# *1 Path regression methods*

The methods presented in previous sections have been extended to structured multi-block models. In the following we shall only show the 'unit setups' and how the optimisation tasks are carried out.

**Optimisation in linear regression**. In section 6 the optimisation task is explained closer. It is repeated here to show that the same principle is used in other optimisation tasks. The task is here, see Figure 1, to find a weight vector **w** such that the **Y**-loading vector **q** is as large as possible, max $|\mathbf{q}|^2 = \max |\mathbf{Y}^T\mathbf{t}|^2 = \max |\mathbf{Y}^T\mathbf{X}\mathbf{w}|^2$, subject to $|\mathbf{w}|=1$. The solution is the eigen vector associated with the largest eigen value of

$$\mathbf{X}^T\mathbf{Y}\mathbf{Y}^T\mathbf{X}\,\mathbf{w} = \lambda\,\mathbf{w}.$$

The matrices **X** and **Y** need to be scaled column-wise before analysis. The regression coefficients are computed as $\mathbf{B}=\mathbf{X}^+\mathbf{Y}$. Thus, the analysis can be viewed as building up $\mathbf{X}^+$ as long as there is support for $\mathbf{X}^+$ in data.

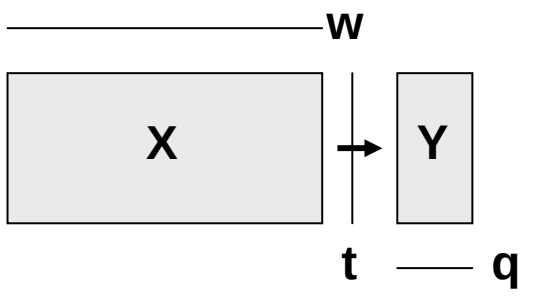


**Figure 1**. Schematic illustration of vectors involved in optimisation.

**Multi-block regression**. Here the **X**-data has been split up into sub-blocks, $\mathbf{X}=(\mathbf{X}_1,\mathbf{X}_2, \ldots, \mathbf{X}_L)$. The response data has also been split into sub-blocks, $\mathbf{Y}=(\mathbf{Y}_1,\ldots,\mathbf{Y}_O)$. In Figure 2 there are two **X**-blocks and two **Y**-blocks. The task is to find weight vectors $\mathbf{w}_1$ for $\mathbf{X}_1$ and $\mathbf{w}_2$ for $\mathbf{X}_2$ such that the resulting **Y**-loadings are as large as possible,

$$\text{maximise } |\mathbf{q}_{11}+\mathbf{q}_{21}+\mathbf{q}_{12}+\mathbf{q}_{22}|^2,$$

subject to $|\mathbf{w}_1|=|\mathbf{w}_2|=1$. Here $\mathbf{q}_{ij}=\mathbf{Y}_j^T\mathbf{X}_i\mathbf{w}_i$. Initially the data matrices **X** and **Y** are scaled column-wise. The optimisation task leads to the following set of equations,

