

Orthogonal polynomials for a population

Orthogonality of functions

We say that two vectors are “orthogonal” if the projection of either one onto the other is zero, meaning that the inner product (or “dot product”) between them is zero. For vectors in Euclidean space, this is equivalent to saying that the vectors are at right angles, but the concept of orthogonality extends to cases in which “right angles” has no clear meaning (Jackson, 1941).

In particular, functions can be orthogonal to one another, but here the inner product is slightly more complicated than it is for vectors. Specifically, the inner product of two functions is defined relative to a separate “weight function” (the biological meaning of which will be made clear below). The inner product of two functions, $f(\phi)$ and $g(\phi)$, with respect to a weight function, $\omega(\phi)$, is written $\langle f, g \rangle_\omega$ (note the single angle brackets, distinguishing this from probability covariance). This is calculated as:

$$\langle f, g \rangle_\omega = \int_{-\infty}^{\infty} f(\phi)g(\phi)\omega(\phi)d\phi \quad (1)$$

The functions f and g are said to be “orthogonal” to one another given weight function ω if $\langle f, g \rangle_\omega = 0$.

Figure 1 illustrates one way to visualize orthogonality of two functions. If either function has a mean of zero (when weighted by ω), then saying that the functions are orthogonal is equivalent to saying that they are uncorrelated. In Figure 1A, the function $f = \phi^3$ is clearly correlated with ϕ (the red line is the regression) over the interval from -1 to 1. Figure 1B shows that a different cubic polynomial, $h = \phi^3 - \frac{3}{5}\phi$, is uncorrelated with ϕ over this interval. Since h (as well as ϕ) has a mean of zero on $[-1, 1]$, this means that $h = \phi^3 - \frac{3}{5}\phi$ is orthogonal to $g = \phi$.

The figure also shows the importance of the weight function, one role of which is to specify the range over which we compare the other functions. Figure 1C shows that if we change the weight function (ω) so that we are comparing the functions over the interval $[-\frac{1}{2}, \frac{1}{2}]$, then they are not orthogonal over this interval.

Different sets of orthogonal polynomials are defined by using different weight functions. For example, the Legendre polynomials, used to study the evolution of continuous growth processes (Kirkpatrick et al., 1990) as well as to solve certain differential equations in physics, use a uniform distribution on the interval $[-1, 1]$ as their weight function. By contrast, the Hermite Polynomials (in probability) have the weight function $\omega = e^{-\frac{\phi^2}{2}}$. The weight function, ω , need not be continuous; meaning that we can use the distribution of phenotypic variation to construct orthogonal polynomials that are tailored to a population.

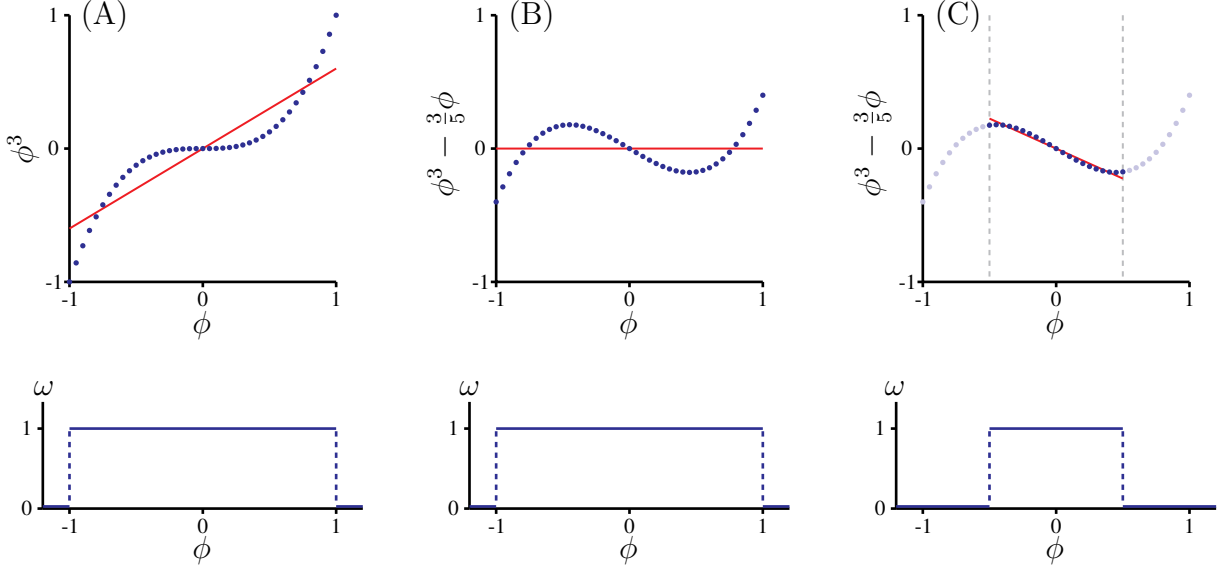


Figure 1: **Orthogonality of functions:** (A) shows that the function $f = \phi^3$ is not orthogonal to the function $g = \phi$ on the interval $[-1, 1]$ (i.e. with the weight function, ω , being 1 on the interval $[-1, 1]$ and 0 everywhere else). The red line is the regression of ϕ^3 on ϕ . Since $\bar{g} = 0$ on this interval, the slope of this regression ($\left[\left[\frac{\phi^3}{\phi}\right]\right]$) would be zero if the functions were orthogonal. (B) shows that the function $h = \phi^3 - \frac{3}{5}\phi$ is orthogonal to $g = \phi$ on this interval, since the slope of the regression is zero. (C) demonstrates the importance of the weight function. If we consider the interval $[-\frac{1}{2}, \frac{1}{2}]$ ($\omega = 1$ on $[-\frac{1}{2}, \frac{1}{2}]$ and 0 elsewhere), we find the regression slope to be clearly nonzero, meaning that h and g are not orthogonal with respect to this weight function.

