# Chapter 1
# General Introduction

**Abstract** We describe the main characteristics of two approaches to the linear selection indices theory. The first approach is called *standard linear selection indices* whereas the second of them is called *eigen selection index methods.* In the first approach, the economic weights are fixed and known, whereas in the second approach the economic weights are fixed but unknown. This is the main difference between both approaches and implies that the eigen selection index methods include to the standard linear selection indices because they do not require that the economic weights be known. Both types of indices predict the net genetic merit and maximize the selection response, and they give the breeder an objective criterion to select individuals as parents for the next selection cycle. In addition, in the prediction they can use phenotypic, markers, and genomic information. In both approaches, the indices can be unrestricted, null restricted or predetermined proportional gains and can be used in the context of single-stage or multistage breeding selection schemes. We describe the main characteristics of the two approaches to the linear selection indices theory and we finish this chapter describing the Lagrange multiplier method, which is the main tool to maximize the selection index responses.

Linear selection indices that assume that economic weights are fixed and known to predict the net genetic merit are based on the linear selection index theory originally developed by Smith (1936), Hazel and Lush (1942), and Hazel (1943). They are called *standard linear selection indices* in this introduction. Linear selection indices that assume that economic weights are fixed but unknown are based on the linear selection index theory developed by Cerón-Rojas et al. (2008a, 2016) and are called *Eigen selection index methods*. The Eigen selection index methods include the standard linear selection indices as a particular case because they do not require the economic weights to be known. To understand the Eigen selection index methods theory, the point is to see that this is an application of the canonical correlation theory to the standard linear selection index context. The multistage linear selection index theory will be described only in the context of the standard linear selection indices. As we shall see, there are three main types of LSI: phenotypic, marker, and genomic. Each can be unrestricted, null restricted or



J. J. Céron-Rojas, J. Crossa, *Linear Selection Indices in Modern Plant Breeding*,
https://doi.org/10.1007/978-3-319-91223-3_1

predetermined proportional gains and can be used in the context of single-stage or multistage breeding selection schemes.

For each specific selection index described in this book, we have used an acronym. For example, the Smith (1936), Hazel and Lush (1942), and Hazel (1943) index was denoted LPSI (linear phenotypic selection index), whereas the Cerón-Rojas et al. (2008a) index was denoted ESIM (Eigen selection index method), etc. For additional details, see Table 1.1 and the Preface of this book. We think that such notation gives the reader a more general point of view of the relationship that exists among all the indices described in this book.

**Table 1.1** Chapter where the index was described, authors who developed the selection index, acronym of the index used in this book, and description of the acronym

| Chapter | Authors who developed the index | Acronym | Description |
|---|---|---|---|
| 2 | Smith(1936), Hazel and Lush (1942), Hazel (1943) | LPSI[a] | Linear phenotypic selection index |
| | Williams (1962a) | BLPSI[a] | Base linear phenotypic selection index |
| 3 | Kempthorne and Nordskog (1959) | RLPSI[a] | Restricted linear phenotypic selection index |
| | Mallard (1972), Harville (1975), Tallis (1985), Itoh and Yamada (1987) | PPG-LPSI[a] | Predetermined proportional gain linear phenotypic selection index |
| | Pesek and Baker (1969), Yamada et al. (1975), Itoh and Yamada (1986) | DG-LPSI[a] | Desired gains linear phenotypic selection index |
| 4 | Lande and Thompson (1990) | LMSI[b] | Linear marker selection index |
| | Lange and Whittaker (2001) | GW-LMSI[b] | Genome-wide linear marker selection index |
| 5 | Togashi et al. (2011), Ceron-Rojas et al. (2015) | LGSI[c] | Linear genomic selection index |
| | Dekkers (2007) | CLGSI[d] | Combined linear genomic selection index |
| 6 | Kempthorne and Nordskog (1959), Ceron-Rojas et al. (2015) | RLGSI[c] | Restricted linear genomic selection index |
| | Tallis(1985), Ceron-Rojas et al. (2015) | PPG-LGSI[c] | Predetermined proportional gain linear genomic selection index |
| | Kempthorne and Nordskog (1959), Dekker (2007) | CRLGSI[d] | Combined restricted linear genomic selection index |
| | Tallis (1985), Dekker (2007) | PPG-CLGSI[d] | Predetermined proportional gain combined linear genomic selection index |
| 7 | Cerón-Rojas et al. (2008a) | ESIM[a] | Eigen selection index method |
| | Cerón-Rojas et al. (2008a) | RESIM[a] | Restricted eigen selection index method |
| | Cerón-Rojas et al. (2016) | PPG-ESIM[a] | Predetermined proportional gain eigen selection index method |

