



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Institut Supérieur de l'Aéronautique et de l'Espace (ISAE)*

Présentée et soutenue le 07/12/2017 par :

ANKIT CHIPLUNKAR

Incorporating Prior Information from Engineering Design into Gaussian Process Regression: with applications to Aeronautical Engineering

JURY

NICOLAS GAYTON
ALEXANDER FORRESTER
RODOLPHE LE RICHE
NATHALIE BARTOLI
JOSEPH MORLIER
EMMANUEL RACHELSON

IFMA, Clermont-Ferrand
University of Southampton
CNRS, Saint-Étienne
ONERA, Toulouse
ISAE-Supaero, Toulouse
ISAE-Supaero, Toulouse

Rapporteur
Rapporteur
Examineur
Examineur
Directeur de thèse
Co-directeur de thèse

École doctorale et spécialité :

AA : Aéronautique et Astronautique

Unité de Recherche :

Institut Clément Ader

Directeur(s) de Thèse :

Prof. Joseph MORLIER (directeur de thèse), Dr. Emmanuel RACHELSON (co-directeur de thèse) et M. Michele COLOMBO (superviseur industriel)

Rapporteurs :

Prof. Nicolas GAYTON et Prof. Alexander FORRESTER

Contents

I	Introduction	1
1	Context	3
1.1	Introduction	4
1.2	Aircraft design cycle	5
1.3	Machine Learning	7
1.4	Bayesian linear regression	9
1.5	Layout	13
1.6	Contributions	15
2	Gaussian Process Regression	17
2.1	Introduction	18
2.2	Prior	19
2.2.1	Hyper-parameters	19
2.2.2	Mean function	20
2.2.3	Covariance function	20
2.2.4	Sampling functions from GP priors	21
2.3	Posterior	24
2.3.1	Posterior with Noise-free observations	24
2.3.2	Posterior with Noisy observations	27
2.3.3	Interpretation of posterior	29
2.4	Choosing Hyper-parameters	31
2.5	Summary and discussion	35
3	Scaling up Gaussian Process Regression	37

3.1	Introduction	38
3.2	Sparse Approximations	38
3.2.1	Nyström Approximation	38
3.2.2	Variational Approximation	41
3.2.3	Experiments	42
3.3	Distributed Gaussian Process	44
3.3.1	Combining experts	46
3.3.2	Experiments	49
3.4	Summary and discussion	49
II	Incorporating structure in Gaussian Process Regression	53
4	Basic Covariance Functions	55
4.1	Introduction	56
4.2	Properties	57
4.3	Non-stationary kernels	58
4.3.1	Linear Kernel	58
4.3.2	Neural Network Kernel	59
4.3.3	Constant and Noise kernel	61
4.4	Stationary kernels	62
4.4.1	Squared Exponential Kernel	63
4.4.2	Matérn Kernel	63
4.4.3	Experiments	66
4.4.4	Spectral Mixture Kernels	69
4.5	Application: Identifying Structural Dynamics Parameters	70
4.5.1	Results on a toy data-set	73
4.5.2	Results on a HCT building data-set	75
4.6	Summary and discussion	75
5	Combining Basic Covariance Functions	79
5.1	Introduction	80

5.2	One dimensional inputs	81
5.2.1	Multiplying Kernels	81
5.2.2	Adding Kernels	81
5.2.3	Change-Point kernels	83
5.2.4	Application: Identifying onset of non-linearity using CP kernel . . .	84
5.3	Multi-dimensional kernels	86
5.3.1	Adding across dimensions	86
5.3.2	Multiplying across dimensions	87
5.3.3	Sensitivity analysis	87
5.3.4	Low dimensional structure	89
5.3.5	Application: Interpolation of aerodynamic pressures	90
5.4	Summary and discussion	98

III Incorporating multiple outputs in Gaussian Process Regression 101

6	Multi-task learning for extrapolation	103
6.1	Introduction	104
6.2	Multi-task Gaussian Process	105
6.2.1	An extra dimension	106
6.2.2	Simple Multi-task kernel	107
6.2.3	Linear Model of Coregionalization	109
6.2.4	Multi-fidelity GP	110
6.2.5	Posterior distribution	112
6.3	Experiments	113
6.4	Extrapolation of Flight Loads	116
6.4.1	Current Approach	117
6.4.2	Proposed Approach	119
6.4.3	Proposed method	119
6.4.4	Experiments on toy data-set	121
6.5	Summary and discussion	123

