Expectation Propagation for Scalable Inverse Problems in Imaging

Dan Yao

Heriot-Watt University

UDRC summer school, University of Edinburgh
June, 27th 2022

This talk is about

This talk is about

Expectation Propagation

This talk is about

Expectation Propagation

1. Problem formulation and challenges

- Imaging inverse problems
- Bayesian estimation strategy
- challenges

2. Solution – EP for approximate Bayesian inference

- basic idea
- KL divergence minimization
- factor graph

3. Applications – EP for scalable imaging inverse problems

- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- toy examples and applications

4. Conclusion

This talk is about

Expectation Propagation

1. Problem formulation and challenges

- Imaging inverse problems
- Bayesian estimation strategy
- challenges

2. Solution – EP for approximate Bayesian inference

- basic idea
- KL divergence minimization
- factor graph

3. Applications – EP for scalable imaging inverse problems

- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- toy examples and applications

4. Conclusion

By the end of this talk,

you will know how to implement your own EP algorithm to:

This talk is about

Expectation Propagation

1. Problem formulation and challenges

- Imaging inverse problems
- Bayesian estimation strategy
- challenges

2. Solution – EP for approximate Bayesian inference

- basic idea
- KL divergence minimization
- factor graph

3. Applications – EP for scalable imaging inverse problems

- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- toy examples and applications

4. Conclusion

By the end of this talk,

you will know how to implement your own EP algorithm to:

This talk is about

Expectation Propagation

1. Problem formulation and challenges

- Imaging inverse problems
- Bayesian estimation strategy
- challenges

2. Solution – EP for approximate Bayesian inference

- basic idea
- KL divergence minimization
- factor graph

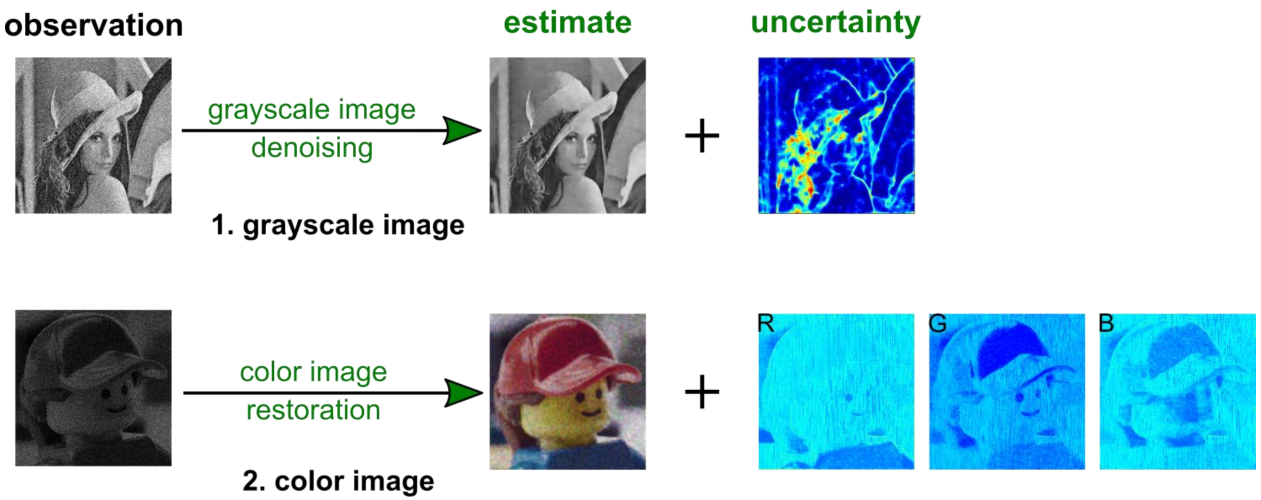
3. Applications – EP for scalable imaging inverse problems

- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- toy examples and applications

4. Conclusion

By the end of this talk,

you will know how to implement your own EP algorithm to:



This talk is about

Expectation Propagation

1. Problem formulation and challenges

- Imaging inverse problems
- Bayesian estimation strategy
- challenges

2. Solution – EP for approximate Bayesian inference

- basic idea
- KL divergence minimization
- factor graph

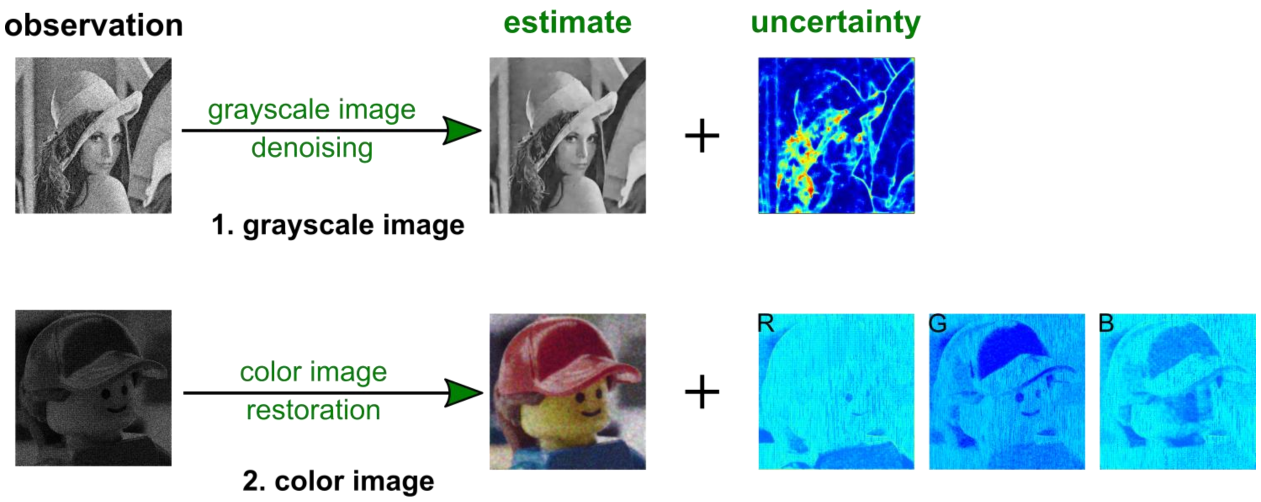
3. Applications – EP for scalable imaging inverse problems

- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- toy examples and applications

4. Conclusion

By the end of this talk,

you will know how to implement your own EP algorithm to:



This talk is about

Expectation Propagation

1. Problem formulation and challenges

- Imaging inverse problems
- Bayesian estimation strategy
- challenges

2. Solution – EP for approximate Bayesian inference

- basic idea
- KL divergence minimization
- factor graph

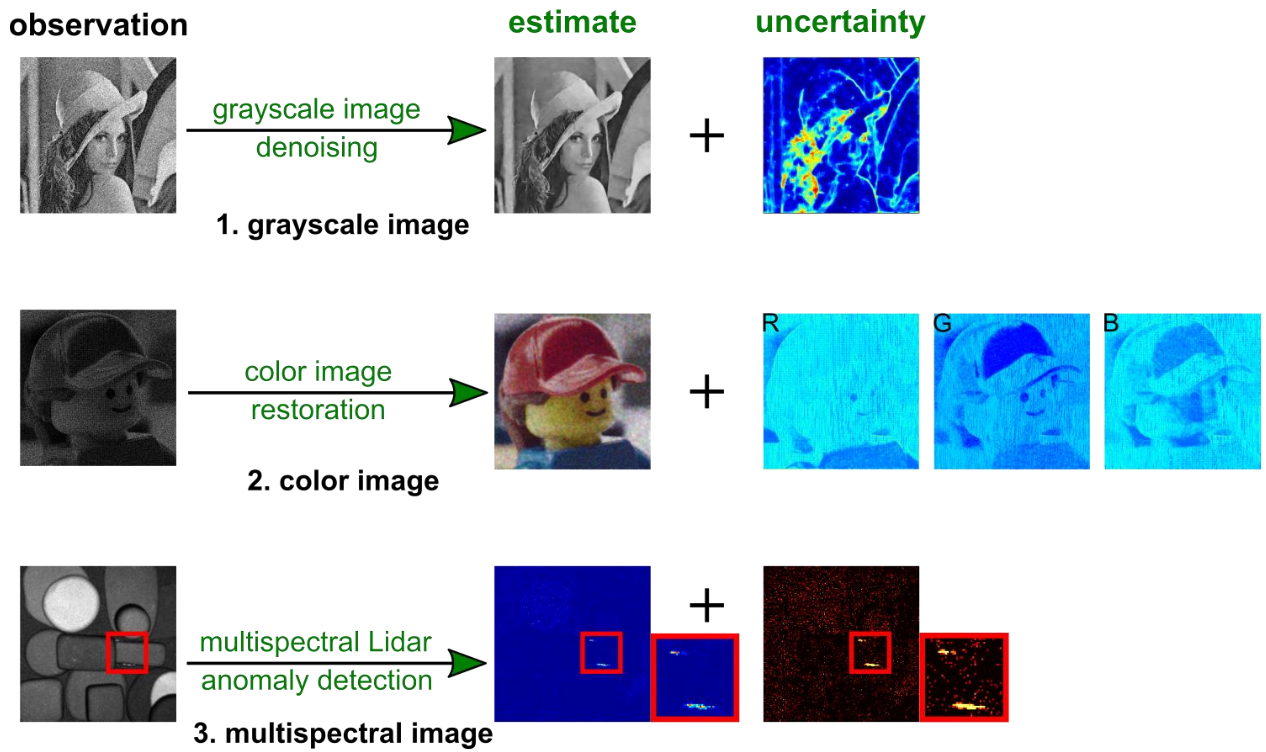
3. Applications – EP for scalable imaging inverse problems

- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- toy examples and applications

4. Conclusion

By the end of this talk,

you will know how to implement your own EP algorithm to:



This talk is about

Expectation Propagation

1. Problem formulation and challenges

- Imaging inverse problems
- Bayesian estimation strategy
- challenges

2. Solution – EP for approximate Bayesian inference

- basic idea
- KL divergence minimization
- factor graph

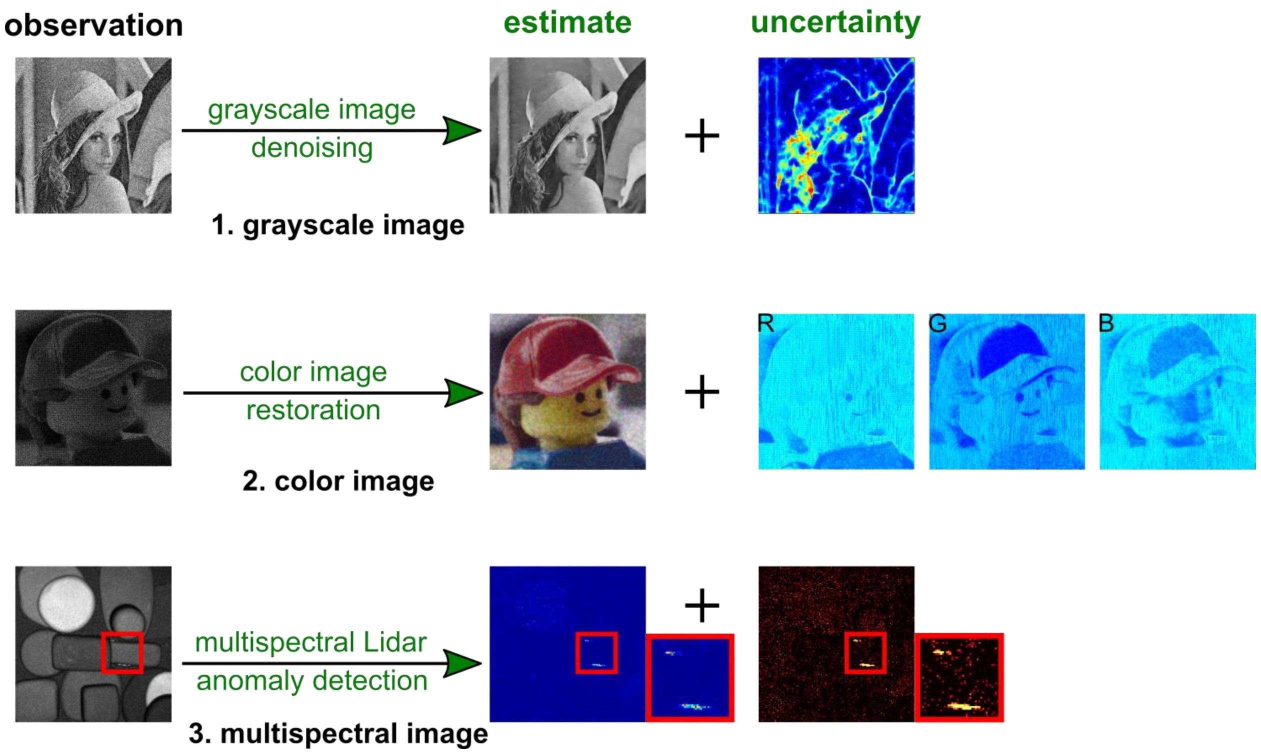
3. Applications – EP for scalable imaging inverse problems

- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- toy examples and applications

4. Conclusion

By the end of this talk,

you will know how to implement your own EP algorithm to:



This talk is about

Expectation Propagation

1. Problem formulation and challenges

- Imaging inverse problems
- Bayesian estimation strategy
- challenges

2. Solution – EP for approximate Bayesian inference

- basic idea
- KL divergence minimization
- factor graph

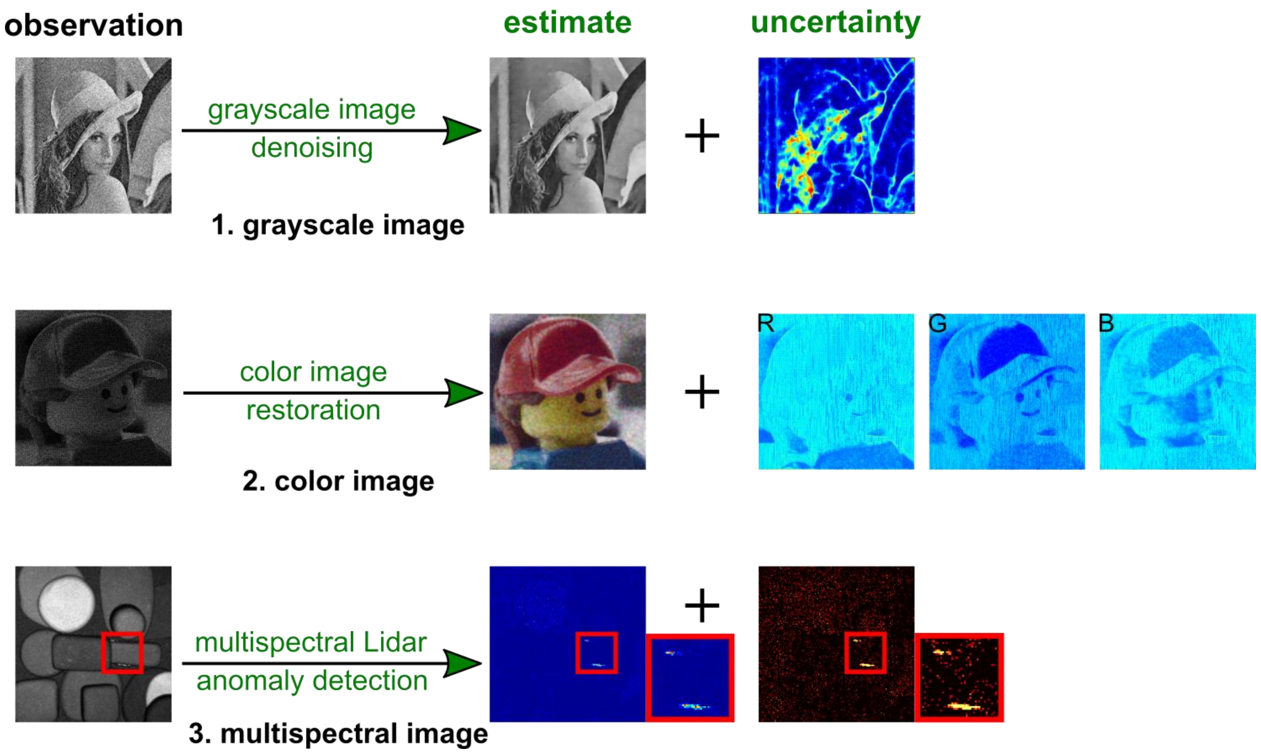
3. Applications – EP for scalable imaging inverse problems

- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- toy examples and applications

4. Conclusion

By the end of this talk,

you will know how to implement your own EP algorithm to:



1. Problem formulation and challenges

- **Imaging inverse problems**
- Bayesian estimation strategy
- challenges

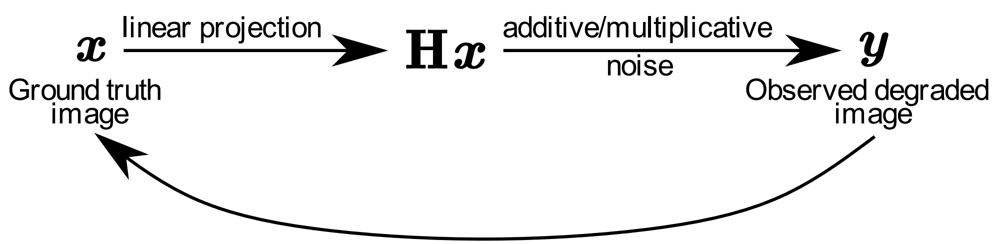
- **Model**:

$$
\begin{tikzcd}
\underset{\text{Ground truth image}}{\boldsymbol{x}} \arrow[r, "\text{linear projection}"] & \mathbf{H}\boldsymbol{x} \arrow[r, "\text{additive/multiplicative}", "\text{noise}"'] & \underset{\text{Observed degraded image}}{\boldsymbol{y}} \arrow[ll, bend left, "\text{Linear inverse problems in imaging}"]
\end{tikzcd}
$$

1. Problem formulation and challenges

- **Imaging inverse problems**
- Bayesian estimation strategy
- challenges

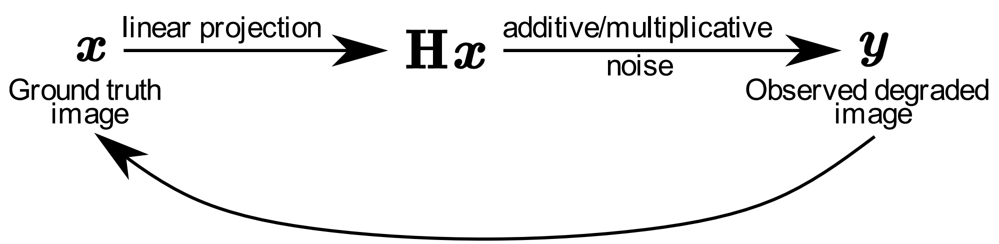
- **Model**:



Linear inverse problems in imaging

- **Goal:**

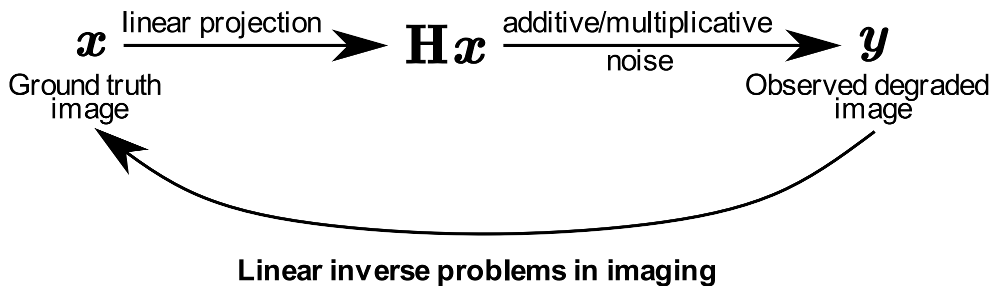
- **Model**:



Linear inverse problems in imaging

- **Goal:** $\boldsymbol{y}, \mathbf{H} \longrightarrow \hat{\boldsymbol{x}}$

- **Model**:



- **Goal:** $\boldsymbol{y}, \mathbf{H} \longrightarrow \hat{\boldsymbol{x}} + \textbf{uncertainty}$

- **Model**:

$$\underset{\text{Ground truth image}}{\boldsymbol{x}} \xrightarrow{\text{linear projection}} \mathbf{H}\boldsymbol{x} \xrightarrow[\text{noise}]{\text{additive/multiplicative}} \underset{\text{Observed degraded image}}{\boldsymbol{y}}$$

Linear inverse problems in imaging

- **Goal:**

$$\boldsymbol{y}, \mathbf{H} \xrightarrow{\textit{scalable}} \hat{\boldsymbol{x}} + \textbf{uncertainty}$$

- **Model**:

$$\underset{\text{Ground truth image}}{\boldsymbol{x}} \xrightarrow{\text{linear projection}} \mathbf{H}\boldsymbol{x} \xrightarrow[\text{noise}]{\text{additive/multiplicative}} \underset{\text{Observed degraded image}}{\boldsymbol{y}}$$

Linear inverse problems in imaging

- **Goal:** $\boldsymbol{y}, \mathbf{H} \xrightarrow{\textit{scalable}} \hat{\boldsymbol{x}} + \textbf{uncertainty}$

- **Examples:**

- **Model**:

$$\underset{\text{Ground truth image}}{\boldsymbol{x}} \xrightarrow{\text{linear projection}} \mathbf{H}\boldsymbol{x} \xrightarrow[\text{noise}]{\text{additive/multiplicative}} \underset{\text{Observed degraded image}}{\boldsymbol{y}}$$

Linear inverse problems in imaging

- **Goal:** $\boldsymbol{y}, \mathbf{H} \xrightarrow{\textit{scalable}} \hat{\boldsymbol{x}} + \textbf{uncertainty}$

- **Examples:**

- **Model**:

$$\underset{\text{Ground truth image}}{\boldsymbol{x}} \xrightarrow{\text{linear projection}} \mathbf{H}\boldsymbol{x} \xrightarrow[\text{noise}]{\text{additive/multiplicative}} \underset{\text{Observed degraded image}}{\boldsymbol{y}}$$

Linear inverse problems in imaging

- **Goal:**

$$\boldsymbol{y}, \mathbf{H} \xrightarrow{\textit{scalable}} \hat{\boldsymbol{x}} + \textbf{uncertainty}$$

- **Examples:**

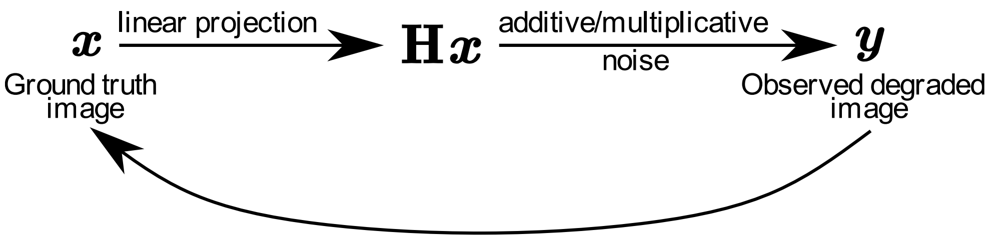
$\boldsymbol{x} \in \mathbb{R}^{N\times 1}$

Ground truth image

1. Problem formulation and challenges

- ➢ **Imaging inverse problems**
- ➢ Bayesian estimation strategy
- ➢ challenges

- **Model**:



Linear inverse problems in imaging

- **Goal:** $\boldsymbol{y}, \mathbf{H} \xrightarrow{\textit{scalable}} \hat{\boldsymbol{x}} + \textbf{uncertainty}$

- **Examples:**

$\boldsymbol{x} \in \mathbb{R}^{N \times 1}$

Ground truth image

$\boldsymbol{y} = \mathcal{N}(\boldsymbol{x}, \sigma^2 \mathbf{I}) \in \mathbb{R}^{N \times 1}$

Noisy observation

- **Model**:

$$\underset{\text{Ground truth image}}{\boldsymbol{x}} \xrightarrow{\text{linear projection}} \mathbf{H}\boldsymbol{x} \xrightarrow[\text{noise}]{\text{additive/multiplicative}} \underset{\text{Observed degraded image}}{\boldsymbol{y}}$$

Linear inverse problems in imaging

- **Goal:**

$$\boldsymbol{y}, \mathbf{H} \xrightarrow{\textit{scalable}} \hat{\boldsymbol{x}} + \textbf{uncertainty}$$

- **Examples:**

$\boldsymbol{x} \in \mathbb{R}^{N\times 1}$

Ground truth image

$\boldsymbol{y} = \mathcal{N}(\boldsymbol{x}, \sigma^2\mathbf{I}) \in \mathbb{R}^{N\times 1}$

Noisy observation

- **Model**:

$$\underset{\text{Ground truth image}}{\boldsymbol{x}} \xrightarrow{\text{linear projection}} \mathbf{H}\boldsymbol{x} \xrightarrow[\text{noise}]{\text{additive/multiplicative}} \underset{\text{Observed degraded image}}{\boldsymbol{y}}$$

Linear inverse problems in imaging

- **Goal:**

$$\boldsymbol{y}, \mathbf{H} \xrightarrow{\textit{scalable}} \hat{\boldsymbol{x}} + \textbf{uncertainty}$$

- **Examples:**

$\boldsymbol{x} \in \mathbb{R}^{N \times 1}$

Ground truth image

$\boldsymbol{y} = \mathcal{N}(\boldsymbol{x}, \sigma^2 \mathbf{I}) \in \mathbb{R}^{N \times 1}$

Noisy observation

$\hat{\boldsymbol{x}}$

1. Problem formulation and challenges

- **Imaging inverse problems**
- Bayesian estimation strategy
- challenges

- **Model**:

$$\underset{\text{Ground truth image}}{\boldsymbol{x}} \xrightarrow{\text{linear projection}} \mathbf{H}\boldsymbol{x} \xrightarrow[\text{noise}]{\text{additive/multiplicative}} \underset{\text{Observed degraded image}}{\boldsymbol{y}}$$

Linear inverse problems in imaging

- **Goal:** $\boldsymbol{y}, \mathbf{H} \xrightarrow{\textit{scalable}} \hat{\boldsymbol{x}} + \textbf{uncertainty}$

- **Examples:**

$\boldsymbol{x} \in \mathbb{R}^{N \times 1}$

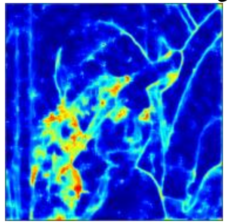
Ground truth image

$\boldsymbol{y} = \mathcal{N}(\boldsymbol{x}, \sigma^2 \mathbf{I}) \in \mathbb{R}^{N \times 1}$

Noisy observation

$\hat{\boldsymbol{x}}$

uncertainty



- **Model:**

$$\boldsymbol{x} \xrightarrow{\text{linear projection}} \mathbf{H}\boldsymbol{x} \xrightarrow[\text{noise}]{\text{additive/multiplicative}} \boldsymbol{y}$$

$\boldsymbol{x}$: Ground truth image; $\boldsymbol{y}$: Observed degraded image

Linear inverse problems in imaging

- **Goal:**

$$\boldsymbol{y}, \mathbf{H} \xrightarrow{\textit{scalable}} \hat{\boldsymbol{x}} + \textbf{uncertainty}$$

- **Examples:**

$\boldsymbol{x} \in \mathbb{R}^{N\times 1}$

Ground truth image

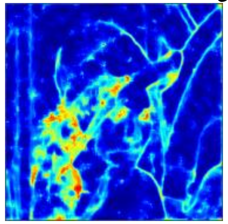
$\boldsymbol{y} = \mathcal{N}(\boldsymbol{x}, \sigma^2\mathbf{I}) \in \mathbb{R}^{N\times 1}$

Noisy observation

$\xrightarrow{\textit{scalable}}$

$\hat{\boldsymbol{x}}$

uncertainty



- **Model**:

$$\boldsymbol{x} \xrightarrow{\text{linear projection}} \mathbf{H}\boldsymbol{x} \xrightarrow[\text{noise}]{\text{additive/multiplicative}} \boldsymbol{y}$$

Ground truth image — Observed degraded image

Linear inverse problems in imaging

- **Goal:**

$$\boldsymbol{y}, \mathbf{H} \xrightarrow{\textit{scalable}} \hat{\boldsymbol{x}} + \textbf{uncertainty}$$

- **Examples:**

- **Model**:

$$\boldsymbol{x} \xrightarrow{\text{linear projection}} \mathbf{H}\boldsymbol{x} \xrightarrow[\text{noise}]{\text{additive/multiplicative}} \boldsymbol{y}$$

Ground truth image — Observed degraded image

Linear inverse problems in imaging

- **Goal:**

$$\boldsymbol{y}, \mathbf{H} \xrightarrow{\textit{scalable}} \hat{\boldsymbol{x}} + \textbf{uncertainty}$$

- **Examples:**

- **Model**:

$$\boldsymbol{x} \xrightarrow{\text{linear projection}} \mathbf{H}\boldsymbol{x} \xrightarrow[\text{noise}]{\text{additive/multiplicative}} \boldsymbol{y}$$

Ground truth image — Observed degraded image

Linear inverse problems in imaging

- **Goal:**

$$\boldsymbol{y}, \mathbf{H} \xrightarrow{\textit{scalable}} \hat{\boldsymbol{x}} + \textbf{uncertainty}$$

- **Examples:**

$\boldsymbol{x} \in \mathbb{R}^{N \times 3}$

Ground truth image

- **Model**:

$$\underset{\text{Ground truth image}}{\boldsymbol{x}} \xrightarrow{\text{linear projection}} \mathbf{H}\boldsymbol{x} \xrightarrow[\text{noise}]{\text{additive/multiplicative}} \underset{\text{Observed degraded image}}{\boldsymbol{y}}$$

Linear inverse problems in imaging

- **Goal:**

$$\boldsymbol{y}, \mathbf{H} \xrightarrow{\textit{scalable}} \hat{\boldsymbol{x}} + \textbf{uncertainty}$$

- **Examples:**

$\boldsymbol{x} \in \mathbb{R}^{N\times 3}$

Ground truth image

$\boldsymbol{y} = \mathcal{P}(\mathbf{H}\boldsymbol{x}) \in \mathbb{R}^{N\times 1}$

Noisy observation

- **Model**:

$$\underset{\text{Ground truth image}}{\boldsymbol{x}} \xrightarrow{\text{linear projection}} \mathbf{H}\boldsymbol{x} \xrightarrow[\text{noise}]{\text{additive/multiplicative}} \underset{\text{Observed degraded image}}{\boldsymbol{y}}$$

Linear inverse problems in imaging

- **Goal:**

$$\boldsymbol{y}, \mathbf{H} \xrightarrow{\textit{scalable}} \hat{\boldsymbol{x}} + \textbf{uncertainty}$$

- **Examples:**

$\boldsymbol{x} \in \mathbb{R}^{N \times 3}$

Ground truth image

$\boldsymbol{y} = \mathcal{P}(\mathbf{H}\boldsymbol{x}) \in \mathbb{R}^{N \times 1}$

Noisy observation

- **Model**:

$$\boldsymbol{x} \xrightarrow{\text{linear projection}} \mathbf{H}\boldsymbol{x} \xrightarrow[\text{noise}]{\text{additive/multiplicative}} \boldsymbol{y}$$

Ground truth image — Observed degraded image

Linear inverse problems in imaging

- **Goal:**

$$\boldsymbol{y}, \mathbf{H} \xrightarrow{\textit{scalable}} \hat{\boldsymbol{x}} + \textbf{uncertainty}$$

- **Examples:**

$\boldsymbol{x} \in \mathbb{R}^{N\times 3}$

Ground truth image

$\boldsymbol{y} = \mathcal{P}(\mathbf{H}\boldsymbol{x}) \in \mathbb{R}^{N\times 1}$

Noisy observation

$\hat{\boldsymbol{x}} \in \mathbb{R}^{N\times 3}$

- **Model**:

$$\underset{\text{Ground truth image}}{\boldsymbol{x}} \xrightarrow{\text{linear projection}} \mathbf{H}\boldsymbol{x} \xrightarrow[\text{noise}]{\text{additive/multiplicative}} \underset{\text{Observed degraded image}}{\boldsymbol{y}}$$

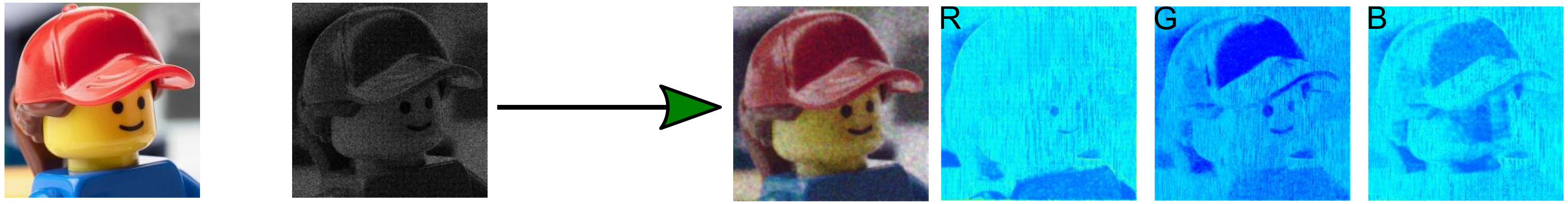
Linear inverse problems in imaging

- **Goal:**

$$\boldsymbol{y}, \mathbf{H} \xrightarrow{\textit{scalable}} \hat{\boldsymbol{x}} + \textbf{uncertainty}$$

- **Examples:**

$\boldsymbol{x} \in \mathbb{R}^{N\times 3}$ $\quad$ $\boldsymbol{y} = \mathcal{P}(\mathbf{H}\boldsymbol{x}) \in \mathbb{R}^{N\times 1}$ $\quad$ $\hat{\boldsymbol{x}} \in \mathbb{R}^{N\times 3}$ $\quad$ **uncertainty**



Ground truth image $\quad$ **Noisy observation**

- **Model**:

$$\underset{\text{Ground truth image}}{\boldsymbol{x}} \xrightarrow{\text{linear projection}} \mathbf{H}\boldsymbol{x} \xrightarrow[\text{noise}]{\text{additive/multiplicative}} \underset{\text{Observed degraded image}}{\boldsymbol{y}}$$

Linear inverse problems in imaging

- **Goal:**

$$\boldsymbol{y}, \mathbf{H} \xrightarrow{\textit{scalable}} \hat{\boldsymbol{x}} + \textbf{uncertainty}$$

- **Examples:**

$\boldsymbol{x} \in \mathbb{R}^{N\times 3}$

Ground truth image

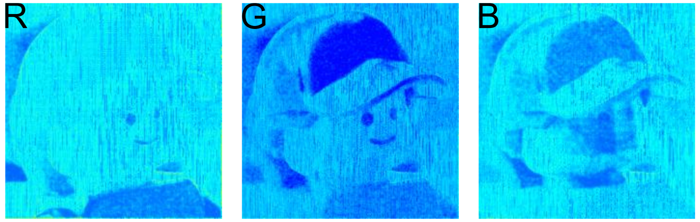
$\boldsymbol{y} = \mathcal{P}(\mathbf{H}\boldsymbol{x}) \in \mathbb{R}^{N\times 1}$

Noisy observation

$\xrightarrow{\textit{scalable}}$

$\hat{\boldsymbol{x}} \in \mathbb{R}^{N\times 3}$

uncertainty



- **Bayesian model**:

- **Bayesian model**:

likelihood: $f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})$

- **Bayesian model**:

likelihood: $f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})$

prior: $f_{x}(\boldsymbol{x}|\boldsymbol{\theta})$

- **Bayesian model**:

likelihood: $f_{y|x}\left(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}\right)$

prior: $f_{x}\left(\boldsymbol{x}|\boldsymbol{\theta}\right)$

posterior: $p\left(\boldsymbol{x}|\boldsymbol{y}\right) = \dfrac{f_{y|x}\left(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}\right) f_{x}\left(\boldsymbol{x}|\boldsymbol{\theta}\right)}{\int f_{y|x}\left(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}\right) f_{x}\left(\boldsymbol{x}|\boldsymbol{\theta}\right) \mathrm{d}\boldsymbol{x}}$

- **Bayesian model**:

likelihood: $f_{y|x}\left(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}\right)$

prior: $f_{x}\left(\boldsymbol{x}|\boldsymbol{\theta}\right)$

posterior: $p\left(\boldsymbol{x}|\boldsymbol{y}\right) = \dfrac{f_{y|x}\left(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}\right) f_{x}\left(\boldsymbol{x}|\boldsymbol{\theta}\right)}{\int f_{y|x}\left(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}\right) f_{x}\left(\boldsymbol{x}|\boldsymbol{\theta}\right) \mathrm{d}\boldsymbol{x}}$

- **Estimation strategy**:

- **Bayesian model**:

likelihood: $f_{y|x}\left(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}\right)$

prior: $f_{x}\left(\boldsymbol{x}|\boldsymbol{\theta}\right)$

posterior: $p\left(\boldsymbol{x}|\boldsymbol{y}\right) = \dfrac{f_{y|x}\left(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}\right) f_{x}\left(\boldsymbol{x}|\boldsymbol{\theta}\right)}{\int f_{y|x}\left(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}\right) f_{x}\left(\boldsymbol{x}|\boldsymbol{\theta}\right) \mathrm{d}\boldsymbol{x}}$

- **Estimation strategy**:

point estimate:

- **Bayesian model**:

likelihood: $f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})$

prior: $f_x(\boldsymbol{x}|\boldsymbol{\theta})$

posterior: $p(\boldsymbol{x}|\boldsymbol{y}) = \dfrac{f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})}{\int f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x}}$

- **Estimation strategy**:

point estimate: $\hat{\boldsymbol{x}}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{x}} \; p(\boldsymbol{x}|\boldsymbol{y})$

- **Bayesian model**:

likelihood: $f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})$

prior: $f_x(\boldsymbol{x}|\boldsymbol{\theta})$

posterior: $p(\boldsymbol{x}|\boldsymbol{y}) = \dfrac{f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})}{\int f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x}}$

- **Estimation strategy**:

point estimate: $\hat{\boldsymbol{x}}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{x}} \; p(\boldsymbol{x}|\boldsymbol{y})$

or

$\hat{\boldsymbol{x}}_{\mathrm{MMSE}} = \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}]$

- **Bayesian model**:

likelihood: $f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})$

prior: $f_x(\boldsymbol{x}|\boldsymbol{\theta})$

posterior: $p(\boldsymbol{x}|\boldsymbol{y}) = \dfrac{f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) f_x(\boldsymbol{x}|\boldsymbol{\theta})}{\int f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) f_x(\boldsymbol{x}|\boldsymbol{\theta}) \mathrm{d}\boldsymbol{x}}$

- **Estimation strategy**:

point estimate: $\hat{\boldsymbol{x}}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{x}} \; p(\boldsymbol{x}|\boldsymbol{y})$

or

$\hat{\boldsymbol{x}}_{\mathrm{MMSE}} = \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}]$

uncertainty:

- **Bayesian model**:

likelihood: $f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})$

prior: $f_x(\boldsymbol{x}|\boldsymbol{\theta})$

posterior: $p(\boldsymbol{x}|\boldsymbol{y}) = \dfrac{f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})}{\int f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{x}}$

- **Estimation strategy**:

point estimate: $\hat{\boldsymbol{x}}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{x}} \; p(\boldsymbol{x}|\boldsymbol{y})$

or

$\hat{\boldsymbol{x}}_{\mathrm{MMSE}} = \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}]$

uncertainty: $\mathrm{Cov}_{p(\boldsymbol{x}|\boldsymbol{y})}(\boldsymbol{x})$

- **Bayesian model**:

likelihood: $f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})$

prior: $f_x(\boldsymbol{x}|\boldsymbol{\theta})$

posterior: $p(\boldsymbol{x}|\boldsymbol{y}) = \dfrac{f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})}{\int f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{x}}$

