Deriving a Joint Interference Detection and Channel Estimation for WB-OFDM from EM-MAP Theory

Niels Hadaschik, and Gerd Ascheid
Institute for Integrated Signal Processing Systems, RWTH Aachen University, Templergraben 55, 52056 Aachen, Germany
{hadaschik, ascheid}@iss.rwth-aachen.de

Abstract—For modern wideband transmission systems narrowband interference and their identification becomes an important issue especially for weak interferers. This paper theoretically derives an interference detection scheme from EM theory in conjunction with a MAP channel estimation. The approach is related to a previously introduced heuristic scheme [1] and does not require mute periods or carriers, but works on a known burst preamble. Beyond the derivation of the detector-estimator, this work analytically evaluate the estimator variance, showing the value of the scheme.

I. INTRODUCTION

In this paper, an optimization criterion based derivation of the heuristic approach in [1] is introduced for a data-aided (DA) joint narrowband interference (NBI) detection and channel estimation for wideband OFDM. Basis is the maximum a-posteriori (MAP) version of the expectation maximization (EM) algorithm. Usually, the EM algorithm [2] iteratively solves maximum likelihood (ML) estimation problems, but is extendable to the MAP problems. This is advantageous, because ML estimates do not exploit correlations between the elements of a random parameter vector.

If coexisting narrowband systems operate in the same spectral range, interference (IF) occurs. Coexisting systems may be operating in the same freely usable spectrum like e.g. Bluetooth in the ISM band, or may be licensed systems in their frequency range. Our system is then tolerated with low power spectral densities. The latter one is the UWB approach for which MB-OFDM [3] is an implementation.

In [4], the IF is combated with estimation and cancellation techniques in frequency domain and the approach in [5] is based on adaptive filtering in time domain. However, reference [4] relies on measurements on scattered null tones, ignores a first IF acquisition and assumes the center frequency of the interferer to be known. The concepts in [5] needs mute periods where no signal is transmitted in order to adapt to the IF. [6] discusses the topic of IF with the focus on ADC performance in conjunction with an analogue notch filter and assumes partial IF knowledge. The work [7] introduces a blind IF detection which averages data over a whole burst, but leaves the transmitted data unprocessed.

IF can easily be detected, if sensing the otherwise void channel before transmission, even for relatively low IF power by using simple power detection schemes in frequency domain. However, these mute periods are often neither wanted nor feasible. [1] bases the IF detection on the received preamble or data symbols and shows that power detection of weak interferers in frequency domain is severely harmed by OFDM symbols and channel variations.

Hence, this text derives an iterative multistage approach from EM theory in a similar fashion as [10], [8], [9] for multiuser detection, which combines the IF detector with a MAP channel estimator working on the burst preamble. Therefor, a first step (pre-)detects the IF and, afterwards, the scheme iteratively estimates channel and IF considering the acquired knowledge on the other, and considering channel covariance information. Beneath the derivation, the focus of this paper lies within a variance analysis of the estimators.

This work is outlined as follows: in section II the models of the OFDM signals, and of the NBI are introduced. Section III derives the detector schemes, and then section IV analyzes the performance. Exemplary results are provided in section V, before VI concludes the text.

II. SIGNAL AND INTERFERENCE MODEL

Identifiers for scalars are normal letters, vectors are denoted by bold, lower case letters, matrices by bold, upper case letters. \( \mathbf{I} \) is the unit matrix with respective dimension.

We assume an OFDM based transmission system with a preamble of \( M \) OFDM symbols each consisting of \( K \) subcarriers. The IF in this context is sufficiently accurately modelled by a vector \( \mathbf{b} \) of \( N_{\text{IF}} \) complex degrees of freedom, covering its magnitude and phase and an \( (MK \times N_{\text{IF}}) \) matrix \( D \) mapping the degrees of freedom on distinct frequency and time bins of the received signal. Thus, the received signal vector in frequency domain of the whole preamble has the length \( MK \) and becomes

\[
\mathbf{y} = \mathbf{Xh} + \mathbf{Db} + \mathbf{w} \quad (1)
\]

