Chalmers Publication Library

(article starts on next page)

On the Use of Factor Graphs in Optical Communications

Henk Wymeersch[1], Naga V. Irukulapati[1], Domenico Marsella[2], Pontus Johannisson[3],
Erik Agrell[1], Marco Secondini[2]

[1]*Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden*
[2]*Institute of Communication, Information, and Perception Technologies, Scuola Superiore Sant'Anna, Pisa, Italy*
[3]*Photonics Laboratory, Chalmers University of Technology, Gothenburg, Sweden*

henkw@chalmers.se

Abstract: Factor graphs and message passing allow the near-automated development of algorithms in many engineering disciplines, including digital communications. This paper gives an overview of their possible use in optical communications.

1. Introduction

A factor graph (FG) is a general type of graphical model that expresses how a function (e.g., a probability density function) factorizes in local functions [1]. This factorization is then harnessed to compute certain properties of the original function, such as its mode or marginals. These computations are performed using a variety of computationally efficient message passing algorithms, the most well-known of which is the sum-product algorithm (SPA, also known as belief propagation). It can be argued that the power of FGs lies in their ability to nearly automatically solve a wide variety of problems without supervision, once the problem has been formalized and factorized. In the particular context where the function in question is a distribution $p(\mathbf{x}, \mathbf{y})$, in which $\mathbf{y}$ is an observation and $\mathbf{x}$ is a random unknown, the factorization is based on conditional independence among the unknown variables. Given a factorization that gives rise to a cycle-free FG,[1] message passing algorithms are able to exactly and efficiently determine the following fundamental quantities in inference problems:

- The marginal posteriors: $p(x_i|\mathbf{y})$, which can be used to make optimal (in terms of minimizing the probability of making an error) decisions on the components of $\mathbf{x}$;
- The posterior mode: $\hat{\mathbf{x}} = \arg\max_\mathbf{x} p(\mathbf{x}|\mathbf{y})$, which is the optimal decision on the entire vector $\mathbf{x}$; and
- The (unconditional) likelihood: $p(\mathbf{y})$, which is used to compare different models.

FGs find their roots in coding theory, statistics, and expert systems, combining the idea of representing a problem as a graph and solving the problem on the graph. Since the early 2000s, FGs have gained in popularity, and, due to their ability to deal with extremely complex and large-scale problems, are now being applied in fields ranging from bioinformatics over cooperative localization to user interface design. FGs are not only important for their ability to automatically generate new algorithms for problems but also for the fact that they generalize many important optimal algorithms in the fields of signal processing and statistics. These include the Viterbi algorithm, the Kalman filter, as well as many soft decoding algorithms [2].

2. An Example of an FG and Messages for a Simple Communication Scheme

To make the concepts described in the previous section more concrete, we provide a small artificial example. Consider a communication system where 4 independently generated bits $\mathbf{b} = [b_1, b_2, b_3, b_4]$ are mapped onto two QPSK symbols using a mapping function $\phi(\cdot)$: $\mathbf{s} = (s_1, s_2)^T = (\phi(b_1, b_2), \phi(b_3, b_4))^T$, sent over a channel $\mathbf{H} \in \mathbb{R}^{2\times2}$, and subjected to additive white Gaussian noise $\mathbf{n}$, giving rise to an observation $\mathbf{y} = \mathbf{Hs} + \mathbf{n}$. Assuming the receiver knows $\mathbf{H}$ and the noise variance σ^2 per real dimension, it can factorize (with $\mathbf{x} = (\mathbf{b}, \mathbf{s})$)

$$p(\mathbf{b}, \mathbf{s}, \mathbf{y}) = p(\mathbf{b})p(\mathbf{s}|\mathbf{b})p(\mathbf{y}|\mathbf{s}) = \left(\prod_{i=1}^{4} p(b_i)\right)\mathbb{I}\{s_1 = \phi(b_1, b_2)\}\mathbb{I}\{s_2 = \phi(b_3, b_4)\}\frac{1}{(2\pi\sigma^2)^2}\exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{Hs}\|^2\right), \quad (1)$$

in which $\mathbb{I}\{\cdot\}$ is the indicator function (i.e., $\mathbb{I}\{S\} = 0$ when the statement S is false and 1 otherwise). The corresponding FG is depicted in Fig. 1. In order to compute, for example, $p(b_3|\mathbf{y})$, the SPA computes messages[2]

---

[1]In the case the FG has cycles, the message passing algorithms only give approximate solutions. The quality of these solutions depends on the lengths of the cycles, the factors along the cycle, and the message passing schedule. In many practical applications, the resulting algorithms still perform very well and include common decoding algorithms of low density parity check codes and turbo codes.

[2]Superscripts R and L denote rightward and leftward messages. Messages are distributions of variables (e.g., a message associated with a binary variable is a vector of length 2 with non-negative entries that sum up to one). An outgoing message from a factor vertex is obtained by multiplying the factor with the incoming message, and summing over all variables, except the outgoing variable.