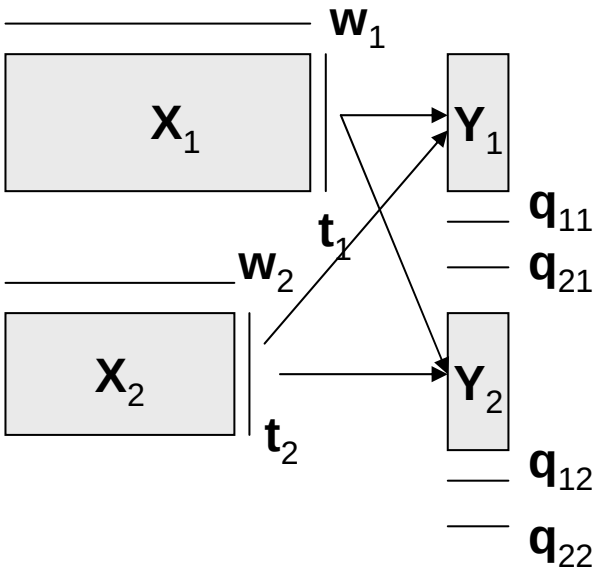


**Figure 2**. Schematic illustration of vectors in optimisation in multi-block regression.

$$\mathbf{X}_1^T(\delta_{11}\mathbf{Y}_1\mathbf{Y}_1^T+\delta_{12}\mathbf{Y}_2\mathbf{Y}_2^T)\mathbf{X}_1\mathbf{w}_1 + \mathbf{X}_1^T(\delta_{11}\mathbf{Y}_1\mathbf{Y}_1^T+\delta_{12}\mathbf{Y}_2\mathbf{Y}_2^T)\mathbf{X}_2\mathbf{w}_2 = \lambda_1\,\mathbf{w}_1$$

$$\mathbf{X}_2^T(\delta_{21}\mathbf{Y}_1\mathbf{Y}_1^T+\delta_{22}\mathbf{Y}_2\mathbf{Y}_2^T)\mathbf{X}_1\mathbf{w}_1 + \mathbf{X}_2^T(\delta_{21}\mathbf{Y}_1\mathbf{Y}_1^T+\delta_{22}\mathbf{Y}_2\mathbf{Y}_2^T)\mathbf{X}_2\mathbf{w}_2 = \lambda_2\,\mathbf{w}_2$$

Initially, $\delta_{11}=\delta_{12}=\delta_{21}=\delta_{22}=1$. The weight vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ are found, when all $\delta$'s are 1. Then, for each $\mathbf{Y}_i$, the score vectors are ranked according to importance in describing $\mathbf{Y}_i$. E.g., suppose that the score vectors, $\mathbf{t}_1$ and $\mathbf{t}_2$, are ranked according to how well they describe $\mathbf{Y}_1$. Let $\mathbf{t}_2$ be the most important. Then $\mathbf{Y}_1$ is adjusted for $\mathbf{t}_2$. If $\mathbf{t}_1$ is not significant in describing the reduced $\mathbf{Y}_1$, it is

dropped and $\delta_{11}$ is sat to zero, $\delta_{11}=0$. Similarly, the modelling of $\mathbf{Y}_2$ is evaluated. On the basis of this evaluation the $\delta$'s are revised, and a new set of weight vectors are found. This is one step of the analysis. When all appropriate vectors have been found, both **X**'s and **Y**'s are adjusted for what has been found and the analysis starts over again. If no significant score vectors are found, the analysis stops. The analysis can be viewed as building up four sets of $\mathbf{X}^+$, which is used in the regression coefficients, $\mathbf{B}=\mathbf{X}^+\mathbf{Y}$. The estimated equations are $\hat{\mathbf{Y}}_1=\mathbf{X}_1\mathbf{B}_{11}+\mathbf{X}_2\mathbf{B}_{21}$ and $\hat{\mathbf{Y}}_2=\mathbf{X}_1\mathbf{B}_{12}+\mathbf{X}_2\mathbf{B}_{22}$, with $\mathbf{B}_{ij}=\mathbf{X}_{ij}^+\mathbf{Y}_j$. For a presentation of Multi-block methods see Ref 8.

**Path modelling**. A path model is a directional set of data block. There can be any number of input or starting data blocks. There can also be several output or ending data blocks. Each data block can lead to one or more data blocks. In Figure 3 is shown three data blocks in a path. The task is to find a weight vector **w** such that the $\mathbf{X}_3$-loading vector $\mathbf{q}_3$ has maximal size. It is computed as $\mathbf{q}_3=\mathbf{X}_3^T\mathbf{t}_2=\mathbf{X}_3^T\mathbf{X}_2\mathbf{q}_2=\mathbf{X}_3^T\mathbf{X}_2\mathbf{X}_2^T\mathbf{t}_1=\mathbf{X}_3^T\mathbf{X}_2\mathbf{X}_2^T\mathbf{X}_1\mathbf{w}$. The solution is the eigen vector associated with the largest eigen value of

