Short communication: Methods to compute genomic inbreeding for ungenotyped individuals

A. Legarra,^{1*} I. Aguilar,² and J. J. Colleau³

¹UMR GenPhySE, INRA, Castanet-Tolosan 31320, France

²Instituto Nacional de Investigación Agropecuaria (INIA), 11100 Montevideo, Uruguay

³UMR GABI, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France

ABSTRACT

The genomic measure of inbreeding is closer to the actual inbreeding than the pedigree-based measure. However, it cannot be computed for ungenotyped animals. An estimate of genomic inbreeding comes from the diagonal of matrix $\mathbf{H}$ used in single-step methods. This matrix projects genomic relationships to all ungenotyped members of the pedigree. The diagonal element of $\mathbf{H}^{-1}$ gives an estimate of the genomic inbreeding coefficient. However, so far no computational methods are available to compute the diagonal of $\mathbf{H}$. Here we propose 3 exact methods to compute this diagonal. The first uses an already-existing algorithm to compute, for each ungenotyped individual, products of the form $\mathbf{H}\mathbf{x}$ to obtain the corresponding diagonal element of $\mathbf{H}$. The second method computes, for each ungenotyped individual, a term that can be written as a quadratic form involving pedigree and genomic relationships. For both methods, the computational burden is linear in the number of ungenotyped animals. The last method reorders the computations of the second method so that they become linear in the number of genotyped animals, which is usually much smaller. We tested the methods in 3 small data sets (with ~2,000 genotyped animals and 30,000–500,000 animals in pedigree) and in a large simulated population (with 1,220,000 animals in pedigree and 36,000 genotyped animals). Tests resulted in satisfactory computing times (<10 min in the largest example using 10 parallel threads). Computing times were much shorter for the third method, as expected. Using these methods, estimates of genomic inbreeding in ungenotyped animals can be obtained on a regular basis.

Key words: single step, inbreeding, genetic variability, genomic selection

Short Communication

Wright (1922) conceived the inbreeding coefficient to measure the degree of inbreeding per individual. In pedigree-based analysis of populations, inbreeding coefficients or their summaries (i.e., averages per year of birth) are typically used (1) to describe the evolution of genetic diversity in populations (Falconer and Mackay, 1996), (2) as a covariate to model inbreeding depression (de Boer and Hoeschele, 1993), and (3) to derive individual accuracies from genetic evaluations (Van Vleck, 1993).

Because the genome is of finite size, pedigree-based inbreeding coefficients are only expectations of the realized inbreeding coefficients (VanRaden, 2008; Hill and Weir, 2011). Information from genotypes allows more accurate genomic estimates of realized inbreeding coefficients (VanRaden, 2007, 2008). For genotyped animal i , genomic inbreeding is obtained as $F_{Gi} = G_{ii} - 1$, where $\mathbf{G}$ is the genomic relationship matrix (VanRaden, 2008). The inbreeding coefficient describes deviations from Hardy-Weinberg, and therefore values of $F_G < 0$ are interpreted as an individual more heterozygous than the average of the population. If base allele frequencies are used in the computation of $\mathbf{G}$, F_G is an estimate of inbreeding relative to the base population (VanRaden, 2008). If arbitrary frequencies of 0.5 are used, then F_G is related to molecular self-coancestry f_M (or homozygosity, or identity by state at the markers) as $2f_M = \frac{1}{2}(1 + F_G) + 1$ (Garcia-Baccino et al., 2017).

When animals are not genotyped, a linear projection of genomic relationships ($\mathbf{G}$) via pedigree relationships ($\mathbf{A}$) is achieved in the single-step matrix $\mathbf{H}$ (Legarra et al., 2009; Christensen and Lund, 2010). With subscripts 1 and 2 pertaining to ungenotyped and genotyped individuals, respectively, this matrix has the form

Received October 15, 2019.

Accepted December 18, 2019.

