Update Rules for a Weighted Non-negative FH*G Factorization

Pieter Peers
Philip Dutré

Report CW440, April 2006

Katholieke Universiteit Leuven
Department of Computer Science
Celestijnenlaan 200A – B-3001 Heverlee (Belgium)

Update Rules for a Weighted Non-negative FH*G Factorization

Pieter Peers
Philip Dutré

Report CW440, April 2006

Department of Computer Science, K.U.Leuven

Abstract

In this report the derivation and prove of convergence of a weighted non-negative factorization of the form FH*G is discussed. The derivation and proof is based on Lee and Seung's original derivation of Non-negative Matrix Factorization . This form is particularly suited to factor matrices exhibiting a band-diagonal structure with horizontal and vertical discontinuities.

Keywords : Non-negative Matrix Factorization.

Update Rules for a Weighted Non-negative FH*G Factorization

Pieter Peers and Philip Dutré

1 Introduction

This document is intended to support [2]: "A Compact Factored Representation for Heterogeneous Subsurface Scattering". However, the derived factorization can be used in other more general contexts.

More specifally, in this technical report we dicuss the derivation and proof of the convergence of the following non-negative factorization form:

$$R_d = FH \star G.$$

The notation $\star$ indicates a component-wise matrix multiplication:

$$\begin{aligned} c_{ij} &= (A \star B)_{ij} \\ &= a_{ij} \star a_{ij}. \end{aligned}$$

We will follow the proof of Lee and Seung [1] closely, and make necessary adjustments where needed. This particular factorization form is well suited for compactly representing band-diagonal matrices with horizontal, and vertical structured discontinuities. Furthermore we will include a weighting matrix W that can be used to place more or less confidence in particular elements of the original matrix R_d.

2 Update Rules for the Factorization

Using the following update rules ensures a non-negative factorization of R_d:

$$h_{a\mu} \leftarrow h_{a\mu} \frac{(F^T(W^2 \star G \star R_d))_{a\mu}}{(F^T(W^2 \star G \star (FH \star G)))_{a\mu}}, \quad (1)$$

$$f_{ia} \leftarrow f_{ia} \frac{((W^2 \star G \star R_d)H^T)_{ia}}{(W^2 \star G \star (FH \star G))H^T)_{ia}}. \quad (2)$$

3 Proof of Convergence

We follow a similar approach as in [1] to prove the convergence of these update rules. We would like to prove that the update rules (1) and (2) minimize the following cost function:

$$\mathcal{F}(h) = \frac{1}{2}\sum_i (w_i(r_i - \sum_a (F_{ia}h_a g_i)))^2, \tag{3}$$

where r, g, w and h are corresponding columns in respectively R_d, G, W and H. As in [1] we will define an auxiliary function $\mathcal{G}(h, h^t)$ which adheres to the following:

$$\mathcal{G}(h, h) = \mathcal{F}(h), \tag{4}$$
$$\mathcal{G}(h, h^t) \geq \mathcal{F}(h). \tag{5}$$

In this case we can update h by:

$$h^{t+1} = \arg\min_h \mathcal{G}(h, h^t). \tag{6}$$

If $t \to \infty$, then $h^t \to h$, thus it is sufficient to prove that $\mathcal{G}(h, h^t)$ is an auxiliary function for the cost function (3) in order to prove convergence. $\mathcal{G}$ can be freely chosen, as long as the conditions (4), and (5) are met. Keeping the update rules (1), and (2) in mind, we define $\mathcal{G}(h, h^t)$ by:

$$\mathcal{G}(h, h^t) = \mathcal{F}(h^t) - (h - h^t)^T \nabla\mathcal{F}(h^t) + \frac{1}{2}(h - h^t)^T K(h^t)(h - h^t), \tag{7}$$

where K is a diagonal matrix defined by:

$$K_{ab}(h^t) = \frac{\Delta_{ab}(F^T(Fh^t \star g^2 \star w^2))_a}{h_a^t}. \tag{8}$$

It is obvious that if $h = h^t$ in formula (7) condition (4) is met. Thus, we only need to prove the second condition (5) in order to prove that $\mathcal{G}(h, h^t)$ is a valid auxiliary function.

First rewrite $\mathcal{F}(h)$ using a Taylor expansion. Note that $\mathcal{F}(h)$ is a second degree polynomial, and thus the Taylor expansion only has 3 terms:

$$\mathcal{F}(h) = \mathcal{F}(h^t) + (h - h^t)^T \nabla\mathcal{F}(h^t) + \frac{1}{2}(h - h^t)^T \Delta\mathcal{F}(h^t)(h - h^t). \tag{9}$$

Note that:

$$
\begin{aligned}
\nabla\mathcal{F}(h) &= \nabla(\frac{1}{2}\sum_i(w_i(r_i-\sum_a(F_{ia}h_ag_i)))^2) \\
&= \nabla(\frac{1}{2}\sum_i(w_i^2r_i^2-2w^2s_ig_i\sum_a(F_iah_a)+(w_i^2\sum_a(F_iah_ag_i))^2)) \\
&= -\sum_i w_i^2r_ig_iF_i+\frac{1}{2}\sum_i w_i^2g_i^2\nabla(\sum_{ab}F_iaF_ibh_ah_b) \\
&= -F^T(s\star g\star w^2)+(F^T(Fh\star g^2\star w^2)),
\end{aligned}
$$

and

$$
\begin{aligned}
\Delta\mathcal{F}(h) &= \sum_i w_i^2g_i^2\sum_{ab}F_iaF_ib \\
&= F^T(F\star g^2\star w^2).
\end{aligned}
$$