Defining the weight function as the distribution of phenotypic variation in a population

The key to this approach is to define the weight function as the distribution of phenotypic variation in a population, assigning a weight of $\frac{1}{N}$ for each individual present in the population. For any finite population, this means that the integral in Equation 1 becomes a summation over all individuals in the population, as in Equation 2. (The formal transition from integral equation 1 to Equation 2 is given in the appendix.)

$$\langle f, g \rangle_\omega = \frac{1}{N} \sum_{i=1}^N f(\phi_i)g(\phi_i) = \overline{fg} \quad (2)$$

If either of our functions, f or g , has a mean value of zero in the population, then the inner product becomes simply the frequency covariance:

$$\langle f, g \rangle_\omega = \llbracket f, g \rrbracket \quad \text{if} \quad \overline{f(\phi)} = 0 \quad \text{or} \quad \overline{g(\phi)} = 0 \quad (3)$$

Recalling that two functions are orthogonal if their inner product is zero, this justifies our visualization of orthogonality as lack of covariation, illustrated in Figure 1.

(Note: Throughout this discussion, we denote the covariance between f and g as $\llbracket f, g \rrbracket$, the variance of g as $\llbracket^2 g \rrbracket$, and the regression of f on g as $\left\llbracket \frac{f}{g} \right\rrbracket$.)

Note that, according to Equation 2, the inner product over a population of any function, f , with the constant 1 gives us \bar{f} , the mean value of the function in the population. This is why we set the zeroth order polynomial equal to 1:

$$\mathbb{P}^{[0]} = 1 \quad (4)$$

The inner product of 1 with itself is also 1, so the projection of f onto $\mathbb{P}^{[0]}$ (Equation 5) is also \bar{f} .

Constructing orthogonal functions using the Gram-Schmidt process

Figure 2 illustrates how we can construct orthogonal vectors via the Gram-Schmidt process.

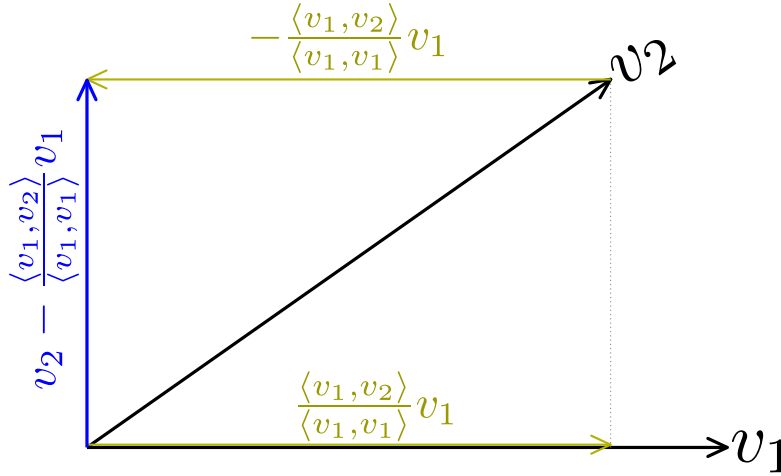


Figure 2:

Given two vectors, v_1 and v_2 , that are not orthogonal to one another, we can construct a third vector capturing the component of v_2 that is orthogonal to v_1 . This is achieved by projecting v_2 onto v_1 , then subtracting this projection from v_2 .

The projection of v_2 onto v_1 is found by taking the inner product of v_1 and v_2 ($\langle v_1, v_2 \rangle$), then dividing this by the inner product of v_1 with itself ($\langle v_1, v_1 \rangle$), and finally multiplying this ratio by v_1 . This projection is the yellow vector pointing along v_1 in Figure 2.

The key to constructing orthogonal polynomials is that we can use the population inner product in Equation 2 in the same way that we use the vector inner product in Figure 2. Given the weight function in Equation 27, and that we set $\bar{g} = 0$, the projection of f onto g is:

$$\frac{\langle f, g \rangle_\omega}{\langle g, g \rangle_\omega} = \frac{\llbracket f, g \rrbracket}{\llbracket^2 g \rrbracket} = \left\llbracket \begin{matrix} f \\ g \end{matrix} \right\rrbracket \quad (5)$$

Thus, if $\bar{g} = 0$, the projection of f onto g is just the slope of the (least squares linear) regression of f on g . This is why all polynomials of order ≥ 1 are set so that their mean is zero - doing so converts inner products into moments, and projections into regressions.

Starting with first order terms, we use Gram-Schmidt orthogonalization to construct new polynomials that are orthogonal to these. Under this approach, a function for f independent of g is constructed by starting with f and subtracting out the projection of f on g .