*Corresponding author: andres.legarra@inra.fr

$$\mathbf{H} = \begin{pmatrix} \mathbf{A}_{11} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G} \\ \mathbf{GA}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G} \end{pmatrix}.$$

Throughout this paper, we assume that matrices $\mathbf{G}$ and $\mathbf{A}$ are compatible, either using base allele frequencies in the computation of $\mathbf{G}$ (VanRaden, 2008) or, if these are not available, integrating over the unknown distribution of allele frequencies (Christensen, 2012) through the theory of metafounders (Legarra et al., 2015; Garcia-Baccino et al., 2017). Such compatibility is needed for correct estimates, although the algorithms below do not assume such compatibility.

It is generally accepted that relationships across all individuals are better described by $\mathbf{H}$ than by $\mathbf{A}$ (Meuwissen et al., 2016). Thus, inbreeding can be estimated as $F_{Hi} = H_{ii} - 1$ (i.e., through the diagonal of $\mathbf{H}$). For genotyped individuals, $F_{Hi} = F_{Gi} = G_{ii} - 1$ [i.e., VanRaden's (2008) estimator of inbreeding]. For ungenotyped individuals, the estimator is obtained as the diagonal of section $\mathbf{H}_{11} = \mathbf{A}_{11} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$ minus 1. It is of interest to obtain these elements (i.e., $F_{Hi} = H_{ii} - 1$ for ungenotyped individuals) first, to use comparable statistics of inbreeding for both genotyped and ungenotyped animals, and second, to extract the maximum information from the data (e.g., ungenotyped offspring from genomically related mates should be inbred).

Computational evaluation procedures of single-step genomic BLUP use the inverse of $\mathbf{H}$, so that the diagonal elements of $\mathbf{H}$ are not explicitly obtained. Recently, for computing evaluation accuracies, Xiang et al. (2017) computed the diagonal of $\mathbf{H}_{11}$ based on the sparse inverse (Takahashi et al., 1973) of the sparse matrix $\mathbf{H}^{-1}$. This procedure is too time and memory consuming for large data sets. For management of genomic variability in populations with partial genotyping (optimizing contributions of candidates for breeding), Colleau et al. (2017) presented a fast method to compute an operational section of $\mathbf{H}$ vector-wise, where the diagonal elements of this section are by-products. This method was an extension of Colleau's (2002) indirect method concerning matrix $\mathbf{A}$ (the pedigree relationship matrix) and involved the calculation of vector products $\mathbf{G}\mathbf{x}$ and $\mathbf{A}_{22}^{-1}\mathbf{x}$, where $\mathbf{x}$ vectors are working vectors. However, Colleau et al.'s (2017) method might be time consuming if the objective were to compute F_{Hi} for every ungenotyped individual in large populations. Thus, a fast method for computing the genomic inbreeding of ungenotyped individuals in large populations is needed.

Here we present 3 exact methods to compute the diagonal of $\mathbf{H}$. The first method is a straightforward

application of Colleau et al. (2017). The second method computes the full algebraic expression of the diagonal term for every ungenotyped individual. The third method is a vector extension of the second method, considerably reducing computation time. Three examples are given from dairy goat, dairy sheep, and a simulated dairy cattle population.

Let us consider a population pedigree of size n , where n_1 individuals are ungenotyped and n_2 individuals are genotyped. Computing the estimated genomic inbreeding coefficients of the ungenotyped individuals requires computing the diagonal of section $\mathbf{H}_{11}$ of the full genomic relationship matrix $\mathbf{H}$. This section is equal to section $\mathbf{A}_{11}$ of relationship matrix $\mathbf{A}$ plus the correction matrix $\mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$.

We will talk in the following about *procedures*. These can be seen as subroutines or functions, programmed efficiently. Procedure $\mathbf{Ax}$ involves the computation of product $\mathbf{Ax}$ with an appropriate algorithm (Colleau, 2002). Procedures $\mathbf{Gt}$ and $\mathbf{A}_{22}^{-1}\mathbf{t}$ are used for the computation of 2 matrix products, $\mathbf{Gt}$ and $\mathbf{A}_{22}^{-1}\mathbf{t}$, where vectors $\mathbf{t}$ are working vectors of size n_2 . These products can be computed using various numerical methods (Colleau et al., 2017). Finally, procedure $\mathbf{Hx}$ uses the 3 preceding procedures to compute the product $\mathbf{Hx}$ (Colleau et al., 2017).

