Lecture 15: Exact Tensor Completion

Joint Work with David Steurer

Lecture Outline

- Part I: Matrix Completion Problem
- Part II: Matrix Completion via Nuclear Norm Minimization
- Part III: Generalization to Tensor Completion
- Part IV: SOS-symmetry to the Rescue
- Part V: Finding Dual Certificate for Matrix Completion
- Part VI: Open Problems

Part I: Matrix Completion Problem

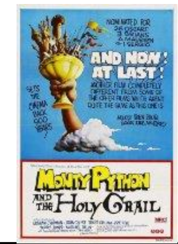
Matrix Completion

- Matrix Completion: Let Ω be a set of entries sampled at random. Given the entries $\{M_{ab}: (a, b) \in \Omega\}$ from a matrix M, can we determine the remaining entries of M?
- Impossible in general, tractable if M is low rank i.e. $M = \sum_{i=1}^{r} \lambda_i u_i v_i^T$ where r is not too large.

Netflix Challenge

- Canonical example of matrix completion: Netflix Challenge
- Can we predict users’ preferences on other movies from their previous ratings?

Netflix Challenge

					
	?	5	6	?	6
	10	8	9	?	?
	?	9	?	6	7.5
	9	?	8	9	?

Solving Matrix Completion

- Current best method in practice: Alternating minimization
- Idea: Write $M = \sum_{i=1}^{r} u_i v_i^T$, alternate between optimizing $\{u_i\}$ and $\{v_i\}$
- Best known theoretical guarantees: Nuclear norm minimization
- This lecture: We'll describe nuclear norm minimization and how it generalizes to tensor completion via SOS.

Part II: Nuclear Norm Minimization

Theorem Statement

- Theorem [Rec11]: If $M = \sum_{i=1}^{r} \lambda_i \, u_i v_i^T$ is an $n \times n$ matrix then nuclear norm minimization requires $O(nr\mu_0(logn)^2)$ random samples to complete M with high probability
- Note: μ_0 is a parameter related to how coherent the $\{u_i\}$ and the $\{v_i\}$ (see appendix for the definition)
- Example of why this is needed: If $u_i = e_j$ then $u_i v_i^T = e_j v_i^T$ can only be fully detected by sampling all of row j, which requires sampling almost everything!

Nuclear Norm

- Recall the singular value decomposition (SVD) of a matrix M
- $M = \sum_{i=1}^{r} \lambda_i u_i v_i^T$ where the $\{u_i\}$ are orthonormal, the $\{v_i\}$ are orthonormal, and $\lambda_i \geq 0$ for all i.
- The nuclear norm of M is $||M||_* = \sum_{i=1}^{r} \lambda_i$

Nuclear Norm Minimization

- Matrix completion problem: Recover M given randomly sampled entries $\{M_{ab}: (a,b) \in \Omega\}$
- Nuclear norm minimization: Find the matrix X which minimizes $\|X\|_*$ while satisfying $X_{ab} = M_{ab}$ whenever $(a,b) \in \Omega$.
- How do we minimize $\|X\|_*$?

Semidefinite Program

- We can implement nuclear norm minimization with the following semidefinite program:
- Minimize the trace of $\begin{pmatrix} U & X \\ X^T & V \end{pmatrix} \succcurlyeq 0$ where $X_{ab} = M_{ab}$ whenever $(a, b) \in \Omega$
- Why does this work? We'll first show that the true solution is a good solution. We'll then describe how to show the true solution is the optimal solution

True Solution

- Program: Minimize the trace of $\begin{pmatrix} U & X \\ X^T & V \end{pmatrix} \succcurlyeq 0$ where $X_{ab} = M_{ab}$ whenever $(a, b) \in \Omega$
- True solution: $\begin{pmatrix} U & X \\ X^T & V \end{pmatrix} = \sum_i \lambda_i \begin{pmatrix} u_i \\ v_i \end{pmatrix} \begin{pmatrix} u_i^T & v_i^T \end{pmatrix}$ (recall that $M = \sum_i \lambda_i u_i v_i^T$)
- Since for all i, $tr\left(u_i u_i^T\right) = tr\left(v_i v_i^T\right) = 1$, $tr \begin{pmatrix} U & X \\ X^T & V \end{pmatrix} = 2 \sum_i \lambda_i$