Using $f_{\perp g}$ to denote a function in f that is orthogonal to g , and setting $\bar{g} = 0$ so that we can use Equation 5 for the projection, we get:

$$f_{\perp g} = f - \left\llbracket \begin{matrix} f \\ g \end{matrix} \right\rrbracket g \quad (6)$$

We can apply this method sequentially to construct bases in other variables that are orthogonal to these. Given a third variable, h , we find the function of h orthogonal to both f and g as:

$$\begin{aligned} h_{\perp f, g} &= h - \left\llbracket \begin{matrix} h \\ f_{\perp g} \end{matrix} \right\rrbracket f_{\perp g} - \left\llbracket \begin{matrix} h \\ g \end{matrix} \right\rrbracket g \\ &= h - \left\llbracket \begin{matrix} h \\ g_{\perp f} \end{matrix} \right\rrbracket g_{\perp f} - \left\llbracket \begin{matrix} h \\ f \end{matrix} \right\rrbracket f \end{aligned} \quad (7)$$

Note that there are (at least) two orthogonal bases spanning the $f - g$ space, $(f, g_{\perp f})$ and $(f_{\perp g}, g)$. Either will suffice for calculating $h_{\perp f, g}$. What we may not do is subtract the projection of h on f and the projection of h on g (unless f and g happen to be orthogonal already).

$$h_{\perp f, g} \neq h - \left\llbracket \begin{matrix} h \\ f \end{matrix} \right\rrbracket f - \left\llbracket \begin{matrix} h \\ g \end{matrix} \right\rrbracket g \quad (8)$$

In other words, the Gram-Schmidt method works sequentially. At each step, we construct an orthogonal basis for a set of functions, then include the next function by subtracting its projection into the orthogonal basis that we already have.

Univariate orthogonal polynomials

In this case, there is a single first order polynomial:

$$\mathbb{P}^{[1]} = \phi - \bar{\phi} \quad (9)$$

We subtract out $\bar{\phi}$ so that $\mathbb{P}^{[1]}$ will have a mean of zero. The second order polynomial, $\mathbb{P}^{[2]}$, is constructed so that it satisfies the following criteria:

1. The leading (first) term is $(\mathbb{P}^{[1]})^2$.
2. It is orthogonal to $\mathbb{P}^{[1]}$.
3. It has a mean of zero.

Using Equation 6, and the fact that $\overline{(\mathbb{P}^{[1]})^2} = \llbracket^2 \phi \rrbracket$, these conditions give us:

$$\mathbb{P}^{[2]} = (\mathbb{P}^{[1]})^2 - \left[\left[\frac{(\mathbb{P}^{[1]})^2}{\mathbb{P}^{[1]}} \right] \right] \mathbb{P}^{[1]} - \llbracket^2 \phi \rrbracket \quad (10)$$

The regression of one polynomial in ϕ on another can be written in terms of central moments of ϕ . For example, to get the equation for $\mathbb{P}^{[2]}$ used in the text, note that:

$$\left[\left[\frac{(\mathbb{P}^{[1]})^2}{\mathbb{P}^{[1]}} \right] \right] = \frac{\llbracket^3 \phi \rrbracket}{\llbracket^2 \phi \rrbracket} \quad (11)$$

We construct the third order polynomial using the same approach. Following Equation 7, and subtracting $(\mathbb{P}^{[1]})^3 = \llbracket^3 \phi \rrbracket$ so that the mean will be zero, we get:

$$\mathbb{P}^{[3]} = (\mathbb{P}^{[1]})^3 - \left[\left[\frac{(\mathbb{P}^{[1]})^3}{\mathbb{P}^{[2]}} \right] \right] \mathbb{P}^{[2]} - \left[\left[\frac{(\mathbb{P}^{[1]})^3}{\mathbb{P}^{[1]}} \right] \right] \mathbb{P}^{[1]} - \llbracket^3 \phi \rrbracket \quad (12)$$

Repeating this procedure produces sequentially higher order orthogonal polynomials. The entire set of such polynomials forms the basis of a function space. Any variable can be projected into this space. Equivalently, we can write any variable of interest as a sum of terms, each containing the projection (regression) of the variable on one of the polynomials. The sum of the first $N - 1$ terms, where N is the number of distinct values on the ϕ axis, yields a function that passes through each point (or the mean, if multiple traits share the same ϕ value) (Figure 3).

Multivariate orthogonal polynomials

First order

For the i^{th} trait, ϕ_i , we write the first order polynomial as:

$$\mathbb{P}^i = \phi_i - \bar{\phi}_i \quad (13)$$

Note that there are no brackets around the superscript, indicating that this is the polynomial for trait i , rather than the i^{th} order polynomial.

For two traits, \mathbb{P}^1 and \mathbb{P}^2 span a two dimensional space that includes all values of ϕ_1 and ϕ_2 , but they are not likely to be orthogonal. We thus need to construct first order orthogonal polynomials that span this space.

