

Glue Factors, Likelihood Computation, and Filtering in State Space Models

Christoph Reller
ETH Zurich
reller@isi.ee.ethz.ch

Murthy V. R. S. Devarakonda
murthyds@ieee.org

Hans-Andrea Loeliger
ETH Zurich
loeliger@isi.ee.ethz.ch

Abstract—Factor graphs of statistical models can be augmented by a glue factor that expresses some additional (initial, final, or otherwise “local”) condition. That applies, in particular, to (otherwise time-invariant) linear Gaussian state space models, which are thus generalized to pulse-like models that are localized anywhere in time. The model likelihood can then be computed by (forward-backward or forward-only) sum-product message passing, which leads to the concept of a likelihood filter. We propose to build (forward-only) likelihood filters from a bank of second-order linear systems. We also observe that such likelihood filters can be cascaded into a new sort of neural network that works naturally with multichannel time signals at multiple time scales.

I. INTRODUCTION

In this paper, we pursue further two notions that were introduced in [1], viz., glue factors in factor graphs and likelihood filters. While these concepts are independent, it turns out that glue factors are particularly useful for likelihood filters. The two main new contributions of this paper are the following:

- 1) The decomposition of a likelihood filter into a bank of second-order filters—essentially a discrete-time Fourier transform with an exponential window—and a glue factor.
- 2) The observation that likelihood filters can be cascaded into a new kind of neural network that works naturally with multichannel time signals at multiple time scales.

The paper is structured as follows. The necessary background in factor graphs is briefly summarized in Section II. Glue factors and their use for likelihood computations are discussed in Section III. Beginning with Section IV, we focus on online, or forward-only, likelihood filtering. The decomposition of such a likelihood filter into a bank of second-order components/filters is described in Section V, and hierarchical networks of likelihood filters are proposed in Section VI.

II. STATE SPACE MODELS AND FACTOR GRAPH NOTATION

Throughout the paper, we will be concerned with statistical models of signals (i.e., time series) with a state space representation, i.e., variations of models of the form

$$p(x_0, \dots, x_n, y_1, \dots, y_n) = g_0(x_0) \prod_{k=1}^n g_k(x_{k-1}, y_k, x_k) \quad (1)$$

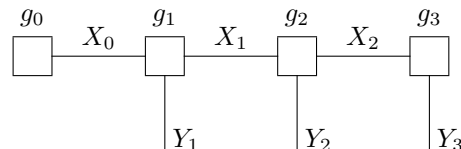


Fig. 1. Forney-style factor graph of (1) for $n = 3$.

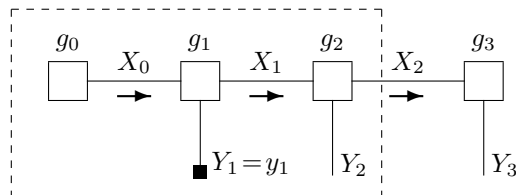


Fig. 2. Sum-product messages.

where Y_1, \dots, Y_n are the observable variables, where X_0, \dots, X_n are auxiliary variables (state variables), where $p(\cdot)$ is a probability mass function or a probability density, and where g_0, g_1, \dots, g_n are real-valued functions. Clearly, a probability law of the form (1) is a hidden Markov model. We will be particularly interested in linear Gaussian models of this form.

We will use the notation of Forney-style factor graphs as in [2], where nodes/boxes represent factors and edges represent variables. For example, the (Forney-style) factor graph of (1) is shown in Fig. 1.

A *sum-product message* in a factor graph without cycles is a quantity such as (see Fig. 2)

$$\vec{\mu}_{X_2}(x_2) = \sum_{\text{variables inside box}} \prod_{\text{factors inside box}} \quad (2)$$

$$= \sum_{x_0, x_1, y_2} g_0(x_0) g_1(x_0, y_1, x_1) g_2(x_1, y_2, x_2), \quad (3)$$

which may be viewed as the result of closing the dashed box in Fig. 2 by summing (or integrating) over its internal variables. Such messages can be computed recursively as, e.g.,

$$\vec{\mu}_{X_2}(x_2) = \sum_{x_1, y_2} \vec{\mu}_{X_1}(x_1) g_2(x_1, y_2, x_2), \quad (4)$$

which is known as *sum-product message passing* or *belief propagation*.

