Linear Operators and Stochastic Partial Differential Equations in Gaussian Process Regression

Simo Särkkä

Department of Biomedical Engineering and Computational Science
Aalto University, Finland
simo.sarkka@tkk.fi

Abstract. In this paper we shall discuss an extension to Gaussian process (GP) regression models, where the measurements are modeled as linear functionals of the underlying GP and the estimation objective is a general linear operator of the process. We shall show how this framework can be used for modeling physical processes involved in measurement of the GP and for encoding physical prior information into regression models in form of stochastic partial differential equations (SPDE). We shall also illustrate the practical applicability of the theory in a simulated application.

Keywords: Gaussian process regression, linear operator, stochastic partial differential equation, inverse problem.

1 Introduction

This paper is concerned with Gaussian process (GP) regression [1,2], which refers to a Bayesian machine learning approach, where the regression functions are modeled non-parametrically as Gaussian processes. In the approach, we first postulate a Gaussian process prior for the unknown function $\mathbf{f}(\mathbf{x})$ and then compute the posterior distribution of the function by conditioning to observed measurements $D = \{(\mathbf{x}_i, \mathbf{y}_i) \ : \ i = 1, \ldots, n\}$. The predicted function values at arbitrary points $\mathbf{x}^*$ are then estimated by computing the posterior predictive distributions of the unknown function values.

Linear operators and functionals, which act on functions in Hilbert spaces [3] are commonly used models in engineering and physics applications. In telecommunications [4], statistical signal processing [5] and image processing [6] the linear operators are typically integral operators in the form of convolution kernels, which model the effect of a linear system to the input or measurement signal. In physics [7,8] the linear operators are typically differential operators, which appear as parts of partial differential equation (PDE) models of the Nature. When PDE models include stochastic terms, they become stochastic partial differential equations (SPDE) [9], which can be used for modeling spatial physical processes with unknown sub-phenomena.

T. Honkela et al. (Eds.): ICANN 2011, Part II, LNCS 6792, pp. 151–158, 2011.

In this paper we shall show how linear operators can be naturally included into GP regression and used for encoding physical and other background information into the models. The information can be encoded either by including a linear operator to the measurement model or by forming the prior covariance function as a solution to a stochastic partial differential equation (SPDE). Similar ideas have been previously used, for example, in Kriging [10,11,12,13,14], image processing [15,16,6], Kalman filtering [17,18], and physical inverse problems [19,20,21]. In the machine learning context, the inclusion of linear operators into GP regression models has also been previously proposed. For example, convolved multi-output Gaussian processes [22,23], latent force models [24], noisy operator equations [25,26], as well as measurement and estimation of derivatives and integrals [27,28] fall into this class of models. In this paper we shall formulate a general model, which includes all of these models and present its solution.

2 Linear Operators and Functionals of GPs

In this article we shall denote the application of linear operator $\mathcal{L}_x$ to the function $\mathbf{f}(\mathbf{x})$, which results in another function $\mathbf{g}(\mathbf{x})$ as $\mathbf{g}(\mathbf{x}) = \mathcal{L}_x\,\mathbf{f}(\mathbf{x})$. We shall assume that the operator produces functions with the same input domain as its input function has – if the input function is $\mathbf{f} : \mathbb{R}^d \mapsto \mathbb{R}^{m_f}$ then the output function is $\mathbf{g} : \mathbb{R}^d \mapsto \mathbb{R}^{m_g}$. That is, the output dimensionality might change, but input not. We shall explicitly allow vector and matrix valued operators, because scalar models are too restricted for most modeling purposes. Examples of allowed linear operators are, for example integration $\mathcal{L}_x\,\mathbf{f}(\mathbf{x}) = \int_{-\infty}^{x_i} \mathbf{f}(\mathbf{x})\,dx_i$, and computation of the Jacobian matrix $[\mathcal{L}_x\,\mathbf{f}(\mathbf{x})]_{ij} = \partial f_i(\mathbf{x})/\partial x_j$.

