Gaussian Process Modeling of Derivative Curves

Tracy Holsclaw, Bruno Sansó, and Herbert Lee
Department of Applied Mathematics and Statistics
University of California Santa Cruz
Santa Cruz, CA 95064
(tholscla@ams.ucsc.edu)

David Higdon, Katrin Heitmann,Ujjaini Alam, and Salman Habib
Los Alamos National Laboratory
Los Alamos, NM 87545

December 23, 2010

Abstract

Gaussian process (GP) models provide non-parametric methods to fit continuous curves observed with noise. In this paper, we develop a GP based inverse method that allows for the estimation of the derivative of a curve, avoiding direct estimation from the data. In principle, a GP model may be fit to the data directly, then the derivatives obtained by means of differentiation of the correlation function. However, it is known that this approach can be inadequate due to loss of information when differentiating. We present a new method of obtaining the derivative process by viewing this procedure as an inverse problem that does not lose information. We use the properties of a GP to obtain a computationally efficient fit. We illustrate our method with simulated data as well as apply it to an important cosmological application. We include a discussion on model comparison techniques for assessing the quality of the fit of this alternative method.
(LA-UR 10 − 08095)

Keywords: Bayesian statistics; cosmology; dark energy equation of state; stochastic process models.

1 INTRODUCTION

While most statistical curve fitting methods focus on the observed data directly, some aim to fit a more complex function of the data. Here we develop methodology when the object of primary interest is the derivative of a process, or a complex function of a derivative. The motivation for this work is a problem in cosmology where the interest is in understanding the behavior of a function of the second derivative of the data process. In a larger context, this problem can be viewed as an *inverse problem*, where the data we observe are a non-linear transformation of the quantity of interest. Inverse problems are a challenging category of problems, because the likelihoods typically cannot be written down in closed form, and often suffer from issues of identifiability (Higdon et al. 2003; Kaipio and Somersalo 2004). As with most inverse problems, it is better to model the unknown quantity of interest directly, rather than to fit the data and transform the fit to obtain an estimate of the unknown quantity. Thus here we propose modeling the derivative process with a Gaussian process (GP), rather than fitting a curve to the data and attempting to differentiate it. Taking a derivative results in a loss of information, it propagates noise, and it can lead to estimation degeneracies for the curve and its probability bands. Sharing another similarity with the general class of inverse problems, we find that a Bayesian approach allows regularization, which reduces issues with identifiability.

GP models provide a class of flexible non-parametric fits that can be used to analyze a wide range of data sets (Banerjee et al. 2004; Rasmussen and Williams 2006). A GP is a stochastic process $w(z)$ indexed by z, such that for any finite collection of

locations $z_1, ..., z_n$, the vector $(w(z_1), ..., w(z_n))$ follows a multivariate Gaussian distribution. A GP is fully defined by its mean and covariance functions. In this paper, we restrict our attention to a constant (but possibly unknown) mean function, although the methodology is readily extensible to other mean functions. Similarly, for the covariance function, we focus on the widely used power-exponential family, parametrized as $\text{cov}(w(z), w(z')) = \Sigma(z, z') = \kappa^2 \rho^{|z-z'|^\alpha}$, where $\kappa > 0$ is a scaling (variance) parameter, $\rho \in [0, 1)$ is a correlation parameter, and $\alpha \in (0, 2]$ is a smoothness parameter. Again, other covariance functions could be used instead. Thus the models herein are stationary, but there is no inherent need for stationarity with our proposed methods. This paper focuses on processes in one dimension, but the methodology is generalizable to higher dimensions as well.

The main difficulties of the proposed approach involve the integration of the derivative process, so as to be able to evaluate the likelihood of the observed data. Section 2 discusses these issues and our solutions. Section 3 demonstrates our methods on a simple simulated example, and Section 4 applies our methods to an important problem in cosmology. We conclude with a brief discussion.

2 INTEGRATED GAUSSIAN PROCESS MODEL

Consider the problem of estimating the derivative of a curve from noisy observations. Let the observations correspond to the pairs (z_i, x_i), $i = 1, \ldots, n$ where z_i is a univariate predictor and x_i a univariate response. The fundamental assumption is that x_i corresponds to noisy evaluations of an unknown function $f(\cdot)$ at point z_i. We are interested in estimating $f'(z)$ without making parametric assumptions. A possible approach is to model $f(\cdot)$ with a Gaussian process and then take the first derivative of the resulting fit. Thus, letting ϵ_i be observational error, $x_i = f(z_i) + \epsilon_i = \theta + y(z_i) + \epsilon_i$

where θ is the mean, $y(z)$ is a GP with mean function zero and covariance function $\kappa^2 K(\cdot, \cdot; \boldsymbol{\psi})$, depending on the parameter vector $\boldsymbol{\psi}$. Additionally, we assume that all ϵ_i are independently distributed according to a $N(0, \tau^2)$. It is known (Yaglom 1962) that $y(z)$ is mean square differentiable if K is two times differentiable at the origin. Furthermore, the process $w(z) \sim GP(0, \kappa^2 \frac{dK(z,z';\boldsymbol{\psi})}{dzdz'})$ is the first derivative process of $y(z)$. Thus, given a parametric form for K, we use the data (z_i, x_i) to estimate the parameters in K and then use $K'' = \frac{dK(z,z';\boldsymbol{\psi})}{dzdz'}$ to make inferences about f'. Overall, this approach results in a non-parametric fit for the data and for the derivative curve.

