EFFICIENT VARIATIONAL INFERENCE FOR THE DYNAMIC HARMONIC MODEL

Ali Taylan Cemgil and Simon J. Godsill

Signal Processing Group, University of Cambridge
Department of Engineering, Trumpington Street, Cambridge, CB2 1PZ, UK
{atc27,sjg}@eng.cam.ac.uk

ABSTRACT

In this paper, we develop a class of probability models that are potentially useful for various music applications such as polyphonic transcription, source separation, restoration or denoising. This class unifies and extends several models such as sinusoidal and harmonic models, additive synthesis model, Gabor regression and probabilistic phase vocoder. We overcome computational intractability issues by introducing structured variational (mean-field) approximations that lead to efficient local message passing algorithms.

1. INTRODUCTION

When excited, vast majority of physical systems (human vocal tract, many acoustical musical instruments e.t.c.) respond with quasi-periodic oscillations. Hence, in short time scales, the energy of signals from these sources is concentrated to a few narrow frequency bands. In modelling music or speech, this property is exploited extensively and it is a well established practice to model audio as a superposition of sinusoidals. This provides a compact representation of the audio waveform using a few amplitude and phase coefficients. However, since the frequency content of music or speech changes over time, the analysis needs to be applied to subsequent short (and possibly overlapping) time frames. This is, of course, the underlying principle behind many well known methods such as short time Fourier transform (STFT) analysis, phase vocoder [1, 2], the sinusoidal model [3], or various harmonic-plus-noise models [4, 5, 6, 7].

However, in many applications, the most "interesting" regions of an audio signal are transient regions associated with *changepoints* such as onsets or offsets, yet it is exactly these points where the stationarity assumptions of a sinusoidal model are violated. Correct characterisation of changepoints is important in various applications: in transcription we wish to estimate note onsets, offset or frequency modulations precisely, in coding or time scaling unfaithful reconstruction of attacks introduces annoying perceptual artifacts such as "flanging" or "phasiness", e.t.c [2].

In this paper, we approach signal analysis from a Bayesian perspective. Central to our approach is a probabilistic hierarchical (three layer) signal model. The bottom layer $p(\mathbf{x}|\mathbf{s})$ describes how the audio signal $\mathbf{x}$ is generated from a latent dynamical process with state variable $\mathbf{s}$. The intermediate layer, denoted by $p(\mathbf{s}|\boldsymbol{\theta})$, is non-stationary and its dynamics is governed by a sequence of unknown parameters $\boldsymbol{\theta}$. The basic idea is to view the parameters $\boldsymbol{\theta}$ as describing the characteristics of a hidden "excitation" sequence that drives the dynamical process. In signal analysis, we wish to estimate $\boldsymbol{\theta}$ (possibly using some prior knowledge encoded by a top layer process $p(\boldsymbol{\theta})$). As a general inference problem, the posterior distribution is given by the Bayes' rule

$$p(\boldsymbol{\theta}|\mathbf{x}) \quad = \quad \frac{1}{p(\mathbf{x})} \int d\mathbf{s} p(\mathbf{x}|\mathbf{s}) p(\mathbf{s}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) \tag{1}$$

In the next section, we introduce this model class, that we name as *dynamic harmonic model* (DHM). An attractive property of DHM is that transient signal characteristics, onsets or frequency fluctuations can be estimated to sample precision, if desired. Moreover, this class unifies and extends several models such as sinusoidal and harmonic models[8, 9], additive synthesis model, Gabor regression [10] and probabilistic phase vocoder[11].

Our purpose in this paper is twofold: First, we introduce DHM using the language of directed graphical models as a generic model for harmonic signal modelling. Our second aim is to illustrate inference methods based on variational (structured mean-field type) approximations [12, 13, 14]. We will describe the method as an iterative local message passing algorithm on a factor graph[15]. Our focus here is to describe a particular variational inference schema on DHM and to demonstrate effectiveness.

