilbert spaces are *reflexive*. In this precise sense, taking the continuous dual twice recovers the original space.

1.3 Dual Cones

Given $C \subseteq V$, define its *dual cone* by

$$C^* := \{v \in V^* : \langle v, x \rangle \geq 0 \quad \text{for every } x \in C\},$$

where $\langle v, x \rangle := v(x)$ applies the dual functional to x. Again, this duality reverses the natural inclusion order. The bipolar theorem makes the recovery statement precise:

$$C^{**} = \overline{\operatorname{cone}(\operatorname{conv} C)}.$$

In particular, $C^{**} = C$ whenever C is a closed convex cone.

Dual cones arise naturally in constrained optimization, particularly in the Karush–Kuhn–Tucker (KKT) conditions discussed below. Near a candidate solution x, the primal constraints determine a cone of feasible directions in which x may move while remaining feasible. Passing to the dual cone translates these constraints on the input directions into constraints on the gradients at x. This is the geometric mechanism by which the KKT conditions convert a constrained primal problem into conditions involving the objective and constraint gradients.

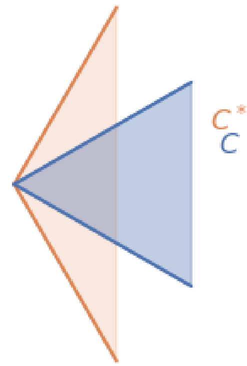

Figure 1: A cone and its dual cone.

1.4 Dual Graphs

Any planar graph G divides two-dimensional space into a collection of closed regions, called *faces*, together with a possibly unbounded exterior region. We define another graph G^* by adding one vertex for each face and an edge between two vertices whenever their corresponding faces are adjacent along an edge of G. Generally, G^* depends on the chosen planar embedding, but many of its properties are invariant under the embedding. Here duality is a genuine involution for connected plane graphs: once the embedding is fixed, $G^{**} \cong G$, with bridges and loops interpreted in the usual plane-multigraph sense.

The graph G^* encodes connectivity in the original graph. One manifestation of this is the relationship between cycles and cut sets. A cut set is a subset of the edges of G obtained by partitioning its vertices into two parts and collecting the edges between those parts. Since there is a bijection between the edges of G and G^*, a cycle in G induces a cut set in G^*, partitioning the dual vertices into the faces inside and outside the cycle. Thus minimal cycles in G correspond to minimal cut sets in G^*, and vice versa.

One classical example comes from bond percolation on the lattice $\mathbb{Z}^2$. Suppose each edge is independently retained with probability p to form a random graph G. We are interested in the probability that there is a path in G from the origin to infinity. Such a path exists if and only if the dual graph contains no closed dual cycle separating the origin from infinity.

Lemma 1. *There exists $p < 1$ such that $\mathbb{P}(0 \leftrightarrow \infty) > 0$.*

Proof. This is a classical Peierls-style argument. For a fixed dual circuit γ of length n, every crossed primal edge must be absent, an event of probability $(1-p)^n$. A circuit of length n surrounding the origin must cross a fixed ray from the origin within distance at most n. After choosing such a crossing edge, there are at most three non-backtracking choices at each subsequent step. Thus, for some absolute constant C, the number of candidate circuits is at most $Cn3^n$. If the origin is not connected to infinity, some closed dual circuit must surround it. Therefore, by the union bound,

$$\mathbb{P}(0 \not\leftrightarrow \infty) \le C \sum_{n \ge 4} n[3(1-p)]^n,$$

which tends to zero as $p \uparrow 1$ and is therefore less than one for p sufficiently close to one. $\square$

1.5 Duality in Population Genetics

In population genetics, we study how the genetic makeup of a population evolves over time due to randomness in reproduction, natural selection, and mutation. In the simplest case, suppose a gene has two alleles, a and A. The same process can often be viewed under two lenses: forward in time, by tracking how alleles are passed from one generation to the next, or backward in time, by taking present-day individuals and tracing their ancestral lineages.

