Non-uniform Deblurring for Shaken Images: Derivation of parameter update equations for blind de-blurring

Oliver Whyte

Abstract

This note outlines the derivation of the parameter update formulas for the variational non-uniform blind deblurring algorithm described in Whyte et al. *[4]. First, using the calculus of variations, we find the optimal forms of the factorized approximating distributions and arrive at the same formulas as in the uniform blind deblurring of Miskin & MacKay [3] and Fergus* et al. *[2]. Next, we derive the parameter update equations, which differ significantly from [3, 2].*

1. Summary

We derive here the optimal forms and parameters of the approximating distributions $q(\mathbf{f})$, $q(\mathbf{w})$ and $q(\beta_\sigma)$ used in the variational inference of the blind deblurring algorithms of Miskin & MacKay [3] and Fergus *et al*. [2]. Using the calculus of variations, we find that the optimal distributions for the latent variables are (the same as Equations (42, 43, 17) of [3]):

$$q(w_k) \propto p(w_k) \exp\left(-\frac{1}{2} w_k^{(2)} \left(w_k - w_k^{(1)}\right)^2\right) \tag{1}$$

$$q(f_j) \propto p(f_j) \exp\left(-\frac{1}{2} f_j^{(2)} \left(f_j - f_j^{(1)}\right)^2\right) \tag{2}$$

$$q(\beta_\sigma) = \Gamma\left(\beta_\sigma; \frac{1}{2} \sum_i \left\langle (g_i - \hat{g}_i(\mathbf{f}, \mathbf{w}))^2 \right\rangle_{q(\mathbf{f},\mathbf{w})}, \frac{N_g}{2}\right), \tag{3}$$

where $w_k^{(1)}$, $w_k^{(2)}$, $f_j^{(1)}$, $f_j^{(2)}$ are parameters of the distributions, $\hat{g}_i(\mathbf{f}, \mathbf{w})$ is the "reconstruction" of the i^{th} blurry pixel using the model in Equation (7) of Whyte *et al*. [4]. Note that $\mathbf{f}$ and $\mathbf{w}$ are random variables, so in this context $\hat{g}_i(\mathbf{f}, \mathbf{w})$ is also a random variable, which for simplicity we denote by $\hat{g}_i$ from now on. N_g is the number of observed blurry pixels, and $\langle \cdot \rangle_q$ represents the expectation with respect to the the distribution q. For each latent variable, the parameters of its distribution depend on the distributions of all the other latent variables, *e.g.* $w_k^{(1)}$ and $w_k^{(2)}$ depend on $q(\beta_\sigma)$, $q(f_j)$ for all j and $q(w_{k'})$ for all $k' \neq k$. For our non-uniform blur model, we find the following optimal values for the parameters (*c.f*. Equations (46–49) of [3]):

$$w_k^{(2)} = \langle \beta_\sigma \rangle \sum_i \left\langle \Big(\sum_j C_{ijk} f_j \Big)^2 \right\rangle_{q(\mathbf{f})} \tag{4}$$

$$w_k^{(1)} w_k^{(2)} = \langle \beta_\sigma \rangle \sum_i \Big(g_i \sum_j C_{ijk} \langle f_j \rangle_{q(f_j)} - \sum_{k' \neq k} \left\langle \Big(\sum_j C_{ijk} f_j \Big) \Big(\sum_j C_{ijk'} f_j \Big) \right\rangle_{q(\mathbf{f})} \langle w_{k'} \rangle_{q(w_{k'})} \Big) \tag{5}$$

$$f_j^{(2)} = \langle \beta_\sigma \rangle \sum_i \left\langle \Big(\sum_k C_{ijk} w_k \Big)^2 \right\rangle_{q(\mathbf{w})} \tag{6}$$

$$f_j^{(1)} f_j^{(2)} = \langle \beta_\sigma \rangle \sum_i \Big(g_i \sum_k C_{ijk} \langle w_k \rangle_{q(w_k)} - \sum_{j' \neq j} \langle f_{j'} \rangle_{q(f_{j'})} \left\langle \Big(\sum_k C_{ij'k} w_k \Big) \Big(\sum_k C_{ijk} w_k \Big) \right\rangle_{q(\mathbf{w})} \Big). \tag{7}$$

The details of the derivation are given next.