2. MODEL

Consider a single channel audio signal $\mathbf{x} = (x_0, \ldots, x_t, \ldots, x_{T-1})^\top$. The harmonic model represents the signal as

$$\mathbf{x}|\mathbf{s} \quad \sim \quad \mathcal{N}(\mathbf{x}; C\mathbf{s}, R) \tag{2}$$

The symbol $\mathcal{N}(x; \mu, \Sigma)$ denotes a Gaussian distribution on x, with mean μ and covariance matrix Σ. In particular, we take $R = \sigma_R I$, where I is the identity matrix. C is a $T \times 2W$ matrix

$$C\mathbf{s} \quad \equiv \quad \begin{pmatrix} C_0^0 & \ldots & C_0^{W-1} \\ \vdots & C_t^\nu & \vdots \\ C_{T-1}^0 & \ldots & C_{T-1}^{W-1} \end{pmatrix} \begin{pmatrix} s^0 \\ \vdots \\ s^{W-1} \end{pmatrix}$$

with time (row) index $t = 0, \ldots, T-1$ and frequency (column) index $\nu = 0, \ldots, W-1$. The block entries are defined by

$$\begin{aligned} C_t^\nu \quad &\equiv \quad \begin{pmatrix} \cos(t\nu\theta) & \sin(t\nu\theta) \end{pmatrix} \\ x_t \quad &= \quad \sum_\nu C_t^\nu s^\nu \end{aligned}$$

where $\theta = 2\pi\Delta_f$ is the angular frequency, $\Delta_f = F_s/L$ is the frequency resolution with F_s being the sampling frequency and $L \geq W$. Note when $\sigma_R = 0$, $T = W = L$, then C is equivalent (up to normalisation) to the inverse Fourier transform matrix and $\mathbf{s}$ are alternating real and imaginary parts of Fourier coefficients.

By slight abuse of notation, we will refer to ν'th coefficient by s^ν, keeping in mind that it is actually not a scalar but a 2×1 vector[1].

The above interpretation highlights the well-known fact that we can view Fourier analysis (or any basis expansion) as a particular instance of the *generalised linear model* [16]. Note that, this is a *static* interpretation, whilst suitable for fast transformations, does not reflect the fact that data is arriving online. One observation is that we can define the model for arbitrary frame lengths. Let N be the length of a frame in samples and define a frame index k such that $k = 0 \dots T/N - 1$. We denote

$$\mathbf{x}_k \quad \equiv \quad (x_{kN}, \dots, x_{(k+1)N-1})$$

The alternative (and arguably physically more realistic) *dynamic state-space* interpretation hinges on the well known trigonometric identity (which is more familiar in complex arithmetic as $e^{j(t+N)\theta} = e^{jN\theta}e^{jt\theta}$)

$$\begin{pmatrix} \cos((t+N)\theta) \\ \sin((t+N)\theta) \end{pmatrix} \quad = \quad B(N\theta) \begin{pmatrix} \cos(t\theta) \\ \sin(t\theta) \end{pmatrix}$$

where $B(\cdot)$ is a Givens rotations matrix.

$$B(\omega) \quad \equiv \quad \begin{pmatrix} \cos(\omega) & -\sin(\omega) \\ \sin(\omega) & \cos(\omega) \end{pmatrix}$$

Using this identity, we see that the basis matrix can be "generated" by the recursion

$$C^\nu_{t+N} \quad = \quad C^\nu_t B(N\nu\theta)^\top$$

Multiplying both sides by s^ν, we obtain for $t = 0, \dots, N-1$

$$C^\nu_{t+N} s^\nu \quad = \quad C^\nu_t B(N\nu\theta)^\top s^\nu$$

We now define $s^\nu_k \equiv B(Nk\nu\theta)^\top s^\nu$. According to this definition, we can represent each frame $\mathbf{x}_k$ by the recursion

$$\begin{aligned} s^\nu_k &= B(N\nu\theta)^\top s^\nu_{k-1} && \nu = 0 \dots W-1 \\ \mathbf{x}_k &= \sum_\nu C^\nu_{0:N-1} s^\nu_k \equiv C^{0:W-1}_{0:N-1} \mathbf{s}_k \end{aligned}$$

