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

\[ g(x) = h(x) \ast f(x) + n(x), \quad x = (x_1, x_2) \in I, \quad (1) \]

\( I \subset \mathbb{R}^2 \) 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.

\( (\ast) \) 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 $f$

$$p(f|\alpha_{\text{im}}) \propto \alpha_{\text{im}}^{N/2} \exp\left\{-\frac{1}{2} \alpha_{\text{im}} \parallel C f \parallel^2\right\}, \quad (2)$$

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

$N$ or $N - 1$?
Prior model for $h$

$$p(h|\alpha_{bl}) \propto \alpha_{bl}^{M/2} \exp\left\{-\frac{1}{2}\alpha_{bl} \| Ch \|^2\right\}, \quad (3)$$

- $C$ denotes the laplacian operator
- $M = U \times V$ is the known size of the support of the blur
- $h$ is the column vector of size $N$ lexicographically ordered by rows denoting the blur (this vector has all its components equal to zero outside its support)
- $\alpha_{bl}^{-1}$ 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 $\beta^{-1}$. That is,

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

(4)

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

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

When $\alpha_{im}$, $\alpha_{bl}$ and $\beta$ are unknown, the hierarchical Bayesian paradigm uses two stages

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

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

GOAL
Calculating (or estimating)

\[ p(\alpha_{im}, \alpha_{bl}, \beta, f, h | g) \]  \hspace{1cm} (6)

to perform inference on the unknown hyperparameters, blur and image.
Remember that in the classical restoration problem (blur known) we had

\[ p(\alpha_{im}, \beta, f|g) \]

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

\[
\hat{\alpha}_{im}, \hat{\beta} = \arg \max_{\alpha_{im}, \beta} p(\alpha_{im}, \beta|g)
\]

\[
= \arg \max_{\alpha_{im}, \beta} \int f p(\alpha_{im}, \beta, f|g) df
\]

Then we calculated

\[
\hat{f} = \arg \max_{f} p(f|g, \hat{\alpha}_{im}, \hat{\beta})
\]
For the blind deconvolution problem we cannot easily calculate

\[ p(\alpha_{im}, \alpha_{bl}, \beta | g) \]

and

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

WHY?

And we will have to approximate these distributions.

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

\[ \Theta = (\alpha_{\text{im}}, \alpha_{\text{bl}}, \beta, f, h) \quad (7) \]

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

we approximate \( p(\Theta | g) \) by \( q(\Theta) = q(\Omega)q(f)q(h) \quad (8) \)

\( q(f) \) and \( q(h) \) denote distributions on \( f \) and \( h \)
\( q(\Omega) \) is a distribution on

\[ \Omega = (\alpha_{\text{im}}, \alpha_{\text{bl}}, \beta) \quad (9) \]

where usually

\[ q(\Omega) = q(\alpha_{\text{im}})q(\alpha_{\text{bl}})q(\beta) \quad (10) \]
The criterion we use to find $q(\Theta)$ is to minimize the Kullback-Leibler divergence:

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

(11)

We note that

$$
C_{KL}(q(\Theta) \parallel p(\Theta|g)) \geq 0 \quad \forall \ q(\Theta)
$$

$$
C_{KL}(q(\Theta) \parallel p(\Theta|g)) = 0 \quad \text{iff} \ q(\Theta) = p(\Theta|g)
$$
\[ C_{KL}(q(\Theta) \parallel p(\Theta|g)) = C_{KL}(q(\Omega)q(f)q(h) \parallel p(\Theta|g)) \]

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

+ const (12)

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

+ const (13)

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

+ const (14)
For \( \omega \in \{ \Omega, f, h \} \), \( \Theta_\omega \) is the subset of \( \Theta \) with \( \omega \) removed.

