

A penalized likelihood approach to pooling estimates of covariance components from analyses by parts

Karin Meyer

Animal Genetics and Breeding Unit^A, University of New England, Armidale NSW 2351, Australia

Short title: Pooling part analyses

1 Introduction

2 Estimation of variance and covariance components due to genetic effects and other sources of variation
3 is one of the prime tasks in quantitative genetics. Increasingly, there are substantial numbers of traits
4 that need to be considered simultaneously and corresponding estimates of covariance matrices are
5 required. For instance, genetic evaluation schemes for beef cattle and sheep in Australia currently
6 involve multivariate analyses with up to 60 traits [14, 9]. Technical advances have made direct
7 estimation of covariance matrices comprising some 20 to 30 traits feasible for simple models [e.g 42].
8 However, such undertakings are computationally highly demanding, can be afflicted by convergence
9 problems and limits on the number of traits are generally lower when additional random factors –
10 such as maternal genetic and permanent environmental effects – need to be taken into account.

11 Hence complete covariance matrices for all traits of interest are routinely constructed from estimates
12 obtained by considering overlapping subsets of traits. In the simplest scenario, this may involve
13 analyses for all pairs of traits, yielding a single estimate of each covariance but multiple estimates for
14 all variance components. In other cases, some traits are measured on selected individuals only with
15 selection based on other traits of interest and tri- or higher-variate analyses including the selection
16 criteria as ‘anchor’ traits are needed to counteract the effects of selection bias. Procedures commonly
17 used to pool the resulting sets of estimates are typically somewhat *ad hoc*, and differ in their scope to
18 account for unequal reliabilities of estimates based on diverse numbers of traits, to allow for sampling
19 covariances and to ensure that the resulting, pooled matrix is within the parameter space, i.e. does
20 not have any negative eigenvalues. Few comparisons of the sampling properties of pooled matrices
21 exist and no guidelines are available on which approaches are preferable for different scenarios.

22 The need to modify non-positive definite or ill-conditioned covariance or correlation matrices arises
23 in a number of areas outside of quantitative genetics, in particular behavioural sciences and finance.
24 Most methods in use to perform this task consider a single matrix at a time. Techniques utilised
25 range from simple manipulation of matrix eigenvalues to generalized least-squares and maximum
26 likelihood (ML) based procedures. Generally, the aim is to replace the original matrix with its ‘nearest’
27 approximation.

28 It is well known that sampling variances increase dramatically with the number of parameters to be
29 estimated. For quantitative genetics problems, this is exacerbated by the need to partition the overall,

^Aa joint venture with NSW Department of Primary Industries

phenotypic covariances into their causal components. There has been considerable interest recently in the use of so-called regularization techniques in the estimation of covariance matrices. Typically, these yield ‘improved’ estimators, i.e. estimators which are on average closer to the population values than their non-regularized counterparts, by reducing sampling variation at the expense of introducing some additional bias. Again, applications span a number of areas; see Meyer and Kirkpatrick [30] for a review and references.

For genetic problems, penalized restricted maximum likelihood (REML) estimation for multivariate analyses comprising more than a few traits has been shown to provide substantially improved estimates of covariance matrices. Penalties employed were designed to ‘borrow strength’ from the phenotypic covariance matrix, with the rationale that, due to strong, negative sampling correlations, this sum is estimated much more accurately than its constituents [30, 29]. In a similar vein, we propose to employ a maximum likelihood approach, subject to corresponding penalties, to pool estimates from part analyses whilst simultaneously applying some regularization. A simulation study examining the sampling properties of the resulting estimates of covariance matrices is presented, showing that such procedure can yield considerably ‘better’ pooled estimates than methods currently in use.

Background

Likelihood analyses by parts

Multivariate analyses by parts typically involve overlapping subsets of traits and the resulting estimates from individual analyses are thus correlated. However, when pooling results from part analyses, estimates from different subsets are generally treated as independent. Recently, there has been considerable interest in so-called composite likelihood (CL) estimation for complex analyses where calculation of the likelihood is difficult or computationally infeasible; see Varin *et al.* [43] for a comprehensive review. In brief, a composite likelihood function approximates the likelihood from a weighted product of partial likelihoods, where the latter may be marginal or conditional likelihoods of subsets, ignoring dependencies between parts. A particular, marginal CL for multivariate analyses of q traits is the ‘pair-wise’ likelihood, consisting of the product of likelihoods for all $q(q-1)/2$ pairs of traits. Some types of CL are also referred to as pseudo-likelihood.

Maximum pseudo or composite likelihood estimators are generally consistent and asymptotically normal, though they tend to make less efficient use of the data than standard likelihood estimators [43]. A number of simulation studies are available investigating the loss in efficiency for estimates of covariance components, showing that reductions in efficiency are often relatively small [26, 10, 3].

While most CL applications approximate the likelihood by parts and estimate the same parameters as in a full analysis, others invoke the same principles for scenarios in which a number of separate analyses yield estimates for overlapping subsets of parameters. Fieuws and Verbeke [13] considered estimation of a covariance matrix using bivariate analyses for all pairs of traits. They emphasized that this fitted into the pseudo-likelihood framework, and showed that approximate sampling variances could be obtained combining derivatives of the likelihood from part analyses. However, when

67 combining estimates, Fieuws and Verbeke [13] relied on simple averages over pairs, ignoring the
68 possibility that the resulting pooled covariance matrix may be non-positive definite.

69 Ducrocq [12] summarized (unpublished) work in a quantitative genetics context by Yerex [45],
70 which showed that ignoring dependencies between subsets of data in the estimation of variance
71 components by maximum likelihood yielded similar results than attempts to take these into account.
72 This has been used in a number of studies to justify both accumulating contributions to the likelihood
73 from subsets of data and pooling of results from subsets of traits.

74 *Procedures to make covariance matrices positive-definite*

75 Most literature on replacing a non-positive definite covariance matrix with a positive definite counter-
76 part deals with a single covariance matrix only. Early methods suggested range from augmentation
77 of the diagonal elements by addition a small multiple of an identity matrix to ‘smoothing’ procedures
78 using a generalised or ordinary least squares approach [36, 23, 39]. Others relied on the modification
79 of eigenvalues. Higham [18] showed that the ‘nearest’ positive semi-definite matrix – defined as
80 the matrix for which the Frobenius norm of the difference to the original matrix is minimized – is
81 obtained by replacing any eigenvalues less than zero with this value.

82 Similar techniques have been applied to correlation matrices [35, 34, 19, 4]. For these, modification
83 subject to minimizing a nearness criterion is more complicated as diagonal elements need to remain
84 at unity, and various algorithms for constrained matrix approximation have been described recently
85 [32, 6, 37]. Others dealt with the added problems arising from weighted approximation, adherence
86 to some linear equalities or a given structure of the matrix [8, 7, 33].

87 Considering multivariate maximum likelihood estimation for a one-way classification, Amemiya
88 [1] suggested to truncate the canonical eigenvalues at zero (or a small positive value) to ensure that
89 estimates of the ‘between’ groups covariance matrix were positive (semi-)definite. While this used
90 the canonical eigenvalues of the matrix of within and between mean squares and cross-products, in a
91 genetic context, modification of these values is equivalent to modifying the canonical eigenvalues of
92 the phenotypic and genetic covariance matrix [16]. Amemiya [1] showed that estimates constrained
93 to the parameter space in this way were REML estimates.

94 It is well known that the eigenvalues of estimated covariance matrices are over-dispersed due to
95 sampling variation while their mean is expected to be unbiased [25]. Hence, a number of ‘improved’
96 estimators of covariance matrices have been suggested which reduce sampling variance based on
97 modifications of the estimated eigenvalues. Hayes and Hill [16] proposed to shrink the canonical
98 eigenvalues towards their mean to reduce their spread. This yields a modified estimate of the genetic
99 covariance matrix which is a weighted combination of the unmodified estimate and the phenotypic
100 covariance matrix scaled by the average heritability. The authors thus described their procedure as
101 ‘bending’ the genetic towards the phenotypic covariance matrix. They showed in a simulation study
102 that use of such modified estimates to determine the weights in a selection index could substantially
103 increase the achieved response to selection.

104 However, ‘bending’ as a technique for regularized estimation has found little practical use. One

of the problems in multivariate analyses of variance involving more than a few traits is the high probability of obtaining estimates of the genetic covariance matrix which are not positive definite [20]. ‘Bending’ just far enough so that the smallest, ‘bent’ eigenvalue was equal to a small positive value was thus suggested as a method to modify estimates to be positive definite [16]. While early applications [e.g. 2, 15] modified the canonical eigenvalues as originally proposed, later ‘bending’ became synonymous to shrinkage of eigenvalues of a single matrix [e.g. 22], which is equivalent to ‘bending’ towards an identity matrix.

Likelihood based approaches to pool estimates of covariance components from part analyses have been suggested by Mäntysaari [28] and Thompson *et al.* [41], though neither give details for their underlying rationale. The method of ‘iterative summing of expanded part matrices’ (Mäntysaari [28]; see also Koivula *et al.* [24]) involves an expectation-maximization type of procedure, and is described by

$$\hat{\Sigma}^{t+1} = \sum_{i=1}^S \hat{\Sigma}^t (\mathbf{Q}_i \hat{\Sigma}^t \mathbf{Q}_i)^{-} \mathbf{P}_i \mathbf{S}_i \mathbf{P}_i' (\mathbf{Q}_i \hat{\Sigma}^t \mathbf{Q}_i)^{-} \hat{\Sigma}^t + [(\mathbf{I}_q - \mathbf{Q}_i) (\hat{\Sigma}^t)^{-1} (\mathbf{I}_q - \mathbf{Q}_i)]^{-}$$

with $\hat{\Sigma}^t$ the $q \times q$ pooled covariance matrix at iterate t , \mathbf{S}_i , of size $s_i \times s_i$, the matrix of estimates from the i -th part-analysis, \mathbf{I}_q an identity matrix of size q , \mathbf{P}_i , of size $q \times s_i$ a matrix with elements of unity in position j, k if the k -th trait in the i -th analysis is the j -th trait overall and zero otherwise, and $\mathbf{Q}_i = \mathbf{P}_i \mathbf{P}_i'$. This is readily modified to allow for different weights for estimates from each part analysis [17].