Dual Certificate

- Program: Minimize the trace of $\begin{pmatrix} U & X \\ X^T & V \end{pmatrix} \succcurlyeq 0$ where $X_{ab} = M_{ab}$ whenever $(a, b) \in \Omega$
- Dual Certificate: $\begin{pmatrix} Id & -A \\ -A^T & Id \end{pmatrix} \succcurlyeq 0$
- Recall that if $M_1, M_2 \succcurlyeq 0$ then $M_1 \bullet M_2 \geq 0$ (where $\bullet$ is the entry-wise dot product)
- $\begin{pmatrix} Id & -A \\ -A^T & Id \end{pmatrix} \bullet \begin{pmatrix} U & X \\ X^T & V \end{pmatrix} \geq 0$
- If $A_{ab} = 0$ whenever $(a, b) \notin \Omega$, this lower bounds the trace.

True Solution Optimality

- Dual Certificate: $\begin{pmatrix} Id & -A \\ -A^T & Id \end{pmatrix} \succcurlyeq 0$ where $A_{ab} = 0$ whenever $(a, b) \notin \Omega$
- True solution $\begin{pmatrix} U & X \\ X^T & V \end{pmatrix} = \sum_i \lambda_i \begin{pmatrix} u_i \\ v_i \end{pmatrix} \begin{pmatrix} u_i^T & v_i^T \end{pmatrix}$ is optimal if $\begin{pmatrix} Id & -A \\ -A^T & Id \end{pmatrix} \bullet \begin{pmatrix} U & X \\ X^T & V \end{pmatrix} = 0$
- This occurs if $\begin{pmatrix} Id & -A \\ -A^T & Id \end{pmatrix} \begin{pmatrix} u_i \\ v_i \end{pmatrix} = 0$ for all i

Conditions on A

- We want A such that $\begin{pmatrix} Id & -A \\ -A^T & Id \end{pmatrix} \succcurlyeq 0$, $A_{ab} = 0$ whenever $(a, b) \notin \Omega$, and $\begin{pmatrix} Id & -A \\ -A^T & Id \end{pmatrix} \begin{pmatrix} u_i \\ v_i \end{pmatrix} = 0$ for all i
- Necessary and sufficient conditions on A:
  1. $\|A\| \leq 1$
  2. $A_{ab} = 0$ whenever $(a, b) \notin \Omega$
  3. $Av_i = u_i$ for all i
  4. $A^T u_i = v_i$ for all i

Dual Certificate with all entries

- Necessary and sufficient conditions on A:
  1. $\|A\| \leq 1$
  2. $A_{ab} = 0$ whenever $(a, b) \notin \Omega$
  3. $Av_i = u_i$ for all i
  4. $A^T u_i = v_i$ for all i
- If we have all entries (so we can ignore condition 2), we can take $A = \sum_i u_i v_i^T$
- Challenge: Find A when we don't have all entries
- Remark: This explains why the semidefinite program minimizes the nuclear norm.

Part III: Generalization to Tensor Completion

Tensor Completion

- Tensor Completion: Let Ω be a set of entries sampled at random. Given the entries $\{T_{abc}: (a,b,c) \in \Omega\}$ from a tensor T, can we determine the remaining entries of T?
- More difficult problem: tensor rank is much more complicated

Exact Tensor Completion Theorem

- Theorem [PS17]: If $T = \sum_{i=1}^{r} \lambda_i u_i \otimes v_i \otimes w_i$, the $\{u_i\}$ are orthogonal, the $\{v_i\}$ are orthogonal, and the $\{w_i\}$ are orthogonal then with high probability we can recover T with $O(r\mu n^{\frac{3}{2}} polylog(n))$ random samples
- First algorithm to obtain exact tensor completion
- Remark: The orthogonality condition is very restrictive but this result can likely be extended.
- See appendix for the definition of μ.