(continued)

**Table 1.1** (continued)

| Chapter | Authors who developed the index | Acronym | Description |
|---|---|---|---|
| 8 | Cerón-Rojas et al. (2008b) | MESIM[b] | Molecular eigen selection index method |
| | Crossa and Cerón-Rojas (2011) | GW-ESIM[b] | Genome-wide eigen selection index method |
| | Dekkers (2007), Cerón-Rojas et al. (2008b) | GESIM[d] | Genomic eigen selection index method |
| | Dekkers (2007), Cerón-Rojas et al. (2008a) | RGESIM[d] | Restricted genomic eigen selection index method |
| | Dekkers (2007), Cerón-Rojas et al. (2016) | PPG-GESIM[d] | Predetermined proportional gain genomic eigen selection index method |
| 9 | Cochran (1951), Young (1964) | MLPSI[a] | Multistage linear phenotypic selection index |
| | Cochran (1951), Young (1964), Kempthorne and Nordskog (1959) | MRLPSI[a] | Multistage restricted linear phenotypic selection index |
| | Cochran (1951), Young (1964), Tallis (1985) | MPPG-LPSI[a] | Multistage predetermined proportional gain linear phenotypic selection index |
| | Cochran (1951), Young (1964), Ceron-Rojas et al. (2015) | MLGSI[c] | Multistage linear genomic selection index |
| | Cochran (1951), Young (1964), Kempthorne and Nordskog (1959), Ceron-Rojas et al. (2015) | MRLGSI[c] | Multistage restricted linear genomic selection index |
| | Cochran (1951), Young (1964), Tallis (1985), Ceron-Rojas et al. (2015) | MPPG-LGSI[c] | Multistage predetermined proportional gain linear genomic selection index |

[a]Indices that use only phenotypic information
[b]Indices that use marker and phenotypic information jointly
[c]Indices that use only genomic information
[d]Indices that use genomic and phenotypic information jointly in the prediction of the net genetic merit

## 1.1 Standard Linear Selection Indices

### *1.1.1 Linear Phenotypic Selection Indices*

Three main linear phenotypic selection indices used to predict the net genetic merit and select parents for the next selection cycle are the LPSI, the null restricted LPSI (RLPSI), and the predetermined proportional gains LPSI (PPG-LPSI). The LPSI is an unrestricted index, whereas the RLPSI and the PPG-LPSI allow restrictions to be imposed equal to zero and predetermined proportional gain restrictions respectively, on the trait expected genetic gain per trait values to make some traits change their mean values based on a predetermined level while the rest of the trait means remain without restrictions. All these indices are linear combinations of several observable and optimally weighted phenotypic trait values.

The simplest linear phenotypic selection index (LPSI) can be written as $I_B = \mathbf{w}'\mathbf{y}$, where $\mathbf{w}$ is a known vector of economic values and $\mathbf{y}$ is a vector of phenotypic values. We called this index the *base linear phenotypic selection index* (BLPSI). In this case, the breeder does not need to estimate any parameters, and some authors have indicated that the BLPSI is a good predictor of the net genetic merit ($H = \mathbf{w}'\mathbf{g}$, where $\mathbf{g}$ is a vector of true unobservable breeding values) when no data are available for estimating the phenotypic ($\mathbf{P}$) and genotypic ($\mathbf{G}$) covariance matrices. When the traits are independent and the economic weights are also known, the LPSI can be written as $I = \sum_{i=1}^{t} w_i h_i^2 y_i$, and when the economic weights are not known, the LPSI is $I = \sum_{i=1}^{t} h_i^2 y_i$, where $w_i$ is the $i$th economic weight and $h_i^2$ is the heritability of trait $y_i$. In Chap. 2 (Sects. 2.5.1 and 2.5.2), we will show that the foregoing three indices are particular cases of the more general LPSI, i.e., $I = \mathbf{b}'\mathbf{y}$, where $\mathbf{b}$ is the $I$ vector of coefficients and $\mathbf{y}$ is the vector of observable trait phenotypic values. In the latter case, we need to estimate matrices $\mathbf{P}$ and $\mathbf{G}$.

