

# Multiple-Peak Selection Surface Simulation Methods

Diogo Melo

24/06/2020

## Introduction

This document describes in detail the simulations used to study the interaction between differing levels of genetic integration and different selective surfaces. Our objective is to determine what aspects influence the total magnitude of phenotypic change, and to relate this to a simple observable metric that relates the direction of evolution with the pattern of covariation in the population. There are two main elements to the simulations: (1) The level of integration in the genetic covariance matrix of the population under selection; and (2) the ruggedness, or number of peaks, of the selective surface. We begin by developing a strategy to simulated simple and complex selective surfaces.

## The selective surface

We are interested in creating both very simple selective surfaces, with only a single phenotypic optimum for the population, and complex surfaces, with several optima distributed across the phenotypic space. In order to achieve this, we use a mixture of Gaussian functions. For the single peak landscape, we can simply sample a multivariate mean  $\theta$  from some random distribution and use the diagonal matrix ( $\Sigma$ ) as a covariance for the multivariate normal<sup>1</sup>. This induces the following mean fitness surface, for a  $k$ -dimensional phenotype  $x$ :

$$\bar{W}(x) = \mathcal{N}(x|\theta, \Sigma) = \det(2\pi\Sigma)^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\theta)^T \Sigma^{-1} (x-\theta)}$$

For a bivariate phenotype, we can visualize this surface using a color gradient (fig. 1). The triangle marks the position of the  $\theta$  parameter. The origin of the coordinate system marks where our populations will be in relation to the selective surface.

Creating more complex surfaces can follow basically the same strategy, but using several overlapping Gaussian functions. For this, we have to sample several  $\theta$  parameters and define the fitness as the sum of all the Gaussian functions. For a set of  $N$  Gaussian, each with a different  $\theta$  and the same  $\Sigma$ , we have:

---

<sup>1</sup>The  $\Sigma$  parameter could be any positive definite matrix, but we do not explore this parameter here, and use the identity matrix in all simulations.

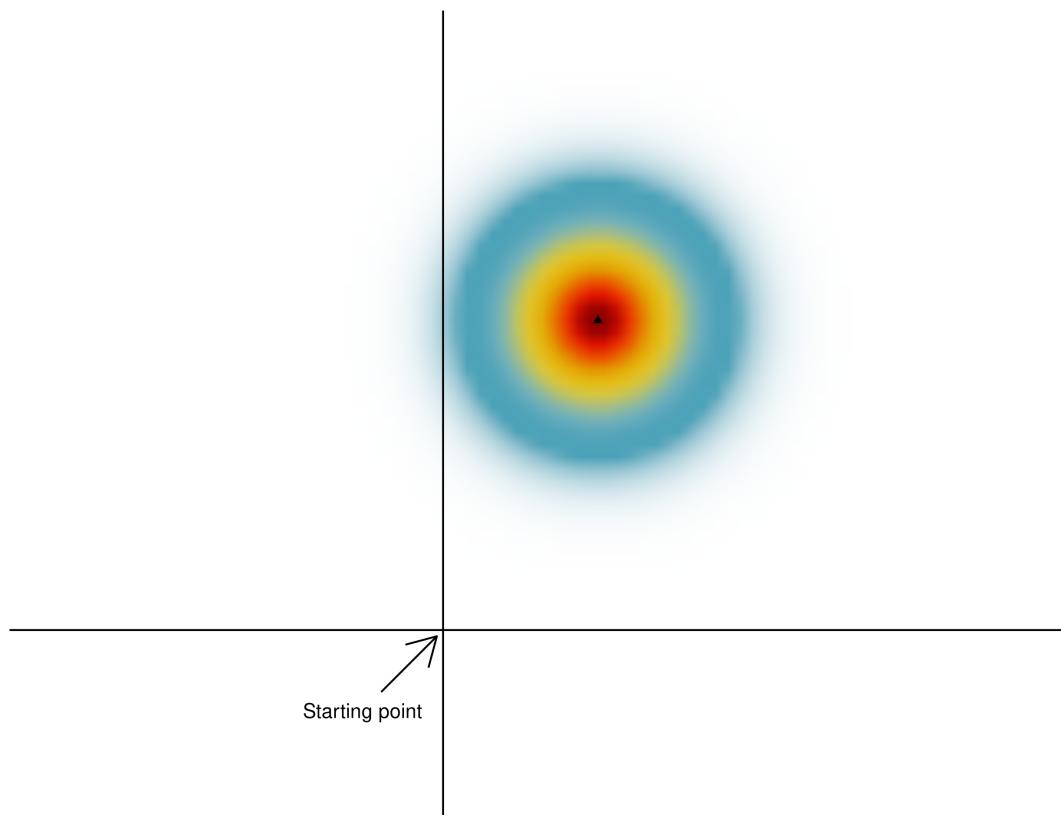


Figure 1: Single peak surface. The triangle marks the position of the  $\theta$  parameter. The origin of the coordinate system marks where our populations will be in relation to the selective surface.

