# Optimizing selection for function-valued traits

Jay H. Beder [*], Richard Gomulkiewicz [†]

**Abstract**

We consider a function-valued trait $z(t)$ whose pre-selection distribution is Gaussian, and a fitness function $W$ that models optimizing selection, subject to certain natural assumptions. We show that the post-selection distribution of $z(t)$ is also Gaussian, compute the selection differential, and derive an equation that expresses the selection gradient in terms of the parameters of $W$ and of the pre-selection distribution. We make no assumptions on the nature of the "time" parameter $t$.

**Key words.** quantitative genetics, finite-dimensional trait, function-valued trait, selection gradient, selection differential, fitness function, Gaussian process, reproducing kernel Hilbert space, weak limits

**AMS(MOS) subject classification.** 47N60, 60G15, 92D15, 46E22

## 1 Introduction

Understanding the adaptive evolution of environmentally-sensitive traits such as thermal performance curves, age-dependent traits such as growth trajectories, and morphological shapes such as wings and leaves is central to addressing many major questions in biology (e.g., [10, 15, 23, 26, 27, 28, 30]). These are all examples of function-valued traits in that the pattern of expression for each trait can be described by a function of a continuous index [16, 14]. The study of function-valued traits in evolutionary biology is a developing field compared to the well-established subject of finite-dimensional traits, including single and multivariate traits, which are described by finite vectors of measurements.

Many finite-dimensional traits have been found or are presumed to be subject to optimizing selection, wherein selection favors an optimal phenotype and fitness is reduced relative to how far an individual's phenotype deviates from the optimum (e.g., [13, 29]). It is natural to expect function-valued traits to be likewise subject to optimizing selection, with the optimal phenotype being a function (such as an ideal morphological shape or an optimal gene expression profile).

At the population level, it has been shown for both finite- and function-valued traits that a population's mean phenotype will evolve according to the pattern of heritable genetic variances and covariances underlying the trait as well as the selection gradient which describes the linear effects of selection [17, 18, 16, 4]. The selection gradient can be determined from the statistical relationship between phenotype and fitness, where the latter may be measured directly or modeled [17, 18, 1]. When phenotypes are normally distributed, Lande has shown that the

[*]Department of Mathematical Sciences; University of Wisconsin; Milwaukee, WI 53201.

[†]School of Biological Sciences and Department of Mathematics; P.O. Box 644236; Washington State University; Pullman, WA 99164

selection gradient can also be computed in terms of the population's mean fitness [17, 18]. This result has been extended to function-valued traits by the present authors [11], who call it Lande's Theorem.

Optimizing selection is often modeled using Gaussian fitness functions (e.g., [17, 18, 5]). The primary objective of this paper is to derive the selection gradient of a function-valued trait for which fitness is Gaussian. We show that the selection gradient for optimizing selection on a function-valued trait contains an unforeseen component that does not arise for finite-dimensional traits.

### 1.1 Finite-dimensional traits

A finite-dimensional quantitative trait is a random vector $\mathbf{z}$ in $\mathbb{R}^n$ which we will assume to be normally distributed among newborns with mean $\overline{\mathbf{z}}$ and phenotypic covariance $\mathbf{P}$, in conformity with notation in the biological literature. The $N(\overline{\mathbf{z}}, \mathbf{P})$ distribution is the *pre-selection* distribution of $\mathbf{z}$, and we denote its probability density by $p_{\overline{\mathbf{z}}}(\mathbf{z})$. Its *post-selection* distribution – the distribution of the trait among surviving adults – has probability density given by

$$p^*_{\overline{\mathbf{z}}}(\mathbf{z}) = \frac{W(\mathbf{z}) p_{\overline{\mathbf{z}}}(\mathbf{z})}{\mathrm{E}_{\overline{\mathbf{z}}} W}, \tag{1}$$

where $W$ is the *fitness function* of the trait $\mathbf{z}$. The *selection differential* is

$$\mathbf{s} = \overline{\mathbf{z}}^* - \overline{\mathbf{z}},$$

where $\overline{\mathbf{z}}^*$ is the post-selection mean of $\mathbf{z}$. The post-selection distribution of $\mathbf{z}$ need not be normal in general.

The *selection gradient* $\boldsymbol{\beta}$ of the trait at $\overline{\mathbf{z}}$ is the matrix product

$$\boldsymbol{\beta} = \mathbf{P}^{-1}\mathbf{s}. \tag{2}$$

According to Lande and Arnold [20], its $i$th component quantifies the force of directional selection acting on the $i$th component of $\mathbf{z}$. Under appropriate assumptions $\boldsymbol{\beta}$ determines the evolutionary (i.e., between-generation) change in the mean according to the *Breeder's equation* $\overline{\mathbf{z}}' - \overline{\mathbf{z}} = \mathbf{G}\boldsymbol{\beta}$, where $\overline{\mathbf{z}}'$ is the mean of the trait $\mathbf{z}$ among newborns of the following generation and $\mathbf{G}$ is the additive-genetic covariance matrix of the trait [18, 31]. Heckman [12] gives very readable mathematical treatment of this equation.

For a finite-dimensional trait $\mathbf{z}$, Lande [18, p.407] discusses a "Gaussian" fitness function of the form

$$W = \exp\{-(1/2)(\mathbf{z} - \boldsymbol{\theta})^{\mathsf{T}} \mathbf{V}^{-1} (\mathbf{z} - \boldsymbol{\theta})\}. \tag{3}$$

Here $\boldsymbol{\theta}$ is an "ideal" phenotype, so that $W$ measures closeness to an ideal, and $\mathbf{V}$ is a symmetric matrix. If $\mathbf{V}$ is positive definite, then $\boldsymbol{\theta}$ functions as an optimal phenotype, and an individual's fitness is higher the more similar its phenotype is to the optimum, which is the essence of optimizing selection [13, 29]. Similarly, if $\mathbf{V}$ is negative definite then phenotypes more dissimilar to $\boldsymbol{\theta}$ will have higher fitness, which characterizes disruptive selection. In this paper we will consider only the case that $\mathbf{V}$ is positive definite.

It is not difficult to show that for the fitness (3) the post-selection distribution of $\mathbf{z}$ is also normal and that we have

$$\mathbf{s} = \mathbf{P}\mathbf{V}^{-1}(\boldsymbol{\theta} - \overline{\mathbf{z}}^*). \tag{4}$$

and

$$(\mathbf{P}+\mathbf{V})\boldsymbol{\beta} = (\boldsymbol{\theta} - \overline{\mathbf{z}}) \tag{5}$$

(see Section 3). The latter equation shows the explicit dependence of the selection gradient on the pre-selection distribution and on the parameters of the fitness function. Our goal will be to generalize these results to the function-valued case.

## 1.2 Function-valued traits

An infinite-dimensional (or *function-valued*) trait is a random function $z(t)$, that is, a family of random variables $z$ defined on a measure space $(\Omega, \mathcal{A})$ and indexed by $t$ in a set $T$. For example, $z(t)$ might be the response of an organism to environmental condition $t$, or its size at age $t$. We will assume that the pre-selection distribution of $z$ is Gaussian with mean function $\overline{z}(t)$ and *phenotypic covariance function* $P(s,t)$. We will denote by $\mathrm{P}_{\overline{z}}$ the pre-selection probability measure on $(\Omega, \mathcal{A})$ corresponding to the mean $\overline{z}$. The study of function-valued traits was initiated in [16]. Here we will follow the mathematical development given in [11].

We will assume that $\overline{z}$ is an element of the *reproducing kernel Hilbert space* $\mathcal{H}(P) = \mathcal{H}(P,T)$ having kernel $P$. This assumes in particular that the constant function 0 is a possible mean, and implies that any two pre-selection distributions are mutually absolutely continuous due to the Gaussian Dichotomy Theorem (see Section 2.3). This allows the measure $\mathrm{P}_0$ (mean zero) to assume the role ordinarily played by Lebesgue measure, which is no longer available in the infinite-dimensional setting.

The probability measure $\mathrm{P}^*_{\overline{z}}$ on $(\Omega, \mathcal{A})$ will denote the post-selection distribution of $z(t)$, and is connected to $\mathrm{P}_{\overline{z}}$ by the analogue of (1), namely

$$d\mathrm{P}^*_{\overline{z}} = (W/\mathrm{E}_{\overline{z}}W)d\mathrm{P}_{\overline{z}}, \tag{6}$$

where $W$ is the *fitness* of the trait $z$ and $\mathrm{E}_{\overline{z}}$ is expectation with respect to $\mathrm{P}_{\overline{z}}$. The fitness is assumed to be a positive random variable belonging to $L^2(\Omega, \mathcal{A}, \mathrm{P}_{\overline{z}})$ for every $\overline{z} \in \mathcal{H}(P)$. That is,

$$\mathrm{Var}_{\overline{z}}W < \infty \quad \text{for all } \overline{z} \in \mathcal{H}(P). \tag{7}$$

The *selection differential* $s$ is the function

$$s = \overline{z}^* - \overline{z}, \tag{8}$$

where $\overline{z}^*$ is the post-selection mean of $z$. It turns out that $\overline{z}^*$ and $s$ belong to $\mathcal{H}(P)$ [11, Proposition 4.1].