2. Variational method

For convenience, we will sometimes collect the latent variables $\mathbf{f}$, $\mathbf{w}$, and β_σ into the "ensemble" $\mathbf{\Theta}$. The aim is to approximate the true posterior $p(\mathbf{\Theta}|\mathbf{g})$ with a simpler factorized distribution $q(\mathbf{\Theta}|\mathbf{g})$, denoted for simplicity as $q(\mathbf{\Theta}) = q(\beta_\sigma)\prod_j q(f_j)\prod_k q(w_k)$. Our model, from Equation (7) in Whyte *et al.* [4], provides the likelihood $p(\mathbf{g}|\mathbf{\Theta})$:

$$g_i = \sum_{j,k} f_j C_{ijk} w_k + \varepsilon, \tag{8}$$

$$g_i = \mathbf{f}^\top C_i \mathbf{w} + \varepsilon \tag{9}$$

$$p(\mathbf{g}|\mathbf{\Theta}) = \prod_i \mathcal{G}\big(g_i; \hat{g}_i, \beta_\sigma^{-1}\big), \qquad \text{see Eqn. (7) of [3]} \tag{10}$$

where $\mathcal{G}(\,\cdot\,; \mu, \sigma^2)$ is a Gaussian with mean μ and variance σ^2, and for each blurry pixel g_i, C_i is a matrix of interpolation co-efficients with elements C_{ijk}. In order to get to the posterior, we also need a prior $p(\mathbf{\Theta})$ for our latent variables. The latent variables are assumed to be independent, so that the prior factorizes:

$$p(\mathbf{\Theta}) = p(\mathbf{f})p(\mathbf{w})p(\beta_\sigma), \tag{11}$$

and furthermore the elements of both $\mathbf{f}$ and $\mathbf{w}$ are assumed to be independent and identically-distributed, *i.e.*

$$p(\mathbf{f}) = \prod_j p(f_j) \tag{12}$$

$$p(\mathbf{w}) = \prod_k p(w_k). \tag{13}$$

From Eqn. (10) of [3], we wish to minimize the following cost function, first using the calculus of variations to find the optimal form of the approximate distributions, then iteratively optimizing their parameters, which is equivalent to minimizing the Kullback-Leibler (KL) divergence between the posterior and the approximating distribution (see [1, Eqn. 10.3, page 463]):

$$C_{\mathrm{KL}} = \int q(\mathbf{\Theta}) \left[\ln \frac{q(\mathbf{\Theta})}{p(\mathbf{\Theta})} - \ln p(\mathbf{g}|\mathbf{\Theta})\right] \mathrm{d}\mathbf{\Theta}. \tag{14}$$

3. Inside the Cost Function

Since $q(\mathbf{\Theta}) = q(\mathbf{f})q(\mathbf{w})q(\beta_\sigma)$ and $p(\mathbf{\Theta}) = p(\mathbf{f})p(\mathbf{w})p(\beta_\sigma)$,

$$C_{\mathrm{KL}} = \int q(\mathbf{\Theta}) \left[\ln \frac{q(\mathbf{f})}{p(\mathbf{f})} + \ln \frac{q(\mathbf{w})}{p(\mathbf{w})} + \ln \frac{q(\beta_\sigma)}{p(\beta_\sigma)} - \ln p(\mathbf{g}|\mathbf{\Theta})\right] \mathrm{d}\mathbf{\Theta}. \tag{15}$$

$$C_{\mathrm{KL}} = \int q(\mathbf{f}) \ln \frac{q(\mathbf{f})}{p(\mathbf{f})} \mathrm{d}\mathbf{f} + \int q(\mathbf{w}) \ln \frac{q(\mathbf{w})}{p(\mathbf{w})} \mathrm{d}\mathbf{w} + \int q(\beta_\sigma) \ln \frac{q(\beta_\sigma)}{p(\beta_\sigma)} \mathrm{d}\beta_\sigma - \int q(\mathbf{\Theta}) \ln p(\mathbf{g}|\mathbf{\Theta})\, \mathrm{d}\mathbf{\Theta}. \tag{16}$$

Similarly, since $q(\mathbf{f}) = \prod_j q(f_j)$, $q(\mathbf{w}) = \prod_k p(w_k)$, $p(\mathbf{f}) = \prod_j p(f_j)$, and $p(\mathbf{w}) = \prod_k p(w_k)$,

$$\begin{aligned} C_{\mathrm{KL}} = \sum_j \int q(f_j) \ln \frac{q(f_j)}{p(f_j)} \mathrm{d}f_j + \sum_k \int q(w_k) \ln \frac{q(w_k)}{p(w_k)} \mathrm{d}w_k + \int q(\beta_\sigma) \ln \frac{q(\beta_\sigma)}{p(\beta_\sigma)} \mathrm{d}\beta_\sigma \\ - \int q(\mathbf{\Theta}) \ln p(\mathbf{g}|\mathbf{\Theta})\, \mathrm{d}\mathbf{\Theta}. \end{aligned} \tag{17}$$

