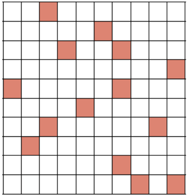
Adaptive sampling in matrix completion: When can it help? How?

Rachel Ward
University of Texas at Austin

November, 2017

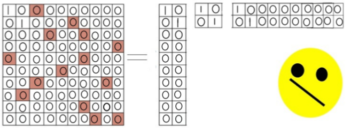
Low-rank matrix completion problem

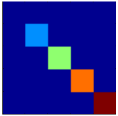
Given some entries of a matrix $\mathbf{M}$, exactly recover ("complete") hidden entries



- Assumption to make well-posed: $\mathbf{M}$ has *low rank*
- $\mathbf{M} \in \mathbb{R}^{n \times n}$ is rank-r means it has a "skinny" *singular value decomposition* $\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$
- An $n \times n$ matrix of rank-r has roughly $2nr$ degrees of freedom. Can we complete a matrix from $\approx nr$ entries given only the knowledge that it is rank-r ?

Need additional structure: incoherence





- Low-rank assumption not enough. Need additional structural assumptions
- Left and right *leverage scores* of $\mathbf{M} = \mathbf{U\Sigma V}^T$ measure angles, or "coherence" of row/columns with coordinate directions:
  - $L_i := \|\mathbf{U}(i,:)\|^2 = \|\mathbf{U}^T e_i\|^2, \quad i \in [n]$
  - $R_j := \|\mathbf{V}(j,:)\|^2 = \|\mathbf{V}^T e_j\|^2, \quad j \in [n]$
- Additional structural assumption for i.i.d. random sampling: uniformly flat leverage scores, or *incoherence*.

$$\max_i L_i, R_i \leq K\frac{r}{n}, \qquad K \geq 1 \text{ is not too big}$$

- Note $1 \leq K \leq \frac{n}{r}$ always.

Uniform-sampling matrix completion

[Candès Recht, 2009; Candès Tao, 2009; Recht 2011; Gross 2011; Chen 2013]

Theorem

Given an $n \times n$ matrix $\mathbf{M}$ of rank r.
Let $\Omega \subset [n]^2$ be a subset of the entries of $\mathbf{M}$, where each entry M_{ij} is observed independently with probability p.

The nuclear norm minimization algorithm

$$\min \|\mathbf{X}\|_* \quad s.t. \quad X_{ij} = M_{ij}, \quad (i,j) \in \Omega$$

will exactly recover $\mathbf{M}$ as its unique solution with probability at least $1 - \frac{1}{n^2}$, provided that

$$C \max\{L_i, R_j\} \log^2(n) \leq p.$$

Here, $C > 1$ is a universal constant.

Uniform-sampling matrix completion

[Candès Recht, 2009; Candès Tao, 2009; Recht 2011; Gross 2011; Chen 2013]

Theorem

Given an $n \times n$ matrix $\mathbf{M}$ of rank r.
Let $\Omega \subset [n]^2$ be a subset of the entries of $\mathbf{M}$, where each entry M_{ij} is observed independently with probability p. The nuclear norm minimization algorithm will exactly recover $\mathbf{M}$ as its unique solution with probability at least $1 - \frac{1}{n^2}$, provided that

$$C \max\{L_i, R_j\} \log^2(n) \leq p.$$

Here, $C > 1$ is a universal constant.

- $\mathbb{E}|\Omega| = pn^2 = C \max\{L_i, R_j\} n^2 \log^2(n)$.
- Incoherence: $\mathbb{E}|\Omega| = CKrn \log^2(n)$. Sharp up to the $\log^2(n)$.
- There are many faster alternative algorithms for matrix completion

A closer look at matrix leverage scores

Recall: $\mathbf{M} = \mathbf{U\Sigma V}^T$ is rank-r. $L_i := \|\mathbf{U}(i,:)\|^2, R_j := \|\mathbf{V}(j,:)\|^2$