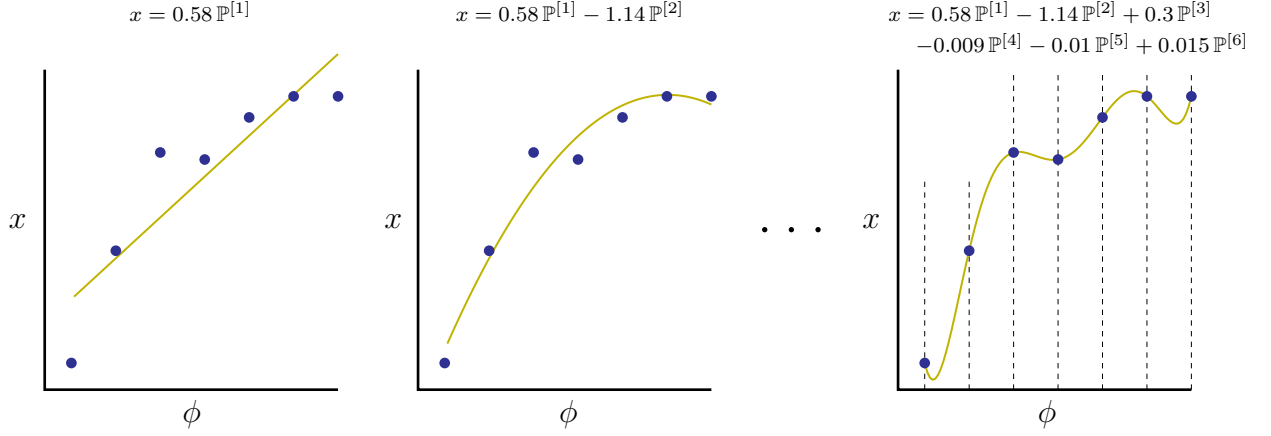


Figure 3: **Capturing points with a series of orthogonal polynomials:** *Using only the first term in the polynomial expansion yields the least squares linear approximation to the data. The sum of the first and second yields the best fit quadratic function. The sum of the first $N - 1$ terms, where N is the number of distinct values on the ϕ axis, yields a function that passes through each point (or the mean, if multiple trait values correspond to a single ϕ value). The combination of this function and the initial values of ϕ (vertical dashed lines) exactly specifies the mean of x for each value of ϕ . Note also that the first and second order coefficients do not change as we add higher order terms, as they would if we were doing an ordinary multiple regression on powers of ϕ .*

We will use \mathbb{P}_j^i to denote a polynomial in ϕ_i that is orthogonal to ϕ_j . Following Equation 6, we calculate:

$$\mathbb{P}_1^2 = \mathbb{P}^2 - \left[\begin{matrix} \mathbb{P}^2 \\ \mathbb{P}^1 \end{matrix} \right] \mathbb{P}^1 \quad (14)$$

For more than two traits, we carry on following Equation 7:

$$\mathbb{P}_{12}^3 = \mathbb{P}^3 - \left[\begin{matrix} \mathbb{P}^3 \\ \mathbb{P}_1^2 \end{matrix} \right] \mathbb{P}_1^2 - \left[\begin{matrix} \mathbb{P}^3 \\ \mathbb{P}^1 \end{matrix} \right] \mathbb{P}^1 \quad (15)$$

$$= \mathbb{P}^3 - \left[\begin{matrix} \mathbb{P}^3 \\ \mathbb{P}_2^1 \end{matrix} \right] \mathbb{P}_2^1 - \left[\begin{matrix} \mathbb{P}^3 \\ \mathbb{P}^2 \end{matrix} \right] \mathbb{P}^2 \quad (16)$$

In this way, we construct a set of first order orthogonal polynomials for our set of traits. We use \mathbb{P}_\bullet^i to denote the polynomial in i that is orthogonal to all others of the same or lower order. Thus, for the case of four traits, the polynomial in ϕ_2 that is orthogonal to ϕ_1 , ϕ_3 , and ϕ_4 is $\mathbb{P}_\bullet^2 = \mathbb{P}_{134}^2$.

Given another variable, a , the regression of a on \mathbb{P}^2 is the same as the simple regression of a on ϕ_2 , while the regression of a on \mathbb{P}_\bullet^2 is equivalent to the *partial* regression of a on ϕ_2 (Saville & Wood, 1991).

Second order

For the multivariate case, we denote second order polynomials using two superscripts, so \mathbb{P}^{11} is the second order polynomial with $(\mathbb{P}^1)^2$ as the leading term, \mathbb{P}^{12} is the second order polynomial with $\mathbb{P}^1\mathbb{P}^2$ as its leading term, and so forth. (If we are considering ≥ 10 traits we can separate them with commas. For this discussion, though, we will omit the commas since we will be working with, at most, three traits. We thus write \mathbb{P}^{11} instead of $\mathbb{P}^{1,1}$).

We want the second order polynomials to be orthogonal to all of the first order polynomials, so we apply the method from Equation 7:

$$\begin{aligned}\mathbb{P}^{11} &= (\mathbb{P}^1)^2 - \left[\begin{smallmatrix} (\mathbb{P}^1)^2 \\ \mathbb{P}_1^2 \end{smallmatrix} \right] \mathbb{P}_1^2 - \left[\begin{smallmatrix} (\mathbb{P}^1)^2 \\ \mathbb{P}^1 \end{smallmatrix} \right] \mathbb{P}^1 - \llbracket^2 \phi_1 \rrbracket \\ &= (\mathbb{P}^1)^2 - \left[\begin{smallmatrix} (\mathbb{P}^1)^2 \\ \mathbb{P}_2^1 \end{smallmatrix} \right] \mathbb{P}_2^1 - \left[\begin{smallmatrix} (\mathbb{P}^1)^2 \\ \mathbb{P}^2 \end{smallmatrix} \right] \mathbb{P}^2 - \llbracket^2 \phi_1 \rrbracket\end{aligned}\tag{17}$$

As in the univariate case, we subtract out $\overline{(\mathbb{P}^1)^2} = \llbracket^2 \phi_1 \rrbracket$ so that \mathbb{P}^{11} will have a mean of zero.

