Experimentally testing the dependence of momentum transport on second derivatives using Gaussian process regression

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1088/1741-4326/AA8387">http://dx.doi.org/10.1088/1741-4326/AA8387</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>IOP Publishing</td>
</tr>
<tr>
<td>Version</td>
<td>Author's final manuscript</td>
</tr>
<tr>
<td>Accessed</td>
<td>Tue Feb 12 07:04:52 EST 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/114540">http://hdl.handle.net/1721.1/114540</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Creative Commons Attribution-Noncommercial-Share Alike</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td><a href="http://creativecommons.org/licenses/by-nc-sa/4.0/">http://creativecommons.org/licenses/by-nc-sa/4.0/</a></td>
</tr>
</tbody>
</table>
Experimentally testing the dependence of momentum transport on second derivatives using Gaussian process regression‡

M A Chilenski¹, M J Greenwald¹, A E Hubbard¹, J W Hughes¹, J P Lee¹, Y M Marzouk², J E Rice¹, and A E White¹

¹ Plasma Science and Fusion Center, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139, USA
² Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139, USA

E-mail: markchil@mit.edu

Abstract. It remains an open question to explain the dramatic change in intrinsic rotation induced by slight changes in electron density (White et al. 2013, Phys. Plasmas 20, 056106). One proposed explanation is that momentum transport is sensitive to the second derivatives of the temperature and density profiles (Lee et al. 2015, Plasma Phys. Controlled Fusion 57, 125006), but it is widely considered to be impossible to measure these higher derivatives. In this paper, we show that it is possible to estimate second derivatives of electron density and temperature using a nonparametric regression technique known as Gaussian process regression (GPR). This technique avoids over-constraining the fit by not assuming an explicit functional form for the fitted curve. The uncertainties, obtained rigorously using Markov chain Monte Carlo (MCMC) sampling, are small enough that it is reasonable to explore hypotheses which depend on second derivatives. It is found that the differences in the second derivatives of $n_e$ and $T_e$ between the peaked and hollow rotation cases are rather small, suggesting that changes in the second derivatives are not likely to explain the experimental results.

PACS numbers: 02.50.Cw, 02.50.Ey, 02.50.Fz, 02.50.Tt, 02.60.Ed, 02.60.Jh, 02.70.Rr, 02.70.Uu, 07.05.Kf, 52.55.Fa, 52.25.Fi, 52.70.Kz

‡ Preprint submitted to Nuclear Fusion.
1. Introduction and motivation: rotation reversals and second derivatives

Explaining the dramatic change of intrinsic rotation with slight changes in density is an open question [1–5]: White et al. [5] present a pair of discharges where the turbulent drive terms are essentially the same within error bars, the only noticeable difference being that the electron density is slightly higher in one discharge. Despite this apparent similarity, the toroidal rotation profile is dramatically different between the two discharges: peaked and strongly co-current in the lower density case and hollow and slightly counter-current on axis in the higher density case. The rotation profiles for these shots are shown in figure 1, the electron density profiles are shown in figure 3 and the electron temperature profiles are shown in figure 4.

Ida et al. [6] and the references therein discuss the experimental connection between the ion temperature gradient and the intrinsic toroidal rotation. Beyond this, there are theoretical reasons to believe that the second derivatives of the profiles play a role in momentum transport [7–18]. Barnes gave a very intuitive description of two of the ways this can occur [13]:

(i) The strength of the turbulence depends on the profile gradient. Therefore, the second derivative of the profile corresponds to a gradient in the strength of the turbulence. Consider two ensembles of trapped electrons which start from the same place at the outboard midplane of an up-down symmetric plasma which have opposite parallel velocities. Supposing the toroidal plasma current is in the same direction as the toroidal magnetic field, the electrons for which \( v_\parallel > 0 \) drift radially outwards when undergoing a banana orbit while the electrons for which \( v_\parallel < 0 \) drift radially inwards. In the absence of turbulence (and ignoring other collisions), the electrons all eventually return to the same poloidal location, forming a pair of closed
banana orbits. With turbulence of constant strength (i.e., the second derivatives of the profiles are zero) the losses due to turbulence for both populations of electrons are equal, and no net momentum transport occurs. But if there is stronger turbulent transport on one side than on the other then more of the electrons going one direction will be lost and there will be net transport of momentum. Therefore, the second derivatives of the plasma profiles must be non-zero for there to be net momentum transport.

