

Circulant matrices and eigenvalue-based p -value adjustment for multiple testing in genetics.

Chris Salahub

University of Waterloo, csalahub@uwaterloo.ca

July 28, 2022

1 Introduction

Suppose each of a collection of M independent variables X_1, \dots, X_M is tested against a common dependent variable Y with a pairwise measure $g(X_i, Y)$ in order to screen X_1, \dots, X_M for association with Y . Though this practice is not ideal for determining the complete dependency, as a subset of the X with individually low g values may collectively be highly associated with Y , when M is particularly large it is common practice.

The combinatorial explosion of all possible subsets of M variables presents a serious computational barrier to such complete modelling, while it is conceptually difficult to think of function which would provide a universal comparison between these subsets. Even if these computational and conceptual burdens are overcome, some repeated measurement of g over thousands of combinations occurs. This introduces an unavoidable multiple testing problem. Values of g which would be significant if observed alone are less meaningful when observed in a collection of thousands of values.

Restricting ourselves to the binary case where we have $\mathbf{g} = \{g(X_i, Y)\}_{i=1, \dots, M}$, a number of adjustments for the problem of M repeated tests exist surveyed in literature such as [3, 2]. The simplest of these is the Bonferroni correction. If the individual significance level is chosen to be α , this correction applies the adjustment

$$\frac{\alpha}{M}$$

and uses this significance level to assess each test. This method is known to be overly conservative and does not take any of the structure of X_1, \dots, X_M into account.

A collection of approaches outlined in [2] attempts to use the correlation structure of X_1, \dots, X_M to relax the Bonferroni correction. The central idea to this relaxation is that strongly dependent tests do not all deserve a weight of one when counting the number of tests for adjustment. Rather, these tests should be combined into an effective count less than the count of variables. Inspired by the use of principal components analysis (PCA) to reduce dimension by selecting the top eigendirections, these methods take

$$M_{\text{eff}} = f(\lambda_1, \lambda_2, \dots, \lambda_M) = f(\boldsymbol{\lambda})$$

where $\boldsymbol{\lambda}$ are the M eigenvalues of $\boldsymbol{\Sigma}$, the correlation matrix of the X_i . This collection includes [1, 12, 10, 4].

Besides inspiration, these test adjustments share a common motivation: the identification of genetic factors in traits and disease through genome-wide association studies (GWAS). [15] provides a good survey of GWAS practice. The typical GWAS will measure M positions on the genome and compare each of these to a single measurable trait Y . These M measured positions will have highly structured correlations between them due to the organization of genetic information in DNA. GWAS are also frequently mentioned in the literature on measures of association, such as [14, 11]. This shared motivation therefore deserves special attention. Indeed, much can be shown about the structure of genetic measurements as collected in a GWAS.

2 Multiple testing

2.1 Notation

Suppose $\tilde{\boldsymbol{\mu}} \sim N_M(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = [\rho_{ij}]$ is a correlation matrix so that $\rho_{ii} = 1$, $\rho_{ij} = \rho_{ji}$, and $|\rho_{ij}| \leq 1$ for all $i = 1, \dots, M$ and $j = 1, \dots, M$.

Let $\boldsymbol{\Sigma} = \mathbf{V}\mathbf{D}_{\lambda}\mathbf{V}^T$ be its eigen decomposition with $\mathbf{D}_{\lambda} = \text{diagonal}(\lambda_1, \lambda_2, \dots, \lambda_M)$ the diagonal matrix of eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \lambda_M \geq 0$ and $M \times M$ orthogonal matrix $\mathbf{V} = [\mathbf{V}_1 | \mathbf{V}_2 | \dots | \mathbf{V}_M]$ having as columns the $M \times 1$ eigenvectors \mathbf{V}_i corresponding to each eigenvalue λ_i for $i = 1, \dots, n$.

Note that for the special case of $\rho_{ij} = \rho > 0$ for all $i \neq j$, $\boldsymbol{\Sigma}$ can be written as

$$\boldsymbol{\Sigma} = \rho \mathbf{1}\mathbf{1}^T + (1 - \rho)\mathbf{I}_M \tag{1}$$

where $\mathbf{1} = (1, 1, \dots, 1)^\top$ is the $M \times 1$ vector of ones and \mathbf{I}_M the $M \times M$ identity matrix. In this case, $\lambda_1 = (M - 1)\rho + 1$ and $\lambda_2 = \lambda_3 = \dots = \lambda_M = (1 - \rho)$. The corresponding eigenvectors are $\mathbf{V}_1 = \frac{1}{\sqrt{M}}\mathbf{1}$ and any linearly independent set of $(M - 1)$ orthonormal vectors $\mathbf{V}_2, \dots, \mathbf{V}_M$ all orthogonal to \mathbf{V}_1 . Note that if $\rho < 0$, the eigen decomposition is the same but $\lambda = (M - 1)\rho + 1$ would be the *smallest* eigenvalue and Σ would be a non-negative definite matrix only when $\rho \geq -1/(M - 1)$.

Of interest is testing the hypothesis $H_0 : \boldsymbol{\mu} = \mathbf{0}$ having observed some realization $\hat{\boldsymbol{\mu}}$ of $\tilde{\boldsymbol{\mu}}$. Equivalently, the M hypotheses $H_{0i} : \mu_i = 0$ might be tested independently and simultaneously based on the marginal distributions $\tilde{\mu}_i \sim N(\mu_i, 1)$ and observed $\hat{\mu}_i$ for $i = 1, \dots, M$. The challenge is combining these as a test for H_0 .

Testing any of these hypotheses for a fixed significance level α is equivalent to forming confidence regions of size $1 - \alpha$ and determining whether the region contains 0 for a univariate test of H_{0i} or the $M \times 1$ vector $\mathbf{0}$ for the multivariate test of H_0 . A fixed level test would reject the hypothesis whenever the corresponding region did not contain the zero value and would accept otherwise.

2.2 The Multivariate Test

For the multivariate test the confidence region would be the set of all $\boldsymbol{\mu}$ which satisfy

$$(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^\top \Sigma^{-1} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \leq C_\alpha \quad (2)$$

where C_α is the critical value found from a χ^2 distribution on M degrees of freedom as

$$Pr(\chi_M^2 \leq C_\alpha) = 1 - \alpha.$$

For example, when $\alpha = 0.05$ and $M = 2$, $C_\alpha \approx 5.99$ and Equation (2) determines the 95% confidence region for $\boldsymbol{\mu}$.

The shape of this confidence region is an M -dimensional ellipsoid centred at $\hat{\boldsymbol{\mu}}$. Points outside the ellipse would be rejected by a level $\alpha = 0.05$ test while those inside would be accepted. Testing $H_0 : \boldsymbol{\mu} = \mathbf{0}$ is equivalent to checking whether the origin $\mathbf{0}$ is outside the ellipse or not.

2.3 Multiple univariate tests

For each individual univariate hypothesis H_{0i} , the confidence region is the set of all values μ_i satisfying

$$(\mu_i - \hat{\mu}_i)^2 \leq c_\alpha \quad (3)$$

where c_α is the critical value from a χ^2 distribution on 1 degree of freedom satisfying

$$Pr(\chi_1^2 \leq c_\alpha) = 1 - \alpha.$$

For example, when $\alpha = 0.05$, $c_\alpha \approx 3.84146$ and Equation (3) determines the 95% confidence region for μ_i . This is simply the standard confidence interval for μ_i with known variance $\sigma_i^2 = 1$

$$\mu_i \in [\hat{\mu}_i - \sqrt{c_\alpha}, \hat{\mu}_i + \sqrt{c_\alpha}]. \quad (4)$$

In the case of $\alpha = 0.05$, $\sqrt{c_\alpha} \approx 1.96$.

Using only the tests of the univariate hypotheses, H_{0i} , it is logically sufficient to reject the multivariate hypothesis H_0 if one or more of the univariate hypotheses H_{0i} is rejected. Indeed, the probability of rejecting the multivariate hypothesis H_0 based on the rejection of one or more of the univariate hypotheses H_{0i} is

$$\begin{aligned} Pr(\text{rejecting } H_0 \mid H_0 \text{ is true}) \\ &= Pr(\text{At least one of } H_{01}, H_{02}, \dots, H_{0M} \text{ is rejected} \mid H_0 \text{ is true}) \\ &= 1 - Pr(\text{None of } H_{01}, H_{02}, \dots, H_{0M} \text{ are rejected} \mid H_0 \text{ is true}) \end{aligned}$$

If the tests are independent, that is $\rho_{ij} = 0$ for all i, j , then

$$\begin{aligned} Pr(\text{rejecting } H_0 \mid H_0 \text{ is true}) &= 1 - \prod_{i=1}^M Pr(H_{0i} \text{ is not rejected} \mid H_0 \text{ is true}) \\ &= 1 - \prod_{i=1}^M (1 - \alpha_i) \end{aligned}$$

where the significance level α_i is used to accept or reject H_{0i} . Supposing that the same significance level, $\alpha_i = \alpha_{\text{individ}}$, is used for each individual hypothesis; the experiment-wide significance level is

$$\alpha = Pr(\text{rejecting } H_0 \mid H_0 \text{ is true}) = 1 - (1 - \alpha_{\text{individ}})^M. \quad (5)$$

This introduces an inflated type one error when using multiple independent tests. In the case where $M = 2$, $\rho = 0$, and $\alpha_{\text{individ}} = 0.05$, the joint hypothesis H_0 will be rejected 9.75% of the time when H_0 is true, rather than the putative 5% of the time. These individual tests produce more false positives than the 5% planned. Alternatively, as a confidence region for $\boldsymbol{\mu}$, the box defined by these marginal tests has an associated confidence level of $(1 - \alpha_{\text{individ}})^M$ rather than $1 - \alpha$.

2.4 Accounting for multiple tests

One way to still use the confidence intervals, or equivalently the independent tests, is to adjust the significance levels for each individual test to get the desired significance level for their combination. Turning Equation (5) around and solving for α_{individ} gives

$$\alpha_{\text{individ}} = 1 - (1 - \alpha)^{\frac{1}{M}}. \quad (6)$$

This indicates that for an experiment-wide significance level of α , each of the M individual significance levels should be adjusted to be the value α_{individ} from Equation (6) and the critical value $c_{\alpha_{\text{individ}}}$ determined accordingly.

Note that these adjusted multiple tests and the multivariate test are matched only on their fixed significance level α or coverage probability $1 - \alpha$. The multiple tests only use marginal distributions whereas the multivariate test uses the joint distribution. Consequently, they may still disagree on which hypotheses to accept and which to reject, though this will occur less often than with the unadjusted test.

Sometimes, an even simpler adjustment commonly called a *Bonferroni correction* is made. Instead of using Equation (6), simply choose

$$\alpha_{\text{individ}} = \frac{\alpha}{M}. \quad (7)$$

For small α and large M , this is very nearly the same value as in Equation (6) since $(1 - \alpha)^{\frac{1}{M}} \approx 1 - \frac{\alpha}{M}$. It is easier to remember in any case.

2.5 Dependent multiple tests

The problem is that the marginal tests are **not** independent whenever $\Sigma \neq \mathbf{I}_M$, and so Equation (6) does not strictly apply. Figure 1 compares the corrected and uncorrected intervals for $M = 2$ and different values of ρ , the correlation between the marginal tests. In each of Figures 1(a), (b), (c), and (d), the coverage probability of the ellipse is 0.95, the correct value for $\alpha = 0.05$. Table 1 gives the coverage probabilities (and experiment wide α levels in parentheses) for the boxes defined by the multiple test intervals of Figures 1(a), (b), and (c). These are correct in only two cases. The case $\rho = 0$ is that for which the correction was designed and α_{individ} was set correctly. When $\rho = 1$, all probability mass lies only along the box diagonal and all marginal tests will reject whenever one does. More generally, as ρ increases, the ellipses narrow and the probability concentrates nearer the

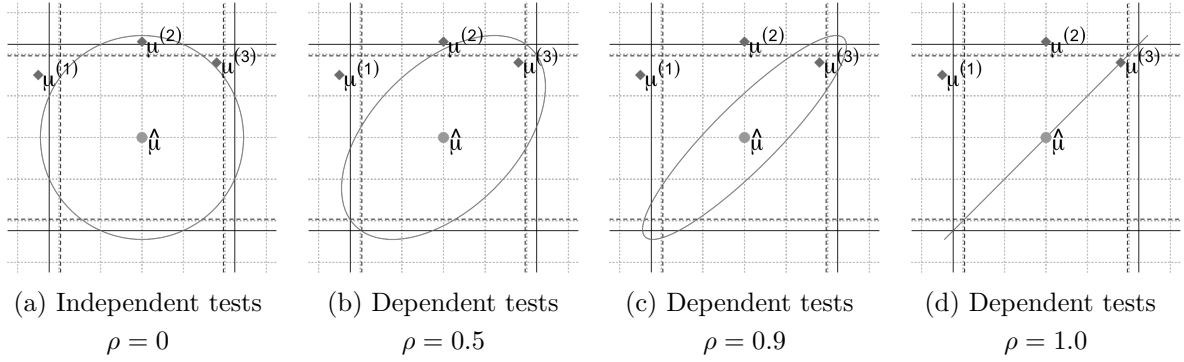


Figure 1: Combining multiple confidence intervals versus a single 95% confidence region for varying ρ . Corrected intervals are solid lines ($\alpha_{\text{individ}} = 0.025$); uncorrected are dashed ($\alpha_{\text{individ}} = 0.05$). Grid shows unit squares.

	$\rho = 0$	$\rho = 0.5$	$\rho = 0.9$	$\rho = 1.0$
Uncorrected	0.902 (0.098)	0.909 (0.091)	0.930 (0.070)	0.950 (0.050)
Corrected	0.950 (0.050)	0.953 (0.047)	0.963 (0.037)	0.975 (0.025)

Table 1: Coverage probabilities and experiment-wide α levels for boxes defined by multiple marginal test intervals.

box diagonal. As a consequence, the probability mass covered by either box increases. The correction, designed for independent tests, is no longer correct when tests are dependent.

When the tests are not independent, it has been suggested that the adjustment given by Equation (6), or by Equation (7), still be used but that the number of independent tests M be replaced by some measure of the *effective number* of independent tests. That is, the adjustments become

$$\alpha_{\text{individ}} = 1 - (1 - \alpha)^{\frac{1}{M_{\text{eff}}}} \quad (8)$$

or simply

$$\alpha_{\text{individ}} = \frac{\alpha}{M_{\text{eff}}} \quad (9)$$

for the Bonferroni correction. M_{eff} is now some measure of the effective number of independent tests, and so need not necessarily be an integer.

