Kernel-based Methods for Forecasting of Waste Heat Recovery in Ships

Mikael Manngård*, Amir H. Shirdel, Jari M. Böling, and Hannu T. Toivonen

Faculty of Science and Engineering, Åbo Akademi University, Åbo, Finland

Abstract

Utilization of waste heat is of great interest to the ship industry, and the need for effective data-based learning methods is steadily increasing. Kernel-based models can easily be fitted to data, but models tend to become large as the number of data points grows. Model complexity is hence a problem, and finding sparse solutions is of great importance. We evaluate two identification methods which promote sparsity in the models. A regularization based approach, using ℓ_2-regularization with iterative reweighting is compared to support vector regression. Both methods are trained and validated on simulated data from a marine diesel-engine fresh-water cooling system, for modelling the recovered waste heat energy based on the engine load. No assumptions on the underlying structure of the system are needed, which make the methods flexible and easy to implement in practice.

Keywords: Kernel regression, sparse optimization, iterative reweighting, support vector regression

1 Introduction

The interest of improving energy efficiency in ships is steadily increasing due to uncertain fuel costs and strict international maritime rules. One way to save energy is to schedule on-board tasks such as fresh water evaporation, air conditioning and general heating to times when excess waste heat is available. By doing so the use of additional heating can be reduced. Good forecasts of both available energy and energy demand are required to schedule such tasks optimally during operations. This is a difficult task, since marine vessels are complex systems, which often vary substantially from ship to ship. Our goal is to develop techniques for modelling ship energy systems at a general (black-box) level, which will allow our methods to be applied for any ship, regardless of exact knowledge of the underlying system. Such models can either be used for tracking of heat flows or forecasting.

In this paper we consider an engine high-temperature fresh water cooling system, where the available waste heat energy is to be modelled. Simulated data were obtained from Zou et.al. [12], and was considered to be actual measurements from a real ship with unknown underlying structure. Kernel functions are used to learn the system behaviour from data. Two methods are evaluated for identification of sparse kernel representations. An ℓ_2-regularization based method with iterative reweighting of the regularization parameters have been compared to support vector regression (SVR) [7].

Various regularization techniques which promote sparsity are proposed in literature [2, 3, 4]. Here the ℓ_1-norm is the most commonly used regularizer, due to it promoting sparsity while having convex properties. However, in this article we are using the ℓ_2-norm as a regularizer instead, which makes the problem quadratic and can hence be solved explicitly. The ℓ_2-norm will not in itself promote sparsity, and iterative reweighting and a small threshold which sets small parameters to zero are needed.

It is well known that support vector machines can be used for classification problems and will result in sparse solutions [7]. Support vector machines can easily be extended to regression problems, while preserving its sparse properties [11, 9, 6]. Support vector regression has been applied on a broad range of problems [5, 7, 10, 8], and is hence considered as a benchmark in this article.

*Corresponding author.

E-mail addresses: mikael.manngard@abo.fi (M. Manngård), amir.shirdel@abo.fi (A. H. Shirdel), jari.boling@abo.fi (J. M. Böling), hannu.toivonen@abo.fi (H. T. Toivonen)

The paper is organized as follows. In the Section 2 we describe the basics of kernel functions and how they can be applied to regression. In the Section 3 we give a short theoretical overview of out proposed methods. In Section 4 we evaluate the methods in a case study for forecasting of waste heat recovery in ships.

2 Kernel Functions

Kernel functions are often used for non-linear black-box identification of systems where the underlying model structure is poorly known. Kernel functions can be interpreted as a measure of similarity or distance between features, and are often defined to be real-valued functions, $K(\mathbf{x}, \mathbf{x}') \in \mathbb{R}^{N \times N}$, of input vectors $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^M$, where N is the number of data points and M is the number of inputs [6]. The simplest form of kernel is the linear kernel which is represented by $K(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T\mathbf{x}'$. Another kernel, commonly user for regression, is the Gaussian kernel. The Gaussian kernel is a radial basis function (RBF), which is expressed by

$$K_{i,j}(\mathbf{x}_i, \mathbf{x}'_j) = \exp\left(-\gamma\|\mathbf{x}_i - \mathbf{x}'_j\|^2\right), i, j = 1, ..., N, \tag{1}$$

where $\|\mathbf{x}_i - \mathbf{x}'_j\|$ is the Euclidean distance between the data points and γ is a design parameter defined as $\gamma = \frac{1}{2\tau^2}$, where τ^2 determines the "width" of the kernel (which can be compared to the variance of a Gaussian probability density function). Given a set of data points $\{\mathbf{x}, \mathbf{y}\}$ and a kernel $K(\mathbf{x}, \mathbf{x}')$, a regression model which is linear in the model parameter θ can be defined as

$$\hat{\mathbf{y}} = K(\mathbf{x}, \mathbf{x}')\theta. \tag{2}$$

Since the number of elements in θ equals the number of data points, is the model expected to fit the training data perfectly, and overfitting is hence expected. In order to prevent overfitting, it is necessary to find sparse representations of the kernel function. One way to achieve this is to find sparse solutions for θ, in which many of the elements are zero, and the model complexity is reduced.