(ii) When there are spatial gradients in the density and temperature profiles there are diamagnetic corrections to the distribution function of the particles which induce a diamagnetic flow. The strength of this flow depends on the first derivatives of the density and temperature profiles. When there is no initial rotation, this diamagnetic flow cancels with the $E \times B$ flow. But there can be different diffusion coefficients for these two flows. So, if the second derivatives of the profiles are non-zero, there will be a gradient in the diamagnetic flow which can cause it to diffuse at a different rate than the $E \times B$ flow, thereby creating a net diffusion of momentum.

Lee et al. [16] puts this in a particularly clear mathematical form: the intrinsic momentum transport is shown to consist of six components (equation (37) of [16]). Of these, the term arising from the different diffusion coefficients for the diamagnetic and $E \times B$ flows, $\Pi^\Delta_{\chi, \phi, \text{int}} \propto \partial \Omega_{\phi,d} / \partial r$, is found to be a major contributor to the momentum transport in GS2 [19] simulations of the Cyclone Base Case [20, 21]. Because the diamagnetic toroidal rotation frequency $\Omega_{\phi,d}$ is, in general, a function of the temperature and density gradients, this is the term which depends on the second derivatives.

In the past it has been widely believed that it is impossible to test hypotheses depending on second derivatives because of how large the uncertainties are expected to grow with each order of derivative, and the risk that an over-constrained fit could mask the true level of uncertainty. In this paper, we show that recent advances in tokamak profile fitting enable this measurement to be made with reasonable uncertainty estimates. It is found that the differences in the second derivative profiles between the peaked and hollow rotation cases are small. This means that, unless the sensitivity to second derivatives is very high, $\Pi^\Delta_{\chi, \phi, \text{int}}$ does not appear to be the dominant momentum transport term in the experimental cases, in contrast to what was observed for the Cyclone Base Case. This paper focuses on the experimental measurements and their interpretation; future work will continue the comparison to theory using GS2 simulations of real Alcator C-Mod conditions.

The rest of this paper is organized as follows: section 2 gives a brief review of Gaussian process regression, the profile fitting technique which enables the present work. Section 3 presents the actual fits to the $n_e$ and $T_e$ profiles, and section 4 uses these results to quantify the differences in the second derivative profiles in a statistically rigorous manner. Section 5 summarizes the work and presents the conclusions reached. Appendix A shows how to obtain the derivatives of the squared exponential covariance kernel used for the present work.
2. Estimating derivatives and their uncertainties using Gaussian process regression

Recent advances in tokamak profile fitting using Gaussian process regression (GPR) allow derivatives to be computed along with statistically rigorous uncertainty estimates [22, 23]. Gaussian process regression is a very powerful general-purpose Bayesian nonparametric regression technique [24–31], and the past few years have seen a number of applications of GPR in plasma physics [22, 23, 32–38]. Open-source software is available for both general-purpose GPR [39, 40] and plasma profile analysis [41, 42].

Refer to [27] for an introduction to the general mathematical details, and [22, 23] for an introduction with plasma physics-specific applications. In short, GPR describes the data and the fit using a multivariate normal distribution rather than assuming a specific functional form. It is this lack of assumed functional form which allows GPR to be flexible enough to rigorously quantify the uncertainty in the fitted curve. The posterior distribution from GPR includes the values and derivatives of the profile at all points in space. The properties of the probability distribution are determined by the covariance kernel, \( k(x_i, x_j) \), which sets how smooth the profile is: a covariance kernel which decays rapidly with the distance \( |x_i - x_j| \) will lead to profiles with fine spatial structure and many wiggles, while a covariance kernel which does not decay rapidly will lead to smooth profiles with little spatial structure. (Here, \( x_i \) and \( x_j \) are two locations at which predictions and/or observations are to be made.) Because a specific functional form is not assumed, GPR enables a complete accounting of the uncertainty in the profile and its derivatives: the error bars on the fitted curve include all of the flexibility permitted by the data and any prior knowledge regarding the macroscopic smoothness of the profile.

