Robust Diagnostics for Bayesian Compressive Sensing with Applications to Structural Health Monitoring

Yong Huang*[a,b], James L Beck[b], Hui Li[a] , Stephen Wu[b]

[a]School of Civil Engineering, Harbin Institute of Technology, Harbin, China150090;

[b]Division of Engineering and Applied Science, California Institute of Technology, Pasadena CA,USA 91125

ABSTRACT

In structural health monitoring (SHM) systems for civil structures, signal compression is often important to reduce the cost of data transfer and storage because of the large volumes of data generated from the monitoring system. Compressive sensing is a novel data compressing method whereby one does not measure the entire signal directly but rather a set of related ("projected") measurements. The length of the required compressive-sensing measurements is typically much smaller than the original signal, therefore increasing the efficiency of data transfer and storage. Recently, a Bayesian formalism has also been employed for optimal compressive sensing, which adopts the ideas in the relevance vector machine (RVM) as a decompression tool, such as the automatic relevance determination prior (ARD). Recently publications illustrate the benefits of using the Bayesian compressive sensing (BCS) method. However, none of these publications have investigated the robustness of the BCS method. We show that the usual RVM optimization algorithm lacks robustness when the number of measurements is a lot less than the length of the signals because it can produce sub-optimal signal representations; as a result, BCS is not robust when high compression efficiency is required. This induces a tradeoff between efficiently compressing data and accurately decompressing it. Based on a study of the robustness of the BCS method, diagnostic tools are proposed to investigate whether the compressed representation of the signal is optimal. With reliable diagnostics, the performance of the BCS method can be monitored effectively. The numerical results show that it is a powerful tool to examine the correctness of reconstruction results without knowing the original signal**.**

Keywords: Bayesian Compressive Sensing, data compression, structural health monitoring, relevance vector machine, automatic relevance determination, robust diagnostics

1. INTRODUCTION

Structural health monitoring (SHM) systems are an active and well-established research area. They seek to detect damage and predict remaining service life in civil structures through structural sensor networks. However, the complexity and large scale of civil structures induce large monitoring systems, often with hundreds to thousands of sensor nodes. For example, For example, approximately 800 sensors were installed on the Wind and Structural Health Monitoring System (WASHMS) on Tsing Ma, Ting Kau, and Kap Shui Mun bridges in Hong Kong, China [1]. A large amount of sensor data is usually produced, especially when the long lifetime of a structure is considered. Therefore, innovative sensor data compression techniques are necessary to reduce the cost of signal transfer and storage generated from such a large-scale sensor network.

Several novel data compression techniques for SHM systems, especially for wireless sensor networks have been developed, including wavelet-based compression techniques [2-4] and Huffman lossless compression technique [5]. However, all these methods belong to traditional framework where the sampling rate satisfies the conditions of the Nyquist–Shannon theorem.

*yonghuan@caltech.edu

Smart Sensor Phenomena, Technology, Networks, and Systems 2011,
edited by Wolfgang Ecke, Kara J. Peters, Theodore E. Matikas, Proc. of SPIE Vol. 7982
79820J · © 2011 SPIE · CCC code: 0277-786X/11/$18 · doi: 10.1117/12.880687

Recent results demonstrate that sparse or compressible signals can be directly acquired at a rate significantly lower than the Nyquist rate, using a novel sampling strategy called compressive sensing (CS) [6-10]. Much smaller number of measurements is obtained by projecting the signal onto a small set of vectors, which are incoherent with respect to the basis vectors that give a sparse representation of the signal. This new technique may come to underlie procedures for sampling and compressing data simultaneously, therefore increasing the efficiency of data transfer and storage. The signals can be subsequently recovered from the acquired measurements by a special decompression algorithm. Recently, there has been much interest to extend the CS method, including Bayesian CS (BCS) [11], distributed CS [12] and 1-BIT CS [13].

As a damage detection and characterization strategy for civil structures, SHM systems should guarantee that the reconstructed signal from data compression is accurate. For compressive sensing based on solving a convex optimization problem, a condition on the number of measurements needed for perfect reconstruction of signal is provided by Candes et al. [14], Donoho et al. [15] and Baraniuk et al. [16]. However, they require the assumption that the sparsity of the original signal is known in advance.