- **Estimation strategy**:

point estimate: $\hat{\boldsymbol{x}}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{x}}\ p(\boldsymbol{x}|\boldsymbol{y})$

or

$\hat{\boldsymbol{x}}_{\mathrm{MMSE}} = \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}]$

uncertainty: $\mathrm{Cov}_{p(\boldsymbol{x}|\boldsymbol{y})}(\boldsymbol{x})$

- **Challenges:**

- **Bayesian model**:

likelihood: $f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})$

prior: $f_x(\boldsymbol{x}|\boldsymbol{\theta})$

posterior: $p(\boldsymbol{x}|\boldsymbol{y}) = \dfrac{f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})}{\int f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{x}}$

- **Estimation strategy**:

point estimate: $\hat{\boldsymbol{x}}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{x}} \ p(\boldsymbol{x}|\boldsymbol{y})$

or

$\hat{\boldsymbol{x}}_{\mathrm{MMSE}} = \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}]$

uncertainty: $\mathrm{Cov}_{p(\boldsymbol{x}|\boldsymbol{y})}(\boldsymbol{x})$

- **Challenges:**

high-dimensional $\boldsymbol{x} = [x_1, \cdots, x_N]^T$

- **Bayesian model**:

likelihood: $f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})$

prior: $f_x(\boldsymbol{x}|\boldsymbol{\theta})$

posterior: $p(\boldsymbol{x}|\boldsymbol{y}) = \dfrac{f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})}{\int f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{x}}$

- **Estimation strategy**:

point estimate: $\hat{\boldsymbol{x}}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{x}} \; p(\boldsymbol{x}|\boldsymbol{y})$

or

$\hat{\boldsymbol{x}}_{\mathrm{MMSE}} = \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}]$

uncertainty: $\mathrm{Cov}_{p(\boldsymbol{x}|\boldsymbol{y})}(\boldsymbol{x})$

1. Problem formulation and challenges
- Imaging inverse problems
- Bayesian estimation strategy
- **challenges**

- **Challenges:**

high-dimensional $\boldsymbol{x} = [x_1, \cdots, x_N]^T$

e.g. $\boldsymbol{x} = [x_1, \ldots, x_{10000}]^T$

1. Problem formulation and challenges
- Imaging inverse problems
- Bayesian estimation strategy
- **challenges**

- **Bayesian model**:

likelihood: $f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})$

prior: $f_x(\boldsymbol{x}|\boldsymbol{\theta})$

posterior: $p(\boldsymbol{x}|\boldsymbol{y}) = \dfrac{f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})}{\int f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{x}}$

- **Estimation strategy**:

point estimate: $\hat{\boldsymbol{x}}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{x}}\ p(\boldsymbol{x}|\boldsymbol{y})$

or

$\hat{\boldsymbol{x}}_{\mathrm{MMSE}} = \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}]$

uncertainty: $\mathrm{Cov}_{p(\boldsymbol{x}|\boldsymbol{y})}(\boldsymbol{x})$

- **Challenges:**

high-dimensional $\boldsymbol{x} = [x_1, \cdots, x_N]^T$

e.g. $\boldsymbol{x} = [x_1, \ldots, x_{10000}]^T$

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})}{\int f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})\,\mathrm{d}[x_1, \ldots, x_{10000}]^T}$$

- **Bayesian model**:

likelihood: $f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})$

prior: $f_x(\boldsymbol{x}|\boldsymbol{\theta})$

posterior: $p(\boldsymbol{x}|\boldsymbol{y}) = \dfrac{f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\,f_x(\boldsymbol{x}|\boldsymbol{\theta})}{\int f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\,f_x(\boldsymbol{x}|\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{x}}$

- **Estimation strategy**:

point estimate: $\hat{\boldsymbol{x}}_{\text{MAP}} = \arg\max_{\boldsymbol{x}}\; p(\boldsymbol{x}|\boldsymbol{y})$

or

$\hat{\boldsymbol{x}}_{\text{MMSE}} = \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}]$

uncertainty: $\mathrm{Cov}_{p(\boldsymbol{x}|\boldsymbol{y})}(\boldsymbol{x})$

1. Problem formulation and challenges

- Imaging inverse problems
- Bayesian estimation strategy
- **challenges**

- **Challenges:**

high-dimensional $\boldsymbol{x} = [x_1, \cdots, x_N]^T$

e.g. $\boldsymbol{x} = [x_1, \ldots, x_{10000}]^T$

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\,f_x(\boldsymbol{x}|\boldsymbol{\theta})}{\int f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\,f_x(\boldsymbol{x}|\boldsymbol{\theta})\,\mathrm{d}[x_1, \ldots, x_{10000}]^T}$$

$$\boldsymbol{\theta}^{(t)} = \arg\max_{\boldsymbol{\theta}} \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(t-1)})}[\log\, f(\boldsymbol{y}|\boldsymbol{\theta})]$$

1. Problem formulation and challenges
- Imaging inverse problems
- Bayesian estimation strategy
- **challenges**

- **Bayesian model**:

likelihood: $f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})$

prior: $f_x(\boldsymbol{x}|\boldsymbol{\theta})$

posterior: $p(\boldsymbol{x}|\boldsymbol{y}) = \dfrac{f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})f_x(\boldsymbol{x}|\boldsymbol{\theta})}{\int f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})f_x(\boldsymbol{x}|\boldsymbol{\theta})\mathrm{d}\boldsymbol{x}}$

- **Estimation strategy**:

point estimate: $\hat{\boldsymbol{x}}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{x}} p(\boldsymbol{x}|\boldsymbol{y})$

or

$\hat{\boldsymbol{x}}_{\mathrm{MMSE}} = \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}]$

uncertainty: $\mathrm{Cov}_{p(\boldsymbol{x}|\boldsymbol{y})}(\boldsymbol{x})$

- **Challenges:**

high-dimensional $\boldsymbol{x} = [x_1, \cdots, x_N]^T$

e.g. $\boldsymbol{x} = [x_1, \ldots, x_{10000}]^T$

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})f_x(\boldsymbol{x}|\boldsymbol{\theta})}{\int f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})f_x(\boldsymbol{x}|\boldsymbol{\theta})\mathrm{d}[x_1, \ldots, x_{10000}]^T}$$

$$\boldsymbol{\theta}^{(t)} = \arg\max_{\boldsymbol{\theta}} \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(t-1)})}[\log f(\boldsymbol{y}|\boldsymbol{\theta})]$$

$$\mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}] = \int \boldsymbol{x}p(\boldsymbol{x}|\boldsymbol{y})\mathrm{d}[x_1, \ldots, x_{10000}]^T$$

1. Problem formulation and challenges
➢ Imaging inverse problems
➢ Bayesian estimation strategy
➢ **challenges**

- **Bayesian model**:

likelihood: $f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})$

prior: $f_x(\boldsymbol{x}|\boldsymbol{\theta})$

posterior: $p(\boldsymbol{x}|\boldsymbol{y}) = \dfrac{f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})f_x(\boldsymbol{x}|\boldsymbol{\theta})}{\int f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})f_x(\boldsymbol{x}|\boldsymbol{\theta})\mathrm{d}\boldsymbol{x}}$

- **Estimation strategy**:

point estimate: $\hat{\boldsymbol{x}}_{\text{MAP}} = \arg\max_{\boldsymbol{x}} \; p(\boldsymbol{x}|\boldsymbol{y})$

or

$\hat{\boldsymbol{x}}_{\text{MMSE}} = \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}]$

uncertainty: $\text{Cov}_{p(\boldsymbol{x}|\boldsymbol{y})}(\boldsymbol{x})$

- **Challenges:**

high-dimensional $\boldsymbol{x} = [x_1, \cdots, x_N]^T$

e.g. $\boldsymbol{x} = [x_1, \ldots, x_{10000}]^T$

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})f_x(\boldsymbol{x}|\boldsymbol{\theta})}{\int f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})f_x(\boldsymbol{x}|\boldsymbol{\theta})\mathrm{d}[x_1, \ldots, x_{10000}]^T}$$

$$\boldsymbol{\theta}^{(t)} = \arg\max_{\boldsymbol{\theta}} \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(t-1)})}[\log f(\boldsymbol{y}|\boldsymbol{\theta})]$$

$$\mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}] = \int \boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y})\mathrm{d}[x_1, \ldots, x_{10000}]^T$$

$$\text{Cov}_{p(\boldsymbol{x}|\boldsymbol{y})}(\boldsymbol{x}) = \int (\boldsymbol{x} - \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}])(\boldsymbol{x} - \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}])^T p(\boldsymbol{x}|\boldsymbol{y})\mathrm{d}[x_1, \ldots, x_{10000}]^T$$

This talk is about

Expectation Propagation

1. Problem formulation and challenges
- ➢ Imaging inverse problems
- ➢ Bayesian estimation strategy
- ➢ challenges

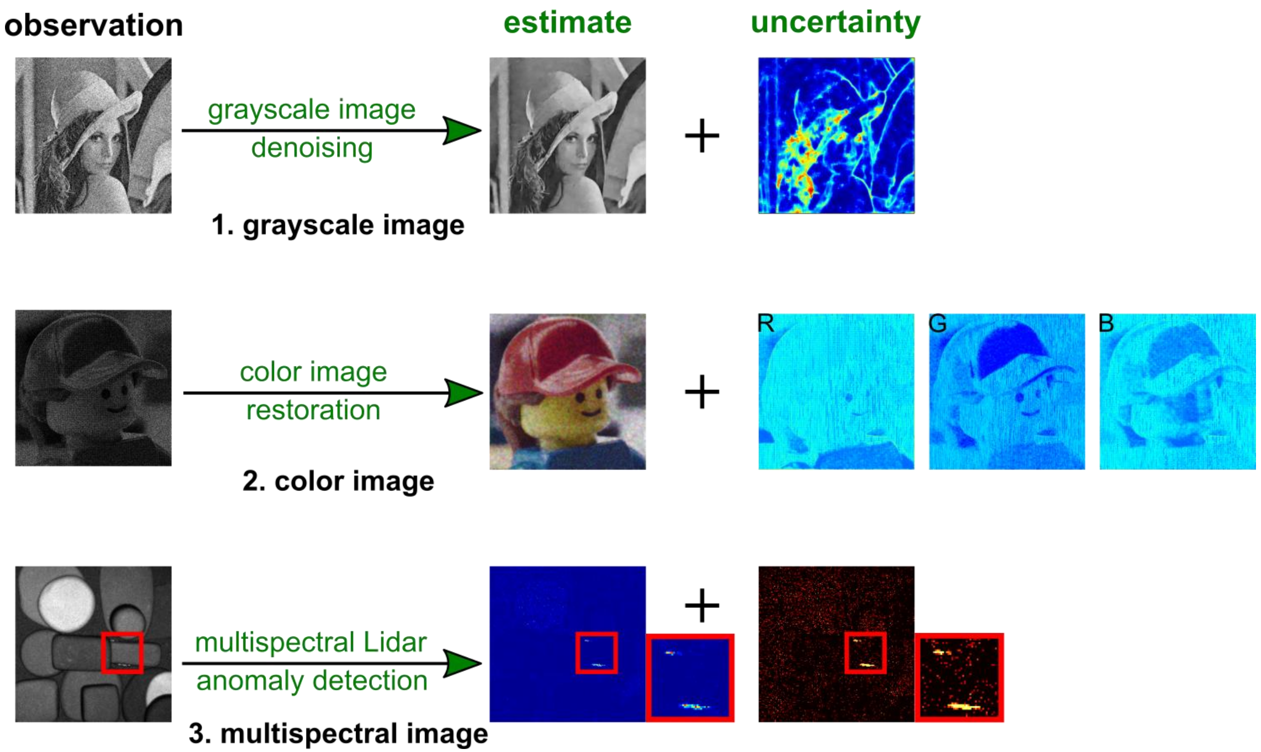
2. Solution – EP for approximate Bayesian inference
- ➢ basic idea
- ➢ KL divergence minimization
- ➢ factor graph

3. Applications – EP for scalable imaging inverse problems
- ➢ how to construct an EP algorithm to solve image inverse problems
- ➢ how to achieve scalable posterior approximation
- ➢ toy examples and applications

4. Conclusion

By the end of this talk,

you will know how to implement your own EP algorithm to:



$$q(\boldsymbol{x}) \approx p(\boldsymbol{x}|\boldsymbol{y})$$

🙂 🥵

$$q(\boldsymbol{x}) \approx p(\boldsymbol{x}|\boldsymbol{y})$$

$q(\boldsymbol{x})$ does not require to compute $\boxed{\cancel{p(\boldsymbol{x}|\boldsymbol{y}) = \frac{f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})}{\int f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x}}}}$

$$q(\boldsymbol{x}) \approx p(\boldsymbol{x}|\boldsymbol{y})$$

$q(\boldsymbol{x})$ does not require to compute $\boxed{\xcancel{p(\boldsymbol{x}|\boldsymbol{y}) = \frac{f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})f_x(\boldsymbol{x}|\boldsymbol{\theta})}{\int f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})f_x(\boldsymbol{x}|\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{x}}}}$

$$\boldsymbol{\theta}^{(t)} = \arg\max_{\boldsymbol{\theta}} \mathbb{E}_{q(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(t-1)})}[\log f(\boldsymbol{y}|\boldsymbol{\theta})] \qquad \boxed{\xcancel{\boldsymbol{\theta}^{(t)} = \arg\max_{\boldsymbol{\theta}} \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(t-1)})}[\log f(\boldsymbol{y}|\boldsymbol{\theta})]}}$$

2. Solution – EP for approximate Bayesian inference

- **basic idea**
- KL divergence minimization
- factor graph

$$q(\boldsymbol{x}) \approx p(\boldsymbol{x}|\boldsymbol{y})$$

$q(\boldsymbol{x})$ does not require to compute $\boxed{\cancel{p(\boldsymbol{x}|\boldsymbol{y}) = \frac{f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})f_x(\boldsymbol{x}|\boldsymbol{\theta})}{\int f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})f_x(\boldsymbol{x}|\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{x}}}}$

$$\boldsymbol{\theta}^{(t)} = \arg\max_{\boldsymbol{\theta}} \mathbb{E}_{q(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(t-1)})}[\log f(\boldsymbol{y}|\boldsymbol{\theta})] \qquad \boxed{\cancel{\boldsymbol{\theta}^{(t)} = \arg\max_{\boldsymbol{\theta}} \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(t-1)})}[\log f(\boldsymbol{y}|\boldsymbol{\theta})]}}$$

$\mathbb{E}_{q(\boldsymbol{x})}[\boldsymbol{x}], \mathrm{Cov}_{q(\boldsymbol{x})}(\boldsymbol{x})$ are easier to compute $\boxed{\cancel{\mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}] = \int \boldsymbol{x}p(\boldsymbol{x}|\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}}}$ $\boxed{\cancel{\mathrm{Cov}_{p(\boldsymbol{x}|\boldsymbol{y})}(\boldsymbol{x}) = \int (\boldsymbol{x} - \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}])(\boldsymbol{x} - \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}])^T p(\boldsymbol{x}|\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}}}$

2. Solution – EP for approximate Bayesian inference

- **basic idea**
- KL divergence minimization
- factor graph

$$q(\boldsymbol{x}) \approx p(\boldsymbol{x}|\boldsymbol{y})$$

$q(\boldsymbol{x})$ does not require to compute $\boxed{\cancel{p(\boldsymbol{x}|\boldsymbol{y}) = \frac{f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})f_x(\boldsymbol{x}|\boldsymbol{\theta})}{\int f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})f_x(\boldsymbol{x}|\boldsymbol{\theta})\mathrm{d}\boldsymbol{x}}}}$

$$\boldsymbol{\theta}^{(t)} = \arg\max_{\boldsymbol{\theta}} \mathbb{E}_{q(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(t-1)})}[\log f(\boldsymbol{y}|\boldsymbol{\theta})] \quad \boxed{\cancel{\boldsymbol{\theta}^{(t)} = \arg\max_{\boldsymbol{\theta}} \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(t-1)})}[\log f(\boldsymbol{y}|\boldsymbol{\theta})]}}$$

$\mathbb{E}_{q(\boldsymbol{x})}[\boldsymbol{x}], \mathrm{Cov}_{q(\boldsymbol{x})}(\boldsymbol{x})$ are easier to compute $\boxed{\cancel{\mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}] = \int \boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y})\mathrm{d}\boldsymbol{x}}}$ $\boxed{\cancel{\mathrm{Cov}_{p(\boldsymbol{x}|\boldsymbol{y})}(\boldsymbol{x}) = \int (\boldsymbol{x} - \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}])(\boldsymbol{x} - \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}])^T p(\boldsymbol{x}|\boldsymbol{y})\mathrm{d}\boldsymbol{x}}}$

- **Approximate Bayesian inference:**

😊 🥵

$$q(\boldsymbol{x}) \approx p(\boldsymbol{x}|\boldsymbol{y})$$

$q(\boldsymbol{x})$ does not require to compute $\boxed{\cancel{p(\boldsymbol{x}|\boldsymbol{y}) = \dfrac{f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})f_x(\boldsymbol{x}|\boldsymbol{\theta})}{\int f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})f_x(\boldsymbol{x}|\boldsymbol{\theta})\mathrm{d}\boldsymbol{x}}}}$

$$\boldsymbol{\theta}^{(t)} = \arg\max_{\boldsymbol{\theta}} \mathbb{E}_{q(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(t-1)})}[\log f(\boldsymbol{y}|\boldsymbol{\theta})] \qquad \boxed{\cancel{\boldsymbol{\theta}^{(t)} = \arg\max_{\boldsymbol{\theta}} \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(t-1)})}[\log f(\boldsymbol{y}|\boldsymbol{\theta})]}}$$

$\mathbb{E}_{q(\boldsymbol{x})}[\boldsymbol{x}], \mathrm{Cov}_{q(\boldsymbol{x})}(\boldsymbol{x})$ are easier to compute $\boxed{\cancel{\mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}] = \int \boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y})\mathrm{d}\boldsymbol{x}}}$ $\boxed{\cancel{\mathrm{Cov}_{p(\boldsymbol{x}|\boldsymbol{y})}(\boldsymbol{x}) = \int (\boldsymbol{x}-\mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}])(\boldsymbol{x}-\mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}])^T p(\boldsymbol{x}|\boldsymbol{y})\mathrm{d}\boldsymbol{x}}}$

- **Approximate Bayesian inference:** $\mathbb{E}_{q(\boldsymbol{x})}[\boldsymbol{x}] \approx \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y})}[\boldsymbol{x}] \qquad \mathrm{Cov}_{q(\boldsymbol{x})}(\boldsymbol{x}) \approx \mathrm{Cov}_{p(\boldsymbol{x}|\boldsymbol{y})}(\boldsymbol{x})$

Now forget about the high-dimensional $\boldsymbol{x} = [\, x_1, \cdots, x_N \,]^T$

- **Expectation Propagation (EP):**

- **Expectation Propagation (EP):**

362 MINKA UAI 2001

Expectation Propagation for Approximate Bayesian Inference

Thomas P. Minka
Statistics Dept.
Carnegie Mellon University
Pittsburgh, PA 15213

- **Expectation Propagation (EP):**

362 MINKA UAI 2001

Expectation Propagation for Approximate Bayesian Inference

Thomas P. Minka
Statistics Dept.
Carnegie Mellon University
Pittsburgh, PA 15213

$$q(x) = \arg\min_{q(x)} KL\left(p(x) || q(x)\right)$$

- **Expectation Propagation (EP):**

362 MINKA UAI 2001

Expectation Propagation for Approximate Bayesian Inference

Thomas P. Minka
Statistics Dept.
Carnegie Mellon University
Pittsburgh, PA 15213

$$q(x) = \arg\min_{q(x)} KL\left(p(x) \| q(x)\right)$$

- **Similarity** between **variational Bayes (VB)** and **EP:**

- **Expectation Propagation (EP):**

362 MINKA UAI 2001

Expectation Propagation for Approximate Bayesian Inference

Thomas P. Minka
Statistics Dept.
Carnegie Mellon University
Pittsburgh, PA 15213

$$q(x) = \arg\min_{q(x)} KL\left(p(x) || q(x)\right)$$

- **Similarity** between **variational Bayes (VB)** and **EP:** $q(x) \overset{KL \text{ divergence minimization}}{\approx} p(x)$

- **Expectation Propagation (EP):**

362 MINKA UAI 2001

Expectation Propagation for Approximate Bayesian Inference

Thomas P. Minka
Statistics Dept.
Carnegie Mellon University
Pittsburgh, PA 15213

$$q(x) = \arg\min_{q(x)} KL\left(p(x) \| q(x)\right)$$

- **Similarity** between **variational Bayes (VB)** and **EP:** $q(x) \overset{KL\ \text{divergence minimization}}{\approx} p(x)$

- **Difference** between **variational Bayes (VB)** and **EP :**

- **Expectation Propagation (EP):**

362 MINKA UAI 2001

Expectation Propagation for Approximate Bayesian Inference

Thomas P. Minka
Statistics Dept.
Carnegie Mellon University
Pittsburgh, PA 15213

$$q(x) = \arg\min_{q(x)} KL\left(p(x)||q(x)\right)$$

- **Similarity** between **variational Bayes (VB)** and **EP:** $q(x) \overset{KL \text{ divergence minimization}}{\approx} p(x)$

- **Difference** between **variational Bayes (VB)** and **EP :** asymmetry: $KL(q||p) \neq KL(p||q)$

- **Expectation Propagation (EP):**

362 MINKA UAI 2001

Expectation Propagation for Approximate Bayesian Inference

Thomas P. Minka
Statistics Dept.
Carnegie Mellon University
Pittsburgh, PA 15213

$$q(x) = \arg\min_{q(x)} KL\left(p(x) || q(x)\right)$$

- **Similarity** between **variational Bayes (VB)** and **EP:** $q(x) \overset{KL \text{ divergence minimization}}{\approx} p(x)$

- **Difference** between **variational Bayes (VB)** and **EP :** asymmetry: $KL(q||p) \neq KL(p||q)$

  - VB: $KL\left(q(x) || p(x)\right)$

- **Expectation Propagation (EP):**

362 MINKA UAI 2001

Expectation Propagation for Approximate Bayesian Inference

Thomas P. Minka
Statistics Dept.
Carnegie Mellon University
Pittsburgh, PA 15213

$$q(x) = \arg\min_{q(x)} KL\left(p(x)||q(x)\right)$$

- **Similarity** between **variational Bayes (VB)** and **EP:** $q(x) \overset{KL\ \text{divergence minimization}}{\approx} p(x)$

- **Difference** between **variational Bayes (VB)** and **EP :** asymmetry: $KL(q||p) \neq KL(p||q)$

  - VB: $KL\left(q(x)||p(x)\right)$
  - EP: $KL\left(p(x)||q(x)\right)$

- **Expectation Propagation (EP):**

362 MINKA UAI 2001

Expectation Propagation for Approximate Bayesian Inference

Thomas P. Minka
Statistics Dept.
Carnegie Mellon University
Pittsburgh, PA 15213

$$q(x) = \underset{q(x)}{\arg\min} KL\left(p(x)||q(x)\right)$$

- **Similarity** between **variational Bayes (VB)** and **EP:** $q(x) \overset{KL \text{ divergence minimization}}{\approx} p(x)$

- **Difference** between **variational Bayes (VB)** and **EP :** asymmetry: $KL(q||p) \neq KL(p||q)$

  - ❑ VB: $KL\left(q(x)||p(x)\right)$
  - ❑ EP: $KL\left(p(x)||q(x)\right)$

  ⟶ different $q(x)$

- **KL divergence minimization in EP** $q(x) = \underset{q(x) \in \mathcal{Q}}{\arg\min} KL(p(x) || q(x))$

- **KL divergence minimization in EP** $q(x) = \arg\min_{q(x) \in \mathcal{Q}} KL\left(p(x) || q(x)\right)$

$$\mathcal{Q} = \{q : q(x) = \boldsymbol{e}^{T(x)^T \boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}\}$$

- **KL divergence minimization in EP** $q(x) = \underset{q(x)\in\mathcal{Q}}{\arg\min}\, KL(p(x)||q(x))$

$$\mathcal{Q} = \{q : q(x) = \boldsymbol{e}^{T(x)^T\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}\}$$

- **Exponential family: nice property**

- **KL divergence minimization in EP** $q(x) = \underset{q(x) \in \mathcal{Q}}{\arg\min} KL(p(x)||q(x))$

$$\mathcal{Q} = \{q : q(x) = \boldsymbol{e}^{T(x)^T \boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}\}$$

- **Exponential family: nice property**

$$KL(p(x)||q(x)) = \int p(x) \log \frac{p(x)}{q(x)} \mathrm{d}x$$

$$= \int p(x) \log \frac{p(x)}{\boldsymbol{e}^{T(x)^T \boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}} \mathrm{d}x$$

- **KL divergence minimization in EP** $q(x) = \underset{q(x)\in\mathcal{Q}}{\arg\min} KL(p(x)||q(x))$

$$\mathcal{Q} = \{q: q(x) = \boldsymbol{e}^{T(x)^T\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}\}$$

- **Exponential family: nice property**

$$KL(p(x)||q(x)) = \int p(x)\log\frac{p(x)}{q(x)}\mathrm{d}x$$

$$= \int p(x)\log\frac{p(x)}{\boldsymbol{e}^{T(x)^T\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}}\mathrm{d}x$$

$$\Longrightarrow \frac{\partial A(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}} = \mathbb{E}_{p(x)}[T(x)]$$

- **KL divergence minimization in EP** $q(x) = \underset{q(x)\in\mathcal{Q}}{\arg\min} KL(p(x)||q(x))$

$$\mathcal{Q} = \{q : q(x) = \boldsymbol{e}^{T(x)^T\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}\}$$

- **Exponential family: nice property**

$$KL(p(x)||q(x)) = \int p(x)\log\frac{p(x)}{q(x)}\mathrm{d}x$$

$$= \int p(x)\log\frac{p(x)}{\boldsymbol{e}^{T(x)^T\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}}\mathrm{d}x$$

$$\Longrightarrow \frac{\partial A(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}} = \mathbb{E}_{p(x)}[T(x)]$$ **moment matching**

$$q(x) \in \mathcal{Q} = \{q : q(x) = \boldsymbol{e}^{T(x)^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}\}$$

e.g. univaritate Gaussian distribution

$$q(x) = \mathcal{N}(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$q(x) \in \mathcal{Q} = \{q : q(x) = \boldsymbol{e}^{T(x)^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}\}$$

e.g. univaritate Gaussian distribution

$$q(x) = \mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$= e^{-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2} + \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)}$$

$$q(x) \in \mathcal{Q} = \{q : q(x) = \boldsymbol{e}^{T(x)^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}\}$$

e.g. univaritate Gaussian distribution

$$q(x) = \mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$= e^{-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2} + \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)}$$

$$= e^{\begin{bmatrix} x \\ x^2 \end{bmatrix}^T \begin{bmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{bmatrix} - \left[-\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2)\right] - \frac{1}{2}\log(2\pi)}$$

$$q(x) \in \mathcal{Q} = \{q : q(x) = \boldsymbol{e}^{T(x)^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}\}$$

e.g. univaritate Gaussian distribution

$$q(x) = \mathcal{N}(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$= e^{-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2} + \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)}$$

$$= e^{\begin{bmatrix} x \\ x^2 \end{bmatrix}^T \begin{bmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{bmatrix} - \left[-\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2)\right] - \frac{1}{2}\log(2\pi)}$$

$$T(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix} \quad \boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{bmatrix} \quad A(\boldsymbol{\eta}) = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2) \quad B(x) = -\frac{1}{2}\log(2\pi)$$

$$q(x) \in \mathcal{Q} = \{q : q(x) = \boldsymbol{e}^{T(x)^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}\}$$

e.g. univaritate Gaussian distribution

$$q(x) = \mathcal{N}(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$= e^{-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2} + \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)}$$

$$= e^{\begin{bmatrix} x \\ x^2 \end{bmatrix}^T \begin{bmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{bmatrix} - \left[-\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2)\right] - \frac{1}{2}\log(2\pi)}$$

$$T(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix} \quad \boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{bmatrix} \quad A(\boldsymbol{\eta}) = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2) \quad B(x) = -\frac{1}{2}\log(2\pi)$$

moment matching $\frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbb{E}_{p(x)}[T(x)]$

$$q(x) \in \mathcal{Q} = \{q : q(x) = \boldsymbol{e}^{T(x)^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}\}$$

e.g. univaritate Gaussian distribution

$$q(x) = \mathcal{N}(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$= e^{-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2} + \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)}$$

$$= e^{\begin{bmatrix} x \\ x^2 \end{bmatrix}^T \begin{bmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{bmatrix} - \left[-\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2)\right] - \frac{1}{2}\log(2\pi)}$$

$$T(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix} \quad \boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{bmatrix} \quad A(\boldsymbol{\eta}) = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2) \quad B(x) = -\frac{1}{2}\log(2\pi)$$

moment matching $\frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbb{E}_{p(x)}[T(x)]$

$$\begin{bmatrix} \frac{\partial A(\boldsymbol{\eta})}{\partial \eta_1} \\ \frac{\partial A(\boldsymbol{\eta})}{\partial \eta_2} \end{bmatrix} = \begin{bmatrix} \mu \\ \mu^2 + \sigma^2 \end{bmatrix}$$

$$q(x) \in \mathcal{Q} = \{q : q(x) = \boldsymbol{e}^{T(x)^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}\}$$

e.g. univaritate Gaussian distribution

$$q(x) = \mathcal{N}(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$= e^{-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2} + \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)}$$

$$= e^{\begin{bmatrix} x \\ x^2 \end{bmatrix}^T \begin{bmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{bmatrix} - \left[-\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2)\right] - \frac{1}{2}\log(2\pi)}$$

$$T(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix} \quad \boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{bmatrix} \quad A(\boldsymbol{\eta}) = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2) \quad B(x) = -\frac{1}{2}\log(2\pi)$$

moment matching $\frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbb{E}_{p(x)}[T(x)]$

$$\begin{bmatrix} \frac{\partial A(\boldsymbol{\eta})}{\partial \eta_1} \\ \frac{\partial A(\boldsymbol{\eta})}{\partial \eta_2} \end{bmatrix} = \begin{bmatrix} \mu \\ \mu^2 + \sigma^2 \end{bmatrix} \quad \Longrightarrow \mathbb{E}_{p(x)}\begin{bmatrix} x \\ x^2 \end{bmatrix} = \begin{bmatrix} \mu \\ \mu^2 + \sigma^2 \end{bmatrix}$$

2. Solution – EP for approximate Bayesian inference

- basic idea
- KL divergence minimization
- **factor graph**

- **Goal of EP:** $q(x) \approx p(x)$

- **Goal of EP:** $q(x) \approx p(x)$ $\quad q(x) = \arg \min_{q(x) \in \mathcal{Q}} KL(p(x)||q(x))$

- **Goal of EP:** $q(x) \approx p(x)$ $\quad q(x) = \underset{q(x) \in \mathcal{Q}}{\arg\min} KL\left(p(x) \| q(x)\right)$ $\quad q(x) = \boldsymbol{e}^{T(x)^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}$

- **Goal of EP:** $q(x) \approx p(x)$ $\quad q(x) = \underset{q(x) \in \mathcal{Q}}{\arg\min} KL\left(p(x) \| q(x)\right)$ $\quad q(x) = \boldsymbol{e}^{T(x)^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}$ $\quad \frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbb{E}_{p(x)}[T(x)]$

- **Goal of EP:** $q(x) \approx p(x)$ $\quad q(x) = \underset{q(x) \in \mathcal{Q}}{\arg\min} KL\left(p(x) \| q(x)\right)$ $\quad q(x) = \boldsymbol{e}^{T(x)^{\mathrm{T}} \boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}$ $\quad \dfrac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbb{E}_{p(x)}[T(x)]$

! $\mathbb{E}_{p(x)}[T(x)]$ **intractable**

- **Goal of EP:** $q(x) \approx p(x)$ $\quad q(x) = \underset{q(x) \in \mathcal{Q}}{\arg\min} KL(p(x)||q(x))$ $\quad q(x) = \boldsymbol{e}^{T(x)^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}$ $\quad \frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbb{E}_{p(x)}[T(x)]$

$$! \ \mathbb{E}_{p(x)}[T(x)] \ \textbf{intractable}$$

- **EP on a factor graph**

- **Goal of EP:** $q(x) \approx p(x)$ $\quad q(x) = \underset{q(x)\in\mathcal{Q}}{\arg\min}\, KL\left(p(x) \| q(x)\right)$ $\quad q(x) = \boldsymbol{e}^{T(x)^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}$ $\quad \frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbb{E}_{p(x)}[T(x)]$

$$!\ \mathbb{E}_{p(x)}[T(x)]\ \textbf{intractable}$$

- **EP on a factor graph** $\textbf{difficult} \ » \ \{\textbf{simple} + \ldots + \textbf{simple}\}$

- **Goal of EP:** $q(x) \approx p(x)$ $\quad q(x) = \underset{q(x)\in\mathcal{Q}}{\arg\min} KL\left(p(x)||q(x)\right)$ $\quad q(x) = \boldsymbol{e}^{T(x)^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}$ $\quad \frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbb{E}_{p(x)}[T(x)]$

$$! \ \mathbb{E}_{p(x)}[T(x)] \ \textbf{intractable}$$

- **EP on a factor graph** $\textbf{difficult} \ » \ \{\textbf{simple} + \ldots + \textbf{simple}\}$

$$p(x) = p_1(x)\, p_2(x) \ldots$$

- **Goal of EP:** $q(x) \approx p(x)$ $\quad q(x) = \underset{q(x)\in\mathcal{Q}}{\arg\min} KL(p(x)||q(x))$ $\quad q(x) = e^{T(x)^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}$ $\quad \frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbb{E}_{p(x)}[T(x)]$

$$! \; \mathbb{E}_{p(x)}[T(x)] \; \textbf{intractable}$$

- **EP on a factor graph** **difficult** » $\{$**simple** $+ \ldots +$ **simple**$\}$

$$p(x) = p_1(x)\, p_2(x) \ldots$$
$$\quad\quad \wr\wr \quad\quad \wr\wr$$

- **Goal of EP:** $q(x) \approx p(x)$ $\quad q(x) = \underset{q(x)\in\mathcal{Q}}{\arg\min}\, KL(p(x)||q(x))$ $\quad q(x) = e^{T(x)^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}$ $\quad \frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbb{E}_{p(x)}[T(x)]$

  $! \ \mathbb{E}_{p(x)}[T(x)]$ **intractable**

- **EP on a factor graph** **difficult** » $\{$**simple** $+ \ldots +$ **simple**$\}$

$$\begin{array}{ccc} p(x) = & p_1(x) & p_2(x)\ldots \\ & \wr\wr & \wr\wr \\ & q_1(x) & q_2(x)\ldots \end{array}$$

- **Goal of EP:** $q(x) \approx p(x)$ $\quad q(x) = \underset{q(x) \in \mathcal{Q}}{\arg\min} KL(p(x) || q(x))$ $\quad q(x) = e^{T(x)^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}$ $\quad \frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbb{E}_{p(x)}[T(x)]$

$$! \ \mathbb{E}_{p(x)}[T(x)] \ \textbf{intractable}$$

- **EP on a factor graph** $\textbf{difficult} \gg \{\textbf{simple} + \ldots + \textbf{simple}\}$

$$\begin{aligned} p(x) = \ & p_1(x) \, p_2(x) \ldots \\ & \ \ \wr\wr \qquad \wr\wr \\ & q_1(x) \, q_2(x) \ldots = q(x) \end{aligned}$$

- **Goal of EP:** $q(x) \approx p(x)$ $\quad q(x) = \underset{q(x) \in \mathcal{Q}}{\arg\min} KL(p(x) || q(x))$ $\quad q(x) = e^{T(x)^{\mathrm{T}} \boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}$ $\quad \frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbb{E}_{p(x)}[T(x)]$

$$! \ \mathbb{E}_{p(x)}[T(x)] \ \textbf{intractable}$$

- **EP on a factor graph** $\textbf{difficult} \ \gg \ \{\textbf{simple} + \ldots + \textbf{simple}\}$

$$\begin{array}{ccc} p(x) = & p_1(x) & p_2(x) \ldots \\ & \wr\wr & \wr\wr \\ & q_1(x) & q_2(x) \ldots \quad = q(x) \end{array}$$

- **a function can be expressed as product of local functions (factors) over a subset of variables**

- **Goal of EP:** $q(x) \approx p(x)$ $\quad q(x) = \underset{q(x) \in \mathcal{Q}}{\arg\min} KL(p(x)||q(x))$ $\quad q(x) = e^{T(x)^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}$ $\quad \frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbb{E}_{p(x)}[T(x)]$

$$! \ \mathbb{E}_{p(x)}[T(x)] \ \textbf{intractable}$$

- **EP on a factor graph** $\textbf{difficult} \ \gg \ \{\textbf{simple} + \ldots + \textbf{simple}\}$

$$p(x) = p_1(x)\, p_2(x) \ldots$$
$$\quad\ \ \wr\wr \qquad \wr\wr$$
$$q_1(x)\, q_2(x) \ldots \ = q(x)$$

- **a function can be expressed as product of local functions (factors) over a subset of variables**

$$f(a,b,c) = f_1(a,b)\, f_2(a,c)\, f_3(b,c)$$

2. Solution – EP for approximate Bayesian inference

- basic idea
- KL divergence minimization
- **factor graph**

- **Goal of EP:** $q(x) \approx p(x)$ $\quad q(x) = \underset{q(x)\in\mathcal{Q}}{\arg\min} KL(p(x)||q(x))$ $\quad q(x) = e^{T(x)^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}$ $\quad \frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbb{E}_{p(x)}[T(x)]$

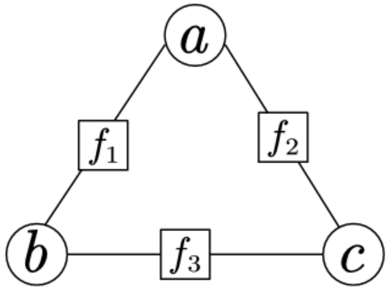
$$! \ \mathbb{E}_{p(x)}[T(x)] \ \textbf{intractable}$$

- **EP on a factor graph** **difficult** » $\{$**simple** $+ \ldots +$ **simple**$\}$

$$\begin{array}{l} p(x) = p_1(x)\, p_2(x) \ldots \\ \qquad\quad \wr\wr \qquad \wr\wr \\ \qquad\quad q_1(x)\, q_2(x) \ldots \quad = q(x) \end{array}$$

- **a function can be expressed as product of local functions (factors) over a subset of variables**

$$f(a,b,c) = f_1(a,b)\, f_2(a,c)\, f_3(b,c)$$



- **Goal of EP:** $q(x) \approx p(x)$ $\quad q(x) = \underset{q(x)\in\mathcal{Q}}{\arg\min} KL(p(x)||q(x))$ $\quad q(x) = e^{T(x)^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}$ $\quad \frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbb{E}_{p(x)}[T(x)]$

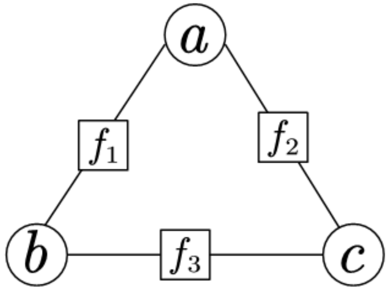
$$! \ \mathbb{E}_{p(x)}[T(x)] \ \textbf{intractable}$$

- **EP on a factor graph** $\textbf{difficult} \gg \{\textbf{simple} + \ldots + \textbf{simple}\}$

$$\begin{array}{l} p(x) = p_1(x)\,p_2(x)\ldots \\ \qquad\quad\ \ \wr\wr \qquad \wr\wr \\ \qquad\quad q_1(x)\,q_2(x)\ldots \ = q(x) \end{array}$$