However, this method may not be adequate. It is known that if we fit the data process directly and then differentiate, we lose information about the derivative process. When observational error is present, the prediction of the derivative process is degraded (Stein 1999). Additionally, the fact that the covariance function K needs to be two times differentiable results in a practical limitation. Very smooth correlations produce numerically singular covariance matrices. This problem is particularly acute for the Gaussian correlation, which is obtained when $\alpha = 2$ in the exponential correlation family. This is the only example of a correlation with any derivatives in the power exponential family. Singularities are also observed for the Matèrn family with smoothness parameter above three. The traditional remedy to the problem of numerical singularity is to add a nugget or jitter to the covariance, but this, of course, produces a discontinuous correlation function. So, it is conceptually wrong to use it for derivative process estimation.

As an alternative to the direct method, we present an inverse approach. We assume that the derivative process $w(z)$ is a GP. The integrated process $y(z) = \int_0^z w(u)du$ is the one for which we have noisy observations. The covariance function of $y(z)$ can be obtained by double integration of the covariance function of $w(z)$. Thus, suppose that

$w(z) \sim GP(0, \kappa^2 K(z, z'; \boldsymbol{\psi}))$, then

$$y(z) \sim GP\left(0, \kappa^2 \int_0^z \int_0^{z'} k(u, u'; \boldsymbol{\psi}) du du'\right). \tag{1}$$

Using Equation (1), the likelihood for $\theta, \boldsymbol{\psi}, \kappa^2$, and σ^2 based on the sample $\boldsymbol{x} = (x_1, \ldots, x_n)$ and the distribution of $\boldsymbol{y} = (y(z_1), \ldots, y(z_n))$, is

$$p(\boldsymbol{x}|\theta, \sigma^2, \boldsymbol{y}) p(\boldsymbol{y}|\boldsymbol{\psi}, \kappa^2) = N_n(\boldsymbol{x}|\mathbf{1}_n\theta + \boldsymbol{y}, \sigma^2 \boldsymbol{I}_n) N_n(\boldsymbol{y}|\mathbf{0}_n, \kappa^2 \boldsymbol{K}_y)$$

where $N_k(\boldsymbol{a}|\boldsymbol{b}, \boldsymbol{C})$ denotes a k-dimensional normal density, evaluated at $\boldsymbol{a}$, with mean $\boldsymbol{b}$ and covariance matrix $\boldsymbol{C}$. $\mathbf{1}_n$ and $\mathbf{0}_n$ denote, respectively, n-dimensional vectors of 1s and 0s, and $\boldsymbol{I}_n$ is the $n \times n$ identity matrix. $\boldsymbol{K}_y$ is the covariance matrix that results from the correlation function in Equation (1). In other words, we use the derivative process's correlation function to obtain the covariance matrix that corresponds to the observed data.

The likelihood obtained from Equation (1) implies that, for every pair of locations (z_i, z_j) we need to calculate a double integral. Thus, evaluating the likelihood for n large has a computational cost that is prohibitive if $K(\cdot, \cdot; \boldsymbol{\psi})$ cannot be integrated explicitly. An alternative is to consider a set of points $z_1^*, \ldots, z_L^*$ and obtain the joint distribution of $\boldsymbol{y}$ conditional on $\boldsymbol{w}_L = (w(z_1^*), \ldots, w(z_L^*))$. The likelihood is then obtained as the product

$$\begin{aligned} &p(\boldsymbol{x}|\theta, \boldsymbol{y}, \sigma^2) p(\boldsymbol{y}|\boldsymbol{w}_L, \boldsymbol{\psi}, \kappa^2) p(\boldsymbol{w}_L|\boldsymbol{\psi}, \kappa^2) = \\ &\quad N_n(\boldsymbol{x}|\mathbf{1}_n\theta + \boldsymbol{y}, \sigma^2 \boldsymbol{I}_n) N_n(\boldsymbol{y}|\boldsymbol{m}, \boldsymbol{V}) N_n(\boldsymbol{w}_L|\mathbf{0}_n, \kappa^2 \boldsymbol{K}_w) \end{aligned}$$

where $\boldsymbol{m} = E(\boldsymbol{y}|\boldsymbol{w}) = \boldsymbol{K}_{yw}\boldsymbol{K}_w^{-1}\boldsymbol{w}$, $\boldsymbol{V} = \text{var}(\boldsymbol{y}|\boldsymbol{w}) = \boldsymbol{K}_y - \boldsymbol{K}_{yw}\boldsymbol{K}_w^{-1}\boldsymbol{K}_{yw}^T$. $\boldsymbol{K}_w$ is a

matrix with elements $k(z_i^*, z_j^*; \boldsymbol{\psi})$, and $\boldsymbol{K}_{yw}$ is a matrix with elements $\int_0^{z_j} k(z_i^*, u; \boldsymbol{\psi})du$. As $\boldsymbol{K}_y$ is needed for the conditional density of $\boldsymbol{y}$ given $\boldsymbol{w}_L$, the alternative likelihood does not present any computational advantage. Nevertheless, the approximation

$$y(z) \approx y_L(z) = \sum_{l=1}^{L} w(z_l')(z_l - z_{l-1}), \;\; 0 = z_0 < z_1 < \ldots < z_L = z, \;\; z_{l-1} \leq z_l' \leq z_l$$

(Yaglom 1962), implies that the integrated process $y(z)$ can be approximated by a linear combination of L values of $w(z)$. Clearly if $w(z)$ is given, then $y_L(z)$ is known exactly for any finite L, thus $\text{var}(y_L(z)|w) = 0$. So, for a large enough L, we can assume that $\boldsymbol{V} \approx 0$. This justifies the use of the likelihood

$$N_n(\boldsymbol{x}|\mathbf{1}_n\theta + \boldsymbol{m}, \sigma^2\boldsymbol{I}_n)N_L(\boldsymbol{w}_L|\mathbf{0}_n, \kappa^2\boldsymbol{K}_w), \tag{2}$$

which does not require the computation of double integrals.

