Joint deconvolution and classification with applications to passive acoustic underwater multipath

Hyrum S. Anderson[a] and Maya R. Gupta[b]

Department of Electrical and Computer Engineering, University of Washington, Seattle, Washington 98195

This paper addresses the problem of classifying signals that have been corrupted by noise and unknown linear time-invariant (LTI) filtering such as multipath, given labeled uncorrupted training signals. A *maximum a posteriori* approach to the deconvolution and classification is considered, which produces estimates of the desired signal, the unknown channel, and the class label. For cases in which only a class label is needed, the classification accuracy can be improved by not committing to an estimate of the channel or signal. A variant of the quadratic discriminant analysis (QDA) classifier is proposed that probabilistically accounts for the unknown LTI filtering, and which avoids deconvolution. The proposed QDA classifier can work either directly on the signal or on features whose transformation by LTI filtering can be analyzed; as an example a classifier for subband-power features is derived. Results on simulated data and real Bowhead whale vocalizations show that jointly considering deconvolution with classification can dramatically improve classification performance over traditional methods over a range of signal-to-noise ratios.
© 2008 Acoustical Society of America. [DOI: 10.1121/1.2981046]

I. INTRODUCTION

Many signal processing applications require classifying a signal $z(t)$ that has been corrupted by an unknown linear time-invariant (LTI) filter $h(t)$,

$$z(t) = h(t) * x(t) + w(t), \tag{1}$$

where $x(t)$ is the signal of interest, $w(t)$ is additive noise, and $*$ denotes convolution. For example, the LTI filter $h(t)$ could model seismic reflections of an impulsive source, blurring of celestial bodies observed through the Earth's atmosphere, or the effects of an underwater channel on sound propagation. We assume that n labeled training pairs $\{x_i(t), y_i\}$, $i = 1, \ldots, n$, are available to classify the observed signal $z(t)$. As is standard in classifier theory, we assume each $x_i(t)$ and its class label y_i are drawn independently and identically (i.i.d.) from the same joint distribution as the test signal $x(t)$ and its unknown label y. This paper presents joint deconvolution and classification methods, in which the existence of training data can inform an otherwise blind-deconvolution problem. From a classification perspective, the challenge is to deal with the mismatch between training pairs $\{x_i(t), y_i\}$ provided in signal space ("$x(t)$ space" or x-space) and the observed signal $z(t)$ in measurement space (z-space).

The framework developed in this paper will apply to any random LTI filtering, but our emphasis will be on multipath, which can be modeled by an impulse response with sparse coefficients that generally decay with time. Multipath affects many sensing modalities (e.g., ultrasound, radar, terahertz imaging); in this work we present experiments for classifying passive acoustic signals corrupted by multipath in a shallow ocean channel using a single hydrophone at low signal-to-noise ratios (SNRs). For the passive sonar problem, $z(t)$ represents the in-channel received signal, $h(t)$ represents the multipath, and $x(t)$ is the free-field signal. Underwater multipath channels are generally time-varying, that is, the multipath $h(t)$ in Eq. (1) changes between successive transmissions, but not during the transmission. Multipath is highly sensitive to the locations of the source and receiver, making it difficult to model effectively.[1,2] In this paper, we account for the uncertainty in the channel by treating $h(t)$ as a random process.

The main contributions of this paper include: (1) a unified *maximum a posteriori* (MAP) framework for deconvolution and classification for multipath, (2) a quadratic discriminant analysis (QDA) classifier that probabilistically takes into account unknown LTI filtering, (3) and a comparison of feature-based classifiers for marine mammal identification using real whale vocalizations in an acoustically accurate multipath environment.

First, we review related research in Sec. I A. Then in Sec. II, we unify deconvolution and classification in a joint MAP framework. This method jointly estimates a clean signal $\hat{x}(t)$, a channel estimate $\hat{h}(t)$ and a class label y^*. In Sec. III, we argue that if signal estimate $\hat{x}(t)$ is not needed, better classification performance can be achieved by not committing to a particular signal or channel estimate. We show how a QDA classifier can be designed to incorporate the effects of uncertain $h(t)$. The joint MAP deconvolution/classification (joint MAP) and joint QDA deconvolution/classification (joint QDA) methods presented in Secs. II and III classify $z(t)$ based on the entire time signal. We show in Sec. IV how to extend the joint QDA classifier for use with subband power features of $z(t)$. We demonstrate the importance of taking into account the second-order statistics using simulated multipath with mock signals and with real marine

---

[a]Electronic mail: hyrum@ee.washington.edu
[b]Electronic mail: gupta@ee.washington.edu

mammal vocalizations. We conclude in Sec. V with a discussion about the techniques presented and suggest directions for future research.

A. Related work on classifying signals corrupted by unknown LTI filtering

Signal processing researchers in underwater passive acoustics have considered the problem of classifying signals corrupted by multipath for over 30 years.[3] Ehrenberg *et al.* demonstrated in an ocean acoustic propagation experiment that multipath effects generally cannot be ignored, and that simple time-gating of the received signal can discard too much of the signal information for classification.[4,5] Multipath induced by a shallow ocean channel presents an additional challenge in that the multipath propagation is generally time varying and the structure of $h(t)$ is sensitive to spatial location, making it difficult to estimate or model effectively.[1,2]

A review of the literature reveals four general strategies for classifying signals corrupted with multipath. The first is to extract features from training signals $\{x_i(t)\}$ and the received signal $z(t)$ that are invariant to multipath distortion, then classify based on the multipath-invariant features. Shin *et al.* consider a number of time-frequency features for clutter rejection.[6] Strausberger *et al.* compare different distance measures for 1-nearest-neighbor (1-NN) classification of signals passed through Rician channels for over-the-horizon radar.[7] Recently, Okopal and Loughlin developed features invariant to channel dispersion and dissipation and demonstrated superior classification performance compared to temporal and spectral moment features.[8] In general, classification using channel invariant features can provide good results inasmuch as the classes are well-separated in the designated feature space.

Blind deconvolution is the basis for a second commonly used approach for classifying $z(t)$: a clean signal $\hat{x}(t)$ is estimated from $z(t)$, then a classifier is used on features of $\hat{x}(t)$. There are many examples of trying to remove multipath by blind deconvolution in order to classify.[9–15] Some researchers exploit the sparseness of the unknown $h(t)$ for producing an estimate $\hat{x}(t)$.[13–15] Broadhead and Pflug report[13] excellent correlation between true signals and signals blindly deconvolved by the minimum entropy method with Cabrelli's sparsity criterion,[16] but did not consider classification. We have shown that these blind deconvolution estimates can be highly correlated to out-of-class training signals, so that nearest neighbor classification on correlation scores performs poorly, particularly at low signal-to-noise ratios.[17,18]

A third approach is to predict the z-space representation of the training signals $\{x_i(t)\}$ using a forward model for the multipath $\hat{h}(t)$. This has the advantage of avoiding deconvolution. A classifier is built using estimated training signals $z_i(t) = x_i(t) * \hat{h}(t)$ for $i = 1, \ldots, n$ to classify $z(t)$. The forward model $\hat{h}(t)$ has been based on geometry or physical assumptions.[9,19] Liu *et al.* first proposed an in-channel classifier based on free-field training data.[9] They build a classifier by assuming a finite number of multipath reflections for near-bottom target classification. Dasgupta and Carin classify

after accounting for multipath via time-reversal imaging, which requires the geometry and sound speed profile of the channel.[19]

We first proposed that to classify a signal corrupted by unknown multipath, jointly considering deconvolution and classification can lead to better performance than traditional approaches that deconvolve then classify in independent steps.[17] Our method leveraged training data to produce a multipath channel candidate $\hat{h}_i(t)$ for each training signal $x_i(t)$ given $z(t)$. Then, a nearest-neighbor classifier chose the class $y^* = y_i$ for which the estimated filter $\hat{h}_i(t)$ was most multipath-like, according to Cabrelli's sparsity criterion.[16] The resulting joint deconvolution and classification method yields the best signal estimate $\hat{x}(t) = x_i(t)$ and filter estimate $\hat{h}(t) = \hat{h}_i(t)$ that may have produced $z(t)$, as well as the optimal class label $y^* = y_i$. Classification performance was markedly better than minimum entropy blind deconvolution followed by classification, particularly at low signal-to-noise ratios. However, the performance of that joint deconvolution/classifier relied on several conditions.[17] First, it required a good criterion for evaluating how well a given $\hat{h}(t)$ represented a multipath filter. Although Cabrelli's sparsity criterion is an intuitive and convenient choice, real multipath filters can violate the maximal sparsity assumption.[9] Second, the proposed nearest-neighbor approach required that the training signals $\{x_i(t)\}$ be plentiful and that the true $x(t)$ be close to a training sample of the correct class in terms of $\|x(t) - x_i(t)\|$. Third, the deconvolution estimate $\hat{x}(t)$ was always restricted to be a member of the set $\{x_i(t)\}$. Last, it is not straightforward to incorporate features in classification.