The diagonal (PSK-)signal matrix \( \mathbf{X} \) is of dimension \( MK \times MK \) and the AWGN is assembled in the vector \( \mathbf{w} \) of length \( MK \) and has the covariance matrix \( \mathbf{R}_w \). The channel vector \( \mathbf{h} \) contains the \( M \) stacked sets of channel coefficients. In order to have the tempo-spectral channel covariance matrix \( \mathbf{R}_h \) invertible, the channel is assumed to be stationary and time varying. The derived estimator itself also allows constant channels. In general, the IF \( \mathbf{b} \) is a random process, but there exists no knowledge on its distribution. An often made assumption is the Gaussian one, but in here, we do not restrict ourselves to this.
III. INTERFERENCE SIGNAL DETECTION

While ML maximizes a-priori probabilities, the MAP estimator does so for a-posteriori probabilities. Already Dempster et al. in [2] refer to the EM-MAP approach, sometimes called Bayesian EM [8], [9], which becomes apparent from the below application of Bayes theorem. In the presence of IF according to the above model, the MAP channel estimator is

\[
\arg \max_h p(h | y) = \arg \max_h \frac{p(y | h) p(h)}{p(y)} = \arg \max_h p(y | h) p(h)
\]

This marginalization represents the main difficulty for MAP estimates. The transmitted signal \( X \) and the matrix \( D \) are assumed to be known, however the condition on them is omitted for ease of notation, i.e. for instance

\[
p(y | h, b) = p(y | h, b, D, X).
\]

In the Gaussian case, the MAP estimator (without IF) equals the LMMSE solution [11]. However, the MAP criterion usually allows a relatively simple derivation of the Bayesian estimate. In contrast to the classical MMSE derivation, it employs no marginalization (integration) but a differentiation.

Hence, the method of choice remains the EM algorithm in its MAP version. However, we simplify the estimate of the IF – in this context, the nuisance parameter – to its ML realization.

Starting with the complete data \( a^T = (y^T, b^T) \), the expectation step initiates the EM algorithm

\[
Q(h | \hat{h}^{(n-1)}) = \int_a \log[p(a | h) p(h)] \cdot p(a | y, \hat{h}^{(n-1)}) da.
\]

and the maximization step renders the channel estimate

\[
\hat{h}^{(n)} = \arg \max_h Q(h | \hat{h}^{(n-1)})
\]

in its \( n \)-th iteration. Realizing independence of channel \( h \) and IF \( b \), after some algebra, (6) expands into

\[
Q(h | \hat{h}^{(n-1)}) = \int_b \log[p(y | h, b)] \cdot p(b | y, \hat{h}^{(n-1)}) \cdot p(b | h) \cdot p(h) db.
\]

This extends the ML solution only by the last summation. When maximizing with respect to \( h \), the second summation vanishes, because it is independent of the channel (the previous estimate \( \hat{h}^{(n-1)} \) is constant). The probability density functions (pdf) are given by

\[
p(y | h, b) = \frac{1}{\pi^{MK} | R_w |} e^{-(y - x_h - Db)^H R_w^{-1} (y - x_h - Db)}
\]

\[
p(h) = \frac{1}{\pi^{MK} | R_h |} e^{-h H R_h^{-1} h}.
\]

Thus, elaborating further on the maximization step yields

\[
\arg \max_h Q(h | \hat{h}^{(n-1)}) = \arg \max_h \left\{ \int_b \log[p(y | h, b)] \cdot p(b | y, \hat{h}^{(n-1)}) db \right\}
\]

\[
= \arg \max_h \left\{ \int_b [-y H R_w^{-1} y - h H X H R_w^{-1} X H + b H D H R_w^{-1} X H - h H X H R_w^{-1} \hat{h}^{(n-1)} - p(b | y, \hat{h}^{(n-1)})] db \right\}
\]

\[
= \arg \max_h \left\{ -h H X H R_w^{-1} X H + b H D H R_w^{-1} X H - h H R_h^{-1} h \right\}
\]

\[
= \arg \max_h \left\{ \int_b [b H D H R_w^{-1} X H] p(b | y, \hat{h}^{(n-1)}) db \right\}.
\]