Thompson *et al.* [41] suggested to transform estimates from part-analyses into pseudo observations which are then used in a linear model analysis to estimate the pooled covariance matrix. This is appealing, as it allows any suitable linear model software to be used. Moreover, any features for weighted estimation or to constrain estimates to have a chosen structure available in such software can be exploited. However, the outline given lacked technical details and applications have thus been limited [44, 38]. In brief, Thompson *et al.* [41]’s procedure can be derived treating estimates \mathbf{S}_i from part analyses as independent matrices of corrected mean squares and crossproducts. The pseudo observations are then chosen so as to recreate this matrix in the data part of the likelihood function. This is done by converting each \mathbf{S}_i into ‘observations’ on s_i unrelated individuals, where records for the j -th individual are comprised of the j -th column of $\sqrt{s_i} \mathbf{S}_i^{\frac{1}{2}}$. These data are then analysed in a linear ‘model’ without any fixed or random effects (not even means) estimating $\hat{\Sigma}$ as the within subject covariance matrix (note also that the software used should not automatically center observations). Thompson *et al.* [41] advocated the Cholesky factorization of \mathbf{S}_i (and omitted the factor $\sqrt{s_i}$ - this factor is needed to mimic the appropriate degrees in the likelihood function), but any matrix square root is suitable.

Penalized maximum likelihood for pooling covariance components

Consider a vector of observations \mathbf{y} for q traits from a multivariate normal distribution, $\mathbf{y} \sim N(\mathbf{X}\mathbf{b}, \mathbf{V})$, with \mathbf{V} the covariance matrix of \mathbf{y} , \mathbf{b} a vector of fixed effects and \mathbf{X} the corresponding design matrix. If \mathbf{y} can be split into independent parts (\mathbf{y}_i), e.g. observations for independent families, \mathbf{V} is block-diagonal and the pertaining REML likelihood on the logarithmic scale ($\log \mathcal{L}$) can be calculated by summing contributions over groups [40]

$$-2 \log \mathcal{L} \propto \sum_i d_i \left(\log |\mathbf{V}_i| + \text{tr}(\mathbf{V}_i^{-1} \mathbf{M}_i) \right) \quad (1)$$

with \mathbf{V}_i denoting the diagonal block of \mathbf{V} for the i -th group, $\mathbf{M}_i = (\mathbf{y}_i - \mathbf{X}_i \hat{\mathbf{b}})(\mathbf{y}_i - \mathbf{X}_i \hat{\mathbf{b}})' / d_i$ the corresponding matrix of corrected mean squares and crossproducts, and d_i the pertaining degrees of freedom.

Following Thompson *et al.* [41], assume that estimates for covariances components from individual, part analyses, represent such matrices of mean squares and products. For genetic analyses, we have several random effects and covariance matrices to be considered. Let \mathbf{S}_i^x denote the estimated covariance for random effect x from the i -th part analysis. For a simple animal model, for instance, $x = G, E$ with G and E standing for genetic and residual effects, respectively. Part estimates \mathbf{S}_i^x have expectations Σ_i^x , the submatrices of the corresponding covariance matrices for all traits, Σ_x , comprised of the covariance components among the traits represented in the i -th subset of traits. Estimates of Σ_x can then be obtained by replacing \mathbf{M}_i with \mathbf{S}_i^x and \mathbf{V}_i with Σ_i^x in (Eq. 1) above, and maximising the resulting log likelihood whilst constraining $\hat{\Sigma}_x$ to be positive definite.

There are several alternatives for dealing with multiple sources of variation. In the simplest scenario, all \mathbf{S}_i^x are treated as independent. We can then maximize $\log \mathcal{L}$ for each Σ_x independently or their sum

$$-2 \log \mathcal{L} \propto \sum_i d_i \sum_x \left(\log |\Sigma_i^x| + \text{tr}((\Sigma_i^x)^{-1} \mathbf{S}_i^x) \right) \quad (2)$$

with d_i the assumed degrees of freedom for the i -th part analysis. Different d_i can be selected for a weighted analysis, for instance, if numbers of records in individual analyses differ substantially. However, this approach ignores strong, negative sampling correlations between estimates \mathbf{S}_i^x from the same analysis.

To mimic some dependence between the \mathbf{S}_i^x for the same i , we propose to assume a simple, pseudo pedigree structure and construct corresponding matrices of mean squares and products and the log likelihood under this assumption. Let this pseudo structure involves families with m members each. Define matrices \mathbf{C}^x (of size $m \times m$) which give the coefficients for the x -th covariance component in the expectation of covariances between family members. This gives

$$-2 \log \mathcal{L} \propto \sum_i d_i \left(\log \left| \sum_x \mathbf{C}^x \otimes \Sigma_i^x \right| + \text{tr} \left(\left[\sum_x \mathbf{C}^x \otimes \Sigma_i^x \right]^{-1} \left[\sum_x \mathbf{C}^x \otimes \mathbf{S}_i^x \right] \right) \right) \quad (3)$$

with \otimes denoting the direct matrix product.

Selecting pseudo pedigree structures

Suitable choices for the pseudo structure depend on the number and type of covariance matrices to be pooled. The structure chosen should comprise a sufficient number of different kinds of covariances between relatives so that, when equating these to their expectations, we can separate all components. For a simple animal model with $x = G, E$, a paternal half-sib design, with each pseudo family consisting of m paternal half-sibs, is sufficient. Coefficient matrices for this case are $\mathbf{C}^E = \mathbf{I}_m$ and $\mathbf{C}^G = \frac{1}{4}\mathbf{J}_m + \frac{3}{4}\mathbf{I}_m$ (with \mathbf{J}_m a matrix of size $m \times m$ with all elements equal to unity). If common environmental covariances between full-sibs need to be considered in addition ($x = G, C, E$ with Σ_C the environmental covariance matrix), a hierarchical full-sib design with n offspring per dam and $k = m/n$ dams per sire may be appropriate. For offspring grouped within dam, coefficient matrices are then $\mathbf{C}^E = \mathbf{I}_m$, $\mathbf{C}^C = \mathbf{I}_k \otimes \mathbf{J}_n$ and $\mathbf{C}^G = \frac{1}{4}\mathbf{J}_m + \mathbf{I}_k \otimes \left(\frac{1}{4}\mathbf{J}_n + \frac{1}{2}\mathbf{I}_n\right)$. For analyses with maternal genetic effects, the pseudo family should include data on at least two generations. e.g. a sire mated to two unrelated dams with two offspring per dam with records on all individuals ($m = 7$) provides sire- and dam-offspring as well as full- and half-sib covariances.

Calculations in (Eq. 3) involve matrices of size $ms_i \times ms_i$. For analysis of variance type pseudo structures, this can be reduced to $s_i \times s_i$ by expressing $\log \mathcal{L}$ in terms of the matrices of mean squares and crossproducts between and within groups and their expectations. For the paternal half-sib pseudo structure with s sires and n progeny per sire,

$$-2 \log \mathcal{L} \propto (s-1) \sum_i \log |\mathbf{V}_{Bi}| + \text{tr}(\mathbf{V}_{Bi}^{-1} \mathbf{M}_{Bi}) + s(n-1) \sum_i \log |\mathbf{V}_{Wi}| + \text{tr}(\mathbf{V}_{Wi}^{-1} \mathbf{M}_{Wi})$$

with $\mathbf{M}_{Wi} = \mathbf{S}_i^E + \frac{3}{4}\mathbf{S}_i^G$, $\mathbf{M}_{Bi} = \mathbf{M}_{Wi} + \frac{1}{4}n\mathbf{S}_i^G$, $\mathbf{V}_{Wi} = \Sigma_i^E + \frac{3}{4}\Sigma_i^G$ and $\mathbf{V}_{Bi} = \mathbf{V}_{Wi} + \frac{1}{4}n\Sigma_i^G$. Similarly, for a hierarchical full-sib design with k dams per sire and n progeny per dam,

$$\begin{aligned} -2 \log \mathcal{L} \propto & (s-1) \sum_i \log |\mathbf{V}_{Bi}| + \text{tr}(\mathbf{V}_{Bi}^{-1} \mathbf{M}_{Bi}) + s(k-1) \sum_i \log |\mathbf{V}_{Di}| + \text{tr}(\mathbf{V}_{Di}^{-1} \mathbf{M}_{Di}) \\ & + sk(n-1) \sum_i \log |\mathbf{V}_{Wi}| + \text{tr}(\mathbf{V}_{Wi}^{-1} \mathbf{M}_{Wi}) \end{aligned}$$

with $\mathbf{M}_{Wi} = \mathbf{S}_i^E + \frac{1}{2}\mathbf{S}_i^G$, $\mathbf{M}_{Di} = \mathbf{M}_{Wi} + n\left(\frac{1}{4}\mathbf{S}_i^G + \mathbf{S}_i^C\right)$, $\mathbf{M}_{Bi} = \mathbf{M}_{Di} + \frac{1}{4}nk\mathbf{S}_i^G$, $\mathbf{V}_{Wi} = \Sigma_i^E + \frac{1}{2}\Sigma_i^G$, $\mathbf{V}_{Di} = \mathbf{V}_{Wi} + n\left(\frac{1}{4}\Sigma_i^G + \Sigma_i^C\right)$ and $\mathbf{V}_{Bi} = \mathbf{V}_{Di} + \frac{1}{4}nk\Sigma_i^G$.

Penalizing the likelihood

Using a likelihood approach not only affords the opportunity to pool estimates for all sources of variation simultaneously with constraints on the parameter space and, if desired, on the structure of the pooled matrices, but it also provides a framework for regularized estimation. This is achieved by maximizing the likelihood subject a penalty aimed at reducing sampling variation and thus to improve estimates. Meyer and Kirkpatrick [30] and Meyer [29] demonstrated in extensive simulation

204 studies of full, multivariate analyses that imposing penalties designed to ‘borrow strength’ from es-
 205 timates of the phenotypic covariance matrix could result in estimates of genetic covariance matrices
 206 with substantially reduced loss, i.e. on average much closer to the population values than unpenal-
 207 ized estimates. Penalties considered acted by either shrinking estimates of covariance matrices for
 208 individual random effects towards their sum or by reducing the spread of estimated canonical eigen-
 209 values, analogous to ‘bending’. Corresponding penalties can be employed when pooling estimates
 210 from part analyses by maximizing

$$211 \quad \log \mathcal{L}_P = \log \mathcal{L} - \frac{1}{2} \psi \mathcal{P} \quad (4)$$

212 instead of $\log \mathcal{L}$, with \mathcal{P} denoting the penalty ($\mathcal{P} > 0$) and ψ a so-called tuning factor ($\psi \geq 0$),
 213 determining the emphasis to be given to \mathcal{P} .

