A general quadratic programming method for the optimisation of genetic contributions using interior point algorithm

R. Pong-Wong and J.A. Woolliams

The Roslin Institute and R(D)SVS, University of Edinburgh, Easter Bush, EH25 9RG, UK. ricardo.pong-wong@roslin.ed.ac.uk (Corresponding author)

Summary

A new method for optimum contribution selection based on quadratic programming was developed. Results from testing the methods using two datasets showed that the proposed method yielded the true optimum solution as observed with the method based on semidefinite programming. The flexibility of the method to allow the inclusion of more than one restriction on coancestry may prove to be a useful tool for a more customised management using dense SNP genotyping to better control the genetic diversity of critical regions of the genome.

Keywords: optimum contribution selection, quadratic programming

Introduction

Optimum contribution selection (OCS) is an effective tool for controlling the rate at which coancestry increases in close managed populations. However, despite the potential benefit of OCS, its practical uptake remains low. A reason for this is the limited number of methods implementing OCS. So far OCS has been implemented using three different approaches based on: (i) relaxed parameter space (Meuwissen, 1997), (ii) evolutionary algorithms (Kinghorn et al., 2002) and (iii) semidefinite programming (Pong-Wong & Woolliams, 2007). Methods based on relaxed parameter space are fast but may yield suboptimal solutions and their modification to add more than one constraint on coancestry is not trivial. Methods based on evolutionary algorithms are flexible but their convergence cannot be ensured. Finally, the semidefinite programming methods guarantee to yield the optimum solution, but they can be computationally demanding. The objective of this study is to propose a new OCS method based on quadratic programming, which guarantees optimum solution, can accommodate for multiple constraints and is faster than the semidefinite programming ones.

Material and methods

Theory and notation

Let $n$ candidates be available for selection and their sex described with the incidence vectors $s$ and $d$, where $s_i = 1$ if candidate $i$ is a male and 0 otherwise and $d_i = 1 - s_i$. Their estimated breeding values are in the vector $g$. Let $c$ be the vector of the candidates’ genetic contributions, where $c_i$ represents half the proportion of offspring from candidate $i$. The values for $c_i$ range between $[0:05]$ and the sum of contributions with a sex group sums to 0.5.

The expected genetic and inbreeding level in the offspring generation is equal to $c'g$ and $c'Gc/2$, respectively (Woolliams & Thompson, 1994), where $G$ is the relationship matrix for
the group of candidates, which can be estimated using either pedigree, high dense marker information or both (Nejati-Javaremi et al., 1997; Villanueva et al., 2005).

**Optimum contribution selection**

OCS considers the genetic relationship among candidates to optimise their genetic contribution \( c \) for maximising response on a given breeding objective \( h(c) \) while controlling the genetic diversity in the offspring generation. More specifically, in populations undergoing directional selection, OCS aims at maximising genetic gain while restricting the rate of loss of the overall genetic diversity to a pre-set rate; and in conservation programmes, OCS can be used to minimise the loss of genetic diversity in the population. Recently, Gómez-Romano et al. (2016) extended OCS to consider separately the genetic diversity of several regions of the genome for a more customised approach imposing greater control for some regions considered to be more critical. A general formulation including all OCS variants described above can be defined as a quadratic programming to optimise \( c \) to:

Minimise: \( h(c) \)  

\[
s.t. \quad s'c = 0.5; \quad d'c = 0.5; \quad c \geq m; \quad c \leq \overline{m}; \quad \frac{c'Gc}{2} \leq F_j^*, j = 1, p
\]

where the objective \( h(c) \) is equal to either \(-c'g\) or \(c'Ge/2\), when the goal is to maximise genetic gain or minimise average coancestry, respectively. The first two restrictions ensure that the contribution within sex group sums to 0.5; the next two impose the restriction on the candidates’ minimum (\( m \)) and maximum (\( \overline{m} \)) contribution; and the last ones are restrictions on average coancestry to be applied separately to \( p \) different regions of the genome, where \( G_j \) is the relationship for region \( j \) and \( F_j^* \) is its maximum average coancestry to be allowed. The overall coancestry, calculated with pedigree or SNP information, may be included in the objective or as one of the restrictions. But, a given region cannot be part of the objective and the restrictions at the same time.

