Distributed Kalman Filter via Gaussian Belief Propagation

Danny Bickson

IBM Haifa Research Lab
Mount Carmel, Haifa 31905, Israel
Email: dannybi@il.ibm.com

9/24/08

Joint work

- Danny Dolev - The Hebrew University, Israel
- Ori Shental - University of San Deigo

- O. Shental acknowledges the partial support of the NSF (Grant CCF-0514859)

Talk outline

$$\begin{pmatrix} -P_{k-1} & A & 0 \\ A^T & Q & H \\ 0 & H^T & R \end{pmatrix} \quad \begin{pmatrix} \Sigma_x & A & 0 \\ A^T & \Sigma_\xi & A\Sigma_{xy} \\ 0 & \Sigma_{yx}A^T & \Sigma_y \end{pmatrix} \quad \begin{pmatrix} 0 & AD & 0 \\ DA^T & I & AD \\ 0 & DA^T & I \end{pmatrix}$$

Outline

Motivation

- Recent work derived a distributed iterative algorithm for computing the MMSE solution for the multiuser detection problem.
- In this work, we extend this construction and investigate what happens when the MMSE solution is applied twice on the matching inputs.
- Using this approach we link several algorithms from the domains of estimation, compression and optimization.
- This allows us to provide an efficient distributed implementation, with numerous applications. For example collaborative signal processing, estimation and resource allocation.

Talk outline

- Formulate Kalman filter algorithm as a matrix inversion problem.
- Show that Kalman filter is a special instance of the GIB algorithm, when the weight parameter $\beta = 1$.
- Show that Affine-scaling interior point method is a special case of Kalman filter.
- Explore the connection two the two step MMSE computation.
- Propose a unified distributed algorithm to compute all of the above methods on a communication network.

Kalman filter

- An efficient iterative algorithm to estimate the state of a discrete-time controlled process $x \in R^n$ that is governed by the linear stochastic difference equation

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1},$$

- A measurement $z \in R^m$ that is $z_k = Hx_k + v_k$.
- The random variables w_k and v_k that represent the process and measurement AWGN noise (respectively). $p(w) \sim \mathcal{N}(0, Q), p(v) \sim \mathcal{N}(0, R)$.
- We further assume that the matrices A, H, B, Q, R are given.
- We assume no external input $\mathbf{u}$.

Kalman Filter

- The prediction step:

$$\begin{aligned} \hat{x}_k^- &= A\hat{x}_{k-1} + Bu_{k-1}\,, \\ P_k^- &= AP_{k-1}A^T + Q. \end{aligned}$$

- The measurement step:

$$\begin{aligned} K_k &= P_k^- H^T (HP_k^- H^T + R)^{-1}\,, \\ \hat{x}_k &= \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-), \\ P_k &= (I - K_k H)P_k^-\,. \end{aligned}$$

  where I is the identity matrix.
- Algorithm operates in rounds.
- The output of the algorithm are the mean vector $\hat{x}_k$ and the covariance matrix P_k.

Kalman Filter as a matrix inversion problem

Theorem: One iteration of the Kalman filter algorithm can be formulated as the following two-step matrix inversion:

$$E = \begin{pmatrix} -P_{k-1} & A & 0 \\ A^T & Q & H \\ 0 & H^T & R \end{pmatrix}$$

Proof sketch

- In the prediction step, given x_k, we compute the MMSE prediction of x_k^-.
- In the measurement step, we compute the MMSE prediction of x_{k+1} given x_k^-, the output of the prediction step.
- Each MMSE computation can be done using the GaBP algorithm. The basic idea, is that given the joint Gaussian distribution $p(\mathbf{x}, \mathbf{y})$ with the covariance matrix $C = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}$, we can compute the MMSE prediction

$$\hat{y} = \arg\max_{y} p(y|x) \propto \mathcal{N}(\mu_{y|x}, \Sigma_{y|x}^{-1})$$

where $\Sigma_{y|x} = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$.

Proof sketch

- Show that P_k^- can be computed by inverting the upper left 2×2 block matrix

$$\begin{pmatrix} X & X & - \\ X & P_k^- & - \\ - & - & - \end{pmatrix}$$

- Show that P_k can be computed by inverting the lower right 2×2 block matrix

$$\begin{pmatrix} - & - & - \\ - & P_k^- & X \\ - & X & X \end{pmatrix}$$

Information Bottleneck

Information Bottleneck

- Given the joint distribution of a source variable X and another relevance variable Y, Information bottleneck (IB) operates to compress X, while preserving information about Y.
- By the following variational problem:

$$\min_{p(t|x)} \mathcal{L} : \mathcal{L} \equiv I(X;T) - \beta I(T;Y)$$

- T represents the compressed representation of X via the conditional distributions $p(t|x)$, while the information that T maintains on Y is captured by the distribution $p(y|t)$.

