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

Tom Hilder

25/08/25

Setup

Take measured data F_{ij} with known variances σ_{ij}^2. We model with low rank:

$$Y_{ij} = a_i^\top g_j, \tag{1}$$

as matrices

$$\begin{aligned} Y &: N \times M, \\ A &: N \times K, \\ G &: K \times M, \end{aligned}$$

such that

$$Y \approx AG. \tag{2}$$

Assume Gaussian heteroskedastic noise gives χ^2 objective:

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

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

a-step

$$\begin{aligned} A_i &\leftarrow G_i F_i, \\ [G_i]_{kk'} &= \sum_{j=1}^{m} \frac{g_{kj} g_{k'j}}{\sigma_{ij}^2}, \\ [F_i]_k &= \sum_{j=1}^{m} \frac{g_{kj} Y_{ij}}{\sigma_{ij}^2}. \end{aligned}$$

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 χ^2 objective. Instead, define

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

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:

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

Auxiliary Form

Claim: the loss can be expressed for any $r > 0$ in the following way

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

where

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

Let's prove that. Define

$$
J\left(w; r\right) = \tfrac{1}{2} w r^2 + \tfrac{Q^2}{2} \left(w - 1 - \log w\right) \tag{8}
$$

Differentiating with respect to w:

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

and

$$
\frac{\partial^2 J}{\partial w^2} = \frac{Q^2}{2} \frac{1}{w^2} > 0, \qquad \forall\, Q, w > 0. \tag{10}
$$

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

$$\hat{w}(r) = \mathrm{argmin}_w \, J(w; r) \tag{11}$$
$$= \frac{1}{1 + (r/Q)^2}, \tag{12}$$

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

$$\hat{w} = \frac{1}{1+t}, \qquad r^2 = Q^2 t, \tag{13}$$

such that

$$J(\hat{w}; r) = \tfrac{1}{2} w r^2 + \phi(\hat{w}) \tag{14}$$
$$\Rightarrow \tfrac{1}{2} w r^2 = \frac{Q^2}{2} \frac{t}{1+t} \tag{15}$$
$$\Rightarrow \phi(\hat{w}) = \frac{Q^2}{2} \left(\hat{w} - 1 - \log \hat{w}\right) \tag{16}$$
$$= \frac{Q^2}{2} \left(-\frac{t}{1+t} + \log(1+t)\right) \tag{17}$$
$$\Rightarrow J(\hat{w}; r) = \frac{Q^2}{2} \left(1 + \left(\tfrac{r}{Q}\right)^2\right). \tag{18}$$

Thus

$$\rho(r) = J(\hat{w}; r) \tag{19}$$
$$= \min_{0 < w \leq 1} \left[\tfrac{1}{2} w r^2 + \phi(w)\right], \tag{20}$$

as claimed.

Three-Step Algorithm

Define new objective:

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

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

$$L(A, G) = \min_W J(A, G, W), \tag{22}$$

and if

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

then

$$\left[\hat{W}\right]_{ij} = \hat{w}(r_{ij}) \tag{24}$$
$$= \frac{1}{1 + (r_{ij}/Q)^2}. \tag{25}$$

This immediately yield's Hogg's procedure

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

where the a-step optises the quadratic

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

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

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

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

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

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

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

and using our result from the previous section gives

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

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

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

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 σ_{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 ρ 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.