

# Robust HMF Notes: proof/sketch of why there *is* an objective

Tom Hilder

25/08/25

## Setup

Say we have measured data  $F_{ij}$  with known variances  $\sigma_{ij}^2$ . We model this with with lower rank matrices:

$$Y_{ij} = a_i^\top g_j, \quad (1)$$

where

$$Y : N \times M,$$

$$A : N \times K,$$

$$G : K \times M,$$

and  $K \leq \min(N, M)$  such that

$$Y \approx AG. \quad (2)$$

Assuming Gaussian (but independent) heteroskedastic noise gives the usual  $\chi^2$  objective:

$$\chi^2(A, G) = \sum_{ij} \frac{(Y_{ij} - a_i^\top g_j)^2}{\sigma_{ij}^2}. \quad (3)$$

This is bilinear in  $(A, G)$  so solvable with alternating least squares.

## a-step

$$A_i \leftarrow G_i F_i,$$

$$[G_i]_{kk'} = \sum_{j=1}^m \frac{g_{kj} g_{k'j}}{\sigma_{ij}^2},$$

$$[F_i]_k = \sum_{j=1}^m \frac{g_{kj} Y_{ij}}{\sigma_{ij}^2}.$$

**g-step**

$$\begin{aligned} G_j &\leftarrow A_j^\top F_j, \\ [A_j]_{kk'} &= \sum_{i=1}^N \frac{a_{ik} a_{ik'}}{\sigma_{ij}^2}, \\ [F_j]_k &= \sum_{i=1}^N \frac{a_{ik} Y_{ij}}{\sigma_{ij}^2}. \end{aligned}$$

This is weighted least squares (WLS) for  $a_i$  or  $g_j$  given fixed  $G$  or  $A$ .

## Adding Robustness

Outliers are not dealt with well by  $\chi^2$  objective because the quadratic loss penalty is too chonky. Instead, define

$$r_{ij} = \frac{Y_{ij} - a_i^\top g_j}{\sigma_{ij}}, \quad L(A, G) = \sum_{ij} \rho(r_{ij}). \quad (4)$$

If  $\rho(r) = \frac{1}{2}r^2$ , we recover the above. Switch to a Cauchy likelihood, taking the negative log likelihood as our loss function:

$$\rho(r) = \frac{c^2}{2} \log \left( 1 + \left( \frac{r}{c} \right)^2 \right). \quad (5)$$

The Cauchy distribution has large tails and so the loss penalty while quadratic for small  $r$ , tends to only  $\rho \propto \log r^2$  for large  $r$ . I think actually the nicest way to do this whole argument would be with student's  $t$  distribution, because the Cauchy and Gaussian distributions are special cases/limits of that. Anyway continuing with Cauchy.

## Auxiliary Form

Claim: the loss can be expressed for any  $r$  in the following way

$$\rho(r) = \min_{0 < w \leq 1} \left[ \frac{1}{2} w r^2 + \phi(w) \right], \quad (6)$$

where

$$\phi(w) = \frac{c^2}{2} (w - 1 - \log w). \quad (7)$$

Let's prove that. Define

$$J(w; r) = \frac{1}{2} w r^2 + \frac{Q^2}{2} (w - 1 - \log w) \quad (8)$$

Differentiating with respect to  $w$ :

$$\frac{\partial J}{\partial w} = \frac{1}{2} r^2 + \frac{c^2}{2} \left( 1 - \frac{1}{w} \right), \quad (9)$$

and

$$\frac{\partial^2 J}{\partial w^2} = \frac{Q^2}{2} \frac{1}{w^2} > 0, \quad \forall Q, w > 0. \quad (10)$$

thus we know that the critical point minimises  $J$ . Setting  $\partial_w J = 0$  yields

$$\hat{w}(r) = \operatorname{argmin}_w J(w; r) \quad (11)$$

$$= \frac{1}{1 + (r/Q)^2}, \quad (12)$$

and note that  $\hat{w} \in (0, 1]$  because  $(r/Q)^2 \geq 0$ . We now prove the claim. First set  $t = (r/c)^2 \geq 0$ , then

$$\hat{w} = \frac{1}{1+t}, \quad r^2 = Q^2 t, \quad (13)$$

such that

$$J(\hat{w}; r) = \frac{1}{2} w r^2 + \phi(\hat{w}) \quad (14)$$

$$\Rightarrow \frac{1}{2} w r^2 = \frac{Q^2}{2} \frac{t}{1+t} \quad (15)$$

$$\Rightarrow \phi(\hat{w}) = \frac{Q^2}{2} (\hat{w} - 1 - \log \hat{w}) \quad (16)$$

$$= \frac{Q^2}{2} \left( -\frac{t}{1+t} + \log(1+t) \right) \quad (17)$$

$$\Rightarrow J(\hat{w}; r) = \frac{Q^2}{2} \left( 1 + \left( \frac{r}{Q} \right)^2 \right). \quad (18)$$

