EFFICIENT MATRIX COMPLETION WITH GAUSSIAN MODELS

By

Flavien Léger

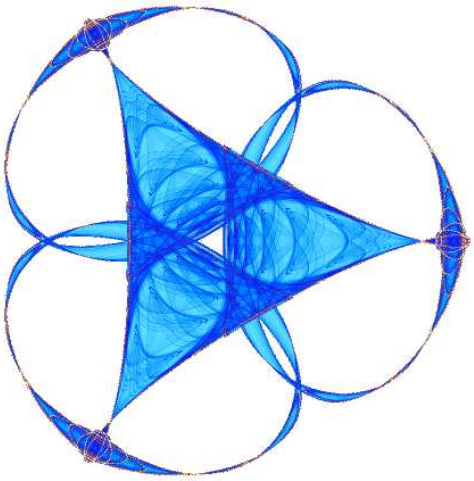
Guoshen Yu

and

Guillermo Sapiro

IMA Preprint Series # 2343

(October 2010)



INSTITUTE FOR MATHEMATICS AND ITS APPLICATIONS

UNIVERSITY OF MINNESOTA

400 Lind Hall

207 Church Street S.E.

Minneapolis, Minnesota 55455–0436

Phone: 612/624-6066 Fax: 612/626-7370

URL: http://www.ima.umn.edu

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
OCT 2010		**00-00-2010 to 00-00-2010**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Efficient Matrix Completion with Gaussian Models	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
University of Minnesota,Institute for Mathematics and Its Application,207 Church Street SE,Minneapolis,MN,55455-0436	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

A general framework based on Gaussian models and a MAPEM algorithm is introduced in this paper for solving matrix/ table completion problems. The numerical experiments with the standard and challenging movie ratings data show that the proposed approach, based on probably one of the simplest probabilistic models, leads to the results in the same ballpark as the state-of-the-art, at a lower computational cost.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**5**	

EFFICIENT MATRIX COMPLETION WITH GAUSSIAN MODELS

Flavien Léger

CMLA, ENS Cachan, 61 av. du
Président Wilson, Cachan 94235, France

Guoshen Yu, Guillermo Sapiro

ECE, University of Minnesota,
Minneapolis, 55455, MN, USA

ABSTRACT

A general framework based on Gaussian models and a MAP-EM algorithm is introduced in this paper for solving matrix/table completion problems. The numerical experiments with the standard and challenging movie ratings data show that the proposed approach, based on probably one of the simplest probabilistic models, leads to the results in the same ballpark as the state-of-the-art, at a lower computational cost.

Index Terms— Matrix completion, inverse problems, collaborative filtering, Gaussian mixture models, MAP estimation, EM algorithm

1. INTRODUCTION

Matrix completion amounts to estimate all the entries in a matrix $\mathbf{F} \in \mathbb{R}^{M\times N}$ from a partial and potentially noisy observation

$$\mathbf{Y} = \mathbf{H} \bullet \mathbf{F} + \mathbf{W}, \tag{1}$$

where $\mathbf{H} \in \mathbb{R}^{M\times N}$ is a binary matrix with 1/0 entries masking a portion of $\mathbf{F}$ through the element-wise multiplication $\bullet$, and $\mathbf{W} \in \mathbb{R}^{M\times N}$ an additive noise. With matrix rows, columns, and entry values assigned to various attributes, matrix completion may have numerous applications. For example, when the rows and the columns of $\mathbf{F}$ are attributed to users and items such as movies and books, and an entry at position (m,n) records a score given by the m-th user to the n-th item, matrix completion predicts users' scores on the items they have not yet rated, based on the available scores recorded in $\mathbf{Y}$, so that personalized item recommendation system becomes possible. This is a classical problem in collaborative filtering.

To solve an ill-posed matrix completion problem, one must rely on some prior information, or in other words, some data model. The most popular family of approaches in the literature assumes that the matrix $\mathbf{F}$ follows approximately a low rank model, and calculates the matrix completion with a matrix factorization [4, 14]. Theoretical results regarding the completion of low-rank matrices have been recently obtained as well, e.g., [3, 13] and references therein. More elaborative probabilistic models and some refinement have been further studied on top of matrix factorization, leading to state-of-the-art results [7, 8, 17] .

