Variational approach to parameter estimation in blind deconvolution

Rafael Molina

Universidad de Granada

Outline

- Problem formulation
- Bayesian modelling
  - Image, blur and formation models
  - Hyperparameter models
- Bayesian Inference
- Variational methods
- Estimation
  - Image and blur estimation
  - Parameter estimation
    - Unrelated variances of prior distributions
    - Relating variances of prior distributions
- Experimental section
- Appendix: Calculating mean values

1 Problem Formulation

Blind deconvolution refers to a class of problems of the form

$$g(x) = h(x) * f(x) + n(x), \quad x = (x_1, x_2) \in I, \quad (1)$$

$I \subset \mathcal{R}^2$ is the support of the image
$f(x)$ unknown original image
$g(x)$ observed image
$h(x)$ unknown blur or psf
$n(x)$ observation noise.
$(*)$ denotes the 2-D convolution

GOAL

To estimate the original image and blur simultaneously.

2 Bayesian modelling

2.1 Image, blur and formation models

Prior model for f

$$\mathrm{p}(f|\alpha_{\mathrm{im}}) \propto \alpha_{\mathrm{im}}^{N/2} \exp\{-\frac{1}{2}\alpha_{\mathrm{im}} \parallel Cf \parallel^2\}, \quad (2)$$

- C denotes the laplacian operator
- $N = P \times Q$ is the size of the column vector denoting the image, lexicographycally ordered by rows
- $\alpha_{\mathrm{im}}^{-1}$ is the variance of the Gaussian distribution

N or $N - 1$?

Prior model for h

$$\mathrm{p}(h|\alpha_{\mathrm{bl}}) \propto \alpha_{\mathrm{bl}}^{M/2} \exp\{-\frac{1}{2}\alpha_{\mathrm{bl}} \parallel Ch \parallel^2\}, \qquad (3)$$

- C denotes the laplacian operator
- $M = U \times V$ is the <u>known</u> size of the support of the blur
- h is the column vector of size N lexicographycally ordered by rows denoting the blur (this vector has all its components equal to zero outside its support)
- $\alpha_{\mathrm{bl}}^{-1}$ is the variance of the Gaussian distribution.

Degradation model of the observed image g, given f and h: the one defined in equation (1) with unknown blur approximated by a block circulant matrix and Gaussian noise $n(x)$ with zero mean and variance β^{-1}. That is,

$$\mathrm{p}(g \mid f, h, \beta) \propto \beta^{N/2} \exp\left[-\frac{1}{2}\beta \left\|g - Hf\right\|^2\right], \quad (4)$$

where we have used H to denote the $N \times N$ blurring matrix corresponding to the blurring vector h.

Important: Through out the paper we will use capital letters to denote convolution matrices corresponding to column vectors, H for h and F for f.
Note that according to this convention the degradation equation (4) can also be written

$$\mathrm{p}(g \mid f, h, \beta) \propto \beta^{N/2} \exp\left[-\frac{1}{2}\beta \left\|g - Fh\right\|^2\right]. \quad (5)$$

2.2 Hyperparameter models

When α_{im}, α_{bl} and β are unknown, the hierarchical Bayesian paradigm uses two stages

- First stage: Formulation of $\text{p}(f|\alpha_{\text{im}})$, $\text{p}(h|\alpha_{\text{bl}})$ and $\text{p}(g|f,h,\beta)$, already studied.
- Second stage: formulation of the hyperprior $\text{p}(\alpha_{\text{im}},\alpha_{\text{bl}},\beta)$

Outcome: $\text{p}(\alpha_{\text{im}},\alpha_{\text{bl}},\beta,f,h,g)$.

3 Bayesian Inference

GOAL

Calculating (or estimating)

$$\mathrm{p}(\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta, f, h|g) \tag{6}$$

to perform inference on the unknown hyperparameters, blur and image.

Remember that in the classical restoration problem (blur known) we had

$$\mathrm{p}(\alpha_{\mathrm{im}}, \beta, f|g)$$

and first we calculated (no approximation was needed in general but see medical images)

$$\begin{aligned}\hat{\alpha}_{\mathrm{im}}, \hat{\beta} &= \arg\max_{\alpha_{\mathrm{im}},\beta} \mathrm{p}(\alpha_{\mathrm{im}}, \beta|g) \\ &= \arg\max_{\alpha_{\mathrm{im}},\beta} \int_f \mathrm{p}(\alpha_{\mathrm{im}}, \beta, f|g)df\end{aligned}$$

Then we calculated

$$\hat{f} = \arg\max_f \mathrm{p}(f|g, \hat{\alpha}_{\mathrm{im}}, \hat{\beta})$$

For the blind deconvolution problem we cannot easily calculate

$$\mathrm{p}(\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta | g)$$

and

$$\mathrm{p}(f, h | g, \alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta).$$

WHY?