$$\bar{W}(x) = \sum_{i=1}^N \mathcal{N}(x|\theta_i, \Sigma) = \det(2\pi\Sigma)^{-\frac{1}{2}} \sum_{i=1}^N e^{-\frac{1}{2}(x-\theta_i)^T \Sigma^{-1} (x-\theta_i)}$$

In order to sample the  $\theta$  parameters for the Gaussian functions, we could use uniform distributions in some interval to sample the  $(x, y)$  components for several optima, like in fig. 2. However, this causes problems in relation to the origin, as some directions have more peaks than others. The directions along the diagonals are longer than along the two axes, and so we end up with more peaks along the diagonals. One solution is to sample the  $(x, y)$  components from a normal distribution, which is spherically symmetrical, and then scale the resulting vectors to a magnitude sampled from a uniform distribution. This gives us control over the distance of the peaks from the origin and guarantees that all directions are equally likely, as in fig. 3. This procedure also generalizes to more dimensions, and avoids the problem of having all the peaks in a thin shell in multivariate space that would happen if we were to only sample all coordinates from a normal distribution and not sample the magnitudes separately.

The final restriction on the multiple peaks is to impose a minimal distance from the origin. This is necessary so that our population can actually evolve, and not start the simulation already at a phenotypic optima. We do this by restricting the distance of the Gaussian optima to be larger than some minimal distance. A set of  $\theta$  parameters that follow these rules is displayed in fig. 4, and this particular set of  $\theta$  results in the complex surface shown in fig. 5. We can see that the optima for the full surface don't necessarily coincide with one particular  $\theta$  and that this method can create complex surfaces, with several peaks, valleys, ridges, and saddle points.

## Interaction between genetic covariation and the selective surface

After we have created a surface, we can place a population at any point and use quantitative genetics theory to predict how the mean phenotype of the population will change over time. Assuming the additive genetic covariance matrix of the population ( $G$ -matrix) stays constant, at each generation the change in the mean phenotype  $\Delta z$  will be given by the Lande equation:

$$\Delta z = G\beta = G \frac{1}{\bar{W}} \nabla \bar{W} = G \nabla \ln \bar{W}$$

In this equation, the selection vector ( $\beta$ ) is given by the gradient of the natural logarithm of the selective surface. The choice of modeling the selective surface with a sum of Gaussian functions allows for an analytical calculation of the gradient, which greatly improves the efficiency of the simulations. Using numerical gradients would make simulations in high dimensional cases impossible.

For a single multivariate Gaussian function  $p(x|\theta, \Sigma)$ , with  $x \in \mathbb{R}^k$ , the gradient of  $p(x|\theta, \Sigma)$  is given by:

$$\frac{\partial p(x|\theta, \Sigma)}{\partial x} = -p(x)\Sigma^{-1}(x - \theta)$$

For a complex surface defined as the sum of several Gaussian functions, with  $p_i(x) = p(x|\theta_i, \Sigma)$ ,