A critical step in fitting profile data using GPR is to infer the hyperparameters\(^\text{§} \) \( \theta \) which determine how rapidly the covariance kernel decays. This can be handled one of two ways: finding a single estimate for the hyperparameters which are most likely given the data, or drawing samples from the posterior distribution for \( \theta \) given the data \( y \):

\[
\tilde{\theta} \sim f_{\theta|y}(\theta|y) = \frac{f_{y|\theta}(y|\theta)f_{\theta}(\theta)}{f_y(y)},
\]

where \( \tilde{\theta} \) is the sample of the hyperparameters, \( f_{\theta|y}(\theta|y) \) is the posterior density containing everything which is known about the hyperparameters once the data have been measured, \( f_{y|\theta}(y|\theta) \) is the likelihood function representing the probability of observing the data given a particular value of the hyperparameters, \( f_{\theta}(\theta) \) is the prior density containing any prior knowledge about typical values for the hyperparameters, and \( f_y(y) \) is a normalization constant. In practice, these samples are drawn using Markov chain Monte Carlo (MCMC) sampling [31, 43, 44], and are then used to estimate

\(^\text{§} \) The term “hyperparameters” is used to denote the fact that these are the parameters of the prior distribution representing the prior knowledge regarding the smoothness of the profile rather than the parameters of a curve used to explicitly fit the profile. See [27, 31] for more details.
Testing the dependence of momentum transport on second derivatives using GPR

Table 1: Prior distributions used for the hyperparameters of the squared exponential covariance kernel when fitting the rotation reversal data.

<table>
<thead>
<tr>
<th>Quantity</th>
<th>$\sigma_f$</th>
<th>$\ell$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_e$</td>
<td>$\mathcal{U}(0, 50 \times 10^{20} \text{ m}^{-3})$</td>
<td>$\mathcal{U}(0, 15)$</td>
</tr>
<tr>
<td>$T_e$</td>
<td>$\mathcal{U}(0, 50 \text{ keV})$</td>
<td>$\mathcal{U}(0, 25)$</td>
</tr>
</tbody>
</table>

the marginalized profile

$$f_{y^* | y}(y^* | y) = \int f_{y^* , \theta | y}(y^* , \theta | y) \, d\theta$$

(2)

$$= \int f_{y^* | y, \theta}(y^* | y, \theta) f_{\theta | y}(\theta | y) \, d\theta$$

(3)

and its uncertainty according to the procedure described in [22, 23], where $y^*$ is the predicted profile and/or its derivatives. As was shown in [22, 23], when inferring derivatives it is necessary to use MCMC sampling to obtain a full accounting of uncertainty.

3. Fitting the second derivative profiles of the rotation reversal data

The profile data from the same shots used in [5] were re-fit using GPR, and the second derivatives and their uncertainties were computed. Both of these L-mode discharges had $I_p = 800$ kA, $B_T = 5.4$ T and were heated with 1.2 MW of ICRF power. For each shot the data were averaged over the period $0.9 \, s \leq t \leq 1.2 \, s$. Data from the core and edge Thomson scattering systems were used for both the $n_e$ and $T_e$ profiles. In addition, the data from two electron cyclotron emission grating polychromator systems were used for the $T_e$ profiles. The equilibrium shape and diagnostic locations for the hollow rotation profile shot are shown in figure 2.

Only the data from inside the LCFS were used, so the rapid change in $n_e$ and $T_e$ at the edge does not need to be fit. This motivates the use of the stationary squared exponential covariance kernel:

$$k_{SE}(x_i , x_j) = \sigma_f^2 \exp \left( -\frac{|x_i - x_j|^2}{2\ell^2} \right),$$

(4)

where $\ell$ is the (constant) covariance length scale which sets how fast the correlation drops off and $\sigma_f^2$ is the signal variance which sets the extent of variation in the fitted curve. It is very important to note that the covariance length scale $\ell$ is not in any way the same thing as the gradient scale length: even if $\ell$ is constant throughout the domain, $a/L$ can still vary. The slope at the magnetic axis was forced to zero and very simple uniform prior distributions were used for the hyperparameters, which are given in table 1. In order to ensure a complete accounting of the uncertainty, the hyperparameters were sampled using MCMC using an affine-invariant ensemble sampler [48, 49]. The sampler was run with 200 walkers for 600 samples, 200 samples were burned and the remaining
Testing the dependence of momentum transport on second derivatives using GPR

Figure 2: Magnetic equilibrium reconstruction and diagnostic locations for the high density shot with hollow rotation profile. The lower density shot with a peaked rotation profile is essentially the same. In the legend, “CTS” is the core Thomson scattering system, “ETS” is the edge Thomson scattering system, and “GPC” and “GPC2” are the two grating polychromator electron cyclotron emission diagnostics. This figure was produced using eqtools [45–47].

samples were thinned by a factor of 200. The computed profiles are given in figures 3 and 4, summary statistics for the posterior distributions of the hyperparameters are given in table 2 and the posterior distributions themselves are shown in figure 5. The posterior distributions for the hyperparameters are very similar between the two shots for both $n_e$ and $T_e$.

