Contributions to the dialogue between Statistics and Numerical Linear Algebra

J-J. Martínez[♭, 1]

(♭) Departamento de Física y Matemáticas, Universidad de Alcalá
Campus Universitario, Alcalá de Henares, Spain.

Abstract

The *linear regression model* is an important issue in Statistics, mathematically related to polynomial least squares fitting. In Lecture 11 (*Least Squares Problems*) of the book of Trefethen and Bau [6] we read: *Least squares data-fitting has been an indispensable tool since its invention by Gauss and Legendre around* 1800. The book of Datta [1] also devotes a chapter (*Chapter 8: Least Squares Solutions to Linear systems*) to this issue, and includes in that chapter an interesting reference ([4]): an interview with G. H. Golub. In that interview a fundamental contribution of Golub is recalled ([3]), and we can read a remarkable sentence by Golub: *So I tried to get the statisticians interested in doing numerical computations. A few people were interested in it, but I don't think it had a heavy influence in statistics. I doubt if today people really use decomposition methods rather than normal equations. Statisticians are fairly fixed in their ways.*

This is the subject we are addressing: the dialogue between Statistics and Numerical Linear Algebra. In the book [2] we find a chapter (*Chapter III.8: (Non)Linear Regression Modeling*) devoted to linear and nonlinear regression, in which we read: *There are many algorithms for constructing a suitable QR decomposition for finding LS estimates, such as the Householder or Givens transformations; see Chapter II.4 for more details.* So the reader interested in that subject is led to Chapter II.4, entitled *Numerical Linear Algebra*, and we see that this book on Computational Statistics naturally pays attention to Numerical Linear Algebra. In the book of Trefethen and Bau [6] we find the important Lecture 19 (*Stability of Least Squares Algorithms*) and in the book of Datta [1] the cited Chapter 8. In both books the advantage of using the *QR factorization* instead of the *normal equations* is stressed. These advantageous algorithms have their roots in the work of Golub [3].

The problem for researchers in Numerical Linear Algebra is that the paper by Golub was published in 1965. So, it would be interesting to look for more recent contributions. In this sense, it is useful to consider the case where the involved matrices are *totally positive*: this is achieved for the *Vandermonde matrices* involved in polynomial least squares fitting when the nodes are nonnegative and ordered in increasing order: $0 \leq x_0 < \ldots < x_n$. Then we can apply results on the QR factorization of a totally positive matrix (as seen in [5] for several classes of totally positive matrices) to obtain more accurate results when computing two important matrices in the linear

[1]jjavier.martinez@uah.es

regression model: the *projection matrix* and the *Moore-Penrose inverse* $A^\dagger$ of the Vandermonde matrix A.

Chapter II.4 of [2] also cites the Moore-Penrose inverse (the pseudo-inverse), but relating it to the *singular value decomposition* (SVD), while our approach computes it by using the QR factorization.

Acknowledgments

This work was partially supported by the Research Grant PID2022-138569NB-I00 from the Spanish Government. The author is member of the Reseach Group ASYNACS (Ref. CT-CE2019/683) of Universidad de Alcalá.

References

[1] Datta, B. N., Numerical Linear Algebra and Applications (Second Edition). SIAM, Philadelphia, 2010.

[2] Gentle, J. E., Härdle, W., Mori, Y., Handbook of Computational Statistics. Concepts and Methods. Springer-Verlag, Berlin, 2004.

[3] Golub, G. H., Numerical Methods for Solving Linear Least Squares Problems. *Numerische Mathematik*, 7:206–216, 1965.

[4] Higham, N. H., An interview with Gene Golub. *MIMS Eprint 2008.8*, Manchester Institute for Mathematical Seciences, Univ. of Manchester, 2008.

[5] Marco, A., Martínez, J. J., Accurate computation of the Moore-Penrose inverse of strictly totally positive matrices. *Journal of Computational and Applied Mathematics*, 350:299–308, 2019.

[6] Trefethen, L. N., Bau III, D., Numerical Linear Algebra. SIAM, Philadelphia, 1997.