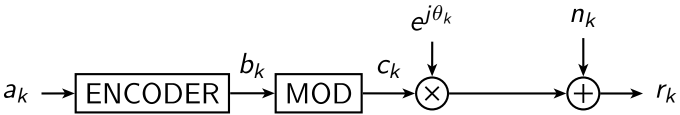
Message Passing Algorithms for Phase Noise Tracking Using Tikhonov Mixtures

Shachar Shayovitz

Tel Aviv University

July 31, 2013

Phase Noise Equivalent Baseband Channel



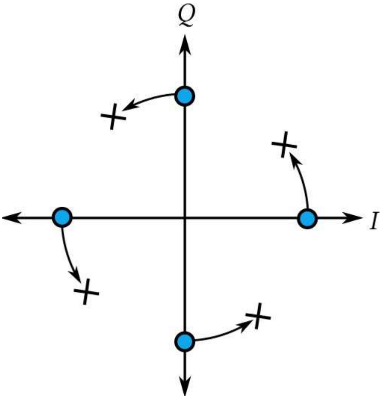
Mathematical Notation

$$r_k = c_k e^{j\theta_k} + n_k \,, \quad n_k \sim \mathcal{CN}(0, \sigma^2)$$

$$\theta_k = \theta_{k-1} + \Delta_k \,, \quad \Delta_k \sim \mathcal{N}(0, \sigma_\Delta^2)$$

Example - Effect on Demodulating QPSK

$$r_k = c_k e^{j\theta_k} + n_k$$



Previous Mitigation Approaches

Method	Pros	Cons
DA/NDA PLL	Low Complexity	Low Phase noise
Noncoherent	No pilots	Intrinsic Loss
Model based	Performs very well	Model dependant
Special codes	Good Performance	Not standard design
LDPC with differential encoding	No pilots	Long convergence time, not better than model based

Thesis Goal

Using Strong FEC

- LDPC and Turbo codes work well in **low SNR regions**
- Standard codes and no need for special design
- MAP detection of code symbols is the optimal scheme in BER sense
- Perform **joint detection and estimation**
- The phase tracker and code decoder will exchange information iteratively

Objective

- Design a **low complexity** algorithm for providing **LLRs to the LDPC/Turbo decoder**

Joint Detection & Estimation

MAP Detection

$$\hat{c}_k = arg \max_{c_k} \Pr(c_k|\mathbf{r})$$

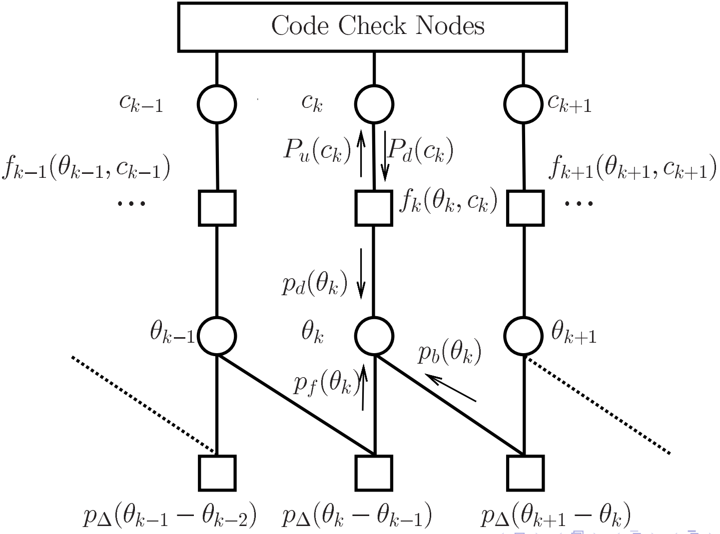
$$\hat{c}_k = arg \max_{c_k} \sum_{\mathbf{c}/c_k, \boldsymbol{\theta}} p(\mathbf{c}, \boldsymbol{\theta}|\mathbf{r})$$

Factorization

$$p(\mathbf{c}, \boldsymbol{\theta}|\mathbf{r}) \propto p(\theta_0) \prod_{k=1}^{K-1} \underbrace{p(\theta_k|\theta_{k-1})}_{p_\Delta(\theta_k - \theta_{k-1})} \prod_{k=0}^{K-1} \underbrace{p(r_k|\theta_k, c_k)}_{f_k(c_k, \theta_k)} \mathbb{1}\{c_0^{K-1} \in \mathcal{C}\}$$

Factor Graph

Barbieri, Colavolpe and Caire (2006)



Sum and Product Algorithm

Forward & Backward Messages

- $p_d(\theta_k) = \sum_{m=0}^{M-1} P_d(c_k = e^{j\frac{2\pi m}{M}}) f_k(c_k, \theta_k)$
- $p_f(\theta_k) = \int_0^{2\pi} p_f(\theta_{k-1}) p_d(\theta_{k-1}) p_\Delta(\theta_k - \theta_{k-1}) d\theta_{k-1}$
- $p_b(\theta_k) = \int_0^{2\pi} p_b(\theta_{k+1}) p_d(\theta_{k+1}) p_\Delta(\theta_{k+1} - \theta_k) d\theta_{k+1}$

LLR

$P_u(c_k) = \int_0^{2\pi} p_f(\theta_k) p_b(\theta_k) f_k(c_k, \theta_k) d\theta_k$

Problem

- Implementation problem - Phase messages are continuous!
- One solution - Quantize the phase and perform approximated SPA

Sum and Product Algorithm

High accuracy requires high complexity

Model Based Approximations

Canonical Model

- SPA messages are approximated using a family of distributions (finite parameters)
- Much lower computational complexity than quantization

Single Tikhonov

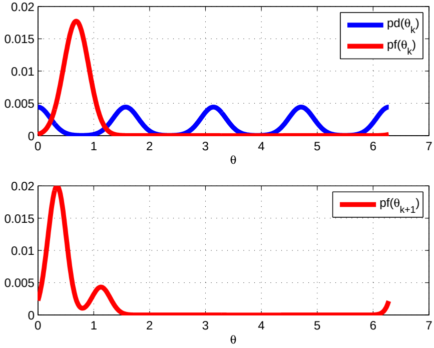
Barbieri, Colavolpe and Caire (2006) used a Single Tikhonov distribution **for all** SPA phase messages

$$p_f(\theta_k) = \frac{e^{Re[z^{k,f} e^{-j\theta_k}]}}{2\pi I_0(|z^{k,f}|)}$$

Single Tikhonov Canonical Model - Problem

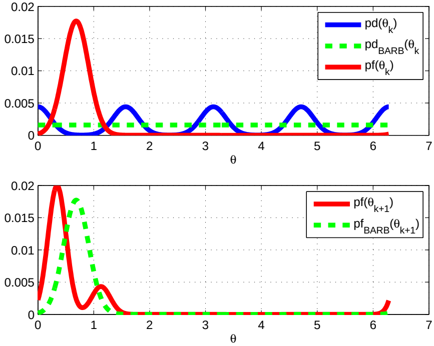
Canonical model is not consistent!

$$\underbrace{p_f(\theta_{k+1})}_{\text{Tikhonov Mixture}} = \int_0^{2\pi} \underbrace{p_f(\theta_k)}_{\text{Tikhonov distribution}} \underbrace{p_d(\theta_k)}_{\text{Tikhonov Mixture}} \underbrace{p_\Delta(\theta_{k+1} - \theta_k)}_{\text{Gaussian}} d\theta_k$$



Single Tikhonov Canonical Model - Problem

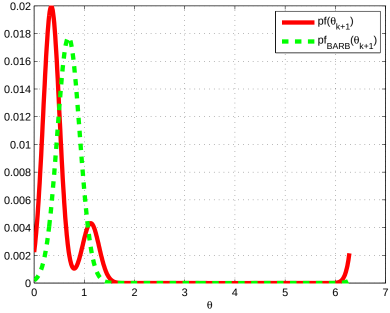
Barbieri, Colavolpe and Caire (2006) proposed to find the closest Gaussian to $p_d(\theta_k)$



Not Good Enough!