Figures 3 and 4 show the unnormalized second derivative profiles as well as two possible normalizations:

$$\frac{a}{L_{\nabla n_e}} = \frac{a d^2 n_e/ dr^2}{dn_e/ dr}$$

$$\frac{a^2 d^2 n_e/ dr^2}{n_e}.$$  \hspace{1cm} (5)

The first version was chosen by analogy with the gradient scale length, $a/L_{n_e} = a (dn_e/ dr)/n_e$, but is made useless in the core by the fact that the first derivative $dn_e/ dr$ must go to zero on axis. Therefore, the second version, chosen purely from dimensional analysis, is likely to be more useful in practice.
Testing the dependence of momentum transport on second derivatives using GPR

4. Quantifying the profile differences

In order to test theories which attempt to explain momentum transport in terms of differences in the second derivatives it is necessary to have an objective way of characterizing the difference between two profiles in the presence of very large, overlapping uncertainty envelopes. One approach could be to apply multivariate hypothesis testing as described in [50]. In this case, the test would be deciding whether or not the data are sufficient to reject the null hypothesis that the mean vectors are equal. This oversimplified choice between “mean vectors match” and “mean vectors do
Figure 4: Data and fitted $T_e$ profile for the two shots with different rotation profiles, presented as in figure 3. With the high resolution ECE measurements, the $T_e$ profiles have a shorter covariance length scale and smaller relative uncertainties than the $n_e$ profiles.

Table 2: Summary statistics of the posterior distributions for the hyperparameters.

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Case</th>
<th>Parameter [units]</th>
<th>Mode</th>
<th>Mean</th>
<th>95% interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_e$</td>
<td>hollow</td>
<td>$\sigma_f [10^{20} \text{ m}^{-3}]$</td>
<td>1.35</td>
<td>13.6</td>
<td>[1.07 , 45.9]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\ell$</td>
<td>1.23</td>
<td>3.10</td>
<td>[0.929, 7.19]</td>
</tr>
<tr>
<td></td>
<td>peaked</td>
<td>$\sigma_f [10^{20} \text{ m}^{-3}]$</td>
<td>1.09</td>
<td>13.1</td>
<td>[0.854, 46.1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\ell$</td>
<td>1.18</td>
<td>3.13</td>
<td>[0.913, 7.40]</td>
</tr>
<tr>
<td>$T_e$</td>
<td>hollow</td>
<td>$\sigma_f [\text{keV}]$</td>
<td>1.20</td>
<td>1.92</td>
<td>[0.835, 5.16]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\ell$</td>
<td>0.317</td>
<td>0.355</td>
<td>[0.255, 0.510]</td>
</tr>
<tr>
<td></td>
<td>peaked</td>
<td>$\sigma_f [\text{keV}]$</td>
<td>1.33</td>
<td>2.09</td>
<td>[0.933, 5.17]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\ell$</td>
<td>0.313</td>
<td>0.344</td>
<td>[0.247, 0.477]</td>
</tr>
</tbody>
</table>
Testing the dependence of momentum transport on second derivatives using GPR

Figure 5: Posterior distributions for the hyperparameters $\sigma_f$ and $\ell$ of the various fits. In each subfigure, the lower left corner shows the bivariate histogram of the MCMC samples and the other two plots show the univariate histograms. The posterior distributions for the peaked and hollow cases are essentially the same. Most of the probability mass is concentrated on the modes at $\sigma_f \approx 10^{20} \text{m}^{-3}$, $\ell \approx 1$ for $n_e$ and $\sigma_f \approx 1 \text{keV}$, $\ell \approx 0.3$ for $T_e$. The posterior distributions are strongly peaked compared to the uniform prior distributions used, which indicates that the sensitivity of the result to the prior distribution is low.

not match” has been noted as a key shortcoming of frequentist hypothesis testing [31]: what really matters is the probability of the profiles being different enough to cause a significant change in the rotation profile. The Bayesian framework allows a more useful test to be made: we can actually use the results of the GPR fit to quantify how likely it is that the profiles differ by a given amount.