II. JOINT MAP DECONVOLUTION AND CLASSIFICATION

A joint deconvolution and classification estimate of the signal, filter, and class label can be constructed using the MAP criterion. Let vectors z, x, h, and w be critically sampled versions of signals $z(t)$, $x(t)$, $h(t)$, and $w(t)$. In this section, we assume that x and w are realizations of random vectors drawn from independent Gaussian distributions. Real signals x certainly may possess non-Gaussian characteristics but the Gaussian assumption is critical to keeping an otherwise formidable deconvolution problem tractable. We assume that the distribution of w is zero-mean with diagonal covariance matrix $\sigma_w^2 I$, where I is the identity matrix; that the probability of x conditioned on the class label y has mean $\mu_{x|y}$ and covariance $\Sigma_{x|y}$; and that the distributions of x, h, and w are mutually independent. We model h using a multivariate Laplacian distribution with independent dimensions, so that the ith element of the random multipath has mean $\mu_h[i]$ and scale parameter $b[i]$. The Laplacian distribution is an appropriate prior model for multipath since it yields sparse realizations. Let $\theta = \{\mu_{x|y}, \Sigma_{x|y}, \mu_h, b, \sigma_w\}$ be the set of parameters for these three distributions, where θ is assumed to have been modeled or estimated *a priori*.

Then, the proposed joint MAP class estimate y^* solves

$$y^* = \arg \max_y (\max_{x,h} p(x,h,y|z,\theta)) \qquad (2)$$

H. S. Anderson and M. R. Gupta: Joint deconvolution and classification

$$=\arg\max_y[\max_{x,h}(p(z|x,h,y,\theta)p(h|\theta)p(x|y,\theta))p(y)] \quad (3)$$

$$=\arg\min_y\left[\min_{x,h}\left(\|z-h*x\|^2+\sigma_w^2\|x-\mu_{x|y}\|^2_{\Sigma_{x|y}^{-1}}\right.\right.$$
$$\left.\left.+2\sigma_w^2\sum_i\frac{|h_i-\mu_i|}{b_i}\right)+\sigma_w^2\log|\Sigma_{x|y}|-2\sigma_w^2\log p(y)\right], \quad (4)$$

where Eq. (3) follows from Eq. (2) using Bayes' rule, the chain rule, and independence assumptions; and Eq. (4) follows from Eq. (3) by taking the negative logarithm of the pdfs, removing constants that do not depend on x, h, or y from the arg min, and scaling each term by $2\sigma_w^2$. Throughout the paper, we use the notation $\|x\|$ to denote the ℓ_2 norm and $\|x\|_A^2$ for $x^T A x$. The $\|z-h*x\|^2$ term in Eq. (4) drives the estimated filter h and test signal x to be consistent with the received signal z in terms of squared error. The next two terms in Eq. (4) drive x to match the *a priori* expected signal via the ℓ_2 norm and drive h to match the *a priori* expected filter via the ℓ_1 norm. Note that these latter two terms are regularized by the noise variance—the greater the noise power, the more the estimate relies on the *a priori* expectations and less on matching the received signal $z(t)$. The fourth term penalizes classes that exhibit high variance, and the fifth term is the class membership prior. Since the noise determines the degree of regularization, a curious behavior of this approach is that it performs poorly for high SNR: the first term will dominate as σ_w^2 goes to zero, and solutions for x and h will no longer depend on $\mu_{x|y}$ and μ_h, respectively.

The objective function in Eq. (4) is not convex since it involves a product of variables in the convolution integral $h*x$. However, the problem is jointly convex in h and x in the limit as $\sigma_w\to\infty$, and is marginally convex in x or h for all σ_w. Therefore, we opt to solve Eq. (4) using an alternating minimization approach as a heuristic for finding the true solution. Using H for the Toeplitz matrix representation of discrete convolution with fixed h,[20] and for fixed y, the objective as a function of x can be written in the form of generalized Tikhonov regularization $\|Hx-z\|^2+\sigma_w^2\|h-\mu_h\|^2_{\Sigma_h^{-1}}$. The solution[21] is

$$\hat{x}=(H^TH+\sigma_n^2\Sigma_{x|y}^{-1})^{-1}(H^Tz+\sigma_w^2\Sigma_{x|y}^{-1}\mu_{x|y}).$$

Next, we solve Eq. (4) for those terms depending on h by fixing x and rewriting as

$$\hat{h}=\arg\min_h\|Xh-z\|_2^2+\sigma_w^2\|D^{-1}(h-\mu_h)\|_1, \quad (5)$$

where X is the Toeplitz matrix representation of discrete convolution of x with h, D is a diagonal matrix with entries $b_i/2$, and $\|\cdot\|_1$ is the ℓ_1 norm. Equation (5) can reformulated as a quadratic program with linear constraints.[22]

Since the optimization problem in Eq. (4) is non-convex, the alternating minimizations strategy is not guaranteed to converge to the global minimum.[23] A common approach is to optimize starting from several initial points, then choose the overall minimizer. The initial guesses could be drawn i.i.d. from the class-conditional distribution $\mathcal{N}(\mu_{x|y},\Sigma_{x|y})$. We use a slightly different approach to take advantage of the fact that we have examples from each class: convex combinations of the training signals are initial guesses.

Experiments and results for the joint MAP classifier are presented in Sec. III.

A. A related MAP deconvolution approach

MAP deconvolution has been explored previously by Lam and Goodman for blind image deblurring (without classification).[24] In that work, Lam and Goodman estimate the point spread function h and the image covariance Σ_x by maximizing $p(z|h,\Sigma_x)p(h)p(\Sigma_x)$. The prior $p(\Sigma_x)$ is replaced with a heuristic smoothness constraint on the covariance, and the prior $p(h)$ is replaced with the hard constraint $h\in\mathcal{H}$ for some convex set $\mathcal{H}$. They propose an EM algorithm implementation that alternates between estimating Σ_x (the E-step) and h (the M-step) in the Fourier domain. The image is finally estimated by Wiener deconvolution using the estimated h and Σ_x. The algorithm results in high-quality deblurred image estimates.[24]

The Lam and Goodman MAP blind deconvolution may be extended to the multipath problem by applying a Laplacian prior for $p(h)$ as we have done in Eq. (4) instead of their hard constraint $h\in\mathcal{H}$. However, their approach cannot be extended to fit in the joint deconvolution/classification paradigm by simply conditioning on the class label y and adding the prior $p(y)$ in the optimization. First, Lam and Goodman assume that the image x (and therefore, z) is a realization of a zero-mean Gaussian distribution. In our framework, the class-conditional mean is an important discriminating feature of the class. Second, their method must estimate Σ_x, but in our framework the class conditional covariance $\Sigma_{x|y}$ is estimated *a priori* from training pairs. Naively replacing Σ_x with $\Sigma_{x|y}$ renders their E-step useless so that iterating does not improve the initial guess. Thus, the training pairs $\{x_i,y_i\}$ offer little advantage to their MAP blind deconvolution technique.

III. PROBABILISTIC DECONVOLUTION AND CLASSIFICATION USING QDA

Estimating the true signal is difficult and unnecessary if only a class label is required. In this section, we explore classifying signals jointly with probabilistic deconvolution, in which a statistical characterization of $x(t)$ and $h(t)$ are used without ever choosing a particular, deterministic signal or channel estimate. Specifically, we consider the maximum likelihood classifier that solves

$$y^*=\arg\max_y p(z|y)$$
$$=\arg\max_y\int\int p(z|x,h,y)p(x|y)p(h)p(y)\,dx\,dh. \quad (6)$$

Assuming a uniform prior, the classifier in Eq. (6) differs from the joint MAP classifier in Eq. (2) in that the $\max_{x,h}$ operator in Eq. (2) is replaced by expectation over x and h in Eq. (6). For the remainder of this paper, we will assume uniform prior $p(y)$ such that $\arg\max_y p(y|z)=\arg\max_y p(z|y)$.