The use of the approximation in Equation (2) can speed up calculations substantially. It allows for an inverse problem approach that does not degrade the GP fit of the derivative process when measurement error is present. Moreover, numerical simulations show that for L as small as 10 we obtain values of $\text{var}(y_L(z)|w)$ that are negligible. We show the power of this method in the following two examples. The first example consists of simulated data where the true curve is known and we are interested in estimating its first derivative using noisy observations. The second example corresponds to the estimation of the dark energy equation of state from supernovae data. This problem is substantially more complex, as we are interested in a derivative process embedded in a non-linear transformation. To assess our method, we consider a comparison between the fit from our model and some of the parametric models favored in the cosmology literature.

3 SIMULATED EXAMPLE

We generate $n = 100$ equally spaced data points from the curve $f(z) = \ln(1+z)$, adding noise $\epsilon \sim N(0, 0.5^2)$, for $z \in [0, 5]$. Thus, the derivative curve is $f'(z) = 1/(1 + z)$. Following the discussion in the previous section, we start by fitting a GP directly to the data and then differentiate to obtain an estimate of the derivative curve (Neal 1997). As stated in the introduction, we focus on the power exponential family, so, in order to have differentiability, we consider a Gaussian correlation function.

Our second fit corresponds to the inverse problem approach based on Equation (1). In this case we use $\alpha = 1$, so that the double integral for the correlation function for $y(z)$ can be calculated in closed form. Our third fit is obtained using the approximation in Equation (2). We use this approach with both the exponential correlation function ($\alpha = 1$) and the Gaussian correlation function ($\alpha = 1.99$). Unfortunately, setting $\alpha = 2$ produces numerical instabilities in the inverse method that prevent us from operating with the induced covariance matrices, so we set $\alpha = 1.99$. For the Gaussian correlation we compute the integrated correlation needed for the matrix $\boldsymbol{K}_{yw}$ numerically, using Gauss-Chebyshev quadrature, as shown in Appendix A.

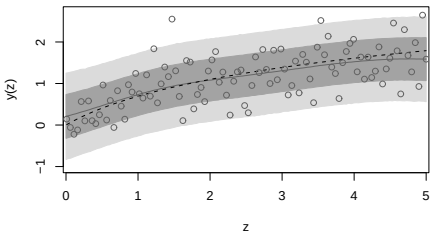
In all methods, we have the same priors and generally avoid non-informative priors because of potential issues with improper posteriors. The parameter ρ is defined on the interval $(0, 1)$ and we want smooth GP realizations that have higher correlation values, so we employ a mildly informative prior on ρ that favors values close to one: $\pi(\rho) \sim Be(6, 1)$. σ^2 is the variance of the observational error. If information on the data is available, we could specify an informative prior, although in this case we can feasibly have a conjugate non-informative prior and that is what we use here: $\pi(\sigma^2) \sim \sigma^{-2}$. κ^2 controls the variance of the GP; we want an informative prior with values away from zero, as this produces a numerical singularity. We use a conjugate

prior: $\pi(\kappa^2) \sim IG(4, 1)$. The last parameter to consider is θ; we assume it to be constant which is equivalent to the conventional method of subtracting off the data mean of the derivative process or a linear trend on the data scale. For this particular problem we have set $\theta = 0.41$. We use an MCMC algorithm based on a combination of Gibbs and Metropolis steps. To ensure good convergence and mixing for the posterior chains we run the chains for 100,000 iterations.

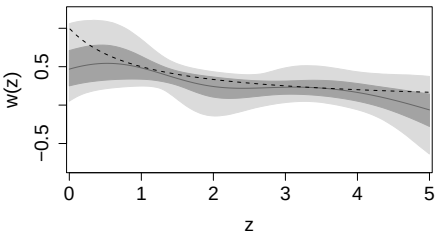
The results for all four fits are presented in Figure 1. Panel 1(a) shows the curve fitted by using a GP directly and 1(b) shows the resulting estimation of the curve derivative. Even though the pointwise probability intervals cover the true values of the derivative for most of the range, it is clear that the estimation is very wiggly. Furthermore, the quality of the fit is poor for values of z in the extremes of the interval. The results obtained for the estimation of the derivative curve using the inverse problem approach are superior to the direct method in all cases. A comparison of the panels in the left column reveals that the method based on the approximate likelihood produces intervals for the data fitting curve that are slightly narrower than the exact method. This is not surprising, as the approximation is based on ignoring some of the variability in the integration of the latent process $w(z)$. On the other hand, the panels on the right column indicate that the derivative curve is estimated quite accurately. This is particularly the case for the fit in panel 1(h), corresponding to the approximated likelihood method with Gaussian correlation function.

