Journal Watch TSP-Aug 15 2017

Akshay Kumar

November 18, 2017

Table of Contents

Fast Algorithms for Demixing Sparse Signals From Nonlinear Observations

- Problem statement

$$y_i = g(< a_i, \Phi w + \Psi z >) + e_i \qquad i = 1, 2, \ldots m \tag{1}$$

where,

- $x = \Phi w + \Psi z$
- $\Phi^{n \times n}$ and $\Psi^{n \times n}$ are orthornormal bases
- a_i is ith row of $A^{m \times n}$ (measurement operator)
- g (link function) is either known or unknown non-linear function

Algorithms

- When g is unknown

Algorithm 1: OneShot.	
Inputs: Φ, Ψ, A, y, s.	
Outputs: Estimates $\widehat{x} = \Phi\widehat{w} + \Psi\widehat{z}$, $\widehat{w} \in K_1$, $\widehat{z} \in K_2$	
$\widehat{x}_{\text{lin}} \leftarrow \frac{1}{m} A^T y$	{form linear estimator}
$b_1 \leftarrow \Phi^* \widehat{x}_{\text{lin}}$	{forming first proxy}
$\widehat{w} \leftarrow \mathcal{P}_s(b_1)$	{sparse projection}
$b_2 \leftarrow \Psi^* \widehat{x}_{\text{lin}}$	{forming second proxy}
$\widehat{z} \leftarrow \mathcal{P}_s(b_2)$	{sparse projection}
$\widehat{x} \leftarrow \Phi\widehat{w} + \Psi\widehat{z}$	{Estimating $\widehat{x}$}

where,

- $\mathcal{P}_s$ is s-sparse projection

Remarks-:

Algorithms

- When g is known
- Let $\Theta^{'}(x) = g(x)$
- $\Gamma = [\Phi \quad \Psi]$, $t = [w \quad z]$

$$\underset{t \in \mathbb{R}^{2n}}{\text{minimize}} \quad F(t) = \frac{1}{m} \sum_{i=1}^{m} (\Theta(a_i^T \Gamma t) - y_i a_i^t \Gamma t)$$

$$\text{subject to} \quad ||t||_0 \leq 2s$$

- Gradient, $\nabla F(t) = \frac{1}{m} \Gamma^T A^T (g(A\Gamma t) - y)$

Algorithm 2: Demixing with Hard Thresholding (DHT).

Inputs: Φ, Ψ, A, g, y, s, η'.
Outputs: Estimates $\widehat{x} = \Phi\widehat{w} + \Psi\widehat{z}$, $\widehat{w}$, $\widehat{z}$
Initialization:
$\left(x^0, w^0, z^0\right) \leftarrow$ ARBITRARY, $k \leftarrow 0$
while $k \leq N$ **do**
  $t^k \leftarrow [w^k; z^k]$ {forming constituent vector}
  $t_1^k \leftarrow \frac{1}{m}\Phi^T A^T (g(Ax^k) - y)$
  $t_2^k \leftarrow \frac{1}{m}\Psi^T A^T (g(Ax^k) - y)$
  $\nabla F^k \leftarrow [t_1^k; t_2^k]$ {forming gradient}
  $\tilde{t}^k = t^k - \eta'\nabla F^k$ {gradient update}
  $[w^k; z^k] \leftarrow \mathcal{P}_{2s}\left(\tilde{t}^k\right)$ {sparse projection}
  $x^k \leftarrow \Phi w^k + \Psi z^k$ {estimating $\widehat{x}$}
  $k \leftarrow k + 1$
end while
Return: $(\widehat{w}, \widehat{z}) \leftarrow \left(w^N, z^N\right)$

Scalable and Flexible Multiview MAX-VAR Canonical Correlation Analysis

- Problem statement- Finding low-dimensional representations from multiple views corresponding to the same entities, termed as Canonical Correlation Analysis (CCA)
- Consider the word 'Akshay'. It has text and audio representation
- A *view* is a high dimensional representation of an entity in some feature space
- Helpful in data fusion. Integrating information acquired from different sources