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

From the previous equations we have

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

(15)

Given \( q(\Theta_\omega) \), in order to obtain

\[
\hat{q}(\omega) = \arg\min_{q(\omega)} C_{KL}(q(\omega)q(\Theta_\omega)) \parallel p(\Theta|g)) ,
\]

we differentiate equation (15) with respect to \( q(\omega) \)

Given q(\Theta_\omega), in order to obtain

\[
\hat{q}(\omega) = \arg\min_{q(\omega)} C_{KL}(q(\omega)q(\Theta_\omega)) \parallel p(\Theta|g)) ,
\]

we differentiate equation (15) with respect to q(\omega)
\[ \frac{\partial C_{KL}(q(\omega)q(\Theta_\omega) \parallel p(\Theta|g))}{\partial q(\omega)} \]

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

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

and so the solution of

\[ \hat{q}(\omega) = \arg \min_{q(\omega)} C_{KL}(q(\omega)q(\Theta_\omega) \parallel p(\Theta|g)), \]

is

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

where

\[ \mathbb{E} \left[ \log p(\Theta)p(g|\Theta) \right]_{q(\Theta_\omega)} = \int_{\Theta_\omega} \log [p(\Theta)p(g | \Theta)] q(\Theta_\omega) d\Theta_\omega. \]
The above equations suggest the following iterative procedure to find $q(\Theta)$.

**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 $\Omega = (\alpha_{im}, \alpha_{bl}, \beta)$, is known and proceed to estimate $q(f)$ and $q(h)$.

Let us denote by $\Omega$ the known value of $\Omega$, knowing $\Omega$ is equivalent to assume

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

In what follows we will not underline $\Omega$ for simplicity but we will assume that remains unchanged during the estimation process.
5.1.1 Approximation of $p(f, h|g)$ by random distributions on $f$ and $h$

Let us now assume that

$$q_{BR}^k(h) = \mathcal{N}(h \mid E_{BR}^k(h), \text{cov}_{BR}^k(h))$$

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

From equation (17) $q_{BR}^k(f)$ satisfies:

$$-2 \log q_{BR}^k(f) = \text{const} + \alpha_{im} \| Cf \|^2 + \beta E[\| g - Hf \|^2]q_{BR}^k(h)$$

$$= \text{const} + \alpha_{im} \| Cf \|^2 + \beta \| g - E_{BR}^k(H)f \|^2$$

$$+ \beta f^t [\text{cov}_{BR}^k(h)] f,$$

(22)
So we have

\[ q^k_{BR}(f) = \mathcal{N} \left( f \mid E^k_{BR}(f), \text{cov}^k_{BR}(f) \right), \quad (23) \]

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

\[ \frac{\partial}{\partial f} 2 \log q^k_{BR}(f) = 0, \quad (24) \]

and the covariance is given by

\[ - \frac{\partial^2}{\partial f^2} 2 \log q^k_{BR}(f) = \left[ \text{cov}^k_{BR}(f) \right]^{-1}. \quad (25) \]
We obtain

\[
E_{BR}^k(f) = \left( \alpha_{im} C^t C + \beta E_{BR}^k(H)^t E_{BR}^k(H) + \beta \text{cov}_{BR}^k(h) \right)^{-1} \\
\times \beta E_{BR}^k(H)^t g,
\]

(26)

\[
cov_{BR}^k(f) = \left( \alpha_{im} C^t C + \beta E_{BR}^k(H)^t E_{BR}^k(H) + \beta \text{cov}_{BR}^k(h) \right)^{-1} \\
\times \beta \text{cov}_{BR}^k(h),
\]

(27)

When \( q^k(f) \) has been calculated, we obtain

\[
q_{BR}^{k+1}(h) = \mathcal{N}(h \mid E_{BR}^{k+1}(h), \text{cov}_{BR}^{k+1}(h)),
\]

(28)

with

\[
E_{BR}^{k+1}(h) = \left( \alpha_{bl} C^t C + \beta E_{BR}^k(F)^t E_{BR}^k(F) + \beta [\text{cov}_{BR}^k(f)] \right)^{-1} \\
\times \beta E_{BR}^k(F)^t g,
\]

(29)

\[
cov_{BR}^{k+1}(h) = \left( \alpha_{bl} C^t C + \beta E_{BR}^k(F)^t E_{BR}^k(F) + \beta [\text{cov}_{BR}^k(f)] \right)^{-1}.
\]

(30)
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 $p(f, h | g)$.

