MULTI-SENSOR ESTIMATION AND DETECTION OF PHASE-LOCKED SINUSOIDS

Christoph Reller, Hans-Andrea Loeliger

ETH Zurich
Dept. of Information Technology & Electr. Eng.
8092 Zürich, Switzerland
Email: {reller, loeliger}@isi.ee.ethz.ch

Stefano Marandò

ETH Zurich
Swiss Seismological Service
8092 Zürich, Switzerland
Email: marano@sed.ethz.ch

ABSTRACT

This paper proposes a method to compute the likelihood function for the amplitudes and phase shifts of noisily observed phase-locked and amplitude-constrained sinusoids. The sinusoids are assumed to be coupled based on a set of parameters as, e.g., measurements of a monochromatic wave field. A factor graph is used to formulate the probability density function of the observations given the parameters. The factor graph consists of one second-order state-space model per signal and one additional factor connecting all the final states. Because the parameters appear only in this latter factor, we are able to formulate a sufficient statistic for parameter estimation and signal detection in terms of messages in the factor graph. In special cases, the general form of the sufficient statistic reduces to the discrete Fourier transform. As extensions we provide iterative algorithms for approximate maximum likelihood estimation of the noise variances and the parameters of superposed waves.

Index Terms— factor graph, sinusoid, detection, estimation, likelihood function

1. INTRODUCTION

Phase-coupled and amplitude-constrained sinusoids play an important role in many fields ranging from sensor arrays in seismology [1], acoustics, and electromagnetics to multiterminal communication. Parameter estimation for, and detection of, uncoupled sinusoids is described e.g. in [2] and [3]. However, as soon as we consider some coupling, a linear model does not apply anymore. This paper provides a unified and general approach to estimation and detection of coupled sinusoids based on a factor graph representation [4].

Consider L discrete-time sinusoidal signals

$$\xi_k^{(\ell)} = \alpha_\ell \cos(\Omega k + \psi_\ell), \tag{1}$$

where $\ell = 1, \ldots, L$ enumerates the signals and $k = 0, \ldots, K-1$ is the time index. All L signals have the same, known frequency Ω but differ in amplitude α_ℓ and phase ψ_ℓ. We observe the noisy signal $Y_k^{(\ell)} = \xi_k^{(\ell)} + Z_k^{(\ell)}$, where $Z_k^{(\ell)}$ are zero-mean white Gaussian noises with noise variances σ_ℓ^2 for $\ell = 1, \ldots, L$. Note that we allow σ_ℓ^2 to differ in each signal.

Unconstrained estimation of α_ℓ and ψ_ℓ is a well known problem [2]. This paper, however, deals with coupled signals. Specifically we assume that the amplitudes α_ℓ and phase shifts ψ_ℓ are constrained by some parameter vector θ via some mapping

$$\Gamma : \theta \mapsto \left((\alpha_1, \psi_1), \ldots, (\alpha_L, \psi_L) \right). \tag{2}$$

As a toy example assume that we have two signals (1) with unconstrained amplitudes α_1 and α_2 but with the same phase $\psi \triangleq \psi_1 = \psi_2$. We are interested in α_1, α_2 and ψ. In this example one possible choice is $\theta = (\rho_0, \phi_0, \beta)$ and $\Gamma(\theta) = \left((\rho_0, \phi_0), (\beta\rho_0, \phi_0) \right)$.

For a more relevant example consider the estimation of seismic wave fields measured by a sensor array. In this example, θ may contain wave field parameters such as wave type, velocity of propagation, angle of arrival, etc. The mapping Γ would include sensor characteristics and positions. (This setting is treated in more detail in [1].)

We use the term *wave* for $\xi \triangleq \{\xi_k^{(\ell)}\}_{\ell=1,\ldots,L, k=0,\ldots,K-1}$ induced by the parameters θ given the mapping (2). For the noisy signal we define $Y^{(\ell)} \triangleq \left(Y_0^{(\ell)}, \ldots, Y_{K-1}^{(\ell)} \right)$ and $Y \triangleq \left(Y^{(1)}, \ldots, Y^{(L)} \right)$. Given observations $Y = y$, we want to formulate the likelihood function $f(y|\theta)$ in order to make maximum likelihood (ML) estimates

$$\hat{\theta} = \operatorname*{argmax}_\theta f(y|\theta) \tag{3}$$

and to compute the generalized log-likelihood ratio (LLR)

$$\ln \frac{f(y|\hat{\theta})}{f(y|\mathcal{H}_0)}, \tag{4}$$

where the hypothesis $\mathcal{H}_0$ is the presence of only noise. (The LLR is usually used for signal detection [3].)

