Bayesian Paradigm

Maximum A Posteriori Estimation

Simple acquisition model

- noise

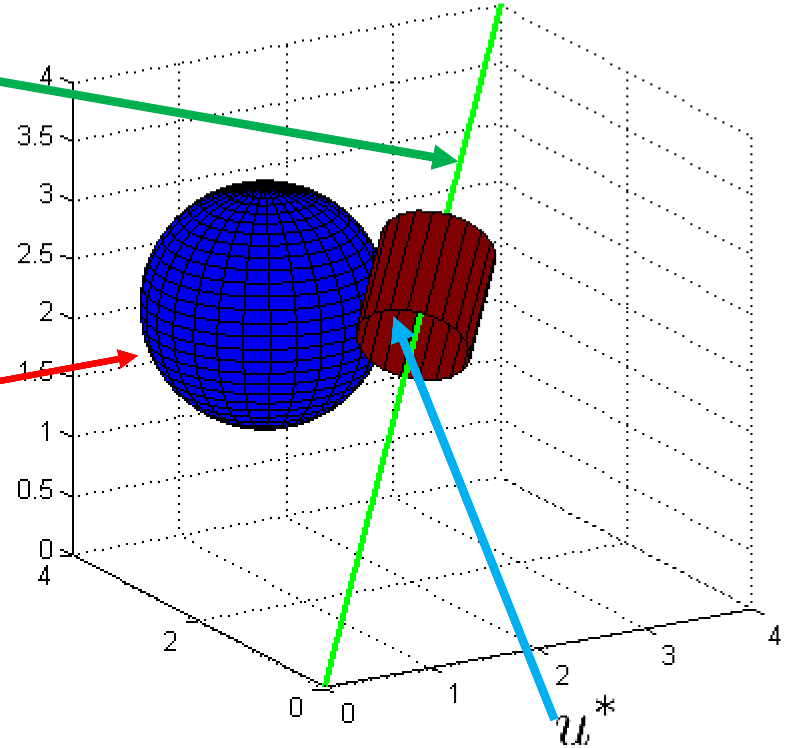
$$z = u + n \qquad n \ldots N(0, \sigma^2)$$

- + degradation

$$z = Hu + n$$

Constraint minimization

$$\min \int |\nabla u|^2$$



$$\text{subject to} \quad \|z - u\|^2 = \sigma^2$$

or

$$\text{subject to} \quad \|z - u\|^2 \leq \sigma^2$$

Equivalent formulation

- Constraint minimization

$$\min \int \phi(|\nabla u|) \quad \text{subject to } \|z-u\|^2 = \sigma^2$$

- Lagrangian (unconstraint minimization)

$$\min \left\{ \frac{1}{2} \int_\Omega |z-u|^2 dx + \lambda \int_\Omega \phi(|\nabla u|) dx \right\}$$

- Maximum Aposteriori (MAP)

……

Discrete representation

$$\mathbf{z} = [z_1, \ldots, z_N]$$

$$\min_{\mathbf{u}} \frac{1}{2} \sum_i (z_i - u_i)^2 + \lambda \sum_i \phi(|\nabla u_i|)$$

- Image is a random field → Each pixel value is a realization of some random variable

Random variables

- Discrete: takes a countable number of distinct values
  - e.g. num. of children in a family
- Continuous: x
  - probability density function ("distribution"): $p(x)$

