On the problem of compensatory mating in animal breeding*

Ana Paula Lüdtke Ferreira[1]

[1]Programa de Pós-graduação em Computação Aplicada
Universidade Federal do Pampa

Abstract. *Animal breeding relies on two processes to achieve its objectives: the selection and the mating systems. Mating systems devise a particular plan to perform one or more breeding goals, which often encompass improving the herd's health and maximising financial gains in animal production systems. Compensatory mating is a strategy to produce animals with more homogeneous selection trait characteristics, discarding the production of exceptional animals in favour of a more balanced herd. This paper defines and investigates the complexity class of the optimal compensatory mating problem, proving that a polynomial-time algorithm can solve it.*

Resumo. *O melhoramento animal depende de dois processos para atingir seus objetivos: os sistemas de seleção e de acasalamento. Os sistemas de acasalamento estabelecem um plano específico para atingir um ou mais objetivos de produção, que muitas vezes incluem a melhora da saúde do rebanho e a maximização dos ganhos financeiros nos sistemas de produção animal. O acasalamento compensatório é uma estratégia para produzir animais com características de seleção mais homogêneas, descartando a produção de animais excepcionais em favor de um rebanho mais equilibrado. Este artigo define e investiga a classe de complexidade do problema de acasalamento compensatório ótimo, provando que um algoritmo de tempo polinomial pode resolvê-lo.*

1. Introduction

Animal breeding involves directing the next generation's genetics towards financial profits for the production system, using essentially two strategies: the selection and the mating systems. The selection process works by choosing the best animals that will reproduce and spread their genes within the population, increasing the frequency of desirable traits and decreasing the undesirable ones. Genetic or phenotypic similarities are the keys to the development of mating systems. Parents can pass their genetic superiority on to their progeny, and the obtained genetic gains are cumulative and long-lasting for later generations [Falconer and Mackay 1996]. Thus, in this case, the goal is to optimize the mating options, considering the animals of better genetic values. For example, some dam (a female parent) may have several sire (male) possibilities to mate, producing similar genetic results. Furthermore, mating systems may have different goals.

The mating system configured so that the best males reproduce with the best females and, consequently, the worst males reproduce with the worst females is called mating between peers. That strategy aims to obtain extreme products, that is, animals that are exceptional in specific characteristics. On the other hand, a compensatory or corrective

*Resultado finalizado.

mating strategy seeks to bring more homogeneity within the population. A compensatory mating system works by mating individuals differing in performance. For instance, an excellent male, regarding a particular trait, mates with a female deficient in that same trait [Eler 2017, Cardoso 2009].

The compensatory mating problem has not been dealt with in the literature from an algorithmic perspective, as far as we can tell. Both selection and mating processes are performed by people, analysing the particular animals involved. However, animal breeding related problems do have described algorithmic approaches [Nayeri et al. 2021]. The focus on those approaches is twofold: maximising expected offspring trait values and minimising co-ancestry and inbreeding. The literature presents several attempts to solve the first problem. Research on that matter focuses mainly on metaheuristics based on genetic algorithms [Carvalheiro et al. 2010, Kinghorn 2011, Barreto Neto et al. 2014, Carvalho et al. 2016, da Fontoura et al. 2020]. However, a recent development has shown an optimal polynomial-time greedy strategy to solve the problem. The second problem is essential for animal breeding programs since breeding exceptional animals will dramatically increase the herd's inbreeding levels. Inbreeding creates issues such as inbreeding depression, which is the reduced survival and fertility of offspring of related individuals in several production traits [Mi et al. 1965, Charlesworth and Willis 2009].

In this paper, we use the optimal algorithm presented in [Ferreira et al. 2021] to tackle the compensatory mating problem. We define optimal compensatory mating as the one that is both maximal in the sense of offspring values and has a minimum variance. We show that the problem belongs to the complexity class **P** [Arora and Barak 2009] by reducing it to the minimum assignment in weighted bipartite graphs (Sec. 2). We also present a discussion on the work presented herein and point to further developments (Sec. 3).