4 COSMOLOGICAL APPLICATION

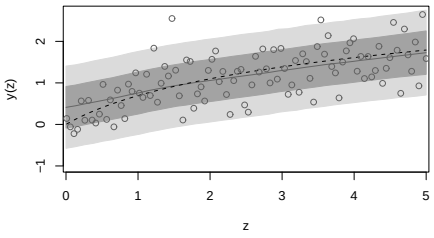
Our motivating example for this work is a problem in cosmology. The dark energy equation of state is at the forefront of the pursuit of knowledge on the cause of the accelerated expansion of the Universe (Riess 1998; Perlmutter et al. 1999). Supernovae



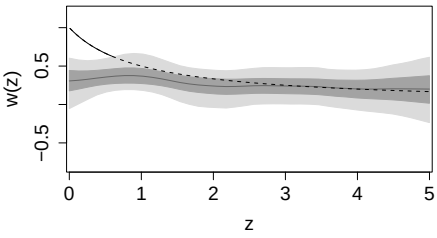
(a) Direct data fit



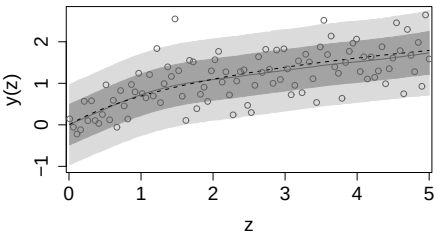
(b) Induced estimation of f'



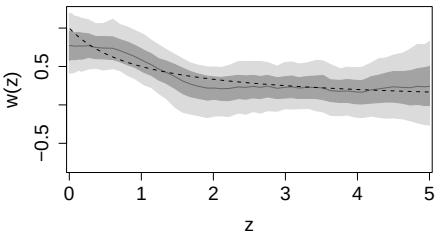
(c) Inverse data fit, $\alpha = 1$



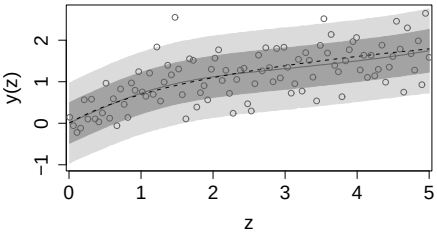
(d) Inverse f' estimation, $\alpha = 1$



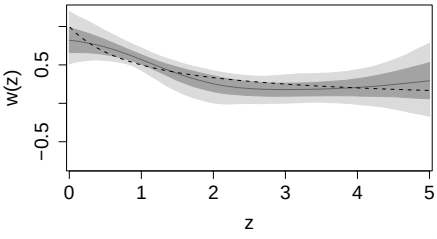
(e) Approx. inv. data fit, $\alpha = 1$



(f) Approx. inv. f' estimation, $\alpha = 1$



(g) Approx. inv. data fit, $\alpha = 1.99$



(h) Approx. inv. f' estimation, $\alpha = 1.99$

Figure 1: Four estimation approaches for $f'(z)$. Left column: simulated data (circles); true curve (dashed line); fitted curve (solid line). Right column: true derivative curve (solid line); estimated derivative (dashed line). 65% and 95% probability interval bands (dark gray and light gray respectively).

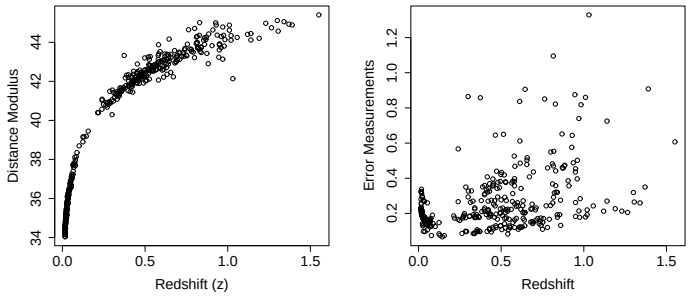


Figure 2: Right panel: distance (μ) versus redshift (z). Left panel: standard errors (τ^2) versus z.

play a key role in studying the equation of state. They are considered standardizable candles, as the peak luminosity during explosion is standard across all supernovae; a feature that allows for good measures of distance. We will denote these as μ. To produce an estimate of μ, measurements of the absolute magnitude as a function of time are taken during the supernova explosion. These measurements are then fitted to produce a light curve for each supernova. From the light curve, μ and the standard measurement error (τ^2) can be extracted. We use the original version of the Spectral Adaptive Lightcurve Template (SALT) fitter (Guy et al. 2005) to obtain μ and τ^2 values. For our analysis, we focus on one of the Constitution supernovae data sets from Hicken (2009); Kowalski (2008), which contains $n = 397$ supernovae measurements. These data are widely available in the literature. A preliminary analysis of them was done in Holsclaw et al. (2010a).

Figure 2 shows the data used in our example; these consist of $n = 397$ (μ, z) pairs. The supernova data are used to infer the the hypothesized dark energy equation of state $w(z)$. This is linked to the supernova distance through a non-linear function $T(w(z))$ involving several unknown cosmological parameters. Genovese et al. (2009)

propose the observation equation

$$\mu(z) = T(w(z)) + \varepsilon(z), \;\; \varepsilon(z) \sim N(0, \tau^2(z)) \; .$$

Here $T(w(z))$ is given as

$$T(w(z)) = 25 + \Delta + 5 \log_{10} \frac{c(1+z)}{H_0} \int_0^z h(s) ds \tag{3}$$

$$h(s) = \left[\Omega_m (1+s)^3 + (1-\Omega_m) (1+s)^3 e^{3 \int_0^s \frac{w(u)}{1+u} du} \right]^{-\frac{1}{2}} \tag{4}$$

indicating that two integrations need to be performed on $w(z)$ before it can be related to the observations. In other words, we are interested in using noisy observations to estimate a function of the second derivative of a curve. This problem has been approached in a direct way, by fitting the observed data and then taking derivatives to obtain $w(z)$, but this led to degeneracies (Sahni and Starobinsky 2006; Saini et al. 2000). Our experience with simulated data is that the direct inference approach produces values for $w(z)$ at the origin that tend to negative infinity, which is not physically plausible.