$$p(x) \geq 0$$

$$\int_{-\infty}^{+\infty} p(x) dx = 1$$

  - Moments: expectation ("average"), variance

Random variables

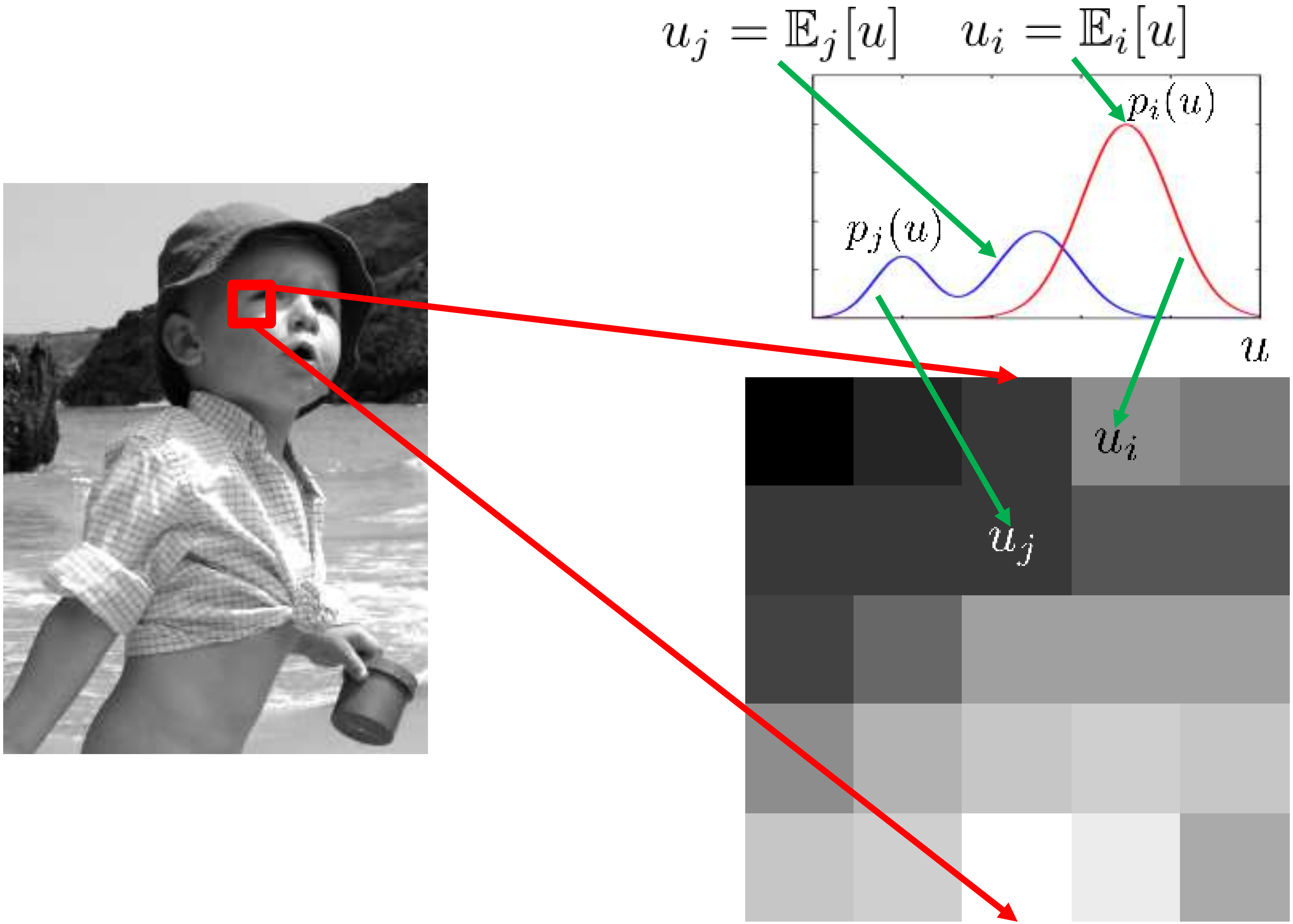
- Expectation

$$\mathbb{E}[f(x)] = \int f(x)p(x)dx \approx \frac{1}{N}\sum_{n=1}^{N} f(x_n)$$

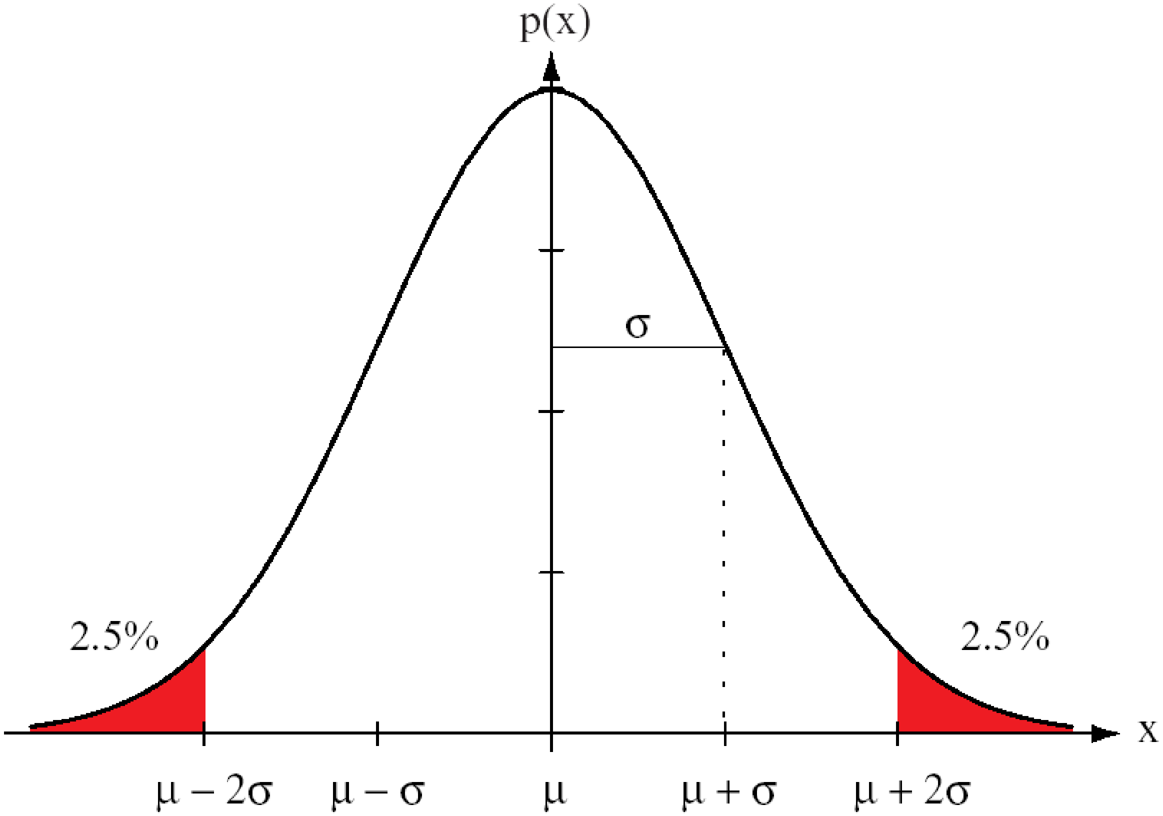
- Variance

$$\mathrm{var}[f(x)] = \mathbb{E}\left[(f(x) - \mathbb{E}[f])^2\right] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2$$

Image as a random field



Normal distribution



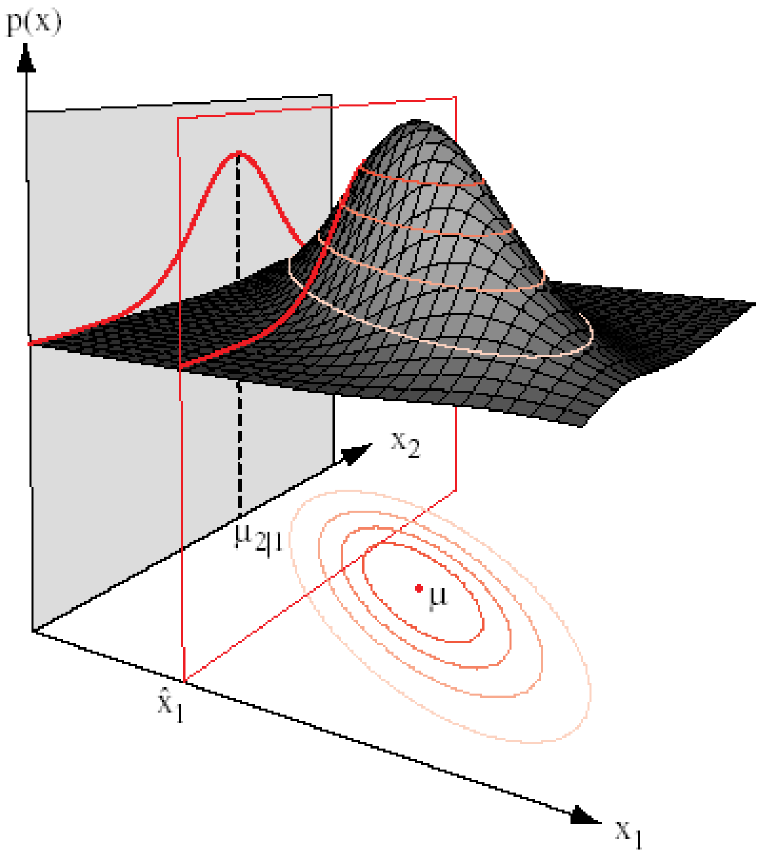
$$\hat{\mu} = \frac{1}{n}\sum_{k=1}^{n} x_k$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n} (x_k - \hat{\mu})^2$$

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \qquad N(x|\mu, \sigma^2)$$

$$p(x) = \sqrt{\frac{\lambda}{2\pi}} \exp\left[-\frac{\lambda}{2}(x-\mu)^2\right] \qquad N(x|\mu, 1/\lambda)$$

Multivariate Normal Distribution



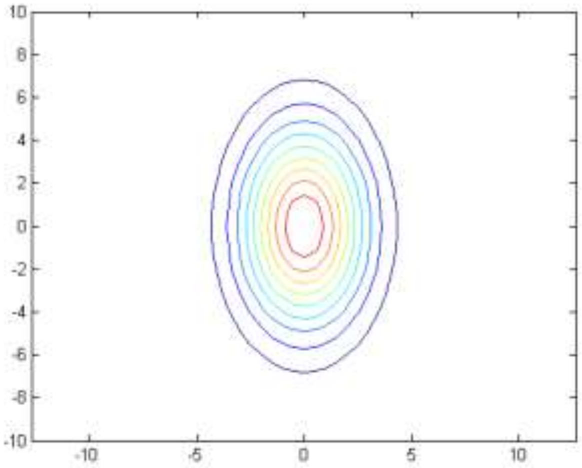
$$\hat{\mu} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{x}_k$$

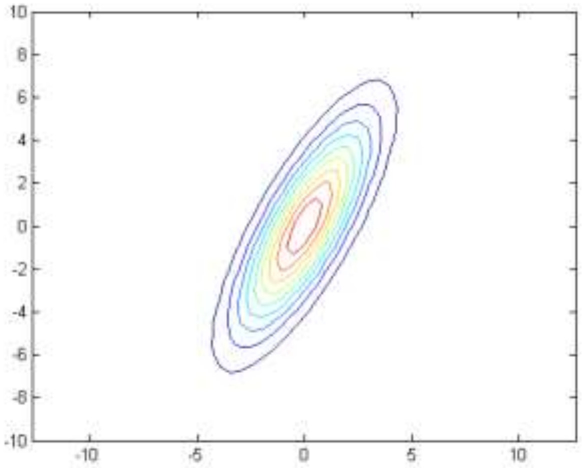
$$\hat{\Sigma} = \frac{1}{n}\sum_{k=1}^{n} (\mathbf{x}_k - \hat{\mu})(\mathbf{x}_k - \hat{\mu})^t.$$

$$N(\mathbf{x}|\mu, \Sigma)$$

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^t\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

Multivariate Normal Distribution





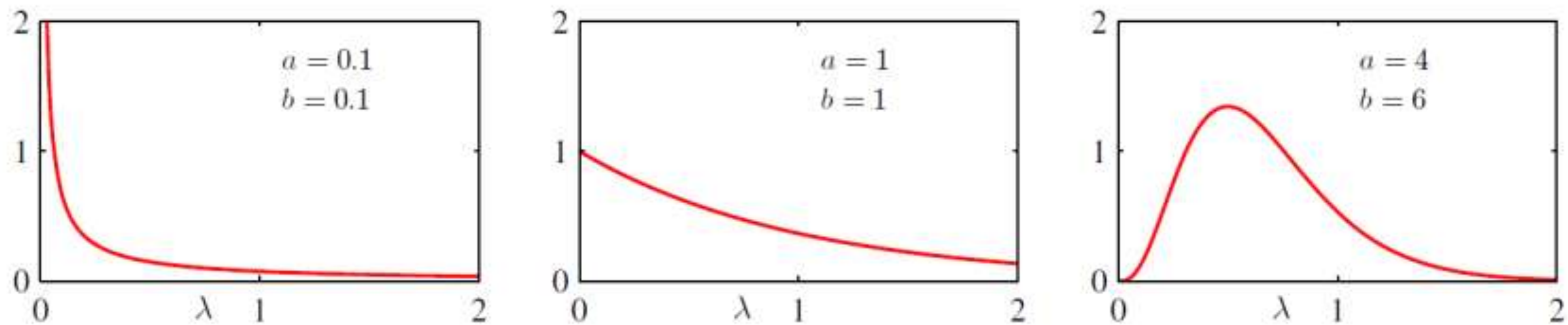
$$\Sigma = \begin{bmatrix} 4 & 0 \\ 0 & 10 \end{bmatrix}$$

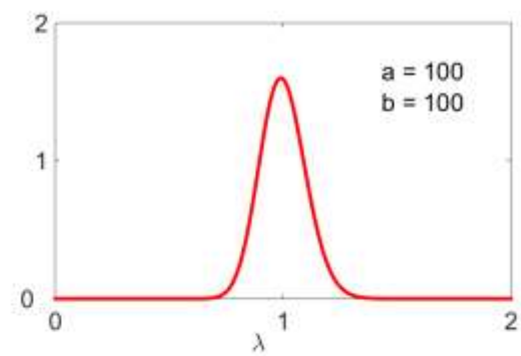
$$\Sigma = \begin{bmatrix} 4 & 5 \\ 5 & 10 \end{bmatrix}$$

