

A method for the dynamic management of genetic variability in dairy cattle

Jean-Jacques COLLEAU^{a*}, Sophie MOUREAUX^{a,b}, Michèle BRIEND^a,
Jérôme BECHU^c

^a Station de génétique quantitative et appliquée, Institut national de la recherche agronomique,
78352 Jouy-en-Josas Cedex, France

^b Institut de l'élevage, 75595 Paris Cedex 12, France

^c Génétique Normande Avenir, 61700 Domfront, France

(Received 18 August 2003; accepted 1 March 2004)

Abstract – According to the general approach developed in this paper, dynamic management of genetic variability in selected populations of dairy cattle is carried out for three simultaneous purposes: procreation of young bulls to be further progeny-tested, use of service bulls already selected and approval of recently progeny-tested bulls for use. At each step, the objective is to minimize the average pairwise relationship coefficient in the future population born from programmed matings and the existing population. As a common constraint, the average estimated breeding value of the new population, for a selection goal including many important traits, is set to a desired value. For the procreation of young bulls, breeding costs are additionally constrained. Optimization is fully analytical and directly considers matings. Corresponding algorithms are presented in detail. The efficiency of these procedures was tested on the current Norman population. Comparisons between optimized and real matings, clearly showed that optimization would have saved substantial genetic variability without reducing short-term genetic gains.

relationship coefficient / mating / optimization / breeding scheme

1. INTRODUCTION

The selection tools currently available for the selection of dairy cattle populations have been shown to be very efficient for generating short and mid-term genetic gains. However, theory has shown that inbreeding and kinship rates are likely to increase very fast. Such predictions can be very easily verified on real populations that exhibit the very narrow gene pool actually available for selection [3,23]. Expected long-term detrimental consequences are reduced ultimate

* Corresponding author: ugenccj@dga2.jouy.inra.fr

genetic gains, a direct impact on performances due to inbreeding depression, especially for functional traits, and an increased expression of genetic defects.

Quantitative geneticists have long been investigating the practical methods to be developed accordingly. They have succeeded in proposing breeding schemes more efficient than the reference one, *i.e.*, a scheme where parents are selected by truncation, used at uniform rates (within each sex) and mated randomly. The first category of research concerns selection methods of parents and determination methods of their contribution to future progeny. The earliest attempts have modified selection indices by inflating genetic parameters [14, 20, 37] or by decreasing the weight of familial information *vs.* the weight of individual information [11, 20, 34, 41] or by including penalties for individual's inbreeding coefficient [20, 37] or for the average coancestry between individual and the rest of the population [4, 5, 41]. The most advanced proposal consists of determining selection of parents and their future contribution after optimizing a decision rule, in general after maximizing genetic gains, based on true estimated breeding values (EBV) and given a certain level of accepted inbreeding rate [4, 5, 15, 16, 21, 22, 26, 27, 35, 36, 40]. Compared to a reference scheme, the last implementation is able to enhance genetic gains by several tens of %, reasoning at the same level of inbreeding coefficients. More rarely [31], authors have proposed to optimize inbreeding for a certain level of desired genetic gain. An additional research area concerns the mating design. First, factorial matings have been shown to be preferable to hierarchical matings [29]. Improvements easy to implement such as compensatory matings have been found to be already effective [6]. The last version consists in optimizing a criterion, *e.g.*, average coancestry between parents given their optimized contributions [25, 28]. As compared to the optimization followed by random matings, this second optimization decreases inbreeding rates and increases selection responses substantially. The theory of long-term contributions provides a consistent understanding of these achievements [29, 38, 39].

Most generally, optimization of selection and optimization of matings are proposed sequentially. However, if the problem under study allows one to merge these two steps into a single step, then the so-called "mate selection" [1] is implemented. It was basically imagined for optimizing some utility function in various genetic contexts, including non-additive genetic models [18]. Examples concerning the management of diversity in selected populations are given in [12, 19, 32, 33] where the best combinations of matings are chosen for fulfilling the objective: here matings determine parents *a posteriori*. The literature does not provide clear indication on whether this procedure differs significantly in terms of efficiency from the previous one [7, 25, 39].

However, Fernandez and Caballero [13] found out that the single step approach was definitely creating more inbreeding than a two-step approach.

The objective of this paper was to present a fully analytical mate selection method, for managing genetic variability in dairy cattle selection. Some operational constraints were accounted for because a major concern was the applicability by practitioners. The theory employed was fully detailed. The potential of the approach was assessed on a real population.

2. GENERAL OUTLINE OF THE APPROACH

Ideally, optimizing matings in real dairy cattle populations, given certain pre-defined constraints, would lead one to program simultaneously the birth of young males to be further progeny-tested and the birth of young females for general use. Meanwhile, since dairy cattle selection occurs in the context of overlapping generations, some cohorts of previous animals should be accounted for. Previous male cohorts are made up of service bulls, already available for use, of young bulls recently progeny-tested and young bulls still waiting for a progeny-test to be completed. Previous cohorts of females are constituted of cohorts of females available for artificial insemination (AI) and of females still too young for breeding. Then, the best solution for matings can be formally established. However, it would be quite difficult to find out the corresponding global solution for real populations, usually of large size, due to the initial huge number of possible matings. Furthermore, except for selection nuclei where matings can be programmed, they are basically dependent on breeders' preferences. Consequently, optimal matings concerning the general population can only be calculated as guidelines for extension services. However, they can be used to some extent for male selection (see further).

Then, a possible practical approach consists of splitting the overall optimization into three distinct steps:

- (i) procreation of young bulls to be progeny-tested (and possibly, procreation of young females within selection nuclei);
- (ii) use of service bulls on non-elite cows;
- (iii) approval of recently progeny-tested bulls for AI use.

The objective of this paper was to present the corresponding analytical approaches in full detail. Despite this division, the methods used for each specific step share common characteristics.

First, the objective was to minimize the average pairwise relationship coefficient (including self-relationships) in the population of individuals to be born

and of existing individuals, so as to maximize the number of founder genes still present [10]. In the same line, Caballero and Toro [7] point out that the average pairwise coancestry coefficient (f according to their notation) of the whole population at a given time indicates the expected fraction (over conceptual replications) of the initial allelic variability which was lost by drift. Consequently, they consider that the difference $1 - f$ is an appropriate measure of diversity. Furthermore, it can be observed that the average relationship coefficient in the generation of progeny is not exactly the same as the average relationship between parents weighted by their contributions to the generation of progeny, because Mendelian sampling should be accounted for. The necessary correction favors inbred parents because they are more protected against within-family drift.