In generalizing the definition (2) of the selection gradient, we replace the matrix $\mathbf{P}$ by an integral operator $\mathcal{P}$. To do this, we will assume that $T$ is a measure space carrying a $\sigma$-finite measure $m$, and we will denote the inner product of $L^2(T)$ by $(\cdot,\cdot)$. We will also assume that the covariance $P$ is a measurable function on $T \times T$ having finite trace (see Section 2.1). $P$ is then the kernel of an integral operator $\mathcal{P}$ defined on $L^2(T)$ which we assume is one-to-one.

If $s$ is in range($\mathcal{P}$), then we define the selection gradient $\beta$ by the formal generalization of (2), that is, $\beta = \mathcal{P}^{-1}s$. To cover the case when $s \notin$ range($\mathcal{P}$), we extend $L^2(T)$ slightly to its *weak P-completion* $\mathcal{L} = \mathcal{L}_P$, whose elements may be viewed as linear functionals on $\mathcal{H}$(P), possibly unbounded with respect to the $L^2(T)$ norm, and we define $\overline{\mathcal{P}}$ to be the *$\mathcal{L}$-extension of* $\mathcal{P}$ (see Section 2.5). Then the *selection gradient $\beta$ is the element of $\mathcal{L}$ satisfying*

$$\overline{\mathcal{P}}\beta = s. \tag{9}$$

The linear functional $\beta$ may be evaluated at any $\eta \in \mathcal{H}(P)$ by

$$(\beta, \eta) = \langle\, s, \eta \,\rangle,$$

$\langle\, \cdot, \cdot \,\rangle$ denoting the inner product in $\mathcal{H}(P)$.

To generalize the fitness function (3), we begin by noting that if the matrix $\mathbf{V}$ in (3) is positive definite, then (apart from the factor $-1/2$) the quadratic exponent is actually the square of a norm given by an inner product on $\mathbb{R}^n$. In fact, this inner product makes $\mathbb{R}^n$ into a finite-dimensional reproducing kernel Hilbert space $\mathcal{H}(V, S)$ where $S = \{1, \ldots, n\}$.

The generalization of the fitness function (3) for infinite-dimensional characters is thus

$$W = \exp\{-(1/2)\|z - \theta\|_V^2\}, \tag{10}$$

where $V(\cdot, \cdot)$ is a positive symmetric kernel and $\|\ \ \|_V$ is the norm in the reproducing kernel Hilbert space $\mathcal{H}(V, T)$. This obviously requires that $\theta \in \mathcal{H}(V, T)$ and that $z$ have its trajectories in $\mathcal{H}(V, T)$ with probability one. The latter requires some conditions relating the kernel $P$ to $V$, as we shall see in Section 2.4.

In [4] $\beta$ was computed for a number of fitness functions of biological interest. In those examples we could call on some relatively straightforward arguments that do not carry over to the present case. Instead, *we will show that the post-selection distribution is Gaussian, find the post-distribution mean $\overline{z}$ and compute $s$, and then show that $\beta$ must satisfy the analog of (5).*

This goal requires a large array of mathematical tools, which we review and develop in Section 2. The reader may prefer to skip this material on first reading, referring back to it as necessary. Section 3 derives the finite-dimensional results referred to above, which we generalize in Section 4.

# 2 Mathematical background

Throughout this section we will use $K$ (rather than $P$) to represent the covariance function of the given process, so as to allow it to stand for either $P$ or $V$, and to make it easier to distinguish from the probability measure P.

## 2.1 Reproducing kernel Hilbert spaces

A kernel $K(s, t)$ on a set $T$ is said to be *positive* resp. *positive definite* if for any $t_1, \ldots, t_n \in T$ the matrix $[K(t_i, t_j)]$ is positive semidefinite resp. positive definite. Since a positive semidefinite matrix is definite iff it is nonsingular, we may refer to positive definite kernels as *nonsingular.*

Note that any covariance function is positive and symmetric.

For any set $T$ and any positive symmetric kernel $K$ on $T$, there exists a *reproducing kernel Hilbert space (RKHS)* $\mathcal{H}(K, T)$ with kernel $K$. We will denote the inner product of $\mathcal{H}(K, T)$ by $\langle\, \cdot, \cdot \,\rangle$. This Hilbert space is a set of functions on $T$ characterized by two properties:

- $K_t \in \mathcal{H}(K,T)$ for all $t \in T$, where $K_t$ is the function defined by

$$K_t(\cdot) = K(t, \cdot); \tag{11}$$

and

- the *reproducing property*

$$\langle K_t, g \rangle = g(t) \text{ for all } t \in T, \; g \in \mathcal{H}(K,T). \tag{12}$$

We will refer to the function $K_t$ as a *section* of $K$. When the underlying set $T$ is understood, we will write $\mathcal{H}(K)$ for $\mathcal{H}(K,T)$. A general reference for such spaces is [2].

Two important cases are the following:

$T$ **finite.** Let $T = \{1, \ldots, n\}$. If the matrix $\mathbf{K} = [K(i,j)]$ is invertible, then this space may be viewed as $\mathbb{R}^n$ with inner product given by

$$\langle \mathbf{c}, \mathbf{d} \rangle = \mathbf{c}^{\mathsf{T}} \mathbf{K}^{-1} \mathbf{d}. \tag{13}$$

$K$ **measurable.** (See [6], [11, Section 3.3].) Suppose $T$ is equipped with a sigma-algebra $\mathcal{T}$ and a sigma-finite measure $\mu$. The inner product of $L^2(T, \mathcal{T}, \mu)$ is given as usual by

$$(f, g) = \int_T f(t) g(t) d\mu(t). \tag{14}$$

A covariance kernel $K$ is said to have *finite trace* if it satisfies

$$\int_T K(t,t) d\mu(t) < \infty. \tag{15}$$

$K$ is then the kernel of an integral operator $\mathcal{K}$ of finite trace defined on $L^2(T) = L^2(T, \mathcal{T}, \mu)$: for all $f \in L^2(T)$,

$$\mathcal{K} f(t) = \int_T K(t,u) f(u) d\mu(u) = (f, K_t), \tag{16}$$

the second equality holding as the functions $K_t$ are square integrable. In fact, (15) implies that all the functions in $\mathcal{H}(K,T)$ are square-integrable, so that $\mathcal{H}(K,T)$ may be viewed as a subset of $L^2(T)$. On the other hand,

$$\operatorname{range}(\mathcal{K}) \subset \mathcal{H}(K,T)$$

[6, Sec. 29]. The relationship of the inner products is given by the following: For all square-integrable $f$ and all $\eta \in \mathcal{H}(K,T)$,

$$\langle \mathcal{K} f, \eta \rangle = (f, \eta). \tag{17}$$

The notation $\mathcal{H}(K,T)$ indicates the dependence of the RKHS on both the kernel $K$ and the index set $T$. Since we will be allowing both to vary, we will index norms and inner products by either the kernel or the set (e.g., $\langle f, g \rangle_V$, $\|f\|_S$). It will be clear from context what the subscript stands for and thus what RKHS is meant.

**Dominance.** Suppose the (positive, symmetric) kernels $K$ and $V$ are defined on the same set $T$ and

$$\mathcal{H}(K) \subset \mathcal{H}(V). \tag{18}$$

Then $\mathcal{H}(K)$ is a sub-vector space of $\mathcal{H}(V)$, though not a Hilbert subspace as the inner products of $\mathcal{H}(K)$ and $\mathcal{H}(V)$ are different. Following [9] we say that the kernel $V$ *dominates* $K$ if (18) holds, and we may write $V \geq K$.

**Theorem 2.1.** *[2, pp. 351-352] Let $V \geq K$. Then $\|g\|_V \leq \|g\|_K$ for all $g \in \mathcal{H}(K)$. Moreover, there exists a unique linear operator $\Psi : \mathcal{H}(V) \to \mathcal{H}(V)$ whose range is contained in $\mathcal{H}(K)$ and such that*

$$\langle f, g \rangle_V = \langle \Psi f, g \rangle_K, \quad \forall f \in \mathcal{H}(V), g \in \mathcal{H}(K).$$

*In particular,*

$$\Psi V_t = K_t \quad \textit{for all}\, t \in T. \tag{19}$$

*As an operator into $\mathcal{H}(V)$, $\Psi$ is bounded, positive and symmetric.*

We will refer to the map $\Psi$ as the *dominance operator of $V$ over $K$*. If $\Psi$ has finite trace, we say [8] that $V$ *n-dominates* (or *nuclear-dominates*) $K$, and we will write $V \gg K$.