Gamma distribution

$$\mathrm{Gam}(\lambda|a,b) = \frac{1}{\Gamma(a)} b^a \lambda^{a-1} \exp(-b\lambda) \qquad \lambda \geq 0, \quad a > 0, \quad b > 0$$

$$\mathbb{E}[\lambda] = \frac{a}{b} \qquad \mathrm{var}[\lambda] = \frac{a}{b^2}$$





Bayesian Paradigm

$$p(u|z) = \frac{p(z|u)p(u)}{p(z)}$$

a posteriori distribution
unknown

likelihood
given by our problem

a priori distribution
our prior knowledge

- Maximum a posteriori (MAP): max p(u|z)
- Maximum likelihood (MLE): max p(z|u)

Likelihood

$$p(u|z) \propto p(z|u)p(u)$$

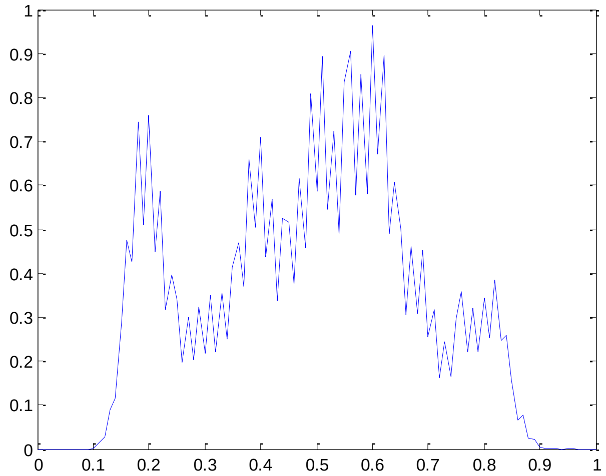
$$-\ln p(u|z) = -\ln p(z|u) - \ln p(u)$$

$$-\ln p(z|u) = -\ln k \prod_i e^{\frac{(z_i - u_i)^2}{2\sigma^2}} = \frac{1}{2\sigma^2} \sum_i (z_i - u_i)^2 + c$$

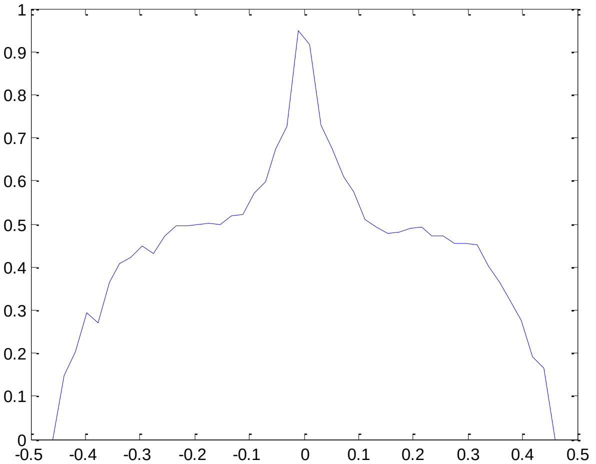
$$n \ldots N(0, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \prod_{i=1}^{N} e^{\frac{n_i^2}{2\sigma^2}}$$

Image Prior

$$\ln p(\mathbf{u}) = \ln \prod_i p(\mathbf{u}_i) = \sum_i \ln p(\mathbf{u}_i)$$



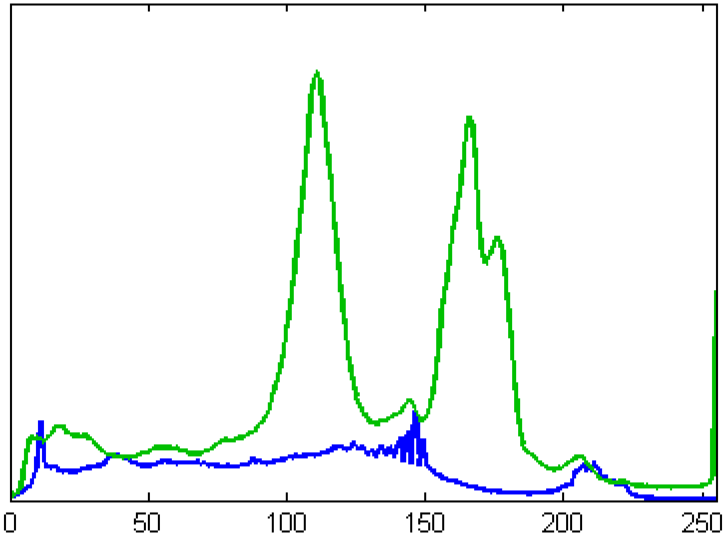
Intensity histogram



Gradient histogram

Image Prior

Intensities



Gradients (log)

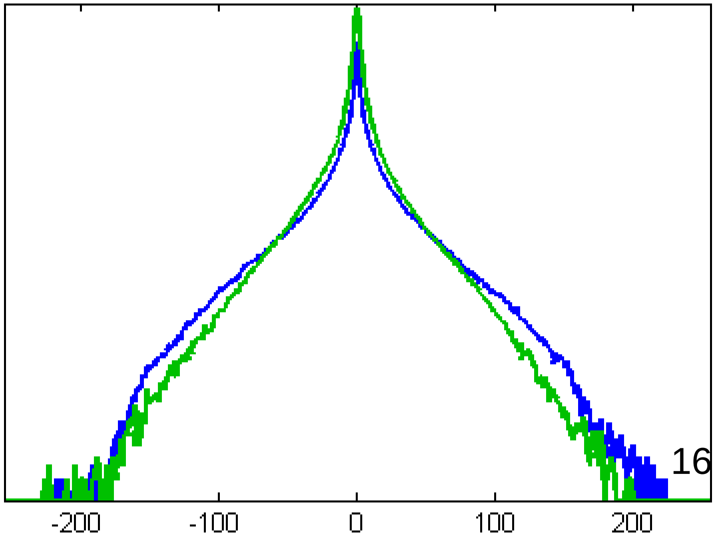
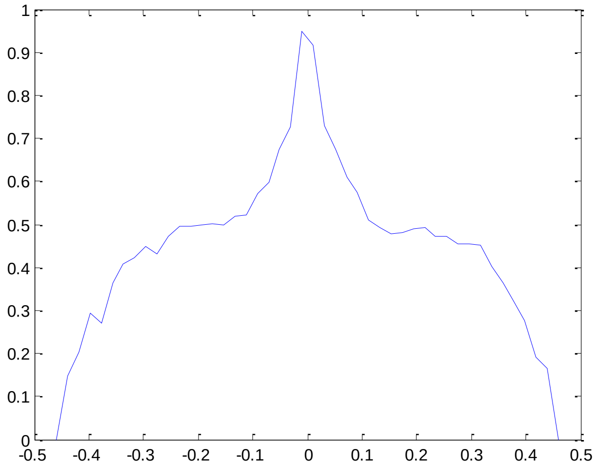


Image Prior

$$p(\mathbf{u}) \propto \prod_i e^{-\lambda \phi(\nabla u_i)}$$



Gradient histogram

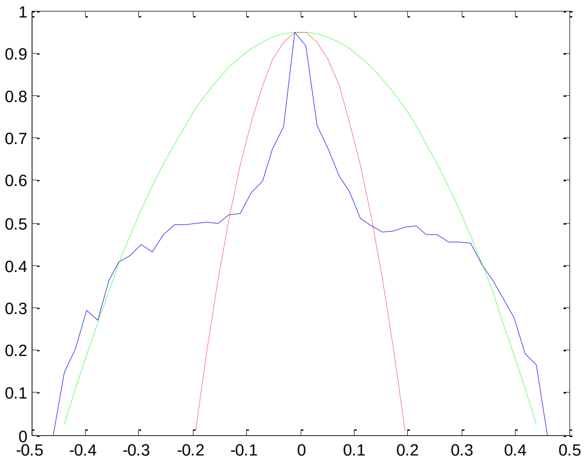
$$\ln p(\mathbf{u}) = -\lambda \sum_i \phi(\nabla u_i) + c$$