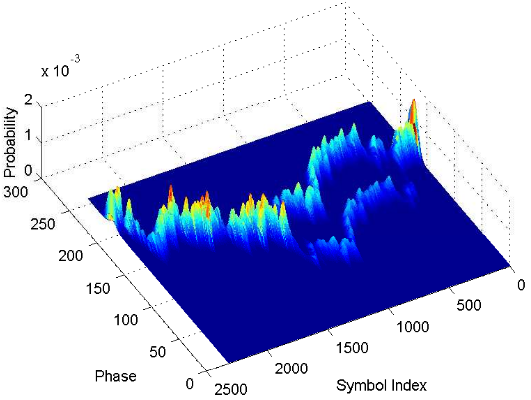
Problem

- In absence of reliable prior information on the code, $p_d(\theta_k)$ is multi modal
- **Single Tikhonov canonical models are not suitable for the first iteration in strong phase noise.**



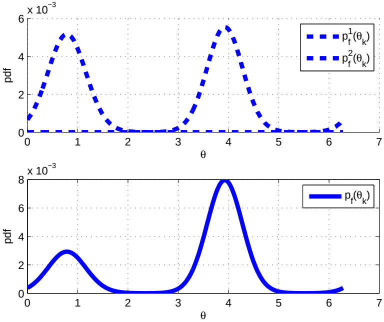
Motivation for Mixtures - Phase Trajectories

Looking at DP, we can see the multi modal dynamics of the phase posterior



Canonical Model - Tikhonov Mixture

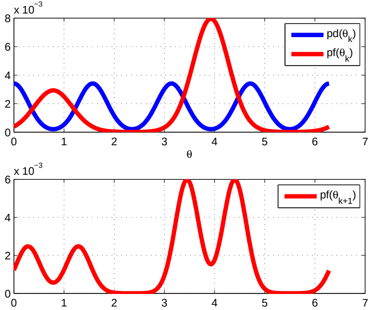
$$p_f(\theta_k) = \sum_{i=1}^{N_f} \alpha_i^f \frac{e^{Re[z_i^{k,f} e^{-j\theta_k}]}}{2\pi I_0(|z_i^{k,f}|)}$$



Problem

Mixture order grows exponentially!

$$\underbrace{p_f(\theta_{k+1})}_{\text{Bigger Tikhonov Mixture}} = \int_0^{2\pi} \underbrace{p_f(\theta_k)}_{\text{Tikhonov Mixture}} \underbrace{p_d(\theta_k)}_{\text{Tikhonov Mixture}} \underbrace{p_\Delta(\theta_{k+1} - \theta_k)}_{\text{Gaussian}} d\theta_k$$



Problem Formulation

Classical Mixture Reduction

Given a Tikhonov mixture,

$$f(\theta) = \sum_{i=1}^{N} \alpha_i f_i(\theta)$$

Find a Tikhonov mixture with $M < N$

$$g(\theta) = \sum_{j=1}^{M} \beta_j g_j(\theta)$$

Which minimizes some distortion criterion,

$$D(f(\theta)||g(\theta))$$

Mixture Reduction Algorithms

Mixture Similarity Metric

- Kullback Leibler divergence (KLD) is more natural for this setting than Integral square error (ISE)
- Known mixture reduction algorithms such as: Salmond (1990),Williams & Maybeck (2003) and Runnalls (2006) don't work well

Why do these algorithms fail?

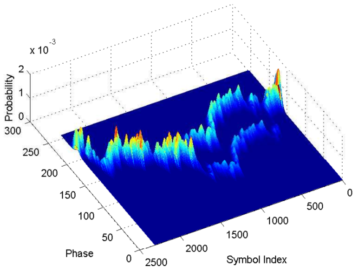
Fixed mixture order and **clustering errors** limit the performance,

- Small order will not be accurate enough and undergo cycle slips and create error floor
- Large order is too computationally demanding

New Approach - Adaptive Mixture Order

Introduction

- Typically, the number of phase trajectories is small → small mixture
- It is important to be very accurate in the mixture reduction in order not to propagate errors → large mixture
- Mixture reduction is performed **for each** symbol → **adaptive mixture order**



New Approach - Adaptive Mixture Order

Objective

Given a Tikhonov mixture,

$$f(\theta) = \sum_{i=1}^{N} \alpha_i f_i(\theta)$$

Find the Tikhonov mixture $g(\theta)$ with a small number of components $M < N$

$$g(\theta) = \sum_{j=1}^{M} \beta_j g_j(\theta)$$

which satisfy,

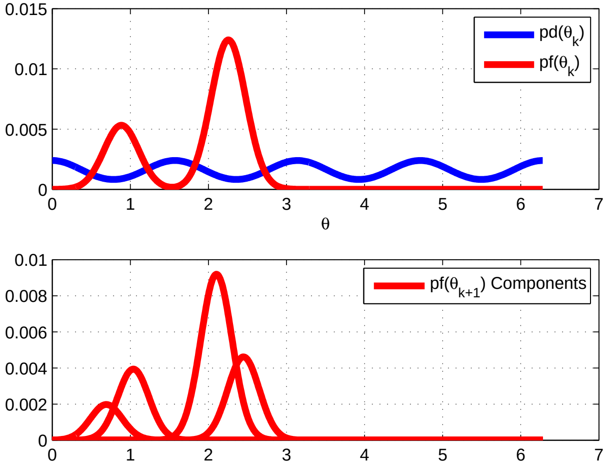
$$D_{KL}(f(\theta)||g(\theta)) \leq \epsilon$$

Unlimited Order Mixture Reduction Algorithm

$j \leftarrow 1$
while $|f(\theta)| > 0$ **do**
  $lead \leftarrow argmax\{\underline{\alpha}\}$
  for $i = 1 \rightarrow |f(\theta)|$ **do**
    if $D_{KL}(f_i(\theta)||f_{lead}(\theta)) \leq \epsilon$ **then**
      $idx \leftarrow [idx, i]$
    end if
  end for
  $\beta_j \leftarrow \sum_{i \in idx} \alpha_i$
  $g_j(\theta) \leftarrow CMVM(\sum_{i \in idx} \frac{\alpha_i}{\beta_j} f_i(\theta))$
  $f(\theta) \leftarrow f(\theta) - \sum_{i \in idx} \alpha_i f_i(\theta)$
  $j \leftarrow j + 1$
end while

