Large Synoptic Survey Telescope

www.lsst.org

Modeling Spatially Varying (De)Convolution Kernels for LSST Difference Imaging

A.C. Becker, N. Silvestri, R. Owen, A. Connolly (U. Wash), R. Lupton (Princeton), T. Axelrod (LSST), LSST Collaboration

A main component of LSST's nightly Image Processing Pipeline is image subtraction. We present advances in the modeling of basis functions for image subtraction convolution kernels, including atomic decomposition and rapid selection of basis functions from a pre-computed dictionary. We also describe methods to model the spatial variation of the kernel, including methods adapted from the Geostatistical community. These techniques are being implemented within the LSST Data Management build system.

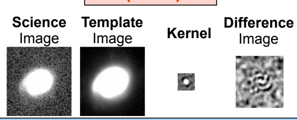
In image subtraction, a **Template** image is subtracted from a nightly **Science** image; all that remains in the resulting **Difference** image are objects that have changed in brightness or position. The **Template** image is generally high signal-to-noise and defect free. To account for the different image qualities, a PSF-matching **Kernel** must be generated. This **Kernel** is typically modeled as a linear combination of basis functions, with the user free to choose the basis set. Ideally this basis set should compactly model the **Kernel**. Given an ensemble of **Kernels** generated from multiple objects across the image (analogous to generating a PSF model from an ensemble of point sources) a spatial model must be built so that the PSF-matching **Kernel** may be applied everywhere in the image.

Convolution vs. Deconvolution

The first consideration is : *which image do you apply the Kernel to?* In general it is good practice to keep the pixels in the **Science** image untouched; (de)convolving with a **Kernel** spreads the impact of bad pixels around, and tends to correlate the noise in the image, which is not desirable. This suggests that we should always apply the **Kernel** to the **Template** image, which is defect-free and has low noise.

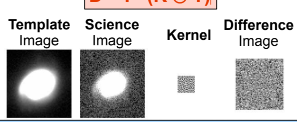
However, a second consideration is : *is the derived Kernel a convolution or deconvolution Kernel?* If the **Kernel** "sharpens" the image, it is a deconvolution **Kernel**; if it "smooths" the image, it is a convolution **Kernel**. We compare and contrast the two operations in the case where the **Template** image has worse seeing than the **Science** image.

$D = (K \otimes I) - T$



In this case, the PSF matching **Kernel** is a convolution **Kernel** applied to the **Science** image. Because the **Science** image has better seeing, the **Kernel** smooths the image. This creates a "compact" **Kernel** with most of its power in the center. However, because the **Science** image is low signal-to-noise, the correlations between the pixels induced by the smoothing become significant, as can be seen from the structure in the resulting difference image.

$D = I - (K \otimes T)$



In this case, the PSF matching **Kernel** is a deconvolution **Kernel** applied to the **Template** image. Because the **Template** image has better seeing, the **Kernel** sharpens the image. This creates a "noisy" **Kernel** with much high frequency power. However, because the **Template** image is high signal-to-noise, deconvolution is feasible and the noise in the difference image appears white.

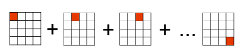
Choice of Basis Set

The next consideration is : *how do you model the Kernel?* Typically, we want to decompose the Kernel using a basis set, such that : $\mathbf{K} = \Sigma\, a_i\, \mathbf{B}_i$. Here "**B**" are the basis vectors, and "a" are the coefficients in front of each. The user is free to choose this basis set. However, an optimal basis will model the **Kernel** *compactly*, meaning you can truncate the approximation using a small number (i) of these functions.

Orthogonal Bases

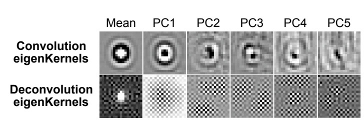
For orthogonal bases, $B_i \bullet B_j = \delta_{ij}$ and the determination of the coefficients "a_i" is trivial : $a_i = K \bullet B_i$. Standard orthogonal bases include :

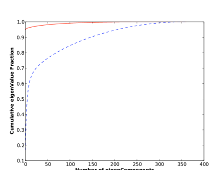
• **Delta functions** : This pixelized, "function free" representation allows for any shape for the **Kernels**, including deconvolution **Kernels**. The downside is that they may be noisy, particularly in the outskirts of the **Kernel**.