We shall also need linear functionals $\mathcal{H}_x$, which are just like linear operators, but produce vectors or matrices instead of functions. Functionals can be considered as operators combined with evaluation of the result function at certain point. For example the derivative $\mathcal{H}_x\,\mathbf{f}(\mathbf{x}) = \partial \mathbf{f}(\mathbf{x})/\partial x_i$ evaluated at fixed point $\hat{\mathbf{x}}$ is a functional as well as integral of $\mathbf{f}(\mathbf{x})$ over a fixed area. The mathematical treatment of functionals is similar to operators, because functionals can be considered as a restricted class of operators.

In this article we are interested in the case where $\mathbf{f}(\mathbf{x})$ is modeled as a zero mean Gaussian process with covariance function $\mathrm{E}[\mathbf{f}(\mathbf{x})\,\mathbf{f}^T(\mathbf{x}')] = \mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}')$, which is denoted as

$$\mathbf{f}(\mathbf{x}) \sim \mathcal{GP}(\mathbf{0}, \mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}')). \tag{1}$$

By applying the well known rules for linear transformations of Gaussian processes [29,2], we get that if $\mathbf{g}(\mathbf{x}) = \mathcal{L}_x\,\mathbf{f}(\mathbf{x})$, then we have

$$\begin{aligned} \mathbf{K}_{\mathrm{gf}}(\mathbf{x}, \mathbf{x}') &= \mathcal{L}_x \mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}') \\ \mathbf{K}_{\mathrm{fg}}(\mathbf{x}, \mathbf{x}') &= \mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}')\,\mathcal{L}_{x'}^T \\ \mathbf{K}_{\mathrm{gg}}(\mathbf{x}, \mathbf{x}') &= \mathcal{L}_x\,\mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}')\,\mathcal{L}_{x'}^T. \end{aligned} \tag{2}$$

Here one has to be careful with the notation – the transpose is just a normal matrix transpose, but by operator application from the right we mean an analogous operation for kernels as right multiplication by matrix in linear algebra

is. The multiplication from right here means operating to the second argument of the kernel $\mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}')$ and this is here emphasized by writing $\mathcal{L}_{x'}$ instead of $\mathcal{L}_x$. This convention is required to be able to consistently cope with vector and matrix operators.

To illustrate the right multiplication, we may consider the operator $\mathcal{L}_x = (1, \partial/\partial x)$ acting on scalar Gaussian process $f(x)$ with scalar input and covariance function $k_{\mathrm{ff}}(x, x')$. The cross covariance $\mathbf{K}_{\mathrm{fg}}(x, x')$ is then given as

$$\mathbf{K}_{\mathrm{fg}}(x, x') = k_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}') \begin{pmatrix} 1 & \partial/\partial x' \end{pmatrix} = \begin{pmatrix} k_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}') & \partial k_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}')/\partial x' \end{pmatrix}. \tag{3}$$

The rules for computation with functionals are analogous to the rules for operators, but of course, the result of applying functional to the function, say, $\mathbf{h} = \mathcal{H}_x \, \mathbf{f}(\mathbf{x})$ is a random variable having the covariance matrix

$$\mathbf{K}_{\mathrm{hh}} = \mathcal{H}_x \, \mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}') \, \mathcal{H}_{x'}^T, \tag{4}$$

not a covariance function.