In image processing, assuming that local image patches follow Gaussian mixture models (GMM), Yu, Sapiro and Mallat have recently reported excellent results in a number of inverse problems [16]. In particular, for inpainting, which is an analogue to matrix completion where the data is an image, the maximum a posteriori expectation-maximization (MAP-EM) algorithm for solving the GMM leads to state-of-the-art results, with a computational complexity considerably reduced with respect to the previous state-of-the-art approaches based on sparse models [9], (which are analogous to the low-rank model assumption for matrices).

In this paper, we investigate Gaussian modeling (a particular case of GMM with only a single Gaussian distribution) for matrix completion. Subparts of the matrix, typically rows or columns, are regarded as a collection of signals that are assumed to follow a Gaussian distribution. An efficient MAP-EM algorithm is introduced to estimate both the Gaussian parameters (mean and covariance) and the signals. We show through numerical experiments that the fast MAP-EM algorithm, based on the Gaussian model which is the simplest probabilistic model one can imagine, leads to results in the same ballpark as the state-of-the-art in movie rating prediction, at significantly lower computational cost. Recent theoretical results [15] further support the consideration of Gaussian models for the recovery of missing data.

Section 2 introduces the Gaussian model and the MAP-EM algorithm. After presenting the numerical experiments in Section 3, Section 4 concludes with some discussions.

2. MODEL AND ALGORITHM

2.1. Gaussian Model

Similar to the local patch decomposition often applied in image processing [16], let us consider each subpart of the matrix $\mathbf{F} \in \mathbb{R}^{M\times N}$, the i-th row $\mathbf{f}_i \in \mathbb{R}^N$ for example, as a signal. Let $\mathbf{h}_i \in \mathbb{R}^N$ denote the i-th row in the binary matrix $\mathbf{H}$, and let $N_i = |\{j|\mathbf{h}_i(j) \neq 0, 1 \leq j \leq N\}|$ be the number of non-zero entries in $\mathbf{h}_i$. Let $\mathbf{U}_i \in \mathbb{R}^{N_i\times N}$ denote a masking operator which maps from $\mathbb{R}^N$ to $\mathbb{R}^{N_i}$ extracting entries of $\mathbf{f}_i$ corresponding to the non-zero entries of $\mathbf{h}_i$, i.e., all but the $(\mathrm{Idx}(\mathbf{h}_i,k))$-th the entries in the k-th row of $\mathbf{U}_i$ are zero, with $(\mathrm{Idx}(\mathbf{h}_i,k))$ the index of the k-th non-zero entry in $\mathbf{h}_i$, $1 \leq k \leq N_i$. Let $\mathbf{y}_i \in \mathbb{R}^{N_i}$ and $\mathbf{w}_i \in \mathbb{R}^{N_i}$ be respectively the *sub-vector* of the i-th row of $\mathbf{Y}$ and $\mathbf{W}$, where the entries of $\mathbf{h}_i$ are non-zero. With this

notation, we can rewrite (1) in a more general linear model

$$\mathbf{y}_i = \mathbf{U}_i \mathbf{f}_i + \mathbf{w}_i, \tag{2}$$

for all the signals $\mathbf{f}_i$, $1 \leq i \leq M$. [1] Note that $\mathbf{f}_i$ can also be columns, or 2D sub-matrices of $\mathbf{F}$ rendered in 1D.

The Gaussian model assumes that each signal $\mathbf{f}_i$ is drawn from a Gaussian distribution, with a probability density function

$$p(\mathbf{f}_i) = \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{f}_i - \mu)^T \Sigma^{-1} (\mathbf{f}_i - \mu)\right), \tag{3}$$

where Σ and μ are the unknown covariance and mean parameters of the Gaussian distribution. The noise $\mathbf{w}_i$ is assumed to follow a Gaussian distribution with zero mean and covariance $\Sigma_{\mathbf{w}}$, here assumed to be known (or calibrated).

Estimating the matrix $\mathbf{F}$ from the partial observation $\mathbf{Y}$ can thus be casted into the following problem:

- Estimate the Gaussian parameters (μ, Σ), from the observation $\{\mathbf{y}_i\}_{1 \leq i \leq M}$.
- Estimate $\mathbf{f}_i$ from $\mathbf{y}_i$, $1 \leq i \leq M$, using the Gaussian distribution $\mathcal{N}(\mu, \Sigma)$.