The Lagrangian function, \( \mathcal{L}(c, \alpha_s, \alpha_d, \lambda_m, \lambda_{\overline{m}}, \lambda_j) \), is:

\[
\mathcal{L}(c, \alpha_s, \alpha_d, \lambda_m, \lambda_{\overline{m}}, \lambda_j) = h(c) - \alpha_s (s'c - 0.5) - \alpha_d (d'c - 0.5) - \lambda_m (c - m) + \lambda_{\overline{m}} (c - \overline{m}) + \sum_{j=1}^{p} \lambda_j (c'G_jc/2 - F_j^*)
\]

(2)

where \( \alpha_s, \alpha_d, \lambda_m, \lambda_{\overline{m}}, \lambda_j \) are Lagrangian multipliers, with size 1, 1, \( n \), \( n \), and \( p \), respectively.

The Karush-Kuhn-Tucker optimality conditions for (1) are:

\[
\begin{align*}
\nabla_c h(c) - \lambda_s s - \lambda_d d - \lambda_m + \lambda_{\overline{m}} + \sum_{j=1}^{p} (\lambda_j G_j c) & = 0 \\
0.5 - s'c & = 0 \\
0.5 - d'c & = 0 \\
y_m - c + m & = 0 \\
y_{\overline{m}} + c - m & = 0 \\
[y_j + c'G_jc/2 - F_j^*] & = 0, j = 1, p \\
\Lambda_m y_m e & = 0 \\
\Lambda_m y_{\overline{m}} e & = 0 \\
\Lambda_j y_j e & = 0
\end{align*}
\]

(3)

and \( \lambda_m, \lambda_{\overline{m}}, \lambda_j, y_m, y_{\overline{m}}, y_j \) must be \( \geq 0 \). The vectors \( y_m, y_{\overline{m}} \) and \( y_j \) are slack variables associated to the inequality constraints, \( \Lambda_x \) and \( Y_x \) are diagonal matrices containing the values of \( \lambda_x \) and \( y_x \) in their diagonal and \( e \) is a vector of ones.
Defining $R(\theta)$ to be the nine LHS terms of the optimality conditions, the optimum solution for (1) is the roots of $R(\theta)$, and it may be searched iteratively using Newton-Raphson (NR). Given a current solution $\theta^i$, this can be updated as $\theta^{i+1} = \theta^i + \alpha \Delta \theta^i$, where $\Delta \theta^i$ is the solution of $[R(\theta^i)d\theta^i]\Delta \theta^i = -[R(\theta^i)]$ and $\alpha$ is a scalar controlling the step length to avoid the Lagrangian multipliers and slack variables for the inequality constraints to become negative. The updating process is repeated until the solution converges. However, the standard NR approach may lead to unfeasible solutions not fulfilling one or more of the restrictions. Here we propose to use the Mehrotra’s predictor-corrector algorithm to solve (1).

**Mehrotra’s predictor-corrector algorithm**

A brief description of the algorithm is given here, but a more detailed explanation can be found in Mehrotra (1992) and Kruth (2008). Basically, it attempts to guide the iterative updates along a central path of strictly feasible points, by finding the search direction from a perturbed NR step. The degree of perturbation gets reduced according the solution approaches the optimum value. The algorithm has two NR steps: the first one is a standard NR to find the degree of perturbation needed (i.e. the corrector and centring parameters) and the second one is a perturbed NR to find the search direction for updating the solution.

To facilitate the description of the algorithm, $R(\theta)$ is grouped into four blocks to distinguish the Lagrangian multipliers and slack variables associated to the equality and inequality constraints, so $\theta' = (c', \lambda'_e, \lambda'_u, y'_u)$, where $\lambda'_e = (\lambda'_s, \lambda'_d)$, $\lambda'_u = (\lambda'_m, \lambda'_r, \lambda'_f)$ and $y'_u = (y'_{m}, y'_{r}, y'_{f})$. Hence, the first block in $R(\theta)$ is the first optimality condition in (3), the second block are the second and third conditions, the third block are the following 3 conditions; and the fourth block the remaining 3 optimality conditions.