Image Prior

$$Q(u) = \lambda \int |\nabla u|^2$$

Tikhonov regularization

$$p(\mathbf{u}) \propto \prod_i e^{-\lambda |\nabla u_i|^2} = e^{-\lambda \mathbf{u}^T \mathbf{L} \mathbf{u}}$$



$$Q(u) = \lambda \int |\nabla u|$$

TV regularization

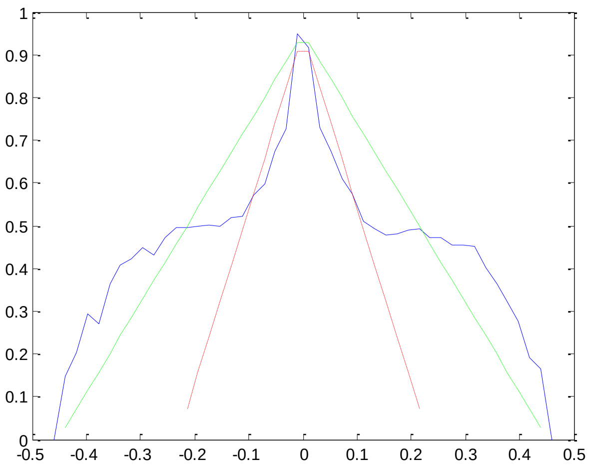
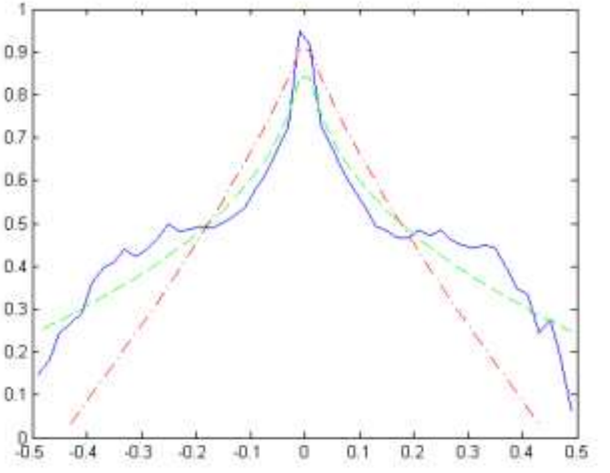


Image Prior

$$Q(u) = \lambda \int |\nabla u|^{0.8}$$

$$Q(u) = \lambda \int |\nabla u|^{0.4}$$

Non-convex regularization



MAP

$$-\ln p(u|z) = -\ln p(z|u) - \ln p(u)$$

$$\frac{1}{2\sigma^2}\sum_i (z_i - u_i)^2 + \lambda \sum_i \phi(|\nabla u_i|)$$

$$\min_{\mathbf{u}} \frac{1}{2}\sum_i (z_i - u_i)^2 + \lambda \sum_i \phi(|\nabla u_i|)$$

Denoising with unknown noise level

- Noise model:

$$z_i = u_i + n_i$$

$$n_i \ldots N(n_i|0,\gamma^{-1}) \propto \gamma^{1/2} \exp\left\{-\frac{\gamma}{2} n_i^2\right\}$$

$$\gamma^{-1} = \sigma^2$$

- Bayes: $p(u|z) \propto p(z|u)p(u)$
- and γ is unknown

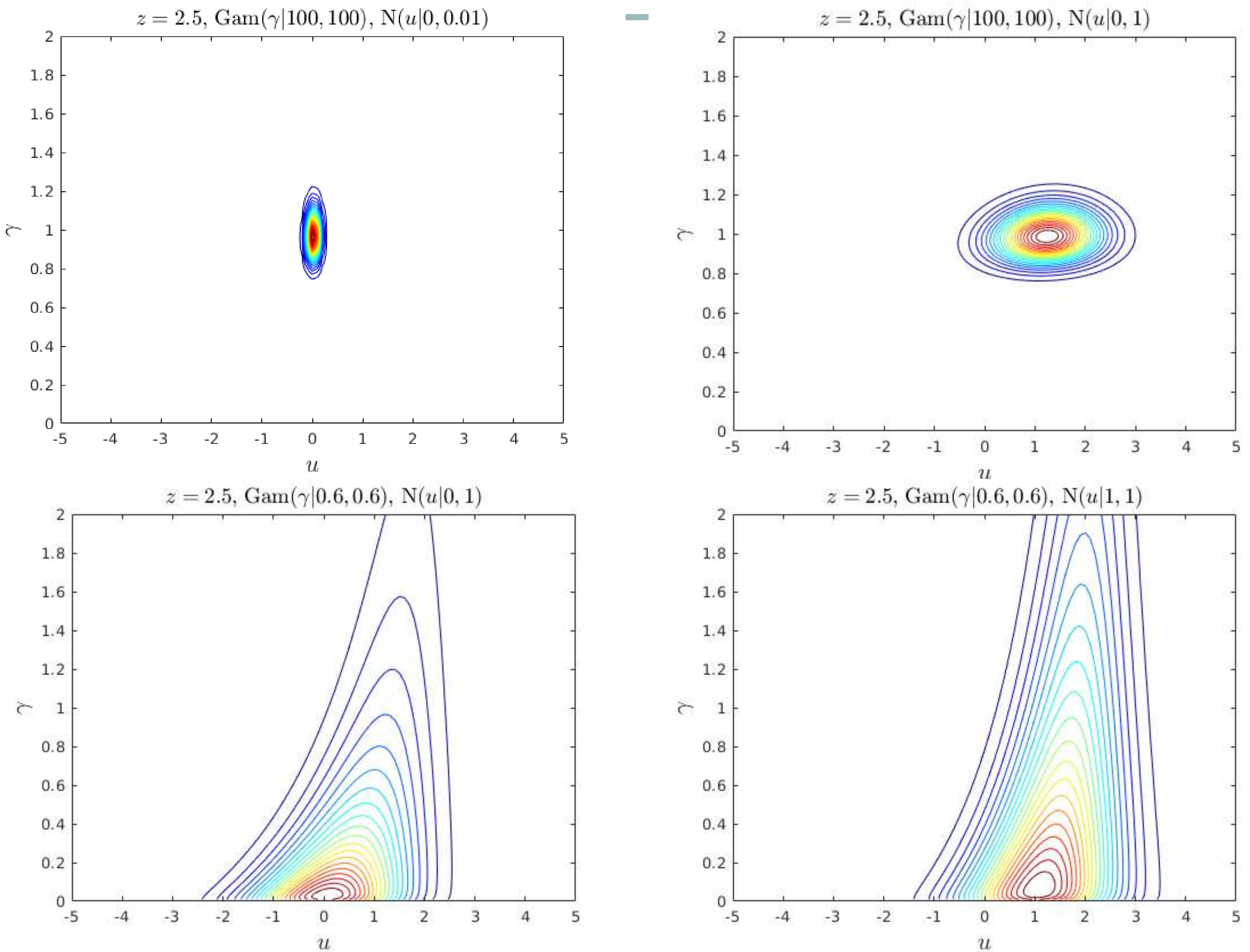
$$p(u,\gamma|z) \propto p(z|u,\gamma)p(u)p(\gamma)$$

$$\ln p(u,\gamma|z) \propto \ln p(z|u,\gamma) + \ln p(u) + \ln p(\gamma)$$

$$\frac{1}{2}\ln(\gamma) - \frac{1}{2}\gamma(u-z)^2 \qquad -\frac{1}{2}u^2 \qquad \ln \mathrm{Gam}(\gamma|a_0,b_0) =$$

$$= (a_0 - 1)\ln(\gamma) - b_0\gamma + \mathrm{const.}$$

A posteriori function



MAP denoising

- Minimize the -log posterior

$$\ln p(u,\gamma|z) \propto \frac{1}{2}\ln\gamma - \frac{1}{2}\gamma(u-z)^2 - \frac{1}{2}u^2 + (a_0-1)\ln\gamma - b_0\gamma$$

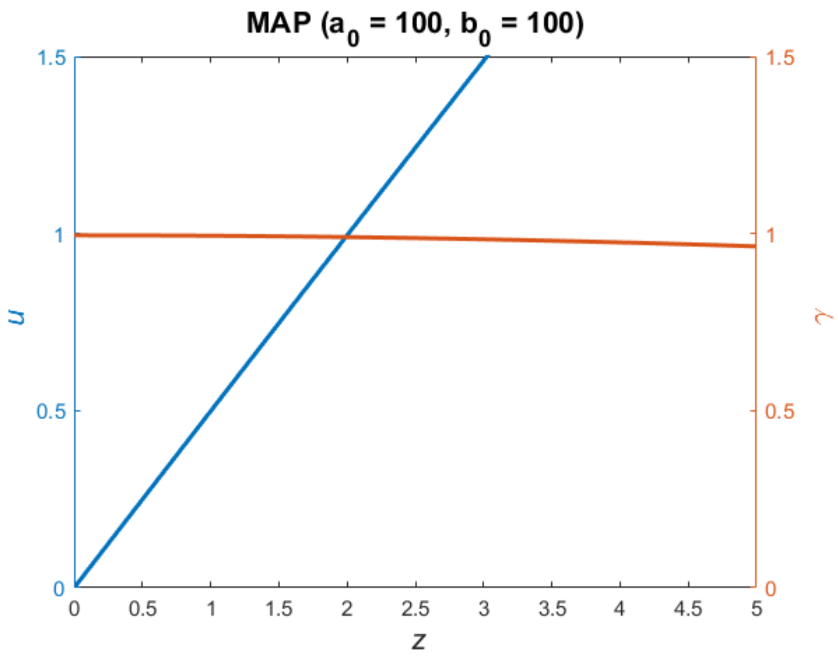
- with respect to u :

