GRAPH SIGNAL PROCESSING: A STATISTICAL VIEWPOINT

Cha Zhang
Joint work with Dinei Florêncio and Philip A. Chou

Microsoft Research
Outline

• Gaussian Markov Random Field
• Graph construction
• Graph transform
• Graph down-sampling
• Graph prediction
• Graph-based regularization
• Conclusions
Gaussian Markov Random Field (GMRF)

A random vector $\mathbf{x} = (x_1, \cdots, x_n)^T$ is called a GMRF with respect to the graph $G = (V = \{1, \cdots, n\}, E)$ with mean $\mu$ and a precision matrix $\mathbf{Q} > 0$ (positive definite), if and only if its density has the form:

$$p(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} |\mathbf{Q}|^{\frac{1}{2}} \exp \left( -\frac{1}{2} (\mathbf{x} - \mu)^T \mathbf{Q} (\mathbf{x} - \mu) \right)$$

and $Q_{ij} \neq 0 \iff \{i, j\} \in E$ for all $i \neq j$. 
Gaussian Markov Random Field (GMRF)

• It is defined on a graph
• It is a multivariate Gaussian distribution
• Regarding the precision matrix $Q$
  – $Q = \Sigma^{-1}$, $\Sigma$ is the covariance matrix
  – Conditional interpretation
    • $E(x_i|x_{-i}) = \mu_i - \frac{1}{Q_{ii}} \sum_{j:j \sim i} Q_{ij} (x_j - \mu_j)$
    • $Prec(x_i|x_{-i}) = Q_{ii}$
    • $Corr(x_i, x_j|x_{-ij}) = \frac{-Q_{ij}}{\sqrt{Q_{ii}Q_{jj}}}, i \neq j$
A Bijective Mapping

- Zero-mean GMRF $\iff$ a subset of bi-directed graphs
  - Random variable $\iff$ node
  - $Q_{ij} \neq 0 \iff$ edge
  - Assign graph weights:
    - $W_{ij} = -Q_{ij}$, for $i \neq j$
    - $W_{ii} = \sum_{j=1}^{n} Q_{ij}$
  - Mapping between $\mathbf{W}$ and $\mathbf{Q}$ is bijective
  - Subset because $\mathbf{Q} > 0$

$$
\mathbf{Q} = 
\begin{bmatrix}
Q_{11} & Q_{12} & Q_{13} \\
Q_{21} & Q_{22} & Q_{23} \\
Q_{31} & Q_{32} & Q_{33}
\end{bmatrix}
$$
Intrinsic GMRF (IGMRF)

A random vector $\mathbf{x} = (x_1, \cdots, x_n)^T$ is called an intrinsic GMRF of $k^{th}$ order if it has density:

$$p(\mathbf{x}) = (2\pi)^{-\frac{n-k}{2}} (|\mathbf{Q}|^*)^{\frac{1}{2}} \exp \left(-\frac{1}{2} (\mathbf{x} - \mathbf{\mu})^T \mathbf{Q} (\mathbf{x} - \mathbf{\mu}) \right)$$

Where $\mathbf{Q}$ is of rank $n - k$, and $|\cdot|^*$ denotes the generalized determinant (product of the non-zero eigenvalues).

IGMRF can be interpreted as limits of proper GMRFs, thus many of the properties of GMRF holds for IGMRF.

A Special GMRF

• Zero mean, and all rows of $Q$ sum to zero
  – For all $i$, $\Sigma_j Q_{ij} = 0$

• When $Q$ is of rank $n - 1$: IGMRF of first order

• No self-loops in the graph $\rightarrow$ Simple Graph

• Local behavior

$$E(x_i|x_{-i}) = -\frac{1}{Q_{ii}} \Sigma_{j:j\sim i} Q_{ij} x_j, \text{ with } -\frac{\Sigma_{j:j\sim i} Q_{ij}}{Q_{ii}} = 1$$

• IGMRF of first order closely related to first-order random walk
So...