The difficulties mentioned above are the reason for the popularity in the literature of inverse problem approaches. The method favored by most authors is to set a parametric functional form for $w(z)$ and perform model comparison to choose the best form. Only one and two parameter forms can be estimated, higher order polynomials cause oscillating behavior and introduce bias in the fit, which degrades for high z values (Linder 2007). Our approach is based on assuming a GP model for $w(z)$. A GP approach provides a non-parametric fit to the data that surpasses the fit of most two dimensional models and does not require a functional form of $w(z)$ to be chosen a priori. Other non-parametric methods exist but assume $w(z)$ to be a piecewise con-

stant function, which is not compatible with the physics of the problem (Huterer and Cooray 2005; Hojjati and Pogosian 2010).

4.1 Modeling

We assume that $w(z)$ is a stationary Gaussian process, with a correlation function in the power exponential family. The integrated process $y(s) = \int_0^s w(u)/(1+u)du$ is non-stationary and so Cauchy-Gauss quadrature is a natural choice to obtain the integrals required for the computation of $\boldsymbol{K}_{yw}$. This is illustrated in Appendix 5. We use the approximation in Equation (2) with $L = 50$. This is a conservative number, as numerical simulations indicate that with about $L = 10$ we can obtain accurate numerical integrations. As for the integral in Equation (3), we approximate it using a trapezoidal rule, evaluating $y(s)$ at 150 points for the range $(0, z_n)$. We use a non-regular grid that places more points at lower values of z to ensure high precision in the numerical integrations for small redshift values. We experimented with grids of different sizes and configurations and found that the current choice provides a good compromise between accuracy and computability.

Equation (3) depends on several cosmological parameters. For these we either take fixed values or assume informative prior distributions. The speed of light (c) is taken to be $3 * 10^5$km/s and Hubble's constant is set at $H_0 = 65$. The offset parameter, M, takes into account any uncertainty associated with Hubble's constant; we assume a flat prior $U(M| -10, 10)$ for it. For the matter density parameter of the Universe, Ω_m, we assume the prior $N(\Omega_m|0.27, 0.04^2)$ (Komatsu 2010). This choice prior is based on other sources of data, including cosmic microwave background radiation and baryon acoustic oscillation data.

In this particular example, we do not have the ability to fit an unknown parameter

in the mean function of the GP (Holsclaw et al. 2010b). As in the first example, the mean parameter θ is fixed. It is initially fixed to -1 which is the cosmologist best guess because we cannot subtract a data mean or linear trend due to the complex non-linear relationship in the data. But we do check the value of θ using the posterior mean of the draws of the GP realizations of $w(z)$. We ran the MCMC for 50,000 iterations and all posterior chains had acceptable mixing and acceptance rates. The priors for the covariance parameters σ^2, κ^2, and ρ are the same as in the first example. Some of the correlated parameters were proposed in joint steps. More details of the MCMC algorithm used for this cosmology example are given in Holsclaw et al. (2010b).

For comparison purposes, we fit six different models, including some of the most popular ones in the cosmology literature. Under Model_0 we assume that $w(z) = -1$. This corresponds to the null hypothesis of the addition of a cosmological constant to the Einstein field equations. Model_1 is a one-parameter model with $w(z) = const$, where the constant is estimated rather than constrained to -1. Model_2 is a two-parameter linear model of the form $w(z) = a+bz$ (Cooray and Huterer 1999; Maor et al. 2001; Weller and Albrecht 2001). Model_3 is the two-parameter model favored in the cosmological literature with $w(z) = a+b(\frac{1}{1+z}-1)$ (Linder 2003; Chevallier and Polarski 2001). It is important to notice that more flexible higher order models generally fail to fit the data, due to lack of identifiability. We fit two GP models: one with exponential correlation (Model_{4e} with $\alpha = 1$) and one with Gaussian like correlation (Model_{4g} with $\alpha = 1.99$). We also fit a model with damped Hermite polynomials that approximates a GP with Gaussian correlation function as proposed in Steinberg and Burstztyn (2004) (Model_5 – see Appendix B for details). Representing $w(z)$ on a set of basis functions has been proposed by other authors. This is an appealing alternative, as it can simplify the integrals involved in the computation of $h(z)$. Unfortunately, it has been found that expansions that include more than two terms have an explosive behavior (Genovese