Therein, we discard several terms not depending on \( h \). As a foresight of a later differentiation with respect to \( h \), we also drop terms just depending on \( h^H \), for \( \frac{\partial}{\partial h^H} h = 0 \). At this point, it is convenient to realize that the prior separates into

\[
p(b | y, \hat{h}^{(n-1)}) = p(b | y, \hat{h}^{(n-1)}, \mathcal{I} \mathcal{F}) P(\mathcal{I} \mathcal{F}) p(\hat{h}^{(n-1)})
\]

\[
+ p(b | y, \hat{h}^{(n-1)}, \mathcal{I} \mathcal{F}) P(\mathcal{I} \mathcal{F}) p(\hat{h}^{(n-1)})
\]

where the discrete probabilities of the events of persistent (\( \mathcal{I} \mathcal{F} \)) and absent interference (\( \mathcal{I} \mathcal{F} \)) are used.

The knowledge of the interference being zero yields

\[
p(b | y, \hat{h}^{(n-1)}, \mathcal{I} \mathcal{F}) = \delta(h), \quad (\delta(\cdot \cdot) \text{ is the multidimensional Dirac-impulse, and } (12) \text{ is rewritten as})
\]

\[
\arg \max_h Q(h | \hat{h}^{(n-1)})
\]

\[
= \arg \max_h \left\{ -h H (X H R_w^{-1} X + R_h^{-1} h) + y H R_w^{-1} X h 
\]

\[
- P(\mathcal{I} \mathcal{F} | y, \hat{h}^{(n-1)}) \int_b b H p(b | y, \hat{h}^{(n-1)}, \mathcal{I} \mathcal{F}) db \cdot D H R_w^{-1} X h \right\}
\]

\[
= \arg \max_h \left\{ -h H (X H R_w^{-1} X + R_h^{-1} h) + y H R_w^{-1} X h 
\]

\[
- P(\mathcal{I} \mathcal{F} | y, \hat{h}^{(n-1)}) E_h \{ b | y, \hat{h}^{(n-1)}, \mathcal{I} \mathcal{F} \} H D H R_w^{-1} X h \right\}
\]

\[
= \arg \max_h \left\{ -h H (X H R_w^{-1} X + R_h^{-1} h) 
\]

\[
+ (y H - \tilde{b}^{(n)} H D H) R_w^{-1} X h \right\},
\]

which interprets expectation and IF probability as iterative IF estimate \( \tilde{b}^{(n)} \) given the channel estimate from the previous iteration \( \hat{h}^{(n-1)} \). Now differentiating (15) with respect to \( h \) and setting the result equal to zero renders

\[
h H (X H R_w^{-1} X + R_h^{-1} h) = (y H - \tilde{b}^{(n)} H D H) R_w^{-1} X.
\]

Thus, we obtain the \( n \)-th channel estimate of the joint EM-MAP channel estimator and IF detector

\[
\hat{h}^{(n)} = (X H R_w^{-1} X + R_h^{-1} X H R_w^{-1} (y - D \tilde{b}^{(n)}) )^{-1} (y - D \tilde{b}^{(n)}).
\]
For later iterations, assume near perfect channel estimation is on (unknown) priors. Nevertheless, note that in convergence, correlations almost cease to exist and iid white gaussian noise interesting range is the transition range between those two the channel estimation error can well be neglected in the SNR however, since the priors of the IF are unknown, this work decisive.

The problem remains that also \( P(b_{\text{IF}}|y, \hat{h}^{(n-1)}) \) depends on (unknown) priors. Nevertheless, note that in convergence, the channel estimation error can well be neglected in the SNR range of interest (0 dB - 10 dB). In this case, residual channel correlations almost cease to exist and iid white gaussian noise distorts the input of the detector.

The actual detector needs to be evaluated based on signal energy \( |y_k|^2 \) due to remaining phase ambiguities. Hence, the input to the energy detector is central \( \chi^2_{2M} \) -distributed in the presence of IF.