And we will have to approximate these distributions.

Note that instead of modes of posterior distributions we could also use mean values.

4 Variational Method

Let us use

$$\Theta = (\alpha_{\text{im}}, \alpha_{\text{bl}}, \beta, f, h) \tag{7}$$

We want to approximate $\mathrm{p}(\Theta|g)$.
Following the variational approach,

we approximate $\mathrm{p}(\Theta \mid g)$ by $\mathrm{q}(\Theta) = \mathrm{q}(\Omega)\mathrm{q}(f)\mathrm{q}(h)$ (8)

$\mathrm{q}(f)$ and $\mathrm{q}(h)$ denote distributions on f and h
$\mathrm{q}(\Omega)$ is a distribution on

$$\Omega = (\alpha_{\text{im}}, \alpha_{\text{bl}}, \beta) \tag{9}$$

where usually

$$\mathrm{q}(\Omega) = \mathrm{q}(\alpha_{\text{im}})\mathrm{q}(\alpha_{\text{bl}})\mathrm{q}(\beta) \tag{10}$$

The criterion we use to find $\mathrm{q}(\Theta)$ is to minimize the Kullback-Leibler divergence:

$$\begin{aligned} C_{KL}(\mathrm{q}(\Theta) \parallel \mathrm{p}(\Theta|g)) &= \int_{\Theta} \mathrm{q}(\Theta) \log \left(\frac{\mathrm{q}(\Theta)}{\mathrm{p}(\Theta|g)} \right) d\Theta \\ &= \int_{\Theta} \mathrm{q}(\Theta) \log \left(\frac{\mathrm{q}(\Theta)}{\mathrm{p}(\Theta, g)} \right) d\Theta \\ &+ \text{const.} \end{aligned} \tag{11}$$

We note that

$$\begin{aligned} C_{KL}(\mathrm{q}(\Theta) \parallel \mathrm{p}(\Theta|g)) &\geq 0 \quad \forall \ \mathrm{q}(\Theta) \\ C_{KL}(\mathrm{q}(\Theta) \parallel \mathrm{p}(\Theta|g)) &= 0 \quad \text{iff} \ \mathrm{q}(\Theta) = \mathrm{p}(\Theta|g) \end{aligned}$$

$$C_{KL}(\mathrm{q}(\Theta) \parallel \mathrm{p}(\Theta|g)) = C_{KL}(\mathrm{q}(\Omega)\mathrm{q}(f)\mathrm{q}(h) \parallel \mathrm{p}(\Theta|g))$$

$$= \int_f \mathrm{q}(f) \left(\int_{\Omega,h} \mathrm{q}(\Omega)\mathrm{q}(h) \log \left(\frac{\mathrm{q}(\Omega)\mathrm{q}(f)\mathrm{q}(h)}{\mathrm{p}(\Omega, f, h, g)} \right) d\Omega dh \right) df$$

$$+\text{const} \tag{12}$$

$$= \int_h \mathrm{q}(h) \left(\int_{\Omega,f} \mathrm{q}(\Omega)\mathrm{q}(f) \log \left(\frac{\mathrm{q}(\Omega)\mathrm{q}(f)\mathrm{q}(h)}{\mathrm{p}(\Omega, f, h, g)} \right) d\Omega df \right) dh$$

$$+ \quad \text{const} \tag{13}$$

$$= \int_\Omega \mathrm{q}(\Omega) \left(\int_{f,h} \mathrm{q}(f)\mathrm{q}(h) \log \left(\frac{\mathrm{q}(\Omega)\mathrm{q}(f)\mathrm{q}(h)}{\mathrm{p}(\Omega, f, h, g)} \right) df dh \right) d\Omega$$

$$+ \quad \text{const} \tag{14}$$

For $\omega \in \{\Omega, f, h\}$, Θ_ω is the subset of Θ with ω removed.

For example, if $\omega = f$ then $\Theta_f = (\Omega, h)$.

