AS-74.4191 Multivariate Regression Methods
Examination December 18, 2012 / OH, VS

You can answer the following questions in Finnish, English or Swedish.

1. Explain briefly the basic idea of the following methods

   a) Principal component analysis (PCA) (2 p.)

   b) Singular spectrum analysis (SSA) (2 p.)

   c) Orthogonal signal correction (OSC) (2 p.)

2. The covariance matrix R_{xx} of the predictor data (X) is

   a) $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, b) $\begin{pmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, c) $\begin{pmatrix} 2 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix}$.

   Describe the properties of the predictor data in each case a)-c). Assuming that a well-behaving response data set Y is also given, what will happen and why if the data (X, Y) is used for calculating a least squares regression model (MLR) in each case a)-c)? (6 p)

3. Suggest a suitable modeling method for the following cases. Justify your answers briefly:

	System	Data variables	Number of samples	Objective of the modeling
a)	Industrial process	20 different measurements (temperatures, pressures, etc.)	1000	To detect if some of the measured variables are redundant and can be omitted
b)	System of 3 reactor tanks	Input: feed flow speed and temperature Output: levels and temperatures in the tanks	a time series of 2000 samples	To create a dynamic simulator of the system
c)	Market research	Answers to 50 questions in a consumer survey, in the scale of 1-5.	200	To identify and characterize different consumer groups (e.g. poor students, rich pensioners)

   (2 p each)

4.   A linear dynamic system is modeled using a state space model of order 10. The controllability and observability gramians of the model are P_C and P_O, respectively. The eigenvalues of the matrix product $P_C P_O$ are shown in Figure 1.
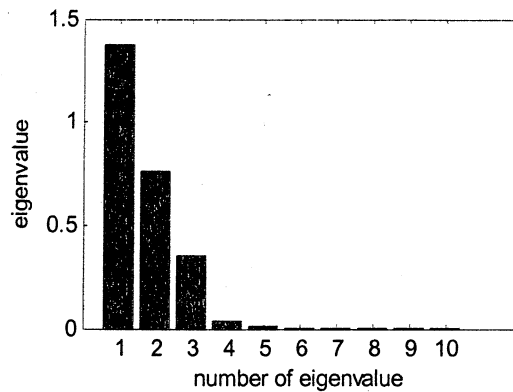

Figure 1.

a) Explain the idea of balanced realization. (4 p)

b) How could you improve the system model? (2 p)

5.   When dealing with multivariable data, *collinearity* is often an issue. What does collinearity mean, and what kind of problems can it cause? How can collinear data be modeled? (6 p)