QDA is a popular classification rule that models each class-conditional distribution in Eq. (6) as Gaussian.[25–27] This assumption can be motivated by the central limit theorem and the fact that the Gaussian is the maximum entropy (least assumptive) distribution given first and second moments. Here, we build a QDA classifier in z-space by assuming $p(z|y)$ in Eq. (6) is Gaussian, and we show that one can calculate the sufficient statistics $\mu_{z|y}$ and $\Sigma_{z|y}$ of $p(z|y)$ from the estimated mean and covariance of the channel $h(t)$ and the estimated means and covariances of the training signals from each class. Note that we do not make any assumptions on the distributions of h or x given y other than that they have finite first and second moments; in fact, the result of the convolution $h*x$ would not be Gaussian if h and x were assumed to be realizations of Gaussian processes. This section expands on a recent workshop paper.[18]

A. QDA classification of signals corrupted by LTI filtering

Let x be a realization of a random process with finite class-conditional mean $\mu_{x|y}$ and finite covariance $\Sigma_{x|y}$; let the noise w be a realization of a zero-mean random process with covariance $\sigma_w^2 I$; and let h be a realization of a random process with mean μ_h and covariance Σ_h. Given these statistics, the full derivations of the first and second moments of the distribution of z are given in the Appendix. In summary, the first moment $\mu_{z|y}$ can be written in terms of the mean class-conditional signal and mean filter as

$$\mu_{z|y} = \mu_{x|y} * \mu_h. \tag{7}$$

Likewise, the covariance $\Sigma_{z|y}$ conditioned on class y can be expressed in terms of the class-conditional signal and filter statistics as

$$\Sigma_{z|y} = (\Sigma_{x|y} + \mu_{x|y}\mu_{x|y}^T) ** (\Sigma_h + \mu_h\mu_h^T) + \sigma_w^2 I - \mu_{z|y}\mu_{z|y}^T, \tag{8}$$

where $**$ denotes two-dimensional discrete convolution.

B. Experiments with signal-based joint QDA and joint MAP classification

We tested the proposed methods with two experiments that differ in how we generate the simulated multipath. In the first we use a Laplacian random process to generate realizations of multipath channels. In the second, a random K-sparse model is used. In both experiments the clean, x-space signals are drawn from class-conditional Gaussian distributions, which means that the received signals in z-space are not actually Gaussian distributed.

1. Signal classification experiment: Laplacian multipath

Each coefficient of a multipath filter was drawn independently from a Laplacian random process with parameters $\mu[n]$, $b[n]$,

TABLE I. Simulation parameters for joint MAP/joint QDA experiments. Note that $\mathrm{square}(n) = \mathrm{sgn}(\sin(n))$.

Parameter	Class 1	Class 2
	Well-separated means	
$\mu_{x\|y}[n]$	$\frac{1}{4}\,\mathrm{square}(6\pi n/100)$	$\frac{1}{4}\,\mathrm{square}(12\pi n/100)$
$\Sigma_{x\|y}[m,n]$	$\frac{1}{100}(\delta[m-n] + e^{-\|m-n\|/20})$	$\frac{1}{100}(\delta[m-n] + e^{-(m-n)^2/10})$
	Close means	
$\mu_{x\|y}[n]$	$\frac{1}{4}\,\mathrm{square}(6\pi n/100)$	$\frac{1}{4}\,\sin(6\pi n/100)$
$\Sigma_{x\|y}[m,n]$	$\frac{1}{100}(\delta[m-n] + e^{-\|m-n\|/20})$	$\frac{1}{100}(\delta[m-n] + e^{-(m-n)^2/10})$

$$p(h[n]|\mu[n],b[n]) = \frac{1}{2b[n]}e^{-(|x-\mu[n]|)/b[n]},$$

for $n = 0, \ldots, 99$, where we set the expected filter to be $\mu_h[n] = \delta[n] - 0.6[n-49] + 0.1\delta[n-99]$, and the scale parameter $b[n]$ decays as n grows: $b[n] = 0.2e^{-0.024n}$. The decay parameter coefficients for this experiment were chosen to model oceanic multipath filtering of sonar signals.

Test and training signals were drawn i.i.d. from a Gaussian distribution $\mathcal{N}(\mu_{x|y}, \Sigma_{x|y})$ where the class y was drawn uniformly between two classes. We considered two classification scenarios to test performance: classes whose mean vectors are well-separated, and close. The mean signals are composed of square and sine waves, and the covariance matrices are Toeplitz with smooth covariance structure. The specific values of $\mu_{x|y}$ and $\Sigma_{x|y}$ for each experiment are shown in Table I. Each test signal z was created by convolving a randomly drawn signal x with randomly drawn multipath h, and adding Gaussian white noise w to achieve different SNRs, where the SNR is with respect to the multipath signal, $20 \log \|x*h\|/\sigma_w$.

We compared the joint QDA classifier to a matched filter that ignores multipath. For the matched filter, the received signal z is tested against $\mu_{x|y}$ for each class. The joint MAP classifier in Eq. (4) is compared to a matched filter on a blind deconvolution signal estimate. For deconvolution, the received signal z is first denoised by Wiener filtering, then $\hat{h}$ is estimated using Cabrelli's blind deconvolution method for signals that have undergone unknown multipath filtering.[16] The estimate $\hat{x}$ is then computed via deconvolution in the Fourier domain. The true signal length was used as a required input to Cabrelli's method. Each of the methods used the true signal and channel statistics, and the true SNR where needed.

2. Signal classification experiment: K-sparse multipath

The K-sparse experiments are the same as described in the previous subsection, except the multipath filters were generated using a sparse model:

H. S. Anderson and M. R. Gupta: Joint deconvolution and classification
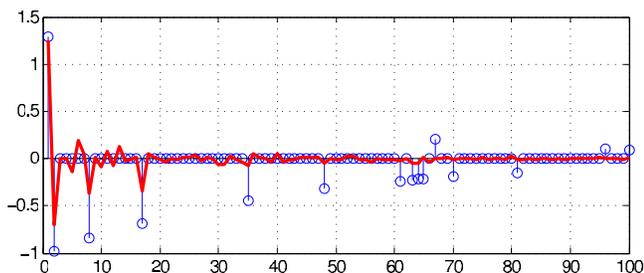
FIG. 1. (Color online) Example multipath realization from the K-sparse model (stem), and the deconvolution estimate produced by the joint MAP deconvolution/classifier (solid) at 10 dB SNR.

$$h[n] = \sum_{i=1}^{K} \alpha_i \delta[n - d_i],$$

with $K = 15$ nonzero coefficients, delays d_i drawn uniformly on [0,99], $\alpha_i = \pm e^{-\beta d_i}$ with randomly chosen sign, and decay parameter $\beta = 0.0240$ chosen to mimic real underwater acoustic channels. An example realization of a filter h drawn from this model is shown in Fig. 1. The diagonal covariance matrix Σ_h is estimated from 1000 samples of the impulse response.

C. Signal-based joint QDA and joint MAP results

Figure 1 shows a reconstructed multipath estimate produced by the joint MAP deconvolution/classifier corresponding to the chosen class for the well-separated means experiment at 10 dB SNR. In this case, joint MAP correctly identified the class label. The recovered filter is a reasonable reconstruction of the true filter, but generally underestimates the amplitude of the first coefficients, and does not reliably reconstruct the tail of h. The gross errors can be ascribed to the fact that the optimization problem in Eq. (4) is not convex, and to the mismatch between the Laplacian prior and K-sparse model.

