L3. The H-principle of mathematical modeling

<u>Agnar Höskuldsson</u>
IPL, Technical University of Denmark, Kgs Lyngby, Denmark

The H-principle (the Heisenberg principle of mathematical modeling) is a recommendation of how mathematical models should be solved, when data are uncertain. It is recommended to carry out the modeling task in steps, where at each step the data (variables) are weighted according to their contribution to the aim of the modeling task. The primary interest in science and industry is the prediction aspect of the model. There are many ways to evaluate the prediction ability of a model. In the present work a cross-validation is used. 10% of the samples are selected at random and eliminated. The estimation of unknown parameters is carried out for the 90% of the samples. The response values of the 10% left out are then estimated using the estimated parameters. This is carried out 30 times. This procedure will tell fairly well of the prediction quality of the model.

Associated with the H-principle there have been developed a large collection of algorithms that have a common name of the 'H-method'. The H-method looks at the weight vectors and chooses the ones that contribute to the modeling task. The remaining weights are set to zero. In the case of linear regression the H-method gives results that have been optimized with respect to the prediction aspect of the model. Thus it gives better prediction than other traditional methods (PLS regression, Principal Component Regression, least squares regression and others).

Many methods in applied sciences use a positive definite matrix a starting point for finding the solution. Examples are Ridge regression, variance components in experimental design, Kalman filter methods and others. The H-method has been extended to these areas. It will be shown by examples that the H-method is superior to the traditional methods that use the full rank solution. Also, it is better to work with the solution obtained by the H-method, when carrying out the significance testing.

Estimation in non-linear models is traditionally carried out by the Gauss-Newton method, supplemented by Marquardt regularization. It will be shown by an example that the H-method gives better results that the traditional method. This holds both for the predictions derived from the model and also for the interpretation of parameters.

In many areas the classification of samples are important. New type of weight vectors have been defined such that the H-method can be used for classification of data. Applications of this method to spectral data have given superior results compared to traditional methods.

There is a tradition to use differentials and differential equations, when it is needed to describe changes of systems in time. The H-method has been extended to Path Modeling, where at each 'knot' (time point, status etc) there is given a data block, a matrix, that describes the situation. The H-method gives regression coefficients between the data blocks in a similar way as in linear regression, where there are only two data blocks. It is more efficient and more informative to use Path Modeling than the traditional approach by differential equations. An example of the usage of the H-method will be shown.

An important aspect of the H-method is that the vectors computed at each step have some special interpretations. Thus, using graphic analysis the user can get important insight into the latent variation of the data. These graphic features are available for all areas, where the H-method has been applied, both linear and non-linear models.

The H-method has been extended to many areas of applied sciences. For orientation visit the author's personal website.