Since this is a non-convex problem, we present an efficient maximum a posteriori expectation-maximization (MAP-EM) algorithm that calculates a local-minimum solution [1, 16].

2.2. Algorithm

Following a simple initialization, addressed in Section 2.2.3, the MAP-EM algorithm is an iterative procedure that alternates between two steps:

1. E-step: signal estimation. Assuming the estimates $(\tilde{\mu}, \tilde{\Sigma})$ are known (following the previous M-step), for each i one computes the maximum a posteriori (MAP) estimate $\tilde{\mathbf{f}}_i$ of $\mathbf{f}_i$.
2. M-step: model estimation. Assuming the signal estimates $\tilde{\mathbf{f}}_i$, $\forall i$, are known (following the previous E-step), one estimates (updates) the Gaussian model parameters $(\tilde{\mu}, \tilde{\Sigma})$.

2.2.1. E-step: Signal Estimation

It is well known that under the Gaussian models assumed in Section 2.1, the MAP estimate that maximizes the log a-posteriori probability

$$\begin{aligned}
\tilde{\mathbf{f}}_i &= \arg\max_{\mathbf{f}_i} \log p(\mathbf{f}_i | \mathbf{y}_i, \tilde{\mu}, \tilde{\Sigma}) \\
&= \arg\max_{\mathbf{f}_i} \left(\log p(\mathbf{y}_i | \mathbf{f}_i) + \log p(\mathbf{f}_i | \tilde{\mu}, \tilde{\Sigma})\right) \\
&= \arg\min_{\mathbf{f}_i} \left((\mathbf{U}_i \mathbf{f}_i - \mathbf{y}_i)^T \Sigma_{\mathbf{w}}^{-1} (\mathbf{U}_i \mathbf{f}_i - \mathbf{y}_i) + (\mathbf{f}_i - \tilde{\mu})^T \tilde{\Sigma}^{-1} (\mathbf{f}_i - \tilde{\mu})\right) \\
&= \tilde{\mu} + \tilde{\Sigma} \mathbf{U}_i^T (\mathbf{U}_i \tilde{\Sigma} \mathbf{U}_i^T + \Sigma_{\mathbf{w}})^{-1} (\mathbf{y}_i - \mathbf{U}_i \tilde{\mu}),
\end{aligned} \tag{4}$$

[1] Writing $\mathbf{y}_i$ in the reduced dimension N_i leads to a calculation in dimension N_i instead of N, which is considerably faster if $N_i \ll N$.

is a linear estimator and is optimal in the sense that it minimizes the mean square error (MSE), i.e., $E_{\mathbf{f}_i, \mathbf{w}_i}[\|\mathbf{f}_i - \tilde{\mathbf{f}}_i\|_2^2] = \min_g E_{\mathbf{f}_i, \mathbf{w}_i}[\|\mathbf{f}_i - g(\mathbf{y}_i)\|_2^2]$, as well as the mean absolute error (MAE), i.e., $E_{\mathbf{f}_i, \mathbf{w}_i}[\|\mathbf{f}_i - \tilde{\mathbf{f}}_i\|_1] = \min_g E_{\mathbf{f}_i, \mathbf{w}_i}[\|\mathbf{f}_i - g(\mathbf{y}_i)\|_1]$, where g is any mapping from $\mathbb{R}^N \to \mathbb{R}^{N_i}$ [6]. The second equality of (4) follows from the Bayes rule, the third follows from the Gaussian models $\mathbf{f}_i \sim \mathcal{N}(\mu, \Sigma)$ and $\mathbf{w}_i \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{w}})$, and the last is obtained by deriving the third line with respect to $\mathbf{f}_i$.

The close-form MAP estimate (4) can be calculated fast. Observe that $\mathbf{U}_i \in \mathbb{R}^{N_i \times N}$ is a sparse extraction matrix, each row containing only one non-zero entry with value 1, whose index corresponds to the non-zero entry in $\mathbf{h}_i$. Therefore, the multiplication operations that involve $\mathbf{U}_i$ or $\mathbf{U}_i^T$ can be realized by extracting the appropriate rows or columns at zero computational cost. The complexity of (4) is therefore dominated by the matrix inversion. As $\mathbf{U}_i \tilde{\Sigma} \mathbf{U}_i^T + \Sigma_{\mathbf{w}}$ is positive-definite, $(\mathbf{U}_i \tilde{\Sigma} \mathbf{U}_i^T + \Sigma_{\mathbf{w}})^{-1}$ can be implemented with $N_i^3/3$ flops through a Cholesky factorization [2].

