IRWIN AND JOAN JACOBS
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES

Minimization of multiplexed unmixing error using gradient descent

Marina Alterman and
Yoav Y. Schechner

CCIT Report #760
February 2010

Electronics
Computers
Communications

DEPARTMENT OF ELECTRICAL ENGINEERING
TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 32000, ISRAEL

CCIT REPORT #760 February 2010

Minimization of multiplexed unmixing error using gradient descent

Marina Alterman , Yoav Y. Schechner
Department of Electrical Engineering
Technion - Israel Inst. Technology,
Haifa 32000, Israel
amarina@tx.technion.ac.il, yoav@ee.technion.ac.il

Ref. [1] presents significant generalizations of the theory of multiplexing. Instead of optimally recovering intensity arrays (which result from a mixture of underlying materials), the optimal recovery of underlying materials is defined as the goal of multiplexed acquisition. As detailed in [1], the mixing/unmixing process is directly incorporated into the optimization of multiplexing codes. This leads to the definition of constrained minimization problem where multiplexing code that yields optimal unmixing in the sense of minimal MSE is sought. The notation here is similar to Ref. [1].

The recovery of the concentrations $\mathbf{c}$ is based on WLS. Thus from [1],

$$\mathrm{MSE}_{\mathbf{c}} = \frac{1}{N_{\mathrm{dyes}}}\mathtt{tr}\left[\left(\mathbf{W}_{\mathrm{x}}{}^{T}\mathbf{\Sigma}_{\mathrm{noise}}^{-1}\mathbf{W}_{\mathrm{x}}\right)^{-1}\right], \tag{1}$$

where $\mathbf{W}_{\mathrm{x}} = \mathbf{W}\mathbf{X}$. Here $\mathbf{X}$ is the mixing matrix. According to [1] we seek to solve

$$\widehat{\mathbf{W}}_{\mathbf{c}} = \arg\min_{\mathbf{W}} \mathrm{MSE}_{\mathbf{c}}, \qquad \text{s.t.} \quad 0 \leq w_{m,s} \leq 1\ . \tag{2}$$

To achieve this, we use the projected gradient descent method [2]: $\mathrm{MSE}_{\mathbf{c}}$ is iteratively minimized as a function of $\mathbf{W}$ in analogy to Ref. [4]. In each iteration k, $\mathbf{W}$ is updated by its gradient $\frac{\partial \mathrm{MSE}_{\mathbf{c}}}{\partial \mathbf{W}}$:

$$\mathbf{W}_{k+1} = \mathbf{W}_{k} - \gamma\frac{\partial \mathrm{MSE}_{\mathbf{c}}}{\partial \mathbf{W}_{k}}\ , \tag{3}$$

where γ is a parameter controlling the step size. This report derives the gradient of $\mathrm{MSE}_{\mathbf{c}}$, with respect to the multiplexing matrix $\mathbf{W}$. The matrix $\mathbf{W}_{k+1}$ is then projected onto the constraint

$$0 \leq w_{m,s} \leq 1\ \ , \tag{4}$$

as a generalization to Ref. [4].

To facilitate Eq. (3), we differentiate $\mathrm{MSE}_{\mathbf{c}}$ with respect to $\mathbf{W}$. To simplify the calculations, define an auxiliary matrix:

$$\mathbf{Z} = \mathbf{W}_{\mathrm{x}}^{T}\mathbf{\Sigma}_{\mathrm{noise}}\mathbf{W}_{\mathrm{x}}. \tag{5}$$

Substituting $\mathbf{Z}$ into Eq. (1), yields

$$\mathrm{MSE}_{\mathbf{c}} = \frac{1}{N_{\mathrm{dyes}}}\mathtt{tr}\left(\mathbf{Z}^{-1}\right)\ \ . \tag{6}$$

The gradient of $\mathrm{MSE}_{\mathbf{c}}$, with respect to $\mathbf{Z}$ [3] is

$$\frac{\partial \mathrm{MSE}_{\mathbf{c}}}{\partial \mathbf{Z}} = -\frac{1}{N_{\mathrm{dyes}}}\left[\left(\mathbf{Z}^{-1}\right)^{2}\right]^{T}\ \ . \tag{7}$$

We use the following chain rule [3] in order to calculate the partial derivatives:

$$\frac{\partial \mathrm{MSE}_{\mathbf{c}}}{\partial w_{m,s}} = \sum_{p=1}^{N_{\mathrm{dyes}}}\sum_{d=1}^{N_{\mathrm{dyes}}}\frac{\partial \mathrm{MSE}_{\mathbf{c}}}{\partial z_{p,d}}\cdot\frac{\partial z_{p,d}}{\partial w_{m,s}}\ \ , \tag{8}$$

where $\frac{\partial \text{MSE}_\mathbf{c}}{\partial z_{p,d}}$ and $z_{p,d}$ are elements in $\frac{\partial \text{MSE}_\mathbf{c}}{\partial \mathbf{Z}}$ and $\mathbf{Z}$ respectively.