Clearly, the likelihood can be written as

$$f(y|\theta) = \prod_{\ell=1}^{L} \prod_{k=0}^{K-1} \frac{1}{\sqrt{2\pi\sigma_\ell^2}} e^{-\left(y_k^{(\ell)} - \xi_k^{(\ell)}(\theta)\right)^2 / (2\sigma_\ell^2)}. \tag{5}$$
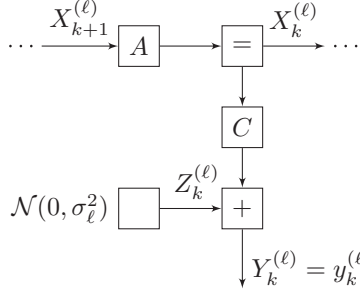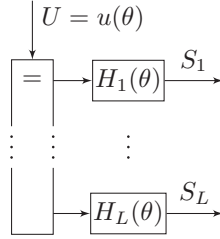
Fig. 1: state-space model for sinusoid

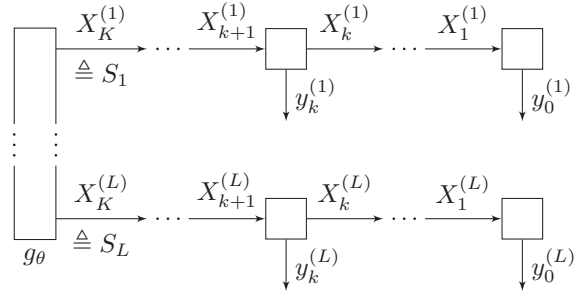**Fig. 2**: Glue factor g_θ

Fig. 3: Overall factor graph

In Section 2 we show that $f(y|\theta)$ is a function of the quantities $H_\ell(\theta) \in \mathbb{R}^{2\times 2}$, $\overleftarrow{W}_{S_\ell}(\sigma_\ell^2) \in \mathbb{R}^{2\times 2}$, and $\overleftarrow{m}_{S_\ell}(y^{(\ell)}) \in \mathbb{R}$ for $\ell = 1, \ldots, L$. The complexity for computing $\overleftarrow{W}_{S_\ell}$ and $\overleftarrow{m}_{S_\ell}$ is linear in K and for fixed H_ℓ, $\overleftarrow{W}_{S_\ell}$, $\overleftarrow{m}_{S_\ell}$ the complexity for computing this function is linear in L. Note that $\overleftarrow{W}_{S_\ell}$ and $\overleftarrow{m}_{S_\ell}$ do not depend on θ and form a sufficient statistic. In Section 3 we show how this sufficient statistic simplifies in restricted settings and we show a connection with the discrete Fourier transform (DFT). In Section 4 an iterative algorithm for approximating the ML estimate of the noise variances in each signal is given. Finally, in Section 5 we formulate an iterative algorithm to compute the likelihood function and LLR of several superposed waves.

2. COMPUTING LIKELIHOODS WITH FACTOR GRAPHS

2.1. State-Space Model

For each signal $Y^{(\ell)}$ we formulate a second-order linear state-space model with state vector X_k as

$$X_k^{(\ell)} = A\, X_{k+1}^{(\ell)}, \tag{6}$$

$$Y_k^{(\ell)} = C\, X_k^{(\ell)} + Z_k^{(\ell)}, \tag{7}$$

where A is a rotation matrix

$$A \triangleq \mathrm{rotm}(-\Omega), \tag{8}$$

$$\mathrm{rotm}\,\alpha \triangleq \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}, \tag{9}$$

and $C \triangleq (1, 0)$. The corresponding (Forney) factor graph is depicted in Fig. 1. Note that the time progression in Figs. 1 and 3 is from right to left. We use the abbreviation $S_\ell \triangleq X_K^{(\ell)}$ for the final system state.

