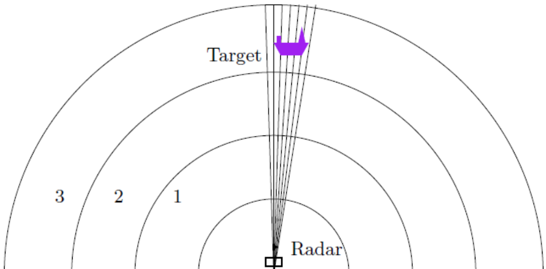
Extended Source Localization Using Variational Methods

Shilpa Rao Chandra R. Murthy

9 April 2016

Introduction

- High resolution array processing - radar, radio astronomy, radio communications.
- Point target assumption is an approximation.
- Target possesses a spatial extent over a continuum of direction of arrivals (DoAs).



System Model

- M_t Tx antennas, spacing Δ_t;
  M_r Rx antennas, spacing Δ_r.
- N_d Doppler bins, N_r range bins, N_a angular bins.
- $\mathbf{s}_i \in \mathbb{C}^{L\times 1}$: waveform transmitted by the ith Tx antenna.
- For the dth Doppler bin
  $\mathbf{s}_i(\omega_d) = \mathbf{s}_i \odot [1, e^{j\omega_d}, \ldots, e^{j(L-1)\omega_d}]^T$,
  $\mathbf{S}_d = [\mathbf{s}_1(\omega_d)\ \mathbf{s}_2(\omega_d)\ \cdots\ \mathbf{s}_{M_t}(\omega_d)]^T$.

System Model

- $x_{d,r}^{(k,p)}(\theta)$: complex angular weighting function of the kth source in direction θ for the radar sweep index p, dth Doppler bin, rth range bin.
- $\mathbf{a}(\theta)$: Tx steering vector, $\mathbf{b}(\theta)$: Rx steering vector.
- Received signal $\mathbf{Y}^{(p)} \in \mathbb{C}^{M_r \times (L+N_r-1)}$:
  $$\mathbf{Y}^{(p)} = \sum_{d=1}^{N_d} \sum_{r=1}^{N_r} \sum_{k=1}^{K} \int_{\theta \in \Theta_k} \{x_{d,r}^{(k,p)}(\theta)\mathbf{b}(\theta)\mathbf{a}^T(\theta)d\theta\}\tilde{\mathbf{S}}_d \mathbf{J}_r + \mathbf{W}^{(p)},$$
- $\tilde{\mathbf{S}}_d = [\mathbf{S}_d \quad \mathbf{0}_{M_t \times N_r - 1}]$ $\mathbf{J}_r = \begin{pmatrix} \overbrace{0 \dots 01}^{r} & & \mathbf{0} \\ & \ddots & \\ & & 1 \\ \mathbf{0} & & \end{pmatrix}$.

System Model

- $\mathbf{Y}^{(p)} = \sum_{d=1}^{N_d} \sum_{r=1}^{N_r} \sum_{a=1}^{N_a} x_{d,r,a}^{(p)} \mathbf{b}(\theta_a)\mathbf{a}^T(\theta_a)\tilde{\mathbf{S}}_d \mathbf{J}_r + \mathbf{W}^{(p)}$.
  [Approximation: $x_{d,r,a}^{(p)} = x_{d,r}^{(k,p)}(\theta_a)\delta\theta$].

- Vectorize to get: $\mathbf{y}^{(p)} = \mathbf{A}\mathbf{x}^{(p)} + \mathbf{w}^{(p)}$.

- $\mathbf{A} = [\mathbf{u}_{1,1,1}\ \mathbf{u}_{1,1,2} \cdots \mathbf{u}_{N_d,N_r,N_a}] \in \mathbb{C}^{M \times N}$,
  $\mathbf{u}_{d,r,a} = \mathrm{vec}(\mathbf{b}(\theta_a)\mathbf{a}^T(\theta_a)\tilde{\mathbf{S}}_d \mathbf{J}_r)$,
  $\mathbf{x}^{(p)} = [x_{1,1,1}^{(p)}, x_{1,1,2}^{(p)}, \ldots, x_{N_d,N_r,N_a}^{(p)}]^T \in \mathbb{C}^{N \times 1}$.