- **a function can be expressed as product of local functions (factors) over a subset of variables**

$$f(a,b,c) = f_1(a,b)\,f_2(a,c)\,f_3(b,c) \qquad\qquad p(x) \propto f(y|x)\,f(x)$$



- **Goal of EP:** $q(x) \approx p(x)$ $\quad q(x) = \underset{q(x)\in\mathcal{Q}}{\arg\min} KL(p(x)||q(x))$ $\quad q(x) = e^{T(x)^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}$ $\quad \frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbb{E}_{p(x)}[T(x)]$

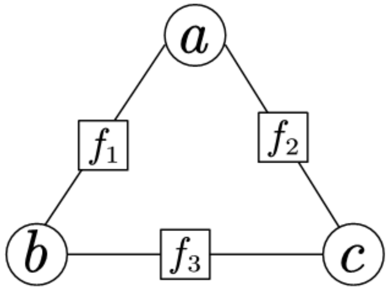
$$! \ \mathbb{E}_{p(x)}[T(x)] \ \textbf{intractable}$$

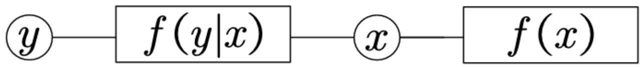
- **EP on a factor graph** $\textbf{difficult} \ \gg \ \{\textbf{simple} + \ldots + \textbf{simple}\}$

$$p(x) = p_1(x)\, p_2(x) \ldots$$
$$\wr\wr \qquad \wr\wr$$
$$q_1(x)\, q_2(x) \ldots \ = q(x)$$

- **a function can be expressed as product of local functions (factors) over a subset of variables**

$$f(a,b,c) = f_1(a,b)\, f_2(a,c)\, f_3(b,c) \qquad\qquad p(x) \propto f(y|x)\, f(x)$$





- **Goal of EP:** $q(x) \approx p(x)$ $\quad q(x) = \underset{q(x) \in \mathcal{Q}}{\arg\min} KL\left(p(x) \| q(x)\right)$ $\quad q(x) = e^{T(x)^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}$ $\quad \dfrac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbb{E}_{p(x)}[T(x)]$

$$! \ \mathbb{E}_{p(x)}[T(x)] \ \textbf{intractable}$$

- **EP on a factor graph** difficult » $\{\textbf{simple} + \ldots + \textbf{simple}\}$

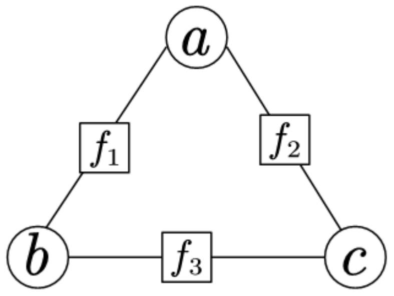
$$p(x) = p_1(x)\, p_2(x) \ldots$$

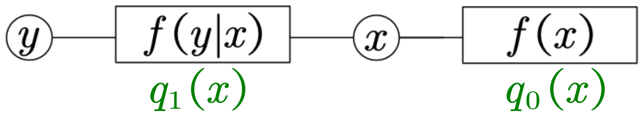
$$\begin{array}{cc} \wr\wr & \wr\wr \end{array}$$

$$q_1(x)\, q_2(x) \ldots = q(x)$$

- **a function can be expressed as product of local functions (factors) over a subset of variables**

$$f(a,b,c) = f_1(a,b)\, f_2(a,c)\, f_3(b,c) \qquad p(x) \propto f(y|x)\, f(x)$$





- **Goal of EP:** $q(x) \approx p(x)$ $\quad q(x) = \underset{q(x)\in\mathcal{Q}}{\arg\min} KL(p(x)||q(x))$ $\quad q(x) = \boldsymbol{e}^{T(x)^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(x)}$ $\quad \frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbb{E}_{p(x)}[T(x)]$

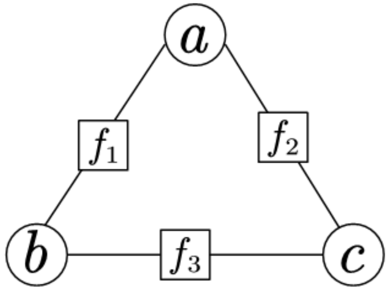
  ! $\mathbb{E}_{p(x)}[T(x)]$ **intractable**

- **EP on a factor graph** **difficult** » $\{$**simple** $+ \ldots +$ **simple**$\}$

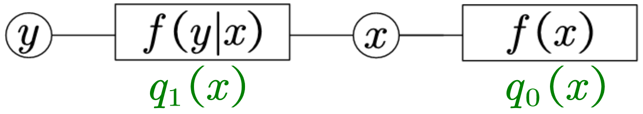
$$\begin{array}{ccc} p(x) = & p_1(x) & p_2(x)\ldots \\ & \wr\wr & \wr\wr \\ & q_1(x) & q_2(x)\ldots \quad = q(x) \end{array}$$

- **a function can be expressed as product of local functions (factors) over a subset of variables**

$$f(a,b,c) = f_1(a,b)\,f_2(a,c)\,f_3(b,c)$$



$$p(x) \propto f(y|x)\,f(x)$$



$$\Longrightarrow q(x) \propto q_1(x)\,q_0(x)$$

- **Approximate Bayesian inference by EP:**

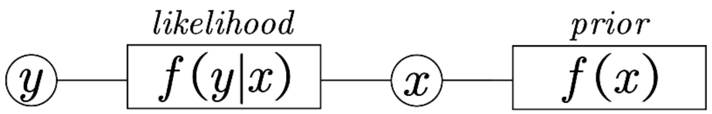
- **Approximate Bayesian inference by EP:**

exact posterior: $$p(x|y) = \frac{f(y|x)\, f(x)}{\int f(y|x)\, f(x)\,\mathrm{d}x}$$

- **Approximate Bayesian inference by EP:**

exact posterior: $$p(x|y) = \frac{f(y|x)\,f(x)}{\int f(y|x)\,f(x)\,\mathrm{d}x}$$

factor graph representation:

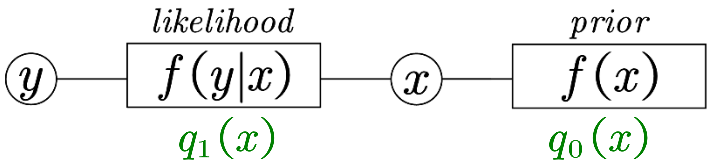


- **Approximate Bayesian inference by EP:**

exact posterior:

$$p\left(x|y\right) = \frac{f\left(y|x\right) f\left(x\right)}{\int f\left(y|x\right) f\left(x\right) \mathrm{d}x}$$

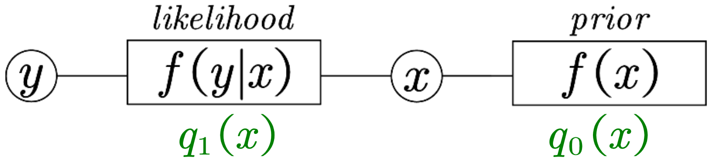
factor graph representation:



- **Approximate Bayesian inference by EP:**

exact posterior: $$p(x|y) = \frac{f(y|x)\,f(x)}{\int f(y|x)\,f(x)\,\mathrm{d}x}$$

factor graph representation:

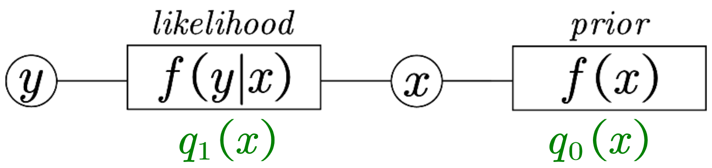


EP sequential KL divergence minimization:

- **Approximate Bayesian inference by EP:**

exact posterior: $$p(x|y) = \frac{f(y|x)\,f(x)}{\int f(y|x)\,f(x)\,\mathrm{d}x}$$

factor graph representation:



EP sequential KL divergence minimization:

$$\begin{cases} q_1(x) = \arg\min\limits_{q_1(x)\in\mathcal{Q}} KL\left(f(y|x)\,q_0(x)\,||\,q(x)\right) \\ q_0(x) = \arg\min\limits_{q_0(x)\in\mathcal{Q}} KL\left(f(x)\,q_1(x)\,||\,q(x)\right) \end{cases}$$

...until convergence

- **Approximate Bayesian inference by EP:**

exact posterior: $$p(x|y) = \frac{f(y|x)\,f(x)}{\int f(y|x)\,f(x)\,\mathrm{d}x}$$

factor graph representation: $$(y) - \overset{\textit{likelihood}}{\boxed{f(y|x)}} - (x) - \overset{\textit{prior}}{\boxed{f(x)}}$$

$$q_1(x) \qquad\qquad q_0(x)$$

EP sequential KL divergence minimization: $$\begin{cases} q_1(x) = \underset{q_1(x)\in\mathcal{Q}}{\arg\min} KL\left(f(y|x)\,q_0(x)\,||\,q(x)\right) \\ q_0(x) = \underset{q_0(x)\in\mathcal{Q}}{\arg\min} KL\left(f(x)\,q_1(x)\,||\,q(x)\right) \end{cases}$$

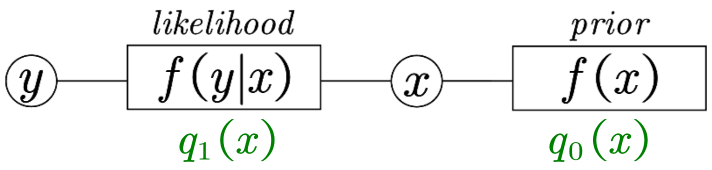
...until convergence

$$\text{tilted distribution: exact factor} \times \text{cavity}$$

- **Approximate Bayesian inference by EP:**

exact posterior: $$p(x|y) = \frac{f(y|x)\,f(x)}{\int f(y|x)\,f(x)\,\mathrm{d}x}$$

factor graph representation:



EP sequential KL divergence minimization:
$$\begin{cases} q_1(x) = \underset{q_1(x)\in\mathcal{Q}}{\arg\min}\, KL\left(f(y|x)\,q_0(x)\,||\,q(x)\right) \\ q_0(x) = \underset{q_0(x)\in\mathcal{Q}}{\arg\min}\, KL\left(f(x)\,q_1(x)\,||\,q(x)\right) \end{cases}$$

...until convergence

$$\text{tilted distribution: exact factor} \times \text{cavity} \qquad \text{cavity: } q_1(x) = \frac{q(x)}{q_0(x)},\; q_0(x) = \frac{q(x)}{q_1(x)}$$

- **Approximate Bayesian inference by EP:**

exact posterior: $$p(x|y) = \frac{f(y|x)\,f(x)}{\int f(y|x)\,f(x)\,\mathrm{d}x}$$

factor graph representation: $$(y) - \underset{q_1(x)}{\boxed{\overset{likelihood}{f(y|x)}}} - (x) - \underset{q_0(x)}{\boxed{\overset{prior}{f(x)}}}$$

EP sequential KL divergence minimization: $$\begin{cases} q_1(x) = \arg\min\limits_{q_1(x)\in\mathcal{Q}} KL\left(f(y|x)\,q_0(x)\,||\,q(x)\right) \\ q_0(x) = \arg\min\limits_{q_0(x)\in\mathcal{Q}} KL\left(f(x)\,q_1(x)\,||\,q(x)\right) \end{cases}$$

...until convergence

$$\text{tilted distribution: exact factor} \times \text{cavity} \qquad \text{cavity: } q_1(x) = \frac{q(x)}{q_0(x)},\ q_0(x) = \frac{q(x)}{q_1(x)}$$

$$\frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbb{E}_{p(x)}\left[T(x)\right]$$

- **Approximate Bayesian inference by EP:**

exact posterior: $$p(x|y) = \frac{f(y|x)\,f(x)}{\int f(y|x)\,f(x)\,\mathrm{d}x}$$

factor graph representation: $$y - \underset{q_1(x)}{\overset{\textit{likelihood}}{\boxed{f(y|x)}}} - x - \underset{q_0(x)}{\overset{\textit{prior}}{\boxed{f(x)}}}$$

EP sequential KL divergence minimization: $$\begin{cases} q_1(x) = \underset{q_1(x)\in\mathcal{Q}}{\arg\min} KL\left(f(y|x)q_0(x)\|q(x)\right) \\ q_0(x) = \underset{q_0(x)\in\mathcal{Q}}{\arg\min} KL\left(f(x)q_1(x)\|q(x)\right) \end{cases}$$

...until convergence

$$\text{tilted distribution: exact factor} \times \text{cavity} \qquad \text{cavity: } q_1(x) = \frac{q(x)}{q_0(x)},\ q_0(x) = \frac{q(x)}{q_1(x)}$$

$$\frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \cancel{\mathbb{E}_{p(x)}}[T(x)]$$

- **Approximate Bayesian inference by EP:**

exact posterior: $$p(x|y) = \frac{f(y|x)\,f(x)}{\int f(y|x)\,f(x)\,\mathrm{d}x}$$

factor graph representation:
$$\textcircled{y} \;-\; \underset{q_1(x)}{\overset{\textit{likelihood}}{\boxed{f(y|x)}}} \;-\; \textcircled{x} \;-\; \underset{q_0(x)}{\overset{\textit{prior}}{\boxed{f(x)}}}$$

EP sequential KL divergence minimization:
$$\begin{cases} q_1(x) = \underset{q_1(x)\in\mathcal{Q}}{\arg\min} KL\left(f(y|x)\,q_0(x)\,||\,q(x)\right) \\ q_0(x) = \underset{q_0(x)\in\mathcal{Q}}{\arg\min} KL\left(f(x)\,q_1(x)\,||\,q(x)\right) \end{cases}$$
...until convergence

$$\text{tilted distribution: exact factor} \times \text{cavity} \qquad \text{cavity: } q_1(x) = \frac{q(x)}{q_0(x)},\; q_0(x) = \frac{q(x)}{q_1(x)}$$

$$\cancel{\frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbb{E}_{p(x)}[T(x)]} \implies \frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbb{E}_{\textit{titled}}[T(x)]$$

This talk is about

Expectation Propagation

1. Problem formulation and challenges

- Imaging inverse problems
- Bayesian estimation strategy
- challenges

2. Solution – EP for approximate Bayesian inference

- basic idea
- KL divergence minimization
- factor graph

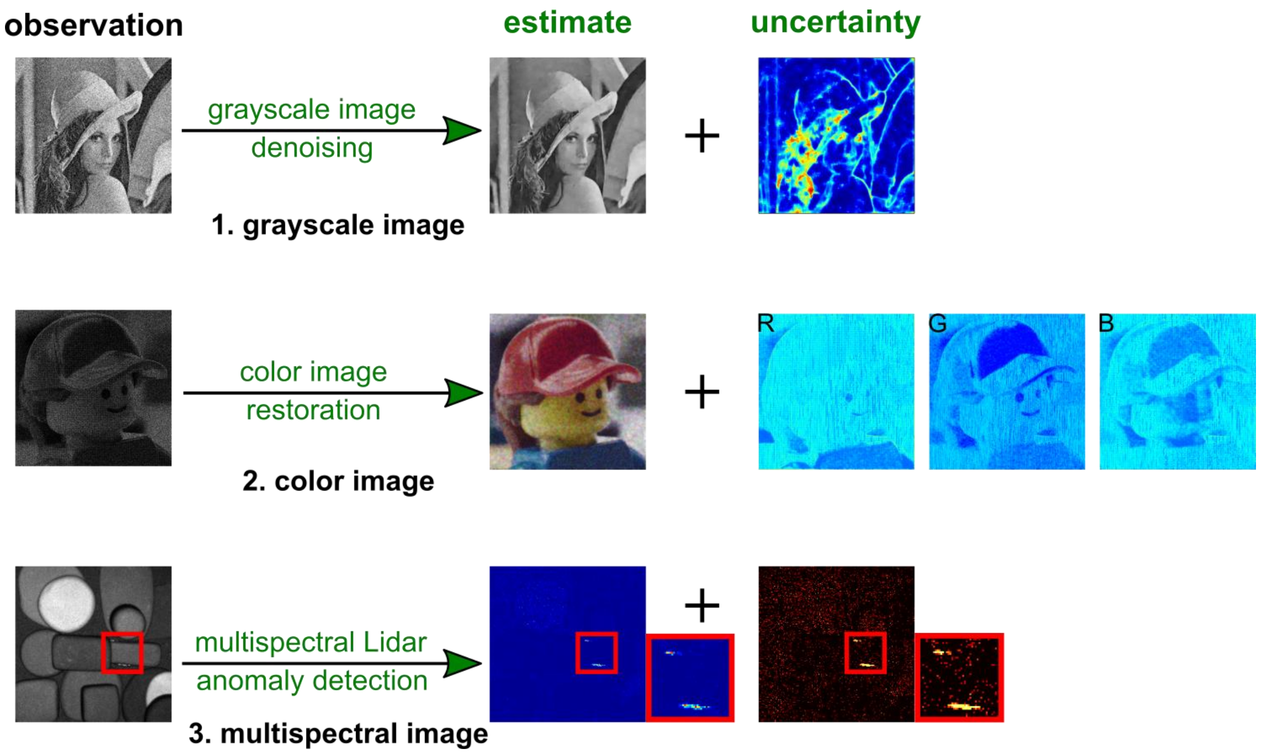
3. Applications – EP for scalable imaging inverse problems

- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- toy examples and applications

4. Conclusion

By the end of this talk,

you will know how to implement your own EP algorithm to:



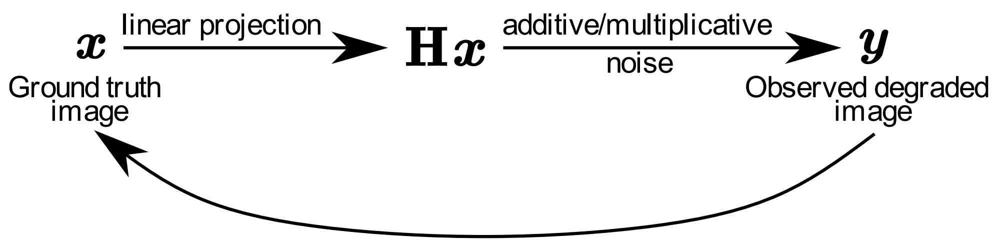
Go back to the high-dimensional $\boldsymbol{x} = [\, x_1, \cdots, x_N \,]^T$

3. Applications – EP for scalable imaging inverse problems

- **how to construct an EP algorithm to solve image inverse problems**
- how to achieve scalable posterior approximation
- toy examples and applications

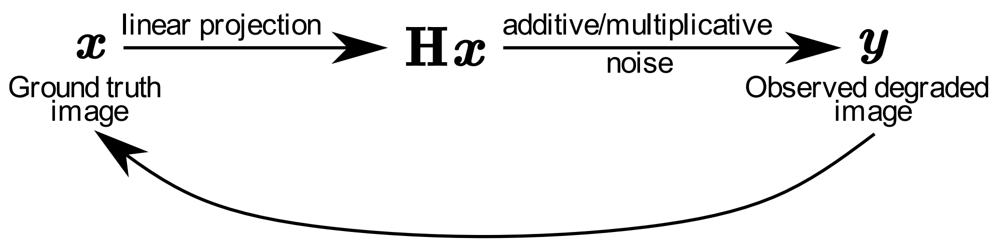
- **how to construct an EP algorithm to solve image inverse problems**
- how to achieve scalable posterior approximation
- toy examples and applications

- **to solve high-dimensional image inverse problems:**



Linear inverse problems in imaging

- **to solve high-dimensional image inverse problems:**

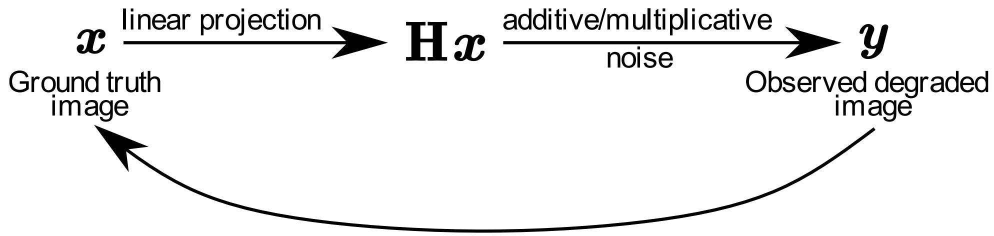


Linear inverse problems in imaging

! intractable high-dimensional posterior in Bayesian estimation : $p(\boldsymbol{x}|\boldsymbol{y}) = \dfrac{f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})}{\int f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x}}$

- **how to construct an EP algorithm to solve image inverse problems**
- how to achieve scalable posterior approximation
- toy examples and applications

- **to solve high-dimensional image inverse problems:**



Linear inverse problems in imaging

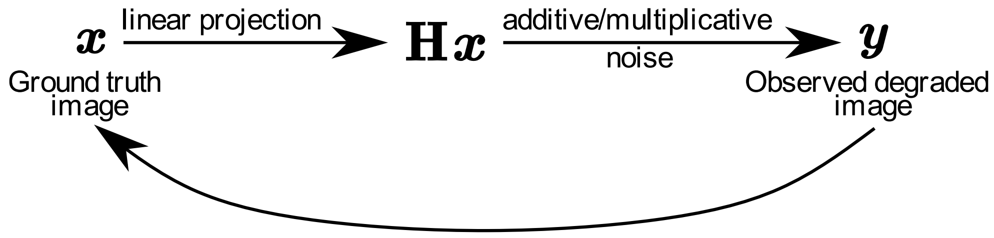
! intractable high-dimensional posterior in Bayesian estimation :
$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})}{\int f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x}}$$

- **approximate Bayesian estimation by EP:**

$$\begin{cases} q_1(\boldsymbol{x}) = \underset{q_1(\boldsymbol{x}) \in \mathcal{Q}}{\arg\min}\ KL\left(f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, q_0(\boldsymbol{x}) || q_1(\boldsymbol{x})\, q_0(\boldsymbol{x})\right) \\ q_0(\boldsymbol{x}) = \underset{q_0(\boldsymbol{x}) \in \mathcal{Q}}{\arg\min}\ KL\left(f_x(\boldsymbol{x}|\boldsymbol{\theta})\, q_1(\boldsymbol{x}) || q_0(\boldsymbol{x})\, q_1(\boldsymbol{x})\right) \\ \vdots \quad = \quad \vdots \end{cases}$$

... until convergence

- **to solve high-dimensional image inverse problems:**



Linear inverse problems in imaging

! intractable high-dimensional posterior in Bayesian estimation : $p(\boldsymbol{x}|\boldsymbol{y}) = \dfrac{f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})}{\int f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x}}$

- **approximate Bayesian estimation by EP:**

$$\begin{cases} q_1(\boldsymbol{x}) = \underset{q_1(\boldsymbol{x}) \in \mathcal{Q}}{\arg\min}\ KL\left(f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, q_0(\boldsymbol{x}) || q_1(\boldsymbol{x})\, q_0(\boldsymbol{x})\right) \\ q_0(\boldsymbol{x}) = \underset{q_0(\boldsymbol{x}) \in \mathcal{Q}}{\arg\min}\ KL\left(f_x(\boldsymbol{x}|\boldsymbol{\theta})\, q_1(\boldsymbol{x}) || q_0(\boldsymbol{x})\, q_1(\boldsymbol{x})\right) \\ \vdots \quad = \quad \vdots \end{cases}$$

… until convergence

$$\Longrightarrow\ q(\boldsymbol{x}) \propto q_1(x)\, q_0(\boldsymbol{x}) \cdots \approx p(\boldsymbol{x}|\boldsymbol{y})$$

3. Applications – EP for scalable imaging inverse problems

- **how to construct an EP algorithm to solve image inverse problems**
- how to achieve scalable posterior approximation
- toy examples and applications

- **how to construct an EP algorithm to solve image inverse problems**
- how to achieve scalable posterior approximation
- toy examples and applications

- **Exponential family :** $\mathcal{Q} = \{q : q(\boldsymbol{x}) = \boldsymbol{e}^{T(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(\boldsymbol{x})}\}$

- **Exponential family :** $\mathcal{Q} = \{q : q(\boldsymbol{x}) = \boldsymbol{e}^{T(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(\boldsymbol{x})}\}$

$q(x)$ univaritate Gaussian distribution

$$q(x) = \mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$T(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix} \quad \boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{bmatrix} \quad A(\boldsymbol{\eta}) = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2)$$

$$\begin{bmatrix} \frac{\partial A(\boldsymbol{\eta})}{\partial \eta_1} \\ \frac{\partial A(\boldsymbol{\eta})}{\partial \eta_2} \end{bmatrix} = \begin{bmatrix} \mu \\ \mu^2 + \sigma^2 \end{bmatrix}$$

- **Exponential family :** $\mathcal{Q} = \{q : q(\boldsymbol{x}) = \boldsymbol{e}^{T(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(\boldsymbol{x})}\}$

$q(x)$ univaritate Gaussian distribution

$$q(x) = \mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$T(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix} \quad \boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{bmatrix} \quad A(\boldsymbol{\eta}) = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2)$$

$$\begin{bmatrix} \frac{\partial A(\boldsymbol{\eta})}{\partial \eta_1} \\ \frac{\partial A(\boldsymbol{\eta})}{\partial \eta_2} \end{bmatrix} = \begin{bmatrix} \mu \\ \mu^2 + \sigma^2 \end{bmatrix}$$

moment matching $\Longrightarrow \begin{bmatrix} \mu \\ \mu^2 + \sigma^2 \end{bmatrix} = \mathbb{E}_{tilted}[T(x)]$

- **how to construct an EP algorithm to solve image inverse problems**
- how to achieve scalable posterior approximation
- toy examples and applications

- **Exponential family :** $\mathcal{Q} = \{q : q(\boldsymbol{x}) = \boldsymbol{e}^{T(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(\boldsymbol{x})}\}$

$q(x)$ univaritate Gaussian distribution

$$q(x) = \mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$T(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix} \quad \boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{bmatrix} \quad A(\boldsymbol{\eta}) = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2)$$

$$\begin{bmatrix} \frac{\partial A(\boldsymbol{\eta})}{\partial \eta_1} \\ \frac{\partial A(\boldsymbol{\eta})}{\partial \eta_2} \end{bmatrix} = \begin{bmatrix} \mu \\ \mu^2 + \sigma^2 \end{bmatrix}$$

moment matching $\Longrightarrow \begin{bmatrix} \mu \\ \mu^2 + \sigma^2 \end{bmatrix} = \mathbb{E}_{tilted}[T(x)]$

$q(\boldsymbol{x})$ multivariate Gaussian distribution

$$q(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{N}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$

- **how to construct an EP algorithm to solve image inverse problems**
- how to achieve scalable posterior approximation
- toy examples and applications

- **Exponential family :** $\mathcal{Q} = \{q : q(\boldsymbol{x}) = \boldsymbol{e}^{T(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(\boldsymbol{x})}\}$

$q(x)$ univaritate Gaussian distribution

$$q(x) = \mathcal{N}(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$T(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix} \quad \boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{bmatrix} \quad A(\boldsymbol{\eta}) = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2)$$

$$\begin{bmatrix} \frac{\partial A(\boldsymbol{\eta})}{\partial \eta_1} \\ \frac{\partial A(\boldsymbol{\eta})}{\partial \eta_2} \end{bmatrix} = \begin{bmatrix} \mu \\ \mu^2 + \sigma^2 \end{bmatrix}$$

moment matching $\Longrightarrow \begin{bmatrix} \mu \\ \mu^2 + \sigma^2 \end{bmatrix} = \mathbb{E}_{tilted}[T(x)]$

$q(\boldsymbol{x})$ multivariate Gaussian distribution

$$q(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = (2\pi)^{-\frac{N}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$

$$T(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{x}\boldsymbol{x}^T \end{bmatrix} \quad \boldsymbol{\eta} = \begin{bmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\ -\frac{1}{2}\boldsymbol{\Sigma}^{-1} \end{bmatrix} \quad A(\boldsymbol{\eta}) = -\frac{1}{4}\boldsymbol{\eta}_1^T\boldsymbol{\eta}_2^{-1}\boldsymbol{\eta}_1 - \frac{1}{2}\log(|-2\boldsymbol{\eta}_2|)$$

- **Exponential family :** $\mathcal{Q} = \{q : q(\boldsymbol{x}) = \boldsymbol{e}^{T(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(\boldsymbol{x})}\}$

$q(x)$ univaritate Gaussian distribution

$$q(x) = \mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$T(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix} \quad \boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{bmatrix} \quad A(\boldsymbol{\eta}) = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2)$$

$$\begin{bmatrix} \frac{\partial A(\boldsymbol{\eta})}{\partial \eta_1} \\ \frac{\partial A(\boldsymbol{\eta})}{\partial \eta_2} \end{bmatrix} = \begin{bmatrix} \mu \\ \mu^2 + \sigma^2 \end{bmatrix}$$

moment matching $\Longrightarrow \begin{bmatrix} \mu \\ \mu^2 + \sigma^2 \end{bmatrix} = \mathbb{E}_{tilted}[T(x)]$

$q(\boldsymbol{x})$ multivariate Gaussian distribution

$$q(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{N}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$

$$T(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{x}\boldsymbol{x}^T \end{bmatrix} \quad \boldsymbol{\eta} = \begin{bmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\ -\frac{1}{2}\boldsymbol{\Sigma}^{-1} \end{bmatrix} \quad A(\boldsymbol{\eta}) = -\frac{1}{4}\boldsymbol{\eta}_1^T \boldsymbol{\eta}_2^{-1} \boldsymbol{\eta}_1 - \frac{1}{2}\log(|-2\boldsymbol{\eta}_2|)$$

$$\begin{bmatrix} \frac{\partial A(\boldsymbol{\eta})}{\partial \eta_1} \\ \frac{\partial A(\boldsymbol{\eta})}{\partial \eta_2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu}\boldsymbol{\mu}^T + \boldsymbol{\Sigma} \end{bmatrix}$$

- **how to construct an EP algorithm to solve image inverse problems**
- how to achieve scalable posterior approximation
- toy examples and applications

- **Exponential family :** $\mathcal{Q} = \{q : q(\boldsymbol{x}) = \boldsymbol{e}^{T(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\eta} - A(\boldsymbol{\eta}) + \boldsymbol{B}(\boldsymbol{x})}\}$

$q(x)$ univaritate Gaussian distribution

$$q(x) = \mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$T(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix} \quad \boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{bmatrix} \quad A(\boldsymbol{\eta}) = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2)$$

$$\begin{bmatrix} \frac{\partial A(\boldsymbol{\eta})}{\partial \eta_1} \\ \frac{\partial A(\boldsymbol{\eta})}{\partial \eta_2} \end{bmatrix} = \begin{bmatrix} \mu \\ \mu^2 + \sigma^2 \end{bmatrix}$$

moment matching $\Longrightarrow \begin{bmatrix} \mu \\ \mu^2 + \sigma^2 \end{bmatrix} = \mathbb{E}_{tilted}[T(x)]$

$q(\boldsymbol{x})$ multivariate Gaussian distribution

$$q(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{N}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})}$$

$$T(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{x}\boldsymbol{x}^T \end{bmatrix} \quad \boldsymbol{\eta} = \begin{bmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\ -\frac{1}{2}\boldsymbol{\Sigma}^{-1} \end{bmatrix} \quad A(\boldsymbol{\eta}) = -\frac{1}{4}\boldsymbol{\eta}_1^T \boldsymbol{\eta}_2^{-1} \boldsymbol{\eta}_1 - \frac{1}{2}\log(|-2\boldsymbol{\eta}_2|)$$

$$\begin{bmatrix} \frac{\partial A(\boldsymbol{\eta})}{\partial \eta_1} \\ \frac{\partial A(\boldsymbol{\eta})}{\partial \eta_2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu}\boldsymbol{\mu}^T + \boldsymbol{\Sigma} \end{bmatrix}$$

moment matching $\Longrightarrow \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu}\boldsymbol{\mu}^T + \boldsymbol{\Sigma} \end{bmatrix} = \mathbb{E}_{tilted}[T(\boldsymbol{x})]$

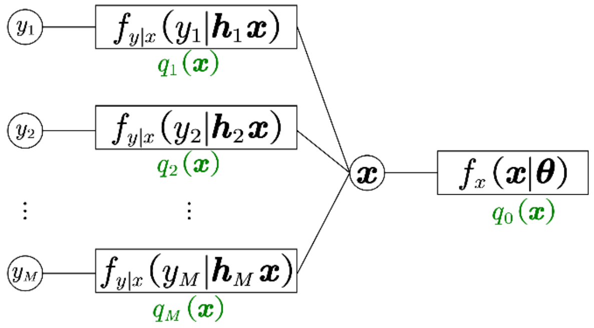
3. Applications – EP for scalable imaging inverse problems

- how to construct an EP algorithm to solve image inverse problems
- **how to achieve scalable posterior approximation**
- toy examples and applications

1. how to factorize?

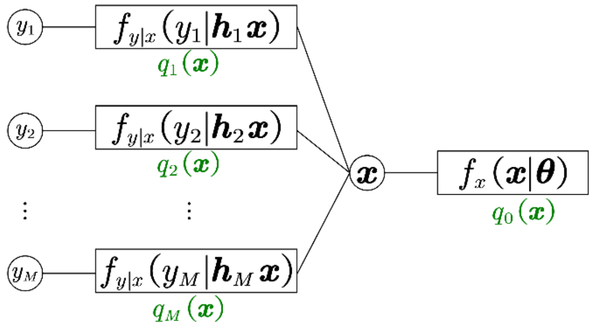
- how to construct an EP algorithm to solve image inverse problems
- **how to achieve scalable posterior approximation**
- toy examples and applications

1. how to factorize?

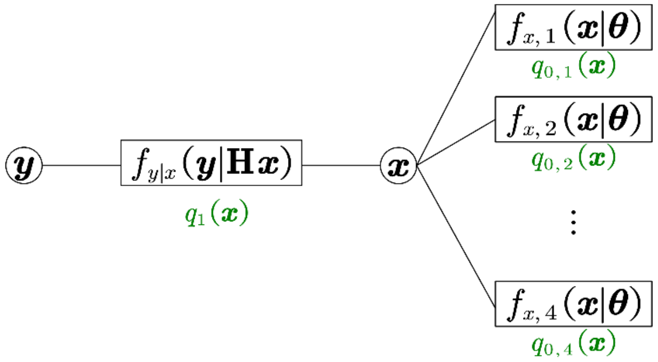


factorize likelihood

1. how to factorize?



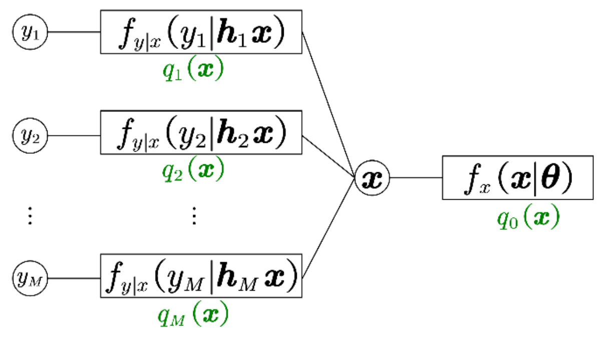
factorize likelihood



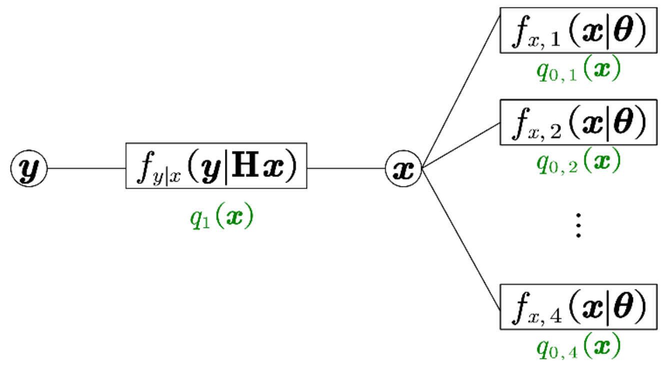
factorize prior

- how to construct an EP algorithm to solve image inverse problems
- **how to achieve scalable posterior approximation**
- toy examples and applications

1. how to factorize?



factorize likelihood



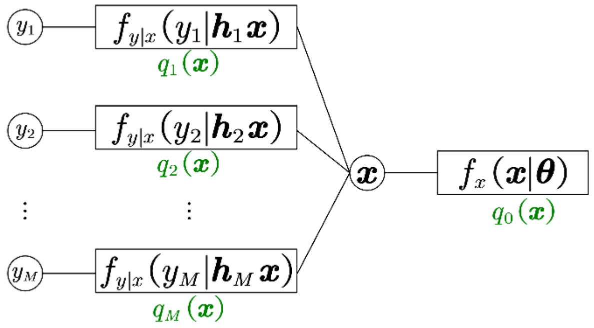
factorize prior

2. how to compute?