Thus

$$\rho(r) = J(\hat{w}; r) \quad (19)$$

$$= \min_{0 < w \leq 1} \left[ \frac{1}{2} w r^2 + \phi(w) \right], \quad (20)$$

as claimed.

## Three-Step Algorithm

Define new objective:

$$J(A, G, W) = \frac{1}{2} \sum_{ij} [w_{ij} r_{ij}^2 + \phi(w_{ij})], \quad (21)$$

with  $r_{ij} = (Y_{ij} - a_i^\top g_j) / \sigma_{ij}$ . By construction,

$$L(A, G) = \min_W J(A, G, W), \quad (22)$$

and if

$$\hat{W} = \operatorname{argmin}_w J(A, G, W), \quad (23)$$

then

$$[\hat{W}]_{ij} = \hat{w}(r_{ij}) \quad (24)$$

$$= \frac{1}{1 + (r_{ij}/Q)^2}. \quad (25)$$

This immediately yield's Hogg's procedure

w-step:  $w_{ij} \leftarrow \hat{w}(r_{ij})$ ,  
a-step: solve WLS for  $A$  with new weights,  
g-step: solve WLS for  $G$  with new weights.

where the a-step optimises the quadratic

$$Q(A|G, W) = \frac{1}{2} \sum_{ij} w_{ij} r_{ij}^2, \quad (26)$$

and the g-step optimises  $Q(G|A, W)$ . It should be pretty apparent now that the procedure gives the MLE with a Cauchy likelihood.

## Extra convincing (showing that the procedure optimises $L$ )

Consider one outer cycle starting at  $(A^{(t)}, G^{(t)})$ . Choose  $W^{(t)} = \hat{w}(r(A^{(t)}, G^{(t)}))$ . Then

$$L(A^{(t)}, G^{(t)}) = J(A^{(t)}, G^{(t)}, W^{(t)}). \quad (27)$$

With frozen  $W^{(t)}$ , the a- and g-steps minimize  $Q(\cdot | W^{(t)})$ . Since our total objective is  $J = Q + \sum \phi(W^{(t)})$ , this implies

$$J(A^{(t+1)}, G^{(t+1)}, W^{(t)}) \leq J(A^{(t)}, G^{(t)}, W^{(t)}). \quad (28)$$

We're guaranteed to be helped by the w-step again now, so setting

$$W^{(t+1)} = \hat{w}(r(A^{(t+1)}, G^{(t+1)})), \quad (29)$$

and using our result from the previous section gives

$$J(A^{(t+1)}, G^{(t+1)}, W^{(t+1)}) \leq J(A^{(t+1)}, G^{(t+1)}, W^{(t)}). \quad (30)$$

Thus chaining the inequalities and  $L(A, G) = \min_W J(A, G, W)$  gives

$$L(A^{(t+1)}, G^{(t+1)}) \leq L(A^{(t)}, G^{(t)}). \quad (31)$$

This is enough to guarantee that robust HMF with Hogg's w-step converges to the Cauchy MLE.

## Extra Notes

- The update  $w_{ij} \leftarrow \hat{w}(r_{ij})$  is exactly Hogg's update rule if one combines  $\sigma_{ij}$  into the weights:  $\tilde{w}_{ij} = w_{ij}/\sigma_{ij}^2$ . I kept the updated weights separate to the data variances since it makes the connection to the MLE clearer (in my mind).
- Different  $\rho$  losses (negative log likelihoods) yield different w-step rules.
- This is basically the standard argument for IRLS (iteratively reweighted least squares), didn't require much extra.
- I accidentally overloaded my notation a bit ( $Q$  is two things). Sorry about that.
- This view gives a very natural interpretation for any regularisation, priors or constraints.