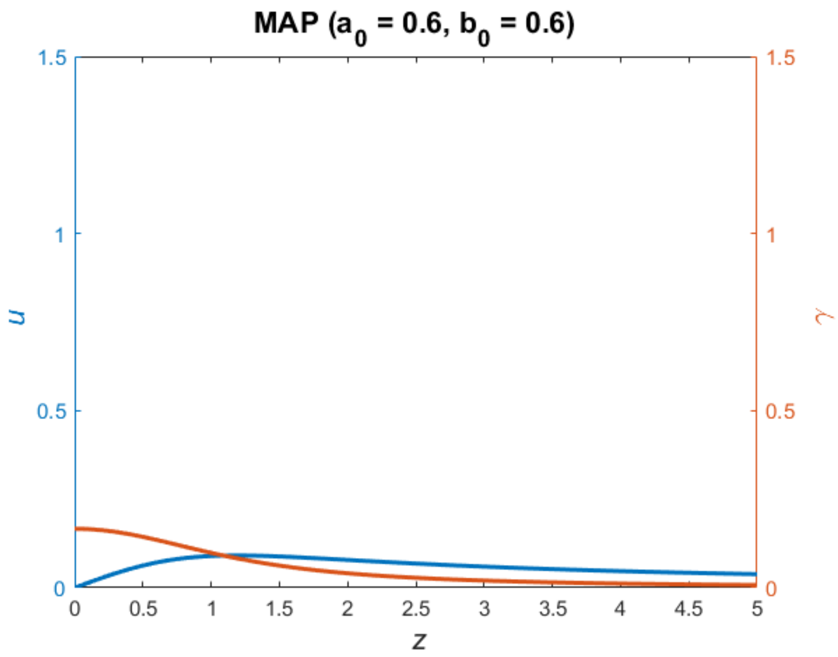
$$u = \frac{\gamma z}{\gamma+1}$$

- with respect to γ :

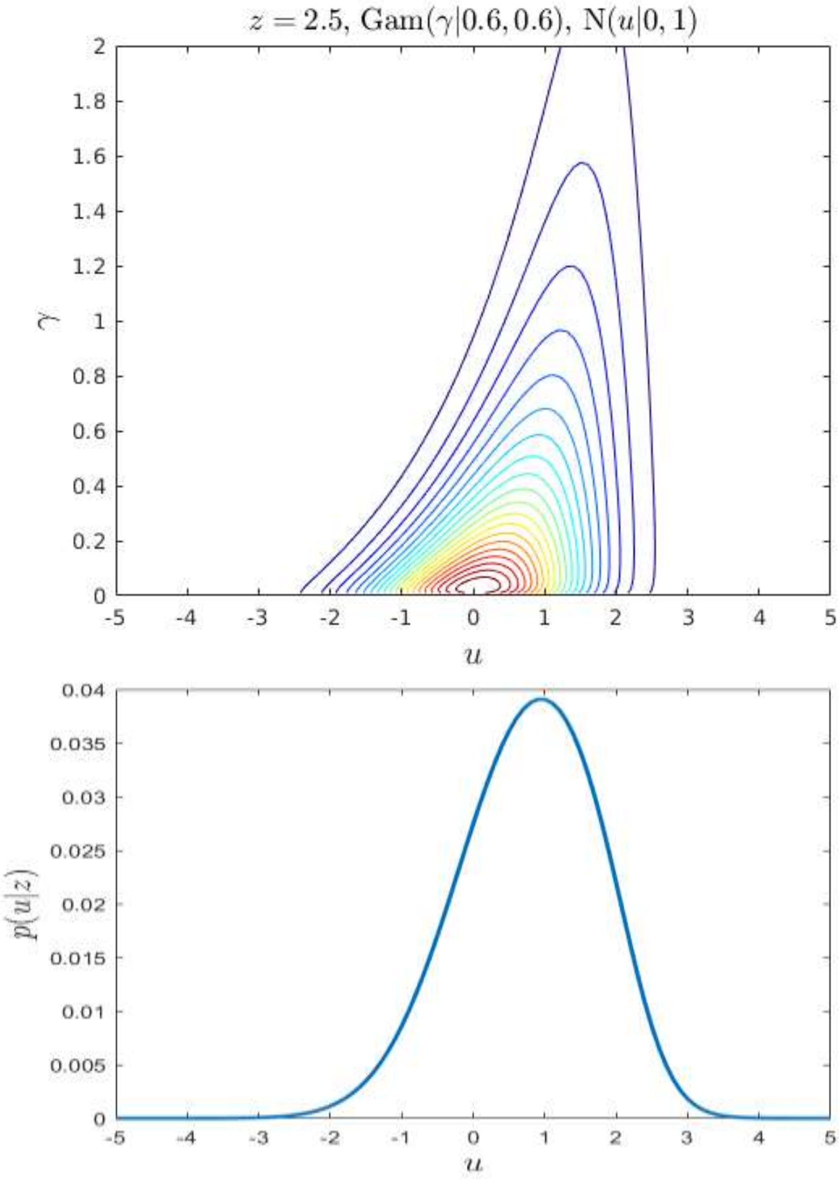
$$\gamma = \frac{a_0 - \frac{1}{2}}{b_0 + \frac{1}{2}(u-z)^2}$$

MAP denoising





Marginalizing the posterior



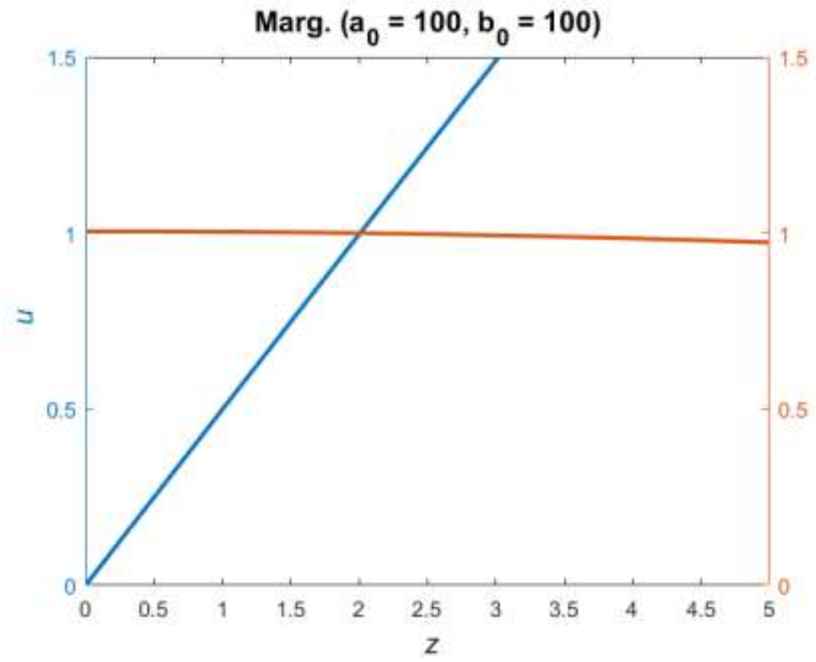
Marginalized denoising

- Expectation of marginalized posterior

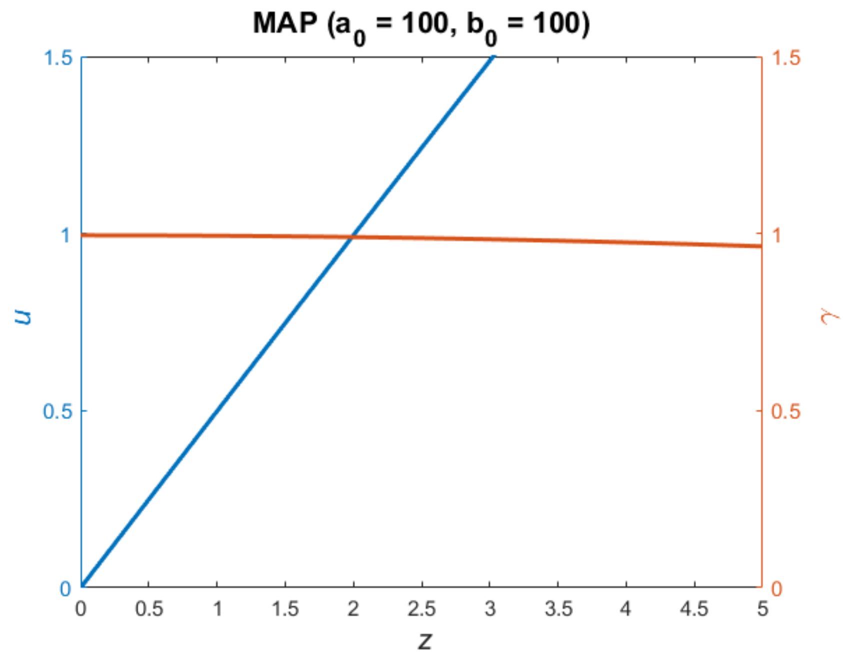
$$p(u|z) = \int p(u, \gamma|z) d\gamma$$