2.6 Variance of the Eigenvalues

The first suggested adjustment was that given by [1] based on the sample variance V_λ of the eigenvalues of the correlation matrix Σ , namely

$$V_\lambda = \frac{\sum_{i=1}^M (\lambda_i - 1)^2}{M - 1}.$$

This variance has been interpreted by [1] as a measure of the “total correlation” among the set of “traits” or tests defining Σ . When all are independent then $\Sigma = \mathbf{I}_M$ and $V_\lambda = 0$, its minimum value; when all are maximally correlated, $\rho_{ij} = 1$ for all $i \neq j$ and $V_\lambda = M$, its maximum value. The ratio (V_λ/M) varies from its minimum 0 when there are M independent tests to its maximum 1 when there is essentially 1 test since they are maximally correlated. A simple location scale change of the ratio (V_λ/M) from $[0, 1]$ to $[M, 1]$ suggests using

$$M_{\text{eff}} = M - (M - 1) \frac{V_\lambda}{M} \quad (10)$$

as the effective number of tests. This is the expression used in [12].

If M_{eff} is used in place of M in the correction formula as in Equation (8), the following table results for the cases illustrated in Figure 1. As can be seen, the correction, the coverage, and the α are correct for the two end cases when $\rho = 0$ and $\rho = 1$. In between, the effective number of cases is a real number between one and two, increasing in size as

	$\rho = 0$	$\rho = 0.5$	$\rho = 0.9$	$\rho = 1.0$
M_{eff}	2	1.75	1.19	1
Coverage	0.950	0.947	0.940	0.950
α	0.050	0.053	0.060	0.050

Table 2: Coverage probabilities and experiment-wide α levels for boxes adjusted by M_{eff} from [1].

the correlation moves toward zero. In these between cases, the coverage probability (or α) is fairly close to its target value of 0.95 (or 0.05).

In terms of the eigenvalues Equation (10) can be rewritten as

$$M_{eff} = M + 1 - \frac{\sum_{i=1}^M \lambda_i^2}{M}. \quad (11)$$

In the special case where $\rho_{ij} = \rho$ for all $i \neq j$, then $V_\lambda = M\rho^2$ and

$$M_{eff} = M - (M - 1)\rho^2. \quad (12)$$

Two edge cases are often considered: the independent case when $\rho = 0$ and the maximally correlated case when $\rho = 1$. In the first of these $M_{eff} = M$ as it should. When $\rho = 1$, it is as if the same test has been applied M times and $M_{eff} = 1$ correctly reflects that only a single test has been applied.

Similarly, we could imagine m copies of k independent tests. This would amount to a block diagonal $(mk) \times (mk)$ matrix Σ having k identical $m \times m$ matrices of ones down the diagonal and zeros elsewhere¹. The resulting eigenvalues are $\lambda_1 = \dots = \lambda_k = m$ and $\lambda_{k+1} = \dots = \lambda_M = 0$, where $M = km$. The number of independent tests here is clearly k . Unfortunately, in this case Equation (11) gives

$$M_{eff} = km + 1 - \frac{km^2}{km} = (k - 1)m + 1 \quad (13)$$

which equals k only when $m = 1$, and is larger otherwise.

When there are differing numbers $m_1 > m_2 > \dots > m_k \geq 1$ of the k tests, then again Equation (11) leads to

$$M_{eff} = M + 1 - \frac{\sum_{i=1}^k m_i^2}{M}$$

¹A similar block diagonal structure is natural in genetic studies.

where $M = \sum_{i=1}^k m_i$. Again this will be very different from k .

Considerations of cases such as these have led [10] to an alternative proposal for M_{eff} which will return k in both cases.

2.7 Separating Integral and Fractional Contributions

Noting the counter-intuitive behaviour of M_{eff} revealed in Equation (13) when independent tests are copied multiple times each, [10] craft a new proposal based entirely on the eigenvalues. They suggest that each eigenvalue λ_i be separated into two parts, its integral part (≥ 1) and its fractional remainder. Its integral part counts as a single independent test and its fractional part as a “partially correlated test” to be counted as some fraction between 0 and 1.

Mathematically, this amounts to a function $g(x)$ defined as

$$g(x) = I_{[1,\infty)}(\lambda_i) + f_i \quad (14)$$

where $I_A(x)$ is the usual indicator function and $f_i = \lambda_i - \lfloor \lambda_i \rfloor$ is the fractional remainder after removing the floor function $\lfloor \cdot \rfloor$. Note that [10] imagine cases where Σ possibly has negative eigenvalues². In that case, they replace λ_i by $|\lambda_i|$ everywhere. A better solution might instead be to replace Σ by its nearest non-negative definite matrix approximation before continuing. **TODO: Justify this statement, seems like the crux of the justification is that the hypersphere of the nearest non-negative definite matrix would better approximate the true coverage region**

Following this intuition, [10] propose M_{eff} be replaced by

$$M_{eff}^* = \sum_{i=1}^M g(\lambda_i) \quad (15)$$

in place of the original by [1] shown in Equations (10) and (11).

Like the definition of Equation (10), this returns $M_{eff}^* = 1$ and $M_{eff}^* = M$ for the edge cases of M maximally correlated tests and M independent tests, respectively. It also addresses the problems raised for the cases where k independent tests are repeated m times each. $M_{eff}^* = k$ no matter how many times each of the k independent tests were repeated.

Applying M_{eff}^* to the examples of Figure 1 gives the results shown in Table 3. Careful

²This situation will arise if the elements of Σ have been separately determined without constraint that Σ satisfy the definition of a correlation matrix and hence be non-negative definite. In many genetic surveys, this is the reality of the correlations measured.

	$\rho = 0$	$\rho = 0.5$	$\rho = 0.9$	$\rho = 1.0$
M_{eff}^*	2	2	2	1
Coverage	0.950	0.953	0.963	0.950
α	0.050	0.047	0.037	0.050

Table 3: Coverage probabilities and experiment-wide α levels for boxes adjusted by M_{eff}^* from [10].

examination of these results reveals a new peculiarity with this definition. In the case when $M = 2$, $M_{eff}^* = 2$ for any $\rho < 1$ and the resulting intervals are necessarily uncorrected!

The problem is that the equal correlation structure of Equation (44) gives $\lambda_1 = 1 + (M - 1)\rho$ and $\lambda_2 = \dots = \lambda_M = 1 - \rho$. For $\rho \geq 0$, $\lambda_1 \geq 1$ and its fractional remainder is $f_1 = (M - 1)\rho - \lfloor (M - 1)\rho \rfloor$. The remaining $\lambda_i = f_i = 1 - \rho < 1$ for all $i \neq 1$.

For $\rho \geq 0$, then,

$$\begin{aligned}
M_{eff}^* &= 1 + f_1 + (M - 1)(1 - \rho) \\
&= M - \lfloor (M - 1)\rho \rfloor.
\end{aligned} \tag{16}$$

When $M = 2$, $M_{eff}^* = 2$ no matter what value $\rho < 1$ takes and so no correction is made. The values for the two edge cases are also easily derived from this equation: when $\rho = 0$, $M_{eff}^* = M$; when $\rho = 1$, then $M_{eff}^* = 1$.

Note also from Equation (16), that M_{eff}^* takes only integer values and is a step function of ρ . The value of $M_{eff}^* = M - k$ is constant for all $\rho \in [\frac{k}{M-1}, \frac{k+1}{M-1})$ with each non-negative integer $k \leq M - 2$. The smaller is M , the wider is the range of ρ returning an identical value for M_{eff}^* . For example, when $M = 3$, all values of $\rho \in [0, 0.5)$ return $M_{eff}^* = 3$.

In the unusual case where $-1/(M - 1) \leq \rho < 0$ (i.e. Σ remains a correlation matrix), the largest eigenvalue will be $1 - \rho = 1 + |\rho|$ in this instance, and there will be $M - 1$ copies of this. The single smaller eigenvalue will be $1 - (M - 1)|\rho|$ which is a positive fraction f_M since $|\rho| \leq 1/(M - 1)$. Together, in the case where $-1/(M - 1) \leq \rho < 0$, these results give

$$\begin{aligned}
M_{eff}^* &= (M - 1) \times 1 + (M - 1)|\rho| + f_M \\
&= (M - 1)(1 + |\rho|) + (1 - (M - 1)|\rho|) \\
&= M
\end{aligned}$$

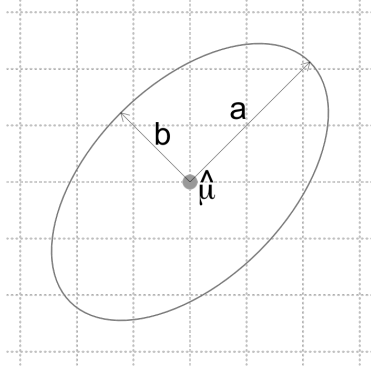
independent of whatever value ρ takes, provided it is negative. Whatever value $\rho < 0$ takes, provided it is negative, no correction will be made using M_{eff}^* .

In contrast, under the same conditions, M_{eff} of [1] is symmetric in ρ about 0 according to Equation (12) provided $|\rho| \leq 1/(M - 1)$. One could argue that this is the sensible choice. Small values of $|\rho|$ indicate M nearly, but not quite, independent tests. It seems unusual to treat the multiple marginal tests as if they are dependent when $\rho > 0$ but as if they are independent when $\rho < 0$, yet this is exactly what M_{eff}^* does.

2.8 Effective Dimension

A final approach is motivated by the shape of the confidence ellipsoids as determined by the relative lengths of their principal axes. For example, when $M = 2$, various confidence ellipses are shown in Figure 1 to demonstrate the change probability mass and so how the box formed by the separate multiple tests must change to match it.

When $M = 2$, a rotated ellipse appears as in Figure 2. As can be seen, the lengths of



Once rotated, the equation of this ellipse in two dimensions is:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

Setting $x = 0$, solve for the limits for y to give $\pm\sqrt{a}$.

Similarly, set $y = 0$ to give limits of x as $\pm\sqrt{b}$

Figure 2: The major and minor semi-axes of the 95% confidence ellipse centred at $\hat{\mu}$ when $M = 2$, $\rho = 0.5$, and $\alpha = 0.05$. Lengths are marked as a and b . Grid shows unit squares.

the principal semi-axes are easily determined, giving $a = \sqrt{C_\alpha \lambda_1}$ and $b = \sqrt{C_\alpha \lambda_2}$. Note that C_α changes the length depending on the size of α and so does not depend on the underlying correlation structure. Whatever the choice of α the major and minor axes will have lengths proportional to $\sqrt{\lambda_1}$ and $\sqrt{\lambda_2}$, respectively. That is, the shape of the ellipse is determined entirely from the eigenvalues of the correlation matrix.

In some sense, each eigenvalue gives the support for the existence of a dimension along that axis. When one of them is zero, the axis and dimension associated with that value disappears as in Figure 1(c).

For M -dimensional data, there will be M such semi-axes of Euclidean length $\sqrt{\lambda_i}$ on the i the axis. A formal definition given by [13] is that the *effective dimensionality* could be measured by the function

$$d(\lambda_1, \dots, \lambda_M; p) = \sum_{i=1}^M \left(\frac{\lambda_i}{\lambda_{max}} \right)^p = \sum_{i=1}^M \left(\frac{\lambda_i}{\lambda_1} \right)^p$$

for any choice of $p \in [0, \infty)$, where different choices of p may be fit for different purposes. The smaller each $\lambda_i \geq 0$ is compared to the $\lambda_{max} = \lambda_1 > 0$, the less relative support there is for that dimension.

The parameter p determines the rate at which the contribution of the shorter directions to the effective dimension are attenuated. When $p = 0$, d_p simply counts the number of non-zero λ_i s and so returns the rank of Σ as the dimension. As p increases, the contribution of the shorter minor axes diminishes, eventually reaching zero as $p \rightarrow \infty$ at which point d_p simply counts the number of equivalent major axes of the ellipsoid. The minimal effective dimension is 1 and the maximum is M for all $p \in [0, \infty)$.

For the multiple testing problem, the attenuation must not be so sharp that it throws away independent tests. One might choose $p = (1/M)$ so that the effective number of tests is estimated by the effective dimension using the M th root of the eigenvalues. That is,

$$M_{eff}^{**} = \sum_{i=1}^M \left(\frac{\lambda_i}{\lambda_{max}} \right)^{1/M} \quad (17)$$

This choice is motivated, in part, by considering the problem of simply repeatedly adding same test over and over to a bank of tests.

For example, suppose Σ is a block diagonal matrix

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \mathbf{0} \\ \mathbf{0}^\top & \Sigma_{22} \end{bmatrix} \quad (18)$$

where the $m_1 \times m_1$ matrix Σ_{11} and the $m_2 \times m_2$ matrix Σ_{22} are both correlation matrices with $M = m_1 + m_2$. These two matrices represent separate independent groups of tests since all correlations between the two groups are zero as shown by the $m_1 \times m_2$ off-diagonal zero-matrix, denoted by $\mathbf{0}$.

Suppose each of Σ_{11} and Σ_{22} have the common correlation structure of Equation (44) but with different correlations $\rho_{11} \geq 0$ and $\rho_{22} \geq 0$. The eigenvalues of Σ are then $(m_1 - 1)\rho_{11} + 1$, $(m_2 - 1)\rho_{22} + 1$, $(1 - \rho_{11})$ repeated $(m_1 - 1)$ times, and $(1 - \rho_{22})$ repeated $(m_2 - 1)$ times. Which is the largest eigenvalue will depend on the values of the pairs m_1 , ρ_{11} , and m_2 , ρ_{22} ; when $\rho_{11} = \rho_{22}$, for example, the largest eigenvalue belongs to the largest group.

To see the effect of increasing M , consider the case when $\rho_{22} = 0$ but $\rho_{11} \neq 0$. The m_2 tests in the second group are independent with unit eigenvalues and the m_1 tests in the first group are dependent with eigenvalues as above. The effective dimension for any p becomes

$$d(\lambda_1, \dots, \lambda_M; p) = 1 + \frac{m_2}{(1 + (m_1 - 1)\rho_{11})^p} + (m_1 - 1) \left(\frac{1 - \rho_{11}}{1 + (m_1 - 1)\rho_{11}} \right)^p. \quad (19)$$

When either $m_1 = 1$ or $\rho_{11} = 0$ also holds, then the effective dimension is $M = m_1 + m_2$, as it should be for M independent tests.

However, when this is not the case, the contribution of the m_2 can be made arbitrarily small for fixed p by simply increasing m_1 . The contribution from the second group of independent tests lies entirely with the second term of Equation (19) while that of the first group is spread over the second and third terms. Increasing m_1 will not only increase the contribution of the first group in the third term but it will also *decrease* the second group's contribution!

This effect is perhaps most dramatic, as well as most worrying, when ρ_{11} is maximal, that is $\rho_{11} = 1$. Then

$$d(\lambda_1, \dots, \lambda_M; p) = 1 + m_2 \left(\frac{1}{m_1} \right)^p$$

and simply increasing m_1 can entirely remove the effect of the second group of independent tests. When $\rho_{11} = 1$ is maximal, then the m_1 tests in the first group are effectively the same test. By simply repeating the same test m_1 times, achieving the same answer each time, the contribution of the m_2 independent tests can be effectively obliterated!