For a given pair of quantities to compare $y_h$, $y_p$ (with corresponding realizations of the hyperparameters $\theta_p$, $\theta_h$) we have $y_h|\theta_h \sim \mathcal{N}(\mu_h, \sigma_h^2)$ and $y_p|\theta_p \sim \mathcal{N}(\mu_p, \sigma_p^2)$. In this context, $y_h$ could be the second derivative at $r/a = 0.5$ of the hollow rotation case and $y_p$ could be the second derivative at $r/a = 0.5$ of the peaked rotation case. The probability

\[ || \text{Note that the normalized forms of the second derivative are actually the ratio of two Gaussian-distributed random variables. As was shown in [22], as long as the denominator is sufficiently far from zero the result can be approximated as being Gaussian.} ||
Testing the dependence of momentum transport on second derivatives using GPR

density function (PDF) for the difference of these two quantities, \( \Delta y = y_h - y_p \), is then
\[
\begin{align*}
    f_{\Delta y|\theta_h, \theta_p}(\Delta y|\theta_h, \theta_p) &= \mathcal{N}(\Delta y; \mu_h - \mu_p, \sigma_h^2 + \sigma_p^2).
\end{align*}
\] (7)
Likewise, the cumulative distribution function (CDF) is
\[
\begin{align*}
    F_{\Delta y|\theta_h, \theta_p}(c|\theta_h, \theta_p) &= \mathbb{P}(\Delta y \leq c|\theta_h, \theta_p) \\
    &= \Phi\left( \frac{c - (\mu_h - \mu_p)}{\sigma_h^2 + \sigma_p^2} \right),
\end{align*}
\] (8)
where \( \Phi(x) \) is the CDF of the standard normal distribution \( \mathcal{N}(0,1) \). What is actually of interest are the PDF and CDF averaged over all values of \( \theta_h \) and \( \theta_p \):
\[
\begin{align*}
    f_{\Delta y}(\Delta y) &= \int f_{\Delta y|\theta_h, \theta_p}(\Delta y|\theta_h, \theta_p) f_{\theta_h, \theta_p}(\theta_h, \theta_p) \, d\theta_h \, d\theta_p \quad \text{(9)}
\end{align*}
\]
\[
\begin{align*}
    F_{\Delta y}(c) &= \int F_{\Delta y|\theta_h, \theta_p}(c|\theta_h, \theta_p) f_{\theta_h, \theta_p}(\theta_h, \theta_p) \, d\theta_h \, d\theta_p. \quad \text{(10)}
\end{align*}
\]
These two functions are estimated from the MCMC samples \( \{\theta_h^{(i)}\}, \{\theta_p^{(i)}\} \) according to
\[
\begin{align*}
    f_{\Delta y}(\Delta y) &\approx \frac{1}{n} \sum_{i=1}^{n} f_{\Delta y|\theta_h^{(i)}, \theta_p^{(i)}}(\Delta y|\theta_h^{(i)}, \theta_p^{(i)}) \quad \text{(11)}
\end{align*}
\]
\[
\begin{align*}
    F_{\Delta y}(c) &\approx \frac{1}{n} \sum_{i=1}^{n} F_{\Delta y|\theta_h^{(i)}, \theta_p^{(i)}}(c|\theta_h^{(i)}, \theta_p^{(i)}).
\end{align*}
\] (12)
These two functions can then be used to compute the posterior intervals which show how large the differences are likely to be.