The following lemma, which will be useful later, may be viewed as a generalization of the reproducing property.

**Lemma 2.1.** *Let $V \geq K$, with dominance operator $\Psi$. For all $f \in \mathcal{H}(V,T)$ and $t \in T$,*

$$\langle f, K_t \rangle_V = \Psi f(t).$$

*Proof.* $\langle f, K_t \rangle_V = \langle f, \Psi V_t \rangle_V = \langle \Psi f, V_t \rangle_V$, which equals $\Psi f(t)$ by the reproducing property of $V$ in $\mathcal{H}(V,T)$. □

When $T = \{t_1, \ldots, t_n\}$, $\Psi$ is given by the matrix $\mathbf{K}\mathbf{V}^{-1}$. When $K$ and $V$ are measurable kernels, then, $\Psi$ might be thought heuristically of as an operator $\mathcal{K}\mathcal{V}^{-1}$. In fact, (19) shows that $\Psi$ satisfies

$$\Psi \mathcal{V} = \mathcal{K},$$

where $\mathcal{V}$ is the integral operator with kernel $V$.

Define

$$d_V(s,t) = \|V_s - V_t\|_V, \quad s,t \in T.$$

**Lemma 2.2.** *[22, Lemmas 3.2 and 3.3] Assume the kernel $V$ is nonsingular (= positive definite). Then*

*a. the set $\{V_t, t \in T\}$ is linearly independent,*

*b. $d_V$ is a metric on $T$, and every element of $\mathcal{H}(V,T)$ is continuous with respect to $d_V$, and*

*c. $(T, d_V)$ is a separable metric space iff $\mathcal{H}(V,T)$ is a separable Hilbert space.*

The fact that the metric $d_V$ makes the map $t \mapsto V_t$ continuous means that the kernel $V$ is itself continuous on $T \times T$ [24, page 41]. In effect, the introduction of $d_V$ allows us to avoid having to *assume* that $T$ is a topological space and $V$ is continuous.

We will henceforth assume that $V$ is nonsingular and that $\mathcal{H}(V,T)$ is separable for simplicity. This is probably a reasonable assumption in practice. The theory may be extended to singular kernels by use of *Hamel sets*; see [22], especially Section 3.

Let $T' = \{t_i : i = 1, 2, \ldots, \}$ be a fixed countable $d_V$-dense set in $T$. For each initial segment $T_n = \{t_i : i = 1, \ldots, n\}$ of $T'$ and any function $f : T \to \mathbb{R}$ let $f_n$ be the restriction of $f$ to $T_n$. Similarly, for any kernel $K$ on $T$ we let $K_n$ be the restriction of $K$ to $T_n \times T_n$, and let $\mathbf{K}_n = [K(t_i, t_j) : i, j \leq n]$. The reproducing kernel Hilbert space $\mathcal{H}(V, T_n)$ may be viewed as $\mathbb{R}^n$ with inner product given by

$$\langle \mathbf{c}, \mathbf{d} \rangle = \mathbf{c}^{\mathsf{T}} \mathbf{V}_n^{-1} \mathbf{d} = \sum \sum a_{ij} c_i d_j, \tag{20}$$

where $\mathbf{V}_n$ is the $n \times n$ matrix $[V(t_i, t_j) : 1 \leq i, j \leq n]$ and $\mathbf{V}_n^{-1} = [a_{ij}]$. (The entries $a_{ij}$ also depend on $n$.)

**Proposition 2.1.** *If $f \in \mathcal{H}(V,T)$, then $f_n \in \mathcal{H}(V,T_n)$, and*

$$\|f_n\|_{T_n} \leq \|f_{n+1}\|_{T_{n+1}} \to \|f\|. \tag{21}$$

*If also $g \in \mathcal{H}(V,T)$, then*

$$\langle f_n, g_n \rangle_{T_n} \to \langle f, g \rangle. \tag{22}$$

*If $K \ll V$ with dominance map $\Psi$ and if $\tau = \operatorname{tr}(\Psi)$, then*

$$\tau = \lim_n \operatorname{tr}(\mathbf{K}_n \mathbf{V}_n^{-1}). \tag{23}$$

Here tr denotes the trace. Proof of (21) and (23) are given in [22, Lemma 3.6 and Proposition 3.10]. The limit (22) may be proved from (21) by the polarization identity.

## 2.2 Hilbert spaces of random variables

In this section we assume that the process $z$, defined on the probability space $(\Omega, \mathcal{A}, \mathrm{P})$, has mean zero and covariance function $K$. Thus we have $z(t) \in L^2(\mathrm{P}) = L^2(\Omega, \mathcal{A}, \mathrm{P})$ for each $t \in T$. We need to introduce the Hilbert spaces $H$ and $H^{2\odot}$ and the map $\Lambda$.

### 2.2.1 The space $H$ and the Loève map $\Lambda$

We let $H \subset L^2(\mathrm{P})$ be the subspace generated by the set $\{z(t), t \in T\}$. It consists of all the random variables $z(t)$ along with finite linear combinations of them and mean-square limits with respect to P. In $H$, covariance = inner product:

$$\operatorname{Cov}(X, Y) = (X, Y). \tag{24}$$

(We use the same notation for the inner product of $H$ as for that of $L^2(\Omega, \mathcal{A}, \mathrm{P})$.) We call $H$ the *Hilbert space spanned by the process.*

The connection between reproducing kernel Hilbert spaces and second-order (finite-variance) stochastic processes is the following. It does not assume anything about the set $T$, and does not assume that the process is Gaussian:

**Lemma 2.3.** *Let the process $z$ have covariance $K$. The map $\Lambda : H \to \mathcal{H}(K,T)$ taking $X$ to the function $g$ defined by*

$$g(t) = (X, z(t))$$

*is a Hilbert space isomorphism.*

This result, due to Loève, is well known; $\Lambda$ has been called the *Loève map* [3]. In particular,

$$\Lambda(z(t)) = K_t \text{ for all } t \in T, \tag{25}$$

where $K_t$ is a section of $K$ (see (11)).

### 2.2.2 The spaces $H^{2\odot}$ and $H^{n\odot}$

In this section, $H$ represents an arbitrary Hilbert space. Its application to stochastic processes will be made clear in Section 2.3.

The tensor product of Hilbert spaces is a general construction that is treated in a number of texts. One takes the vector-space tensor product, defines an inner product, and then takes the closure with respect to the resulting norm. We refer the reader to the literature for details. We will summarize the results we need from Chapters 6–8 of [24].

The *symmetric tensors* in $H \otimes H$ are those like $f \otimes f$ and $f \otimes g + g \otimes f$ that are unaffected by interchanging the two arguments. Formally, we extend the map $f \otimes g \mapsto g \otimes f$ to a linear operator $H \otimes H \to H \otimes H$ and define the symmetric tensors to be the elements of $H \otimes H$ which are invariant under this operator. The set of symmetric tensors is a subspace $H \odot H$, or $H^{2\odot}$, whose inner product in $H^{2\odot}$ can be computed as follows: [24, p. 121]:

$$(U_1 \odot U_2, V_1 \odot V_2) = (U_1, V_1)(U_2, V_2) + (U_1, V_2)(U_2, V_1), \tag{26}$$

if we define $U_1 \odot U_2 = (U_1 \otimes U_2 + U_2 \otimes U_1)/\sqrt{2}$.

A similar construction holds for other "tensor powers". The Hilbert space $H^{n\odot}$ is the *n-th symmetric tensor power* or *n-th Wiener chaos* of $H$.

We may form the *symmetric Hilbert space* of $H$ as the direct Hilbert sum $\oplus_{n\geq 0} H^{n\odot}$, where $H^{n\odot}$ is the *n-th symmetric tensor power* or *n-th Wiener chaos* of $H$ ($H^{0\odot} = \mathbb{R}$).

Given $X \in H^{2\odot}$ there exists a unique symmetric Hilbert-Schmidt operator $\tilde{X} : H \to H$ such that

$$(\tilde{X} U_1, U_2) = (U_1 \odot U_2, X) \tag{27}$$

for any $U_1, U_2$ in $H$ (cf. [24, Proposition 6.16]). (Technically, $\tilde{X}$ maps $H$ to its dual $H'$; it maps $H$ to $H$ if we identify $H$ with its dual in the usual way.)

## 2.3 Gaussian random variables

A process $z$ is Gaussian if any finite linear combination of the variables $z(t)$ has a normal distribution. This is equivalent to saying that these variables are jointly normal. If the process has zero mean under P, then the elements of $H$ are normally distributed with mean zero under P.