2. Problem characterization and analysis

A *selection process* intends to elect the animals to become the next generation parents [Yokoo et al. 2019, Simões et al. 2020, Weigel et al. 2017, Dunne et al. 2021]. The selection process analyses the herd's genetic/economic values and phenotypes of genetic/economic interest. The value of each animal is computed from a selection index, chosen accordingly to the elicited breeding objectives. A *selection index* expresses the relative reward/punishment of each trait to consider in the breeding ahead. Formally, a *selection index* is a pair $\mathcal{I} = (T, w)$ where T is a set of animal traits and $w : T \to \mathbb{R}$ is the index weighting function.

Given a selection index $\mathcal{I}$, the *individual value* of each animal $a \in A$ is the sum of its individual characteristics' values weighted by the relative importance of each characteristic accordingly to $\mathcal{I}$. I.e., the summation given in (1), where $\nu : T \times A \to \mathbb{R}$ outputs $a \in A$'s value for each trait $t \in T$. When the selection index is understood from the context, we write $\iota(a)$ instead of $\iota^{\mathcal{I}}(a)$.

$$\iota^{\mathcal{I}}(a) = \sum_{k \in T} \nu(k, a) \cdot w(k) \tag{1}$$

The mating contribution is the average of both parent's individual values, except when they have a co-ancestry above an acceptable level. The breed of related animals shrinks genetic diversity and lead to undesirable health issues. Inbreeding strategies are helpful to fix a desirable phenotype within a population, but it is usually undesirable in the mating process. Animal breeding programs control the degree of kinship amongst animals, making them available to producers and researchers. As in the case of traits, databases or spreadsheets maintain animal kinship information. We assume that the producer can choose whether inbreeding is acceptable and, if so, to what degree. Definition 1 inserts that idea into the calculation of mating pairs expected offspring contribution value.

Definition 1 (Mating contribution) *Let $A = M \cup F$ be a set of animals, where M is the set of sires and F is the set of dams, with $M \cap F = \emptyset$, $\mathcal{I}$ be a selection index, $r : A \times A \to \mathbb{R}^+$ be the degree of which two animals are related, and l be the maximum inbreeding threshold. The* mating contribution $\pi : M \times F \to \mathbb{R}$ *of a sire $m \in M$ and a dam $f \in F$ is computed as follows:*

$$\pi(m, f) = \begin{cases} (\iota^{\mathcal{I}}(m) + \iota^{\mathcal{I}}(f))/2 & r(m, f) \leq l \\ -\infty & r(m, f) > l \end{cases} \tag{2}$$

where $\iota^{\mathcal{I}}$ is the individual contribution of each animal, given selection index $\mathcal{I}$.

The mating contribution is computed accordingly to the expected offspring trait values of each dam/sire pair, which is the average of both parents' contributions.

Problem 1 (Optimal mating assignment)

Input*: A set of sires M, a set of dams F, a selection index $\mathcal{I}$, a contribution function $\pi : M \times F \to \mathbb{R}$ computed as in Definition 1, a function* $\max : M \to \mathbb{N}$ *expressing the use limit of each sire in the mating process.*

Output*: A function $b^* : F \to M$ obeying the use limit for each sire, i.e. for each $m \in M$ we have that $|\{f \mid b^*(f) = m\}| \leq \max(m)$, and such that the sum $\sum_{f \in F} \pi(b^*(f), f)$ is maximal, i.e., for each other function $b : F \to M$ obeying the sire's use limit we have*

$$\sum_{f \in F} \pi(b(f), f) \leq \sum_{f \in F} \pi(b^*(f), f) \tag{3}$$

In [Ferreira et al. 2021] we present several results concerning strategies for finding optimal mating assignments, including an optimal greedy strategy to solve Problem 1. The proposed algorithm puts the problem within the complexity class **P**, since it runs in worst-case $O(|M|.|F|)$ time. We have also shown that given two mating pairs $(m_1, f_1), (m_2, f_2) \in M \times F$, we have that $\pi(m_1, f_1) + \pi(m_2, f_2) = \pi(m_2, f_1) + \pi(m_1, f_2)$, as long as the maximal accepted inbreeding is not exceeded between pairs (m_2, f_1) and (m_1, f_2). That result can be the basis for an algorithm to build a solution for the compensatory mating problem while retaining the solution's optimality since both total and average mating values are unchanged by the swapping of mated pairs. Theorem 1 proves that claim.