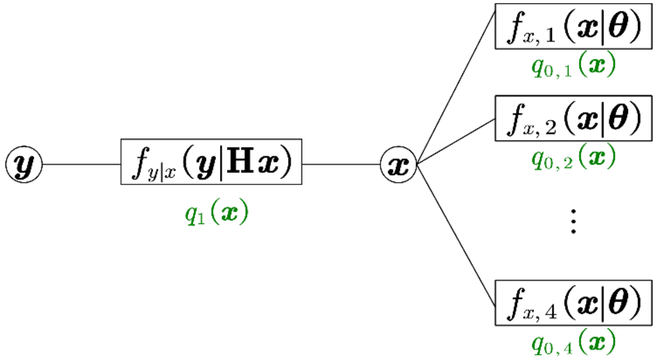
3. Applications – EP for scalable imaging inverse problems

- how to construct an EP algorithm to solve image inverse problems
- **how to achieve scalable posterior approximation**
- toy examples and applications

1. how to factorize?



factorize likelihood



factorize prior

2. how to compute?

$$q_0(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x};\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)$$

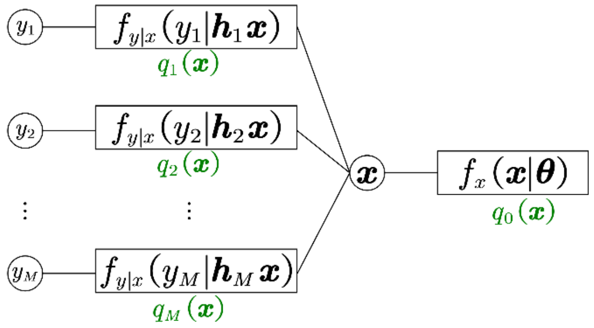
$$q_1(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x};\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1)$$

$$\vdots$$

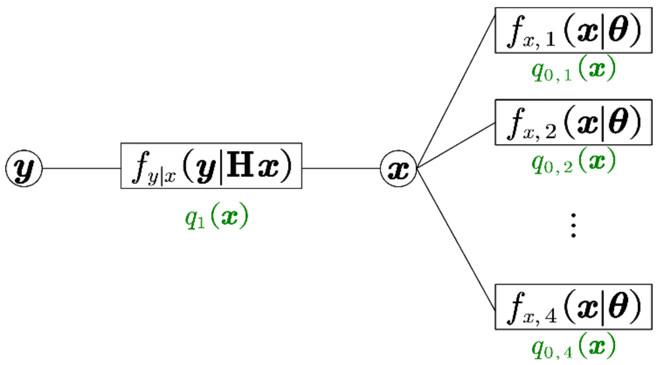
3. Applications – EP for scalable imaging inverse problems

- how to construct an EP algorithm to solve image inverse problems
- **how to achieve scalable posterior approximation**
- toy examples and applications

1. how to factorize?



factorize likelihood



factorize prior

2. how to compute?

$$q_0(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$$
$$q_1(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$$
$$\vdots$$

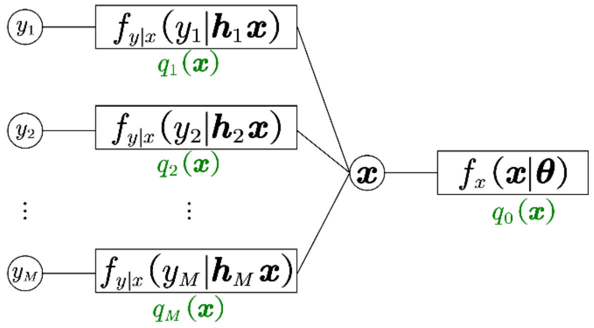
EP update: $q_0(\boldsymbol{x}) \propto \dfrac{q(\boldsymbol{x})}{q_1(\boldsymbol{x})}$ $\quad \boldsymbol{\Sigma}_0 = (\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}_1^{-1})^{-1}$

$q_1(\boldsymbol{x}) \propto \dfrac{q(\boldsymbol{x})}{q_0(\boldsymbol{x})}$ $\quad \boldsymbol{\Sigma}_1 = (\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}_0^{-1})^{-1}$

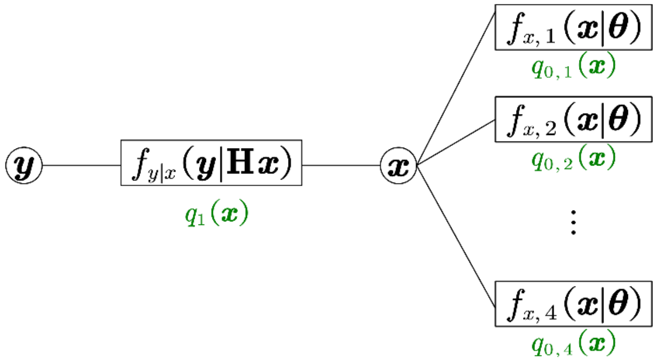
$$\vdots$$

- how to construct an EP algorithm to solve image inverse problems
- **how to achieve scalable posterior approximation**
- toy examples and applications

1. how to factorize?



factorize likelihood



factorize prior

2. how to compute?

$$q_0(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$$
$$q_1(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$$
$$\vdots$$

EP update: $q_0(\boldsymbol{x}) \propto \dfrac{q(\boldsymbol{x})}{q_1(\boldsymbol{x})}$ $\quad \boldsymbol{\Sigma}_0 = (\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}_1^{-1})^{-1}$

$q_1(\boldsymbol{x}) \propto \dfrac{q(\boldsymbol{x})}{q_0(\boldsymbol{x})}$ $\quad \boldsymbol{\Sigma}_1 = (\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}_0^{-1})^{-1}$

$$\vdots$$

$\boldsymbol{x} \in \mathbb{R}^{N \times N}$, $N \times N$ matrix inversion in each iteration

3. Applications – EP for scalable imaging inverse problems

- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- **toy examples and applications**

Toy example 1: 1d clutter problem (GMM likelihood + Gaussian prior)

Toy example 1: 1d clutter problem (GMM likelihood + Gaussian prior)

- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

Toy example 1: 1d clutter problem (GMM likelihood + Gaussian prior)

- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

GMM likelihood: $f(y|x) = 0.5\mathcal{N}(y; x, 1) + 0.5\mathcal{N}(y; 0, 10)$

Toy example 1: 1d clutter problem (GMM likelihood + Gaussian prior)

- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

GMM likelihood: $f(y|x) = 0.5\mathcal{N}(y;x,1) + 0.5\mathcal{N}(y;0,10)$

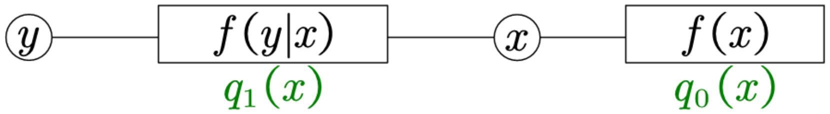
Gaussian prior: $f(x) = \mathcal{N}(x;\mu_0,\sigma_0^2)$

Toy example 1: 1d clutter problem (GMM likelihood + Gaussian prior)

- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

$$\text{GMM likelihood: } f(y|x) = 0.5\mathcal{N}(y;x,1) + 0.5\mathcal{N}(y;0,10)$$

$$\text{Gaussian prior: } f(x) = \mathcal{N}(x;\mu_0,\sigma_0^2)$$



3. Applications – EP for scalable imaging inverse problems

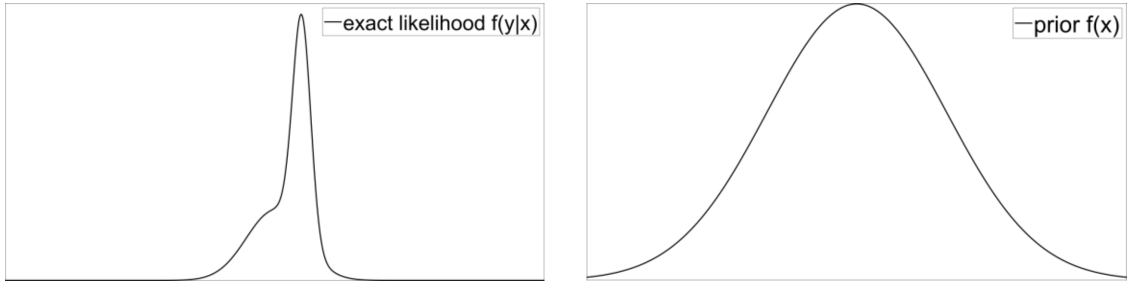
- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- **toy examples and applications**

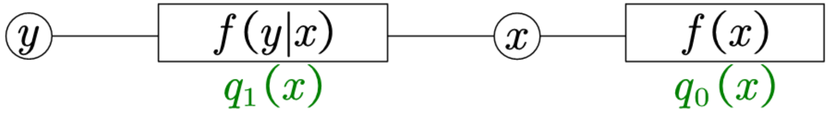
Toy example 1: 1d clutter problem (GMM likelihood + Gaussian prior)

- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

GMM likelihood: $f(y|x) = 0.5\mathcal{N}(y;x,1) + 0.5\mathcal{N}(y;0,10)$

Gaussian prior: $f(x) = \mathcal{N}(x;\mu_0,\sigma_0^2)$



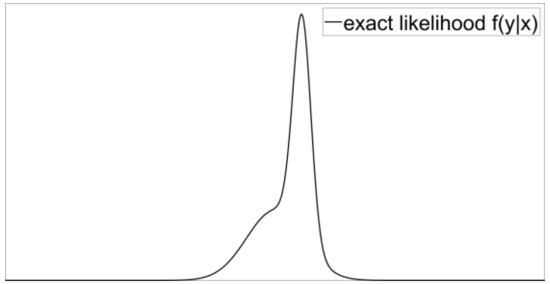


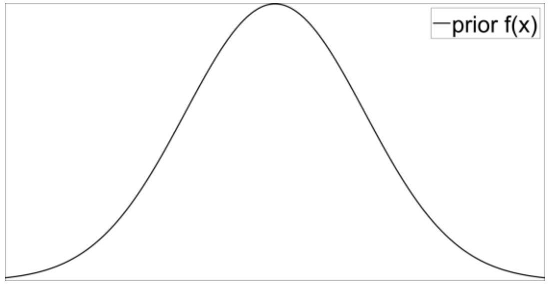
Toy example 1: 1d clutter problem (GMM likelihood + Gaussian prior)

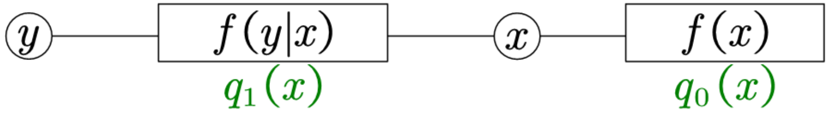
- Bayesian model: $p(x|y) \propto f(y|x) f(x)$

GMM likelihood: $f(y|x) = 0.5\mathcal{N}(y; x, 1) + 0.5\mathcal{N}(y; 0, 10)$

Gaussian prior: $f(x) = \mathcal{N}(x; \mu_0, \sigma_0^2)$







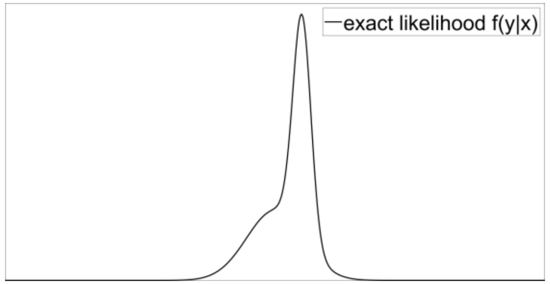
- EP approximation: $q(x) \approx p(x|y)$

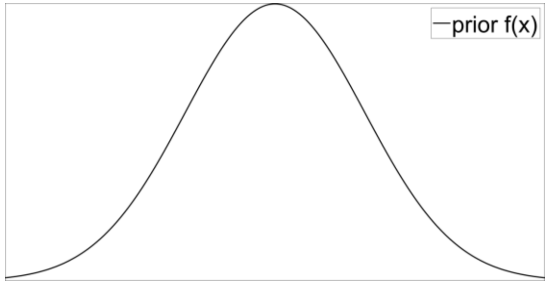
Toy example 1: 1d clutter problem (GMM likelihood + Gaussian prior)

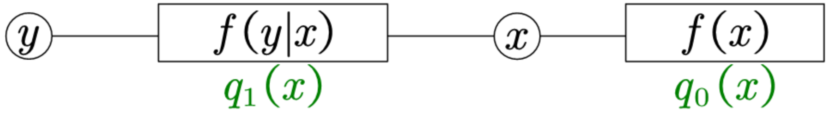
- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

GMM likelihood: $f(y|x) = 0.5\mathcal{N}(y;x,1) + 0.5\mathcal{N}(y;0,10)$

Gaussian prior: $f(x) = \mathcal{N}(x;\mu_0,\sigma_0^2)$







- EP approximation: $q(x) \approx p(x|y)$

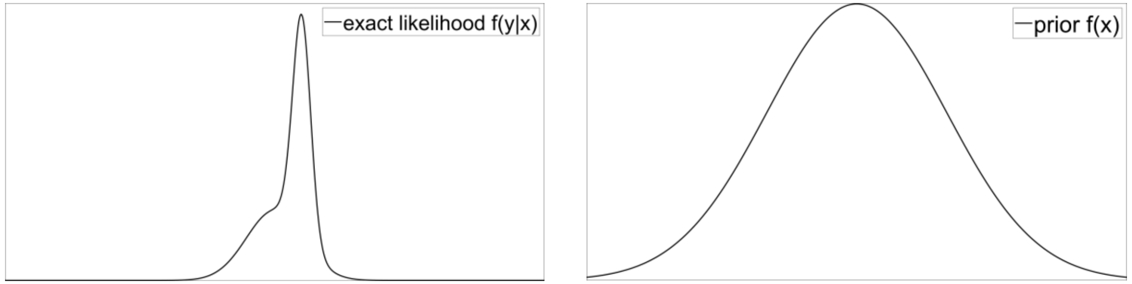
$q(x) = \mathcal{N}(x;\mu,\sigma^2)$

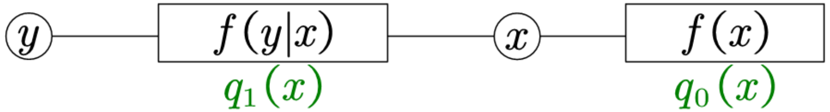
Toy example 1: 1d clutter problem (GMM likelihood + Gaussian prior)

- Bayesian model: $p(x|y) \propto f(y|x) f(x)$

GMM likelihood: $f(y|x) = 0.5\mathcal{N}(y;x,1) + 0.5\mathcal{N}(y;0,10)$

Gaussian prior: $f(x) = \mathcal{N}(x;\mu_0,\sigma_0^2)$





- EP approximation: $q(x) \approx p(x|y)$ $\quad q(x) = \mathcal{N}(x;\mu,\sigma^2)$

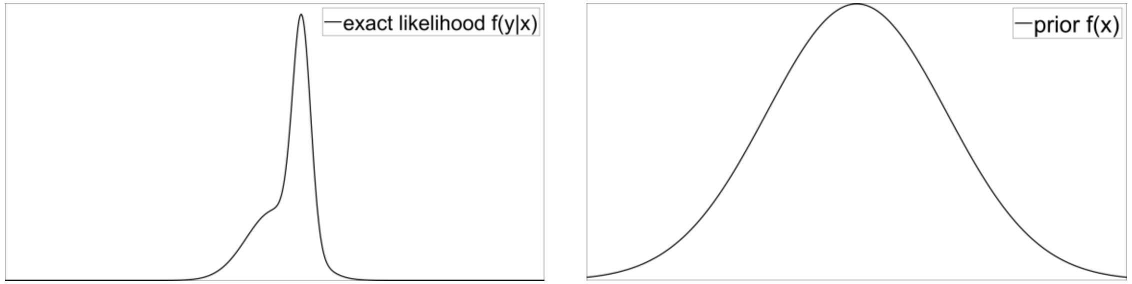
$q_0(x) = f(x) = \mathcal{N}(x;\mu_0,\sigma_0^2)$ (no approximation)

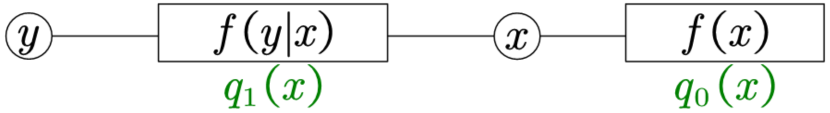
Toy example 1: 1d clutter problem (GMM likelihood + Gaussian prior)

- Bayesian model: $p(x|y) \propto f(y|x) f(x)$

GMM likelihood: $f(y|x) = 0.5\mathcal{N}(y; x, 1) + 0.5\mathcal{N}(y; 0, 10)$

Gaussian prior: $f(x) = \mathcal{N}(x; \mu_0, \sigma_0^2)$





- EP approximation: $q(x) \approx p(x|y)$ $\quad q(x) = \mathcal{N}(x; \mu, \sigma^2)$

$q_0(x) = f(x) = \mathcal{N}(x; \mu_0, \sigma_0^2)$ (no approximation)

$$q_1(x) = \underset{q_1(x) = \mathcal{N}(x; \mu_1, \sigma_1^2)}{\arg\min} KL(f(y|x) q_0(x) || q(x))$$

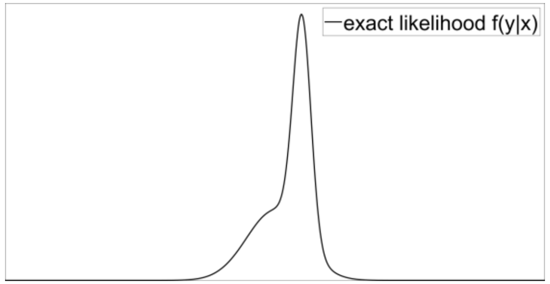
- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- **toy examples and applications**

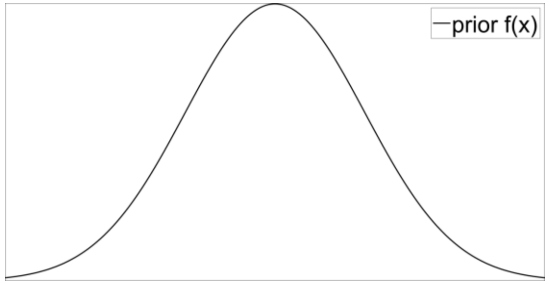
Toy example 1: 1d clutter problem (GMM likelihood + Gaussian prior)

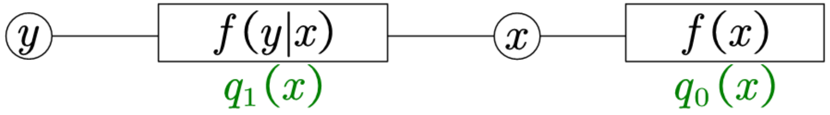
- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

GMM likelihood: $f(y|x) = 0.5\mathcal{N}(y; x, 1) + 0.5\mathcal{N}(y; 0, 10)$

Gaussian prior: $f(x) = \mathcal{N}(x; \mu_0, \sigma_0^2)$







- EP approximation: $q(x) \approx p(x|y)$ $\quad q(x) = \mathcal{N}(x; \mu, \sigma^2)$

$q_0(x) = f(x) = \mathcal{N}(x; \mu_0, \sigma_0^2)$ (no approximation)

$$q_1(x) = \underset{q_1(x) = \mathcal{N}(x; \mu_1, \sigma_1^2)}{\arg\min} KL(f(y|x)q_0(x) || q(x))$$

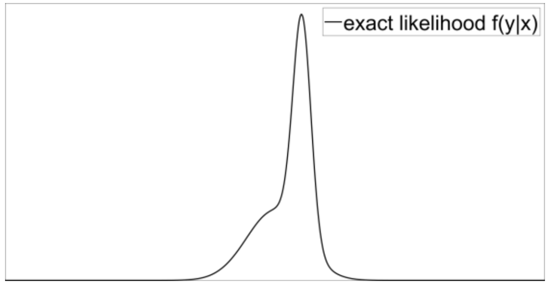
step 1. compute $\mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$ $\quad$ *tilted* $p_1(x) = f(y|x)q_0(x)$

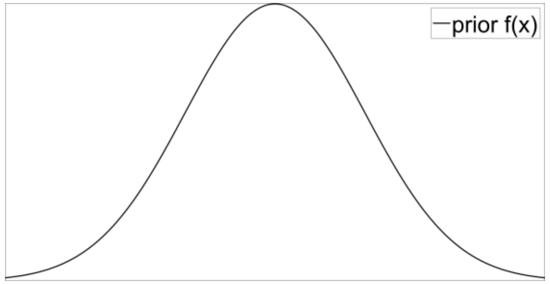
Toy example 1: 1d clutter problem (GMM likelihood + Gaussian prior)

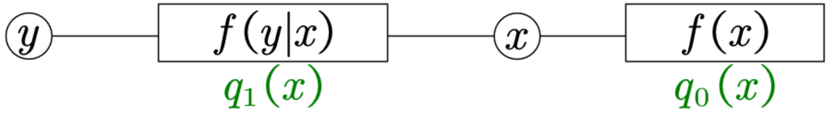
- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

$$\text{GMM likelihood: } f(y|x) = 0.5\mathcal{N}(y;x,1) + 0.5\mathcal{N}(y;0,10)$$

$$\text{Gaussian prior: } f(x) = \mathcal{N}(x;\mu_0,\sigma_0^2)$$







- EP approximation: $q(x) \approx p(x|y)$ $\quad q(x) = \mathcal{N}(x;\mu,\sigma^2)$

$$q_0(x) = f(x) = \mathcal{N}(x;\mu_0,\sigma_0^2) \ \text{(no approximation)}$$

$$q_1(x) = \underset{q_1(x)=\mathcal{N}(x;\mu_1,\sigma_1^2)}{\arg\min} KL(f(y|x)q_0(x)||q(x))$$

step 1. compute $\mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$ $\quad$ *tilted* $p_1(x) = f(y|x)q_0(x)$

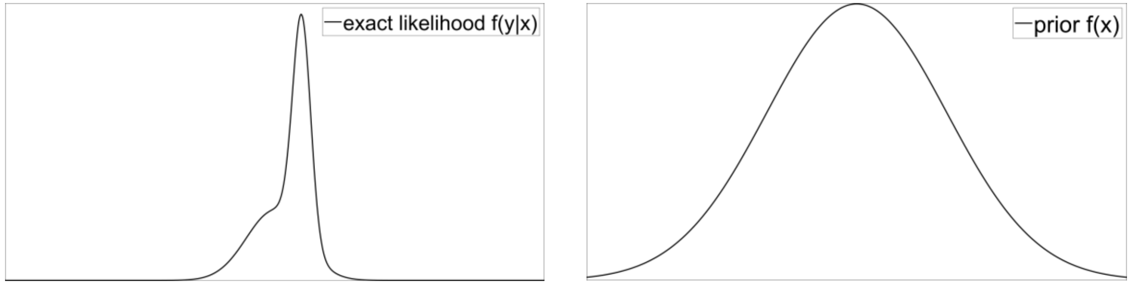
$$p_1(x) = [0.5\mathcal{N}(y;x,1) + 0.5\mathcal{N}(y;0,10)]\mathcal{N}(x;\mu_0,\sigma_0^2)$$
$$= \omega\mathcal{N}(x;\mu_*,\sigma_*^2) + (1-\omega)\mathcal{N}(x;\mu_0,\sigma_0^2)$$

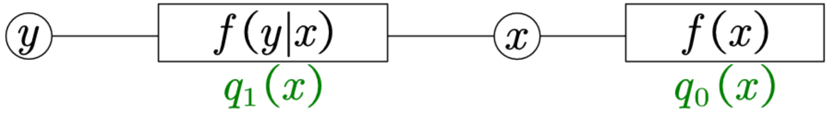
Toy example 1: 1d clutter problem (GMM likelihood + Gaussian prior)

- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

GMM likelihood: $f(y|x) = 0.5\mathcal{N}(y;x,1) + 0.5\mathcal{N}(y;0,10)$

Gaussian prior: $f(x) = \mathcal{N}(x;\mu_0,\sigma_0^2)$





- EP approximation: $q(x) \approx p(x|y)$ $\quad q(x) = \mathcal{N}(x;\mu,\sigma^2)$

$q_0(x) = f(x) = \mathcal{N}(x;\mu_0,\sigma_0^2)$ (no approximation)

$$q_1(x) = \underset{q_1(x)=\mathcal{N}(x;\mu_1,\sigma_1^2)}{\arg\min} KL(f(y|x)q_0(x)||q(x))$$

step 1. compute $\mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$ $\qquad$ *tilted* $p_1(x) = f(y|x)q_0(x)$

$$p_1(x) = [0.5\mathcal{N}(y;x,1) + 0.5\mathcal{N}(y;0,10)]\mathcal{N}(x;\mu_0,\sigma_0^2)$$

$$= \omega\mathcal{N}(x;\mu_*,\sigma_*^2) + (1-\omega)\mathcal{N}(x;\mu_0,\sigma_0^2)$$

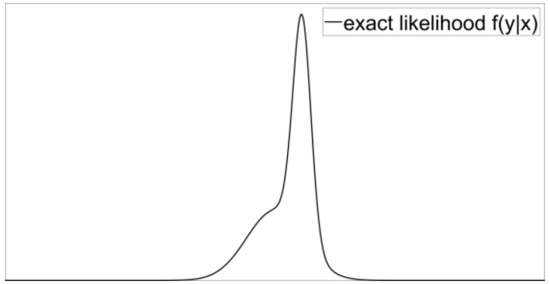
$$\Longrightarrow \mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$$

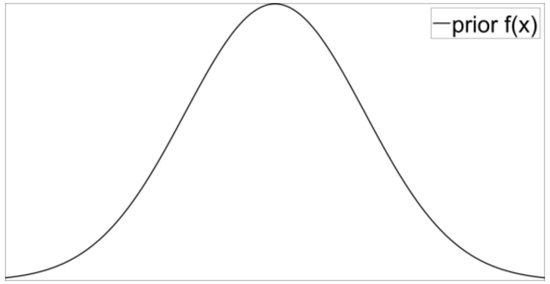
Toy example 1: 1d clutter problem (GMM likelihood + Gaussian prior)

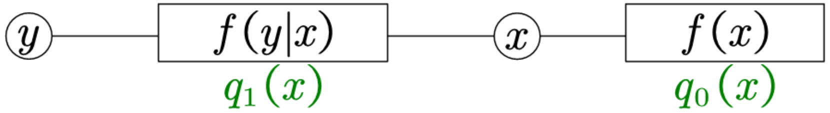
- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

GMM likelihood: $f(y|x) = 0.5\mathcal{N}(y;x,1) + 0.5\mathcal{N}(y;0,10)$

Gaussian prior: $f(x) = \mathcal{N}(x;\mu_0,\sigma_0^2)$







- EP approximation: $q(x) \approx p(x|y)$ $q(x) = \mathcal{N}(x;\mu,\sigma^2)$

$q_0(x) = f(x) = \mathcal{N}(x;\mu_0,\sigma_0^2)$ (no approximation)

$$q_1(x) = \underset{q_1(x)=\mathcal{N}(x;\mu_1,\sigma_1^2)}{\arg\min} KL(f(y|x)q_0(x)||q(x))$$

step 1. compute $\mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$ $tilted\ p_1(x) = f(y|x)q_0(x)$

$$p_1(x) = [0.5\mathcal{N}(y;x,1) + 0.5\mathcal{N}(y;0,10)]\mathcal{N}(x;\mu_0,\sigma_0^2)$$
$$= \omega\mathcal{N}(x;\mu_*,\sigma_*^2) + (1-\omega)\mathcal{N}(x;\mu_0,\sigma_0^2)$$
$$\Longrightarrow \mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$$

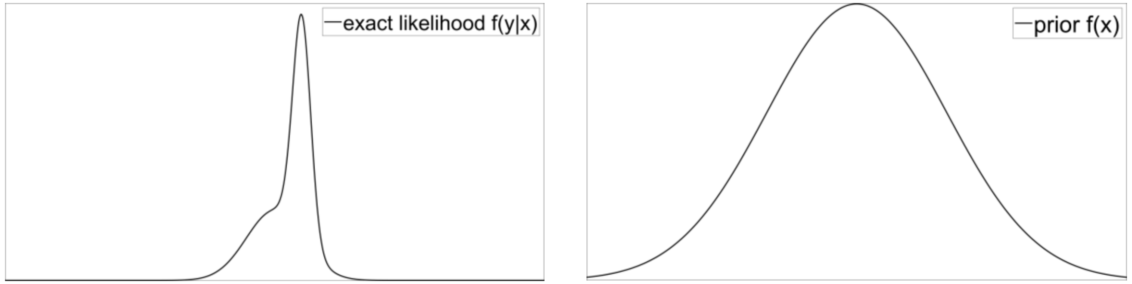
step 2. moment matching $\mathbb{E}_q[x] = \mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$

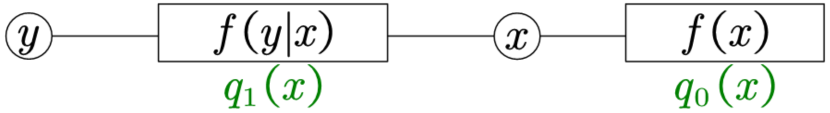
Toy example 1: 1d clutter problem (GMM likelihood + Gaussian prior)

- Bayesian model: $p(x|y) \propto f(y|x) f(x)$

$$\text{GMM likelihood: } f(y|x) = 0.5\mathcal{N}(y; x, 1) + 0.5\mathcal{N}(y; 0, 10)$$

$$\text{Gaussian prior: } f(x) = \mathcal{N}(x; \mu_0, \sigma_0^2)$$





- EP approximation: $q(x) \approx p(x|y)$ $\qquad q(x) = \mathcal{N}(x; \mu, \sigma^2)$

$$q_0(x) = f(x) = \mathcal{N}(x; \mu_0, \sigma_0^2) \ \text{(no approximation)}$$

$$q_1(x) = \underset{q_1(x) = \mathcal{N}(x; \mu_1, \sigma_1^2)}{\arg\min} KL(f(y|x) q_0(x) || q(x))$$

step 1. compute $\mathbb{E}_{tilted} \begin{bmatrix} x \\ x^2 \end{bmatrix}$ $\qquad$ $tilted\ p_1(x) = f(y|x) q_0(x)$

$$p_1(x) = [0.5\mathcal{N}(y; x, 1) + 0.5\mathcal{N}(y; 0, 10)] \mathcal{N}(x; \mu_0, \sigma_0^2)$$

$$= \omega \mathcal{N}(x; \mu_*, \sigma_*^2) + (1 - \omega) \mathcal{N}(x; \mu_0, \sigma_0^2)$$

$$\Longrightarrow \mathbb{E}_{tilted} \begin{bmatrix} x \\ x^2 \end{bmatrix}$$

step 2. moment matching $\mathbb{E}_q[x] = \mathbb{E}_{tilted} \begin{bmatrix} x \\ x^2 \end{bmatrix}$

$$\Longrightarrow \begin{bmatrix} \mu \\ \mu^2 + \sigma^2 \end{bmatrix} = \mathbb{E}_{tilted} \begin{bmatrix} x \\ x^2 \end{bmatrix}$$

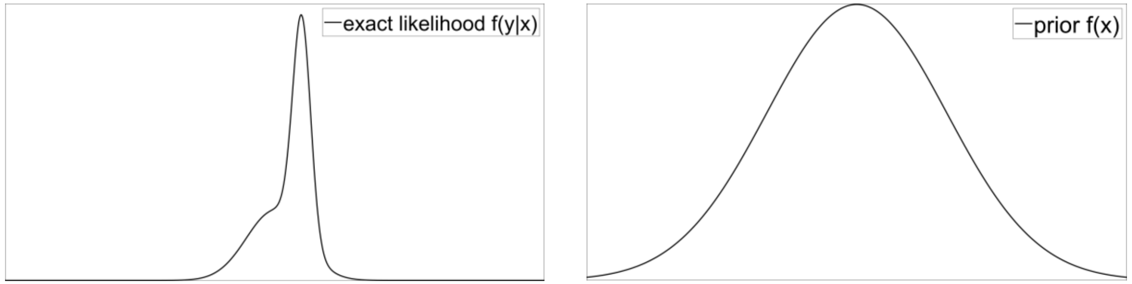
- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- **toy examples and applications**

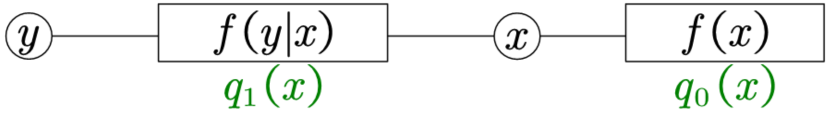
Toy example 1: 1d clutter problem (GMM likelihood + Gaussian prior)

- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

GMM likelihood: $f(y|x) = 0.5\mathcal{N}(y;x,1) + 0.5\mathcal{N}(y;0,10)$

Gaussian prior: $f(x) = \mathcal{N}(x;\mu_0,\sigma_0^2)$





- EP approximation: $q(x) \approx p(x|y)$ $\quad q(x) = \mathcal{N}(x;\mu,\sigma^2)$

$q_0(x) = f(x) = \mathcal{N}(x;\mu_0,\sigma_0^2)$ (no approximation)

$$q_1(x) = \underset{q_1(x)=\mathcal{N}(x;\mu_1,\sigma_1^2)}{\arg\min} KL(f(y|x)q_0(x)||q(x))$$

step 1. compute $\mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$ $\quad$ $tilted\ p_1(x) = f(y|x)q_0(x)$

$$p_1(x) = [0.5\mathcal{N}(y;x,1) + 0.5\mathcal{N}(y;0,10)]\mathcal{N}(x;\mu_0,\sigma_0^2)$$
$$= \omega\mathcal{N}(x;\mu_*,\sigma_*^2) + (1-\omega)\mathcal{N}(x;\mu_0,\sigma_0^2)$$
$$\Longrightarrow \mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$$

step 2. moment matching $\mathbb{E}_q[x] = \mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$

$$\Longrightarrow \begin{bmatrix} \mu \\ \mu^2+\sigma^2 \end{bmatrix} = \mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$$

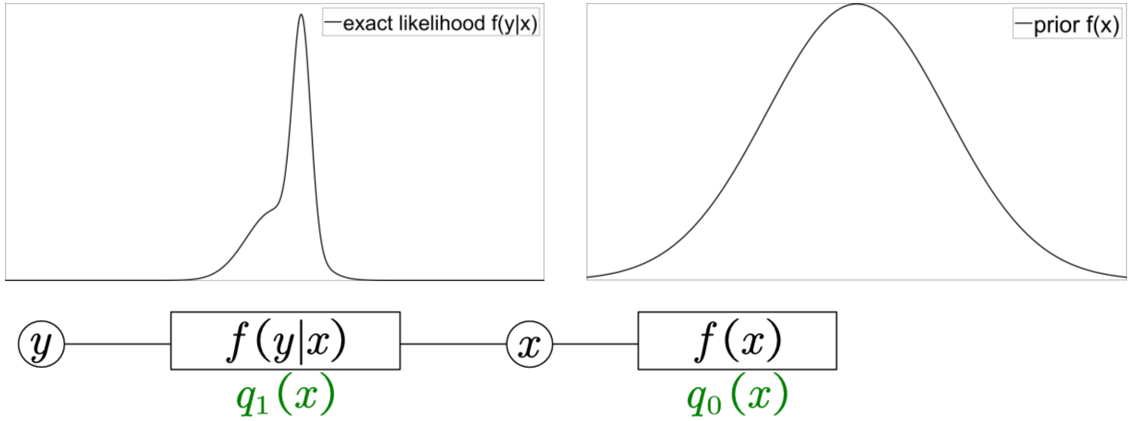
step 3. update $q_1(x) \propto \dfrac{q(x)}{q_0(x)}$

Toy example 1: 1d clutter problem (GMM likelihood + Gaussian prior)

- Bayesian model: $p(x|y) \propto f(y|x) f(x)$

$$\text{GMM likelihood: } f(y|x) = 0.5\mathcal{N}(y;x,1) + 0.5\mathcal{N}(y;0,10)$$

$$\text{Gaussian prior: } f(x) = \mathcal{N}(x;\mu_0,\sigma_0^2)$$



- EP approximation: $q(x) \approx p(x|y)$ $\quad q(x) = \mathcal{N}(x;\mu,\sigma^2)$

$$q_0(x) = f(x) = \mathcal{N}(x;\mu_0,\sigma_0^2) \text{ (no approximation)}$$

$$q_1(x) = \underset{q_1(x)=\mathcal{N}(x;\mu_1,\sigma_1^2)}{\arg\min} KL(f(y|x)q_0(x)||q(x))$$

step 1. compute $\mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$ $\quad$ *tilted* $p_1(x) = f(y|x)q_0(x)$

$$p_1(x) = [0.5\mathcal{N}(y;x,1) + 0.5\mathcal{N}(y;0,10)]\mathcal{N}(x;\mu_0,\sigma_0^2)$$
$$= \omega\mathcal{N}(x;\mu_*,\sigma_*^2) + (1-\omega)\mathcal{N}(x;\mu_0,\sigma_0^2)$$
$$\Longrightarrow \mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$$

step 2. moment matching $\mathbb{E}_q[x] = \mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$

$$\Longrightarrow \begin{bmatrix} \mu \\ \mu^2+\sigma^2 \end{bmatrix} = \mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$$

step 3. update $q_1(x) \propto \dfrac{q(x)}{q_0(x)}$

$$\Longrightarrow (\mu_1,\sigma_1^2)$$

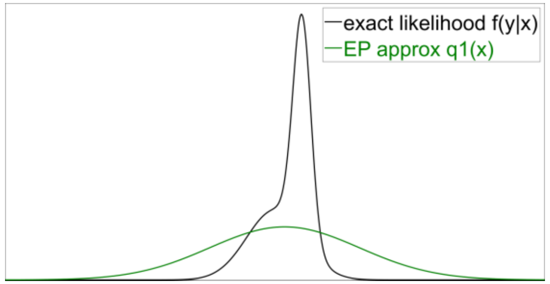
- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- **toy examples and applications**

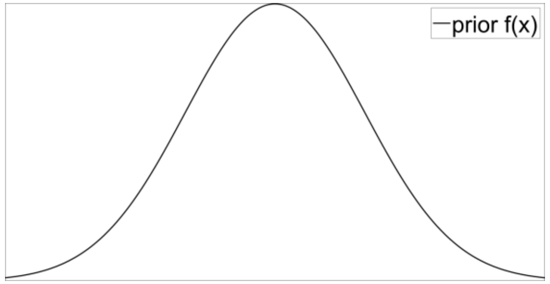
Toy example 1: 1d clutter problem (GMM likelihood + Gaussian prior)

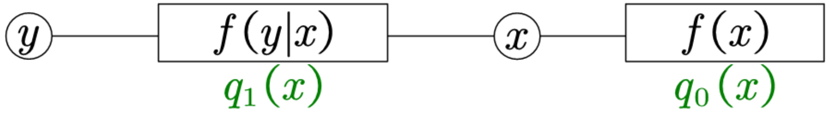
- Bayesian model: $p(x|y) \propto f(y|x) f(x)$

GMM likelihood: $f(y|x) = 0.5\mathcal{N}(y; x, 1) + 0.5\mathcal{N}(y; 0, 10)$

Gaussian prior: $f(x) = \mathcal{N}(x; \mu_0, \sigma_0^2)$







- EP approximation: $q(x) \approx p(x|y)$ $\quad q(x) = \mathcal{N}(x; \mu, \sigma^2)$

$q_0(x) = f(x) = \mathcal{N}(x; \mu_0, \sigma_0^2)$ (no approximation)

$$q_1(x) = \underset{q_1(x) = \mathcal{N}(x; \mu_1, \sigma_1^2)}{\arg\min} KL(f(y|x) q_0(x) || q(x))$$

step 1. compute $\mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$ $\quad$ *tilted* $p_1(x) = f(y|x) q_0(x)$

$$p_1(x) = [0.5\mathcal{N}(y; x, 1) + 0.5\mathcal{N}(y; 0, 10)]\mathcal{N}(x; \mu_0, \sigma_0^2)$$

$$= \omega\mathcal{N}(x; \mu_*, \sigma_*^2) + (1-\omega)\mathcal{N}(x; \mu_0, \sigma_0^2)$$

$$\Longrightarrow \mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$$

step 2. moment matching $\mathbb{E}_q[x] = \mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$

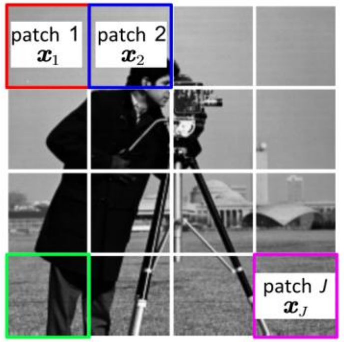
$$\Longrightarrow \begin{bmatrix} \mu \\ \mu^2 + \sigma^2 \end{bmatrix} = \mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$$

step 3. update $q_1(x) \propto \dfrac{q(x)}{q_0(x)}$

$$\Longrightarrow (\mu_1, \sigma_1^2)$$

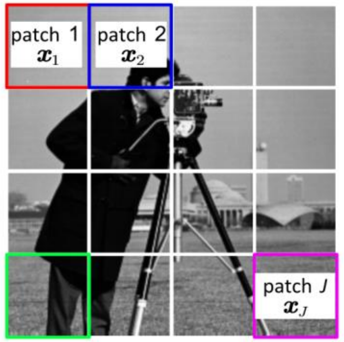
Application 1: scalable image restoration using EP with patch-based GMM prior

Application 1: scalable image restoration using EP with patch-based GMM prior



- **Image denosing/inpainting/deconvolution** $\mathbf{H}$: $\mathbf{I}$/binary mask/convolution matrix

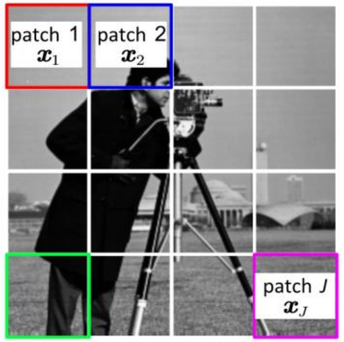
Application 1: scalable image restoration using EP with patch-based GMM prior



- **Image denosing/inpainting/deconvolution** $\mathbf{H}$: $\mathbf{I}$/binary mask/convolution matrix

$$\text{GMM patch} - \text{based prior: } f_x(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{j=1}^{J}\sum_{k=1}^{K}\omega_k \mathcal{N}(\boldsymbol{x}_j; m_0\mathbf{1} + \alpha\boldsymbol{\mu}_k, s^2 11^T + \alpha^2\mathbf{C}_k)$$