A BCS technique has been proposed by employing a Bayesian formalism to estimate the underlying signal based on CS measurements. This technique adopts the ideas of the automatic relevance determination (ARD) prior in the relevance vector machine (RVM) as a decompression tool [21, 22]. One of the benefits of using the BCS method is that measurement of confidence in the inverted signal is given through variances on each reconstructed data point. He and Carin [17] pointed out that these variances could be used to determine whether a sufficient number of CS measurements have been performed, although the exact strategy to implement it has not been presented.

In this paper, the robustness property of the BCS technique is studied; diagnostic tools are proposed to investigate whether the compressed representation of the signal is optimal. A set of numerical results are used to validate the proposed methods.

2. BAYESIAN COMPRESSIVE SENSING

Consider a discrete-time signal $x = [x(1), \cdots x(N)]^T$ in $\mathbb{R}$ represented in terms of an orthogonal basis as

$$x = \sum_{n=1}^{N} w_j \, \Psi_j \text{ or } x = \Psi w \tag{1}$$

where $\Psi = [\Psi_1, \cdots, \Psi_N]$ is the $N \times N$ basis matrix with the orthonormal basis of $N \times 1$ vectors $\{\Psi_n\}_{n=1}^{N}$ as columns; w is a sparse vector, i.e., most of its components are zero or very small (with minimal impact on the signal); the location of the nonzero components of w is referred to as the model indices and the number of them represent the sparsity of the model, representation of the signal x in Equation (1)

In the framework of CS, one infers the coefficients w_j of interest from compressed data instead of directly sampling the signal x. The compressed data vector y is composed of K individual measurements obtained by linearly projecting the signal x using a chosen random projection matrix Φ:

$$y = \Phi x + n = \Phi \Psi w + n = \Theta w + n \tag{2}$$

where $\Theta = \Phi\Psi$ is known and n represents the acquisition noise which is modeled as a zero-mean Gaussian vector with covariance matrix $\sigma^2 I_K$. Since Φ is a $K \times N$ matrix with $K \ll N$, to give good data compression, the inversion to find the signal x is ill-posed.
However, by exploiting the sparsity of the representation of x in basis $\{\Psi_n\}_{n=1}^{N}$, the ill-posed problem can be solved by an optimization formulation to estimate w,

$$\widetilde{w} = arg \ min\{\|y - \Theta w\|_2^2 + \rho\|w\|_1\} \tag{3}$$

where the parameter ρ balances the trade-off between the first and the second expressions in the equation, i.e. between how well the data is fitted and how sparse the signal is [14,15].

The ill-posed data inverse problem can also be tackled using a Bayesian perspective, which has certain distinct advantages compared to previously published CS inversion algorithms such as linear programming [18] and greedy algorithm [19-20]. It also provides a sparse solution to estimate the underlying signal, and it provides a measure of the uncertainty for the reconstructed signal, in particular.

Ji et al. [11] adopt the ideas in the relevance vector machine (RVM) proposed in [21-23] for regression that was RVM method is implemented by using the ARD prior that not only allows regularization of the inversion but also controls model complexity by automatically selecting the relevant regression terms to give a sparse representation of the signal. The characteristic feature of this prior is that there is an independent hyperparameter for each basis coefficient:

$$\mathrm{p}(\mathrm{w}|\alpha_1, \alpha_2, \ldots \alpha_N) = \prod_{n=1}^{N} \left[(2\pi)^{-1/2}(\alpha_n)^{1/2}\exp\left\{-\frac{1}{2}(\alpha_n)\mathrm{w}_n^2\right\}\right] = \mathrm{N}\left(0, \mathrm{diag}(\alpha_1^{-1}, \ldots, \alpha_N^{-1})\right) \quad (4)$$

where hyperparameterα_n is the inverse of the prior variance of each coefficient w_n.

The hyperparameter α are selected by maximizing the Bayesian evidence (marginal likelihood):

$$\mathrm{p}(\mathrm{y}|\alpha) = \int \mathrm{p}\,(\mathrm{y}|\mathrm{w})\mathrm{p}(\mathrm{w}|\alpha)\mathrm{dw} \quad (5)$$

where the likelihood function is given by:

$$\mathrm{p}(\mathrm{y}|\mathrm{w}) = \mathrm{N}(\Theta\mathrm{w}, \sigma^2 \mathrm{I}_K) \quad (6)$$

because of the Gaussian model for the noise in Equation (2). Maximizing the evidence penalized signal models that are too simple or too complex and this has an interesting information-theoretic interpretation [24].