Classification results in Fig. 2 show that the proposed joint QDA classifier dominates the matched filter classifier for both Laplacian multipath in (a) and (b), and for multipath generated by the K-sparse model in (c) and (d). The means for each class used for (a) and (c) (well-separated means) are orthogonal, so the matched filter performs well despite ignoring the multipath. However, for (b) and (d) the means are similar, and the matched filter performs poorly compared to joint QDA. The joint MAP classifier performs well at low SNR, but as predicted, performance degrades as SNR increases. For truly sparse multipath in (c) and (d), the joint MAP approach is unaffected for well-separated means in (c), and affected moderately at high SNR for close means in (d) compared to results using Laplacian multipath. Evidently, the ℓ_1 is an appropriate heuristic for the K-sparse multipath model; using ℓ_1 criteria to obtain sparse solutions is a popular approach.[28]

IV. QDA CLASSIFICATION OF SIGNALS CORRUPTED BY LTI FILTERING USING FEATURES

Blind signal deconvolution and the proposed joint MAP and signal-based joint QDA methods are computationally
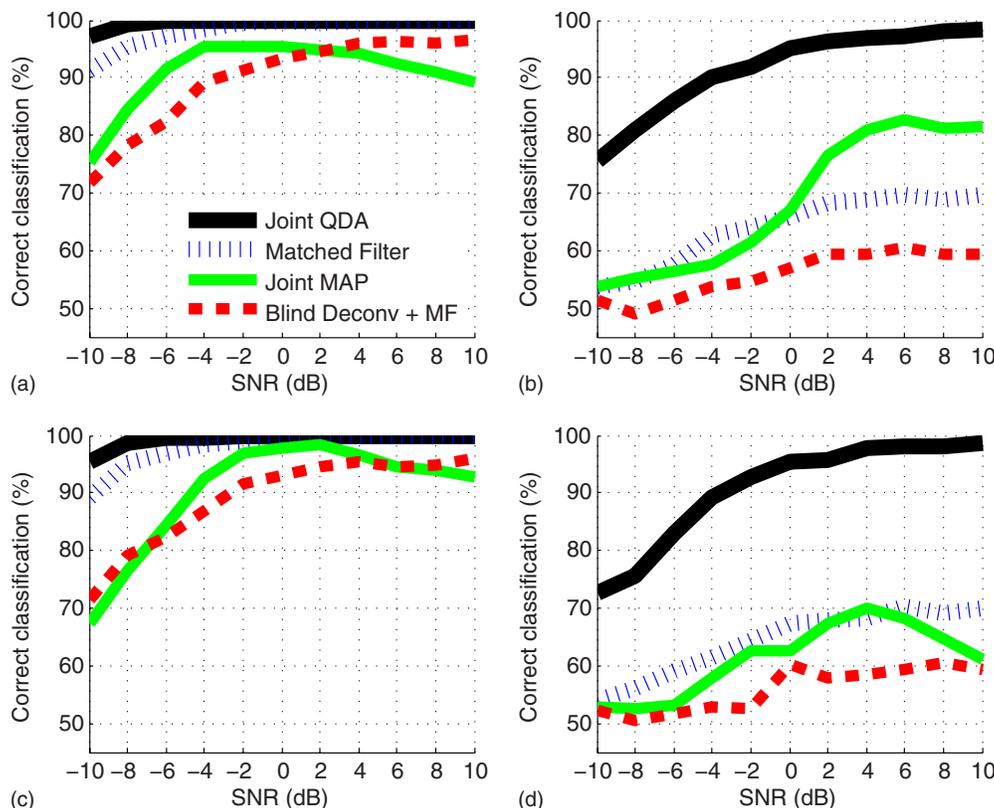
FIG. 2. (Color online) Classification accuracy for four experiments using multipath generated from a Laplacian model in (a) and (b), and a K-sparse model in (c) and (d). The results are averaged over 1000 i.i.d. test signals for each SNR point.

prohibitive for signals captured at high sample rates. For M-length sampled signals, Cabrelli's method requires the inversion of an $M \times M$ Toeplitz matrix, which at best is of complexity $O(M \log M)$. The joint MAP deconvolution requires more iterations to converge as M increases, and requires (as does joint QDA) inversion of an $M \times M$ covariance matrix, which in general is complexity $O(M^3)$. To decrease the computational burden and possibly increase classification performance, an alternative is to classify based on features that represent important characteristics of the signals and provide good class discrimination.[29] The hope is that classes can be well-discriminated by features of significantly smaller dimensionality $d \ll M$.

Unless multipath-invariant features are used, a classifier trained on x-space features will not generally be applicable to classify the features of $z(t)$ directly. However, if a functional relationship can be found that relates x-space features to their images in z-space, then a suitable classifier can be trained and applied in z-space. We show that such a relationship can be derived for subband power features, which represent an important class of discriminating features for many remote-sensing applications. Extending the proposed joint QDA classifier to subband features in z-space requires expressing the mean and covariance of the subband power feature vector for $z(t)$ in terms of statistics of the subband power features of the channel and training data.

Let the d-dimensional feature vector $P_z = [P_z(f_1) \cdots P_z(f_d)]^T$ be composed of the subband powers of $z(t)$ at frequencies $\{f_i\}$ for $i = 1, \ldots, d$. Then for any frequency f, because $P_z(f) = Z(f)Z^*(f)$, and $Z(f) = X(f)H(f) + W(f)$, the subband power can be expressed as

$$P_z(f) = P_x(f)P_h(f) + P_w(f) + 2 \, \text{Re}\{X(f)H(f)W^*(f)\}, \quad (9)$$

where $P_x(f)$, $P_h(f)$, and $P_w(f)$ denote the power of the signal, the channel, and the noise for frequency f, respectively. Based on Eq. (9), derivations for the mean and covariance of the feature vector P_z are given in the Appendix. For these derivations, it is assumed that $w(t)$ is a realization of a Gaussian white noise process. The class-conditional mean feature vector $\mu_{P_z|y}$ can be expressed in terms of the noise power σ_w^2 and mean vectors of the clean signal features and channel features, $\mu_{P_x|y}$ and μ_{P_h}, respectively, as

$$\mu_{P_z|y} = \mu_{P_x|y} \cdot \mu_{P_h} + \sigma_w^2 \mathbf{1}, \quad (10)$$

where $\cdot$ denotes Hadamard (Schur or element-wise) multiplication, and $\mathbf{1}$ is a vector of ones. The class-conditional covariance $\Sigma_{P_z|y}$ can be expressed in terms of σ_w^2 and second-order statistics of $P_x|y$ and P_h as

$$\Sigma_{P_z|y} = \Sigma_{P_x|y} \cdot \Sigma_{P_h} + \sigma_w^4 I + \Sigma_{P_h} \cdot \mu_{P_x|y}\mu_{P_x|y}^T + \Sigma_{P_x} \cdot \mu_{P_h}\mu_{P_h}^T$$
$$+ 2\sigma_w^2 \, \text{diag}\{\mu_{P_x|y} \cdot \mu_{P_h}\}. \quad (11)$$

It is assumed that μ_{P_h}, Σ_{P_h}, and σ_w^2 can be estimated from the channel. The statistics $\mu_{P_x|y}$ and $\Sigma_{P_x|y}$ are estimated from training data for each class. Together, these statistics are used to compute class-conditional QDA model parameters $\mu_{P_z|y}$ and $\Sigma_{P_z|y}$ from Eqs. (10) and (11) and, hence, build a QDA subband power feature classifier in the z-space.

A. Informed and blind classifiers using features

To evaluate the proposed z-space joint QDA using power features, we consider two alternate approaches to classification from power features. First, an "informed" x-space classifier approach uses the expectation of the channel's subband power response $E[P_h]$ and the noise power σ_w^2 to transform the z-space features to x-space by subtracting the noise power and deconvolving by $E[P_h]$: $\hat{P}_x(f_i) = (P_z(f_i) - \sigma_w^2)/E[P_h(f_i)]$. This approach uses the mean but not covariance of the channel.

Second, a "blind" x-space classifier approach uses x-space features, and simply ignores statistics of the channel and noise altogether. Training signals are first normalized by their total power before a classifier is trained. Then, a received signal $z(t)$ is also normalized by its total power before extracting features.