214 *Maximizing the likelihood*

215 Any of the optimization techniques and parameterizations commonly used in standard, ‘full’ REML
 216 estimation is suitable to maximize $\log \mathcal{L}_P$. A method of scoring algorithm is given by

$$217 \quad \hat{\boldsymbol{\theta}}^{t+1} = \hat{\boldsymbol{\theta}}^t - \alpha \mathbf{H}^{-1} \mathbf{g} \quad (5)$$

218 with $\hat{\boldsymbol{\theta}}^t$ the vector of parameters to be estimated at iterate t , \mathbf{g} the vector of gradients and \mathbf{H} the
 219 expected information matrix (both evaluated at $\hat{\boldsymbol{\theta}}^t$). The scaling factor $0 < \alpha \leq 1$ is used select
 220 search steps which increase $\log \mathcal{L}_P$. This can be determined simply by successive step halving or a
 221 back-tracking line search [11]. A suitable parameterisation to avoid estimates out of the parameter
 222 space is to estimate the elements of the Cholesky factors of the covariance matrices, taking logarithms
 223 of the diagonal values.

224 For parameters θ_r and θ_s , elements of \mathbf{g} and \mathbf{H} are

$$225 \quad -\frac{1}{2} \sum_i d_i \left\{ \text{tr} \left(\left[\sum_x \mathbf{C}^x \otimes \boldsymbol{\Sigma}_i^x \right]^{-1} \left[\sum_x \mathbf{C}^x \otimes \frac{\partial \boldsymbol{\Sigma}_i^x}{\partial \theta_r} \right] \right) \right. \\ 226 \quad \left. - \text{tr} \left(\left[\sum_x \mathbf{C}^x \otimes \boldsymbol{\Sigma}_i^x \right]^{-1} \left[\sum_x \mathbf{C}^x \otimes \frac{\partial \boldsymbol{\Sigma}_i^x}{\partial \theta_r} \right] \left[\sum_x \mathbf{C}^x \otimes \boldsymbol{\Sigma}_i^x \right]^{-1} \left[\sum_x \mathbf{C}^x \otimes \mathbf{S}_i^x \right] \right) \right\} - \frac{1}{2} \psi \frac{\partial \mathcal{P}}{\partial \theta_r}$$

228 and

$$229 \quad -\frac{1}{2} \sum_i d_i \text{tr} \left(\left[\sum_x \mathbf{C}^x \otimes \boldsymbol{\Sigma}_i^x \right]^{-1} \left[\sum_x \mathbf{C}^x \otimes \frac{\partial \boldsymbol{\Sigma}_i^x}{\partial \theta_r} \right] \left[\sum_x \mathbf{C}^x \otimes \boldsymbol{\Sigma}_i^x \right]^{-1} \left[\sum_x \mathbf{C}^x \otimes \frac{\partial \boldsymbol{\Sigma}_i^x}{\partial \theta_s} \right] \right) - \frac{1}{2} \psi \frac{\partial^2 \mathcal{P}}{\partial \theta_r \partial \theta_s}$$

230

Material and methods

Simulation set-up

Data for $q = 5, 10$ and 20 traits and three different pedigree structures were simulated by sampling random effects from multivariate normal distributions. No fixed effects were considered. Population values were obtained by combining three levels of heritabilities with four constellations of correlations and variances. For cases I and II, all heritabilities were assumed to be equal, $h_i^2 = 0.4$ and $h_i^2 = 0.2$ for $i = 1, q$, respectively. For case III, values were $h_i^2 = 0.2 + 0.1(5 - \text{mod}(i, 5))$ (with $\text{mod}(a, b)$ the remainder of a/b), i.e. sets of $0.6, 0.5, 0.4, 0.3$ and 0.2 repeated $q/5$ times. Scenario (a) assumed both genetic (r_{Gij}) and residual correlations (r_{Eij}) between traits i and j were zero for all $i \neq j = 1, q$ and that all phenotypic variances were equal, $\sigma_{P_i}^2 = 1$. For scenario (b), correlations among all traits were again equal, with $r_{Gij} = 0.5$ and $r_{Eij} = 0.2$. For scenarios (c) and (d), correlations were modelled depending on the trait numbers, $r_{Gij} = 0.7^{|j-i|}$ and $r_{Eij} = 0.5 + (-0.2)^{|j-i|}$ for case (c) and $r_{Gij} = 0.3 + (-0.3)^{|j-i|}$ and $r_{Eij} = (-0.3)^{|j-i|}$ for case (d). Variances for constellations 2 to 4 were $\sigma_{P_i}^2 = 1 + \text{mod}(3, i)$ for trait i . This yielded 12 sets of population values, referred to as Ia to IIIId in the following.

Data were assumed to originate from f independent families. For the first part, a simple model with additive genetic and residual effects as the only fixed effects were simulated either for a paternal half-sib structure (PHS) with 10 progeny per sire family or Bondari's design (BON). The latter, due to Bondari *et al.* [5], comprises 8 individuals per family: In generation 1, records are taken on two pairs of full-sibs, with one male and one female per pair. In generation 2, two paternal half-sibs of different sex are mated to unrelated individuals, recording two offspring per mating. This provides 9 different types of covariances between relatives. Samples sizes considered were either 1 000, 2 000 or 20 000 records per trait, i.e. $f = 100, f = 200$ or $f = 2000$ for PHS and $f = 125, f = 250$ or $f = 2500$ for BON. This assumed that all traits were recorded on all individuals. In addition, an unbalanced scenario was considered for $q = 10$ traits, where records for 1000 individuals with all traits were combined with measurements for the first five traits only for 20 000 individuals ($s = 2000$ for PHS or $f = 2500$ for BON).

For the second part of the simulations, a permanent environmental covariance between full-sibs was considered in addition. Population parameters were as above with (i) the proportion of the permanent environmental variance (c_i^2) equal to 0.2 for all traits and all corresponding correlations (r_{Cij}) equal to 0 for scenario (a) and equal to 0.4 for scenarios (b) to (d), and (ii) $c_i^2 = 0.15 + 0.05(\text{mod}(i, 3) - 1)$ and $r_{Cij} = 0$ for scenario (a) and $r_{Cij} = -0.4^{|j-i|}$ otherwise. This yielded 24 cases to be examined. To allow estimation of the additional components, PHS was replaced with a hierarchical full-sib design (HFS) with 5 dams per sire and 4 progeny per dam.

Analyses

Restricted maximum likelihood (REML) estimates of covariance matrices were obtained fitting an animal model with means as the only fixed effects. Full multivariate analyses were carried out considering all q traits simultaneously and contrasted to part-analyses, comprised of $q(q-1)/2$ bivariate

analyses for all possible pairs of traits. A method of scoring algorithm followed by derivative-free search steps to ensure convergence had been achieved was used to locate the maximum of the log likelihood function. Starting values for PHS and HFS designs were obtained from appropriate analyses of variance, modifying any non-positive definite estimates as suggested by Amemiya [1]. For Bondari's design, population values were used. A total of 500 replicates were carried out for each case examined.

Results from part-analyses were pooled using five types of methods. For Method 1, preliminary overall covariance matrices were first constructed by simply averaging any components with multiple estimates. After an eigenvalue decomposition, these were modified to ensure they were positive definite, considering one matrix at a time. Let $\tilde{\Sigma}$ represent the 'average' matrix with eigen-decomposition $\tilde{\Sigma} = \mathbf{E}\mathbf{\Gamma}\mathbf{E}'$, where $\mathbf{\Gamma} = \text{Diag}\{\gamma_i\}$ denotes the diagonal matrix of eigenvalues, γ_i , and \mathbf{E} the corresponding matrix of eigenvectors.

For method 1A, any eigenvalues less than $\delta = 0.001$ were replaced by this value, i.e. $\gamma_i^* = \max(\gamma_i, \delta)$ with $\mathbf{\Gamma}^* = \text{Diag}\{\gamma_i^*\}$. Similarly, for method 1B, eigenvalues were modified by shrinking them towards their mean ($\bar{\gamma}$) if the smallest eigenvalue, γ_q , was less than δ

$$\gamma_i^* = \bar{\gamma} + \beta(\gamma_i - \bar{\gamma}) \quad \text{with} \quad \beta = \min\left(1, (\delta - \bar{\gamma})/(\gamma_q - \bar{\gamma})\right) \quad (6)$$

For both, the pooled estimate was then constructed by pre- and post-multiplying the modified matrix of eigenvalues by the matrix of eigenvectors and its transpose, $\hat{\Sigma} = \mathbf{E}\mathbf{\Gamma}^*\mathbf{E}'$.