The LPSI was originally proposed by Smith (1936) in the plant breeding context; later Hazel and Lush (1942) and Hazel (1943) extended the LPSI to the context of animal breeding. These authors made a clear distinction between the LPSI and the net genetic merit. The net genetic merit was defined as a linear combination of the unobservable true breeding values of the traits weighted by their respective economic values. In the LPSI theory, the main assumptions are: the genotypic values that make up the net genetic merit are composed entirely of the additive effects of genes, the LPSI and the net genetic merit have a joint normal distribution, and the regression of the net genetic merit on LPSI values is linear. Two of the main parameters of this index are the selection response and the expected genetic gain per trait or multi-trait selection response. The LPSI selection response is associated with the mean of the net genetic merit and was defined as the mean of the progeny of the selected parents or the mean of the future population (Cochran 1951). The selection response enables breeders to estimate the expected selection progress before carrying it out. This information gives improvement programs a clearer orientation and helps to predict the success of the adopted selection method and choose the option that is technically most effective on a scientific basis (Costa et al. 2008). On the other hand, the LPSI expected genetic gain per trait, or multi-trait selection response, is the population mean of each trait under selection of the progeny of the selected parents. Thus, although the LPSI selection response is associated with the mean of the net genetic merit, the LPSI expected genetic gain per trait is associated with the mean of each trait under selection. The foregoing definition of selection response and the expected genetic gain per trait are valid for all selection indices described in this book.

One of the main problems of the LPSI is that when used to select individuals as parents for the next selection cycle, the expected mean of the traits can increase or decrease in a positive or negative direction without control. This was the main reason why Kempthorne and Nordskog (1959) developed the basics of the restricted LPSI (RLPSI), which allows restrictions to be imposed equal to zero on the expected

genetic gain of some traits whereas the expected genetic gain of other traits increases (or decreases) without any restrictions being imposed. Based on the results of the RLPSI, Tallis (1962) and James (1968) proposed a selection index called predetermined proportional gains LPSI (PPG-LPSI), which attempts to make some traits change their expected genetic gain values based on a predetermined level, while the rest of the traits remain without restrictions. Mallard (1972) pointed out that the PPG-LPSI proposed by Tallis (1962) and James (1968) does not provide optimal genetic gains and was the first to propose an optimal PPG-LPSI based on a slight modification of the RLPSI. Other optimal PPG-LPSIs were proposed by Harville (1975) and Tallis (1985). Itoh and Yamada (1987) showed that the Mallard (1972) index is equal to the Tallis (1985) index and that, except for a proportional constant, the Tallis (1985) index is equal to the Harville (1975) index. Thus, in reality, there is only one optimal PPG-LPSI.

In Chap. 3 (Sect. 3.1.1 and 3.2.1), we show that $\mathbf{b}_R = \mathbf{K}\mathbf{b}$ and $\mathbf{b}_P = \mathbf{K}_P\mathbf{b}$ are the vectors of coefficients of the RLPSI and PPG-LPSI, respectively, where $\mathbf{b}$ is the LPSI vector of coefficients. Matrices $\mathbf{K}$ and $\mathbf{K}_P$ are idempotent ($\mathbf{K} = \mathbf{K}^2$ and $\mathbf{K}_P = \mathbf{K}_P^2$), that is, they are projectors. Matrix $\mathbf{K}$ projects $\mathbf{b}$ into a space smaller than the original space of $\mathbf{b}$ because the restrictions imposed on the expected genetic gains per trait are equal to zero (Sect. 3.1.1). The reduction of the space into which matrix $\mathbf{K}$ projects $\mathbf{b}$ will be equal to the number of null restrictions imposed by the breeder on the expected genetic gain per trait, or multi-trait selection response. In the PPG-LPSI context, matrix $\mathbf{K}_P$ has the same function as $\mathbf{K}$ (see Sect. 3.2.1 for details).

The aims of the LPSI, RLPSI, and PPG-LPSI are to:

1. Predict the unobservable net genetic merit values of the candidates for selection.
2. Maximize the selection response and the expected genetic gain for each trait.
3. Provide the breeder with an objective rule for evaluating and selecting several traits simultaneously (Baker 1974).