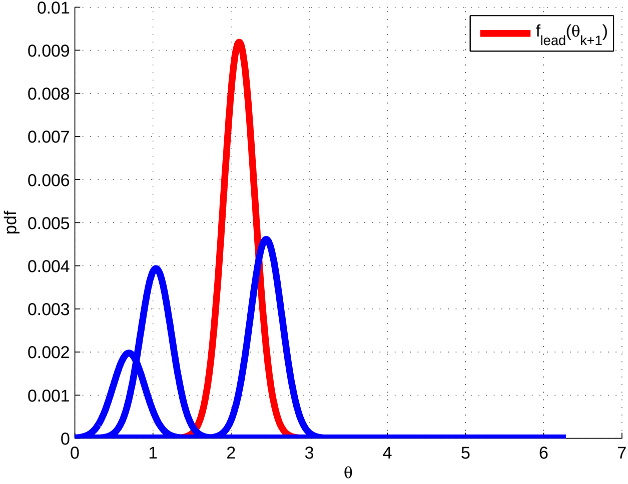
Mixture Reduction Algorithm

Suppose we need to reduce the dimensions of the following message $p_f(\theta_{k+1})$



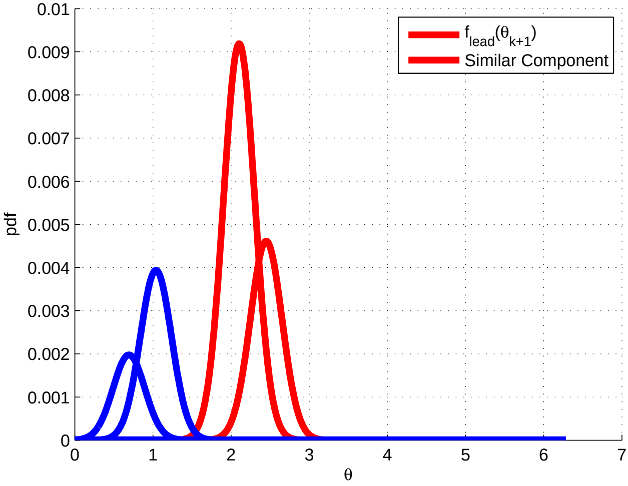
Mixture Reduction Algorithm

Choose the most probable mixture component and name it $f_{lead}(\theta_{k+1})$,



Mixture Reduction Algorithm

Find all other mixture components $f_i(\theta_{k+1})$ for which $D_{KL}(f_i(\theta_{k+1})||f_{lead}(\theta_{k+1})) \leq \epsilon$,



CMVM - Circular Mean and Variance Matching

Theorem (Shayovitz & Raphaeli 2012)

Given a circular distribution $f(\theta)$, the parameters of the Tikhonov distribution $g(\theta)$ which satisfy,

$$[\hat{\mu}, \hat{\sigma}^2] = \underset{\mu,\sigma^2}{\arg\min}\, D_{KL}(f(\theta)||g(\theta))$$

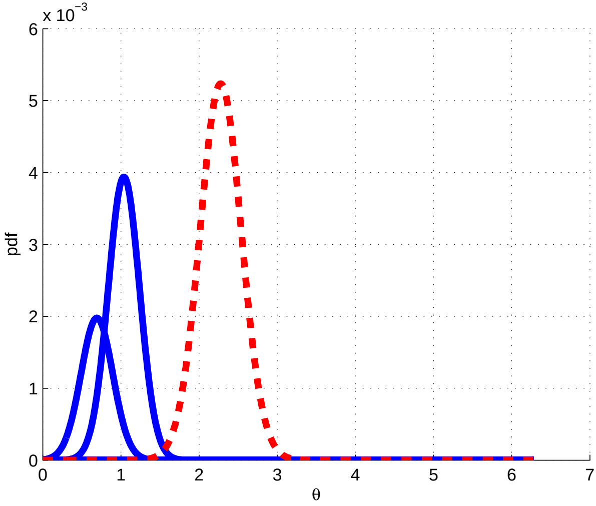
Are given by:

$$\hat{\mu} = \angle \mathbb{E}_f(e^{j\theta})$$

$$\hat{\sigma}^2 = \mathbb{E}_f\left(1 - cos\left(\theta - \hat{\mu}\right)\right)$$

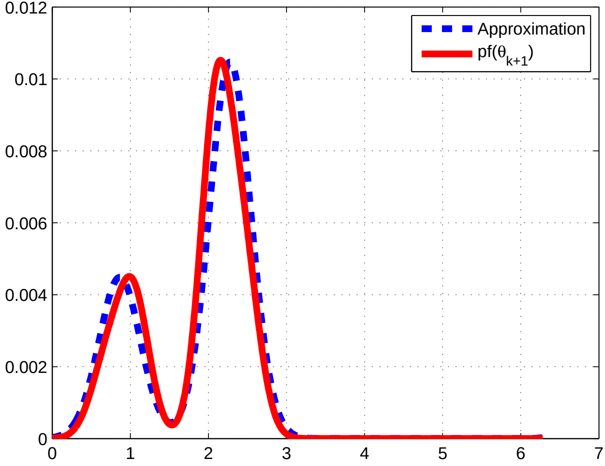
Mixture Reduction Algorithm

Cluster all the chosen mixture components using CMVM and get the first reduced mixture component $g_1(\theta_{k+1})$.



Mixture Reduction Algorithm

Eliminate the clustered components and iterate until there are no original mixture components left...



Accuracy Theorem

Theorem (Shayovitz & Raphaeli 2012)

(Mixture Reduction Accuracy): *Let $f(\theta)$ be a Tikhonov mixture of order L and ϵ be a real positive number.*
Then, applying the mixture reduction algorithm to $f(\theta)$ using ϵ, produces a Tikhonov mixture $g(\theta)$, of order $N < L$ which satisfies,

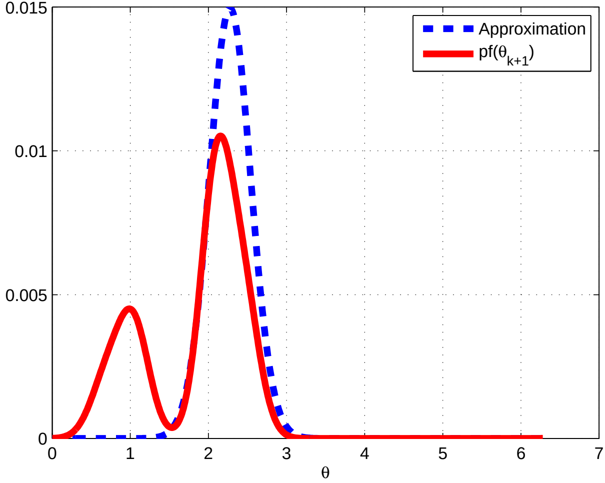
$$D_{KL}(f(\theta)||g(\theta)) \leq \epsilon$$

Implications

- Mixture reduction accuracy is mathematically upper bounded
- Allows to track all significant trajectories and produce accurate LLR
- Shown via simulations to have low

Limited Complexity

What happens if the mixture Order is limited to 1?
We only choose one trajectory!