- If an oracle told us (bounds on) the $2n$ leverage scores for $\mathbf{M}$, we might want to sample entries from a *weighted* probability distribution where entry (i,j) is sampled prop. to its "importance" $L_i + R_j$.
- (one-sided) leverage score sampling has long history in column/row subset selection/matrix sketching [Mahoney and Drineas, 2009; Mahoney 2011; Spielman and Srivastava 2011; Drineas et. al. 2012]

Leveraged sampling

Theorem

Given an $n \times n$ matrix $\mathbf{M}$ of rank-r.
Let $\Omega \subset [n]^2$ be a subset of the entries of $\mathbf{M}$ where each entry M_{ij} is observed independently with probability $P[i,j]$.
Nuclear norm minimization will exactly recover $\mathbf{M}$ as its unique solution with probability at least $1 - \frac{1}{n^2}$, provided that

$$C(L_i + R_j)\log^2(n) \leq P[i,j] \leq 1, \qquad \forall i,j$$

Here, $C > 1$ is a universal constant.

- $\mathbb{E}|\Omega| = \sum_{i,j} P[i,j] = 2Crn\log^2(n)$ is optimal up to $\log^2(n)$.
- Key idea: refined dual certificate proof, concentration w.r.t weighted $L_{2,\infty}$ matrix norm instead of entrywise L_∞ norm.

Y. Chen, S. Bhojanapalli, S. Sanghavi, and R. Ward. Completing any low-rank matrix, provably. JMLR, 2015.

About knowing the leverage scores ...

- Leverage scores could be learned as priors from representative training data
- However, typically leverage scores are not given beforehand
- What about *learning* leverage scores from samples? $2n$ leverage scores compared to $2nr$ degrees of freedom.
- Actually, we only need estimates of *large* leverage scores to apply previous result

2-phase low-rank matrix completion

Given: budget of m samples, parameter $\gamma \in (0, 1)$

- Draw batch of $m_1 = \gamma m$ entries via i.i.d. uniform sampling
- Construct best rank-r approximation of resulting zero-filled sample matrix, and compute its leverage scores $\widehat{L_i}, \widehat{R_j}$.
- Generate new batch of $m_2 = (1 - \gamma)m$ samples according to weighted distribution $\widehat{L_i} + \widehat{R_j}$
- Use all $m = m_1 + m_2$ samples to reconstruct with, e.g. nuclear norm minimization

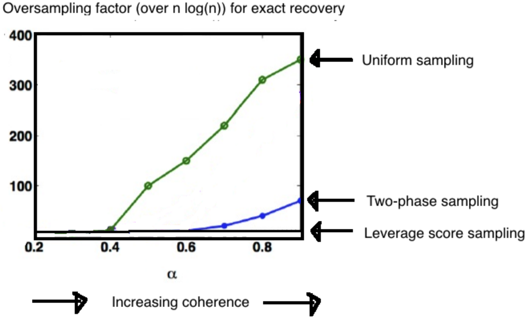
($\gamma = .75$ seems to be good choice)

Y. Chen, S. Bhojanapalli, S. Sanghavi, and R. Ward. Completing any low-rank matrix, provably. JMLR, 2015.

A typical simulation result

Consider 400×400, rank-10 power law matrices of form $\mathbf{M} = \mathbf{D}\mathbf{U}\mathbf{V}^\mathsf{T}\mathbf{D}$;

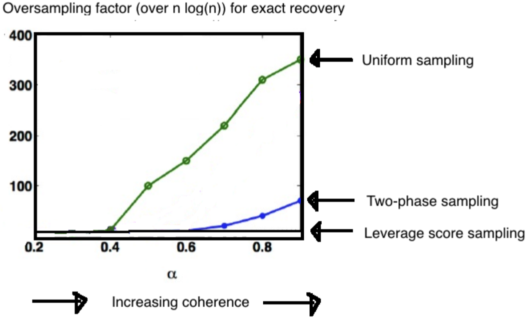
- $\mathbf{U}, \mathbf{V}$ are 400×10 i.i.d. Gaussian.
- $\mathbf{D}$ is diagonal with $D_j = j^{-\alpha}$. $\alpha = 0$: incoherent. $\alpha = 1$: pretty coherent



A typical simulation result

Consider 400×400, rank-10 power law matrices of form $\mathbf{M} = \mathbf{DUV^TD}$;

- $\mathbf{U}, \mathbf{V}$ are 400×10 i.i.d. Gaussian.
- $\mathbf{D}$ is diagonal with $D_j = j^{-\alpha}$. $\alpha = 0$: incoherent. $\alpha = 1$: pretty coherent



- Results are robust to noise
- No theory yet for this.