Semidefinite Program: First Attempt

- Won't quite work, but we'll fix it later.
- Minimize the trace of $\begin{pmatrix} U & X \\ X^T & VW \end{pmatrix} \succcurlyeq 0$ where $X_{abc} = T_{abc}$ whenever $(a, b, c) \in \Omega$
- Here the top and left blocks are indexed by a and the bottom and right blocks are indexed by b, c.

True Solution

- Program: Minimize trace of $\begin{pmatrix} U & X \\ X^T & VW \end{pmatrix} \succcurlyeq 0$ where $X_{abc} = T_{abc}$ whenever $(a, b, c) \in \Omega$
- True solution: $\begin{pmatrix} U & X \\ X^T & VW \end{pmatrix} =$

$\sum_i \lambda_i \begin{pmatrix} u_i \\ v_i \otimes w_i \end{pmatrix} \begin{pmatrix} u_i^T & (v_i \otimes w_i)^T \end{pmatrix}$

(recall that $\mathrm{T} = \sum_i \lambda_i \, u_i (v_i \otimes w_i)^T$)

- $tr \begin{pmatrix} U & X \\ X^T & VW \end{pmatrix} = 2 \sum_i \lambda_i$

Dual Certificate: First Attempt

- Program: Minimize trace of $\begin{pmatrix} U & X \\ X^T & VW \end{pmatrix} \succcurlyeq 0$ where $X_{abc} = T_{abc}$ whenever $(a, b, c) \in \Omega$
- Dual Certificate: $\begin{pmatrix} Id & -A \\ -A^T & Id \end{pmatrix} \succcurlyeq 0$ where $A_{abc} = 0$ whenever $(a, b, c) \notin \Omega$
- We want $\begin{pmatrix} Id & -A \\ -A^T & Id \end{pmatrix} \begin{pmatrix} u_i \\ v_i \otimes w_i \end{pmatrix} = 0$ for all i

Conditions on A

- We want A such that $\begin{pmatrix} Id & -A \\ -A^T & Id \end{pmatrix} \succcurlyeq 0$, $A_{abc} = 0$ whenever $(a,b,c) \notin \Omega$, and $\begin{pmatrix} Id & -A \\ -A^T & Id \end{pmatrix} \begin{pmatrix} u_i \\ v_i \otimes w_i \end{pmatrix} = 0$ for all i
- Necessary and sufficient conditions on A:
  1. $\|A\| \leq 1$
  2. $A_{abc} = 0$ whenever $(a,b,c) \notin \Omega$
  3. $A(v_i \otimes w_i) = u_i$ for all i
  4. $A^T u_i = v_i \otimes w_i$ for all i TOO STRONG, requires $\Omega(n^2)$ samples!

Part IV: SOS-symmetry to the Rescue

SOS Program

- Minimize the trace of $\begin{pmatrix} U & X \\ X^T & VW \end{pmatrix} \succcurlyeq 0$ where $X_{abc} = T_{abc}$ whenever $(a, b, c) \in \Omega$ and VW is SOS-symmetric (i.e. $VW_{bcb'c'} = VW_{b'cbc'}$ for all b, c, b', c')

Review: Matrix Polynomial $q(Q)$

- Definition: Given a symmetric matrix Q indexed by monomials, define

$$\mathrm{q}(Q) = \sum_K \left(\sum_{I,J:K=I \cup J (as\ multisets)} Q_{IJ}\right) x_K$$