Second, as a major constraint, the average EBV of the future individuals for an overall combination of many traits of economical importance, was set to a desired value. This operational choice was preferred to the symmetrical approach (*i.e.*, constraining the average pairwise relationship coefficient while maximizing the average EBV), because it is thought that practitioners might be unefficient, because reluctant, if major emphasis were given to a parameter they are still unfamiliar with. However, this attitude might change in the future, thus allowing one to switch the constraint.

Third, the optimization was formally single-stepped *i.e.*, and directly considered a non-linear function of the frequencies of the full set of possible matings.

Fourth, an implicit penalty against large full-sib families was introduced so as to favor factorial matings, since this type of matings has been generally found able to generate higher potential genetic gains in the progeny [25, 28].

3. PROCREATION OF YOUNG BULLS

3.1. Outline of the strategy

The breeding organization aims at producing N young bulls. These future young bulls are to be compared to N_0 previous young bulls still awaiting completion of the progeny-test. The objective is to minimize the average pairwise relationship coefficient between these $N + N_0$ bulls. Despite the overlapping generation context, all these bulls have not yet started their breeding career: their expected future contributions to the population are the same and consequently, these pairwise comparisons are not to be weighted.

N_s sires and N_d dams were chosen in the current population to be candidates for matings. The techniques described further allows one to use large values

of N_s and N_d , for decreasing the risk of discarding valuable candidates. Then $n = N_s \times N_d$ matings are possible and have to be examined. Let \mathbf{x} be the vector ($n \times 1$) of the internal mating frequencies. Then, $\mathbf{1}'_n \mathbf{x} = 1$. The term corresponding to mating between sire i and dam j is noted x_{ij} . Its position in vector \mathbf{x} can be easily recovered if, for instance, \mathbf{x} is the linearized matrix of mating frequencies (*sire* \times *dam*), *i.e.*, the frequencies of the mating sequence

$$s_1 d_1, \dots, s_1 d_{N_d}, \dots, s_{N_s} d_1, \dots, s_{N_s} d_{N_d}.$$

Then, the position of mating ij is $k = (i - 1)N_d + j$. The corresponding vector of estimated breeding values is of the same dimension and is noted \mathbf{b} . Each element is equal to the average of the EBV of the parents involved. If the average EBV of matings is set to B , the desired value, then $\mathbf{b}'\mathbf{x} = B$. An additional constraint is included so that the overall breeding costs should be equal to some desired value E . This will be detailed further.

Then, optimal solutions for \mathbf{x} are searched in a continuum, using a full analytical approach. The final step consists of taking into account practical constraints. Practitioners are able (or willing) to carry out only a limited number of breeding alternatives per cow, with corresponding cow prolificities, breeding costs and a maximum number of sires allocated. For instance, they may envision a single AI or a single embryo collection after superovulation or one collection followed by AI or two collections. Then, the continuum is progressively destroyed to meet these constraints and to provide solutions ready for practical use, *i.e.* assigning cows to each breeding alternative and appropriate mating(s) to each reproduction step.

3.2. Finding the optimal continuum without an economical constraint

For simplicity, we show how solutions are obtained without the economical constraint, which brings specific difficulties. We consider the population of the N_0 previous animals and of the n individuals corresponding to matings. The vector \mathbf{f} of the individual frequencies is:

$$\mathbf{f} = \frac{1}{N + N_0} \begin{pmatrix} \mathbf{1}_{N_0} \\ N\mathbf{x} \end{pmatrix}.$$

The corresponding relationship matrix \mathcal{A} is equal to

$$\begin{pmatrix} \mathbf{A}_0 & \mathbf{A}_1 \\ \mathbf{A}'_1 & \mathbf{A} \end{pmatrix}.$$

The average pairwise relationship coefficient is equal to $f' \mathcal{A} f$. However, here, matings are not existing individuals, *i.e.*, the Nx 's correspond to the expected sizes of full-sib families. These values may be not integer and may be lower or higher than 1. Using expected full-sib families in the quadratic form, instead of individuals, introduces the penalty against large full-sib families alluded to previously, because full-sibs are considered as sharing the same Mendelian sampling.

Processing further, we can express the quadratic form as

$$\frac{1}{(N + N_0)^2} (\mathbf{1}'_{N_0} \mathbf{A}_0 \mathbf{1}_{N_0} + 2N \mathbf{1}'_{N_0} \mathbf{A}_1 \mathbf{x} + N^2 \mathbf{x}' \mathbf{A} \mathbf{x}).$$

Minimizing this expression leads to the same solutions as minimizing

$$\frac{1}{2} \mathbf{x}' \mathbf{A} \mathbf{x} + \frac{1}{N} \mathbf{1}'_{N_0} \mathbf{A}_1 \mathbf{x}, \quad \text{i.e.,} \quad \frac{1}{2} \mathbf{x}' \mathbf{A} \mathbf{x} + \mathbf{p}' \mathbf{x}$$

where term p_i is the sum of the relationships between mating i and all the previous animals, divided by N . The optima should be found using the corresponding Lagrange function incorporating two constraints, *i.e.*,

$$L(\mathbf{x}) = \frac{1}{2} \mathbf{x}' \mathbf{A} \mathbf{x} + \mathbf{p}' \mathbf{x} - \lambda_1 (\mathbf{b}' \mathbf{x} - B) - \lambda_2 (\mathbf{1}'_n \mathbf{x} - 1)$$

where the λ are Lagrange multipliers. Finding the zeros of derivatives with respect to \mathbf{x} and the λ leads to the following linear system:

$$\begin{pmatrix} \mathbf{A} & -\mathbf{b} & -\mathbf{1}_n \\ \mathbf{b}' & \mathbf{0} & \mathbf{0} \\ \mathbf{1}'_n & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} -\mathbf{p} \\ B \\ 1 \end{pmatrix}.$$

The direct solution is not attempted because matrix \mathbf{A} is usually of very large size (billions of terms can easily be involved even when using the splitting approach described previously). This system is solved iteratively using the conjugate gradient method (CG) [30]. The major task corresponds to calculating $\mathbf{A} \mathbf{x}$ repeatedly, and is executed using the fast exact method described in [8]. Furthermore, this method allows one to deal with very large \mathbf{A} matrices because in reality they are not calculated and stored. In contrast with the situation met when implementing animal model BLUP evaluations, the inverse of this matrix is not sparse. Unfortunately, a counterpart of the fast exact method for calculating products such as $\mathbf{A}^{-1} \mathbf{y}$, without inverting \mathbf{A} , does not exist.