3 Promoting Sparsity

An ideal way to determine a sparse θ-vector from Equation (2), is by minimizing the cost function

$$J = \|\mathbf{y} - K(\mathbf{x}, \mathbf{x}')\theta\|_2^2 + \lambda\|\theta\|_0, \tag{3}$$

where $\|\theta\|_0$ is the ℓ_0-norm, which denotes the number of nonzero elements in θ, and λ is a regularization parameter. This problem is, however, a combinatorial optimization problem which cannot be solved efficiently [3]. To address this problem, various relaxations of the ℓ_0-norm, such as using the ℓ_2- or ℓ_1-norms, can be introduced. Although it is well known that the ℓ_1-regularizer promotes sparsity, have we chosen to relax Equation (3) with an ℓ_2-norm, combined with iterative reweighting, to drive some parameters very close to zero. The advantage of using an ℓ_2-regularization in stead of ℓ_1 is that explicit solutions exist. We have compared the iterative reweighting procedure to support vector regression as a benchmark.

3.1 Iteratively Reweighted Least Squares

To make the problem presented in Equation (3) tractable, we wish to relax the ℓ_0-norm with an ℓ_2-norm. This makes the problem quadratic, which makes it possible to find an explicit solution. We introduce a diagonal weight matrix $\mathbf{\Lambda}$, and state the new problem as

$$\text{argmin}_\theta \ \|\mathbf{y} - K(\mathbf{x}, \mathbf{x}')\theta\|_2^2 + \|\mathbf{\Lambda}\theta\|_2^2, \tag{4}$$

which has the explicit solution

$$\hat{\theta} = [\mathbf{\Lambda}\mathbf{\Lambda}^T + K(\mathbf{x}, \mathbf{x}')K(\mathbf{x}, \mathbf{x}')^T]^{-1}K(\mathbf{x}, \mathbf{x}')^T\mathbf{y}. \tag{5}$$

However, the solution will not promote sparsity by itself, and an iterative reweighting procedure for $\mathbf{\Lambda}$ is hence introduced. In order to enhance sparsity, the weight matrix is set to

$$\mathbf{\Lambda}^{(k)} = \text{diag}\left(\frac{\alpha}{|\hat{\theta}^{(k)}| + \epsilon}\right), \tag{6}$$

where k is the iteration index, α is a scaling parameter and ϵ is a small positive parameter which prevents numerical division-with-zero. Since ℓ_2-regularization will not result in sparse solutions (the only way by which $\hat{\theta}_i = 0$ is when the solution to the unregularized problem is zero), a small threshold δ is introduced, so that if $|\hat{\theta}|_i < \delta$, then $\hat{\theta}_i$ is set to zero. The effects of reweighting are presented in a contour plot in Figure 1.

Remark (A Bayesian perspective)**.** *Solving Equation (4) is equivalent to solving a maximum a posteriori (MAP) problem, which is defined as maximizing the posterior distribution $p(\theta|\mathbf{x}, \mathbf{y})$, with respect to θ. The posterior distribution is given by Bayes' theorem*

$$p(\theta|\mathbf{x}, \mathbf{y}) \propto p(\mathbf{y}|\theta, \mathbf{x})p(\theta), \tag{7}$$

where $p(\mathbf{y}|\theta, \mathbf{x})$ is the likelihood function and $p(\theta)$ is a Gaussian prior distribution with zero mean and variance σ_p^2. If $\sigma_p^2 \to \infty$, the MAP solution will approach the maximum likelihood (ML) solution. [1]