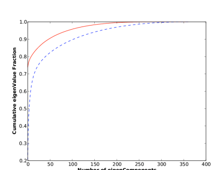


• **Sum of Gaussians** : The commonly used Alard & Lupton (1998) method assumes the **Kernel** may be modeled by the sum of N=3 Gaussians of different widths. This has proven very useful in practice, and made industrial-scale difference imaging feasible. However the widths and number N of the Gaussians are not free to vary in the fit. This restricts the generality of the model. It is also difficult to model **Kernels** where the power is off-center, or noisy **Kernels** that deconvolve.

• **Principal Components** : After you create a set of **Kernels**, you may derive an orthogonal basis set from the data itself. This is an optimal basis in the sense that it *by design* compactly models the input **Kernels**. However, it may not model features that are not present in the input data. This makes it perilous to interpolate between / extrapolate beyond the input data. Example principal components (eigen**Kernels**) are shown below.







Compared to deconvolution **Kernels**, convolution **Kernels** tend to be more self-similar. This means you can approximate a convolution **Kernel** with fewer eigenBases. It also means that you expect the **Kernel** function to vary spatially in a well-behaved (low frequency) manner. This compactness is reflected in the spectrum of eigenValues associated with the eigen**Kernels**. This is demonstrated above, where the solid red line shows the cumulative fraction of eigenValues for convolution **Kernels**, while the dotted blue line shows this for deconvolution **Kernels**. On the left, the convolution **Kernels** are very self-similar; on the right they are less so due to spatial variation of the **Kernel**.

Choice of Spatial Model

The final consideration is, given a set of **Kernels** determined at various points on an image, how do you determine a spatially varying function that describes this **Kernel** at all points in the image? Typically the bases "**B**" are held fixed, and the coefficients "a" are determined as a function of x,y. We are exploring various ways to interpolate in between the **Kernel** constraints. We are exploring application of these functions to the pixels in the **Kernels**, as well as to the coefficients in front of the basis functions.

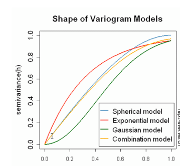
• **Splining** : These piecewise polynomial curves are very fast to evaluate and do not suffer from the Runge phenomena (ringing about the true function; analogous to the Gibbs phenomena in sinusoidal functions) encountered when fitting polynomial functions. Popular variants are cubic splines and B-splines, which use basis functions across the spine.

• **Polynomials** : These functions have historically been used in image subtraction software to model the spatial variation of **Kernels**. However, too-high order polynomials may lead to poor subtraction in the corners of the image, where there are fewer constraints on the function. *Chebyshev* polynomials are less sensitive to the the Runge phenomena; the roots of Chebyshev polynomials are typically used as nodes when splining functions.

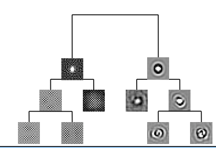
Kriging : This technique is adopted from the Geostatistical community. Kriging is a least squares estimation algorithm that makes use of the Variogram derived from the various constraints on the function. The Variogram is a stochastic model describing the mean squared difference of the **Kernels** as a function of distance (lag). Example Variograms are shown below with different functional forms.

$$V(h) = \Sigma_{i,j}\, |\, K_i - K_j\, |^2$$

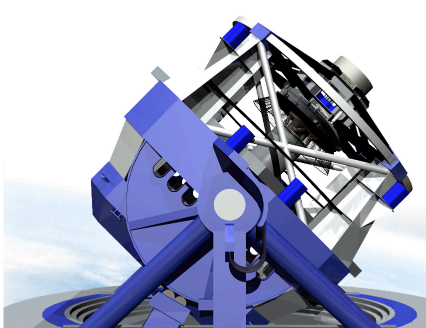
$$h^2 = (x_i - x_j)^2 + (y_i - y_j)^2$$



Non-Orthogonal Bases



Sets of non-orthogonal bases may be put together into an overcomplete dictionary. The challenge here is to quickly find the optimal **Kernel** bases to use for a particular input image. Since the dictionary is overcomplete, the solution will be non-unique. However, metrics such as the minimum L^1 norm of the coefficients "a_i" may be used to define an optimal superposition of basis elements. These methods of "atomic decomposition" are well established in the statistics community. One method to rapidly prune inappropriate bases is to store the dictionary in a tree-based structure using e.g. K-means clustering.



LSST is a public-private partnership. Design and development activity is supported in part by the National Science Foundation. Additional funding comes from private gifts, grants to universities, and in-kind support at Department of Energy laboratories and other LSSTC Institutional Members.