2.2. Glue Factor

Without loss of generality we can assume that θ contains the overall amplitude $\rho_0 \neq 0$ and overall phase ϕ_0 of an unknown

reference sinusoid $\rho_0 \cos(\Omega k + \phi_0)$. By letting

$$u(\theta) \triangleq \rho_0 \begin{pmatrix} \cos(\Omega K + \phi_0) \\ \sin(\Omega K + \phi_0) \end{pmatrix} \tag{10}$$

be the state of this reference sinusoid at time K we can express the coupling between the signals as

$$S_\ell = H_\ell(\theta)\, u(\theta), \tag{11}$$

where $H_\ell(\theta) \triangleq \rho_\ell\, \mathrm{rotm}(\phi_\ell)$, $\rho_\ell \triangleq \alpha_\ell/\rho_0$, and $\phi_\ell \triangleq \psi_\ell - \phi_0$.

Both $H_\ell(\theta)$ and $u(\theta)$ depend on the parameters θ via the mapping Γ in (2). In our toy example we have $H_1 = I_2$ and $H_2 = \beta\, I_2$, where I_2 denotes the 2×2 identity matrix.

We define a *glue factor* (Fig. 2)

$$g_\theta(s_1, \ldots, s_L) = \prod_{\ell=1}^{L} \delta\big(s_\ell - H_\ell(\theta)\, u(\theta)\big), \tag{12}$$

modeling the relations (11) as constraints, where $\delta(\cdot)$ denotes the Dirac delta. (See [5] for the concept of a glue factor.)

2.3. Likelihood Function and LLR

The overall factor graph in Fig. 3 consists of L state-space models (Fig. 1) connected by the glue factor (Fig. 2). This factor graph is tree-shaped and represents the probability density function

$$f(y, x|\theta) = f\big(y, x\,\big|\,u(\theta), H(\theta)\big), \tag{13}$$

with $X \triangleq \{X_k^{(\ell)}\}_{\ell=1,\ldots,L,\, k=1,\ldots,K}$ and $H(\theta) \triangleq \big(H_1(\theta)^{\mathrm{T}}, \ldots, H_\ell(\theta)^{\mathrm{T}}\big)^{\mathrm{T}}$.

Instead of using the brute force calculation (5), we use sum-product message passing [4] in the factor graph to compute the likelihood function. We use arrows to distinguish between forward messages ($\overrightarrow{\cdot}$) in the same direction as the edge and backward messages ($\overleftarrow{\cdot}$) in the opposite direction.

By marginalization and due to the definition of the sum-product rule we can write

$$f(y|\theta) = \int f(y, x|\theta)\, dx = \overleftarrow{\mu}_U(u(\theta)). \tag{14}$$

Since the state-space models (Fig. 1) do not depend on θ, we can compute the messages $\overleftarrow{\mu}_{S_\ell}$ without specifying θ. The likelihood (14) is then calculated from $\overleftarrow{\mu}_{S_\ell}$ and θ. It immediately follows that $\overleftarrow{\mu}_{S_\ell}$ for $\ell = 1, \ldots, L$ is a sufficient statistic.

All messages in the factor graph at hand are (potentially scaled and degenerate) multivariate Gaussian probability density functions. In this paper we prefer to write a message (e.g. for an edge X in forward direction) in the form

$$\overrightarrow{\mu}_X(x) \triangleq \overrightarrow{\gamma}_X \, e^{-x^{\mathrm{T}} \overrightarrow{W}_X x/2 + x^{\mathrm{T}} \overrightarrow{W}_X \overrightarrow{m}_X} \,, \qquad (15)$$

where $\overrightarrow{V}_X = \overrightarrow{W}_X^{-1}$ is the covariance matrix (if it exists), $\overrightarrow{m}_X$ is the mean vector and the scale factor is defined as

$$\overrightarrow{\gamma}_X \triangleq \overrightarrow{\mu}_X(0) \,. \qquad (16)$$

The maximization in (3) can be done by first maximizing over $u(\theta)$. Since for Gaussian messages max-product message passing coincides with sum-product message passing the ML estimate of $U(\theta)$ is $\hat{u} = \overleftarrow{m}_U$. We set $u(\theta) = \hat{u}$ in (14) and get the partially maximized likelihood function from message update rules for Gaussian messages [4] as

$$\ln f\big(y\big|u(\theta) = \hat{u}, H(\theta)\big) = \frac{1}{2}\overleftarrow{m}_U^{\mathrm{T}} \overleftarrow{W}_U \overleftarrow{m}_U + \ln \overleftarrow{\gamma}_U \,, \quad (17)$$