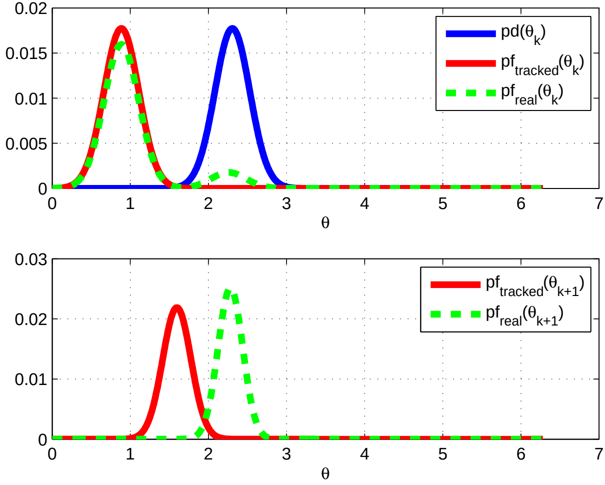


Limited Complexity

Problem

- Can track only a **limited** number of phase trajectories
- If the number of significant phase trajectories is larger than the maximum number of mixture components allowed, then we might **miss the correct trajectory**
- We can never regain tracking, not even using pilots!
- This event is analogous to phase slip in PLL

Limited Complexity



Limited Complexity

Limited Mixture Order

When using the previous algorithm, tracking a limited number of trajectories ⇒ Some components will be ignored
Their cumulative probability is the probability of a phase slip

Online Phase Slip Estimation

We add an additional variable ϕ_k^f (for backward recursions ϕ_k^b) , which online approximates the probability that the tracked trajectories include the correct one.

$$\phi_0^f = 1$$

$$\phi_k^f \leftarrow (\sum_j \beta_j)\phi_{k-1}^f$$

Recovering From Cycle Slips

Problem

In case of a cycle slip, the phase message estimator based on the tracked trajectories is **useless**

Using Pilots

- **Assuming pilots are present**
- One may estimate the message using **only** the pilot symbol, $p_d(\theta_k)$.
- But if a cycle slip has not occurred, then estimating the phase message based **only** on the pilot symbol might damage our tracking.

Recovering From Cycle Slips

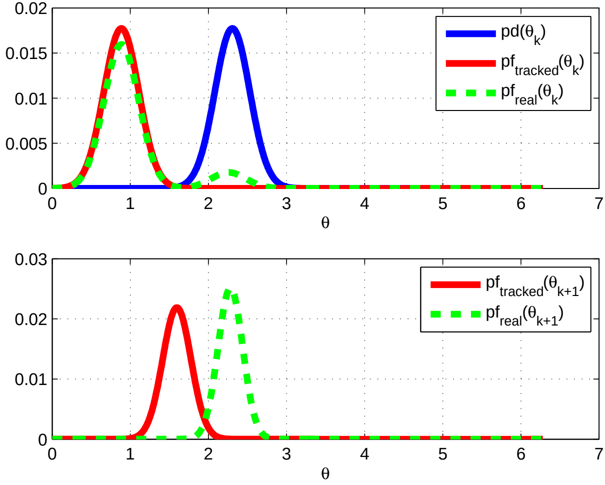
Possible Solution

- Once a pilot symbol arrives, we will **average** the following estimators according to ϕ_k^f,

$$q_f(\theta_k) = \phi_k^f \underbrace{p_f(\theta_k)}_{\text{Tracked Trajectories}} + (1 - \phi_k^f)\frac{1}{2\pi}$$

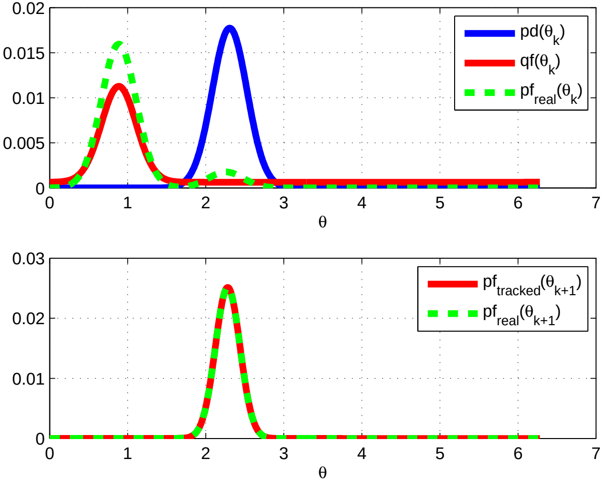
- If a cycle slip has occurred and ϕ_k^f is low, then the pilot will, in high probability, correct the tracking.

Reminder


0.02
0.015
0.01
0.005
0
0 1 2 3 4 5 6 7
θ
pd(θ_k)
pf_tracked(θ_k)
pf_real(θ_k)
0.03
0.02
0.01
0
0 1 2 3 4 5 6 7
θ
pf_tracked(θ_k+1)
pf_real(θ_k+1)

Recovering From Cycle Slips

For $\phi_k^f = 0.6$



Computation of $P_u(c_k)$

$P_u(c_k)$

- Final step in the approximated SPA
- LLR of a code symbol based on the channel part of the factor graph.
- Sent to the LDPC decoder and are crucial for the decoding of the LDPC.

Modified Computation

- We use $q_f(\theta_k) = \phi_k^f p_f(\theta_k) + (1 - \phi_k^f)\frac{1}{2\pi}$
- Forward-Backward scheme coupled with the mixtures based on cycle slip averaging, helps remove wrong trajectories

Complexity Reductions

We use several complexity reduction procedures

- Low complexity approximation of the KLD of two Tikhonov distributions
- All probabilities are in log domain (reduce muls)
- We use the log-sum approximation using maximum operation with LUT correction
- For small ϵ, we can use the leading component instead of using CMVM (tradeoff with mixture order). This saves a lot of computation time.

Complexity

Computational load per code symbol per iteration for MPSK constellations

	DP	BARB	Limited Order
MULS	$4Q^2M^2+2M^2Q+6MQ+M$	$7M+5$	$4M\gamma(i)^2+2M(\gamma(i)+1)$
LUT	QM	$3M$	$3M\gamma(i)^2-\gamma(i)(2M-1)$

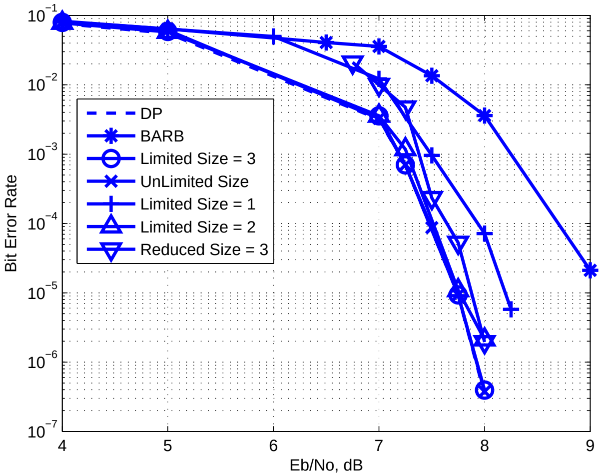
$\gamma(i)$ is the mean mixture order for iteration i, M is the constellation order, L is the number of quantization levels and Q is a parameter for the DP algorithm

Numerical Results

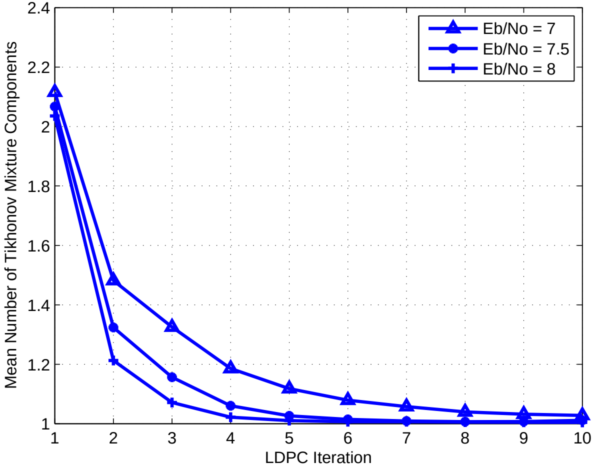
Monte Carlo simulation results for the proposed algorithms with varying mixture order and level of complexity, algorithm proposed by Barbieri et al (2006) and algorithm based on phase quantization (DP).

