Low-rank Approximations and Weighted Low-rank Approximations: an Application to Statistical Genetics

Paulo Canas Rodrigues[*,1,2]

[1]Center for Mathematics and Applications, Faculty of Sciences and Technology, Nova University of Lisbon, Portugal

[2]ISLA Campus Lisboa, Laureate International Universities, Lisbon, Portugal
paulocanas@gmail.com

The increase in the amount of data collected every day, in many disciplines, asks for new statistical strategies to handle it and to evaluate the research hypothesis timely. When dealing with high dimensional data, low-rank approximation based methods, such as singular value decomposition and principal component analysis (PCA) are of great interest. In many of these data sets some variables and/or individuals are more important and should be given higher importance/weight in the analysis. Therefore, algorithms for weighted low-rank approximations must be used.

In this paper we introduce a generalization of one of the most widely used models to study and understand genotype by environment interaction in quantitative genetics and plant breeding: the additive main effects and multiplicative interaction (AMMI) model, which is a class of linear-bilinear models, also called double centered PCA. This generalization—weighted AMMI (WAMMI) model—allows the inclusion of different weights for each column of the two-way data matrix (the environments) and, therefore, accounts for heterogeneity of error variance across environments. The usefulness of the WAMMI model is compared with the AMMI and linear mixed models by analyzing real and simulated data.

Key Words: Additive main effects and multiplicative interaction model, genotype by environment interaction, QTL by environment interaction, linear mixed model