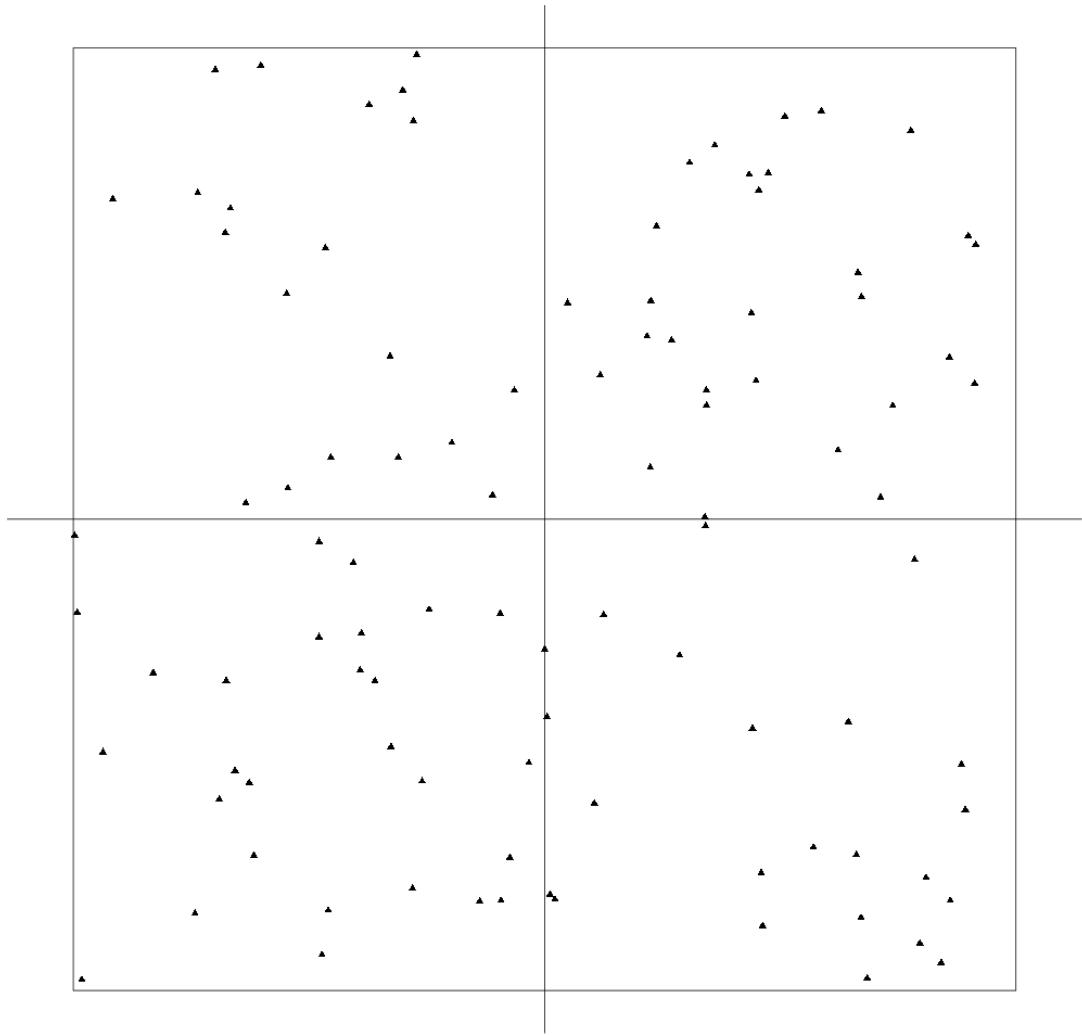


Figure 2: Multiple Gaussian optima with uniform coordinates. All the peaks are restricted to be inside the square, but this causes a bias in direction in relation to the origin. Directions along the diagonal are longer, and so have more peaks along them.

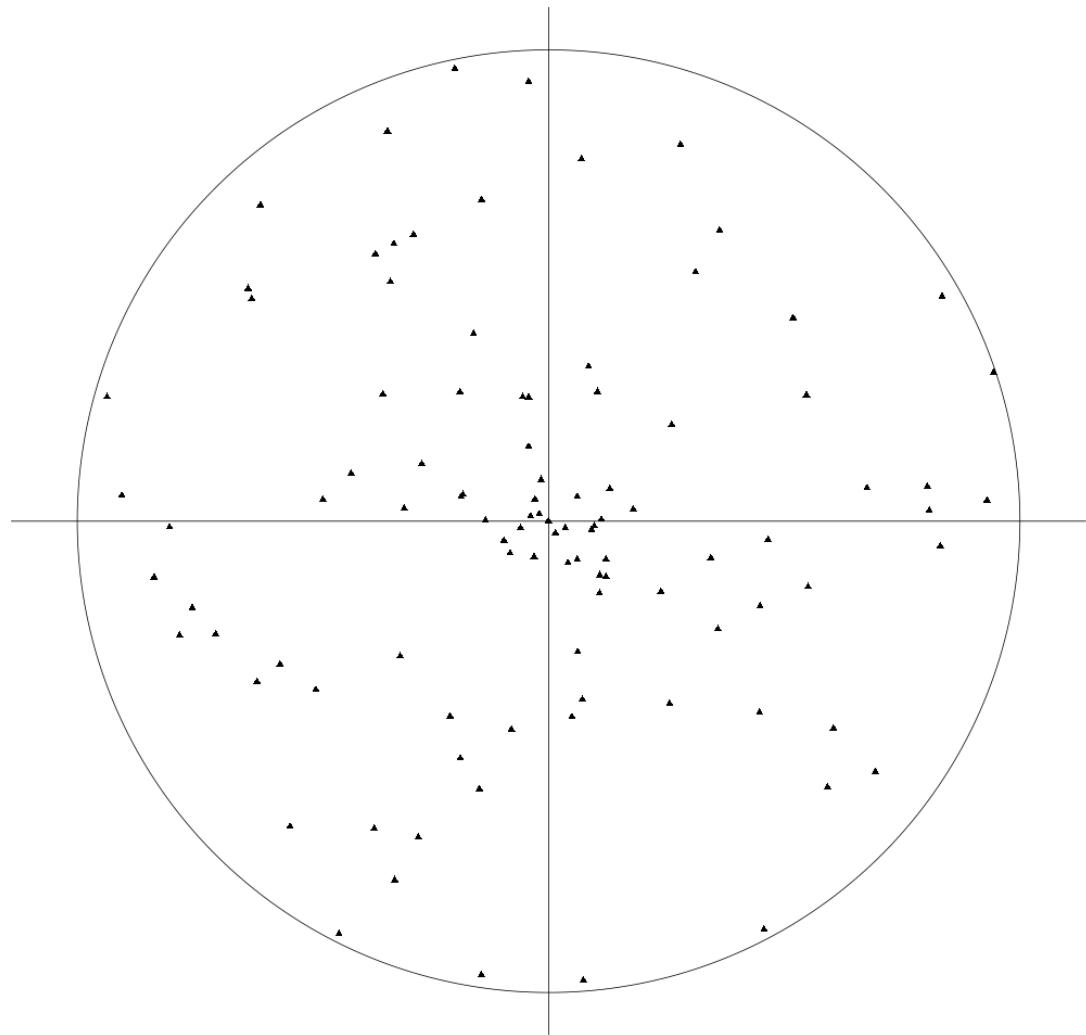


Figure 3: Multiple Gaussian optima with spherical symmetry and uniform distance from origin.