- Length 4608 LDPC code with rate 0.889.
- MPSK constellation.
- Phase noise model with varying σ_Δ [rads/symbol].
- A single pilot was inserted every $\frac{1}{pilotfrequency}$ symbols.
- The DP algorithm was simulated using 16 quantization levels.

8PSK - $\sigma_\Delta = 0.05$, Pilot Frequency $= 0.05$



8PSK Mean Number of Tikhonov Mixture Components - Full Algorihtm, Maximum 3 lobes, $\epsilon = 4$

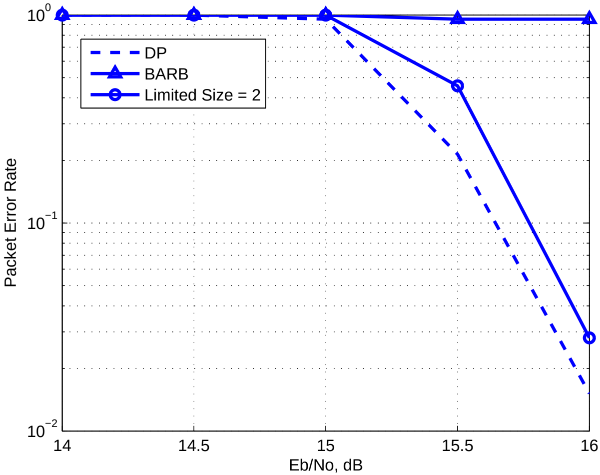


Complexity

Computational load per code symbol per iteration for 8PSK constellation, $\frac{E_b}{N_0} = 8dB$

Algorithm	DP	BARB	Reduced Complexity, Order 3
Iteration	Constant for all iterations	Constant	1 2 3 4
MULS	68360	61	312 292 273 238
LUT	128	24	147 134 123 102

32PSK - $\sigma_\Delta = 0.01$, Pilot Frequency $= 0.025$



Contributions

- A low complexity joint decoding and estimation algorithm for strong phase noise channels with excellent performance in,
  - High code rates
  - Low pilot frequency
  - High order constellations
  - Strong phase noise
- A new approach for mixture dimension reduction (KLD upper bounded).
- A novel approach for combating cycle slips.
- A new theorem in directional statistics for clustering circular mixtures.
- Introduction the field of directional statistics to iterative phase tracking.

Backup

Helpful Results for KL Divergence

We introduce the reader to three results related to the Kullback-Leibler Divergence which will prove helpful in the next sections.

Lemma

Suppose we have two distributions, $f(\theta)$ and $g(\theta)$,

$$f(\theta) = \sum_{i=1}^{M} \alpha_i f_i(\theta)$$

$$D_{KL}\left(\sum_{i=1}^{M} \alpha_i f_i(\theta) || g(\theta)\right) \leq \sum_{i=1}^{M} \alpha_i D_{KL}(f_i(\theta) || g(\theta)) \quad (1)$$

The proof of this bound is based on the Jensen inequality.

Helpful Results for KL Divergence

Lemma

Suppose we have three distributions, $f(\theta)$,$g(\theta)$ and $h(\theta)$. We define the following mixtures,

$$f_1(\theta) = \alpha f(\theta) + (1-\alpha)g(\theta) \tag{2}$$

$$f_2(\theta) = \alpha f(\theta) + (1-\alpha)h(\theta)) \tag{3}$$

for $0 \leq \alpha \leq 1$
Then,

$$D_{KL}(f_1(\theta)||f_2(\theta)) \leq (1-\alpha)D_{KL}(g(\theta)||h(\theta)) \tag{4}$$

Helpful Results for KL Divergence

Lemma

Suppose we have two mixtures, $f(\theta)$ and $g(\theta)$, of the same order M,

$$f(\theta) = \sum_{i=1}^{M} \alpha_i f_i(\theta)$$

and

$$g(\theta) = \sum_{j=1}^{M} \beta_i g_i(\theta)$$

Then the KL divergence between them can be upper bounded by,

$$D_{KL}(f(\theta)||g(\theta)) \leq D_{KL}(\alpha||\beta) + \sum_{i=1}^{M} \alpha_i D_{KL}(f_i(\theta)||g_i(\theta)) \tag{5}$$

Bessel Functions Approximation

Since implementing a modified bessel function is computationally prohibitive, we present the following approximation,

$$\log(I_0(k)) \approx k - \frac{1}{2}\log(k) - \frac{1}{2}\log(2\pi) \tag{6}$$

which holds for $k > 2$, i.e. reasonably narrow distributions. Using the following relation,

$$I_1(x) = \frac{dI_0(x)}{dx} \tag{7}$$

We find that,

$$\frac{I_1(k)}{I_0(k)} = \frac{d}{dk}(\log(I_0(k))) \tag{8}$$

Therefore

$$\frac{I_1(k)}{I_0(k)} \approx 1 - \frac{1}{2k} \tag{9}$$

Computation of $P_u(c_k)$

$$P_u(c_k) \propto \int_0^{2\pi} q_f(\theta_k) q_b(\theta_k) e_k(c_k, \theta_k) d\theta_k$$

We decompose the computation to a summation of four components,

$$P_u(c_k) \propto A + B + C + D$$

and get,

$$A = \sum_{i=1}^{N_f^k} \sum_{j=1}^{N_b^k} \alpha_i^{k,f} \alpha_j^{k,b} \frac{I_0(|z_i^{k,f} + z_j^{k,b} + \frac{r_k c_k^*}{\sigma^2}|)}{2\pi I_0(|z_i^{k,f}|) I_0(|z_j^{k,b}|)}$$

Computation of $P_u(c_k)$

When implementing the algorithm in log domain and for large enough $|z_i^{k,f}|$ and $|z_j^{k,b}|$

$$\log\left(\frac{I_0(|Z_\psi|)}{2\pi I_0(|z_i^{k,f}|)I_0(|z_j^{k,b}|)}\right) \approx |Z_\psi| - |z_i^{k,f}| - |z_j^{k,b}|$$

Approximation of KLD

$$g_1(\theta) = \frac{e^{Re[z_1 e^{-j\theta}]}}{2\pi I_0(|z_1|)} \tag{10}$$

$$g_2(\theta) = \frac{e^{Re[z_2 e^{-j\theta}]}}{2\pi I_0(|z_2|)} \tag{11}$$

We wish to compute the following KL divergence,

$$D_{KL}(g_1(\theta)||g_2(\theta)) \tag{12}$$

which is,

$$D_{KL} = \int_0^{2\pi} g_1(\theta)\log(\frac{e^{Re[z_1 e^{-j\theta}]} I_0(|z_2|)}{e^{Re[z_2 e^{-j\theta}]} I_0(|z_1|)})d\theta \tag{13}$$

Approximation of KLD

Thus,

$$D_{KL} = \log(\frac{I_0(|z_2|)}{I_0(|z_1|)}) + \int_0^{2\pi} g_1(\theta)(Re[z_1 e^{-j\theta}] - Re[z_2 e^{-j\theta}])d\theta \quad (14)$$