Theorem 1 *Let F be a set of dams, M be a set of sires, and $b : F \to M$ be a mating function. Let s_b be the total value of function b, or*

$$s_b = \sum_{f \in F} \pi(b(f), f) \tag{4}$$

Let $f_1, f_2 \in F$ such that neither $\pi(b(f_2), f_1)$ nor $\pi(b(f_1), f_2)$ exceed the inbreeding maximum limit (i.e., neither value equals $-\infty$), and let

$$s'_b = \pi(b(f_2), f_1) + \pi(b(f_1), f_2) + \sum_{f \in F \setminus \{f_1, f_2\}} \pi(b(f), f)$$

then $s_b = s'_b$.

Proof: The proof is straightforward given how each animal pair contribution is computed:

$$\begin{aligned}
s'_b = &\ \pi(b(f_2), f_1) + \pi(b(f_1), f_2) + \textstyle\sum_{f \in F \setminus \{f_1, f_2\}} \pi(b(f), f) \\
= &\ \iota(b(f_2))/2 + \iota(f_1)/2 + \iota(b(f_1))/2 + \iota(f_2)/2 + \textstyle\sum_{f \in F \setminus \{f_1, f_2\}} \pi(b(f), f) \\
= &\ \pi(b(f_1), f_1) + \pi(b(f_2), f_2) + \textstyle\sum_{f \in F \setminus \{f_1, f_2\}} \pi(b(f), f) \\
= &\ \textstyle\sum_{f \in F} \pi(b(f), f) \\
= &\ s_b
\end{aligned}$$

. $\square$

Corollary 1 *Changing mated pairs within an optimal solution for Problem 1 dos not alter the solution average μ.*

Proof: It follows directed from Theorem 1: if $s_b = s'_b$ then $\mu_b = \mu'_b$, given that $\mu_b = s_b/|F|$ and $\mu'_b = s'_b/|F|$. $\square$

Theorem 1 shows that changing mating pairs, as long as each the match is under the inbreeding limit, does not affect neither the total solution value nor its mean.

Problem 2 formalizes the problem we intend to tackle.

Problem 2 (Optimal Compensatory Mating problem)

Input: *A tuple $S = (M, F, \pi, b^*)$ where M is a set of sires, F is a set of dams, π is the contribution function described in (2), and $b^* : F \to M$ is an optimal mating function given as the solution of Problem 1.*

Output: *A function $b^c : F \to M$ with the same total value as b^*, i.e., $\sum_{k \in F} \pi(b^c(k), k) = \sum_{k \in F} \pi(b^*(k), k)$, and such that the solution variance is minimum. I.e, for any other function $b' : F \to M$ with $\sum_{k \in F} \pi(b'(k), k) = \sum_{k \in F} \pi(b^*(k), k)$ we have that*

$$\frac{1}{|F|} \sum_{f \in F} (\pi(b'(f), f) - \mu_{b'})^2 \leq \frac{1}{|F|} \sum_{f \in F} (\pi(b^c(f), f) - \mu_{b^c})^2 \tag{5}$$

The greedy strategy used in [Ferreira et al. 2021] does not work for Problem 2 because searching for minimum variance will not preserve the maximal function value. However, we can take advantage of the fact that changing already mated pairs belonging to the optimal solution will not change its value. We will use a reduction strategy [Garey and Johnson 1979] to show that Problem 2 have a polynomial-time solution. The reduction will also give an algorithm to solve the problem, albeit not necessarily the most efficient one. Specifically, we will prove the existence of a polynomial-time reduction from the optimal compensatory mating problem to the assignment problem of weighted bipartite graphs.