It can be shown that, symmetrically, the solution obtained maximizes the average EBV when the average relationship coefficient is constrained to be the

final average relationship found by this approach (Appendix A). Outer iterations are needed because some negative terms can be found. Then, they are set (fixed) to zero and new optimizations are run on the unfixed (variable) terms until variable terms are all positive. This procedure can be justified based on theoretical grounds (Appendix B). We detail how these outer iterations are carried out. Let \mathbf{x}_F be the vector of the n_F frequencies set to zero and let \mathbf{x}_V be the vector of the n_V frequencies still variable. Matrix A can be subdivided into

$$\begin{pmatrix} \mathbf{A}_{FF} & \mathbf{A}_{FV} \\ \mathbf{A}'_{FV} & \mathbf{A}_{VV} \end{pmatrix}.$$

The problem amounts to minimize

$$\frac{1}{2} \mathbf{x}'_V \mathbf{A}_{VV} \mathbf{x}_V + \mathbf{p}'_V \mathbf{x}_V - \lambda_1 (\mathbf{b}'_V \mathbf{x}_V - B) - \lambda_2 (\mathbf{1}'_{n_V} \mathbf{x}_V - 1)$$

which is obtained by solving the system

$$\begin{pmatrix} \mathbf{A}_{VV} & \mathbf{b}_V & -\mathbf{1}_{n_V} \\ \mathbf{b}'_V & \mathbf{0} & \mathbf{0} \\ \mathbf{1}'_{n_V} & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{x}_V \\ \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} -\mathbf{p}_V \\ B \\ 1 \end{pmatrix}.$$

When implementing CG, product $\mathbf{A}_{VV} \mathbf{x}_V$ is obtained by extracting the appropriate terms from the product

$$A \begin{pmatrix} \mathbf{0} \\ \mathbf{x}_V \end{pmatrix}$$

executed as a whole by the fast method.

3.3. Finding the optimal continuum with an economical constraint

The simplest way of addressing economical constraint would be considering the cost of individual reproduction steps. However, this might be inappropriate due to the system of mating contracts with breeders. For instance, the cost of a calf born from a cow contracted for a single AI may differ from the cost of a calf born from AI following an embryo collection on the contracted cow. Then, we prefer to consider the economical issue reasoning at the cow level, *i.e.*, per type of contract. In this way, the approach is not dependent on the assumption of “additivity” of costs and is still correct if this assumption holds.

Let the vector of reproduction rates per dam be denoted \mathbf{r} , of dimension $(N_d \times 1)$. For dam j , $r_j = \sum_{i=1}^{i=N_s} x_{ij}$. The corresponding vector of breeding costs

is e (like “expenses” or “euros”), of the same dimension. In practice, practitioners can implement l different breeding alternatives. Alternative k leads to the average prolificity ρ_k (*i.e.*, the average absolute prolificity, not necessarily an integer, divided by N). The corresponding cost is ϵ_k . For calculations within a continuum, cost $e(r)$ needs to be rendered continuous. This can be carried out by regression or better, by using the Lagrange interpolation polynomial [30] exact for any r belonging to the allowed set of reproduction levels. We translate the obvious condition that $e = 0$ when $r = 0$, into an extra level $k = 0$, with $\rho_0 = \epsilon_0 = 0$. Then, the degree of polynomial is l . Finally,

$$e(r) = \sum_{k=1}^{k=l} \frac{\prod_{k'=0, k' \neq k}^{k'=l} (r - \rho_{k'})}{\prod_{k'=0, k' \neq k}^{k'=l} (\rho_k - \rho_{k'})} \epsilon_k.$$

This expression yields the coefficients α of the working polynomial

$$e(r) = \sum_{k=1}^{k=l} \alpha_k r^k.$$

Subscript r will be dropped further for simplicity. Then, we have to minimize the Lagrange function

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \frac{1}{2} \mathbf{x}' \mathbf{A} \mathbf{x} + \mathbf{p}' \mathbf{x} - \lambda_1 (\mathbf{b}' \mathbf{x} - B) - \lambda_2 (\mathbf{1}'_n \mathbf{x} - 1) - \lambda_3 (\mathbf{1}'_{N_d} \mathbf{e} - E)$$

i.e.,

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \frac{1}{2} \mathbf{x}' \mathbf{A} \mathbf{x} + \mathbf{p}' \mathbf{x} - \boldsymbol{\lambda}' \mathbf{c}(\mathbf{x}).$$

The chosen iterative resolution method is a projected Lagrangian method [24]. It requires the calculation of the gradient vector

$$\mathbf{g} = \frac{\partial L(\mathbf{x}, \boldsymbol{\lambda})}{\partial \mathbf{x}}$$

and of the Hessian matrix

$$\mathbf{H} = \frac{\partial^2 L(\mathbf{x}, \boldsymbol{\lambda})}{\partial \mathbf{x} \partial \mathbf{x}'}$$

In our case,

$$\mathbf{g} = \mathbf{A} \mathbf{x} + \mathbf{p} - \lambda_1 \mathbf{b} - \lambda_2 \mathbf{1}_n - \lambda_3 \frac{\partial \mathbf{1}'_{N_d} \mathbf{e}}{\partial \mathbf{x}}.$$

The derivative of $\mathbf{1}'_{N_d} \mathbf{e}$ with respect to the frequency of mating ij (involving dam j) is equal to

$$\frac{\partial e_j}{\partial r_j} \frac{\partial r_j}{\partial x_{ij}} = \frac{\partial e_j}{\partial r_j}$$

with

$$\frac{\partial e_j}{\partial r_j} = \sum_{k=1}^{k=l} k\alpha_k r_j^{k-1} = y_j$$

where vector \mathbf{y} has the same dimension as \mathbf{r} or \mathbf{e} . Then, the derivative of $\mathbf{1}'_{N_d} \mathbf{e}$ with respect to \mathbf{x} is vector \mathbf{z} , where terms pertaining to the same dam are identical. Then

$$\mathbf{g} = \mathbf{A}\mathbf{x} + \mathbf{p} - \mathbf{C}\lambda$$

where matrix

$$\mathbf{C} = (\mathbf{b} \ \mathbf{1}_n \ \mathbf{z})$$

depends on the current value of \mathbf{x} through its third column.