The Wright–Fisher model assumes that N individuals in each generation independently sample a parent from the previous generation and inherit its allele. If X_t is the number of individuals carrying A at generation t, then

$$X_{t+1} \sim \text{Binomial}(N, p_t), \qquad p_t := \frac{X_t}{N}.$$

To obtain a nontrivial diffusion limit with selection, introduce a weak bias α/N favoring A and set

$$X_{t+1} \sim \text{Binomial}(N, p_t), \qquad p_t := \frac{(1+\alpha/N)X_t}{N + (\alpha/N)X_t}.$$

After taking $N \to \infty$ and accelerating time by a factor of N, one obtains the Wright–Fisher diffusion

$$dp_t = \alpha p_t(1-p_t)\,dt + \sqrt{p_t(1-p_t)}\,dW_t.$$

Both the drift in allele frequency and the random fluctuations depend on $p(1-p)$, which measures, in a sense, the distance from fixation.

The backward-time dynamics can be simpler. We seek a function $H(p,n)$ relating the forward and backward processes such that

$$\frac{d}{du}\mathbb{E}[H(p(u), n(t-u))] = 0, \qquad 0 \le u \le t,$$

and hence

$$\mathbb{E}[H(p(t), n(0))] = \mathbb{E}[H(p(0), n(t))].$$

If L and G are the generators of p_t and n_t, respectively, this is suggested by finding a duality function $H(p,n)$ satisfying

$$L_p H(p,n) = G_n H(p,n),$$

where the subscripts indicate which variable the generator acts upon.

In the neutral case $\alpha = 0$, p_t has generator

$$Lg(p) = \frac{p(1-p)}{2} g''(p).$$

The corresponding backward-time process is Kingman's coalescent. Its number of remaining ancestral lineages $n(t)$ is a pure-death process on $\mathbb{N}$ with generator

$$Gg(n) = -\binom{n}{2}\big(g(n) - g(n-1)\big).$$

These generators suggest $H(p,n) = p^n$, because

$$L_p H(p,n) = G_n H(p,n) = -\binom{n}{2}(p^n - p^{n-1}).$$

This gives the duality relationship

$$\mathbb{E}[p(t)^{n(0)}] = \mathbb{E}[p(0)^{n(t)}].$$

Both sides give the probability that a sample of $n(0)$ present-day individuals all carry A. The left side samples directly from the present population. The right side evolves the sample backward to its ancestral lineages and asks for the probability that each ancestor carried A.

This duality also determines the fixation probability. Since $n(t) = 1$ almost surely as $t \to \infty$ in Kingman's coalescent, the right-hand expectation converges to $p(0)$. Consequently, the probability that A ultimately fixes is $p(0)$, a conclusion that is much easier to obtain from the dual process than directly from the diffusion.

2 Duality in Convex Analysis

2.1 Convex Sets and Supporting Hyperplanes

A set $X \subseteq V$ is *convex* if it is closed under affine combinations:

$$x, y \in X \quad \Longrightarrow \quad \lambda x + (1-\lambda)y \in X \quad \text{for every } \lambda \in [0, 1].$$

Given X, a *supporting hyperplane* is a hyperplane that intersects ∂X and leaves all of X in one of its two closed half-spaces. More formally, $H = \{x : \ell(x) = b\}$ for $\ell \in V^*$ and $b \in \mathbb{R}$ supports X if

$$\ell(x) \leq b \quad \text{for every } x \in X, \qquad \ell(x_0) = b \quad \text{for some } x_0 \in X.$$

Such a hyperplane is tangent, in a generalized sense, to the boundary of X.

A fundamental result in convex analysis says that if X is closed and convex in a suitably nice topological vector space, then each boundary point admits a supporting hyperplane. In particular, X is the intersection of all closed half-spaces that contain it. Since these half-spaces are defined by dual elements $\ell \in V^*$, this creates a natural relationship between a convex set and the dual space.

