Accepted Manuscript

Optimal projection of observations in a Bayesian setting

Loïc Giraldi, Olivier P. Le Maître, Ibrahim Hoteit, Omar M. Knio

PII: S0167-9473(18)30050-1
DOI: https://doi.org/10.1016/j.csda.2018.03.002
Reference: COMSTA 6574

To appear in: *Computational Statistics and Data Analysis*

Received date : 5 October 2017
Revised date : 11 February 2018
Accepted date : 3 March 2018

Please cite this article as: Giraldi L., Le Maître O.P., Hoteit I., Knio O.M., Optimal projection of observations in a Bayesian setting. *Computational Statistics and Data Analysis* (2018), https://doi.org/10.1016/j.csda.2018.03.002

Optimal projection of observations in a Bayesian setting

Loïc Giraldi[a,*], Olivier P. Le Maître[b], Ibrahim Hoteit[c], Omar M. Knio[a]

[a] *Division of Computer, Electrical and Mathematical Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia*
[b] *LIMSI, CNRS, Université Paris-Saclay, France*
[c] *Division of Physical Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia*

Abstract

Optimal dimensionality reduction methods are proposed for the Bayesian inference of a Gaussian linear model with additive noise in presence of overabundant data. Three different optimal projections of the observations are proposed based on information theory: the projection that minimizes the Kullback-Leibler divergence between the posterior distributions of the original and the projected models, the one that minimizes the expected Kullback-Leibler divergence between the same distributions, and the one that maximizes the mutual information between the parameter of interest and the projected observations. The first two optimization problems are formulated as the determination of an optimal subspace and therefore the solution is computed using Riemannian optimization algorithms on the Grassmann manifold. Regarding the maximization of the mutual information, it is shown that there exists an optimal subspace that minimizes the entropy of the posterior distribution of the reduced model; a basis of the subspace can be computed as the solution to a generalized eigenvalue problem; an a priori error estimate on the mutual information is available for this particular solution; and that the dimensionality of the subspace to exactly conserve the mutual information between the input and the output of the models is less than the number of parameters to be inferred. Numerical applications to linear and nonlinear

*Correspondence to: 4700 King Abdullah University of Science and Technology (KAUST), Thuwal, 23955–6900, Kingdom of Saudi Arabia. Tel.: +966 12 808 0214.
Email address: loic.giraldi@kaust.edu.sa (Loïc Giraldi)