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

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

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

\[ E^k_{OR}(h) = h^k \quad \text{and} \quad cov^k_{OR}(h) = 0 \]  

(32)

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

\[ q^k_{OR}(f) = \mathcal{N}(f \mid E^k_{OR}(f), cov^k_{OR}(f)) \]  

(33)

where

\[ E^k_{OR}(f) = \left( \alpha_{im}C^tC + \beta E^k_{OR}(H)^t E^k_{OR}(H) \right)^{-1} \times \beta E^k_{OR}(H)^t g, \]  

(34)

\[ cov^k_{OR}(f) = \left( \alpha_{im}C^tC + \beta E^k_{OR}(H)^t E^k_{OR}(H) \right)^{-1}. \]  

(35)
Given now the distribution $q_{OR}^k(f)$, the best estimation of the conditional distribution of the blur is given by

$$
q_{OR}^{k+1}(h) = \mathcal{N}(h \mid E_{OR}^{k+1}(h), cov_{OR}^{k+1}(h)),
$$

with

$$
E_{OR}^{k+1}(h) = \left( \alpha_{bl} C^t C + \beta E_{OR}^k(F)^t E_{OR}^k(F) + \beta [cov_{OR}^k(f)] \right)^{-1} \times \beta E_{OR}^k(F)^t g,
$$

$$
cov_{OR}^{k+1}(h) = \left( \alpha_{bl} C^t C + \beta E_{OR}^k(F)^t E_{OR}^k(F) + \beta [cov_{OR}^k(f)] \right)^{-1}.
$$

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

The degenerate distribution $q_{OR}^{k+1}(h)$ is given by

$$
q_{OR}^{k+1}(h) = \begin{cases} 
1 & \text{if } h = E_{OR}^{k+1}(h) \\
0 & \text{otherwise}
\end{cases}
$$
5.1.3 Approximation of \( p(f, h|g) \) by degenerate distributions on \( f \) and \( h \)

Let us assume that \( q(f) \) and \( q(h) \) are degenerate.

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

We have

\[
E_{BD}^k(f) = \left( \alpha_{\text{im}} C^t C + \beta E_{BD}^k(H)^t E_{BD}^k(H) \right)^{-1} \times \beta E_{BD}^k(H)^t g, \tag{40}
\]

\[
E_{BD}^{k+1}(h) = \left( \alpha_{\text{bl}} C^t C + \beta E_{BD}^k(F)^t E_{BD}^k(F) \right)^{-1} \times \beta E_{BD}^k(F)^t g, \tag{41}
\]
Note that this iterative procedure is equivalent to solving

\[
E_{BD}^k(f) = \arg \min_{f} \alpha_{\text{im}} \| Cf \|^2 + \beta \| g - E_{BD}^k(H)f \|^2 ,
\]

(42)

and then

\[
E_{BD}^{k+1}(h) = \arg \min_{h} \alpha_{\text{bl}} \| Ch \|^2 + \beta \| g - E_{BD}^k(F')h \|^2 .
\]

(43)
5.2 Parameter estimation

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

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

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

\( b > 0 \) is the scale parameter and \( a > 0 \) is the shape parameter.

This distribution has the following properties

\[
E[w] = \frac{a}{b} \quad \text{and} \quad Var[w] = \frac{a}{b^2}. \tag{45}
\]
5.2.1 A priori unrelated variances of the distributions

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

\[ p(\omega) = \Gamma(\omega | a_\omega, b_\omega) \]  \hspace{1cm} (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

\[ E \left[ \log p(g | \Theta)p(\Theta) \right] q^k(f)q^{k+1}(h) \]  \hspace{1cm} (47)

see equation (17).
We have

\[
E \left[ \log p(g|\Theta)p(\Theta) \right] q^k(f)q^{k+1}(h)
\]

\[
= \text{const} + \sum_{\omega \in \{\alpha_{\text{bl}}, \alpha_{\text{im}}, \beta\}} \left( (a_\omega - 1) \log \omega - \omega b_\omega \right)
\]

\[
+ \frac{N}{2} \log \alpha_{\text{im}} + \frac{M}{2} \alpha_{\text{bl}} + \frac{N}{2} \log \beta
\]

\[
- \frac{1}{2} \alpha_{\text{im}} E \left[ \| Cf \|^2 \right] q^k(f) - \frac{1}{2} \alpha_{\text{bl}} E \left[ \| Ch \|^2 \right] q^{k+1}(h)
\]

\[
- \frac{1}{2} \beta E \left[ \| g - Hf \|^2 \right] q^k(f)q^{k+1}(h). \]
From the above equations we have

\[ q^{k+1}(\omega) = \Gamma(\omega | a_{\omega}^{k+1}, b_{\omega}^{k+1}) \]  