Where the notation $C^{0:W-1}_{0:N-1}$ denotes the first N rows of the basis matrix and $\mathbf{s}_k \equiv ({s^0_k}^\top, \dots, {s^\nu_k}^\top, \dots, {s^{W-1}_k}^\top)^\top$. This means that by rotating s^ν appropriately, i.e., by phase correction, we can just use the first N samples of each basis vector to represent $\mathbf{x}$. For example, in the extreme case, when $N = 1$,

$$\begin{aligned} s^\nu_k &= B(\nu\theta)^\top s^\nu_{k-1} && \nu = 0 \dots W-1 \\ \mathbf{x}_k &= x_k = \begin{pmatrix} 1 & 0 & \dots & 1 & 0 \end{pmatrix} \mathbf{s}_k \end{aligned}$$

for $k = 0, \dots, T-1$. Here, all information about the signal is conveyed in the latent dynamical process and the (overcomplete) basis C becomes trivial. Using this idea, we can "interpolate" between dynamic and static interpretations. Of course, mathematically speaking, we haven't really gained anything and in the noiseless case, all formulations for any frame length N are equivalent. However, the dynamical system perspective opens up interesting possibilities to extend the basic model in a way that is not apparent in the static formulation.

[1]One could avoid this technicality by describing the model using complex Gaussian distributions, but this approach precludes certain generalisations.

3. DYNAMIC HARMONIC MODEL

Dynamic Harmonic model is a generative model that describes how the audio signal $\mathbf{x}$ is generated from a latent process with state variable $\mathbf{s}_k$, for $k = 0, \dots, T/N - 1$. Each element s^ν_k of $\mathbf{s}_k$ follows its independent piecewise linear dynamics and is subject to non-stationary Gaussian noise as:

$$s^\nu_k | A^\nu_k, Q^\nu_k \quad \sim \quad \mathcal{N}(s^\nu_k; A^\nu_k s^\nu_{k-1}, Q^\nu_k)$$

Here, transition matrix A^ν_k and noise covariance matrix Q^ν_k are 2×2. The k'th frame is generated by

$$\mathbf{x}_k | \mathbf{s}_k \quad \sim \quad \mathcal{N}(\mathbf{x}_k; \mathbf{C}\mathbf{s}_k, R)$$

where observation noise covariance is isotropic with $R = \sigma_R I$. For brevity, we denote the observation matrix $\mathbf{C} \equiv C^{0:W-1}_{0:N-1}$, $\mathbf{A}_k \equiv \mathbf{diag}(A^0_k, \dots, A^{W-1}_k)$ and $\mathbf{Q}_k \equiv \mathbf{diag}(Q^0_k, \dots, Q^{W-1}_k)$. The prior structure on the transition matrices and noises is taken as a hidden Markov model with discrete latent state $r_k \in \{0, 1, \dots, |r|-1\}$ where $|r|$ denotes number of distinct states:

$$\begin{aligned} r^\nu_k | r^\nu_{k-1} &\sim \mathcal{M}(r^\nu_k; \pi_0(r_{k-1}), \dots, \pi_{|r|-1}(r_{k-1})) \\ A^\nu_k | r^\nu_k &\sim \mathcal{N}(A^\nu_k; \Theta(r^\nu_k), \Sigma(r^\nu_k)) \\ Q^\nu_k | r^\nu_k &\sim \mathcal{IG}(Q^\nu_k; \mathbf{a}(r^\nu_k), \mathbf{b}(r^\nu_k)) \end{aligned}$$

The symbols $\mathcal{M}$, $\mathcal{N}$ and $\mathcal{IG}$ denote *multinomial*, *(matrix valued) Gaussian* and *inverse-gamma* distributions respectively and definitions are given at the appendix. The graphical model is shown in Figure 1. In this paper, we will demonstrate DHM with the following parametrisation:

$$\begin{aligned} r_k &\in \{\text{off}, \text{onset}, \text{on}, \text{offset}\} \\ \Theta(r^\nu_k) &= \rho(r) B(\nu\theta)^\top \end{aligned}$$

where $\rho(r)$ is a scalar with $0 = \rho(\text{onset}) \le \rho(\text{offset}) = \rho(\text{off}) < \rho(\text{on}) \le 1$. The interpretation is that the oscillations are damped more during "off" state and less in "on" state. We let $\Sigma(r^\nu_k) = 0$ to ensure A_k is a fixed rotation matrix. Only the following state transitions have nonzero probability

$$\begin{aligned} \pi_{\text{on}} &= p(\text{on} \to \text{on}) = 1 - p(\text{on} \to \text{offset}) \\ \pi_{\text{off}} &= p(\text{off} \to \text{off}) = 1 - p(\text{off} \to \text{onset}) \\ 1 &= p(\text{offset} \to \text{off}) = p(\text{onset} \to \text{on}) \end{aligned}$$