Methods 2A and 2B applied corresponding truncation and shrinkage procedures to canonical eigenvalues. For the simple animal model with genetic (G) and residual (E) covariance matrices $\tilde{\Sigma}_G$ and $\tilde{\Sigma}_E$, these were the eigenvalues of $\hat{\Sigma}_P^{-1}\tilde{\Sigma}_G$. $\hat{\Sigma}_P$, the estimate of the overall phenotypic covariance matrix, was obtained as $\tilde{\Sigma}_G + \tilde{\Sigma}_E$, modified using method 1A if this sum has any eigenvalues less than δ . For $\hat{\Sigma}_G$ to be positive definite, canonical eigenvalues (λ_i) need to be greater than zero and less than unity. Hence, for method 2A, any λ_i less than δ were replaced by δ and any values greater than $1 - \delta$ were set to the latter, $\lambda_i^* = \max(\lambda_i, \delta)$ and $\lambda_i^* = \min(\lambda_i, 1 - \delta)$. Analogously, for method 2B, the two separate shrinkage factors required to ensure all modified eigenvalues λ_i^* fell in this range were determined, selecting the smaller to shrink the canonical eigenvalues towards their mean ($\bar{\lambda}$), $\beta = \min(\beta_1, \beta_2)$ for $\beta_1 = (\delta - \bar{\lambda})/(\lambda_q - \bar{\lambda})$ and $\beta_2 = (1 - \delta - \bar{\lambda})/(\lambda_1 - \bar{\lambda})$ and λ_1 and λ_q the largest and smallest canonical eigenvalue, respectively. Estimates of the pooled covariance matrices were then obtained as

$$\hat{\Sigma}_G = \mathbf{F}\mathbf{\Lambda}^*\mathbf{F}' \quad \text{and} \quad \hat{\Sigma}_E = \mathbf{F}(\mathbf{I} - \mathbf{\Lambda}^*)\mathbf{F}' \quad (7)$$

for $\mathbf{\Lambda}^* = \text{Diag}\{\lambda_i^*\}$, \mathbf{F} the matrix of eigenvectors of $\hat{\Sigma}_P^{-1}\tilde{\Sigma}_G$ and \mathbf{I} denoting an identity matrix. For models fitting an additional random effect with covariance matrix Σ_C , $\hat{\Sigma}_P$ was obtained from $\tilde{\Sigma}_G + \tilde{\Sigma}_C + \tilde{\Sigma}_E$ and the above procedures were applied to $\hat{\Sigma}_P^{-1}\tilde{\Sigma}_G$ and $\hat{\Sigma}_P^{-1}\tilde{\Sigma}_C$ to obtain estimates $\hat{\Sigma}_G$ and $\hat{\Sigma}_C$. The pooled residual covariance matrix was then estimated using $\hat{\Sigma}_P - \hat{\Sigma}_G - \hat{\Sigma}_C$. As the difference between positive definite matrices is not guaranteed to be positive definite, this was again modified, if necessary, using method 1A to give a positive definite estimate $\hat{\Sigma}_E$. Method 2C was like method 2B, but applied

more shrinkage than needed to constrain the canonical eigenvalues to the interval $[\delta, 1 - \delta]$. This was obtained by scaling the shrinkage factor β from method 2B with a coefficient less than unity, using values of 0.98 to 0.90.

The third procedure considered was the ‘iterative summation of expanded part matrices’ [28], using pooled estimates from method 1A as starting values. Method 3A applied this technique individually to Σ_G , Σ_C and Σ_E . Method 3B pooled phenotypic rather than residual components and again obtained $\hat{\Sigma}_E$ from $\hat{\Sigma}_P - \hat{\Sigma}_G - \hat{\Sigma}_C$, modified using method 1A if required.

Method 4 used the ML approach to pool part results, applied to a single covariance matrix at a time. Analogous to methods 3A and 3B, methods 4A and 4B pooled residual and phenotypic covariances, respectively. Method 4C extended method 4B by applying a penalty aimed at shrinking the estimate of the individual matrices towards their phenotypic counterpart, $\hat{\Sigma}_P$ (obtained by method 4B, i.e. pooling estimates of phenotypic components from part analyses by ML). Details of the penalties considered are given below.

Finally, method 5 combined part results for all sources of variation, Σ_G , Σ_E and Σ_C (if fitted), simultaneously using the likelihood approach. First, this was done by simple maximizing the sum of likelihood contributions from individual matrices, either without (method 5A) or with (method 5C) penalties. Secondly, a pseudo-pedigree structure was invoked as described above (see xxx), again with and without penalization of the likelihood. Pseudo-structures considered were a simple balanced, paternal half-sib design with 2 progeny for 2 sires each, a hierarchical full-sib design with 2 sires, 2 dams per sire and 2 progeny per dam, and Bondari’s design. For the unbalanced scenario, results for part analyses involving 21 000 records for both traits, 21 000 records for one trait and 1 000 records for the second trait, and 1 000 records for both traits were assigned weights (i.e. degrees of freedom) of 1 : 1 : 1 (unweighted), 4 : 2 : 1, 20 : 10 : 1, 100 : 10 : 1 and 400 : 20 : 1. Different constellations are denoted by the first weight in the following.

Penalties

The first type of penalty considered aimed at shrinking individual covariance matrices towards their sum, i.e. the estimate of the phenotypic covariance matrix. As outlined by Meyer *et al.* [31], this can be derived assuming an Inverse Wishart prior with $\hat{\Sigma}_P$ taking the rôle of the scale matrix

$$\mathcal{P}_\Sigma \propto \sum_x \log |\hat{\Sigma}_x| + \text{tr} \left(\hat{\Sigma}_x^{-1} \hat{\Sigma}_P \right) \quad (8)$$

For the simple animal model, this was done imposing a penalty on the genetic covariance matrix only, $x = G$, denoted as \mathcal{P}_Σ , and on both genetic and residual components, $x = G, E$, denoted as \mathcal{P}_Σ^2 . Similarly, when fitting common environmental effects in addition, combinations $x = G$ (\mathcal{P}_Σ), $x = G, C$ (\mathcal{P}_Σ^{2b}) and $x = G, C, E$ (\mathcal{P}_Σ^3) were examined.

Corresponding penalties on correlation rather than covariance matrices ($\mathcal{P}_R, \mathcal{P}_R^2, \dots$) were obtained by substituting correlation matrices $\hat{\mathbf{R}}_x$ for $\hat{\Sigma}_x$ in (Eq. 8).

Secondly, penalties encouraging shrinkage of canonical eigenvalues, λ_{xi} , towards their mean ($\bar{\lambda}_x$)

341 were used.

$$342 \quad \mathcal{P}_\lambda \propto \sum_x \sum_i (\lambda_{xi} - \bar{\lambda}_x)^2 \quad (9)$$

343 These were the eigenvalues of $\hat{\Sigma}_p^{-1} \hat{\Sigma}_x$, with $x = G$ for the simple animal model and $x = G$ and
 344 $x = G, C$ for a model fitting permanent environmental effects. In addition, corresponding penalties on
 345 canonical eigenvalues transformed to logarithmic scale were considered, penalizing either $\log(\lambda_{xi})$
 346 alone (denoted as $\mathcal{P}_\lambda^{\ell}$) or both $\log(\lambda_{xi})$ and $\log(1 - \lambda_{xi})$ (denoted as $\mathcal{P}_\lambda^{\ell 2}$), as done by Meyer [29].

347 Fixed tuning factors of $\psi = 0.1$ or $\psi = 0.02$ were used throughout.

348 *Summary statistics*

349 The quality of estimates of covariance matrices from different methods was assessed as the deviation
 350 from population values, summarized by the entropy loss [21]

$$351 \quad L_1(\Sigma, \hat{\Sigma}) = \text{tr}(\Sigma^{-1} \hat{\Sigma}) - \log|\Sigma^{-1} \hat{\Sigma}| - q \quad (10)$$

352 Comparisons to estimates from full multivariate results were made using the percentage reduction
 353 in average loss (PRIAL) [27]

$$354 \quad 100 \left[1 - \bar{L}_1(\Sigma_x, \hat{\Sigma}_x) / \bar{L}_1(\Sigma_x, \hat{\Sigma}_x^0) \right] \quad (11)$$

355 with $\bar{L}_1(\cdot)$ the average entropy loss over replicates and $\hat{\Sigma}_x^0$ denoting the estimate from a multivariate
 356 analysis of all q traits.

357 In addition, pooled estimates for all sources of variation were used to calculate the corresponding
 358 log likelihood in the full, multivariate model. This was expressed as deviation from the maximum
 359 log likelihood from the latter analysis, $\Delta \mathcal{L}$.

360 **Results**

361 Average losses in estimates of pooled covariance matrices, compared to MUV analyses, are summa-
 362 rized in Table 1 for several commonly employed methods, $q = 10$ traits and 1 000 records per trait.
 363 For this relatively small sample size, 80% of ‘averaged’ matrices $\hat{\Sigma}_G$ and 19% of $\hat{\Sigma}_E$ for PHS were
 364 not positive definite. With more covariances between relatives, corresponding figures for BON were
 365 considerably lower, 56% and 5%, respectively, and differences in PRIAL can, to a large extent, be
 366 attributed to these different proportions. Modifying a single matrix at a time resulted in substan-
 367 tial increases in loss (i.e. negative PRIAL), especially when pooling genetic and residual matrices
 368 separately and obtaining $\hat{\Sigma}_p$ as their sum (methods 1A, 1B, 3A, 4A). Pooling genetic and phenotypic
 369 matrices and determining $\hat{\Sigma}_E$ as their difference performed somewhat better (methods 3B and 4B) and
 370 yielded estimates and $\log \mathcal{L}$ closer to those from corresponding multivariate analyses (MUV). While

371 resulting in less additional loss in $\hat{\Sigma}_G$, the frequently used approach of shrinking the eigenvalues of
 372 $\hat{\Sigma}_G$ towards their mean (method 1B) resulted in the worst estimates $\hat{\Sigma}_P$ with the largest deviation in
 373 $\log \mathcal{L}$ from MUV.

374 Canonical eigenvalues are a function of both $\hat{\Sigma}_G$ and $\hat{\Sigma}_P$, estimated from $\hat{\Sigma}_G + \hat{\Sigma}_E$. Hence modification
 375 of these values (methods 2) gave $\hat{\Sigma}_P$ with PRIALs close to zero, i.e. the virtually same average loss
 376 as MUV. As expected, truncating canonical eigenvalues at an operational zero (2A) yielded $\log \mathcal{L}$
 377 very close to the maximum for MUV – this is the mechanism implicit in multivariate REML analyses
 378 constraining estimates to the parameter space [1]. Shrinking the canonical eigenvalues towards their
 379 mean (method 2B) improved pooled estimates substantially, and further reductions in loss were
 380 obtained when ‘bending’ somewhat further than needed to make estimates positive definite (method
 381 2C). In contrast, placing penalties on $\hat{\Sigma}_G$ and $\hat{\Sigma}_E$ when pooling these separately (method 4C) yielded
 382 estimates of Σ_P and $\log \mathcal{L}$ comparable to method 1B. Figure 1 shows the distribution of losses in
 383 estimates of individual covariance matrices across replicates for one of the 12 cases examined.