- $\mathbf{x}^{(p)}$ is block-sparse.

- P radar sweeps: $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{W}$.

Variational Methods - Introduction

- Bayesian models have become increasingly important to address long-standing theoretical questions.
- Problem of probabilistic inference: computing a conditional probability distribution over the values of some of the nodes (the "hidden nodes"), given the values of other nodes ("evidence" nodes).
  $P(H|E) = P(H, E)/P(E)$.
- Exact algorithms provide a satisfactory solution to inference problems, but there are cases when time and space complexity of the exact calculation is unacceptable. - Variational approximations.
- Markov chain Monte Carlo (MCMC): requires massive computing resources, converge slowly and might approximate the wrong posterior.

Variational Methods - Introduction

- Variational method- deterministic approximation procedures that generally provide bounds on probabilities of interest.
- Intuition- complex graphs can be probabilistically simple.
- If $\mathbf{y}$ are the observations, $\mathbf{x}$ are the hidden variables, $\boldsymbol{\theta}$ are the unknown parameters ,EM involves:

  E-step: Compute $p(\mathbf{x}|\mathbf{y};\boldsymbol{\theta}_{\mathrm{old}})$.
  M-step: Evaluate $\boldsymbol{\theta}_{\mathrm{new}} = \arg\max_{\boldsymbol{\theta}} \mathbb{E}_{\mathbf{x}|\mathbf{y};\boldsymbol{\theta}_{\mathrm{old}}}[\ln p(\mathbf{y},\mathbf{x};\boldsymbol{\theta})]$

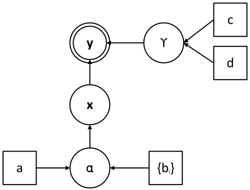
Variational Methods- Introduction

- EM requires that $p(\mathbf{x}|\mathbf{y};\boldsymbol{\theta})$ be explicitly known, or be able compute the conditional expectation.
- Variational EM- Bypasses knowledge of $p(\mathbf{x}|\mathbf{y};\boldsymbol{\theta})$ by assuming an appropriate $q(\mathbf{x})$ and lower-bounding the log-likelihood $(F(q,\boldsymbol{\theta}))$.

  Variational E-step: Evaluate $q_{\mathrm{new}}(\mathbf{z})$ to maximize $F(q,\boldsymbol{\theta}_{\mathrm{old}})$.
  Variational M-step: Find $\boldsymbol{\theta}_{\mathrm{new}} = \arg\max_{\boldsymbol{\theta}} F(q_{\mathrm{new}},\boldsymbol{\theta})$.
- Variational Garrote: Principaled approach to feature subset selection based on variational approximation of posterior through an alternate means of specifying prior to encourage sparsity.

Variational EM

- Hierarchical prior model.
- Estimation of $\mathbf{X}$ is viewed as the estimation of $\{\mathbf{x}^{(p)}\}_{p=1}^{P}$
- $\mathbf{x}^{(p)}$ is, in turn, viewed as a concatenation of several smaller blocks $\mathbf{x}_{b,r}^{(p)} \in \mathbb{R}^{h_r}$ corresponding to bth block of range bin r; h_r: size of the block in range bin r.
- [Z. Zhang & B.D. Rao, TSP Apr. '13] $\mathbf{x}_{b,r}^{(p)}$ satisfies $p(\mathbf{x}_{b,r}^{(p)}|\alpha_{b,r}, \mathbf{B}_{b,r}) \sim \mathcal{N}(0, \alpha_{b,r}^{-1}\mathbf{B}_{b,r})$; $\alpha_{b,r}$: hyperparameter controlling sparsity $\mathbf{B}_{b,r}$: positive definite matrix captures the correlations between the elements of $\mathbf{x}_{b,r}^{(p)}$.