The *support function* $\sigma_X : V^* \to \mathbb{R} \cup \{+\infty\}$ is

$$\sigma_X(\ell) := \sup_{x \in X} \ell(x).$$

It encodes the supporting level associated with each dual direction, and

$$X = \bigcap_{\ell \in V^*} \{x \in V : \ell(x) \leq \sigma_X(\ell)\}.$$

Thus convex analysis contains a duality between convex sets and linear functionals, with σ_X encoding the relationship.

2.2 Convex Conjugates

A set X is closed and convex if and only if its indicator function $\iota_X : V \to \overline{\mathbb{R}}$, where $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$,

$$\iota_X(x) := \begin{cases} 0, & x \in X, \\ +\infty, & x \notin X, \end{cases}$$

is a closed convex function. The support function can then be written as

$$\sigma_X(\ell) = \sup_{x \in V} \{\ell(x) - \iota_X(x)\}.$$

Here ι_X encodes the hard constraint $x \in X$, while $\sigma_X(\ell)$ records the maximal value of the objective ℓ subject to that constraint. For example, if $X = \{x : \|x\|_2 \le 1\}$ is the Euclidean unit ball, then

$$\sigma_X(\ell) = \sup_{x \in X} \ell(x) = \|\ell\|_2,$$

where in the last equality we identify ℓ with its Riesz representative.

For a general convex function $\phi : V \to \overline{\mathbb{R}}$, define its *convex conjugate* $\phi^* : V^* \to \overline{\mathbb{R}}$ by

$$\phi^*(\ell) := \sup_{x \in V}\{\ell(x) - \phi(x)\}.$$

Instead of imposing a hard constraint, we introduce a convex potential ϕ. The conjugate records the optimal balance between maximizing the linear objective ℓ and minimizing the potential ϕ.

There is also a geometric interpretation. A function $f : V \to \mathbb{R}$ is convex if and only if its epigraph

$$\mathrm{epi}(f) := \{(x, y) \in V \times \mathbb{R} : y \ge f(x)\}$$

is convex. The convex conjugate is defined so that the line

$$g_\ell(x) := \langle \ell, x\rangle - f^*(\ell) \le f(x) \qquad \text{for every } x \in V.$$

It therefore selects the highest affine function $g_\ell(x)$ of slope ℓ that lies entirely below f. In one dimension these supporting hyperplanes are simply supporting lines.

In general, $f^{**} \le f$, but if f is a proper, lower semi-continuous, convex function then $f^{**} = f$ by the Fenchel–Moreau theorem. Geometrically, this is almost obvious: the convex conjugate of a function encodes the set of affine functions, that lie completely below f, and the biconjugate then takes the supremum of all these lines. Since any nice enough convex function equals the supremum of all affine functions that lie below it, we get equality. Hence again we see duality manifests itself as an involution, at least when restricted to the proper class of objects.

Algebraically, after identifying V^{**} with V, we see

$$f^{**}(x) = \sup_{\ell \in V^*}\{\ell(x) - f^*(\ell)\} = \sup_{\ell \in V^*} g_\ell(x) \le f(x).$$

2.3 Duality in Convex Optimization

Suppose we want to solve the convex optimization problem

$$\min_x f(x) \qquad \text{subject to} \qquad g_i(x) \le 0, \quad i \in [m], \tag{1}$$

where f and the g_i are convex. For simplicity, assume all these functions are differentiable. Below, we discuss two geometric interpretations of duality in this setting.

2.3.1 Lagrangian Duality

Lagrangian duality begins by observing if x^* minimizes (1), then

$$\langle -\nabla f(x^*), v\rangle \le 0$$

for every feasible direction v: there is no direction in which we can both remain feasible and decrease the objective. The discussion below gives the geometric intuition; turning these linearized conditions into necessary optimality conditions requires an appropriate constraint qualification.