Next, we explain the computation of the derivatives in Eq. (8). Since $\mathbf{Z}$ is a function of both $\mathbf{W}_\text{x}$ and $\mathbf{\Sigma}_\text{noise}$ (Eq. 5), we use the chain rule [3] again in order to calculate their partial derivatives. Recall the noise covariance matrix $\mathbf{\Sigma}_\text{noise}$. It is diagonal. Thus, its elements are

$$\sigma_{m,q}^2 = \delta_{m,q}\sigma_m^2 \quad , \tag{9}$$

where $\delta_{m,q}$ is Kronecker's delta. Therefore,

$$\frac{\partial z_{p,d}}{\partial w_{m,s}} = \sum_{q=1}^{N_\text{measure}} \sum_{r=1}^{N_\text{dyes}} \frac{\partial z_{p,d}}{\partial w_{q,r}^\text{x}} \frac{\partial w_{q,r}^\text{x}}{\partial w_{m,s}} + \sum_{m'=1}^{N_\text{measure}} \frac{\partial z_{p,d}}{\partial \sigma_{m'}^2} \frac{\partial \sigma_{m'}^2}{\partial w_{m,s}} . \tag{10}$$

According to [1],

$$\sigma_m^2 \approx \sigma_\text{gray}^2 + \sum_{d=1}^{N_\text{dyes}} w_{m,d}^\text{x} \nu_d^2 \, . \tag{11}$$

Based on Eq. (11), the elements of $\mathbf{\Sigma}_\text{noise}$ depend on elements of $\mathbf{W}_\text{x}$. Therefore, the partial derivatives of $\sigma_{m'}^2$ in Eq. (10) can also be computed using the chain rule:

$$\frac{\partial \sigma_{m'}^2}{\partial w_{m,s}} = \sum_{k=1}^{N_\text{measure}} \sum_{l=1}^{N_\text{dyes}} \frac{\partial \sigma_{m'}^2}{\partial w_{k,l}^\text{x}} \cdot \frac{\partial w_{k,l}^\text{x}}{\partial w_{m,s}} \quad . \tag{12}$$

Based on Eq. (5),

$$\frac{\partial z_{p,d}}{\partial w_{q,r}^\text{x}} = \frac{\delta_{d,r} w_{q,p}^\text{x}}{\sigma_q^2} + \frac{\delta_{p,r} w_{q,d}^\text{x}}{\sigma_q^2} \, , \tag{13}$$

$$\frac{\partial z_{p,d}}{\partial \sigma_{m'}^2} = -\frac{w_{m',p}^\text{x} w_{m',d}^\text{x}}{(\sigma_{m'}^2)^2} \, . \tag{14}$$

From Eq. (11)

$$\frac{\partial \sigma_{m'}^2}{\partial w_{k,l}^\text{x}} = \delta_{m',k} \, \nu_l^2 \quad . \tag{15}$$

Finally, recall from [1] that $\mathbf{W}_\text{x} = \mathbf{W}\mathbf{X}$. Then,

$$\frac{\partial w_{q,r}^\text{x}}{\partial w_{m,s}} = \delta_{q,m} x_{s,r} \quad . \tag{16}$$

Substituting Eqs. (15,16) into Eq. (12) we get,

$$\frac{\partial \sigma_{m'}^2}{\partial w_{m,s}} = \delta_{m',m} \sum_{l=1}^{N_\text{dyes}} \nu_l^2 x_{s,l} \quad . \tag{17}$$

Substituting Eqs. (13,14,16,17) into Eq. (10), yields

$$\frac{\partial z_{p,d}}{\partial w_{m,s}} = \frac{1}{\sigma_m^2}(w_{m,p}^\text{x} \sigma_q^2 x_{s,d} + w_{m,d}^\text{x} \sigma_q^2 x_{s,p}) - \frac{1}{(\sigma_m^2)^2} w_{m,p}^\text{x} w_{m,d}^\text{x} \sum_{l=1}^{N_\text{dyes}} \nu_l^2 x_{s,l}. \tag{18}$$