An alternative 2-phase algorithm

- For rank-r $\mathbf{M} = \mathbf{U\Sigma V}^T$ with condition number $\kappa(\mathbf{M}) = \sigma_1/\sigma_r$, it holds:

$$\sigma_r^2 L_i \leq \sum_{j=1}^{n} M_{i,j}^2 \leq \sigma_1^2 L_i, \quad \sigma_r^2 R_j \leq \sum_{i=1}^{n} M_{i,j}^2 \leq \sigma_1^2 R_j$$

- Implication: If $\mathbf{M}$ is well-conditioned, we can estimate its leverage scores from sample row and column norms. More amenable to theoretical analysis.

MC^2: two-phase algorithm

Given a fixed budget of m samples:

- (**Phase 1**) Observe each entry of $\mathbf{M}$ with probability p. Let Y be zero-filled sample matrix.
- (**Estimate leverage scores**) Set

$$\widehat{L}_i \leftarrow \frac{\kappa^2 \| Y[i,:] \|_2^2}{\| Y \|_F^2}, \quad \widehat{R}_j \leftarrow \frac{\kappa^2 \| Y[:,j] \|_2^2}{\| Y \|_F^2}, \quad i,j \in [1:n].$$

- (**Phase 2: Leveraged sampling**) Set

$$P[i,j] \leftarrow \min\{1, C \log^2(n) \left(\widehat{L}_i + \widehat{R}_j \right) \}$$

  Observe (i,j)th entry of $\mathbf{M}$ with probability $P[i,j]$.
- (**Completion**) Using all samples, complete to matrix $\widehat{\mathbf{M}}$ using e.g. nuclear norm minimization

A. Eftekhari, M. Wakin, and R. Ward. MC^2: A two-phase algorithm for leveraged matrix completion. Information and Inference, 2017.

Theory for two-phase algorithm

Let

$$L_{(1)} \geq L_{(2)} \geq \cdots \geq L_{(n)}, \quad R_{(1)} \geq R_{(2)} \geq \cdots \geq R_{(n)}$$

be row/column leverage scores of rank-r $\mathbf{M}$ in decreasing order.

Theorem

Suppose the Phase 1 sampling probability satisfies

$$p \geq C\tau^{-1}\kappa^4 \log^2(n) \min_{T \in [1:n]} T \left(\sum_{j=1}^{T} L_{(j)}^2 + \sum_{j=1}^{T} R_{(j)}^2 + L_{(T+1)} + R_{(T+1)} \right)$$

Then with probability $\geq 1 - \tau$,

$$\begin{aligned} \frac{1}{3} L_{(i)} \leq \widehat{L}_{(i)} \leq 3\kappa^4 L_{(i)}, \qquad & i \in [1:T] \\ \frac{1}{3} R_{(j)} \leq \widehat{R}_{(j)} \leq 3\kappa^4 R_{(j)}. \qquad & j \in [1:T] \end{aligned}$$

Theory for two-phase algorithm

Let

$$L_{(1)} \geq L_{(2)} \geq \cdots \geq L_{(n)}, \quad R_{(1)} \geq R_{(2)} \geq \cdots \geq R_{(n)}$$

be row/column leverage scores of rank-r $\mathbf{M}$ in decreasing order.