3 GP Regression Model with Linear Operators

If we model the unknown regression function $\mathbf{f} : \mathbb{R}^d \mapsto \mathbb{R}^m$ as a zero mean Gaussian process with covariance function $\mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}')$, then a basic multi–input multi–output Gaussian process regression model can be stated as

$$\begin{aligned} \mathbf{f}(\mathbf{x}) &\sim \mathcal{GP}(\mathbf{0}, \mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}')) \\ \mathbf{y}_i &= \mathbf{f}(\mathbf{x}_i) + \mathbf{e}_i, \end{aligned} \tag{5}$$

where $\mathbf{y}_i, \mathbf{e}_i \in \mathbb{R}^m$. The joint covariance of errors $\mathbf{e} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ is assumed to be given by the matrix Σ. The equations for the posterior mean and covariance given the concatenated vector of measurements $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ are then given by the well-known GP regression equations (see, e.g., [1,2]):

$$\begin{aligned} \mathrm{E}[\mathbf{f}(\mathbf{x}) \,|\, \mathbf{y}] &= \mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}_{1:n}) \, [\mathbf{K}_{\mathrm{ff}}(\mathbf{x}_{1:n}, \mathbf{x}_{1:n}) + \Sigma]^{-1} \, \mathbf{y} \\ \mathrm{Cov}[\mathbf{f}(\mathbf{x}) \,|\, \mathbf{y}] &= \mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}) - \mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}_{1:n}) \, [\mathbf{K}_{\mathrm{ff}}(\mathbf{x}_{1:n}, \mathbf{x}_{1:n}) + \Sigma]^{-1} \, \mathbf{K}_{\mathrm{ff}}^T(\mathbf{x}, \mathbf{x}_{1:n}), \end{aligned} \tag{6}$$

where $\mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}_{1:n})$ is a block matrix formed of blocks $\mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}_j)$ where $j = 1, \ldots, n$ and $\mathbf{K}_{\mathrm{ff}}(\mathbf{x}_{1:n}, \mathbf{x}_{1:n})$ is a block matrix with blocks $\mathbf{K}_{\mathrm{ff}}(\mathbf{x}_i, \mathbf{x}_j)$ where $i = 1, \ldots, n$ and $j = 1, \ldots, n$.

In image processing [6] and physical inverse problems [19,20] the Gaussian process based models are often stated in more general form than Equation (5). In order to model linear processes such as motion blur or physical phenomena occurring in the measurement process, the measurements are assumed to be linear functionals of the underlying Gaussian process, not direct (noisy) values of the process. Furthermore, the estimation objective often is not the underlying Gaussian process, but some linear operator transformation of it, such as its

derivative or integral. With this motivation, a suitably generalized Gaussian process regression model can be stated as

$$\begin{aligned} \mathbf{f}(\mathbf{x}) &\sim \mathcal{GP}(\mathbf{0}, \mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}')) \\ \mathbf{y}_i &= \mathcal{H}_{x,i}\, \mathbf{f}(\mathbf{x}) + \mathbf{e}_i, \end{aligned} \tag{7}$$

where $\mathcal{H}_{x,i}$ is a deterministic linear functional and the objective is to estimate a linear operator transformation of the signal $\mathbf{d}(\mathbf{x}) = \mathcal{L}_x\, \mathbf{f}(\mathbf{x})$, where $\mathcal{L}_x$ is a linear operator. As in the basic model in Equation (5), $\{\mathbf{y}_i \ : \ i = 1, \ldots, n\}$ are the m-dimensional measurements and the errors $\mathbf{e} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ have zero mean and joint covariance matrix Σ.

With these assumptions the posterior process $\mathbf{d}(\mathbf{x})$, after conditioning to measurements, will be a Gaussian process and the mean and covariance can be derived as follows. The joint distribution of $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ and $\mathbf{d}(\mathbf{x})$ is multidimensional Gaussian with zero mean and covariance

$$\mathrm{Cov}\left[\begin{pmatrix} \mathbf{y} \\ \mathbf{d}(\mathbf{x}) \end{pmatrix}\right] = \begin{pmatrix} \mathcal{H}_x\, \mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}')\, \mathcal{H}_{x'}^T + \Sigma & \mathcal{H}_x\, \mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}')\, \mathcal{L}_{x'}^T \\ \mathcal{L}_x\, \mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}')\, \mathcal{H}_{x'}^T & \mathcal{L}_x\, \mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}')\, \mathcal{L}_{x'}^T \end{pmatrix} \Bigg|_{\mathbf{x}'=\mathbf{x}}, \tag{8}$$

