Nonnegative Matrix Factorizations with intercepts

Figure 1 below presents a comparison between the factorizations produced by the standard NMF algorithm and its improvement $NMF_{intercept}$ on a synthetic dataset in which columns 36 to 85 are constant “genes”. As can be easily seen in the Figure, these “genes” are reconstructed by the standard NMF algorithm from combinations of clusters, while $NMF_{intercept}$ uses the additional degrees of freedom So to produce null cluster membership degrees S_{cg} for the constant genes. Moreover, $NMF_{intercept}$ recovers with much more accuracy than standard *NMF* the original sample clusters, the standard NMF algorithm being confused by the cluster overlaps. This improvement in recovery of the original clusters is very important in our application, where we aim at a correct sub-classification of samples.

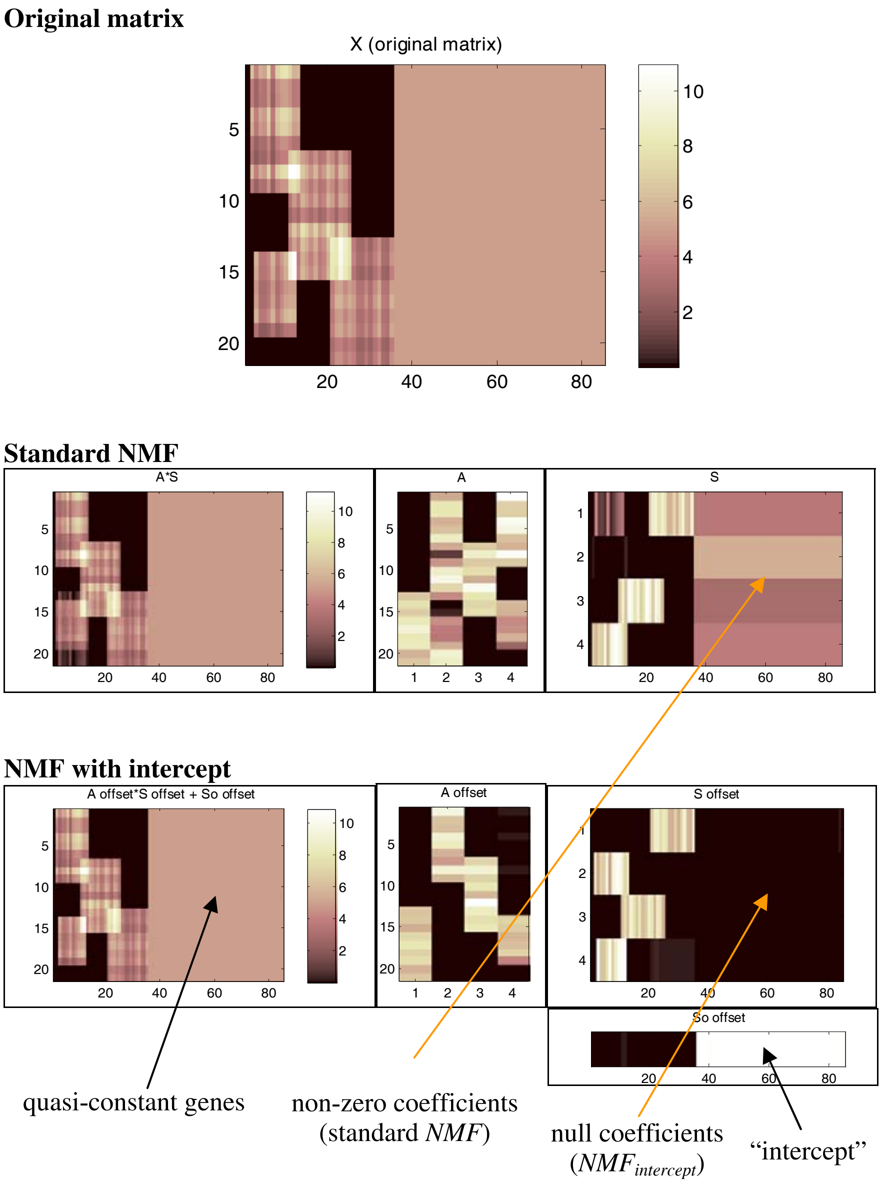

Figure 1. Comparing standard *NMF* with $NMF_{intercept}$