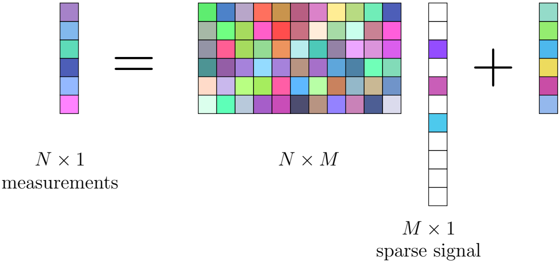
Sparse processing / Compressed sensing

Model : $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$, $\quad$ $\mathbf{x}$ is sparse



- Problem : Solve for $\mathbf{x}$
- Basis pursuit, LASSO (convex objective function)
- Matching pursuit (greedy method)
- Sparse Bayesian Learning (non-convex objective function)

Bayesian interpretation of LASSO

MAP estimate via the unconstrained -LASSO- formulation

$$\widehat{\mathbf{x}}_{\mathrm{LASSO}}(\mu) = \underset{\mathbf{x}\in\mathbb{C}^N}{\arg\min}\ \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \mu\|\mathbf{x}\|_1$$

Bayes rule:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}$$

MAP estimate:

$$\begin{aligned}
\widehat{\mathbf{x}}_{\mathrm{MAP}} &= \underset{\mathbf{x}}{\arg\max}\ \ln p(\mathbf{x}|\mathbf{y}) \\
&= \underset{\mathbf{x}}{\arg\max}\ [\ln p(\mathbf{y}|\mathbf{x}) + \ln p(\mathbf{x})] \\
&= \underset{\mathbf{x}}{\arg\min}\ [-\ln p(\mathbf{y}|\mathbf{x}) - \ln p(\mathbf{x})]
\end{aligned}$$

MAP estimate via the unconstrained -LASSO- formulation

Bayes rule:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}$$

MAP estimate:

$$\widehat{\mathbf{x}}_{\text{MAP}} = \arg\min_{\mathbf{x}} \left[-\ln p(\mathbf{y}|\mathbf{x}) - \ln p(\mathbf{x}) \right]$$

Gaussian likelihood:

$$p(\mathbf{y}|\mathbf{x}) \propto \mathrm{e}^{-\frac{\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2}{\sigma^2}}$$

Laplace-like prior:

$$p(\mathbf{x}) \propto \prod_{i=1}^{N} \mathrm{e}^{-\frac{\sqrt{(\Re x_i)^2+(\Im x_i)^2}}{\nu}} = \mathrm{e}^{-\frac{\|\mathbf{x}\|_1}{\nu}}$$

MAP estimate (LASSO):

$$\widehat{\mathbf{x}}_{\text{MAP}} = \arg\min_{\mathbf{x}} \left[\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \mu\|\mathbf{x}\|_1 \right] = \widehat{\mathbf{x}}_{\text{LASSO}}(\mu), \;\; \mu = \frac{\sigma^2}{\nu}$$

Prior and Posterior densities (Ex. Murphy)

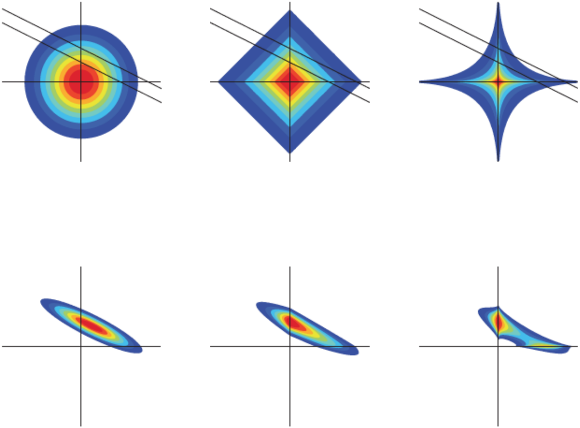


Figure 13.17 Top: plot of log *prior* for three different distributions with unit variance: Gaussian, Laplace and exponential power. Bottom: plot of log *posterior* after observing a single observation, corresponding to a single linear constraint. The precision of this observation is shown by the diagonal lines in the top figure. In the case of the Gaussian prior, the posterior is unimodal and symmetric. In the case of the Laplace prior, the posterior is unimodal and asymmetric (skewed). In the case of the exponential prior, the posterior is bimodal. Based on Figure 1 of (Seeger 2008). Figure generated by `sparsePostPlot`, written by Florian Steinke.