A bipartite graph is a graph $G = (V, E)$ where the vertex set V has a partition $\{L, R\}$, such that $L \cap R = \emptyset$ and $L \cup R = V$, induced by the edge set E, as follows: for each $(v_1, v_2) \in E$ we have that $v_1 \in L$ and $v_2 \in R$. For that reason, bipartite graphs are commonly referred as a tuple $G = (L, R, E \subseteq L \times R)$. The definition of weighted graphs extends naturally to bipartite graphs, where a weighting function $w : E \to \mathbb{R}$ attributes a value to each edge $e \in E$. The *minimum assignment problem* for weighted bipartite graph can be defined as follows:

Problem 3 (Minimum assignment in weighted bipartite graphs)

Input: *A weighted bipartite graph* $G = (L, R, E, w : E \to \mathbb{R})$

Output: *An* assignment *from L to R which is minimum, i.e., a function* $a^* : L \to R$ *such that for each* $v \in L$ *we have that* $(v, a^*(v)) \in E$*, and for any other function* $a : L \to R$ *with* $(v, a(v)) \in E$*, the following holds:*

$$\sum_{v \in L} w(v, a^*(v)) \leq \sum_{v \in L} w(v, a(v)) \tag{6}$$

The assignment problem in weighted bipartite graphs can be trivially modelled as an integer linear programming problem, which is **NP**-complete in the general case [Garey and Johnson 1979]. However, the literature presents several polynomial-time algorithms to solve the problem faster. When the graph is *balanced*, i.e., when $|L| = |R|$, the Kuhn-Munkres algorithm (also known as the Hungarian method) [Kuhn 1955, Munkres 1957] can solve the problem in $O(mn + n^2 \log n)$ [Ramshaw and Tarjan 2012] where $n = |L| = |R|$ and $m = |E|$, proving it belongs to the complexity class **P**.

A *reduction* from a problem P_1 to a problem P_2, with (respectively) inputs I_1 and I_2, and outputs O_1 and O_2 is a pair of functions $f_I : I_1 \to I_2$ and $f_O : O_1 \to O_2$ such that for any problem instance $x \in I_1$ we have that $P_1(x) = f_O(P_2(f_I(x)))$. The function notation to represent problems P_1 and P_2 derives from the problem definition as a mapping between input and outputs instance values (i.e., a function). A reduction gives us an alternative algorithm for solving the problem P_1 if an algorithm for solving P_2 is known. Furthermore, the complexity of P_1 is bounded by the sum of f_I, P_2, and f_O complexities [Arora and Barak 2009].

Problem 1 does not make any assumptions regarding the relative number of sires or dams nor it requires an equal number of animals in each set. Actually, the number of dams usually exceeds the number of sires by a large amount. However, the problem solution delivered by the greedy algorithm described in [Ferreira et al. 2021] is a surjective function from F to the actual image of the mating function in M. We will use the solution of Problem 1, given as input to Problem 2 to build two sets of equal size. The reduction from the Compensatory Mating problem to the assignment problem for weighted bipartite graph appears in Definition 2.

Definition 2 (Problem Reduction) *Let* F *be a set of dams,* M *be a set of sires,* π *be the pair contribution function and* $b : F \to M$ *be an optimal solution to the Mating Selection*

Maximization problem. Let $n_b : M \to \mathbb{N}$ be the function $n_b(m) = |\{f \in F \mid b(f) = m\}|$, i.e., the function that for each sire returns the number of dams it was mated with. Let $G = (F, M', E, w : E \to \mathbb{R})$ be a weighted bipartite graph built as follows:

- $M' = \cup_{m \in M}\{[m, i] \mid 1 \leq i \leq n_b(m)\}$
- $E = \{(f, [m, i]) \mid 1 \leq i \leq n_b(m), \pi(m, f) \neq -\infty\}$
- $w(f, [m, i]) = (\pi(m, f) - \mu_b)^2$, *for each* $w([m, i], f) \in E$, *with* $1 \leq i \leq n_b(m)$

The assignment problem for weighted bipartite graph output when presented with G as input is the function $a^* : F \to M'$, which serves as basis to build the Compensatory Mating problem solution $b^c : F \to M$ as follows: $b_c(f) = m$, for each $a^*(f) = [m, i]$, for some $1 \leq i \leq n_b(m)$.