Then

$$\mathbf{H} = \mathbf{A} - \lambda_3 \frac{\partial \mathbf{z}}{\partial \mathbf{x}'}$$

The last derivative is matrix \mathbf{W} , block diagonal. For block j corresponding to dam j , all the terms are equal, because

$$\frac{\partial z_{ij}}{\partial x_{ij}} = \frac{\partial y_j}{\partial r_j} \frac{\partial r_j}{\partial x_{ij}} = \frac{\partial y_j}{\partial r_j} = \sum_{k=2}^{k=l} k(k-1)\alpha_k r_j^{k-2}.$$

Before giving the detailed resolution algorithms, the major characteristics of the projected Lagrangian method are recalled. First, current estimates of Lagrange multipliers ($\tilde{\lambda}$) are used and second, constraints are linearized locally, conditionally on the current value $\tilde{\mathbf{x}}$ for unknowns. Then, the vector of constraint functions becomes

$$\mathbf{c}^*(\mathbf{x}) = \mathbf{c}(\tilde{\mathbf{x}}) + \mathbf{C}(\tilde{\mathbf{x}})(\mathbf{x} - \tilde{\mathbf{x}}).$$

It has been shown that the correct corresponding Lagrange function is

$$L(\mathbf{x}, \tilde{\lambda}) + \tilde{\lambda}' \mathbf{C}(\tilde{\mathbf{x}})\mathbf{x}.$$

This new Lagrange function is approximated by a second order Taylor expansion and finally, the optimal search direction $\Delta \mathbf{x}$ turns out to be equal to:

$$\Delta \mathbf{x} = \mathbf{C}\mathbf{u} + \mathbf{T}(\mathbf{T}'\mathbf{H}\mathbf{T})^{-1} \mathbf{T}'(\mathbf{H}\mathbf{C}\mathbf{u} + \mathbf{g})$$

where

$$\mathbf{u} = (\mathbf{C}'\mathbf{C})^{-1} \mathbf{c}$$

$$\mathbf{T} = \begin{pmatrix} \mathbf{I}_{n-m} \\ -\mathbf{C}'_m^{-1} \mathbf{C}'_{n-m} \end{pmatrix}$$

after dropping subscript $\tilde{\mathbf{x}}$ for simplicity. \mathbf{C}_m is the part of \mathbf{C} pertaining to m “dependent” solutions (as many as constraints) and \mathbf{C}_{n-m} the part pertaining to the $n - m$ “independent” solutions. The updated value for vector $\boldsymbol{\lambda}$ is finally set to

$$(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'(\mathbf{g} + \mathbf{H}\Delta\mathbf{x}).$$

This defines outer iterations, run until constraints are met and each term of \mathbf{x} may be either a positive value or 0. However, inner iterations, through CG, are needed when direct inversion is not possible, *i.e.*, when calculating $\Delta\mathbf{x}$. The final result corresponds to a continuum of mating frequencies and of reproduction levels for females.

3.4. Assigning cows to discrete levels of reproduction

We have to find the optimum group sizes $N_{d1} \dots N_{dl}$ of dams assigned to breeding alternatives $1, \dots, l$. These integers should verify

$$\sum_{k=1}^{k=l} N_{dk}\epsilon_k = E \quad \text{and} \quad \sum_{k=1}^{k=l} N_{dk}\rho_k = 1.$$

For the following, the levels of reproduction are ranked downwards: level 1 corresponds to the highest level of reproduction.

The full set of combinations meeting these conditions can be calculated in a simple way because we have

$$N_{dl} = \frac{1 - \sum_{k=1}^{k=l-1} N_{dk}\rho_k}{\rho_l} \quad \text{and} \quad \sum_{k=1}^{k=l-1} N_{dk} \left(\epsilon_k - \frac{\epsilon_l}{\rho_l} \rho_k \right) = E - \frac{\epsilon_l}{\rho_l}.$$

Then, N_{dl} is allowed to vary by integer values from 0 to the maximum integer possible based on reproduction. The same is done for levels $2, \dots, l - 2$, given the values obtained for previous levels. Values of $N_{d,l-1}$ are obtained from the second equation above. If this value is positive, then the current combination is accepted after calculating N_{dl} from the first equation. Otherwise, combination is rejected. If cows were ranked by decreasing reproduction rate and if appropriate numbers N_d were chosen, then the average reproduction rates of subpopulations would be close to the corresponding ρ 's. Hence, the idea of choosing the combination able to minimize a norm $\|\mathbf{q} - \tilde{\mathbf{q}}\|$ (for instance, $\sum_{k=1}^{k=l} (q_k - \tilde{q}_k)^2$), where \mathbf{q} is the vector of theoretical overall reproduction rates per group of dams, for a given combination of N_{dk} 's ($q_k = N_{dk}\rho_k$) and $\tilde{\mathbf{q}}$ is the

observed vector with

$$\tilde{q}_k = \frac{\sum_{j=N_{d1}+\dots+N_{dk-1}+1}^{j=N_{d1}+\dots+N_{dk}} r_j}{\sum_{j=1}^{j=N_{d1}+\dots+N_{dl}} r_j}.$$

Finally, the system to be solved becomes analogous to the system described in Section 3.2, after adding $N_d^* = \sum_{k=1}^{k=l} N_{dk}$ constraints for reproduction and deleting the redundant constraint $\mathbf{1}'_n \mathbf{x} = 1$.

3.5. Final mating selection

Cow j belonging to the final set of N_d^* cows can be mated to a maximum of n_j different sires. For instance, this value is equal to 1 for cows with a single AI or a single embryo collection and to 2, for cows with two collections or with one collection + AI.

The method used is an iterative selection of matings, through dropping or fixation, followed by re-optimization, so as to keep as much efficiency as possible.

The basic step consists of considering the list of matings still variable (subject to optimization) and the list of the n_j most frequent matings of the cows still involved in the current optimization (“protected” matings). Then, the globally least frequent matings, are set to 0 and added to the list of fixed matings. For not losing efficiency too fast, they represent only a small part (typically 5%, based on trial and error) of the “free” matings, *i.e.*, the unprotected variable matings. After this elimination, the number of different matings remaining for each cow still under optimization is updated and compared to the corresponding n_j . If both numbers differ, then the cow is maintained in the list of cows constrained for further optimization. Otherwise, the cow is removed from this list and her matings are fixed according to the following method. If n_j is equal to 1, then the frequency of the remaining mating is obviously set to the reproduction rate of cow j . For n_j larger than 1, an optimization procedure is needed. Case $n_j = 2$ provides a simple illustration. For cow j , ρ_{j1} and ρ_{j2} are the reproduction potentials allowed by breeding action 1 and 2 respectively ($\rho_j = \rho_{j1} + \rho_{j2}$). Matings aj and mating bj remain for final consideration. Then, the possible matings assigned to the ordered pair (action 1, action 2) are: aj and aj , bj and bj , aj and bj , bj and aj . Each combination is tested through the last Lagrange function optimized, where current values of \mathbf{x} are unchanged except for cow j . The selected combination minimizes the value of

this function. The end of the step consists in updating the constraint for breeding values, that becomes $B - \mathbf{b}'_F \mathbf{x}_F$, where subscript F refers to matings fixed to 0 or to some positive value.