Looking again at the second term in Equation (19), the ideal solution in this case would have the denominator be equal to one. As a function of $m = m_1$ and p this denominator behaves roughly like

$$(1 + m)^p.$$

For fixed m , the smaller is p , the closer this is to one. For fixed p , the larger is m , the farther this moves from one, getting larger and larger. The obvious solution is to choose $p = 1/m$, for then one corrects the other and $\lim_{m \rightarrow \infty} (1 + m)^{1/m} = 1$.

This suggests the choice $p = 1/M$ in general; all limits still hold since $M = m_1 + m_2$ (including the original denominator going to 1). Moreover, this choice amplifies small ratios of $\frac{\lambda_i}{\lambda_{max}}$ more as M increases. This choice leads to

$$M_{eff}^{**} = \sum_{i=1}^M \left(\frac{\lambda_i}{\lambda_{max}} \right)^{1/M}$$

as suggested earlier.

Each λ_i is proportional to the Euclidean length of a principal axis of the confidence ellipsoid determined by Σ with proportionality constant determined by the choice of α . Ratios of axis lengths remove the proportionality constant allowing the effective number of tests to be independent of the selected significance level α .

As with the other choices of the effective number of tests M_{eff}^{**} , we can examine the coverage probabilities for M_{eff}^{**} defined as the effective dimensionality of Equation (17) for the cases of Figure 1. The results are given in the table below.

	$\rho = 0$	$\rho = 0.5$	$\rho = 0.9$	$\rho = 1.0$
M_{eff}^{**}	2	1.57735	1.229416	1
Coverage	0.950	0.941	0.942	0.950
α	0.050	0.059	0.058	0.050

Similar to the other methods, it agrees in the effective number of tests and gets the coverage probabilities exactly right for the extreme cases of $\rho = 0$ and $\rho = 1$. The M th root effective dimension is similar to the original proposal by [1] in coverage probabilities and in effective number of tests being non-integral for these examples. It differs from the indicator function and fractional part method of [10] which, for any values of $\rho \in (0, 1)$, has $M_{eff}^{**} = M = 2$ and so simply reproduces the original independent tests adjustment in terms of effective number of tests and coverage probability (or experiment-wide significance level).

Although not introduced in terms of effective dimension, [4] proposes a similar measure in

$$M_{eff}^{***} = \frac{\left(\sum_{i=1}^M \sqrt{\lambda_i} \right)^2}{\sum_{i=1}^M \lambda_i} \quad (20)$$

which is simply the ratio

$$\frac{d(\lambda_1, \dots, \lambda_M; 1/2)^2}{d(\lambda_1, \dots, \lambda_M; 1)}.$$

2.9 ‘Typical’ Values of M

While the above methods for adjusting individual tests could be applied to any data, all are motivated by the analysis of genetic data. It is therefore worthwhile to consider the typical values of M which are encountered in genetic investigations. Historically, these have been smaller, but technological and computational developments have increased the possible values of M considerably. While [9], [1], [12], and [10] perform analyses of only several dozen tests, with M almost always less than one hundred, modern advances have facilitated studies of hundreds of thousands to millions of potential tests [8]. This trend can be seen in [4], which presents results for the pairwise testing of 2,000 variables in a genetic data set. So, while the strange behaviour of Equation (15) for $M = 2$ is certainly not desirable, it may not be relevant in practice.

3 Theoretical Eigenvalues

Stepping away from the particular structure of Equation (44) as a description of repeated tests used to motivate the development of the above methods, consider the actual correlation structure of genetic data. For a first principles derivation of this structure and a complete discussion of the model it is based on, refer to Appendix ??.

Simply, genetic studies look at a selection of locations on the genome. For the i^{th} location let h_i be the chromosome of the location and z_i be the observed summarized genetic information: a realization of the random variable Z_i . Between locations i and j , a final relevant feature is $d(i, j)$, the *genetic distance* motivated by the joint behaviour of the locations in inheritance.

The summary $\mathbf{Z} = (Z_1, \dots, Z_M)^\top$ used in [1, 12, 10] and [4] is the *additive summary*. This amounts to the count of a particular variant over the two copies of the genome in an individual, and so $Z_i \in \{0, 1, 2\}$. Under the model from Appendix ?? the correlation between Z_i and Z_j under an additive summary is given by

$$\text{Corr}(Z_i, Z_j) = I_{\{h_i\}}(h_j) \gamma e^{-2\beta d(i, j)}, \quad (21)$$

where β is a constant determined by the particular genetic distance chosen and γ is a constant determined by the distribution of variants in the population. Commonly β is set to $1/100^3$ and γ is 1.

³In which case $d(i, j)$ is called the *Haldane map distance* measured in *centiMorgans* after [7].

For simplicity, consider letting $\gamma = 1$ and $h_i = 1$ for all $i \in \{1, \dots, M\}$. This is the case of a typical population measured on a single chromosome. Under these settings, the correlation matrix of \mathbf{Z} encoded using the additive map has an i, j entry

$$\text{Corr}(Z_i, Z_j) = e^{-2\beta d(i,j)}.$$

For the Haldane map with sequential positions $i < j < k$, the model in Appendix ?? gives $d(i, j) + d(j, k) = d(i, k)$, and so

$$\begin{aligned} \text{Corr}(Z_i, Z_k) &= e^{-2\beta d(i,k)} \\ &= e^{-2\beta(d(i,j)+d(j,k))} \\ &= e^{-2\beta d(i,j)} e^{-2\beta d(j,k)} \\ &= \text{Corr}(Z_i, Z_j) \text{Corr}(Z_j, Z_k). \end{aligned}$$

Take a final simplification and assume that $d(i, j) = c$ for all adjacent pairs j and k . Letting $e^{-2\beta c} = \rho < 1$, the correlation structure for M markers is given by

$$\Sigma = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{M-1} \\ \rho & 1 & \rho & \dots & \rho^{M-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{M-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{M-1} & \rho^{M-2} & \rho^{M-3} & \dots & 1 \end{bmatrix}. \quad (22)$$

As $M \rightarrow \infty$, [6] show that the eigenvalues of Equation (43), $\lambda_1, \dots, \lambda_M$ with $\lambda_1 \geq \lambda_2 \geq$

$\dots \geq \lambda_M$, approach

$$\begin{aligned}
f(x; \rho) &= \sum_{n=-\infty}^{\infty} \rho^{|n|} e^{inx} \\
&= \sum_{n=1}^{\infty} (\rho e^{ix})^n + 1 + \sum_{n=-\infty}^{-1} (\rho e^{-ix})^{-n} \\
&= \left(\frac{1}{1 - \rho e^{ix}} - 1 \right) + 1 + \left(\frac{1}{1 - \rho e^{-ix}} - 1 \right) \\
&= \frac{1 - \rho^2}{1 - \rho(e^{ix} + e^{-ix}) + \rho^2} \\
&= \frac{1 - \rho^2}{1 - 2\rho \cos x + \rho^2}
\end{aligned} \tag{23}$$

evaluated at x_1, \dots, x_M respectively, where $x_m = \frac{m\pi}{M+1}$. That is,

$$\lambda_m \xrightarrow{M \rightarrow \infty} f(x_m; \rho) \tag{24}$$

for $f(x; \rho) = \frac{1 - \rho^2}{1 - 2\rho \cos x + \rho^2}$ and $x_m = \frac{m\pi}{M+1}$. Note that i in Equation (23) is the imaginary unit, that is $i = \sqrt{-1}$. For the finite case, they further prove that

$$f(t_m; \rho) = \lambda_m \tag{25}$$

for t_1, \dots, t_M which satisfy

$$0 < t_1 < x_1 < t_2 < x_2 < \dots < t_M < x_M < \pi.$$

Noting that $f(x)$ is monotonically decreasing over $[0, \pi]$ and taking $x_0 = 0$, this bounds the eigenvalue λ_m in the interval $(f(x_m; \rho), f(x_{m-1}; \rho))$, or rather

$$\frac{1 - \rho^2}{1 - 2\rho \cos \frac{m\pi}{M+1} + \rho^2} < \lambda_m < \frac{1 - \rho^2}{1 - 2\rho \cos \frac{(m-1)\pi}{M+1} + \rho^2}. \tag{26}$$

Consider plotting these bounds for a given M over a sequence of ρ values to see their behaviour. Bounds for $\rho = 0, 0.5, 0.9$ and 0.99 are displayed for $M = 10, 50$, and 250 . The upper limit of $\rho = 1$ has been excluded because the upper bound in this case is divergent.

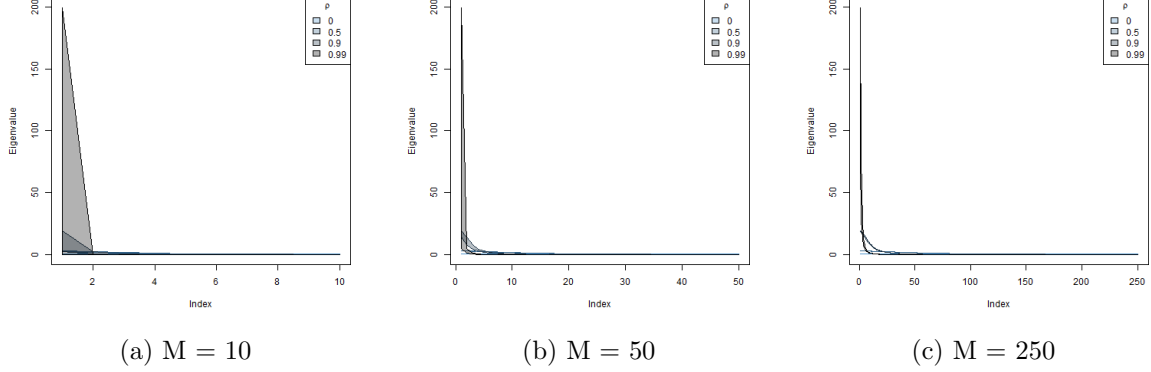


Figure 3: The bounds for each eigenvalue by index given by Equation (26) for different settings of M and ρ .

These bounds are based on $\rho < 1$ in any case, and when $\rho = 1$ the eigensystem is known. Figure 3 displays these bounds.

TODO: Reproduce these figures on a log scale

Note that the bounds get tighter as M increases, but the upper bound at the index $m = 1$ is constant for a given ρ regardless of M . This is because the upper bound on the largest eigenvalue is given by Equation (26) with $m = 1$, or rather

$$\frac{1 - \rho^2}{1 - 2\rho \cos \frac{(1-1)\pi}{M+1} + \rho^2} = \frac{1 - \rho^2}{1 - 2\rho + \rho^2} = \frac{1 + \rho}{1 - \rho}.$$

So, for $\rho = 0.99$ the upper bound is 199, for $\rho = 0.9$ it is 19, for $\rho = 0.5$ it is 3, and for $\rho = 0$ it is 1. Noting that the exceptionally large value for $\rho = 0.99$ dwarfs all other patterns, images with a limit of 20 on the vertical axis are displayed in Figure 4. The pattern in the bounds is clearer. As expected for an asymptotic result, the bounds grow tighter as M increases. Additionally, it is much clearer in this picture that as ρ approaches 1, the bounds grow wider for the large eigenvalues and the curve of eigenvalues approaches a step function which is zero everywhere except for the first index. That is, the shape of the curve approaches that of the known structure for $\rho = 1$.

It would therefore be reasonable to expect that any measure of the effective number of tests would approach the value of 1 for these eigenvalues. Indeed, Figure 4(c) would suggest the effective number of tests for $M = 250$ and $\rho = 0.99$ should already be quite

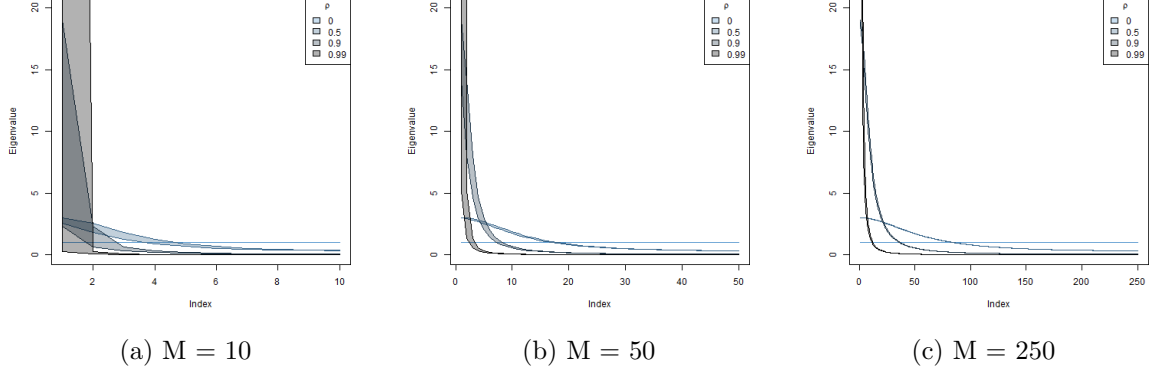


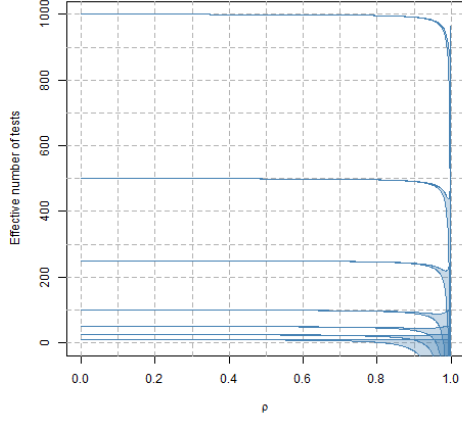
Figure 4: The eigenvalue bounds given by Equation (26) for different settings of M and ρ with the vertical axis restricted to show the patterns more clearly.

small. To test this, consider plugging the eigenvalue bounds displayed into Equations (11), (15), (17), and (20). This will produce bounds for the effective number of tests which can be plotted over a greater range of ρ values. Plotting these bounds for $M = 10, 25, 50, 100, 250, 500, 1000$ across all methods gives Figure 5.

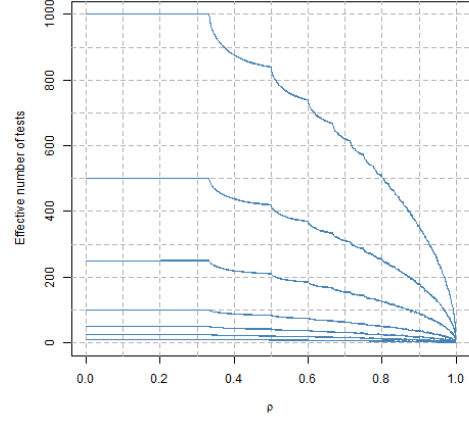
Of the four methods, Equation (20) proposed by [4] seems to have the intuitively best properties. It declines smoothly and monotonically towards 1 as ρ approaches 1. Equation (17) seems to fare the worst, as it fails to change appreciably with ρ , likely due to the way it suppresses large relative eigenvalues and magnifies small ones with the exponent $1/M$. Equation (11), meanwhile, only really begins to decrease when $\rho \approx 0.9$. Equation (15) has a similar overall behaviour to Equation (20), but with additional local patterns which seem spurious.