7	Adding relationships in GP Regression	125
7.1	Introduction	126
7.2	Encoding relationships in GP	126
7.2.1	Case of Linear Operators	127
7.2.2	Case of non-linear operators	131
7.2.3	Calculating posterior	133
7.3	Experiments	134
7.3.1	Quadratic relation on Synthetic Data	135
7.3.2	Flight Mechanics on Flight Test Data	137
7.4	Scaling up MTGP	139
7.4.1	Variational Approximation on Multi-output GP	139
7.4.2	Distributed Inference on Multi-output GP	141
7.5	Experiments	142
7.5.1	Experiments on Theoretical Data	142
7.5.2	Experiments on Flight Test Data	145
7.6	Summary and discussion	148
IV	Conclusions	149
8	Conclusions and perspectives	151
8.1	Perspectives	153
A	Gaussian Identities	157
A.1	Joint distribution	158
A.2	Sum of Gaussians	158
A.3	Affine property of Gaussians	158
A.4	Marginal Distribution	158
A.5	Conditional Distribution	159
B	Proper Orthogonal Decomposition for pressure snapshots	161
B.1	Pressure Snapshot	161
B.2	POD for aerodynamic snapshots	162

B.3 Interpolation	163
C Gaussian Process Regression enforcing Non-linear operators	165
Bibliography	169

List of Figures

1.1	Phases of an aircraft design cycle	5
1.2	Induction vs Deduction	6
1.3	Bias vs Variance trade-off	8
1.4	Prior: Contours of probability density for a prior on parameters w as defined by equation 1.7	11
1.5	Posterior and Prediction in Bayesian linear regression	12
2.1	Covariance matrix for a SE kernel with different hyper-parameters at the input points $\mathbf{X}^* = \{[0 : 0.02 : 1]\}$. The SE kernel of figure 2.1(a) has a lower length-scale than figure 2.1(b). Notice how the covariance values are more spread out for figure 2.1(b).	23
2.2	The solid black line defines the mean function, shaded blue region defines 95% confidence interval (2σ) distance away from the mean. The dashed lines represent five functions drawn at random from a GP prior. We can observe that figure 2.2(a) varies faster when compared to figure 2.2(b) due to smaller length scale hyper-parameter.	25
2.3	Prediction in the case of noiseless observations. The solid black line defines the mean function, blue region defines 95% confidence interval (2σ) distance away from the mean. The dashed lines represent three functions drawn at random from a GP posterior. We can observe that Bayes Theorem eliminates all the functions that do not pass through the observed data-set.	28
2.4	Prediction in the case of noisy observations. The solid black line defines the mean function, blue region defines 95% confidence interval (2σ) distance away from the mean. The dashed lines represent three functions drawn at random from a GP posterior. The mean and the draws do not pass exactly from the observation points.	30

2.5	Posteriors for 2 different sets of hyper-parameters. Solid black line defines the mean function, blue region defines 95% confidence interval (2σ) distance away from mean.	32
2.6	Maximizing marginal likelihood	35
3.1	Approximate Gram matrix for a SE kernel using Nyström approximation. .	40
3.2	Results of Nyström Approximation on a toy data-set of size $N = 1000$. .	44
3.3	Approximate Gram matrix for a SE kernel using mixture of experts. . . .	46
3.4	Results of Distributed GPs approximation on a toy data-set of size $N = 1000$	50
4.1	Prior and posterior from a GP prior of linear kernel. The solid black line defines the mean function, shaded blue region defines 95% confidence interval (2σ) distance away from the mean. The dashed lines represent five functions drawn at random from a GP prior.	60
4.2	Maximizing Marginal Likelihood Linear kernel	60
4.3	Draws from Neural Network kernels having different hyper-parameters. The solid black line defines the mean function, shaded blue region defines 95% confidence interval (2σ) distance away from the mean. The dashed lines represent five functions drawn at random from a GP prior.	61
4.4	Covariance functions and Power spectrums for SE kernel with three different length-scales	64
4.5	Covariance functions and Power spectrums for three different kernels . . .	65
4.6	Prior distribution and five random draws from 3 different covariance functions. The solid black line defines the mean function, shaded blue region defines 95% confidence interval (2σ) distance away from the mean. The dashed lines represent five functions drawn at random from a GP prior. . .	66
4.7	Posterior distribution and three random draws from 3 different covariance functions. The solid black line defines the mean function, shaded blue region defines 95% confidence interval (2σ) distance away from the mean. The dashed lines represent five functions drawn at random from a GP prior. . .	67
4.8	Posterior distributions from three different covariance functions after maximizing the hyper-parameters.	68
4.9	Different types of measurements for estimation of Modal parameters in OMA	71
4.10	Results of Spectral Mixture kernels on a toy data-set	74
4.11	Results of Spectral Mixture kernels on real data from HTC building	76