(50)

where \( a_{\omega}^{k+1} \) and \( b_{\omega}^{k+1} \) are

\[ a_{\alpha, \text{im}}^{k+1} = a_{\alpha, \text{im}} + \frac{N}{2} \]  

(51)

\[ b_{\alpha, \text{im}}^{k+1} = b_{\alpha, \text{im}} + \frac{1}{2} E \left[ \| Cf \|^{2} \right] q^{k}(f) \]  

(52)

\[ a_{\alpha, \text{bl}}^{k+1} = a_{\alpha, \text{bl}} + \frac{M}{2} \]  

(53)

\[ b_{\alpha, \text{bl}}^{k+1} = b_{\alpha, \text{bl}} + \frac{1}{2} E \left[ \| Ch \|^{2} \right] q^{k+1}(h) \]  

(54)

\[ a_{\beta}^{k+1} = a_{\beta} + \frac{N}{2} \]  

(55)

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

(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 $q^{k+1}(\omega)$ which are, see equation (45), $a^{k+1}_\omega / b^{k+1}_\omega$.

The only remaining problem is the calculation of $E [\| Cf \|^2] q^k(f)$, $E [\| Ch \|^2] q^{k+1}(h)$, and $E [\| g - H f \|^2] q^k(f) q^{k+1}(h)$ 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 $\alpha_{\text{im}}$ and $\alpha_{\text{bl}}$ vary in very different ranges.

Let

$$\sum_{g} = \sum_{i} g_{i},$$

we could assume a priori that $f/\sum_{g}$ and $h$ have the same probabilistic properties.
A priori

\[ [Cf]_i = \varepsilon_i \quad \text{and} \quad [Ch]_j = \mu_j \quad (58) \]

where all \( \varepsilon_i, i = 1, \ldots, N \) and \( \mu_j, j = 1, \ldots, M \) are independent with variances \( \alpha_{im}^{-1} \) and \( \alpha_{bl}^{-1} \) respectively.

We could write

\[
\frac{1}{(\sum g)^2} \frac{1}{\alpha_{im}} = \frac{1}{\alpha_{bl}} \quad (59)
\]

which is equivalent to

\[
\alpha_{bl} \approx (\sum g)^2 \alpha_{im}. \quad (60)
\]
Following a regularization based approach, You and Kaveh suggest the following relationship between the prior variances of the image and blur

\[ \alpha_{bl} \approx (\max_j f_j) \sum_g \alpha_{im} \]  

(61)

where we are assuming that \( \sum_g = \sum_f \) and usually \( \max_j f_j = 255 \).

What it is clear from the above discussion is that we can write

\[ \alpha_{bl} \approx \gamma \alpha_{im} \]  

(62)

where \( \gamma \) 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 \( \beta \) and \( \alpha_{\text{im}} \) and we could use the previously described methodology to estimate the \( \beta, \alpha_{\text{im}}, \) 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

\[ \Psi = (\beta, \mu, \nu) \]  \hspace{1cm} (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

\begin{align}
p(f|\mu) &\propto \mu^{N/2} \exp \left[ -\frac{\mu}{2} \| Cf \|^2 \right] \quad (64) \\
p(h|\mu, \nu) &\propto (\mu \nu)^{M/2} \exp \left[ -\frac{\mu \nu}{2} \| Ch \|^2 \right]. \quad (65)
\end{align}

From the above definitions,

$$\alpha_{\text{im}} = \mu \quad \text{and} \quad \alpha_{\text{bl}} = \mu \nu$$

(66)

$\nu$ is used to relate the priors image and blur variances.

