Random Thoughts on Fast Inference with Gaussian Likelihoods

Landman Bester

Aug 2019

Dimensionality Reduction vs. Data Compression

Dimensionality Reduction $\Rightarrow$ Embed into $\tilde{d}$ s.t.

- Fast mapping between parameters and $\tilde{\boldsymbol{d}}$
- Covariance of $\tilde{\boldsymbol{d}}$ should be close to diagonal
- Approximations can lead to loss of information
- If parameter space larger than data space might be more relveant to apply embedding to parameter space

Data Compression $\Rightarrow$ Embed s.t. $\tilde{d} << d$ and:

- As little as possible information is lost
- Secure format for storage and decompression

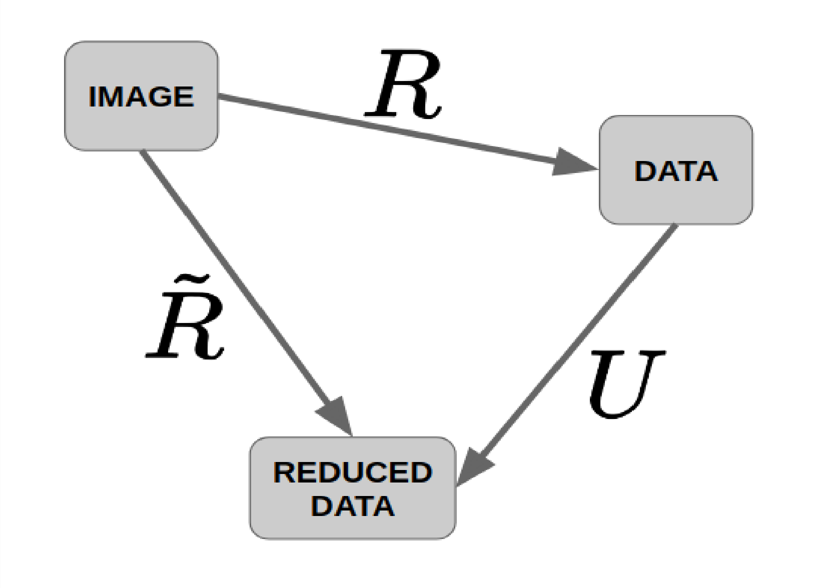
IMAGE
R
DATA
$\tilde{R}$
U
REDUCED
DATA

Fourier Dimensionality Reduction

Measurement model

$$V = RI + \epsilon, \quad \text{where} \quad \epsilon \sim \mathcal{N}\left(\mathbf{0}, \Sigma\right)$$

Embedding operator

$$U = \mathcal{F} Z \Psi^{-1} R^{\dagger} \Sigma^{-1}$$

Response implemented as

$$R = G \mathcal{F} Z \Psi$$

Dimensionally reduced problem

$$\tilde{V} = \tilde{R} I + \tilde{\epsilon}, \quad \text{where} \quad \begin{cases} \tilde{V} = G^{\dagger} \Sigma^{-1} V \\ \tilde{R} = G^{\dagger} \Sigma^{-1} G \mathcal{F} Z \Psi \\ \tilde{\Sigma} = \langle \tilde{\epsilon}, \tilde{\epsilon}^{\dagger} \rangle = G^{\dagger} \Sigma^{-1} G \end{cases}$$

Fourier Dimensionality Reduction

Measurement model

$$V = RI + \epsilon, \quad \text{where} \quad \epsilon \sim \mathcal{N}\left(\mathbf{0}, \Sigma\right)$$

Natural Weighting

Embedding operator

$$U = \mathcal{F} Z \Psi^{-1} R^{\dagger} \Sigma^{-1}$$

Response implemented as

$$R = G \mathcal{F} Z \Psi$$

Dimensionally reduced problem

$$\tilde{V} = \tilde{R} I + \tilde{\epsilon}, \quad \text{where} \quad \begin{cases} \tilde{V} = G^{\dagger} \Sigma^{-1} V \\ \tilde{R} = G^{\dagger} \Sigma^{-1} G \mathcal{F} Z \Psi \\ \tilde{\Sigma} = \langle \tilde{\epsilon}, \tilde{\epsilon}^{\dagger} \rangle = G^{\dagger} \Sigma^{-1} G \end{cases}$$