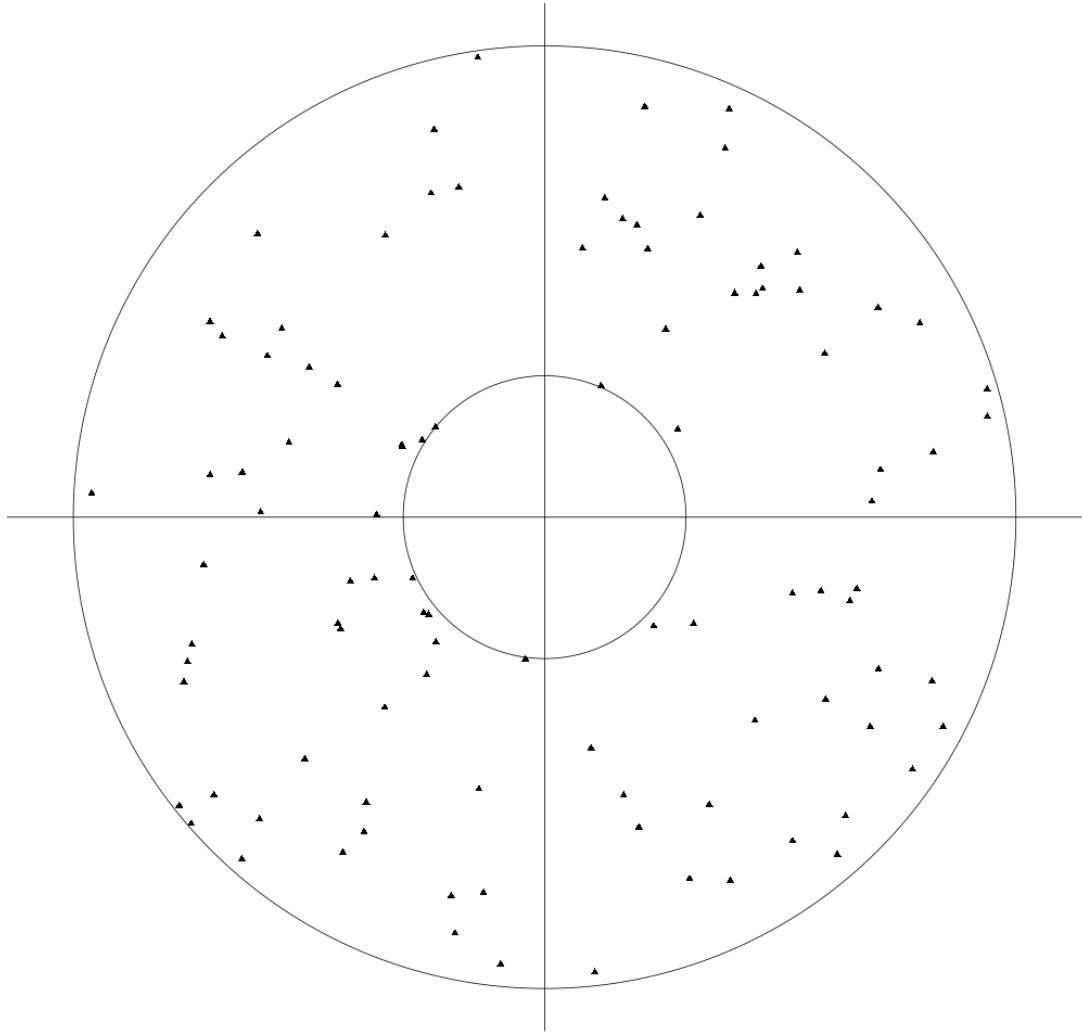


Figure 4: Multiple Gaussian optima with spherical symmetry and a minimal distance from the origin.