- Idea: $\mathrm{M} \cdot Q = \tilde{E}[q(Q)]$

Dual Certificate

- Program: Minimize trace of $\begin{pmatrix} U & X \\ X^T & VW \end{pmatrix} \succcurlyeq 0$ where $X_{abc} = T_{abc}$ whenever $(a, b, c) \in \Omega$ and VW is SOS-symmetric
- Dual Certificate: $\begin{pmatrix} Id & -A \\ -A^T & B \end{pmatrix} \succcurlyeq 0$ where $A_{abc} = 0$ whenever $(a, b, c) \notin \Omega$ and $q(B) = q(Id)$
- We want $\begin{pmatrix} Id & -A \\ -A^T & B \end{pmatrix} \begin{pmatrix} u_i \\ v_i \otimes w_i \end{pmatrix} = 0$ for all i

Dual Certificate Tightness Condition

- Write $B = A^T A + Id - R$
- Dual Certificate: $\begin{pmatrix} Id & -A \\ -A^T & A^T A + Id - R \end{pmatrix} \succcurlyeq 0$ where $A_{abc} = 0$ whenever $(a, b, c) \notin \Omega$ and q$(B) = q(Id)$
- This dual certificate is tight for the true solution if

$$\begin{pmatrix} Id & -A \\ -A^T & A^T A + Id - R \end{pmatrix} \begin{pmatrix} u_i \\ v_i \otimes w_i \end{pmatrix} = 0 \text{ for all } i$$

Dual Certificate Conditions

- This gives us the following conditions on A, R
  1. $A_{abc} = 0$ whenever $(a, b, c) \notin \Omega$
  2. $\forall i, A(v_i \otimes w_i) = u_i$
  3. $\|R\| \leq 1$
  4. $\forall i, R(v_i \otimes w_i) = v_i \otimes w_i$
  5. $q(R) = q(A^T A)$ (so that $q(B) = q(Id) = \sum_{b,c} y_b^2 z_c^2$)
- Remark: These conditions are sufficient even if T is not orthogonal. We only prove the theorem for orthogonal tensors because that's what our current analysis can handle.

Part V: Finding Dual Certificate for Matrix Completion

Conditions on A

- Necessary and sufficient conditions on A:
  1. $||A|| \leq 1$
  2. $A_{ab} = 0$ whenever $(a, b) \notin \Omega$
  3. $Av_i = u_i$ for all i
  4. $A^T u_i = v_i$ for all i
- How can we find such an A?
- Idea: Alternate between satisfying condition 2 and conditions 3,4, converging to a final solution.

Definition of $\mathrm{P_U}, \mathrm{P_V}, \mathrm{P_T}$

- Define P_U to be the projection to $span\{u_i\}$. The equation for this is $P_U(x) = \sum_i (x \cdot u_i)u_i$
- Define P_V to be the projection to $span\{v_i\}$. The equation for this is $P_V(y) = \sum_i (y \cdot v_i)v_i$
- Define P_T to be the projection (on the space of matrices) to $span\{xv_i^T, u_i^T y\}$ (for arbitrary x, y). The equation for this is

$$P_T M = P_U M + P_V M - P_U M P_V$$

Restatement of Conditions 3,4

- Necessary and sufficient conditions on A:
  1. $||A|| \leq 1$
  2. $A_{ab} = 0$ whenever $(a, b) \notin \Omega$
  3. $Av_i = u_i$ for all i
  4. $A^T u_i = v_i$ for all i
- Without loss of generality, assume $M = \sum_i u_i v_i^T$ (the values of the λ_i don't affect the dual certificate)
- Assuming $M = \sum_i u_i v_i^T$, conditions 3,4 are equivalent to $P_T A = M$