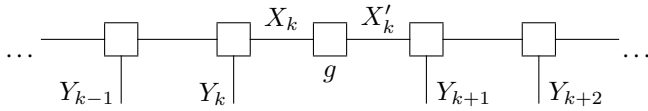


Fig. 3. Time invariant factor graph with a glue factor g .

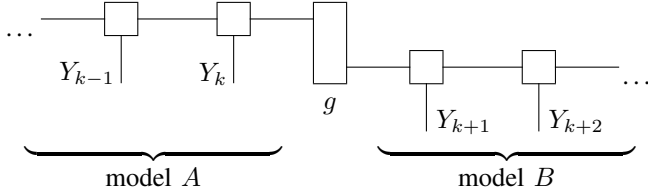


Fig. 4. Factor graph with a glue factor at a model switch.

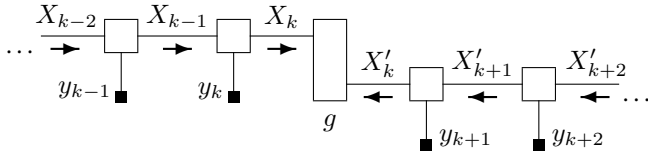


Fig. 5. Computation of likelihoods in Fig. 4 according to (5).

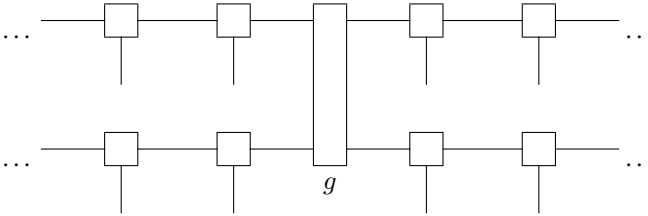


Fig. 6. Glue factor across two otherwise independent models.

Variables whose value is known (such as $Y_1 = y_1$ in Fig. 2, as indicated by the small black square) are, of course, not summed over in (2)–(4).

In many applications of sum-product message passing, scale factors in messages may be ignored. In this paper, however, scale factors cannot be ignored.

III. GLUE FACTORS AND LIKELIHOOD COMPUTATION

Consider the factor graph in Fig. 3, where all factors except g are identical functions; except for the factor g , we thus have a time invariant model. The factor g may be used, e.g., to express some constraint on the state $X_k = X'_k$, or to allow some discontinuity at time k by allowing $X_k \neq X'_k$. For linear Gaussian models, g may express a jump in the state space, in which case the impulse response of the system will appear in Y_{k+1}, Y_{k+2}, \dots . We will refer to g as a *glue factor*, since it specifies how the states X_k and X'_k are glued together.

More generally, in the factor graph of Fig. 4, the glue factor g at time k specifies whatever conditions apply for a switch from some model A for Y_1, \dots, Y_k to some model B for

Y_{k+1}, \dots, Y_n . The likelihood of any fixed signal/sequence $y_1, \dots, y_k, \dots, y_n$ can then be written as

$$p(y_1, \dots, y_n | H_k) = \sum_{x_k, x'_k} \vec{\mu}_{X_k}(x_k) g(x_k, x'_k) \overleftarrow{\mu}_{X'_k}(x'_k), \quad (5)$$

where H_k is the hypothesis that the switch occurs at time k , $\vec{\mu}_{X_k}$ is the forward sum-product message along the edge X_k , and $\overleftarrow{\mu}_{X'_k}$ is the backward sum-product message along the edge X'_k as illustrated in Fig. 5.

The likelihood (5) can thus be efficiently computed for all values of k simultaneously: the messages $\vec{\mu}_{X_k}$ are obtained by forward sum-product message passing in model A , the messages $\overleftarrow{\mu}_{X'_k}$ are obtained by backward sum-product message passing in model B , and from these messages, (5) can be computed for all k . This computation of (5) for $k = 1, 2, \dots, n$ is a generic example of (offline, forward-backward) likelihood filtering.

For large n , the likelihood (5) is normally a very small number, and even $\log p(y_1, \dots, y_n | H_k)$ typically decreases linearly with n . Practical implementations thus need to rescale the sum-product messages and keep track of the scale factors in logarithmic form.