Method 1

Consider vector $\mathbf{x}' = [\mathbf{x}'_1 \ \mathbf{x}'_2]$ of size $n = n_1 + n_2$. Vector $\mathbf{x}$ contains 0 values except for a single 1 for the ungenotyped individual considered. Then, method 1 consists of computing $\mathbf{y} = \mathbf{Hx}$ n_1 times (one time for each ungenotyped individual) using procedure $\mathbf{Hx}$ and collecting each time the appropriate element of vector $\mathbf{y}$, which is the corresponding element of the diagonal of $\mathbf{H}$. Finally, products $\mathbf{Ax}$, $\mathbf{Gt}$, and $\mathbf{A}_{22}^{-1}\mathbf{t}$ are used $2n_1$, n_1 , and $2n_1$ times, respectively.

Method 2

For ungenotyped animals, $H_{ii} = A_{ii} + c_i$ and $F_{Hi} = F_{Ai} + c_i$, where F_{Ai} is the pedigree inbreeding coefficient. The correction term c_i is equal to the quadratic $c_i = \mathbf{a}_{i,2}\mathbf{A}_{22}^{-1}\mathbf{GA}_{22}^{-1}\mathbf{a}_{2,i} - \mathbf{a}_{i,2}\mathbf{A}_{22}^{-1}\mathbf{a}_{2,i} = \mathbf{a}'_{2,i}\mathbf{Ma}_{2,i}$, where $\mathbf{a}'_{2,i} = \mathbf{a}_{2,i}$ (the i th column of matrix $\mathbf{A}_{21}$) is the vector of relationships between ungenotyped individual i with all the genotyped individuals and $\mathbf{M} = \mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}$.

As a result, procedure $\mathbf{Ax}$, procedure $\mathbf{Gt}$, and procedure $\mathbf{A}_{22}^{-1}\mathbf{t}$ are used n_1 times, which almost halves the

computation load of method 1. Therefore, the algorithm for each ungenotyped individual i is as follows:

1. Compute and store vector $\mathbf{a}_{2,i}$ using procedure $\mathbf{Ax}$.
2. Compute
 - a. $\mathbf{w} = \mathbf{A}_{22}^{-1} \mathbf{a}_{2,i}$
 - b. $c_i = \mathbf{w}'(\mathbf{Gw}) - \mathbf{a}_{2,i}'\mathbf{w}$.

Method 3

The correction term of method 2 pertaining to ungenotyped individual i is the quadratic $c_i = \mathbf{a}_{2,i}'\mathbf{M}\mathbf{a}_{2,i}$. This correction term can be rewritten, for each individual, as the sum of all the terms of the Hadamard product between matrix $\mathbf{M}$ and matrix $\mathbf{a}_{2,i}\mathbf{a}_{2,i}'$, as follows:

$$c_i = \sum_{j \in \text{genotyped}} \sum_{k \in \text{genotyped}} m_{j,k} a_{j,i} a_{k,i} = \sum_{j \in \text{genotyped}} a_{j,i} \sum_{k \in \text{genotyped}} m_{j,k} a_{k,i},$$

where $a_{j,i}$ and $a_{k,i}$ are the pedigree relationships between individual i (ungenotyped) and individuals j and k (genotyped). The fact that the summation above can be reordered suggests computing correction terms by accumulation of the individual contributions of the genotyped individuals. Computations can be carried out by 2 successive $\mathbf{Ax}$ procedures implemented by the indirect method (Colleau, 2002) and executed for each genotyped individual j . The role of each procedure is to provide the appropriate terms under each summation symbol. The whole algorithmic sequence is as follows. Due to the method of construction, parallel threading is allowed. Set vector $\mathbf{c}$ to 0. Set up working vectors $\mathbf{u}$ and $\mathbf{v}$. For each genotyped individual j :