Application 1: scalable image restoration using EP with patch-based GMM prior

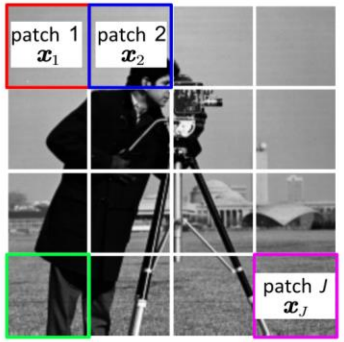


- **Image denosing/inpainting/deconvolution** $\mathbf{H}$: $\mathbf{I}$/binary mask/convolution matrix

$$\text{GMM patch}-\text{based prior}: f_x(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{j=1}^{J}\sum_{k=1}^{K}\omega_k\mathcal{N}(\boldsymbol{x}_j; m_0\mathbf{1}+\alpha\boldsymbol{\mu}_k, s^2 11^T+\alpha^2\mathbf{C}_k)$$

$$\text{Gaussian/Poisson likelihood}: f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y};\mathbf{H}\boldsymbol{x},\sigma^2)\ or\ \mathcal{P}(\mathbf{H}\boldsymbol{x})$$

Application 1: scalable image restoration using EP with patch-based GMM prior



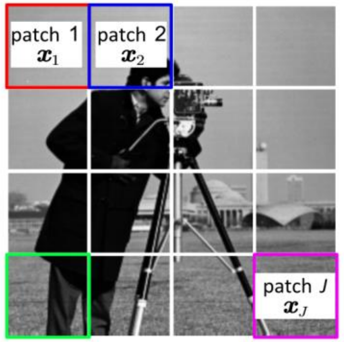
- **Image denosing/inpainting/deconvolution** $\mathbf{H}$: $\mathbf{I}$/binary mask/convolution matrix

$$\text{GMM patch}-\text{based prior:}\ f_x(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{j=1}^{J}\sum_{k=1}^{K}\omega_k\mathcal{N}(\boldsymbol{x}_j; m_0\mathbf{1}+\alpha\boldsymbol{\mu}_k, s^2 11^T+\alpha^2\mathbf{C}_k)$$

$$\text{Gaussian/Poisson likelihood:}\ f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y};\mathbf{H}\boldsymbol{x},\sigma^2)\ or\ \mathcal{P}(\mathbf{H}\boldsymbol{x})$$

$$\text{exact posterior:}\ p(\boldsymbol{x}|\boldsymbol{y}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) f_x(\boldsymbol{x}|\boldsymbol{\theta})$$

Application 1: scalable image restoration using EP with patch-based GMM prior



- **Image denosing/inpainting/deconvolution** $\mathbf{H}$: $\mathbf{I}$/binary mask/convolution matrix

$$\text{GMM patch}-\text{based prior: } f_x(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{j=1}^{J}\sum_{k=1}^{K}\omega_k \mathcal{N}(\boldsymbol{x}_j; m_0\mathbf{1}+\alpha\boldsymbol{\mu}_k, s^2 11^T+\alpha^2\mathbf{C}_k)$$

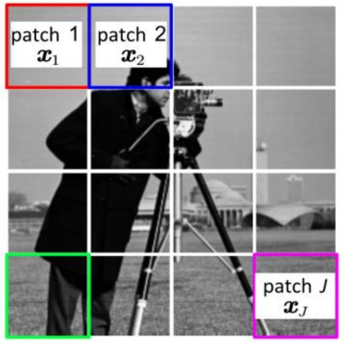
$$\text{Gaussian/Poisson likelihood: } f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y};\mathbf{H}\boldsymbol{x},\sigma^2)\ or\ \mathcal{P}(\mathbf{H}\boldsymbol{x})$$

$$\text{exact posterior: } p(\boldsymbol{x}|\boldsymbol{y}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})f_x(\boldsymbol{x}|\boldsymbol{\theta})$$

how to factorize: $\boldsymbol{y}$ — $\boxed{\prod_{i=1}^{N} f_{y|x}(y_i|\boldsymbol{h}_i\boldsymbol{x})}$ — $\boldsymbol{x}$ — $\boxed{\prod_{j=1}^{J}\sum_{k=1}^{K}\omega_k\mathcal{N}(\boldsymbol{x}_j;\boldsymbol{m}_k,\mathbf{C}_k)}$

$$q_1(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x};\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1) \qquad q_0(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x};\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)$$

Application 1: scalable image restoration using EP with patch-based GMM prior



- **Image denosing/inpainting/deconvolution** $\mathbf{H}$: $\mathbf{I}$/binary mask/convolution matrix

$$\text{GMM patch}-\text{based prior: } f_x(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{j=1}^{J}\sum_{k=1}^{K}\omega_k\mathcal{N}(\boldsymbol{x}_j; m_0\mathbf{1}+\alpha\boldsymbol{\mu}_k, s^2 11^T+\alpha^2\mathbf{C}_k)$$

$$\text{Gaussian/Poisson likelihood: } f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y}; \mathbf{H}\boldsymbol{x}, \sigma^2) \ or\ \mathcal{P}(\mathbf{H}\boldsymbol{x})$$

$$\text{exact posterior: } p(\boldsymbol{x}|\boldsymbol{y}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})f_x(\boldsymbol{x}|\boldsymbol{\theta})$$

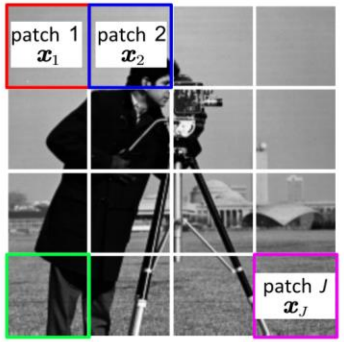
how to factorize:

$$\boldsymbol{y} \; - \; \boxed{\prod_{i=1}^{N} f_{y|x}(y_i|\boldsymbol{h}_i\boldsymbol{x})} \; - \; \boldsymbol{x} \; - \; \boxed{\prod_{j=1}^{J}\sum_{k=1}^{K}\omega_k\mathcal{N}(\boldsymbol{x}_j; \boldsymbol{m}_k, \mathbf{C}_k)}$$

$$q_1(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \qquad q_0(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$$

how to compute:

Application 1: scalable image restoration using EP with patch-based GMM prior



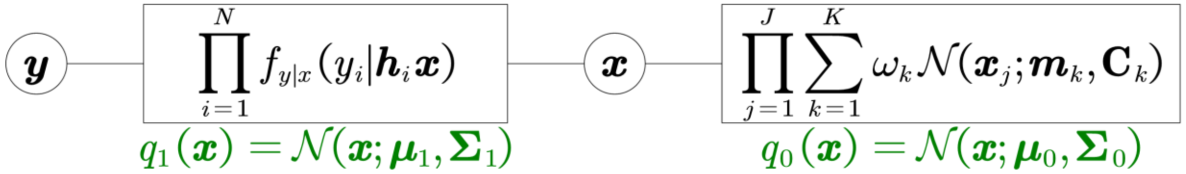
- **Image denosing/inpainting/deconvolution** $\mathbf{H}$: $\mathbf{I}$/binary mask/convolution matrix

$$\text{GMM patch} - \text{based prior: } f_x(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{j=1}^{J}\sum_{k=1}^{K}\omega_k\mathcal{N}(\boldsymbol{x}_j; m_0\mathbf{1} + \alpha\boldsymbol{\mu}_k, s^2 11^T + \alpha^2\mathbf{C}_k)$$

$$\text{Gaussian/Poisson likelihood: } f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y}; \mathbf{H}\boldsymbol{x}, \sigma^2) \ or\ \mathcal{P}(\mathbf{H}\boldsymbol{x})$$

$$\text{exact posterior: } p(\boldsymbol{x}|\boldsymbol{y}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) f_x(\boldsymbol{x}|\boldsymbol{\theta})$$

how to factorize:



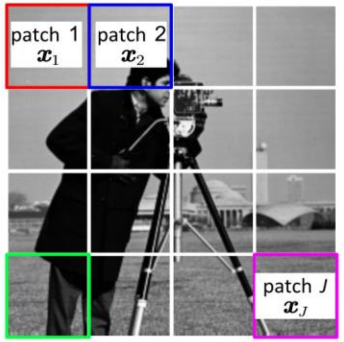
how to compute: 1. $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_0$ are constrained to be **block-diagonal**

$$\begin{bmatrix} [\boldsymbol{\Sigma}_1]_1 & \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \\ \vdots & [\boldsymbol{\Sigma}_1]_2 & \cdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \cdots & \mathbf{0} & [\boldsymbol{\Sigma}_1]_J \end{bmatrix} \quad \begin{bmatrix} [\boldsymbol{\Sigma}_0]_1 & \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \\ \vdots & [\boldsymbol{\Sigma}_0]_1 & \cdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \cdots & \mathbf{0} & [\boldsymbol{\Sigma}_0]_J \end{bmatrix}$$

3. Applications – EP for scalable imaging inverse problems

- ➢ how to construct an EP algorithm to solve image inverse problems
- ➢ how to achieve scalable posterior approximation
- ➢ **toy examples and applications**

Application 1: scalable image restoration using EP with patch-based GMM prior



- **Image denosing/inpainting/deconvolution** $\mathbf{H}$: $\mathbf{I}$/binary mask/convolution matrix

$$\text{GMM patch} - \text{based prior: } f_x(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{j=1}^{J}\sum_{k=1}^{K}\omega_k\mathcal{N}(\boldsymbol{x}_j; m_0\mathbf{1} + \alpha\boldsymbol{\mu}_k, s^2 11^T + \alpha^2\mathbf{C}_k)$$

$$\text{Gaussian/Poisson likelihood: } f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y}; \mathbf{H}\boldsymbol{x}, \sigma^2) \ or \ \mathcal{P}(\mathbf{H}\boldsymbol{x})$$

$$\text{exact posterior: } p(\boldsymbol{x}|\boldsymbol{y}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) f_x(\boldsymbol{x}|\boldsymbol{\theta})$$

how to factorize:

$$\boldsymbol{y} \;—\; \boxed{\prod_{i=1}^{N} f_{y|x}(y_i|\boldsymbol{h}_i\boldsymbol{x})} \;—\; \boldsymbol{x} \;—\; \boxed{\prod_{j=1}^{J}\sum_{k=1}^{K}\omega_k\mathcal{N}(\boldsymbol{x}_j; \boldsymbol{m}_k, \mathbf{C}_k)}$$

$$q_1(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \qquad q_0(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$$

how to compute: 1. $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_0$ are constrained to be **block-diagonal**

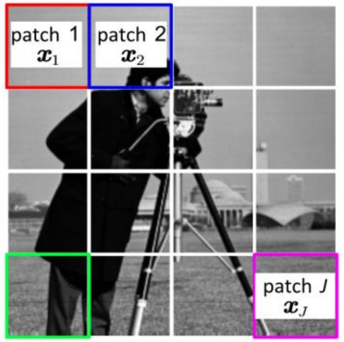
$$\begin{bmatrix} [\boldsymbol{\Sigma}_1]_1 & \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \\ \vdots & [\boldsymbol{\Sigma}_1]_2 & \cdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \cdots & \mathbf{0} & [\boldsymbol{\Sigma}_1]_J \end{bmatrix} \quad \begin{bmatrix} [\boldsymbol{\Sigma}_0]_1 & \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \\ \vdots & [\boldsymbol{\Sigma}_0]_1 & \cdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \cdots & \mathbf{0} & [\boldsymbol{\Sigma}_0]_J \end{bmatrix}$$

$$N \times N \implies J \ r \times r \ \text{matrix } \textbf{parallel} \text{ inversion } (r \ll N)$$

3. Applications – EP for scalable imaging inverse problems

- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- **toy examples and applications**

Application 1: scalable image restoration using EP with patch-based GMM prior



- **Image denosing/inpainting/deconvolution** $\mathbf{H}$: $\mathbf{I}$/binary mask/convolution matrix

$$\text{GMM patch}-\text{based prior:}\ f_x(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{j=1}^{J}\sum_{k=1}^{K}\omega_k\mathcal{N}(\boldsymbol{x}_j; m_0\mathbf{1}+\alpha\boldsymbol{\mu}_k, s^2 11^T+\alpha^2\mathbf{C}_k)$$

$$\text{Gaussian/Poisson likelihood:}\ f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y};\mathbf{H}\boldsymbol{x},\sigma^2)\ or\ \mathcal{P}(\mathbf{H}\boldsymbol{x})$$

$$\text{exact posterior:}\ p(\boldsymbol{x}|\boldsymbol{y}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})f_x(\boldsymbol{x}|\boldsymbol{\theta})$$

how to factorize:

$$\boldsymbol{y} \ —\ \boxed{\prod_{i=1}^{N} f_{y|x}(y_i|\boldsymbol{h}_i\boldsymbol{x})}\ —\ \boldsymbol{x}\ —\ \boxed{\prod_{j=1}^{J}\sum_{k=1}^{K}\omega_k\mathcal{N}(\boldsymbol{x}_j;\boldsymbol{m}_k,\mathbf{C}_k)}$$

$$q_1(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x};\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1) \qquad q_0(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x};\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)$$

how to compute: 1. $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_0$ are constrained to be **block-diagonal**

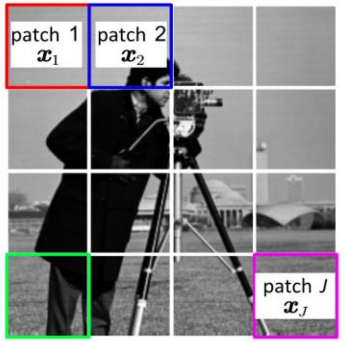
$$\begin{bmatrix} [\boldsymbol{\Sigma}_1]_1 & \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \\ \vdots & [\boldsymbol{\Sigma}_1]_2 & \cdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \cdots & \mathbf{0} & [\boldsymbol{\Sigma}_1]_J \end{bmatrix} \quad \begin{bmatrix} [\boldsymbol{\Sigma}_0]_1 & \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \\ \vdots & [\boldsymbol{\Sigma}_0]_1 & \cdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \cdots & \mathbf{0} & [\boldsymbol{\Sigma}_0]_J \end{bmatrix}$$

$$N\times N \implies J\ r\times r\ \text{matrix}\ \textbf{parallel}\ \text{inversion}\ (r\ll N)$$

2. automatic estimation of hyperparameter $\boldsymbol{\theta} = (m_0, s^2, \alpha)$

- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- **toy examples and applications**

Application 1: scalable image restoration using EP with patch-based GMM prior



- **Image denosing/inpainting/deconvolution** $\mathbf{H}$: $\mathbf{I}$/binary mask/convolution matrix

$$\text{GMM patch}-\text{based prior: } f_x(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{j=1}^{J}\sum_{k=1}^{K}\omega_k\mathcal{N}(\boldsymbol{x}_j; m_0\mathbf{1}+\alpha\boldsymbol{\mu}_k, s^2 11^T+\alpha^2\mathbf{C}_k)$$

$$\text{Gaussian/Poisson likelihood: } f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y};\mathbf{H}\boldsymbol{x},\sigma^2) \; or \; \mathcal{P}(\mathbf{H}\boldsymbol{x})$$

$$\text{exact posterior: } p(\boldsymbol{x}|\boldsymbol{y}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})f_x(\boldsymbol{x}|\boldsymbol{\theta})$$

how to factorize:

$$\boldsymbol{y} \; — \; \boxed{\prod_{i=1}^{N} f_{y|x}(y_i|\boldsymbol{h}_i\boldsymbol{x})} \; — \; \boldsymbol{x} \; — \; \boxed{\prod_{j=1}^{J}\sum_{k=1}^{K}\omega_k\mathcal{N}(\boldsymbol{x}_j;\boldsymbol{m}_k,\mathbf{C}_k)}$$

$$q_1(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x};\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1) \qquad q_0(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x};\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)$$

how to compute: 1. $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_0$ are constrained to be **block-diagonal**

$$\begin{bmatrix} [\boldsymbol{\Sigma}_1]_1 & \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \\ \vdots & [\boldsymbol{\Sigma}_1]_2 & \cdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \cdots & \mathbf{0} & [\boldsymbol{\Sigma}_1]_J \end{bmatrix} \quad \begin{bmatrix} [\boldsymbol{\Sigma}_0]_1 & \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \\ \vdots & [\boldsymbol{\Sigma}_0]_1 & \cdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \cdots & \mathbf{0} & [\boldsymbol{\Sigma}_0]_J \end{bmatrix}$$

$$N \times N \implies J \; r \times r \text{ matrix } \textbf{parallel} \text{ inversion } (r \ll N)$$

2. automatic estimation of hyperparameter $\boldsymbol{\theta} = (m_0, s^2, \alpha)$

$$\text{EP scalable posterior approximation: } q(\boldsymbol{x}) \propto q_1(\boldsymbol{x})q_0(\boldsymbol{x}) \propto \mathcal{N}(;\boldsymbol{\mu},\boldsymbol{\Sigma})$$

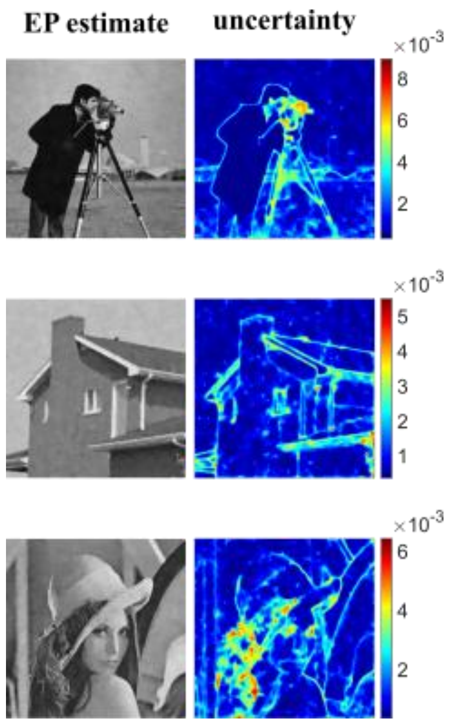
3. Applications – EP for scalable imaging inverse problems

- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- **toy examples and applications**

Application 1: scalable image restoration using EP with patch-based GMM prior

Image denoising $\boldsymbol{x} \in \mathbb{R}^{256 \times 256}$

	LIDIA [1]	BM3D [2]	MAP-GMM [3]	MAP – GMM [4]	EP-GMM
			Cameraman		
10/255	–	**34.12**	33.99	33.94	34.04
15/255	**32.41**	31.90	31.79	31.65	31.71
20/255	–	**30.45**	30.36	30.10	30.19
25/255	**29.91**	29.21	29.04	28.77	28.83
30/255	–	**28.61**	28.34	27.99	28.09
50/255	**26.83**	25.39	25.08	24.55	24.52
			House		
10/255	–	**36.79**	35.77	35.79	35.82
15/255	**35.09**	34.97	34.18	34.06	34.12
20/255	–	**33.83**	33.05	32.75	32.82
25/255	**33.08**	32.91	32.14	31.66	31.72
30/255	–	**32.08**	31.25	30.60	30.68
50/255	**30.14**	29.45	28.91	27.91	27.82
			Lena		
10/255	–	**33.95**	33.66	33.66	33.67
15/255	**32.27**	31.93	31.61	31.53	31.56
20/255	–	**30.41**	30.18	30.04	30.05
25/255	**29.91**	29.45	29.28	29.05	29.05
30/255	–	**28.62**	28.44	28.15	28.13
50/255	**26.86**	26.18	25.99	25.55	25.44



[1] G. Vaksman, M. Elad, and P. Milanfar. "*Lidia: Lightweight learned image denoising with instance adaptation*", IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 524–525, 2020.
[2] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. "*Image denoising by sparse 3-D transform-domain collaborative filtering*", IEEE Transactions on image processing, ol. 16, no. 8, pp. 2080–2095, 2007.
[3] D. Zoran and Y. Weiss. "*From learning models of natural image patches to whole image restoration*", IEEE International Conference on Computer Vision (ICCV), pp. 479–486, 2011.
[4] A. M. Teodoro, M. S. Almeida, and M. A. Figueiredo. "*Single-frame image denoising and inpainting using Gaussian mixtures*",ICPRAM (2), pp. 283–288, 2015.

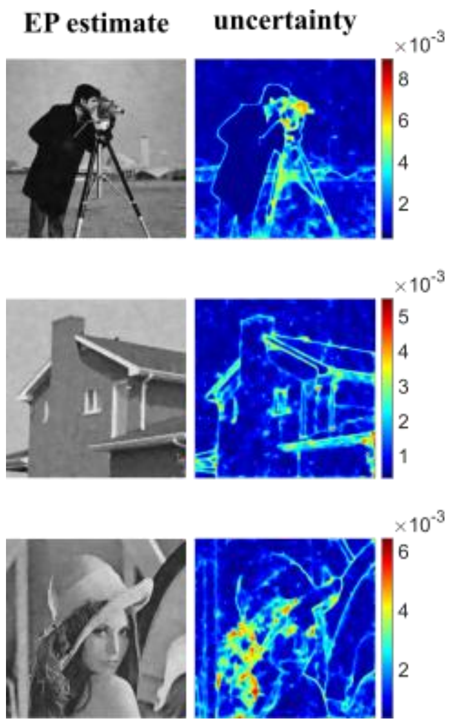
Dan Yao, Stephen McLaughlin, and Yoann Altmann. "*Patch-based Image Restoration using Expectation Propagation,*" SIAM Journal on Imaging Sciences, vol. 15, no. 1, pp. 192–227, 2022. https://doi.org/10.1137/21M1427541

- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- **toy examples and applications**

Application 1: scalable image restoration using EP with patch-based GMM prior

Image denoising $\boldsymbol{x} \in \mathbb{R}^{256 \times 256}$

	LIDIA [1]	BM3D [2]	MAP-GMM [3]	MAP–GMM [4]	EP-GMM
			Cameraman		
10/255	–	**34.12**	33.99	33.94	34.04
15/255	**32.41**	31.90	31.79	31.65	31.71
20/255	–	**30.45**	30.36	30.10	30.19
25/255	**29.91**	29.21	29.04	28.77	28.83
30/255	–	**28.61**	28.34	27.99	28.09
50/255	**26.83**	25.39	25.08	24.55	24.52
			House		
10/255	–	**36.79**	35.77	35.79	35.82
15/255	**35.09**	34.97	34.18	34.06	34.12
20/255	–	**33.83**	33.05	32.75	32.82
25/255	**33.08**	32.91	32.14	31.66	31.72
30/255	–	**32.08**	31.25	30.60	30.68
50/255	**30.14**	29.45	28.91	27.91	27.82
			Lena		
10/255	–	**33.95**	33.66	33.66	33.67
15/255	**32.27**	31.93	31.61	31.53	31.56
20/255	–	**30.41**	30.18	30.04	30.05
25/255	**29.91**	29.45	29.28	29.05	29.05
30/255	–	**28.62**	28.44	28.15	28.13
50/255	**26.86**	26.18	25.99	25.55	25.44



[1] G. Vaksman, M. Elad, and P. Milanfar. "*Lidia: Lightweight learned image denoising with instance adaptation*", IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 524–525, 2020.
[2] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. "*Image denoising by sparse 3-D transform-domain collaborative filtering*", IEEE Transactions on image processing, ol. 16, no. 8, pp. 2080–2095, 2007.
[3] D. Zoran and Y. Weiss. "*From learning models of natural image patches to whole image restoration*", IEEE International Conference on Computer Vision (ICCV), pp. 479–486, 2011.
[4] A. M. Teodoro, M. S. Almeida, and M. A. Figueiredo. "*Single-frame image denoising and inpainting using Gaussian mixtures*",ICPRAM (2), pp. 283–288, 2015.

Dan Yao, Stephen McLaughlin, and Yoann Altmann. "*Patch-based Image Restoration using Expectation Propagation,*" SIAM Journal on Imaging Sciences, vol. 15, no. 1, pp. 192–227, 2022. https://doi.org/10.1137/21M1427541

Toy example 2: 1d sparse prior (Gaussian likelihood + Laplace prior)

Toy example 2: 1d sparse prior (Gaussian likelihood + Laplace prior)

- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

Toy example 2: 1d sparse prior (Gaussian likelihood + Laplace prior)

- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

Gaussian likelihood: $f(y|x) = \mathcal{N}(y; x, \sigma^2)$

Toy example 2: 1d sparse prior (Gaussian likelihood + Laplace prior)

- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

$$\text{Gaussian likelihood: } f(y|x) = \mathcal{N}(y; x, \sigma^2)$$

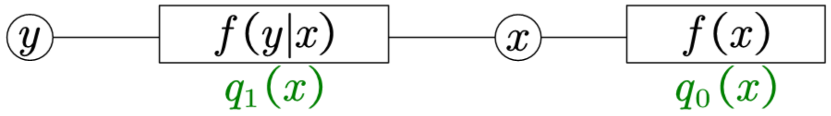
$$\text{Laplace prior: } f(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}$$

Toy example 2: 1d sparse prior (Gaussian likelihood + Laplace prior)

- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

Gaussian likelihood: $f(y|x) = \mathcal{N}(y; x, \sigma^2)$

Laplace prior: $f(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}$

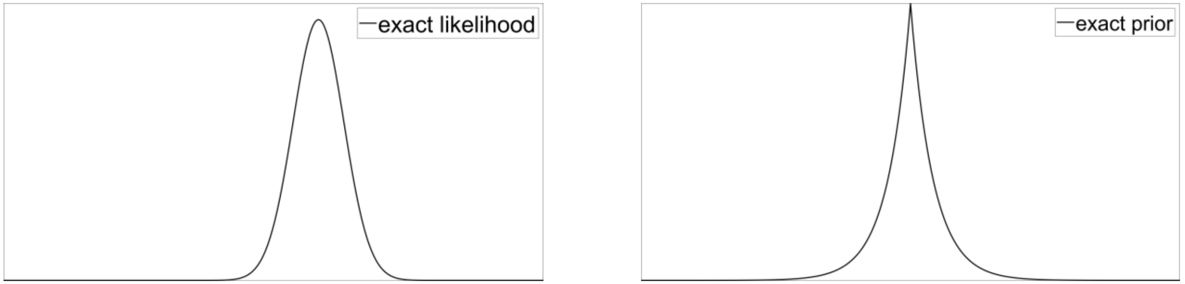


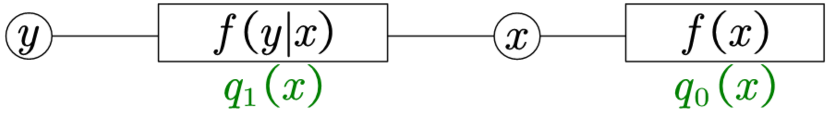
Toy example 2: 1d sparse prior (Gaussian likelihood + Laplace prior)

- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

Gaussian likelihood: $f(y|x) = \mathcal{N}(y; x, \sigma^2)$

Laplace prior: $f(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}$



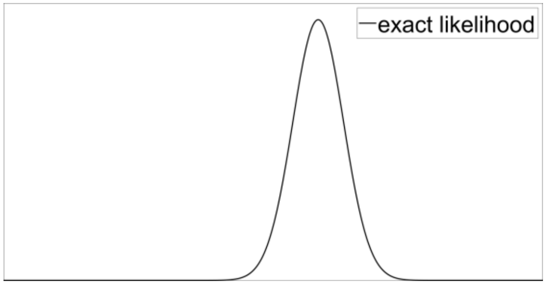


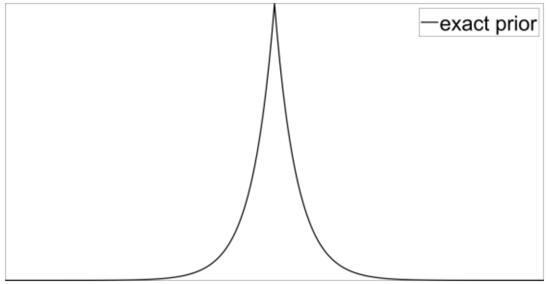
Toy example 2: 1d sparse prior (Gaussian likelihood + Laplace prior)

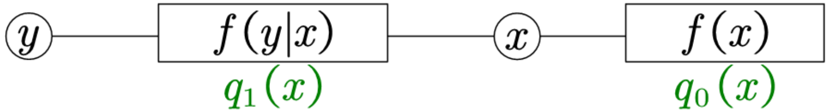
- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

Gaussian likelihood: $f(y|x) = \mathcal{N}(y; x, \sigma^2)$

Laplace prior: $f(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}$







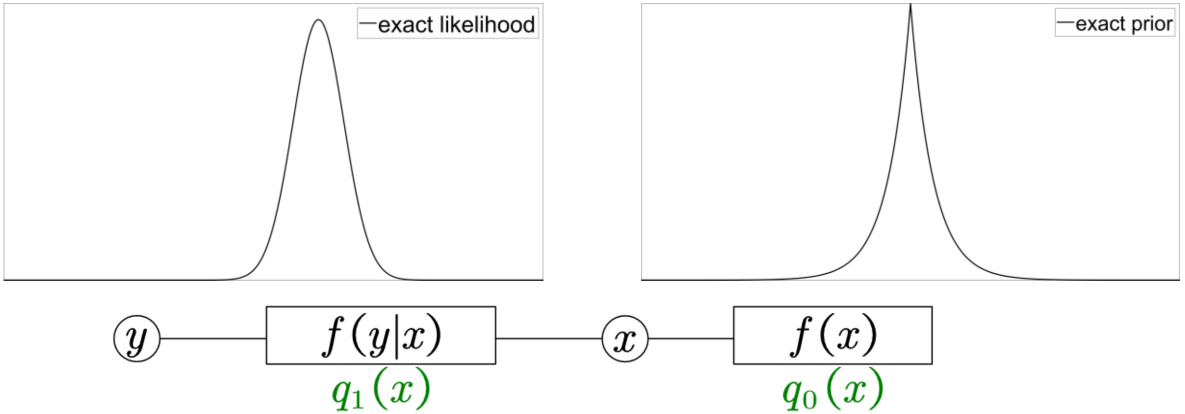
- EP approximation: $q(x) \approx p(x|y)$

Toy example 2: 1d sparse prior (Gaussian likelihood + Laplace prior)

- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

$$\text{Gaussian likelihood: } f(y|x) = \mathcal{N}(y; x, \sigma^2)$$

$$\text{Laplace prior: } f(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}$$



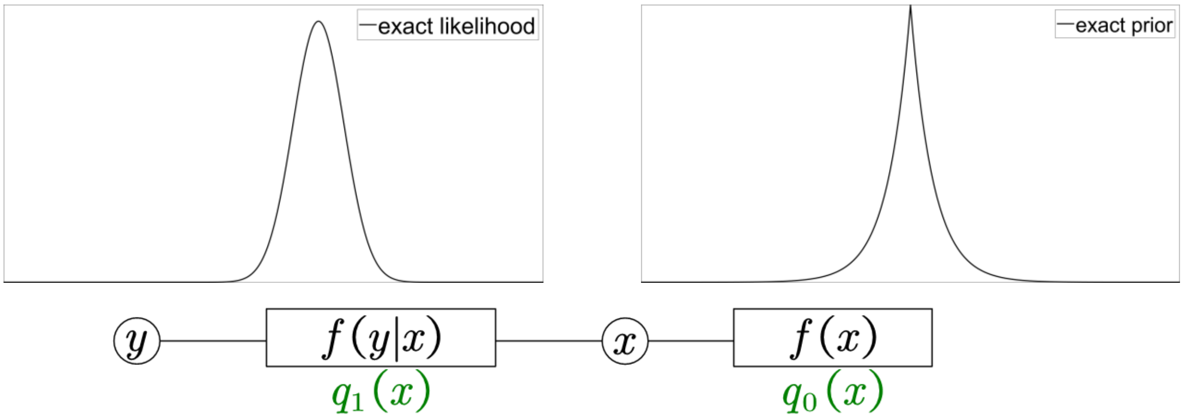
- EP approximation: $q(x) \approx p(x|y)$ $\quad q(x) = \mathcal{N}(x; \mu, \sigma^2)$

Toy example 2: 1d sparse prior (Gaussian likelihood + Laplace prior)

- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

Gaussian likelihood: $f(y|x) = \mathcal{N}(y; x, \sigma^2)$

Laplace prior: $f(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}$



- EP approximation: $q(x) \approx p(x|y)$ $\quad q(x) = \mathcal{N}(x; \mu, \sigma^2)$

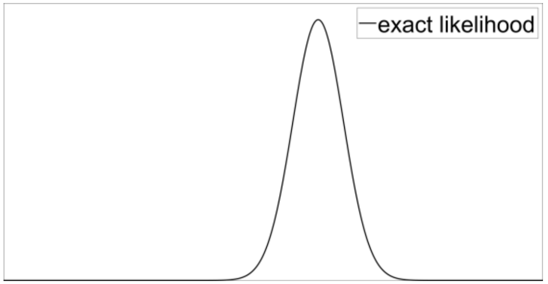
$q_1(x) = f(x) = \mathcal{N}(x; y, \sigma^2)$ (no approximation)

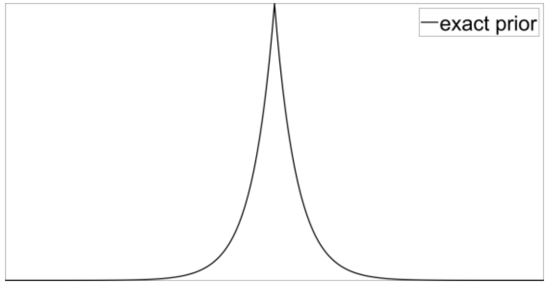
Toy example 2: 1d sparse prior (Gaussian likelihood + Laplace prior)

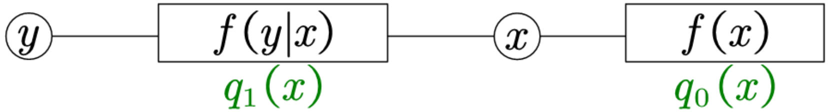
- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

Gaussian likelihood: $f(y|x) = \mathcal{N}(y; x, \sigma^2)$

Laplace prior: $f(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}$







- EP approximation: $q(x) \approx p(x|y)$ $\quad q(x) = \mathcal{N}(x; \mu, \sigma^2)$

$$q_1(x) = f(x) = \mathcal{N}(x; y, \sigma^2) \text{ (no approximation)}$$

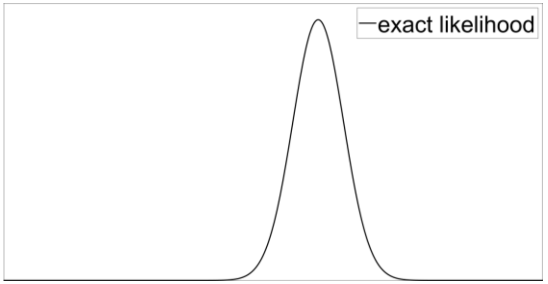
$$q_0(x) = \underset{q_0(x) = \mathcal{N}(x; \mu_0, \sigma_0^2)}{\arg\min} KL(f(x) q_1(x) || q(x))$$

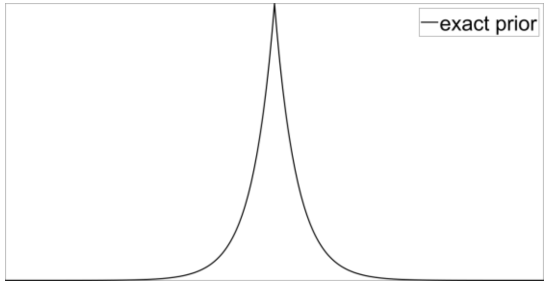
Toy example 2: 1d sparse prior (Gaussian likelihood + Laplace prior)

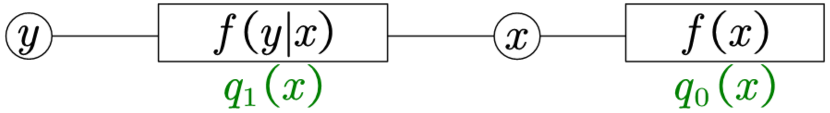
- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

Gaussian likelihood: $f(y|x) = \mathcal{N}(y; x, \sigma^2)$

Laplace prior: $f(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}$







- EP approximation: $q(x) \approx p(x|y)$ $\quad q(x) = \mathcal{N}(x; \mu, \sigma^2)$

$q_1(x) = f(x) = \mathcal{N}(x; y, \sigma^2)$ (no approximation)

$$q_0(x) = \underset{q_0(x) = \mathcal{N}(x; \mu_0, \sigma_0^2)}{\arg\min} KL(f(x) q_1(x) || q(x))$$

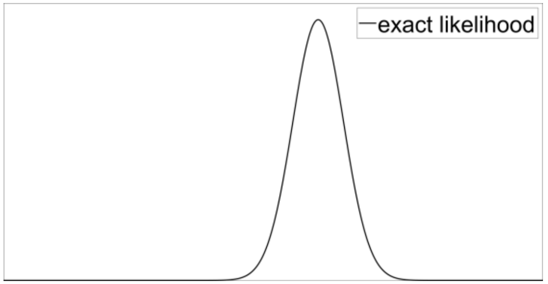
step 1. compute $\mathbb{E}_{tilted} \begin{bmatrix} x \\ x^2 \end{bmatrix}$ $\quad$ $tilted\ p_0(x) = f(x) q_1(x)$

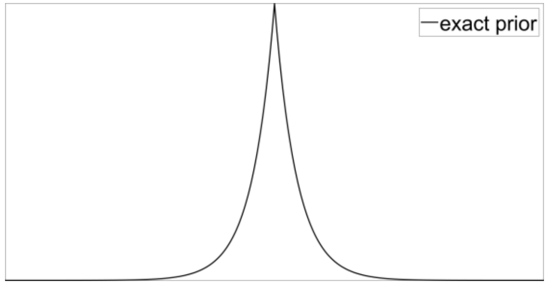
Toy example 2: 1d sparse prior (Gaussian likelihood + Laplace prior)

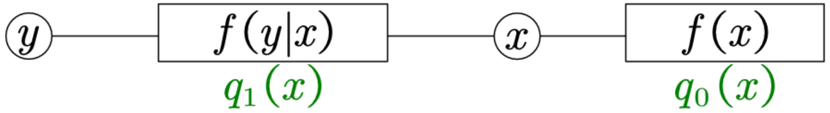
- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

Gaussian likelihood: $f(y|x) = \mathcal{N}(y; x, \sigma^2)$

Laplace prior: $f(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}$







- EP approximation: $q(x) \approx p(x|y)$ $\quad q(x) = \mathcal{N}(x; \mu, \sigma^2)$

$q_1(x) = f(x) = \mathcal{N}(x; y, \sigma^2)$ (no approximation)

$$q_0(x) = \underset{q_0(x) = \mathcal{N}(x; \mu_0, \sigma_0^2)}{\arg\min} KL(f(x) q_1(x) || q(x))$$

step 1. compute $\mathbb{E}_{tilted} \begin{bmatrix} x \\ x^2 \end{bmatrix}$ $\quad$ *tilted* $p_0(x) = f(x) q_1(x)$

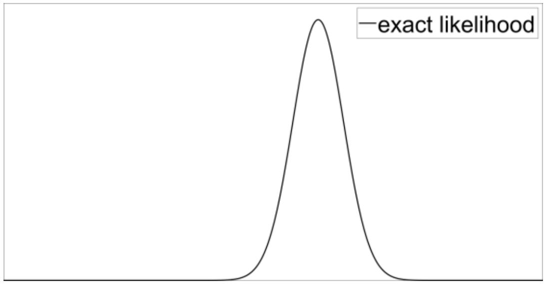
$$p_0(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}} \mathcal{N}(x; y, \sigma^2)$$
$$= \omega_- \mathcal{N}_{\mathbb{R}^-}(x; \mu_-, \sigma_-^2) + \omega_+ \mathcal{N}_{\mathbb{R}^+}(x; \mu_+, \sigma_+^2)$$