We compare joint QDA to both blind and informed approaches for QDA, 1-NN, and a support vector machine (SVM).[25]

B. Simulated classification experiments using joint QDA with power features

We consider the task of classifying narrow-band signals corrupted by unknown multipath due to propagation in a
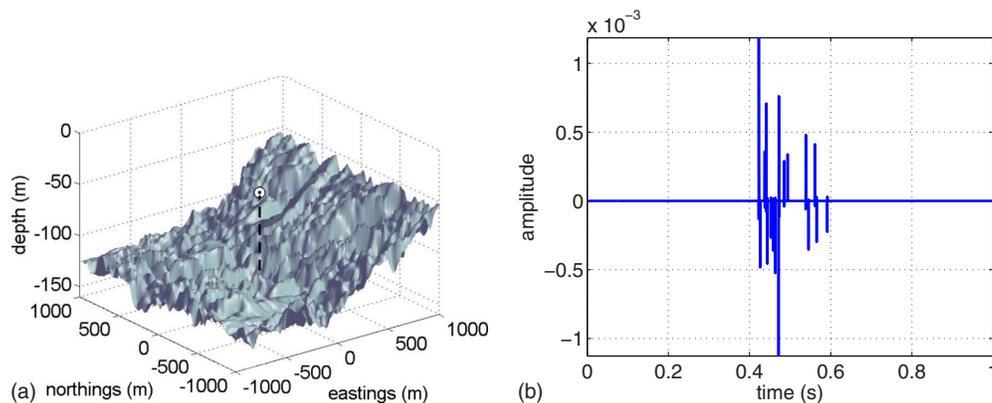
FIG. 3. (Color online) (a) Simulated ocean bathymetry with a single receiver (marked by ⊙) at $(0, 0, -50)$ m, and (b) a sample channel impulse response for a source located at $(460, 250, -70)$ m, generated by the Sonar Simulation Toolset (Refs. 30 and 31).

H. S. Anderson and M. R. Gupta: Joint deconvolution and classification

TABLE II. Pole magnitude distribution for feature-based classification experiments.

Parameter	Class 1	Class 2
$\Sigma_{a\|y}$	$\begin{pmatrix} 1.00 & 0.99 \\ 0.99 & 9.00 \end{pmatrix} \times 10^{-6}$	$\begin{pmatrix} 6.00 & -0.80 \\ -0.80 & 1.00 \end{pmatrix} \times 10^{-4}$
	Close poles	
$\mu_{a\|y}$	(0.945 0.905)	(0.909 0.948)
	Moderately-separated poles	
$\mu_{a\|y}$	(0.945 0.875)	(0.879 0.948)
	Well-separated poles	
$\mu_{a\|y}$	(0.965 0.875)	(0.875 0.948)

shallow ocean channel. To simulate two classes of narrowband signals with two subband power features, training and test signals are generated i.i.d. using the z-domain model

$$X|y[z] = \frac{(z-1)(z+1)}{\Pi_{\ell=1}^2 (z-p_{\ell|y})(z-p_{\ell|y}^*)}, \quad (12)$$

where the location of each class-conditional pole $p_{\ell|y}$, $\ell \in \{1,2\}$ is drawn randomly from the model $a_{\ell|y} \exp(j\theta_\ell)$, where θ_ℓ is fixed, and $a_{\ell|y} = [a_{1|y}, a_{2|y}]^T$ is multivariate Gaussian distributed with mean $\mu_{a|y}$ and covariance matrix $\Sigma_{a|y}$. Although the vector a for each class is Gaussian distributed, the signals in feature space are not, as per Eq. (12). Figure 4 shows an example pole-zero plot and corresponding log-feature space scatterplot for a well-separated case. We consider three instances of the experiment for choices of $\mu_{a|y}$ that result in different class separation. The parameters $\mu_{a|y}$ and $\Sigma_{a|y}$ for each instance of the experiment are shown in Table II, and $\theta_1 = \frac{1}{50}$ and $\theta_2 = \frac{1}{5}$. Note that since all poles and zeros lie within the unit circle, for each case the selected parameters correspond to a realization of a minimum phase signal, which could be produced from natural sources.

Test and training signals were generated by taking i.i.d. draws of poles as described above, and taking 5000 evenly spaced samples around the unit circle of the complex z-plane, so that the length of each signal corresponds to

1.25 s, sampled at 4 kHz. At frequencies θ_1 and θ_2 the subband power is extracted from each signal and used as classification features. The parameters in Table II were chosen such that the generated test and training signals were linearly separable in the subband power feature space.

Channel impulse responses h were drawn i.i.d. in the following manner. A receiver is placed at a depth of 50 m in a simulated shallow water channel, as shown in Fig. 3. Source locations were drawn uniformly from the cube 2 km across north and east and 150 m deep; locations falling below the ocean floor are discarded and redrawn. Channel impulse responses h were generated by propagating an impulsive source from the random source locations to the receiver using the *CASS Eigenray* routine provided in the Sonar Simulation Toolset.[30] Impulse responses were sampled at 4 kHz. The ocean environment is set up to be fairly extreme, but static. We have imposed a prototypical sound speed profile (ranging from 1477 to 1492 m/s), and have modeled the ocean bottom to contain sandy gravel with mean grain size 2 mm. Surface roughness is governed by the wind speed, which is set to 15 km/h. The channel geometry and a sample channel impulse response are shown in Fig. 3.

Maximum likelihood estimates of $\mu_{P_x|y}$ and $\Sigma_{P_x|y}$ were computed from 1000 randomly drawn training signals, and maximum likelihood estimates of μ_{P_h} and Σ_{P_h} were computed from 1000 randomly drawn channel impulse responses. The test samples were corrupted with randomly drawn multipath, and then i.i.d. white noise with variance σ_w^2 was added, where the multipath-corrupted-signal to noise ratio was varied between -10 and 10 dB. Classification results were averaged over 5000 trials for each SNR.

We compare the proposed z-space joint QDA classifier to x-space QDA, SVM with a linear kernel, and 1-NN classifiers using both the informed and blind approaches described in Sec. IV A. In each of the three simulations, the training and test data were linearly separable. Therefore, we set the regularization term C in the linear SVM (Ref. 32) to a large value, $C = 10^7$. Then, none of the classifiers requires cross-validation.
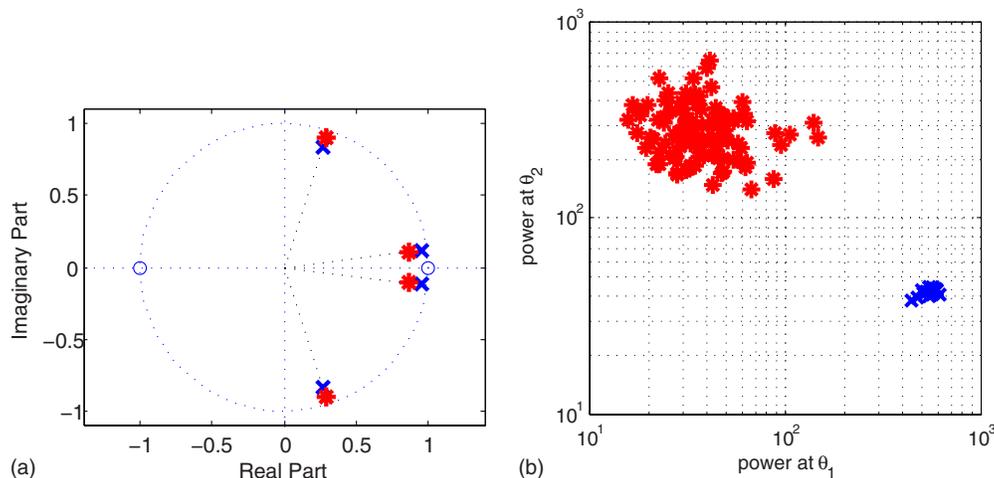
FIG. 4. (Color online) (a) Pole-zero plot showing the mean location of the poles for each class 1 ($\times$) and class 2 ($*$) for the "well-separated" case, and (b) scatter-plot of the classes in log-feature space.
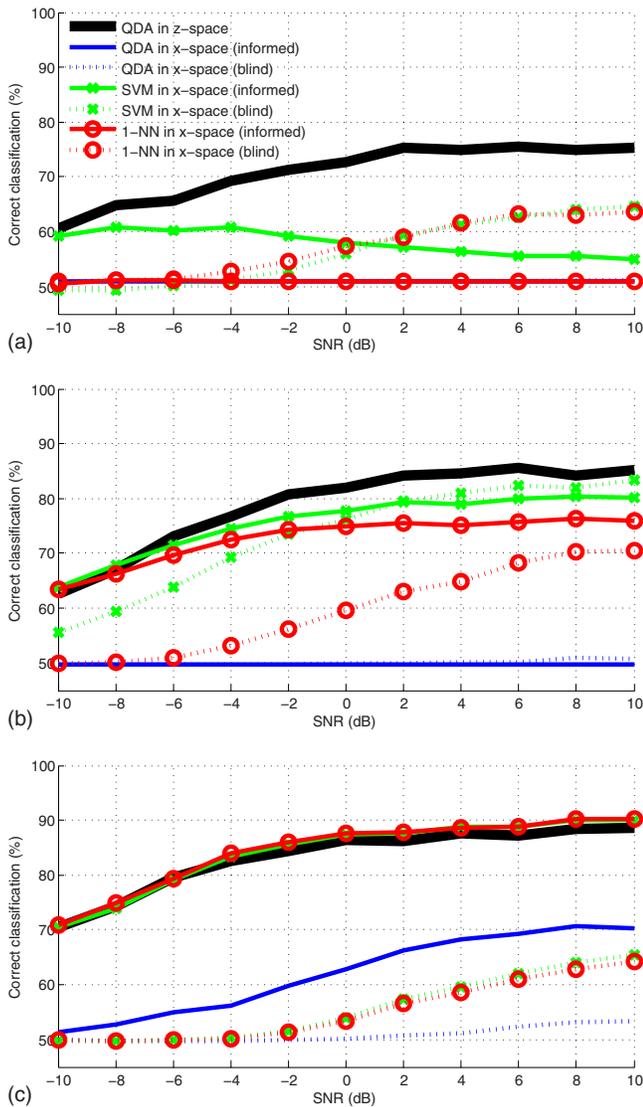
FIG. 5. (Color online) Results for feature-based classification on simulated data for the experiments where classes are (top) close in feature space, (middle) moderately separated in feature space, and (bottom) well-separated in feature space.