Definition of R_Ω and $\bar{R}_\Omega$

- Definition: Define $R_\Omega(X) = \frac{n_1 n_2 n_3}{m} X_{abc}$ if $(a, b, c) \in \Omega$ and 0 otherwise where $n_1 \times n_2 \times n_3$ are the dimensions of the tensor and each entry is sampled indepently with probability $\frac{m}{n_1 n_2 n_3}$.
- Define $\bar{R}_\Omega(X) = \left(\frac{n_1 n_2 n_3}{m} - 1\right) X_{abc}$ if $(a, b, c) \in \Omega$ and $-X_{abc}$ if $(a, b, c) \notin \Omega$
- $R_\Omega(X)_{abc} = 0$ whenever $(a, b, c) \notin \Omega$
- $E\left[\bar{R}_\Omega(X)\right] = 0$ (over the choice of Ω)

First Iteration

- Start with M. $\mathrm{P_T}M = M$ but M has nonzero entries outside the sampled entries
- $R_\Omega(M)$ is zero outside the sampled entries, but $P_T R_\Omega(M) \neq M$
- We take $\mathrm{A}_1 = R_\Omega(M)$ as the first approximation, we'll need to correct for the difference

$$P_T R_\Omega M - M = P_T \bar{R}_\Omega M$$

Technical Note

- For the analysis, actually need to resample independently for each iteration, obtaining sets of samples $\Omega_1, \Omega_2, \ldots$. This is the source of the $(logn)^2$ in the upper bound (the lower bound only has $\log n$ (reference to be added))

Iterative Equation

- Take

$$A^k = \sum_{j=0}^{k-1} (-1)^j R_{\Omega_{j+1}} (P_T \bar{R}_{\Omega_j}) \dots (P_T \bar{R}_{\Omega_1}) M$$

- Claim:

$$P_T A^k = M + (-1)^{k-1} (P_T \bar{R}_{\Omega_k}) \dots (P_T \bar{R}_{\Omega_1}) M$$

- Proof idea: Use the facts that $R_\Omega = 1 + \bar{R}_\Omega$, $P_T^2 = P_T$, and $P_T M = M$.

Convergence and Final Step

- Take

$$A^k = \sum_{j=0}^{k-1}(-1)^j R_{\Omega_{j+1}}(P_T\bar{R}_{\Omega_j}) \dots (P_T\bar{R}_{\Omega_1})M$$

- Claim:

$$P_T A^k = M + (-1)^{k-1}(P_T\bar{R}_{\Omega_k}) \dots (P_T\bar{R}_{\Omega_1})M$$

- To show that $P_T A^k$ converges to M w.h.p., it is sufficient to show that the $P_T\bar{R}_\Omega$ operation makes matrices “smaller” with high probability.
- Once the error is small enough, we then take one final step to satisfy all conditions simultaneously. For details, see [Rec11].

Part VI: Open Problems

Open Problems

- For which tensors T can we show that SOS gives exact tensor completion? We've shown it when T is orthogonal, but this can very likely be extended.
- Important subproblem: When can we find A such that $A(v_i \otimes w_i) = u_i$ for all i and $|A(u, v, w)| \leq 1$ for all unit u, v, w?
- Barak and Moitra [BM16] show that SOS solves the approximate tensor completion problem in a somewhat broader setting with a different analysis. Can these analyses assist each other?

References

- [BM16] B. Barak and A. Moitra, Noisy tensor completion via the sum-of-squares hierarchy, COLT, JMLR Workshop and Conference Proceedings, vol. 49, JMLR.org p. 417–445, 2016
- [PS17] A. Potechin and D. Steurer. Exact tensor completion with sum-of-squares. COLT 2017
- [Rec11] B. Recht. A Simpler Approach to Matrix Completion. JMLR Volume 12, p. 3413-3430. 2011

Appendix: μ_0 and μ Definitions

μ_0 and μ Definitions

- Definition:

$$\mu_0 = \frac{n}{r} \cdot \max\{\max_a \|P_U e_a\|^2, \max_b \|P_V e_b\|^2\}$$

- Definition:

$$\mu = \mathrm{n} \cdot \max\{\max_{i,a} u_{ia}^2, \max_{j,b} v_{jb}^2, \max_{k,c} w_{kc}^2\}$$