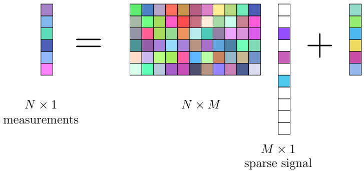
Sparse Bayesian Learning (SBL)

$$\text{Model}: \mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$$

$$\text{Prior}: \mathbf{x} \sim \mathcal{N}(\mathbf{x}; 0, \boldsymbol{\Gamma})$$

$$\boldsymbol{\Gamma} = \mathrm{diag}(\gamma_1, \ldots, \gamma_M)$$

$$\text{Likelihood}: p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}; \mathbf{A}\mathbf{x}, \sigma^2 \mathbf{I}_N)$$



$$\text{Evidence}: p(\mathbf{y}) = \int_{\mathbf{x}} p(\mathbf{y}|\mathbf{x}) p(\mathbf{x}) d\mathbf{x} = \mathcal{N}(\mathbf{y}; 0, \boldsymbol{\Sigma}_{\mathbf{y}})$$

$$\boldsymbol{\Sigma}_{\mathbf{y}} = \sigma^2 \mathbf{I}_N + \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^H$$

$$\begin{aligned} \text{SBL solution}: \hat{\boldsymbol{\Gamma}} &= \arg\max_{\boldsymbol{\Gamma}} p(\mathbf{y}) \\ &= \arg\min_{\boldsymbol{\Gamma}} \left\{ \log |\boldsymbol{\Sigma}_{\mathbf{y}}| + \mathbf{y}^H \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} \mathbf{y} \right\} \end{aligned}$$

M.E.Tipping, "Sparse Bayesian learning and the relevance vector machine," Journal of Machine Learning Research, June 2001.

SBL overview

- SBL solution : $\hat{\mathbf{\Gamma}} = \arg\min_{\mathbf{\Gamma}} \left\{ \log |\mathbf{\Sigma_y}| + \mathbf{y}^H \mathbf{\Sigma_y}^{-1} \mathbf{y} \right\}$

- SBL objective function is non-convex
- Optimization solution is non-unique
- Fixed point update using derivatives, works in practice
- $\mathbf{\Gamma} = \text{diag}(\gamma_1, \ldots, \gamma_M)$

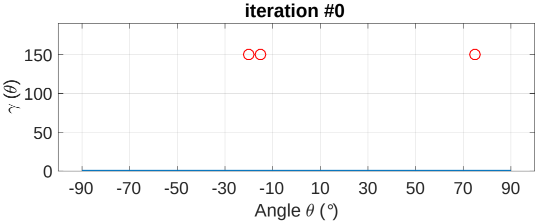
$$\text{Update rule : } \gamma_m^{\text{new}} = \gamma_m^{\text{old}} \left(\frac{||\mathbf{y}^H \mathbf{\Sigma_y}^{-1} \mathbf{a}_m||_2^2}{\mathbf{a}_m^H \mathbf{\Sigma_y}^{-1} \mathbf{a}_m} \right)^r$$

$$\mathbf{\Sigma_y} = \sigma^2 \mathbf{I}_N + \mathbf{A}\mathbf{\Gamma}\mathbf{A}^H$$

- Multi snapshot extension : same $\mathbf{\Gamma}$ across snapshots

SBL overview

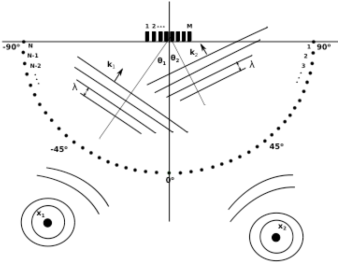
- Posterior : $\mathbf{x}_{\text{post}} = \boldsymbol{\Gamma}\mathbf{A}^H\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{y}$
- At convergence, $\gamma_m \to 0$ for most γ_m
- $\boldsymbol{\Gamma}$ controls sparsity, $\mathsf{E}(|x_m|^2) = \gamma_m$