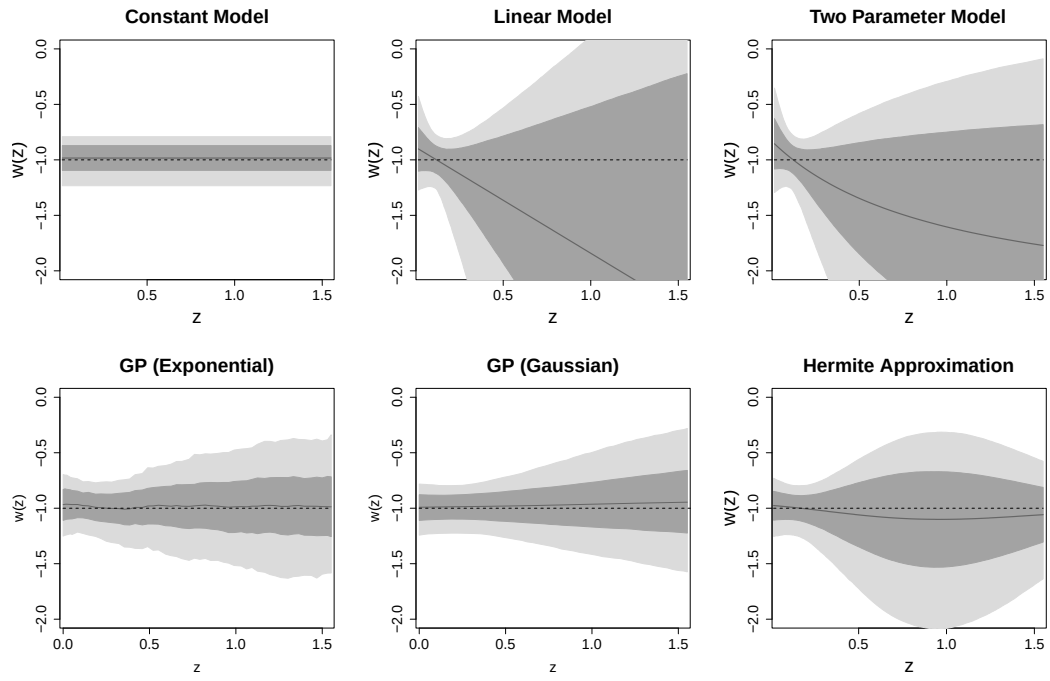


Figure 3: The top row shows the fit of Models$_{1,2,3}$ which are the most common parametric models. The second row shows GP based Models$_{4e,4g,5}$. The dark gray line is the mean fit, the medium grey area is the 68% PI bands and the light gray is the 95% PI bands.

et al. 2009). The expansion proposed in Steinberg and Burstztyn (2004) is appealing, as it includes a damping term that is linked to the correlation length of the GP. The results for all six models are presented in Figure 3. For each model the mean fitted curve is shown with a dark gray continuous line. A dashed line for $w(z) = -1$ is drawn for reference.

We see in Figure 3 that Model$_0$ (the constant function) has the tightest probability intervals (PIs), followed by the GP fits. The GP not only allows for a more flexible fit than any of the parametric forms but it also has tighter PIs than the two-parameter models. The model based on a damped Hermite polynomial expansion was truncated at two terms, as, contrary to our expectations but similar to other expansion approaches, it shows an explosive behavior when more terms are added despite the damping terms. The truncation results in a model that is similar to a two-parameter model. The

results in Figure 3 show that the GP provides a good solution for fitting the dark energy equation of state, as it is flexible without the penalty of increased uncertainty.

4.2 Model Comparison

As there are multiple possible models under consideration, we want to formally quantify how the GP method here compares to other methods. Other studies have relied on the BIC (Bayesian Information Criterion) (Schwarz 1978) to distinguish between models (Genovese et al. 2009). In several trials with simulated data where the true $w(z)$ was known, we found BIC to incorrectly choose the model with the least number of parameters, particularly when the true $w(z)$ had some curvature. Additionally, the typical selection criteria like BIC do not allow for a direct comparison with our non-parametric model, as the GP presented here does not have a well-defined number of parameters to use in the penalty terms for these criteria. Bayes factors were considered and not used for model comparison because of known issues with non-informative or flat priors we are using in our models (March et al. 2010). If the priors are changed, the results of the model comparison also change. This makes the model comparison somewhat incoherent, and therefore this approach was not pursued further. Overall, we found the DIC (Deviance Information Criterion)(Gelman et al. 2004) to be useful here as our decision criterion for model selection. The DIC is defined as $DIC = 2\overline{D}(\psi) - D(\overline{\psi})$, where ψ is the set of unknown parameters in the model, $D(\psi) = -2\log(p(y|\psi))$ and $\overline{D}(\psi)$ is the average of each iterations $D(\psi)$ value in the MCMC and $D(\overline{\psi})$ is $D(\psi)$ evaluated at the average value of ψ from the iterations. We display the DIC value of each model in Table 1, together with the values of the BIC and AIC (Szydlowski and Wlodzimierz 2006).

The DIC provides another useful feature in that we can estimate the number of

effective parameters: $p_D = \overline{D}(\psi) - D(\overline{\psi})$. The p_D values tend to be smaller than the actual number of parameters in the model. This shrinkage happens when some of the parameters are correlated and do not add another full parameter worth of information or when there is an informative prior on the parameter. We know that the parameters in our cosmology example are correlated and have informative priors, so we expect shrinkage in p_D.