This indicates that Equations (11) and (17) will generally under-adjust the individual confidence intervals in real genetic data. While the results gained here are only valid for a single chromosome, the block diagonal structure of the entire genome suggests that this under-adjustment would still hold for genome-wide tests.

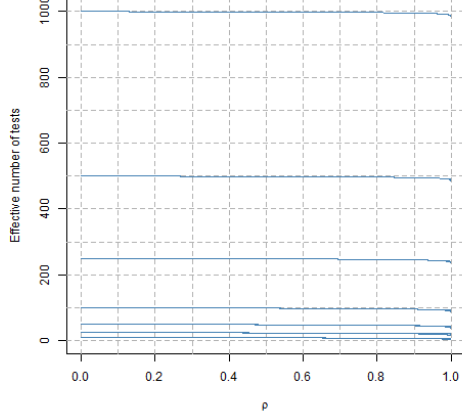
What remains unclear is what pattern is desirable for data of this type. The patterns of Figure 5(b) and (d) are intuitively appealing, but perhaps there is a theoretically optimal curve to adjust for testing in the genetic context. Such a curve could then be used to devise a method of calculating M_{eff} .



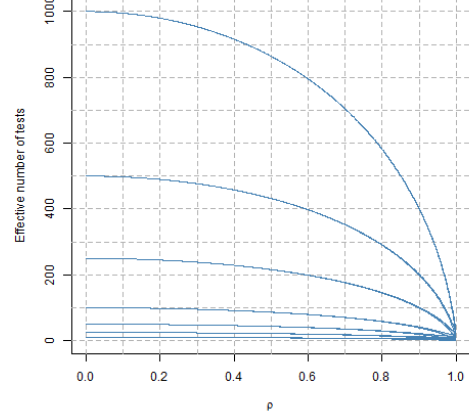
(a) Equation (11)



(b) Equation (15)



(c) Equation (17)



(d) Equation (20)

Figure 5: Bounds on the effective dimension for ρ computed by different means. For all bounds, the dimension M is given by the starting value when $\rho = 0$.

3.1 Circulant Matrices

Equation (43) is a rather particular correlation structure, however. Ideally, similar bounds to Equation (26) could be found for a broader class of matrices. [5] suggests an intuitive way to relax the form of Equation (43) based on the work of [6]. Rather than forcing an exponential decay away from the diagonal, consider the symmetric Toeplitz matrix

$$\mathbf{\Sigma} = \begin{bmatrix} \rho_0 & \rho_1 & \rho_2 & \cdots & \rho_{M-1} \\ \rho_1 & \rho_0 & \rho_1 & \cdots & \rho_{M-2} \\ \rho_2 & \rho_1 & \rho_0 & \cdots & \rho_{M-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{M-1} & \rho_{M-2} & \rho_{M-3} & \cdots & \rho_0 \end{bmatrix}. \quad (27)$$

This more general class of matrices does not admit simple bounds. Approximating the eigensystem of this general $\mathbf{\Sigma}$ is achieved through the use of *circulant matrices*, which have an asymptotically equivalent eigensystem. Unlike Equation (42), however, the eigensystem of circulant matrices is known.

This intuitive explanation additionally suggests a different framing of the problem. Instead of directly trying to gain the eigenvalues of Equation (42), consider attempting to optimize the circulant approximation used to be the best possible according to some metric. Note that in this section the eigenvalues of numerous different matrices are addressed, so let the function $\lambda_k(\mathbf{A})$ return the k^{th} eigenvalue of \mathbf{A} , which may not be ordered by magnitude.

A complex matrix $\mathbf{C} \in \mathbb{C}^{M \times M}$ is called circulant if the i^{th} row is given by the cyclic shift i elements rightward of a vector of M elements, typically denoted $(c_0, c_1, c_2, \dots, c_{M-1})$. Explicitly

$$\mathbf{C} = \begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_{M-2} & c_{M-1} \\ c_{M-1} & c_0 & c_1 & \cdots & c_{M-3} & c_{M-2} \\ c_{M-2} & c_{M-1} & c_0 & \cdots & c_{M-4} & c_{M-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ c_2 & c_3 & c_4 & \cdots & c_0 & c_1 \\ c_1 & c_2 & c_3 & \cdots & c_{M-1} & c_0 \end{bmatrix} \quad (28)$$

So every circulant matrix \mathbf{C} can be specified by its first row alone. Moreover, this first row corresponds to the coefficients in a convenient expression of \mathbf{C} as a matrix polynomial. Let \mathbf{P} be the circulant matrix with $c_0 = c_2 = c_3 = \cdots = c_{M-1} = 0$ and $c_1 = 1$. That is,

$$\mathbf{P} = [\mathbf{e}_M | \mathbf{e}_1 | \cdots | \mathbf{e}_{M-1}] \quad (29)$$

where \mathbf{e}_i is the i^{th} basis vector. Then \mathbf{P} is the permutation matrix corresponding to a cyclic shift of all elements of a vector $\mathbf{x} \in \mathbb{C}^M$ one to the right. Due to this cyclic shift property, it is also straightforward to note that

$$\mathbf{P}^m = \mathbf{P}\mathbf{P} \dots \mathbf{P} = [\mathbf{e}_{M-m+1} | \mathbf{e}_{M-m+2} | \dots | \mathbf{e}_M | \mathbf{e}_1 | \dots | \mathbf{e}_{M-m}]. \quad (30)$$

Using these \mathbf{P}^m , \mathbf{C} can be written

$$\mathbf{C} = c_0 \mathbf{I}_M + c_1 \mathbf{P} + c_2 \mathbf{P}^2 + \dots + c_{M-1} \mathbf{P}^{M-1}, \quad (31)$$

from which the eigensystem of \mathbf{C} can be simply derived from the eigensystem of \mathbf{P} . Using a cofactor expansion of $\det(\mathbf{P} - \lambda \mathbf{I})$ it can be shown that the eigenvalues of \mathbf{P} are the M^{th} roots of unity, that is

$$\lambda_k(\mathbf{P}) = \left(e^{\frac{2\pi i}{M}} \right)^k = \omega^k$$

where $k = 0, \dots, M-1$. The corresponding eigenvectors \mathbf{x}_k are then

$$\mathbf{x}_k = \begin{bmatrix} 1 \\ \omega^k \\ \omega^{2k} \\ \vdots \\ \omega^{(M-1)k} \end{bmatrix}, \quad (32)$$

which can be seen by considering $\mathbf{P}\mathbf{x}_k = \lambda_k(\mathbf{P})\mathbf{x}_k$. Note that any eigenvector of \mathbf{P} with eigenvalue λ is also an eigenvector of \mathbf{P}^m with eigenvalue λ^m , and so the eigenvalues of \mathbf{C} are given by

$$\lambda_k(\mathbf{C}) = c_0 + \sum_{m=1}^{M-1} c_m \omega^{mk} \quad (33)$$

with corresponding eigenvectors \mathbf{x}_k as above for $k = 0, \dots, M-1$. A particular circulant matrix structure is of interest.

Definition 1 (Symmetric circulant). *A circulant matrix $\mathbf{C} \in \mathbb{C}^{M \times M}$ is symmetric if elements in its first row $c_0, c_1, c_2, \dots, c_{M-1}$ satisfy $c_m = c_{M-m}$ for all $m \geq 1$.*

To provide a visual example of such a matrix, consider the circulant with $c_m = \min\{m, M-m\}$. When $M = 6$ we have

$$\begin{bmatrix} 0 & 1 & 2 & 3 & 2 & 1 \\ 1 & 0 & 1 & 2 & 3 & 2 \\ 2 & 1 & 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 & 1 & 2 \\ 2 & 3 & 2 & 1 & 0 & 1 \\ 1 & 2 & 3 & 2 & 1 & 0 \end{bmatrix}$$

This circulant is symmetric, and any symmetric circulant matrix will have an analogous structure. By basic results of linear algebra, it follows that any symmetric circulant matrix will have only real eigenvalues. Indeed, a circulant matrix will have only real eigenvalues for any M if and only if it is symmetric.

Theorem 1 (Real eigenvalues of symmetric circulants). *A circulant matrix $\mathbf{C} \in \mathbb{C}^{M \times M}$ has real eigenvalues if and only if it is a symmetric circulant matrix.*

Proof. Consider the eigensystem of \mathbf{C} . As it is circulant, it has eigenvalues

$$\lambda_k(\mathbf{C}) = c_0 + \sum_{m=1}^{M-1} c_m \omega^{mk},$$

or rather

$$\lambda_k(\mathbf{C}) = c_0 + \sum_{m=1}^{M-1} c_m \left(\cos \frac{2\pi mk}{M} + i \sin \frac{2\pi mk}{M} \right)$$

We can rewrite this to emphasize the real and imaginary components as

$$\lambda_k(\mathbf{C}) = \left(\sum_{m=0}^{M-1} c_m \cos \frac{2\pi mk}{M} \right) + i \left(\sum_{m=1}^{M-1} c_m \sin \frac{2\pi mk}{M} \right).$$

If $\lambda_k(\mathbf{C}) \in \mathbb{R}$ for all $k \in \{0, 1, \dots, M-1\}$, we must have

$$\sum_{m=1}^{M-1} c_m i \sin \frac{2\pi mk}{M} = 0 \quad \forall k \in \{0, 1, \dots, M-1\}.$$

But note that

$$i \sin \frac{2\pi mk}{M} = \frac{1}{2} \left(e^{\frac{2\pi mk}{M} i} - e^{-\frac{2\pi mk}{M} i} \right)$$

and

$$e^{-\frac{2\pi mk}{M} i} = e^{-\frac{2\pi mk}{M} i + 2\pi k i} = e^{\frac{2\pi(M-m)k}{M} i},$$

and so we require

$$\sum_{m=1}^{M-1} c_m \frac{1}{2} \left(e^{\frac{2\pi mk}{M} i} - e^{\frac{2\pi(M-m)k}{M} i} \right) = 0 \quad \forall k \in \{0, 1, 2, \dots, M-1\}.$$

However,

$$\sum_{m=1}^{M-1} c_m \frac{1}{2} \left(e^{\frac{2\pi mk}{M} i} - e^{\frac{2\pi(M-m)k}{M} i} \right) = 0$$

$$\iff \sum_{m=1}^{M-1} c_m e^{\frac{2\pi mk}{M}i} - \sum_{m=1}^{M-1} c_m e^{\frac{2\pi(M-m)k}{M}i} = 0$$

$$\iff \sum_{m=1}^{M-1} c_m e^{\frac{2\pi mk}{M}i} - \sum_{m=1}^{M-1} c_{M-m} e^{\frac{2\pi mk}{M}i} = 0$$

$$\iff \sum_{m=1}^{M-1} (c_m - c_{M-m}) e^{\frac{2\pi mk}{M}i} = \sum_{m=1}^{M-1} (c_m - c_{M-m}) \omega^{mk} = 0$$

for all $k \in \{0, 1, \dots, M-1\}$. In other words, the eigenvalues of \mathbf{C} are all real if and only if the vector of differences $(c_m - c_{M-m})_{m=1, \dots, M-1}$ is a vector in the null space of

$$\begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ \omega & \omega^2 & \omega^3 & \dots & \omega^{M-1} \\ \omega^2 & \omega^4 & \omega^6 & \dots & \omega^{2(M-1)} \\ \omega^3 & \omega^6 & \omega^9 & \dots & \omega^{3(M-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \omega^{M-1} & \omega^{2(M-1)} & \omega^{3(M-1)} & \dots & \omega^{(M-1)^2} \end{bmatrix}.$$

Noting that these columns are eigenvectors of \mathbf{C} , it can quickly be recognized that the null space of this matrix is the line $x_1 = x_2 = \dots = x_M$, as this is the final orthogonal eigenvector of \mathbf{C} . Therefore, our differences must satisfy

$$c_m - c_{M-m} = a$$

for some $a \in \mathbb{C}$ for all $m \in \{1, 2, \dots, M-1\}$. If M is even, then $M/2 \in \{1, 2, \dots, M-1\}$, and so when $m = M/2$ we get the difference $c_{M/2} - c_{M-M/2} = c_{M/2} - c_{M/2} = 0$. Hence, $a = 0$ is the only solution. If M is odd, then we have in particular $c_{\lfloor M/2 \rfloor} - c_{M-\lfloor M/2 \rfloor} = c_{\lfloor M/2 \rfloor} - c_{\lfloor M/2 \rfloor+1} = a = c_{\lfloor M/2 \rfloor+1} - c_{\lfloor M/2 \rfloor} = c_{\lfloor M/2 \rfloor+1} - c_{M-\lfloor M/2 \rfloor-1}$, which is true only if $c_{\lfloor M/2 \rfloor+1} = c_{\lfloor M/2 \rfloor}$ and $a = 0$. Therefore, real eigenvalues are ensured if and only if

$$c_m - c_{M-m} = 0 \iff c_m = c_{M-m},$$

the definition of a symmetric circulant. □

With this proof, we can move to a statistical framing of the problem of the eigenvalue distribution of Σ . Consider the decomposition

$$\Sigma = \mathbf{C} + \mathbf{R} \quad (34)$$

where \mathbf{C} is a circulant matrix and $\mathbf{R} = \Sigma - \mathbf{C}$ is a matrix of the element-wise residuals between \mathbf{C} and Σ . This reframing moves the discussion from the space of asymptotic results to the features of \mathbf{C} and \mathbf{R} , and places the result in a familiar framework for the statistician accustomed to considering the residuals of a given approximation.