Variational EM

- $p(\boldsymbol{\alpha}) = \prod_{r=1}^{N_r} \prod_{b=1}^{B_r} \mathrm{Gamma}(\alpha_{b,r}|c,d)$

- Marginal prob. : $\ln p(\mathbf{Y}) = L(q) + \mathrm{KL}(q \parallel p)$
  $L(q) = \int q(\boldsymbol{\Theta}) \ln \frac{p(\mathbf{Y},\boldsymbol{\Theta})}{q(\boldsymbol{\Theta})} d\boldsymbol{\Theta}$
  $\mathrm{KL}(q \parallel p) = -\int q(\boldsymbol{\Theta}) \ln \frac{p(\boldsymbol{\Theta}|\mathbf{Y})}{q(\boldsymbol{\Theta})} d\boldsymbol{\Theta}$
  Hidden variables $\boldsymbol{\Theta} = \{\mathbf{X}, \boldsymbol{\alpha}\}$

- $q(\boldsymbol{\Theta}) = \prod_i q_i(\Theta_i)$

- $L(q)$ maximized when $p(\boldsymbol{\Theta}|\mathbf{Y}) = q(\boldsymbol{\Theta})$.

Variational EM

- Posterior distribution of each hidden variable computed by maximizing $L(q)$ while keeping other variables fixed using their most recent distributions.

- $\ln q_X(\mathbf{X}) = \langle \ln p(\mathbf{Y}, \mathbf{X}, \boldsymbol{\alpha}) \rangle_{q_\alpha(\boldsymbol{\alpha})} + \text{constant}$
  $\ln q_\alpha(\boldsymbol{\alpha}) = \langle \ln p(\mathbf{Y}, \mathbf{X}, \boldsymbol{\alpha}) \rangle_{q_X(\mathbf{X})} + \text{constant}$

- Variational EM:
  - Variational E-step: Given $\mathbf{X}$ from $q_X(\mathbf{X})$, compute $q_\alpha(\boldsymbol{\alpha})$.
  - Variational M-step: Given $q_\alpha(\boldsymbol{\alpha})$, compute $\mathbf{X}$ that maximizes $L(q)$.

- Then

$$\ln q_X(\mathbf{X}) \propto \langle \ln p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\alpha}) + \ln p(\mathbf{X}|\boldsymbol{\alpha}) \rangle_{q_\alpha(\boldsymbol{\alpha})},$$

$$\propto -\frac{\lambda}{2} \sum_{p=1}^{P} \left[(\mathbf{y}^{(p)} - \mathbf{A}\mathbf{x}^{(p)})^T (\mathbf{y}^{(p)} - \mathbf{A}\mathbf{x}^{(p)}) - \frac{1}{2} \left(\mathbf{x}^{(p)} \right)^T \langle \boldsymbol{\Sigma}_0^{-1} \rangle \mathbf{x}^{(p)} \right],$$

$$\ln q_\alpha(\boldsymbol{\alpha}) \propto \sum_{r=1}^{N_r} \sum_{b=1}^{B_r} \left[\left(c + \frac{P}{2} \right) \ln \alpha_{b,r} - \left(d + \sum_{p=1}^{P} \langle \mathbf{x}_{b,r}^{(p)} \mathbf{B}_{b,r}^{-1} \mathbf{x}_{b,r}^{(p)} \rangle \right) \alpha_{b,r} \right].$$

Variational EM - Algorithm

Algorithm 1 Block VB

1: **Input:**

Data $\{\mathbf{y}^{(p)}, \mathbf{A}\}, p = 1, 2, \ldots, P$, and block sizes $\{h_1, h_2, \ldots, h_{N_r}\}$.