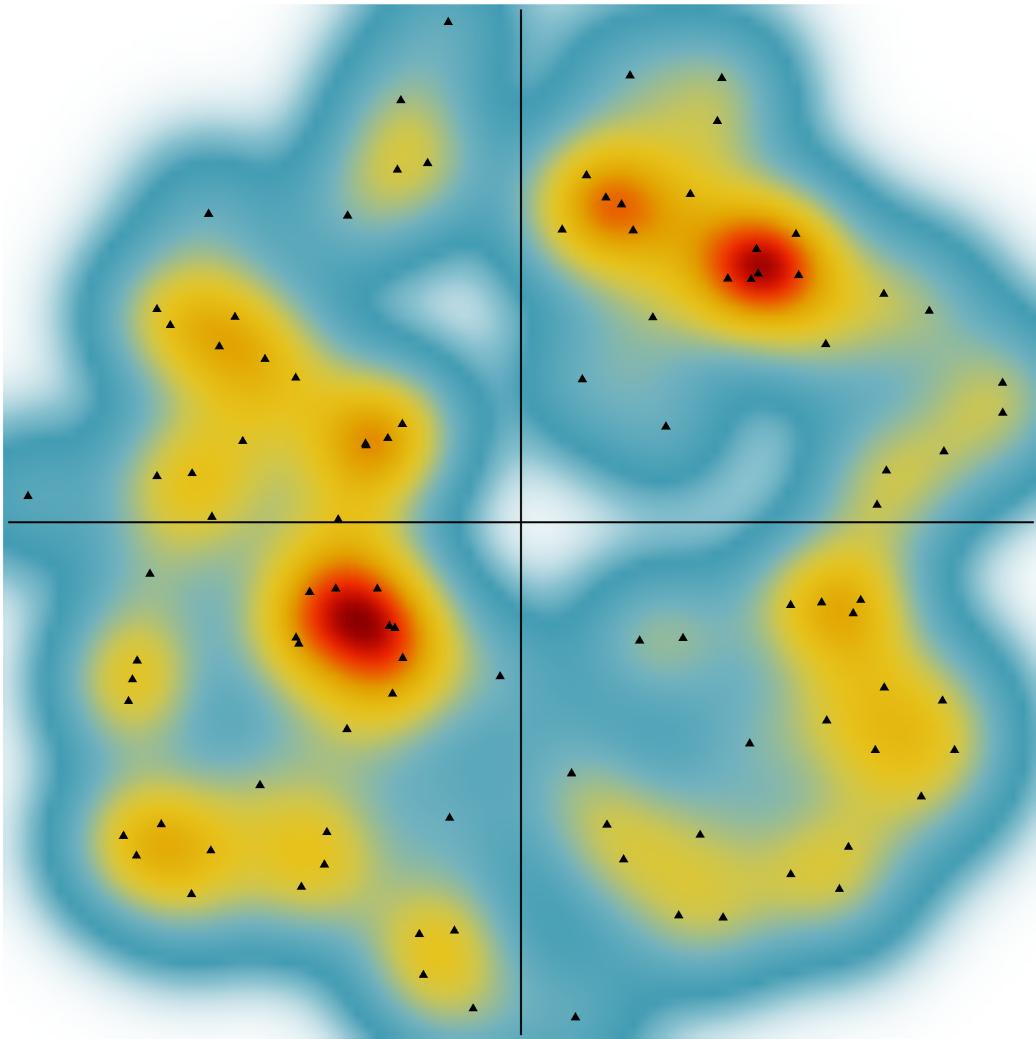


Figure 5: Surface generated by the multiple Gaussian optima in +fig:peaks\_ring.

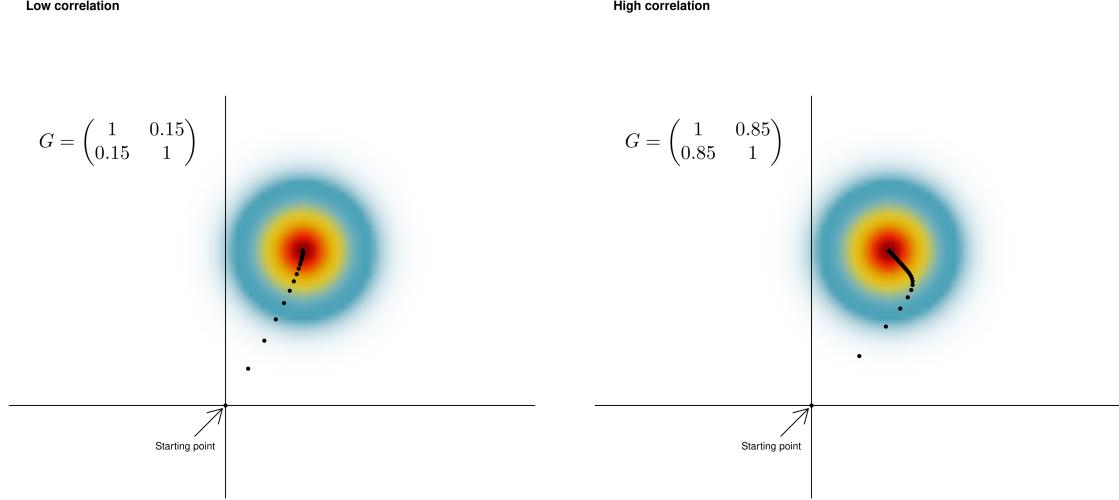


Figure 6: Different G-matrices produce different trajectories for the surface in fig. 1.