Fourier Dimensionality Reduction

Measurement model

$$V = RI + \epsilon, \quad \text{where} \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \Sigma)$$

Embedding operator

Grid Corrector

$$U = \mathcal{F}Z\Psi^{-1}R^{\dagger}\Sigma^{-1}$$

Response implemented as

$$R = G\mathcal{F}Z\Psi$$

Dimensionally reduced problem

$$\tilde{V} = \tilde{R}I + \tilde{\epsilon}, \quad \text{where} \quad \begin{cases} \tilde{V} = G^{\dagger}\Sigma^{-1}V \\ \tilde{R} = G^{\dagger}\Sigma^{-1}G\mathcal{F}Z\Psi \\ \tilde{\Sigma} = \langle \tilde{\epsilon}, \tilde{\epsilon}^{\dagger} \rangle = G^{\dagger}\Sigma^{-1}G \end{cases}$$

Fourier Dimensionality Reduction

Measurement model

$$V = RI + \epsilon, \quad \text{where} \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \Sigma)$$

Embedding operator

Zero Padding

$$U = \mathcal{F} Z \Psi^{-1} R^{\dagger} \Sigma^{-1}$$

Response implemented as

$$R = G \mathcal{F} Z \Psi$$

Dimensionally reduced problem

$$\tilde{V} = \tilde{R} I + \tilde{\epsilon}, \quad \text{where} \quad \begin{cases} \tilde{V} = G^{\dagger} \Sigma^{-1} V \\ \tilde{R} = G^{\dagger} \Sigma^{-1} G \mathcal{F} Z \Psi \\ \tilde{\Sigma} = \langle \tilde{\epsilon}, \tilde{\epsilon}^{\dagger} \rangle = G^{\dagger} \Sigma^{-1} G \end{cases}$$

Fourier Dimensionality Reduction

Measurement model

$$V = RI + \epsilon, \quad \text{where} \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \Sigma)$$

Embedding operator

FFT

$$U = \mathcal{F}Z\Psi^{-1}R^{\dagger}\Sigma^{-1}$$

Response implemented as

$$R = G\mathcal{F}Z\Psi$$

Dimensionally reduced problem

$$\tilde{V} = \tilde{R}I + \tilde{\epsilon}, \quad \text{where} \quad \begin{cases} \tilde{V} = G^{\dagger}\Sigma^{-1}V \\ \tilde{R} = G^{\dagger}\Sigma^{-1}G\mathcal{F}Z\Psi \\ \tilde{\Sigma} = \langle \tilde{\epsilon}, \tilde{\epsilon}^{\dagger} \rangle = G^{\dagger}\Sigma^{-1}G \end{cases}$$

Fourier Dimensionality Reduction

Measurement model

$$V = RI + \epsilon, \quad \text{where} \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \Sigma)$$

Degridding Operator

Embedding operator

$$U = \mathcal{F} Z \Psi^{-1} R^{\dagger} \Sigma^{-1}$$

Response implemented as

$$R = G \mathcal{F} Z \Psi$$

Dimensionally reduced problem

$$\tilde{V} = \tilde{R} I + \tilde{\epsilon}, \quad \text{where} \quad \begin{cases} \tilde{V} = G^{\dagger} \Sigma^{-1} V \\ \tilde{R} = G^{\dagger} \Sigma^{-1} G \mathcal{F} Z \Psi \\ \tilde{\Sigma} = \langle \tilde{\epsilon}, \tilde{\epsilon}^{\dagger} \rangle = G^{\dagger} \Sigma^{-1} G \end{cases}$$

Fourier Dimensionality Reduction

Measurement model

$$V = RI + \epsilon, \quad \text{where} \quad \epsilon \sim \mathcal{N}\left(\mathbf{0}, \Sigma\right)$$

Embedding operator

$$U = \mathcal{F} Z \Psi^{-1} R^{\dagger} \Sigma^{-1}$$

Response implemented as

$$R = G \mathcal{F} Z \Psi$$

Dimensionally reduced problem

$$\tilde{V} = \tilde{R} I + \tilde{\epsilon}, \quad \text{where} \quad \begin{cases} \tilde{V} = G^{\dagger} \Sigma^{-1} V \\ \tilde{R} = G^{\dagger} \Sigma^{-1} G \mathcal{F} Z \Psi \\ \tilde{\Sigma} = \langle \tilde{\epsilon}, \tilde{\epsilon}^{\dagger} \rangle = G^{\dagger} \Sigma^{-1} G \end{cases}$$