Gaussian information bottleneck

GIB

- Deals with the special case where the underlying distributions are Gaussian.
- The computed distribution $p(t)$ is Gaussian as well, represented by a linear transformation $T_k = A_k X + \xi_k$ where A_k is a joint covariance matrix of X and T, $\xi_k \sim \mathcal{N}(0, \Sigma_{\xi_k})$ is a multivariate Gaussian independent of X.
- The outputs of the algorithm are the covariance matrices representing the linear transformation T: A_k, Σ_{ξ_k}.

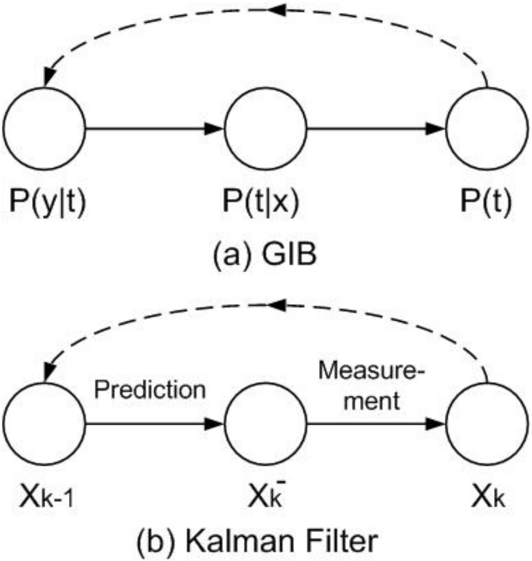
Gaussian information bottleneck

Iterative algorithm

An iterative algorithm is derived by substituting Gaussian distributions into the IB algorithm, resulting in the following update rules:

$$\begin{aligned}\Sigma_{\xi+1} &= (\beta\Sigma_{t_k|y} - (\beta-1)\Sigma_{t_k}^{-1}),\\ A_{k+1} &= \beta\Sigma_{\xi_k+1}\Sigma_{t_k|y}^{-1}A_k(I - \Sigma_{y|x}\Sigma_x^{-1}).\end{aligned}$$

Relation to Kalman filter



The relation between Kalman filter and GIB

- **Theorem**: Regarding the computation of the posterior covariance, one iteration in the GIB algorithm when $\beta = 1$ is equivalent to one iteration of the Kalman filter algorithm.
- **Theorem**: When $\beta > 1$, The posterior covariance in GIB can be computed by a modified Kalman filter iteration.

Some differences

- Kalman filter has input observations z_k while GIB does not. (Observations does not affect the covariance, only the posterior mean $\hat{x}_k$).
- The GIB algorithm computes an additional covariance A_k which is assumed to be known in the Kalman filter.
- Information theoretic interpretation: we extend the ideas of predictive information, the information between the past and the future in a time series.

The Affine-scaling algorithm

LP in its canonical form

$$\begin{aligned} \text{minimize} \quad & \mathbf{c}^T\mathbf{x} && \text{(4a)} \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{b}, \quad \mathbf{x} \geq 0. && \text{(4b)} \end{aligned}$$

where $A \in \mathbb{R}^{n \times p}$ with $\textbf{rank}\{A\} = p < n$. We assume the problem is solvable with an optimal $\mathbf{x}^*$. We also assume that the problem is strictly feasible, in other words there exists $\mathbf{x} \in \mathbb{R}^n$ that satisfies $A\mathbf{x} = \mathbf{b}$ and $\mathbf{x} > 0$.

The Affine-scaling algorithm

Affine-scaling

Let $\mathbf{x}_0$ is an interior feasible point. Let $D = diag(\mathbf{x}_0)$. Iterate:

$$\mathbf{x}_1 = \mathbf{x}_0 - \frac{\alpha}{\gamma} D^2 \mathbf{r}$$

where $0 < \alpha < 1$ is the step size, $\mathbf{r}$ is the step direction.

$$\begin{aligned} \mathbf{r} &= (\mathbf{c} - A^T \mathbf{w}), \\ \mathbf{w} &= (AD^2A^T)^{-1} AD^2 \mathbf{c}, \\ \gamma &= \max_i (\mathbf{e}_i PD\mathbf{c}). \end{aligned}$$

Where $\mathbf{e}_i$ is the i^{th} unit vector and P is a projection matrix given by:

$$P = I - DA^T (AD^2A^T)^{-1} AD.$$

Relation to Kalman filter

- **Theorem** The Affine-scaling algorithm iteration is an instance of the Kalman filter algorithm iteration.
- **Corollary** The Affine-scaling algorithm can be expressed the following matrix inversion problem:

$$\begin{pmatrix} 0 & AD & 0 \\ DA^T & I & AD \\ 0 & DA^T & I \end{pmatrix}.$$

.