**Theorem 2.2** (N. Wiener)**.** *Assume the process $z$ is Gaussian, and let $\mathcal{A}$ be the $\sigma$-algebra generated by the random variables $z(t)$, $t \in T$, and the sets of probability zero. Then there is an isometric isomorphism $\phi$ mapping $\oplus_{n\geq 0} H^{n\odot}$ onto $L^2(\Omega, \mathcal{A}, P)$.*

We refer the reader to [24, Proposition 7.3] for the definition of the map $\phi$. Evaluation of $\phi$ at a specific element of the component $H^{n\odot}$ is given in [24, Proposition 7.5]. In particular, $\phi(X) = X$ for any $X \in H$, and for an element $U_1 \odot U_2 \in H^{2\odot}$ we have

$$\phi(U_1 \odot U_2) = U_1 U_2 - (U_1, U_2). \tag{28}$$

Thus the random variable corresponding to $z(t_1) \odot z(t_2)$ is $z(t_1)z(t_2) - K(t_1, t_2)$. In particular, $\phi(z(t)^{2\odot}) = z(t)^2 - K(t,t)$.

The so-called *Gaussian Dichotomy Theorem* [7], or GDT, asserts that if a process is Gaussian with respect to probability measures P and Q, then the measures are either mutually singular ($\mathrm{P} \perp \mathrm{Q}$) or equivalent ($\mathrm{P} \sim \mathrm{Q}$). Here *equivalent* means *mutually absolutely continuous*. The theorem is essentially composed of two cases. In the following, the subscript P or Q means "with respect to the measure" P or Q. The Hilbert space $H$ is defined with respect to P.

**Theorem 2.3** (GDT for means)**.** *Suppose the process $z$ is Gaussian with covariance $K$ with respect to both* P *and* Q*, with mean functions* 0 *and $m$, respectively. Then $\mathrm{P} \sim \mathrm{Q}$ or $\mathrm{P} \perp \mathrm{Q}$. We have $\mathrm{P} \sim \mathrm{Q}$ iff $m \in \mathcal{H}(K)$, in which case*

*a. the density of* Q *with respect to* P *is $d\mathrm{Q}/d\mathrm{P} = \exp Y / \mathrm{E}_{\mathrm{P}}(\exp Y)$, and*

*b. $\mathrm{E}_{\mathrm{Q}} X = (X, Y)_{\mathrm{P}}$ for every $X \in H$,*

*where $Y = \Lambda^{-1} m$. Conversely, if Q satisfies (a) for some $Y \in H$ then $z$ is Gaussian with covariance $K$ with respect to Q, and $\mathrm{E}_{\mathrm{Q}}$ is given by (b).*

Theorem 2.3 is adapted from [24, Proposition 8.1 and Corollary 8.3]. The following is taken from [24, Propositions 8.4 and 8.6]; $\phi$ is the map given above.

**Theorem 2.4** (GDT for covariances)**.** *Suppose the process $z$ is Gaussian with mean zero with respect to both* P *and* Q*. Then $\mathrm{P} \sim \mathrm{Q}$ or $\mathrm{P} \perp \mathrm{Q}$. We have $\mathrm{P} \sim \mathrm{Q}$ iff there is a $U \in H^{2\odot}$ such that*

$$\mathrm{Cov}_{\mathrm{Q}}(Z, Y) - \mathrm{Cov}_{\mathrm{P}}(Z, Y) = (Z \odot Y, U)_{\mathrm{P}} \quad \textit{for every } Z, Y \in H, \tag{29}$$

*and such that the eigenvalues of the operator $\tilde{U}$ are all $\geq c$ for some $c > -1$. In this case, the density of* Q *with respect to* P *is*

$$d\mathrm{Q}/d\mathrm{P} = \exp X / \mathrm{E}_{\mathrm{P}}(\exp X), \tag{30}$$

*where $X$ is the element of $\phi(H^{2\odot})$ whose Hilbert-Schmidt operator $\tilde{X}$ satisfies*

$$(I - \tilde{X})(I + \tilde{U}) = (I + \tilde{U})(I - \tilde{X}) = I. \tag{31}$$

*Conversely, if* Q *is a measure on $(\Omega, \mathcal{A})$ satisfying (30) where $X$ is an element of $\phi(H^{2\odot})$ such that the eigenvalues of $\tilde{X}$ are all less than 1, then the process $z$ is Gaussian with respect to Q, and $\mathrm{Cov}_{\mathrm{Q}}(Z, Y)$ satisfies (29) where $U$ is an element of $H^{2\odot}$ satisfying (31).*

If $z$ has covariances $K_{\mathrm{P}}$ and $K_{\mathrm{Q}}$ with respect to P and Q, then (29) means that $K_{\mathrm{Q}}(s,t) - K_{\mathrm{P}}(s,t) = (z(s) \odot z(t), U)$.

Technically the operator $\tilde{X}$ is attached not to $X$ (an element of $L^2(\mathrm{P})$) but to $\phi^{-1}(X)$ (an element of $H^{2\odot}$), and so the notation $\tilde{X}$ is slightly at variance with its use in the defining equation (27). Thus (27) should be understood here as

$$(\tilde{X}U_1, U_2) = (U_1 \odot U_2, \phi^{-1}(X)) \tag{32}$$

for any $U_1, U_2$ in $H$. We note that Neveu [24] drops all references to $\phi$ once he has established Theorem 2.2, evidently viewing $\phi$ as an identification. It seems prudent for us to make references to $\phi$ explicit, as needed.

### 2.4 Some sample-path results.

When a stochastic process $\{z(t), t \in T\}$ has its sample paths almost surely in a Hilbert space $\mathcal{H}$, a sample path may be viewed as a random element $z$ in $\mathcal{H}$. That is, the inner product $\langle z, h \rangle$ is a random variable for every $h \in \mathcal{H}$ (see [22, Lemma 2.1]). When $\mathcal{H} = \mathcal{H}(V,T)$ is a RKHS, we have in particular that $z(t) = \langle z, V_t \rangle$ for all $t \in T$.

A random element in $\mathcal{H}$ is *Gaussian* if the random variable $\langle z, h \rangle$ is normally distributed for every $h \in \mathcal{H}$.

**Theorem 2.5.** *[22, Theorem 7.1] Let $\{z(t), t \in T\}$ be a Gaussian process with mean function $\overline{z}$ and covariance function $K$. Let $\mathcal{H} = \mathcal{H}(V,T)$ be a RKHS with $\overline{z} \in \mathcal{H}$. If the sample paths of $z$ belong almost surely to $\mathcal{H}$, then the random element defined by the process $z$ is Gaussian. In particular, $V \gg K$.*

Conversely, if $\overline{z} \in \mathcal{H}$ and $V \gg K$, then a version of $z$ has its sample paths almost surely in $\mathcal{H}$, even without the Gaussian assumption. See [22, Theorem 5.1].

**Theorem 2.6.** *[21, Theorem 7.2.1] If $z$ is a zero-mean Gaussian process with sample paths in a separable RKHS $\mathcal{H}$, then $\mathrm{E}(\|z\|^k) < \infty$ for $k = 1, 2, 3 \ldots$.*

We now assume that $V$ is a positive definite kernel and that $\mathcal{H}(V,T)$ is separable. Let $\|\cdot\|$ and $\langle \cdot, \cdot \rangle$ denote its norm and inner product. Following Lemma 2.2 we let $d_V$ be the metric on $T$ defined by $V$, and let $T' = \{t_1, t_2, \ldots,\}$ be a countable $d_V$-dense set $T'$ in $T$. As in Section 2.1, for each initial segment $T_n = \{t_1, \ldots, t_n\}$ of $T'$ and any function $f : T \to \mathbb{R}$ we let $f_n$ be the restriction of $f$ to $T_n$.

**Lemma 2.4.** *Let $z(t)$ be Gaussian with mean zero with respect to the probability measure* P. *Assume the sample paths of $z$ are in $\mathcal{H}(V,T)$ with probability 1. Then the random variables $\|z\|^2, \langle z, \theta \rangle$, and (for each $n$) $\|z_n\|^2_{T_n}$ and $\langle z_n, \theta_n \rangle_{T_n}$ are in $L^2(\mathrm{P})$. Moreover, the following limits hold pointwise and in $L^2(\mathrm{P})$:*

$$\|z_n\|^2_{T_n} \quad \to \quad \|z\|^2 \tag{33}$$

*and*

$$\langle z_n, \theta_n \rangle_{T_n} \quad \to \quad \langle z, \theta \rangle. \tag{34}$$

*Proof.* From Theorem 2.6 we see that $\|z\|$ has moments of all orders. In particular, $\|z\|^2 \in L^2(\mathrm{P})$, while $|\langle z, \theta \rangle| \le \|z\|\|\theta\|$ implies that $\langle z, \theta \rangle \in L^2(\mathrm{P})$. From

$$\|z_n\|_{T_n}^2 \le \|z\|^2 \tag{35}$$

(Proposition 2.1) and

$$|\langle z_n, \theta_n \rangle_{T_n}| \le \|z_n\|_{T_n}\|\theta_n\|_{T_n} \le \|z\|\|\theta\| \tag{36}$$

we see that both $\|z_n\|_{T_n}^2$ and $\langle z_n, \theta_n \rangle_{T_n}$ are in $L^2(\mathrm{P})$ as well.