5.1	Random draws from combining a Linear and SE kernel. The solid black line defines the mean function, shaded blue region defines 95% confidence interval (2σ) distance away from the mean. The dashed lines represent five functions drawn at random from a GP prior. Random functions drawn from a linear GP are linear.	82
5.2	Random draws by having a change-point between a Linear and SE kernel. The solid black line defines the mean function, shaded blue region defines 95% confidence interval (2σ) distance away from the mean. The dashed lines represent five functions drawn at random from a GP prior. Random functions drawn from a linear GP are linear.	84
5.3	Estimation of linear regimes using a change-point kernels	86
5.4	Random draw from a 2 dimensional prior.	88
5.5	Random draw from a 2 dimensional prior which encodes a low-dimensional structure. The hyper-parameters of the kernel are $\theta_{amplitude} = 1$ and $\theta_{lengthscale} = 1$	89
5.6	Details of the FTF	92
5.7	Results for elsA interpolation	93
5.8	CRM shape for aerodynamic simulations	95
5.9	Results for CRM interpolation	96
5.10	Comparison of pressure interpolations for first cut $y/b = 0.105$. Here we compare the accuracy of prediction for interpolation and extrapolation cases.	97
5.11	Performance of Distributed GPs across cuts	98
5.12	Pressure reconstructions for constant $Mach = 0.845$ and sweeping $\alpha \in [1, 3]$	99
6.1	The outputs can also be represented as an extra dimension.	106
6.2	Joint draws from a simple multi-task kernel function between outputs y^1 and y^2 . The matrix $\mathbf{K}_{output} = [4, 0.2; 0.2, 1]$ while the covariance function between inputs $k_{input}(x_1, x_2)$ is a SE kernel with hyper-parameters ($\theta = [1, 0.2]$). The solid black line defines the mean function, shaded region defines 95% confidence interval (2σ) distance away from the mean. The dashed lines represent five functions drawn at random from a GP prior. We can observe that figure 6.3 varies faster when compared to figure 6.2 due to smaller length scale hyper-parameter in one of the latent functions.	108

6.3	Joint draws from a ‘Linear Model of Coregionalization’ kernel function between outputs y^1 and y^2 . We use two latent functions for figure; the kernels between output dimensions are $k_{output}^1 = [4, 0.2; 0.2, 1]$ and $k_{output}^2 = [1, 0.2; 0.2, 1]$, while the covariance function between inputs are SE kernel with hyper-parameters ($\theta = [1, 0.2]$) for the first latent process and ($\theta = [1, 0.1]$) for the second latent process. The solid black line defines the mean function, shaded region defines 95% confidence interval (2σ) distance away from the mean. The dashed lines represent five functions drawn at random from a GP prior. We can observe that figure 6.3 varies faster when compared to figure 6.2 due to smaller length scale hyper-parameter in one of the latent functions.	110
6.4	Joint Posterior distribution for two outputs with missing data at $x^2 = [-1 : -0.75]$. The solid black line defines the mean function, shaded region defines 95% confidence interval (2σ) distance away from the mean.	115
6.5	10-fold RMSE box-plots for increasing number of input points. The box-plots in red are cases when outputs are assumed independent, the box-plots in green are cases when covariance is assumed to be multi-task, and the box-plots in blue are cases when the output covariance is assumed to be multi-fidelity. The error improves with increasing data-points, the multi-fidelity covariance is best fir for all the three cases.	116
6.6	Flight-loads identification	117
6.7	Current Flight-Loads theoretical model	118
6.8	Example of difference between simulated Coefficient of lift (f^s) and experimental Coefficient of lift (y^e) (coming from equation 6.36), with respect to angle of attack.	122
6.9	Boxplots of extrapolation for the four different cases; experimental set with no error (equations 6.33), with a white noise error (equations 6.35), with a sinusoidal error (equations 6.35) and with translation (equations 6.36). . .	122
7.1	Gram matrices for the joint kernel, which encodes the relation between two outputs y^1 and y^2	130
7.2	Joint draws defined by priors on related-outputs. The solid black line defines the mean function, shaded blue region defines 95% confidence interval (2σ) distance away from the mean. The dashed lines represent five functions drawn at random from a GP prior.	131