1. Compute $\mathbf{u} = \mathbf{Ax}$, where $\mathbf{x}_1 = 0$ and $\mathbf{x}_2 = 0$ except for 1 at position j and extract $\mathbf{u}_1$ (the part corresponding to ungenotyped individuals). Vector $\mathbf{u}_1$ gathers the terms $a_{j,i}$ under the first summation symbol, simultaneously for all ungenotyped individuals.
2. Either extract $\mathbf{m}_j$ from $\mathbf{M} = \mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}$ or compute $\mathbf{m}_j$ as $\mathbf{m}_j = \mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{x}$, where vector $\mathbf{x}$ contains 1 for individual j and 0 otherwise. Vector $\mathbf{m}_j$ can be efficiently computed using the sparse decomposition of $\mathbf{A}_{22}^{-1}$ (Masuda et al., 2017) and the representation of $\mathbf{Gx}$ as $\mathbf{Z}(\mathbf{Z}'\mathbf{x})/2\Sigma(p_i q_i)$.

3. Compute $\mathbf{v}_1$ from $\mathbf{v} = \mathbf{Ax}$, where $\mathbf{x}_1 = 0$ and $\mathbf{x}_2 = \mathbf{m}_j$ and extract $\mathbf{v}_1$. This vector gathers the terms $m_{jk}a_{k,i}$ under the second summation symbol, simultaneously for all ungenotyped individuals.
4. Update $\mathbf{c}$ by $\mathbf{c} = \mathbf{c} + \mathbf{u}_1 \odot \mathbf{v}_1$, where $\odot$ is the Hadamard product. In other words, $c_i = c_i + u_{1,i}v_{1,i}$. This computes contributions of j to vector $\mathbf{c}$ based on vectors $\mathbf{u}_1$ and $\mathbf{v}_1$.

Equivalent algebra leading to this algorithm is given in the Appendix. Finally, procedure $\mathbf{Ax}$, procedure $\mathbf{Gt}$, and procedure $\mathbf{A}_{22}^{-1}\mathbf{t}$ are used $2n_2$, n_2 , and $2n_2$ times, respectively, and full storage of matrix $\mathbf{M}$ is needed only for the current column during the update. Then, the overall computation time is equal to n_2/n_1 times the one of method 1. The cost relative to method 2 is not a simple function of n_2 and n_1 because it depends on the relative costs of the 3 basic procedures.

Performance of the preceding algorithms was tested in 4 data sets, 3 of them of moderate size and 1 of large size. Two dairy goat data sets were used with $n_1 = 30,067$, $n_2 = 1,643$ and $n_1 = 34,818$, $n_2 = 2,477$ for breeds Alpine and Saanen, respectively. In dairy sheep (Manech Tete Rousse breed), $n_1 = 498,515$, $n_2 = 2,111$. Finally, we simulated a large dairy cattle data set with 20 generations including $n_1 = 1,220,889$, $n_2 = 36,518$ animals genotyped (sires from generation 11 to 19 and all males in the last generation). This population was composed to a large extent of half-sib families (1 progeny per female). In all cases, $\mathbf{G}$ and $\mathbf{A}_{22}^{-1}$ were explicitly computed. Computing time is reported without considering computation of these matrices because this computation time is highly sensitive to the use of parallel libraries and subroutines. Two different programs were used: homemade Fortran programs for small ruminants (developed with no threading) and preGSf90 (part of the blupf90 suite of programs) in the simulated data (<http://nce.ads.uga.edu/wiki/doku.php>). In particular, preGSf90 uses a parallel implementation of method 3. For this particular data set, 10 threads were used. Computation times (including computation of $\mathbf{M}$) are shown in Table 1. It can be seen that, as expected, method 2 halves computation times of method 1, but computation times of method 2 are still far from acceptable for large data files such as Manech Tete Rousse. Computation time reduces (theoretically) from method 1 to method 3 by n_2/n_1 , and this prediction agrees well with Table 1.