The partial derivatives that were computed in Eq. (18) are substituted into Eq. (8), yielding

$$\frac{\partial \text{MSE}_\mathbf{c}}{\partial w_{m,s}} = \sum_{p=1}^{N_\text{dyes}} \sum_{d=1}^{N_\text{dyes}} \frac{\partial \text{MSE}_\mathbf{c}}{\partial z_{p,d}} \cdot \left[\frac{1}{\sigma_m^2} \left(w_{m,p}^\text{x} \sigma_q^2 x_{s,d} + + w_{m,d}^\text{x} \sigma_q^2 x_{s,p} \right) - \frac{1}{(\sigma_m^2)^2} w_{m,p}^\text{x} w_{m,d}^\text{x} \sum_{l=1}^{N_\text{dyes}} \nu_l^2 x_{s,l} \right]. \tag{19}$$

This is the gradient of $\text{MSE}_\mathbf{c}$ with respect to the multiplexing matrix $\mathbf{W}$.

Recall that in [1] we made the following approximation

$$\sigma_m^2 \approx \sigma_\text{gray}^2 + \nu^2 \sum_{d=1}^{N_\text{dyes}} w_{m,d}^\text{x} \quad . \tag{20}$$

Substituting the approximation in Eq. (20) into Eq. (19) yields

$$\frac{\partial \mathrm{MSE}_{\mathbf{c}}}{\partial w_{m,s}} \approx \frac{\sum_{p=1}^{N_{\mathrm{dyes}}} \sum_{d=1}^{N_{\mathrm{dyes}}} \frac{\partial \mathrm{MSE}_{\mathbf{c}}}{\partial z_{p,d}} (w^{\mathrm{x}}_{m,p} x_{s,d} + w^{\mathrm{x}}_{m,d} x_{s,p})}{\sigma^2_{\mathrm{gray}} + \nu^2 \sum_{l=1}^{N_{\mathrm{dyes}}} w^{\mathrm{x}}_{m,l}} - \frac{\nu^2 \sum_{l=1}^{N_{\mathrm{dyes}}} x_{s,l} \sum_{p=1}^{N_{\mathrm{dyes}}} \sum_{d=1}^{N_{\mathrm{dyes}}} \frac{\partial \mathrm{MSE}_{\mathbf{c}}}{\partial z_{p,d}} w^{\mathrm{x}}_{m,p} w^{\mathrm{x}}_{m,d}}{(\sigma^2_{\mathrm{gray}} + \nu^2 \sum_{l=1}^{N_{\mathrm{dyes}}} x_{s,l})^2} \,. \tag{21}$$

Define a row vector

$$\mathbf{1} = (1 \;\; 1 \;\; 1 \;\; 1 \;\; \cdots \;\; 1) \tag{22}$$

of length N_{dyes}. Now we can rewrite Eq. (21) using matrix form as

$$\frac{\partial \mathrm{MSE}_{\mathbf{c}}}{\partial \mathbf{W}} \approx \mathbf{\Sigma}^{-1}_{\mathrm{noise}} \left[\mathbf{W}_{\mathrm{x}} \frac{\partial \mathrm{MSE}_{\mathbf{c}}}{\partial \mathbf{Z}} \mathbf{X}^T + \mathbf{W}_{\mathrm{x}} \left(\frac{\partial \mathrm{MSE}_{\mathbf{c}}}{\partial \mathbf{Z}}\right)^T \mathbf{X}^T \right] - \nu^2 \mathtt{diag}\left(\mathbf{W}_{\mathrm{x}} \frac{\partial \mathrm{MSE}_{\mathbf{c}}}{\partial \mathbf{Z}} \mathbf{W}^T_{\mathrm{x}} \right) \cdot \mathbf{1} \cdot \mathbf{X}^T \cdot (\mathbf{\Sigma}^{-1}_{\mathrm{noise}})^2 \,. \tag{23}$$

Here `diag` returns a column vector formed by the elements of the main diagonal and $\mathbf{\Sigma}_{\mathrm{noise}}$ is approximated by Eq (20). Consequently we use Eq. (23) to express the gradient of $\mathrm{MSE}_{\mathbf{c}}$ with respect to the multiplexing matrix $\mathbf{W}$.

References

[1] M. Alterman, Y. Y. Schechner and A. Weiss. Multiplexed fluorescence unmixing. *IEEE ICCP*, 2010.

[2] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1999.

[3] K. B. Petersen and M. S. Pedersen. The matrix cookbook. http://matrixcookbook.com/ , 2008.

[4] N. Ratner and Y. Y. Schechner. Illumination multiplexing within fundamental limits. In *Proc. IEEE CVPR*, 2007.