- For many bi-directed graphs, we can give them a GMRF (IGMRF) interpretation
- It’s a Gaussian, isn’t it too simple?
  - Real-world signals are hardly Gaussian...
  - What can we do with such a simple model?
So...

• For many bi-directed graphs, we can give them a GMRF (IGMRF) interpretation

• It’s a Gaussian, isn’t it too simple?
  – Real-world signals are hardly Gaussian...
  – What can we do with such a simple model?
    • A lot!
      • Many are already using it without realizing it...
Graph Construction

• Data-driven
  – Collect enough examples to estimate $\mu$ and $Q$
  – $Q$ is unlikely sparse $\Rightarrow$ densely connected graph

• Intuitive model-based
  – Widely used in 1D/2D regular grid signals
  – IGMRF, no self-loops: nice local behavior
  – $Q$ identical to the Laplacian matrix
Graph Construction

• Model-constrained data-driven approach
• Maximum a-posterior estimation
  \[ \hat{Q} = \arg \max_Q p(Q | x_1, \ldots, x_N) \]
  \[ = \arg \max_Q p(Q) \prod_{n=1}^{N} p(x_n | Q) \]
• \( p(x_n | Q) \): likelihood, GMRF model
• \( p(Q) \): prior knowledge
  – Application dependent
  – For example, \( Q \) is sparse, know graph structure, etc.

Graph Transform

• Karhunen-Loève transform (KLT)
  \[ \Sigma = \Phi \Lambda \Phi^T \]
  
  Where \( \Lambda = \text{diag}(\lambda_0, \cdots, \lambda_n) \) is the diagonal matrix of eigenvalues of \( \Sigma \).

• Since:
  \[ Q = \Sigma^{-1} = \Phi \Lambda^{-1} \Phi^T \]
  
  The eigenvector matrix of \( Q \) is the same as the KLT, optimal for decorrelation.

• Since GMRF is defined on a graph, we call the eigenvector matrix of the precision matrix \( Q \) the Graph Transform.
Graph Transform and DCT

- When $Q$ is the Laplacian of a regular grid graph
  - 1D signal $\Rightarrow$ Graph transform = 1D DCT
  - 2D signal on regular grid $\Rightarrow$ 2D DCT is **one** of the eigenvector matrices of $Q$
  - Thus under such an image model, 1D DCT and 2D DCT are optimal for decorrelation
  - **First known proof of 2D DCT optimality!**

Graph Down-sampling

Given random vector $\mathbf{x} = (x_1, \ldots, x_n)^T$ on graph $G = (V = \{1, \ldots, n\}, E)$, keep $\mathbf{x}_1 = (x_1, \ldots, x_{n-k})^T$, remove $\mathbf{x}_2 = (x_{n-k+1}, \ldots, x_n)^T$, what is the graph for $\mathbf{x}_1$?
Graph Down-sampling

Given GMRF random vector \( \mathbf{x} = (x_1, \cdots, x_n)^T \) on graph \( G = (V = \{1, \cdots, n\}, E) \), keep \( x_1 = (x_1, \cdots, x_{n-k})^T \), remove \( x_2 = (x_{n-k+1}, \cdots, x_n)^T \), what is the graph for \( x_1 \)?

\[
\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = Q^{-1} = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}^{-1}
\]

\[
\rightarrow Q_1 = \Sigma_{11}^{-1} = Q_{11} - Q_{12} Q_{22}^{-1} Q_{21}
\]

Identical to *Kron Reduction*, a well-know technique for graph down-sampling originally derived for electronic networks.