2: **Initialize:**

Set $\alpha_{b,r}$ to random values, $c = d = 10^{-6}$.

3: **Repeat until** $\|\hat{\mathbf{X}}^{(t+1)} - \hat{\mathbf{X}}^{(t)}\|_F < \epsilon$:

(a) Form $\langle \mathbf{\Sigma}_0 \rangle = \mathrm{diag}\{\langle \mathbf{\Sigma}_1 \rangle, \langle \mathbf{\Sigma}_2 \rangle, \ldots, \langle \mathbf{\Sigma}_{N_r} \rangle\}$.

(b) Compute $\mathbf{\Sigma}^{t+1} = (\mathbf{A}^H \mathbf{A} + \mathbf{\Sigma}_0^{-1})^{-1}$.

(c) Compute $\hat{\mathbf{X}} = \mathbf{\Sigma}^{t+1} \mathbf{A}^H \mathbf{Y}$.

(d) Compute $\alpha_{b,r} = \frac{2c+P}{d + \sum_{p=1}^{P} \langle \mathbf{x}_{b,r}^{(p)} \mathbf{B}_{b,r}^{-1} \mathbf{x}_{b,r}^{(p)} \rangle}$

Variational Garrote

- Alternate Bayesian approach- uses a variational approximation for feature subset selection.
- Computationally efficient, provides more accurate predictions than methods like Lasso, ridge regression and the paired mean field.
- A binary variable for each unknown- provides an adaptive description of the support.
  Due to the decoupling of the estimation of the support and the unknown vector, the VG provides excellent estimates.
- VG extended to a block-sparse recovery problem by associating a binary selector variable with a block of the unknown vector.

Variational Garrote

- Re-write: $\mathbf{y}^{(p)} = \sum_{r=1}^{N_r} \sum_{b=1}^{B_r} s_{b,r} \mathbf{A}_{b,r} \mathbf{x}_{b,r}^{(p)} + \mathbf{w}^{(p)}$
  $s_{b,r} \in \{0, 1\}$.

- Prior distribution on $\mathbf{s}$

$$p(\mathbf{s}|\gamma) = \prod_{r=1}^{N_r} \prod_{b=1}^{B_r} p(s_{b,r}|\gamma), \qquad p(s_{b,r}|\gamma) = \frac{\exp(\gamma s_{b,r})}{1 + \exp(\gamma)},$$

  $\gamma < 0$: sparse solutions.

- Likelihood of measurements:

$$p(\mathbf{Y}|\mathbf{s}, \mathbf{X}; \lambda) = \left(\frac{\lambda}{2\pi} \right)^{\frac{PM}{2}} \exp \left\{ \frac{-\lambda M}{2} \sum_p \left((\sigma_y^{(p)})^2 - \sum_{r=1}^{N_r} \sum_{b=1}^{B_r} s_{b,r} \left((\mathbf{v}_{b,r}^{(p)})^H \mathbf{x}_{b,r}^{(p)} + (\mathbf{x}_{b,r}^{(p)})^H \mathbf{v}_{b,r}^{(p)} \right) + \sum_{r,t=1}^{N_r} \sum_{b,c=1}^{B_r,B_t} s_{b,r} s_{c,t} (\mathbf{x}_{b,r}^{(p)})^H \mathbf{D}_{bc,rt} \mathbf{x}_{c,t}^{(p)} \right) \right\}$$

$(\sigma_y^{(p)})^2 = \frac{1}{M} (\mathbf{y}^{(p)})^H \mathbf{y}^{(p)}$, $\mathbf{v}_{b,r}^{(p)} = \frac{1}{M} \mathbf{A}_{b,r}^H \mathbf{y}^{(p)}$ and $\mathbf{D}_{bc,rt} = \frac{1}{M} \mathbf{A}_{b,r}^H \mathbf{A}_{c,t}$.