By Proposition 2.1 the limits (33) and (34) hold pointwise. But convergence in $L^2(\mathrm{P})$ follows from inequalities (35) and (36) and a slight variant of the Dominated Convergence Theorem. □

## 2.5 Weak limits

In this section we summarize and extend some results from [11].

Let $X$ be a vector space and let $\mathcal{F}$ be a set of linear functionals on $X$. The $\sigma(X, \mathcal{F})$ topology on $X$ is the weakest topology with respect to which all the functionals in $\mathcal{F}$ are continuous. If $X$ is a Hilbert space and $\mathcal{F}$ its dual, then $\sigma(X, \mathcal{F})$ is called *the* weak topology on $X$.

In the weak topology on any reproducing kernel Hilbert space, convergence turns out to be pointwise convergence [2, p. 344]. We let $\overline{\mathcal{H}}(K,T)$ denote the completion of $\mathcal{H}(K,T)$ in its weak topology.

Assume now that $K$ is a measurable kernel on $T$ with finite trace. We define $\mathcal{L} = \mathcal{L}_K$ to be the completion of $L^2(T, \mathcal{T}, \mu)$ in the $\sigma(L^2(T), \mathcal{F})$ topology where $\mathcal{F}$ is the set of linear functionals $\mathcal{F} = \{(\cdot, \eta) : \eta \in \mathcal{H}(K,T)\}$. Let us call $\mathcal{L}_K$ the *weak $K$-completion of $L^2(T)$*. A sequence $\{\beta_n\}$ in $L^2(T)$ is convergent in this topology if

$$(\beta_n, \eta) \text{ is convergent for every } \eta \in \mathcal{H}(K,T).$$

A limit $\beta$ of such a sequence may be viewed as a linear functional (possibly unbounded) on $\mathcal{H}(K,T)$. We may write such a functional as $(\beta, \cdot)$.

The integral operator $\mathcal{K}$ has a unique extension to a linear operator $\overline{\mathcal{K}}$ on $\mathcal{L}$ as follows: if $\beta_n \in L^2(T)$ and $\beta_n \to \beta \in \mathcal{L}$, then $s_n = \mathcal{K}\beta_n$ is weakly Cauchy in $\mathcal{H}(K,T)$ with limit $s \in \overline{\mathcal{H}}(K,T)$, and we define $\overline{\mathcal{K}}\beta = s$. The range of $\overline{\mathcal{K}}$ is larger than $\mathcal{H}(K,T)$, as the following proposition shows.

**Proposition 2.2** ([11], Proposition 2)**.** *$\overline{\mathcal{K}}$ maps $\mathcal{L}$ into $\overline{\mathcal{H}}(K,T)$, and satisfies*

$$\langle \overline{\mathcal{K}}\beta, \eta \rangle = (\beta, \eta) \tag{37}$$

*for all $\beta \in \mathcal{L}$ and $\eta \in \mathcal{H}(K,T)$. If $\mathcal{K}$ is one-to-one, so is $\overline{\mathcal{K}}$. We have* $\mathrm{range}(\overline{\mathcal{K}}) \supset \mathcal{H}(K,T)$.

$\mathcal{K}$ satisfies (16); $\overline{\mathcal{K}}$ satisfies an extension of it, namely, for all $\beta \in \mathcal{L}$,

$$\overline{\mathcal{K}}\beta(t) = (\beta, \mathcal{K}_t). \tag{38}$$

Now let $V$ and $P$ be measurable kernels of finite trace, defining integral operators $\mathcal{V}$ and $\mathcal{P}$. Let $\mathcal{L}_V$ and $\mathcal{L}_P$ be the weak completions of $L^2(T)$ that they define. It is easy to see that if

$\mathcal{H}(V,T) \supset \mathcal{H}(P,T)$ then $\mathcal{L}_V \subset \mathcal{L}_P$. $\mathcal{P}$ has an extension $\overline{\mathcal{P}}$ to $\mathcal{L}_P$. We investigate the extension of $\mathcal{V}$ to $\mathcal{L}_P$.

Let $\beta \in \mathcal{L}_P$ and let $\{\beta_n\}$ be a sequence in $L^2(T)$ converging in $\mathcal{L}_P$ to $\beta$. Certainly $r_n = \mathcal{V}\beta_n$ is defined and is an element of $\mathcal{H}(V,T)$ as range$\mathcal{V} \subset \mathcal{H}(V,T)$. Moreover, for all $\eta \in \mathcal{H}(V,T)$ we have

$$\langle r_n, \eta \rangle_V = \langle \mathcal{V}\beta_n, \eta \rangle_V = (\beta_n, \eta),$$

by (17). Thus this holds for all $\eta \in \mathcal{H}(P,T)$; but for such $\eta$ the sequence $(\beta_n, \eta)$ is Cauchy (converging to $(\beta, \eta)$). Thus

$$\langle r_n, \eta \rangle_V \text{ is Cauchy for all } \eta \in \mathcal{H}(P,T). \tag{39}$$

In effect we are defining a new weak topology on $\mathcal{H}(V,T)$ in which a sequence $\{r_n\}$ is Cauchy if it satisfies (39). (It is the $\sigma(\mathcal{H}(V,T), \mathcal{F})$ topology of $\mathcal{H}(V,T)$ where $\mathcal{F}$ is the set of linear functionals $\mathcal{F} = \{\langle \cdot, \eta \rangle_V : \eta \in \mathcal{H}(P,T)\}$.) We denote the completion of $\mathcal{H}(V,T)$ in this topology by $\overline{\overline{\mathcal{H}}}(V,T)$. If $r = \lim_n r_n$ in this topology, then we define the desired extension $\overline{\overline{\mathcal{V}}}$ of $\mathcal{V}$ to $\mathcal{B}$ by

$$\overline{\overline{\mathcal{V}}}\beta = r$$

*Note:* We use the double bar to distinguish $\overline{\overline{\mathcal{H}}}(V,T)$ from "the" weak completion of $\mathcal{H}(V,T)$, and $\overline{\overline{\mathcal{V}}}$ from the extension of $\mathcal{V}$ to $\mathcal{L}_V$.

Now as we saw following Theorem 2.1,

$$\Psi\mathcal{V} = \mathcal{P} \tag{40}$$

is interpreted to mean $\Psi V_s = P_s$ for each $s \in T$. We have seen how to extend $\mathcal{V}$ and $\mathcal{P}$ to $\mathcal{L}_P$, and we wish to extend $\Psi$ so that equation (40) is still valid, that is, so that

$$\overline{\Psi}\overline{\overline{\mathcal{V}}}\beta = \overline{\mathcal{P}}\beta \tag{41}$$

for all $\beta \in \mathcal{L}_P$. To this end, let $\beta_n \in L^2(T)$ converging in $\mathcal{L}_P$ to $\beta$. then $\Psi\mathcal{V}\beta_n = \mathcal{P}\beta_n$. Say $\mathcal{V}\beta_n = r_n \in \mathcal{H}(V)$ and $\mathcal{P}\beta_n = s_n \in \mathcal{H}(P)$ (in fact, $\in$ range($\mathcal{P}$)). Then $\Psi r_n = s_n$. But $s_n \to s \in \overline{\mathcal{H}}(P)$ and $r_n \to r \in \overline{\overline{\mathcal{H}}}(V)$, so we define $\overline{\Psi}r = s$. Thus $\overline{\Psi}$ maps $\overline{\overline{\mathcal{H}}}(V)$ to $\overline{\mathcal{H}}(P)$.