It often happens that the factor graph represents not $p(y_1, \dots, y_n | H_k)$ but $\xi p(y_1, \dots, y_n | H_k)$, where $\xi \in \mathbb{R}$ is an unknown positive scale factor. In this case, ξ can itself be computed by sum-product message passing in Fig. 4 according to

$$\xi = \sum_{x_k, x'_k} \vec{\mu}_{X_k}(x_k) g(x_k, x'_k) \overleftarrow{\mu}_{X'_k}(x'_k). \quad (6)$$

The right-hand side of (6) looks like the right-hand side of (5), but the former refers to Fig. 4, where Y_1, \dots, Y_n are variables to be summed over, while the latter refers to Fig. 5, where $Y_1 = y_1, \dots, Y_n = y_n$ are known.

Finally, we note that glue factors can also express constraints across (otherwise independent) models of a multi-channel signal, as illustrated in Fig. 6. This situation may occur if some event or waveform is to be detected via several independent sensors, as, e.g., in [5], [6].

IV. ON-LINE (FORWARD-ONLY) LIKELIHOOD FILTERING

We now consider problems whose nature is illustrated by questions of the following type:

- *Is my phone ringing?*
- *Did I just hear “stop”?*

Each of these questions refers to a condition that involves some unspecified duration. Nonetheless, both questions strictly refer to the present time; we do not care for how long the respective condition has been satisfied.

In mathematical terms, we are interested in defining and computing quantities of the form

$$p(\dots, y_{k-1}, y_k | H) \quad (7)$$

and

$$p(H | \dots, y_{k-1}, y_k) \quad (8)$$

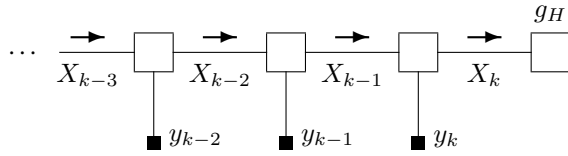


Fig. 7. Forward-only likelihood filtering.

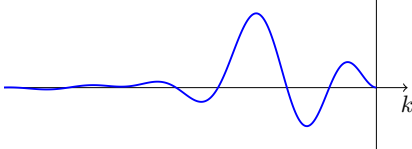


Fig. 8. A pulse that ends at time k .

for $k = 1, 2, 3, \dots$, especially for $k \gg 1$ when the effect of any initial conditions has disappeared. Here H means some hypothesis or statistical model and $p(\cdot|H)$ denotes the corresponding probability mass function (or probability density).

We begin by assuming a state space representation as in Fig. 7 (i.e., a hidden Markov model), where all factors are assumed to be identical functions (i.e., we have a time invariant model) except for the glue factor g_H , which expresses some terminating condition.

We are particularly interested in time invariant linear Gaussian models, where the glue factor allows us to model pulses as in Fig. 8 that end at time k . (Note that a continuous-time linear system with an impulse response as in Fig. 8 has its poles outside the unit circle, i.e., its forward impulse response is unstable. This is no problem at all, since we are interested in the backward impulse response, which is stable.) While Fig. 8 shows a clean deterministic pulse, the linear Gaussian case includes not only such pulses with additive Gaussian noise but also stochastic versions of the pulse shape itself.

In the model of Fig. 7, we clearly have

$$p(y_1, \dots, y_k | H) = \sum_{x_k} \vec{\mu}_{X_k}(x_k) g_H(x_k), \quad (9)$$

which can be computed for $k = 1, 2, 3, \dots$ by forward sum-product message passing as illustrated in Fig. 7.

The logarithm of (9) typically decreases linearly with k . It may not be immediately obvious what this quantity can be compared with, and how something like (8) might be obtained from (9). A natural way to deal with this difficulty is to use a reference model/hypothesis H_0 that agrees with H for sufficiently old data, i.e., the sum-product messages $\vec{\mu}_{X_\ell}$ in the factor graphs of H and H_0 , respectively, coincide for $\ell \ll k$. In this case, the log-likelihood ratio

$$L_k \triangleq \log \frac{p(\dots, y_{k-1}, y_k | H)}{p(\dots, y_{k-1}, y_k | H_0)} \quad (10)$$

is normally a finite number. In the linear Gaussian case (if H is a pulse as in Fig. 8 that decays to zero towards the past), such a reference model H_0 may be obtained by using the same time invariant linear Gaussian model as H , but