Variational Garrote

- Posterior of $\mathbf{X}$: $p(\mathbf{X}|\mathbf{Y},\gamma;\lambda) \propto \sum_{\mathbf{s}} p(\mathbf{Y}|\mathbf{s},\mathbf{X};\lambda)p(\mathbf{s}|\gamma)$.

- Approximation:
$\log \sum_{\mathbf{s}} p(\mathbf{Y}|\mathbf{s},\mathbf{X};\lambda)p(\mathbf{s}|\gamma) \geq -\sum_{\mathbf{s}} q(\mathbf{s}) \log \frac{q(\mathbf{s})}{p(\mathbf{s}|\gamma)p(\mathbf{Y}|\mathbf{s},\mathbf{X};\lambda)} = -F(q,\mathbf{X},\lambda)$.
$q(\mathbf{s}) = \prod_{r=1}^{N_r}\prod_{b=1}^{B} q(s_{b,r})$ with $q(s_{b,r}) = m_{b,r}s_{b,r} + (1-m_{b,r})(1-s_{b,r})$.

- Solve for F:

$$F = \frac{\lambda M}{2}\sum_p \left(\sum_{r,t=1}^{N_r}\sum_{b,c=1}^{B_r,B_t} m_{b,r}m_{c,t}(\mathbf{x}_{b,r}^{(p)})^H \mathbf{D}_{bc,rt}\mathbf{x}_{c,t}^{(p)} \right.$$
$$+ \sum_{r=1}^{N_r}\sum_{b=1}^{B_r} m_{b,r}(1-m_{b,r})(\mathbf{x}_{b,r}^{(p)})^H \mathbf{D}_{bb,rr}\mathbf{x}_{b,r}^{(p)}$$
$$\left. -\sum_{r=1}^{N_r}\sum_{b=1}^{B_r} m_{b,r}\left((\mathbf{v}_{b,r}^{(p)})^H \mathbf{x}_{b,r}^{(p)} + (\mathbf{x}_{b,r}^{(p)})^H \mathbf{v}_{b,r}^{(p)} \right) + (\sigma_y^{(p)})^2 \right)$$
$$+ \sum_{r=1}^{N_r}\sum_{b=1}^{B_r} (m_{b,r}\log m_{b,r} + (1-m_{b,r})\log(1-m_{b,r})) - \sum_{r=1}^{N_r}\sum_{b=1}^{B_r} \gamma m_{b,r}.$$

Variational Garrote

- Updates for $\mathbf{m}$ and $\mathbf{x}^{(p)}$:

$$\mathbf{x}^{(p)} = (\mathbf{D}^{'})^{-1}\mathbf{v}^{(p)} \ \forall p,$$

$$m_{b,r} = \sigma\left(\gamma + \frac{\lambda M}{2}\sum_{p}(\mathbf{x}_{b,r}^{(p)})^{H}\mathbf{D}_{bb,rr}\mathbf{x}_{b,r}^{(p)}\right),$$

$\mathbf{D}^{'}$: matrix with $(t-1)N_r + (c-1)B_t + 1 : (t-1)N_r + cB_t$ rows and $(r-1)N_r + (b-1)B_r + 1 : (r-1)N_r + bB_r$ columns are $m_{b,r}\mathbf{D}_{bc,rt} + (1-m_{b,r})\mathbf{D}_{cc,tt}\delta_{bc}\delta_{rt}$.

- To learn γ, we see that the probability of $s_{b,r} = 1$ is $p(s_{b,r} = 1|\gamma) = \frac{\exp(\gamma)}{1+\exp(\gamma)}, \quad q(s_{b,r} = 1) = m_{b,r}$.
$\gamma = \frac{1}{(\sum_{r=1}^{N_r} B_r)}\sum_{r=1}^{N_r}\sum_{b=1}^{B_r}\ln\left(\frac{m_{b,r}}{1-m_{b,r}}\right)$.