where

$$\overleftarrow{W}_U = \sum_{\ell=1}^{L} H_\ell(\theta)^{\mathrm{T}} \, \overleftarrow{W}_{S_\ell} \, H_\ell(\theta) \,, \qquad (18)$$

$$\overleftarrow{W}_U \overleftarrow{m}_U = \sum_{\ell=1}^{L} H_\ell(\theta)^{\mathrm{T}} \, \overleftarrow{W}_{S_\ell} \, \overleftarrow{m}_{S_\ell} \,, \qquad (19)$$

$$\ln \overleftarrow{\gamma}_U = \sum_{\ell=1}^{L} \ln \overleftarrow{\gamma}_{S_\ell} \,. \qquad (20)$$

The remaining maximization over $H_\ell(\theta)$ is in general non-convex and depends largely on the mapping Γ. We do not treated it here.

For the noise hypothesis $\mathcal{H}_0$ we constrain X to be zero by setting $u = 0$ in (12). From (14) and (16) we get

$$\ln f(y|\mathcal{H}_0) = \ln \overleftarrow{\gamma}_U \,, \qquad (21)$$

so that the partially maximized LLR is

$$\mathrm{LLR}(\theta) \triangleq \ln \frac{f\big(y\big|u(\theta) = \hat{u}, H(\theta)\big)}{f(y|\mathcal{H}_0)} = \frac{1}{2}\overleftarrow{m}_U^{\mathrm{T}} \overleftarrow{W}_U \overleftarrow{m}_U \,. \quad (22)$$

Note that (20) does not depend on θ and therefore can be neglected when finding ML estimates $\hat{\theta}$. The sufficient statistic thus consists of $\big\{\overleftarrow{m}_{S_\ell}, \overleftarrow{W}_{S_\ell}\big\}_{\ell=1,\ldots,L}$.

It is easy to generalize the state-space models to non-uniform sampling. For this we substitute A in (6) by time-varying matrices $A_k^{(\ell)} \triangleq \mathrm{rotm}\big((t_{k-1}^{(\ell)} - t_k^{(\ell)})\,\omega\big)$, where ω is the continuous time frequency and $t_k^{(\ell)}$ are time stamps of $y_k^{(\ell)}$. In the same fashion we can easily accommodate time varying noise variances. The likelihood function and the LLR can still be computed by message passing as in (17) and (22).

3. CONNECTION WITH THE DFT

In the case of uniform sampling as assumed in (1), the following analytic solution for the messages $\overleftarrow{\mu}_{S_\ell}$ can be proven.

$$\overleftarrow{W}_{S_\ell} = \frac{K}{2\sigma_\ell^2} I_2 + \frac{\sin(\Omega K)}{2\sigma_\ell^2 \sin \Omega} \, \mathrm{rotm}(\Omega)\,R \,, \qquad (23)$$

$$\overleftarrow{W}_{S_\ell}\overleftarrow{m}_{S_\ell} = \frac{1}{\sigma_\ell^2} R \sum_{k=0}^{K-1} y_k^{(\ell)} \begin{pmatrix} \cos(k\Omega) \\ \sin(k\Omega) \end{pmatrix} \,, \qquad (24)$$

where $R \triangleq \mathrm{rotm}(\Omega K)\big(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\big)$. When viewing the sum in (24) as a function of Ω we recognize the real and imaginary parts of the finite-length discrete-time Fourier transform of $\big(y_0^{(\ell)}, \ldots, y_{K-1}^{(\ell)}\big)$.

If we further restrict the frequency Ω to be $\Omega_n \triangleq 2\pi n/K$ for $n = 0, \ldots, K-1$, then $R = \big(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\big)$ and the expressions above simplify to

$$\overleftarrow{W}_{S_\ell} = \frac{K}{2\sigma_\ell^2} I_2 \,, \qquad (25)$$

$$\overleftarrow{m}_{S_\ell} = \frac{2}{K} \sum_{k=0}^{K-1} y_k^{(\ell)} \begin{pmatrix} \cos(2\pi kn/K) \\ -\sin(2\pi kn/K) \end{pmatrix} \,. \qquad (26)$$

