Potsdam Data Assimilation Days 2025

University of Potsdam, Campus Griebnitzsee, Haus 6, Hörsaal 01

Tuesday, 23rd September

14:00-15:00 Daniela Calvetti (Case Western Reserve University, USA)

Kalman Lecture: Dictionary learning at the crossroads of inverse problems and data science

In the current era of big data, a lot of effort is spent in organizing and querying data sets. These efforts are particularly central for inverse problems, in which the goal is to estimate quantities depending indirectly on the data. Previously collected or simulated data sets provide insight into how typical data should look like, and how much variability in the data can be expected as the unknown of interest varies. In this context, it is natural to think of data sets as entries in a dictionary. Intrinsic knowledge about the data combined with data science methods can be used to partition the dictionary into subdictionaries. Matching previously unseen data to labeled dictionary entries can provide an interpretation of the data. Dictionary matching/learning methods provide a flexible and versatile framework for traditional classification problems as well as for solving inverse problems where traditional techniques fail either because the forward model is complex, ill-defined, or difficult to parametrize, or the data are insufficient for standard methods. To increase computational efficiency and accuracy dictionary matching can be preceded by a dictionary learning step, yielding a reduced dictionary.

Sparsity of the solutions can greatly speed up the calculation and facilitate the interpretation of the dictionary matching process. Hierarchical Bayesian methods leading to very effective computations with sparsity-promoting prior models are very naturally suited for dictionary learning applications. In this talk, we review the sparsity promoting methods for solving inverse problems as a dictionary matching problem, and discuss how Bayesian modeling error methods and matrix factorization techniques can be used to learn compressed dictionaries.

Wednesday, 24th September

09:30-10:15 Harrie-Jan Hendricks-Franssen (Forschungszentrum Jülich)

Data assimilation with terrestrial system models

Terrestrial system models simulate fluxes of water, energy, carbon and nutrients for the terrestrial part of Earth including atmosphere, land surface (e.g., vegetation, urban areas, surface water bodies) and subsurface (soil and aquifers). They also calculate states like temperature, soil moisture content, air humidity and size of carbon pools. Terrestrial system models typically have many unknowns related to many grid cells and many unknown variables, including model parameters. The uncertainty of terrestrial system model simulations can be reduced by assimilating remote sensing and in situ measurements. These measurements are often indirectly related to the model variables, and a measurement operator model is needed to link the observed and simulated value. Uncertainty for land surface measurements is considerable, in situ measurements are more precise but very sparse, even more for the subsurface.

The high dimensionality of the inverse problem and the elevated compute times condition the data assimilation methodology to be used. Therefore Ensemble Kalman Filters have long been the method of choice. We coupled the parallel data assimilation framework (PDAF) to the terrestrial systems model platform (including ICON, eCLM and ParFlow for atmosphere, land and subsurface) and developed capacity to assimilate in situ and remotely sensed soil moisture, in situ groundwater levels, remotely sensed total water storage, in situ evapotranspiration, in situ snow depth, in situ leaf area index and remotely sensed land surface temperature. The implemented methodology also allows to estimate unknown soil and vegetation parameters. Aim is to improve characterization of states and fluxes of the terrestrial system with the multivariate multiscale data assimilation system. Applications range from the point scale to the continental scale (Europe and Africa) and stress the importance of assimilating different data types. We want to increase further the types of observations to be assimilated, and speed up simulations using machine learning derived surrogate models.

Wednesday, 24th September

10:15-11:00 Elisabeth Ullmann (Technical University of Munich)

Uncertainty quantification analysis of bifurcations of the Allen-Cahn equations with random coefficients

In this work we consider the Allen-Cahn equation, a prototypical model problem in nonlinear dynamics that exhibits bifurcations corresponding to variations of a deterministic bifurcation parameter. Going beyond the state-of-the-art, we introduce a random coefficient in the linear reaction part of the equation, thereby accounting for random, spatially-heterogeneous effects. Importantly, we assume a spatially constant, deterministic mean value of the random coefficient. We show that this mean value is in fact a bifurcation parameter in the Allen-Cahn equation with random coefficients. Moreover, we show that the bifurcation points and bifurcation curves become random objects. We consider two distinct modelling situations: (i) for a spatially homogeneous coefficient we derive analytical expressions for the distribution of the bifurcation points and show that the bifurcation curves are random shifts of a fixed reference curve; (ii) for a spatially heterogeneous coefficient we employ a generalized polynomial chaos expansion to approximate the statistical properties of the random bifurcation points and bifurcation curves. Our exposition addresses both, dynamical systems and uncertainty quantification, highlighting how analytical and numerical tools from both areas can be combined efficiently for the challenging uncertainty quantification analysis of bifurcations in random differential equations. This is joint work with Chiara Piazzola (TUM) and Christian Kuehn (TUM).

11:30-12:15 Daniel Rudolf (University of Passau)

Slice sampling

For approximate sampling of a partially known distribution, the slice sampling methodology provides a machinery for the design and simulation of a Markov chain with desirable properties. In the machine learning community it is a frequently used approach, which appears not only there as a standard sampling tool. In particular, the elliptical slice sampler attracted considerable attention in the last decade, as a tuning-free and dimension-robust algorithm. However, from a theoretical point of view it is not well understood. In general, the theoretical results, which promise qualitatively robust and “good” convergence properties of classical slice sampling methods, are mostly not applicable because of idealized implementation assumptions. Motivated by that, the aim of the talk is

1. to provide an introduction into the slice sampling methodology;
2. to discuss different interpretations;
3. to talk about convergence results; as well as
4. to point to open questions.

Wednesday, 24th September

12:15-12:40 Maia Tienstra (University of Potsdam)

Early stopping for prior selection in Bayesian inverse problems

We present an extension of early stopping as a prior selection method for Bayesian inverse problems. In the nonparametric setting the prior plays an important role. If the prior is tuned correctly one can derive a posterior which is optimal, and has conservative coverage. We will show how early stopping under certain conditions provides a data-driven way to tune the prior distribution. We will consider a variety of examples starting with the linear case and then the nonlinear setting.

12:40-13:05 Josie König (University of Potsdam)

Balancing-based model order reduction for data assimilation

Solving data assimilation problems typically requires expensive forward model simulations to compute the expected output for multiple initial conditions. Since this quickly becomes infeasible in high dimensions, reduced models can play a key role in rendering such computations tractable. In particular, model reduction methods from control (systems) theory are well established for obtaining efficient reduced models while preserving the map from control input to observed output. In this talk, we present how balancing-based model order reduction can be adapted for use in data assimilation. This includes a new interpretation of the Bayesian inverse problem — essential in data assimilation — as a control problem and its reduction.