$$\mathbb{E}[p(u|z)] = \int u p(u|z) du$$

Marginalized denoising

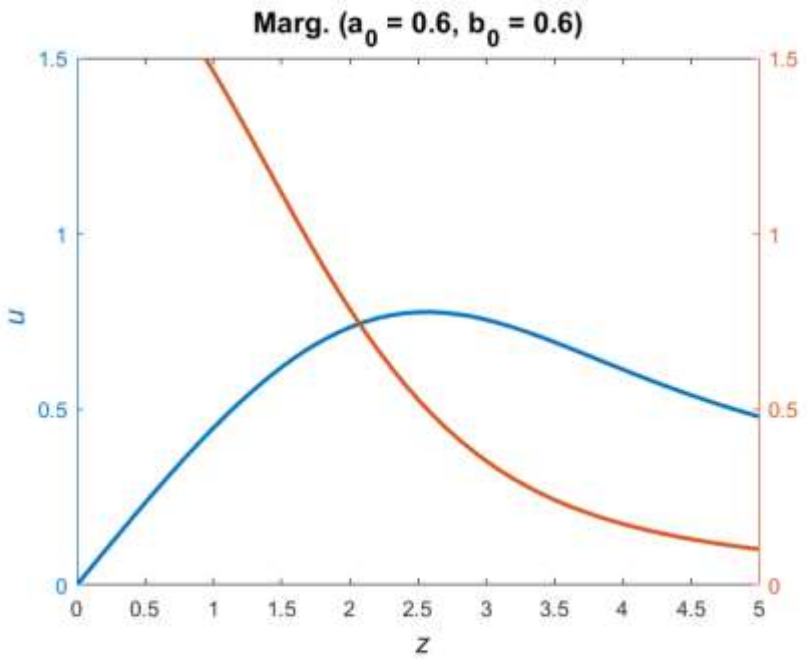


Marginalized

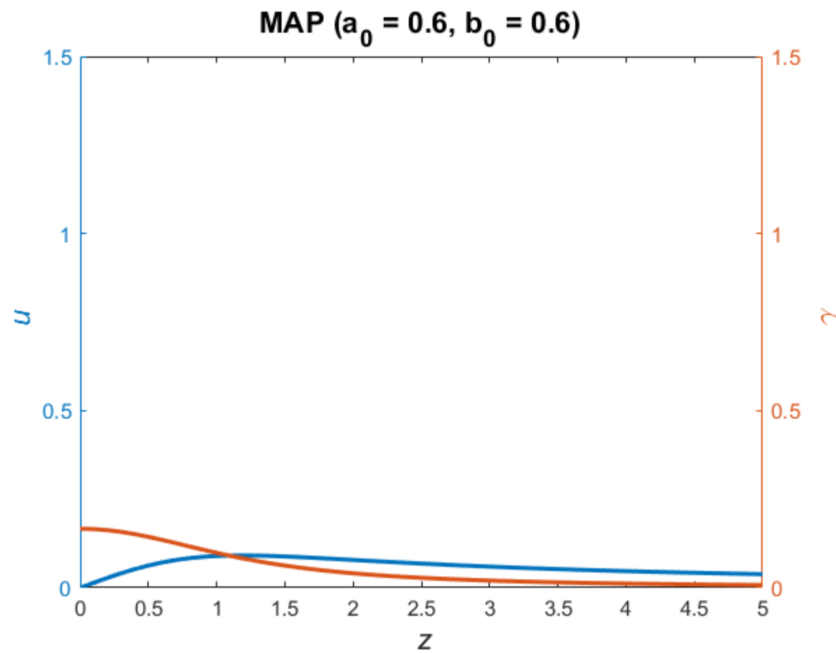


MAP

Marginalized denoising



Marginalized



MAP

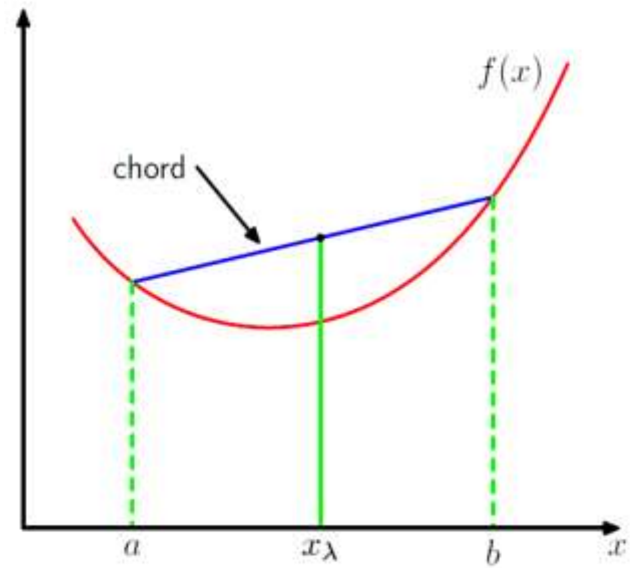
Kullback-Leibler divergence

- Measures similarity between two distributions

$$KL(p||q) = -\int p(x) \ln \left\{ \frac{q(x)}{p(x)} \right\} dx$$

- $KL(p||q) \neq KL(q||p)$
- $KL(p||q) \geq 0$
- $KL(p||q) = 0 \quad \Leftrightarrow \quad p(x) = q(x)$

Convex function $f(x)$



$$f(\lambda a + (1-\lambda)b) \leq \lambda f(a) + (1-\lambda) f(b)$$

Jensen's inequality: $$f\left(\sum_i \lambda_i x_i\right) \leq \sum_i \lambda_i f(x_i)$$

$$\lambda_i \geq 0 \quad \sum_i \lambda_i = 1$$

$$f\left(\int x p(x) dx\right) \leq \int f(x) p(x) dx$$

$$f\left(\mathbb{E}[x]\right) \leq \mathbb{E}[f(x)]$$

Kullback-Leibler divergence

$$\int p(x) f(x) dx \geq f\left(\int x p(x) dx\right)$$

$$KL(p||q) = -\int p(x) \ln\left\{\frac{q(x)}{p(x)}\right\} dx \geq -\ln \int q(x) dx$$

$$\int q(x) dx = 1 \quad \Rightarrow \quad KL(p||q) \geq 0$$

$-\ln()$ is strictly convex $\Rightarrow$ $KL(p||q) = 0 \Leftrightarrow p(x) = q(x)$

Variational Bayes

- Approximate posterior with a simpler form

$$p(x_1, x_2) \approx q_1(x_1) q_2(x_2)$$

- Minimize $KL(q_1 q_2 || p) = -\int\int q_1 q_2 \ln \frac{p}{q_1 q_2} dx_1 dx_2$

with respect to q_1

E-L: $$\frac{\partial}{\partial q_1} \int \underbrace{\int q_1 q_2 \ln \frac{q_1 q_2}{p} dx_2}_{f(x_1, q_1(x_1), \nabla q_1(x_1))} dx_1 = \frac{\partial f}{\partial q_1}$$

Variational Bayes

$$\frac{\partial f}{\partial q_1} = \frac{\partial}{\partial q_1} \int q_1 q_2 \ln \frac{q_1 q_2}{p} dx_2 =$$

$$= \int q_2 \ln \frac{q_1 q_2}{p} dx_2 + \int q_1 q_2 \frac{p}{q_1 q_2} \frac{q_2}{p} dx_2 =$$

$$= \int \left[q_2 (\ln q_2 + 1) + q_2 \ln q_1 - q_2 \ln p \right] dx_2 =$$

$$= \ln q_1 - \mathbb{E}_{x_2}[\ln p] + \text{const} = 0 \quad \Rightarrow$$

$$\ln q_1 = \mathbb{E}_{x_2}[\ln p] + \text{const.}$$

$$\boxed{q_1(x_1) \propto \exp\left\{\mathbb{E}_{x_2}[\ln p(x_1, x_2)]\right\}}$$

Variational Bayes

- Minimize with respect to all other factors q_i

$$q_i(x_i) \propto \exp\left\{\mathbb{E}_{x_{j\neq i}}[\ln p(x_1,\ldots,x_j)]\right\}$$