The latter consists of the real and the imaginary parts of the DFT of $\big(y_0^{(\ell)}, \ldots, y_{K-1}^{(\ell)}\big)$, scaled by $2/K$. If, in addition, the noise variances are the same in all signals, i.e., if $\sigma_\ell^2 = \sigma^2$ for $\ell = 1, \ldots, L$, then (22) simplifies to

$$\mathrm{LLR}_n(\theta) = \frac{\breve{y}_n^{\mathrm{T}}\, H(\theta)\, H(\theta)^{\mathrm{T}}\, \breve{y}_n}{K\sigma^2 \sum_{\ell=1}^{L} \rho_\ell^2} \,, \qquad (27)$$

where $\breve{y}_n \triangleq \frac{K}{2}\big(\overleftarrow{m}_{S_1}^{\mathrm{T}}, \ldots, \overleftarrow{m}_{S_L}^{\mathrm{T}}\big)^{\mathrm{T}}$ contains the n'th component of the DFTs of the signals. Equation (27) is a standard beam-forming result. (See e.g. [6].)

4. NOISE VARIANCE ESTIMATION

In this section we consider the case where the noise variances $\eta \triangleq \big(\sigma_1^2, \ldots, \sigma_L^2\big)$ are not given a-priori but are estimated in an ML sense for both the "signal present" ($\mathcal{H}_1$) and the "noise present" ($\mathcal{H}_0$) hypothesis.

The joint maximization for $\mathcal{H}_1$ over (θ, η) is non-convex. We propose to use cyclic maximization [7] by alternating

$$\hat{\theta} = \underset{\theta}{\mathrm{argmax}}\; f\big(y\big|\theta, \hat{\eta}\big) \,, \qquad (28)$$

$$\hat{\eta} = \underset{\eta}{\mathrm{argmax}}\; f\big(y\big|\hat{\theta}, \eta\big) \,. \qquad (29)$$

Since the likelihood in every iteration cannot decrease, cyclic maximization algorithms are guaranteed to converge.

The maximization (28) is the same as (3) and hence the procedure in Section 2 applies. To start the algorithm we propose an initial estimate $\hat{\eta}$ based on assuming that the signals

are decoupled, i.e., the glue factor (12) is replaced by

$$g_\theta(s_1, \ldots, s_L) = \prod_{\ell=1}^{L} \delta(s_\ell - \overleftarrow{m}_{s_\ell}). \quad (30)$$

With (30), these initial ML noise variance estimates are

$$\hat{\sigma}_\ell^2 = \frac{1}{K} \sum_{k=0}^{K-1} \left(y_k^{(\ell)} - C A^{K-k} \overleftarrow{m}_{S_\ell} \right)^2. \quad (31)$$

Once we have an estimate $\hat{\theta}$ we can calculate $\vec{m}_{S_\ell}$ by apply $\hat{\theta}$ in the glue factor (12) and get the coupled ML noise variance estimates as

$$\hat{\sigma}_\ell^2 = \frac{1}{K} \sum_{k=0}^{K-1} \left(y_k^{(\ell)} - C A^{K-k} \vec{m}_{S_\ell} \right)^2. \quad (32)$$

If the signals are long we might want to avoid the direct computation of $\hat{\sigma}_\ell^2$. We can approximate (31) and (32) by

$$\hat{\sigma}_\ell^2 \approx \zeta_\ell^2 - \overleftarrow{m}_{S_\ell}^{\mathrm{T}} \overleftarrow{m}_{S_\ell}/2, \quad (33)$$
$$\hat{\sigma}_\ell^2 \approx \zeta_\ell^2 + \vec{m}_{S_\ell}^{\mathrm{T}} \vec{m}_{S_\ell}/2 - \vec{m}_{S_\ell}^{\mathrm{T}} \overleftarrow{m}_{S_\ell} \quad (34)$$

respectively, where $\zeta_\ell^2 \triangleq \frac{1}{K} \sum_{k=0}^{K-1} \left(y_k^{(\ell)} \right)^2$ are the signal powers. Using these approximations, the only input to the algorithm is $\{\overleftarrow{m}_{S_\ell}, \overleftarrow{W}_{S_\ell}, \zeta_\ell^2\}_{\ell=1,\ldots,L}$. It can be shown that $\overleftarrow{m}_{S_\ell}$ does not depend on σ_ℓ^2 and $\overleftarrow{W}_{S_\ell}$ depends linearly on σ_ℓ^2.