Table 1: Model Comparison for Example 2

Model	Form of $w(z)$	BIC	AIC	p	DIC	p_D
Model_0	w(z)=-1	9.997*	-5.939*	4	-8.452*	2.7
Model_1	w(z)=const	15.984	-3.936	5	-7.732	3.0
Model_2	w(z)=a+bz	18.685	-5.407	6	-7.400	2.9
Model_3	$\text{w(z)} = \text{a} + \text{b}(\frac{1}{1+z} - 1)$	18.685	-5.218	6	-7.049	3.4
$\text{Model}_4 e$	$\text{w(z)}\sim \text{GP}(\alpha = 1)$	NA	NA	NA	-7.604	3.2
$\text{Model}_4 g$	$\text{w(z)}\sim \text{GP}(\alpha = 1.99)$	NA	NA	NA	-7.510	3.2
Model_5	w(z)= Hermite Poly	32.356	0.484	8	-7.325	3.2

We marked the favored method for each model choice criterion with a star. The null hypothesis model with $w(z) = -1$ is the one chosen by all the criteria considered. We note that this finding is consistent with the findings in the literature. Interestingly, the DIC shows the $w(z) = const$ next and then the GP model with exponential correlation function. Moreover, the two-parameter models rank at the bottom of the group. It seems that for future exploration of the adequacy of the hypothesis $w(z) \equiv -1$, as the data quantity and quality improve, a GP based model should be the method of choice.

5 CONCLUSIONS

We desired to find a method of non-parametric reconstruction of a derivative process; we have shown that this can be done by viewing the process as an inverse problem. We assume the derivative process is a GP and therefore the resulting integrated process

is also a GP with an unnamed correlation function and models the data well. This method presents advantages over previous methods as it does not lose information by taking derivatives of a fit of the data. It outperforms most parametric models in our example while providing a flexible fit to the underlying derivative process. The GP method constrains the uncertainty about the derivative process of interest. We have also shown that this method is computationally efficient in the way it approaches integration. Furthermore, there is no penalty for having a non-stationary or non-integrable correlation function.

In our motivating example, we find that the currently available data are consistent with a cosmological constant, and thus the flexibility of a GP may be overkill. However, it is still unknown if this result is correct or if it is due to limitations in the available data. Current efforts are underway to collect more, and more accurate data, and our methodology is better-placed than its parametric competitors to help refine the answers when more data arrive.

ACKNOWLEDGEMENTS

We thank the Institute for Scalable Scientific Data Management for supporting this work. Part of this research was supported by the DOE under contract W-7405-ENG-36. HL acknowledges support from NSF grant DMS-0906720. UA, SH, KH, and DH acknowledge support from the LDRD program at Los Alamos National Laboratory.

APPENDIX A: CHEBYSHEV-GAUSS QUADRATURE

Evaluations of the approximate likelihood in Equation (2) require computations of the matrix $\boldsymbol{K}_{yw}$, which involve a one dimensional integral over a finite interval. When the integrals cannot be obtained in closed form, we use Chebyshev-Gauss quadrature to approximate the integration with finite summations. This consists of the approximation $\int_{-1}^{1} \frac{f(z)}{\sqrt{1-z^2}} dz \sim \sum_{i=1}^{I} \gamma_i f(z_i)$, $z_i = \cos\left(\frac{(2i-1)\pi}{2I}\right)$, and $\gamma_i = \frac{\pi}{I}$. In our cosmology example we fixed $I = 100$. A much smaller value of I would be sufficient, but to avoid the propagation of biases, we decided on a relatively large value of I. For the cosmology application, we have the following approximate correlation function:

$$\begin{aligned} K(z, z') &= \int_0^{z'} \frac{\rho^{|x-z|^\alpha}}{(1+x)} dx = \frac{z'}{2} \int_{-1}^{1} \frac{\rho^{|\frac{z'x}{2} + \frac{z'}{2} - z|^\alpha}}{(1 + \frac{z'x}{2} + \frac{z'}{2})} \sqrt{1-x^2} \frac{1}{\sqrt{1-x^2}} dx \\ &\approx \frac{z'}{2} \sum_{i=1}^{L} \gamma_i \frac{\sqrt{1-z_i^2}}{(1 + \frac{z'z_i}{2} + \frac{z'}{2})} \rho^{|\frac{z'z_i}{2} + \frac{z'}{2} - z|^\alpha} . \end{aligned}$$

APPENDIX B: DAMPED HERMITE POLYNOMIAL APPROXIMATION OF A GP

Steinberg and Burstztyn (2004) propose a representation of a GP with Gaussian correlation function based on an infinite linear combination of basis functions obtained from Hermite polynomials. For our cosmology example we consider the process $W(u) = \frac{w(u)}{1+u}$ and assume that it is a Gaussian process with Gaussian correlation function. We assume that θ is the mean of $w(u)$ and, following Steinberg and Burstztyn (2004), we

write

$$W(u) = \gamma(u) + \sum_{s=0}^{\infty} \beta_s J_s(u) \ , \tag{5}$$

where $\gamma(u) = -\theta/(1+u)$ is the mean of $W(u)$, $J_s(u) = H_s^* e^{-\frac{\nu u^2}{2(1+\nu)}}$ are damped polynomials, and $H_s^*(u) = H_s(u/\sqrt(2))/(2^s s!)^{1/2}$ is a physicist Hermite polynomial. The Hermite polynomials, H_s^*, are orthogonal with respect to a standard Normal, as they satisfy $E[H_s^*(W)H_t^*(W)] = \int_{-\infty}^{\infty} e^{-u^2/2} H_s^*(u) H_t^*(u) du = \delta_{s,t}$, where $\delta_{s,t}$ is the Kronecker delta. Clearly, the damped polynomials $J_s(u)$ do not satisfy this property. Thus, the expansion in Equation (5) is not an expansion on orthogonal basis functions.

