Supplementary Materials for "Going off the Grid: Iterative Model Selection for Biclustered Matrix Completion"

Eric C. Chi, Liuyi Hu, Arvind K. Saibaba, and Arvind U. K. Rao

1 Gradient of BIC

Consider the BIC objective function

$$\text{BIC}(\gamma_r, \gamma_c) \equiv |\Omega| \log\left(\|\mathbf{P}_\Omega \mathbf{x} - \mathbf{P}_\Omega \mathbf{S}^{-1} \mathbf{P}_\Omega \mathbf{x}\|_2^2\right) + \log(|\Omega|)\,\text{tr}(\mathbf{S}^{-1})$$

where $\mathbf{S} \equiv [\mathbf{P}_\Omega + \gamma_c(\mathbf{L}_c \otimes \mathbf{I}) + \gamma_r(\mathbf{I} \otimes \mathbf{L}_r)]$. Define $\mathbf{z}$ as the solution of the system $\mathbf{S}\mathbf{z} = \mathbf{P}_\Omega \mathbf{x}$ and define the residual $\mathbf{r} \equiv \mathbf{P}_\Omega(\mathbf{z} - \mathbf{x})$. We first note that the partial derivatives of $\mathbf{S}^{-1}$ are

$$\frac{\partial \mathbf{S}^{-1}}{\partial \gamma_r} = -\mathbf{S}^{-1}(\mathbf{I} \otimes \mathbf{L}_r)\mathbf{S}^{-1} \equiv -\mathbf{S}_r \qquad \text{and} \qquad \frac{\partial \mathbf{S}^{-1}}{\partial \gamma_c} = -\mathbf{S}^{-1}(\mathbf{L}_c \otimes \mathbf{I})\mathbf{S}^{-1} \equiv -\mathbf{S}_c.$$

The partial derivatives with respect to $\|\mathbf{P}_\Omega \mathbf{r}\|^2$ can be computed as

$$\frac{\partial \|\mathbf{P}_\Omega \mathbf{r}\|^2}{\partial \gamma_r} = -2\mathbf{x}^\mathsf{T} \mathbf{P}_\Omega \mathbf{S}_r \mathbf{P}_\Omega \mathbf{r}, \qquad \frac{\partial \|\mathbf{P}_\Omega \mathbf{r}\|^2}{\partial \gamma_c} = -2\mathbf{x}^\mathsf{T} \mathbf{P}_\Omega \mathbf{S}_c \mathbf{P}_\Omega \mathbf{r}.$$

The derivatives with respect to $\text{tr}(\mathbf{S}^{-1})$ can be computed as

$$\frac{\partial\,\text{tr}(\mathbf{S}^{-1})}{\partial \gamma_r} = -\,\text{tr}(\mathbf{S}_r) \qquad \frac{\partial\,\text{tr}(\mathbf{S}^{-1})}{\partial \gamma_c} = -\,\text{tr}(\mathbf{S}_c).$$

Combining together this gives us the gradient

$$\nabla \text{BIC} = \begin{pmatrix} -\frac{2|\Omega|}{\|\mathbf{P}_\Omega \mathbf{r}\|^2} \mathbf{x}^\mathsf{T} \mathbf{P}_\Omega \mathbf{S}_r \mathbf{P}_\Omega \mathbf{r} - \log(|\Omega|)\,\text{tr}(\mathbf{S}_r) \\ -\frac{2|\Omega|}{\|\mathbf{P}_\Omega \mathbf{r}\|^2} \mathbf{x}^\mathsf{T} \mathbf{P}_\Omega \mathbf{S}_c \mathbf{P}_\Omega \mathbf{r} - \log(|\Omega|)\,\text{tr}(\mathbf{S}_c) \end{pmatrix}.$$

The Hutchinson approximation can be used to evaluate the gradient. For example, note that

$$\text{tr}(\mathbf{S}_c) = \frac{1}{N} \sum_{k=1}^{N} \mathbf{w}_k^\top \mathbf{S}^{-1}(\mathbf{L}_c \otimes \mathbf{I})\mathbf{S}^{-1}\mathbf{w}_k.$$

By using the same set of vectors w_k, as needed for the objective function evaluation, it can be readily seen that the gradient function can be evaluated using all the information already available. The only other computation needed is $\mathbf{S}^{-1}\mathbf{P}_\Omega \mathbf{x}$.