The bipartite graph built by the reduction process contains the set of dams F as its left side all the elements in M belonging to the actual domain of the mating function b, multiplied by the number of times they appear in the mating function. Therefore, both sides of G, namely F and M', have equal sizes. The graph edges connect each dam-sire pair below the inbreeding limit. Since it is a minimisation problem, the missing edges could exist with a very large weight between them (represented as ∞). We have omitted them for the sake of clarity, knowing that graph algorithms that require complete graphs represent missing edges in that way. The weighting function informs how far the pair's contribution is from the mean function value μ_b or, in other words, how the mating of those animals contributes to the solution's variance. Therefore, minimising the total weight of the graph assignment problem actually indicates the mating function that delivers the lowest variance value. We prove that claim in Theorem 2.

Theorem 2 *Definition 2 presents a correct reduction from Problem 2 to Problem 3.*

Proof: Let $G = (F, M', E, w : E \to \mathbb{R})$ be the weighted bipartite graph built as in Definition 2. Theorem 1 assures that changing already mated pairs within an optimal solution does not change its value. Therefore, the requirement of Problem 2 is met by construction: the only sires within the set M' are those belonging to the optimal solution b in Definition 2. Now, let the solution for the assignment problem for weighted bipartite graph for G be a function $a^* : F \to M'$ such that $\sum_{f \in F} w(f, a^*(f))$ is minimum. Since $w(f, a^*(f)) = (\pi(a^*(f), f) - \mu_b)^2$, we have that $\sum_{f \in F}(\pi(a^*(f), f) - \mu_b)^2$ is minimum, where μ_b is the average value of b. The output reduction function maps each $a^*(f) = [m, i]$, for some $1 \leq i \leq n_b(m)$, to the pair (m, f). Therefore, the output function $b_c : F \to M$ where $b_c(f) = m$ whenever $a^*(f) = [m, i]$ assures that the value $s_c = \sum_{f \in F}(\pi(b_c(f), f) - \mu_b)^2$ is minimum. The variance of b_c, $\sigma_c^2 = s_c/|F|$ is also minimum, since $|F|$ is a constant value. □

Corollary 2 *The Optimal Compensatory Mating problem (Problem 2) belongs to the complexity class **P**.*

Proof: Direct from the fact it can be solved by the reduction $f_O(P_2(f_I(x)))$, which has worst-case running time $O(n^2) + O(mn + n^2 \log n) + O(n) = O(mn + n^2 \log n)$ [Ramshaw and Tarjan 2012]. In the case of the assignment problem for weighted bipartite graph, that expression translates into $O((|M|.|F|)|F| + |F|^2 \log |F|)$, which is polynomial in the number of animals involved. □

3. Conclusion

Mating systems devise a particular strategy to perform one or more breeding goals. Breeding goals are usually related to health issues and economic gains. Producers use selection indexes to measure each animal value. Arranging the pairing of sires and dams can lead to different results, such as the production of exceptional animals or to a more homogeneous herd. In either case, one wants to maximise the total herd next generation's expected value.

This paper shows a reduction from the compensatory mating problem to the minimal assignment for weighted bipartite graphs. A compensatory mating intends to produce a more homogeneous offspring, given the breeding goals expressed as a selection index. The literature on breeding problems does not appear to discuss their complexity classes thoroughly. This paper gives a sound contribution to the matter by showing a polynomial-time algorithm to solve it.

We can devise further developments from the results presented here. First and foremost, the problem appears to admit a more efficient solution than the Hungarian algorithm. The same happened to the maximisation problem for the mating assignment: although we could have used the same reduction presented in this paper, we have developed an optimal greedy strategy with worst-case quadratic execution time [Ferreira et al. 2021].

Another critical problem, which is the minimisation of the herd inbreeding levels, will be carried on in the future. All algorithms developed so far rely on a parameter specifying the maximum number of times a sire can participate in the breeding process. That parameter is a rule of thumb, not assuring the co-ancestry relations within the generations produced over time. That information is critical for animal breeding programs. We intend to explore the results reached so far to create a model of long-term strategies to achieve breeding goals whilst preserving the co-ancestry levels.

References

Arora, S. and Barak, B. (2009). *Computational Complexity: a Modern Approach*. Cambridge University Press, Cambridge, UK.

Barreto Neto, A. D. et al. (2014). Estrutura populacional e otimização de esquemas de acasalamento em ovinos com uso de algoritmos evolucionários.