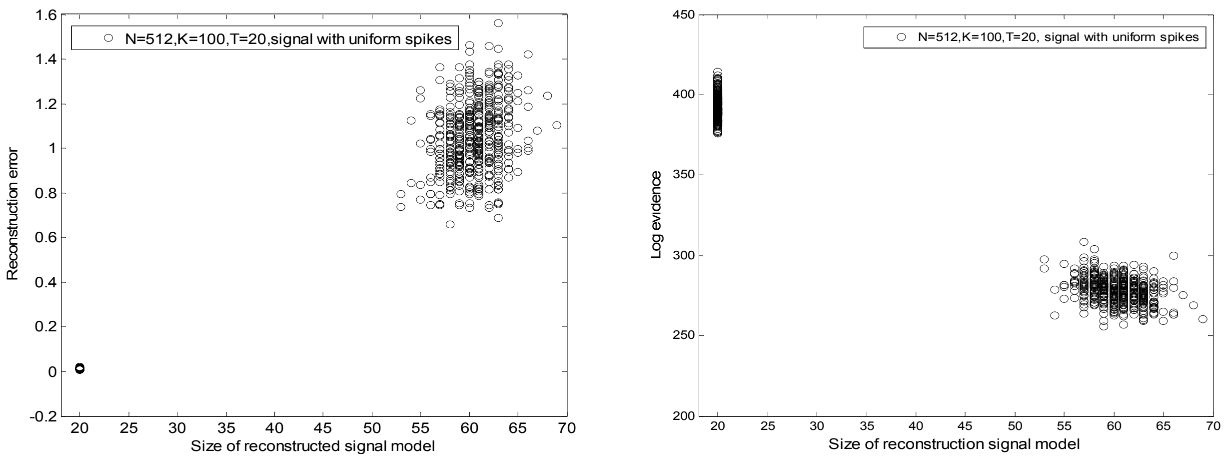
3. ROBUSTNESS PROBLEM FOR BCS RECONSTRUCTION

In order to increase the efficiency of compressive sensing, the number of the measurements must be reduced to be much smaller than the length of the original signal. As a result, there is a large number of local maximas in the evidence over α that can trap the optimization [25].
Figure 1 illustrates the problem by showing some results from the fast optimization algorithm in [23] to different samples of the data y from the same signal x by choosing different random projection matrices Φ, while the orthogonal basis Ψ was held fixed, the signal of length $N = 512$ consisted of $T = 20$ uniformly spaced spikes. The reconstruction error and converged value of the log evidence are plotted against the size of the final reconstructed signal (a constant noise variance σ^2 was used that was 10% of the 2-norm of the measurement y). The figure shows only a few of the optimization runs produce the correct signal size of 20 corresponding to the global maximum of the log evidence near 400. Most of the runs give local maxima of the evidence that correspond to larger amounts of non-zero signal components.

It is also found that the global maximum of the evidence corresponds to the most stable signal model; this is almost the only one model in the adjacent area. The signal models with local maximas of the evidence are much less stable; they have the potential to converge to other models if perturbed slightly and re-run.

4. ROBUST DIAGNOSTIC METHODS FOR BCS RECONSTRUCTION

We conclude that the BCS technique lacks robustness when high compression efficiency is required, and so a robustness diagnostic which works without knowing of the original signal is necessary. Such robust diagnosing methods are proposed here that utilize the different level of stability between the reconstructed signals that correspond to global and local maxima of the evidence. The signal indices of the previous final reconstructed signal are used to initialize a new diagnostic process. This process tries to perturb the original converged result to another local maximum based on the various strategies described next. If the current local maximum is actually the global maximum, then this process will not make any changes in the results, revealing that the optimum reconstructed signal has been achieved.



(a) (b)

Figure 1. Reconstruction error and log evidence versus size of the final models

4.1 Change in noise variance σ^2

The selection of the noise variance σ^2 may affect algorithm significantly due to the underdetermined nature of the inverse problem. The noise variance σ^2 has significant influence on the trade-off between how well the signal model fits the data and how sparse it is.

The sparsity of the final reconstructed signals versus noise variance σ^2 is investigated as shown in Figure 2. The larger σ^2, sparser is the final reconstructed signal. This is because larger noise variance σ^2 reduces the information to support the more complicated reconstructed signals.