From the previous equations we have

$$C_{KL}(\mathrm{q}(\omega)\mathrm{q}(\Theta_\omega)) \parallel \mathrm{p}(\Theta|g))$$
$$= \int_\omega \mathrm{q}(\omega) \left(\int_{\Theta_\omega} \mathrm{q}(\Theta_\omega) \log \left(\frac{\mathrm{q}(\omega)\mathrm{q}(\Theta_\omega)}{\mathrm{p}(\omega, \Theta_\omega, g)} \right) d\Theta_\omega \right) d\omega + \text{const} \tag{15}$$

Given $\mathrm{q}(\Theta_\omega)$, in order to obtain

$$\hat{\mathrm{q}}(\omega) = \arg\min_{\mathrm{q}(\omega)} C_{KL} \left(\mathrm{q}(\omega)\mathrm{q}(\Theta_\omega)) \parallel \mathrm{p}(\Theta|g) \right), \tag{16}$$

we differentiate equation (15) with respect to $\mathrm{q}(\omega)$

$$\frac{\partial C_{KL}\left(\mathrm{q}(\omega)\mathrm{q}(\Theta_\omega) \parallel \mathrm{p}(\Theta|g)\right)}{\partial \mathrm{q}(\omega)}$$

$$= \int_{\Theta_\omega} \mathrm{q}(\Theta_\omega) \log\left(\frac{\mathrm{q}(\omega)\mathrm{q}(\Theta_\omega)}{\mathrm{p}(\omega, \Theta_\omega, g)}\right) d\Theta_\omega + \text{constant}$$

$$= \log \mathrm{q}(\omega) - \mathrm{E}\left[\, \log \mathrm{p}(\Theta)\mathrm{p}(g|\Theta) \,\right]_{\mathrm{q}(\Theta_\omega)} + \text{cst}$$

and so the solution of

$$\hat{\mathrm{q}}(\omega) = \arg\min_{\mathrm{q}(\omega)} C_{KL}\left(\mathrm{q}(\omega)\mathrm{q}(\Theta_\omega)\right) \parallel \mathrm{p}(\Theta|g)\right),$$

is

$$\hat{\mathrm{q}}(\omega) = \text{const} \times \exp\left(\mathrm{E}\left[\, \log \mathrm{p}(\Theta)\mathrm{p}(g|\Theta) \,\right]_{\mathrm{q}(\Theta_\omega)}\right), \qquad (17)$$

where

$$\mathrm{E}\left[\, \log \mathrm{p}(\Theta)\mathrm{p}(g|\Theta) \,\right]_{\mathrm{q}(\Theta_\omega)} = \int_{\Theta_\omega} \log\left[\mathrm{p}(\Theta)\mathrm{p}(g \mid \Theta)\right] \mathrm{q}(\Theta_\omega) d\Theta_\omega.$$

The above equations suggest the following iterative procedure to find q(Θ).

Algorithm 1

Given $q^1(h)$ *and* $q^1(\Omega)$, *current estimates of the distributions* $q(h)$, $q(\Omega)$,

For $k = 1, 2, \ldots$ *until an stopping criterion is met:*

1. Find

$$q^k(f) = \arg\min_{q(f)} C_{KL}(q^k(\Omega)q(f)q^k(h) \parallel p(\Omega, f, h|g)), \quad (18)$$

2. Find

$$q^{k+1}(h) = \arg\min_{q(h)} C_{KL}(q^k(\Omega)q^k(f)q(h) \parallel p(\Omega, f, h|g)), \quad (19)$$

3. Find

$$q^{k+1}(\Omega) = \arg\min_{q(\Omega)} C_{KL}(q(\Omega)q^k(f)q^{k+1}(h) \parallel p(\Omega, f, h|g)) \quad (20)$$

5 Variational approaches of the posterior distribution

5.1 Known hyperparameters

Let us first assume that $\Omega = (\alpha_{\text{im}}, \alpha_{\text{bl}}, \beta)$, is known and proceed to estimate $\mathrm{q}(f)$ and $\mathrm{q}(h)$.
Let us denote by $\underline{\Omega}$ the known value of Ω, knowing Ω is equivalent to assume

$$\mathrm{q}^k(\Omega) = \begin{cases} 1 & \text{if } \Omega = \underline{\Omega} \\ 0 & \text{otherwise} \end{cases}, \tag{21}$$

In what follows we will not underline Ω for simplicity but we will assume that remains unchanged during the estimation process

5.1.1 Approximation of $\mathbf{p}(f, h|g)$ by random distributions on f and h

Let us now assume that

$$\mathrm{q}^k_{BR}(h) = \mathcal{N}(h \mid E^k_{BR}(h), cov^k_{BR}(h))$$

We are using the subscript BR to denote that the distributions on f and h are random (non-degenerate).