Theorem

Suppose the Phase 1 sampling probability satisfies

$$p \geq C\tau^{-1}\kappa^4 \log^2(n) \min_{T \in [1:n]} T \left(\sum_{j=1}^{T} L_{(j)}^2 + \sum_{j=1}^{T} R_{(j)}^2 + L_{(T+1)} + R_{(T+1)} \right)$$

Then with probability $\geq 1 - \tau$,

$$\begin{aligned} \frac{1}{3} L_{(i)} \leq \widehat{L}_{(i)} \leq 3\kappa^4 L_{(i)}, \qquad & i \in [1:T] \\ \frac{1}{3} R_{(j)} \leq \widehat{R}_{(j)} \leq 3\kappa^4 R_{(j)}. \qquad & j \in [1:T] \end{aligned}$$

- Fewer samples needed for estimating large leverage scores (at

Theory for two-phase algorithm

Theorem

Using

$$C\kappa^4 \log^2(n) n^2 \min_{T \in [1:n]} T \left(\sum_{j=1}^{T} L_{(j)}^2 + \sum_{j=1}^{T} R_{(j)}^2 + L_{(T+1)} + R_{(T+1)} \right)$$

samples for Phase 1, and $C' rn\kappa^2 \log^2(n)$ *samples for Phase 2,* MC^2 *will recover the rank-r matrix* $\mathbf{M}$ *with probability* $\geq 1 - \tau$.

Theory for two-phase algorithm

Theorem

Using

$$C\kappa^4 \log^2(n) n^2 \min_{T\in[1:n]} T \left(\sum_{j=1}^{T} L_{(j)}^2 + \sum_{j=1}^{T} R_{(j)}^2 + L_{(T+1)} + R_{(T+1)} \right)$$

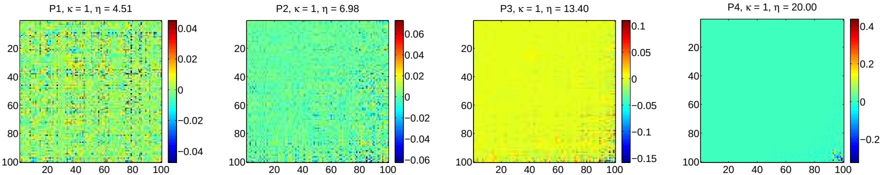
samples for Phase 1, and $C' rn\kappa^2 \log^2(n)$ *samples for Phase 2,* MC^2 *will recover the rank-r matrix* $\mathbf{M}$ *with probability* $\geq 1-\tau$.

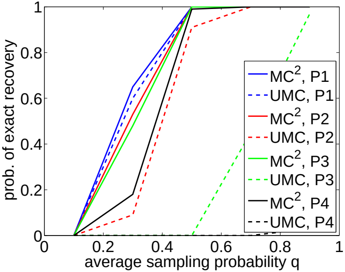
Cases:

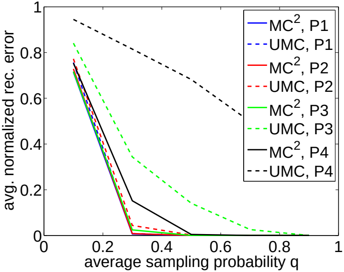
- $L_{(1)} = R_{(1)} = \frac{r}{n}$. Take $m=1$ to recover (up to κ) standard $O(nr\log^2(n))$ result.
- $L_{(1):(T)}, R_{(1):(T)} = \sqrt{\frac{r}{n}}$, and $L_{(T+1)}, R_{(T+1)} = \frac{r}{n}$, and m is small. Need $O(Tnr\log^2(n))$ samples
- $L_{(i)} \leq L_{(1)} i^{3/2}$ for $i = 1:T$ and $L_{(1)} \geq \sqrt{\frac{r}{n}}$. Need $O(r^{2/3} n^{4/3} \log^2(n))$ samples
  - compare to $O(r^{1/2} n^{3/2} \log^2(n))$ samples from uniform sampling