Under the noise hypothesis $\mathcal{H}_0$ the ML estimate of the noise variances are $\hat{\eta}^{(0)} \triangleq (\zeta_1^2, \ldots, \zeta_L^2)$. The generalized LLR can easily be calculated from (5) as [3]

$$\ln \frac{f(y | \hat{\theta}, \hat{\eta}, \mathcal{H}_1)}{f(y | \hat{\eta}^{(0)}, \mathcal{H}_0)} = \frac{K}{2} \sum_{\ell=1}^{L} \ln \frac{\zeta_\ell^2}{\hat{\sigma}_\ell^2}. \quad (35)$$

5. EXTENSION TO WAVE SUPERPOSITION

Assume that we observe a linear superposition of M waves $\xi^{(m)}$ with same frequency Ω, parameters θ_m, and mappings $\Gamma_m : \theta_m \mapsto \left((\alpha_1^{(m)}, \psi_1^{(m)}), \ldots, (\alpha_L^{(m)}, \psi_L^{(m)}) \right)$ for $m = 1, \ldots, M$. Our signal model now is

$$Y_k^{(\ell)} = \sum_{m=1}^{M} \left(\alpha_\ell^{(m)} \cos(\Omega k + \psi_\ell^{(m)}) \right) + Z_k^{(\ell)}. \quad (36)$$

We collect the parameters in a vector $\theta \triangleq (\theta_1, \ldots, \theta_M)$. It is straight forward to model all waves simultaneously by using extended matrices $\tilde{H}_\ell(\theta) \triangleq (H_\ell(\theta_1), \ldots, H_\ell(\theta_M))$ and state vectors $\tilde{u}(\theta) \triangleq (u(\theta_1)^{\mathrm{T}}, \ldots, u(\theta_M)^{\mathrm{T}})^{\mathrm{T}}$ in (11) and (12). However, the space over which to maximize in (17) increases approximately M fold.

As an alternative we propose an iterative algorithm based on cyclic maximization [7]. Assume that we have an estimate

$\hat{\theta}$. We pick some $m \in \{1, \ldots, M\}$ and update the estimate of θ_m while fixing $\{\theta_j = \hat{\theta}_j\}_{j \in \{1, \ldots, M\} \backslash \{m\}}$. This leads to the following glue factor

$$g_{\theta_m}(s_1, \ldots, s_L) = \prod_{\ell=1}^{L} \delta\big(s_\ell - \hat{s}_\ell^{\cancel{m}} - H_\ell(\theta_m) u(\theta_m)\big), \quad (37)$$

where $\hat{s}_\ell^{\cancel{m}} \triangleq \sum_{j \in \{1, \ldots, M\} \backslash \{m\}} H_\ell(\hat{\theta}_j) u(\hat{\theta}_j)$ is the estimated state due to all the waves except for the m-th. The corresponding $\hat{\theta}_m$ can again be found by maximizing (17) where $\overleftarrow{\mu}_U$ is calculated using the glue factor (37).

To apply this algorithm we propose the following greedy-type procedure. Initially, set $M = 1$ and use the glue factor (12) to find $\hat{\theta}_1$. Then repeatedly do the following. Increase M, use the glue factor (37) with $m = M$ to find $\hat{\theta}_M$, and iterate finding $\hat{\theta}_m$ for $m \in \{1, \ldots, M\}$ until convergence. This algorithm is applied successfully in [1].

6. CONCLUSION

We have used a factor graph to derive a sufficient statistic (in simple cases this is the DFT) for the ML estimation of wave parameters. The sufficient statistic can be used to devise iterative algorithms for the estimation of superposed waves and of the noise variances.

7. REFERENCES

[1] S. Maranò, C. Reller, D. Fäh, and H.-A. Loeliger, "Seismic waves estimation and wave field decomposition with factor graphs," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, May 2011.

[2] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993, vol. 1.

[3] ——, *Fundamentals of Statistical Signal Processing: Detection Theory*. Prentice Hall, 1998, vol. 2.

[4] 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*, vol. 95, no. 6, pp. 1295–1322, 2007.

[5] H.-A. Loeliger, L. Bolliger, C. Reller, and S. Korl, "Localizing, forgetting, and likelihood filtering in state-space models," in *Proc. Information Theory and Applications Workshop*, 2009, pp. 184–186.

[6] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, 1996.

[7] P. Stoica and Y. Selen, "Cyclic minimizers, majorization techniques, and the expectation-maximization algorithm: a refresher," *IEEE Signal Process. Mag.*, vol. 21, no. 1, pp. 112–114, 2004.