The two versions of Equation 17 correspond to two different ways of projecting $(\mathbb{P}^1)^2$ into the space spanned by the first order polynomials. Since any of the many ways of doing this will suffice, we use $\mathcal{P}^i\{g\}$ to denote the projection of g into the space spanned by all of the i^{th} order polynomials (note that $\mathcal{P}^{[n]}$ is orthogonal to $\mathcal{P}^{[m]}$ for all $m < n$). With this notation, we can write the second order polynomials for two traits as:

$$\begin{aligned}\mathbb{P}^{11} &= (\mathbb{P}^1)^2 - \mathcal{P}^1\{(\mathbb{P}^1)^2\} - \llbracket^2 \phi_1 \rrbracket \\ \mathbb{P}^{12} &= \mathbb{P}^1\mathbb{P}^2 - \mathcal{P}^1\{(\mathbb{P}^1\mathbb{P}^2)\} - \llbracket \phi_1, \phi_2 \rrbracket \\ \mathbb{P}^{22} &= (\mathbb{P}^2)^2 - \mathcal{P}^1\{(\mathbb{P}^2)^2\} - \llbracket^2 \phi_2 \rrbracket\end{aligned}\tag{18}$$

Though the second order polynomials in Equations 18 are orthogonal to all first order polynomials, they are not, in general, orthogonal to one another. To construct second order polynomials that are orthogonal to one another, we work sequentially using the Gram-Schmidt method. The polynomial in $\phi_1\phi_2$ that is orthogonal to \mathbb{P}^{11} and to all first order polynomials is found as:

$$\mathbb{P}_{11}^{12} = \mathbb{P}^1\mathbb{P}^2 - \left[\begin{smallmatrix} \mathbb{P}^1\mathbb{P}^2 \\ \mathbb{P}^{11} \end{smallmatrix} \right] \mathbb{P}^{11} - \mathcal{P}^1\{(\mathbb{P}^1\mathbb{P}^2)\} - \llbracket \phi_1, \phi_2 \rrbracket\tag{19}$$

The second term on the righthand side of Equation 19 makes it orthogonal to \mathbb{P}^{11} , the third term makes it orthogonal to all first order terms, and the fourth term makes it so that the mean of \mathbb{P}_{11}^{12} is zero.

Note that the first, third, and fourth terms on the righthand side of Equation 19 are the same at the righthand side of the equation for \mathbb{P}^{12} in Equation 18. This means that we can write Equation 19 more compactly as:

$$\mathbb{P}_{11}^{12} = \mathbb{P}^{12} - \left[\begin{smallmatrix} \mathbb{P}^1\mathbb{P}^2 \\ \mathbb{P}^{11} \end{smallmatrix} \right] \mathbb{P}^{11}\tag{20}$$

Given this term, we find the polynomial in $(\mathbb{P}^2)^2$ that is orthogonal to all other second order terms as well as all first order terms as:

$$\mathbb{P}_{11,12}^{22} = \mathbb{P}_{\bullet}^{22} = \mathbb{P}^{22} - \left[\begin{array}{c} (\mathbb{P}^2)^2 \\ \mathbb{P}_{11}^{12} \end{array} \right] \mathbb{P}_{11}^{12} - \left[\begin{array}{c} (\mathbb{P}^2)^2 \\ \mathbb{P}^{11} \end{array} \right] \mathbb{P}^{11} \quad (21)$$

Following this approach, the remaining conditional second order terms can be written as:

$$\begin{aligned} \mathbb{P}_{\bullet}^{11} &= \mathbb{P}^{11} - \left[\begin{array}{c} (\mathbb{P}^1)^2 \\ \mathbb{P}_{22}^{12} \end{array} \right] \mathbb{P}_{22}^{12} - \left[\begin{array}{c} (\mathbb{P}^1)^2 \\ \mathbb{P}^{22} \end{array} \right] \mathbb{P}^{22} \\ \mathbb{P}_{\bullet}^{12} &= \mathbb{P}^{12} - \left[\begin{array}{c} \mathbb{P}^1 \mathbb{P}^2 \\ \mathbb{P}_{11}^{22} \end{array} \right] \mathbb{P}_{11}^{22} - \left[\begin{array}{c} \mathbb{P}^1 \mathbb{P}^2 \\ \mathbb{P}^{11} \end{array} \right] \mathbb{P}^{11} \end{aligned} \quad (22)$$

Biorthogonal Bases

All of the moments and regressions that we use are calculated using inner products. In fact, the principle reason that we seek orthogonal bases is that it is easy to calculate inner products in them. In a non-orthogonal basis, the standard methods for calculating inner products (and thus for calculating covariances, regressions, etc.) fail because different coordinate vectors (or functions) are correlated with one another.

Multivariate orthogonal polynomials pose a substantial challenge that does not appear in the univariate case. As shown above, finding a set of orthogonal polynomials for a multivariate distribution is relatively easy, the problem is that there are many such sets, and which one we get depends on how we choose to order our initial variables (Dunkl & Xu, 2001). Figure 4 illustrates this: The three pairs of colored vectors on the left represent three different bases that are orthogonal with respect to the two dimensional distribution of points shown. (Note that these bases are orthogonal with respect to a discrete distribution of points, not with respect to a uniform distribution over the space. This is why the pairs of orthogonal vectors are not at right angles. Vectors that were at right angles in the drawing would not be orthogonal with respect to the distribution of points shown). The small plots on the right of Figure 4 show the points projected into each of the bases shown on the left. The fact that these bases are orthogonal with respect to these points is manifest in the fact that there is no correlation in any of the plots on the right.

Because each of the colored bases in the figure are orthogonal with respect to the distribution, we could project fitness and offspring phenotype into any one of them and get the right answer for change in mean phenotype. Unfortunately, although the answers would be mathematically equivalent, they would be written in terms of different sets of parameters - none of which capture all of the terms that we want in our model.

