Master thesis proposal: Forecasting the number of pupils and students in the Netherlands

Lisa Wobbes*, Kees Vuik†, and Sahar Haj Kasem‡

July 2020

Keywords: constrained optimization, conjugate gradient, convergence

Description

The Dutch Ministry of Education, Culture, and Science[1] creates annual forecasts of the number of pupils and students within the Dutch educational system [1]. Within the ministry these forecasts form the main reference point for other forecasts and are widely used for budgetary purposes.

The current mathematical model behind the forecasts relies on the Markov assumption but can be extended so that major trends from previous years are taken into account. To ensure stability of the results with regard to the last prediction the model includes a constrained optimization problem that can be stated as follows:

$$\min_{(X,\hat{\mathbf{x}})} \sum_{ijk} \omega_{ijk} \left(x_{ijk} - a_{ijk}\right)^2 + \sum_{q} \hat{\omega}_q (\hat{x}_q - b_q)^2,$$

$$\text{subject to } \sum_j x_{ijk} = \begin{cases} h_{ik} & \text{if } i \neq i_d, \\ \sum_l x_{li_dk} & \text{otherwise} \end{cases} \qquad \forall i,j,k,$$

$$x_{ijk} \geq 0 \qquad \forall i,j,k.$$

*lisa.wobbes@duo.nl
†c.vuik@tudelft.nl
‡s.hajkasem@minocw.nl

[1] Dutch: Ministerie van Onderwijs, Cultuur en Wetenschap (OCW)

Here, $X = [x_{ijk}]$ is a so-called education tensor for year t that contains the number of pupils and students grouped by the educational level i they were at in year $t-1$, the educational level j they are at in year t, and their age k in year t, $\hat{\mathbf{x}} = [\hat{x}_q]$ is a vector of control variables, $[a_{ijk}]$ is the education tensor for year $t-1$, b_q is the desired value of the control variable x_q, h_{ik} is the total number of students enrolled in a particular educational level, ω_{ijk} and $\hat{\omega}_q$ represent the weights, and i_d represents an obtained diploma of educational level i. The education tensor describes approximately 10^3 educational levels and 65 age categories for each forecast year. Since many combinations of these levels and categories are meaningless, the total number of non-zero entries is typically below $1.5 \cdot 10^5$.

The solution to the above optimization problem is found in a number of stages, all of which solve a least squares problem. Due to the sparsity and symmetry of the considered system the Conjugate Gradient (CG) method [2] with diagonal preconditioning is used to find the solution of the least squares problem. Although the CG algorithm generally converges within several iterations, occasionally a lack of convergence is observed. For instance, problems occur when the desired values of the control variables fall below a certain threshold.

The goal of this project is to thoroughly investigate and possibly improve the performance of the CG algorithm within the forecast methodology. The new insights could significantly contribute to further development of the mathematical model behind the forecasts. One such development is related to the integration of international students into the predictions. Although the current model takes into account the demographic changes, it is unable to distinguish between Dutch and international students enrolled in higher education. Since the internationalization of higher education is increasing and the forecasts are used to estimate the annual budget for colleges and universities, this shortcoming of the forecast model has become a major issue.

Requirements

- Since the documentation related to the forecasts is only available in Dutch, a good knowledge of Dutch language is required for this project.
- The model is implemented in C#. Therefore, the applicant should also posses a basic understanding of compiled programming languages.

If you are interested in applying your knowledge of numerical methods to this socially relevant research, then please send us an email.

References

[1] Referentieraming OCW 2019. https://www.rijksoverheid.nl/documenten/rapporten/2019/09/17/referentieraming-ocw-2019, Accessed: 06-07-2020.

[2] Yousef Saad. *Iterative methods for sparse linear systems*. SIAM, 2003.