Note that the distribution on $\nu$ will be used to assess our confidence on the parameter $\gamma$ in equation (62).
Our blind deconvolution modelling is now

\[
p(\beta, \mu, \nu, f, h, g) = p(\beta)p(\mu)p(\nu)p(f|\mu)p(h|\mu, \nu)p(g|f, h, \beta)
\]

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

\[
a^{k+1}_\mu = a_\mu + \frac{N + M}{2}
\]

\[
b^{k+1}_\mu = b_\mu + \frac{1}{2}E[\| Cf \|^2]q^k(f) + \frac{1}{2}E[\nu]q^k(\nu)E[\| Ch \|^2]q^{k+1}(h)
\]

\[
a^{k+1}_\nu = a_\nu + \frac{M}{2}
\]

\[
b^{k+1}_\nu = b_\nu + \frac{1}{2}E[\mu]q^k(\mu)E[\| Ch \|^2]q^{k+1}(h)
\]

\[
a^{k+1}_\beta = a_\beta + \frac{N}{2}
\]

\[
b^{k+1}_\beta = b_\beta + \frac{1}{2}E[\| g - Hf \|^2]q^k(f)q^{k+1}(h)
\]
From the obtained updated posterior distributions of the hyperparameters we have

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

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

\[
E[\beta]_{q^{k+1}(\beta)} = \frac{a_\beta + \frac{N}{2}}{b_\beta + \frac{1}{2}E[\| g - Hf \|^2]q^{k}(f)q^{k+1}(h)}
\]
while, using equations (52)-(56) the posterior distributions of the hyperparameters in section 5.2.1 were

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

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

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

$$a_\nu = \gamma S/2 \quad \text{and} \quad b_\nu = S/2$$

(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 $E[\nu]_{q^k(\nu)} = \gamma \quad \forall k$

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

In our experiments we will use $S = N^0, N^1, N^2$ as measure of our prior confidence on the provided estimate of $\gamma$. 
A  Calculating the mean values

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

\[
E[\| Cf \|^2]q^k_{BR}(f) = \| CE^k_{BR}(f) \|^2 + \text{trace}(C^t C \text{cov}^k_{BR}(f))
\]

\[
E[\| Ch \|^2]q^{k+1}_{BR}(h) = \| CE^{k+1}_{BR}(h) \|^2 + \text{trace}(C^t C \text{cov}^k_{BR}(h))
\]

\[
E[\| g - Hf \|^2]q^k_{BR}(f)q^{k+1}_{BR}(h) = \| g - E^{k+1}_{BR}(h)E^k_{BR}(f) \|^2 + \text{trace}(\text{cov}^k_{BR}(f) \text{cov}^{k+1}_{BR}(h))
\]

\[+ \text{trace}(E^k_{BR}(F)^t E^k_{BR}(F) \text{cov}^{k+1}_{BR}(h))
\]

\[+ \text{trace}(E^{k+1}_{BR}(H)^t E^{k+1}_{BR}(H) \text{cov}^k_{BR}(f))
\]
\[ E \left[ \| \mathbf{Cf} \|_2^2 \right] q_{OR}^k(f) = \| CE_{OR}^k(f) \|^2 + \text{trace}(C^t Cov_{OR}^k(f)) \]

\[ E \left[ \| \mathbf{Ch} \|_2^2 \right] q_{OR}^{k+1}(h) = \| CE_{OR}^{k+1}(h) \|^2 \]

\[ E \left[ \| \mathbf{g} - \mathbf{Hf} \|_2^2 \right] q_{OR}^k(f) q_{OR}^{k+1}(h) = \| g - E_{OR}^{k+1}(H) E_{OR}^k(f) \|^2 \]

\[ + \text{trace}(E_{OR}^{k+1}(H)^t E_{OR}^{k+1}(H) Cov_{OR}^k(f)) \]

\[ E \left[ \| \mathbf{Cf} \|_2^2 \right] q_{BD}^k(f) = \| CE_{BD}^k(f) \|^2 \]

\[ E \left[ \| \mathbf{Ch} \|_2^2 \right] q_{BD}^{k+1}(h) = \| CE_{BD}^{k+1}(h) \|^2 \]

\[ E \left[ \| \mathbf{g} - \mathbf{Hf} \|_2^2 \right] q_{BD}^k(f) q_{BD}^{k+1}(h) = \| g - E_{BD}^{k+1}(H) E_{BD}^k(f) \|^2 \]