where $\mathcal{H}_x = (\mathcal{H}_{x,1}, \ldots, \mathcal{H}_{x,n})$. The substitution $\mathbf{x}' = \mathbf{x}$ after performing all the operations is needed just to get the notation consistent, that is, to make it clear that to which variables do the operators operate to. By the elementary rules for Gaussian random variables we can compute the conditional mean and covariance of $\mathbf{d}(\mathbf{x})$:

$$\begin{aligned} \mathrm{E}[\mathbf{d}(\mathbf{x}) \,|\, \mathbf{y}] &= \mathcal{L}_x\, \mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}')\, \mathcal{H}_{x'}^T \left[\mathcal{H}_x\, \mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}')\, \mathcal{H}_{x'}^T + \Sigma\right]^{-1} \mathbf{y} \\ \mathrm{Cov}[\mathbf{d}(\mathbf{x}) \,|\, \mathbf{y}] &= \mathcal{L}_x\, \mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}')\, \mathcal{L}_{x'}^T \\ &\quad - \mathcal{L}_x\, \mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}')\, \mathcal{H}_{x'}^T \left[\mathcal{H}_x\, \mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}')\, \mathcal{H}_{x'}^T + \Sigma\right]^{-1} \mathcal{H}_x\, \mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}')\, \mathcal{L}_{x'}^T \Big|_{\mathbf{x}'=\mathbf{x}}, \end{aligned} \tag{9}$$

where the substitution $\mathbf{x}' = \mathbf{x}$ is understood to be done for all instances of $\mathbf{x}'$ after all the operations in the expression have been performed.

4 Stochastic Partial Differential Equations

Linear stochastic partial differential equation (SPDE) is an operator equation of the form

$$\mathcal{D}_x\, \mathbf{g}(\mathbf{x}) = \mathbf{n}(\mathbf{x}), \tag{10}$$

where $\mathcal{D}_x$ is a linear differential operator and $\mathbf{n}(\mathbf{x})$ is a Gaussian process with zero mean and covariance function $\mathbf{K}_{\mathrm{nn}}(\mathbf{x}, \mathbf{x}')$. Note that often SPDE refers to the special case where $\mathbf{n}(\mathbf{x})$ is a Gaussian white noise process [9], that is, a process with covariance function $\mathbf{K}_{\mathrm{nn}}(\mathbf{x}, \mathbf{x}') = \mathbf{Q}\, \delta(\mathbf{x} - \mathbf{x}')$, but here we shall define SPDE in the above generalized sense.

Assume that we now want to solve the corresponding inverse problem, that is, estimate $\mathbf{n}(\mathbf{x})$ in the Equation (10), based on the measurements of $\mathbf{g}(\mathbf{x})$. If we convert this into Gaussian process regression model, denote the unknown function as $\mathbf{f}(\mathbf{x})$, and include the measurement functional, the general model can be written in form

$$\begin{aligned} \mathbf{f}(\mathbf{x}) &\sim \mathcal{GP}(\mathbf{0}, \mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}')) \\ \mathcal{D}_x\, \mathbf{g}(\mathbf{x}) &= \mathbf{f}(\mathbf{x}) \\ \mathbf{y}_i &= \mathcal{H}_{x,i}\, \mathbf{g}(\mathbf{x}) + \mathbf{e}_i. \end{aligned} \tag{11}$$

Unfortunately this model is incompatible with the GP model in Equation (7).