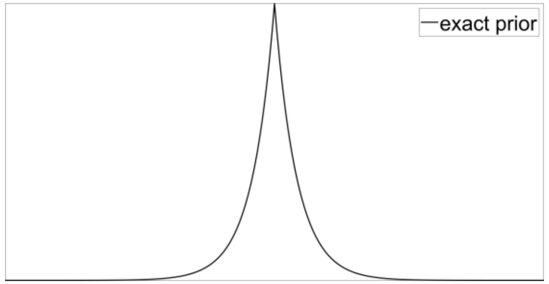
Toy example 2: 1d sparse prior (Gaussian likelihood + Laplace prior)

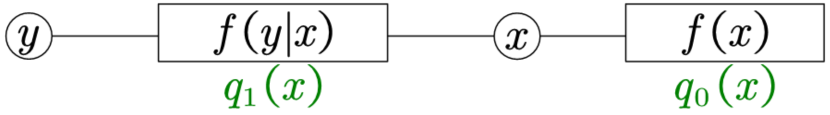
- Bayesian model: $p(x|y) \propto f(y|x) f(x)$

Gaussian likelihood: $f(y|x) = \mathcal{N}(y; x, \sigma^2)$

$$\text{Laplace prior: } f(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}$$







- EP approximation: $q(x) \approx p(x|y)$ $\quad q(x) = \mathcal{N}(x; \mu, \sigma^2)$

$q_1(x) = f(x) = \mathcal{N}(x; y, \sigma^2)$ (no approximation)

$$q_0(x) = \underset{q_0(x) = \mathcal{N}(x;\mu_0,\sigma_0^2)}{\arg\min} KL(f(x) q_1(x) || q(x))$$

step 1. compute $\mathbb{E}_{tilted} \begin{bmatrix} x \\ x^2 \end{bmatrix}$ $\quad$ *tilted* $p_0(x) = f(x) q_1(x)$

$$p_0(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}} \mathcal{N}(x; y, \sigma^2)$$

$$= \omega_- \mathcal{N}_{\mathbb{R}^-}(x; \mu_-, \sigma_-^2) + \omega_+ \mathcal{N}_{\mathbb{R}^+}(x; \mu_+, \sigma_+^2)$$

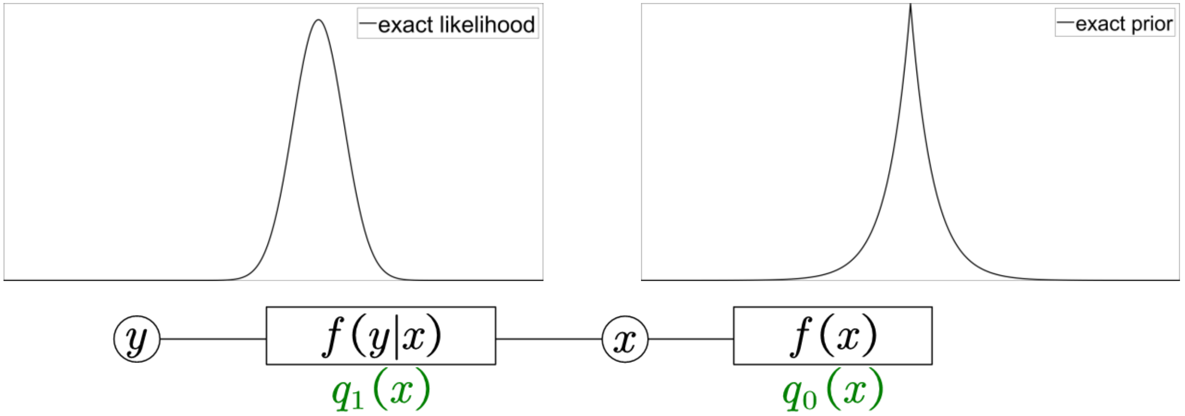
$$\Longrightarrow \mathbb{E}_{tilted} \begin{bmatrix} x \\ x^2 \end{bmatrix}$$

Toy example 2: 1d sparse prior (Gaussian likelihood + Laplace prior)

- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

Gaussian likelihood: $f(y|x) = \mathcal{N}(y; x, \sigma^2)$

Laplace prior: $f(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}$



- EP approximation: $q(x) \approx p(x|y)$ $\quad q(x) = \mathcal{N}(x; \mu, \sigma^2)$

$q_1(x) = f(x) = \mathcal{N}(x; y, \sigma^2)$ (no approximation)

$$q_0(x) = \underset{q_0(x) = \mathcal{N}(x;\mu_0,\sigma_0^2)}{\arg\min} KL(f(x)q_1(x) || q(x))$$

step 1. compute $\mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$ $\quad$ *tilted* $p_0(x) = f(x)q_1(x)$

$$p_0(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}} \mathcal{N}(x; y, \sigma^2)$$

$$= \omega_- \mathcal{N}_{\mathbb{R}^-}(x; \mu_-, \sigma_-^2) + \omega_+ \mathcal{N}_{\mathbb{R}^+}(x; \mu_+, \sigma_+^2)$$

$$\Longrightarrow \mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$$

step 2. moment matching $\mathbb{E}_q[x] = \mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$

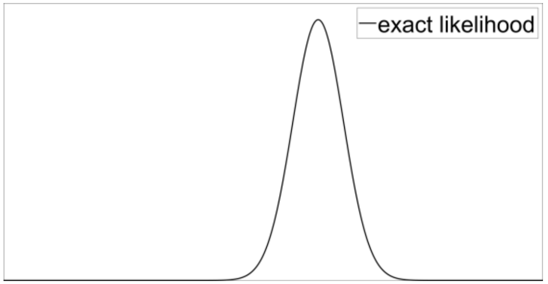
- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- **toy examples and applications**

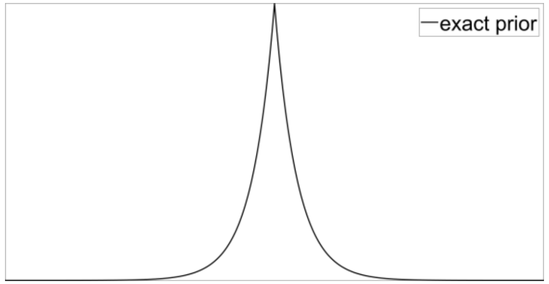
Toy example 2: 1d sparse prior (Gaussian likelihood + Laplace prior)

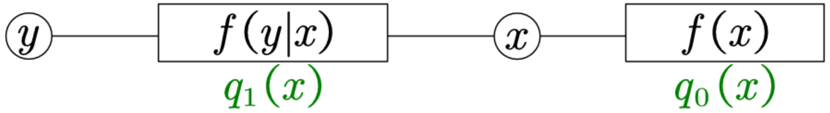
- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

$$\text{Gaussian likelihood: } f(y|x) = \mathcal{N}(y; x, \sigma^2)$$

$$\text{Laplace prior: } f(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}$$







- EP approximation: $q(x) \approx p(x|y)$ $q(x) = \mathcal{N}(x; \mu, \sigma^2)$

$$q_1(x) = f(x) = \mathcal{N}(x; y, \sigma^2) \text{ (no approximation)}$$

$$q_0(x) = \underset{q_0(x) = \mathcal{N}(x; \mu_0, \sigma_0^2)}{\arg\min} KL(f(x) q_1(x) || q(x))$$

step 1. compute $\mathbb{E}_{tilted} \begin{bmatrix} x \\ x^2 \end{bmatrix}$ *tilted* $p_0(x) = f(x) q_1(x)$

$$p_0(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}} \mathcal{N}(x; y, \sigma^2)$$

$$= \omega_- \mathcal{N}_{\mathbb{R}^-}(x; \mu_-, \sigma_-^2) + \omega_+ \mathcal{N}_{\mathbb{R}^+}(x; \mu_+, \sigma_+^2)$$

$$\Longrightarrow \mathbb{E}_{tilted} \begin{bmatrix} x \\ x^2 \end{bmatrix}$$

step 2. moment matching $\mathbb{E}_q[x] = \mathbb{E}_{tilted} \begin{bmatrix} x \\ x^2 \end{bmatrix}$

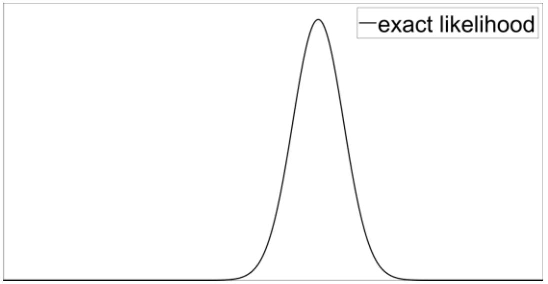
$$\Longrightarrow \begin{bmatrix} \mu \\ \mu^2 + \sigma^2 \end{bmatrix} = \mathbb{E}_{tilted} \begin{bmatrix} x \\ x^2 \end{bmatrix}$$

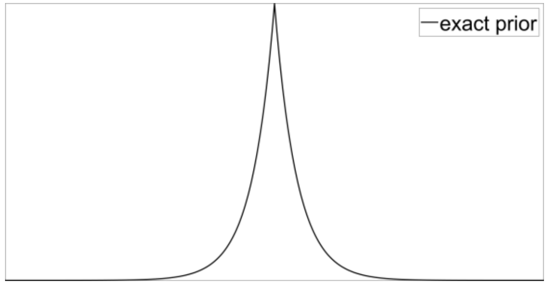
Toy example 2: 1d sparse prior (Gaussian likelihood + Laplace prior)

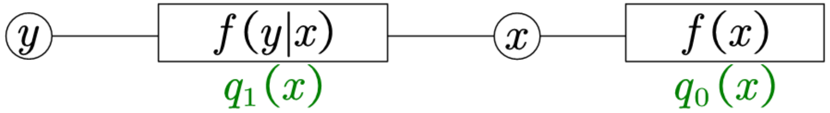
- Bayesian model: $p(x|y) \propto f(y|x)f(x)$

$$\text{Gaussian likelihood: } f(y|x) = \mathcal{N}(y; x, \sigma^2)$$

$$\text{Laplace prior: } f(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}$$







- EP approximation: $q(x) \approx p(x|y)$ $\quad q(x) = \mathcal{N}(x; \mu, \sigma^2)$

$$q_1(x) = f(x) = \mathcal{N}(x; y, \sigma^2) \ \text{(no approximation)}$$

$$q_0(x) = \underset{q_0(x) = \mathcal{N}(x; \mu_0, \sigma_0^2)}{\arg\min} KL(f(x) q_1(x) || q(x))$$

step 1. compute $\mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$ $\quad$ *tilted* $p_0(x) = f(x) q_1(x)$

$$p_0(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}} \mathcal{N}(x; y, \sigma^2)$$
$$= \omega_- \mathcal{N}_{\mathbb{R}^-}(x; \mu_-, \sigma_-^2) + \omega_+ \mathcal{N}_{\mathbb{R}^+}(x; \mu_+, \sigma_+^2)$$
$$\Longrightarrow \mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$$

step 2. moment matching $\mathbb{E}_q[x] = \mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$

$$\Longrightarrow \begin{bmatrix} \mu \\ \mu^2 + \sigma^2 \end{bmatrix} = \mathbb{E}_{tilted}\begin{bmatrix} x \\ x^2 \end{bmatrix}$$

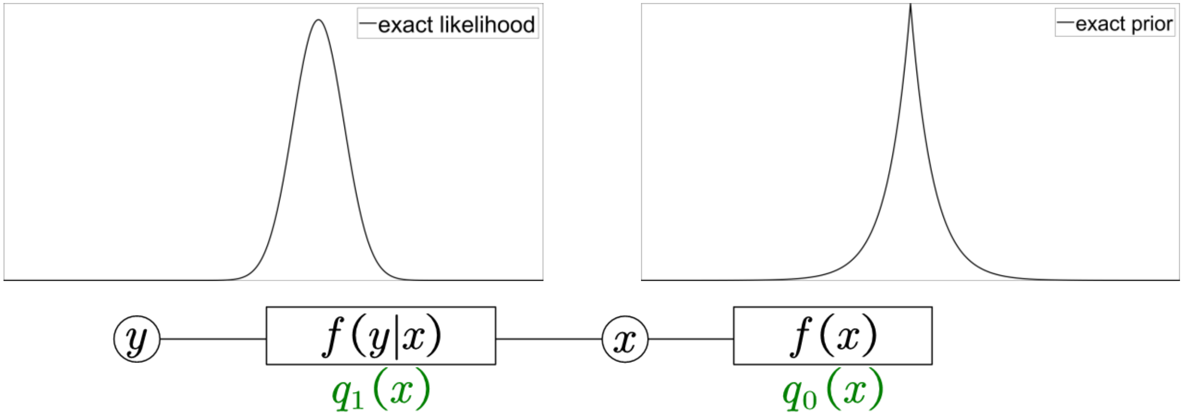
step 3. update $q_0(x) \propto \dfrac{q(x)}{q_1(x)}$

Toy example 2: 1d sparse prior (Gaussian likelihood + Laplace prior)

- Bayesian model: $p(x|y) \propto f(y|x) f(x)$

Gaussian likelihood: $f(y|x) = \mathcal{N}(y; x, \sigma^2)$

Laplace prior: $f(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}$



- EP approximation: $q(x) \approx p(x|y)$ $\quad q(x) = \mathcal{N}(x; \mu, \sigma^2)$

$q_1(x) = f(x) = \mathcal{N}(x; y, \sigma^2)$ (no approximation)

$$q_0(x) = \underset{q_0(x) = \mathcal{N}(x; \mu_0, \sigma_0^2)}{\arg\min} KL(f(x) q_1(x) || q(x))$$

step 1. compute $\mathbb{E}_{tilted} \begin{bmatrix} x \\ x^2 \end{bmatrix}$ $\quad$ *tilted* $p_0(x) = f(x) q_1(x)$

$$p_0(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}} \mathcal{N}(x; y, \sigma^2)$$

$$= \omega_- \mathcal{N}_{\mathbb{R}^-}(x; \mu_-, \sigma_-^2) + \omega_+ \mathcal{N}_{\mathbb{R}^+}(x; \mu_+, \sigma_+^2)$$

$$\Longrightarrow \mathbb{E}_{tilted} \begin{bmatrix} x \\ x^2 \end{bmatrix}$$

step 2. moment matching $\mathbb{E}_q[x] = \mathbb{E}_{tilted} \begin{bmatrix} x \\ x^2 \end{bmatrix}$

$$\Longrightarrow \begin{bmatrix} \mu \\ \mu^2 + \sigma^2 \end{bmatrix} = \mathbb{E}_{tilted} \begin{bmatrix} x \\ x^2 \end{bmatrix}$$

step 3. update $q_0(x) \propto \frac{q(x)}{q_1(x)}$

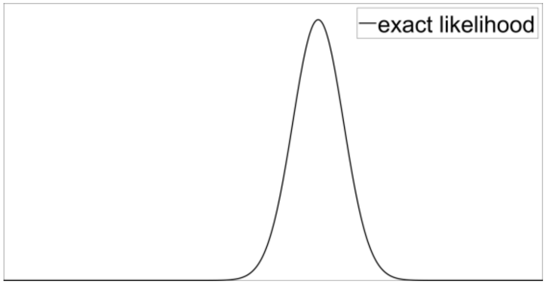
$$\Longrightarrow (\mu_0, \sigma_0^2)$$

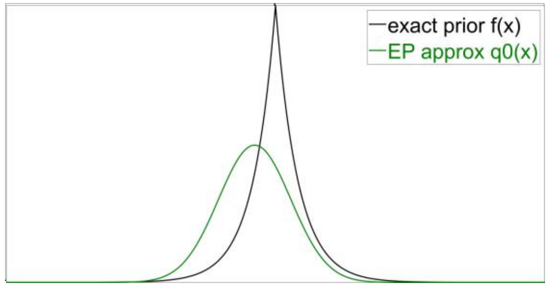
Toy example 2: 1d sparse prior (Gaussian likelihood + Laplace prior)

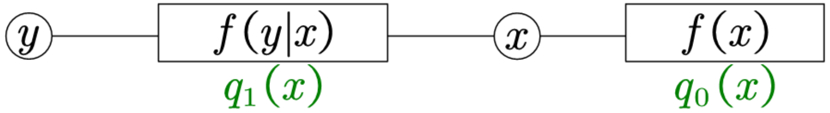
- Bayesian model: $p(x|y) \propto f(y|x) f(x)$

$$\text{Gaussian likelihood: } f(y|x) = \mathcal{N}(y; x, \sigma^2)$$

$$\text{Laplace prior: } f(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}$$







- EP approximation: $q(x) \approx p(x|y)$ $\qquad q(x) = \mathcal{N}(x; \mu, \sigma^2)$

$$q_1(x) = f(x) = \mathcal{N}(x; y, \sigma^2) \ \text{(no approximation)}$$

$$q_0(x) = \underset{q_0(x) = \mathcal{N}(x; \mu_0, \sigma_0^2)}{\arg\min} KL(f(x) q_1(x) || q(x))$$

step 1. compute $\mathbb{E}_{tilted} \begin{bmatrix} x \\ x^2 \end{bmatrix}$ $\qquad$ *tilted* $p_0(x) = f(x) q_1(x)$

$$p_0(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}} \mathcal{N}(x; y, \sigma^2)$$

$$= \omega_- \mathcal{N}_{\mathbb{R}^-}(x; \mu_-, \sigma_-^2) + \omega_+ \mathcal{N}_{\mathbb{R}^+}(x; \mu_+, \sigma_+^2)$$

$$\Longrightarrow \mathbb{E}_{tilted} \begin{bmatrix} x \\ x^2 \end{bmatrix}$$

step 2. moment matching $\mathbb{E}_q[x] = \mathbb{E}_{tilted} \begin{bmatrix} x \\ x^2 \end{bmatrix}$

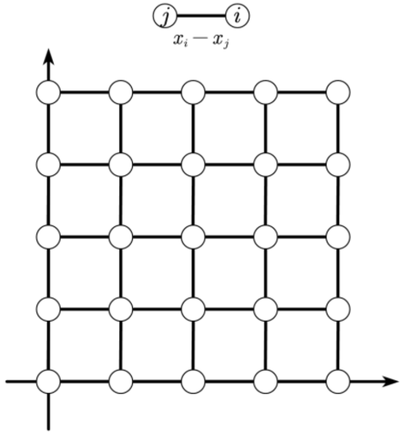
$$\Longrightarrow \begin{bmatrix} \mu \\ \mu^2 + \sigma^2 \end{bmatrix} = \mathbb{E}_{tilted} \begin{bmatrix} x \\ x^2 \end{bmatrix}$$

step 3. update $q_0(x) \propto \dfrac{q(x)}{q_1(x)}$

$$\Longrightarrow (\mu_0, \sigma_0^2)$$

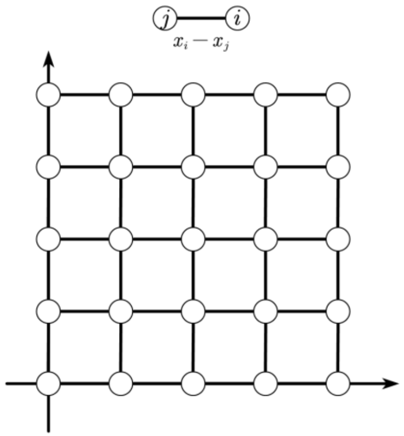
Application 2: fast scalable image restoration using EP with TV prior

Application 2: fast scalable image restoration using EP with TV prior



- **Image denosing/deconvolution/CS**

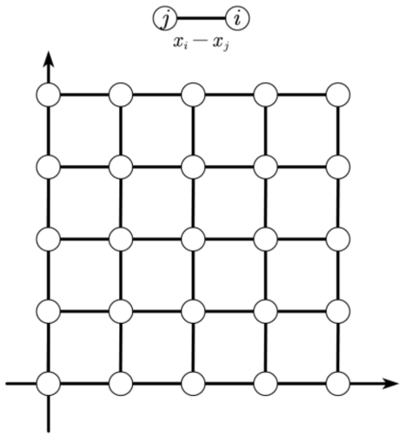
Application 2: fast scalable image restoration using EP with TV prior



- **Image denosing/deconvolution/CS**

$$\text{total variation prior: } f_x(\boldsymbol{x}|\boldsymbol{\theta}) \propto e^{-\lambda\, TV(\boldsymbol{x})} \qquad TV(\boldsymbol{x}) = \sum_{i,j} |x_{i+1,j} - x_{i,j}| + |x_{i,j+1} - x_{i,j}|$$

Application 2: fast scalable image restoration using EP with TV prior

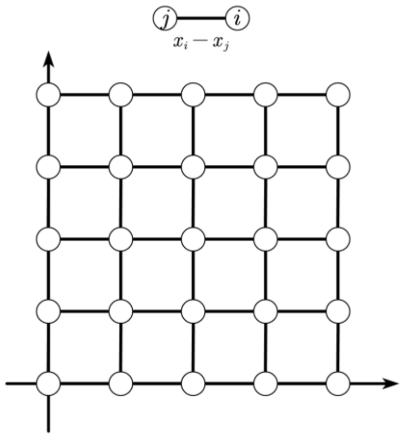


- **Image denosing/deconvolution/CS**

$$\text{total variation prior: } f_x(\boldsymbol{x}|\boldsymbol{\theta}) \propto e^{-\lambda\, TV(\boldsymbol{x})} \qquad TV(\boldsymbol{x}) = \sum_{i,j} |x_{i+1,j} - x_{i,j}| + |x_{i,j+1} - x_{i,j}|$$

$$\text{Gaussian likelihood: } f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y}; \mathbf{H}\boldsymbol{x}, \sigma^2) \text{ or } \mathcal{P}(\mathbf{H}\boldsymbol{x})$$

Application 2: fast scalable image restoration using EP with TV prior



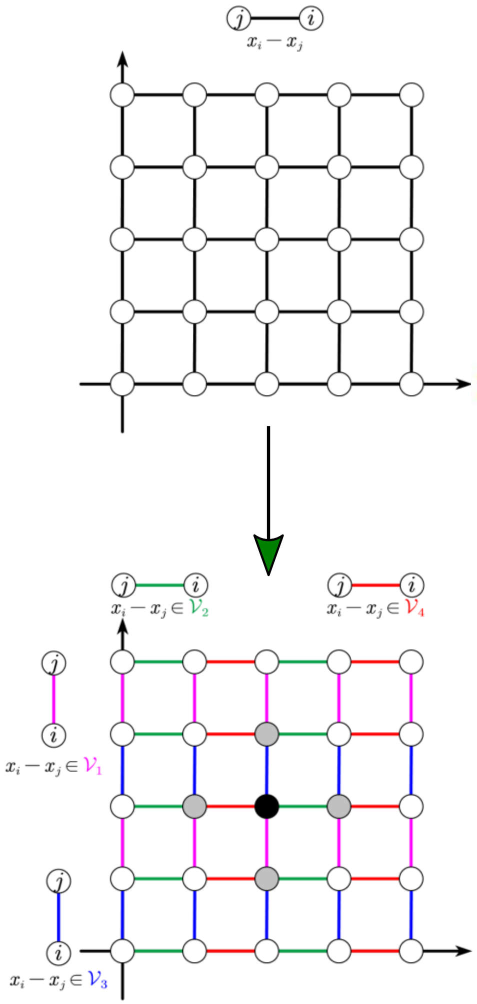
- **Image denosing/deconvolution/CS**

$$\text{total variation prior: } f_x(\boldsymbol{x}|\boldsymbol{\theta}) \propto e^{-\lambda\, TV(\boldsymbol{x})} \qquad TV(\boldsymbol{x}) = \sum_{i,j} |x_{i+1,j} - x_{i,j}| + |x_{i,j+1} - x_{i,j}|$$

$$\text{Gaussian likelihood: } f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y}; \mathbf{H}\boldsymbol{x}, \sigma^2) \ or \ \mathcal{P}(\mathbf{H}\boldsymbol{x})$$

$$\text{exact posterior: } p(\boldsymbol{x}|\boldsymbol{y}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) f_x(\boldsymbol{x}|\boldsymbol{\theta})$$

Application 2: fast scalable image restoration using EP with TV prior



- **Image denosing/deconvolution/CS**

$$\text{total variation prior: } f_x(\boldsymbol{x}|\boldsymbol{\theta}) \propto e^{-\lambda\, TV(\boldsymbol{x})} \qquad TV(\boldsymbol{x}) = \sum_{i,j} |x_{i+1,j} - x_{i,j}| + |x_{i,j+1} - x_{i,j}|$$

$$\text{Gaussian likelihood: } f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y}; \mathbf{H}\boldsymbol{x}, \sigma^2) \ or \ \mathcal{P}(\mathbf{H}\boldsymbol{x})$$

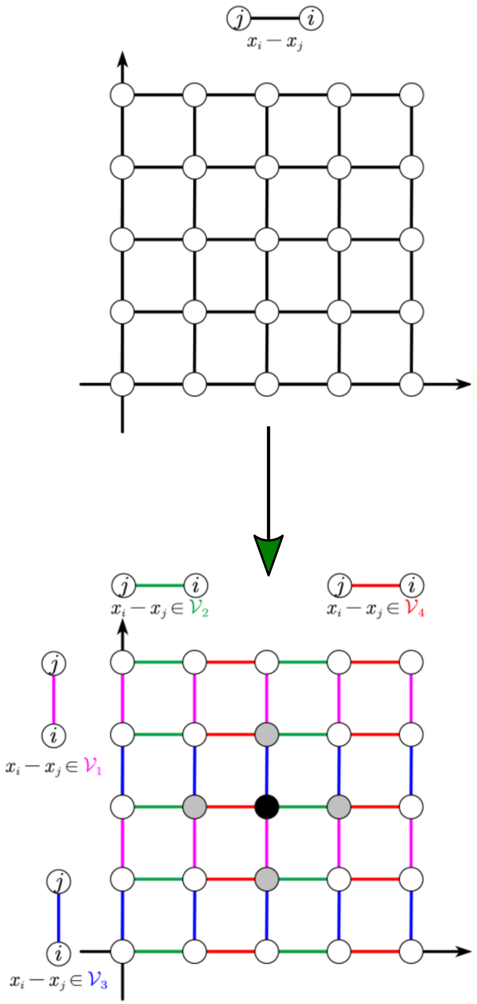
$$\text{exact posterior: } p(\boldsymbol{x}|\boldsymbol{y}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) f_x(\boldsymbol{x}|\boldsymbol{\theta})$$

how to factorize: $f_x(\boldsymbol{x}|\boldsymbol{\theta}) \propto \prod_{(i,j)\in\mathcal{V}} e^{-\lambda|x_i - x_j|}$

$$\propto \prod_{(i,j)\in\mathcal{V}_1} e^{-\lambda|x_i - x_j|} \prod_{(i,j)\in\mathcal{V}_2} e^{-\lambda|x_i - x_j|} \prod_{(i,j)\in\mathcal{V}_3} e^{-\lambda|x_i - x_j|} \prod_{(i,j)\in\mathcal{V}_4} e^{-\lambda|x_i - x_j|}$$

- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- **toy examples and applications**

Application 2: fast scalable image restoration using EP with TV prior



- **Image denosing/deconvolution/CS**

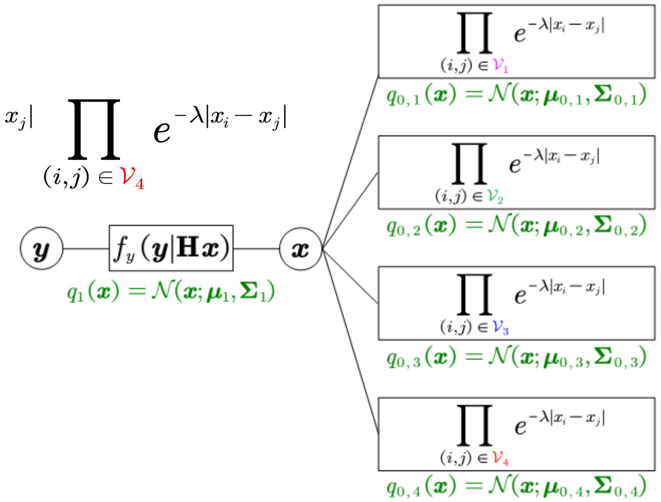
$$\text{total variation prior: } f_x(\boldsymbol{x}|\boldsymbol{\theta}) \propto e^{-\lambda TV(\boldsymbol{x})} \quad TV(\boldsymbol{x}) = \sum_{i,j} |x_{i+1,j} - x_{i,j}| + |x_{i,j+1} - x_{i,j}|$$

$$\text{Gaussian likelihood: } f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y}; \mathbf{H}\boldsymbol{x}, \sigma^2) \ or \ \mathcal{P}(\mathbf{H}\boldsymbol{x})$$

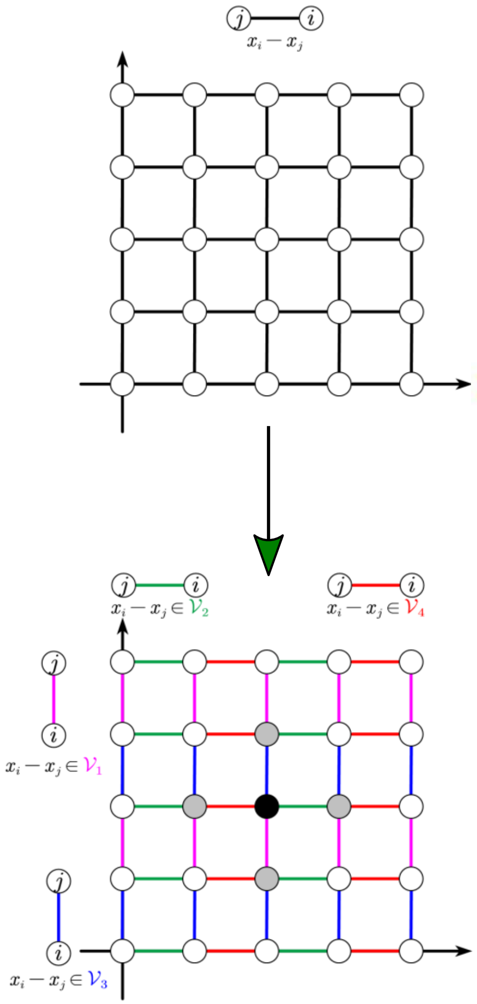
$$\text{exact posterior: } p(\boldsymbol{x}|\boldsymbol{y}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) f_x(\boldsymbol{x}|\boldsymbol{\theta})$$

how to factorize: $$f_x(\boldsymbol{x}|\boldsymbol{\theta}) \propto \prod_{(i,j)\in\mathcal{V}} e^{-\lambda|x_i - x_j|}$$

$$\propto \prod_{(i,j)\in\mathcal{V}_1} e^{-\lambda|x_i - x_j|} \prod_{(i,j)\in\mathcal{V}_2} e^{-\lambda|x_i - x_j|} \prod_{(i,j)\in\mathcal{V}_3} e^{-\lambda|x_i - x_j|} \prod_{(i,j)\in\mathcal{V}_4} e^{-\lambda|x_i - x_j|}$$



Application 2: fast scalable image restoration using EP with TV prior



- **Image denosing/deconvolution/CS**

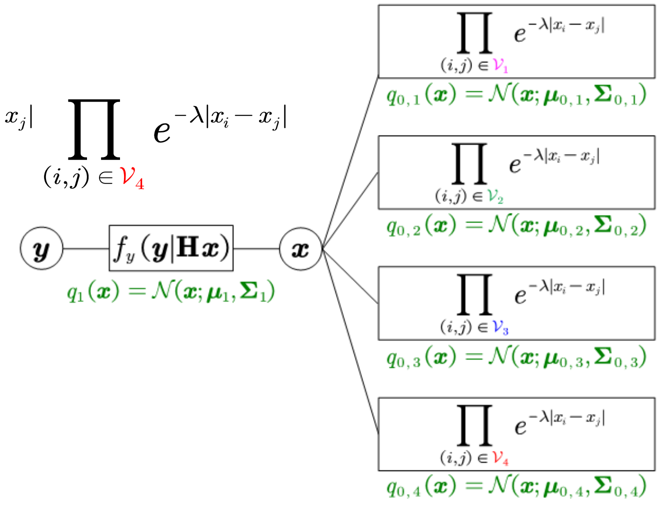
$$\text{total variation prior: } f_x(\boldsymbol{x}|\boldsymbol{\theta}) \propto e^{-\lambda TV(\boldsymbol{x})} \quad TV(\boldsymbol{x}) = \sum_{i,j} |x_{i+1,j} - x_{i,j}| + |x_{i,j+1} - x_{i,j}|$$

$$\text{Gaussian likelihood: } f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y}; \mathbf{H}\boldsymbol{x}, \sigma^2) \text{ or } \mathcal{P}(\mathbf{H}\boldsymbol{x})$$

$$\text{exact posterior: } p(\boldsymbol{x}|\boldsymbol{y}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) f_x(\boldsymbol{x}|\boldsymbol{\theta})$$

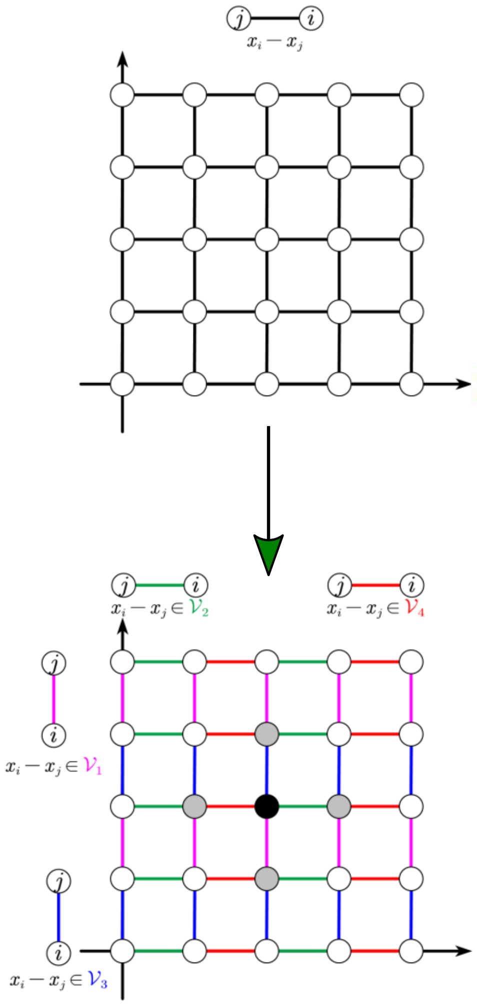
how to factorize: $f_x(\boldsymbol{x}|\boldsymbol{\theta}) \propto \prod_{(i,j)\in\mathcal{V}} e^{-\lambda|x_i - x_j|}$

$$\propto \prod_{(i,j)\in\mathcal{V}_1} e^{-\lambda|x_i - x_j|} \prod_{(i,j)\in\mathcal{V}_2} e^{-\lambda|x_i - x_j|} \prod_{(i,j)\in\mathcal{V}_3} e^{-\lambda|x_i - x_j|} \prod_{(i,j)\in\mathcal{V}_4} e^{-\lambda|x_i - x_j|}$$



how to compute:

Application 2: fast scalable image restoration using EP with TV prior



- **Image denosing/deconvolution/CS**

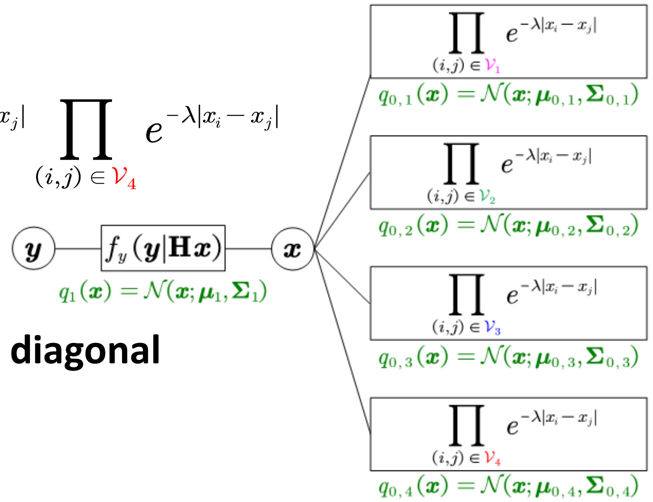
$$\text{total variation prior: } f_x(\boldsymbol{x}|\boldsymbol{\theta}) \propto e^{-\lambda\, TV(\boldsymbol{x})} \quad TV(\boldsymbol{x}) = \sum_{i,j} |x_{i+1,j} - x_{i,j}| + |x_{i,j+1} - x_{i,j}|$$

$$\text{Gaussian likelihood: } f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y}; \mathbf{H}\boldsymbol{x}, \sigma^2) \; or \; \mathcal{P}(\mathbf{H}\boldsymbol{x})$$

$$\text{exact posterior: } p(\boldsymbol{x}|\boldsymbol{y}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) f_x(\boldsymbol{x}|\boldsymbol{\theta})$$

how to factorize: $f_x(\boldsymbol{x}|\boldsymbol{\theta}) \propto \prod_{(i,j)\in\mathcal{V}} e^{-\lambda|x_i - x_j|}$

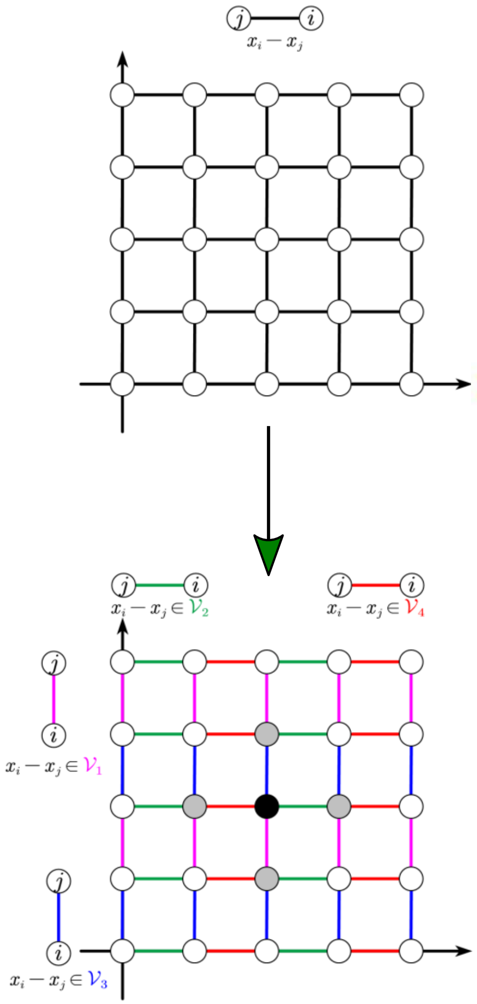
$$\propto \prod_{(i,j)\in\mathcal{V}_1} e^{-\lambda|x_i - x_j|} \prod_{(i,j)\in\mathcal{V}_2} e^{-\lambda|x_i - x_j|} \prod_{(i,j)\in\mathcal{V}_3} e^{-\lambda|x_i - x_j|} \prod_{(i,j)\in\mathcal{V}_4} e^{-\lambda|x_i - x_j|}$$



how to compute: 1. $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_{0,1}, \boldsymbol{\Sigma}_{0,2}, \boldsymbol{\Sigma}_{0,3}, \boldsymbol{\Sigma}_{0,4}$ are constrained to be **diagonal**

- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- **toy examples and applications**

