APPLICATION OF THE MINIMUM FUEL NEURAL NETWORK TO MUSIC SIGNALS

Anders la Cour-Harbo, member, IEEE

Department of Control Engineering
Aalborg University, Denmark
alc@control.auc.dk

ABSTRACT

Finding an optimal representation of a signal in an over-complete dictionary is often quite difficult. Since general results in this field are not very application friendly it truly helps to specify the framework as much as possible. We investigate the method Minimum Fuel Neural Network (MFNN) for finding sparse representations of music signals. This method is a set of two ordinary differential equations. We argue that the most important parameter for optimal use of this method is the discretization step size, and we demonstrate that this can be a priori determined. This significantly speeds up the convergence of the MFNN to the optimal sparse solution.

1. INTRODUCTION

Automatic processing of music signals is a key component in applications such as recognition, classification, thumb-nailing, watermarking, and transcription of music. This processing forms the basis for making high level decisions such as what type or category the music belongs to, which 10 seconds of the music is most descriptive, which instruments or notes are present in the music, and so on. The common factor in all these applications is the need for feature extraction, which basically means some sort conversion from the music signal to a feature vector.

1.1. Linear transforms

There exists quite a few different methods for feature extraction. A significant part of the methods is the linear transforms, which include the Fourier and wavelet transforms as well as all sorts of filtering.

In recent years a new tool has emerged in this category; redundant representations or over-complete dictionaries. The basic idea is to represent the signal by M elements chosen from a dictionary with N atoms, where M is the length of the signal and $N \gg M$. Since this allows for many different valid choices the optimality is given by a measure of 'goodness' of the choice. While this idea is an extension of the traditional 'sufficient' linear transform, and therefore in theory is at least as good, it is often difficult (and extremely measure dependent) to determine which choice of atoms is optimal.

1.2. Sparseness measures

An often useful measure is sparseness. That means to what extend the majority of signal energy can be represented by few atoms. This can be measured in several ways, and a typical measure is ℓ_p norm, which is also the one used in this presentation. That is, we are interested in solving the problem

$$\min \|\mathbf{x}\|_p \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}\,, \quad 0 \le p \le 1 \tag{1}$$

with $\mathbf{A} \in \mathbb{R}^{M\times N}$ being the dictionary (atoms as columns), $\mathbf{b} \in \mathbb{R}^M$ the signal to be represented, and $\operatorname{rank}(\mathbf{A}) = M \le N$. The ratio N/M is the redundancy factor.

The perhaps most obvious choice for the sparseness measure is the ℓ_0 norm since this measures the number of non-zero entries. Unfortunately, finding the solution to (1) for $p = 0$ is in general NP hard, and thus not feasible for even moderately sized problems. However, there is a series of results on the relations between solutions to (1) for varying p. In particular, it has been shown [1, 2] that under some conditions imposed on the dictionary the solution to (1) is the same for $p = 0$ and $p = 1$ (similar results exists for more general sparseness measures [3]). While these conditions are rarely met these results are nonetheless encouraging, because while no known method exists for analytically determining the solution to (1), there do exists methods for iteratively approximating the solution $\mathbf{x}$ when $p = 1$. Four examples are linear programming [4], quadratic programming [5], minimum fuel neural networks [6], and FOCUSS [7]. Sub-optimal solutions can be obtained for instance by the pseudo inverse (also called Moore Penrose inverse, method of frames [8]; they solve the problem for $p = 2$), various types of matching pursuit [9, 10], and best orthogonal basis (like cosine packets [11], wavelet packets [12] etc.).

As mentioned previously it is often a challenge to actually find the minimizer $\mathbf{x}$, and the challenge varies with the choice of method, dictionary, measure, and signals. In this presentation the focus is on how to efficiently apply the MFNN with a variety of dictionaries for finding the minimizer when the sparseness measure is the ℓ_1 norm and the signal class is music signals.