The lack of a priori knowledge is circumvented when transiting from soft decisions on the (non-) occurrence of IF to hard ones. Therefore, a threshold \( P_0 \) for the conditional probability is introduced. Then, with the mapping

\[
M\{P\} = \begin{cases} 
1 & \text{for } P > P_0 \\
0 & \text{otherwise},
\end{cases}
\]

an estimate based on hard decisions results

\[
\hat{b}^{(n)} = E\{b|y, \hat{h}^{(n-1)}, IF \} : M\{P(IF|y, \hat{h}^{(n-1)})\}.
\]

Ant error covariance derivation

This section analyzes the variance of the above scheme in an iterative manner and by (in-)finite matrix power series. Therefore, the variance of the two partial estimators is determined each considering the error covariance of its respective counterpart; the estimator matrices

\[
Q_b = (D^H D)^{-1} D^H \text{ and } Q_h = R_h^H X^H (R_w + X R_h X^H)^{-1}
\]

symbolize the estimators for interference and channel h, respectively. The covariance of the interference estimator in (19), i.e. only considering correct hard decisions, is determined by

\[
R_b^{(n)} = E\{[b - Q_b(y - X h^{(n-1)})][b - Q_b(y - X h^{(n-1)})]^H\}
\]

\[
= E\{Q_b([w + X e^{(n-1)}])(w + X e^{(n-1)})^H Q_b^H\}
\]

\[
\approx Q_b(R_w + X R_h^{(n-1)} X^H)Q_b^H
\]

\[
= R_b + Q_b(X R_h^{(n-1)} X^H)Q_b^H.
\]

The error covariance of the channel estimator in the \( n \)-th iteration yields

\[
R_h^{(n)} = E\{|h - Q_h(y - D b^{(n)})|[h - Q_h(y - D b^{(n)})]^H\}
\]

\[
\approx R_h + Q_h(D R_b^{(n)} D^H)Q_h^H,
\]

which drops the covariance between \( e^{(n)} \) and \( w \). Thus, appropriately initializing \( R_b^{(0)} \) or \( R_h^{(0)} \), and then, iterating \( n \)-times according to (26) and (27) obtains the \( n \)-th error variances. The result converges vs. a fixed point, the approximate minimum error covariance. Irrespective of the initialization, this fixed point remains the same. A worst case initialization is e.g. no channel estimate is available prior to IF detection, such that the detection noise consists of AWGN and received OFDM symbol with joint covariance \( R_g = R_w + X R_h X^H \). A best case arises from assuming perfect channel knowledge prior to interference detection such that the detection noise has the covariance \( R_w \).

The channel estimation error covariance practically converges to the case of an LMMSE channel estimator with perfect knowledge of interference statistics of a Gaussian interferer with the same power. Simulations with prior perfect knowledge of the matrix \( D \) and pure hard interference
Better analytical error covariance terms for the \( n \)-th iteration are derived in Appendix B after initializing with \( R_b^{(0)} = R_h \) – i.e. no channel estimate is initially available. Introducing therein the shortcuts \( \mathcal{A} = Q_h D Q_b X \) and \( \mathcal{B} = Q_b X Q_h D \), the error covariances in (44) and (45) are based on finite sums of powers of \( \mathcal{A} \) and \( \mathcal{B} \). Letting the iteration index \( n \) go versus infinity, the minimum achievable covariances for correct detection (\( P_b^{(n=n_0)} = 1 \)) result in

\[
\begin{align*}
R_b^{(\infty)} &= (I - B)^{-1} \left[ R_b + Q_h (X R_b X^H - R_w Q_h X^H - X Q_h R_w Q_b^H) \right] (I - B^H)^{-1} \\
R_h^{(\infty)} &= (I - A)^{-1} \left[ R_h + Q_h (D R_b D^H - R_w Q_b^H D^H - D Q_b R_w Q_b^H) \right] (I - A^H)^{-1}
\end{align*}
\]

(28)

In order to prove that these comprise the mean square estimation error (MSEE), it is to show, under which conditions on \( \mathcal{A} \) and \( \mathcal{B} \) the infinite sums converge, and the initialization terms comprising \( R_h \) in (44) and (45) vanish. Afterwards, it is shown that \( \mathcal{A} \) and \( \mathcal{B} \) fulfill this condition by definition. This indicates for correct IF detection, the convergence of the original joint channel estimation and IF detection algorithm.