Of immediate and obvious interest is the “closest” circulant matrix to a given Σ , call it \mathbf{C}_Σ . Consider using the weak, or Frobenius, norm on matrices. First, introduce the *vectorization* operator on a matrix $\mathbf{A} \in \mathbb{C}^{M \times N}$, denoted $\text{vec } \mathbf{A}$. This operator takes $\mathbf{A} \in \mathbb{C}^{M \times N}$ and converts it to a vector in \mathbb{C}^{MN} by appending columns in order. So, for example,

$$\text{vec} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 2 \\ 5 \\ 3 \\ 6 \end{bmatrix}.$$

The Frobenius norm is then given by

$$\|\mathbf{A}\|_F = \sqrt{(\text{vec } \mathbf{A})^* (\text{vec } \mathbf{A})}$$

or equivalently

$$\|\mathbf{A}\|_F = \sqrt{\text{trace}(\mathbf{A}^* \mathbf{A})}$$

for a matrix $\mathbf{A} \in \mathbb{C}^{M \times M}$ with complex conjugate \mathbf{A}^* . Note that for any real pairwise measure $\Sigma, \mathbf{C}, \mathbf{R} \in \mathbb{R}^{M \times M}$, and so for our purpose

$$\|\mathbf{A}\|_F = \sqrt{(\text{vec } \mathbf{A})^\top (\text{vec } \mathbf{A})} = \sqrt{\text{trace}(\mathbf{A}^\top \mathbf{A})}.$$

Let \mathbf{C} be an $M \times M$ circulant matrix with first row $(c_0, c_1, c_2, \dots, c_{M-1})$. Then by definition \mathbf{C}_Σ is given by $\arg \min_{\mathbf{C}} \|\Sigma - \mathbf{C}\|_F$, or equivalently $\arg \min_{\mathbf{C}} \|\Sigma - \mathbf{C}\|_F^2$. Taking the second of these, we have

$$\mathbf{C}_\Sigma = \arg \min_{\mathbf{C}} \|\Sigma - \mathbf{C}\|_F^2. \quad (35)$$

Considering that

$$\begin{aligned} \|\Sigma - \mathbf{C}\|_F^2 &= \text{trace} \left([\Sigma - \mathbf{C}]^\top [\Sigma - \mathbf{C}] \right) \\ &= \left(\text{trace} \Sigma^\top \Sigma - \text{trace} \Sigma^\top \mathbf{C} - \text{trace} \mathbf{C}^\top \Sigma + \text{trace} \mathbf{C}^\top \mathbf{C} \right) \end{aligned}$$

and $\text{trace} \Sigma^\top \Sigma$ is constant in \mathbf{C} , we need only consider minimizing

$$F(\mathbf{C}) = \text{trace} \mathbf{C}^\top \mathbf{C} - \text{trace} \Sigma^\top \mathbf{C} - \text{trace} \mathbf{C}^\top \Sigma. \quad (36)$$

The first term of Equation (53) is straightforward to express in terms of the c_m . As \mathbf{C} is circulant,

$$\text{trace} \mathbf{C}^\top \mathbf{C} = M \sum_{m=0}^{M-1} c_m^2.$$

The other terms can be evaluated by considering \mathbf{C} as a matrix polynomial.

Equation (48) and the symmetry of Σ allow us to write

$$\Sigma^\top \mathbf{C} = \Sigma \left(c_0 \mathbf{I} + \sum_{m=1}^{M-1} c_m \mathbf{P}^m \right) = c_0 \Sigma + \sum_{m=1}^{M-1} c_m \Sigma \mathbf{P}^m$$

and similarly

$$\mathbf{C}^\top \Sigma = c_0 \Sigma + \sum_{m=1}^{M-1} c_m (\mathbf{P}^m)^\top \Sigma = c_0 \Sigma + \sum_{m=1}^{M-1} c_m (\Sigma \mathbf{P}^m)^\top.$$

Next, consider

$$\text{trace} \Sigma \mathbf{P}^m = \text{trace} \left(\Sigma [\mathbf{e}_{M-m+1} | \mathbf{e}_{M-m+2} | \dots | \mathbf{e}_{M-m}] \right),$$

which can be evaluated by considering the k^{th} row of Σ , σ_k . The first k elements of this row are the descending sequence $\rho_{k-1}, \rho_{k-2}, \dots, \rho_0$, and the remaining elements are the ascending sequence $\rho_1, \rho_2, \dots, \rho_{M-k}$. Noting that the trace of a product of two matrices is simply a sum of the inner products of the rows of the first with the columns of the second,

we obtain

$$\begin{aligned}
\text{trace } \mathbf{\Sigma} \mathbf{P}^m &= \sum_{k=1}^m \boldsymbol{\sigma}_k^\top \mathbf{e}_{M-m+k} + \sum_{k=1}^{M-m} \boldsymbol{\sigma}_{m+k}^\top \mathbf{e}_k \\
&= \sum_{k=1}^m \rho_{M-m} + \sum_{k=1}^{M-m} \rho_m \\
&= m\rho_{M-m} + (M-m)\rho_m.
\end{aligned} \tag{37}$$

Equation (54) can then be substituted into Equation (53) using the decomposition of Equation (48) to give

$$F(\mathbf{C}) = M \sum_{m=0}^{M-1} c_m^2 - 2 \left(M c_0 \rho_0 + \sum_{m=1}^{M-1} c_m (m\rho_{M-m} + (M-m)\rho_m) \right). \tag{38}$$

As $\arg \min_{\mathbf{C}} F(\mathbf{C})$ is the same as $\arg \min_{\mathbf{C}} \|\mathbf{\Sigma} - \mathbf{C}\|_F$, we can now consider the values which minimize Equation (55) in order to find the nearest circulant matrix to $\mathbf{\Sigma}$. Taking

$$\frac{\partial}{\partial c_m} F(\mathbf{C}) = \begin{cases} 2Mc_0 - 2M\rho_0 & \text{for } m = 0, \\ 2Mc_m - 2(m\rho_{M-m} + (M-m)\rho_m) & \text{otherwise,} \end{cases}$$

and noting that the Hessian matrix is $2M\mathbf{I}$ and hence is positive definite so any solutions to $\arg \min F(\mathbf{C})$ must be minima, we obtain

$$c_m = \begin{cases} \rho_0 & \text{for } m = 0, \\ \frac{m}{M}\rho_{M-m} + \frac{M-m}{M}\rho_m & \text{otherwise,} \end{cases}$$

which can be re-expressed as

$$c_m = \begin{cases} \rho_0 & \text{for } m = 0, \\ \rho_m + \frac{m}{M}(\rho_{M-m} - \rho_m) & \text{otherwise,} \end{cases} \tag{39}$$

to make the relationship between c_m and ρ_m clearer. An important consequence of this system of equations is that $c_m = \frac{m}{M}\rho_{M-m} + \frac{M-m}{M}\rho_m = \frac{M-(M-m)}{M}\rho_{M-m} + \frac{M-m}{M}\rho_{M-(M-m)} = c_{M-m}$, and so \mathbf{C}_Σ is a symmetric circulant matrix with c_m defined as in Equation (56). Therefore, this nearest circulant will have only real eigenvalues.

So the optimal decomposition in the Frobenius norm is

$$\mathbf{\Sigma} = \mathbf{C}_\Sigma + \mathbf{R}_\Sigma \quad (40)$$

where \mathbf{C}_Σ is the circulant matrix defined by Equation (56), that is

$$c_m = \begin{cases} \rho_0 & \text{for } m = 0, \\ \rho_m + \frac{m}{M}(\rho_{M-m} - \rho_m) & \text{otherwise,} \end{cases}$$

and $\mathbf{R}_\Sigma = \mathbf{\Sigma} - \mathbf{C}_\Sigma$, and so the value in the m^{th} off-diagonal of \mathbf{R}_Σ is $\rho_m - c_m = \rho_m - (\rho_m + \frac{m}{M}(\rho_{M-m} - \rho_m)) = \frac{m}{M}(\rho_m - \rho_{M-m})$.

The eigenvalues of \mathbf{C}_Σ are given by a substitution of Equation (56) into Equation (50), giving

$$\lambda_k(\mathbf{C}_\Sigma) = \rho_0 + 2 \sum_{m=1}^{M-1} \frac{M-m}{M} \rho_m \cos \frac{2\pi mk}{M}. \quad (41)$$

The corresponding eigenvectors are given by Equation (49).

Consider \mathbf{R}_Σ briefly. For large M and small m , $\frac{m}{M}(\rho_{M-m} - \rho_m) \approx 0$ while when m is close to M , $\frac{m}{M}(\rho_{M-m} - \rho_m) \approx \rho_{M-m} - \rho_m$. This implies that in the case of large M , \mathbf{R}_Σ will have vanishingly small values for the central off-diagonals, and values in the corners of approximately $\rho_m - \rho_{M-m}$.

Note that the approximate matrix \mathbf{C}_Σ has been derived here based purely on minimizing $\|\mathbf{\Sigma} - \mathbf{C}\|_F$ without any of the asymptotic guarantees of [6]. It is therefore worthwhile to compare the eigenvalues of \mathbf{C}_Σ in the special case of Equation (43) to the bounds in Equation (26) to evaluate this approximation. That is, to check the approximation of the eigenvalues given by \mathbf{C}_Σ in the case where $\rho_m = \rho^m$.

A simple inspection involves choosing M and ρ and plotting the eigenvalue by index alongside the bounds by index as in Figure 3. Suppose $M = 100$ and $\rho = 0.45$ are the chosen values. Then Figure 6 results. While there are some patterns in the eigenvalue distribution of note, particularly its roughness, there is an encouraging agreement of both the general shape and the general value. Choosing other M and ρ gives similar agreement.

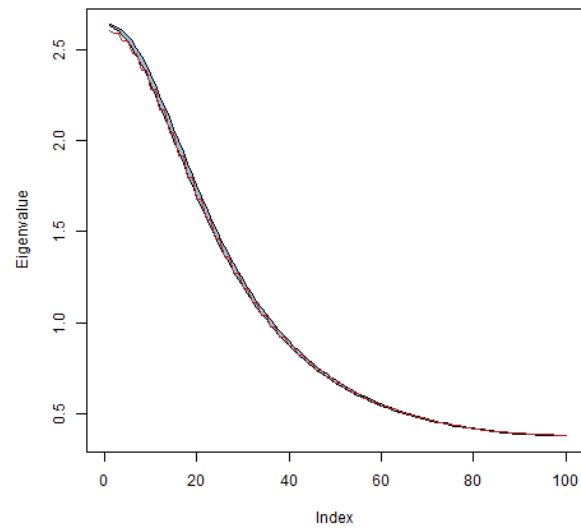


Figure 6: The eigenvalue distribution of \mathbf{C}_Σ and bounds of Equation (26) for $M = 100$ and $\rho = 0.45$.

More systematically, consider each of the bounds shown in Figure 3. For each of the M and ρ settings displayed, generate the approximate circulant eigenvalues using \mathbf{C}_Σ . These approximate eigenvalues can then be subtracted from the corresponding set of bounds, and the difference scaled by the approximate eigenvalue. The relative differences of the upper and lower bound can then be used to define an analogous polygon where the line of zero indicates perfect agreement. Note that for $\rho = 0$, the agreement is perfect for all M . Figure 7 displays relative differences for $M = 10, 50$, and 250 and $\rho = 0.5, 0.9$, and 0.99 .

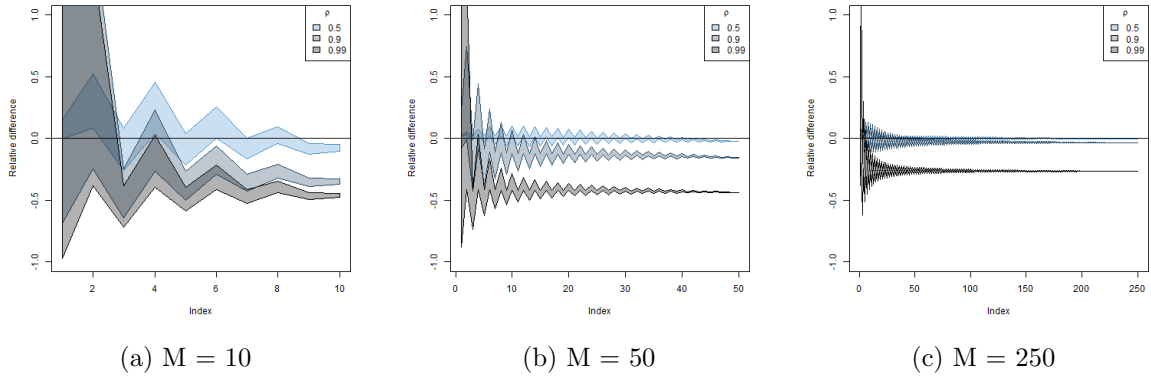


Figure 7: The relative difference of eigenvalue bounds from Equation (26) and the eigenvalues of \mathbf{C}_Σ for different M and ρ values.

Two patterns stand out. First, there is an obvious cyclical pattern for all M and ρ . This suggests that the roughness seen in Figure 6 is typical. Second, there is a clear pattern of bias which depends on ρ and M . For small M and large ρ , when the bounds are widest and the asymptotic approximation worst, the eigenvalues of \mathbf{C}_Σ tend to be larger than the bounds. As M increases, the magnitude of this bias decreases.

A number of methods exist to approximate the symmetric Toeplitz matrix

$$\Sigma = \begin{bmatrix} \rho_0 & \rho_1 & \rho_2 & \cdots & \rho_{M-1} \\ \rho_1 & \rho_0 & \rho_1 & \cdots & \rho_{M-2} \\ \rho_2 & \rho_1 & \rho_0 & \cdots & \rho_{M-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{M-1} & \rho_{M-2} & \rho_{M-3} & \cdots & \rho_0 \end{bmatrix}, \quad (42)$$

where $\rho_0, \dots, \rho_{M-1} \in \mathbb{R}$. Σ appears regularly in statistical applications such as the correlation matrix of the discrete autoregressive process of order one, in information theory

in certain filtering tasks [5], and in genetics in the correlation between disjoint measured sequences [?]. In the case of the equidistant genetic survey of [9] and the autoregressive time series, the matrix has the more specific form

$$\mathbf{\Sigma}_e = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{M-1} \\ \rho & 1 & \rho & \dots & \rho^{M-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{M-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{M-1} & \rho^{M-2} & \rho^{M-3} & \dots & 1 \end{bmatrix} \quad (43)$$

where $\rho \in [0, 1]$ is a real constant.

Of interest in many applications are the eigenvalues and eigenvectors of $\mathbf{\Sigma}$, call them $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_M$ and $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$ respectively. For brevity, the combination of these vectors and values is here referred to as the *eigensystem* of $\mathbf{\Sigma}$. [5] demonstrates the application of this particular eigensystem in signal processing, while [1], [10], and [4] use the eigenvalues to adjust for dependent multiple tests in the genomic context.

The typical approach to obtain is eigensystem is to approximate $\mathbf{\Sigma}$ with a matrix with a solved eigensystem. [1], [10], and [4] present different approximations based on the matrix

$$\mathbf{A}(\rho) = \rho \mathbf{1}\mathbf{1}^\top + (1 - \rho)\mathbf{I}_M. \quad (44)$$

$\mathbf{A}(\rho)$ has a known eigensystem for any ρ and matches the eigensystem of $\mathbf{\Sigma}$ for the edge cases where $\rho_1 = \rho_2 = \dots = 0$ or $\rho_1 = \rho_2 = \dots = 0$ and $\rho_0 = 1$.

[5] and [6] instead utilize circulant matrices to approximate the eigensystem of $\mathbf{\Sigma}$. Asymptotically, $\mathbf{\Sigma}$ has the same eigensystem as certain circulant matrices and the eigensystem of any circulant matrix is known.