The LPSI is described in Chap. 2, and the RLPSI and PPG-LPSI are described in Chap. 3. As we will be see in this book, the RLPSI and PPG-LPSI theories can be extended to all selection indices described in this book. Also, the main objectives of all selection indices described in this book are the same as those of the LPSI, RLPSI, and PPG-LPSI.

### *1.1.2 Linear Marker Selection Indices*

The linear marker selection index (LMSI) and the genome-wide LMSI (GW-LMSI) are employed in marker-assisted selection (MAS) and are useful in training populations when there is phenotypic and marker information; both are a direct application of the LPSI theory to the MAS context. The LMSI was originally proposed by Lande and Thompson (1990), and the GW-LMSI was proposed by Lange and Whittaker (2001). The fundamental idea of these authors is based on the fact that crossing two inbred lines generates linkage disequilibrium between markers

and quantitative trait loci (QTL), which is useful for identifying markers correlated with the traits of interest and estimating the correlation between each of the selected markers and the trait; the selection criteria are then based upon this marker information (Moreau et al. 2007). The LMSI combines information on markers linked to QTL and the phenotypic values of the traits to predict the net genetic merit of the candidates for selection because it is not possible to identify all QTL affecting the economically important traits (Li 1998). That is, unless all QTL affecting the traits of interest can be identified, phenotypic values should be combined with the marker scores to increase LMSI efficiency (Dekkers and Settar 2004).

Moreau et al. (2000) and Whittaker (2003) found that the LMSI is more effective than LPSI only in early generation testing and that LMSI increased costs because of molecular marker evaluation. The LMSI assumes that favorable alleles are known, as are their average effects on phenotype (Lande and Thompson 1990; Hospital et al. 1997). This assumption is valid for major gene traits but not for quantitative traits that are influenced by the environment and many QTLs with small effects interacting among them and with the environment. The LMSI requires regressing phenotypic values on marker-coded values and, with this information, constructing the marker score for each individual candidate for selection, and then combining the marker score with phenotypic information using the LMSI to obtain a final prediction of the net genetic merit. Several authors (Lange and Whittaker 2001; Meuwissen et al. 2001; Dekkers 2007; Heffner et al. 2009) have criticized the LMSI approach because it makes inefficient use of the available data. It would be preferable to use all the available data in a single step to achieve maximally accurate estimates of marker effects. In addition, because the LMSI is based on only a few large QTL effects, it violates the selection index assumptions of multivariate normality and small changes in allele frequencies.

Lange and Whittaker (2001) proposed the genome-wide LMSI (GW-LMSI) as a possible solution to LMSI problems. The GW-LMSI is a single-stage procedure that treats information at each individual marker as a separate trait. Thus, all marker information can be entered together with phenotypic information into the GW-LMSI, which is then used to predict the net genetic merit and select candidates. Both selection indices are described in Chap. 4.

### *1.1.3 Linear Genomic Selection Indices*

The linear genomic selection index (LGSI) is a linear combination of genomic estimated breeding values (GEBVs) and was originally proposed by Togashi et al. (2011); however, Ceron-Rojas et al. (2015) developed the LGSI theory completely. The advantage of the LGSI over the other indices lies in the possibility of reducing the intervals between selection cycles by more than two thirds. A 4-year breeding cycle (including 3 years of field testing) is thus reduced to only 4 months, i.e., the time required to grow and cross a plant. As a result, thousands of candidates for selection can be evaluated without ever taking them out to the field (Lorenz et al. 2011).

In the LGSI, phenotypic and marker data from the training population are fitted in a statistical model to estimate all available marker effects; these estimates are then used to obtain GEBVs that are predictors of breeding values in a testing population for which there is only marker information. The GEBV can be obtained by multiplying the genomic best linear unbiased predictor (GBLUP) of the estimated marker effects in the training population (Van Raden 2008) by the coded marker values obtained in the testing population in each selection cycle. Applying the LGSI in plant or animal breeding requires genotyping the candidates for selection to obtain the GEBV, and predicting and ranking the net genetic merit of the candidates for selection using the LGSI. An additional genomic selection index was given by Dekkers (2007); however, this index can only be used in training populations because GEBV and phenotypic information are jointly used to predict the net genetic merit. Both indices are described in Chap. 5 and in Chap. 6, we describe both indices in the context of the restricted selection indices.