# 3 Optimizing selection for finite-dimensional traits

Assume a finite-dimensional trait $\mathbf{z}$ has a $N(\overline{\mathbf{z}}, \mathbf{P})$ pre-selection distribution, so that its density is $p_{\overline{\mathbf{z}}}(\mathbf{z}) = K \exp\{-(1/2)(\mathbf{z}-\overline{\mathbf{z}})^\mathsf{T}\mathbf{P}^{-1}(\mathbf{z}-\overline{\mathbf{z}})\}$, where $K$ is the normalizing constant. Assuming that the fitness function $W$ has the form (3), let us find the post-selection distribution of $\mathbf{z}$. The post-selection density function is

$$\begin{aligned}
p^*_{\overline{\mathbf{z}}}(\mathbf{z}) &\propto W(\mathbf{z})p_{\overline{\mathbf{z}}}(\mathbf{z}) \\
&\propto \exp{(-1/2)}[(\mathbf{z}-\boldsymbol{\theta})^\mathsf{T}\mathbf{V}^{-1}(\mathbf{z}-\boldsymbol{\theta}) + (\mathbf{z}-\overline{\mathbf{z}})^\mathsf{T}\mathbf{P}^{-1}(\mathbf{z}-\overline{\mathbf{z}})] \\
&\propto \exp{(-1/2)}[\mathbf{z}^\mathsf{T}\mathbf{Q}^{-1}\mathbf{z} - 2\mathbf{z}^\mathsf{T}\mathbf{c}],
\end{aligned}$$

where $\propto$ denotes "proportional to" and where

$$\mathbf{Q}^{-1} = \mathbf{V}^{-1} + \mathbf{P}^{-1} \tag{42}$$

and

$$\mathbf{c} = \mathbf{V}^{-1}\boldsymbol{\theta} + \mathbf{P}^{-1}\overline{\mathbf{z}}.$$

In writing (42), we are assuming that $\mathbf{V}^{-1} + \mathbf{P}^{-1}$ is invertible (it is symmetric and at least positive semidefinite), with inverse $\mathbf{Q}$. Completing the square in the exponent, then, we have

$$p^*_{\overline{\mathbf{z}}}(\mathbf{z}) \propto \exp{(-1/2)}[(\mathbf{z} - \mathbf{Qc})^{\mathsf{T}}\mathbf{Q}^{-1}(\mathbf{z} - \mathbf{Qc})],$$

so that the post-selection distribution of $\mathbf{z}$ is $N(\mathbf{Qc}, \mathbf{Q})$. In particular, the selection differential is

$$\mathbf{s} = \mathbf{Qc} - \overline{\mathbf{z}}. \tag{43}$$

We can get a simpler form for $\mathbf{s}$ by multiplying through by $\mathbf{Q}^{-1}$ and simplifying, so that

$$\mathbf{V}^{-1}\mathbf{s} + \mathbf{P}^{-1}\mathbf{s} = \mathbf{V}^{-1}\boldsymbol{\theta} - \mathbf{V}^{-1}\overline{\mathbf{z}}.$$

Multiplying through by $\mathbf{V}$ and using $\mathbf{s} = \overline{\mathbf{z}}^* - \overline{\mathbf{z}}$, we have $\mathbf{VP}^{-1}\mathbf{s} = \boldsymbol{\theta} - \overline{\mathbf{z}}^*$, or

$$\mathbf{s} = \mathbf{PV}^{-1}(\boldsymbol{\theta} - \overline{\mathbf{z}}^*)$$

(equation (4) above).

Multiplying both sides by $\mathbf{VP}^{-1}$ and using the definition $\boldsymbol{\beta} = \mathbf{P}^{-1}\mathbf{s}$, we may rewrite the last equation as $\mathbf{V}\boldsymbol{\beta} = \boldsymbol{\theta} - \overline{\mathbf{z}}^*$, and adding $\mathbf{s} = \mathbf{P}\boldsymbol{\beta}$ to both sides shows that the selection gradient $\boldsymbol{\beta}$ satisfies equation (5), as claimed.

The expression (5) can be derived directly from (43) without first getting a simplified equation for $\mathbf{s}$. It was first given by Lande [19, equation (7)], who derived it by calculating the (ordinary) gradient of $\log \mathrm{E}_{\overline{\mathbf{z}}}W$ with respect to $\overline{\mathbf{z}}$, using a result he had proved in [18].

The expressions (42) and (4) for the post-selection covariance and the selection differential were first given by Karl Pearson [25]; (42) is his (xcii), and (4) is essentially his (xciii), where he has apparently assumed that the pre-selection mean is zero. Pearson was actually pursuing a variation of our problem, namely, finding the fitness function (the "probability of survival") assuming that the pre- and post-selection distributions are multivariate normal.

# 4 Optimizing selection for function-valued traits

We recall the notation and assumptions of Section 1.2. The trait $z(t), t \in T$, is assumed to be a Gaussian process with pre-selection mean function $\overline{z}$ and covariance function $P$, where $\overline{z} \in \mathcal{H}(P, T)$. We denote the pre-selection measure by $\mathrm{P}_{\overline{z}}$.

The post-selection measure $\mathrm{P}^*_{\overline{z}}$ is given by (6), where the fitness $W$ has the form (10). As noted earlier, this definition requires that

A1. the phenotype $\theta$ belong to $\mathcal{H}(V, T)$; and

A2. the sample paths $z(\cdot)$ of the trait, considered as functions of $t$, belong to $\mathcal{H}(V, T)$ with probability 1.

According to Theorem 2.5, under condition (A2) and the additional assumption

A3. $\overline{z} \in \mathcal{H}(V, T)$

we must have $\mathcal{H}(P) \subset \mathcal{H}(V)$, and the dominance operator must have finite trace $\tau$. We will also assume that the kernel $V$ is nonsingular and that $\mathcal{H}(V,T)$ is separable.

We let $d_V$ be the metric on $T$ defined by $V$ (Section 2.1), and let $T' = \{t_1, t_2, \dots, \}$ be a countable $d_V$-dense set $T'$ in $T$. For each initial segment $T_n = \{t_1, t_2, \dots, t_n\}$ of $T'$ and any function $f : T \to \mathbb{R}$ we let $f_n$ be the restriction of $f$ to $T_n$. The norm and inner product of $\mathcal{H}(V, T_n)$ will be indexed by $T_n$.

Consider the exponent of $W$ in (10). Expanding the squared norm and adding and subtracting $\tau$, we easily see that the exponent has the form

$$W = e^{X+Y+c} \tag{44}$$

where

$$\begin{aligned} X &= (-1/2)(\|z\|^2 - \tau), \\ Y &= \langle\, z, \theta \,\rangle, \end{aligned} \tag{45}$$

and $c$ is a constant (the norm and inner product are in $\mathcal{H}(V,T)$). Let us define

$$\begin{aligned} X_n &= (-1/2)(\|z_n\|^2_{T_n} - \mathrm{tr}(\mathbf{P}_n \mathbf{V}_n^{-1})) \\ Y_n &= \langle\, z_n, \theta_n \,\rangle_{T_n}, \end{aligned} \tag{46}$$

where $\mathbf{P}_n$ and $\mathbf{V}_n$ are as defined in Section 2.1.

**For the remainder of this section let** $\mathrm{P} = \mathrm{P}_0$, the measure which gives the trait $z$ a Gaussian distribution with mean 0 and covariance $P$ (the pre-selection phenotypic covariance). Let $L^2(\mathrm{P}) = L^2(\Omega, \mathcal{A}, \mathrm{P})$. The following is essentially a restatement of Lemma 2.4:

**Lemma 4.1.** *The random variables $X$, $Y$, $X_n$ and $Y_n$ are in $L^2(\mathrm{P})$. Moreover, $X_n \to X$ and $Y_n \to Y$ both pointwise and in $L^2(\mathrm{P})$-norm.*

In fact, we can make a stronger statement about $X_n$ and $Y_n$:

**Lemma 4.2.** *Let $\mathbf{V}_n^{-1} = [a_{ij}]$, and let $\phi$ be the map given by Theorem 2.2. Then*

$$Y_n = \sum\sum a_{ij} z(t_i)\theta(t_j) \tag{47}$$

*and*

$$X_n = \phi(\,(-1/2)\sum_i\sum_j a_{ij} z(t_i) \odot z(t_j)\,). \tag{48}$$

*In particular, $Y_n \in H$ and $X_n \in \phi(H^{2\odot})$.*

*Proof.* The form of $Y_n = \langle\, z_n, \theta_n \,\rangle_{T_n}$ is simply the inner product (20) in $\mathcal{H}(V, T_n)$, while

$$\begin{aligned} -2X_n = \|z_n\|^2_{T_n} - \mathrm{tr}(\mathbf{P}_n\mathbf{V}_n^{-1}) &= \sum_i\sum_j a_{ij} z(t_i) z(t_j) - \sum_i\sum_j P(t_i,t_j) a_{ij} \\ &= \sum_i\sum_j a_{ij}\{z(t_i)z(t_j) - P(t_i,t_j)\} \\ &= \phi(\sum_i\sum_j a_{ij} z(t_i) \odot z(t_j)\,). \end{aligned}$$

□

**Proposition 4.1.** *Assume that the kernel $V$ is nonsingular, that $\mathcal{H}(V)$ is separable, and that assumptions (A1)–(A3) above are satisfied. Let $H$ be the Hilbert space spanned by the process $\{z(t), t \in T\}$ under the probability measure $\mathrm{P} = \mathrm{P}_0$, and let $\phi$ be the map given in Theorem 2.2. Then $X \in \phi(H^{2\odot})$ and $Y \in H$.*

*Proof.* By Lemma 4.2, the random variables $X_n$ and $Y_n$ are elements of $\phi(H^{2\odot})$ and $H$, respectively. But Lemma 4.1 asserts that $X$ and $Y$ are in $L^2(\mathrm{P})$, and that the limits $X_n \to X$ and $Y_n \to Y$ hold in $L^2(\mathrm{P})$. Since $\phi(H^{2\odot})$ and $H$ are closed in $L^2(\mathrm{P})$, this implies that $X \in \phi(H^{2\odot})$ and $Y \in H$. $\square$

We now compute the post-selection distribution of the trait $z$ for a fitness of form (44). Let $\tilde{X}$ be the symmetric Hilbert-Schmidt operator associated to $X$ – or more precisely, to $\phi^{-1}(X)$ – defined by (27).