4. OPTIMAL USE OF SERVICE BULLS

4.1. Outline of the strategy

The reproduction regime within the general population is AI. Then, the optimal use of service bulls corresponds to assigning a single (optimal) sire to each cow. Solving this problem by using the approach previously adopted for reproduction of bull dams would require the manipulation of linear systems of huge size.

A feasible approximate approach consists in considering the general population of existing females as constituted by sire \times maternal grand-sire groups, where relationships between and within groups are calculated only based on the exact relationships between the males involved (either sires or MGS or both). As a result, the problem amounts to find out, for each group, the optimal proportions of females to be served by the different AI sires. Then, the overall optimal use of a given AI sire is obtained after considering group frequencies in the population and specific within-group optimal use of this sire. These indications allow extension services of AI organizations to orientate the effective use by breeders, according to the groups their cows belong to.

Technically speaking, the analytical developments are much more simple than for young bull procreation. First, discretization of matings is no longer needed and second, the uniform use of AI removes the need of calculating optimal reproduction rates. Then, the analytical approach is basically very similar to the one described in Section 3.3.

4.2. Finding the optimal continuum

Breeders aim to produce M young females for further reproduction. These future young females are to be compared to N_g previous groups of females already existing in the population. The number of females belonging to group j is M_j . Breeding females, served by AI sires, belong to only N_d groups, because some j 's are not represented in this population. The number of breeding females belonging to group j is μ_j . If N_s is the number of service bulls, then $n = N_s \times N_d$ matings are possible and have to be examined. As for young bull

procreation, \mathbf{x} is the vector ($n \times 1$) of the internal mating frequencies ($\mathbf{1}'_n \mathbf{x} = 1$). The frequency of the breeding group j is

$$r_j = \frac{\mu_j}{\sum_{j=1}^{N_d} \mu_j}.$$

The sums of mating frequencies per group correspond to vector \mathbf{Kx} where matrix \mathbf{K} is of dimension $N_d \times n$. Then, in system 3.2, constraint $\mathbf{1}'\mathbf{x}$ is deleted and replaced by $\theta'(\mathbf{Kx} - \mathbf{r})$, where θ 's are Lagrange multipliers. The constraint on EBV is chosen so that the weighted average EBV of AI sires be equal to a desired value B (constraint on the sire-daughter gene transmission path). Then, vector \mathbf{b} describing the EBV of matings only refers to the sires involved.

In comparison with the problem addressed in Section 3.3, the g groups are analogous to the N_0 previous young bulls. However, the diversity of their future contribution to the population should be accounted for. Based on the gene-flow theory [17], w_j , the weight to be given to group j is, for simplification, the proportion of the original group j still surviving the next breeding year (the year after the current breeding year being examined). Exact weight would be rather difficult to calculate on real selected populations, because future culling and replacement decisions are still unknown. Finally, using a reasoning similar to the one of Section 3.3, we still find that the function to minimize is $\frac{1}{2}\mathbf{x}'\mathbf{Ax} + \mathbf{p}'\mathbf{x}$ but with $\mathbf{p}' = \frac{1}{M}(M_1w_1, \dots, M_gw_g)\mathbf{A}_1$ where matrices \mathbf{A} and \mathbf{A}_1 have the same meaning as in Section 3.2. Finally, the system to be solved is:

$$\begin{pmatrix} \mathbf{A} & -\mathbf{b} & -\mathbf{K}' \\ \mathbf{b}' & \mathbf{0} & \mathbf{0} \\ \mathbf{K} & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \lambda \\ \boldsymbol{\theta} \end{pmatrix} = \begin{pmatrix} -\mathbf{p} \\ B \\ \mathbf{r} \end{pmatrix}.$$

5. SELECTION OF YOUNG BULLS FOR SERVICE

In some countries, official approval of bulls, mainly based on EBV, is needed. Afterwards, through their individual decisions on the farm, breeders decide about the real use of these approved bulls. The current section deals with the problem of additionally considering genetic variability during this selection step. The corresponding approach is clearly related to that of the previous section. Indeed, the selection of young bulls for service should be carried out based on their optimal use, when competing with the current AI bulls, for service on current sire \times MGS groups of females. However, a decision should be made about what should be the minimal optimal frequency of use, after considering the cost of maintaining bulls for a very limited use. A simple approach is to dismiss the subgroup of young bulls with a null optimal contribution.

6. THE RESULTS OBTAINED ON A FULL SIZE DATA SET

The efficiency of the previous approach was not tested on simulated populations due to the amount of calculations required (especially for young bull procreation) and to the urgent need of making practitioners aware both of the situation and of the possible solutions. The real data were provided by the breeding scheme of the Norman cattle breed, centralized to some extent. The general idea was to examine recent decisions effectively carried out and to compare them to optimized decisions, using the approaches described in Sections 3, 4 and 5. The B values for constraints on EBV were simply the observed ones, *i.e.*, we were investigating the issue of preserving most genetic variability without any short term sacrifice.

Finally, the objective was at the same time: (i) to examine the numerical feasibility of the approach on a large population; and (ii) to measure the extent of sub-optimality of current practices.

6.1. Procreation of young bulls

In the Norman breeding scheme, about 400 bulls entered a performance-test station for selection on growth rate, muscularity, general fitness and ability to produce semen of good quality. Eventually, only about 150 bulls were progeny-tested. The overall selection index is ISU, combining milk yield, milk composition, functional and type traits [9]. We considered the group entering the station between March 1, 2001 and February 28, 2002 (401 young bulls) and the 4 previous annual groups of animals still being progeny-tested (626 bulls). Station bulls were born from 338 dams and 21 sires. Their average EBV for ISU was equal to 136.8 (2001/2 evaluation). Thirty-four per cent were born from embryo transfer and the overall breeding cost was equal to 560 k€ after including the cost of contracts with the breeders.

For the optimization, the number of candidates was increased very much: based on ISU only, 2112 cows were candidates for dams and 22 AI bulls were candidates for sire. The allowed reproduction levels corresponded to a single AI or an embryo collection followed by AI or two collections followed by AI: corresponding expected numbers of male calves were 0.5, 1.75 and 3, respectively, and corresponding costs were 0.50, 2.75 and 5 k€, respectively. Given this economical constraint, the search algorithm assigned 273, 94 and 33 cows respectively to the different reproduction regimes. Finally, the algorithm selected the different bulls to be used on these cows at each reproduction step. Only 11 bulls out of the 22 candidates were effectively used,

Table I. Young bull procreation: average kinship coefficients.