Finally, we expand the last term:

$$p(\mathbf{g}|\mathbf{\Theta}) = \prod_i \mathcal{G}(g_i; \hat{g}_i, \beta_\sigma^{-1}), \tag{18}$$

$$\ln p(\mathbf{g}|\mathbf{\Theta}) = \sum_i \ln \mathcal{G}(g_i; \hat{g}_i, \beta_\sigma^{-1}) \tag{19}$$

$$= \frac{1}{2} \sum_i \big(\ln \beta_\sigma - \beta_\sigma (g_i - \hat{g}_i)^2 - \ln 2\pi\big) \tag{20}$$

$$\int q(\mathbf{\Theta}) \ln p(\mathbf{g}|\mathbf{\Theta})\, \mathrm{d}\mathbf{\Theta} = \frac{1}{2} \int q(\mathbf{\Theta}) \sum_i \big(\ln \beta_\sigma - \beta_\sigma (g_i - \hat{g}_i)^2 - \ln 2\pi\big)\, \mathrm{d}\mathbf{\Theta} \tag{21}$$

$$= \frac{1}{2} \int q(\beta_\sigma) \sum_i \left(\ln \beta_\sigma - \beta_\sigma \int q(\mathbf{f}) q(\mathbf{w}) (g_i - \hat{g}_i)^2 \,\mathrm{d}\mathbf{f}\, \mathrm{d}\mathbf{w}\right) \mathrm{d}\beta_\sigma - \frac{1}{2} \sum_i \ln 2\pi. \tag{22}$$

Putting (22) into (17) and ignoring terms independent of $\mathbf{\Theta}$,

$$C_{\mathrm{KL}} = \sum_j \int q(f_j) \ln \frac{q(f_j)}{p(f_j)} \mathrm{d}f_j + \sum_k \int q(w_k) \ln \frac{q(w_k)}{p(w_k)} \mathrm{d}w_k + \int q(\beta_\sigma) \ln \frac{q(\beta_\sigma)}{p(\beta_\sigma)} \mathrm{d}\beta_\sigma$$
$$- \frac{1}{2} \int q(\beta_\sigma) \sum_i \left(\ln \beta_\sigma - \beta_\sigma \int q(\mathbf{f}) q(\mathbf{w}) (g_i - \hat{g}_i)^2 \,\mathrm{d}\mathbf{f}\, \mathrm{d}\mathbf{w}\right) \mathrm{d}\beta_\sigma \tag{23}$$

4. Optimal Distributions

4.1. Optimal $q(\beta_\sigma)$

To derive the optimal form of $q(\beta_\sigma)$, we ignore terms in C_{KL} independent of β_σ, add a Lagrange multiplier for the constraint that $\int q(\beta_\sigma)\, \mathrm{d}\beta_\sigma = 1$, and differentiate with respect to $q(\beta_\sigma)$:

$$C_{\mathrm{KL}}\big(q(\beta_\sigma)\big) = \int q(\beta_\sigma) \left[\ln \frac{q(\beta_\sigma)}{p(\beta_\sigma)} - \frac{1}{2} \sum_i \big(\ln \beta_\sigma - \beta_\sigma \langle (g_i - \hat{g}_i)^2 \rangle\big)\right] \mathrm{d}\beta_\sigma + \lambda_\sigma \left(\int q(\beta_\sigma)\, \mathrm{d}\beta_\sigma - 1\right), \tag{24}$$

where $\langle \cdot \rangle$ denotes the expectation under the approximating distribution $q(\mathbf{\Theta})$.