In a typical case where $\mathbf{f}_i$ is the i-th row of the matrix $\mathbf{F}$, $1 \leq i \leq M$, to estimate $\mathbf{F}$ the total complexity of the E-step is therefore dominated by $\sum_{i=1}^{M} N_i^3/3$ flops. For typical rating prediction datasets that are highly sparse, among a large number of items N, most users have rated more or less only a small number $N_i \approx N/C$ of items, where C is large. The total complexity of the E-step is thus dominated by $\frac{1}{3C^3} MN^3$ flops.

2.2.2. M-step: Model Estimation

The parameters of the model are estimated/updated with the maximum likelihood estimate,

$$(\tilde{\mu}, \tilde{\Sigma}) = \arg\max_{\mu, \Sigma} \log p(\{\tilde{\mathbf{f}}_i\}_{1 \leq i \leq M} | \mu, \Sigma). \tag{5}$$

With the Gaussian model $\mathbf{f}_i \sim \mathcal{N}(\mu, \Sigma)$, it is well known that the resulting estimate is the empirical one,

$$\tilde{\mu} = \frac{1}{M} \sum_{i=1}^{M} \tilde{\mathbf{f}}_i \;\text{ and }\; \tilde{\Sigma} = \frac{1}{M} \sum_{i=1}^{M} (\tilde{\mathbf{f}}_i - \tilde{\mu})(\tilde{\mathbf{f}}_i - \tilde{\mu})^T. \tag{6}$$

The empirical covariance estimate may be improved through regularization when there is lack of data (let us take an example of standard rating prediction, where there are $N \sim 10^3$ items and $M \sim 10^4$ users, the dimension of the covariance matrix Σ_k is $N \times N \sim 10^6$). A simple and standard eigenvalue-based regularization is used here,

$$\tilde{\Sigma} \leftarrow \tilde{\Sigma} + \varepsilon Id, \tag{7}$$

where ε is a small constant. The regularization also guarantees that the estimate $\tilde{\Sigma}$ of the covariance matrix is full-rank, so that (4) is always well defined.

To estimate $\mathbf{F}$, the computational complexity of the M-step is dominated by the calculation of the empirical covariance estimate requiring MN^2 flops, which is negligible with respect to the E-step.

As the MAP-EM algorithm iterates, the MAP probability of the observed signals $p(\tilde{\mathbf{f}}_i|\mathbf{y}_i,\tilde{\mu},\tilde{\Sigma})$ increases. This can be observed by interpreting the E- and M-steps as a coordinate descent optimization [5]. In the experiments, the algorithm converges within 10 iterations.

2.2.3. Initialization

The MAP-EM algorithm is initialized with an initial guess of $\mathbf{f}_i$, $\forall i$. The experiments show that the result is insensitive to the initialization for movie rating prediction. In the numerical experiments, all the unknown entries are initialized to 3 for datasets containing ratings ranging from 1 to 5 or 6.

2.2.4. Computational and Memory Requirements

In a typical case where the matrix row decomposition in (2) is considered, the overall computational complexity of the MAP-EM to estimate an $M \times N$ matrix is dominated by $\frac{1}{3C^3}LMN^3$, with L the number of iterations (typically < 10) and $1/C$ the available data ratio, with C typically large (≈ 25 for the standard movie ratings data). The algorithm is thus very fast. As each row $\mathbf{f}_i$ is treated as a signal and the signals can be estimated in sequence, the memory requirement is dominated by N^2 (to store the covariance matrix Σ).

3. NUMERICAL EXPERIMENTS

3.1. Experimental Protocols

The experimental protocols strictly follow those described in the literature [4, 7, 8, 10, 12, 14, 17]. The proposed method is evaluated on two movie rating prediction benchmark datasets, the *EachMovie* dataset and the *1M MovieLens* dataset. [2] Two test protocols, the so-called "weak generalization" and "strong generalization," [10] are applied.

- **Weak generalization** measures the ability of a method to generalize to other items rated by the *same* users used for training the method. One known rating is randomly held out from each user's rating set to form a test set, the rest known ratings form a training set. The method is trained using the data in the training set, and its performance is evaluated over the test set.