From equation (17) $\mathrm{q}^k_{BR}(f)$ satisfies:

$$\begin{aligned} & -2\log \mathrm{q}^k_{BR}(f) \\ &= \text{const} + \alpha_{\text{im}} \parallel Cf \parallel^2 + \beta E[\parallel g - Hf \parallel^2]_{\mathrm{q}^k_{BR}(h)} \\ &= \text{const} + \alpha_{\text{im}} \parallel Cf \parallel^2 + \beta \parallel g - E^k_{BR}(H)f \parallel^2 \\ &+ \beta f^t [cov^k_{BR}(h)] f, \end{aligned} \tag{22}$$

So we have

$$\mathrm{q}_{BR}^{k}(f) = \mathcal{N}\left(f \mid E_{BR}^{k}(f), cov_{BR}^{k}(f)\right), \tag{23}$$

In order to calculate the mean and covariance of the normal distribution we note that the mean is the solution of

$$\frac{\partial\, 2 \log \mathrm{q}_{BR}^{k}(f)}{\partial f} = 0, \tag{24}$$

and the covariance is given by

$$-\frac{\partial^2\, 2 \log \mathrm{q}_{BR}^{k}(f)}{\partial f^2} = [cov_{BR}^{k}(f)]^{-1}. \tag{25}$$

We obtain

$$
\begin{aligned}
E^k_{BR}(f) &= \left(\alpha_{\text{im}} C^t C + \beta E^k_{BR}(H)^t E^k_{BR}(H) + \beta cov^k_{BR}(h)\right)^{-1} \\
&\times \beta E^k_{BR}(H)^t g, \qquad (26) \\
cov^k_{BR}(f) &= \left(\alpha_{\text{im}} C^t C + \beta E^k_{BR}(H)^t E^k_{BR}(H) + \beta cov^k_{BR}(h)\right)^{-1} \qquad (27)
\end{aligned}
$$

When $\mathrm{q}^k(f)$ has been calculated, we obtain

$$
\mathrm{q}^{k+1}_{BR}(h) = \mathcal{N}\left(h \mid E^{k+1}_{BR}(h), cov^{k+1}_{BR}(h)\right), \qquad (28)
$$

with

$$
\begin{aligned}
E^{k+1}_{BR}(h) &= \left(\alpha_{\text{bl}} C^t C + \beta E^k_{BR}(F)^t E^k_{BR}(F) + \beta [cov^k_{BR}(f)]\right)^{-1} \\
&\times \beta E^k_{BR}(F)^t g, \qquad (29 \\
cov^{k+1}_{BR}(h) &= \left(\alpha_{\text{bl}} C^t C + \beta E^k_{BR}(F)^t E^k_{BR}(F) + \beta [cov^k_{BR}(f)]\right)^{-1} \qquad (30
\end{aligned}
$$

5.1.2 Approximation of p$(f, h|g)$ by a degenerate distribution on h and a random distribution on f

Let us now try to find the best combination of distributions q(f), random on f, and q(h) deterministic on h that best approximate p$(f, h \mid g)$.

Let us assume that h^k is the value of the blurring function where q$^k_{OR}(h)$ is degenerate

$$\mathrm{q}^k_{OR}(h) = \begin{cases} 1 & \text{if } h = h^k \\ 0 & \text{otherwise} \end{cases}, \tag{31}$$

we are using the subscript OR to denote that only the distribution on f is non-degenerate.

We then have

$$E^k_{OR}(h) = h^k \text{ and } cov^k_{OR}(h) = \mathbf{0} \tag{32}$$

We then obtain from equation (17) and using equation (32) in equations (26) and (27) that

$$\mathrm{q}^k_{OR}(f) = \mathcal{N}\left(f \mid E^k_{OR}(f), cov^k_{OR}(f)\right) \tag{33}$$

where

$$\begin{aligned} E^k_{OR}(f) &= \left(\alpha_{\text{im}} C^t C + \beta E^k_{OR}(H)^t E^k_{OR}(H)\right)^{-1} \\ &\times \beta E^k_{OR}(H)^t g, \end{aligned} \tag{34}$$

$$cov^k_{OR}(f) = \left(\alpha_{\text{im}} C^t C + \beta E^k_{OR}(H)^t E^k_{OR}(H)\right)^{-1}. \tag{35}$$