For instance, if we choose to represent fitness in the blue coordinates in figure 4, we will end up with a term for the regression of fitness on ϕ_1 (since b_1 points along the ϕ_1 axis), but there will be no corresponding term for the regression of fitness on ϕ_2 , only on b_2 , which is orthogonal to b_1 . In other words; we would have the simple regression of fitness on ϕ_1 , and the *partial* regression of fitness on ϕ_2 .

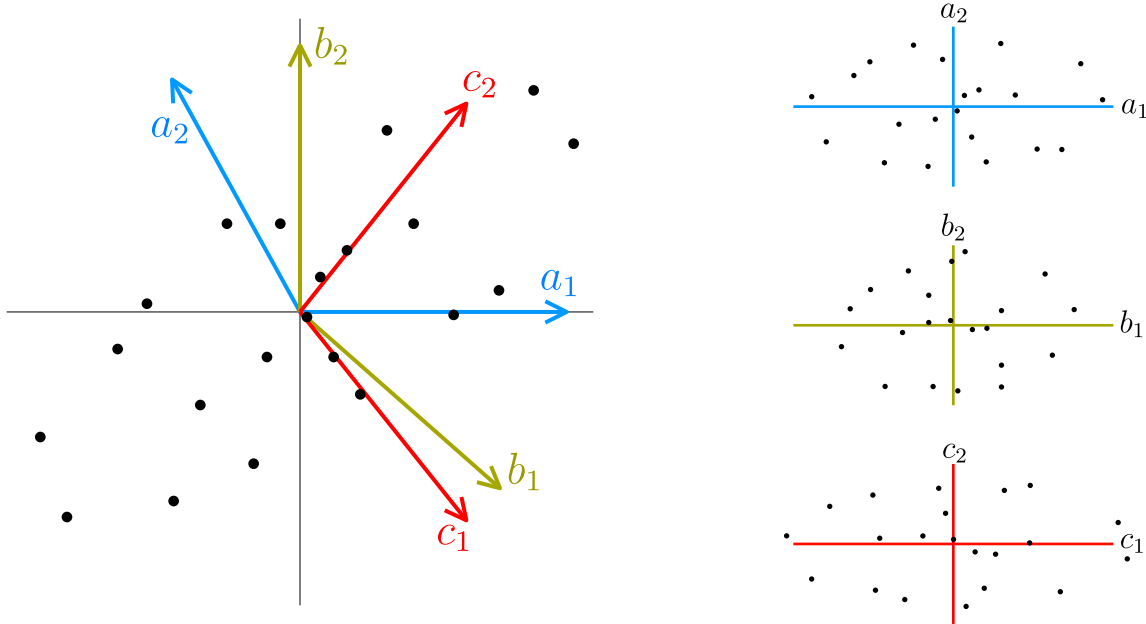


Figure 4: **Three different pairs of orthogonal bases for a two dimensional distribution of points.** *Axes of the same color are orthogonal to one another with respect to this particular distribution of points. The plots on the right show the points projected into each of the bases shown on the left.*

None of the obvious sets of orthogonal bases in a multivariate space is satisfactory for our purposes. However, we can construct a *pair* of bases that do the job. The trick is to find a pair of bases with the property that, for each basis, each axis in it is orthogonal to all but one of the axes in the other basis. A pair of bases with this property is said to be “biorthogonal”. Formally: for two bases in an n dimensional space, $X = \{x_1, x_2, \dots, x_n\}$ and $Y = \{y_1, y_2, \dots, y_n\}$, we say that X and Y are biorthogonal if:

$$\langle x_i, y_j \rangle = \begin{cases} c & j = i \\ 0 & j \neq i \end{cases} \quad (23)$$

where $c \neq 0$ is a constant. (In the special case of $c = 1$, we say that the bases are “dual”.)

Given a pair of biorthogonal bases, we can find the inner product of two functions if we first project one function into one of the pair of bases, and project the other function into the other basis (scaled by the projection of the second basis on the first). In fact, we have already done all of the work needed to get a pair of biorthogonal bases – we need only rearrange the terms that we have.

Figure 5 shows how, in two dimensions, we can construct a pair of biorthogonal bases from the two sets of orthogonal bases that we already have. We refer to the basis formed by \mathbb{P}^1 and \mathbb{P}^2 (which just measure the original ϕ values) as the simple basis, and that formed