There is no doubt about the potential interest of computing genomic inbreeding coefficients in animal breeding. However, so far, no clear computation method is available in the now-frequent case of mixtures of

Table 1. Number of ungenotyped (n_1) and genotyped (n_2) animals and wall-clock computation time (min) of inbreeding coefficients from matrix $\mathbf{H}$ for different data sets: dairy goat (Alpine, Saanen), dairy sheep (Manech Tete Rousse), and simulated dairy cattle

Data set	n_1	n_2	Method 1	Method 2	Method 3
Saanen	34,818	2,477	9.6	6	0.4
Alpine	30,067	1,643	25	19	1.2
Manech Tete Rousse	498,515	2,111	327	169	1.4
Simulated dairy cattle	1,220,889	36,518	—	—	9

genotyped and ungenotyped individuals. Here we have addressed this issue. In most ruminant populations, the number of ungenotyped animals (n_1) vastly outnumbers the number of genotyped animals (n_2). We have shown that the more straightforward approaches are linear in the number of ungenotyped animals and acceptable for small to medium data sets, whereas method 3 is linear in the number of genotyped animals and consequently appropriate for large data files.

The methodology presented here can readily be extended to the use of metafounders to account for missing pedigrees or different origins or lines. Metafounders have relationships γ , including also self-relationships (Legarra et al., 2015). In that sense, an animal born from metafounder i has a pedigree inbreeding coefficient $\gamma_{i,i}/2$ and an animal born from metafounders i,j has a pedigree inbreeding coefficient $\gamma_{i,j}/2$ instead of 0. Legarra et al. (2015) presented methods for computation of $\mathbf{A}\mathbf{x}$ and $\mathbf{A}_{22}^{-1}\mathbf{t}$ considering metafounders.

All methods give the same results as they all compute the diagonal of matrix $\mathbf{H}$ exactly. We conclude that method 3, which uses contributions of genotyped animals to update pedigree inbreeding of ungenotyped animals, is the best performing and can be used in very large databases. Methods 1 and 2 may be simpler to program and may be used in small to medium databases.

ACKNOWLEDGMENTS

This study was partially funded by the INRA (Paris, France) SELGEN funding metaprogram (Project GDivSelGen) and FEDER (Poctefa funds, project ARDI). This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement no. 772787-SMARTER. The authors have not stated any conflicts of interest.