## 1.2 Eigen Selection Index Methods

The eigen selection index methods are described in Chaps. 7 and 8. As we shall see, these indices are only used in training populations and can be unrestricted, restricted, and predetermined proportional gains selection indices; they can also use phenotypic and/or marker information to predict the net genetic merit. In the context of this linear selection index theory, it is assumed that economic weights are fixed but unknown. The eigen selection index methods is based on the canonical correlation theory and applied to the LPSI, RLSPI, etc., selection indices's context.

### *1.2.1 Linear Phenotypic Eigen Selection Index Method*

Cerón-Rojas and Sahagún-Castellanos (2005) and Cerón-Rojas et al. (2006) proposed a phenotypic selection index in the principal component context that has low accuracy; later, Cerón-Rojas et al. (2008a, 2016) developed the eigen selection index method (ESIM), the restricted ESIM (RESIM) and the predetermined proportional gain ESIM (PPG-ESIM) in the canonical correlations context (Hotelling 1935, 1936). The ESIM is an unrestricted index, but the RESIM and PPG-ESIM allow null and predetermined restrictions respectively to be imposed on the expected genetic gains of some traits, whereas the rest remain without restrictions. The latter three indices use only phenotypic information to predict the individual net genetic merit of the candidate for selection and use the elements of the first eigenvector of the multi-trait heritability as the index vector of coefficients and the first eigenvalue of the multi-trait heritability in their selection response. The main objectives of the three indices are to predict the unobservable net genetic merit values of the candidates for selection, maximize the selection response and the expected genetic gain

per trait, and provide the breeder with an objective rule for evaluating and selecting several traits simultaneously. Their main characteristics are:

1. They do not require the economic weights to be known.
2. The first eigenvector of the multi-trait heritability is used as their vector of coefficients, and the first eigenvalue of the multi-trait heritability is used in the selection response.
3. Owing to the properties associated with eigen analysis, it is possible to use the theory of similar matrices (Harville 1997) to change the direction and proportion of the expected genetic gain values without affecting the accuracy.
4. The sampling statistical properties of ESIM are known.
5. The PPG-ESIM does not require a proportional constant.

Finally, the main theory describe in Chapter 7 was developed by Cerón-Rojas et al.(2008a, 2016) based on the canonical correlation framework. That is, ESIM and its variants (RESIM, MESIM, PPG-ESIM) are applications of the canonical correlation theory to the LPSI context.

### *1.2.2 Linear Marker and Genomic Eigen Selection Index Methods*

Cerón-Rojas et al. (2008b) and Crossa and Cerón-Rojas (2011) extended the ESIM to a molecular ESIM (MESIM) and to a genome-wide ESIM (GW-ESIM), respectively, similar to the linear molecular selection index (LMSI) and to the genome-wide LMSI (GW-LMSI). The MESIM and GW-ESIM have problems similar to those associated with the LMSI and GW-LMSI respectively (Chap. 4 for details). The MESIM and GW-ESIM use phenotypic information and markers linked to QTL to predict the net genetic merit, but the GW-ESIM omits the molecular selection step in the prediction. The main difference among the MESIM, the GW-ESIM, the LMSI, and the GW-LMSI is how they obtain the vector of coefficients: while the LMSI and GW-LMSI obtain the vector of coefficients according to the LPSI theory, the MESIM and the GW-ESIM obtain the vector of coefficients based on canonical correlation analysis and the singular value decomposition theory.

It is possible to extend the ESIM to a genomic ESIM (GESIM), and the restricted RESIM and the PPG-ESIM can be extended to a restricted genomic ESIM (RGESIM) and to a predetermined proportional gain genomic ESIM (PPG-GESIM) that use phenotypic and GEBV information jointly to predict the net genetic merit of the candidates for selection, maximizing the selection response and optimizing the expected genetic gain per trait; but although the GESIM is not constrained, the RGESIM and the PPG-GESIM allow null and predetermined restrictions respectively to be imposed on the expected genetic gain to make some traits change their mean values based on a predetermined level, while the rest of the traits remain without any restriction.

## 1.3 Multistage Linear Selection Indices

Multistage linear selection indices are methods of selecting one or more individual traits available at different times or stages and are applied mainly in animals and tree breeding where the traits under consideration become evident at different ages. The theory of these indices is based on the independent culling level method and the standard linear selection index theory. There are two main approaches associated with these indices:

1. The *optimal* multistage linear selection index, which takes into consideration the correlation among indices at different stages when makes selection.
2. The s*election index updating* or *decorrelated* multistage linear selection index, in which the correlation among indices at different stages is zero when makes selection.