7.3	Multi-Output Gaussian Process Regression Predictions. The solid black line defines the mean function, shaded region defines 95% confidence interval (2σ) distance away from the mean. Dotted lines represent the functions randomly drawn from the posterior distributions.	134
7.4	GP Regression on Quadratic relationship. The solid black line represents the predicted mean while the shaded area denotes 2σ uncertainty region. The dashed black line represents the real value of f^1 or f^2 . For y^1 the data is hidden from section $x = [-0.2, 0.2]$	136
7.5	Joint multi-output GPR on K_z and C_z relationship. The solid black line represents the predicted mean while the shaded area denotes 2σ uncertainty region. The figure demonstrates the impact of the covariance function for relationship enabled regression.	138
7.6	Approximate Gram matrix for a Joint MTGP kernel using Nyström approximation	140
7.7	Approximate Gram matrix for a SE kernel using mixture of experts.	141
7.8	Experimental results for differential relationship while using Variational approximation. Predicted mean is represented by solid black line. 2σ confidence band is represented by light red for f^1 and light blue for f^2 . The dashed black line represents the true latent function values	144
7.9	Comparison of run time and RMSE between Distributed GPs and Variational Inference	145
7.10	Wing Load Diagram	146
7.11	Experimental results for aircraft flight loads	147

List of Matlab Codes

2.1	A zero mean function	20
2.2	A SE covariance function	21
2.3	Plotting the Gram Matrix	22
2.4	Sampling a random function from the prior	23
2.5	Calculating and plotting the mean, the variance and a sample of the posterior	26
2.6	Matlab code for a noisy SE covariance function	28
2.7	Code for data-set D2	32
2.8	Optimizing the Log Marginal Likelihood	34
3.1	Gram Matrix using Nyström Approximation	39
3.2	Code for toy data-set 3	42
3.3	Randomly clustering points into experts	45
3.4	log Marginal Likelihood for Distributed GPs	48
6.1	Code for data-set D4	114
7.1	Create joint covariance functions using Matlab Symbolic Toolbox	129
7.2	Create joint covariance functions for non-linear operators using Matlab Symbolic Toolbox	132

Nomenclature

$GP[m(x), k(x, x')]$ Gaussian Process

$\Pr[a \mid b]$ Conditional probability distribution of a given b

$\Pr[a, b]$ Joint probability distribution of a and b

$\Pr[a]$ Probability distribution of a

\mathbf{a} Vector \mathbf{a} , written in bold lowercase

\mathbf{a}^T Transpose of vector \mathbf{a}

$\mathcal{N}[\boldsymbol{\mu}, \boldsymbol{\Sigma}]$ Multivariate Gaussian Distribution