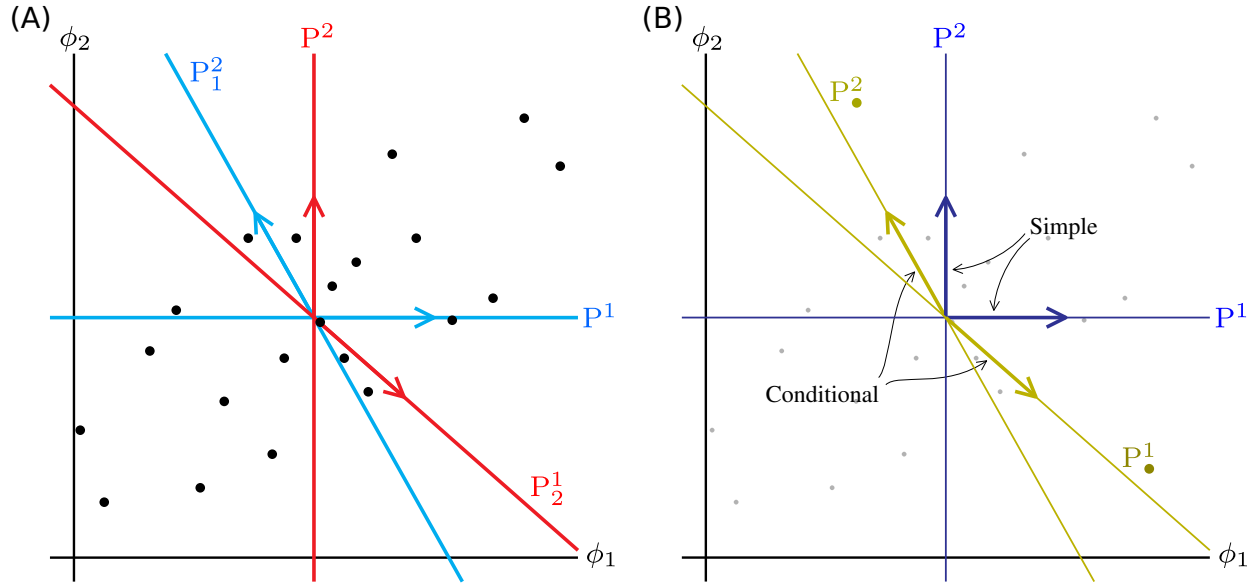


Figure 5: **Constructing biorthogonal bases from different sets of orthogonal bases:** The basis formed by vectors \mathbb{P}^1 and \mathbb{P}^2 (dark blue) and the basis formed by \mathbb{P}_\bullet^1 and \mathbb{P}_\bullet^2 (yellow) are biorthogonal to one another. \mathbb{P}_\bullet^1 and \mathbb{P}_\bullet^2 are the same as \mathbb{P}_2^1 and \mathbb{P}_1^2 from Figure 4. In our notation, \mathbb{P}^1 and \mathbb{P}^2 comprise the simple basis, and \mathbb{P}_2^1 and \mathbb{P}_1^2 the conditional basis.

by \mathbb{P}_\bullet^1 and \mathbb{P}_\bullet^2 (each constructed to be orthogonal to all but one of the simple vectors) as the conditional basis.

Table 4 shows how we construct simple and conditional basis that include higher order polynomials. Note that all polynomials, simple and conditional, of a given order are orthogonal to all lower orders.

When we project another variable into the space, we always end up with regressions on the conditional basis. The Conditional Regression Theorem (see the document “Proof of the conditional regression theorem”) shows that using biorthogonal bases yields – for first order terms – the same result as using standard multiple (partial) regression analysis. For higher order terms, the biorthogonal approach yields terms that show the effects of each order independent of lower orders. It thus does not lead to lower order terms changing as we add higher orders (as standard multiple regression using powers of ϕ would).

If we look at terms of order j , the biorthogonal analysis including j and higher orders will yield the same j^{th} order coefficients (but not the same coefficients for orders $< j$) as would a multiple regression analysis that went only up to order j .

Writing a variable in biorthogonal bases

Given a variable, f , we can write it in terms of our biorthogonal bases by regressing f on each conditional basis, then multiplying this regression by the corresponding simple basis.

Simple Basis	Conditional Basis
First Order	
$\mathbb{P}^1 = \phi_1 - \bar{\phi}_1$	$\mathbb{P}_{\bullet}^1 = \mathbb{P}^1 - \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} \mathbb{P}^2$
$\mathbb{P}^2 = \phi_2 - \bar{\phi}_2$	$\mathbb{P}_{\bullet}^2 = \mathbb{P}^2 - \begin{bmatrix} \phi_2 \\ \phi_1 \end{bmatrix} \mathbb{P}^1$
Second Order	
$\mathbb{P}^{11} = (\mathbb{P}^1)^2 - \mathcal{P}^1\{(\mathbb{P}^1)^2\} - \llbracket^2 \phi_1 \rrbracket$	$\mathbb{P}_{\bullet}^{11} = \mathbb{P}^{11} - \begin{bmatrix} \phi_1^2 \\ \mathbb{P}_{12}^{22} \end{bmatrix} \mathbb{P}_{12}^{22} - \begin{bmatrix} \phi_1^2 \\ \mathbb{P}_{12}^{12} \end{bmatrix} \mathbb{P}^{12}$
$\mathbb{P}^{12} = \mathbb{P}^1 \mathbb{P}^2 - \mathcal{P}^1\{\mathbb{P}^1 \mathbb{P}^2\} - \llbracket \phi_1, \phi_2 \rrbracket$	$\mathbb{P}_{\bullet}^{12} = \mathbb{P}^{12} - \begin{bmatrix} \phi_1 \phi_2 \\ \mathbb{P}_{11}^{22} \end{bmatrix} \mathbb{P}_{11}^{22} - \begin{bmatrix} \phi_1 \phi_2 \\ \mathbb{P}_{11}^{12} \end{bmatrix} \mathbb{P}^{11}$
$\mathbb{P}^{22} = (\mathbb{P}^2)^2 - \mathcal{P}^1\{(\mathbb{P}^2)^2\} - \llbracket^2 \phi_2 \rrbracket$	$\mathbb{P}_{\bullet}^{22} = \mathbb{P}^{22} - \begin{bmatrix} \phi_2^2 \\ \mathbb{P}_{11}^{12} \end{bmatrix} \mathbb{P}_{11}^{12} - \begin{bmatrix} \phi_2^2 \\ \mathbb{P}_{11}^{11} \end{bmatrix} \mathbb{P}^{11}$
Third Order	
$\mathbb{P}^{111} = (\mathbb{P}^1)^3 - \mathcal{P}^2\{(\mathbb{P}^1)^3\} - \mathcal{P}^1\{(\mathbb{P}^1)^3\} - \llbracket^3 \phi_1 \rrbracket$	$\mathbb{P}_{\bullet}^{111} = \mathbb{P}^{111} - \begin{bmatrix} \phi_1^3 \\ \mathbb{P}_{122}^{222} \end{bmatrix} \mathbb{P}_{122}^{222} - \begin{bmatrix} \phi_1^3 \\ \mathbb{P}_{122}^{122} \end{bmatrix} \mathbb{P}_{112}^{122} - \begin{bmatrix} \phi_1^3 \\ \mathbb{P}_{112}^{112} \end{bmatrix} \mathbb{P}^{112}$
\mathbb{P}^{112}	$\mathbb{P}_{\bullet}^{112}$
\vdots	\vdots