$\bar{W}(x|\theta_1, \dots, \theta_n, \Sigma) = \sum_{i=1}^n p_i(x)$ , the gradient of the logarithm of  $\bar{W}(x)$  is given by:

$$\nabla \ln \bar{W}(x|\theta_1, \dots, \theta_n, \Sigma) = \frac{1}{\bar{W}(x)} \sum_{i=1}^n -p_i(x) \Sigma^{-1}(x - \theta_i)$$

Multiplying the gradient at the position the population currently occupies by the G-matrix gives the phenotypic change, which can be added to the current position to obtain the new position of the population after selection. Iterating this processes gives a trajectory in phenotype space, which stops when the gradient is zero. Different covariance matrices will create different trajectories. For example, fig. 6 shows the difference between the trajectories of two populations with high and low integration. In this example, both populations end up on the same optimum. However, if the surface is more complex, with several optima, they can end up on different optima. Figure 7 shows a slightly more complex surface where the different correlations lead the populations to different optima on the same surface. As the surfaces become more complex, these cases become more frequent.

## Simulations

We use this framework of generating random selective surfaces to investigate the relation between the magnitude of the total phenotypic change ( $\|\Delta z\|$ ) and the alignment between the direction of phenotypic change and the main axis of genetic variation in the populations (given by the cosine of the angle between these vectors:  $\text{Cos}(gmax, \Delta z)$ , see fig. 8).

We use 4 different scenarios, presenting high and low integration populations with single- and multi-peaked selective surfaces. By generating several different surfaces, we can investigate the relation between these variables.

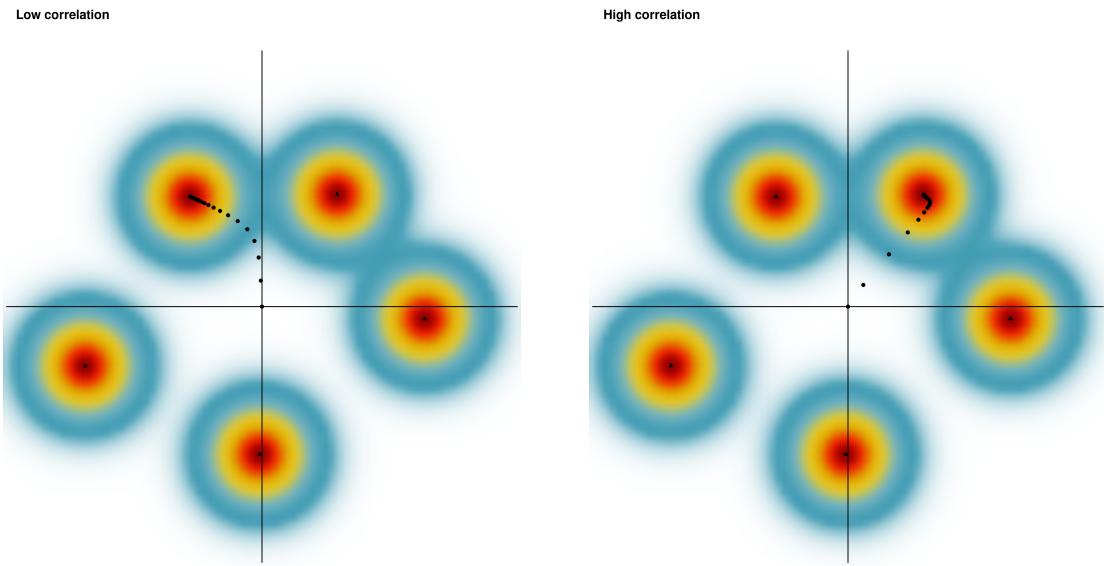


Figure 7: Different G-matrices can alter the final position in the landscape.