After some algebraic manipulations, we get

$$D_{KL} = \log(\frac{I_0(|z_2|)}{I_0(|z_1|)}) + \frac{I_1(|z_1|)}{I_0(|z_1|)}(|z_1| - |z_2|cos(\angle z_1 - \angle z_2)) \quad (15)$$

Using (9) and (6) we get

$$D_{KL} \approx |z_2|(1 - cos(\angle z_1 - \angle z_2)) - \frac{1}{2}\log(\frac{|z_2|}{|z_1|}) + \frac{|z_2|}{2|z_1|}cos(\angle z_1 - \angle z_2) \quad (16)$$

CMVM Proof

Let $f(\theta)$ be any circular distribution defined on $[0, 2\pi)$ and $g(\theta)$ a Tikhonov distribution.

$$g(\theta) = \frac{e^{Re[\kappa e^{-j(\theta-\mu)}]}}{2\pi I_0(\kappa)} \tag{17}$$

We wish to find,

$$[\mu^*, \kappa^*] = arg \min_{\mu,\kappa} D_{KL}(f||g) \tag{18}$$

According to the definition of the KL divergence,

$$D_{KL}(f||g) = -h(f) - \int_0^{2\pi} f(\theta) \log g(\theta) d\theta \tag{19}$$

CMVM Proof

where the differential entropy of the circular distribution $f(\theta)$, $h(f)$ does not affect the optimization,

$$[\mu^*, \kappa^*] = arg \max_{\mu,\kappa} \int_0^{2\pi} f(\theta) \log g(\theta) d\theta \tag{20}$$

After the insertion of the Tikhonov form into (20), we get

$$[\mu^*, \kappa^*] = arg \max_{\mu,\kappa} \int_0^{2\pi} f(\theta) Re[\kappa e^{-j(\theta-\mu)}] d\theta \; - \log 2\pi I_0(\kappa) \tag{21}$$

Rewriting (21) as an expectation and maximizing over μ only,

$$\mu^* = arg \max_{\mu} \kappa \mathbb{E}(Re[e^{-j(\theta-\mu)}]) \tag{22}$$

CMVM Proof

Using the linearity of the expectation and real operators,

$$\mu^* = arg \max_{\mu} \kappa Re[\mathbb{E}(e^{j(\theta-\mu)})] \tag{23}$$

We can view (23) as an inner product operation and therefore, the maximal value of μ is obtained, according to the Cauchy-Schwartz inequality, for

$$\mu^* = \angle \mathbb{E}(e^{j(\theta)}) \tag{24}$$

CMVM Proof

Now we move on to finding the optimal κ, using the fact that we found the optimal μ. For μ^*, the optimal $g(\theta)$ needs to satisfy

$$\frac{\partial D(f||g)}{\partial \kappa} = 0 \tag{25}$$

After applying the partial derivative to (21), and using

$$\frac{dI_0(\kappa)}{d\kappa} = \frac{I_1(\kappa)}{I_0(\kappa)} \tag{26}$$

CMVM Proof

We get,

$$\mathbb{E}(Re[e^{-j(\theta-\mu^*)}]) = \frac{I_1(\kappa^*)}{I_0(\kappa^*)} \tag{27}$$

Recalling the definitions of circular moments, we get that the optimal Tikhonov distribution $g(\theta)$ is given by matching its circular mean and variance to the circular mean and circular variance of the distribution $f(\theta)$.

Apply CMVM

At each clustering iteration, a set J of mixture components indices of the input Tikhonov mixture is selected. The corresponding mixture components are clustered using the CMVM operator. In this appendix we will explicitly compute the application of the CMVM operator and introduce several approximations to speed up the computational complexity. For simplicity, assume that the mixture components in the set J are,

$$f^J(\theta_k) = \sum_{l \in J}^{|J|} \alpha_l \frac{e^{Re[Z_l e^{-j\theta_k}]}}{2\pi I_0(|Z_l|)} \tag{28}$$

Apply CMVM

Using CMVM theorem and skipping the algebraic details, the CMVM operator for (28), is:

$$\text{CMVM}(f^J(\theta_k)) = \frac{e^{Re[Z_k^f e^{-j\theta_k}]}}{2\pi I_0(|Z_k^f|)} \tag{29}$$

where

$$Z_k^f = \hat{k} e^{j\hat{\mu}} \tag{30}$$

and

$$\hat{\mu} = \arg \sum_{l \in J}^{|J|} \alpha_l \frac{I_1(|Z_l|)}{I_0(|Z_l|)} e^{j \arg(Z_l)} \tag{31}$$

$$\frac{1}{2\hat{k}} = 1 - \sum_{l \in J}^{|J|} \alpha_l \frac{I_1(|Z_l|)}{I_0(|Z_l|)} Re[e^{j(\hat{\mu} - \arg(Z_l))}] \tag{32}$$

Apply CMVM

Thus, the approximated versions of (32) and (31) are

$$\hat{\mu} = \arg[\sum_{l \in J}^{|J|} \alpha_l (1 - \frac{1}{2|Z_l|}) e^{j \arg(Z_l)}] \tag{33}$$

$$\frac{1}{2\hat{k}} = 1 - \sum_{l \in J}^{|J|} \alpha_l (1 - \frac{1}{2|Z_l|}) \cos(\hat{\mu} - \arg(Z_l)) \tag{34}$$

We also use the approximation for the modified bessel function in the computation of α_l.

Apply CMVM

For a small enough ϵ, $\cos(\hat{\mu} - \arg(Z_l)) \approx 1$, thus one can further reduce the complexity of (34)

$$\frac{1}{\hat{k}} = \sum_{l \in J}^{|J|} \alpha_l \frac{1}{|Z_l|} \tag{35}$$

which coincides with the computation of a variance of a Gaussian mixture.

Mixture Reduction As Phase Noise Tracking

Multiple PLLs Equivalence

Assuming slowly varying phase noise and high SNR, the mixture reduction tracking loop i, $\hat{\theta}_k^i$ for each trajectory can be computed in the following manner,

$$\hat{\theta}_k^i \approx \hat{\theta}_{k-1}^i + \frac{|r_{k-1}||c_t|}{G_{k-1}\sigma^2}(\angle r_{k-1} + \angle c_t - \hat{\theta}_{k-1}^i)$$

where, c_t and G_{k-1} are a soft decision of the constellation symbol and the inverse conditional MSE for $\hat{\theta}_{k-1}$,

Multiple PLL Equivalence Theorem

Under certain channel conditions, the mixture reduction algorithms can be viewed as multiple PLLs tracking the different phase trajectories. For reasons of simplicity, will only show the case where the mixture reduction algorithm converges to a single PLL (the generalization for more than one PLL is trivial, as long as there are no splits). As described earlier, we model the forward messages as Tikhonov mixtures. Suppose the m^{th} component is,

$$p_f^m(\theta_{k-1}) = \frac{e^{Re[z_m^{k-1,f} e^{-j\theta_{k-1}}]}}{2\pi I_0(|z_m^{k-1,f}|)} \quad (36)$$

Multiple PLL Equivalence Theorem

then we get a Tikhonov mixture $f(\theta_k)$,

$$f(\theta_k) = \sum_{i=1}^{M} \alpha_i f_i(\theta_k) \tag{37}$$

where,

$$f_i(\theta_k) = \frac{e^{Re[\tilde{z}_{m,i}^{k-1,f} e^{-j\theta_k}]}}{2\pi I_0(|\tilde{z}_{m,i}^{k-1,f}|)} \tag{38}$$

$$\tilde{z}_{m,i}^{k-1,f} = \frac{(z_m^{k-1,f} + \frac{r_{k-1}x_i^*}{\sigma^2})}{1 + \sigma_\Delta^2 |(z_m^{k-1,f} + \frac{r_{k-1}x_i^*}{\sigma^2})|} \tag{39}$$

and x_i is the i^{th} constellation symbol.