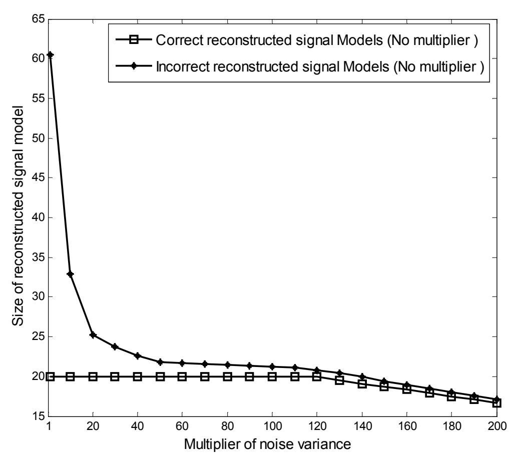


Figure 2. Plot of sparsity as a function of increasing noise variance σ^2 (Different training process with fixed σ^2, the same original signal and the same projection matrix)

4.2 Change in indices of the signal model

Another method to perturb the converged solution for the reconstructed signal is to delete or add some indices to the signal model (i.e. add or delete non-zero coefficients in w). Since the signals corresponding to local maximum of evidence are less stable than that for the global maximum, changing the indices of the signal with a local maximum will lead to convergence to a new local maximum, showing that the reconstructed signal is sub-optimal.

4.3 Summary of robust diagnostic procedure

The method employs signal indices of the previous final reconstructed model to initialize a new diagnostic learning process and the initial values of the hyperparameter α for every signal index are re-estimated. There are four diagnostic steps shown below (the value of the noise variance for the last iteration of original training process is σ_{or}^2):

(1). Initialize with 80% of the signal indices obtained from first trial (randomly choosing), take noise variance $\sigma_{\mathrm{new},1}^2$ to be twice σ_{or}^2, then perform the second trial with constant $\sigma_{\mathrm{new},1}^2$.

(2). Initialize with all signal indices obtained from first trial and 20% of extra random signal indices, take noise variance $\sigma_{\mathrm{new},2}^2$ as half of σ_{or}^2, then perform the second trial with constant $\sigma_{\mathrm{new},2}^2$.

(3). Set the final signal of the second trial of step 1 as the original signal model and perform step 2 again.

(4). Set the final signal of the second trial of step 2 as the original signal model and perform step 1 again.

This procedure for diagnosis detects incorrect reconstruction signals when either the original signal model (non-zero coefficients in w) is not included in the final signal model or the final signal model is not included in the original one for the diagnosing steps. If there is at least one diagnosis step that shows the reconstructed signal is "incorrect", the results can be judged as an incorrect reconstructed signal.

4.4 Numerical validations for robust diagnosis

For the numerical validations of the robust diagnosis procedure, 1000 runs are performed, as shown in Figure 4. Signals with length $\mathrm{N} = 512$ and non-zero spikes created by choosing $\mathrm{T} = 20$ discrete times at random are considered; the non-zero spikes of the signals are drawn from two different probability distributions, one is uniform ± 1 random spikes, and the other one is zero-mean unit variance Gaussian spikes (Figure 3). Uniform random projection is employed to construct the projection matrix Φ. In the experiment, the number of measurements is $\mathrm{K} = 60$ for signals with non-uniform spikes and $\mathrm{K} = 90$ for signals with uniform spikes. For the BCS implementation [11], we used the bcs_ver0.1 package available online at http://people.ee.duke.edu/~lcarin/BCS.html, and the BCS parameters were set as those suggested by bcs_ver0.1. Original signal reconstruction is defined as incorrect if the actual reconstruction error is larger than 0.01, which we can compute because in the test the original signal is known. The trials are sorted to better visualize the accuracy of different diagnostic processes. Figure 4(a) and (b) show that the robust diagnostic procedure is a powerful tool to examine the correctness of reconstruction results without knowing the original signal.