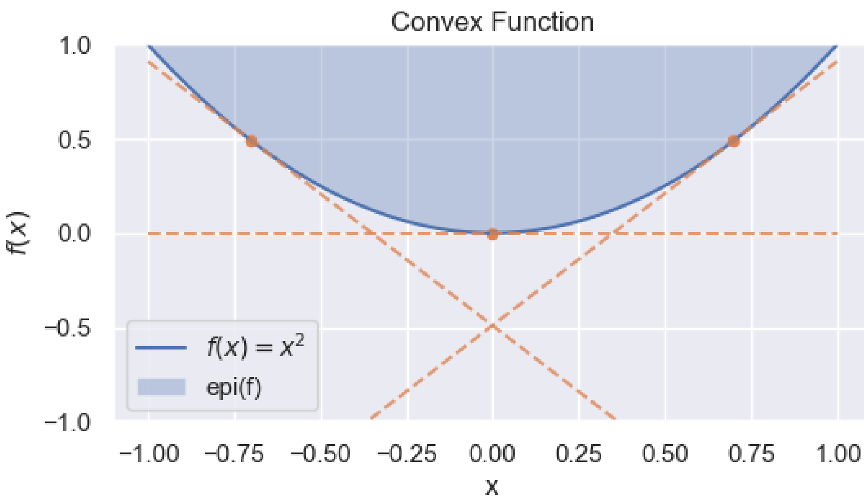

Figure 2: Supporting lines below the convex function $f(x) = x^2$.

First consider a single equality constraint,

$$\min_x f(x) \qquad \text{subject to} \qquad g(x) = 0.$$

At a point x, $\nabla g(x)$ is perpendicular to the level set $\{g = 0\}$, so the first-order optimality condition becomes

$$\nabla f(x) = \lambda \nabla g(x), \qquad \lambda \in \mathbb{R}.$$

With several equality constraints,

$$\min_x f(x) \qquad \text{subject to} \qquad g_i(x) = 0, \quad i \in [m],$$

the feasible directions are perpendicular to every $\nabla g_i(x)$. Thus

$$\nabla f(x) = \sum_{i=1}^{m} \lambda_i \nabla g_i(x), \qquad \lambda_i \in \mathbb{R}.$$

Now consider inequality constraints $g_i(x) \leq 0$. At x^*, a constraint is *inactive* when $g_i(x^*) < 0$ and *active* when $g_i(x^*) = 0$. Writing $A(x) = \{i : g_i(x) = 0\}$, the local feasible directions are

$$F_x = \{v \in \mathbb{R}^n : \langle v, \nabla g_i(x^*) \rangle \leq 0 \text{ for every } i \in A(x^*)\}.$$

The optimality condition can therefore be expressed as

$$-\nabla f(x^*) = \sum_{i \in A(x^*)} \lambda_i \nabla g_i(x^*), \qquad \lambda_i \geq 0.$$

Complementary slackness lets us avoid naming the active set explicitly:

$$-\nabla f(x^*) = \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*), \qquad \lambda_i g_i(x^*) = 0, \qquad \lambda_i \geq 0.$$

These observations are encoded in the *Lagrangian*

$$L(x, \lambda) := f(x) + \sum_{i=1}^{m} \lambda_i g_i(x), \qquad \lambda_i \geq 0.$$

Stationarity in x gives

$$\nabla_x L(x, \lambda) = \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) = 0.$$

For fixed λ, the Lagrangian can also be viewed as a soft version of the original constrained problem. If x is feasible, then $L(x, \lambda) \leq f(x)$ for every $\lambda \geq 0$. Define the dual objective

$$h(\lambda) := \inf_x L(x, \lambda).$$

If x^* solves the primal problem, then

$$h(\lambda) \leq \inf_{x \text{ feasible}} L(x, \lambda) \leq f(x^*).$$

This motivates the dual problem

$$\sup_{\lambda \geq 0} h(\lambda).$$

The inequality $h(\lambda) \leq f(x^*)$ for every $\lambda \geq 0$, and hence

$$\sup_{\lambda \geq 0} h(\lambda) \leq f(x^*),$$

is *weak duality*. Under additional assumptions, such as Slater's condition, equality holds and we obtain *strong duality*:

$$\sup_{\lambda \geq 0} h(\lambda) = \min_{x \text{ feasible}} f(x).$$

Application: Duality in Linear Programming. A linear program is a convex optimization problem whose objective and constraints are linear:

$$\min_x c^\top x \qquad \text{subject to} \qquad Ax \leq b.$$

Its Lagrangian is

$$L(x, \lambda) = c^\top x + \lambda^\top (Ax - b), \qquad \lambda \geq 0.$$

Stationarity enforces the dual constraint

$$\nabla_x L(x, \lambda) = 0 \quad \Longrightarrow \quad A^\top \lambda = -c.$$

Under this constraint,

$$h(\lambda) := \inf_x L(x, \lambda) = -\lambda^\top b.$$

Hence the dual problem is

$$\max_\lambda -\lambda^\top b \qquad \text{subject to} \qquad A^\top \lambda = -c, \quad \lambda \geq 0.$$

2.3.2 Duality Through Perturbation

The key idea is this: given an optimization problem, we don't obtain a dual problem until we specify how to perturb the optimization problem. This is why equivalent formulations of an optimization problem can lead to different dual problems: we simply perturb the system differently.

To be a bit more concrete, suppose our optimaztion problem is of the form

$$\min_x \ \phi(x, 0),$$

for some convex function $\phi : X \times Y \to \overline{R}$. Here, we essentially use $\phi(x, 0) = \infty$ to encode constraints on x. We can then define the perturbed optimization problem as

$$h(y) := \min_x \ \phi(x, y).$$

Our original problem then seeks to calculate $h(0)$. Recall that the convex biconjugate gives a lower bound: $h^{**}(0) \leq h(0)$ (*weak duality*)) and for convex functions we often have equality (*strong duality*) by the Fenchel-Moreau theorem. The dual problem is then to calculate

$$h^{**}(0) = \sup_{\ell \in Y^*} \{-h^*(\ell)\} = -\inf_{\ell \in Y^*} h^*(\ell) = -\inf_{\ell \in Y^*} \phi^*(0, \ell).$$

For example, if we define the perturbated problem by allowing less slack in the constraints

$$\phi(x, y) = \begin{cases} f(x) & \text{if } g_i(x) + y_i \leq 0, \ i = 1, \ldots, m, \\ +\infty & \text{otherwise,} \end{cases}$$

then by introducing slack variables $s_i = -y_i - g_i(x) \geq 0$, we see

$$\begin{aligned} -\phi^*(0, \lambda) &:= -\sup_{x,y} \{\langle 0, x\rangle + \langle \lambda, y\rangle - \phi(x, y)\} \\ &= -\sup_{s \geq 0, x} \{-\langle \lambda, s + g(x)\rangle - f(x)\} \\ &= \begin{cases} \inf_x \{f(x) + \langle \lambda, g_i(x)\rangle\} & \lambda \geq 0, \\ -\infty & \text{otherwise.} \end{cases} \end{aligned}$$

Thus the dual problem becomes

$$\sup_{\lambda \geq 0} \inf_x L(x, \lambda), \quad L(x, \lambda) := f(x) + \langle \lambda, g_i(x)\rangle,$$

so we recover Lagrangian duality. This makes two things immediately clear:

(i) Assuming strong duality and attainment, a primal-dual optimal pair (x^*, λ^*) is a saddle point of the Lagrangian. In particular,

$$L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*)$$

for every x and every $\lambda \geq 0$. Thus x^* minimizes $x \mapsto L(x, \lambda^*)$, while λ^* maximizes $\lambda \mapsto L(x^*, \lambda)$ over $\lambda \geq 0$.

(ii) Since $\partial L / \partial g_i = \lambda_i$, the dual variable λ_i^* measures the sensitivity of the optimal value to perturbations of the ith constraint. Under suitable regularity conditions, it gives the marginal cost of tightening that constraint, or equivalently the marginal benefit of relaxing it.

3 Useful References

Here I collect some useful references if you are interested in reading more about these topics.

1. Stephen Boyd and Lieven Vandenberghe, *Convex Optimization*
2. Cédric Villani, *Optimal Transport: Old and New* (Chapter 5)
3. Dual Cones in Lagrangian Duality
4. Perturbation Interpretation of Duality.