$\mu^R(b_1) = p(b_1)$ (and similarly $\mu^R(b_2)$), then $\mu^R(s_1) = \sum_{b_1,b_2} \mathbb{I}\{s_1 = \phi(b_1,b_2)\}\mu^R(b_1)\mu^R(b_2)$, followed by $\mu^R(s_2) = \sum_{s_1} \mu^R(s_1)\exp(-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{Hs}\|^2)$, then $\mu^R(b_3) = \sum_{b_4,s_2} \mathbb{I}\{s_2 = \phi(b_3,b_4)\}\mu^R(s_2)\mu^L(b_4)$, in which $\mu^L(b_4) = p(b_4)$. Finally, the desired marginal is $p(b_3|\mathbf{y}) \propto \mu^R(b_3)\mu^L(b_3)$, based on which we can make an optimal decision regarding b_3.
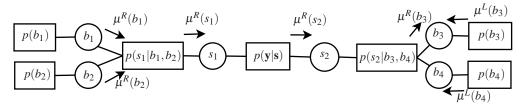

Fig. 1: An FG of the factorization of $p(\mathbf{b},\mathbf{s},\mathbf{y})$, comprising one variable node for each variable in $\mathbf{x} = (b_1,b_2,b_3,b_4,s_1,s_2)$ and one factor node for each factor in the factorization. Several messages, which are distributions of the associated variables, are also depicted.

3. Factor Graphs in Digital Communications

The flexible nature of FGs was utilized in receiver design for digital communication systems in [3], in particular focusing on iterative receivers, where probabilistic information flows between the different parts of the receiver. While FGs can be applied to perform inference over entire networks, we constrain ourselves to single links between a transmitter and a receiver. In such scenarios, FGs have been used in the following aspects: computing information rates [4], channel estimation [5], channel equalization [6], demodulation [7], decoding [1], phase noise tracking [8], and synchronization [9]. This enables the use of probabilistic (soft) information provided by the decoder to be used to improve other receiver tasks (e.g., equalization, synchronization). In most of these aspects, the FG solution is among the most powerful known approaches. However, this often comes at a high complexity cost, which has spurred intense research in low-complexity approximations [6].

4. Factor Graphs in Optical Communications

FGs have had a more limited use in optical communications. This partly because, in contrast to generic digital communication systems, there is no explicit data-aided channel estimation thus making an FG solution extremely complex. Nevertheless, FGs can still be utilized in demodulation [10, 11] (offering improved performance over the state of the art) and decoding [12, 13] (where the need for proper code construction to accommodate the properties of message passing is highlighted). Iterative schemes, with information flow from the decoder to other receiver blocks (e.g., the equalizer) have also been considered [14], but could lead to excessive delays in current architectures. Other interesting applications of FGs arise in offline processing, such as in computation of information rates [15] or the design of near-optimal receivers [16].

5. An Application Example: Nonlinear Impairment Compensation

To demonstrate the potential of FGs, we consider a detection problem for which no optimal receiver is known: detection in the presence of fiber nonlinearity, dispersion, and noise. In particular, consider a scenario where a $2 \times K$ matrix of K PM-QPSK symbols $\mathbf{X}_0 = (\mathbf{x}_{0,1},\ldots,\mathbf{x}_{0,K})$ is transmitted over a system, comprising multiple fiber spans (each of which can be modeled through the Manakov model), separated by amplifiers. The received $2 \times L$ matrix $(L > K)$ after oversampling is denoted by $\mathbf{Y}$. Due to the interaction between the nonlinearity, dispersion, and noise, there is no closed-form expression for $p(\mathbf{X}_0,\mathbf{Y})$, precluding making optimal decisions directly. However, due to the Manakov model, we have a closed-form expression of the operation affecting the signal at each fiber segment, as well as models for the amplifiers. By including the corresponding intermediate variables (say, $\mathbf{X}_n$, modeling the sampled version of an intermediate waveform after passing through a segment ordered according to the index n from transmitter to receiver), we can factorize $p(\mathbf{X}_0,\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_N,\mathbf{Y}) = p(\mathbf{X}_0)[\prod_{i=1}^N p(\mathbf{X}_i|\mathbf{X}_{i-1})]p(\mathbf{Y}|\mathbf{X}_N)$. The corresponding FG is a linear graph, on which we can execute the sum-product algorithm. Using a particle representation of the messages with P particles, as detailed in [16], we end up with a description of $\mu^L(\mathbf{X}_0)\mu^R(\mathbf{X}_0) = p(\mathbf{X}_0,\mathbf{Y})$ in the form of P matrices of length $2 \times K$, say, $\mathbf{X}_0^{(1)},\ldots,\mathbf{X}_0^{(P)}$. From this description, we can detect the k-th symbol $\mathbf{x}_{0,k}$ as follows: approximate $p(\mathbf{x}_{0,k-W,\ldots,k+W},\mathbf{Y})$ with a multivariate Gaussian distribution, in which $0 \leq W \ll P$ is the window size. Find, by trying all combinations, the sequence $\mathbf{x}_{0,k-W},\ldots,\mathbf{x}_{0,k+W}$ that maximizes $p(\mathbf{x}_{0,k-W},\ldots,\mathbf{x}_{0,k+W},\mathbf{Y})$, and take the center entry $\mathbf{x}_{0,k}$ as the decision $\hat{\mathbf{x}}_{0,k}$. In Fig. 2, we show the performance of this approach, for a 28 GBaud system comprising