Absent from the asymptotic approach of [5] or the myriad approaches of [1], [10], and [4] is a consideration of the finite, non-edge cases of $\mathbf{\Sigma}$ and how it might be best approximated. This work derives \mathbf{C}_Σ , the nearest circulant matrix to $\mathbf{\Sigma}$ in the weak, or Frobenius, norm. \mathbf{C}_Σ is symmetric circulant, and so has only real eigenvalues. Indeed, it is proven that a real circulant matrix has real eigenvalues if and only if it is a symmetric circulant. \mathbf{C}_{Σ_e} is additionally shown to be asymptotically equivalent to $\mathbf{\Sigma}_e$ by considering the limit of the remainder $\mathbf{\Sigma}_e - \mathbf{C}_{\Sigma_e}$. The approximation of \mathbf{C}_Σ and the classic approximation of [5] are compared, with the nearest circulant provided a much better approximation in the cases of ρ close to 1 and small M .

Section 4 presents an introduction to circulant matrices and the proof that the only real circulants with all real eigenvalues are symmetric. Section 5 derives \mathbf{C}_Σ and \mathbf{C}_{Σ_e} and

states their eigensystems. The asymptotic equivalence of Σ_e and \mathbf{C}_Σ are then present in Section 6 before a comparison of the approximations is present in Section 7.

4 Circulant matrices

For clarity working with different matrices, let the function $\lambda_k(\mathbf{C})$ return the k^{th} eigenvalue of \mathbf{C} , which may not be ordered by magnitude.

A complex matrix $\mathbf{C} \in \mathbb{C}^{M \times M}$ is called circulant if the i^{th} row is given by the cyclic shift i elements rightward of a vector of M elements, typically denoted $(c_0, c_1, c_2, \dots, c_{M-1})$. Explicitly

$$\mathbf{C} = \begin{bmatrix} c_0 & c_1 & c_2 & \dots & c_{M-2} & c_{M-1} \\ c_{M-1} & c_0 & c_1 & \dots & c_{M-3} & c_{M-2} \\ c_{M-2} & c_{M-1} & c_0 & \dots & c_{M-4} & c_{M-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ c_2 & c_3 & c_4 & \dots & c_0 & c_1 \\ c_1 & c_2 & c_3 & \dots & c_{M-1} & c_0 \end{bmatrix} \quad (45)$$

So every circulant matrix \mathbf{C} can be specified by its first row alone. Moreover, this first row corresponds to the coefficients in a convenient expression of \mathbf{C} as a matrix polynomial. Let \mathbf{P} be the circulant matrix with $c_0 = c_2 = c_3 = \dots = c_{M-1} = 0$ and $c_1 = 1$. That is,

$$\mathbf{P} = [\mathbf{e}_M | \mathbf{e}_1 | \dots | \mathbf{e}_{M-1}] \quad (46)$$

where \mathbf{e}_i is the i^{th} basis vector. Then \mathbf{P} is the permutation matrix corresponding to a cyclic shift of all elements of a vector $\mathbf{x} \in \mathbb{C}^M$ one to the right. Due to this cyclic shift property, it is also straightforward to note that

$$\mathbf{P}^m = \mathbf{P}\mathbf{P} \dots \mathbf{P} = [\mathbf{e}_{M-m+1} | \mathbf{e}_{M-m+2} | \dots | \mathbf{e}_M | \mathbf{e}_1 | \dots | \mathbf{e}_{M-m}]. \quad (47)$$

Using these \mathbf{P}^m , \mathbf{C} can be written

$$\mathbf{C} = c_0 \mathbf{I}_M + c_1 \mathbf{P} + c_2 \mathbf{P}^2 + \dots + c_{M-1} \mathbf{P}^{M-1}, \quad (48)$$

from which the eigensystem of \mathbf{C} can be derived from the eigensystem of \mathbf{P} . Using a cofactor expansion of $\det(\mathbf{P} - \lambda \mathbf{I})$ it can be shown that the eigenvalues of \mathbf{P} are the M^{th} roots of unity, that is

$$\lambda_k(\mathbf{P}) = \left(e^{\frac{2\pi i}{M}} \right)^k = \omega^k$$

where $k = 0, \dots, M - 1$. The corresponding eigenvectors \mathbf{x}_k are then

$$\mathbf{x}_k = \begin{bmatrix} 1 \\ \omega^k \\ \omega^{2k} \\ \vdots \\ \omega^{(M-1)k} \end{bmatrix}, \quad (49)$$

which can be seen by considering $\mathbf{P}\mathbf{x}_k = \lambda_k(\mathbf{P})\mathbf{x}_k$. Note that any eigenvector of \mathbf{P} with eigenvalue λ is also an eigenvector of \mathbf{P}^m with eigenvalue λ^m , and so the eigenvalues of \mathbf{C} are given by

$$\lambda_k(\mathbf{C}) = c_0 + \sum_{m=1}^{M-1} c_m \omega^{mk} \quad (50)$$

with corresponding eigenvectors \mathbf{x}_k as above for $k = 0, \dots, M - 1$. A particular circulant matrix structure is of interest.

Definition 2 (Symmetric circulant). *A circulant matrix $\mathbf{C} \in \mathbb{C}^{M \times M}$ is symmetric if elements in its first row $c_0, c_1, c_2, \dots, c_{M-1}$ satisfy $c_m = c_{M-m}$ for all $m \geq 1$.*

To provide a visual example of such a matrix, consider the circulant with $c_m = \min\{m, M - m\}$. When $M = 6$ we have

$$\begin{bmatrix} 0 & 1 & 2 & 3 & 2 & 1 \\ 1 & 0 & 1 & 2 & 3 & 2 \\ 2 & 1 & 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 & 1 & 2 \\ 2 & 3 & 2 & 1 & 0 & 1 \\ 1 & 2 & 3 & 2 & 1 & 0 \end{bmatrix}$$

This circulant is symmetric, and any symmetric circulant matrix will have an analogous structure. By basic results of linear algebra, it follows that any symmetric circulant matrix will have only real eigenvalues. Indeed, a circulant matrix will have only real eigenvalues for any M if and only if it is symmetric.

Theorem 2 (Real eigenvalues of symmetric circulants). *A circulant matrix $\mathbf{C} \in \mathbb{R}^{M \times M}$ has real eigenvalues if and only if it is a symmetric circulant matrix.*

Proof. Consider the eigensystem of \mathbf{C} . As it is circulant, it has eigenvalues

$$\lambda_k(\mathbf{C}) = c_0 + \sum_{m=1}^{M-1} c_m \omega^{mk},$$

or rather

$$\lambda_k(\mathbf{C}) = c_0 + \sum_{m=1}^{M-1} c_m \left(\cos \frac{2\pi mk}{M} + i \sin \frac{2\pi mk}{M} \right)$$

We can rewrite this to emphasize the real and imaginary components as

$$\lambda_k(\mathbf{C}) = \left(\sum_{m=0}^{M-1} c_m \cos \frac{2\pi mk}{M} \right) + i \left(\sum_{m=1}^{M-1} c_m \sin \frac{2\pi mk}{M} \right).$$

If $\lambda_k(\mathbf{C}) \in \mathbb{R}$ for all $k \in \{0, 1, \dots, M-1\}$, we must have

$$\sum_{m=1}^{M-1} c_m i \sin \frac{2\pi mk}{M} = 0 \quad \forall k \in \{0, 1, \dots, M-1\}.$$

But note that

$$i \sin \frac{2\pi mk}{M} = \frac{1}{2} \left(e^{\frac{2\pi mk}{M}i} - e^{-\frac{2\pi mk}{M}i} \right)$$

and

$$e^{-\frac{2\pi mk}{M}i} = e^{-\frac{2\pi mk}{M}i + 2\pi ki} = e^{\frac{2\pi(M-m)k}{M}i},$$

and so we require

$$\sum_{m=1}^{M-1} c_m \frac{1}{2} \left(e^{\frac{2\pi mk}{M}i} - e^{\frac{2\pi(M-m)k}{M}i} \right) = 0 \quad \forall k \in \{0, 1, 2, \dots, M-1\}.$$

However,

$$\begin{aligned} & \sum_{m=1}^{M-1} c_m \frac{1}{2} \left(e^{\frac{2\pi mk}{M}i} - e^{\frac{2\pi(M-m)k}{M}i} \right) = 0 \\ \iff & \sum_{m=1}^{M-1} c_m e^{\frac{2\pi mk}{M}i} - \sum_{m=1}^{M-1} c_m e^{\frac{2\pi(M-m)k}{M}i} = 0 \end{aligned}$$

$$\iff \sum_{m=1}^{M-1} c_m e^{\frac{2\pi mk}{M}i} - \sum_{m=1}^{M-1} c_{M-m} e^{\frac{2\pi mk}{M}i} = 0$$

$$\iff \sum_{m=1}^{M-1} (c_m - c_{M-m}) e^{\frac{2\pi mk}{M}i} = \sum_{m=1}^{M-1} (c_m - c_{M-m}) \omega^{mk} = 0$$

for all $k \in \{0, 1, \dots, M-1\}$. In other words, the eigenvalues of \mathbf{C} are all real if and only if the vector of differences $(c_m - c_{M-m})_{m=1, \dots, M-1}$ is a vector in the null space of

$$\begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ \omega & \omega^2 & \omega^3 & \dots & \omega^{M-1} \\ \omega^2 & \omega^4 & \omega^6 & \dots & \omega^{2(M-1)} \\ \omega^3 & \omega^6 & \omega^9 & \dots & \omega^{3(M-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \omega^{M-1} & \omega^{2(M-1)} & \omega^{3(M-1)} & \dots & \omega^{(M-1)^2} \end{bmatrix}.$$

Noting that these columns are eigenvectors of \mathbf{C} , it can quickly be recognized that the null space of this matrix is the line $x_1 = x_2 = \dots = x_M$, as this is the final orthogonal eigenvector of \mathbf{C} . Therefore, our differences must satisfy

$$c_m - c_{M-m} = a$$

for some $a \in \mathbb{C}$ for all $m \in \{1, 2, \dots, M-1\}$. If M is even, then $M/2 \in \{1, 2, \dots, M-1\}$, and so when $m = M/2$ we get the difference $c_{M/2} - c_{M-M/2} = c_{M/2} - c_{M/2} = 0$. Hence, $a = 0$ is the only solution. If M is odd, then we have in particular $c_{\lfloor M/2 \rfloor} - c_{M-\lfloor M/2 \rfloor} = c_{\lfloor M/2 \rfloor} - c_{\lfloor M/2 \rfloor + 1} = a = c_{\lfloor M/2 \rfloor + 1} - c_{\lfloor M/2 \rfloor} = c_{\lfloor M/2 \rfloor + 1} - c_{M-\lfloor M/2 \rfloor - 1}$, which is true only if $c_{\lfloor M/2 \rfloor + 1} = c_{\lfloor M/2 \rfloor}$ and $a = 0$. Therefore, real eigenvalues are ensured if and only if

$$c_m - c_{M-m} = 0 \iff c_m = c_{M-m},$$

the definition of a symmetric circulant. □

5 The nearest circulant to Σ

Now consider the decomposition

$$\Sigma = \mathbf{C} + \mathbf{R} \tag{51}$$

where \mathbf{C} is a circulant matrix and $\mathbf{R} = \mathbf{\Sigma} - \mathbf{C}$ is a matrix of the element-wise residuals between \mathbf{C} and $\mathbf{\Sigma}$. This reframing moves the discussion from the space of asymptotic results to the features of \mathbf{C} and \mathbf{R} , a familiar framework for anyone accustomed to considering the residuals of a given approximation.

Of immediate and obvious interest is the closest circulant matrix to $\mathbf{\Sigma}$, call it \mathbf{C}_{Σ} . Consider using the weak, or Frobenius, norm on matrices. First, introduce the *vectorization* operator on a matrix $\mathbf{A} \in \mathbb{C}^{M \times N}$, denoted $\text{vec } \mathbf{A}$. This operator takes $\mathbf{A} \in \mathbb{C}^{M \times N}$ and converts it to a vector in \mathbb{C}^{MN} by appending columns in order. So, for example,

$$\text{vec} \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 2 \\ 5 \end{bmatrix}.$$

The Frobenius norm is then given by

$$\|\mathbf{A}\|_F = \sqrt{(\text{vec } \mathbf{A})^* (\text{vec } \mathbf{A})}$$

or equivalently

$$\|\mathbf{A}\|_F = \sqrt{\text{trace}(\mathbf{A}^* \mathbf{A})}$$

for a matrix $\mathbf{A} \in \mathbb{C}^{M \times M}$ with complex conjugate \mathbf{A}^* . In the real case, this is

$$\|\mathbf{A}\|_F = \sqrt{\text{trace}(\mathbf{A}^T \mathbf{A})}$$

Let \mathbf{C} be an $M \times M$ circulant matrix with first row $(c_0, c_1, c_2, \dots, c_{M-1}) \in \mathbb{R}^M$. Then by definition \mathbf{C}_{Σ} is given by $\arg \min_{\mathbf{C}} \|\mathbf{\Sigma} - \mathbf{C}\|_F$, or equivalently $\arg \min_{\mathbf{C}} \|\mathbf{\Sigma} - \mathbf{C}\|_F^2$. Taking the second of these, we have

$$\mathbf{C}_{\Sigma} = \arg \min_{\mathbf{C}} \|\mathbf{\Sigma} - \mathbf{C}\|_F^2. \quad (52)$$

Considering that

$$\begin{aligned} \|\mathbf{\Sigma} - \mathbf{C}\|_F^2 &= \text{trace} \left([\mathbf{\Sigma} - \mathbf{C}]^T [\mathbf{\Sigma} - \mathbf{C}] \right) \\ &= (\text{trace } \mathbf{\Sigma}^T \mathbf{\Sigma} - \text{trace } \mathbf{\Sigma}^T \mathbf{C} - \text{trace } \mathbf{C}^T \mathbf{\Sigma} + \text{trace } \mathbf{C}^T \mathbf{C}) \end{aligned}$$

and $\text{trace } \mathbf{\Sigma}^T \mathbf{\Sigma}$ is constant in \mathbf{C} , we need only consider minimizing

$$F(\mathbf{C}) = \text{trace } \mathbf{C}^T \mathbf{C} - \text{trace } \mathbf{\Sigma}^T \mathbf{C} - \text{trace } \mathbf{C}^T \mathbf{\Sigma}. \quad (53)$$

The first term of Equation (53) is straightforward to express in terms of the c_m . As \mathbf{C} is circulant,

$$\text{trace } \mathbf{C}^\top \mathbf{C} = M \sum_{m=0}^{M-1} c_m^2.$$

The other terms can be evaluated by expressing \mathbf{C} as a matrix polynomial.