C. Simulation results

Results for each experiment are shown in Fig. 5. The joint QDA method using features (QDA in z-space) performs markedly better than the other approaches when the classes are difficult to separate. As class separation increases, the performance of the z-space QDA and informed SVM and 1-NN become similar.

D. Classifying bowhead whale songs in multipath channel

We employ the feature-based joint QDA method to identify individual Bowhead whales in a multipath environment by classifying the end notes of their songs. Several end notes of Bowhead whale vocalizations for two individuals were extracted from the MobySound archive.[33] Fifteen vocalizations are available for whale 1, and nine vocalizations are available for whale 2. According to the metadata, the end notes of Bowhead whale songs are relatively stable from

year to year. Therefore, we hope to be able to acoustically discriminate between two individuals based on previously recorded vocalizations. Our experimental setup simulates a shallow ocean channel (in comparison to the observation distances), and low SNR. Each of the signals has non-negligible interfering noise from bearded seals, sea ice and banging hydrophone cables.[33] The vocalizations were recorded in April 1988 near the coast of Point Barrow, Alaska, but for these experiments, we inject the signals into randomly drawn locations in the simulated bathymetry shown in Fig. 3(a). Example vocalizations for each whale are shown in Fig. 6.

Classifiers were trained on five training signals drawn at random for each class. The remaining 14 signals were propagated from a random source location in the bathymetry to the receiver using the *CASS Eigenray* routine. Gaussian white noise is added to the multipath signal to achieve a specified SNR. To increase the statistical significance of the results, classification results were averaged over 1000 iterations of the random training/test partitioning with random source locations.

For classification, four peak power features were selected: the two largest amplitude peaks averaged over signals in class 1 (163 and 258 Hz) and the two largest amplitude peaks for class 2 (588 and 207 Hz). Features do not correspond to strong interfering noise, and result in classes that, prior to the channel effects, are 100% linearly separable. As before, we compare the z-space joint QDA classifier to x-space QDA, linear SVM, and 1-NN classifiers using both informed and blind approaches.

E. Bowhead whale song results

Results for the Bowhead whale songs are shown in Fig. 7. The z-space joint QDA classifier using power features consistently achieves roughly 4% higher accuracy than other methods across the range of SNRs. This can be ascribed to the fact that the multipath channel distorts the relative covariance structures of class 1 and class 2 whale vocalizations.

V. CONCLUSIONS AND OPEN QUESTIONS

We have presented classification methods that jointly consider the effects of multipath distortion with classification. In particular, we have investigated a joint MAP deconvolution/classifier that incorporates first- and second-order statistics of the channel and yields a MAP solution for the recovered signal $\hat{x}$, the recovered filter $\hat{h}$, and the class estimate y^*. Two drawbacks of the joint MAP algorithm are that it is not convex, and that it theoretically performs poorly at high SNR. The first problem might be addressed by maximizing the marginal $p(h, y | z)$ which yields a convex expression, but requires more complicated optimization approaches. The second problem arises since regularization scales with the noise power σ_w^2, which may be replaced by a fixed penalty that can be chosen via cross-validation.

We hypothesized that better classification performance can be gained by marginalizing over x and h. To that end, we presented a joint QDA classifier that accounts for the LTI corruption probabilistically. Experiments showed that the joint QDA classifier outperformed the joint MAP classifier
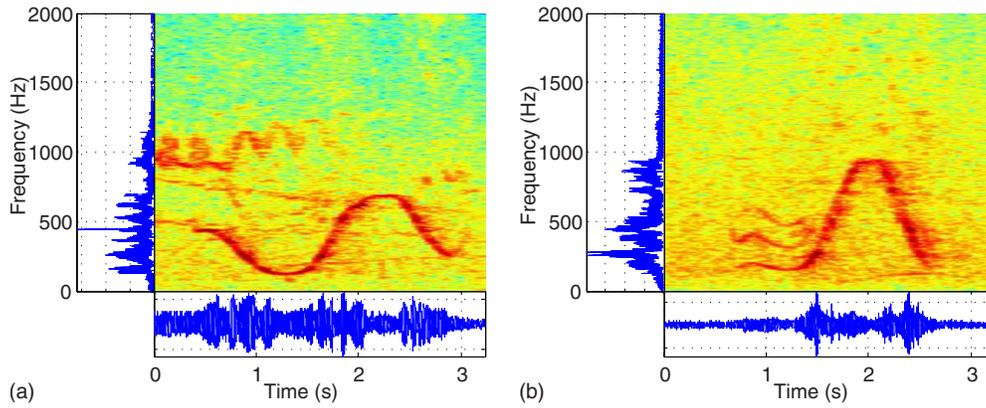
FIG. 6. (Color online) Spectrograms of whale song-endnotes for (a) the first Bowhead whale and (b) the second Bowhead whale. The vocalizations of the second whale tend to be more variable, cover a greater dynamic range, and contain stronger harmonic components than the first whale. Notice that the vocalization in (a) contains interfering calls from a bearded seal from about 800 to 1200 Hz.

and a classifier based on blind deconvolution. Further, we derived a joint QDA classifier that classifies based on subband power features. Experiments on both simulated signals and real Bowhead whale vocalizations demonstrated the efficacy of the joint QDA classifier using features. Although we have derived the joint QDA classifier for subband power features, there are many other important classes of features for signal processing including real cepstrum, wavelets, and wavepacket decomposition.[12] In cases where the features are linear functions of the data, such as the shift invariant wavepackets described in Ref. 12, the derivation of a joint QDA classifier is straightforward. However, features that are nonlinear functions of the received signal (e.g., real cepstrum) may require low-order approximate solutions.

We compared the joint QDA classifier on subband features to other classifiers that also took into account the mean multipath using subband power features. We showed that using this first-order multipath information significantly improved performance over ignoring the multipath altogether, except in the simulation experiment with well-separated classes and low noise. An open question in this line of research is how to take advantage of further information about the multipath for other classifiers, such as the nearest-neighbor and support vector machine classifiers.
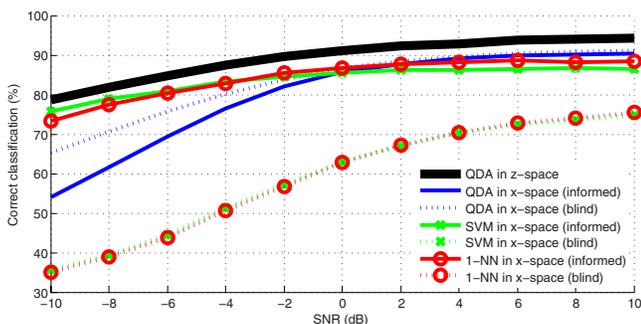
FIG. 7. (Color online) Classification results for identifying Bowhead whales by the end-notes of their songs. The SVM in x-space (blind) and 1-NN in x-space (blind) mostly overlap one another. Note that unlike previous figures, the ordinate axis shows results down to 30% correct classification to better show the trend over SNR.