Kinship %	Without optimization	With optimization	% decrease
P*P	6.07%	idem	0
P*B	5.87%	4.13%	28
B*B	7.03%	5.10%	27
(P B)*(P B)	6.07%	5.00%	18

B = new bulls; P = bulls under progeny-testing.

including 3 of them with little use (about 2%). Seventy-six per cent of the cows with 1.75 calves and 100% of the cows with 3 calves were mated to a single bull. The last result was not found again during subsequent optimizations (for preparing the real future matings of late 2003). All these steps required extensive calculations because the function $\mathbf{x}'\mathbf{A}\mathbf{x}$ was evaluated about 40 000 times.

The results of Table I clearly show that a substantial decrease (18%) of the average kinship coefficient would have been possible, without damaging the average EBV for ISU nor increasing breeding costs. This decrease was even more clear when considering the average kinship coefficient between bulls and progeny-test bulls (28%).

6.2. Optimal use of AI bulls

Inseminations between October 1, 2001 and September 30, 2002 were examined. Selected cows (201 692) were born from known sires and MGS (40 676 sire-MGS group) and were inseminated by 42 approved bulls. For the optimization, 2075 groups of at least 10 cows were selected to avoid returning to the individual mating situation. Forty different bulls were used on these groups with an average weighted ISU equal to 128.2 (2002/2 evaluation). For completing the current female population, 194 300 young females not yet inseminated and split into 9261 groups of at least 3 individuals were considered.

Optimization results are shown in Table II. As for young bull procreation, substantial decrease of kinship coefficients was obtained, except for the overall population due to the very high proportion of existing females.

For simplicity, details about the use of the 40 bulls are not shown. The correlation coefficient between use frequencies in alternative situations, real or optimized, was very small ($r = 0.11$). Furthermore, optimized use was slightly more correlated with ISU ($r = 0.37$) than real use ($r = 0.28$). The discrepancy between both uses can be illustrated by dramatic examples: the most used bull (15% of AI) was totally eliminated by the optimization whereas another one,

Table II. Use of service bulls: average kinship coefficients.

Kinship %	Without optimization	With optimization	% decrease
P*P	6.20%	idem	0
P*H	3.82%	3.30%	14
H*H	5.20%	4.08%	21
(P H)*(P H)	5.60%	5.46%	3

H = new females; P = existing females.

almost unused (0.6%) was recommended to 6%. However, in some instances, both uses were very similar.

6.3. Selection of young bulls for use

We considered the batch of 152 progeny-tested bulls born in 1996. Approval for use was essentially decided in July 2002. Nineteen bulls were really selected and their ISU ranged from 121 to 152 (1002/2 evaluation). If these bulls had been set in competition with service bulls of the previous year, after updating the constrained average ISU by the amount of desired genetic gain (132.2), then their overall use would have been 29%. However, 9 bulls out of these 19 (range of ISU: 122–141) would have been totally eliminated. When considering the 19 best bulls for ISU instead, virtually the same bulls would have been selected.

An optimization considering the 52 best bulls for ISU led to selection of 12 young bulls with an overall use of 33%, due to the inclusion of bulls previously dismissed. For instance two bulls with ISU as low as 119 and 120 were recommended for 3% and 1% use, respectively.

7. DISCUSSION AND CONCLUSION

7.1. Methodological issues

Based on the considerations of Caballero and Toro [7] on diversity, an appropriate method for keeping most genetic diversity seems to be the dynamical minimization of the average pairwise kinship (or relationship) coefficient, including self-comparisons. This was definitely the approach adopted in this paper. Furthermore, the optimization was performed in a single step, *i.e.*, selection of matings was directly targeted and selection and contribution of parents

were post-determined. Both points made the current approach to differ substantially from the usual optimizations mentioned earlier.

The essential difference was that the priority was given to maintaining diversity without paying direct attention to inbreeding coefficients. Further research work comparing the diversity-oriented and the inbreeding-oriented approaches based on simple theoretical populations would be useful for determining whether the diversity-oriented approach is really rewarding, especially within a reasonable time horizon. Otherwise, the price to pay in terms of inbreeding might be considered as deterring. A minor difference was that mate selection was implemented. If the evaluation criterion is unique, such as here, then the one-step approach is more efficient than the two-step approach, where the second step is constrained for the optimal parental contributions found after the first step (the difference was found minor in our simplified test populations). However, two-step approaches are generally multi-purpose and consider at the same time inbreeding rates and inbreeding coefficients. Then, for a comparison between methods to make sense, one should balance several coefficients (diversity and inbreeding coefficients) and corresponding rates.

The current optimization might have suffered from efficiency losses. First, introducing separate steps was clearly suboptimal but this was carried out for the sake of feasibility in order to avoid manipulating a huge amount of matings for examination, while incorporating a large list of very various constraints. However, an improvement to be tested further, would be to consider as a “previous” population, the existing bulls and the existing females altogether. However, a specific problem would arise from the very large number of females and their collapse into genetic groups. The effect of mixing true relationships and pseudo-relationships is still unknown.

7.2. Practical implications

Sub-optimality of current practices was clearly demonstrated because substantial savings of genetic variability would have been possible, especially for young bull procreation. This finding for diversity reminds the statement of Bijma *et al.* [2] for inbreeding: “by using these (optimized) procedures, breeding organizations can make the same ΔG as they do at present whilst reducing the rate of inbreeding generated”. Additional work, not research indeed, is needed to understand how exactly diversity is lost. This might be due to disregarding the remote relationships when mating, the excessive use of some breeding animals (especially bulls) or threshold selection traits (an approach still popular among practitioners). The use of this last approach instead of

using an overall EBV, as in the current paper, would certainly have prevented the optimization algorithm from considering some matings potentially interesting for genetic variability and would have led to a smaller efficiency gap between practice and “optimized” matings. The ultimate practical objective is not only making breeders to stick strongly to mating plans designed for the long term but also to give up habits already well-known as harmful for genetic gains and additionally detrimental for a good management of genetic variability.

Because of the future challenging situation for dairy cattle selection, practitioners should clearly modify their practices. First, they should put trust in mathematical algorithms able to detect and to exploit real relationships between animals, including remote relationships over generations, very difficult to quickly assess by traditional methods. Second, when selecting on many traits, they should put more and more emphasis on the overall EBV and give up accordingly the independent culling approach, due to its inefficiency for creating overall genetic gains while canalizing selection too much towards “ideal” animals.