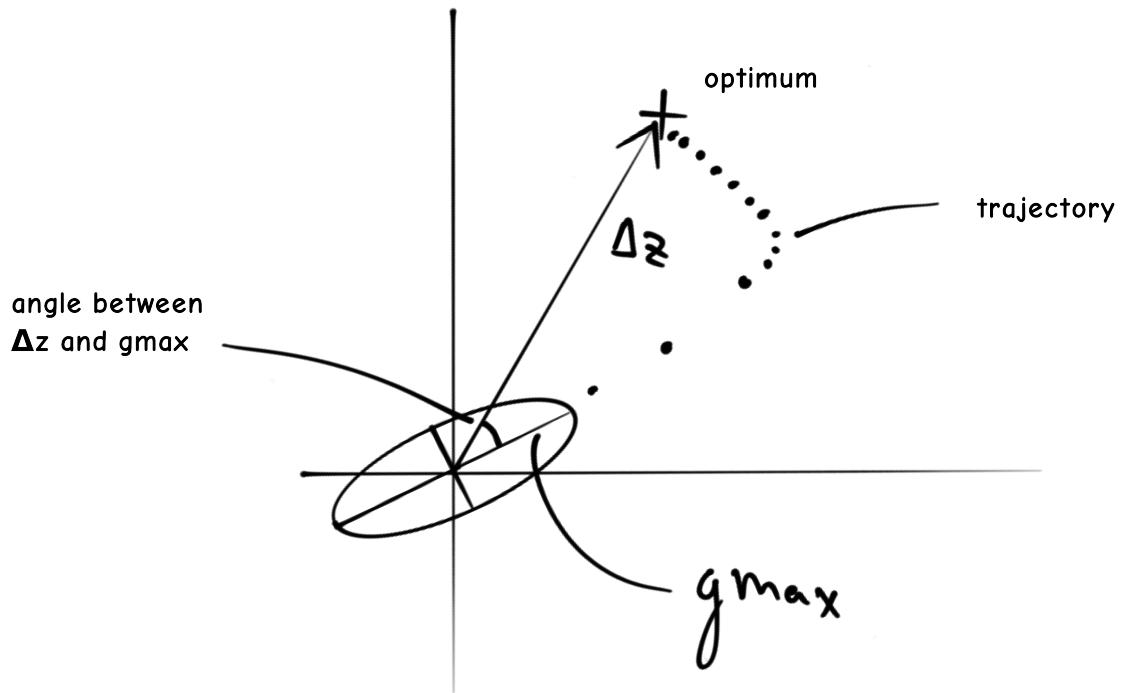
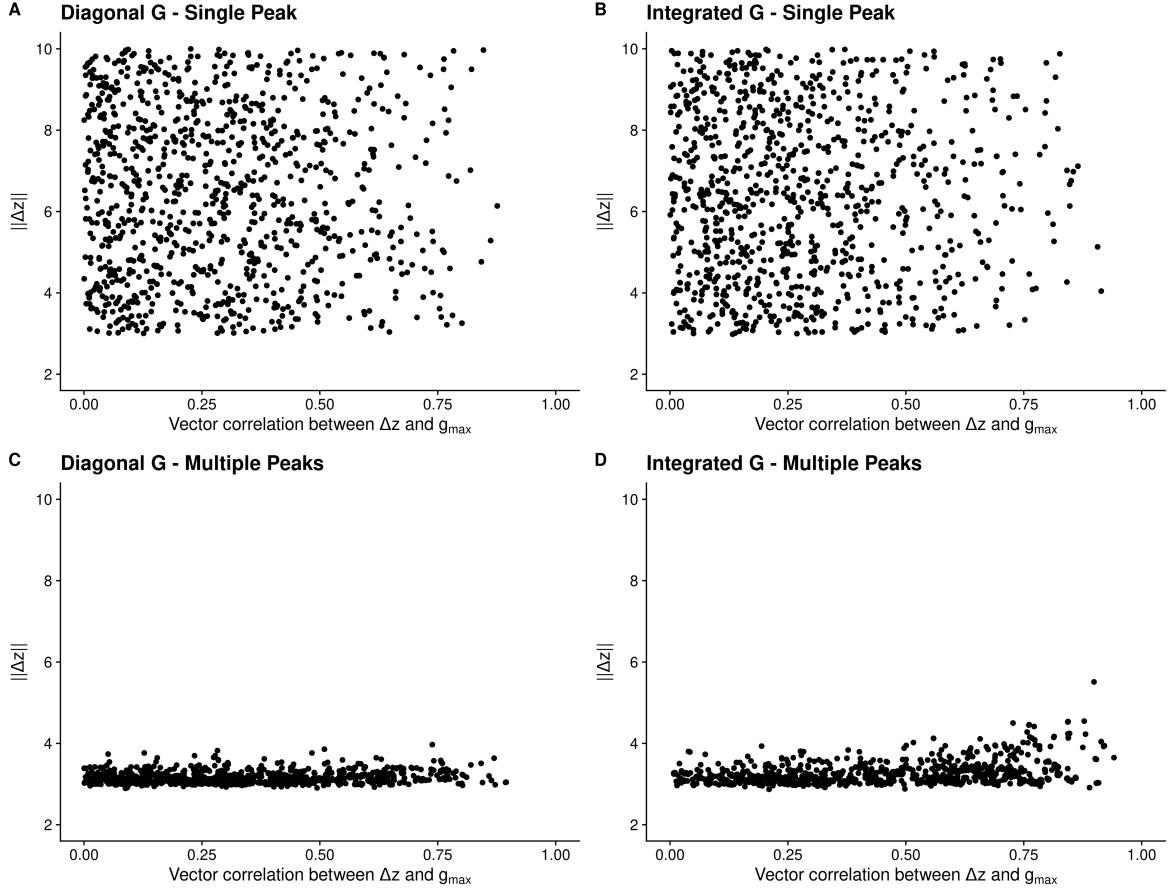


Figure 8: Scheme of the measurements in a simulation run. The total length of the phenotypic change is given by the norm of the  $\Delta z$  vector. The angle between the direction of most genetic variation and the  $\Delta z$  vector measure the alignment between the phenotypic change and genetic variation.