APPENDIX: DERIVATIONS

In the Appendix, we use x, h, w, and z (and hence, all functions of them, for example P_x) to denote random signals, not their realizations as done in the main body.

Derivation of the first and second moments of $z(t)$

This section details the derivation of Eqs. (7) and (8). Since the noise is zero-mean, the nth component of the mean time signal $\mu_{z|y}$ is

$$\mu_{z|y}[n] = E[(x*h)[n] + w[n]|y] = E\left[\sum_k x[k]h[n-k]\Big|y\right]$$

$$= \sum_k E[x[k]|y]E[h[n-k]] = \sum_k \mu_{x|y}[k]\mu_h[n-k],$$

and thus $\mu_{z|y} = \mu_{x|y} * \mu_h$.

The covariance matrix is derived:

$$\Sigma_{z|y} = E[zz^T|y] - E[z|y]E[z|y]^T$$

$$\overset{(a)}{=} E[(x*h)(x*h)^T|y] + E[ww^T] - \mu_{z|y}\mu_{z|y}^T$$

$$\overset{(b)}{=} E[(xx^T) ** (hh^T)|y] + \sigma_w^2 I - \mu_{z|y}\mu_{z|y}^T$$

$$\overset{(c)}{=} E[xx^T|y] ** E[hh^T] + \sigma_w^2 I - \mu_{z|y}\mu_{z|y}^T,$$

where the expectations are taken with respect to the appropriate distributions. In the above, (b) follows from (a) because the (m,n)th component of the outer product $(x*h) \times (x*h)^T$ can be expressed as

$$\left(\sum_k x[k]h[n-k]\right)\left(\sum_i x[i]h[m-i]\right)$$

$$= \sum_k \sum_i x[k]h[n-k]x[i]h[m-i]$$

$$= \sum_k \sum_i (x[k]x[i])(h[n-k]h[m-i]),$$

and thus $(x*h)(x*h)^T = (xx^T) ** (hh^T)$. Then, (c) follows from (b) due to the linearity of both expectation and convolution. Finally, (c) can be rewritten as Eq. (8).

Derivation of the first and second moments of $P_z(f)$

This section details the derivation of Eqs. (10) and (11). Let $\mu_{P_z|y} = E[P_z|y]$, $\mu_{P_x|y} = E[P_x|y]$, $\mu_{P_h} = E[P_h]$ and $E[P_w] = E[WW^*] = \sigma_w^2 \mathbf{1}$ by assumption, where $\mathbf{1}$ is a vector of ones. The first moment given by Eq. (10) follows from Eq. (9) since $E[\text{Re}\{a\}] = E[\frac{1}{2}(a+a^*)] = \frac{1}{2}(E[a]+E[a]^*) = \text{Re}\{E[a]\}$, and $X(f)$, $H(f)$, and $W^*(f)$ are independent, so $E[\text{Re}\{X(f)H(f)W^*(f)\}] = 0$ because $E[W^*(f)] = 0$. The fact that $E[W^*(f)] = 0$ follows from $E[w(t)] = 0$ and the independence of the noise random process over time.

For the class-conditional covariance $\Sigma_{P_z|y}$, we derive the second moment and cross-correlation separately. For notational simplicity, we denote $E[x|y]$ by $E[x]$, that is, the class-conditional membership is implied in the expectation. For the second moment, it follows from Eq. (9) that

$$E[P_z(f)^2] \overset{(a)}{=} E[P_x^2 P_h^2 + P_w^2 + 4\,\text{Re}(XHW^*)^2$$
$$+ 4(P_x P_h + P_w)\text{Re}(XHW^*) + 2P_x P_h P_w]$$

$$\overset{(b)}{=} E[P_x^2]E[P_h^2] + E[P_w^2] + 4E[\text{Re}(XHW^*)^2]$$
$$+ 2E[P_x]E[P_h]E[P_w]$$

$$\overset{(c)}{=} E[P_x^2]E[P_h^2] + E[P_w^2] + 2E[XX^*HH^*WW^*$$
$$+ \text{Re}((XHW^*)^2)] + 2E[P_x]E[P_h]E[P_w]$$

$$\overset{(d)}{=} E[P_x^2]E[P_h^2] + E[P_w^2] + 4E[P_x]E[P_h]E[P_w]$$
$$+ 2\,\text{Re}(E[XX]E[HH]E[W^*W^*])$$

$$\overset{(e)}{=} E[P_x^2]E[P_h^2] + E[P_w^2] + 4E[P_x]E[P_h]E[P_w]. \quad \text{(A1)}$$

In the above, (b) follows from (a) because $4E[(P_x P_h + P_w)\text{Re}(XHW^*)]$ can be expanded into the sum of two terms containing $E[W^*]$ and $E[W^*WW^*]$, and since these terms equal zero, $4E[(P_x P_h + P_w)\text{Re}(XHW^*)] = 0$. For (c), we employed the identity $(\text{Re}\{c\})^2 = 1/2(cc^* + \text{Re}(c^2))$. Then (d) follows from (c) by the definition of power and the interchangeability of expectation and the real operator, explained earlier in this section. The step from (d) to (e) holds because the Fourier transform $W(f) = \mathcal{F}[w(t)]$ of a zero-mean Gaussian white noise process is a complex zero-mean Gaussian white process with $\text{Re}(W)$ and $\text{Im}(W)$ uncorrelated, $E[\text{Re}(W)] = E[\text{Im}(W)] = 0$, and $E[\text{Re}(W)^2] = E[\text{Im}(W)^2] = \frac{1}{2}\sigma_w^2$, so that $E[W^*W^*] = E[\text{Re}(W)^2 - 2j\,\text{Re}(W)\text{Im}(W) - \text{Im}(W)^2] = 0$.[34]

Last, we derive the cross-correlation using as short-hand P_{z_i} to denote $P_z(f_i)$, P_{h_i} to denote $P_h(f_i)$, H_i to denote $H(f_i)$, etc. From Eq. (9), for $i \neq j$,

$$E[P_{z_i} P_{z_j}] = E[P_{x_i} P_{h_i} P_{x_j} P_{h_j}] + E[P_{x_i} P_{h_i} P_{w_j}]$$
$$+ E[P_{x_j} P_{h_j} P_{w_i}] + E[P_{w_i} P_{w_j}] + 2E[\text{Re}(X_j H_j W_j)$$
$$\times (P_{x_i} P_{h_i} + P_{w_i})] + 2E[\text{Re}(X_i H_i W_i)(P_{x_j} P_{h_j}$$
$$+ P_{w_j})] + 4E[\text{Re}(X_i H_i W_i)\text{Re}(X_j H_j W_j)], \quad \text{(A2)}$$

Consider the fifth term of Eq. (A2),

$$2E[\text{Re}(X_j H_j W_j)(P_{x_i} P_{h_i} + P_{w_i})] = E[(X_j H_j W_j + X_j^* H_j^* W_j^*)$$
$$\times (X_i X_i^* H_i H_i^* + W_i W_i^*)].$$

After multiplication, the above consists of four terms, each containing a single W_j. Since W_j is uncorrelated with every other random variable and $E[W_j] = 0$, each of the four terms in the expansion has zero mean, and thus the entire expression is zero. By the same logic, the sixth term of of Eq. (A2) is

$$2E[\text{Re}(X_i H_i W_i)(P_{x_j} P_{h_j} + P_{w_j})] = 0.$$

The last term of Eq. (A2) can be rewritten

$$4E[\text{Re}(X_i H_i W_i)\text{Re}(X_j H_j W_j)] = E[(X_i H_i W_i + X_i^* H_i^* W_i^*)$$
$$\times (X_j H_j W_j + X_j^* H_j^* W_j^*)].$$

Taking the product results in four terms which each have a single W_i and W_j. Since W_i and W_j are uncorrelated and each has zero-mean, each of the four terms is zero. Thus, the cross-correlation of any two frequencies $f_i \neq f_j$ is

$$E[P_{z_i} P_{z_j}] = E[P_{x_i} P_{x_j}]E[P_{h_i} P_{h_j}] + E[P_{x_i}]E[P_{h_i}]E[P_{w_j}]$$
$$+ E[P_{x_j}]E[P_{h_j}]E[P_{w_i}] + E[P_{w_i}]E[P_{w_j}]. \quad \text{(A3)}$$