The transition noise is constrained to be isotropic $Q^\nu_k = \sigma^\nu_k I$ with

$$\sigma^\nu_k | r^\nu_k \quad \sim \quad \mathcal{IG}(\sigma^\nu_k; a(r^\nu_k), b(r^\nu_k))$$

The parameters a and b should be chosen such that the expected transition noise precision $\langle 1/\sigma^\nu_k \rangle_{\mathcal{IG}}$ is small when $r^\nu_k = \text{onset}$ and very large when $r^\nu_k \neq \text{onset}$. This mechanism models the scenario where the state variable s^ν_k is reinitialised during an onset by injecting a random amount of energy and otherwise follows an almost deterministic dynamics. In Figure 2, we demonstrate a typical sample from the model.

4. INFERENCE

To solve the problem in (1) exactly, we need to first infer the posterior distribution

$$p(\mathbf{s}, \boldsymbol{\theta} | \mathbf{x}) = \frac{1}{Z_x} p(\mathbf{x}|\mathbf{s}) p(\mathbf{s}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) \equiv \frac{1}{Z_x} \phi(\mathbf{s}, \boldsymbol{\theta}) \equiv \mathcal{P} \qquad (3)$$

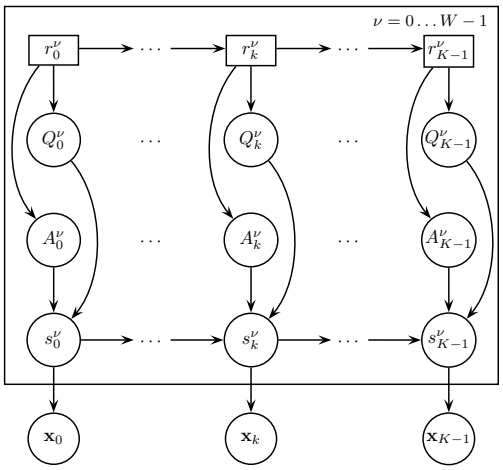

Figure 1: Graphical model for the DHM. The rectangle denotes a plate, W copies of the nodes inside.

where $\mathbf{s} = \{s_k^\nu\}$, $\boldsymbol{\theta} = \{(A_k^\nu, Q_k^\nu)\}$ for all pairs (k, ν). Here, $Z_x = p(\mathbf{x})$ is a normalising constant (also known as the evidence or data likelihood). The parameter posterior is obtained by the marginal $p(\boldsymbol{\theta}|\mathbf{x}) = \int d\mathbf{s} p(\mathbf{s}, \boldsymbol{\theta}|\mathbf{x})$. However, exact evaluation of the posterior distribution in (3) is intractable due to couplings between $\boldsymbol{\theta}$ and $\mathbf{s}$, so we will resort to approximations.

4.1. Structured Mean Field

One possible approximation method, that leads to a practical optimisation procedure is *structured mean field*, also known as *variational Bayes*, see [12, 13, 14] and references herein. In the particular case of (1), mean field boils down to approximating the exact posterior $\mathcal{P}$ in (3) with a simple distribution $\mathcal{Q}$ in such a way that the integrand in (1) becomes tractable. An intuitive interpretation of mean field is minimising the KL divergence with respect to (the parameters of) $\mathcal{Q}$ where

$$KL(\mathcal{Q}||\mathcal{P}) \quad = \quad \langle \log \mathcal{Q} \rangle_{\mathcal{Q}} - \left\langle \log \frac{1}{Z_x} \phi(\mathbf{s}, \boldsymbol{\theta}) \right\rangle_{\mathcal{Q}} \qquad (4)$$