384 Corresponding results obtained when pooling genetic and residual components simultaneously
 385 using the penalized likelihood (PL) approach are given in Table 2. Clearly, when treating estimates \mathbf{S}_i^G
 386 and \mathbf{S}_i^E from the same part analysis as independent, PL performed no better than methods combining
 387 estimates for each source of variation separately. Large (absolute) average values of $\Delta \mathcal{L}$, accompanied
 388 by dramatic increases in loss for $\hat{\Sigma}_P$ for the PHS design were due to cases Iib, Iic and Iid, i.e. cases
 389 with population heritabilities of $h_i^2 = 0.2$. For $s = 100$ families, the combined matrix $\hat{\Sigma}_G$ for these
 390 cases was non-positive definite in all replicates. Allowing for a pseudo pedigree structure, however,
 391 yielded pooled estimates of Σ_P with the same average loss than the full, multivariate analysis (PRIAL
 392 close to zero). Similarly, PRIAL in $\hat{\Sigma}_G$ and $\hat{\Sigma}_E$ increased substantially (i.e. average losses in estimates
 393 were reduced). While there was little difference between pseudo pedigree structures for PRIALs of
 394 $\hat{\Sigma}_P$, the structure with the strongest implied sampling correlation (PHS) resulted in estimates of Σ_G
 395 and Σ_E with least losses, even for data sampled under Bondari’s design. The PHS design shown used
 396 the lowest possible numbers for s and n . Other values were investigated and yielded comparable
 397 results, though with some tendency for PRIALs to decrease as s and n increased.

398 Imposing penalties aimed at reducing sampling variation greatly increased PRIALs further. As
 399 reported for full, multivariate analyses, a penalty on both $\hat{\Sigma}_G$ and $\hat{\Sigma}_E$ (\mathcal{P}_Σ^2) improved estimates of
 400 both covariance matrices with little change in the PRIAL for $\hat{\Sigma}_G$ compared to a penalty on this matrix
 401 only (\mathcal{P}_Σ) [29]. Similar patterns were obtained for penalties on correlation matrices, \mathcal{P}_R and \mathcal{P}_R^2 (not
 402 shown), though with somewhat lower PRIALs because these penalties had a lower values and, at the
 403 same tuning factor, thus resulted in less stringent regularization. Shrinking canonical eigenvalues
 404 towards their mean (\mathcal{P}_λ) appeared to be most effective for estimates of Σ_E , while corresponding
 405 penalties on the logarithmic scale (\mathcal{P}_λ^ℓ (not shown) and $\mathcal{P}_\lambda^{\ell 2}$) resulted in substantial improvements for
 406 both Σ_G and Σ_E . Again, similar relationships between penalties have been reported for penalized
 407 estimation in full, multivariate analyses [30, 29].

Varying sample size

Results for selected penalties and a range of sample sizes are contrasted in Table 3. For $q = 5$ traits, there were few replicates ($< 7\%$) for which combined matrices $\hat{\Sigma}_G$ and $\hat{\Sigma}_E$ had eigenvalues less than δ , so that few modifications were required, losses in $\hat{\Sigma}_P$ were very similar to those for MUV and reductions in $\log \mathcal{L}$ were negligible. For larger numbers of traits, there were increasing proportions of ‘problematic’ replicates, and differences in results largely reflect variation in the latter. Similarly, these were pronounced differences between sets of population values considered – generally cases for correlation scenario (c) or low levels of heritabilities were most afflicted. For $q = 20$ traits, for instance, substantial negative PRIALs, especially for method 1B, were due to cases Ic and IIc, for which almost 100% of replicates for the paternal half-sib design required modification of both $\hat{\Sigma}_G$ and $\hat{\Sigma}_E$. Further, for all 8 cases with heritabilities $h_i^2 = 0.2$, all replicates produced $\hat{\Sigma}_G$ with at least one eigenvalue below δ .

However, ranking and relative performance of the different methods across numbers of traits, sample sizes, designs and sets population parameters was highly consistent. Truncating canonical eigenvalues (method 2A) yielded pooled estimates ‘nearest’ (in terms of $\Delta \mathcal{L}$) to multivariate results, closely followed by the PL approach without penalties but assuming a pseudo pedigree structure (method 5A). Generally, the PL approach (5A) gave higher PRIALs for both $\hat{\Sigma}_G$ and $\hat{\Sigma}_E$ than the former (2A), indicating that, even without penalties, PL involved more regularization. Additional investigations (not presented here) showed that this was due to some modifications of the highest as well as the lowest eigenvalues. Yet, the PL approach gave estimates substantially closer to MUV (lower absolute values of $\Delta \mathcal{L}$) than ‘bending’ just far enough that the smallest canonical eigenvalue was equal to δ (method 2B).

Imposing a penalty increased PRIALs throughout, accompanied by a corresponding decrease in $\Delta \mathcal{L}$. As outlined above, the various penalties considered acted in somewhat different fashions. For relatively small sample sizes comprising only 1 000 or 2 000 records per trait, the tuning factor of $\psi = 0.1$ chosen resulted in substantial reductions in loss, especially for estimates of Σ_G . In practice, a somewhat smaller value of ψ , resulting in less regularization and reduction in loss but estimates closer to those from multivariate analyses may be preferred.

Corresponding results for large samples (20 000 records per trait) are presented in Table 4. With eigenvalues of $\hat{\Sigma}_G$ less than δ only occurring for 1% and 2% of replicates for cases IIc and IIIc for the PHS design, results for methods 1A, 1B, 2A and 2B were virtually identical. Moreover, these simple ‘average’ estimates agreed closely with those from full multivariate analyses, especially for the paternal half-sib design, supporting the claim that maximum composite likelihood estimates are consistent [26]. Again, estimates using the PL approach without penalization (method 5A) were overall comparable, with a tendency for the loss in $\hat{\Sigma}_G$ to be slightly reduced. Imposing a stringent penalty, however, proved detrimental, in particular for data with Bondari’s design. Similarly, ‘bending’ further than required to obtain a positive definite matrix increased losses in $\hat{\Sigma}_E$ substantially. Results emphasize that regularized estimation should be used with care – while some penalization appears invariably advantageous for small to medium samples, it may be counter-productive for large, highly informative samples.

448 Extensions

449 One concern when pooling estimates is that, due to different amounts of data or relationship informa-
 450 tion available, estimates from various part analyses differ in their reliability. The PL approach readily
 451 allows for weighted pooling by choosing the degrees of freedom for individual parts accordingly.
 452 Figure 2 summarizes results for Bondari’s design with $f = 125$ families with records for all $q = 10$
 453 traits and an additional $f = 2500$ families with records for the first five traits only. Without penaliza-
 454 tion, $\Delta\mathcal{L}$ were comparable for all weightings examined and similar to that obtained when truncating
 455 canonical eigenvalues at δ (method 2A). However, PRIALs for individual sources of variation were
 456 consistently lower than for the corresponding balanced design with $f = 125$ families (see Table 3).
 457 In particular, losses in $\hat{\Sigma}_P$ were about 4% higher than for full, multivariate analyses. With more than
 458 20 times more records for part analyses involving traits 1 to 5 than those involving traits 6 to 10,
 459 corresponding weights of 20 : 1 seemed a reasonable choice. However, PRIALs were surprisingly
 460 robust against the wide range of weights considered, with unweighted pooling generally resulting
 461 in the least loss in estimates. This tendency was more pronounced when imposing penalties, as, for
 462 constant ψ , the higher degrees of freedom used reduced the relative emphasis given to the penalties.

463 Results for Bondari’s design when records are subject to a permanent environmental covariance
 464 between full-sibs are shown in Figure 3. Overall, differences between methods for pooling were
 465 analogous to those for a simple animal model. While truncating canonical eigenvalues (method 2A)
 466 again yielded pooled estimates closest to MUV, this was associated with substantially increased losses
 467 for $\hat{\Sigma}_E$. As above, ‘bending’ (method 2B) and the unpenalized PL approach (5A) resulted in estimates
 468 with comparable PRIALs, but the latter had $\Delta\mathcal{L}$ closer to zero. For method 5C, it was advantageous
 469 to place penalties on all three covariance matrices to be estimated (\mathcal{P}_Σ^3) or to penalize the spread of
 470 the canonical eigenvalues of both of $\hat{\Sigma}_P^{-1}\hat{\Sigma}_G$ and $\hat{\Sigma}_P^{-1}\hat{\Sigma}_C$ on the logarithmic scale ($\mathcal{P}_\lambda^{\ell^2}$).

471 Discussion

472 Combining estimates of covariance components from different analyses is a long-standing problem
 473 which generally is not been given sufficient attention. Though mostly motivated by simple necessity,
 474 the common practice of obtaining estimates in a series of analyses of subsets appears to fit under
 475 the umbrella of maximum composite or pseudo likelihood estimation. While ignoring covariances
 476 between subsets is expected to make less efficient use of the data than full, multivariate analyses (i.e.
 477 to result in larger sampling variances), the resulting estimators share some desirable properties with
 478 full ML estimators, as they are generally consistent and asymptotically normal [26, 43].

479 With more than one covariance matrix to be estimated, sampling variation arising from partitioning
 480 phenotypic covariances into their causal components is substantial and a simple ‘averaging’ strategies
 481 to combine estimates from part analyses are likely to yield non-positive definite or ill-conditioned
 482 matrices. Results shown provide a strong argument for combining part estimates for all sources of
 483 variation considered simultaneously whilst, approximately at least, accounting for negative sampling
 484 covariances. This provides estimates of phenotypic covariances close to those from multivariate

485 analyses. Moreover, with numerous traits and correspondingly many parameters to be estimated,
486 average reductions in the log likelihood are often not significant (at an error probability of 5% or less).
487 Modification of the canonical eigenvalues of covariances matrices for individual random effects and
488 the phenotypic covariance matrix provides a simple procedure to do so. Simulation results suggest
489 that this is effective even if the model of analysis comprises additional random effects, though it
490 should be borne in mind that an estimate of the residual covariance matrix, obtained as the difference
491 between the phenotypic and the other covariance matrices, is not guaranteed to be positive definite.