- **Strong generalization** measures the ability of the method to generalize to some items rated by *novel* users that have *not* been used for training. The set of users is first divided into training users and test users. Learning is performed with all available ratings from the training set. To test the method, the ratings of each test users are further split into an observed set and a held-out set. The method is shown the observed ratings, and is evaluated by predicting the held-out ratings.

[2] http://www.grouplens.org/

The EachMovie dataset contains 2.8 million ratings in the range $\{1,\ldots,6\}$ for 1,648 movies (columns) and 74,424 users (rows). Following the standard procedure [10], users with fewer than 20 ratings and movies with less than 2 ratings are removed. This leaves us 36,656 users, 1,621 movies, and 2.5 million ratings (available data ratio 4.3%). We randomly select 30,000 users for the weak generalization, and 5,000 users for the strong generalization. The 1M MovieLens dataset contains 1 million ratings in the range $\{1,\ldots,5\}$ for 3,900 movies (columns) and 6,040 users (rows). The same filtering leaves us 6,040 users, 3,592 movies, and 1 million ratings (available data ratio 4.6%). We randomly select 5,000 users for the weak generalization, and 1,000 users for the strong generalization. Each experiment is run 3 times and the average reported.

The performance of the method is measured by the standard Normalized Mean Absolute Error (NMAE), computed by normalizing the mean absolute error by a factor for which random guessing produces a score of 1. The factor is 1.944 for EachMovie, and 1.6 for MovieLens.

3.2. Proposed Method Setup

In contrast to most exiting algorithms in the literature, the proposed method, thanks to its simplicity, enjoys the advantage of having very few intuitive parameters. The covariance regularization parameter ε in (7) is set equal to 0.3 (whose square root is one order of magnitude smaller than the maximum rating), the results being insensitive to this value as shown by the experiments. The number of iterations of the MAP-EM algorithm is fixed at $L = 10$, beyond which the convergence of the algorithm is always observed. The noise $\mathbf{w}_i$ in (2) is neglected, i.e., $\Sigma_{\mathbf{w}}$ is set to $\mathbf{0}$, as the movie rating datasets mainly involve missing data, the noise being implicit and assumed negligible.

The experiments show that considering row $\mathbf{f}_i \in \mathbb{R}^N$ of the matrix $\mathbf{F} \in \mathbb{R}^{M\times N}$ as signals leads to slightly better results than taking columns or 2D sub-matrices. This means that each user is a signal, whose dimension is the number of movies.

As in previous works [17], a post-processing that projects the estimated rating to an integer within the rating range is performed.

A Matlab code of the proposed algorithm is available at *http://www.cmap.polytechnique.fr/∼yu/research/MC/demo.html*.

3.3. Results

The results of the proposed method are compared with the best published ones including User Rating Profile (URP) [10], Attitude [11], Maximum Margin Matrix Factorization (MMMF) [14], Ensemble of MMMF (E-MMMF) [4], Item Proximity based Collaborative Filtering (IPCF) [12], Gaussian Process Latent Variable Models (GPLVM) [7], Mixed Membership Matrix Factorization (M^3F) [8], and Nonparametric Bayesian Matrix Completion (NBMC) [17]. For each of these methods, more than one results produced with different configurations are

often reported, among which we systematically cite the best one. All these methods are significantly more complex than the one here proposed.

Tables 1 and 2 presents the results of various methods for both weak and strong generalizations on the two datasets. NBMC generates most often the best results, followed closely by the proposed method referred to as GM (Gaussian Model) and GPLVM, all of them outperforming the other methods. The results produced by the proposed GM, with a by far simpler model and faster algorithm, is in the same ballpark as those of NBMC and GPLVM, the difference with NMAE being smaller than about 0.005, marginal in the rating range that goes from 1 to 5 or 6.

EachMovie	Weak NMAE	Strong NMAE
URP [10]	0.4422	0.4557
Attitude [11]	0.4520	0.4550
MMMF [14]	0.4397	0.4341
IPCF [12]	0.4382	0.4365
E-MMMF [4]	0.4287	0.4301
GPLVM [7]	0.4179	0.4134
M^3F [8]	0.4293	n/a
NBMC [17]	**0.4109**	**0.4091**
GM (proposed)	0.4164	0.4163

Table 1. NMAEs generated by different methods for EachMovie database. The smallest NMAE is in boldface.