Fortunately, we can convert the model into compatible form by introducing the Green's function $\mathbf{G}(\mathbf{x}, \mathbf{x}')$ of the operator, which can be interpreted as inverse of the operator $\mathcal{D}_x$. Thus we get the formal solution for $\mathbf{g}(\mathbf{x}) = \int_{\mathscr{X}} \mathbf{G}(\mathbf{x}, \mathbf{x}')\, \mathbf{f}(\mathbf{x}')\, d\mathbf{x}'$ and the model can be rewritten as

$$\begin{aligned} \mathbf{f}(\mathbf{x}) &\sim \mathcal{GP}(\mathbf{0}, \mathbf{K}_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}')) \\ \mathbf{y}_i &= \mathcal{H}_{x,i} \int_{\mathscr{X}} \mathbf{G}(\mathbf{x}, \mathbf{x}')\, \mathbf{f}(\mathbf{x}')\, d\mathbf{x}' + \mathbf{e}_i, \end{aligned} \tag{12}$$

which is compatible with the model (7).

Physical laws such as the laws of electromagnetism are often linear PDEs. When linear PDE is driven by Gaussian process, we get a SPDE, whose solution is a Gaussian process. By suitable formulation of the Gaussian process regression model and the covariance functions is it also possible to encode physical laws into the model.

5 Simulation: Electrostatic Inverse Problem

Assume that we are measuring electrostatic potential $\phi(\mathbf{x})$ of a stationary charge density at some predefined points $\mathbf{x}_i$. If we denote the unknown charge density as $f(\mathbf{x})$, then from Maxwell's equations (see, e.g., [8]) we know that the potential is the solution to the Poisson equation $\nabla^2 \varphi(\mathbf{x}) = -f(\mathbf{x})/\epsilon_0$. If we model the unknown charge density as a Gaussian process, then by using the procedure presented in Section 4 the model can be converted into the form

$$\begin{aligned} f(\mathbf{x}) &\sim \mathcal{GP}(0, k_{\mathrm{ff}}(\mathbf{x}, \mathbf{x}')) \\ y_i &= \frac{1}{4\pi\epsilon_0} \int_{\mathbb{R}^3} \frac{f(\mathbf{x}')}{||\mathbf{x}_i - \mathbf{x}'||}\, d\mathbf{x}' + e_i, \end{aligned} \tag{13}$$

which indeed has the form of the extended GP regression model (7) with the measurement model functional defined as

$$\mathcal{H}_{x,i}\, f(\mathbf{x}) = \frac{1}{4\pi\epsilon_0} \int_{\mathbb{R}^3} \frac{f(\mathbf{x}')}{||\mathbf{x}_i - \mathbf{x}'||}\, d\mathbf{x}'. \tag{14}$$