492 A more flexible alternative to pool estimates is provided by the likelihood approach proposed.
493 We have demonstrated for a range of scenarios that it performs well, yielding estimates close to
494 multivariate results and less subject to increases in loss than many other procedures in use. Like most
495 methods available to replace non-positive definite matrices with a positive definite equivalent, it is *ad*
496 *hoc* and we do not claim any specific statistical properties. Estimation by parts appears to be the less
497 efficient, i.e. the more subject to increased losses, the higher the incidence of non-positive 'averaged'
498 matrices and the degree of modifications required. Simply making such matrices positive definite
499 then implies some regularization and thus reduces losses. Moreover, there is usually considerable
500 scope to 'improve' estimates further by applying some additional regularization to reduce sampling
501 variation. The likelihood approach readily allows for this through a penalty on the likelihood
502 function. Simulation results show that penalties aimed at borrowing strength from the estimate of
503 the phenotypic covariance matrix can reduce losses in estimates of covariance matrices substantially,
504 similar to applications of penalized estimation in full, multivariate analyses.

505 The combination of estimation by parts and our penalized likelihood approach to pool estimates
506 provides a pragmatic solution to routine estimation of high-dimensional covariance matrices required
507 in quantitative genetics application. We have not discussed choice of tuning factors - simulation results
508 suggest that penalties applied should be 'mild', in particular when sample sizes are large. A way to
509 evaluate changes due to penalization may be to monitor the Frobenius norm of the matrix differences
510 between penalized and unpenalized estimates. While computationally slightly more demanding
511 than 'bending' or related techniques, requirements are trivial compared to those for the individual
512 part analyses.

513 Acknowledgments

514 This work was supported by Meat and Livestock Australia under grant B.BFG.0050.

References

- 515
- 516 [1] Amemiya Y (1985) What should be done when an estimated between-group covariance matrix
517 is not nonnegative definite ? *American Statistician* **39**, 112–117.
- 518 [2] Arnason T (1984) Genetic studies on conformation and performance of Icelandic Toelter horses.
519 *Acta Agriculturae Scandinavica* **34**, 428–439. doi: 10.1080/00015128409435411.
- 520 [3] Bellio R, Varin C (2005) A pairwise likelihood approach to generalized linear models with crossed
521 random effects. *Statistical Modelling* **5**, 217–227. doi: 10.1191/1471082X05st095oa.
- 522 [4] Bentler P, Yuan KH (2011) Positive definiteness via off-diagonal scaling of a symmetric indefinite
523 matrix. *Psychometrika* **76**, 119–123. doi: 10.1007/s11336-010-9191-3.
- 524 [5] Bondari K, Willham RL, Freeman AE (1978) Estimates of direct and maternal genetic correlations
525 for pupa weight and family size of *Tribolium*. *Journal of Animal Science* **47**, 358–365.
- 526 [6] Borsdorf R, Higham NJ (2010) A preconditioned Newton algorithm for the nearest correlation
527 matrix. *IMA Journal of Numerical Analysis* **30**, 94–107. doi: 10.1093/imanum/drn085.
- 528 [7] Borsdorf R, Higham NJ, Raydan M (2010) Computing a nearest correlation matrix with factor
529 structure. *SIAM Journal of Matrix Analysis and Applications* **31**, 2603–2622.
- 530 [8] Boyd S, Xiao L (2005) Least-squares covariance matrix adjustment. *SIAM Journal of Matrix*
531 *Analysis and Applications* **27**, 532–546.
- 532 [9] Brown DJ, Ball AJ, Huisman AE, Swan AA, Atkins KD, Graser HU, Banks R, Swan P, Woolaston
533 RR (2006) Sheep Genetics Australia: A national genetic evaluation system for Australian sheep.
534 *Proceedings of the Eighth World Congress on Genetics Applied to Livestock Production* Communication
535 No. 05–03.
- 536 [10] Cox DR, Reid N (2004) A note on pseudolikelihood constructed from marginal densities.
537 *Biometrika* **91**, 729–737.
- 538 [11] Dennis JE, Schnabel RB (1996) Numerical methods for Unconstrained Optimization and Non-
539 linear Equations. SIAM Classics in Applied Mathematics (Philadelphia: Society for Industrial
540 and Applied Mathematics).
- 541 [12] Ducrocq V (1993) Genetic parameters for type traits in the French Holstein breed based on
542 a multiple-trait animal model. *Livestock Production Science* **36**, 143 – 156. doi: 10.1016/0301-
543 6226(93)90148-B.
- 544 [13] Fieuws S, Verbeke G (2006) Pairwise fitting of mixed models for the joint modeling of multivariate
545 longitudinal profiles. *Biometrics* **62**, 424–431. doi: 10.1111/j.1541-0420.2006.00507.x.
- 546 [14] Graser HU, Tier B, Johnston DJ, Barwick SA (2005) Genetic evaluation for the beef industry in
547 Australia. *Australian Journal of Experimental Agriculture* **45**, 913–921. doi: 10.1071/EA05075.
- 548 [15] Gunawan B, James JW (1986) The use of ‘bending’ in multiple trait selection of Border Leicester-
549 Merino synthetic populations. *Australian Journal of Agricultural Research* **37**, 539–547.

- 550 [16] Hayes JF, Hill WG (1981) Modifications of estimates of parameters in the construction of genetic
551 selection indices ('bending'). *Biometrics* **37**, 483–493.
- 552 [17] Henshall JM, Meyer K (2002) "PDMATRIX" – Programs to make matrices positive definite.
553 *Proceedings of the Seventh World Congress on Genetics Applied to Livestock Production* Communication
554 No. 28–12.
- 555 [18] Higham NJ (1988) Computing a nearest symmetric positive semi-definite matrix. *Linear Algebra*
556 *and its Applications* **103**, 103–118.
- 557 [19] Higham NJ (2002) Computing the nearest correlation matrix – a problem from finance. *IMA*
558 *Journal of Numerical Analysis* **22**, 329–343. doi: 10.1093/imanum/22.3.329.
- 559 [20] Hill WG, Thompson R (1978) Probabilities of non-positive definite between-group or genetic
560 covariance matrices. *Biometrics* **34**, 429–439.
- 561 [21] James W, Stein C (1961) Estimation with quadratic loss. J Neiman, editor, Proceedings of the
562 Fourth Berkeley Symposium on Mathematical Statistics and Probability (University of California
563 Press), pp. 361–379.
- 564 [22] Jorjani H, Klei L, Emanuelson U (2003) A simple method for weighted bending of genetic
565 (co)variance matrices. *Journal of Dairy Science* **86**, 677–679.
- 566 [23] Knol DL, ten Berge JMF (1989) Least-squares approximation of an improper correlation matrix
567 by a proper one. *Psychometrika* **54**, 53–61. doi: 10.1007/bf02294448.
- 568 [24] Koivula M, Negussie E, Mäntysaari EA (2004) Genetic parameters for test-day somatic cell count
569 at different lactation stages of Finnish dairy cattle. *Livestock Production Science* **90**, 145–157.
- 570 [25] Lawley DN (1956) Tests of significance of the latent roots of covariance and correlation matrices.
571 *Biometrika* **43**, 128–136.
- 572 [26] Lele S, Taper ML (2002) A composite likelihood approach to (co) variance components estimation.
573 *Journal of Statistical Planning and Inference* **103**, 117–135.
- 574 [27] Lin SP, Perlman MD (1985) A Monte Carlo comparison of four estimators of a covariance matrix.
575 PR Krishnaish, editor, *Multivariate Analysis* (Amsterdam: North-Holland), volume 6, pp. 411–
576 428.
- 577 [28] Mäntysaari EA (1999) Derivation of multiple trait reduced random regression (RR) model for the
578 first lactation test day records of milk, protein and fat. Proceedings of the 50th Annual Meeting
579 of the European Association of Animal Production (Europ. Ass. Anim. Prod.).
- 580 [29] Meyer K (2011) Performance of penalized maximum likelihood in estimation of genetic covari-
581 ances matrices. *Genetics Selection Evolution* **43**, 39. doi: 10.1186/1297-9686-43-39.
- 582 [30] Meyer K, Kirkpatrick M (2010) Better estimates of genetic covariance matrices by 'bending' using
583 penalized maximum likelihood. *Genetics* **185**, 1097–1110. doi: 10.1534/genetics.109.113381.

- 584 [31] Meyer K, Kirkpatrick M, Gianola D (2011) Penalized maximum likelihood estimates of genetic
585 covariance matrices with shrinkage towards phenotypic dispersion. *Proceedings of the Association*
586 *for Advancement of Animal Breeding and Genetics* **19**, 87–90.
- 587 [32] Qi H, Sun D (2006) A quadratically convergent Newton method for computing the near-
588 est correlation matrix. *SIAM Journal of Matrix Analysis and Applications* **28**, 360–385. doi:
589 10.1137/050624509.
- 590 [33] Qi H, Sun D (2011) An augmented Lagrangian dual approach for the H-weighted nearest cor-
591 relation matrix problem. *IMA Journal of Numerical Analysis* **31**, 491. ISSN 0272-4979. doi:
592 10.1093/imanum/drp031.
- 593 [34] Rebonato R, Jaeckel P (2000) The most general methodology to create a valid correlation matrix
594 for risk management and option pricing purposes. *The Journal of Risk* **2**, 17–27.
- 595 [35] Rousseeuw PJ, Molenberghs G (1993) Transformation of non positive semidefinite correlation
596 matrices. *Communications in Statistics - Theory and Methods* **22**, 965–984.
- 597 [36] Schwertman NC, Allen DM (1979) Smoothing an indefinite variance-covariance matrix. *Journal*
598 *of Statistical Computation and Simulation* **9**, 183–194.
- 599 [37] Simon D, Abell J (2010) A majorization algorithm for constrained correlation matrix approxima-
600 tion. *Linear Algebra and its Applications* **432**, 1152–1164. doi: 10.1016/j.laa.2009.10.025.
- 601 [38] Stewart ID, White IMS, Gilmour AR, Woolliams JA, Brotherstone S (2012) Estimating variance
602 components and predicting breeding values for eventing disciplines and grades in sport horses.
603 *Animal* **00**, 000–000.
- 604 [39] ten Berge JMF, Kiers HAL (1993) An alternating least squares method for the weighted approxi-
605 mation of a symmetric matrix. *Psychometrika* **58**, 115–118. doi: 10.1007/bf02294475.
- 606 [40] Thompson R (1976) The estimation of maternal genetic variance. *Biometrics* **32**, 903–917.
- 607 [41] Thompson R, Brotherstone S, White IMS (2005) Estimation of quantitative genetic parameters.
608 *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**, 1469–1477. doi: 10.1098/rb-
609 stb.2005.1676.
- 610 [42] Tyrisevä AM, Meyer K, Fikse F, Ducrocq V, Jakobsen J, Lidauer MH, Mäntysaari EA (2011) Prin-
611 cipal component approach in variance component estimation for international sire evaluation.
612 *Genetics Selection Evolution* **43**, 21. doi: 10.1186/1297-9686-43-21.
- 613 [43] Varin C, Reid N, Firth D (2011) An overview of composite likelihood methods. *Statistica Sinica*
614 **21**, 5–42.
- 615 [44] Wall E, White IMS, Coffey MP, Brotherstone S (2005) The relationship between fertility, rump
616 angle, and selected type information in Holstein-Friesian cows. *Journal of Dairy Science* **88**,
617 1521–1528. doi: 10.3168/jds.S0022-0302(05)72821-6.
- 618 [45] Yerex RP (1988) Pooling restricted maximum likelihood estimates from data subsets under the
619 animal model. Ph.D. thesis, Cornell University, Ithaca, NY. 224pp.