$\mathcal{N}[\mu, \sigma^2]$ Scalar Gaussian distribution

\mathcal{R} Set of all real numbers

\mathbf{A} Matrix \mathbf{A} , written in bold uppercase

a Scalar a , written in lowercase

A_i i^{th} mode shape

$Cov[x, x']$ Covariance between two random variables

D_{inputs} Number of dimensions in input data

$D_{outputs}$ Number of dimensions in the output data

M Number of inducing points

M_x Bending moment on the wing

N Number of points in training dataset

N_*	Number of points in test dataset
$N_{experiments}$	Number of experiments
$N_{experts}$	Number of experts in distributed GP
N_{hyp}	Number of hyper-parameters
N_{joint}	Total number of points for MTGP
N_j	Number of points in the j^{th} output for MTGP
N_{nodes}	Number of nodes in aerodynamic mesh
P	Number of points in an expert
Q	Number of Gaussians in a SM kernel
T_z	Shear Load on the wing
Ω	Full spatial coordinates of the aerodynamic mesh
θ	Vector of hyper-parameters
\mathbf{y}_{joint}	Joint output dataset for MTGP
\mathbf{y}	Full training output dataset
\mathbf{y}^j	Full set of outputs for the j^{th} output for MTGP
δ	Dirac-delta function
ϵ	Independent white noise
η_{edge}	Wing span at the edge of the wing
η_{root}	Wing span at the root of the wing
λ_i	i^{th} modal frequency
$\mathbf{E}[x]$	Expectation of the random variable x
\mathcal{D}_i	i^{th} Data-set
μ	Mean value of a Gaussian Distribution
\mathbf{C}	Damping matrix
\mathbf{H}	Low-rank matrix

\mathbf{I}_N	Identity matrix of size N
$\mathbf{K}(\mathbf{X}_1, \mathbf{X}_2)$	Gram matrix between \mathbf{X}_1 and \mathbf{X}_2
\mathbf{K}_{FITC}	Gram matrix using FITC approximation
$\mathbf{K}_{Nyström}$	Gram matrix using Nyström approximation
\mathbf{K}_{lower}	Lower Cholesky decomposition of Gram matrix
$\mathbf{K}_{outputs}$	Covariance matrix between output dimensions
$\mathbf{K}_{stiffness}$	Stiffness matrix
\mathbf{M}	Mass matrix
\mathbf{P}	Full Pressure data for the aerodynamic mesh
\mathbf{X}^M	Set of inducing points
\mathbf{X}^j	Full set of inputs for the j^{th} output for MTGP
\mathbf{X}_{joint}	Joint input dataset for MTGP
\mathbf{X}	Full training input dataset
Σ_{Prior}	Covariance matrix of prior over parameters
\mathbf{x}_*	Full test input dataset
ν	Degrees of freedom of student distribution
ω_i	i^{th} spatial coordinate of the aerodynamic node
$\phi(x)$	Basis functions
$\sigma^{(i)}(x)$	Predicted covariance of the i^{th} expert
σ_n^2	Variance of the white noise
τ	Time lag between two points
$\theta_{amplitude}$	Amplitude hyper-parameter
$\theta_{changeLocation}$	Change location hyper-parameter
$\theta_{intensity}$	Intensity hyper-parameter for k_{CP}
$\theta_{lengthScale}$	Length scale hyper-parameter

b	Wing span
d	Distance between two input points
$diag(\mathbf{K})$	Diagonal of the matrix
f	Function mapping transformation from x to y
$f_{pressure}$	Pressure function mapping transformation from ω_i to p_i
$k(x, x')$	Covariance function a GP
k^δ	Covariance function of the difference between multi-fidelity GPs
$m(x)$	Mean function a GP
$m^{(i)}(x)$	Predicted mean of the i^{th} expert
p_i	Pressure at the i^{th} aerodynamic node
w	Parameters of functions
x_*	Single test input point
x_i	i^{th} observational input point
y_i	i^{th} observational output point
<i>ANOVA</i>	Analysis Of Variance
<i>ARD</i>	Automatic Relevance Determination
<i>BCM</i>	Bayesian Committee Machine
<i>BIC</i>	Bayesian Information Criterion
<i>CFD</i>	Computational Fluid Dynamics
<i>CP</i>	Change-Point
<i>CRM</i>	Common Research Model
<i>CV</i>	Cross Validation
<i>EMA</i>	Experimental Modal Analysis
<i>FEM</i>	Finite Element Method
<i>FITC</i>	Fully Independent Training Conditional

FTF Flap Track Fairing
GP Gaussian Process
HCT Heritage Court Tower
IT Information Technology
LOO Leave One Out
MDO Multi Disciplinary Optimization
MDOF Multi Degree Of Freedom
NE \acute{x} T Natural Excitation Technique
OMA Operational Modal Analysis
PCA Principal Component Analysis
POD Proper Orthogonal Decomposition
POE Product Of Experts
PSD Positive Semi Definite
RANS Reynolds Averaged Navier-Stokes
RFP Rational Fractional Polynomial
RMSE Root Mean Square Error
ROM Reduced Order Models
SE Standard Exponential
SVD Singular Value Decomposition
gPOE generalized Product Of Experts
rBCM robust Bayesian Committee Machine