Subtracting the Taylor expansion (9) from the auxiliary function (7) results in an equivalent equation for the second condition condition (5):

$$
0\leq(h-h^t)^T[K(h^t)-\Delta\mathcal{F}(h^t)](h-h^t). \tag{10}
$$

This inequality can be proven by considering the following matrix $M(h^t)$:

$$
M_{ab}(h^t)=h_a^t(K(h^t)-\Delta\mathcal{F}(h^t))h_b^t.
$$

Then for any vector v we have:

$$
\begin{aligned}
v^TMv &= \sum_{ab}v_aM_{ab}v_b \\
&= \sum_{ab}v_av_a^t(\Delta_{ab}(F^T(F\star g^2\star w^2))_{ab}-F^T(F\star g^2\star w^2))_{ab}v_bh_b^t \\
&= \sum_{ab}(v_a^2h_a^th_b^t(F^T(F\star g^2\star w^2))_{ab}-v_ah_a^t(F^T(F\star g^2\star w^2))_{ab}h_a^tv_b) \\
&= \sum_{ab}(F^T(F\star g^2\star w^2))_{ab}h_a^th_b^t[v_a^2-v_av_b] \\
&= \sum_{ab}(F^T(F\star g^2\star w^2))_{ab}h_a^th_b^t[\frac{1}{2}v_a^2+\frac{1}{2}v_b^2-v_av_b] \\
&= \sum_{ab}(F^T(F\star g^2\star w^2))_{ab}h_a^th_b^t\frac{1}{2}(v_a-v_b)^2 \\
&\geq 0.
\end{aligned}
$$

Thus $v^T M v$ is non-negative for any v since F, g^2, w^2 and h are non-negative. To understand the step from the first to the second equation consider:

$$
\begin{aligned}
\sum_{ab} v_a h_a^t K(h^t) h_b^t v_b &= \sum_{ab} v_a h_a^t \frac{\Delta_{ab}(F^T(Fh^t \star g^2 \star w^2))_a}{h_a^t} h_b^t v_b \\
&= \sum_a v_a^2 h_a^t (F^T(Fh^t \star g^2 \star w^2))_a \\
&= \sum_a v_a^2 h_a^t (\sum_i F_{ai} (\sum_b F_{ib} h_b^t g_i^2 w_i^2))_a \\
&= \sum_a v_a^2 h_a^t (\sum_b h_b^t \sum_i F_{ia} F_{ib} g_i^2 w_i^2)_a \\
&= \sum_{ab} v_a^2 h_a^t h_b^t (\sum_i F_{ia} F_{ib} g_i^2 w_i^2)_a \\
&= \sum_{ab} v_a^2 h_a^t h_b^t (F^T(F \star g^2 \star w^2))_{ab}.
\end{aligned}
$$

Thus, since $v^T M v$ is non-negative for any v, we have proven (10), from which follows that $\mathcal{G}(h, h^t)$ is an auxiliary function for $\mathcal{F}$.

4 Derivation of the Update Rules

We can now derive the multiplicative update rules by computing:

$$h^t = \arg\min_h \mathcal{G}(h, h^t),$$

thus, by taking the derivative of $\mathcal{G}$ in terms of h, and setting this equal to 0:

$$0 = \frac{\nabla \mathcal{G}(h, h^t)}{\nabla h^{t+1}} = \nabla \mathcal{F}(h^t) + (h^{t+1} - h^t) K(h^t).$$

Rewrite this as:

$$
\begin{aligned}
-\nabla \mathcal{F}(h^t) &= (h^{t+1} - h^t) K(h^t) \\
h^{t+1} &= h^1 - \nabla \mathcal{F}(h^t) K^{-1}(h^t).
\end{aligned}
$$

and thus:

$$
\begin{aligned}
h_a^{t+1} &= h_a^t + \frac{F^T(r \star g \star w^2)_a - (F^T(Fh^t \star g^2 \star w^2))_a}{(F^T(Fh^t \star g^2 \star w^2))_a} h_a^t \\
&= h_a^t \frac{(F^T(r \star g \star w^2)_a}{(F^T(Fh^t \star g^2 \star w^2))_a}.
\end{aligned}
$$

Resulting in the update rule for H. A similar result can be obtained by interchanging F and H to obtain an update rule for F.

Note, that the update rules guarantee convergence, but not necessary to a global minimum. This is a well-known problem with the multiplicative update rules for non-negative matrix factorization.

5 Conclusion

We have derived and proved the convergence of the update rules of a weighted matrix factorization $R_d = FH \star G$, by trivially extending the derivation and proof of the original non-negative matrix factorization paper of Lee and Seung [1].

References

[1] Daniel D. Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. In *NIPS*, pages 556–562, 2000.

[2] Pieter Peers, Karl vom Berge, Wojciech Matusik, Ravi Ramamoorthi, Jason Lawrence, Szymon Rusinkiewicz, and Philip Dutré. A comact factored representation of heterogeneous subsurface scattering. *ACM Transactions on Graphics*, 2006.