For convergence, it is necessary that \( \mathcal{A} \) and \( \mathcal{B} \) have the absolute value of their eigenvalues smaller than 1. Proof: \( \mathcal{A} \) and \( \mathcal{B} \) are decomposed into unitary matrices \( U_A I_U \) and \( U_B I_U \) assembling their right sided eigenvectors and the respective diagonal matrix \( \Lambda_A I_U \) of their eigenvalues \( \lambda_{A,B,i} \) such that for the example of \( \mathcal{A} \)

\[
\mathcal{A} = U_A \Lambda_A U_A^H.
\]

(30)

Since \( \mathcal{B} \) is gained by product commutation of \( \mathcal{A} \), both share the same non-zero eigenvalues according to theorem 1.3.20 from [12]. That means considering their powers of \( n \) in (44) and (45)

\[
\begin{align*}
\mathcal{A}^n &= (U_A \Lambda_A U_A^H)^n \\
&= U_A \Lambda_A^n U_A^H
\end{align*}
\]

(31)

(32)

that convergence is assured if and only if all \( |\lambda_{A,i}| \) and \( |\lambda_{B,j}| \) are strictly smaller than 1 [12].

Thus in order to prove convergence, it suffices to show that the \( |\lambda_{A,i}| \) are strictly smaller than 1 – corresponding to a spectral radius \( \rho(\mathcal{A}) < 1 \). Therefore, it is helpful to realize that the spectral radius is smaller or equal than any matrix norm [12]. The spectral radius itself is not a matrix norm, and hence, does not obey the rules of matrix norms: it fails to provide the submultiplicative characteristic of matrix norms stating that for an arbitrary norm \( \| \cdot \| \) and two matrices \( F, G \in M_n \) that \( \| F G \| \leq \| F \| \cdot \| G \| \). However here without a proof, for positive semidefinite matrices the spectral radius is submultiplicative.

According to the above theorem on matrix commutations and eigenvalues, the matrices commute without changing the non-zero eigenvalues, and hence, the spectral radius

\[
\rho(\mathcal{A}) = \rho(X Q_h D Q_b)
\]

\[
\leq \rho(X R_h X^H R_w)^{-1}) \rho(D(D^H D)^{-1}) \rho(1)
\]

\[
= \rho(X R_h X^H (X R_h X^H + R_w)^{-1}).
\]

(33)

(34)

(35)

The partial matrices \( X R_h X^H \) and \( D(D^H D)^{-1} D^H \) are positive semidefinite, and \( (X R_h X^H + R_w)^{-1} \) is positive definite.

Hence, according to theorem 7.6.3 from [12] the product of a positive definite matrix and a positive semidefinite matrix is positive semidefinite. Thus together with the submultiplicativity, (34) follows. The spectral radius \( \rho(\mathcal{B}) = 1 \), because it shares the eigenvalues with \( (D(D^H D)^{-1} D^H \).

Intuitively, it is clear that due to the inherent damping according to the noise covariance, (35) has a spectral range \( \rho(X Q_h) < 1 \). Using the shortcut \( \mathcal{R} = X R_h X^H \), an prove follows from (35) if \( R_w = \sigma_w^2 I \) approximates the noise covariance

\[
\mathcal{R}(\mathcal{R} + R_w)^{-1} = U_R \Lambda_R U_R^H [U_R \Lambda_R U_R^H + \sigma_w^2 I]^{-1} = U_R \Lambda_R U_R^H [U_R \Lambda_R + \sigma_w^2 I] U_R^H]
\]

\[
\mathcal{R}(\mathcal{R} + R_w)^{-1} = U_R \Lambda_R [\Lambda_R + \sigma_w^2 I]^{-1} U_R^H
\]

(36)

such that the eigenvalues directly emerge as

\[
\lambda_{X Q_h,i} = \frac{\lambda_{R,i}}{\lambda_{R,i} + \sigma_w^2} < 1.
\]

(37)

Together with (35), this proves convergence.