2 Numerical Experiments: AIC

We perform the exact same comparisons that are conducted in Section 7.1 replacing the BIC with AIC. Figure 1 corresponds to Figure 3 in the manuscript. Figure 2 corresponds to Figure 4 in the manuscript. Figure 3 corresponds to Figure 5 in the manuscript. The results and interpretations are essentially the same.

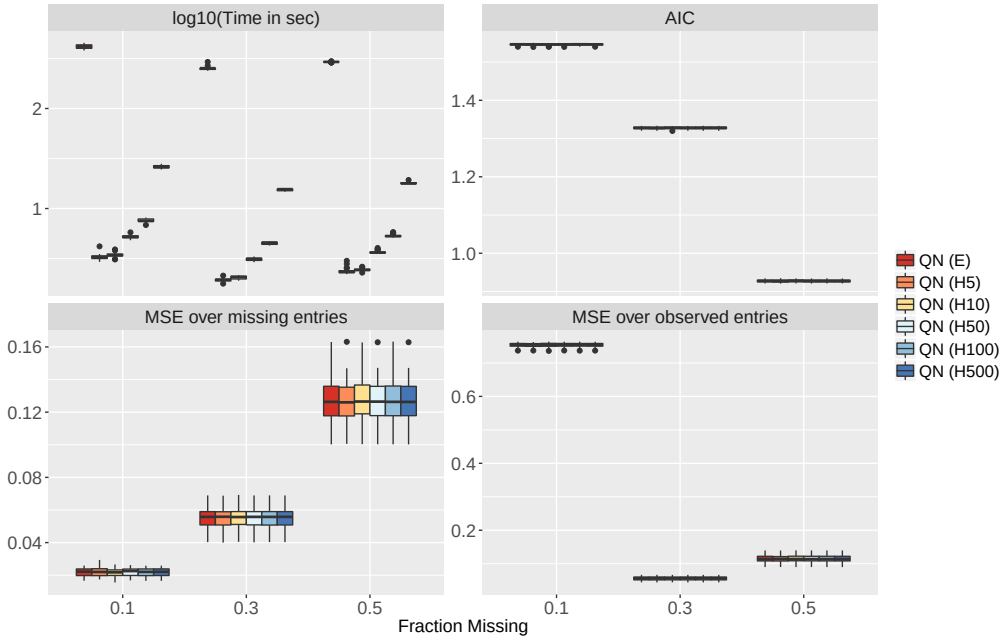

Figure 1: Comparison between IMS via Quasi-Newton with exact computation (E) and IMS via Quasi-Newton with Hutchinson estimation (HN indicates N samples), under different missing fractions.

3 Mondrian Example in Section 2

We give details on the experiment comparing noisy matrix completion of Composition A by LRMC and BMC. The data matrix is 370-by-380. Each element takes on an integer value between 0 and 255. We added i.i.d. $\mathcal{N}(0, \sigma^2)$ noise where $\sigma = 50$ and removed 50% of the entries completely at random. Note that $\sigma = 50$ corresponds to noise that is on the order of 20% the dynamic range of the pixel intensities in the image. We created 25 instances of corrupted Composition A in this way.

For each of the 25 replicates we performed LRMC and BMC. We used the singular value thresholding (SVT) algorithm (Cai et al., 2010) to perform LRMC. To choose the regularization parameter, we employed hold-out validation on 10% of the observed data over a grid of regularization parameters γ_n in (2.2). The estimated model that gave the best least squares prediction over the validation set was chosen. We used IMS to select the parameters γ_c and γ_r in the BMC formulation.

To quantify the quality of the estimates, we also computed the average squared error

