Gaussian Processes for Bayesian Deconvolution and Source Separation:

Applications to XFEL Spectroscopy

SLAC NATIONAL ACCELERATOR LABORATORY

Anton Loukianov[1+], Franklin Fuller[1*]

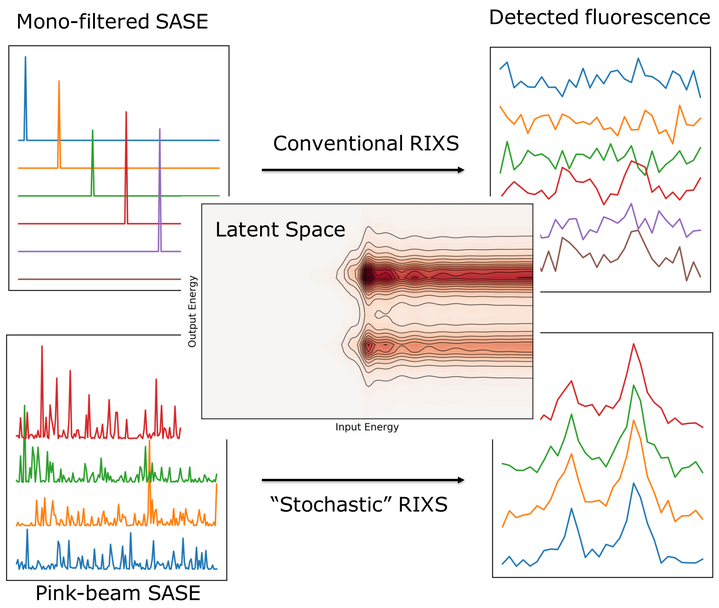
[1]Linac Coherent Light Source, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA. [+]aloukian@slac.stanford.edu [*]fdfuller@slac.stanford.edu

Introduction

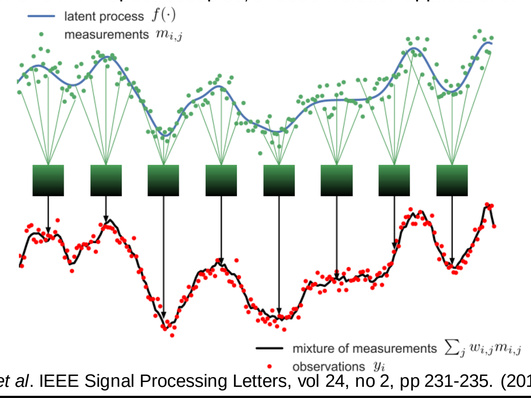
In many experiments, we can only measure the state of the system indirectly. The signal of interest is latent (hidden) behind a transformation and obscured by noise. To remove the transformation and obtain the hidden signal, we directly model the transformation and noise processes while using a flexible function (Gaussian Process) to represent the desired signal. Thus, unlike many ML applications, where the latent function is often useful only as a means for inference, we obtain a meaningful latent representation. Gaussian Processes (GP) additionally allow us to rapidly and easily characterize the uncertainty of our learned signal. We demonstrate the advantages of GPs over conventional methods for rapidly collecting Resonant Inelastic X-ray Scattering (RIXS) spectra using the chaotic SASE source. Further we show that we can separate a sum of linear and non-linear X-ray signals, which is a critical problem in non-linear X-ray science, where the non-linear signal is measured with a strong linear background.

GPs in Spectroscopy

In linear spectroscopies, the true signal ("latent space") acts as a linear operator that converts a set of input vectors (left) to an output vectors (right). The output vectors are often detected in the presence of noise. The example here depicts the case of RIXS, measured on the top with a monochromatic light source and the bottom with a polychromatic spectrum. At an XFEL the "pink" SASE beam is much more intense than the filtered monochromatic beam, resulting in favorable counting statistics of the output signal.