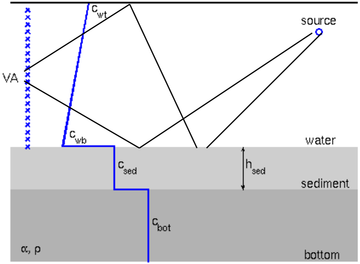


- Different ways to show that SBL gives sparse output
- Automatic determination of sparsity
- Also provides noise estimate σ^2

Applications to acoustics - Beamforming

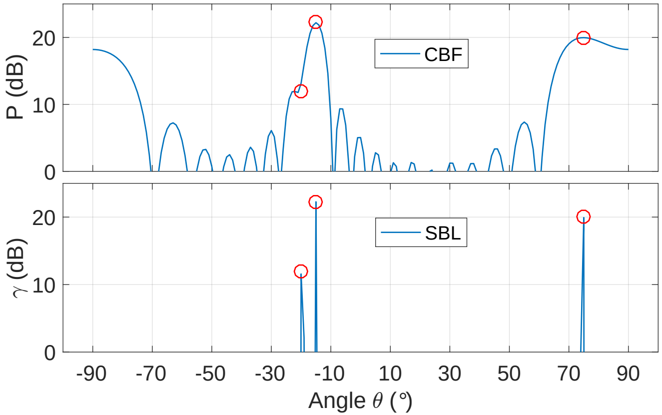
- Beamforming
- Direction of arrivals (DOAs)



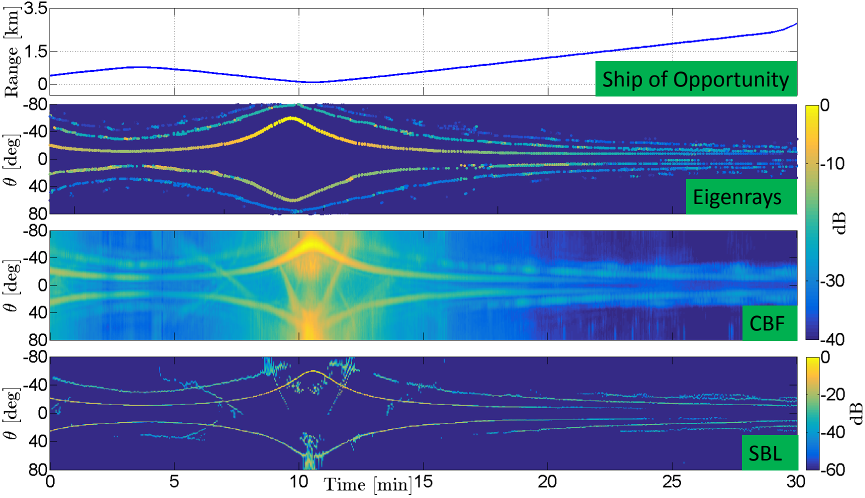


SBL - Beamforming example

- N $= 20$ sensors, uniform linear array
- Discretize angle space: $\{-90:1:90\}$, $M = 181$
- Dictionary $\mathbf{A}$: columns consist of steering vectors
- K $= 3$ sources, DOAs, $[-20, -15, 75]°$, $[12, 22, 20]$ dB
- $M \gg N > K$



SBL - Acoustic hydrophone data processing (from Kai)



Problem with Degrees of Freedom

- As the number of snapshots (=observations) increases, so does the number of unknown complex source amplitudes

- PROBLEM: LASSO for multiple snapshots estimates the realizations of the random complex source amplitudes.

- However, we would be satisfied if we just estimated their power

$$\gamma_m = E\{ |x_{ml}|^2 \}$$

- Note that γ_m does not depend on snapshot index l.

Thus SBL is much faster than LASSO for more snapshots.

Example CPU Time

LASSO use CVX, $\mathrm{CPU} \propto L^2$

SBL nearly independent on snapshots