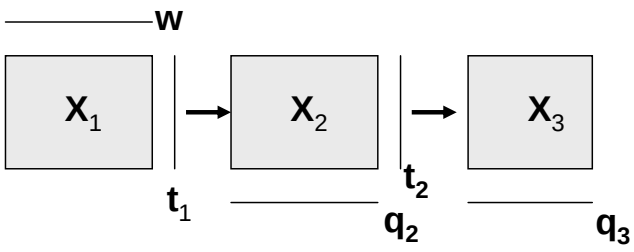


**Figure 3**. Schematic illustration of vectors in optimisation of a path of three data blocks.

$$\mathbf{X}_1^T\mathbf{X}_2\mathbf{X}_2^T\mathbf{X}_3\mathbf{X}_3^T\mathbf{X}_2\mathbf{X}_2^T\mathbf{X}_1\,\mathbf{w}=\lambda\,\mathbf{w}.$$

When **w** has been found, the associated vectors are computed. The analysis can be viewed as building generalised inverses for $\mathbf{X}_1$, $\mathbf{X}_1^+$, and for $\mathbf{X}_2$, $\mathbf{X}_2^+$. The estimated $\mathbf{X}_2$-bloc, $\mathbf{X}_{2,est}$, is computed as $\mathbf{X}_{2,est}=\mathbf{X}_1\mathbf{B}_1$, where $\mathbf{B}_1=\mathbf{X}_1^+\mathbf{X}_2$. Similarly, the estimated $\mathbf{X}_3$-block is $\mathbf{X}_{3,est}=\mathbf{X}_2\mathbf{B}_2$, where $\mathbf{B}_2=\mathbf{X}_2^+\mathbf{X}_3$. When a new $\mathbf{X}_1$-sample, $\mathbf{x}_{10}$, is available, the regression coefficients $\mathbf{B}_1$ are used to estimate an $\mathbf{X}_2$-sample. $\mathbf{B}_2$ is used to take this estimated sample for estimating an $\mathbf{X}_3$-sample.

In path models new samples from the input data blocks are propagated along the path. The path models, thus, provides with estimates of samples for all later data blocks. The uncertainties of the estimated samples in the path are larger the later the block is in the path. There can also be a data block, which is not functioning well in the path. E.g., in Figure 20 it might be that $\mathbf{X}_2$ is not good in explaining $\mathbf{X}_3$, while $\mathbf{X}_1$ may be. When working with path models it may be important to detect, if some data blocks are blocking in this way the modelling task.

In natural sciences it is common to explain phenomenon by a system of differential equations. But the measurement situations often give many measurement values due to cheap measurement equipments like sensors, optical devices etc. When there are many measurement values (variables) at each time point, it may be more efficient to model the situation as one path, $\mathbf{X}_1 \rightarrow \mathbf{X}_2 \rightarrow \ldots \rightarrow \mathbf{X}_L$.

Process data today are basically multi-block data. Process factories carry out numerous types of measurements to control the processes. These processes can often be organised as stages, where data from one stage make a data block. An example of an application of such kind is presented in Ref 9.

The processes can be organised in stages, where there is a path model for each stage. When the processes of a stage are completed, there will be estimates for the results later in the path. Then there can a path model for the next stage that can provide with estimates of processes that succeed the stage. Thus, path modelling is a flexible way of modelling large interconnected systems, where many possible paths can be modelled.

**Two independent sources. Growth curve models.** Denote by **Y** the data matrix in the example in section 22. The rows refer to the judges. Often there is available experimental data that show the 'properties' of the judges. A regression model, **Y**=**XB**+residual, will show how these properties explain the ratings. Columns of **Y** represent the food type. Taste and other features of the food may influence on the ratings. Thus, a regression model, $\mathbf{Y}^T$=**ZB**+residual, may be used to study the influence of these features. Often there are interactions between the properties of the judges and features of the food. It may be better to combine these two models into one, **Y=XBZ**+residual. Element-wise the model is $y_{ij}=\mathbf{x}^i\mathbf{B}\mathbf{z}_j+\varepsilon_{ij}$.