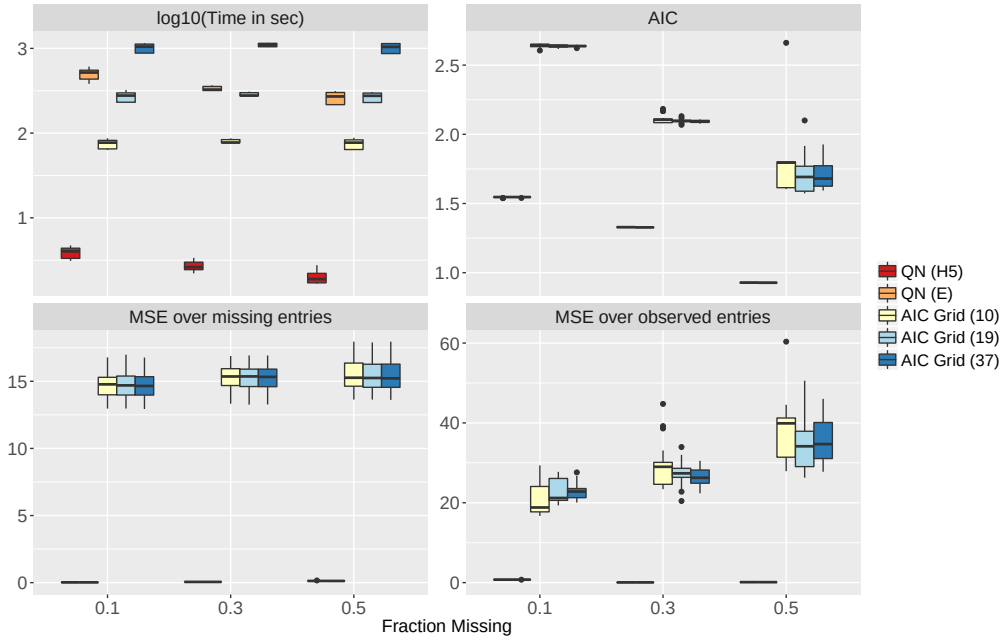

Figure 2: Comparison of (i) IMS via Quasi-Newton with exact computation (E), (ii) IMS via Quasi-Newton with Hutchinson estimation (HN indicates N samples), and (iii) cross-validation grid-search, under different missing fractions.

over the true and estimated values over the missing entries. Averaged over the 25 replicates we observed an estimated MSE of 501.23 for LRMC and 433.32 for BMC.

4 Composition A by Piet Mondrian

We now revisit the problem of completing Composition A by Piet Mondrian introduced at the end of Section 2. Recall the painting is an example with an underlying checkerboard pattern (Figure 1 in the manuscript).

We performed the following illustrative experiment to compare the performance using Quasi-Newton with Hutchinson estimation and conjugate gradient with Hutchinson estimation. The experiment were repeated for 30 times under missing fractions of 0.1, 0.3, and 0.5. Figure 4 and Figure 5 show the timing and prediction accuracy results using the BIC and AIC respectively.

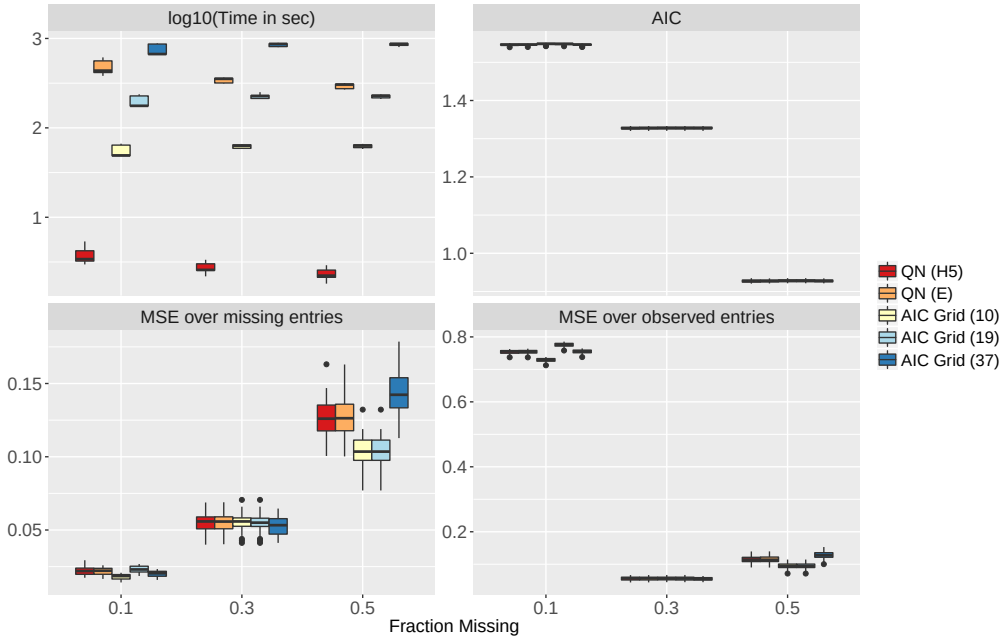

Figure 3: Comparison of (i) IMS via Quasi-Newton with exact computation (E), (ii) IMS via Quasi-Newton with Hutchinson estimation (HN indicates N samples), and (iii) AIC grid-search, under different missing fractions.

References

Cai, J.-F., Candès, E. J., and Shen, Z. (2010), "A Singular Value Thresholding Algorithm for Matrix Completion," *SIAM Journal on Optimization*, 20, 1956–1982.