Given now the distribution $\mathrm{q}_{OR}^{k}(f)$, the best estimation of the conditional distribution of the blur is given by

$$\underline{q}_{OR}^{k+1}(h) = \mathcal{N}\left(h \mid E_{OR}^{k+1}(h), cov_{OR}^{k+1}(h)\right), \tag{36}$$

with

$$\begin{aligned} E_{OR}^{k+1}(h) &= \left(\alpha_{\mathrm{bl}} C^t C + \beta E_{OR}^{k}(F)^t E_{OR}^{k}(F) + \beta[cov_{OR}^{k}(f)]\right)^{-1} \\ &\times \beta E_{OR}^{k}(F)^t g, \qquad (37 \\ cov_{OR}^{k+1}(h) &= \left(\alpha_{\mathrm{bl}} C^t C + \beta E_{OR}^{k}(F)^t E_{OR}^{k}(F) + \beta[cov_{OR}^{k}(f)]\right)^{-1} \qquad (38 \end{aligned}$$

note that $\underline{q}_{OR}^{k+1}$ is not a degenerate distribution.

The degenerate distribution $\mathrm{q}_{OR}^{k+1}(h)$ is given by

$$\mathrm{q}_{OR}^{k+1}(h) = \begin{cases} 1 & \text{if } h = E_{OR}^{k+1}(h) \\ 0 & \text{otherwise} \end{cases}. \tag{39}$$

5.1.3 Approximation of $\mathbf{p}(f, h|g)$ by degenerate distributions on f and h

Let us assume that $\mathrm{q}(f)$ and $\mathrm{q}(h)$ are degenerate.

Let $E^k_{BD}(h)$ be the current blur estimation where we assume that the degenerate distribution $\mathrm{q}^k_{BD}(h)$ is located (we use the subscript BD to denote that the distributions of f and h are both deterministic).

We have

$$E^k_{BD}(f) = \left(\alpha_{\text{im}} C^t C + \beta E^k_{BD}(H)^t E^k_{BD}(H)\right)^{-1} \times \beta E^k_{BD}(H)^t g, \tag{40}$$

$$E^{k+1}_{BD}(h) = \left(\alpha_{\text{bl}} C^t C + \beta E^k_{BD}(F)^t E^k_{BD}(F)\right)^{-1} \times \beta E^k_{BD}(F)^t g, \tag{41}$$

Note that this iterative procedure is equivalent to solving

$$E_{BD}^{k}(f) = \arg\min_{f} \alpha_{\text{im}} \parallel Cf \parallel^2 + \beta \parallel g - E_{BD}^{k}(H)f \parallel^2, \tag{42}$$

and then

$$E_{BD}^{k+1}(h) = \arg\min_{h} \alpha_{\text{bl}} \parallel Ch \parallel^2 + \beta \parallel g - E_{BD}^{k}(F)h \parallel^2 . \tag{43}$$

5.2 Parameter estimation

Let us now assume that the hyperparameter vector $\Omega = (\alpha_{\text{bl}}, \alpha_{\text{im}}, \beta)$ is unknown.

For $\omega \in \Omega$ we will use gamma distributions, $\Gamma(a, b)$, as hyperpriors, that is,

$$\mathrm{p}(\omega) = \frac{b^a}{\Gamma(a)} \omega^{a-1} \exp[-b\,\omega], \tag{44}$$

$b > 0$ is the scale parameter and $a > 0$ is the shape parameter.

This distribution has the following properties

$$E[w] = \frac{a}{b} \qquad \text{and} \qquad Var[w] = \frac{a}{b^2}. \tag{45}$$

5.2.1 A priori unrelated variances of the distributions

Let us assume that the hyperparameters have the following independent prior distributions

$$\mathrm{p}(\omega) = \Gamma(\omega | a_\omega, b_\omega) \tag{46}$$

$\omega \in \Omega$

In order to find $q^{k+1}(\omega)$, $\omega \in \Omega$ in equation (20) of Algorithm 1 we have to calculate

$$E\left[\, \log \mathrm{p}(g|\Theta)\mathrm{p}(\Theta) \,\right]_{\mathbf{q}^k(f)\mathbf{q}^{k+1}(h)} \tag{47}$$

see equation (17).