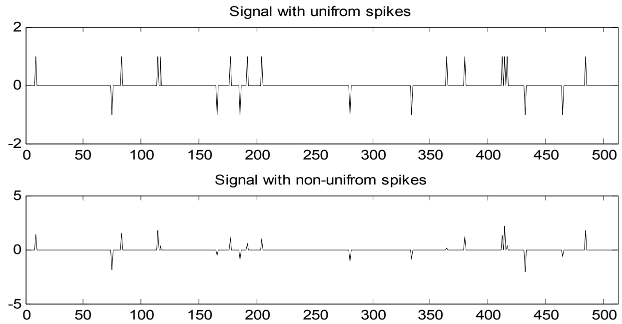
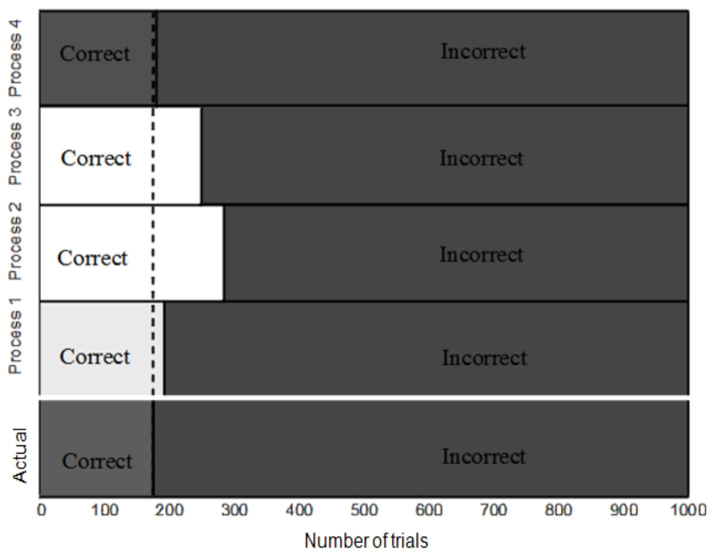
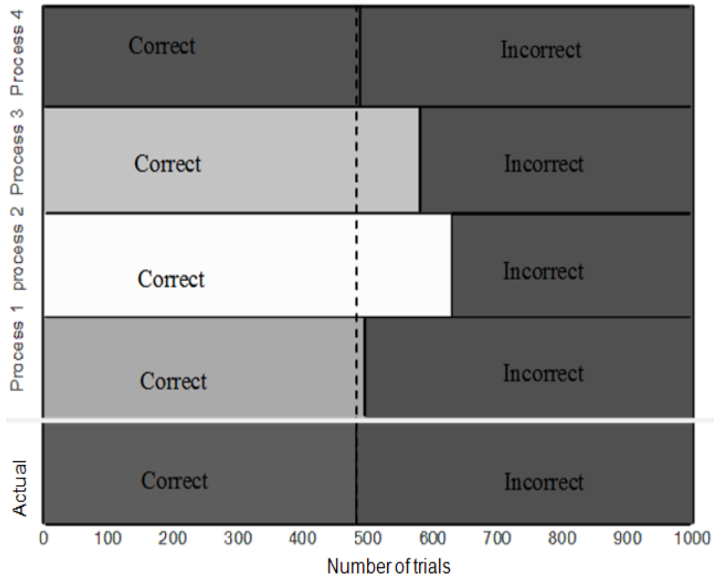


Figure 3. Signal with uniform spikes and non-uniform spikes



(a) N=512, T=20, K=90; Signal with uniform spike



(b) N=512, T=20, K=60; Signal with non-uniform spike

Figure 4. Results of the robust diagnosis

5. CONCLUSION

The robustness of the BCS method is investigated. We show that the optimization algorithm of RVM lacks robustness when the number of measurements is a lot less than the length of the signals because it often converges to sub-optimal signal representations that are local maxima of evidence, so BCS is not robust when high compression efficiency is required.

Based on a study of the robustness of the BCS method, diagnostic tools are proposed to investigate whether the compressed representation of the signal is optimal. With reliable diagnostics, the performance of the BCS method can be monitored. Numerical results based on simulating SHM signals show that it is a powerful tool to examine the correctness of reconstruction results without knowing the original signal.

ACKNOWLEDGMENTS

One of the authors (Yong Huang) acknowledges the support provided by the China Scholarship Council while he was a Visiting Student Researcher at the California Institute of Technology. This research is also supported by grants from National Natural Science Foundation of China (NSFC grant nos. 50538020, 50278029 and 50525823), which supported the first and third authors (Yong Huang and Hui Li).

REFERENCES

[1] Ko, J. M., "Health monitoring and intelligent control of cable-supported bridges". Proceedings of the International Workshop on Advanced Sensors, Structural Health Monitoring and Smart Structures, Keio University, Tokyo, Japan, 10–11 (2003).

[2] Zhang, Y. and Li, J., "Wavelet-based vibration sensor data compression technique for civil infrastructure condition monitoring," Journal of Computing in Civil Engineering 20(6), 390-399 (2006).