If assuming knowledge on the interference occurrence \( P_b^{(\infty)} = 1 \) and an initialization \( \hat{h}^{(0)} = 0 \) the channel estimate is evaluated (see appendix A), the non-iterative solutions

\[
\hat{h}^{(\infty)} = (I - A)^{-1} Q_h (I - D Q_b) y
\]

\[
\hat{b}^{(\infty)} = (I - B)^{-1} Q_b (I - X Q_h) y
\]

(38)

(39)

result. Although the solutions (38) and (39) appear to be much simpler than the original iterative way in solving the problem, it is important to notice that the inherent matrix inversions to be calculated at runtime is at computational order \( O(K^3) \) compared to \( O(K^2) \) of the iterative solution, as \( D \) must still be assumed unknown. Additionally, the approach in (38) and (39) would completely abstract from the underlying interference detection problem.

However according to [2] and [13] given a sufficiently accurate initialization, convergence of EM estimates is assured anyway, such that the limits in (38) and (39) must exist even without the upper theorem \( \rho(\mathcal{A}) = \rho(\mathcal{B}) < 1 \).
802.15.3a CM4). Simulations will be based on hard decisions on the IF i.e. for the interference estimates, the equation (21) applies. The interferers are implemented (a) as a tone interferer at an arbitrarily chosen carrier frequency within the employed spectrum and (b) as modulated root-raised cosine (RRC) interferer with a bandwidth $B_i = 20$ MHz and roll off 0.25. Note that the losses with respect to a correct model tone and modulated interferer with a bandwidth this is no problem, for higher bandwidths the amount of critical IF spectra is low such that they can be computed and stored in advance.

**VI. CONCLUSION**

A joint interference detection and channel estimation scheme has been derived from EM-MAP theory, which gives foundation to a prior heuristic approach [1]. As main contribution, the estimator and its variance are analyzed theoretically, and simulations reconcile the theoretical result, proving the value of the approach. Space limitations hinder BER comparisons in here, interested readers shall refer to [1].

**APPENDIX**

A. Recursive Estimation

The derivation for $P_b^{(n)} \leq 1$ of the (in-)finite sums for the channel estimates goes as follows

\[
\hat{h}^{(0)} = 0 \quad \text{(initialization)}
\]

\[
\hat{h}^{(1)} = Q_h y - P_b^{(1)} A Q_h D Q_b y
\]

\[
\hat{h}^{(2)} = [(I + P_b^{(2)} A) Q_h + P_b^{(2)} (I + P_b^{(1)} A) Q_h D Q_b] y
\]

\[
\vdots
\]

\[
\hat{h}^{(n)} = \left[ I + \sum_{l=1}^{n-1} \prod_{m=n-l+1}^{n} P_b^{(m)} A^l \right] Q_h y
\]

\[
- \sum_{l=0}^{n-1} \prod_{m=n-l}^{n} P_b^{(m)} A^l Q_h D Q_b y
\]

(40)

and for the interference estimates in the same manner

\[
\hat{b}^{(1)} = P_b^{(1)} Q_b y
\]

\[
\hat{b}^{(2)} = P_b^{(2)} [(I + P_b^{(1)} B) Q_b - Q_b X Q_h] y
\]

\[
\vdots
\]

\[
\hat{b}^{(n)} = \left( \prod_{l=1}^{n-2} P_b^{(l)} B \right) Q_b (I - X Q_h) y
\]

\[
+ \prod_{m=1}^{n} P_b^{(m)} B^{n-1} Q_b y.
\]

(41)

Induction easily proves (40) and (41). For $P_b^{(n)} = 1$, i.e. interference presence is known for sure, the final formulas simplify significantly

\[
\hat{h}^{(n)} = \left( \sum_{l=0}^{n-1} A^l \right) Q_h (I - D Q_b) y
\]

(42)

\[
\hat{b}^{(n)} = \left( \sum_{l=0}^{n-2} B^l \right) Q_b (I - X Q_h) y + B^{n-1} Q_b y.
\]

(43)
variances become estimate (40) simplifies directly to the LMMSE channel estimator. Together with the convergence of channel... For the case $\hat{\beta}(0) = 0$ i.e. unlikely interference, (40) simplifies directly to the LMMSE channel estimator without interference. For $X = 0$, the simple, non-iterative ML interference estimator $Q_b y$ results. Both observations coincide with intuition.