We have

$$
\begin{aligned}
& E\left[\,\log \mathrm{p}(g|\Theta)\mathrm{p}(\Theta)\,\right]_{\mathrm{q}^k(f)\mathrm{q}^{k+1}(h)} \qquad (48)\\
&= \text{const} + \sum_{\omega\in\{\alpha_{\mathrm{bl}},\alpha_{\mathrm{im}},\beta\}} \left((a_\omega - 1)\log\omega - \omega b_\omega\right)\\
&+ \frac{N}{2}\log\alpha_{\mathrm{im}} + \frac{M}{2}\alpha_{\mathrm{bl}} + \frac{N}{2}\log\beta\\
&- \frac{1}{2}\alpha_{\mathrm{im}} E\left[\|\, Cf \,\|^2\right]_{\mathrm{q}^k(f)} - \frac{1}{2}\alpha_{\mathrm{bl}} E\left[\|\, Ch \,\|^2\right]_{\mathrm{q}^{k+1}(h)}\\
&- \frac{1}{2}\beta E\left[\|\, g - Hf \,\|^2\right]_{\mathrm{q}^k(f)\mathrm{q}^{k+1}(h)}. \qquad (49)
\end{aligned}
$$

From the above equations we have

$$\mathrm{q}^{k+1}(\omega) = \Gamma(\omega | a_\omega^{k+1}, b_\omega^{k+1}) \tag{50}$$

where a_ω^{k+1} and b_ω^{k+1} are

$$a_{\alpha_{\mathrm{im}}}^{k+1} = a_{\alpha_{\mathrm{im}}} + \frac{N}{2} \tag{51}$$

$$b_{\alpha_{\mathrm{im}}}^{k+1} = b_{\alpha_{\mathrm{im}}} + \frac{1}{2} E\left[\| Cf \|^2\right]_{\mathrm{q}^k(f)} \tag{52}$$

$$a_{\alpha_{\mathrm{bl}}}^{k+1} = a_{\alpha_{\mathrm{bl}}} + \frac{M}{2} \tag{53}$$

$$b_{\alpha_{\mathrm{bl}}}^{k+1} = b_{\alpha_{\mathrm{bl}}} + \frac{1}{2} E\left[\| Ch \|^2\right]_{\mathrm{q}^{k+1}(h)} \tag{54}$$

$$a_\beta^{k+1} = a_\beta + \frac{N}{2} \tag{55}$$

$$b_\beta^{k+1} = b_\beta + \frac{1}{2} E\left[\| g - Hf \|^2\right]_{\mathrm{q}^k(f)\mathrm{q}^{k+1}(h)} \tag{56}$$

The process to re-estimate the image and blur once $q^{k+1}(\omega)$, $\omega \in \Omega$ has been found is the same as the one described in section 5.1 using now as known hyperparameters the mean values of the distributions $\mathrm{q}^{k+1}(\omega)$ which are, see equation (45), $a_\omega^{k+1}/b_\omega^{k+1}$.

The only remaining problem is the calculation of $E\left[\| Cf \|^2\right]_{\mathrm{q}^k(f)}$, $E\left[\| Ch \|^2\right]_{\mathrm{q}^{k+1}(h)}$, and $E\left[\| g - Hf \|^2\right]_{\mathrm{q}^k(f)\mathrm{q}^{k+1}(h)}$ this is done in Appendix A.

5.2.2 Relating variances of the prior distributions

A simple observation of the problem makes us aware of the fact that at least α_{im} and α_{bl} vary in very different ranges.

Let

$$\text{sum}_g = \sum_i g_i, \tag{57}$$

we could assume a priori that f/sum_g and h have the same probabilistic properties.

A priori

$$[Cf]_i = \epsilon_i \quad \text{and} \quad [Ch]_j = \mu_j \tag{58}$$

where all ϵ_i, $i = 1, \ldots, N$ and μ_j, $j = 1, \ldots, M$ are independent with variances α_{im}^{-1} and α_{bl}^{-1} respectively. We could write

$$\frac{1}{(\text{sum}_g)^2} \frac{1}{\alpha_{\text{im}}} = \frac{1}{\alpha_{\text{bl}}} \tag{59}$$

which is equivalent to

$$\alpha_{\text{bl}} \approx (\text{sum}_g)^2 \alpha_{\text{im}}. \tag{60}$$

Following a regularization based approach, You and Kaveh suggest the following relationship between the prior variances of the image and blur

$$\alpha_{\text{bl}} \approx (\max_j f_j)\, \text{sum}_g\, \alpha_{\text{im}} \tag{61}$$

where we are assuming that $\text{sum}_g = \text{sum}_f$ and usually $\max_j f_j = 255$.
What it is clear from the above discussion is that we can write

$$\alpha_{\text{bl}} \approx \gamma \alpha_{\text{im}}, \tag{62}$$

where γ is a known a priori value.