1M MovieLens	Weak NMAE	Strong NMAE
URP [10]	0.4341	0.4444
Attitude [11]	0.4320	0.4375
MMMF [14]	0.4156	0.4203
IPCF [12]	0.4096	0.4113
E-MMMF [4]	0.4029	0.4071
GPLVM [7]	0.4026	0.3994
M^3F [8]	n/a	n/a
NBMC [17]	**0.3916**	0.3992
GM (proposed)	0.3959	**0.3928**

Table 2. NMAEs generated by different methods for 1M MovieLens database. The smallest NMAE is in boldface.

4. CONCLUSION AND DISCUSSION

We have shown that a Gaussian model and a MAP-EM algorithm provide a simple and computational efficient solution for matrix completion, leading to results in the same ballpark as state-of-the-art ones for movie rating prediction.

Future work may go in several directions. On the one hand, the proposed conceptually simple and computationally efficient method may provide a good baseline for further refinement, for example by incorporating user and item bias or meta information [17]. On the other hand, Gaussian mixture models (GMM) that have been shown to bring dramatic improvements over single Gaussian models in image inpainting [16], are expected to better capture different characteristics of various categories of movies (comedy, action, etc.) and classes of users (age, gender, etc.). However, no significant improvement by GMM over Gaussian model has yet been observed for movie rating prediction. This needs to be further investigated, and such improvement might come from proper grouping and initialization.

Acknowledgments: Work partially supported by NSF, ONR, NGA, ARO, and NSSEFF. We thank S. Mallat for co-developing the proposed framework, and M. Zhou and L. Carin for their comments and help with the data.

5. REFERENCES

[1] S. Allassonniere, Y. Amit, and A. Trouvé. Towards a coherent statistical framework for dense deformable template estimation. *J.R. Statist. Soc. B*, 69(1):3–29, 2007.

[2] S.P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[3] E.J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009.

[4] D. DeCoste. Collaborative prediction using ensembles of maximum margin matrix factorizations. In *Proc. ICML*, pages 249–256, 2006.

[5] R.J. Hathaway. Another interpretation of the EM algorithm for mixture distributions. *Stat. & Prob. Letters*, 4(2):53–56, 1986.

[6] S.M. Kay. *Fundamentals of Statistical signal processing, Volume 1: Estimation theory*. Prentice Hall, 1998.

[7] N.D. Lawrence and R. Urtasun. Non-linear matrix factorization with Gaussian processes. In *Proc. ICML*, pages 601–608, 2009.

[8] L. Mackey, D. Weiss, and M.I. Jordan. Mixed membership matrix mactorization. In *Proc. ICML*, 2010.

[9] S. Mallat. *A Wavelet Tour of Signal Proc.: The Sparse Way, 3rd edition*. Academic Press, 2008.

[10] B. Marlin. *Collaborative Filtering: A Machine Learning Perspective*. Master thesis, University of Toronto, 2004.

[11] B. Marlin. Modeling user rating profiles for collaborative filtering. In *Proc. NIPS*, volume 16, pages 627–634, 2004.

[12] S.T. Park and D.M. Pennock. Applying collaborative filtering techniques to movie search for better ranking and browsing. In *Proc. ACM SIGKDD*, page 559, 2007.

[13] B. Recht, W. Xu, and B. Hassibi. Necessary and sufficient conditions for success of the nuclear norm heuristic for rank minimization. In *Proc. IEEE CDC*, pages 3065–3070. IEEE, 2009.

[14] J.D.M. Rennie and N. Srebro. Fast maximum margin matrix factorization for collaborative prediction. In *Proc. ICML*, pages 713–719, 2005.

[15] G. Yu and Sapiro. G. Statistical compressive sensing with Gaussian models. *submitted*, 2010.

[16] G. Yu, G. Sapiro, and S. Mallat. Solving inverse problems with piecewise linear estimators: from Gaussian mixture models to structured sparsity. *submitted, http://arxiv.org/abs/1006.3056*, June, 2010.

[17] M. Zhou, C. Wang, M. Chen, J. Paisley, D. Dunson, and L. Carin. Nonparametric Bayesian matrix completion. In *Proc. IEEE SAM*, 2010.