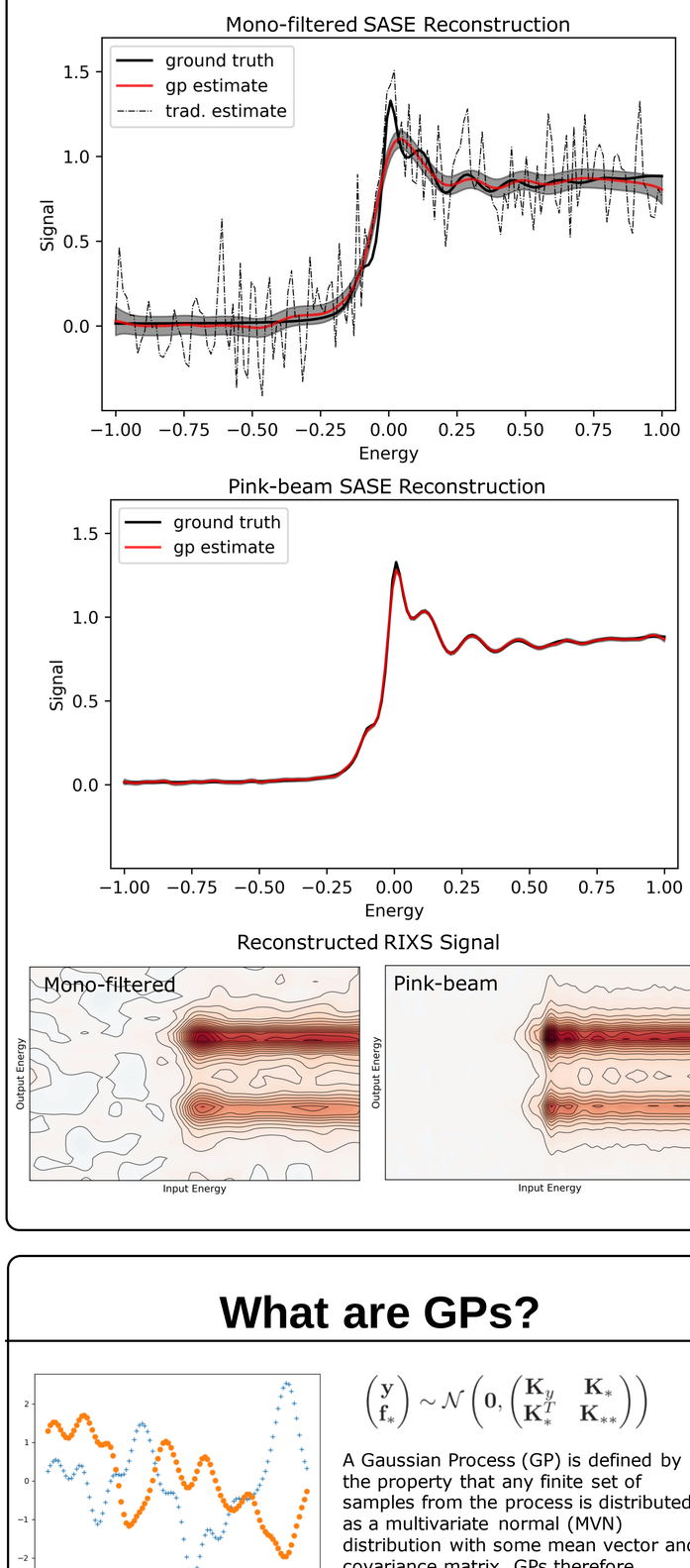
To handle both of these detection modalities in the Gaussian Process framework, a Gaussian Process prior was placed on the latent space (see What are GPs?) and weighed by either the pink SASE matrix or a mono-scanned SASE matrix. GPs are closed under affine transformations; this implies that the weighing by the SASE spectrum produces another GP (the fluorescence) that is then sampled in the presence of noise. Expert knowledge can be put into the covariance kernel to capture non-stationary behavior; alternatively, mild assumptions like twice-differentiability can be used. Because of the affine transformation property, the latent space may be convolved with some function, differentiated, or Fourier-transformed and still retain GP property. This allows the use of GPs in FT spectroscopies, or deconvolution applications.



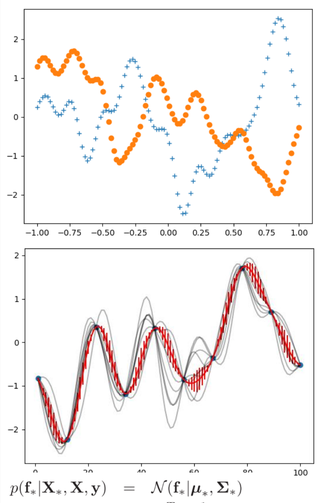
Tobar, F. *et al*. IEEE Signal Processing Letters, vol 24, no 2, pp 231-235. (2017)

Bayesian Deconvolution

We compared the sample-efficiency of the mono-scanned and pink-beam RIXS detection methods on a simulated spectrum detected in the presence of Gaussian noise. For small latent spaces, iid Gaussian noise, and under the assumption of repeated measurements, it is possible to efficiently run exact GP regression on a million SASE shots; the latent space is ~10k points. Shown here is the 16:1 peak SNR case for 100K shots. The hyperparameters of the regression problem were optimized using Adam. In all noise levels, the pink beam method achieves better MSE loss and is more sample-efficient. For larger latent spaces, compressive or inducing point methods must be used. Poisson noise is also possible, but much more expensive to implement. Shown below are slices through the output axis at the most intense peak for both the mono and pink-beam experiments. For the mono case, the traditional estimate involves binning each energy and averaging the values.