42 spans of single-mode fiber (each of length 80 km with dispersion parameter 16 ps/(nm km)) followed by periodic optical dispersion compensation using a fiber Bragg grating (with 3 dB insertion loss) and an amplifier. We observe that the FG-based receiver significantly outperforms a receiver based on digital backpropagation.
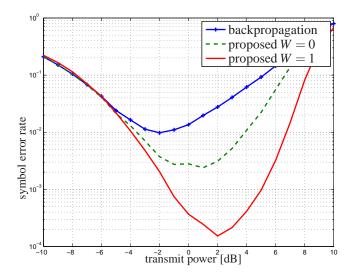

Fig. 2: Symbol error rate performance for detection in the presence of nonlinearity, dispersion, and noise. As a reference we show digital back propagation, along with the FG solution for $W = 0$ and $W = 1$. For complexity reduction in our simulation, we only evaluated sequences $\mathbf{x}_{0,k-1}, \mathbf{x}_{0,k}, \mathbf{x}_{0,k+1}$ for which $\mathbf{x}_{0,k-1}$ and $\mathbf{x}_{0,k+1}$ were fixed to the actual transmitted symbols. Actual gains may be smaller.

6. Conclusions

Factor graphs present a unified, consistent, and near-automatic framework for the design of inference algorithms. They have had a significant impact in the design of digital communication receivers, and are a promising technique in optical communications, especially in cases where traditional techniques are severely suboptimal and where complexity is a secondary constraint.

References

1. F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," IEEE Trans. Inf. Theory **47**, 498–519 (2001).
2. H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang, "The factor graph approach to model-based signal processing," Proc. IEEE **95**, 1295–1322 (2007).
3. A. P. Worthen and W. E. Stark, "Unified design of iterative receivers using factor graphs," IEEE Trans. Inf. Theory **47**, 843–849 (2001).
4. D.-M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," IEEE Trans. Inf. Theory **52**, 3498–3508 (2006).
5. Q. Guo, L. Ping, and D. Huang, "A low-complexity iterative channel estimation and detection technique for doubly selective channels," IEEE Trans. Wireless Commun. **8**, 4340–4349 (2009).
6. M. Tüchler and A. C. Singer, "Turbo equalization: An overview," IEEE Trans. Inf. Theory **57**, 920–952 (2011).
7. A. G. i Fàbregas, A. Martinez, and G. Caire, *Bit-interleaved coded modulation* (Now Publishers Inc, 2008).
8. G. Colavolpe, A. Barbieri, and G. Caire, "Algorithms for iterative decoding in the presence of strong phase noise," IEEE J. Sel. Areas Commun. **23**, 1748–1757 (2005).
9. C. Herzet, N. Noels, V. Lottici, H. Wymeersch, M. Luise, M. Moeneclaey, and L. Vandendorpe, "Code-aided turbo synchronization," Proc. IEEE **95**, 1255–1271 (2007).
10. M. A. Castrillon, D. A. Morero, and M. R. Hueda, "Joint demapping and decoding for DQPSK optical coherent receivers," arXiv preprint arXiv:1206.4914 (2012).
11. F. Yu *et al.*, "Soft-decision LDPC turbo decoding for DQPSK modulation in coherent optical receivers," in ECOC, We.10.P1.70 (2011).
12. B. P. Smith and F. R. Kschischang, "Future prospects for FEC in fiber-optic communications," IEEE J. Quantum Electron. **16**, 1245–1257 (2010).
13. I. B. Djordjevic, M. Arabaci, and L. L. Minkov, "Next generation FEC for high-capacity communication in optical transport networks," J. Lightw. Technol. **27**, 3518–3530 (2009).
14. I. B. Djordjevic, L. L. Minkov, and H. G. Batshon, "Mitigation of linear and nonlinear impairments in high-speed optical networks by using LDPC-coded turbo equalization," IEEE J. Sel. Areas Commun. **26**, 73–83 (2008).
15. I. B. Djordjevic, B. Vasic, M. Ivkovic, and I. Gabitov, "Achievable information rates for high-speed long-haul optical transmission," J. Lightw. Technol. **23**, 3755–3763 (2005).
16. N. V. Irukulapati, H. Wymeersch, P. Johannisson, and E. Agrell, "Stochastic digital backpropagation," IEEE Trans. Commun. – (2014). To appear.