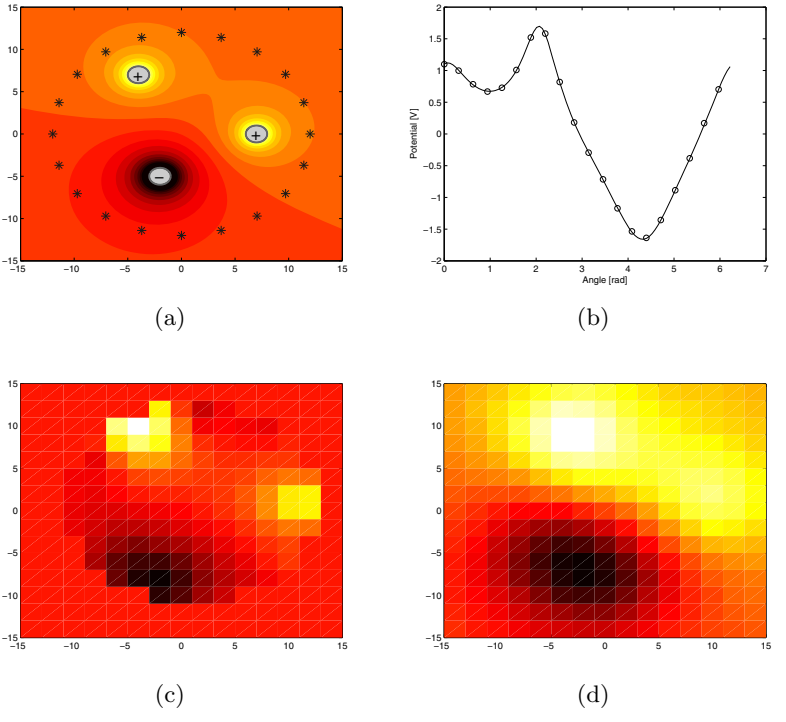

Fig. 1. (a) Charges, potential and measurements points. The measurement points are located at radius $r = 12$ around the origin. (b) Measured point potentials. (c) Estimated charge density. (d) Estimated potential.

In this case we assume that due to the physical setup we know that the charge density is confined into the area of radius $r_0 = 12$ centered at origin and outside that area the charge density is identically zero. If we assume that the prior covariance function of $f(\mathbf{x})$ is a squared exponential inside the specified area, the suitable non-stationary covariance function is

$$k_{\text{ff}}(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{1}{2l^2}||\mathbf{x} - \mathbf{x}'||^2\right) \mathcal{I}(||\mathbf{x}|| < r_0)\, \mathcal{I}(||\mathbf{x}'|| < r_0), \tag{15}$$

where the $\mathcal{I}$ denotes the indicator function, which takes the value 1 when the argument is true, zero otherwise.

In the simulation, three charged objects were placed on the xy plane and electrostatic potential measurements were obtained on circle of radius $r = 12$ around the origin in xy-plane, as shown in Figure 1 (a). The measurements were corrupted by Gaussian noise with standard deviation 0.01. Note that the selection of the finite support covariance function also ensures that the singularities in the integrands disappear, because the measurements are strictly on the area with no charge. The electrostatic potential measurements obtained from the simulated measurement points are shown in Figure 1 (b).

The computation was performed by simple Euler discretization of the integral operator in Equation (14), but more efficient quadrature or cubature integration methods could be applied instead. With this approximation the mean and covariance Equations (9) reduced to matrix expressions for the grid points. The estimated charge density with prior parameter values $\sigma^2 = 10^{-12}, l = 2$ is shown in Figure 1 (c) and the corresponding predicted potential is shown in Figure 1 (d). As can be seen in the figures, both the estimated charge density and potential indeed visually resemble the true quantities, and the estimates are quite reasonable given that we have only observed the 20 measurements in Figure 1 (b). However, due to the ambiguities in the physical setup, the prior covariance defines the size and shape of the charge areas and with the current selection the change areas and the corresponding potentials are always smoother and flatter than the true ones.

6 Conclusion

In this article we have discussed an extended Gaussian process regression model for machine learning problems, which allows modeling of linear functionals in measurement process and estimation of the result of a linear operator applied to the Gaussian process. We have also demonstrated how this extended model can be used for formulating Gaussian process based solutions to physical inverse problems. The practical applicability of the theory was illustrated in a simulation of a classical electrostatic inversion problem.

Acknowledgments. The author is grateful to the Centre of Excellence in Computational Complex Systems Research of Academy of Finland for the financial support and also likes to thank Aki Vehtari and Jouko Lampinen for helpful comments on the manuscript.

References

1. O'Hagan, A.: Curve fitting and optimal design for prediction (with discussion). Journal of the Royal Statistical Society B 40(1), 1–42 (1978)
2. Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning. MIT Press, Cambridge (2006)
3. Akhiezer, N.I., Glazman, I.M.: Theory of Linear Operators in Hilbert Space. Dover (1993)
4. Carlson, A.B.: Communication Systems: An Introduction to Signals and Noise in Electrical Communication, 3rd edn. McGraw-Hill, New York (1986)
5. Hayes, M.H.: Statistical Digital Signal Processing and Modeling. John Wiley & Sons, Chichester (1996)
6. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Prentice-Hall, Englewood Cliffs (2002)
7. Guenther, R.B., Lee, J.W.: Partial Differential Equations of Mathematical Physics and Integral Equations. Dover, New York (1988)
8. Griffiths, D.J.: Introduction to Electrodynamics, 3rd edn. Pearson, London (2008)