Simulations for two-phase algorithm

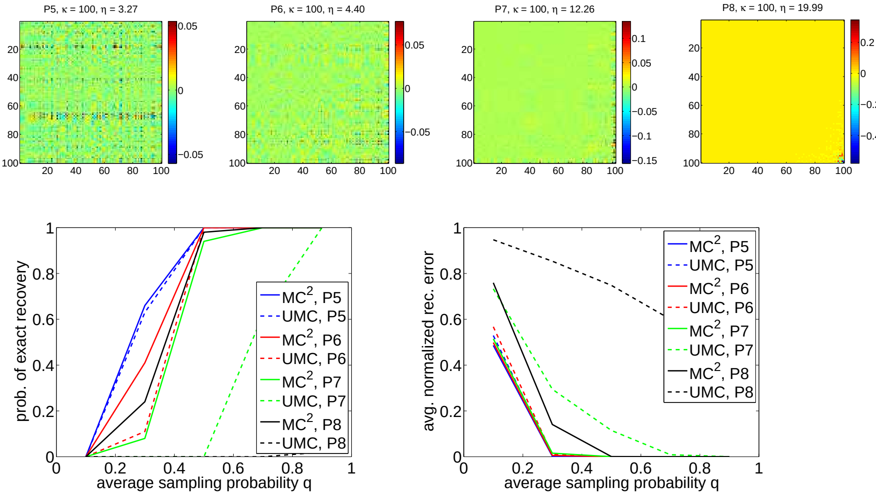
(50%/50% sample split between Phase 1 and Phase 2)



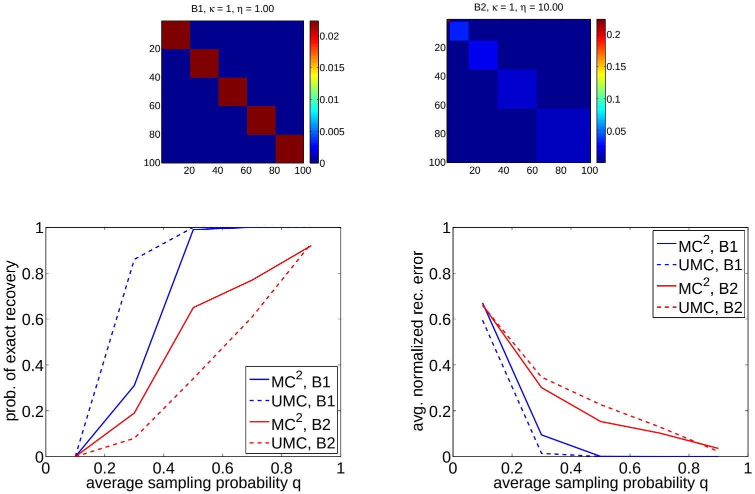




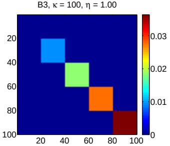
Simulations for two-phase algorithm

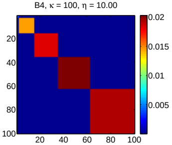


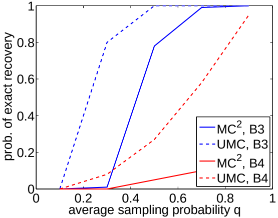
Simulations for two-phase algorithm

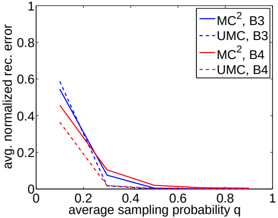


Simulations for two-phase algorithm









Summary

- Any rank-r matrix can be completed from $O(n^2(L_{(1)} + R_{(1)})\log^2(n))$ uniform samples. Good when $L_{(1)}, R_{(1)} \in [\frac{r}{n}, 1]$ are small.
- Any rank-r matrix can be completed from $O(nr\log^2(n))$ samples using a weighted sampling strategy which depends on the row and column leverage scores.
- A two-phase sampling procedure which first samples entries uniformly, then estimates leverage scores and draws a second batch of samples from estimated weighted sampling strategy works well empirically, and has provably better sample complexity compared to uniform sampling for e.g. well-conditioned matrices with power-law decaying leverage scores.

Future directions

- Diminish dependence on condition number in two-phase sampling theory (using different algorithm?).
- Theory for noisy and/or nearly low-rank matrices
- Extensions to other types of “reduced sample complexity if sparse and incoherent” type problems

Future directions

- Diminish dependence on condition number in two-phase sampling theory (using different algorithm?).
- Theory for noisy and/or nearly low-rank matrices
- Extensions to other types of "reduced sample complexity if sparse and incoherent" type problems

Thanks!