Here, $\langle f(x) \rangle_{p(x)} \equiv \int dx p(x) f(x)$ denotes the expectation of f w.r.t. p. Using non-negativity of KL [17] we obtain a lower bound on the evidence $\log Z_x \geq \langle \log \phi(\mathbf{s}, \boldsymbol{\theta}) \rangle_{\mathcal{Q}} - \langle \log \mathcal{Q} \rangle_{\mathcal{Q}}$ where maximising this lower bound is equivalent to finding the "nearest" $\mathcal{Q}$ to $\mathcal{P}$ in terms of KL. In this paper, we choose the approximating distribution $\mathcal{Q}$ of form

$$\begin{aligned} \mathcal{Q} \quad &\equiv \quad \mathcal{Q}_{\mathbf{s}} \prod_{\nu=0}^{W-1} \mathcal{Q}_r^\nu \prod_{k=1}^{K-1} \mathcal{Q}_{A,k}^\nu \mathcal{Q}_{Q,k}^\nu \\ &= \quad \prod_{k=1}^{K-1} \left(q(\mathbf{s}_k|\mathbf{s}_{k-1}) \prod_{\nu=0}^{W-1} q(r_k^\nu|r_{k-1}^\nu) q(A_k^\nu) q(Q_k^\nu) \right) \end{aligned} \qquad (5)$$

Although a closed form solution for $\mathcal{Q}$ still can not be found, it can be easily shown, e.g. see [18], that each factor $\mathcal{Q}_a$ of the optimal approximating distribution should satisfy the following fixed point equation

$$\mathcal{Q}_a \quad \propto \quad \exp\left(\langle \log \phi(\mathbf{s}, \boldsymbol{\theta}) \rangle_{\mathcal{Q}/\mathcal{Q}_a} \right) \qquad (6)$$

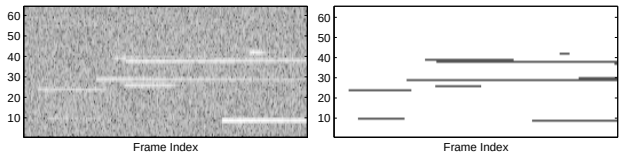

Figure 2: (left) Short Time Fourier Transform log-magnitude of a signal generated from the model (right) Corresponding indicators r_k^ν.

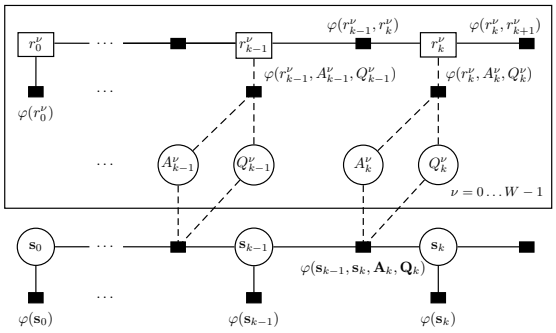

Figure 3: Factor graph that corresponds to the variational approximation in (5). Solid links correspond to substructures where it is feasible to carry out exact inference via belief propagation and dotted links correspond to the links "broken" by the variational approximation.

where $\mathcal{Q}/\mathcal{Q}_a$ denotes product of all factors excluding $\mathcal{Q}_a$. Hence, the mean field approach leads to a set of fixed point equations that need to be iterated.

Right hand side of this fixed point iteration can be computed efficiently since $\log \phi(\mathbf{s}, \boldsymbol{\theta})$ is of form $\sum_\xi \varphi(ne(\xi))$ where φ are the *factors* (local functions) defined on a small set of variables denoted by $ne(\xi)$ and ξ is an index that runs over factors. The structure of this expression for the DHM can be conveniently "visualised" using a factor graph[15] depicted in Figure 3. Here, each factor potential $\varphi(ne(\xi))$ is shown as a black node adjacent to variable nodes $a \in ne(\xi)$. Using this notation, we see that (6) simplifies to

$$\mathcal{Q}_a \quad \propto \quad \exp\left(\sum_{\xi \in ne(a)} \langle \varphi(ne(\xi)) \rangle_{\mathcal{Q}_{ne(\xi)}/\mathcal{Q}_a} \right)$$

where $ne(a)$ denotes set of all factors where a occurs as an argument. The expectations $\langle \varphi \rangle$ can be computed easily if all distributions are chosen to be in a conjugate-exponential family [13], which is the case for the DHM.