Variational Bayes

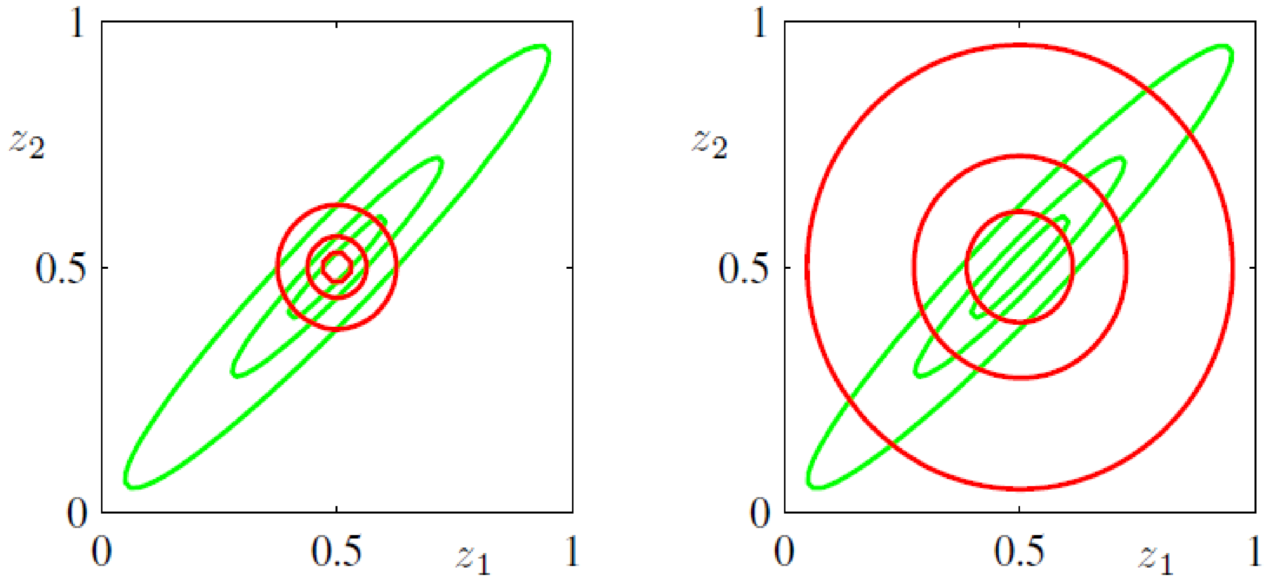
- Minimize $KL(p||q_1 q_2) = -\int\int p \ln \frac{q_1 q_2}{p} dx_1 dx_2$

with respect to q_1

……

$$q_1(x_1) = \int p(x_1, x_2) dx_2$$

Factorized approximation



$$\min_{q_1, q_2} KL\left(q_1(z_1)q_2(z_2)||p(z_1, z_2)\right) \qquad \min_{q_1, q_2} KL\left(p(z_1, z_2)||q_1(z_1)q_2(z_2)\right)$$

VB denoising

- Find the factorized approximation of the posterior:

$$p(u,\gamma|z) \approx q_1(u)q_2(\gamma)$$

- image: $$\ln p(u,\gamma|z) \propto \frac{1}{2}\ln\gamma - \frac{1}{2}\gamma(u-z)^2 - \frac{1}{2}u^2 + (a_0-1)\ln\gamma - b_0\gamma$$

$$q_1(u) \propto \exp\{\mathbb{E}_\gamma[\ln p(u,\gamma|z)]\} = N(u|\overline{u},(\overline{\gamma}+1)^{-1})$$

$$(\overline{\gamma}+1)\overline{u} = \overline{\gamma}z$$

VB denoising

- noise precision:

$$\ln p(u,\gamma|z) \propto \frac{1}{2}\ln\gamma - \frac{1}{2}\gamma(u-z)^2 - \frac{1}{2}u^2 + (a_0-1)\ln\gamma - b_0\gamma$$

$$= (a_0 - \frac{1}{2})\ln\gamma - (b_0 + \frac{1}{2}(u-z)^2)\gamma + \text{const.}$$

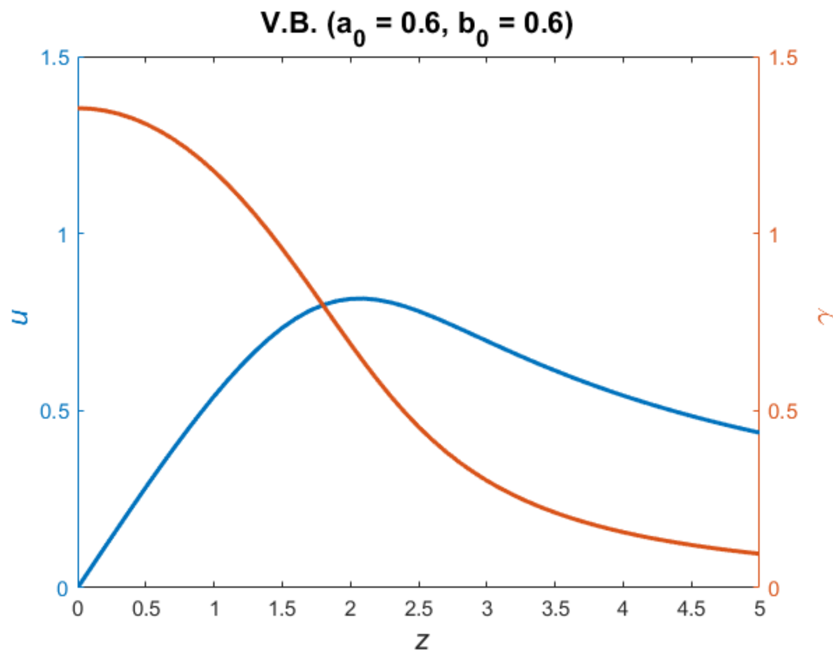
$$q_2(\gamma) \propto \exp\{\mathbb{E}_u[\ln p(u,\gamma|z)]\} = \text{Gam}(\gamma|a_\gamma, b_\gamma)$$

$$\overline{\gamma} = \frac{a_\gamma}{b_\gamma} = \frac{a_0 + \frac{1}{2}}{b_0 + \frac{1}{2}((\overline{u}-z)^2 + \text{var}[u])}$$

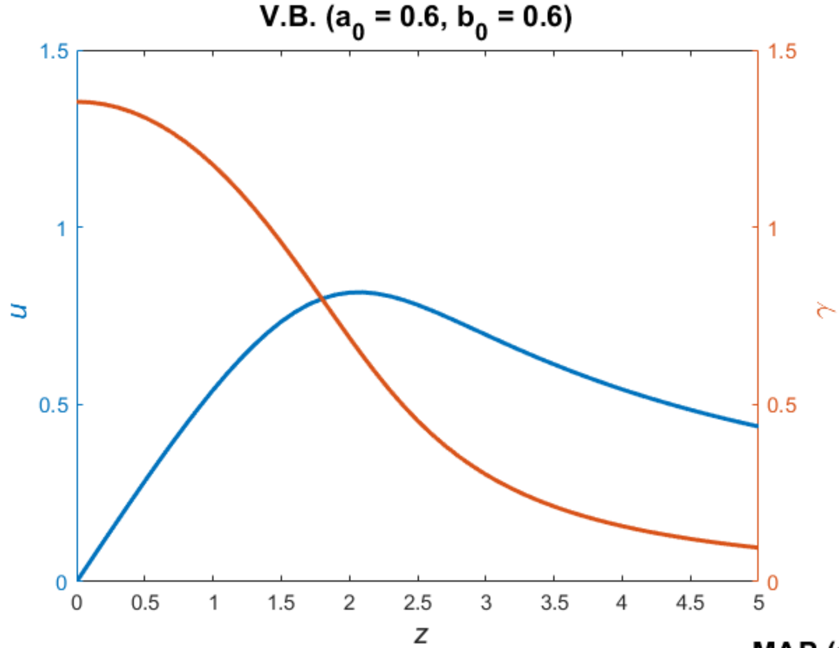
$$a_\gamma = a_0 + \frac{1}{2}$$

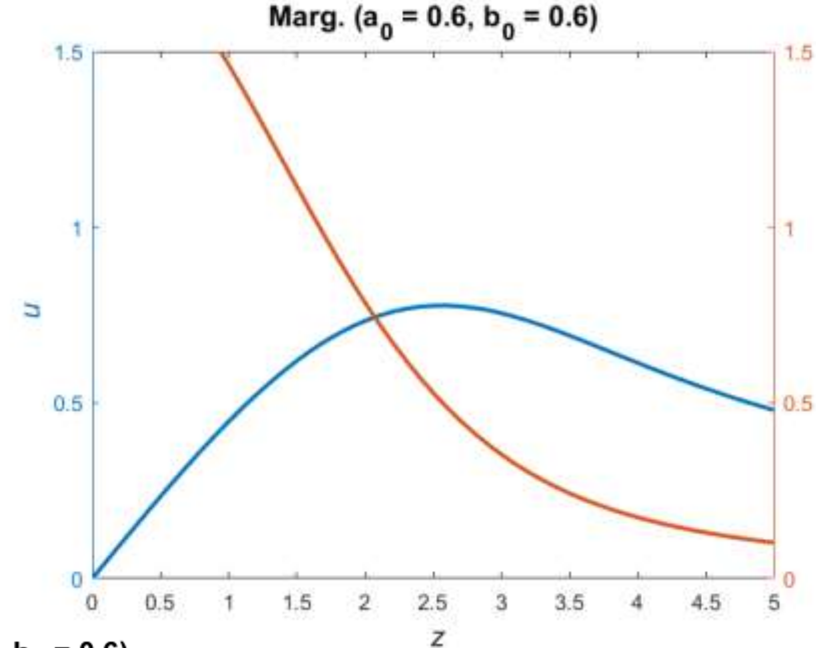
$$b_\gamma = b_0 + \frac{1}{2}\mathbb{E}_u[(u-z)^2]$$