$$C_{\mathrm{KL}}\big(q(\beta_\sigma) + \delta q(\beta_\sigma)\big) = \int q(\beta_\sigma) \left[\ln \frac{q(\beta_\sigma) + \delta q(\beta_\sigma)}{p(\beta_\sigma)} - \frac{1}{2} \sum_i \big(\ln \beta_\sigma - \beta_\sigma \langle (g_i - \hat{g}_i)^2 \rangle\big)\right] \mathrm{d}\beta_\sigma$$
$$+ \int \delta q(\beta_\sigma) \left[\ln \frac{q(\beta_\sigma) + \delta q(\beta_\sigma)}{p(\beta_\sigma)} - \frac{1}{2} \sum_i \big(\ln \beta_\sigma - \beta_\sigma \langle (g_i - \hat{g}_i)^2 \rangle\big)\right] \mathrm{d}\beta_\sigma$$
$$+ \lambda_\sigma \left(\int q(\beta_\sigma)\, \mathrm{d}\beta_\sigma + \int \delta q(\beta_\sigma)\, \mathrm{d}\beta_\sigma - 1\right) \tag{25}$$

$$\ln\big(q(\beta_\sigma) + \delta q(\beta_\sigma)\big) \simeq \ln q(\beta_\sigma) + \frac{\delta q(\beta_\sigma)}{q(\beta_\sigma)} \qquad \text{to first order, so} \tag{26}$$

$$C_{\mathrm{KL}}\big(q(\beta_\sigma) + \delta q(\beta_\sigma)\big) = C_{\mathrm{KL}}\big(q(\beta_\sigma)\big) + \int \delta q(\beta_\sigma)\, \mathrm{d}\beta_\sigma$$
$$+ \int \delta q(\beta_\sigma) \left[\ln \frac{q(\beta_\sigma)}{p(\beta_\sigma)} + \frac{\delta q(\beta_\sigma)}{q(\beta_\sigma)} - \frac{1}{2} \sum_i \big(\ln \beta_\sigma - \beta_\sigma \langle (g_i - \hat{g}_i)^2 \rangle\big)\right] \mathrm{d}\beta_\sigma$$
$$+ \lambda_\sigma \int \delta q(\beta_\sigma)\, \mathrm{d}\beta_\sigma. \tag{27}$$

Discarding higher order terms in δq,

$$\delta C_{\mathrm{KL}} = \int \delta q(\beta_\sigma) \left[1 + \ln \frac{q(\beta_\sigma)}{p(\beta_\sigma)} - \frac{1}{2} \sum_i \left(\ln \beta_\sigma - \beta_\sigma \left\langle (g_i - \hat{g}_i)^2 \right\rangle \right) + \lambda_\sigma \right] \mathrm{d}\beta_\sigma \tag{28}$$

$$\frac{\partial C_{\mathrm{KL}}}{\partial q(\beta_\sigma)} = 1 + \ln \frac{q(\beta_\sigma)}{p(\beta_\sigma)} - \frac{1}{2} \sum_i \left(\ln \beta_\sigma - \beta_\sigma \left\langle (g_i - \hat{g}_i)^2 \right\rangle \right) + \lambda_\sigma. \tag{29}$$

Setting this derivative to zero, we obtain an relation similar to Equation (14) in [3]:

$$\ln q(\beta_\sigma) = \ln p(\beta_\sigma) + \frac{1}{2} \sum_i \left(\ln \beta_\sigma - \beta_\sigma \left\langle (g_i - \hat{g}_i)^2 \right\rangle \right) - 1 - \lambda_\sigma. \tag{30}$$

Thus the optimal distribution is

$$q(\beta_\sigma) \propto p(\beta_\sigma) \beta_\sigma^{\frac{N_g}{2}} \exp \left(-\frac{1}{2} \beta_\sigma \sum_i \left\langle (g_i - \hat{g}_i)^2 \right\rangle \right), \tag{31}$$

which, given that $p(\ln \beta_\sigma) = 1 \Rightarrow p(\beta_\sigma) = \Gamma(\beta_\sigma; \epsilon, \epsilon)$ with $\epsilon \to 0$ (Eqn. (8) of [3]), gives Eqn. (17) of [3]:

$$q(\beta_\sigma) = \Gamma \left(\beta_\sigma; \frac{1}{2} \sum_i \left\langle (g_i - \hat{g}_i)^2 \right\rangle, \frac{N_g}{2} \right), \tag{32}$$

where the Γ distribution is given by Eqn. (15) of [3]:

$$\Gamma(x; a, b) = \frac{1}{\Gamma(b)} a^b x^{(b-1)} \exp(-ax). \tag{33}$$

4.2. Optimal $q(f_j)$

Starting from Equation (23), and isolating the relevant terms,

$$C_{\mathrm{KL}}\big(q(f_j)\big) = \int q(f_j) \ln \frac{q(f_j)}{p(f_j)} \, \mathrm{d}f_j + \frac{1}{2} \langle \beta_\sigma \rangle \sum_i \int q(\mathbf{f}) q(\mathbf{w}) (g_i - \hat{g}_i)^2 \, \mathrm{d}\mathbf{f} \, \mathrm{d}\mathbf{w} + \lambda_j \left(\int q(f_j) \, \mathrm{d}f_j - 1 \right). \tag{34}$$