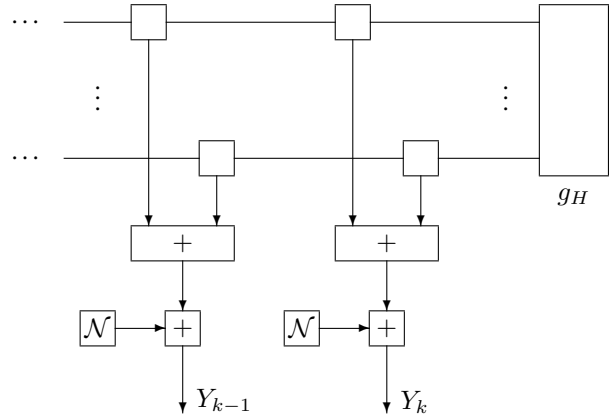


Fig. 9. Superposition of second-order models (damped sinusoids) in additive Gaussian noise. The unlabeled boxes represent the constraints (13) and (14); the boxes labeled \mathcal{N} are Gaussian factors.

terminating it in state $X_k = 0$, which can be expressed by an appropriate glue factor g_{H_0} .

Finally, a posterior probability (8) can be defined by the (admittedly wild) assumption

$$p(H) + p(H_0) = 1, \quad (11)$$

which implies

$$p(H | \dots, y_{k-1}, y_k) = \frac{1}{1 + \frac{p(H_0)}{p(H)} e^{-L_k}} \quad (12)$$

V. DECOMPOSITION INTO SECOND-ORDER SYSTEMS

We now focus on linear Gaussian models that can be decomposed into second-order models of the following special form. The state $X_k \in \mathbb{R}^2$ of such a second-order model evolves deterministically according to

$$X_k = \rho \begin{pmatrix} \cos \Omega & -\sin \Omega \\ \sin \Omega & \cos \Omega \end{pmatrix} X_{k-1} \quad (13)$$

and produces the scalar output signal

$$\tilde{Y}_k = (1, 0) X_k. \quad (14)$$

Clearly, these output signals have the form

$$\tilde{Y}_k = a \rho^k \cos(\Omega k + \varphi) \quad (15)$$

for arbitrary amplitude a and phase φ (which might be fixed by a glue factor at the end). We will assume $\rho > 1$ (but typically $\rho \approx 1$), which makes the signals (15) decay towards the past. (Alternatively, we could assume $\rho = 1$ and work with a forgetting factor as in [1].)

We now proceed to the superposition of such signals as shown in Fig. 9, where the unlabeled boxes represent the constraints (13) and (14) of the individual second-order systems. The observable output signal is the sum of the output signals of the component systems plus additive white Gaussian noise.

We will typically choose the frequencies Ω of the constituent second-order systems to be equidistant. With a suitable choice of the frequency spacing, the resulting output signals (15) will be approximately orthogonal (for $\rho \approx 1$)

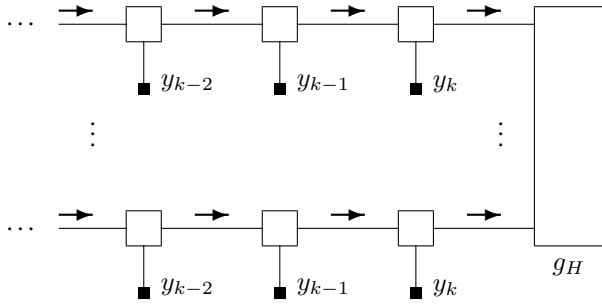


Fig. 10. Factor graph of a likelihood filter comprising a bank of second-order models/filters and a glue factor g_H .

and yet allow to represent essentially any smooth pulse-like signal (as in Fig. 8) by a suitable choice of amplitudes and phases.

The glue factor g_H in Fig. 9 can be used to fix a particular combination of amplitudes and phases, which results in a deterministic pulse. However, much more general glue factors can be used, e.g., to fix only certain ratios of phases and amplitudes, or to impose a Gaussian prior on the joint state space.