Equation (48) and the symmetry of Σ allow us to write

$$\Sigma^\top \mathbf{C} = \Sigma \left(c_0 \mathbf{I} + \sum_{m=1}^{M-1} c_m \mathbf{P}^m \right) = c_0 \Sigma + \sum_{m=1}^{M-1} c_m \Sigma \mathbf{P}^m$$

and similarly

$$\mathbf{C}^\top \Sigma = c_0 \Sigma + \sum_{m=1}^{M-1} c_m (\mathbf{P}^m)^\top \Sigma = c_0 \Sigma + \sum_{m=1}^{M-1} c_m (\Sigma \mathbf{P}^m)^\top.$$

Next, consider

$$\text{trace } \Sigma \mathbf{P}^m = \text{trace } (\Sigma [\mathbf{e}_{M-m+1} | \mathbf{e}_{M-m+2} | \dots | \mathbf{e}_{M-m}]),$$

which can be evaluated by considering the k^{th} row of Σ , σ_k . The first k elements of this row are the descending sequence $\rho_{k-1}, \rho_{k-2}, \dots, \rho_0$, and the remaining elements are the ascending sequence $\rho_1, \rho_2, \dots, \rho_{M-k}$. Noting that the trace of a product of two matrices is simply a sum of the inner products of the rows of the first with the columns of the second, we obtain

$$\begin{aligned} \text{trace } \Sigma \mathbf{P}^m &= \sum_{k=1}^m \sigma_k^\top \mathbf{e}_{M-m+k} + \sum_{k=1}^{M-m} \sigma_{m+k}^\top \mathbf{e}_k \\ &= \sum_{k=1}^m \rho_{M-m} + \sum_{k=1}^{M-m} \rho_m \\ &= m \rho_{M-m} + (M-m) \rho_m. \end{aligned} \tag{54}$$

Equation (54) can then be substituted into Equation (53) using the decomposition of Equation (48) to give

$$F(\mathbf{C}) = M \sum_{m=0}^{M-1} c_m^2 - 2 \left(M c_0 \rho_0 + \sum_{m=1}^{M-1} c_m (m \rho_{M-m} + (M-m) \rho_m) \right). \tag{55}$$

As $\arg \min_{\mathbf{C}} F(\mathbf{C})$ is the same as $\arg \min_{\mathbf{C}} \|\mathbf{\Sigma} - \mathbf{C}\|_F$, we can now consider the values which minimize Equation (55) in order to find the nearest circulant matrix to $\mathbf{\Sigma}$. Taking

$$\frac{\partial}{\partial c_m} F(\mathbf{C}) = \begin{cases} 2Mc_0 - 2M\rho_0 & \text{for } m = 0, \\ 2Mc_m - 2(m\rho_{M-m} + (M-m)\rho_m) & \text{otherwise,} \end{cases}$$

and noting that the Hessian matrix is $2M\mathbf{I}$ and hence is positive definite so any solutions to $\arg \min F(\mathbf{C})$ must be minima, we obtain

$$c_m = \begin{cases} \rho_0 & \text{for } m = 0, \\ \frac{m}{M}\rho_{M-m} + \frac{M-m}{M}\rho_m & \text{otherwise,} \end{cases}$$

which can be re-expressed as

$$c_m = \begin{cases} \rho_0 & \text{for } m = 0, \\ \rho_m + \frac{m}{M}(\rho_{M-m} - \rho_m) & \text{otherwise,} \end{cases} \quad (56)$$

to make the relationship between c_m and ρ_m clearer. An important consequence of this system of equations is that $c_m = \frac{m}{M}\rho_{M-m} + \frac{M-m}{M}\rho_m = \frac{M-(M-m)}{M}\rho_{M-m} + \frac{M-m}{M}\rho_{M-(M-m)} = c_{M-m}$, and so \mathbf{C}_{Σ} is a symmetric circulant matrix with c_m defined as in Equation (56). Therefore, this nearest circulant will have only real eigenvalues. This is desirable for the multiple testing adjustment case of [1], [10], and [4], where the effective number of tests is a real number that is some function of the eigenvalues.

So the optimal decomposition in the Frobenius norm is

$$\mathbf{\Sigma} = \mathbf{C}_{\Sigma} + \mathbf{R}_{\Sigma} \quad (57)$$

where \mathbf{C}_{Σ} is the circulant matrix defined by Equation (56), that is

$$c_m = \begin{cases} \rho_0 & \text{for } m = 0, \\ \rho_m + \frac{m}{M}(\rho_{M-m} - \rho_m) & \text{otherwise,} \end{cases}$$

and $\mathbf{R}_{\Sigma} = \mathbf{\Sigma} - \mathbf{C}_{\Sigma}$, and so the value in the m^{th} off-diagonal of \mathbf{R}_{Σ} is $\rho_m - c_m = \rho_m - (\rho_m + \frac{m}{M}(\rho_{M-m} - \rho_m)) = \frac{m}{M}(\rho_m - \rho_{M-m})$.

The eigenvalues of \mathbf{C}_Σ are given by a substitution of Equation (56) into Equation (50), giving

$$\lambda_k(\mathbf{C}_\Sigma) = \rho_0 + 2 \sum_{m=1}^{M-1} \frac{M-m}{M} \rho_m \cos \frac{2\pi mk}{M}. \quad (58)$$

The corresponding eigenvectors are given by Equation (49).

Consider \mathbf{R}_Σ briefly. For large M and small m , $\frac{m}{M}(\rho_{M-m} - \rho_m) \approx 0$ while when m is close to M , $\frac{m}{M}(\rho_m - \rho_{M-m}) \approx \rho_m - \rho_{M-m}$. This implies that in the case of large M , \mathbf{R}_Σ will have vanishingly small values for the central off-diagonals, and values in the corners of approximately $\rho_m - \rho_{M-m}$.

In the particular case of Σ_e from Equation (43), $\rho_m = \rho^m$. In this case, \mathbf{C}_{Σ_e} has entries

$$c_m = \rho^m + \frac{m}{M}(\rho^{M-m} - \rho^m)$$

and \mathbf{R}_{Σ_e} is $\frac{m}{M}(\rho^m - \rho^{M-m})$ for the m^{th} off-diagonal where $m \in \{0, 1, 2, \dots, M-1\}$. The eigenvalues of \mathbf{C}_{Σ_e} are therefore

$$\lambda_k(\mathbf{C}_{\Sigma_e}) = 1 + 2 \sum_{m=1}^{M-1} \frac{M-m}{M} \rho^m \cos \frac{2\pi mk}{M}.$$

6 Asymptotic equivalence of \mathbf{C}_{Σ_e} and Σ_e

The approximation matrix \mathbf{C}_{Σ_e} has been derived here based purely on minimizing $\|\Sigma_e - \mathbf{C}_{\Sigma_e}\|_F$ without any of the asymptotic guarantees of [6]. [5] derives a similar result by considering the asymptotic equivalence of matrices in the weak norm. Matrices \mathbf{A} and \mathbf{B} in $\mathbb{C}^{M \times M}$ are said to be asymptotically equivalent in the weak norm if

$$\lim_{M \rightarrow \infty} \frac{1}{\sqrt{M}} \|A - B\|_F = 0.$$

Therefore, a natural consideration is the difference

$$\lim_{M \rightarrow \infty} \frac{1}{\sqrt{M}} \|\Sigma_e - \mathbf{C}_{\Sigma_e}\|_F = \lim_{M \rightarrow \infty} \frac{1}{\sqrt{M}} \|\mathbf{R}_{\Sigma_e}\|_F. \quad (59)$$

Before taking the limit, note

$$\frac{1}{\sqrt{M}} \|\mathbf{R}_{\Sigma_e}\|_F = \sqrt{\frac{1}{M} \text{trace } \mathbf{R}_{\Sigma_e}^\top \mathbf{R}_{\Sigma_e}},$$

which has the square

$$\begin{aligned}
\frac{1}{M} \|\mathbf{R}_{\Sigma_e}\|_F^2 &= \frac{1}{M} \text{trace } \mathbf{R}_{\Sigma_e}^\top \mathbf{R}_{\Sigma_e} \\
&= \frac{1}{M} \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} \frac{|i-j|^2}{M^2} (\rho^{2|i-j|} - 2\rho^{|i-j|+M-|i-j|} + \rho^{2(M-|i-j|)}) \\
&= \frac{2}{M^3} \sum_{m=1}^{M-1} (M-m)m^2 (\rho^{2m} - 2\rho^M + \rho^{2(M-m)})
\end{aligned} \tag{60}$$

Evaluating this expression for $\rho < 1$ is made easier by considering the general sum

$$\sum_{m=1}^M m^k \rho^m.$$

Theorem 3 (Truncated geometric power series). *The finite sum*

$$\sum_{n=1}^N n^k p^n$$

with $|p| < 1$ can be expressed as

$$\frac{p}{1-p} \left[(1-p^N) G^{(k)}(0) - p^N \sum_{l=1}^k \binom{k}{l} N^l G^{(k-l)}(0) \right]$$

where

$$G^{(k)}(0) = \left. \frac{d^k}{dt^k} E[e^{tX}] \right|_{t=0}$$

is the k^{th} derivative of $G(t) = E[e^{tX}]$ evaluated at zero for $X \sim \text{Geo}(1-p)$.

Proof. Multiplying by $\frac{1-p}{1-p}$, we obtain

$$\sum_{n=1}^N n^k p^n = \frac{p}{1-p} \sum_{n=1}^N n^k p^{n-1} (1-p).$$

Now

$$\begin{aligned}
\sum_{n=1}^N n^k p^{n-1} (1-p) &= \sum_{n=1}^{\infty} n^k p^{n-1} (1-p) - p^N \sum_{n=N+1}^{\infty} n^k p^{n-N-1} (1-p) \\
&= \sum_{n=1}^{\infty} n^k p^{n-1} (1-p) - p^N \sum_{n=1}^{\infty} (n+N)^k p^{n-1} (1-p) \\
&= \sum_{n=1}^{\infty} n^k p^{n-1} (1-p) - p^N \sum_{n=1}^{\infty} \sum_{l=0}^k \binom{k}{l} n^{k-l} N^l p^{n-1} (1-p) \\
&= (1-p^N) \sum_{n=1}^{\infty} n^k p^{n-1} (1-p) \\
&\quad - p^N \sum_{l=1}^k \binom{k}{l} N^l \sum_{m=1}^{\infty} n^{k-l} p^{n-1} (1-p), \tag{61}
\end{aligned}$$

but

$$\sum_{n=1}^{\infty} n^k p^{n-1} (1-p)$$

is just the k^{th} moment of a geometric distribution with a probability of success of $1-p$. This distribution has a moment-generating function

$$G(t) = \frac{1-p}{1-e^t p} \tag{62}$$

which can be used to evaluate the k^{th} moment by taking the k^{th} derivative and evaluating it at $t=0$, denoted

$$G^{(k)}(0)$$

and substituted into Equation 61 to give the more succinct

$$\sum_{n=1}^N n^k p^{n-1} (1-p) = (1-p^N) G^{(k)}(0) - p^N \sum_{l=1}^k \binom{k}{l} N^l G^{(k-l)}(0). \tag{63}$$

This gives the result for the original sum

$$\sum_{n=1}^N n^k p^n = \frac{p}{1-p} \left[(1-p^N) G^{(k)}(0) - p^N \sum_{l=1}^k \binom{k}{l} N^l G^{(k-l)}(0) \right]. \tag{64}$$

□

This result provides a representation of the sum $\sum_{n=1}^N n^k p^n$ in terms of the first k moments of the geometric distribution with parameter $1 - p$. Additionally, taking the limit $N \rightarrow \infty$ reduces this expression to the k^{th} moment of the geometric distribution with parameter $1 - p$ multiplied by $p/(1 - p)$. As the original sum is the k^{th} geometric moment truncated at N and scaled in the same way, this result is expected.

We can use this theorem by expanding Equation 60 to give

$$\begin{aligned}
\frac{1}{M} \|\mathbf{R}_{\Sigma_e}\|_F^2 &= \frac{2}{M^3} \sum_{m=1}^{M-1} (M - m) m^2 (\rho^{2m} - 2\rho^M + \rho^{2(M-m)}), \\
&= \frac{2}{M^2} \left[\sum_{m=1}^{M-1} (M^2 - 2Mm + 2m^2) \rho^{2m} - 2\rho^M \sum_{m=1}^{M-1} m^2 \right] \\
&\quad - \frac{2}{M^3} \left[\sum_{m=1}^{M-1} (M^3 - 3M^2m + 3Mm^2) \rho^{2m} - 2\rho^M \sum_{m=1}^{M-1} m^3 \right], \\
&= 2 \left[\sum_{m=1}^{M-1} \left(\frac{m}{M} - \frac{m^2}{M^2} \right) \rho^{2m} - 2\rho^M \left(\sum_{m=1}^{M-1} \frac{m^2}{M^2} - \frac{m^3}{M^3} \right) \right] \tag{65}
\end{aligned}$$

which has terms proportional to Equation 64 for $k = 1$ and 2. No further evaluation is necessary to see

$$\lim_{M \rightarrow \infty} \frac{1}{M} \|\mathbf{R}_{\Sigma_e}\|_F^2 = 0,$$

as these first two moments are finite for the geometric distribution and the final term is strictly less than $4M\rho^M$, which is zero in the limit. As a result we also get

$$\lim_{M \rightarrow \infty} \frac{1}{\sqrt{M}} \|\mathbf{R}_{\Sigma_e}\|_F = 0$$

and so \mathbf{C}_{Σ_e} is asymptotically equivalent to Σ_e .

The case of \mathbf{C}_{Σ} and Σ is less straightforward, as Equation 60 becomes

$$\frac{1}{M} \|\mathbf{R}_{\Sigma}\|_F^2 = \frac{2}{M^3} \sum_{m=1}^{M-1} (M - m) m^2 (\rho_m^2 - 2\rho_{M-m}\rho_m + \rho_{M-m}^2). \tag{66}$$

This expression will be zero in the limit $M \rightarrow \infty$ if the series ρ_m is absolutely convergent and the partial sums $\sum_{m=1}^{M-1} m^3 \rho_m$ and $\sum_{m=1}^{M-1} m^2 \rho_m$ are $O(M^2)$ and $O(M)$, respectively. If these conditions are met, then \mathbf{C}_{Σ} and Σ are asymptotically equivalent.