For convenience, we partition $\mathbf{f}$ into f_j, the pixel of interest, and $\mathbf{f}_{j^*}$, the remaining pixels:

$$C_{\mathrm{KL}}\big(q(f_j)\big) = \int q(f_j) \left[\ln \frac{q(f_j)}{p(f_j)} + \frac{1}{2} \langle \beta_\sigma \rangle \sum_i \int q(\mathbf{f}_{j^*}) q(\mathbf{w}) (g_i - \hat{g}_i)^2 \, \mathrm{d}\mathbf{f}_{j^*} \, \mathrm{d}\mathbf{w} \right] \mathrm{d}f_j + \lambda_j \left(\int q(f_j) \, \mathrm{d}f_j - 1 \right) \tag{35}$$

$$\begin{aligned} C_{\mathrm{KL}}\big(q(f_j) + \delta q(f_j)\big) &= \int q(f_j) \left[\ln \frac{q(f_j)}{p(f_j)} + \frac{\delta q(f_j)}{q(f_j)} + \frac{1}{2} \langle \beta_\sigma \rangle \sum_i \int q(\mathbf{f}_{j^*}) q(\mathbf{w}) (g_i - \hat{g}_i)^2 \, \mathrm{d}\mathbf{f}_{j^*} \, \mathrm{d}\mathbf{w} \right] \mathrm{d}f_j \\ &\quad + \int \delta q(f_j) \left[\ln \frac{q(f_j)}{p(f_j)} + \frac{\delta q(f_j)}{q(f_j)} + \frac{1}{2} \langle \beta_\sigma \rangle \sum_i \int q(\mathbf{f}_{j^*}) q(\mathbf{w}) (g_i - \hat{g}_i)^2 \, \mathrm{d}\mathbf{f}_{j^*} \, \mathrm{d}\mathbf{w} \right] \mathrm{d}f_j \\ &\quad + \lambda_j \left(\int q(f_j) \, \mathrm{d}f_j + \int \delta q(f_j) \, \mathrm{d}f_j - 1 \right) \\ &= C_{\mathrm{KL}}\big(q(f_j)\big) + (1 + \lambda_j) \int \delta q(f_j) \, \mathrm{d}f_j \end{aligned} \tag{36}$$

$$+\int \delta q(f_j)\left[\ln\frac{q(f_j)}{p(f_j)}+\frac{1}{2}\langle\beta_\sigma\rangle\sum_i\int q(\mathbf{f}_{j^*})q(\mathbf{w})(g_i-\hat{g}_i)^2\,\mathrm{d}\mathbf{f}_{j^*}\,\mathrm{d}\mathbf{w}\right]\mathrm{d}f_j \tag{37}$$

$$\frac{\partial C_{\mathrm{KL}}}{\partial q(f_j)}=\ln\frac{q(f_j)}{p(f_j)}+\frac{1}{2}\langle\beta_\sigma\rangle\sum_i\int q(\mathbf{f}_{j^*})q(\mathbf{w})(g_i-\hat{g}_i)^2\,\mathrm{d}\mathbf{f}_{j^*}\,\mathrm{d}\mathbf{w}+1+\lambda_j. \tag{38}$$

Setting this equal to zero, we obtain the optimal form

$$\ln q(f_j)=\ln p(f_j)-\frac{1}{2}\langle\beta_\sigma\rangle\sum_i\int q(\mathbf{f}_{j^*})q(\mathbf{w})(g_i-\hat{g}_i)^2\,\mathrm{d}\mathbf{f}_{j^*}\,\mathrm{d}\mathbf{w}-1-\lambda_j. \tag{39}$$

Here we need to make some simplifications to obtain a function of f_j.

$$(g_i-\hat{g}_i)^2=g_i^2-2g_i\hat{g}_i+\hat{g}_i^2 \tag{40}$$

$$\hat{g}_i=\mathbf{f}^\top C_i\mathbf{w}, \tag{41}$$

$$=\mathbf{f}_{j^*}^\top C_{ij^*}\mathbf{w}+f_j\mathbf{c}_{ij}^\top\mathbf{w} \tag{42}$$

where $\mathbf{c}_{ij}^\top$ is the j^{th} row of C_i, and C_{ij^*} is C_i with this row removed.