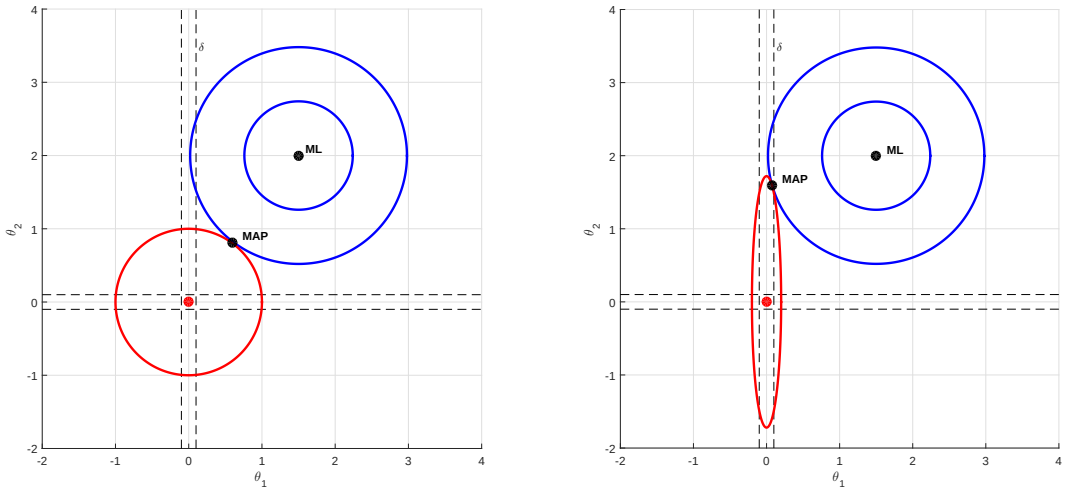

Figure 1: Contours for the unregularized error function (blue) and the regularization function (red). Optimal values for both the regularized (MAP) and unregularized (ML) cases are marked out with black dots. The figure to the right shows the effects of reweighting

It is also worth mentioning that when setting a weight to zero, can the whole corresponding column in $K(\mathbf{x}, \mathbf{x}')$ be removed, and hence reducing the complexity of the model. Our proposed method can be considered a variation of an iteratively reweighted least-squares (IRLS) approach [3]. The complete algorithm is presented below.

Algorithm 1 Iteratively reweighted least squares

Initialize $\epsilon, \delta, \alpha, \mathbf{\Lambda} = \mathbf{0}$
for iteration $k = 1, ..., K$ **do**
  $\hat{\theta}^{(k)} = [\mathbf{\Lambda}^{(k)}\mathbf{\Lambda}^{(k)T} + K(\mathbf{x}, \mathbf{x}')K(\mathbf{x}, \mathbf{x}')^T]^{-1}K(\mathbf{x}, \mathbf{x}')^T\mathbf{y}$
  $\mathbf{\Lambda}^{(k)} = \text{diag}\left(\frac{\alpha}{|\hat{\theta}^{(k)}|+\epsilon}\right)$
  if $\hat{\theta}_i^{(k)} < \delta$ **then**
    $\hat{\theta}_i^{(k)} = 0$
  end if
end for

Although there are several design parameters to initialize, most of them are fairly easy to determine. The parameter ϵ can simply be chosen small enough. A logical way to initialize the $\mathbf{\Lambda}$-parameter is to initialize it to be zero, which would correspond to unregularized solution. The parameter δ is chosen so that small θ_i are cut off. Parameter α is a scaling parameter, which ensure that all weights will not get too heavily regularized, this parameter can be selected through cross validation. Note that the ratio α/ϵ is the largest value an element in $\mathbf{\Lambda}$ can obtain. Depending on which kernel is used, additional parameters may occur. Such parameters are also typically determined through cross validation.

3.2 Support Vector Regression

Support vector machines are powerful techniques for data classification and regression [11], which has been applied to robust black-box system identification. In support vector regression, a regularization term, which is taken as a quadratic function of the parameters, is minimized subject to linear constraints on the prediction error magnitudes. The solution is usually posed in terms of a dual maximization problem, in which a Lagrange function is maximized with respect to the Lagrange multipliers associated with the linear constraints. In support vector regression, a nonlinear model is represented by defining a feature space, which can be expressed in terms of an associated kernel with the same dimension as the data set, and it is therefore not necessary to compute the nonlinear mappings explicitly. Support vector regression has several appealing properties for black-box identification. The solution is defined in terms of a convex quadratic minimization problem, for which convergence to the global solution can be guaranteed. In SVR, the parameter vector $\hat{\theta}$ is determined by solving the problem

$$\min \frac{1}{2}\|\hat{\theta}\|^2 + C\sum_{k=1}^{N}(\xi_k + \xi_k^*) \tag{8}$$