7 Comparing approximations

\mathbf{R}_{Σ_e} can be evaluated further by considering Equation 64 for $k = 1$ and 2 with $p = \rho^2$ and $N = M - 1$:

$$\sum_{m=1}^{M-1} m \rho^{2m} = \frac{\rho^2}{1 - \rho^2} \left[\frac{1 - \rho^{2(M-1)}}{1 - \rho^2} - \rho^{2(M-1)}(M - 1) \right], \quad (67)$$

$$\begin{aligned} \sum_{m=1}^{M-1} m^2 \rho^{2m} &= \frac{\rho^2}{1 - \rho^2} \left[\frac{(1 - \rho^{2(M-1)})(1 + \rho^2)}{(1 - \rho^2)^2} - \frac{2\rho^{2(M-1)}}{1 - \rho^2}(M - 1) \right. \\ &\quad \left. - \rho^{2(M-1)}(M - 1)^2 \right]. \end{aligned} \quad (68)$$

Substituting Equations 67 and 68 into Equation 65 gives

$$\begin{aligned} \sum_{m=1}^{M-1} \left(\frac{m}{M} - \frac{m^2}{M^2} \right) \rho^{2m} &= \frac{\rho^2}{M(1 - \rho^2)} \left[\frac{1 + \rho^{2(M-1)}}{1 - \rho^2} - \frac{1 + \rho^2 + \rho^{2(M-1)} - 3\rho^{2M}}{(1 - \rho^2)^2} \frac{1}{M} \right. \\ &\quad \left. - \rho^{2(M-1)} \left(1 - \frac{1}{M} \right) \right]. \end{aligned} \quad (69)$$

Simplifying

$$2\rho^M \left(\sum_{m=1}^{M-1} \frac{m^2}{M^2} - \frac{m^3}{M^3} \right)$$

comes by substituting the equations for the sums of squares and cubes of the first $M - 1$ integers:

$$\begin{aligned} 2\rho^M \left(\sum_{m=1}^{M-1} \frac{m^2}{M^2} - \frac{m^3}{M^3} \right) &= 2\rho^M \left(\frac{(M - 1)M(2M - 1)}{6M^2} - \frac{((M - 1)M)^2}{4M^3} \right) \\ &= 2\rho^M \left(\frac{2M^3 - 3M^2 + M}{6M^2} - \frac{M^4 - 2M^3 + M^2}{4M^3} \right) \\ &= \frac{1}{6}\rho^M \left(M - \frac{1}{M} \right) \end{aligned} \quad (70)$$

Taking the difference of Equations 69 and 70 gives an expression for the squared Frobenius norm of \mathbf{R}_{Σ_e} :

$$\begin{aligned} \frac{1}{M} \|\mathbf{R}_{\Sigma_e}\|_F^2 &= \frac{2\rho^2}{M(1-\rho^2)} \left[\frac{1+\rho^{2(M-1)}}{1-\rho^2} - \frac{1+\rho^2+\rho^{2(M-1)}-3\rho^{2M}}{M(1-\rho^2)^2} - \rho^{2(M-1)} \frac{M-1}{M} \right] \\ &\quad - \frac{1}{3} \rho^M \left(M - \frac{1}{M} \right) \end{aligned} \quad (71)$$

[5] suggests a slightly different approximation. Following [6], the sum

$$f(x) = \sum_{k=-\infty}^{\infty} \rho_k e^{ikx} = \sum_{k=-\infty}^{\infty} \rho^{|k|} e^{ikx}$$

is known to be critical to the approximation of Σ . [5] considers circulant entries generated using the expression

$$c_m = \begin{cases} \rho_0 & \text{for } m = 0, \\ \frac{1}{M} \sum_{j=0}^{M-1} f\left(\frac{2\pi j}{M}\right) e^{\frac{2\pi i j m}{M}} & \text{otherwise,} \end{cases} \quad (72)$$

the second case can be expressed

$$\begin{aligned} \frac{1}{M} \sum_{j=0}^{M-1} f\left(\frac{2\pi j}{M}\right) e^{\frac{2\pi i j m}{M}} &= \frac{1}{M} \sum_{j=0}^{M-1} \sum_{k=-\infty}^{\infty} \rho^{|k|} e^{ik \frac{2\pi j}{M}} e^{\frac{2\pi i j m}{M}} \\ &= \sum_{k=-\infty}^{\infty} \rho^{|k|} \frac{1}{M} \sum_{j=0}^{M-1} e^{\frac{2\pi i j}{M} (m+k)} \\ &= \sum_{k=-\infty}^{\infty} \rho^{|k|} I_{(m+k) \bmod M}(0) \end{aligned}$$

as the second sum is the sum of the squared M^{th} roots of unity, which are orthonormal.

Now

$$\begin{aligned}
\sum_{k=-\infty}^{\infty} \rho^{|k|} I_{(m+k) \bmod M}(0) &= \sum_{k=-\infty}^{\infty} \rho^{|-m+kM|} \\
&= \sum_{k=-\infty}^0 \rho^{m-kM} + \sum_{k=1}^{\infty} \rho^{-m+kM} \\
&= \rho^m \frac{1}{1-\rho^M} + \rho^{-m} \frac{\rho^M}{1-\rho^M} \\
&= \frac{1}{1-\rho^M} (\rho^m + \rho^{M-m})
\end{aligned}$$

when $\rho < 1$. Therefore, Equation 72 becomes

$$c_m = \begin{cases} \rho_0 & \text{for } m = 0, \\ \frac{1}{1-\rho^M} (\rho^m + \rho^{M-m}) & \text{otherwise,} \end{cases} \quad (73)$$

in the particular case of $\rho_m = \rho^m$. So, while \mathbf{C}_{Σ_e} has entries which are a weighted average of ρ^m and ρ^{M-m} , this approximation instead takes a sum scaled by $1 - \rho^M$. Define \mathbf{C}_{GS} as the circulant matrix with these entries, and let \mathbf{R}_{GS} be $\Sigma_e - \mathbf{C}_{GS}$. Then

$$\frac{1}{M} \|\mathbf{R}_{GS}\|_F^2 = \frac{2}{(1-\rho^M)^2} \left[\frac{\rho^2(1-\rho^{2M})^2}{(1-\rho^2)^2} \frac{1}{M} + (M-2)\rho^{2M} \right]. \quad (74)$$

As $\rho^M = O(1/M^k)$ for $0 \leq \rho < 1$ and $k < \infty$, Equation 74 gives

$$\frac{1}{\sqrt{M}} \|\mathbf{R}_{GS}\|_F = \frac{\sqrt{2}\rho}{(1-\rho^2)(1-\rho^M)} \frac{1}{\sqrt{M}} + O\left(\frac{1}{M}\right) \quad (75)$$

while the same logic applied to Equation 71 gives

$$\frac{1}{\sqrt{M}} \|\mathbf{R}_{\Sigma_e}\|_F = \frac{\sqrt{2}\rho}{1-\rho^2} \frac{1}{\sqrt{M}} + O\left(\frac{1}{M}\right) \quad (76)$$

The leading terms of which give the relationship for large M

$$\frac{1}{\sqrt{M}} \|\mathbf{R}_{\Sigma_e}\|_F \approx (1-\rho^M) \frac{1}{\sqrt{M}} \|\mathbf{R}_{GS}\|_F.$$

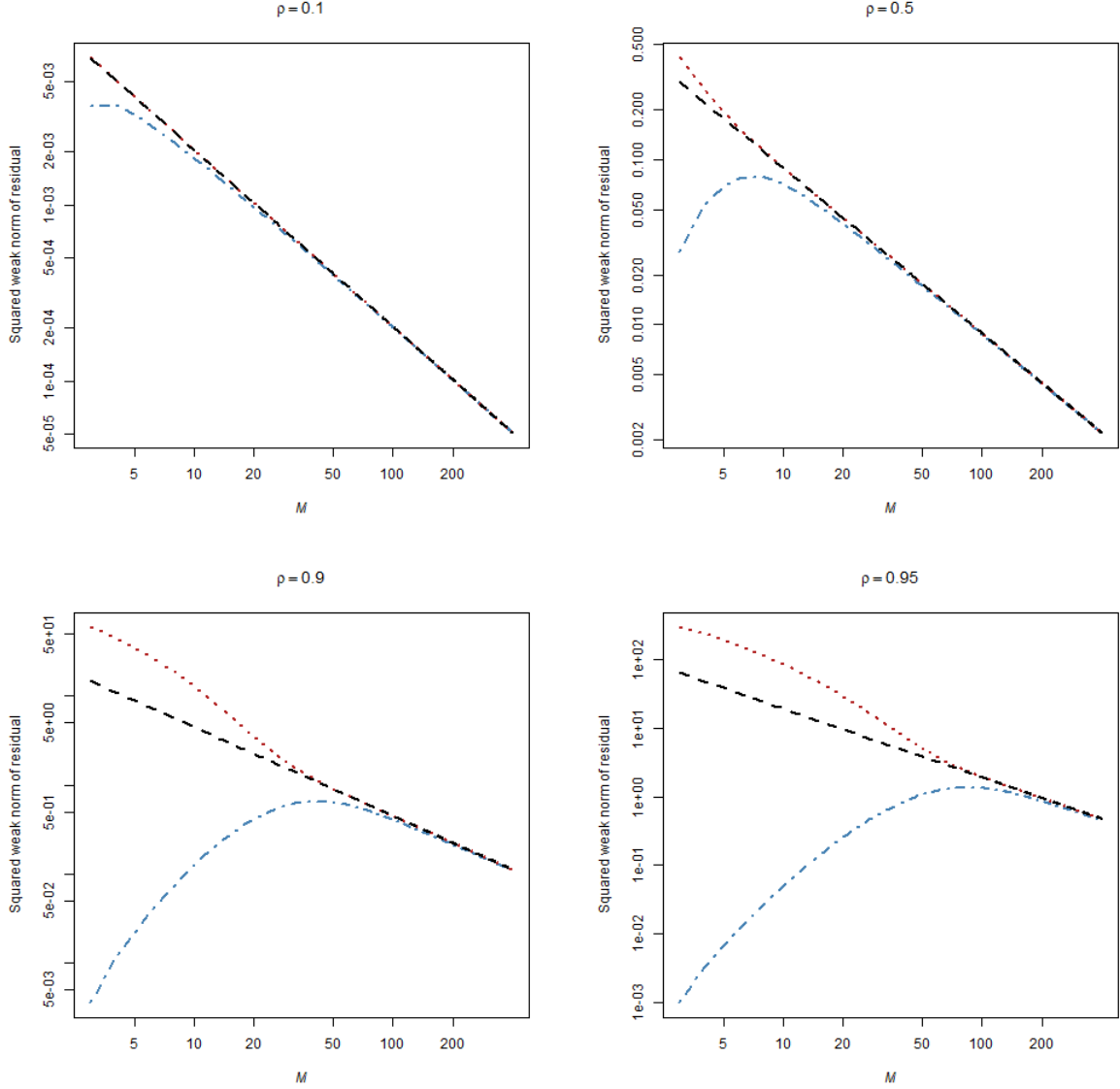


Figure 8: Log-scale plots of the squared Frobenius norms of the residuals \mathbf{R}_{Σ_e} and \mathbf{R}_{GS} divided by M for a selection of ρ values over $M = 3$ to 400. The blue dotted line corresponds to $\log \frac{1}{M} \|\mathbf{R}_{\Sigma_e}\|_F^2$, the red dot-dashed line to $\log \frac{1}{M} \|\mathbf{R}_{GS}\|_F^2$, and the black dashed line to the log of the leading term given by Equation 76

As $1 - \rho^M < 1$, \mathbf{R}_{Σ_e} converges to zero faster than \mathbf{R}_{GS} with a greater difference in convergence for ρ values close to 1. A comparison of the values of Equations 71 and 74 is shown in Figure 8.

The nearest circulant approximation performs considerably better than the approximation suggested by [5]. Using the leading term line as a guide, the nearest circulant matrix is uniformly less than this leading term, while the other is greater than the leading term. Moreover, while the nearest circulant improves in quality for small M as ρ nears 1, the approximation of [5] performs poorer relative to the leading term. This makes sense, as for ρ close to one, the absolute difference between ρ^m and $\rho^{(M-m)}$ decreases, making the weighted average entries of \mathbf{C}_{Σ_e} closer to the true values. In contrast, the other approximation adds these values and scales them, and so generally has residual terms with larger magnitude. The nearest circulant is orders of magnitude closer in the Frobenius norm for small M in these cases.

8 Conclusion

We have derived here a circulant approximation of Σ , \mathbf{C}_{Σ} , which is optimal in the sense of the Frobenius norm. \mathbf{C}_{Σ} is also symmetric, and so is guaranteed to have real eigenvalues. For applications such as multiple test adjustment, this is a highly desirable quality. Additionally, it was proven that only symmetric real circulants are guaranteed to have this property. Beyond multiple test adjustment, any application requiring the eigenvalues of a Toeplitz matrix can use \mathbf{C}_{Σ} for its optimality and guaranteed real eigenvalues.

\mathbf{C}_{Σ} is also much better than classic approximation provided in [6] in the case of Σ_e , where $\rho_m = \rho^m$. In particular, as ρ approaches one, \mathbf{C}_{Σ_e} performs orders of magnitude better. In the limit of large M , both approximations have the same behaviour.

In demonstrating this improved behaviour, another general result was demonstrated. It was proven that the sum $\sum_{n=1}^N n^k p^n$ can be expressed as a linear combination of the first k geometric moments. This expression was provided explicitly for interested readers.

References

- [1] James M Cheverud. A simple correction for multiple comparisons in interval mapping genome scans. *Heredity*, 87(1):52–58, 2001.

- [2] Ozan Cinar and Wolfgang Viechtbauer. The poolr package for combining independent and dependent p values. *Journal of Statistical Software*, 101:1–42, 2022.
- [3] Robert D. Cousins. Annotated bibliography of some papers on combining significances or p -values, 2008.
- [4] Nicholas W Galwey. A new measure of the effective number of tests, a practical tool for comparing families of non-independent significance tests. *Genetic Epidemiology*, 33(7):559–568, 2009.
- [5] Robert M Gray. *Toeplitz and circulant matrices: A review*. now Publishers Inc., 2006.
- [6] Ulf Grenander and Gabor Szegö. *Toeplitz forms and their applications*. University of California Press, 1958.
- [7] JBS Haldane. The combination of linkage values and the calculation of distances between the loci of linked factors. *Journal of Genetics*, 8(29):299–309, 1919.
- [8] Thomas LaFramboise. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic acids research*, 37(13):4181–4193, 2009.
- [9] Eric S Lander and David Botstein. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, 121(1):185–199, 1989.
- [10] J Li and L Ji. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity*, 95(3):221–227, 2005.
- [11] Ying Liu, Victor de la Pena, and Tian Zheng. Kernel-based measures of association. *WIREs Computational Statistics*, 10(2), 2018.
- [12] Dale R Nyholt. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *The American Journal of Human Genetics*, 74(4):765–769, 2004.
- [13] R.W. Oldford. On the n -dimensional geometry of regression diagnostics. *Communications in Statistics-Theory and Methods*, 16(9):2517–2540, 1987.
- [14] David N Reshef, Yakir A Reshef, Hilary K Finucane, Sharon R Grossman, Gilean McVean, Peter J Turnbaugh, Eric S Lander, Michael Mitzenmacher, and Pardis C Sabeti. Detecting novel associations in large data sets. *Science*, 334(6062):1518–1524, 2011.

- [15] Christopher Salathé. A structural model of genome-wide association studies, 2022.
- [16] Emil Uffelmann, Qin Qin Huang, Nchangwi Syntia Munung, Jantina de Vries, Yukinori Okada, Alicia R Martin, Hilary C Martin, Tuuli Lappalainen, and Danielle Posthuma. Genome-wide association studies. *Nature Reviews Methods Primers*, 1(1):1–21, 2021.