9. Holden, H., Øksendal, B., Ubøe, J., Zhang, T. (eds.): Stochastic Partial Differential Equations: A Modeling, White Noise Functional Approach. Birkhäuser, Basel (1996)
10. Whittle, P.: On stationary processes in the plane. Biometrica 41(3/4), 434–449 (1954)
11. Matérn, B.: Spatial variation. Technical report, Meddelanden från Statens Skogforskningsinstitut, Band 49 - Nr 5 (1960)
12. Christakos, G.: Random Field Models in Earth Sciences. Academic Press, London (1992)
13. Cressie, N.A.C.: Statistics for Spatial Data. Wiley, Chichester (1993)
14. Gelfand, A.E., Diggle, P.J., Fuentes, M., Guttorp, P.: Handbook of Spatial Statistics. Chapman & Hall/CRC (2010)
15. Jain, A.K.: Partial differential equations and finite-difference methods in image processing, part 1: Image representation. Journal of Optimization Theory and Applications 23(1) (1977)
16. Jain, A.K., Jain, J.R.: Partial differential equations and finite difference methods in image processing – part II: Image restoration. IEEE Transactions on Automatic Control 23(5) (1978)
17. Curtain, R.: A survey of infinite-dimensional filtering. SIAM Review 17(3), 395–411 (1975)
18. Ray, W.H., Lainiotis, D.G.: Distributed Parameter Systems. Marcel Dekker, New York (1978)
19. Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation. SIAM, Philadelphia (2004)
20. Kaipio, J., Somersalo, E.: Statistical and Computational Inverse Problems. Applied mathematical Sciences, vol. 160. Springer, Heidelberg (2005)
21. Hiltunen, P., Särkkä, S., Nissila, I., Lajunen, A., Lampinen, J.: State space regularization in the nonstationary inverse problem for diffuse optical tomography. Inverse Problems 27(2) (2011)
22. Alvarez, M., Lawrence, N.D.: Sparse convolved Gaussian processes for multi-output regression. In: Koller, D., Schuurmans, D., Bengio, Y., Bottou, L. (eds.) Advances in Neural Information Processing Systems, vol. 21, pp. 57–64. The MIT Press, Cambridge (2009)
23. Alvarez, M., Luengo, D., Titsias, M.K., Lawrence, N.D.: Efficient multioutput Gaussian processes through variational inducing kernels. In: Teh, Y.W., Titterington, M. (eds.) Proceedings of the 13th International Workshop on Artificial Intelligence and Statistics, pp. 25–32 (2010)
24. Alvarez, M., Lawrence, N.D.: Latent force models. In: van Dyk, D., Welling, M. (eds.) Proceedings of the 12th International Workshop on Artificial Intelligence and Statistics, pp. 9–16 (2009)
25. Vapnik, V.: Statistical Learning Theory. Wiley, Chichester (1998)
26. Graepel, T.: Solving noisy linear operator equations by Gaussian processes: Application to ordinary and partial differential equations. In: Proceedings of 20th International Conference on Machine Learning (2003)
27. Solak, E., Murray-Smith, R., Leithead, W.E., Leith, D., Rasmussen, C.E.: Derivative observations in Gaussian process models of dynamic systems. In: Advances in Neural Information Processing Systems, vol. 15, pp. 1033–1040. MIT Press, Cambridge (2003)
28. O'Hagan, A.: Bayes-Hermite quadrature. Journal of Statistical Planning and Inference 29, 245–260 (1991)
29. Papoulis, A.: Probability, Random Variables, and Stochastic Processes. McGraw-Hill, New York (1984)