These indices can use phenotypic or GEBV information to predict the net genetic merit or combine phenotypic and GEBV in the prediction. These indices can also be unrestricted, null restricted or predetermined proportional gains. In this book, we describe only the *optimal* multistage linear selection index in Chap. 9 and, in this book, it is called simply multistage linear selection index.

Multistage linear selection indices are a cost-saving strategy for improving multiple traits, because not all traits need to be measured at each stage. Thus, when traits have a developmental sequence in ontogeny or there are large differences in the costs of measuring several traits, the efficiency of this index over LPSI efficiency can be substantial (Xu et al. 1995). Xu and Muir (1992) have indicated that the optimal multistage linear phenotypic selection index (MLPSI) increases selection intensity on traits measured at an earlier age, and, with fixed facilities, a greater number of individuals can be selected at an earlier age. For example, if some individuals can be culled before final traits are measured (e.g., weaning weights in swine and beef cattle breeding), savings are realized in terms of feed, labor, and facilities. With the LPSI, the same individuals must be measured for each trait; thus, the number of traits measured per mature individual is the same as that for an immature individual.

The original MLPSI was developed by Cochran (1951) in the two-stage context and later, Young (1964) and Cunningham (1975) combined the LPSI theory with the independent culling method to simultaneously select more than one trait in the multistage selection context. This selection method was called *multistage selection* by Cochran (1951) and Young (1964) and *multistage index selection* by Cunningham (1975).

The MLPSI theory can also be adapted to the genomic selection context, where it is possible to develop an optimal multistage unrestricted, restricted, and predetermined proportional gains linear genomic selection index. The latter indices are linear combinations of estimated breeding values (GEBV) used to predict the individual net genetic merit and select individual traits available at different stages in a non-phenotyped testing population and are called *multistage linear genomic selection indices*. The advantage of these indices over the other selection indices

lies in the possibility of reducing the intervals between selection cycles or stages by more than two thirds.

One of the main problems of all the multistage selection indices is that after the first selection stage their values could be non-normally distributed. In addition, for more than two stages, those indices require computationally sophisticated multiple integration techniques to derive selection intensities, and there are problems of convergence when the traits and the index values of successive stages are highly correlated. Furthermore, the computational time could be unacceptable if the number of selection stages becomes too high (Börner and Reinsch 2012). One possible solution to these problems was given by Xu and Muir (1992) in the s*election index updating* or *decorrelated* multistage linear phenotypic selection index context. However, one problem with the *decorrelated* multistage selection index is that its accuracy and selection response is generally lower than the accuracy and selection response of the multistage selection index described in this book.

## 1.4 Stochastic Simulation of Four Linear Phenotypic Selection Indices

Chapter 10 describes a stochastic simulation of four linear indices: LPSI, ESIM, RLPSI, and RESIM. We think that stochastic simulation can contribute to a better understanding of the relationship between these indices and their accuracies to predict the net genetic merit.

## 1.5 RIndSel: Selection Indices with R

Chapter 11 describes how RIndSel can be used to determine individual candidates as parents for the next cycle of improvement. RIndSel is a graphical unit interface that uses the selection index theory to make selection. The index can be a linear combination of phenotypic values, genomic estimated breeding values or a linear combination of phenotypic values and marker scores.

## 1.6 The Lagrange Multiplier Method

To obtain the constrained linear selection indices (e.g., RLPSI, PPG-LPSI, RESIM) described in Chaps. 3, 6, 7, 8, and 9, we used the method of Lagrange multipliers. This is a powerful method for finding extreme values (maxima or minima) of constrained functions. For example, the covariance between the breeding value vector ($\mathbf{g}$) and the LPSI ($I = \mathbf{b}'\mathbf{y}$) is $Cov(I, \mathbf{g}) = \mathbf{Gb}$. In the LPSI context, the $\mathbf{Gb}$ vector can

take any value (positive or negative) which could be a problem for some breeding objectives. That is, the breeder could be interested in improving only $(t - r)$ of $t$ $(r < t)$ traits, leaving $r$ of them fixed; that is, the expected genetic gains of $r$ traits will be equal to zero for a specific selection cycle. In such cases, we want $r$ covariances between the linear combinations of $\mathbf{g}$ $(\mathbf{U}'\mathbf{g})$ and the $I = \mathbf{b}'\mathbf{y}$ to be zero, i.e., $Cov(I, \mathbf{U}'\mathbf{g}) = \mathbf{U}'\mathbf{G}\mathbf{b} = \mathbf{0}$, where $\mathbf{U}'$ is a matrix with $r$ 1's and $(t - r)$ 0's; 1 indicates that the trait is restricted and 0 that the trait is not restricted. This is the main problem of the RLPSI, and the method of Lagrange multipliers is useful for solving that problem.