2. METHODS

The MFNN approach is a set of two non-linear ordinary differential equations (ODE). This was first introduced in [13], and further developed to the following equations in [6].

$$\begin{aligned} \frac{d\mathbf{x}}{dt} &= -\mathbf{A}^\top(\mathbf{A}\mathbf{x} - \mathbf{z} - \mathbf{b}) - P(\mathbf{x} + \mathbf{A}^\top\mathbf{b}) \\ \frac{d\mathbf{z}}{dt} &= -\mathbf{A}\big(\mathbf{x} + \mathbf{A}^\top\mathbf{z} - P(\mathbf{x} + \mathbf{A}^\top\mathbf{b})\big) + \mathbf{b}\,, \end{aligned} \tag{2}$$

This work is supported by the Danish Technical Science Foundation (STVF) grant no. 56-00-0143 (WAVES)

where $\mathbf{z} \in \mathbb{R}^M$, and

$$P(\mathbf{x})_n = \begin{cases} 1 & \text{if } x_n > 1 \\ x_n & \text{otherwise} \\ -1 & \text{if } x_n < -1 \ . \end{cases}$$

Finding the solution to these equations is equivalent to finding the solution to (1) for $p = 1$. Moreover, (2) is globally Lyapunov stable and globally converges to the exact solution.

The ODE are referred to as a neural network because they can be implemented as a number of adders with weighting functions. In other words the equations contain only multiplication and addition. This means that not only is matrix inversion avoided (in contrast to some of the other methods mentioned in the introduction), but it also allows for the use of any fast implementation such as fast trigonometric transforms, fast wavelet transforms, etc.

2.1. Discretization

The main challenge in solving (2) is to find the optimal discretization step for iteratively determining a solution to the continuous time ODE. Or equivalently, to determine the largest step size for which the iterative procedure converges.

A selection of ODE solvers are readily available in various numerical processing tools, and in the authors experience they do solve (2). However, such solvers apply to much more general ODE, and consequently converge much slower (typically one or two magnitudes) than a brute force implementation of the two equations with the right choice of step size. The speed of convergence is of interest in this context because the music signals are often quite long, and the processing time for each iterations grows as $O(N^2)$ (or $O(N \log N)$ for most fast implementations).

The process of obtaining convergence by adjusting the step size can be quite time consuming (it can easily take up a majority of the total processing time), and it would therefore be quite useful to be able to determine a good step size prior to iterating for a solution. This is because iterating with an a priori given step size, though not optimal, is often faster in terms of convergence than letting any given solver determine the optimal step size. This is even more so if we are able to a priori determine a near-optimal step size.

2.2. General step size

The class of music signals is a very small subset of the set of structured signals. And although this subset eludes any rigorous characterization it is nonetheless more sparse when represented in some joint time-frequency-related dictionary than the music signal itself. The hope is therefore that the behaviour of (2) on this subset with TF dictionaries is fairly even.

To test this hypothesis and to provide simple guidelines for determining the near-optimal step size we have performed a test involving excerpts from ten music pieces. We have determined an approximate optimal step size for a total of 2800 excerpts. This involves ten music pieces, ten different starting locations distributed evenly throughout each music signal, seven different signal length at each location, and four different dictionaries. The hypothesis is that the step size various with signal length and choice of dictionary, but not with choice of music and location. Excerpts are taken evenly throughout each music signal.

Table 1. The 10 music pieces.

1	Aqua	Freaky Friday
2	Beethoven	Piano sonate no. 8, 3rd mov
3	Bjork	Army of me
4	Eurythmics	Love is a stranger
5	Jean-Michel Jarre	Chronologie part 4
6	Joe Satriani	The Extremist
7	Lisa Ekdahl	Sunny Weather
8	Mike Oldfield	Islands
9	Orff	Carmina Burana
10	Paul Simon	Bridge over Troubled Water

Table 2. Description and redundancy factor of the dictionaries.