The structured mean field for DHM has a particularly intuitive interpretation: each chain structured distribution $\mathcal{Q}_r^\nu$ on $r_{0:K-1}^\nu$ corresponds to an HMM, and similarly, the chain structure $\mathcal{Q}_{\mathbf{s}}$ on $\mathbf{s}_{0:K-1}$ corresponds to a Kalman filter model. The observations of the HMM's and transition model of the Kalman filter at each time slice are determined by sufficient statistics of A_k^ν and Q_k^ν. In other words, the approximation boils down to iteratively running a sequence of Kalman and HMM smoothing algorithms.

Further computational savings can be obtained by considering the expressions for individual factors. For example, when frame length N is chosen such that C is the inverse Fourier transform matrix and $R = \sigma_R I$

$$\begin{aligned}\varphi(\mathbf{s}_k) &= -\frac{1}{2}\,\mathbf{Tr}\,\mathbf{C}^T R^{-1}\mathbf{C}s_k s_k^T + \mathbf{Tr}\,\mathbf{x}_k^T R^{-1}\mathbf{C}s_k \\ &= -\frac{1}{2\sigma_R}\,\mathbf{Tr}\,s_k s_k^T + \frac{1}{\sigma_R}\,\mathbf{Tr}(C^T\mathbf{x})s_k\end{aligned}$$

where the terms $C^T\mathbf{x}$ can be computed by the FFT. Other savings (for any N) can be obtained for $\varphi(s_k, s_{k-1}, \mathbf{A}_k, \mathbf{Q}_k)$ when $\mathbf{A}_k$ is a block-diagonal rotation matrix and $\mathbf{Q}_k$ is diagonal. Due to page limitations, we will report these details in a future technical report.

5. CONCLUSIONS

A very attractive feature of the message passing schema is that the approximating structures can be freely chosen to trade of accuracy versus computation time. Moreover, computer code can be modularised and organised in such a way to make distributed and parallel execution possible. Moreover, by imposing mild restrictions on model parameters, inference can be further speeded up by making use of fast transforms such as the fast Fourier transform. The convergence properties of resulting iterative algorithms can be further improved using deterministic annealing procedures.

We have recently applied a particular DHM successfully for signal restoration tasks[11]. Restoration results can be listened at http://www-sigproc.eng.cam.ac.uk/∼atc27/em-restore/. In future work, we will investigate further applications such as polyphonic transcription or audio source separation and compare variational inference with stochastic methods (e.g. MCMC).

A. APPENDIX

In this section, we define standard distributions in exponential form. For each case, we denote the log-partition function (log normaliser) by $\psi(\cdot)$.

Matrix valued Gaussian: We let X be a $M_r \times M_c$ matrix. Then $\mathbf{vec}\,X$ is a $M_r M_c \times 1$ vector obtained by concatenation of columns of X. The mean Θ has the same size as X and the covariance Σ is $M_r M_c \times M_r M_c$. Then a matrix valued Gaussian

$$\mathcal{N}(X;\Theta,\Sigma) \equiv \mathcal{N}(\mathbf{vec}\,X;\mathbf{vec}\,\Theta,\Sigma)$$

Multinomial: We let $\boldsymbol{\pi} \equiv (\pi_0,\ldots,\pi_i,\ldots,\pi_{|r|-1})$ a vector such that $\pi_i \geq 0$ for all $i = 1,\ldots,|r|$. We denote a multinomial distribution in exponential form as

$$\mathcal{M}(r;\boldsymbol{\pi}) \equiv \exp\Big(\sum_{i=0}^{|r|-1}[r=i]\log\pi_i - \psi(\boldsymbol{\pi})\Big)$$

where $\psi(\boldsymbol{\pi}) = \log\sum_j \pi_j$ and $[r=i]$ is the indicator of the event that $r = i$. Note that the expectation $\langle[r=i]\rangle_{\mathcal{M}} = \pi_1/\sum_j \pi_j$ of the indicator gives simply the probability of the event.