REFERENCES

- Christensen, O. F. 2012. Compatibility of pedigree-based and marker-based relationship matrices for single-step genetic evaluation. *Genet. Sel. Evol.* 44:37. <https://doi.org/10.1186/1297-9686-44-37>.
- Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42:2. <https://doi.org/10.1186/1297-9686-42-2>.
- Colleau, J. J. 2002. An indirect approach to the extensive calculation of relationship coefficients. *Genet. Sel. Evol.* 34:409–421. <https://doi.org/10.1186/1297-9686-34-4-409>.
- Colleau, J.-J., I. Palhière, S. T. Rodríguez-Ramilo, and A. Legarra. 2017. A fast indirect method to compute functions of genomic relationships concerning genotyped and ungenotyped individuals, for diversity management. *Genet. Sel. Evol.* 49:87. <https://doi.org/10.1186/s12711-017-0363-9>.
- de Boer, I. J., and I. Hoeschele. 1993. Genetic evaluation methods for populations with dominance and inbreeding. *Theor. Appl. Genet.* 86:245–258. <https://doi.org/10.1007/BF00222086>.
- Falconer, D. S., and T. F. C. Mackay. 1996. *Introduction to Quantitative Genetics*. Longman, New York, NY.
- Garcia-Baccino, C. A., A. Legarra, O. F. Christensen, I. Misztal, I. Pocrnic, Z. G. Vitezica, and R. J. C. Cantet. 2017. Metafounders are related to Fst fixation indices and reduce bias in single-step genomic evaluations. *Genet. Sel. Evol.* 49:34. <https://doi.org/10.1186/s12711-017-0309-2>.
- Hill, W. G., and B. S. Weir. 2011. Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genet. Res. (Camb.)* 93:47–64. <https://doi.org/10.1017/S0016672310000480>.
- Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92:4656–4663. <https://doi.org/10.3168/jds.2009-2061>.
- Legarra, A., O. F. Christensen, Z. G. Vitezica, I. Aguilar, and I. Misztal. 2015. Ancestral relationships using metafounders: Finite ancestral populations and across population relationships. *Genetics* 200:455–468. <https://doi.org/10.1534/genetics.115.177014>.
- Masuda, Y., I. Misztal, A. Legarra, S. Tsuruta, D. A. L. Lourenco, B. O. Fragomeni, and I. Aguilar. 2017. Avoiding the direct inversion of the numerator relationship matrix for genotyped animals in single-step genomic best linear unbiased prediction solved with the preconditioned conjugate gradient. *J. Anim. Sci.* 95:49–52.
- Meuwissen, T., B. Hayes, and M. Goddard. 2016. Genomic selection: A paradigm shift in animal breeding. *Anim. Front.* 6:6–14. <https://doi.org/10.2527/af.2016-0002>.
- Minka, T. P. 2000. Old and new matrix algebra useful for statistics. Accessed Sep. 30, 2019. <https://tminka.github.io/papers/matrix/>.
- Takahashi, K., J. Fagan, and M. Chen. 1973. Formation of sparse bus impedance matrix and its application to short circuit study. Pages 63–69 in 8th Power Industry Computer Applications Conference. IEEE, New York, NY.
- Van Vleck, L. D. 1993. Variance of prediction error with mixed model equations when relationships are ignored. *Theor. Appl. Genet.* 85:545–549. <https://doi.org/10.1007/BF00220912>.
- VanRaden, P. M. 2007. Genomic measures of relationship and inbreeding. *Interbull Bull.* 37:33–36.
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423. <https://doi.org/10.3168/jds.2007-0980>.
- Wright, S. 1922. Coefficients of inbreeding and relationship. *Am. Nat.* 56:330–338. <https://doi.org/10.1086/279872>.
- Xiang, T., O. F. Christensen, and A. Legarra. 2017. Technical note: Genomic evaluation for crossbred performance in a single-step ap-

proach with metafounders. J. Anim. Sci. 95:1472–1480. <https://doi.org/10.2527/jas2016.1155>.

ORCIDS

A. Legarra <https://orcid.org/0000-0001-8893-7620>

I. Aguilar <https://orcid.org/0000-0002-1038-4752>

APPENDIX

As another derivation of method 3, Minka (2000) showed that for any two matrices $\mathbf{A}$, $\mathbf{B}$ $diag(\mathbf{AB}') = (\mathbf{A} \odot \mathbf{B})\mathbf{1}$, where $diag(\mathbf{K})$ is a vector with

the diagonal elements of matrix $\mathbf{K}$ and $\odot$ is the Hadamard product. Applied to our problem, $\mathbf{c} = diag(\mathbf{H} - \mathbf{A}) = diag(\mathbf{A}_{12}\mathbf{MA}_{21}) = (\mathbf{A}_{12}\mathbf{M} \odot \mathbf{A}_{12})\mathbf{1}$. The product by 1 is simply a summation, whereas $\mathbf{A}_{12}\mathbf{M}$ can be computed as $\mathbf{A}_{12}\mathbf{M} = \sum_i \mathbf{A}_{12}\mathbf{M}_i^*$, where $\mathbf{M}_i^*$ contains the i th column of $\mathbf{M}$ and 0 otherwise. Putting all these steps together, we can compute $(\mathbf{A}_{12}\mathbf{M} \odot \mathbf{A}_{12})\mathbf{1}$, in form of contributions due to each genotyped individual, as the cumulated sum of the direct product of vectors $\mathbf{v}_1$ and $\mathbf{u}_1$, where $\mathbf{v}_1$ is columns of $\mathbf{A}_{12}\mathbf{M}$ and $\mathbf{u}_1$ is columns of $\mathbf{A}_{12}$. This is the algorithm in method 3.