Efficient distributed solution using the GaBP algorithm

- We have shown that the discussed algorithms can be formulated as a two-step MMSE computation.
- Following recent work on GaBP, we propose to use it as an iterative algorithm for distributively computing the MMSE solution over a communication network.

The GaBP algorithm

#	Stage	Operation
1.	*Initialize*	Compute $P_{ii} = A_{ii}$ and $\mu_{ii} = b_i/A_{ii}$. Set $P_{ki} = 0$ and $\mu_{ki} = 0$, $\forall k \neq i$.
2.	*Iterate*	Propagate P_{ki} and μ_{ki}, $\forall k \neq i$ such that $A_{ki} \neq 0$. Compute $P_{i\backslash j} = P_{ii} + \sum_{k\in \mathrm{N}(i)\backslash j} P_{ki}$ and $\mu_{i\backslash j} = P_{i\backslash j}^{-1}(P_{ii}\mu_{ii} + \sum_{k\in \mathrm{N}(i)\backslash j} P_{ki}\mu_{ki})$. Compute $P_{ij} = -A_{ij}P_{i\backslash j}^{-1}A_{ji}$ and $\mu_{ij} = -P_{ij}^{-1}A_{ij}\mu_{i\backslash j}$.
3.	*Check*	If P_{ij} and μ_{ij} did not converge, return to #2. Else, continue to #4.
4.	*Infer*	$P_i = P_{ii} + \sum_{k\in \mathrm{N}(i)} P_{ki}$, $\mu_i = P_i^{-1}(P_{ii}\mu_{ii} + \sum_{k\in \mathrm{N}(i)} P_{ki}\mu_{ki})$.
5.	*Output*	$x_i = \mu_i$

GaBP Convergence and Exactness

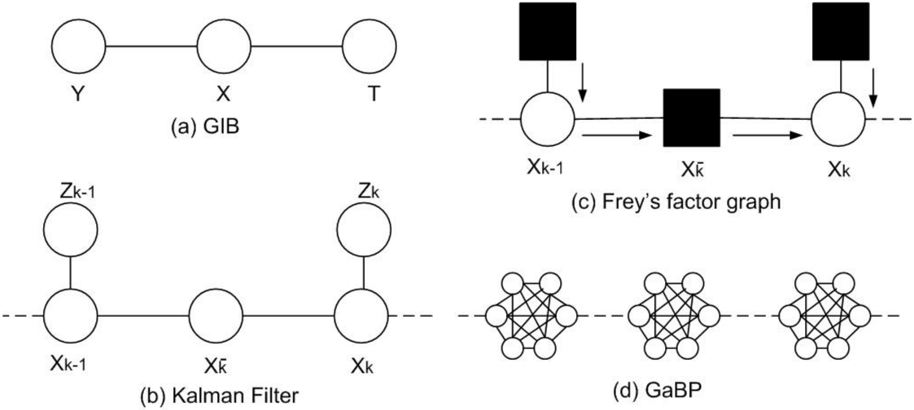
Theorem [Weiss and Freeman,'01,Claim 4]

If the matrix $\mathbf{A}$ is strictly diagonally dominant (*i.e.* , $|A_{ii}| > \sum_{j \neq i} |A_{ij}|, \forall i$), then the GaBP solver converges and the marginal means converge to the true solution.

Theorem [Johnson *et al.* ,'06,Proposition 2]

If the spectral radius (maximum of the absolute values of the eigenvalues) ρ of the matrix $|\mathbf{I}_n - \mathbf{A}|$ satisfies $\rho(|\mathbf{I}_n - \mathbf{A}|) < 1$, then the GaBP solver converges and the marginal means converge to the true solution.

Different graphical models



Conclusion

- We have found a common underlying similarity between the Kalman filter, GIB and Affine-scaling algorithms.
- We propose to use GaBP for distributively and iteratively computing all above methods.
- Future work is to investigate implications of our findings.

THANK YOU!