Minor Cycle

Fundamental approximation

$$R^\dagger \Sigma^{-1} R \approx I^{PSF} *$$

Minor Cycle

Fundamental approximation

$$R^{\dagger}\Sigma^{-1}R \approx I^{PSF} *$$

- Inputs are I^D, I^{PSF} and Ψ

Minor Cycle

Fundamental approximation

$$R^\dagger \Sigma^{-1} R \approx I^{PSF} *$$

- Inputs are I^D, I^{PSF} and Ψ
- Notion of locality (partially) preserved in image space

Minor Cycle

Fundamental approximation

$$R^\dagger \Sigma^{-1} R \approx I^{PSF} *$$

- Inputs are I^D, I^{PSF} and Ψ
- Notion of locality (partially) preserved in image space
- Deconvolve different parts of the image with different algorithms/priors

Minor Cycle

Fundamental approximation

$$R^{\dagger}\Sigma^{-1}R \approx I^{PSF}*$$

- Inputs are I^D, I^{PSF} and Ψ
- Notion of locality (partially) preserved in image space
- Deconvolve different parts of the image with different algorithms/priors
- Sub-Space Deconvolution (SSD) in DDFacet illustrates proof of concept

Minor Cycle

Fundamental approximation

$$R^\dagger \Sigma^{-1} R \approx I^{PSF} *$$

- Inputs are I^D, I^{PSF} and Ψ
- Notion of locality (partially) preserved in image space
- Deconvolve different parts of the image with different algorithms/priors
- Sub-Space Deconvolution (SSD) in DDFacet illustrates proof of concept
- Variable resolution in different sub-spaces

Minor Cycle

Fundamental approximation

$$R^\dagger \Sigma^{-1} R \approx I^{PSF} *$$

- Inputs are I^D, I^{PSF} and Ψ
- Notion of locality (partially) preserved in image space
- Deconvolve different parts of the image with different algorithms/priors
- Sub-Space Deconvolution (SSD) in DDFacet illustrates proof of concept
- Variable resolution in different sub-spaces
- Combine with "beamforming"?

Minor Cycle

Fundamental approximation

$$R^{\dagger}\Sigma^{-1}R \approx I^{PSF} *$$

- Inputs are I^D, I^{PSF} and Ψ
- Notion of locality (partially) preserved in image space
- Deconvolve different parts of the image with different algorithms/priors
- Sub-Space Deconvolution (SSD) in DDFacet illustrates proof of concept
- Variable resolution in different sub-spaces
- Combine with "beamforming"?
- How to represent fundamental assumption in $\tilde{\boldsymbol{d}}$ space?

Deconvolving LMC

Assumption:

$$\text{diag}(G^\dagger \Sigma^{-1} G) = \mathcal{F} Z I^{PSF}$$

- Direction independent phase only calibration applied
- Input - $8k \times 8k$ MFS Stokes I dirty and PSF (missing Ψ in this experiment)
- No spectral model
- Single major cycle

Likelihood Considerations for Calibration

Measurement model

$$\mathbf{V}_{pq} = \sum_{s} \mathbf{G}_{p,s}\mathbf{X}_{pq,s}\mathbf{G}_{q,s} + \boldsymbol{\epsilon}_{\boldsymbol{pq}}, \quad \mathrm{vec}(\boldsymbol{\epsilon}_{\boldsymbol{pq}}) \sim \mathcal{N}\left(\mathbf{0}, \Sigma_{pq}\right)$$

Assumptions:

- Gains are independent accross antenna and correlation
- Can neglect Fisher Info arrising from direction axis
  $\Rightarrow$ The cheese ($\boldsymbol{J}^{\dagger}\Sigma^{-1}\boldsymbol{J}$) is (block) diagonal!
- Cubical = the pudding