The current situation in many populations is so damaged and evolving so fast that simple procedures would not be enough for containing increase of kinship and inbreeding coefficients. Simple procedures are for instance stopping the excessive use of some sires for young bull procreation or selecting breeding animals on EBV penalized for kinship. The current literature clearly shows that the careful preparation of optimized matings is the least detrimental approach for saving the future.

Anyhow, in the long term, kinship and inbreeding are certain to accumulate. This fact will influence selection procedures increasingly. For instance, even with the approach developed here, breeders will be led to progressively decrease the level of desired genetic gains in order to avoid reaching dangerous rates of increase for these key parameters (high inbreeding depressions and a high risk of expression for genetic defects).

ACKNOWLEDGEMENTS

The reviewers chosen by the journal are acknowledged for providing useful comments.

REFERENCES

- [1] Allaire F.R., Mate selection by selection index theory, *Theor. Appl. Genet.* 57 (1980) 267–272.

- [2] Bijma P., Meuwissen T.H.E., Woolliams J.A., Design of sustainable breeding programs in developed countries, in: 7th World Congress on Genetics Applied to Livestock Production, 19-23 August 2002, Montpellier, CD ROM communication 19-23, Inra, Castanet-Tolosan.
- [3] Boichard D., Maignel L., Verrier E., The value of using probabilities of gene origin to measure genetic variability in a population, *Genet. Sel. Evol.* 29 (1997) 5–23.
- [4] Brisbane J.R., Gibson J.P., Balancing selection response and rate of inbreeding by including genetic relationships in selection decisions, *Theor. Appl. Genet.* 91 (1995) 421–431.
- [5] Brisbane J.R., Gibson J.P., Balancing selection response and rate inbreeding by including predicted stabilised genetic contributions in selection decisions, *Genet. Sel. Evol.* 27 (1995) 541–549.
- [6] Caballero A., Santiago E., Toro M.A., Systems of mating to reduce inbreeding in selected populations, *Anim. Sci.* (1996) 431–442.
- [7] Caballero A., Toro M.A., Interrelations between effective population size and other pedigree tools for the management of conserved populations, *Genet. Res.* (2000) 331–343.
- [8] Colleau J.J., An indirect approach to the extensive calculation of relationship coefficients, *Genet. Sel. Evol.* 34 (2002) 409–421.
- [9] Colleau J.J., Regaldo D., Définition de l'objectif de sélection dans les races bovines laitières, *Renc. Rech. Ruminants* 8 (2001) 329–332.
- [10] Crow J.F., Kimura M., An introduction to population genetics theory, Burgess publishing company Minneapolis USA, 1970.
- [11] Dempfle L., A note on increasing the limit of selection through selection within families, *Genet. Res.* 24 (1975) 127–135.
- [12] Fernandez J., Toro M.A., The use of mathematical programming to control inbreeding in selection schemes, *J. Anim. Breed. Genet.* 166 (1999) 447–466.
- [13] Fernandez J., Caballero A., A comparison of management strategies for conservation with regard to population fitness, *Conserv. Genet.* 2 (2001) 121–131.
- [14] Grundy B., Caballero A., Santiago E., Hill W.G., A note on using biased parameter values and non-random mating to reduce rates of inbreeding in selection programmes, *Anim. Prod.* 59 (1994) 465–468.
- [15] Grundy B., Villanueva B., Woolliams J.A., Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development, *Genet. Res.* 72 (1998) 159–168.
- [16] Grundy B., Villanueva B., Woolliams J.A., Dynamic selection for maximizing response with constrained inbreeding in schemes with overlapping generations, *Anim. Sci.*, 70 (2000) 373–382.
- [17] Hill W.G., Prediction and evaluation of response to selection with overlapping generations, *Anim. Prod.* 18 (1974) 117–139.
- [18] Kinghorn B., On computing strategies for mate allocation, *J. Anim. Breed. Genet.* 104 (1987) 12–22.
- [19] Klieve H.M., Kinghorn B.P., Barwick S.A., The joint regulation of genetic gain and inbreeding under mate selection, *J. Anim. Breed. Genet.* 111 (1994) 81–88.

- [20] Luo Z.W., Woolliams J.A., Thompson R., Controlling inbreeding in dairy MOET nucleus schemes, *Anim. Sci.* 60 (1995) 379–387.
- [21] Meuwissen T.H.E., Maximizing the response of selection with a predefined rate of inbreeding, *J. Anim. Sci.* (1997) 934–940.
- [22] Meuwissen T.H.E., Sonesson A.K., Maximizing the response of selection with a predefined rate of inbreeding: overlapping generations, *J. Anim. Sci.* 76 (1998) 2575–2583.
- [23] Moureaux S., Boichard D., Verrier E., Utilisation de l'information généalogique pour l'estimation de la variabilité génétique de huit races bovines laitières françaises d'expression nationale ou régionale, *Renc. Rech. Ruminants* 7 (2000) 149–152.
- [24] Scales L.E., *Introduction to Non-Linear optimization*, Mac Millan Publishers, London, 1985.
- [25] Sonesson A.K., Meuwissen T.H.E., Mating schemes for optimum contribution selection with constrained rates of inbreeding, *Genet. Sel. Evol.* 32 (2000) 231–248.
- [26] Sonesson A.K., Grundy B., Woolliams J.A., Meuwissen T.H.E., Selection with control of inbreeding in populations with overlapping generations: a comparison of methods, *Anim. Sci.* 70 (2000) 1–8.
- [27] Sonesson A.K., Meuwissen T.H.E., Minimisation of rate of inbreeding for small populations with overlapping generations, *Genet. Res.* 77 (2001) 285–292.
- [28] Sonesson A.K., Meuwissen T.H.E., Non-random mating for selection with restricted rates of inbreeding and overlapping generations, *Genet. Sel. Evol.* 34 (2002) 23–39.
- [29] Sorensen A.C., Berg P., Woolliams J.A., Predicting rates of gains and inbreeding with factorial mating and the use of mating for optimal designs, in: 7th World Congress on Genetics Applied to Livestock Production, 19-23 August 2002, Montpellier, CD ROM communication 19-08, Inra, Castanet-Tolosan.
- [30] Stoer J., Bulirsch R., *Introduction to numerical analysis*, Springer-Verlag, New York, 1980.
- [31] Toro M.A., Nieto B.M., A simple method for increasing the response to artificial selection, *Genet. Res.* 44 (1984) 347–349.
- [32] Toro M.A., Nieto B., Salgado C., A note on minimization of inbreeding in small scale selection programmes, *Livest. Prod. Sci.* 20 (1988) 317–323.
- [33] Toro M.A., Perez-Enciso M., Optimisation of selection response under restricted inbreeding, *Genet. Sel. Evol.* 22 (1990) 347–349.
- [34] Verrier E., Colleau J.J., Foulley J.L., Long-term effects of selection based on the animal model BLUP in a finite population, *Theor. Appl. Genet.* 87 (1993) 446–454.
- [35] Villanueva B., Woolliams J.A., Optimization of breeding programmes under index selection and constrained inbreeding, *Genet. Res.* 69 (1997) 145–158.
- [36] Villanueva B., Bijma P., Woolliams J.A., Optimal mass selection policies for schemes with overlapping generations and restricted inbreeding, *Genet. Sel. Evol.* 32 (2000) 339–345.