Inverse Gamma: Let $\Gamma(a)$ denote the *gamma* function. We define *digamma* function as $\Psi(a) \equiv \partial\log\Gamma(a)/\partial a$. For $x \geq 0$, the inverse Gamma distribution in exponential form is given as We let $\Sigma = \mathbf{diag}(\sigma_1,\ldots,\sigma_M)$ is a diagonal $M\times M$ matrix with $\sigma_i \geq 0$, we define $\mathbf{a} \equiv \mathbf{diag}(a_1,\ldots,a_M)$ and $\mathbf{b} \equiv \mathbf{diag}(b_1,\ldots,b_M)$

$$\begin{aligned}\mathcal{IG}(\Sigma;\mathbf{a},\mathbf{b}) &\equiv \exp\big(-\mathbf{Tr}(I+\mathbf{a})\log\Sigma - \mathbf{Tr}\,\mathbf{b}^{-1}\Sigma^{-1} - \psi\big) \\ \psi = \psi(\mathbf{a},\mathbf{b}) &\equiv \mathbf{Tr}\log\Gamma(\mathbf{a}) + \mathbf{Tr}\,\mathbf{a}\log\mathbf{b}\end{aligned}$$

$$\begin{aligned}\big\langle\Sigma^{-1}\big\rangle_{\mathcal{IG}} &= \mathbf{diag}(b_1a_1,\ldots,b_Ma_M) \equiv \mathbf{ba} \\ \big\langle\log|\Sigma^{-1}|\big\rangle_{\mathcal{IG}} &= \sum_{j=1}^{M}(\Psi(a_j)+\log b_j) = \mathbf{Tr}(\Psi(\mathbf{a})+\log|\mathbf{b}|)\end{aligned}$$

Note that $\mathcal{IG}(\Sigma;\mathbf{a},\mathbf{b}) = \prod_{j=1}^{M}\mathcal{IG}(\sigma_j;a_j,b_j)$.

B. REFERENCES

[1] J. L. Flanagan and R. M. Golden, "Phase vocoder," *Bell System Technical Journal*, pp. 1493–1509, November 1966.

[2] M. Dolson, "The phase vocoder: A tutorial," *Computer Music Journal*, vol. 10, no. 4, pp. 14–27, 1986.

[3] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 4, pp. 744–754, 1986.

[4] X. Serra and J. O. Smith, "Spectral modeling synthesis: A sound analysis/synthesis system based on deterministic plus stochastic decomposition," *Computer Music Journal*, vol. 14, no. 4, pp. 12–24, 1991.

[5] X. Rodet, "Musical sound signals analysis/synthesis: Sinusoidal + residual and elementary waveform models," *Applied Signal Processing*, 1998.

[6] L. Parra and U. Jain, "Approximate Kalman filtering for the harmonic plus noise model," in *Proc. of IEEE WASPAA*, New Paltz, 2001.

[7] R. A. Irizarry, "Weighted estimation of harmonic components in a musical sound signal," *Journal of Time Series Analysis*, vol. 23, 2002.

[8] P. J. Walmsley, "Signal separation of musical instruments," Ph.D. dissertation, University of Cambridge, 2000.

[9] M. Davy and S. J. Godsill, "Bayesian harmonic models for musical signal analysis," in *Bayesian Statistics 7*, 2003.

[10] P. J. Wolfe, S. J. Godsill, and W. Ng, "Bayesian variable selection and regularisation for time-frequency surface estimation," *Journal of the Royal Statistical Society series B*, 2004, read paper (with discussion).

[11] A. T. Cemgil and S. J. Godsill, "Probabilistic Phase Vocoder and its application to Interpolation of Missing Values in Audio Signals," in *Accepted to 13th European Signal Processing Conference*. Antalya/Turkey: EURASIP, 2005.

[12] W. Wiegerinck, "Variational approximations between mean field theory and the junction tree algorithm," in *UAI-2000 (16-th conference)*, pp. 626–633.

[13] Z. Ghahramani and M. Beal, "Propagation algorithms for variational Bayesian learning," in *Neural Information Processing Systems 13*, 2000.

[14] M. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," Department of Statistics, UC Berkeley, Tech. Rep. 649, September 2003.

[15] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm." *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, February 2001.

[16] A. J. Dobson, *An Introduction to Generalized Linear Models*, 2nd ed. London: Chapman and Hall/CRC, November 2001.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, Inc., 1991.

[18] J. Winn and C. Bishop, "Variational message passing," *Submitted to Journal of Machine Learning Research*, 2004.