Taking the prior $\pi(\beta_s) \sim N(0, \sigma^2 \nu^s)$, the expansion in Equation (5) corresponds to a GP with Gaussian correlation with range $\nu/[2(1-\nu^2)]$ and variance $\tau^2(1-\nu^2)^{-1/2}$ where $0 \leq \nu \leq 1$. For practical use, we need to re-scale the functions, so that the damping takes effect within the range of the data. In our case this amounts to multiplying u by a factor of 3. The cosmology example requires $y(s) = \int_0^s W(u)du$. Using Equation (5) we obtain an expansion of $y(s)$ whose coefficients depend on the cumulative standard normal distribution. For more than two terms in the expansion we get poor mixing in the MCMC, wide probability bands, and a poor fitting, especially for high redshift values. This behavior is similar to the one reported in the literature for other basis expansions.

References

Banerjee, S., Carlin, B., and Gelfand, A. E. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, New York: Chapman and Hall.

Chevallier, M. and Polarski, D. (2001), "Accelerating Universes with Scaling Dark Matter," *Int.J.Mod.Phys.D.*, **10**.

Cooray, A. and Huterer, D. (1999), "Gravitational Lensing as a Probe of Quintessence," *Astrophys. J.*, **513**.

Gelman, A., Carlin, B., Stern, H., and Rubin, D. (2004), *Bayesian Data Analysis*, New York: Chapman and Hall.

Genovese, C. R., Freeman, P., Wasserman, L., Nichol, R., and Miller, C. (2009), "Inference for the Dark Energy Equation of State Using Type Ia Supernova Data," *Annals of Applied Statistics*, **3**, 144–178.

Guy, J., Astier, P., Nobili, S., Regnault, N., and Pain, R. (2005), "SALT: A Spectral Adaptive Light Curve Template for Type Ia Supernovae," *Astrophysics*.

Hicken, M. *et al.* (2009), "Improved Dark Energy Constraints from $\sim$ 100 New CfA Supernova Type Ia Light Curves," *Astrophys.J.*, **700**.

Higdon, D., Lee, H., and Holloman, C. (2003), "Markov chain Monte Carlo-Based Approaches for Inference in Computationally Intensive Inverse Problems," in *Bayesian Statistics 7. Proceedings of the Seventh Valencia International Meeting*, eds. Bernardo, J., Bayarri, M., Berger, J., David, A., Heckerman, D., Smith, A., and West, M., Oxford University Press.

Hojjati, A. and Pogosian, L. (2010), "Detecting Features in the Dark Energy Equation of State: a Wavelet Approach," *Journal of Cosmology and Astroparticle Physics*.

Holsclaw, T., Alam, U., Sansó, B., Lee, H., Heitmann, K., Habib, S., and Higdon, D. (2010a), "Nonparametric Dark Energy Reconstruction from Supernova Data," *Physical Review Letters*, **105**.

— (2010b), "Nonparametric Reconstruction of the Dark Energy Equation of State," *Physical Review D*, **82**.

Huterer, D. and Cooray, A. (2005), "Uncorrelated Estimates of Dark Energy Evolution," *Phys. Rev. D*, **71**.

Kaipio, J. and Somersalo, E. (2004), *Statistical and Computational Inverse Problems*, Springer-Verlag.

Komatsu, E. *et al.* (2010), "Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation," *Astrophysical Journal Supplement Series*.

Kowalski, M. *et al.* (2008), "Supernova Cosmology Project, Collaboration," *Astrophys.J.*, **686**.

Linder, E. V. (2003), "Exploring the Expansion History of the Universe," *Phys.Rev.Lett.*, **90**.

— (2007), "The Mirage of w=-1," *Astrophysics*.

Maor, I., Brustein, R., and Steinhardt, P. (2001), "Limitations in Using Luminosity Distance to Determine the Equation of State of the Universe," *Phys.Rev.Lett.*, **87**.

March, M., Starkman, G., Trotta, R., and Vaudrevange, P. (2010), "Should We Doubt the Cosmological Constant?" *Mon.Not.R.Astron.Soc.*

Neal, R. (1997), "Monte Carlo Implementation of Gaussian Process Models for Bayesian Regression and Classification," *Technical Report No. 9702, Department of Statistics, University of Toronto*.

Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R. A., and Nugent, P. *et al.* (1999), "Measurements of Omega and Lambda for 42 High-Redshift Supernovae," *Astrophysics Journal*, **517**, 565–586.

Rasmussen, C. and Williams, K. (2006), *Gaussian Processes for Machine Learning*, MIT Press.

Riess, A. *et al.* (1998), "Supernova Search Team Collaboration," *Astron. J.*, **116**, 1009–1038.

Sahni, V. and Starobinsky, A. (2006), "Reconstructing Dark Energy," *Int.J.Mod.Phys.D*, **15**, 2105.

Saini, T. D., Raychaudhury, S., Sahni, V., and Starobinsky, A. A. (2000), "Reconstructing the Cosmic Equation of State from Supernova Distances," *Phys.Rev.Lett*, **85**, 1162–1165.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–464.

Stein, M. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, New York: Springer.

Steinberg, D. and Burstztyn, D. (2004), "Data Analytic Tools for Understanding Random Field Regression Models," *Technometrics*, **46**.

Szydlowski, M. and Wlodzimierz, G. (2006), "Which Cosmological Model - with Dark Energy or Modified FRW Dynamics?" *Physics Letters B*, **633**, 427–432.

Weller, J. and Albrecht, A. (2001), "Opportunities for Future Supernova Studies of Cosmic Acceleration," *Phys.Rev.Lett.*, **86**.

Yaglom, A. (1962), *Stationary Random Functions*, New York: Dover.