STATISTICAL BLIND SOURCE SEPARATION

AXEL MUNK
Institute for Mathematical Stochastics, Georg-August-University and Max Planck Institute for Biophysical Chemistry, Göttingen, Germany
e-mail: munk@math.uni-goettingen.de

MERLE BEHR
Institute for Mathematical Stochastics, Georg-August-University
e-mail: merle.behr@stud.uni-goettingen.de

We provide a new methodology for statistical recovery of single linear mixtures of signals (sources) with unknown mixing weights in a multiscale fashion. In general, this is an impossible task as sources and weights cannot be separated. However, under the assumption that the signals attain only a few distinct values, we will show that this becomes doable in large generality. Such an assumption is justified in many applications, e.g. in digital communications, with mixtures of multi-level PAM signals or in cancer genetics, with copy number variation as a mixture of cancer clones.

We discuss identifiability [1] and show exact recovery of such signals within a neighborhood of the mixture. Based on this we provide the SLAM [2] (Separates Linear Alphabet Mixtures) estimators for the mixing weights and sources in a statistical setting. For Gaussian error, we provide uniform confidence sets and optimal rates (up to log-factors) for all quantities. SLAM is efficiently computed as a nonconvex optimization problem by a dynamic program tailored to the finite alphabet assumption. Its complexity is addressed and its performance is investigated in a simulation study.

Our methodology is applied to assign copy-number aberrations from genetic sequencing data to different clones and to estimate their proportions.

Finally, we discuss some extensions to linear models, such as statistical seriation [3].

Parts of this is joint work with Chris Holmes (Oxford).

Acknowledgement Support of DFG RTN 2088 and CRC 803, Z2 is gratefully acknowledged.

References