Figure 6 shows the marginalized PDFs for the normalized and unnormalized second derivatives of \( n_e \) and \( T_e \) at \( r/a = 0.5 \). In all six cases the 68\% equal-tailed posterior density interval contains zero, suggesting that there is very little difference in both the normalized and unnormalized second derivatives of both \( n_e \) and \( T_e \) between these two cases. This implies that, unless there is a very strong sensitivity to second derivatives, the momentum diffusion term \( \Pi_{\text{int}}^{\Delta \chi \phi} \) cannot account for the change in rotation between these two discharges. This is in contrast to GS2 predictions for the Cyclone Base Case, which suggest that \( \Pi_{\text{int}}^{\Delta \chi \phi} \) is one of the dominant momentum flux components [16]. That being said, table 3 shows that there are key differences between the Cyclone Base Case and the real experimental conditions. In particular, the Cyclone Base Case is constructed to be far from marginal stability so as to exhibit robust ion temperature gradient (ITG) turbulence [51]. The results presented here suggest that GS2 simulations of the real experimental conditions should be performed to check whether or not the momentum diffusion term is still expected to be dominant in this pair of discharges.

5. Summary and conclusions

This paper has presented the use of Gaussian process regression (GPR) to infer second derivative profiles from noisy, discrete profile data. A complete accounting of uncertainty was obtained through use of Markov chain Monte Carlo (MCMC) sampling. This
Testing the dependence of momentum transport on second derivatives using GPR

Figure 6: Probability density functions for the differences in the normalized and unnormalized second derivatives of $n_e$ and $T_e$ at $r/a = 0.5$ between the hollow and peaked rotation profile shots. The dark and light shaded regions are the 68% and 95% equal-tailed posterior intervals, which indicate the regions where the differences are most likely to lie. The PDFs are asymmetric as a result of having marginalized over the hyperparameters.

work shows that the uncertainties on second derivatives are not as large as might be feared, and hence we can realistically attempt to test hypotheses which depend on these quantities. This technique was used to test the hypothesis that changes in the second derivatives can explain the dramatic change in intrinsic rotation induced by slight changes in density. The present analysis indicates that there is not a substantial difference between the second derivatives of $n_e$ and $T_e$ in the higher density, hollow rotation and lower density, peaked rotation shots. Therefore, unless a theory is put forth which predicts an exceptionally strong sensitivity to the second derivatives, another explanation is needed.

Acknowledgments

The authors are very grateful for the clear description of the relationship between second derivatives and momentum transport provided by M. Barnes. This material is based
Table 3: Comparison of the simulation parameters for the Cyclone Base Case \cite{21} and the peaked rotation discharge (low power shot from \cite{52}).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Cyclone Base Case</th>
<th>Peaked rotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>geometry</td>
<td>shifted-circle</td>
<td>Miller geometry</td>
</tr>
<tr>
<td>electrons</td>
<td>adiabatic</td>
<td>drift-kinetic</td>
</tr>
<tr>
<td>$R/L_{Te}$</td>
<td>6.9</td>
<td>6.0</td>
</tr>
<tr>
<td>$R/L_{Te}$</td>
<td>N/A</td>
<td>8.1</td>
</tr>
<tr>
<td>$R/L_{ne}$</td>
<td>2.2</td>
<td>2.3</td>
</tr>
<tr>
<td>$r/R$</td>
<td>0.18</td>
<td>0.16</td>
</tr>
<tr>
<td>$q$</td>
<td>1.4</td>
<td>1.3</td>
</tr>
<tr>
<td>$s$</td>
<td>0.8</td>
<td>0.91</td>
</tr>
<tr>
<td>$T_i/T_e$</td>
<td>1.0</td>
<td>0.89</td>
</tr>
<tr>
<td>$Z_{\text{eff}}$</td>
<td>1.0</td>
<td>1.9</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>$\nu$</td>
<td>0.0</td>
<td>0.11</td>
</tr>
</tbody>
</table>

upon work conducted using the Alcator C-Mod tokamak, a DOE Office of Science user facility. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Fusion Energy Sciences under Award Number DE-FC02-99ER54512. This material is based upon work supported in part by the U.S. Department of Energy Office of Science Graduate Research Fellowship Program (DOE SCGF), made possible in part by the American Recovery and Reinvestment Act of 2009, administered by ORISE-ORAU under contract number DE-AC05-06OR23100.

Appendix A. Derivatives of the squared exponential covariance kernel

This section obtains an expression for the $n^{\text{th}}$ derivative of the $D$-dimensional squared exponential (SE) covariance kernel. The $n = 2$, $D = 1$ case is what was used for the present work. For this section, $x_i \in \mathbb{R}^D$ is the location an observation or prediction $y_i$ is to be made at, while in the rest of this paper $x_i = r_i/a \in \mathbb{R}$ was a single coordinate.