subject to ε-insensitive constraints

$$y(k) - \sum_{l=1}^{N} K_{k,l}(\mathbf{x}_k, \mathbf{x}_l')\hat{\theta}_l \;\leq\; \varepsilon + \xi_k \tag{9}$$

$$-y(k) + \sum_{l=1}^{N} K_{k,l}(\mathbf{x}_k, \mathbf{x}_l')\hat{\theta}_l \;\leq\; \varepsilon + \xi_k^* \tag{10}$$

$$\xi_k, \xi_k^* \;\geq\; 0$$

Here ξ_k, ξ_k^* are slack variables, and C is a design parameter which determines the weight of the prediction errors relative to the first term, which acts as a regularization term. The Lagrange function has a saddle point with respect to the primal and dual variables at the solution associated with Equations (8) and (9) [7]. The support vectors are those data points that lie on the boundary or outside of the ε-tube (i.e. the data points which satisfy constraints (9) and (10) with equality). Those points are the only ones contributing to the prediction model, and hence we have a sparse solution [1]. The number of support vectors is equal to $\|\hat{\theta}\|_0$.

4 Case Study: Forecasting of Waste Heat Recovery in Ships

Kernel functions can be used as black-box models for describing systems where the underlying models structure is either complex or is poorly known. We have evaluated our proposed methods for identifying sparse kernel functions on data from a fresh water cooling system for a marine diesel engine. The data we have used is obtained from a simulator developed by Zou et.al. [12]. Since the structure of engine cooling systems may vary significantly from ship to ship, and information about the exact structures may be hard to obtain, there is a great interest of finding general methods for identifying black-box models based on data.

The main purpose for an engine cooling system is to keep the engine temperature within an acceptable temperature range. Excess heat is extracted from the system through heat exchangers, and it can as an example be used for fresh water generation. Utilizing excess heat from a system is referred to as waste heat recovery. A possible configuration for such a process is presented in Figure 2.

The goal is to describe the input-output dependence between the engine load (input) and recovered heat (output). Approximately 7 hours of data (sampling time 100 seconds) were used for training and validation. A total number of $N = 500$ data points were used. Dynamics are neglected due to the long time horizon and large sampling time. A Gaussian RBF-kernel, described by Equation (1) was used. The iteratively reweighted least-squares approach was initialized with parameters $\gamma = 100, \epsilon = 10^{-16}, \delta = 0.005$ and $\alpha = 0.001$. Parameters γ and α were determined through cross validation. The support vector regression was initialized with parameters $\gamma = 1$, $\varepsilon = 0.004$ and $C = 100$. Results are presented in Figure 3 and Table 1.

Both methods show excellent fit on both training and validation data, and are hence well suited for forecasting of waste heat recovery, given different scenarios of future engine loads. Although the SVR algorithm shows slightly higher validation error, is it likely to be less overfitted to the data due to a significantly lower number of support vectors and smaller gap between training and validation RMSE.

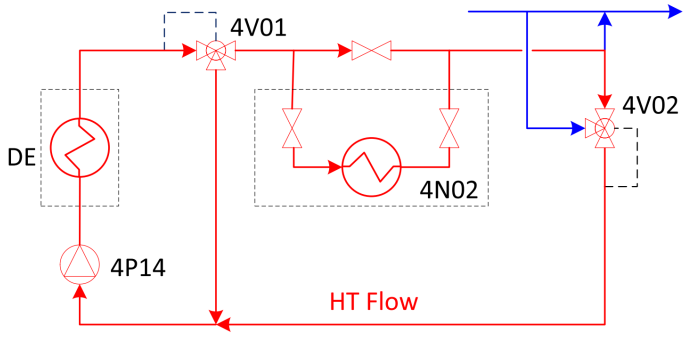

Figure 2: Schematics of a high temperature (HT), engine cooling system, illustrating a possible configuration of a diesel engine (DE), heat exchanger (4N02), and mixing valves (4V01,4V02) [12].