Table 1:

Defining \tilde{f} as the mean value of f for a particular combination of ϕ values (so $\tilde{f} = f$ if each individual has a unique phenotype), we can write \tilde{f} as:

$$\tilde{f} = \bar{f} + \sum_{i=1}^n \begin{bmatrix} f \\ \mathbb{P}_{\bullet}^i \end{bmatrix} \mathbb{P}^i + \sum_{i=1}^n \sum_{j \geq i} \begin{bmatrix} f \\ \mathbb{P}_{\bullet}^{ij} \end{bmatrix} \mathbb{P}^{ij} + \sum_{i=1}^n \sum_{j \geq i} \sum_{k \geq j} \begin{bmatrix} f \\ \mathbb{P}_{\bullet}^{ijk} \end{bmatrix} \mathbb{P}^{ijk} + \dots \quad (24)$$

Note that we are summing first over all traits, then over all pairs of traits, and so on. The summations in Equation 24 can obscure the basic pattern. We thus will use a summation convention: Writing $\mathbf{P}^{()}$ (or $\mathbf{P}_{\bullet}^{()}$) will indicate that we are summing over all orders and all combinations of variables within each order. For two variables, ϕ_1 and ϕ_2 :

$$\mathbf{P}^{()} = \mathbb{P}^1 + \mathbb{P}^2 + \mathbb{P}^{11} + \mathbb{P}^{12} + \mathbb{P}^{22} + \mathbb{P}^{111} + \mathbb{P}^{112} + \mathbb{P}^{122} + \mathbb{P}^{222} + \dots \quad (25)$$

The order of terms within a superscript does not matter, so $\mathbb{P}^{12} = \mathbb{P}^{21}$. Using this summation convention, we can write Equation 24 as:

$$\tilde{f} = \bar{f} + \begin{bmatrix} f \\ \mathbf{P}_{\bullet}^{()} \end{bmatrix} \mathbf{P}^{()} \quad (26)$$

References

- Dunkl, C. F. & Xu, Y. 2001: Orthogonal polynomials of several variables. Cambridge Univ. Press, New York.
- Jackson, D. 1941: Fourier series and orthogonal polynomials. Mathematical Association of America, Oberlin, OH.
- Kirkpatrick, M.; Lofsvold, D. & Bulmer, M. 1990: Analysis of the Inheritance, Selection and Evolution of Growth Trajectories. *Genetics* 124(4):979–993.
- Saville, D. J. & Wood, G. R. 1991: Statistical methods: the geometric approach. Springer-Verlag, New York.

Appendix

For a population of size N , in which individual i has phenotype ϕ_i , we define the weight function over the space of all possible phenotypic values as:

$$\omega(\phi) = \frac{1}{N} \sum_{i=1}^N \delta(\phi - \phi_i) \quad (27)$$

Where $\delta(\phi - \phi_i)$ is the Dirac delta function, which has a value of zero everywhere except at the point $\phi = \phi_i$, where its value is “infinite”. (Specifically, the magnitude of $\delta(\phi - \phi_i)$ at $\phi = \phi_i$ is defined such that $\int_{-\infty}^{\infty} \delta(\phi - \phi_i) = 1$). The weight function is thus basically a set of Dirac delta functions, each corresponding to one individual. We divide by N so that the weight function is itself a distribution that sums to 1. Figure 6 illustrates the weight function for a hypothetical population.

With this weight function, the inner product of any two functions of phenotype can be written as the mean of their product (Equation 2).

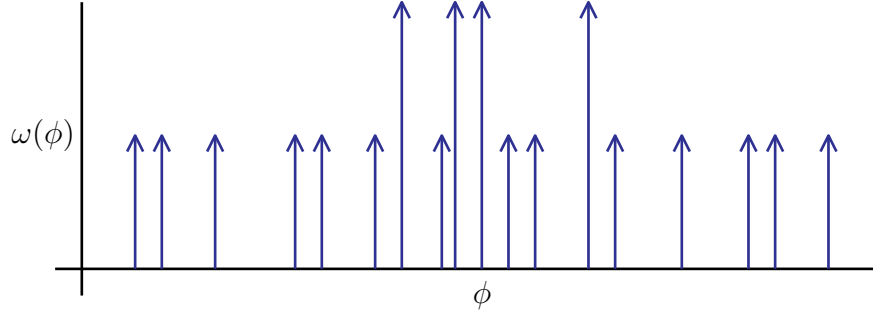


Figure 6: **The population weight function:** *A population represented as a set of Dirac delta functions, each corresponding to a phenotypic value present in the population. Arrows drawn twice as long indicate two individuals with the same phenotypic value. With this weight function, the product of the functions f and g in Equation 1 is evaluated only at phenotypic values that are present in the population – turning the integral in Equation 1 into the sum in Equation 2.*