Recall that the covariance kernel is defined as $k(x_i, x_j) \equiv \text{cov}[y(x_i), y(x_j)]$. Next, note that the relationship between a Gaussian process and its derivatives is

$$\text{cov} \begin{bmatrix} y_i, \frac{\partial y_j}{\partial x_{jd}} \end{bmatrix} = \frac{\partial k(x_i, x_j)}{\partial x_{jd}}$$ \hspace{1cm} (A.1)

$$\text{cov} \begin{bmatrix} \frac{\partial y_i}{\partial x_{ic}}, \frac{\partial y_j}{\partial x_{jd}} \end{bmatrix} = \frac{\partial^2 k(x_i, x_j)}{\partial x_{ic} \partial x_{jd}},$$ \hspace{1cm} (A.2)

where the notation $\partial / \partial x_{jd}$ refers to a derivative with respect to the $d^{\text{th}}$ component of the input $x_j$ to the covariance kernel $k(x_i, x_j)$. These expressions can be extended to arbitrary derivatives simply by taking the relevant partial derivatives of $k(x_i, x_j)$. The SE covariance kernel given in equation (4) only depends on $x_i, x_j$ through the quantity...
Testing the dependence of momentum transport on second derivatives using GPR

\[ r = |\tau| = |x_i - x_j|; \]covariance kernels with this property are said to be \textit{isotropic}. But, if \( x_i \) consists of position and time, for instance, we have no reason to expect the curve to behave the same in every direction. In general, one can make the substitution \( r^2 \to \tau^T M \tau \) to introduce anisotropy into an isotropic covariance kernel, where \( M \) is any positive semidefinite matrix. This work takes \( M \) to be diagonal and uses the substitution

\[
r^2 \to \sum_{d=1}^{D} \frac{r_d^2}{\ell_d^2} \to \sum_{d=1}^{D} \frac{(x_{id} - x_{jd})^2}{\ell_d^2},
\]

where each dimension has its own covariance length scale \( \ell_d \).

Substituting equation (A.3), the general form of the SE covariance kernel is

\[
k_{SE}(x_i, x_j) = k_{SE}(\tau) = \sigma_f^2 \exp \left(-\frac{1}{2} \sum_{d=1}^{D} \frac{r_d^2}{\ell_d^2} \right),
\]

Turning the sum in the middle expression in equation (A.5) into a product yields

\[
k_{SE}(x_i, x_j) = \sigma_f^2 \prod_{d=1}^{D} \exp \left(-\frac{\tau_d^2}{2\ell_d^2} \right).
\]

Recall the Rodrigues representation of the (physicists') Hermite polynomials \( H_n(x) \) [53]:

\[
H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2},
\]

where [54, 55]:

\[
H_n(x) = n! \sum_{l=0}^{\left\lfloor n/2 \right\rfloor} \frac{(-1)^l (2x)^{n-2l}}{l!(n-2l)!}.
\]

Rearranging,

\[
\frac{d^n}{dx^n} e^{-x^2} = (-1)^n H_n(x) e^{-x^2}.
\]

To find the derivatives of equation (A.6), first find

\[
\frac{d^n}{dx^n} \exp \left(-\frac{\tau_d^2}{2\ell_d^2} \right) = \left( \frac{1}{\sqrt{2\ell_d}} \right)^n \frac{d^n}{dx^n} e^{-x^2} = \left( \frac{-1}{\sqrt{2\ell_d}} \right)^n H_n \left( \frac{\tau_d}{\sqrt{2\ell_d}} \right) \exp \left(-\frac{\tau_d^2}{2\ell_d^2} \right),
\]

where the substitution \( x = \tau_d/(\sqrt{2}\ell_d) \) was used. The derivative of equation (A.6) to arbitrary orders for each dimension is then

\[
\frac{\partial^n k_{SE}(\tau)}{\prod_{d=1}^{D} \partial \tau_d^{n_d}} = \sigma_f^2 \prod_{d=1}^{D} \left( \frac{-1}{\sqrt{2\ell_d}} \right)^{n_d} H_{n_d} \left( \frac{\tau_d}{\sqrt{2\ell_d}} \right) \exp \left(-\frac{\tau_d^2}{2\ell_d^2} \right)
\]