Block VG

Algorithm 2 Block VG

1: **Input:**
Data $\{\mathbf{y}^{(p)}, \mathbf{A}\}, p = 1, 2, \ldots, P$, and block sizes $\{h_1, h_2, \ldots, h_{N_r}\}$.

2: **Initialize:**
Compute $\mathbf{v}_{b,r}^{(p)}$ and $\mathbf{D}_{bc,rt}$ for $r, t = 1, 2, \ldots, N_r, b = 1, 2, \ldots, B_r$ for each r, where $B_r = N_a N_d / h_r$ and $c = 1, 2, \ldots, B_t$ for each t, where $B_t = N_a N_d / h_t$; set $m_{b,r}$ to random values. Set the initial value of $\mathbf{D}^{'}$ from $m_{b,r}$

3: **Repeat until** $\|\mathbf{m}^{(t+1)} - \mathbf{m}^{(t)}\|_2 < \epsilon$:

(a) Update $\mathbf{x}^{(p)}$ and $m_{b,r}$.

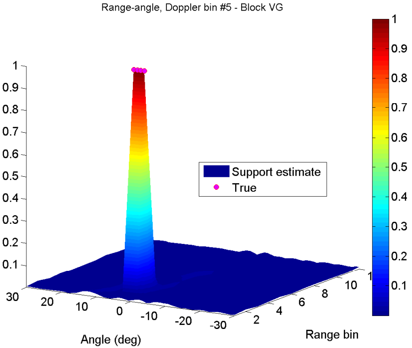
(b) Update γ.

(c) Compute the matrix $\mathbf{D}^{'}$ using the latest values of $m_{b,r}$.

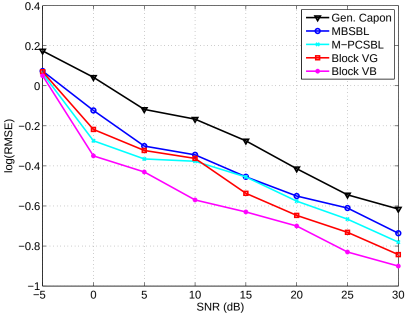
(d) Update $\mathbf{m}$ for the current iteration: $\mathbf{m}^{(t+1)} = [m_{1,1}, m_{2,1}, \ldots, m_{B_{N_r},N_r}]$.

Results

$M_t = 5,\ M_r = 5,\ N_r = 12,\ N_a = 61(-30^\circ : 30^\circ),\ N_d = 11,\ P = 50.$

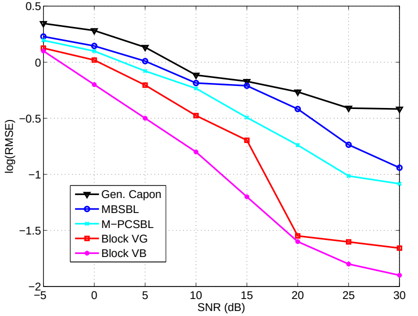


(i) Support estimate.

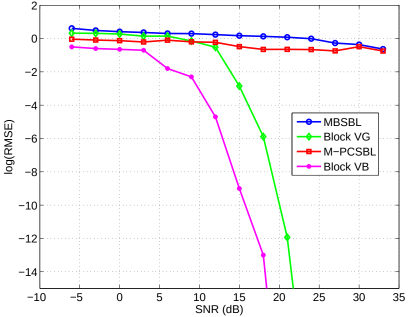


(ii) RMSE of central angle estimate versus SNR.

Results



(i) RMSE of angular spread versus SNR.



(ii) RMSE of range versus SNR.

Conclusion

- Extended source localization problem in radar/sonar - joint estimation of angle, spread, Doppler and range.
- Block-sparse MMV problem with common support across radar sweeps.
- Two methods - variational EM and variational Garrote.
- Future work - plot CRB-type bounds for the two variational methods
  - analysis of convergence of these algorithms.