What are GPs?



Data fitting cast as conditional probability estimation

A Gaussian Process (GP) is defined by the property that any finite set of samples from the process is distributed as a multivariate normal (MVN) distribution with some mean vector and covariance matrix. GPs therefore generalize MVNs to a continuous function space. Constraints on the basis functions are parameterized by the covariance kernel, which encode prior knowledge about the signal's smoothness and other properties. Top left are two finite dimensional draws from a multi-variate normal with a square exponential (RBF) covariance. Bottom left shows how the basis functions can be constrained (conditioned upon) observations, shown in blue. Analytic manipulation of gaussians feasible and enables rapid uncertainty quantification, among other nice properties.

Variational Free Energy Approximations

Gaussian Processes are non-parametric, i.e. their capacity grows proportionally with the data. The variational free energy approximation to GPs permits us to find a finite rank approximation that no longer grows in size with increased data. The objective function for VFE is shown below.

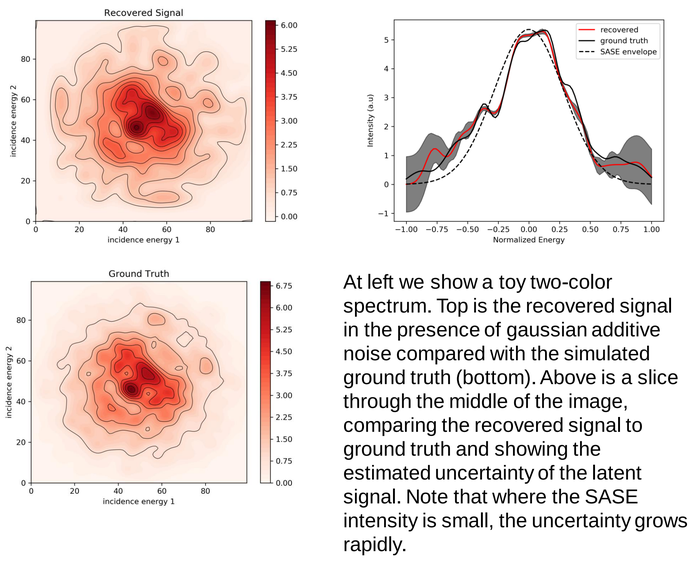
$$\mathcal{F}_{\text{vfe}}[q, \boldsymbol{\theta}] = \sum_n \mathbb{E}_{q(f_n|\boldsymbol{\theta})}\left[-\log p(y_n|f_n)\right] + \text{KL}\left(q(\boldsymbol{u}|\boldsymbol{\theta})||p(\boldsymbol{u}|\boldsymbol{\theta})\right)$$

With the VFE framework, GP methods can handle non-Gaussian likelihoods, scale to millions of data points, allow minibatch training and streaming data.

Separation of Linear from Non-linear Signals

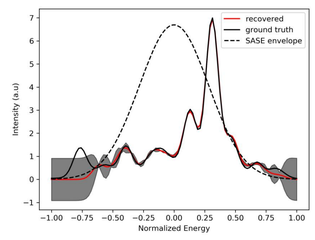
For a single SASE shot (left), the corresponding input to the two color response is shown at the right.

Multi-dimensional X-ray Spectroscopy: We consider here the case of regressing the two color response in the presence of linear signal corruption. Two color signals are those which depend quadratically on the SASE intensity at two energies (colors) of the incidence light. For two-color SASE, all possible color combinations need be considered. The above image on the right is the outer product of the spectrum on the left.



At left we show a toy two-color spectrum. Top is the recovered signal in the presence of gaussian additive noise compared with the simulated ground truth (bottom). Above is a slice through the middle of the image, comparing the recovered signal to ground truth and showing the estimated uncertainty of the latent signal. Note that where the SASE intensity is small, the uncertainty grows rapidly.

As we regress both the linear and non-linear signals jointly, the linear absorptive signal can also be recovered. At the right, the recovered linear signal compared with ground truth is shown. As for the quadratic signal, the uncertainty of the recovered signal grows as the SASE intensity diminishes.



Conclusions

Gaussian processes are an elegant framework for solving latent-variable problems occurring in spectroscopy applications. We have demonstrated that the "stochastic" RIXS problem is more sample-efficient than a mono in Gaussian noise. Then we showed that GPs provide a way to separate superpositions of linear and nonlinear signals, enabling novel nonlinear spectroscopies. In an upcoming beamtime, we will test these methods on experimental data. In the future, we plan on using more complicated neural network kernels with the GP framework to enable transfer learning when studying similar chemical compounds.

Acknowledgments

Franklin Fuller and Anton Loukianov are supported by the W.K.H Panofsky Fellowship and LCLS Hard X-ray Department.

Use of the Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515.