**Proposition 4.2.** *Assume the fitness function $W$ has the form given in (44), and let $\mathrm{Q}^* = \mathrm{P}^*_{\overline{z}}$ be the post-selection probability measure (6). Assume further that*

$$\text{the eigenvalues of } \tilde{X} \text{ are less than 1.} \tag{49}$$

*Let $Y'' = Y + Y'$ where $Y' = \Lambda^{-1}(\overline{z})$ and $\Lambda$ is the Loeve map from $H$ to $\mathcal{H}(\mathrm{P})$. Then the distribution of $\{z(t), t \in T\}$ under $\mathrm{Q}^*$ is Gaussian with mean function*

$$\overline{z}^* = \Lambda(S) \tag{50}$$

*where $S \in H$ is the solution of the operator equation*

$$(I - \tilde{X})S = Y''. \tag{51}$$

*Proof.* From (44) and Theorem 2.3 we see that with respect to the measure P,

$$\frac{d\mathrm{Q}^*}{d\mathrm{P}} = \frac{d\mathrm{Q}^*}{d\mathrm{P}_{\overline{z}}} \frac{d\mathrm{P}_{\overline{z}}}{d\mathrm{P}} \propto e^{X+Y} e^{Y'} = e^{X+Y''}.$$

For the moment, introduce the measure $\mathrm{P}^*$ such that $d\mathrm{P}^*/d\mathrm{P} = e^X/\mathrm{E}_\mathrm{P}(e^X)$, so that

$$\frac{d\mathrm{Q}^*}{d\mathrm{P}} = \frac{d\mathrm{Q}^*}{d\mathrm{P}^*} \frac{d\mathrm{P}^*}{d\mathrm{P}}.$$

By the "converse" part of Theorem 2.4 the process $z$ is zero-mean Gaussian with respect to $\mathrm{P}^*$, and the covariances under P and $\mathrm{P}^*$ are related by (29); that is, there is a $U \in H^{2\odot}$ such that

$$(Z, Y'')_{\mathrm{P}^*} = (Z, Y'') + (Z \odot Y'', U) \quad \text{for all } Z \in H$$

(inner products on the right-hand-side computed with respect to P).

On the other hand, $d\mathrm{Q}^*/d\mathrm{P}^* = e^{Y''}/\mathrm{E}_{\mathrm{P}^*}(e^{Y''})$, and so by Theorem 2.3 the process is Gaussian under $\mathrm{Q}^*$ with the same covariance as under $\mathrm{P}^*$, and with

$$\mathrm{E}_{\mathrm{Q}^*}(Z) = (Z, Y'')_{\mathrm{P}^*} \quad \text{for all } Z \in H.$$

Thus the mean of any $Z \in H$ with respect to $\mathrm{Q}^*$ is given by

$$\mathrm{E}_{Q^*}(Z) = (Z, Y'') + (Z \odot Y'', U) = (Y'', Z) + (Y'' \odot Z, U).$$

By (27) this

$$= (Y'', Z) + (\tilde{U}Y'', Z) = ([I + \tilde{U}]Y'', Z) = (S, Z),$$

where

$$S = (I + \tilde{U})Y''.$$

But (31) asserts that the inverse of $I + \tilde{U}$ is $I - \tilde{X}$, from which we see that $S$ satisfies (51). Finally, letting $Z = z(t)$, we have $\overline{z}^*(t) = \mathrm{E}_{Q^*}[z(t)] = (S, z(t)) = \Lambda(S)(t)$ by Lemma 2.3, so that (50) holds. □

We now find expressions for the post-selection mean $\overline{z}^*$ and the selection differential $s$ in terms of the population and selection parameters. We will assume that the random variable $X$ given by (45) satisfies the eigenvalue condition (49).

Proposition 4.2 tells us that the post-selection distribution of $z(t)$ is Gaussian with mean function $\overline{z}^* = \Lambda(S)$ where $S$ is given by (51). We seek an equivalent form of (51) in terms of functions of $t$.

We begin by rewriting equation (51) in the form

$$S - \tilde{X}(S) = Y + Y'.$$

Applying $\Lambda$ to both sides, we have

$$\overline{z}^* - \Lambda(\tilde{X}(S)) = \Lambda(Y) + \overline{z}. \tag{52}$$

Thus we need to evaluate $\Lambda(\tilde{X}(S))$ and $\Lambda(Y)$.

**Proposition 4.3.** *We have $\Lambda(\tilde{X}(S)) = -\Psi\overline{z}^*$ and $\Lambda(Y) = \Psi\theta$, where $\Psi$ is the dominance operator of $V$ over $P$.*

*Proof.* The value of $\Lambda(\tilde{X}(S))$ at $t$ is defined to be $(\tilde{X}(S), z(t))$ (Lemma 2.3). But since $X_n \to X$ in $L^2(\mathrm{P})$ (Lemma 4.1), the property (32) of $\tilde{X}$ implies that

$$(\tilde{X}(S), z(t)) = (S \odot z(t), \phi^{-1}(X)) = \lim_n (S \odot z(t), \phi^{-1}(X_n)),$$

where $\phi$ is the map given in Theorem 2.2. From the expansion for $X_n$ given by (48), the inner products in this sequence may be evaluated using (26), (50), (25) and (20):

$$\begin{aligned}
(S \odot z(t), \phi^{-1}(X_n)) &= (-1/2) \sum\sum a_{ij}(S \odot z(t), z(t_i) \odot z(t_j)) \\
&= (-1/2) \sum\sum a_{ij}[(S, z(t_i))(z(t), z(t_j)) + (S, z(t_j))(z(t_i), z(t))] \\
&= (-1/2) \sum\sum a_{ij}[\overline{z}^*(t_i)P(t, t_j) + \overline{z}^*(t_j)P(t_i, t)] \\
&= -\sum\sum a_{ij}\overline{z}^*(t_i)P(t, t_j) \\
&= -\langle \overline{z}^*_n, P_t \rangle_{T_n}.
\end{aligned}$$

Thus, taking the limit as $n \to \infty$, we have

$$(\tilde{X}(S), z(t)) \;=\; -\langle\, \overline{z}^{*}, P_t \,\rangle_V \;=\; -\Psi\overline{z}^{*}(t),$$

by Proposition 2.1 and Lemma 2.1. That is,

$$\Lambda(\tilde{X}(S)) = -\Psi\overline{z}^{*}.$$

The value of $\Lambda(Y)$ at $t$ is given similarly by

$$(Y, z(t)) \;=\; (\langle\, z, \theta \,\rangle_V, z(t)) \;=\; \lim_n(\langle\, z_n, \theta_n \,\rangle_{T_n}, z(t)).$$

From the expansion (47) for $Y_n$ and application of (20) and (25), we have

$$\begin{aligned}
(\langle\, z_n, \theta_n \,\rangle_{T_n}, z(t)) &= (\sum\sum a_{ij} z(t_i)\theta(t_j), z(t)) \\
&= \sum\sum a_{ij}\theta(t_j)(z(t_i), z(t)) \\
&= \sum\sum a_{ij}\theta(t_j)P_t(t_i) \\
&= \langle\, \theta_n, P_t \,\rangle_{T_n}.
\end{aligned}$$

Thus, taking the limit as $n \to \infty$, we have

$$(Y, z(t)) \;=\; \langle\, \theta, P_t \,\rangle_V \;=\; \Psi\theta(t)$$

again by Proposition 2.1 and Lemma 2.1. That is,

$$\Lambda(Y) = \Psi\theta.$$

□

**Theorem 4.1.** *Let the pre-selection distribution of the trait $z(t), t \in T$, be Gaussian with mean function $\overline{z}$ and covariance function $P$, where $\overline{z} \in \mathcal{H}(P, T)$. Let the fitness function be given by (10), where we assume that the kernel $V$ is nonsingular, that $\mathcal{H}(V, T)$ is separable, and that conditions (A1)–(A3) are satisfied. Let $\Psi$ be the dominance operator of $V$ over $P$, with trace $\tau$. Finally, let $X$ be given by (45), and assume that the operator $\tilde{X}$ satisfies the eigenvalue condition (49).*

*Then the post-selection distribution of $z$ is also Gaussian and the selection differential is*

$$s = \Psi(\theta - \overline{z}^{*}). \tag{53}$$

*Moreover, the selection gradient $\beta$ satisfies*

$$(\overline{\mathcal{P}} + \overline{\overline{\mathcal{V}}})\beta = \theta - \overline{z} + \eta, \tag{54}$$

*where $\overline{\mathcal{P}}$ and $\overline{\overline{\mathcal{V}}}$ are the extensions of the integral operators $\mathcal{P}$ and $\mathcal{V}$ to $\mathcal{L}_P$, the weak $P$-completion of $L^2(T)$, and where $\eta \in$ nullspace$(\Psi)$.*

*Proof.* From (52) and Proposition 4.3, we see that equation (51) can be rewritten as

$$(I + \Psi)\overline{z}^* = \overline{z} + \Psi\theta. \tag{55}$$

Rearranging (55) shows that $s$ satisfies (53).