Simulations using random peaks. We see a clear lack of  $\Delta z$  with high correlation to  $g_{\max}$ . In this simulation, dimensionality of the phenotype is  $k = 8$ , and in the multi-peak landscape we have  $N = 50$  peaks. Simulations are repeated 1000 times.

### Random peaks

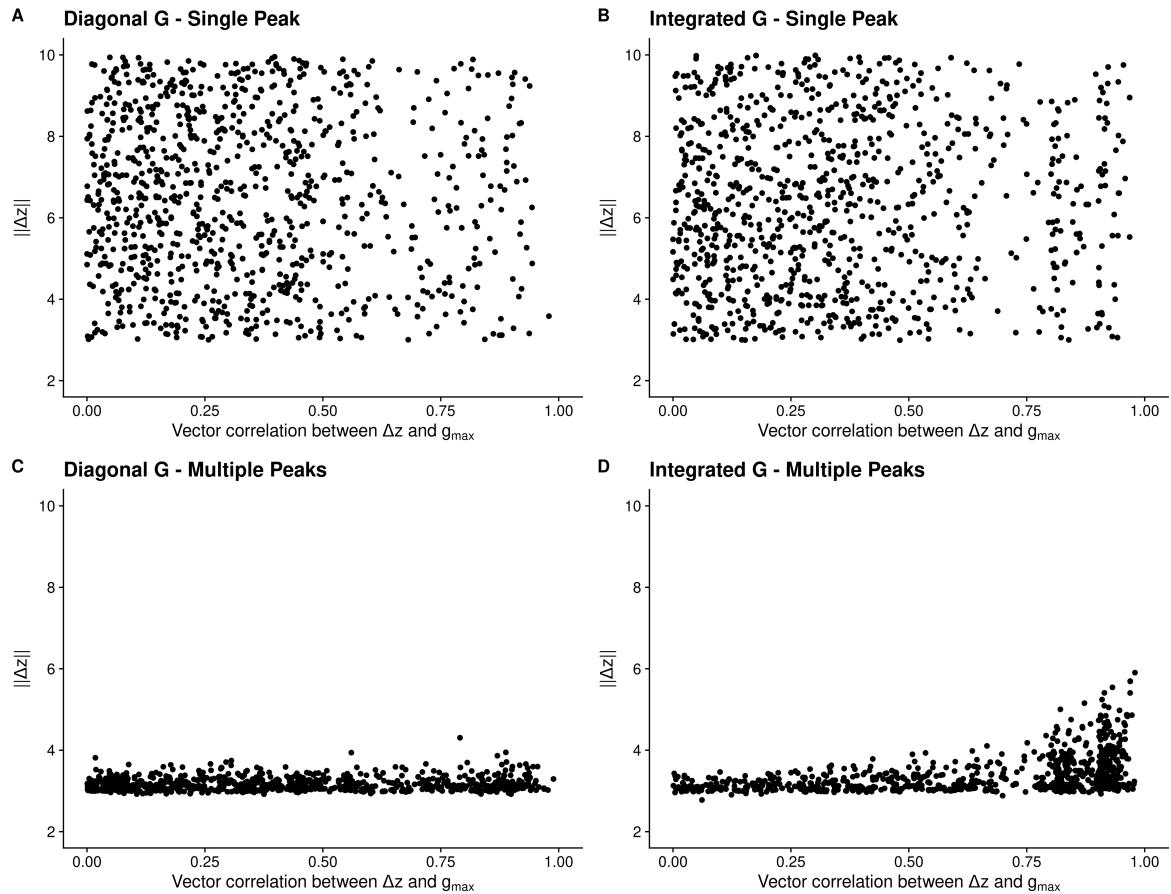
We start with a random placement of possible peaks, like in figure 4. This leads to the following relations between total amount of phenotypic change and the correlation between phenotypic change and genetic variation:

We see a clear difference between the single-peak and multi-peak surfaces, but not much difference between diagonal and integrated matrices. There is also a clear lack of  $\Delta z$  aligned with  $g_{\max}$ .

### Enriched peaks

We can use our peak generating process to increase the amount of peaks aligned with  $g_{\max}$ . This is necessary in high dimensions, as the probability that there are any peaks aligned with  $g_{\max}$  at random goes to zero in high dimensions. A slight enrichment of the peaks aligned with  $g_{\max}$  gives the following pattern:

Under these conditions, we see that the magnitude of  $\Delta z$  increases as a function of the alignment with  $g_{\max}$ . This pattern is similar to the one observed in natural populations.



Simulations using random peaks. We see a clear lack of  $\Delta z$  with high correlation to  $g_{\max}$ . In this simulation, dimensionality of the phenotype is  $k = 8$ , and in the multi peak landscape we have  $N = 50$  peaks. Simulations are repeated 1000 times.