$$(g_i-\hat{g}_i)^2=g_i^2-2g_i(\mathbf{f}^\top C_i\mathbf{w})+(\mathbf{f}^\top C_i\mathbf{w})^2 \tag{43}$$

$$=g_i^2-2g_i(\mathbf{f}_{j^*}^\top C_{ij^*}\mathbf{w})-2g_i(f_j\mathbf{c}_{ij}^\top\mathbf{w})+(\mathbf{f}_{j^*}^\top C_{ij^*}\mathbf{w})^2+2(\mathbf{f}_{j^*}^\top C_{ij^*}\mathbf{w})(f_j\mathbf{c}_{ij}^\top\mathbf{w})+(f_j\mathbf{c}_{ij}^\top\mathbf{w})^2 \tag{44}$$

$$=-2g_i(f_j\mathbf{c}_{ij}^\top\mathbf{w})+2(\mathbf{f}_{j^*}^\top C_{ij^*}\mathbf{w})(f_j\mathbf{c}_{ij}^\top\mathbf{w})+(f_j\mathbf{c}_{ij}^\top\mathbf{w})^2+\text{const.} \tag{45}$$

$$\left\langle(g_i-\hat{g}_i)^2\right\rangle_{q(\mathbf{f}_{j^*},\mathbf{w})}=\int q(\mathbf{f}_{j^*})q(\mathbf{w})(g_i-\hat{g}_i)^2\,\mathrm{d}\mathbf{f}_{j^*}\,\mathrm{d}\mathbf{w} \tag{46}$$

$$=\left\langle-2g_i(f_j\mathbf{c}_{ij}^\top\mathbf{w})+2(\mathbf{f}_{j^*}^\top C_{ij^*}\mathbf{w})(f_j\mathbf{c}_{ij}^\top\mathbf{w})+(f_j\mathbf{c}_{ij}^\top\mathbf{w})^2+\text{const.}\right\rangle_{q(\mathbf{f}_{j^*},\mathbf{w})} \tag{47}$$

$$=-2f_j\left(g_i\mathbf{c}_{ij}^\top\langle\mathbf{w}\rangle_{q(\mathbf{w})}-\langle\mathbf{f}_{j^*}\rangle_{q(\mathbf{f}_{j^*})}^\top\left\langle(C_{ij^*}\mathbf{w})(\mathbf{c}_{ij}^\top\mathbf{w})\right\rangle_{q(\mathbf{w})}\right)+f_j^2\left\langle(\mathbf{c}_{ij}^\top\mathbf{w})^2\right\rangle_{q(\mathbf{w})}+\text{const.} \tag{48}$$

$$\sum_i\left\langle(g_i-\hat{g}_i)^2\right\rangle_{q(\mathbf{f}_{j^*},\mathbf{w})}=-2f_j\sum_i\left(g_i\mathbf{c}_{ij}^\top\langle\mathbf{w}\rangle_{q(\mathbf{w})}-\langle\mathbf{f}_{j^*}\rangle_{q(\mathbf{f}_{j^*})}^\top\left\langle(C_{ij^*}\mathbf{w})(\mathbf{c}_{ij}^\top\mathbf{w})\right\rangle_{q(\mathbf{w})}\right)+f_j^2\sum_i\left\langle(\mathbf{c}_{ij}^\top\mathbf{w})^2\right\rangle_{q(\mathbf{w})}+\text{const.} \tag{49}$$

which is just a quadratic in f_j. Replacing the coefficients with a_j and b_j,

$$\sum_i\left\langle(g_i-\hat{g}_i)^2\right\rangle_{q(\mathbf{f}_{j^*},\mathbf{w})}=a_jf_j^2-b_jf_j+\text{const.} \tag{50}$$

$$=a_j\left(f_j-\frac{b_j}{2a_j}\right)^2+\text{const.} \tag{51}$$

$$\ln q(f_j)=\ln p(f_j)-\frac{1}{2}\langle\beta_\sigma\rangle a_j\left(f_j-\frac{b_j}{2a_j}\right)^2+\text{const.} \tag{52}$$

$$q(f_j)\propto p(f_j)\exp\left(-\frac{1}{2}\langle\beta_\sigma\rangle a_j\left(f_j-\frac{b_j}{2a_j}\right)^2\right) \tag{53}$$