Mathematical Formulation

- Consider L entities have different representation in I views
- $\mathbf{X}_i \in \mathbb{R}^{L \times M_i}$ is the feature matrix for L entities in ith view

$$\begin{aligned} \underset{\{\mathbf{Q}_i\}_{i=1}^{I}, \mathbf{G}}{\text{minimize}} \quad & \sum_{i=1}^{I} ||\mathbf{X}_i \mathbf{Q}_i - \mathbf{G}||_F^2 \\ \text{subject to} \quad & \mathbf{G}^T \mathbf{G} = \mathbb{I} \end{aligned}$$

where, $G \in \mathbb{R}^{L \times K}$, $K (\ll min(M_i, L))$ is number of canonical components

- The above problem has closed form expression
- Solving wrt $\mathbf{Q}_i$, $\mathbf{Q}_i = \mathbf{X}_i^{\dagger} \mathbf{G}$, $\mathbf{X}_i^{\dagger} = (\mathbf{X}_i^T \mathbf{X}_i)^{-1} \mathbf{X}_i^T$
- Substituting back, estimating $\mathbf{G}$ reduces to

$$\underset{\mathbf{G}^T \mathbf{G} = \mathbb{I}}{\text{maximize}} \quad Tr \left(\mathbf{G}^T \left(\sum_{i=1}^{I} \mathbf{X}_i \mathbf{X}_i^{\dagger} \right) \mathbf{G} \right)$$

- Solution $\rightarrow$ First K principal eigenvectors of $\sum_{i=1}^{I} \mathbf{X}_i \mathbf{X}_i^{\dagger}$

Major challenges -:

- Implementing the solution to large-scale data
- Incorporating structure in $\mathbf{Q}_i$

To circumvent the second issue, add regularizer $h_i(\mathbf{Q}_i)$

- $h_i(\mathbf{Q}_i) = \frac{\mu_i}{2} \cdot ||\mathbf{Q}_i||_F^2$
- $h_i(\mathbf{Q}_i) = \frac{\mu_i}{2} \cdot ||\mathbf{Q}_i||_{2,1}$
- $h_i(\mathbf{Q}_i) = \frac{\mu_i}{2} \cdot ||\mathbf{Q}_i||_F^2 + \beta_i \cdot ||\mathbf{Q}_i||_{2,1}$
- $h_i(\mathbf{Q}_i) = \mathbf{1}_+(\mathbf{Q}_i)$

The reulting objective,

$$\begin{aligned} \underset{\{\mathbf{Q}_i\}_{i=1}^I, \mathbf{G}}{\text{minimize}} \quad & \sum_{i=1}^{I} ||\mathbf{X}_i\mathbf{Q}_i - \mathbf{G}||_F^2 + h_i(\mathbf{Q}_i) \\ \text{subject to} \quad & \mathbf{G}^T\mathbf{G} = \mathbb{I} \end{aligned}$$

The authors use alternating optimization; solve two subproblems wrt $\{\mathbf{Q}_i\}$ and $\mathbf{G}$

After r iterations we have $\mathbf{Q}^{(r)}, \mathbf{G}^{(r)}$

- Solving for $\mathbf{Q}_i^{(r+1)}$

$$\underset{\mathbf{Q}_i}{\text{minimize}} \quad ||\mathbf{X}_i\mathbf{Q}_i - \mathbf{G}^{(r)}||_F^2 + h_i(\mathbf{Q}_i)$$

- Rewritten as,

$$\underset{\mathbf{Q}_i}{\text{minimize}} \quad f_i(\mathbf{Q}_i, \mathbf{G}^{(r)}) + g_i(\mathbf{Q}_i)$$

where, f_i is continuously differentiable part and g_i is non-smooth part of the objective

- Use proximal gradient to solve and get $\mathbf{Q}_i^{(r+1)}$

- Solving for $\mathbf{G}^{(r+1)}$

$$\begin{aligned} \underset{\mathbf{G}}{\text{minimize}} \quad & \sum_{i=1}^{I} ||\mathbf{X}_i \mathbf{Q}_i^{(r+1)} - \mathbf{G}||_F^2 \\ \text{subject to} \quad & \mathbf{G}^T \mathbf{G} = \mathbb{I} \end{aligned}$$

- Rewritten as,

$$\underset{\mathbf{G}^T\mathbf{G}=\mathbb{I}}{\text{maximize}} \quad Tr\left(\mathbf{G}^T \left(\sum_{i=1}^{I} \mathbf{X}_i \mathbf{Q}_i^{(r+1)}\right)\right)$$