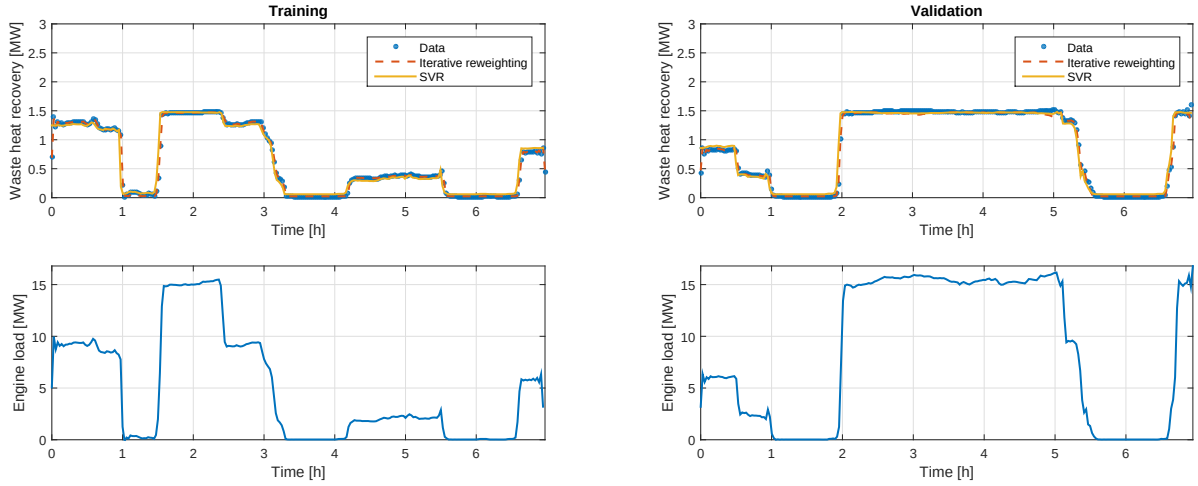

Figure 3: Results on training and validation data for both the iterative reweighting procedure and support vector regression

Table 1: Number of support vectors $\|\hat{\theta}\|_0$ and RMSE on training and validation, for both the iteratively reweighted least squares approach and support vector regression.

	$\|\hat{\theta}\|_0$	RMSE Training	RMSE Validation
IRLS	8	0.0033	0.0085
SVR	3	0.0094	0.0095

5 Conclusions

Two methods for identification of sparse kernel functions have been validated on a simulation study for forecasting of waste heat recovery in ships. An iteratively reweighted least-squares approach have been proposed for finding sparsity in the parameter vector. The approach was able to reduce the total number of parameters from initially 250 to only 8, while maintaining a good fit to the data. The reweighting procedure was compared to support vector regression, which can be considered as a benchmark approach for this type of problem. The support vector regression reduced the total number of parameters to only 3, while maintaining good fit. The obtained sparse solutions are expected not to be overfitted to the training data, and can hence be used for forecasting waste heat recovery in ships, given different scenarios for the engine load. Further work will be directed towards evaluating these methods on larger ship systems containing multiple diesel engines. Since the ℓ_2-norm is clearly not ideal for promoting sparsity, are other relaxations of the ℓ_0-norm to be investigated.

Acknowledgements

This work was carried out in the Efficient Energy Use (EFEU) research program coordinated by CLEEN Ltd. with funding from the Finnish Funding Agency for Technology and Innovation, Tekes. Financial support from Finnish Graduate School in Chemical Engineering (GSCE) is also acknowledged.

References

[1] Bishop, C. M. et al. (2006). *Pattern recognition and machine learning*. Springer.

[2] Candes, E. J., Wakin, M. B., and Boyd, S. P. (2008). Enhancing sparsity by reweighted l_1 minimization. *Journal of Fourier analysis and applications*, 14(5-6):877–905.

[3] Elad, M. (2010). *Sparse and redundant representations: from theory to applications in signal and image processing*. Springer Science & Business Media.

[4] Gorodnitsky, I. F. and Rao, B. D. (1997). Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. *Signal Processing, IEEE Transactions on*, 45(3):600–616.

[5] Müller, K.-R., Smola, A. J., Rätsch, G., Schölkopf, B., Kohlmorgen, J., and Vapnik, V. (1997). Predicting time series with support vector machines. In *Proceedings, International Conference on Artificial Neural Networks*, pages 999–1004. Springer.

[6] Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

[7] Schölkopf, B. and Smola, A. J. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.

[8] Shirdel, A. H., Björk, K.-M., and Toivonen, H. T. (2014). Identification of linear switching system with unknown dimensions. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on*, pages 1344–1352. IEEE.

[9] Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222.

[10] Tötterman, S. and Toivonen, H. T. (2009). Support vector method for identification of wiener models. *Journal of Process Control*, 19(7):1174–1181.

[11] Vapnik, V. (1995). *The nature of statistical learning theory*. Springer.

[12] Zou, G., Elg, M., Kinnunen, A., Kovanen, P., Tammi, K., and Tervo, K. (2013). Modeling ship energy flow with multi-domain simulation. In *CIMAC Congress, Shanghai*. CIMAC.