$$\propto p(f_j)\exp\left(-\frac{1}{2}f_j^{(2)}\left(f_j-f_j^{(1)}\right)^2\right)\qquad \textit{c.f.}\text{ Eqn. (43) of [3]} \tag{54}$$

where

$$
\begin{aligned}
f_j^{(2)} &= \langle\beta_\sigma\rangle a_j && (55)\\
&= \langle\beta_\sigma\rangle \sum_i \big\langle (\mathbf{c}_{ij}^\top \mathbf{w})^2 \big\rangle_{q(\mathbf{w})} \qquad \textit{c.f.}\text{ Eqn. (48) of [3]} && (56)\\
&= \langle\beta_\sigma\rangle \sum_i \Big\langle \Big(\sum_k C_{ijk} w_k\Big)^2 \Big\rangle_{q(\mathbf{w})} && (57)\\
f_j^{(1)} f_j^{(2)} &= \frac{1}{2}\langle\beta_\sigma\rangle b_j && (58)\\
&= \langle\beta_\sigma\rangle \sum_i \Big(g_i \mathbf{c}_{ij}^\top \langle\mathbf{w}\rangle_{q(\mathbf{w})} - \langle\mathbf{f}_{j^*}\rangle_{q(\mathbf{f}_{j^*})}^\top \big\langle (C_{ij^*}\mathbf{w})(\mathbf{c}_{ij}^\top\mathbf{w}) \big\rangle_{q(\mathbf{w})} \Big) \quad \textit{c.f.}\text{ Eqn. (49) of [3]} && (59)\\
&= \langle\beta_\sigma\rangle \sum_i \Big(g_i \sum_k C_{ijk}\langle w_k\rangle_{q(w_k)} - \sum_{j'\neq j} \langle f_{j'}\rangle_{q(f_{j'})} \Big\langle \Big(\sum_k C_{ij'k} w_k\Big)\Big(\sum_k C_{ijk} w_k\Big) \Big\rangle_{q(\mathbf{w})} \Big). && (60)
\end{aligned}
$$

4.3. Optimal $q(w_k)$

We proceed much the same as for $q(f_j)$, starting from Equation (23), and isolating the relevant terms,

$$
\begin{aligned}
C_{\mathrm{KL}}\big(q(w_k)\big) = &\int q(w_k) \ln \frac{q(w_k)}{p(w_k)}\, \mathrm{d}w_k + \frac{1}{2}\langle\beta_\sigma\rangle \sum_i \int q(\mathbf{f}) q(\mathbf{w}) (g_i - \hat{g}_i)^2\, \mathrm{d}\mathbf{f}\, \mathrm{d}\mathbf{w}\\
&+ \lambda_k \Big(\int q(w_k)\, \mathrm{d}w_k - 1 \Big). \qquad (61)
\end{aligned}
$$

similarly, we partition $\mathbf{w}$ into w_k, the element of interest, and $\mathbf{w}_{k^*}$, the remaining elements:

$$
\begin{aligned}
C_{\mathrm{KL}}\big(q(w_k)\big) = &\int q(w_k) \left[\ln \frac{q(w_k)}{p(w_k)} + \frac{1}{2}\langle\beta_\sigma\rangle \sum_i \int q(\mathbf{f}) q(\mathbf{w}_{k^*}) (g_i - \hat{g}_i)^2\, \mathrm{d}\mathbf{f}\, \mathrm{d}\mathbf{w}_{k^*} \right] \mathrm{d}w_k\\
&+ \lambda_k \Big(\int q(w_k)\, \mathrm{d}w_k - 1 \Big) \qquad (62)
\end{aligned}
$$

and obtain the optimal form

$$
\ln q(w_k) = \ln p(w_k) - \frac{1}{2}\langle\beta_\sigma\rangle \sum_i \int q(\mathbf{f}) q(\mathbf{w}_{k^*}) (g_i - \hat{g}_i)^2\, \mathrm{d}\mathbf{f}\, \mathrm{d}\mathbf{w}_{k^*} - 1 - \lambda_k. \qquad (63)
$$

Here we need to make some simplifications to obtain a function of w_k.

$$
\begin{aligned}
\hat{g}_i &= \mathbf{f}^\top C_i \mathbf{w}, && (64)\\
&= \mathbf{f}^\top \mathbf{c}_{ik} w_k + \mathbf{f}^\top C_{ik^*} \mathbf{w}_{k^*} && (65)
\end{aligned}
$$

where $\mathbf{c}_{ik}$ is the k^{th} column of C_i, and C_{ik^*} is C_i with this column removed.