Application 2: fast scalable image restoration using EP with TV prior



- **Image denosing/deconvolution/CS**

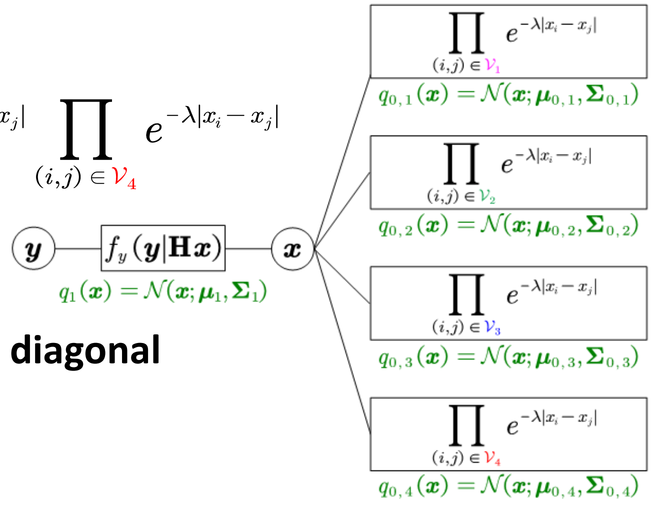
$$\text{total variation prior: } f_x(\boldsymbol{x}|\boldsymbol{\theta}) \propto e^{-\lambda\, TV(\boldsymbol{x})} \qquad TV(\boldsymbol{x}) = \sum_{i,j} |x_{i+1,j} - x_{i,j}| + |x_{i,j+1} - x_{i,j}|$$

$$\text{Gaussian likelihood: } f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y}; \mathbf{H}\boldsymbol{x}, \sigma^2) \ \ or\ \ \mathcal{P}(\mathbf{H}\boldsymbol{x})$$

$$\text{exact posterior: } p(\boldsymbol{x}|\boldsymbol{y}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) f_x(\boldsymbol{x}|\boldsymbol{\theta})$$

how to factorize: $f_x(\boldsymbol{x}|\boldsymbol{\theta}) \propto \prod_{(i,j)\in\mathcal{V}} e^{-\lambda|x_i - x_j|}$

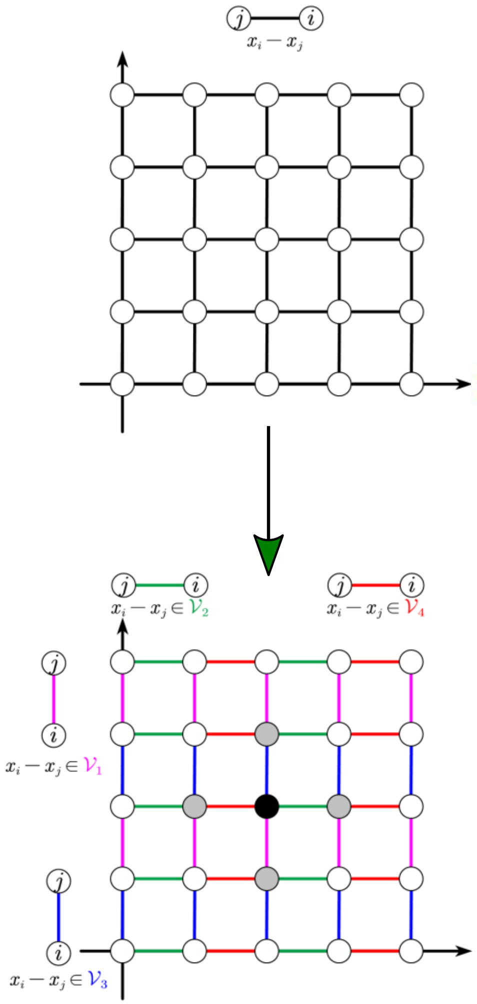
$$\propto \prod_{(i,j)\in\mathcal{V}_1} e^{-\lambda|x_i - x_j|} \prod_{(i,j)\in\mathcal{V}_2} e^{-\lambda|x_i - x_j|} \prod_{(i,j)\in\mathcal{V}_3} e^{-\lambda|x_i - x_j|} \prod_{(i,j)\in\mathcal{V}_4} e^{-\lambda|x_i - x_j|}$$



how to compute:

1. $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_{0,1}, \boldsymbol{\Sigma}_{0,2}, \boldsymbol{\Sigma}_{0,3}, \boldsymbol{\Sigma}_{0,4}$ are constrained to be **diagonal**
2. automatic estimation of hyperparameter $\boldsymbol{\theta} = \lambda$

Application 2: fast scalable image restoration using EP with TV prior



- **Image denosing/deconvolution/CS**

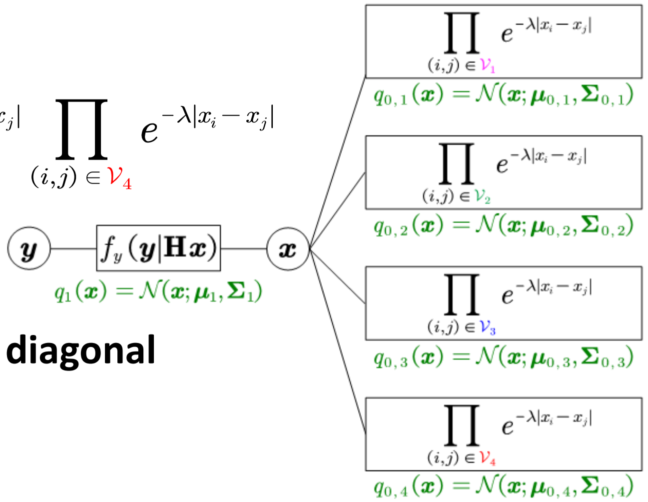
$$\text{total variation prior: } f_x(\boldsymbol{x}|\boldsymbol{\theta}) \propto e^{-\lambda\, TV(\boldsymbol{x})} \quad TV(\boldsymbol{x}) = \sum_{i,j} |x_{i+1,j} - x_{i,j}| + |x_{i,j+1} - x_{i,j}|$$

$$\text{Gaussian likelihood: } f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y}; \mathbf{H}\boldsymbol{x}, \sigma^2) \ or\ \mathcal{P}(\mathbf{H}\boldsymbol{x})$$

$$\text{exact posterior: } p(\boldsymbol{x}|\boldsymbol{y}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) f_x(\boldsymbol{x}|\boldsymbol{\theta})$$

how to factorize: $f_x(\boldsymbol{x}|\boldsymbol{\theta}) \propto \prod_{(i,j)\in\mathcal{V}} e^{-\lambda|x_i - x_j|}$

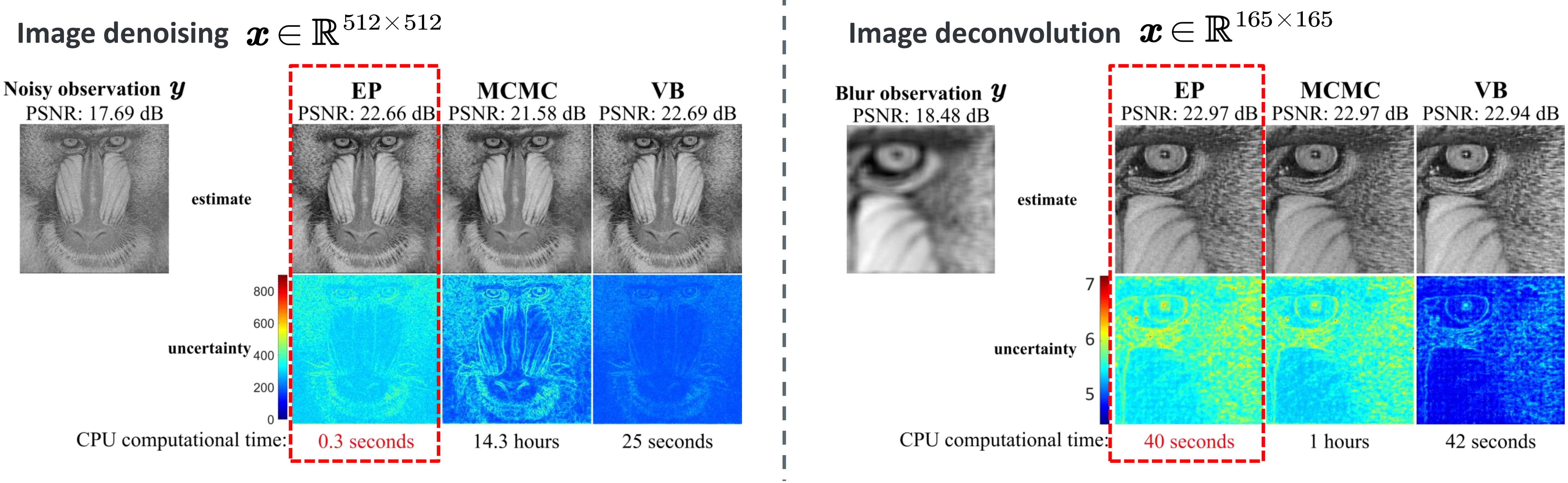
$$\propto \prod_{(i,j)\in\mathcal{V}_1} e^{-\lambda|x_i - x_j|} \prod_{(i,j)\in\mathcal{V}_2} e^{-\lambda|x_i - x_j|} \prod_{(i,j)\in\mathcal{V}_3} e^{-\lambda|x_i - x_j|} \prod_{(i,j)\in\mathcal{V}_4} e^{-\lambda|x_i - x_j|}$$



how to compute:
1. $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_{0,1}, \boldsymbol{\Sigma}_{0,2}, \boldsymbol{\Sigma}_{0,3}, \boldsymbol{\Sigma}_{0,4}$ are constrained to be **diagonal**
2. automatic estimation of hyperparameter $\boldsymbol{\theta} = \lambda$

EP scalable posterior approximation: $q(\boldsymbol{x}) \propto q_1(\boldsymbol{x}) q_{0,1}(\boldsymbol{x}) q_{0,2}(\boldsymbol{x}) q_{0,3}(\boldsymbol{x}) q_{0,4}(\boldsymbol{x}) \propto \mathcal{N}(;\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Application 2: fast scalable image restoration using EP with TV prior



Dan Yao, Stephen McLaughlin, and Yoann Altmann. "*Fast Scalable Image Restoration using Total Variation Priors and Expectation Propagation,*" Arxiv. https://doi.org/10.48550/arxiv.2110.01585

- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- **toy examples and applications**

Application 3: color image restoration in the low-photon count regime

Application 3: color image restoration in the low-photon count regime

observation $\boldsymbol{y}$

Application 3: color image restoration in the low-photon count regime

observation $\boldsymbol{y}$
single band

Application 3: color image restoration in the low-photon count regime

observation $\boldsymbol{y}$
single band
low photon-count

Application 3: color image restoration in the low-photon count regime

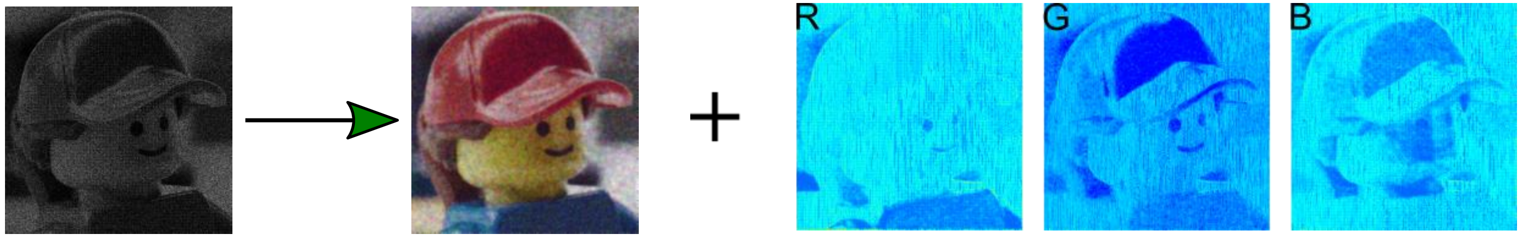
observation $\boldsymbol{y}$
single band
low photon-count
Poisson noise

Application 3: color image restoration in the low-photon count regime

observation $\boldsymbol{y}$
single band
low photon-count
Poisson noise

recovered color image

Application 3: color image restoration in the low-photon count regime



observation $\boldsymbol{y}$
single band
low photon-count
Poisson noise

recovered color image

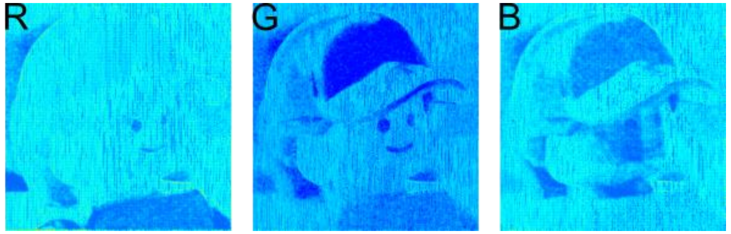
uncertainty

Application 3: color image restoration in the low-photon count regime

observation $\boldsymbol{y}$
single band
low photon-count
Poisson noise

recovered color image

+



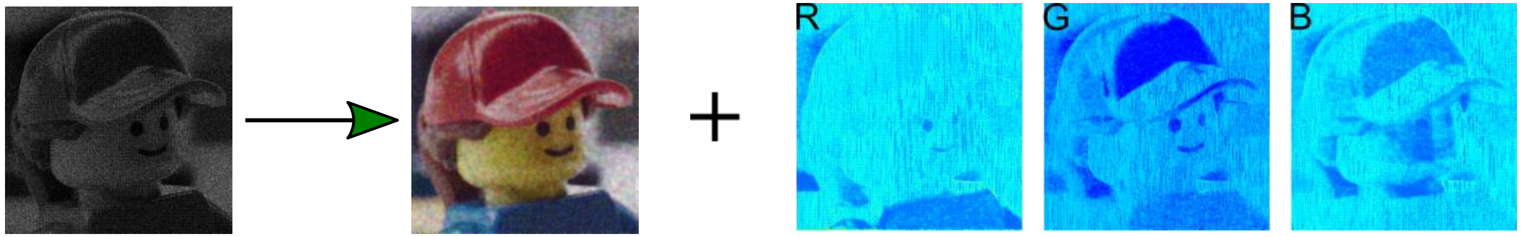
uncertainty

$$\ell_1-\text{norm TV prior: } f_x(\boldsymbol{x}|\boldsymbol{\theta}) \propto f_x(\boldsymbol{x}_R|\boldsymbol{\theta}_R)\, f_x(\boldsymbol{x}_G|\boldsymbol{\theta}_G)\, f_x(\boldsymbol{x}_B|\boldsymbol{\theta}_B)$$

3. Applications – EP for scalable imaging inverse problems

- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- **toy examples and applications**

Application 3: color image restoration in the low-photon count regime



observation $\boldsymbol{y}$
single band
low photon-count
Poisson noise

recovered color image

uncertainty

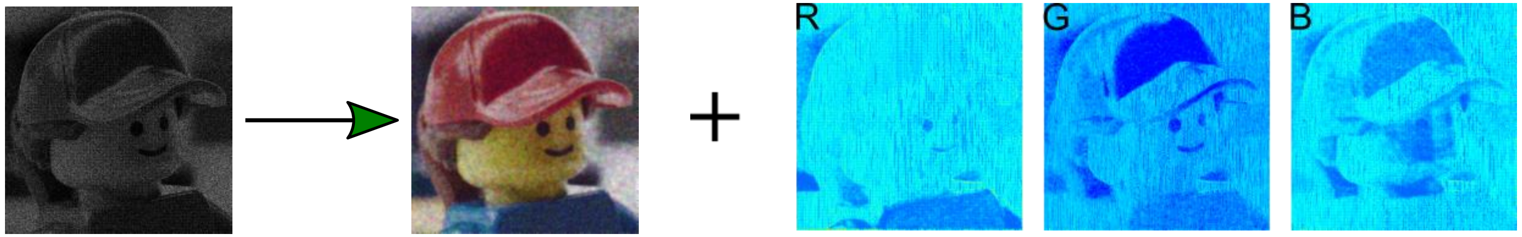
$$\ell_1-\text{norm TV prior}: f_x(\boldsymbol{x}|\boldsymbol{\theta}) \propto f_x(\boldsymbol{x}_R|\boldsymbol{\theta}_R)\, f_x(\boldsymbol{x}_G|\boldsymbol{\theta}_G)\, f_x(\boldsymbol{x}_B|\boldsymbol{\theta}_B)$$

$$\text{Poisson likelihood}: f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) = \mathcal{P}(\mathbf{H}_R\boldsymbol{x}_R + \mathbf{H}_G\boldsymbol{x}_G + \mathbf{H}_B\boldsymbol{x}_B)$$

3. Applications – EP for scalable imaging inverse problems

- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- **toy examples and applications**

Application 3: color image restoration in the low-photon count regime



observation $\boldsymbol{y}$
single band
low photon-count
Poisson noise

recovered color image

uncertainty

$$\ell_1-\text{norm TV prior:}\ f_x(\boldsymbol{x}|\boldsymbol{\theta}) \propto f_x(\boldsymbol{x}_R|\boldsymbol{\theta}_R)\, f_x(\boldsymbol{x}_G|\boldsymbol{\theta}_G)\, f_x(\boldsymbol{x}_B|\boldsymbol{\theta}_B)$$

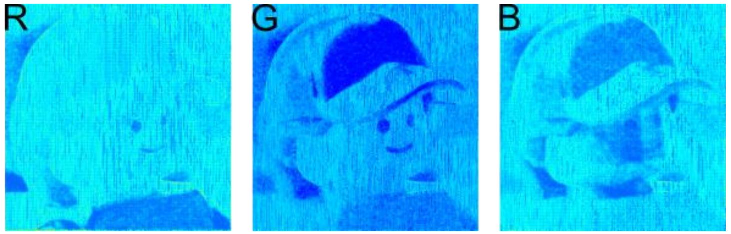
$$\text{Poisson likelihood:}\ f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) = \mathcal{P}(\mathbf{H}_R\boldsymbol{x}_R + \mathbf{H}_G\boldsymbol{x}_G + \mathbf{H}_B\boldsymbol{x}_B)$$

$$\text{exact posterior:}\ p(\boldsymbol{x}_R, \boldsymbol{x}_G, \boldsymbol{x}_B|\boldsymbol{y}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})$$

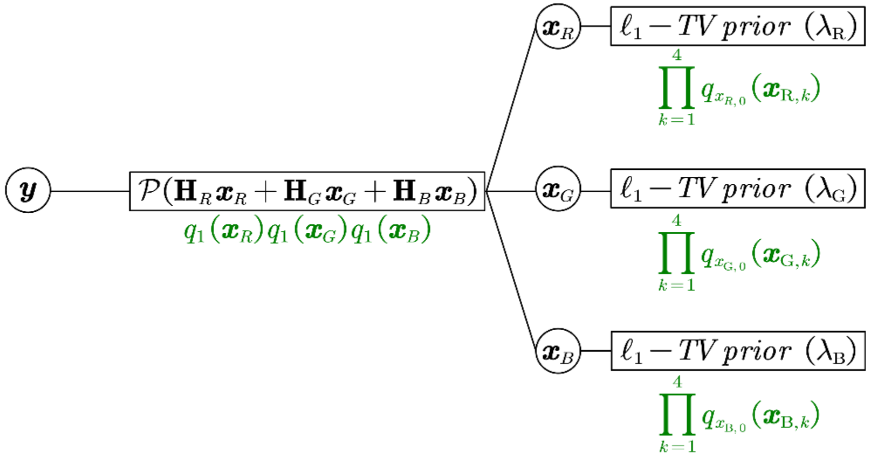
Application 3: color image restoration in the low-photon count regime

observation $\boldsymbol{y}$
single band
low photon-count
Poisson noise

recovered color image



uncertainty



$$\ell_1 - \text{norm TV prior}: f_x(\boldsymbol{x}|\boldsymbol{\theta}) \propto f_x(\boldsymbol{x}_R|\boldsymbol{\theta}_R) f_x(\boldsymbol{x}_G|\boldsymbol{\theta}_G) f_x(\boldsymbol{x}_B|\boldsymbol{\theta}_B)$$

$$\text{Poisson likelihood}: f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) = \mathcal{P}(\mathbf{H}_R\boldsymbol{x}_R + \mathbf{H}_G\boldsymbol{x}_G + \mathbf{H}_B\boldsymbol{x}_B)$$

$$\text{exact posterior}: p(\boldsymbol{x}_R, \boldsymbol{x}_G, \boldsymbol{x}_B|\boldsymbol{y}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) f_x(\boldsymbol{x}|\boldsymbol{\theta})$$

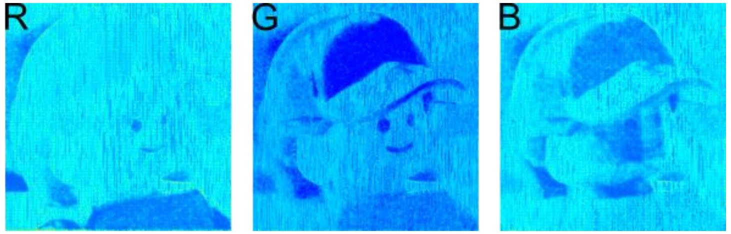
- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- **toy examples and applications**

Application 3: color image restoration in the low-photon count regime

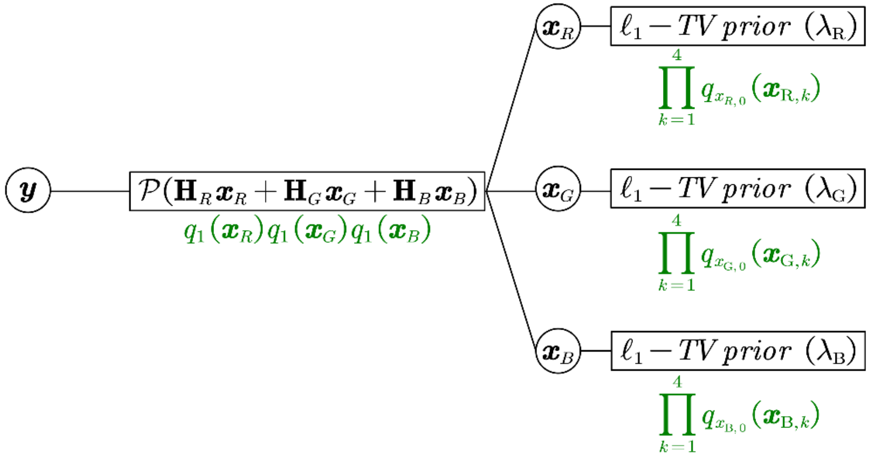
observation $\boldsymbol{y}$
single band
low photon-count
Poisson noise

recovered color image

+



uncertainty



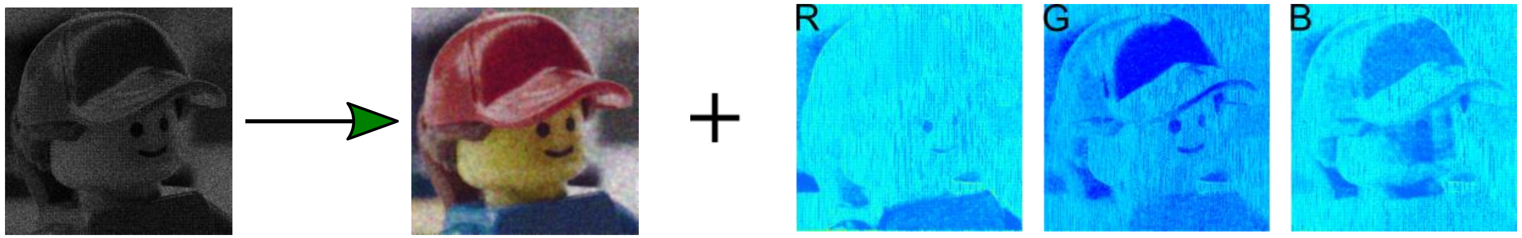
$$\ell_1 - \text{norm TV prior:}\ f_x(\boldsymbol{x}|\boldsymbol{\theta}) \propto f_x(\boldsymbol{x}_R|\boldsymbol{\theta}_R) f_x(\boldsymbol{x}_G|\boldsymbol{\theta}_G) f_x(\boldsymbol{x}_B|\boldsymbol{\theta}_B)$$

$$\text{Poisson likelihood:}\ f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) = \mathcal{P}(\mathbf{H}_R\boldsymbol{x}_R + \mathbf{H}_G\boldsymbol{x}_G + \mathbf{H}_B\boldsymbol{x}_B)$$

$$\text{exact posterior:}\ p(\boldsymbol{x}_R, \boldsymbol{x}_G, \boldsymbol{x}_B|\boldsymbol{y}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) f_x(\boldsymbol{x}|\boldsymbol{\theta})$$

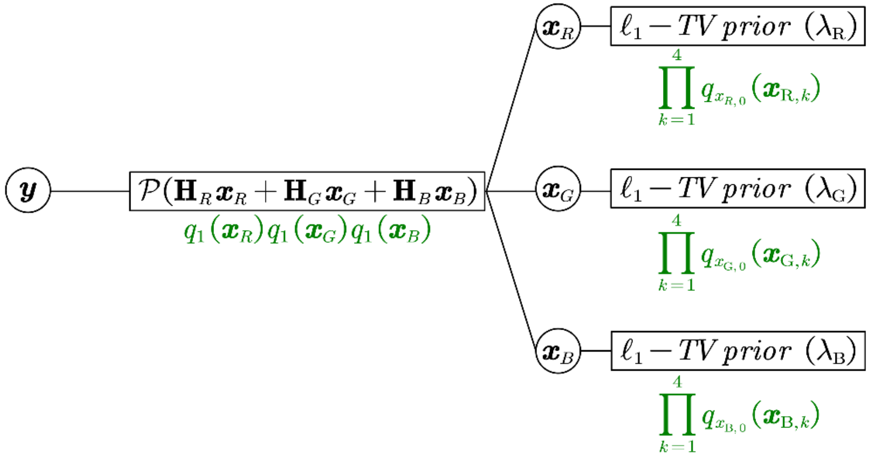
scalability:

Application 3: color image restoration in the low-photon count regime



observation $\boldsymbol{y}$ | recovered color image | uncertainty

single band
low photon-count
Poisson noise



$$\ell_1 - \text{norm TV prior: } f_x(\boldsymbol{x}|\boldsymbol{\theta}) \propto f_x(\boldsymbol{x}_R|\boldsymbol{\theta}_R)\, f_x(\boldsymbol{x}_G|\boldsymbol{\theta}_G)\, f_x(\boldsymbol{x}_B|\boldsymbol{\theta}_B)$$

$$\text{Poisson likelihood: } f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) = \mathcal{P}(\mathbf{H}_R\boldsymbol{x}_R + \mathbf{H}_G\boldsymbol{x}_G + \mathbf{H}_B\boldsymbol{x}_B)$$

$$\text{exact posterior: } p(\boldsymbol{x}_R, \boldsymbol{x}_G, \boldsymbol{x}_B|\boldsymbol{y}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\boldsymbol{\theta})$$

scalability:

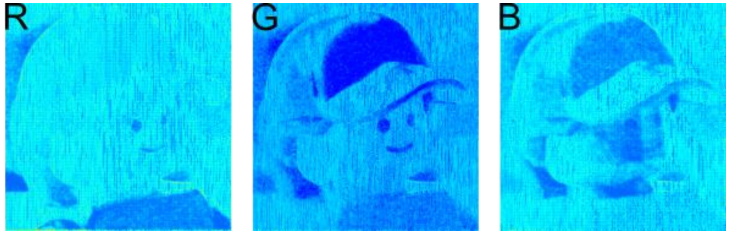
1. parallel update of approximating factors over R,G,B channels in likelihood
2. parallel update of approximating factor over R,G,B channels in the prior
3. automatic hyperparameter estimation over R,G,B channels

Application 3: color image restoration in the low-photon count regime

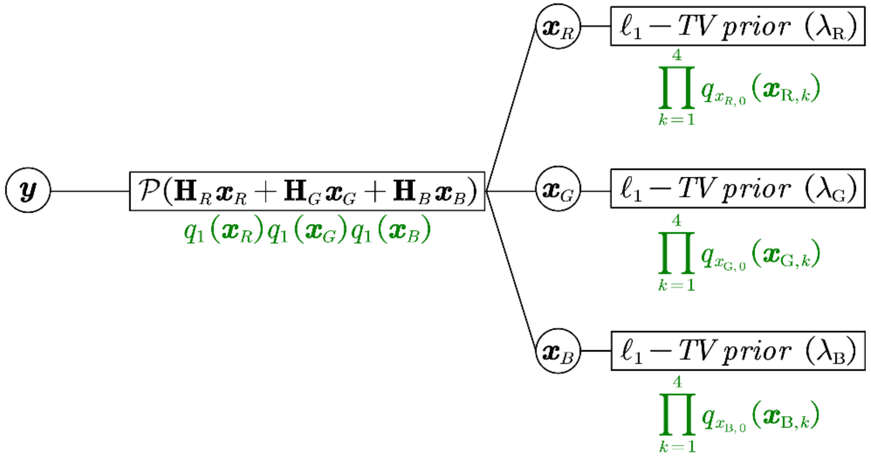
observation $\boldsymbol{y}$
single band
low photon-count
Poisson noise

recovered color image

+



uncertainty



$$\ell_1 - \text{norm TV prior:}\ f_x(\boldsymbol{x}|\boldsymbol{\theta}) \propto f_x(\boldsymbol{x}_R|\boldsymbol{\theta}_R) f_x(\boldsymbol{x}_G|\boldsymbol{\theta}_G) f_x(\boldsymbol{x}_B|\boldsymbol{\theta}_B)$$

$$\text{Poisson likelihood:}\ f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) = \mathcal{P}(\mathbf{H}_R\boldsymbol{x}_R + \mathbf{H}_G\boldsymbol{x}_G + \mathbf{H}_B\boldsymbol{x}_B)$$

$$\text{exact posterior:}\ p(\boldsymbol{x}_R, \boldsymbol{x}_G, \boldsymbol{x}_B|\boldsymbol{y}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}) f_x(\boldsymbol{x}|\boldsymbol{\theta})$$

scalability:

1. parallel update of approximating factors over R,G,B channels in likelihood
2. parallel update of approximating factor over R,G,B channels in the prior
3. automatic hyperparameter estimation over R,G,B channels

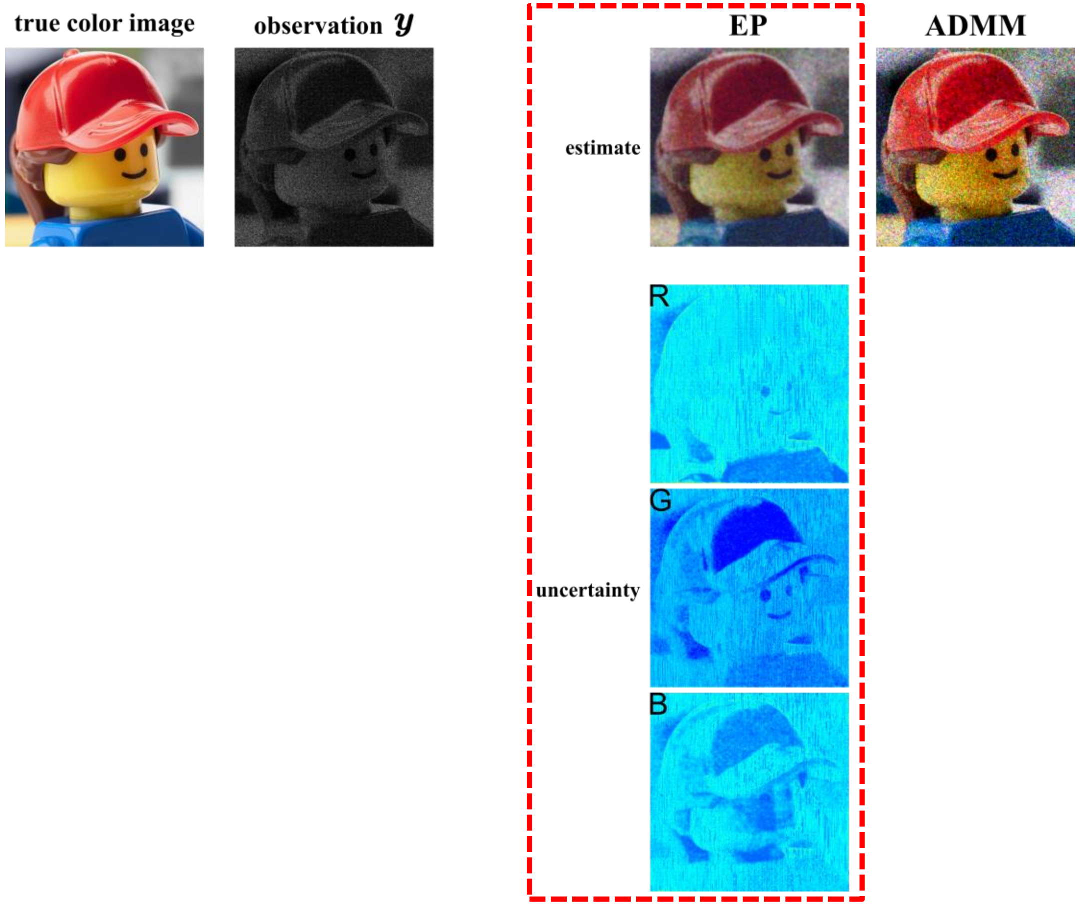
$$\text{EP scalable posterior approximation:}\ q(\boldsymbol{x}_R) \propto q_1(x_R) \prod_{k=1}^{4} q_{R,0}(\boldsymbol{x}_{R,k}) \propto \mathcal{N}(;\boldsymbol{\mu}_R, \boldsymbol{\Sigma}_R)$$

$$q(\boldsymbol{x}_G) \propto q_1(x_G) \prod_{k=1}^{4} q_{G,0}(\boldsymbol{x}_{G,k}) \propto \mathcal{N}(;\boldsymbol{\mu}_G, \boldsymbol{\Sigma}_G)$$

$$q(\boldsymbol{x}_B) \propto q_1(x_B) \prod_{k=1}^{4} q_{B,0}(\boldsymbol{x}_{B,k}) \propto \mathcal{N}(;\boldsymbol{\mu}_B, \boldsymbol{\Sigma}_B)$$

Application 3: color image restoration in the low-photon count regime

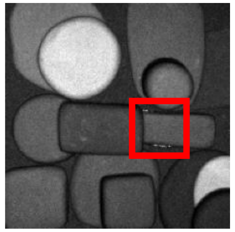
$$\boldsymbol{x} \in \mathbb{R}^{512\times512\times3} \text{ (channels)}$$



Dan Yao, Stephen McLaughlin, and Yoann Altmann. "*Color Image Restoration in the Low Photon-Count Regime using Expectation Propagation,*" IEEE International Conference in Image Processing 2022 (accepted).