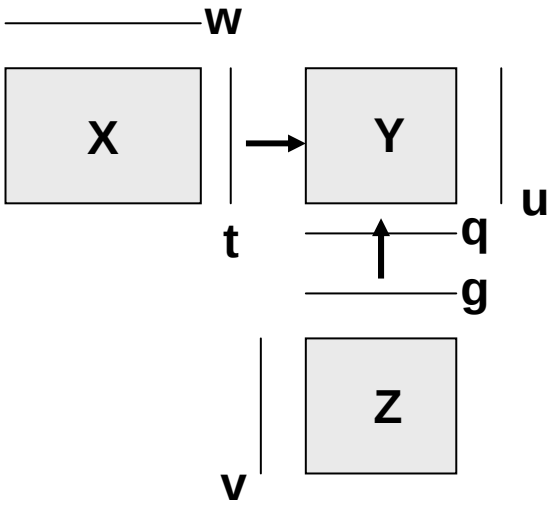


**Figure 4**. Schematic illustration of vectors in optimsation in growth models.

Growth curve models can be formulated in this way, see e.g. Ref 10. The matrix **X** can be the growth curves and experimental data. **Z** can be design conditions or similar ones.

The data can be depicted as shown in Figure 4. We are looking for a weight vector **w** for **X** and a weight vector **v** for **Z**, such that

$$\max |\mathbf{q}|^2 = \max |\mathbf{Y}^T\mathbf{X}\mathbf{w}|^2$$

$$\max |\mathbf{g}|^2 = \max |\mathbf{Y}\mathbf{Z}^T\mathbf{v}|^2$$

The weight vectors **w** and **v** are found by solving the eigen value systems,

$$\mathbf{X}^T\mathbf{Y}\mathbf{Y}^T\mathbf{X}\,\mathbf{w} = \lambda\,\mathbf{w} \qquad \text{and} \qquad \mathbf{Z}\mathbf{Y}^T\mathbf{Y}\mathbf{Z}^T\,\mathbf{v} = \mu\,\mathbf{v}.$$

When the weight vectors have been computed, the resulting vectors are computed, **t**=**Xw**, **q**=$\mathbf{Y}^T$**t**, **g**=$\mathbf{Z}^T$**v** and **u**=**Yg**. Then the matrices are adjusted

$$\mathbf{X} \leftarrow \mathbf{X} - \mathrm{d}\,\mathbf{t}\,\mathbf{p}^T, \qquad \text{with } \mathbf{p}=\mathbf{X}^T\mathbf{t},\ \mathrm{d}=1/(\mathbf{t}^T\mathbf{t})$$

$$\mathbf{Z} \leftarrow \mathbf{Z} - \mathrm{e}\,\mathbf{b}\,\mathbf{g}^T, \qquad \text{with } \mathbf{b}=\mathbf{Z}\mathbf{g},\ \mathrm{e}=1/(\mathbf{g}^T\mathbf{g})$$

It may happen that at this step **X** or **Z** do not contribute to the modelling of **Y**. It may also happen that there are conflicting results from **X** and **Z**, which gives small values of ($\mathbf{t}^T\mathbf{Y}\mathbf{g}$). The practical issues in the computations are not considered closer here. The generalised inverses, $\mathbf{X}^+$ and $\mathbf{Z}^+$, are computed as described above. The solution $\mathbf{B}=\mathbf{X}^+\mathbf{Y}\mathbf{Z}^+$ is computed using these inverses. Note that there can be different number of terms obtained for **X** than for **Z**. The dimension used for **X** may be say 5, while for **Z** only 2. Studying the score and loading vectors for **X** and **Z** often give good insight into the results obtained for **Y**.

The models of this type are traditionally analysed by maximum likelihood methods, Ref 10. They typically assume full rank models. But data is usually not of full rank except the model is small, only few variables and samples. The maximum likelihood methods break down, when there is reduced rank in data, while present method gives reliable results.