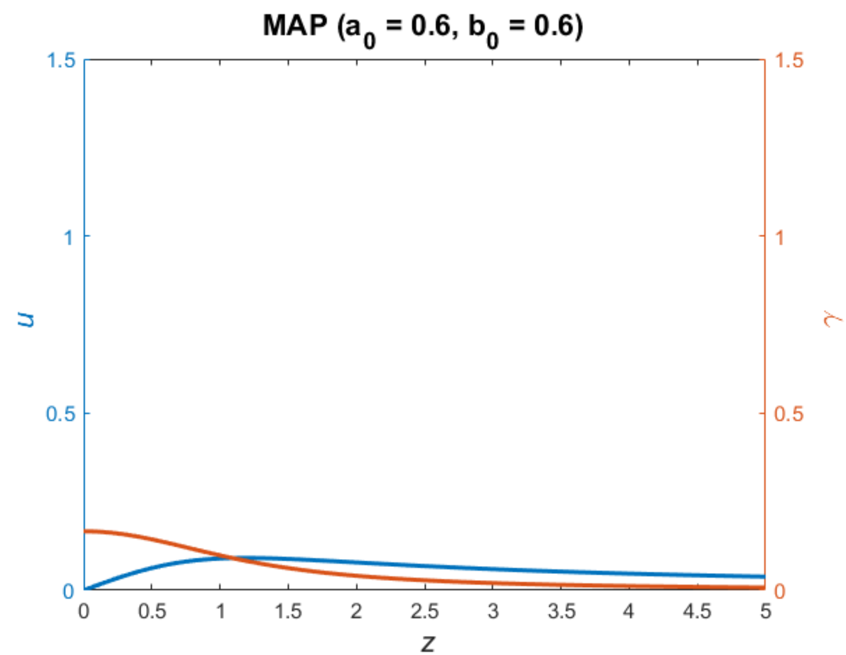
VB denoising



Comparison


V.B. (a_0 = 0.6, b_0 = 0.6)
u
γ
z
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5
0 0.5 1 1.5


Marg. (a_0 = 0.6, b_0 = 0.6)
u
z
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5
0 0.5 1 1.5


MAP (a_0 = 0.6, b_0 = 0.6)
u
γ
z
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5
0 0.5 1 1.5

Bayesian Paradigm for BD

$$p(u,h|z) = \frac{p(z|u,h)\,p(u)\,p(h)}{p(z)}$$

***a posteriori* distribution**
unknown

likelihood
given by our problem

***a priori* distribution**
our prior knowledge

- Maximum a posteriori (MAP): $\max p(u,h|z)$

Blind deconvolution with MAP

- max *a posteriori* probability $p(u, h|z)$

$$\Rightarrow \quad \min - \log p(u, h|z)$$

$$- \log p(u, h|z) \propto - \log p(z|u, h) - \log p(u) - \log p(h)$$

- Exponential family

$$E(u, h) = \frac{\lambda}{2} \|u * h - z\|^2 + Q(u) + R(h)$$

Bayesian Paradigm revisited

- Marginalize the posterior

$$p(h|z) = \int p(u, h|z) du$$

- Maximize the marginalized prob.

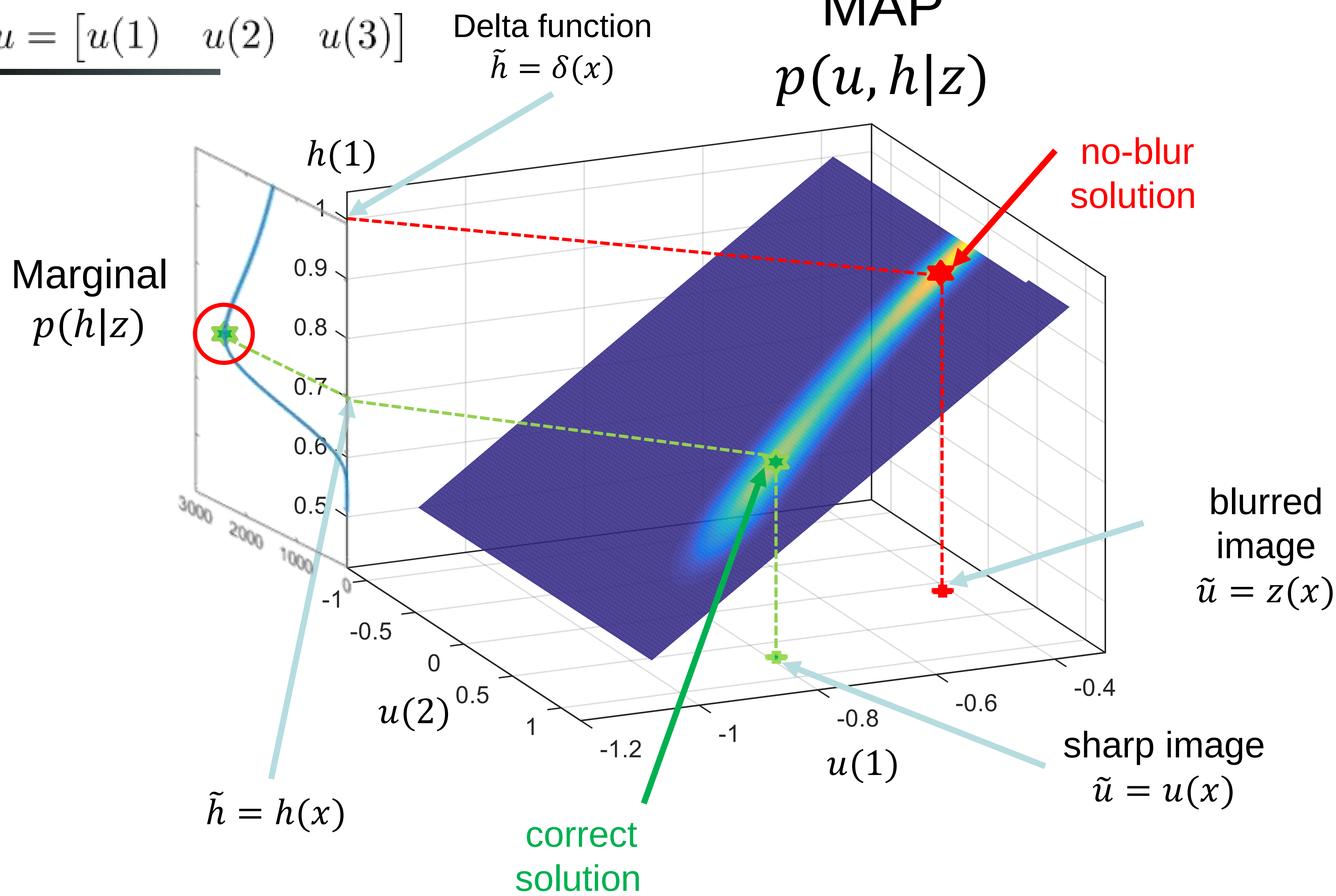
$$\hat{h} = \arg\max_{h} p(h|z)$$

- and then maximize the posterior

$$\hat{u} = \arg\max_{u} p(u, \hat{h}|z)$$

$h = [h(1) \quad h(2)]$

$u = [u(1) \quad u(2) \quad u(3)]$



How to marginalize?

$$p(h|z) = \int p(u, h|z) du$$

- If Gaussian distributions → analytic solution exists in the form of Gaussian distribution
- If not (our case) → approximation
  - Laplace approximation
  - Factorization with Variational Bayes

Variational Bayes

- Factorization of the posterior

$$p(u, h|z) \approx q(u)q(h)$$

and then marginalization is trivial.

Miskin 01
Fergus 06
Whyte 10
Levin 11
Babacan 12
Wipf 14

- Every factor q depends on moments of other variables => must be solved iteratively.

Example of blind deconvolution

Blurred image
$z(x)$

Reconstructed image
$\tilde{u}(x)$