$$
\begin{aligned}
(g_i - \hat{g}_i)^2 &= g_i^2 - 2g_i(\mathbf{f}^\top C_i \mathbf{w}) + (\mathbf{f}^\top C_i \mathbf{w})^2 && (66)\\
&= g_i^2 - 2g_i(\mathbf{f}^\top \mathbf{c}_{ik} w_k) - 2g_i(\mathbf{f}^\top C_{ik^*}\mathbf{w}_{k^*}) + (\mathbf{f}^\top \mathbf{c}_{ik} w_k)^2\\
&\quad + 2(\mathbf{f}^\top \mathbf{c}_{ik} w_k)(\mathbf{f}^\top C_{ik^*}\mathbf{w}_{k^*}) + (\mathbf{f}^\top C_{ik^*}\mathbf{w}_{k^*})^2 && (67)\\
&= -2g_i(\mathbf{f}^\top \mathbf{c}_{ik} w_k) + 2(\mathbf{f}^\top \mathbf{c}_{ik} w_k)(\mathbf{f}^\top C_{ik^*}\mathbf{w}_{k^*}) + (\mathbf{f}^\top \mathbf{c}_{ik} w_k)^2 + \text{const.} && (68)\\
\sum_i \big\langle (g_i - \hat{g}_i)^2 \big\rangle_{q(\mathbf{f}, \mathbf{w}_{k^*})} &= -2w_k \sum_i \Big(g_i \langle\mathbf{f}\rangle_{q(\mathbf{f})}^\top \mathbf{c}_{ik} - \big\langle (\mathbf{f}^\top \mathbf{c}_{ik})(\mathbf{f}^\top C_{ik^*}) \big\rangle_{q(\mathbf{f})} \langle\mathbf{w}_{k^*}\rangle_{q(\mathbf{w}_{k^*})} \Big)
\end{aligned}
$$

$$+ w_k^2 \sum_i \left\langle (\mathbf{f}^\top \mathbf{c}_{ik})^2 \right\rangle_{q(\mathbf{f})} + \text{const.} \tag{69}$$

$$q(w_k) \propto p(w_k) \exp\left(-\frac{1}{2} w_k^{(2)} \left(w_k - w_k^{(1)}\right)^2\right) \qquad \textit{c.f.}\text{ Eqn. (42) of [3]} \tag{70}$$

where

$$w_k^{(2)} = \langle \beta_\sigma \rangle \sum_i \left\langle (\mathbf{f}^\top \mathbf{c}_{ik})^2 \right\rangle_{q(\mathbf{f})} \qquad \textit{c.f.}\text{ Eqn. (46) of [3]} \tag{71}$$

$$= \langle \beta_\sigma \rangle \sum_i \left\langle \left(\sum_j C_{ijk} f_j \right)^2 \right\rangle_{q(\mathbf{f})} \tag{72}$$

$$w_k^{(1)} w_k^{(2)} = \langle \beta_\sigma \rangle \sum_i \left(g_i \langle \mathbf{f} \rangle_{q(\mathbf{f})}^\top \mathbf{c}_{ik} - \left\langle (\mathbf{f}^\top \mathbf{c}_{ik})(\mathbf{f}^\top C_{ik^*}) \right\rangle_{q(\mathbf{f})} \left\langle \mathbf{w}_{k^*} \right\rangle_{q(\mathbf{w}_{k^*})} \right) \qquad \textit{c.f.}\text{ Eqn. (47) of [3].} \tag{73}$$

$$= \langle \beta_\sigma \rangle \sum_i \left(g_i \sum_j C_{ijk} \langle f_j \rangle_{q(f_j)} - \sum_{k' \neq k} \left\langle \left(\sum_j C_{ijk} f_j \right) \left(\sum_j C_{ijk'} f_j \right) \right\rangle_{q(\mathbf{f})} \langle w_{k'} \rangle_{q(w_{k'})} \right) \tag{74}$$

References

[1] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 2006. 2

[2] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman. Removing camera shake from a single photograph. *SIGGRAPH*, 2006. 1

[3] J. W. Miskin and D. J. C. MacKay. Ensemble learning for blind image separation and deconvolution. In *Advances in Independent Component Analysis*. Springer-Verlag, 2000. 1, 2, 4, 5, 6, 7

[4] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce. Non-uniform deblurring for shaken images. In *CVPR*, 2010. 1, 2