Likelihood Considerations for Calibration

Measurement model

$$\mathbf{V}_{pq} = \sum_s \mathbf{G}_{p,s}\mathbf{X}_{pq,s}\mathbf{G}_{q,s} + \boldsymbol{\epsilon_{pq}}, \quad \mathrm{vec}(\boldsymbol{\epsilon_{pq}}) \sim \mathcal{N}\left(\mathbf{0}, \Sigma_{pq}\right)$$

Maximum Likelihood Jacobian

Assumptions:

- Gains are independent accross antenna and correlation
- Can neglect Fisher Info arrising from direction axis
  $\Rightarrow$ The cheese $(\boldsymbol{J}^\dagger\Sigma^{-1}\boldsymbol{J})$ is (block) diagonal!
- Cubical = the pudding

Likelihood Considerations for Calibration

Measurement model

$$\mathbf{V}_{pq} = \sum_s \mathbf{G}_{p,s}\mathbf{X}_{pq,s}\mathbf{G}_{q,s} + \boldsymbol{\epsilon_{pq}}, \quad \mathrm{vec}(\boldsymbol{\epsilon_{pq}}) \sim \mathcal{N}\left(\mathbf{0}, \Sigma_{pq}\right)$$

Assumptions:

- Gains are independent accross antenna and correlation
- Can neglect Fisher Info arrising from direction axis
  $\Rightarrow$ The cheese ($\boldsymbol{J}^\dagger \Sigma^{-1} \boldsymbol{J}$) is (block) diagonal!
- Cubical = the pudding

Likelihood Considerations for Calibration

Measurement model

$$\mathbf{V}_{pq} = \sum_s \mathbf{G}_{p,s}\mathbf{X}_{pq,s}\mathbf{G}_{q,s} + \boldsymbol{\epsilon_{pq}}, \quad \mathrm{vec}(\boldsymbol{\epsilon_{pq}}) \sim \mathcal{N}\left(\mathbf{0}, \Sigma_{pq}\right)$$

Assumptions:

- Gains are independent accross antenna and correlation
- Can neglect Fisher Info arrising from direction axis
  $\Rightarrow$ The cheese ($\boldsymbol{J}^\dagger\Sigma^{-1}\boldsymbol{J}$) is (block) diagonal!
- Cubical = the pudding

- If phase-only $\boldsymbol{J}^\dagger\Sigma^{-1}\boldsymbol{J}$ only depends on Sky Model
  $\Rightarrow$ only need to compute it once (for calibration)

MGVI Considerations

- Metric operator

$$\Xi^{-1} = \left(A^\dagger \boldsymbol{J}^\dagger \Sigma^{-1} \boldsymbol{J} A + I \right)$$

- Gradient term

$$A^\dagger \boldsymbol{J}^\dagger \Sigma^{-1} (\mathbf{V}_{pq} - \sum_s \mathbf{G}_{p,s} \mathbf{X}_{pq,s} \mathbf{G}_{q,s})$$

- Back-projected noise

$$A^\dagger \boldsymbol{J}^\dagger \Sigma^{-1} \xi, \quad \text{where} \quad \xi \sim \mathcal{G}(0, \Sigma)$$

Explicit sparse representation of $\boldsymbol{J}^\dagger \Sigma^{-1} \boldsymbol{J}$ $\Rightarrow$ don't need action of $\boldsymbol{J}$ only $\boldsymbol{J}^\dagger$ $\Rightarrow$ chunked up operators

MGVI Considerations

- Metric operator

$$\Xi^{-1} = \left(A^\dagger \boldsymbol{J}^\dagger \Sigma^{-1} \boldsymbol{J} A + I\right)$$

Standardisation Term

- Gradient term

$$A^\dagger \boldsymbol{J}^\dagger \Sigma^{-1} (\mathbf{V}_{pq} - \sum_s \mathbf{G}_{p,s} \mathbf{X}_{pq,s} \mathbf{G}_{q,s})$$

- Back-projected noise

$$A^\dagger \boldsymbol{J}^\dagger \Sigma^{-1} \xi, \quad \text{where} \quad \xi \sim \mathcal{G}(0, \Sigma)$$

Explicit sparse representation of $\boldsymbol{J}^\dagger \Sigma^{-1} \boldsymbol{J} \Rightarrow$ don't need action of $\boldsymbol{J}$ only $\boldsymbol{J}^\dagger \Rightarrow$ chunked up operators