1	Cosines packets with six levels (incl. original signal)	6
2	Wavelet packets with seven levels (incl. original signal) based on Coiflets with three vanishing moments [14]	7
3	Union of the CP and WP dictionaries	13
4	Union of the frequency critically sampled sinusoid IV and the Dirac (Kronecker) basis.	2

The ten music pieces are listed in Table 1, and the four dictionaries are listed in Table 2. The seven different signal lengths are 128, 256, 512, 1024, 2048, 4096, and 8192.

3. RESULTS

The test set described above produces a total of $4 \times 10 \times 10 \times 7 = 2800$ step sizes. To test whether the step size is correlated with choice of dictionary and length of excerpt the values have been divided into 28 groups where each group corresponds to a particular choice of dictionary and excerpt length. The result of this grouping is shown in Figure 1. To test whether the step size is correlated with choice of music the values have been divided into 10 groups, one for each music signal, and the result is shown in Figure 2. Finally, to test for correlation with excerpt location the step sizes have been plotted as a regular graph for two dictionaries and three excerpt lengths. These graphs are shown in Figure 3. Note that no grouping is performed due to the fact that equal excerpt locations in two different music signals does not qualitatively relate the two excerpts.

3.1. Interpretation

It is not hard to tell from Figure 1 that the optimal step size varies with in a fairly structured and predictable way with the length of the music signal. In particular, there seems to be a logarithmic relationship between median and excerpt length, at least for the longer excerpts. This might be expected since this indicates that the stability of the MFNN is related linearly to the energy of $d\mathbf{x}/dt$.

Comparing the scaling of the four plots also reveals that the step size depends on the choice of dictionary, at least in the sense that the maximum step size for each dictionary and for any of the given lengths is roughly inversely proportional to the redundancy factor.