Conditioning on the class label y, Eqs. (A1) and (A3) can be combined into a single covariance matrix, where the (i,j)th element is

$$\Sigma_{P_z|y}[i,j] = (\Sigma_{P_x|y}[i,j] + \mu_{P_x|y}[i]\mu_{P_x|y}[j])(\Sigma_{P_h}[i,j]$$
$$+ \mu_{P_h}[i]\mu_{P_h}[j]) + (\Sigma_{P_w}[i,j] + \mu_{P_w}[i]\mu_{P_w}[j])$$
$$+ (\mu_{P_x|y}[i]\mu_{P_h}[i]\mu_{P_w}[j]$$
$$+ \mu_{P_x|y}[j]\mu_{P_h}[j]\mu_{P_w}[i])(1 + \delta_{ij})$$
$$- \mu_{P_z|y}[i]\mu_{P_z|y}[j]$$
$$= \Sigma_{P_x|y}[i,j]\Sigma_{P_h}[i,j] + \Sigma_{P_w}[i,j]$$
$$+ \mu_{P_x|y}[i]\mu_{P_x|y}[j]\Sigma_{P_h}[i,j]$$
$$+ \Sigma_{P_x|y}[i,j]\mu_{P_h}[i]\mu_{P_h}[j]$$
$$+ 2\mu_{P_x|y}[i]\mu_{P_h}[i]\mu_{P_w}[i]\delta_{ij}$$
$$= \Sigma_{P_x|y}[i,j]\Sigma_{P_h}[i,j] + \sigma_w^4 \delta_{ij}$$
$$+ \mu_{P_x|y}[i]\mu_{P_x|y}[j]\Sigma_{P_h}[i,j]$$
$$+ \Sigma_{P_x|y}[i,j]\mu_{P_h}[i]\mu_{P_h}[j]$$
$$+ 2\sigma_w^2 \mu_{P_x|y}[i]\mu_{P_h}[i]\delta_{ij},$$

where $\delta_{ij} = 1$ if $i = j$ or 0 otherwise. The last step in the derivation holds since $E[P_w(f_i)P_w(f_j)] = E[W_i W_i^* W_j W_j^*]$ is $2(\sigma_w^2)^2$ for $i = j$ or is $(\sigma_w^2)^2$ for $i \neq j$ by Isserlis' Gaussian moment theorem,[35] and $\mu_{P_w}[i] = \sigma_w^2$. The result is rewritten in compact form in Eq. (11).

[1] R. L. Field, "Transient signal distortion in a multipath environment," Proc. OCEANS 1990, pp. 111–114.
[2] D. W. Tufts, I. Kirsteins, R. J. Vaccaro, C. S. Ramalingam, and A. Shah, "Improvements in signal processing for time-varying multipath propaga-

tion environments," Proc. OCEANS, 1991, Vol. **1**, pp. 586–593.

[3]S. Senmoto and D. Childers, "Signal resolution via digital inverse filtering," IEEE Trans. Aerosp. Electron. Syst. **8**, 644–640 (1972).

[4]J. E. Ehrenberg, T. E. Ewart, and R. D. Morris, "Signal-processing techniques for resolving individual pulses in a multipath signal," J. Acoust. Soc. Am. **63**, 1861–1865 (1978).

[5]T. E. Ewart, J. E. Ehrenberg, and S. A. Reynolds, "Observations of the phase and amplitude of individual Fermat paths in a multipath environment," J. Acoust. Soc. Am. **63**, 1801–1808 (1978).

[6]F. B. Shin, D. H. Kil, and R. Wayland, "IER clutter reduction in shallow water," Proc. IEEE ICASSP, 1996, pp. 3149–3152.

[7]D. J. Strausberger, E. D. Garber, N. F. Chamberlain, and E. K. Walton, "Modeling and performance of HF/OTH radar target classification systems," IEEE Trans. Aerosp. Electron. Syst. **28**, 396–403 (1992).

[8]G. Okopal and P. Loughlin, "Feature extraction for classification of signals propagating in channels with dispersion and dissipation," Proc. SPIE **6566**, April 9–13, 2007.

[9]H. Liu, P. Runkle, and L. Carin, "Classification of distant targets situated near channel bottoms," J. Acoust. Soc. Am. **115**, 1185–1197 (2004).

[10]D. A. Caughey and R. L. Kirlin, "Blind deconvolution of echosounder envelopes," Proc. IEEE ICASSP, 1996, pp. 3149–3152.

[11]M. J. Roan, M. R. Gramann, J. G. Erling, and L. H. Sibuld, "Blind deconvolution applied to acoustical systems identification with supporting experimental results," J. Acoust. Soc. Am. **114**, 1988–1996 (2003).

[12]D. H. Kil and F. B. Shin, *Pattern Recognition and Prediction with Applications to Signal Characterization* (American Institute of Physics Press, Woodbury, NY, 1996).

[13]M. K. Broadhead and L. A. Pflug, "Performance of some sparseness criterion blind deconvolution methods in the presence of noise," J. Acoust. Soc. Am. **107**, 885–893 (2000).

[14]Q. Zhu and B. Steinberg, "Correction of multipath interference using clean and spatial location diversity," Proc. IEEE Ultrasonics Symp., 1995, pp. 1367–1370.

[15]R. A. Wiggins, "Minimum entropy deconvolution," Geoexploration **16**, 21–35 (1978).

[16]C. A. Cabrelli, "Minimum entropy deconvolution and simplicity: A noniterative algorithm," Geophysics **50**, 394–413 (1984).

[17]M. R. Gupta, H. S. Anderson, and Y. Chen, "Joint deconvolution and classification for signals with multipath," Proc. IEEE ICASSP, 2007.

[18]M. R. Gupta and H. Anderson, "Maximum likelihood signal classification using second-order blind deconvolution probability model," Proc. IEEE Wkshp. Stat. Sig. Proc., 2007.

[19]N. Dasgupta and L. Carin, "Time-reversal imaging for classification of submerged elastic targets via Gibbs sampling and the relevance vector machine," J. Acoust. Soc. Am. **117**, 1999–2011 (2005).

[20]A. K. Jain, *Fundamentals of Digital Image Processing* (Prentice Hall, Englewood Cliffs, NJ, 1989).

[21]A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems* (Springer, New York, 1996).

[22]S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2004).

[23]R. W. Yeung, *A First Course in Information Theory* (Springer, New York, 2002).

[24]E. Y. Lam and J. W. Goodman, "Iterative statistical approach to blind image deconvolution," J. Opt. Soc. Am. A **17**, 1177–1184 (2000).

[25]T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning* (Springer, New York, 2001).

[26]J. H. Friedman, "Regularized discriminant analysis," J. Am. Stat. Assoc. **84**, 165–175 (1989).

[27]S. Srivastava, M. R. Gupta, and B. A. Frigyik, "Bayesian quadratic discriminant analysis," J. Mach. Learn. Res. **8**, 1287–1314 (2007).

[28]E. Candes, "Compressive sampling," Proc. Int. Congress of Math., 2006, Vol. **3**, pp. 1433–1452.

[29]R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. (Wiley, New York, 2001).

[30]R. P. Goddard, "The Sonar Simulation Toolset," Proc. OCEANS, 1989, Vol. **4**, pp. 1217–1222.

[31]R. P. Goddard, "The Sonar Simulation Toolset, Release 4.1: Science, mathematics and algorithms," Technical Report APL-UW TR 0404, Applied Phyics Laboratory, University of Washington, 2005; URL http://stinet.dtic.mil/cgi-bin/GetTRDoc?AD=ADA430617&Location=U2&doc=GetTRDoc.pdf (Last viewed August 2008).

[32]R. E. Fan, P. H. Chen, and C. J. Lin, "Working set selection using the second order information for training SVM," J. Mach. Learn. Res. **6**, 1889–1918 (2005).

[33]D. K. Mellinger and C. W. Clark, "Mobysound: A reference archive for studying automatic recognition of marine mammal sounds," Appl. Acoust. **67**, 1226–1242 (2006); http://hmsc.oregonstate.edu./projects/MobySound/ (Last viewed: August 2008).

[34]C. Therrien, *Discrete Random Signals and Statistical Signal Processing* (Prentice Hall, Upper Saddle River, NJ, 1992).

[35]L. Isserlis, "On a formula for the product-mean coefficient of any order of a normal frequency distribution in any number of variables," Biometrika **12**, 134–139 (1918).