We are not writing

$$\alpha_{\text{bl}} = \gamma \alpha_{\text{im}}$$

in this case we would have only two hyperparameters β and α_{im} and we could use the previously described methodology to estimate the β, α_{im}, image and blur.

To take into account the existent relationship between the image and blur prior variances, see equation (62), we define a new set of hyperparameters

$$\Psi = (\beta, \mu, \nu) \tag{63}$$

the precise meaning of these hyperparameters will be made clear shortly.

Then use $\rho \sim \Gamma(a_\rho, b_\rho))$ as hyperpriors for $\rho \in \Psi$ and the following priors for image and blur

$$\mathrm{p}(f|\mu) \;\propto\; \mu^{N/2} \exp\left[-\frac{\mu}{2} \parallel Cf \parallel^2\right] \tag{64}$$

$$\mathrm{p}(h|\mu,\nu) \;\propto\; (\mu\nu)^{M/2} \exp\left[-\frac{\mu\nu}{2} \parallel Ch \parallel^2\right]. \tag{65}$$

From the above definitions,

$$\alpha_{\mathrm{im}} = \mu \;\text{ and }\; \alpha_{\mathrm{bl}} = \mu\nu \tag{66}$$

ν is used to relate the priors image and blur variances.

Note that the distribution on ν will be used to asses our confidence on the parameter γ in equation (62).

Our blind deconvolution modelling is now

$$\mathrm{p}(\beta, \mu, \nu, f, h, g) = \mathrm{p}(\beta)\mathrm{p}(\mu)\mathrm{p}(\nu)\mathrm{p}(f|\mu)\mathrm{p}(h|\mu, \nu)\mathrm{p}(g|f, h, \beta)$$

Updating the distributions $\mathrm{q}^{k+1}(\rho)$, $\rho \in \Psi$ in parallel

$$\begin{aligned}
a_\mu^{k+1} &= a_\mu + \frac{N+M}{2} \\
b_\mu^{k+1} &= b_\mu + \frac{1}{2} E\left[\| Cf \|^2\right]_{\mathrm{q}^k(f)} + \frac{1}{2} E\left[\nu\right]_{\mathrm{q}^k(\nu)} E\left[\| Ch \|^2\right]_{\mathrm{q}^{k+1}} \\
a_\nu^{k+1} &= a_\nu + \frac{M}{2} \\
b_\nu^{k+1} &= b_\nu + \frac{1}{2} E\left[\mu\right]_{\mathrm{q}^k(\mu)} E\left[\| Ch \|^2\right]_{\mathrm{q}^{k+1}(h)} \\
a_\beta^{k+1} &= a_\beta + \frac{N}{2} \\
b_\beta^{k+1} &= b_\beta + \frac{1}{2} E\left[\| g - Hf \|^2\right]_{\mathrm{q}^k(f)\mathrm{q}^{k+1}(h)}
\end{aligned}$$

From the obtained updated posterior distributions of the hyperparameters we have

$$E[\mu]_{q^{k+1}(\mu)} = \frac{a_\mu + \frac{N+M}{2}}{b_\mu + \frac{1}{2}E\left[\| Cf \|^2\right]_{\mathbf{q}^k(f)} + \frac{1}{2}E\left[\nu\right]_{\mathbf{q}^k(\nu)} E\left[\| Ch \|^2\right]_{\mathbf{q}^{k+1}(h)}}$$

$$E[\nu]_{q^{k+1}(\nu)} = \frac{a_\nu + \frac{M}{2}}{b_\nu + \frac{1}{2}E\left[\mu\right]_{\mathbf{q}^k(\mu)} E\left[\| Ch \|^2\right]_{\mathbf{q}^{k+1}(h)}}$$

$$E[\beta]_{q^{k+1}(\beta)} = \frac{a_\beta + \frac{N}{2}}{b_\beta + \frac{1}{2}E\left[\| g - Hf \|^2\right]_{\mathbf{q}^k(f)\mathbf{q}^{k+1}(h)}}$$

while, using equations (52)-(56) the posterior distributions of the hyperparameters in section 5.2.1 were

$$E[\alpha_{\text{im}}]_{q^{k+1}(\alpha_{\text{im}})} = \frac{a_{\alpha_{\text{im}}} + \frac{N}{2}}{b_{\alpha_{\text{im}}} + \frac{1}{2}E\left[\| Cf \|^2\right]_{\text{q}^k(f)}}$$

$$E[\alpha_{\text{bl}}]_{q^{k+1}(\alpha_{\text{bl}})} = \frac{a_{\alpha_{\text{bl}}} + \frac{M}{2}}{b_{\alpha_{\text{bl}}} + \frac{1}{2}E\left[\| Ch \|^2\right]_{\text{q}^{k+1}(h)}}$$

$$E[\beta]_{q^{k+1}(\beta)} = \frac{a_{\beta} + \frac{N}{2}}{b_{\beta} + \frac{1}{2}E\left[\| g - Hf \|^2\right]_{\text{q}^k(f)\text{q}^{k+1}(h)}}$$

In our experiments we will use non-informative priors for all our hyperparameters $\eta \in \Omega \cup \Psi - \{\nu\}$, that is, $a_\eta \approx 0$ and $b_\eta \approx 0$.