- Closed form update can be found using Procrustes projection
- $\mathbf{G}^{(r+1)} \rightarrow UV^T$, where, $[U, :, V] = SVD\left(\sum_{i=1}^{I} \mathbf{X}_i \mathbf{Q}_i^{(r+1)}\right)$, $\mathcal{O}(LK^2)$

Adaptive Subspace Signal Detection with Uncertain Partial Prior Knowledge

- Consider the hypothesis testing problem,

$$H_0 : \mathbf{y} = \mathbf{d}$$
$$H_1 : \mathbf{y} = \kappa\mathbf{s} + \mathbf{d}$$

where,

- $\mathbf{y} \in \mathbb{R}^n$ is test data, $\mathbf{s}$ is known signal with unknown amplitude κ
- $\mathbf{d}$ is the disturbance signal with low-rank subspace representation

$$\mathbf{d} = \mathbf{H}\beta + \mathbf{n}$$

- $\mathbf{H} \in \mathbb{R}^{N\times L}$ consists of $L(< N)$ independent basis vectors, $\mathbf{n}$ follows $\mathcal{N}(\mathbf{0}, \sigma^2\mathbb{I})$
- $\mathbf{s} \notin \text{span}(\mathbf{H})$

- Authors consider the case where **H** is partially known, i.e.,

$$\mathbf{H}\beta = \mathcal{H}\mathbf{x}$$

where, $\mathcal{H} \in \mathbb{R}^{N \times M}$ is an overcomplete dictionary, **x** is a sparse vector with sparsity L

- We know which columns of $\mathcal{H}$ spans the column space of **H**
- But, that information is not completely accurate
- May contain erroneous columns or may miss some columns

- The likelihood functions under the H_0 and H_1 hypotheses given observation **y** are

$$p_0(\beta, \mathbf{H}, \sigma^2; \mathbf{y}) = \mathcal{N}(\mathbf{y}; \mathbf{H}\beta, \sigma^2\mathbb{I})$$
$$p_1(\kappa, \beta, \mathbf{H}, \sigma^2; \mathbf{y}) = \mathcal{N}(\mathbf{y}; \kappa\mathbf{s} + \mathbf{H}\beta, \sigma^2\mathbb{I})$$

- Under H_1, MLE of κ conditioned on $\mathbf{H}$, β is,

$$\hat{\kappa} = \frac{\mathbf{s}^H(\mathbf{y} - \mathbf{H}\beta)}{\mathbf{s}^H\mathbf{s}}$$

- Substituting it in p_1 then MLE of noise variance under H_1 is,

$$\sigma_1^2 = \frac{1}{N}||\mathbf{P}_s^{\perp}\mathbf{y} - \mathbf{P}_s^{\perp}\mathbf{H}\beta||^2$$

where, $\mathbf{P}_s^{\perp} = \mathbb{I} - \mathbf{s}(\mathbf{s}^H\mathbf{s})^{-1}\mathbf{s}^H$

- MLEs of $\mathbf{H}, \beta$ under H_1 ,

$$\{\hat{\mathbf{H}}_\mathbf{1}, \hat{\beta}\} = \text{arg min}||\mathbf{P}_s^{\perp}\mathbf{y} - \mathbf{P}_s^{\perp}\mathbf{H}\beta||^2$$

Under H_0,

$$\sigma_0^2 = \frac{1}{N}||\mathbf{y} - \mathbf{H}\beta||^2$$

$$\{\hat{\mathbf{H}}_\mathbf{0}, \hat{\beta}\} = \text{arg min}||\mathbf{y} - \mathbf{H}\beta||^2$$

We need to solve sparse recovery problem,

$$min_x ||\mathbf{z} - \mathbf{Ax}||^2$$

where, $\mathbf{z} = \mathbf{P}_s^{\perp}\mathbf{y}$ and $\mathbf{A} = \mathbf{P}_s^{\perp}\mathcal{H}$ under H_1 and $\mathbf{z} = \mathbf{y}$ and $\mathbf{A} = \mathcal{H}$ under H_0

- Once $\{\mathbf{H_1}, \beta\}$ and $\{\mathbf{H_0}, \beta\}$ are obtained we can substitute them back to obtain κ and σ^2 .
- Then perform GLRT