Multiple PLL Equivalence Theorem

We insert (37) into the mixture reduction algorithms. Assuming slowly varying phase noise and high SNR, such that the mixture reduction will cluster all the mixture components, with non negligible probability, to one Tikhonov distribution. Then, the **circular** mean, $\hat{\theta}_k$, of the clustered Tikhonov distribution is computed according to,

$$\hat{\theta}_k = \angle \mathbb{E}(e^{j\theta_k}) \tag{40}$$

where the expectation is over the distribution $f(\theta_k)$. We note that for every complex valued scalar z, the following holds

$$\angle z = \Im(\log z) \tag{41}$$

Multiple PLL Equivalence Theorem

where $\Im$ denotes the imaginary part of a complex scalar. If we apply (41) to (40) we get,

$$\hat{\theta}_k = \Im\left(\log \sum_{i=1}^{M} \alpha_i \frac{\tilde{z}_{m,i}^{k-1,f}}{|\tilde{z}_{m,i}^{k-1,f}|}\right) \tag{42}$$

which can be rewritten as,

$$\hat{\theta}_k = \Im\left(\log \sum_{i=1}^{M} \alpha_i \frac{z_m^{k-1,f} + \frac{r_{k-1}x_i^*}{\sigma^2}}{|z_m^{k-1,f} + \frac{r_{k-1}x_i^*}{\sigma^2}|}\right) \tag{43}$$

we denote,

$$G_{k-1} = |z_m^{k-1,f} + \frac{r_{k-1}x_i^*}{\sigma^2}| \tag{44}$$

and assume that G_{k-1}, the conditional causal MSE of the phase estimation under mixture component $f_i(\theta_k)$, is constant for all significant components.

Multiple PLL Equivalence Theorem

Then,

$$\hat{\theta}_k \approx \hat{\theta}_{k-1} + \Im\left(\log\left(1 + \frac{r_{k-1}}{G_{k-1} z_m^{k-1,f} \sigma^2}\left(\sum_{i=1}^{M} \alpha_i x_i^*\right)\right)\right) \tag{45}$$

where,

$$\hat{\theta}_{k-1} = \angle z_m^{k-1,f} \tag{46}$$

We will define c_{soft} as the soft decision symbol using the significant components,

$$c_{soft} = \sum_{i=1}^{M} \alpha_i x_i \tag{47}$$

Multiple PLL Equivalence Theorem

Since we assume high SNR and small phase noise variance, then the tracking conditional MSE will be low, i.e $|z_1^{k,f}|$ will be high. Using the fact that for small angles ϕ,

$$\angle(1+\phi) \approx \Im(\phi) \tag{48}$$

Therefore,

$$\hat{\theta}_k \approx \hat{\theta}_{k-1} + \Im\left(\frac{r_{k-1} c^*_{soft}}{G_{k-1} z_m^{k-1,f} \sigma^2}\right) \tag{49}$$

Which, again for small angles x, $sin(x) \approx x$,

$$\hat{\theta}_k \approx \hat{\theta}_{k-1} + \frac{|r_{k-1}||c^*_{soft}|}{G_{k-1}|z_m^{k-1,f}|\sigma^2}(\angle r_{k-1} + \angle c^*_{soft} - \hat{\theta}_{k-1}) \tag{50}$$

Accuracy Theorem

In the first iteration, the algorithm selects the highest probability mixture component and denotes it as $f_{lead}(\theta)$. Let M_0, be the set of mixture components $f_i(\theta)$ selected for clustering,

$$M_0 = \{f_i(\theta) \mid D_{KL}(f_i(\theta)||f_{lead}(\theta)) \leq \epsilon\} \quad (51)$$

and M_1 be the set of mixture components which were not selected,

$$M_1 = \{f_i(\theta) \mid D_{KL}(f_i(\theta)||f_{lead}(\theta)) > \epsilon\} \quad (52)$$

Accuracy Theorem

Thus,

$$\sum_{i \in M_0} \frac{\alpha_i}{\beta_1} D_{KL}(f_i(\theta)||f_{lead}(\theta)) \leq \epsilon \tag{53}$$

where,

$$\beta_1 = \sum_{i \in M_0} \alpha_i \tag{54}$$

Using Lemma (3),

$$D_{KL}\left(\sum_{i \in M_0} \frac{\alpha_i}{\beta_1} f_i(\theta)||f_{lead}(\theta)\right) \leq \epsilon \tag{55}$$

Accuracy Theorem

The algorithm then clusters all the distributions in M_0 using CMVM,

$$g_1(\theta) = CMVM\left(\sum_{i \in M_0} \frac{\alpha_i}{\beta_1} f_i(\theta)\right) \quad (56)$$

then,

$$D_{KL}\left(\sum_{i \in M_0} \frac{\alpha_i}{\beta_1} f_i(\theta) || g_1(\theta)\right) \leq D_{KL}\left(\sum_{i \in M_0} \frac{\alpha_i}{\beta_1} f_i(\theta) || f_{lead}(\theta)\right) \quad (57)$$

which means that,

$$D_{KL}\left(\sum_{i \in M_0} \frac{\alpha_i}{\beta_1} f_i(\theta) || g_1(\theta)\right) \leq \epsilon \quad (58)$$

Accuracy Theorem

We can rewrite the mixtures $f(\theta)$ and $g(\theta)$ in the following way,

$$f(\theta) = \alpha_{M_0} f_{M_0}(\theta) + \alpha_{M_1} f_{M_1}(\theta) \tag{59}$$

$$g(\theta) = \beta_1 g_1(\theta) + (1 - \beta_1) h(\theta) \tag{60}$$

where,

$$\alpha_{M_0} = \sum_{i \in M_0} \alpha_i \tag{61}$$

$$\alpha_{M_1} = \sum_{i \in M_1} \alpha_i \tag{62}$$

$$f_{M_i}(\theta) = \sum_{j \in M_i} \frac{\alpha_j}{\alpha_{M_i}} f_j(\theta) \tag{63}$$

Accuracy Theorem

Using (54),

$$\alpha_{M_i} = \beta_i \tag{64}$$

Therefore (59) and (60) are two mixtures of the same size and have exactly the same coefficients, thus the KL of the probability mass functions induced by the coefficients of both mixtures is zero. Using Lemma (5),

$$D_{KL}(f(\theta)||g(\theta)) \leq \beta_1 D_{KL}(f_{M_0}(\theta)||g_1(\theta) + (1-\beta_1)D_{KL}(f_{M_1}(\theta)||h(\theta)) \tag{65}$$

using (57) we get,

$$D_{KL}(f(\theta)||g(\theta)) \leq \beta_1 \epsilon + (1-\beta_1)D_{KL}(f_{M_1}(\theta)||h(\theta)) \tag{66}$$