Graph Prediction

- Random vector $\mathbf{x} = (x_1, \cdots, x_n, x_{n+1}, \cdots, x_m)^T$ follows GMRF with precision matrix $\mathbf{Q}$.
  - $x_1 = (x_1, \cdots, x_n)^T$ unknown
  - $x_2 = (x_{n+1}, \cdots, x_m)^T$ known
Graph Prediction

- Random vector \( \mathbf{x} = (x_1, \ldots, x_n, x_{n+1}, \ldots, x_m)^T \) follows GMRF with precision matrix \( \mathbf{Q} \).
  - \( x_1 = (x_1, \ldots, x_n)^T \) unknown
  - \( x_2 = (x_{n+1}, \ldots, x_m)^T \) known
  - Partition \( \mathbf{Q} \):
    \[
    \mathbf{Q} = \begin{bmatrix}
    Q_{11} & Q_{12} \\
    Q_{21} & Q_{22}
    \end{bmatrix}
    \]
- \( x_1|x_2 \) is also a GMRF with respect to its own subgraph with mean \( \mu_{x_1|x_2} \) and precision matrix \( Q_{x_1|x_2} > 0 \):
  \[
  \begin{align*}
  \mu_{x_1|x_2} &= \mu_{x_1} - Q_{11}^{-1} Q_{12} (x_2 - \mu_{x_2}) \\
  Q_{x_1|x_2} &= Q_{11}
  \end{align*}
  \]

Predictive Transform Coding

• In the above formulation:
  – $x_1$: pixels to be encoded
  – $x_2$: know pixels nearby

• New conditional mean: best prediction
  \[ \mu_{x_1|x_2} = \mu_{x_1} - Q_{11}^{-1}Q_{12}(x_2 - \mu_{x_2}) \]

• New best decorrelation transform after applying best prediction:
  – Eigenvector matrix of $Q_{x_1|x_2} = Q_{11}$
Motion Prediction

• Reference block and current block, both zero mean GMRF, with $Q_{\text{ref}}$ and $Q_c$
• Form a 3D GMRF model
• Precision matrix:

$$Q = \begin{bmatrix} Q_c + I & -I \\ -I & Q_{\text{ref}} + I \end{bmatrix}$$

• Regarding $x_1|x_2$:

$$\mu_{x_1|x_2} = -(Q_c + I)^{-1} x_2$$
$$Q_{x_1|x_2} = Q_c + I$$
Motion Prediction

• When \( Q_c \) is the Laplacian matrix \( Q_c = \delta L \)
  
  – New mean: \( \mu_{x_1|x_2} = -(\delta L + I)^{-1} x_2 \)
    
    • Low pass filter on \( x_2 \)
    
    • (Partial) reason that half-pixel prediction is better!
  
  – New precision matrix: \( Q_{x_1|x_2} = \delta L + I \)
    
    • Its eigenvector matrix is same as the eigenvector matrix of \( L \)
    
    • First known proof that DCT is still optimal for residue compression in motion compensation!
Intra-Frame Predictive Coding

- Zero mean GMRF
- New mean: \( \mu_{x_1|x_2} = -Q^{-1}_{11} Q_{12} x_2 \)
  - Optimal prediction given the GMRF model
- New precision matrix: \( Q_{x_1|x_2} = Q_{11} \)
  - DCT no longer optimal!
  - First general analysis of the optimal transform for intra-frame predictive coding

Graph-Based Regularization

- Estimate a signal $x$ on graph, such that a cost function is minimized, subject to certain regularization

$$\hat{x} = \arg \min_x f(x) + \lambda S_p(x)$$

A popular smoothness term (p-Dirichlet form):

$$S_p(x) := \frac{1}{p} \left[ \sum_i \sum_{j:j \sim i} W_{ij} (x_i - x_j)^2 \right]^{\frac{p}{2}}$$
A Prob. Viewpoint on Regularization

• Target function:

\[ \hat{x} = \arg\min_x f(x) + \lambda S_p(x) \]
\[ = \arg\max_x e^{-\frac{f(x)}{\lambda} - S_p(x)} \]

• Mapping to MAP:

  – Likelihood: \( e^{-\frac{f(x)}{\lambda}} \)
  – Prior: \( e^{-S_p(x)} \)

    • When \( p = 2 \) → Laplacian GMRF prior
    • When \( p \neq 2 \) → Generalized GMRF prior
Conclusions

• GMRF provides a statistical viewpoint for GSP
• Provides valuable insights for
  – Graph construction
  – Graph transform
  – Graph down-sampling
  – Graph prediction
  – Graph-based regularization
• Always be aware of the statistical implications!