### B. Error Covariance Computation

In the same manner, the derivation of the joint estimation error covariance starts from the initialization with a channel estimate $\hat{h}(0) = 0$. If we remember the identities $R_b = Q_b R_w Q_b^H$ and $R_h = Q_h (X R_h X^H + R_w) Q_h^H$, the covariances become

$$
\begin{align*}
R_b^{(0)} &= R_h \\
R_b^{(1)} &= Q_h X R_h X^H Q_h^H + R_w \\
R_h^{(1)} &= \mathcal{A} R_h \mathcal{A}^H + R_h + Q_h (D R_b D^H \\
&- D Q_b R_w - R_w Q_b^H D^H) Q_h^H \\
&+ (1 - Q_h X) R_h \mathcal{A}^H + \mathcal{A} R_h (1 - Q_h X)^H \\
R_b^{(2)} &= B Q_b X R_h X^H Q_b^H + (I + \mathcal{B}) R_b (I + \mathcal{B})^H \\
&+ Q_h X R_h X^H Q_h^H - (I + \mathcal{B}) Q_b R_w Q_b^H X^H Q_h^H Q_b^H \\
&- Q_h X^H Q_h^H R_w Q_b^H (I + \mathcal{B}) \\
&+ Q_b X (1 - Q_h X) R_h X^H Q_b^H B^H \\
&+ B Q_b X R_h (1 - Q_h X)^H X^H Q_b^H \\
R_h^{(2)} &= \mathcal{A}^2 R_h (\mathcal{A}^H)^H + (I + \mathcal{A}) R_h + Q_h (D R_b D^H \\
&- R_w Q_b^H D^H - D Q_b R_w) Q_b^H (I + \mathcal{A})^H \\
&+ (I + \mathcal{A}) (1 - Q_h X) R_h (\mathcal{A}^H)^H \\
&+ \mathcal{A}^2 R_h (1 - Q_h X)^H (I + \mathcal{A})^H \\
R_b^{(n)} &= B^{n-1} Q_b X R_h X^H Q_b^H (B^{n-1})^H + (\sum_{l=0}^{n-1} B^l) R_b (\sum_{l=0}^{n-1} B^l)^H \\
&+ (\sum_{l=0}^{n-2} B^l) Q_b X R_h X^H Q_b^H (\sum_{l=0}^{n-2} B^l)^H \\
&- (\sum_{l=0}^{n-2} B^l) Q_b R_w Q_b^H X^H Q_b^H (\sum_{l=0}^{n-2} B^l)^H \\
&- (\sum_{l=0}^{n-2} B^l) Q_b X Q_h R_w Q_b^H (\sum_{l=0}^{n-2} B^l)^H \\
&+ (\sum_{l=0}^{n-2} B^l) (1 - Q_h X) R_h (B^{n-1})^H \\
&+ (B^{n-1})^H R_h (1 - Q_h X)^H (\sum_{l=0}^{n-2} B^l)^H \\
R_h^{(n)} &= \mathcal{A}^n R_h (\mathcal{A}^n)^H + \left( \sum_{l=0}^{n-1} \mathcal{A}^l \right) [R_h + Q_h (D R_b D^H \\
&- R_w Q_b^H D^H - D Q_b R_w) Q_b^H] \left( \sum_{l=0}^{n-1} \mathcal{A}^l \right)^H \\
&+ \left( \sum_{l=0}^{n-1} \mathcal{A}^l \right) (1 - Q_h X) R_h (\mathcal{A}^{n-1})^H \\
&+ \mathcal{A}^{n-1} R_h (1 - Q_h X)^H \left( \sum_{l=0}^{n-1} \mathcal{A}^l \right)^H .
\end{align*}
$$

These error covariances are generally valid expressions. For $n \to \infty$, (28) and (29) express them in finite closed forms, if furthermore exploiting

$$
\sum_{l=0}^{\infty} T^l = (I - T)^{-1} .
$$

This is valid, if the sum converges, i.e. if the spectral radius $\rho(T) < 1$ (cf. section IV for $\rho(\mathcal{A}) < 1$).

### REFERENCES