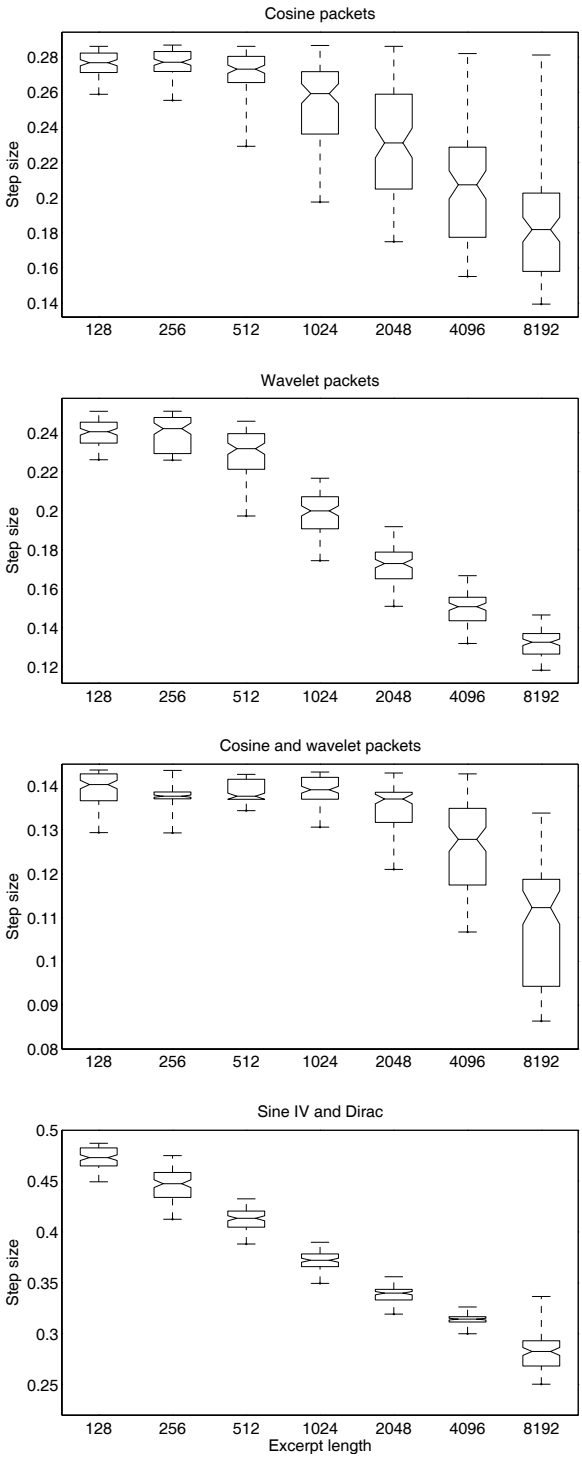

Fig. 1. Box plots for each dictionary. The boxes show the median (the middle notch) and the second and third quartiles for the distribution of the step sizes. The whiskers show the first and fourth quartiles. Each box is based on step sizes for all music signals and all locations (a total of 100 step sizes for each box).

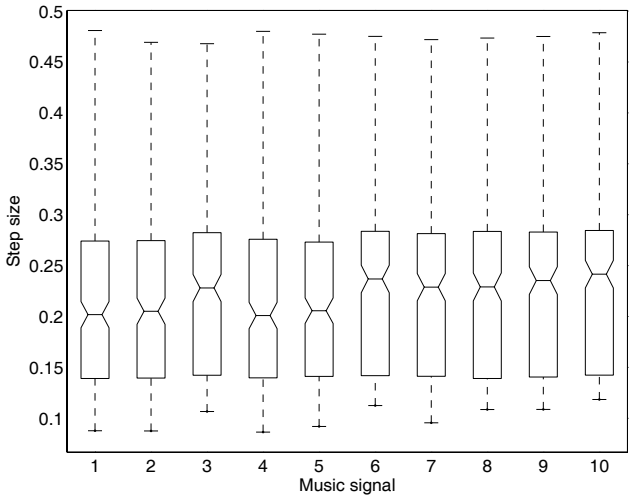

Fig. 2. Box plots for each music signals. The boxes have the same meaning as in Figure 1. Each box is based on step sizes for all locations, lengths, and dictionaries (a total of 280 step sizes for each box)

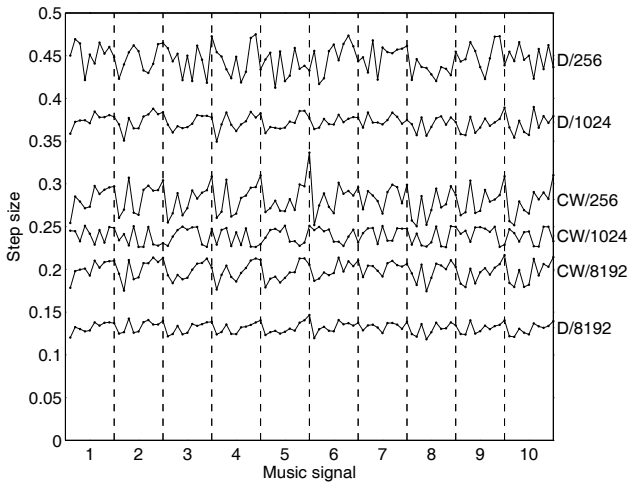

Fig. 3. The step size for three chosen lengths from the two dictionaries CP/WP, and Dirac/sine IV. For each music signal, numbered along the first axis, the step sizes corresponding to the 10 excerpt locations are shown sequentially from left to right in each column.

Turning to Figure 2 it is clear that the dependency of the step size on the choice of music is much less significant than on the choice of length and dictionary.

Note that the the side-by-side comparison of the boxes in Figure 1 and 2 is a graphical equivalent of a t test. Since the step sizes are fairly close to being normally distributed (this cannot be verified visually in these plots, though) it is reasonable to apply the t test, which clearly shows that for Figure 1 virtually all 28 groups are drawn from different normal distributions, while for Figure 2 the samples are drawn from the same distribution.