Let $\theta^i$ be an arbitrary starting solution with values for $\lambda^i_e$, $\lambda^i_u$ and $y^i_u$ being greater than zero, compute $R(\theta^i)$ and the complementary measure $u^i = \frac{(\lambda^i_u)'(y^i_u)}{n}$. The update protocol is, then, the following:

1. Compute $R(\theta^i)d\theta^i$
2. Solve the NR step $[R(\theta^i)d\theta^i]\Delta \theta^i = -[R(\theta^i)]$
3. Calculate the step length $\alpha^{aff}$ such as $(\lambda^i_u + \alpha^{aff}\Delta \lambda^i_u) and (y^i_u + \alpha^{aff}\Delta y^i_u) \geq 0$
4. Compute complementary measure $u^{aff} = \frac{(\lambda^i_u + \alpha^{aff}\Delta \lambda^i_u)'(y^i_u + \alpha^{aff}\Delta y^i_u)}{n}$
5. Compute centring parameter $\sigma = \left(\frac{u^{aff}}{u^i}\right)^3$
6. Add corrector $(\Delta \lambda^i_u, \Delta y^i_u e)$ and centring parameters $(-\sigma u^i e)$ to the fourth block of $R(\theta^i)$ to produce the perturbed $R(\theta^i)$
7. Solve a NR step using $R(\theta^i)$ as the RHS: $[R(\theta^i)d\theta^i]\Delta \theta^i = -[R(\theta^i)]$
8. Calculate the step length $\alpha$ with new values from 7
9. Update new solution $\theta^{i+1} = \theta^i + \alpha \Delta \theta^i$
10. Compute $R(\theta^{i+1})$ and the new complementary measure $u^{i+1} = \frac{(\lambda^{i+1}_u)'(y^{i+1}_u)}{n}$
11. Test for convergence. If it has not, go back to step 1 and start a new update cycle.

**Results and discussion**

The method proposed here was tested using two populations. The first one (small population) is the same example B used by Pong-Wong & Woolliams (2007) to demonstrate the potential
problem with suboptimal solutions with methods based on relaxed parameter space. The number of candidates to be optimised was 6 and the relationship matrix was calculated from pedigree information. The second population (large population) consists of 200 candidates with a genome of two chromosomes in which SNP were simulated using mutation-drift equilibrium. Relationship matrix was calculated for each chromosome using the method from Nejati-Javaremi et al. (1997) with 1000 SNP for each chromosome. The small population was used to test the performance of the proposed method when restriction on one F* was added to the optimisation. On the other hand, the optimisation in the large population included two separate restrictions on coancestry, one for each chromosome.

Figure 1 shows the results for the small population for several scenarios varying F* from 0.178 to 0.30. When the restriction on F* was less than 0.271, the proposed method yielded solutions in which their expected inbreeding was the same as the imposed restriction. Restriction greater than 0.271 resulted in solution where only one male was selected. Additionally, including a restriction on maximum contribution of 0.3, yielded the same optimum results observed with semidefinite programming (but the relaxed parameter space method results in a suboptimal solution).

Table 1 shows the results from optimising the large population with several levels of restriction on the chromosomal F*. The same as in the small population, the expected genetic gain was larger for scenarios allowing for greater value for F*. In all scenarios, the proposed method yielded valid solutions where the expected level of chromosomal inbreeding was always smaller than F*. The results are the same as the ones observed using the method based on semidefinite programing.

Conclusions

A new OCS method based on quadratic programming was proposed here. The testing of the method using two populations always yielded optimum solutions similar to those observed with semidefinite programming. The proposed OCS method managed to find the optimum solutions in situations where methods based on relaxed parameter space failed to do.

List of References


Figure 1: Expected genetic gain and level of inbreeding for the solution of several scenarios optimising the small population with different degree of restriction of F*.

Table 1. Expected gain and level of inbreeding from solution of several scenarios optimising the large population with two separate restrictions of the diversity of each chromosomal diversity.

<table>
<thead>
<tr>
<th>Restriction on F*</th>
<th>Results in optimum solution</th>
<th>Genetic gain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>chr 1</td>
<td>chr 2</td>
</tr>
<tr>
<td>0.63</td>
<td>0.58</td>
<td>0.59</td>
</tr>
<tr>
<td>0.64</td>
<td>0.59</td>
<td>0.60</td>
</tr>
<tr>
<td>0.65</td>
<td>0.60</td>
<td>0.61</td>
</tr>
<tr>
<td>0.66</td>
<td>0.61</td>
<td>0.62</td>
</tr>
<tr>
<td>0.67</td>
<td>0.62</td>
<td>0.63</td>
</tr>
<tr>
<td>0.68</td>
<td>0.63</td>
<td>0.64</td>
</tr>
<tr>
<td>0.69</td>
<td>0.64</td>
<td>0.66</td>
</tr>
<tr>
<td>0.70</td>
<td>0.65</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>0.58</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>0.59</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>0.61</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>0.62</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>0.63</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>0.64</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>0.65</td>
<td>0.70</td>
</tr>
</tbody>
</table>