\[
= k_{SE}(\tau) \prod_{d=1}^{D} \left( \frac{-1}{\sqrt{2\ell_d}} \right)^{n_d} H_{n_d} \left( \frac{\tau_d}{\sqrt{2\ell_d}} \right),
\]
where $n = \sum_{d=1}^{D} n_d$ and $n_d$ is the order of derivative with respect to dimension $d$. What is needed to apply equations (A.1) and (A.2), however, are the derivatives with respect to $x_{id}$ and $x_{jd}$. Note that

$$\tau_d = x_{id} - x_{jd}, \quad \frac{\partial \tau_d}{\partial x_{id}} = 1, \quad \frac{\partial \tau_d}{\partial x_{jd}} = -1$$

(A.13)

$$\frac{\partial^{n_i+n_j} \tau_d}{\partial x_{id}^{n_i} \partial x_{jd}^{n_j}} = 0, \text{ for } n_i + n_j \geq 2.$$  

(A.14)

Applying the chain rule to equation (A.12) then gives

$$\frac{\partial^n k_{SE}(\tau(x_i, x_j))}{\prod_{d=1}^{D} \partial x_{id}^{n_i} \prod_{d=1}^{D} \partial x_{jd}^{n_j}}$$

(A.15)

$$= \frac{\partial^n k_{SE}(\tau)}{\prod_{d=1}^{D} \partial \tau_d^{n_i+n_j}} \prod_{d=1}^{D} \left( \frac{\partial \tau_d}{\partial x_{id}} \right)^{n_i} \prod_{d=1}^{D} \left( \frac{\partial \tau_d}{\partial x_{jd}} \right)^{n_j}$$

(A.16)

$$= k_{SE}(\tau) \prod_{d=1}^{D} \left( \frac{-1}{\sqrt{2\ell_d}} \right)^{n_i+n_j} \quad H_{n_i+n_j} \left( \frac{\tau_d}{\sqrt{2\ell_d}} \right) \cdot \prod_{d=1}^{D} (-1)^{n_j}$$

(A.17)

$$= k_{SE}(\tau) \prod_{d=1}^{D} \left( \frac{-1}{\sqrt{2\ell_d}} \right)^{n_i+n_j+2n_j} \quad H_{n_i+n_j+n_j} \left( \frac{\tau_d}{\sqrt{2\ell_d}} \right)$$

(A.18)

where $n = \sum_{d=1}^{D} n_{id} + \sum_{d=1}^{D} n_{jd}$, $n_{id}$ is the order of derivative with respect to dimension $d$ of $x_i$ and $n_{jd}$ is the order of derivative with respect to dimension $d$ of $x_j$. Equation (A.18) enables the computation of derivatives of arbitrary order for data of arbitrary dimensionality. A plot of the 1D SE covariance kernel and its first few derivatives is provided in figure A1.

References


Figure A1: One-dimensional squared exponential covariance kernel and a few of its derivatives. The horizontal scale has been scaled by $\ell$ and the vertical scale has been scaled by $\sigma_f^2$ to remove the dependence on the hyperparameters. The solid blue curve gives $k_{SE}(0,x)$ itself. The long-dashed curves give the functions necessary for dealing with first derivatives. The green dashed curve is $\partial k_{SE}/\partial x_j = \text{cov}[y(0), y'(x)]$. The red dashed curve is $\partial k_{SE}/\partial x_i = \text{cov}[y'(0), y(x)]$. The teal dashed curve is $\partial^2 k_{SE}/\partial x_i \partial x_j = \text{cov}[y'(0), y'(x)]$. The dotted curves give the functions necessary for dealing with second derivatives. The magenta dotted curve is $\partial^2 k_{SE}/\partial x_i^2 = \partial^2 k_{SE}/\partial x_j^2 = \text{cov}[y''(0), y(x)] = \text{cov}[y(0), y''(x)]$. The yellow dotted curve is $\partial^3 k_{SE}/\partial x_i \partial x_j^2 = \text{cov}[y'(0), y''(x)]$. The black dotted curve is $\partial^3 k_{SE}/\partial x_i^2 \partial x_j = \text{cov}[y''(0), y'(x)]$. The blue dotted curve is $\partial^4 k_{SE}/\partial x_i^2 \partial x_j^2 = \text{cov}[y''(0), y''(x)]$.


REFERENCES


REFERENCES


REFERENCES