Finally, Figure 3 shows the step size for each location for each music signal for a selection of three lengths from two dictionaries. While the step size do vary with locations, it is not significant compared to the variations between dictionaries, or indeed between different excerpt lengths. The shown graph is representative for the majority of the remaining 2200 step sizes.

3.2. Available implementation

Prior to this test the MFNN was implemented in Matlab. This implementation takes advantage of the possibility for fast transforms. It also includes some means for detecting divergence, and thus allows for determination of the optimal step size. The Matlab code of the MFNN is available at `www.control.auc.dk/~alc` under Homemade.

4. CONCLUSION

The challenge is to find a solution to (1) for $p = 1$ by using the MFNN. This requires solving a set of two non-linear ODE. It was argued that in terms of convergence speed it it quite useful to be able to a priori determine the near-optimal step size for the discretization of the MFNN ODE. It has been demonstrated in this presentation that at least for musical signals this is possible to some extent. In particular, it was demonstrated that the step size does not dependent on the choice of music, or on which part of the music the solution of (1) is desired. Further, an implementation of a solver for the MFNN ODE is made available.

It is important to realize that there is no minimal-optimal step size for music signals. The step size estimated above will be useful in the sense that in a majority of music excerpts the convergence of the MFNN is near-optimal. And in those few cases where the algorithm diverges nonetheless the proposed implementation will detect this and act accordingly, but at the expense of significantly longer processing time.

5. REFERENCES

[1] R. Gribonval and M. Nielsen, "Sparse decomposition in "incoherent" dictionaries," in *Proc. IEEE Intl. Conf. on Image Proc.*, 2003.

[2] D.L. Donoho and M. Elad, "Optimally sparse representations in general (non-orthogonal) dictionaries via ℓ^1 minimization," *Proc. National Academy of Sciences of USA*, vol. 100, no. 5, pp. 2197 – 2202, March 4 2003.

[3] R. Gribonval and M. Nielsen, "Highly sparse representations from dictionaries are unique and independent of the sparseness measure," Preprint, 2003.

[4] S.S. Chen, D.L. Donoho, and M.A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[5] J.-J. Fuchs, "On sparse representations in arbitrary redundant bases," Tech. Rep., IRISA, Dec. 2002, submitted to IEEE Trans. Inform. Theory.

[6] Z.S. Wang, J.Y. Cheung, Y.S. Xia, and J.D.Z. Chen, "Minimum fuel neural network and their applications to overcomplete signal representations," *IEEE Trans. Circuits and Systems*, vol. 47, no. 8, pp. 1146 – 1159, Aug 2000.

[7] I.F. Gorodnitsky and B.D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Sig. Proc.*, vol. 45, no. 3, pp. 600 – 616, March 1997.

[8] I. Daubechies, "Time-frequency localization operators: A geometric phase space approach," *IEEE Trans. Inform. Theory*, , no. 34, pp. 605 – 612, 1988.

[9] S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Trans. on Sig. Proc.*, vol. 41, no. 12, pp. 3397 – 3415, 1993.

[10] S. Jaggi, W.C. Karl, S. Mallat, and A.S. Willsky, "High resolution pursuit for feature extraction," *Applied and Computational Harmonic Analysis*, vol. 5, pp. 428, October 1998.

[11] E. Hernández and G. Weiss, *A first course on wavelets*, CRC Press, Boca Raton, FL, 1996, With a foreword by Yves Meyer.

[12] M.V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*, A K Peters, May 1994.

[13] A. Cichocki and R. Unbehauen, "Neural networks for solving systems of linear equations. II. Minimax and least absolute value problems," *IEEE Trans. on Circ. and Sys. II: Analog and Digital Signal Processing*, vol. 39, no. 9, pp. 619 – 633, Sept 1992.

[14] I. Daubechies, "Orthonormal bases of compactly supported wavelets. II. Variations on a theme," *SIAM J. Math. Anal.*, vol. 24, no. 2, pp. 499 – 519, 1993.