Accuracy Theorem

If we find a Tikhonov mixture $h(\theta)$,which satisfies,

$$D_{KL}(f_{M_1}(\theta)||h(\theta)) \leq \epsilon \tag{67}$$

then we will prove the theorem. But (67) is exactly the same as the original problem, thus applying the same clustering steps as described earlier on the new mixture $f_{M_1}(\theta)$ will ultimately satisfy,

$$D_{KL}(f(\theta)||g(\theta)) \leq \epsilon \tag{68}$$

Directional Statistics

Introduction

Directional statistics is a branch of mathematics which studies random variables defined on circles and spheres.
The **circular** mean and variance of a circular random variable θ, are defined as

$$\mu_C = \angle \mathbb{E}(e^{j\theta})$$

$$\sigma_C^2 = \mathbb{E}(1 - cos(\theta - \mu_C))$$

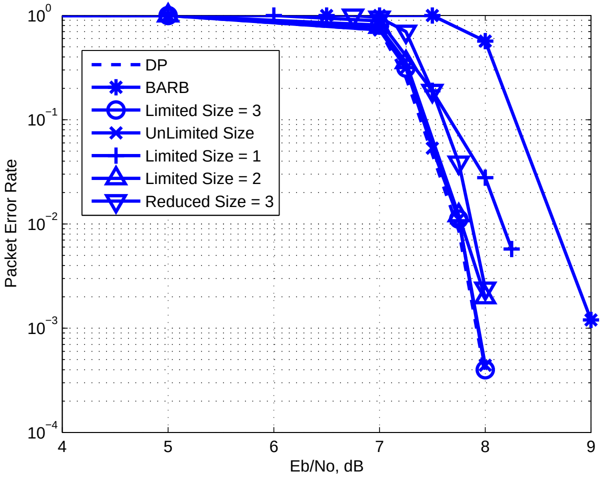
Mixture Reduction Algorithm with Cycle Slip Estimation

$j \leftarrow 1$
while $j \leq L$ or $|f(\theta)| > 0$ **do**
  $lead \leftarrow argmax\{\underline{\alpha}\}$
  for $i = 1 \rightarrow |f(\theta)|$ **do**
    if $D_{KL}(f_i(\theta)||f_{lead}(\theta)) \leq \epsilon$ **then**
      $idx \leftarrow [idx, i]$
    end if
  end for
  $\beta_j \leftarrow \sum_{i \in idx} \alpha_i$
  $g_j(\theta) \leftarrow CMVM(\sum_{i \in idx} \frac{\alpha_i}{\beta_j} f_i(\theta))$
  $f(\theta) \leftarrow f(\theta) - \sum_{i \in idx} \alpha_i f_i(\theta)$
  $j \leftarrow j + 1$
end while
$\phi_k^f \leftarrow (\sum_j \beta_j)\phi_{k-1}^f$

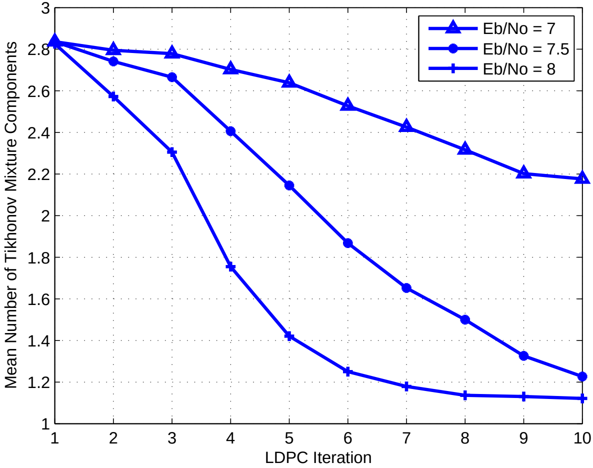
Recovering From Cycle Slips

$p_f(\theta_0) \leftarrow \frac{1}{2\pi}$
$\phi_0^f \leftarrow 1$
$k \leftarrow 1$
while $k \leq K$ **do**
  Compute $p_d(\theta_{k-1})$
  if c_{k-1} is a pilot **then**
    $q_f(\theta_{k-1}) \leftarrow \phi_{k-1}^f p_f(\theta_{k-1}) + (1 - \phi_{k-1}^f)\frac{1}{2\pi}$
    $t \leftarrow 1$
  else
    $q_f(\theta_{k-1}) \leftarrow p_f(\theta_{k-1})$
    $t \leftarrow \phi_{k-1}^f$
  end if
  $\tilde{p}_f(\theta_k) \leftarrow \int_0^{2\pi} q_f(\theta_{k-1}) p_d(\theta_{k-1}) p_\Delta(\theta_k - \theta_{k-1}) d\theta_{k-1}$
  $[p_f(\theta_k), \phi_k^f] \leftarrow MixReductionAlgo(\tilde{p}_f(\theta_k), t)$
  $k \leftarrow k + 1$
end while

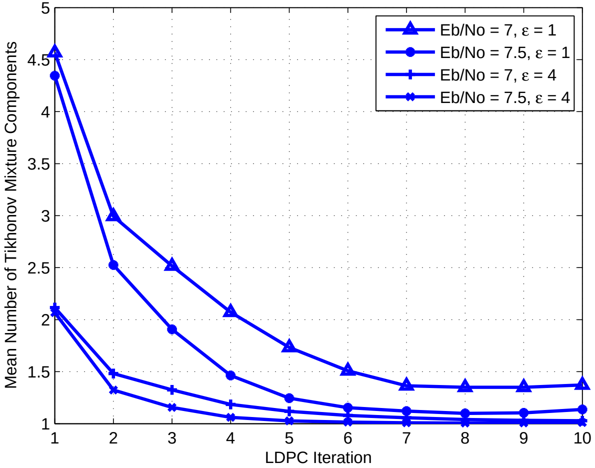
8PSK - $\sigma_\Delta = 0.05$, Pilot Frequency $= 0.05$



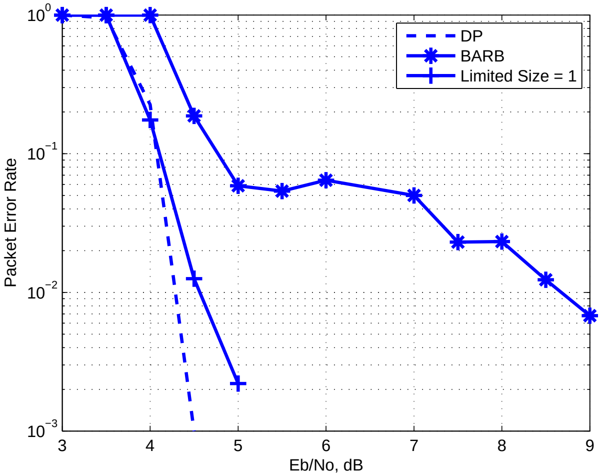
8PSK Mean Number of Tikhonov Mixture Components - Reduced Complexity Algorithm, Maximum 3 lobes, $\epsilon = 1$



8PSK Mean Number of Tikhonov Mixture Components - Unlimited Algorithm



BPSK - $\sigma_\Delta = 0.1$, Pilot Frequency = 0.0125



Future Research

- Use the mixture model framework for asymmetrical constellation analysis (Accepted to GlobeCom 2013)
- Compute the mean number of mixture components for given channel conditions