Table 1. Mean percentage reduction in average loss in estimates of genetic (G), residual (E) and phenotypic (P) covariance matrices when pooling estimates from pairwise analyses

Data for 10 traits and 1000 records per trait

Method	Pen. ^A	Half-sib design				Bondari's design			
		G	E	P	$\log \mathcal{L}^B$	G	E	P	$\log \mathcal{L}$
1A		-33	-9	-10	-3.6	-13	-9	-2	-1.7
1B		-3	-9	-47	-18.7	4	-9	-10	-6.6
2A		-28	-2	0	-0.1	-10	-5	0	-0.5
2B		8	30	0	-1.8	9	10	0	-2.1
2C	0.98	32	40	0	-2.1	30	17	0	-2.4
	0.90	54	51	0	-3.5	52	24	0	-3.8
3A		5	15	-32	-13.3	18	7	-4	-4.2
3B		5	-2	-2	-3.5	18	-5	4	-2.1
4A		-30	-3	-38	-13.8	-8	-2	-10	-4.5
4B		-30	-5	-6	-3.4	-8	-7	0	-1.9
4C	\mathcal{P}_Σ	19	-2	-6	-4.1	30	-4	0	-2.5
	\mathcal{P}_Σ^2	19	20	-45	-17.0	30	10	-13	-6.2

^APenalty. 2C: Scaling factor for β used, 4C: Type of penalty^BAverage change in log likelihood from maximum for multivariate analysis of all traits

Table 2. Mean percentage reduction in average loss in estimates of covariance matrices obtained by jointly pooling estimates from pairwise analyses using penalized maximum likelihood (methods 5) with different pseudo pedigree structures

Data for $q = 10$ traits and 1000 records per trait; penalties applied using a tuning factor of $\psi = 0.1$. Pseudo pedigree structures: PHS - paternal half sib design with 2 sires and 2 progeny per sire, HFS - hierarchical full sib design with 2 sires, 2 dams per sire and 2 progeny per dam and BON - Bondari's design

Penalty	Genetic				Residual				Phenotypic			
	None	PHS	HFS	BON	None	PHS	HFS	BON	None	PHS	HFS	BON
<i>Paternal half-sib design</i>												
None	-38	-11	-15	-21	-20	8	3	2	-182	0	0	0
\mathcal{P}_Σ	19	70	56	50	-4	29	12	9	-52	1	1	1
\mathcal{P}_Σ^2	19	71	57	50	19	50	35	32	-55	2	1	1
\mathcal{P}_λ	-36	-4	-13	-21	2	35	27	25	-186	0	0	0
\mathcal{P}_λ^2	-10	61	38	27	3	45	31	28	-190	1	1	0
<i>Bondari's design</i>												
None	-15	5	2	-2	-19	3	0	0	-70	0	0	0
\mathcal{P}_Σ	29	70	58	52	-6	22	10	7	-16	1	1	1
\mathcal{P}_Σ^2	29	70	58	52	8	27	21	18	-17	2	1	1
\mathcal{P}_λ	-13	10	3	-2	-7	17	14	12	-81	0	0	0
\mathcal{P}_λ^2	7	62	42	34	-6	25	17	15	-83	1	0	0

Table 3. Mean percentage reduction in average loss in pooled estimates of genetic (G), residual (E) and phenotypic (P) covariance matrices for selected sample sizes

Methods 5A and 5C: assuming a paternal half-sib pseudo pedigree structure comprising 2 sires and 2 progeny per sire; 5C: Penalties applied using a tuning factor of $\psi = 0.1$

q ^A	f ^B	Method of pooling												
		1A	1B	2A	2B	2C	3A	4A	5A	5C				
					0.98 ^C					\mathcal{P}_{Σ}^2	$\mathcal{P}_{\mathbf{R}}^2$	\mathcal{P}_{λ}	$\mathcal{P}_{\lambda}^{\ell 2}$	
<i>Paternal half-sib design</i>														
10	200	G	-11	0	-8	7	25	17	-5	7	65	55	12	57
		E	-4	-4	0	10	22	14	3	8	39	33	26	35
		P	-4	-20	0	0	0	-6	-10	0	1	1	0	1
		$\Delta\mathcal{L}^{\text{D}}$	-1.3	-7.7	0.0	-0.7	-0.9	-3.8	-3.6	-0.2	-7.6	-2.9	-0.6	-3.4
20	200	G	-45	8	-39	21	37	-6	-31	-21	68	56	-12	56
		E	-14	-15	-6	23	28	6	-4	5	36	31	24	31
		P	-15	-157	-1	-1	-1	-21	-30	0	2	1	0	1
		$\Delta\mathcal{L}$	-18.5	-185.9	-0.3	-10.7	-12.1	-32.1	-39.7	-2.7	-30.7	-15.4	-4.6	-15.4
<i>Bondari's design</i>														
5	125	G	2	3	3	4	16	9	5	11	54	45	15	50
		E	-1	-1	-1	0	4	1	-1	1	22	14	6	19
		P	0	0	0	0	0	0	-1	0	1	0	0	1
		$\Delta\mathcal{L}$	-0.1	-0.1	-0.1	-0.1	-0.1	-0.2	-0.2	-0.1	-1.8	-0.5	-0.1	-0.7
10	125	G	-13	4	-10	9	30	18	-8	5	70	60	10	62
		E	-9	-9	-5	10	17	7	-2	3	27	24	17	25
		P	-2	-10	0	0	0	-4	-10	0	2	1	0	1
		$\Delta\mathcal{L}$	-1.7	-6.6	-0.5	-2.1	-2.4	-4.2	-4.5	-0.8	-7.8	-3.4	-1.2	-3.9
10	250	G	-10	-4	-8	-1	17	15	-4	6	58	49	12	51
		E	-5	-5	-3	-1	7	7	2	2	8	15	10	11
		P	-1	-3	0	0	0	2	-2	0	1	0	0	0
		$\Delta\mathcal{L}$	-0.7	-2.3	-0.4	-1.0	-1.2	-1.6	-1.4	-0.6	-9.6	-3.4	-1.2	-4.4
20	250	G	-42	-6	-37	3	22	3	-23	-12	61	50	-4	51
		E	-11	-11	-5	3	6	11	4	3	12	14	7	10
		P	-3	-39	0	0	0	1	-7	0	2	1	0	1
		$\Delta\mathcal{L}$	-8.2	-66.0	-2.5	-12.3	-13.7	-12.1	-13.4	-4.3	-32.9	-15.6	-7.4	-17.5

^ANumber of traits

^BNumber of families

^CScaling factor for β

^DAverage change in log likelihood from maximum for multivariate analysis of all traits

Table 4. Mean percentage reduction in average loss in pooled estimates of genetic (G), residual (E) and phenotypic (P) covariance matrices for large samples

Data for $q = 10$ traits, using a paternal half sib pseudo pedigree structure with 2 sires and 2 progeny per sire for method 5 and a tuning factor of ψ for method 5C.

f^A		Method of pooling												
		2B			2C			5A	5C ($\psi = 0.02$)			5C ($\psi = 0.10$)		
		0.98 ^B	0.94 ^B	0.90 ^B		\mathcal{P}_Σ^2	\mathcal{P}_λ	$\mathcal{P}_\lambda^{\ell^2}$	\mathcal{P}_Σ^2	\mathcal{P}_λ	$\mathcal{P}_\lambda^{\ell^2}$			
<i>Paternal half-sib design</i>														
2000	G	0	6	13	15	6	14	7	15	-36	8	7		
	E	0	9	11	1	3	2	12	10	-78	0	-32		
	P	0	0	0	0	0	0	0	0	-2	0	0		
	$\Delta\mathcal{L}^C$	0.0	-0.1	-1.3	-3.5	-0.3	-6.4	-0.6	-1.8	-28.4	-1.9	-10.3		
<i>Bondari's design</i>														
2500	G	-1	4	9	6	4	1	5	10	-92	4	-13		
	E	-3	-1	-28	-86	-1	-65	-9	-31	-346	-82	-217		
	P	0	0	0	0	0	0	0	0	-3	0	-1		
	$\Delta\mathcal{L}$	-0.3	-0.7	-3.5	-9.0	-0.9	-11.5	-1.5	-3.9	-53.8	-5.7	-22.6		

^ANumber of families

^BScaling factor for β

^CAverage change in log likelihood from maximum for multivariate analysis of all traits

Figure 1. Distribution of loss ($\times 100$) in estimates of genetic (G), residual (E) and phenotypic (P) covariance matrices together with corresponding log likelihood

Case Ic, data for $q = 10$ traits, Bondari's design with $f = 125$ families. Log likelihood shown as deviation from value for full multivariate analysis. Methods as described in text and: For 2C digits after hyphen give scaling factor used ($\times 100$). For 4C, penalties \mathcal{P}_Σ (-G) and \mathcal{P}_Σ^2 (-GE) were used. For method 5, none and a paternal half-sib pseudo-pedigree structure were assumed, denoted by '-0' and '-P', respectively. For method 5C, letters 'a', 'b' and 'c' denote use of penalties \mathcal{P}_Σ , \mathcal{P}_Σ^2 and \mathcal{P}_R^2 , and 'd' and 'e' of penalties \mathcal{P}_λ and \mathcal{P}_λ^2 , respectively.