It should be noted, however, that the factor graph of Fig. 9 has cycles and is thus not amenable to forward-only likelihood filtering as in Section IV. Noting, however, that the constituent second-order systems are approximately orthogonal (for suitable frequency spacing), it can be shown that the forward sum-product messages are approximately independent and can be computed as shown in Fig. 10, where each constituent second-order system (with its own additive noise) is fed directly with the observed signal.

The computation of the sum-product messages in Fig. 10 amounts essentially to a bank of second-order filters. Moreover, it can be shown that the mean vectors of the messages form a discrete-time Fourier transform with exponential window [4].

Clearly, the same sum-product messages (filter bank) can be used for many different glue factors.

Finally, the architecture of Fig. 10 is easily extended to multichannel signals, where each signal is fed into its own filter bank and glue factors extend across multiple filter banks.

VI. HIERARCHICAL PROBABILITY FILTER NETWORKS

The probability output signal (8) or (12) of a likelihood filter can itself be analyzed by likelihood filtering. (The probability signal (8) or (12), which is bounded between 0 and 1, is much better suited for this purpose than (7) or (10).) We thus arrive at the architecture shown in Fig. 11, where a multichannel signal is analyzed by a first set of likelihood filters with cross-channel glue factors, which produce a collection of probability signals, which in turn are analyzed by a second set of likelihood filters with cross-channel glue factors, and so on.

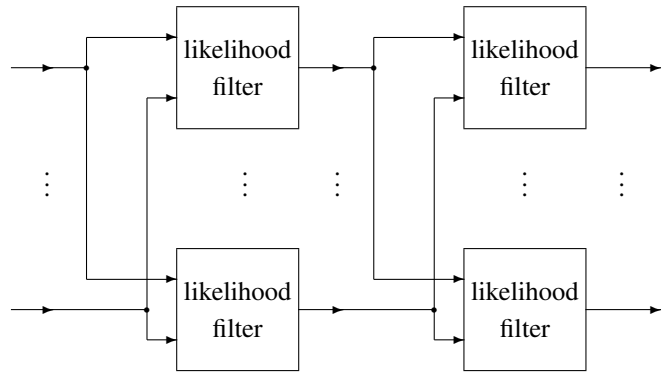


Fig. 11. Probability filter network (block diagram, not a factor graph).

Note that, in the actual implementation, the likelihood filters on the same level may share the same set of filter banks, i.e., one filter bank per signal may be sufficient.

Such a network of likelihood filters, or rather probability filters, is naturally suited for analyzing signals at multiple time scales. The intermediate probability signals may be viewed as features that are used by the next level. The computation of the output signals follows the same principles as the computation of the feature signals at each level.

VII. CONCLUSION

We discussed glue factors and likelihood computations in factor graphs of statistical state space models with emphasis on the linear Gaussian case. In particular, we showed a decomposition of a likelihood filter into a bank of second-order models/filters and a glue factor. We finally arrived at an architecture for a network of likelihood filters (or rather probability filters) that is naturally suited for extracting information at multiple time scales.

It is obvious that the present paper is only an outline. While a wealth of pertinent details is given in [3] and [4], the subject is still in its infancy.

REFERENCES

- [1] H.-A. Loeliger, L. Bolliger, S. Korl, and Ch. Reller, "Localizing, forgetting, and likelihood filtering in state-space models," 2009 Information Theory & Applications Workshop, UCSD, La Jolla, CA, Feb. 8-13, 2009.
- [2] H.-A. Loeliger, J. Dauwels, Junli Hu, S. Korl, Li Ping, and F. R. Kschischang, "The factor graph approach to model-based signal processing," *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1295-1322, June 2007.
- [3] Murti Devarakonda, *Joint Matched Filtering, Decoding, and Timing Synchronization*. PhD thesis at ETH Zurich No 19882, 2011.
- [4] Christoph Reller, *State-Space Methods in Statistical Signal Processing: New Ideas and Applications*. PhD thesis at ETH Zurich No 20584, 2012.
- [5] Ch. Reller, H.-A. Loeliger, and S. Marandò, "Multi-sensor estimation and detection of phase-locked sinusoids," *Proc. 2011 Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 22-27, 2011, pp. 3872-3875.
- [6] S. Marandò, Ch. Reller, D. Fäh, and H.-A. Loeliger, "Seismic waves estimation and wave field decomposition with factor graphs," *Proc. 2011 Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 22-27, 2011, pp. 2748-2751.