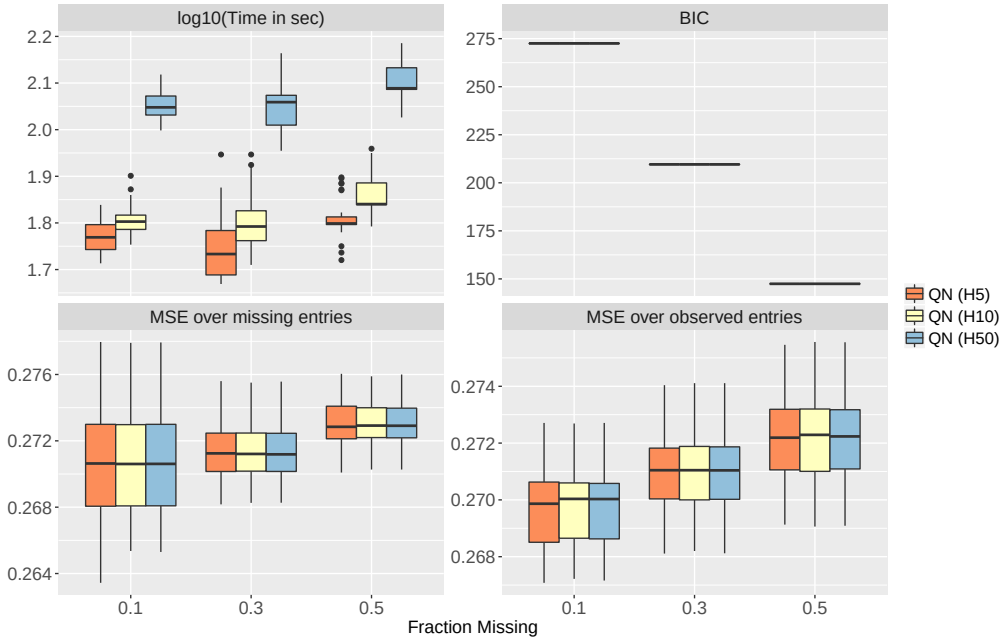

Figure 4: Mondrian: Comparison between Quasi-Newton with Hutchinson estimation (QN, size=N) under different sample size ($N = 5, 10, 50$) and different missing fractions.

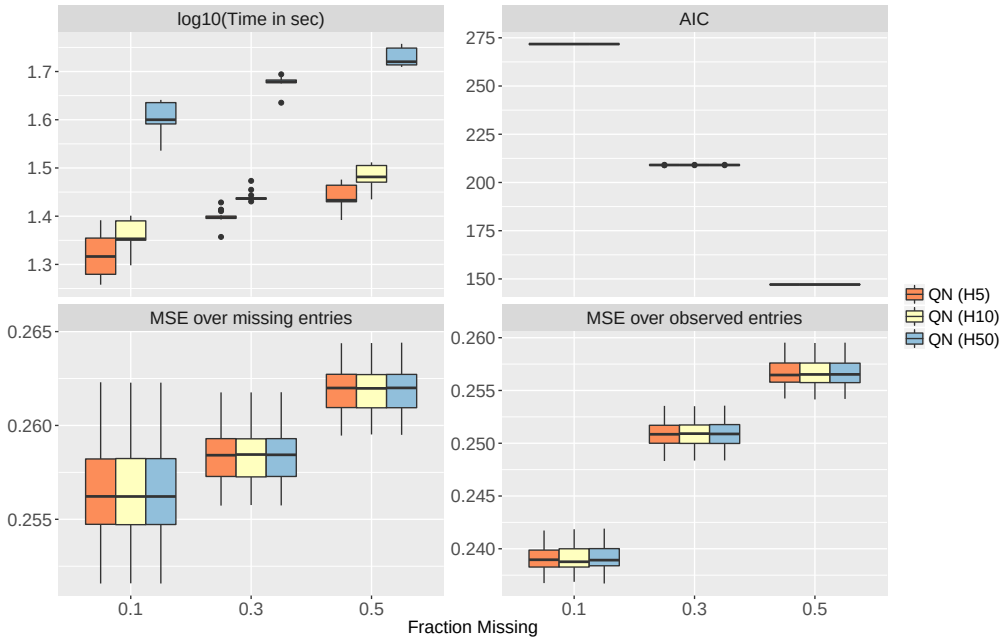

Figure 5: Mondrian: Comparison between Quasi-Newton with Hutchinson estimation (QN, size=N) under different sample size ($N = 5, 10, 50$) and different missing fractions.