In the constrained linear selection indices context, the method of Lagrange multipliers involves maximizing (or minimizing) the Lagrange function: $L[H, I, \mathbf{g}, \mathbf{v}] = f(H, I) + \mathbf{v}'g(\mathbf{g}, I)$, where the elements of vector $\mathbf{v}'$ are called Lagrange multipliers. In the RLPSI context, $f(H, I) = E[(H - I)^2] = \mathbf{w}'\mathbf{G}\mathbf{w} + \mathbf{b}'\mathbf{P}\mathbf{b} - 2\mathbf{w}'\mathbf{G}\mathbf{b}$ is the mean squared difference between $I$ and $H$. Let $g(\mathbf{g}, I) = Cov(I, \mathbf{U}'\mathbf{g}) = \mathbf{U}'\mathbf{G}\mathbf{b}$ be the covariances between the linear combinations of $\mathbf{g}$ $(\mathbf{U}'\mathbf{g})$, and $I = \mathbf{b}'\mathbf{y}$, the LPSI. Then, to find the RLPSI vector of coefficients $\mathbf{b}_R = \mathbf{K}\mathbf{b}$, we need to minimize the Lagrange function: $\mathbf{b}'\mathbf{P}\mathbf{b} + \mathbf{w}'\mathbf{G}\mathbf{w} - 2\mathbf{w}'\mathbf{G}\mathbf{b} + 2\mathbf{v}'\mathbf{C}'\mathbf{b}$, with respect to vectors $\mathbf{b}$ and $\mathbf{v}' = [v_1 \;\; v_2 \;\; \cdots \;\; v_{r-1}]$, where $\mathbf{v}$ is a vector of Lagrange multipliers (see Chap. 3, Sect. 3.1.1 for details). Schott (2005) has given additional details associated with the method of Lagrange multipliers.

## References


Baker RJ (1974) Selection indexes without economic weights for animal breeding. Can J Anim Sci 54:1–8

Börner V, Reinsch N (2012) Optimising multistage dairy cattle breeding schemes including genomic selection using decorrelated or optimum selection indices. Genet Sel Evol 44(1):11

Cerón-Rojas JJ, Sahagún-Castellanos J (2005) A selection index based on principal components. Agrociencia 39:667–677

Cerón-Rojas JJ, Crossa J, Sahagún-Castellanos J, Castillo-González F, Santacruz-Varela A (2006) A selection index method based on eigenanalysis. Crop Sci 46:1711–1721

Cerón-Rojas JJ, Sahagún-Castellanos J, Castillo-González F, Santacruz-Varela A, Crossa J (2008a) A restricted selection index method based on eigenanalysis. J Agric Biol Environ Stat 13 (4):421–438

Cerón-Rojas JJ, Sahagún-Castellanos J, Castillo-González F, Santacruz-Varela A, Benítez-Riquelme I, Crossa J (2008b) A molecular selection index method based on eigenanalysis. Genetics 180:547–557

Ceron-Rojas JJ, Crossa J, Arief VN, Basford K, Rutkoski J, Jarquín D, Alvarado G, Beyene Y, Semagn K, DeLacy I (2015) A genomic selection index applied to simulated and real data. G3 (Bethesda) 5:2155–2164

Cerón-Rojas JJ, Crossa J, Toledo FH, Sahagún-Castellanos J (2016) A predetermined proportional gains eigen selection index method. Crop Sci 56:2436–2447

Cochran WG (1951) Improvement by means of selection. In: Neyman J (ed) Proceedings of the second Berkeley symposium on mathematical statistics and probability. University of California Press, Berkeley, CA, pp 449–470

Costa MM, Di Mauro AO, Unêda-Trevisoli SH, Castro Arriel NH, Bárbaro IM, Dias da Silveira G, Silva Muniz FR (2008) Analysis of direct and indirect selection and indices in soybean segregating populations. Crop Breed Appl Biotechnol 8:47–55