[3] Xu, N., Rangwala, S., Chintalapudi, K., Ganesan, D., Broad, A., Govindan, R. and Estrin, D.. "A wireless sensor network for structural monitoring," Proceedings of the ACM Conference on Embedded Networked Sensor Systems. Baltimore, MD, USA (2004).

[4] Caffrey, J., Govindan, R., Johnson, E., Krishnamachari, B., Masri, S.,Sukhatme, G., Chintalapudi, K., Dantu, K., Rangwala, S., Sridharan, A., Xu, N. and Zuniga, M., "Networked Sensing for Structural Health Monitoring," Proceedings of the 4th International Workshop on Structural Control, New York, NY, 57–66 (2004).

[5] Lynch, J. P., Sundararajan, A., Law, K. H, Kiremidjian, A. S., and Carryer, E. "Power-efficient data management for a wireless structural monitoring system," Proceedings of the 4th International Workshop on Structural Health Monitoring, Stanford, CA, 1177–1184 (2003).

[6] Candes, E. J. "Compressive sampling," Proceedings of the International Congress of Mathematicians, Madrid, Spain, 1433–1452 (2006).

[7] Donoho, D., "Compressed sensing," IEEE Transactions on Information Theory 52(4), 1289–1306 (2006).

[8] Candes, E. J. and Tao, T., "Near-optimal signal recovery from random projections: Universal encoding strategies?" IEEE Transactions on Information Theory 52(12), 5406–5425 (2006).

[9] Candes, E. J. and Tao, T., "Decoding by linear programming," IEEE Transactions on Information Theory 51, 4203–4215 (2005).

[10] Donoho, D. "For most large underdetermined systems of linear equations, the minimal l_1 norm solution is also the sparsest solution," Communications on Pure and Applied Mathematics, 59(6), 797–829 (2006).

[11] Ji, S., Xue, Y. and Carin L., "Bayesian compressive sensing," IEEE Transactions on Signal Processing, 56(6), 2346 - 2356 (2008)

[12] Baron, D., Duarte, M. F., Wakin, M. B., Sarvotham, S. and Baraniuk, R.G., "Distributed compressive sensing," Preprint (2005)

[13] Boufounos, P. and Baraniuk, R., "1-Bit compressive sensing," Conf. on Info. Sciences and Systems (CISS), Princeton, New Jersey (2008)

[14] Candes, E. J., Romberg, J. and Tao, T., "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," IEEE Transactions on Information Theory, 52(2), 489-509 (2006).

[15] Donoho, D.,"Compressed sensing," IEEE Trans Inform Theory 52(4), 1289-1306 (2006).

[16] Baraniuk, R., Davenport, M., DeVore, R. and Wakin, M.. "A simple proof of the restricted isometry property for random matrices". Constructive Approximation, 28(3),253-263, (2008).

[17] He, L. and Carin L., "Exploiting Structure in Wavelet-Based Bayesian Compressive Sensing," IEEE Transactions on Signal Processing, 57(9), 3488-3497 (2009).

[18] Chen, S. S., Donoho, D. L. and Saunders, M. A., "Atomic decomposition by basis pursuit," SIAM Journal on Scientific and Statistical Computing, 20(1), 33–61(1999).

[19] Tropp, J. A. and Gilbert, A. C., "Signal recovery from random measurements via orthogonal matching pursuit, " IEEE Transactions on Information Theory, 53(12), 4655-4666 (2007).

[20] Donoho, D. L. and Tanner, J., "Sparse nonnegative solution of underdetermined linear equations by linear programming," Proceedings of the National Academy of Sciences, 102(27), 9446-9451 (2005).

[21] Tipping, M. E., "Sparse Bayesian learning and the relevance vector machine". Journal of Machine Learning Research 1, 211-244 (2001).

[22] Tipping, M. E., "Bayesian inference: An introduction to Principles and practice in machine learning." Advanced Lectures on Machine Learning 3176, 41-62 (2004)

[23] Tipping, M. E. and Faul, A. C., "Fast marginal likelihood maximisation for sparse Bayesian models," Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics, Key West, FL (2003).

[24] Beck, J. L., “Bayesian system identification based on probability logic,” Structural Control and Health Monitoring, 17 (7), 825-847(2010).

[25] Faul, A. C. and M. E. Tipping, "Analysis of sparse Bayesian learning," Advances in Neural Information Processing Systems 14, 383–389 (2002).