To introduce the relationship between the prior variances in equation (62) we assume that the parameters defining the hyperprior distribution of ν are

$$a_\nu = \gamma S/2 \;\text{ and }\; b_\nu = S/2 \tag{67}$$

Since $E[\nu] = \gamma$ and $\text{var}[\nu] = 2\gamma/S$, S measures the confidence on the relationship between the image and blur variances.

If S tends to infinity $E\left[\nu\right]_{\mathrm{q}^k(\nu)} = \gamma \;\; \forall k$

which is equivalent to assume $\alpha_{\mathrm{bl}} = \gamma \alpha_{\mathrm{im}}$

In our experiments we will use $S = N^0, N^1, N^2$ as measure of our prior confidence on the provided estimate of γ.

A Calculating the mean values

In this appendix we calculate the values of $E\left[\| Cf \|^2\right]_{\mathrm{q}^{k(f)}}$, $E\left[\| Ch \|^2\right]_{\mathrm{q}^{k+1(h)}}$, and $E\left[\| g - Hf \|^2\right]_{\mathrm{q}^{k(f)}\mathrm{q}^{k+1(h)}}$ for the distributions of f and h discussed in section 5.1.

$$
\begin{aligned}
E\left[\| Cf \|^2\right]_{\mathrm{q}^k_{BR}(f)} &= \| CE^k_{BR}(f) \|^2 + \mathrm{trace}(C^t C cov^k_{BR}(f)) \\
E\left[\| Ch \|^2\right]_{\mathrm{q}^{k+1}_{BR}(h)} &= \| CE^{k+1}_{BR}(h) \|^2 + \mathrm{trace}(C^t C cov^k_{BR}(h)) \\
E\left[\| g - Hf \|^2\right]_{\mathrm{q}^k_{BR}(f)\mathrm{q}^{k+1}_{BR}(h)} & \\
&= \| g - E^{k+1}_{BR}(h) E^k_{BR}(f) \|^2 + \mathrm{trace}(cov^k_{BR}(f) cov^{k+1}_{BR}(h)) \\
&+ \mathrm{trace}(E^k_{BR}(F)^t E^k_{BR}(F) cov^{k+1}_{BR}(h)) \\
&+ \mathrm{trace}(E^{k+1}_{BR}(H)^t E^{k+1}_{BR}(H) cov^k_{BR}(f))
\end{aligned}
$$

$$
\begin{aligned}
E\left[\| Cf \|^2\right]_{\mathrm{q}^k_{OR}(f)} &= \| CE^k_{OR}(f) \|^2 + \mathrm{trace}(C^t C cov^k_{OR}(f)) \\
E\left[\| Ch \|^2\right]_{\mathrm{q}^{k+1}_{OR}(h)} &= \| CE^{k+1}_{OR}(h) \|^2 \\
E\left[\| g - Hf \|^2\right]_{\mathrm{q}^k_{OR}(f)\mathrm{q}^{k+1}_{OR}(h)} & \\
&= \| g - E^{k+1}_{OR}(H) E^k_{OR}(f) \|^2 \\
&+ \mathrm{trace}(E^{k+1}_{OR}(H)^t E^{k+1}_{OR}(H) cov^k_{OR}(f))
\end{aligned}
$$

$$
\begin{aligned}
E\left[\| Cf \|^2\right]_{\mathrm{q}^k_{BD}(f)} &= \| CE^k_{BD}(f) \|^2 \\
E\left[\| Ch \|^2\right]_{\mathrm{q}^{k+1}_{BD}(h)} &= \| CE^{k+1}_{BD}(h) \|^2 \\
E\left[\| g - Hf \|^2\right]_{\mathrm{q}^k_{BD}(f)\mathrm{q}^{k+1}_{BD}(h)} &= \| g - E^{k+1}_{BD}(H) E^k_{BD}(f) \|^2
\end{aligned}
$$