Crossa J, Cerón-Rojas JJ (2011) Multi-trait multi-environment genome-wide molecular marker selection indices. J Indian Soc Agric Stat 62(2):125–142
Cunningham EP (1975) Multi-stage index selection. Theor Appl Genet 46:55–61
Dekkers JCM (2007) Prediction of response to marker-assisted and genomic selection using selection index theory. J Anim Breed Genet 124:331–341
Dekkers JCM, Settar P (2004) Long-term selection with known quantitative trait loci. Plant Breed Rev 24:311–335
Harville DA (1975) Index selection with proportionality constraints. Biometrics 31(1):223–225
Harville DA (1997) Matrix Algebra from a statistician's perspective. Springer, New York
Hazel LN (1943) The genetic basis for constructing selection indexes. Genetics 8:476–490
Hazel LN, Lush JL (1942) The efficiency of three methods of selection. J Hered 33:393–399
Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. Crop Sci 49 (1):12
Hospital F, Moreau L, Lacoudre F, Charcosset A, Gallais A (1997) More on the efficiency of marker-assisted selection. Theor Appl Genet 95:1181–1189
Hotelling H (1935) The most predictable criterion. J Educ Psychol 26:139–142
Hotelling H (1936) Relations between two sets of variables. Biometrika 28:321–377
Itoh Y, Yamada Y (1986) Re-examination of selection index for desired gains. Genet Sel Evol 18 (4):499–504
Itoh Y, Yamada Y (1987) Comparisons of selection indices achieving predetermined proportional gains. Genet Sel Evol 19(1):69–82
James JW (1968) Index selection with restriction. Biometrics 24:1015–1018
Kempthorne O, Nordskog AW (1959) Restricted selection indices. Biometrics 15:10–19
Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743–756
Lange C, Whittaker JC (2001) On prediction of genetic values in marker-assisted selection. Genetics 159:1375–1381
Li Z (1998) Molecular analysis of epistasis affecting complex traits. In: Paterson AH (ed) Molecular dissection of complex traits. CRC Press, Boca Raton, New York, pp 119–1130
Lorenz AJ, Chao S, Asoro FG, Heffner EL, Hayashi T et al (2011) Genomic selection in plant breeding: knowledge and prospects. Adv Agron 110:77–123
Mallard J (1972) The theory and computation of selection indices with constraints: a critical synthesis. Biometrics 28:713–735
Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829
Moreau L, Lemarie S, Charcosset A, Gallais A (2000) Economic efficiency of one cycle of marker-assisted selection efficiency. Crop Sci 40:329–337
Moreau L, Hospital F, Whittaker J (2007) Marker-assisted selection and introgression. In: Balding DJ, Bishop M, Cannings C (eds) Handbook of statistical genetics, vol 1, 3rd edn. Wiley, New York, pp 718–751
Pesek J, Baker RJ (1969) Desired improvement in relation to selection indices. Can J Plant Sci 49:803–804
Schott JR (2005) Matrix analysis for statistics, 2nd edn. Wiley, Hoboken, NJ
Smith HF (1936) A discriminant function for plant selection. In: Papers on quantitative genetics and related topics. Department of Genetics, North Carolina State College, Raleigh, NC, pp 466–476
Tallis GM (1962) A selection index for optimum genotype. Biometrics 18:120–122
Tallis GM (1985) Constrained selection. Jpn J Genet 60(2):151–155
Togashi K, Lin CY, Yamazaki T (2011) The efficiency of genome-wide selection for genetic improvement of net merit. J Anim Sci 89:2972–2980
Van Raden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423

Whittaker JC (2003) Marker-assisted selection and introgression. In: Balding DJ, Bishop M, Cannings C (eds) Handbook of statistical genetics, vol 1, 2nd edn. Wiley, New York, pp 554–574

Xu S, Muir WM (1992) Selection index updating. Theor Appl Genet 83:451–458

Xu S, Martin TG, Muir WM (1995) Multistage selection for maximum economic return with an application to beef cattle breeding. J Anim Sci 73:699–710

Yamada Y, Yokouchi K, Nishida A (1975) Selection index when genetic gains of individual traits are of primary concern. Jpn J Genet 50(1):33–41

Young SSY (1964) Multi-stage selection for genetic gain. Heredity 19:131–143