The steps from this to (54) are more delicate than in the finite-dimensional case, as the map $\Psi$ is not invertible in general. From (53) and the defining equation of the selection gradient (9), we have $\overline{\mathcal{P}}\beta = \Psi(\theta - \overline{z}^*)$. By (41), we can write this as

$$\overline{\Psi}\overline{\overline{\mathcal{V}}}\beta = \overline{\Psi}(\theta - \overline{z}^*),$$

which implies that

$$\overline{\overline{\mathcal{V}}}\beta = \theta - \overline{z}^* + \eta$$

where $\eta$ is in the nullspace of $\overline{\Psi}$. Adding $\overline{z}^* - \overline{z} = s = \overline{\mathcal{P}}\beta$ to both sides gives us (54), as desired. □

# 5 Conclusion

Phenotypes subject to optimizing selection experience directional selection whenever a population's mean phenotype deviates from the optimum. For a function-valued trait subject to Gaussian optimizing selection we have derived the selection gradient (54), which quantifies this directional selection. In the course of doing so, we have also derived the corresponding selection differential (53), which describes the within-generation change in mean phenotype. Equations (53) and (54) are the function-valued generalizations of the corresponding equations (4) and (5) for a finite-dimensional trait. In particular, (54) expresses the selection gradient completely in terms of the given population and selection parameters, as desired.

If we delete the eigenvalue assumption (49), the post-selection distribution may no longer be Gaussian. In this case, it may be possible to derive (53) and (54) by application of the generalization of Lande's Theorem given in [11], as may be done in the finite-dimensional case.

As noted after Lemma 2.1, the operator $\Psi$ in (53) is the function-valued analog of the matrix $\mathbf{PV}^{-1}$ in (4), so that the selection differentials for finite-dimensional and function-valued traits have essentially the same form. This suggests that the latter could have been postulated from the former (cf. [4]). However, the function-valued selection gradient (54) contains an ingredient that does not appear in the finite-dimensional case, since the element $\eta$ belonging to the nullspace of $\Psi$ has no counterpart for finite-dimensional traits. The biological interpretation of $\eta$ is not yet clear and hence, neither is its biological significance.

Nevertheless, (54) shows that the component of linear selection on a function-valued trait under optimizing selection is determined by more than just the simple difference between the optimal and mean phenotypes. In particular, it may be possible for two populations with mean functions that lie at different distances from the optimum to experience exactly the same directional selection. Important challenges for future work will be to characterize the component $\eta$ more fully and, indeed, to develop statistical methods to detect it in empirical studies of function-valued traits.

**Acknowledgment**. Research partially supported by NSF grant EF0328594. Thanks to Russell Lande for identifying the sources for the results of Section 3. Thanks also go to Milan Lukić for questions and comments that led to a substantially improved exposition.

## References

[1] S. J. Arnold and M. J. Wade. On the measurement of natural and sexual selection: Theory. *Evolution*, 38:709–720, 1984.

[2] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:337–404, 1950.

[3] J. H. Beder. A sieve estimator for the mean of a Gaussian process. *Annals of Statistics*, 15:59–78, 1987.

[4] J. H. Beder and R. Gomulkiewicz. Computing the selection gradient and evolutionary response of an infinite-dimensional trait. *Journal of Mathematical Biology*, 36:299–319, 1998.

[5] R. Bürger. *The Mathematical Theory of Selection, Recombination, and Mutation*. John Wiley & Sons, Ltd., West Sussex, England, 2000.

[6] P. Cartier. Une étude des covariances mesurables. In L. Nachbin, editor, *Mathematical Analysis and Applications, Part A*. Academic Press, New York, 1981.

[7] J. Feldman. Equivalence and perpendicularity of Gaussian processes. *Pacific Journal of Mathematics*, 9:699–708, 1958. Correction, **10** (1959), 1295–1296.

[8] R. Fortet. Espaces à noyau reproduisant et lois de probabilités des fonctions aléatoires. *Annales de l'Institut Henri Poincaré. Section B, Calcul des Probabilités et Statistique*, IX(1):41–58, 1973.

[9] R. Fortet. Espaces à noyau reproduisant et lois de probabilités des fonctions aléatoires. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences. Série A, Sciences Mathematiques*, 278:1439–1440, 1974.

[10] G. W. Gilchrist, R. B. Huey, J. Balanyà, M. Pascual, and L. Serra. A time series of evolution in action: Latitudinal cline in wing size in South American Drosophila subobscura. *Evolution*, 58:768–780, 2004.

[11] R. Gomulkiewicz and J. H. Beder. The selection gradient of an infinite-dimensional trait. *SIAM Journal of Applied Mathematics*, 56:509–523, 1996.

[12] N. E. Heckman. Functional data analysis in evolutionary biology. In M. G. Akritas and D. N. Politis, editors, *Recent Advances and Trends in Nonparametric Statistics*. Elsevier, 2003.

[13] J.G.Kingsolver, H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. H. Hill, and A. Hoang. The strength of phenotypic selection in natural populations. *American Naturalist*, 157:245–261, 2001.

[14] J. G. Kingsolver, R. Gomulkiewicz, and P. A. Carter. Variation, selection and evolution of function-valued traits. *Genetica*, 112:87–104, 2001.

[15] J. G. Kingsolver, G. J. Ragland, and J. G. Shlichta. Quantitative genetics of continuous reaction norms: Thermal sensitivity of caterpillar growth rates. *Evolution*, 58:1521–1529, 2004.

[16] M. Kirkpatrick and N. Heckman. A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. *J. Math. Biol.*, 27:429–450, 1989.

[17] R. Lande. Natural selection and random genetic drift in phenotypic evolution. *Evolution*, 30:314–334, 1976.

[18] R. Lande. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution*, 33:402–416, 1979.

[19] R. Lande. Genetic variation and phenotypic evolution during allopatric speciation. *The American Naturalist*, 116:463–479, 1980.

[20] R. Lande and S. J. Arnold. The measurement of selection on correlated characters. *Evolution*, 37:1210–1226, 1983.

[21] M. N. Lukić. *Stochastic processes having sample paths in reproducing kernel Hilbert spaces with an application to white noise analysis.* PhD thesis, University of Wisconsin–Milwaukee, 1996.

[22] M. N. Lukić and J. H. Beder. Stochastic processes with sample paths in reproducing kernel Hilbert spaces. *Transactions of the American Mathematical Society*, 353(10):3945–3969, 2001.

[23] J. G. Mezey, D. Houle, and S. V. Nuzhdin. Naturally segregating quantitative trait loci affecting wing shape of Drosophila melanogaster. *Genetics*, 169:2101–2113, 2005.

[24] J. Neveu. *Processus Aléatoires Gaussiens.* Publications du Séminaire de Mathématiques Supérieures. Les Presses de l'Université de Montréal, 1968.

[25] K. Pearson. Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. *Philosophical Transactions of the Royal Society of London, Series A*, 200:1–66, 1903.

[26] S. D. Pletcher and C. J. Geyer. The genetic analysis of age-dependent traits: Modeling the character process. *Genetics*, 153:825–835, 1999.

[27] G. J. Ragland and P. A. Carter. Genetic covariance structure of growth in the salamander Ambystoma macrodactylum. *Heredity*, 92:569–578, 2004.

[28] J. Schmitt, J. R. Stinchcombe, M. S. Heschel, and H. Huber. The adaptive evolution of plasticity: Phytochrome-mediated shade avoidance responses. *Integrative and Comparative Biology*, 43:459–469, 2003.

[29] J. Travis. The role of optimizing selection in natural populations. *Annual Review of Ecology and Systematics*, 20:279296, 1989.

[30] J. K. Ward and J. K. Kelly. Scaling up evolutionary responses to elevated $CO_2$: Lessons from Arabidopsis. *Ecology Letters*, 7:427–440, 2004.

[31] S. S. Y. Young and H. Weiler. Selection for two correlated traits by independent culling levels. *Journal of Genetics*, 57:329–338, 1960.