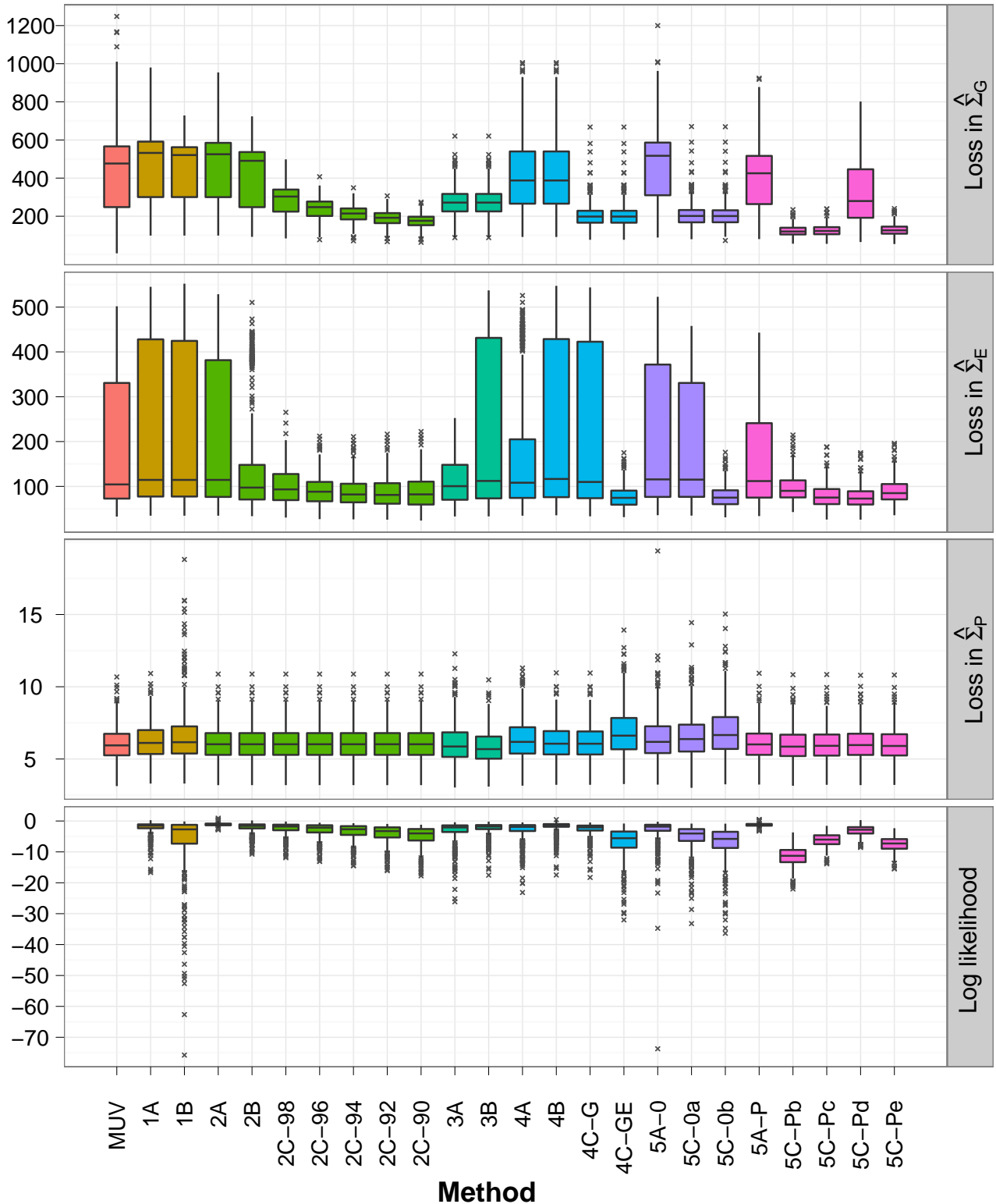


Figure 2. Mean percentage reduction in average loss for estimates of genetic (G), residual (E) and phenotypic (P) covariance matrices together with corresponding log likelihood for data with unequal numbers of records per trait, using different weights when combining part analyses

Data for $q = 10$ traits and Bondari's design with $f = 125$ families with records for all traits and $f = 2500$ families with records for traits 1 to 5 only.

■ method 2A, ● method 5A, ▼ method 5C with $\psi = 0.1$, and ▲ method 5C with $\psi = 0.02$, assuming a paternal half-sib pseudo pedigree structure for method 5. Penalty on $\hat{\Sigma}$ applied to both G and E (\mathcal{P}_{Σ}^2), and penalty on $\log \hat{\lambda}$ applied to both $\log \hat{\lambda}$ and $\log(1 - \hat{\lambda})$ ($\mathcal{P}_{\lambda}^{\ell_2}$).

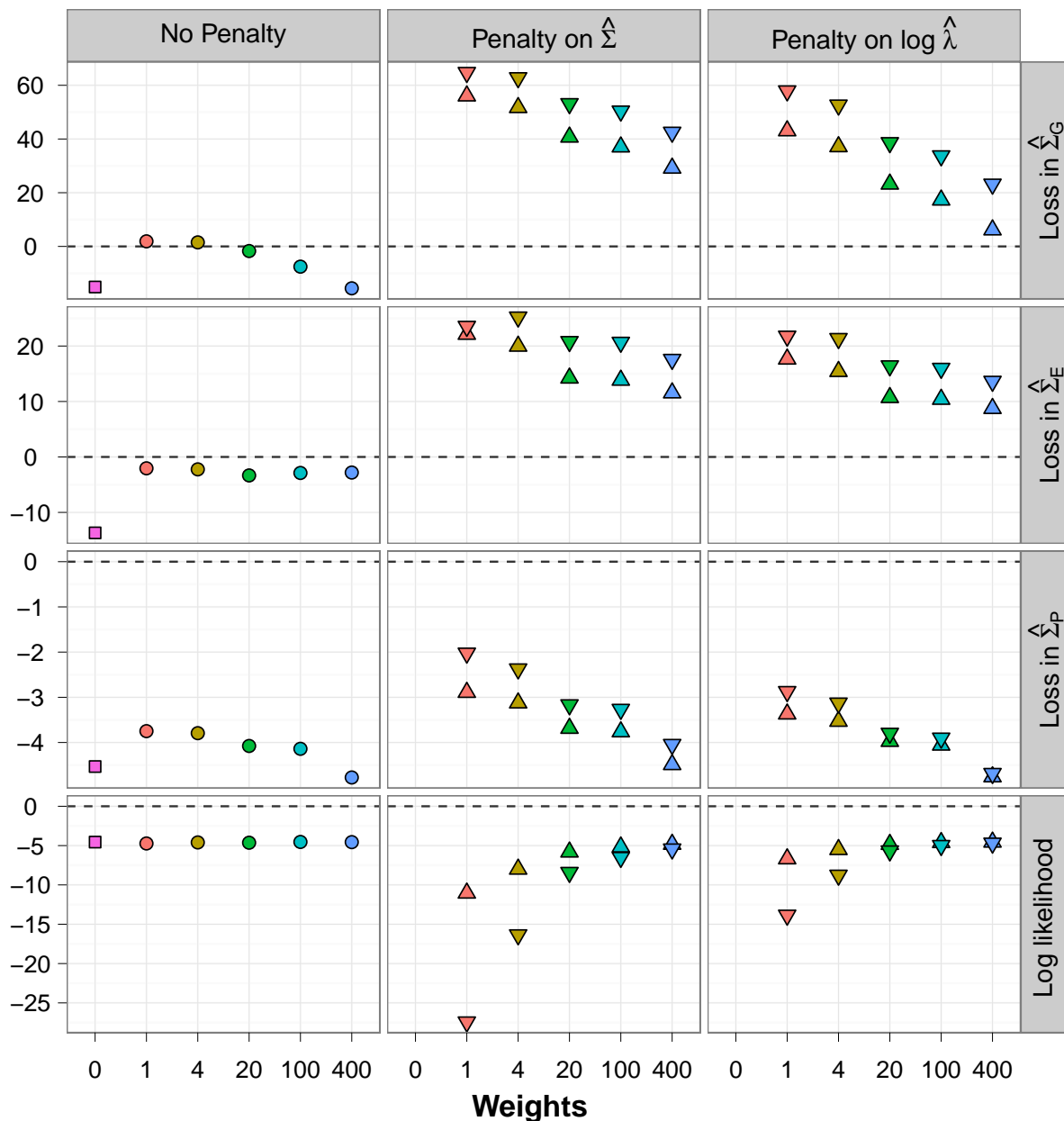


Figure 3. Mean percentage reduction in average loss for estimates of genetic (G), common environmental (C) residual (E) and phenotypic (P) covariance matrices together with corresponding log likelihood (log L)

Data for $q = 10$ traits, Bondari's design with $f = 250$ families. Methods as described in text: For 2C digits after hyphen give scaling factor used ($\times 100$). For methods 5, a hierarchical full-sib pseudo-pedigree structure was assumed throughout. For 5C, letters 'a', 'b' and 'c' define use of penalties \mathcal{P}_Σ , \mathcal{P}_Σ^{2b} and \mathcal{P}_Σ^3 , respectively, and 'd' denotes application of penalty $\mathcal{P}_\lambda^{\ell_2}$ to both the canonical eigenvectors of $\hat{\Sigma}_P^{-1}\hat{\Sigma}_G$ and $\hat{\Sigma}_P^{-1}\hat{\Sigma}_C$, using a tuning factor of $\psi = 0.02$ throughout.

▼ genetic, ▲ permanent environmental, ● residual and ◆ phenotypic covariance, and ■ log likelihood as deviation from value for multivariate analysis.