- [37] Villanueva B., Woolliams J.A., Simm G., Strategies for controlling rates of inbreeding in MOET nucleus schemes for beef cattle, *Genet. Sel. Evol.* 26 (1994) 517–535.
- [38] Woolliams J.A., Bijma P., Predicting rates of inbreeding in populations undergoing selection, *Genetics* 154 (2000) 1851–1864.
- [39] Woolliams J.A., Pong-Wong R., Villanueva B., Strategic optimisation of short-and-long-term gain and inbreeding in MAS and non-MAS schemes, in: 7th World Congress on Genetics Applied to Livestock Production, 19-23 August 2002, Montpellier, CD ROM communication 19-23, Inra, Castanet-Tolosan.
- [40] Wray N.R., Goddard M.E., Increasing long-term response to selection, *Genet. Sel. Evol.* 26 (1994) 431–451.
- [41] Wu L., Schaeffer L.R., Reducing the effect of parent averages from animal solutions in mixed model equations, *J. Anim. Breed. Genet.* 117 (2000) 361–374.

APPENDIX A: THE SYMMETRICAL OPTIMIZATION

If we wish to maximize (second optimization) $\mathbf{b}'\mathbf{x}$ with the constraints $\frac{1}{2}\mathbf{x}'\mathbf{A}\mathbf{x} = C$, $\mathbf{1}'\mathbf{x} = 1$, then we have the Lagrange function

$$\mathbf{b}'\mathbf{x} - \lambda_1^* \left(\frac{1}{2}\mathbf{x}'\mathbf{A}\mathbf{x} - C \right) - \lambda_2^* (\mathbf{1}'\mathbf{x} - 1).$$

Solutions are given by

$$\mathbf{b} - \lambda_1^* \mathbf{A}\mathbf{x} - \lambda_2^* \mathbf{1} = \mathbf{0}$$

with the constraints being satisfied. Let us imagine that $\tilde{\mathbf{x}}$ is the solution of the first optimization with constraints $\mathbf{b}'\mathbf{x} = B$ (multiplier λ_1), $\mathbf{1}'\mathbf{x} = 1$ (multiplier λ_2). Then $\tilde{\mathbf{x}}$ is also the solution of the second optimization problem where $C = \frac{1}{2}\tilde{\mathbf{x}}'\mathbf{A}\tilde{\mathbf{x}}$. The first optimization yields

$$\mathbf{A}\tilde{\mathbf{x}} - \lambda_1 \mathbf{b} - \lambda_2 \mathbf{1} = \mathbf{0}$$

i.e.,

$$\mathbf{b} - \frac{1}{\lambda_1} \mathbf{A}\tilde{\mathbf{x}} + \frac{\lambda_2}{\lambda_1} \mathbf{1} = \mathbf{0}.$$

If we set $\lambda_1^* = \frac{1}{\lambda_1}$ and $\lambda_2^* = -\frac{\lambda_2}{\lambda_1}$, then $\tilde{\mathbf{x}}$ clearly fulfills the requirements of the second optimization.

Interestingly, it can be observed that in the second optimization problem, calculation of solutions for λ_1^* and λ_2^* involves the calculation of the quadratics $\mathbf{b}'\mathbf{A}^{-1}\mathbf{b}$, $\mathbf{1}'\mathbf{A}^{-1}\mathbf{b}$, $\mathbf{1}'\mathbf{A}^{-1}\mathbf{1}$. Hence, the main difficulty is to calculate vectors $\mathbf{A}^{-1}\mathbf{b}$ and $\mathbf{A}^{-1}\mathbf{1}$. If direct inversion of matrix \mathbf{A} is excluded, CG could be resorted to for each vector. Finally, a third CG iteration step would be needed in order to calculate the solution of \mathbf{x} , given the values of the λ^* . This is to be compared to a single CG step for the first optimization problem.

APPENDIX B: NEGATIVE OPTIMAL FREQUENCIES MAY BE PERMANENTLY SET TO ZERO

First minimization of $\frac{1}{2}\mathbf{x}'\mathbf{A}\mathbf{x}$ with constraints $\mathbf{b}'\mathbf{x} = B$ and $\mathbf{1}'_n\mathbf{x} = 1$ generally yields negative values for some x_i 's, *i.e.*, outside the domain of feasibility. The most likely feasible values are 0 and consequently, they are set to 0 during the second optimization. If this procedure is carried out repeatedly, then the final $\tilde{\mathbf{x}} = \begin{pmatrix} \tilde{\mathbf{x}}_0 \\ \tilde{\mathbf{x}}_1 \end{pmatrix}$ where the first sub-vector is constituted of n_0 zeros and the second subvector is constituted of $n_1 = n - n_0$ positive values. Let \mathbf{b}_1 be the subvector of \mathbf{b} corresponding to these positive values. Then, $\tilde{\mathbf{x}}$ corresponds to the optimum of the Lagrange function

$$L(\mathbf{x}) = \frac{1}{2}\mathbf{x}'\mathbf{A}\mathbf{x} - \lambda_1(\mathbf{b}'_1\mathbf{x}_1 - B) - \lambda_2(\mathbf{1}'_{n_1}\mathbf{x} - 1) - \theta'(\mathbf{x}_0 - \mathbf{0}_{n_0})$$

where λ' and θ' are Lagrange multipliers.

If we let

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{00} & \mathbf{A}_{01} \\ \mathbf{A}_{10} & \mathbf{A}_{11} \end{pmatrix}$$

then $\theta = \mathbf{A}_{01}\tilde{\mathbf{x}}_1$. Because all the terms of \mathbf{A}_{01} and $\tilde{\mathbf{x}}_1$ are positive, all of θ are positive.

Theory [24] has shown that, if an inequality constraint (feasibility) is in fact an equality constraint at the optimum, then the corresponding Lagrange multiplier is positive. Conversely, if this multiplier is positive, then the solution of the corresponding Lagrange function is really the optimal solution.