Cardoso, F. (2009). Ferramentas e estratégias para o melhoramento genético de bovinos de corte. *Embrapa Pecuária Sul-Documentos (INFOTECA-E)*.

Carvalheiro, R., Queiroz, S. A. d., and Kinghorn, B. (2010). Optimum contribution selection using differential evolution. *Revista Brasileira de Zootecnia*, 39(7):1429–1436.

Carvalho, T., Santos, N., Lira, W., Oliveira, P. A., Neto, P. S., Lindenberg, J., and Rabêlo, R. (2016). Um sistema de informação para melhoramento genético de caprinos e ovinos. In *Anais Principais do XII Simpósio Brasileiro de Sistemas de Informação*, pages 100–107. SBC.

Charlesworth, D. and Willis, J. H. (2009). The genetics of inbreeding depression. *Nature Reviews Genetics*, 10:783–796.

da Fontoura, D. C. N., Camargo, S. S., de Almeida Torres Junior, R. A., Carvalho, H. G., and Cardoso, F. F. (2020). Optimizing mate selection: a genetic algorithm approach. In *Proceedings of the ICAR Conference*, ICAR Technical Series no. 24 - NEW TRAITS AND ADDING VALUE TO THE RECORDING AND ID SERVICES IN THE ANIMAL PRODUCTION, Prague, CZ.

Dunne, F. L., Evans, R. D., Kelleher, M. M., Walsh, S. W., and Berry, D. P. (2021). Formulation of a decision support tool incorporating both genetic and non-genetic effects to rank young growing cattle on expected market value. *Animal*, 15:100077.

Eler, J. P. (2017). *Teorias e métodos em melhoramento genético animal: sistemas de acasalamento*. Faculdade de Zootecnia e Engenharia de Alimentos da USP.

Falconer, D. S. and Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. Longman Group Ltd., Harlow, 4 edition.

Ferreira, A. P. L., Yokoo, M. J.-I., and Motta, B. E. T. (2021). On the problem of optimal mating in animal breeding. In *Simposio Latinoamericano de Teoría Computacional, Proceedings of the XLVII Conferencia Latinoamericana de Informática (CLEI 2021) (to be pulished in IEEE Xplore)*, San José, Costa Rica.

Garey, M. R. and Johnson, D. S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness (Series of Books in the Mathematical Sciences)*. W. H. Freeman, 1st edition.

Kinghorn, B. P. (2011). An algorithm for efficient constrained mate selection. *Genetics Selection Evolution*, 43(1):4.

Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2:83–97.

Mi, M. P., Chapman, A. R., and Tyler, W. J. (1965). Effect of mating system on production traits in dairy cattle. *J. Dairy Sci.*, 48:77–84.

Munkres, J. (1957). Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics*, 5(1):32–38.

Nayeri, S., Sargolzaei, M., and Tulpan, D. (2021). A review of traditional and machine learning methods applied to animal breeding. *Animal Health Research Reviews*, 20:31–46.

Ramshaw, L. and Tarjan, R. E. (2012). On minimum-cost assignments in unbalanced bipartite graphs. Technical Report HPL-2012-40.

Simões, M. R. S., Leal, J. J. B., Minho, A. P., Gomes, C. C., MacNeil, M. D., Costa, R. F., Junqueira, V. S., Schmidt, P. I., Cardoso, F. F., Boligon, A. A., and Yokoo, M. J. (2020). Breeding objectives of brangus cattle in brazil. *Journal of Animal Breeding and Genetics*, 137(2):177–188.

Weigel, K. A., VanRaden, P. M., Norman, H. D., and Grosu, H. (2017). A 100-year review: Methods and impact of genetic selection in dairy cattle—from daughter–dam comparisons to deep learning algorithms. *J. Dairy Sci.*, 100(12):10234–10250.

Yokoo, M. J.-I., de Oliveira Seno, L., Oliveira, L. C., da Costa, P. U., da Silva, G. M., Suñé, R. W., and Cardoso, F. F. (2019). Economic selection index in small rural dairy farms. *The Journal of Dairy Research*, 86(1):25–33.