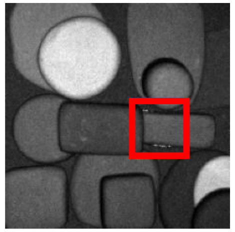
Application 4: multispectral Lidar data anomaly detection

Application 4: multispectral Lidar data anomaly detection



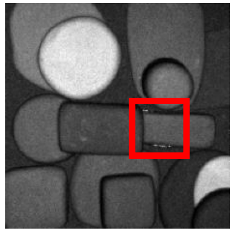
observation $\boldsymbol{y}$

Application 4: multispectral Lidar data anomaly detection



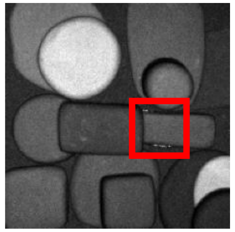
observation $\boldsymbol{y}$
multispectral Lidar data

Application 4: multispectral Lidar data anomaly detection



observation $\boldsymbol{y}$
multispectral Lidar data
low photon-count

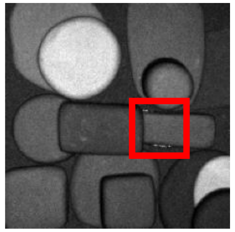
Application 4: multispectral Lidar data anomaly detection



observation $\boldsymbol{y}$
multispectral Lidar data

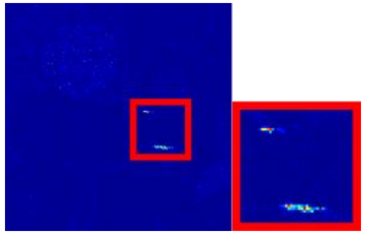
low photon-count
Poisson noise

Application 4: multispectral Lidar data anomaly detection



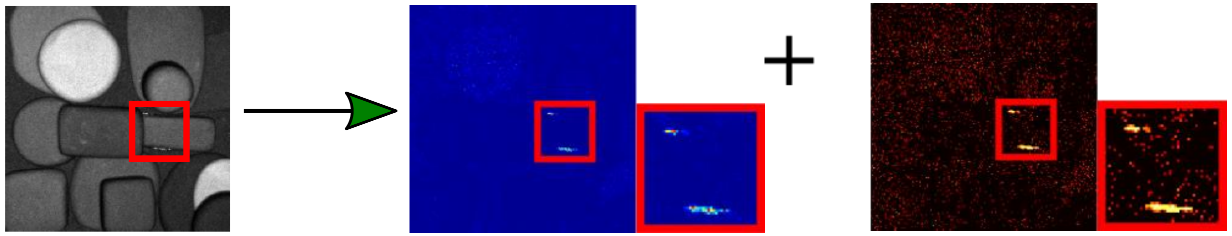
observation $\boldsymbol{y}$
multispectral Lidar data

low photon-count
Poisson noise



detected anomalies

Application 4: multispectral Lidar data anomaly detection



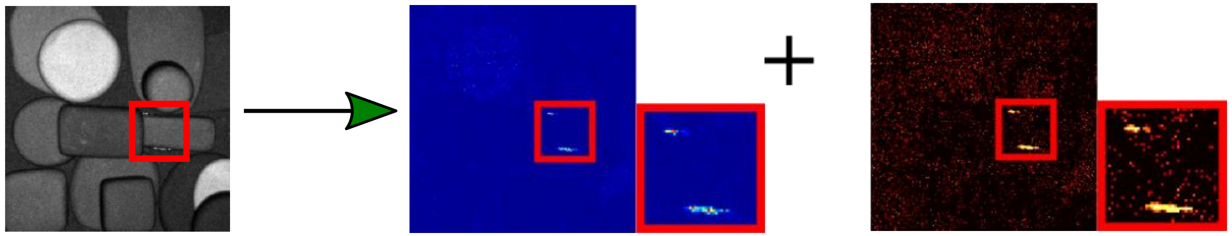
observation $\boldsymbol{y}$
multispectral Lidar data

low photon-count
Poisson noise

detected anomalies

uncertainty

Application 4: multispectral Lidar data anomaly detection



observation $\boldsymbol{y}$ — detected anomalies — uncertainty

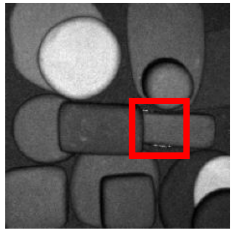
multispectral Lidar data

low photon-count

Poisson noise

$$\text{Poisson likelihood: } f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x},\boldsymbol{z},\boldsymbol{r}) = \mathcal{P}((1-\boldsymbol{z})\odot\mathbf{H}\boldsymbol{x}+\boldsymbol{z}\odot\boldsymbol{r})$$

Application 4: multispectral Lidar data anomaly detection

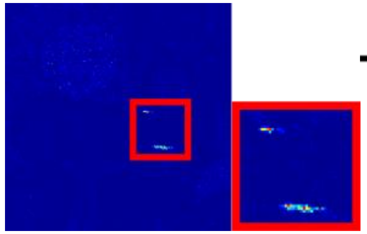


observation $\boldsymbol{y}$

multispectral Lidar data

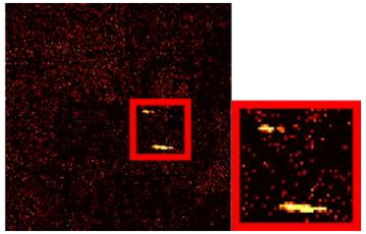
low photon-count

Poisson noise



detected anomalies

+

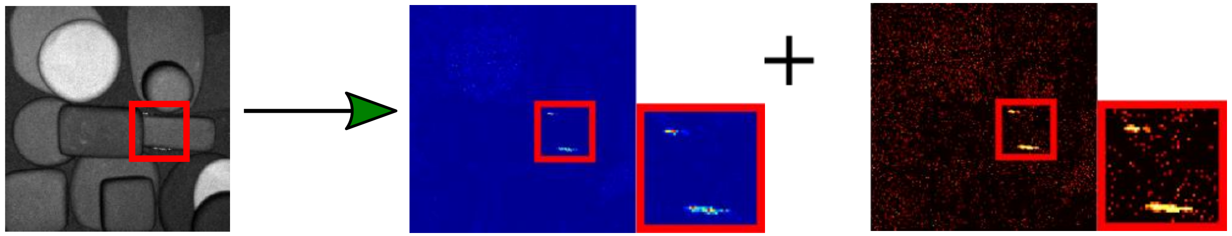


uncertainty

$$\text{Poisson likelihood: } f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x},\boldsymbol{z},\boldsymbol{r}) = \mathcal{P}((1-\boldsymbol{z})\odot\mathbf{H}\boldsymbol{x}+\boldsymbol{z}\odot\boldsymbol{r})$$

$$\text{positive \& sparse \& Bernoulli priors: } f_x(\boldsymbol{x}|\theta_x),\ f_r(\boldsymbol{r}|\theta_r),\ f_z(\boldsymbol{z}|\theta_z)$$

Application 4: multispectral Lidar data anomaly detection



observation $\boldsymbol{y}$ — detected anomalies — uncertainty

multispectral Lidar data

low photon-count

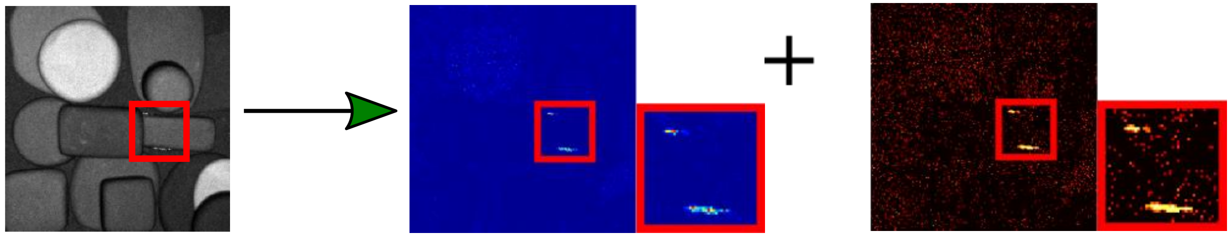
Poisson noise

$$\text{Poisson likelihood: } f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x},\boldsymbol{z},\boldsymbol{r}) = \mathcal{P}((1-\boldsymbol{z})\odot\mathbf{H}\boldsymbol{x}+\boldsymbol{z}\odot\boldsymbol{r})$$

$$\text{positive \& sparse \& Bernoulli priors: } f_x(\boldsymbol{x}|\theta_x),\ f_r(\boldsymbol{r}|\theta_r),\ f_z(\boldsymbol{z}|\theta_z)$$

$$\text{exact posterior: } p(\boldsymbol{x},\boldsymbol{v},\boldsymbol{z}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\theta_x)\, f_{v,z}(\boldsymbol{v},\boldsymbol{z}|\theta_v,\theta_z)$$

Application 4: multispectral Lidar data anomaly detection

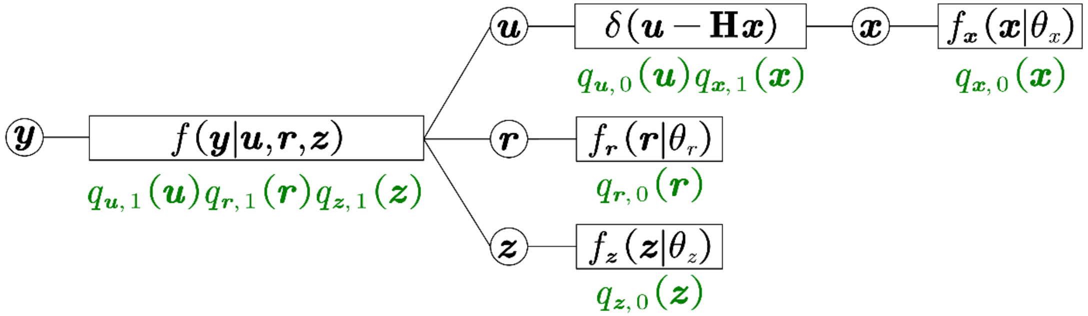


observation $\boldsymbol{y}$ — detected anomalies + uncertainty

multispectral Lidar data

low photon-count

Poisson noise

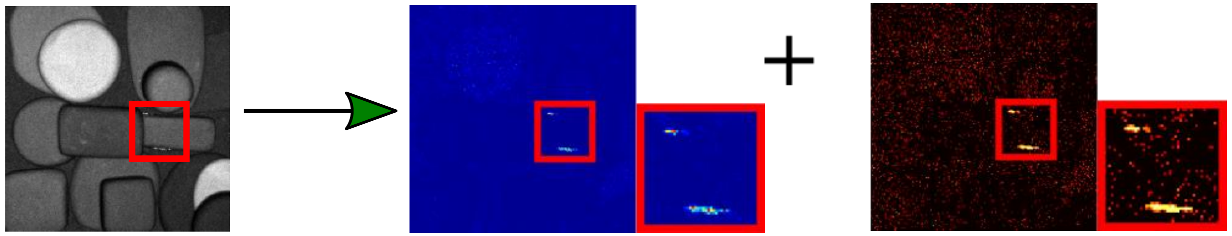


$$\text{Poisson likelihood: } f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x},\boldsymbol{z},\boldsymbol{r}) = \mathcal{P}((1-\boldsymbol{z})\odot\mathbf{H}\boldsymbol{x}+\boldsymbol{z}\odot\boldsymbol{r})$$

$$\text{positive \& sparse \& Bernoulli priors: } f_x(\boldsymbol{x}|\theta_x),\ f_r(\boldsymbol{r}|\theta_r),\ f_z(\boldsymbol{z}|\theta_z)$$

$$\text{exact posterior: } p(\boldsymbol{x},\boldsymbol{v},\boldsymbol{z}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\,f_x(\boldsymbol{x}|\theta_x)\,f_{v,z}(\boldsymbol{v},\boldsymbol{z}|\theta_v,\theta_z)$$

Application 4: multispectral Lidar data anomaly detection

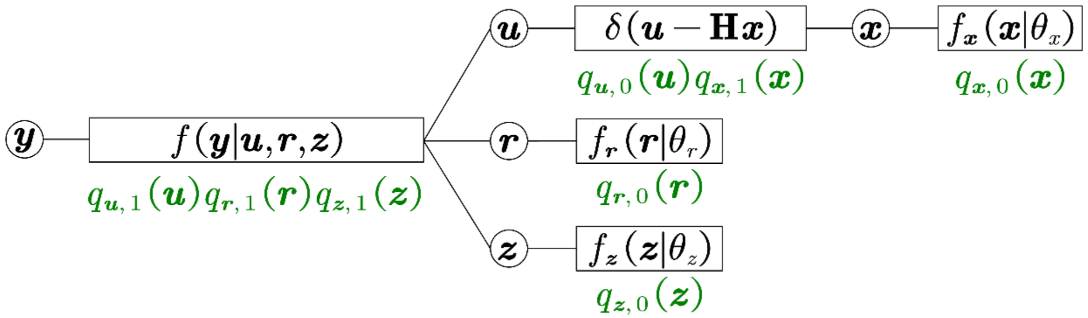


observation $\boldsymbol{y}$ — detected anomalies + uncertainty

multispectral Lidar data

low photon-count

Poisson noise



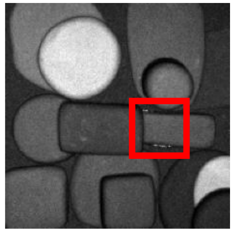
Poisson likelihood: $f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x},\boldsymbol{z},\boldsymbol{r}) = \mathcal{P}((1-\boldsymbol{z})\odot\mathbf{H}\boldsymbol{x}+\boldsymbol{z}\odot\boldsymbol{r})$

positive & sparse & Bernoulli priors: $f_x(\boldsymbol{x}|\theta_x),\ f_r(\boldsymbol{r}|\theta_r),\ f_z(\boldsymbol{z}|\theta_z)$

exact posterior: $p(\boldsymbol{x},\boldsymbol{v},\boldsymbol{z}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})f_x(\boldsymbol{x}|\theta_x)f_{v,z}(\boldsymbol{v},\boldsymbol{z}|\theta_v,\theta_z)$

scalability:

Application 4: multispectral Lidar data anomaly detection

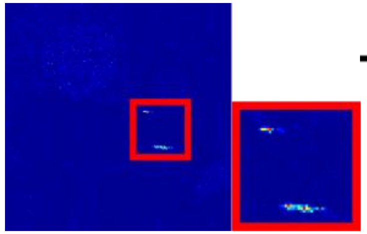


observation $\boldsymbol{y}$

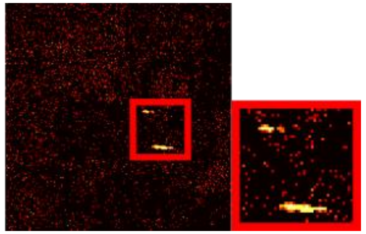
multispectral Lidar data

low photon-count

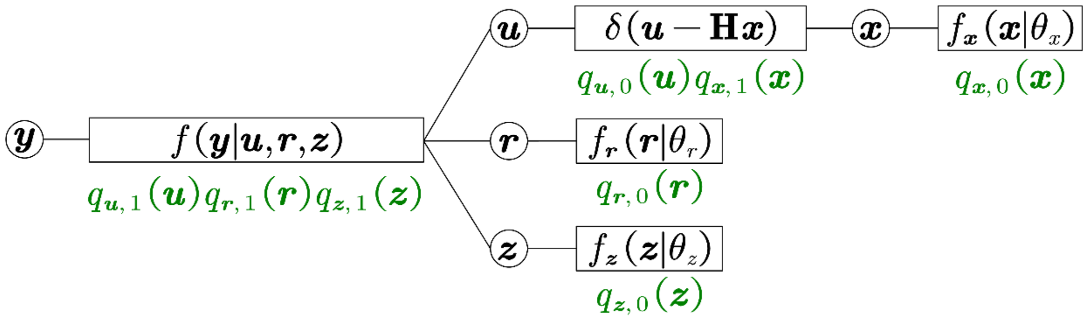
Poisson noise



detected anomalies



uncertainty



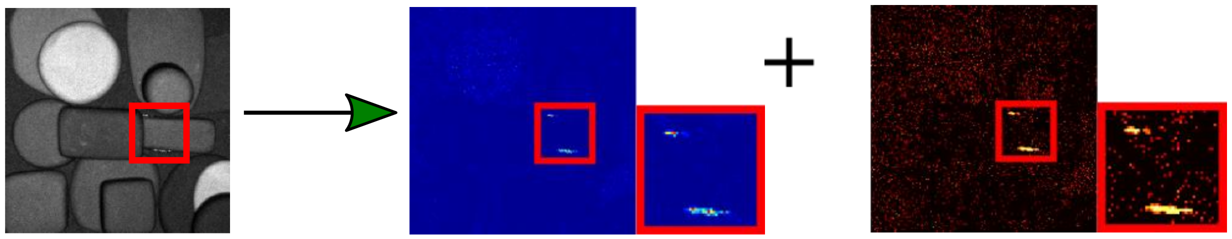
$$\text{Poisson likelihood: } f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x},\boldsymbol{z},\boldsymbol{r}) = \mathcal{P}((1-\boldsymbol{z})\odot\mathbf{H}\boldsymbol{x}+\boldsymbol{z}\odot\boldsymbol{r})$$

$$\text{positive \& sparse \& Bernoulli priors: } f_x(\boldsymbol{x}|\theta_x),\ f_r(\boldsymbol{r}|\theta_r),\ f_z(\boldsymbol{z}|\theta_z)$$

$$\text{exact posterior: } p(\boldsymbol{x},\boldsymbol{v},\boldsymbol{z}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\theta_x)\, f_{v,z}(\boldsymbol{v},\boldsymbol{z}|\theta_v,\theta_z)$$

scalability:
1. combining full, diagonal, isotropic covariance matrices for flexible and efficient approximation
2. automatic hyperparameter estimation

Application 4: multispectral Lidar data anomaly detection

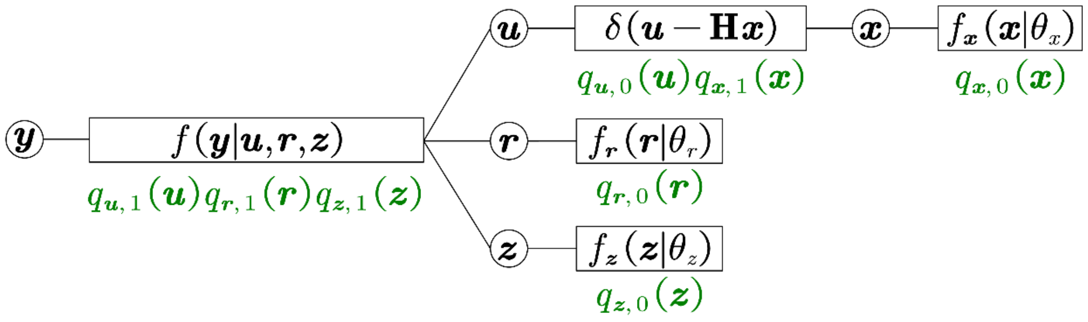


observation $\boldsymbol{y}$ — detected anomalies + uncertainty

multispectral Lidar data

low photon-count

Poisson noise



$$\text{Poisson likelihood: } f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x},\boldsymbol{z},\boldsymbol{r}) = \mathcal{P}((1-\boldsymbol{z})\odot\mathbf{H}\boldsymbol{x}+\boldsymbol{z}\odot\boldsymbol{r})$$

$$\text{positive \& sparse \& Bernoulli priors: } f_x(\boldsymbol{x}|\theta_x),\ f_r(\boldsymbol{r}|\theta_r),\ f_z(\boldsymbol{z}|\theta_z)$$

$$\text{exact posterior: } p(\boldsymbol{x},\boldsymbol{v},\boldsymbol{z}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})\, f_x(\boldsymbol{x}|\theta_x)\, f_{v,z}(\boldsymbol{v},\boldsymbol{z}|\theta_v,\theta_z)$$

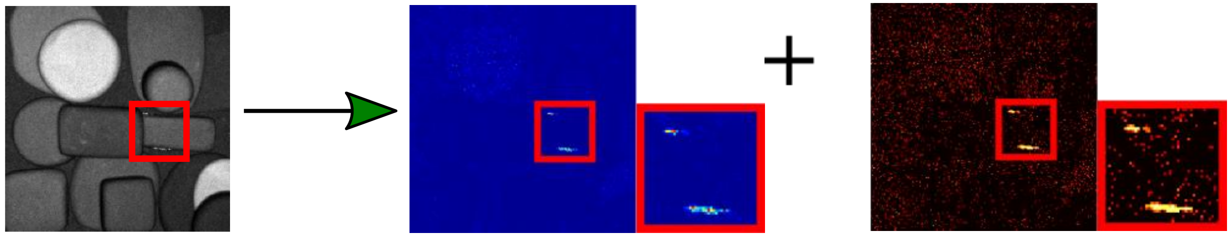
scalability:
1. combining full, diagonal, isotropic covariance matrices for flexible and efficient approximation
2. automatic hyperparameter estimation

EP scalable posterior approximation: $q(\boldsymbol{x}) \propto q_{x,1}(\boldsymbol{x})q_{x,0}(\boldsymbol{x}) \propto \mathcal{N}(\,;\boldsymbol{\mu}_x,\boldsymbol{\Sigma}_x)$

$$q(\boldsymbol{r}) \propto q_{r,1}(\boldsymbol{r})q_{r,0}(\boldsymbol{r}) \propto \mathcal{N}(\,;\boldsymbol{\mu}_r,\boldsymbol{\Sigma}_r)$$

$$q(\boldsymbol{z}) \propto q_{z,1}(\boldsymbol{z})q_{z,0}(\boldsymbol{z}) \propto Bern(\boldsymbol{z})$$

Application 4: multispectral Lidar data anomaly detection

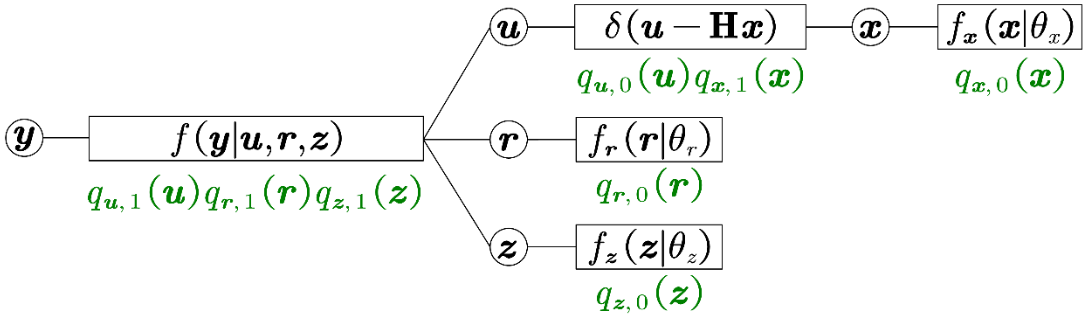


observation $\boldsymbol{y}$ — detected anomalies + uncertainty

multispectral Lidar data

low photon-count

Poisson noise



$$\text{Poisson likelihood: } f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x},\boldsymbol{z},\boldsymbol{r}) = \mathcal{P}((1-\boldsymbol{z})\odot\mathbf{H}\boldsymbol{x}+\boldsymbol{z}\odot\boldsymbol{r})$$

$$\text{positive \& sparse \& Bernoulli priors: } f_x(\boldsymbol{x}|\theta_x),\ f_r(\boldsymbol{r}|\theta_r),\ f_z(\boldsymbol{z}|\theta_z)$$

$$\text{exact posterior: } p(\boldsymbol{x},\boldsymbol{v},\boldsymbol{z}) \propto f_{y|x}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x})f_x(\boldsymbol{x}|\theta_x)f_{v,z}(\boldsymbol{v},\boldsymbol{z}|\theta_v,\theta_z)$$

scalability:
1. combining full, diagonal, isotropic covariance matrices for flexible and efficient approximation
2. automatic hyperparameter estimation

EP scalable posterior approximation:

$$q(\boldsymbol{x}) \propto q_{x,1}(\boldsymbol{x})q_{x,0}(\boldsymbol{x}) \propto \mathcal{N}(\ ;\boldsymbol{\mu}_x,\boldsymbol{\Sigma}_x)$$

$$q(\boldsymbol{r}) \propto q_{r,1}(\boldsymbol{r})q_{r,0}(\boldsymbol{r}) \propto \mathcal{N}(\ ;\boldsymbol{\mu}_r,\boldsymbol{\Sigma}_r)$$

$$q(\boldsymbol{z}) \propto q_{z,1}(\boldsymbol{z})q_{z,0}(\boldsymbol{z}) \propto Bern(\boldsymbol{z})$$

anomaly amplitude: $q(\boldsymbol{r})$

anomaly presence: $q(\boldsymbol{z})$

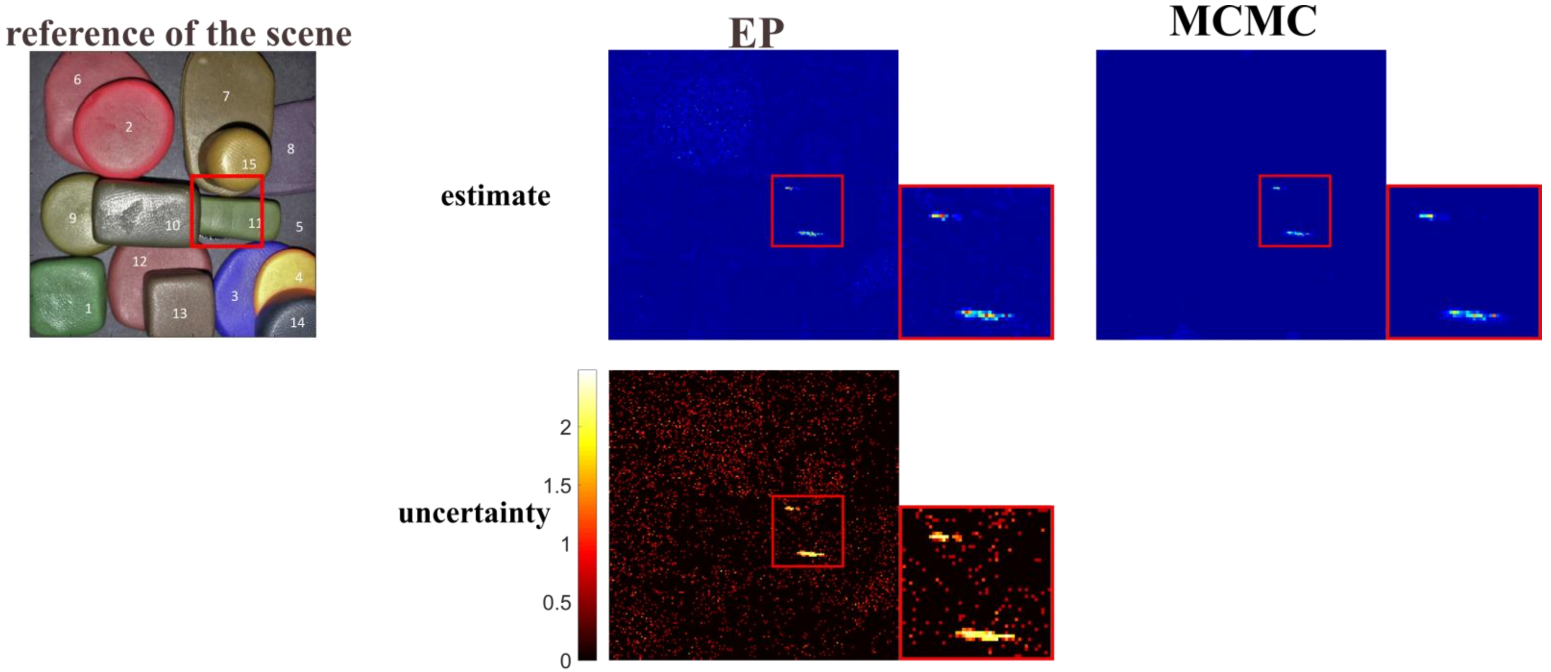
3. Applications – EP for scalable imaging inverse problems

- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- **toy examples and applications**

Application 4: multispectral Lidar data anomaly detection

$$\boldsymbol{x} \in \mathbb{R}^{190\times190\times15} \text{ (materials)}$$

$$\boldsymbol{r}, \boldsymbol{z} \in \mathbb{R}^{190\times190\times33} \text{ (bands)}$$



CPU computational time: 0.3 seconds per pixel — 14.3 hours

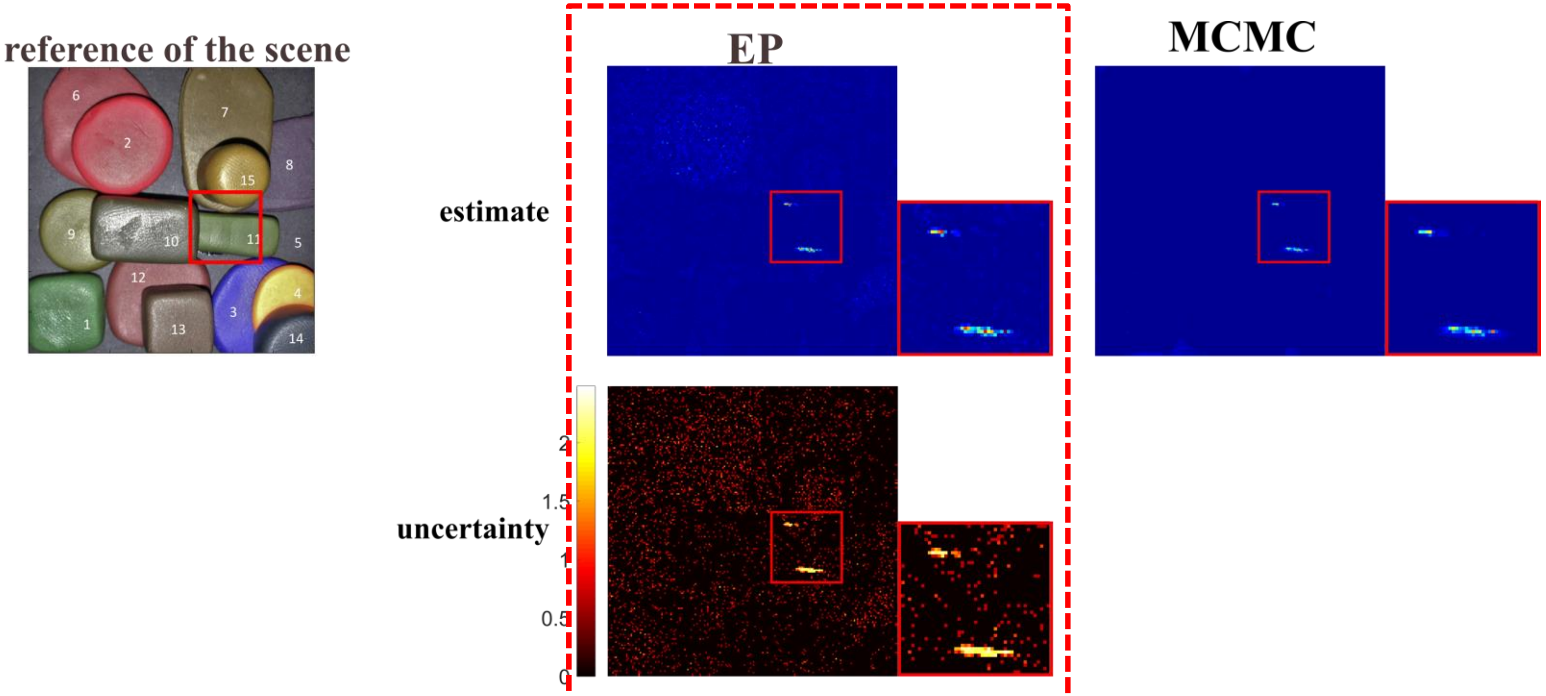
Dan Yao, Stephen McLaughlin, Yoann Altmann, Michael E Davies. "*Joint Robust Linear Regression and Anomaly Detection in Poisson noise using Expectation-Propagation*", 28th European Signal Processing Conference (EUSIPCO). 2021. pp. 2463-2467. https://doi.org/10.23919/Eusipco47968.2020.9287355

Yoann Altmann, Dan Yao, Stephen McLaughlin, Michael E Davies. "*Robust Linear Regression and Anomaly Detection in the Presence of Poisson Noise Using Expectation-Propagation*", Advances in Condition Monitoring and Structural Health Monitoring: WCCM 2019 (pp. 143-158). Springer. https://doi.org/10.1007/978-981-15-9199-0_14

Application 4: multispectral Lidar data anomaly detection

$$\boldsymbol{x} \in \mathbb{R}^{190\times190\times15} \text{ (materials)}$$

$$\boldsymbol{r}, \boldsymbol{z} \in \mathbb{R}^{190\times190\times33} \text{ (bands)}$$



CPU computational time: 0.3 seconds per pixel | 14.3 hours

Dan Yao, Stephen McLaughlin, Yoann Altmann, Michael E Davies. "*Joint Robust Linear Regression and Anomaly Detection in Poisson noise using Expectation-Propagation*", 28th European Signal Processing Conference (EUSIPCO). 2021. pp. 2463-2467. https://doi.org/10.23919/Eusipco47968.2020.9287355

Yoann Altmann, Dan Yao, Stephen McLaughlin, Michael E Davies. "*Robust Linear Regression and Anomaly Detection in the Presence of Poisson Noise Using Expectation-Propagation*", Advances in Condition Monitoring and Structural Health Monitoring: WCCM 2019 (pp. 143-158). Springer. https://doi.org/10.1007/978-981-15-9199-0_14

Other applications:

Model selection:
K. Drummond, D. Yao, S. McLaughlin, A. Pawlikowska, R. Lamb, Y. Altmann. '*Efficient joint surface detection and depth estimation of single-photon Lidar data using assumed density filtering*' , Submitted to SSPD 2022.

Online processing:
Y. Altmann, S. McLaughlin, Michael E Davies. '*Fast Online 3D Reconstruction of Dynamic Scenes from Individual Single-Photon Detection Events*', IEEE Transactions on Image Processing, vol. 29, pp. 2666-2675, 2020, doi: 10.1109/TIP.2019.2952008.

This talk is about

Expectation Propagation

1. Problem formulation and challenges

- Imaging inverse problems
- Bayesian estimation strategy
- challenges

2. Solution – EP for approximate Bayesian inference

- basic idea
- KL divergence minimization
- factor graph

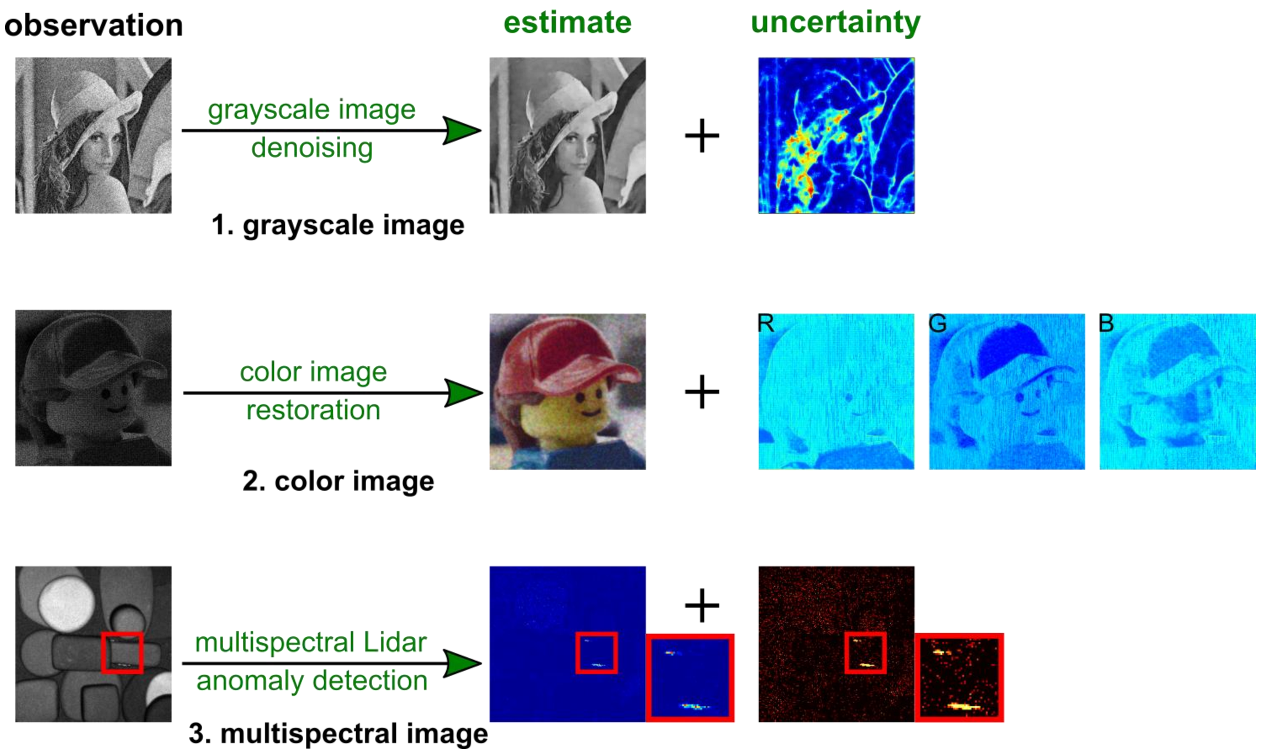
3. Applications – EP for scalable imaging inverse problems

- how to construct an EP algorithm to solve image inverse problems
- how to achieve scalable posterior approximation
- toy examples and applications

4. Conclusion

By the end of this talk,

you will know how to implement your own EP algorithm to:



4. Conclusion

- **Goal:** $\underset{\text{Ground truth image}}{\boldsymbol{x}} \xrightarrow{\text{linear projection}} \mathbf{H}\boldsymbol{x} \xrightarrow[\text{noise}]{\text{additive/multiplicative}} \underset{\text{Observed degraded image}}{\boldsymbol{y}}$

  Linear inverse problems in imaging

- **Goal:**

$$\boldsymbol{x} \xrightarrow{\text{linear projection}} \mathbf{H}\boldsymbol{x} \xrightarrow[\text{noise}]{\text{additive/multiplicative}} \boldsymbol{y}$$

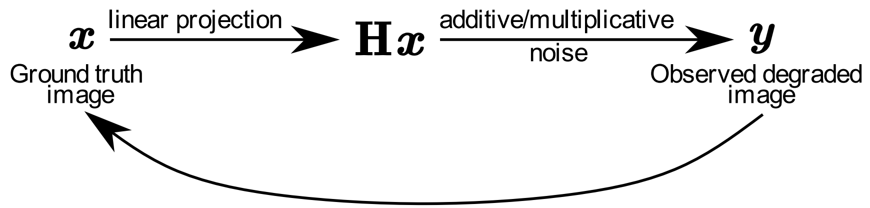
Ground truth image — Observed degraded image

Linear inverse problems in imaging

$$\boldsymbol{y}, \mathbf{H} \xRightarrow{\textit{scalable}} \textit{high dimensional } \hat{\boldsymbol{x}}$$

4. Conclusion

- **Goal:**



Linear inverse problems in imaging

$$\boldsymbol{y}, \mathbf{H} \xrightarrow{\textit{scalable}} \textit{high dimensional } \hat{\boldsymbol{x}}$$

- **Scalable solution by Expectation Propagation:**

- **Goal:**

$$\underset{\text{Ground truth image}}{\boldsymbol{x}} \xrightarrow{\text{linear projection}} \mathbf{H}\boldsymbol{x} \xrightarrow[\text{noise}]{\text{additive/multiplicative}} \underset{\text{Observed degraded image}}{\boldsymbol{y}}$$

Linear inverse problems in imaging

$$\boldsymbol{y}, \mathbf{H} \xrightarrow{\textit{scalable}} \textit{high dimensional } \hat{\boldsymbol{x}}$$

- **Scalable solution by Expectation Propagation:**

$$\boldsymbol{y}, \mathbf{H} \xrightarrow{\text{Input}} \begin{matrix}\text{scalable EP} \\ \text{posterior approximation}\end{matrix}$$

- **Goal:**

$$\underset{\text{Ground truth image}}{\boldsymbol{x}} \xrightarrow{\text{linear projection}} \mathbf{H}\boldsymbol{x} \xrightarrow[\text{noise}]{\text{additive/multiplicative}} \underset{\text{Observed degraded image}}{\boldsymbol{y}}$$

Linear inverse problems in imaging

$$\boldsymbol{y}, \mathbf{H} \xrightarrow{\textit{scalable}} \textit{high dimensional } \hat{\boldsymbol{x}}$$

- **Scalable solution by Expectation Propagation:**

$$\boldsymbol{y}, \mathbf{H} \xrightarrow{\text{Input}} \begin{array}{c}\text{scalable EP}\\ \text{posterior approximation}\end{array} \xrightarrow{\text{Output}} \hat{\boldsymbol{x}} \begin{cases} \text{point estimate: approximate } \textbf{MMSE} \text{ estimate} \\ \text{uncertainty: approximate posterior } \textbf{covariance} \end{cases}$$

- **Goal:**

$$\underset{\text{Ground truth image}}{\boldsymbol{x}} \xrightarrow{\text{linear projection}} \mathbf{H}\boldsymbol{x} \xrightarrow[\text{noise}]{\text{additive/multiplicative}} \underset{\text{Observed degraded image}}{\boldsymbol{y}}$$

Linear inverse problems in imaging

$$\boldsymbol{y}, \mathbf{H} \xrightarrow{\textit{scalable}} \textit{high dimensional } \hat{\boldsymbol{x}}$$

- **Scalable solution by Expectation Propagation:**

$$\boldsymbol{y}, \mathbf{H} \xrightarrow{\text{Input}} \begin{matrix}\text{scalable EP} \\ \text{posterior approximation}\end{matrix} \xrightarrow{\text{Output}} \hat{\boldsymbol{x}} \begin{cases} \text{point estimate: approximate } \textbf{MMSE} \text{ estimate} \\ \text{uncertainty: approximate posterior } \textbf{covariance} \end{cases}$$

- **Applications:**

- **Goal:**

$$\underset{\text{Ground truth image}}{\boldsymbol{x}} \xrightarrow{\text{linear projection}} \mathbf{H}\boldsymbol{x} \xrightarrow[\text{noise}]{\text{additive/multiplicative}} \underset{\text{Observed degraded image}}{\boldsymbol{y}}$$

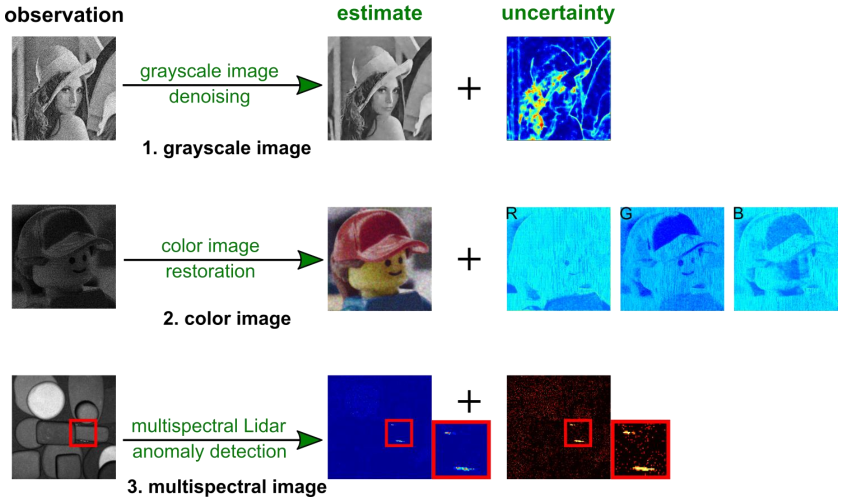
Linear inverse problems in imaging

$$\boldsymbol{y}, \mathbf{H} \xrightarrow{\textit{scalable}} \textit{high dimensional } \hat{\boldsymbol{x}}$$

- **Scalable solution by Expectation Propagation:**

$$\boldsymbol{y}, \mathbf{H} \xrightarrow{\text{Input}} \begin{matrix}\text{scalable EP}\\ \text{posterior approximation}\end{matrix} \xrightarrow{\text{Output}} \hat{\boldsymbol{x}} \begin{cases} \text{point estimate: approximate } \textbf{MMSE} \text{ estimate} \\ \text{uncertainty: approximate posterior } \textbf{covariance} \end{cases}$$

- **Applications:**



Some EP references

- **EP tutorial videos:**

1. Thomas Minka: **Approximate Inference** http://videolectures.net/mlss09uk_minka_ai/

2. Simon Barthelmé: **The Expectation-Propagation algorithm: a tutorial - Part 1** https://youtu.be/0tomU1q3AdY

- **Homepages:**

1. Thomas Minka
   A roadmap to research on EP https://tminka.github.io/papers/ep/roadmap.html

2. Matt Wand
   Statistics Methodology and Theory http://matt-wand.utsacademics.info/statsPapers.html

3. José Miguel Hernández-Lobato
   Scalable methods for approximate inference https://jmhl.org/publications/

4. Matthias Seeger, Young-Jun Ko
   Scalable variational approximate inference algorithms https://mseeger.github.io/

5. Yoann Altmann
   Our group https://yoannaltmann.weebly.com/publications.html

...

Thanks for you attention !

Expectation Propagation for
Scalable Inverse Problems in Imaging

Dan Yao

Heriot-Watt University

dy2008@hw.ac.uk

Y.Altmann@hw.ac.uk