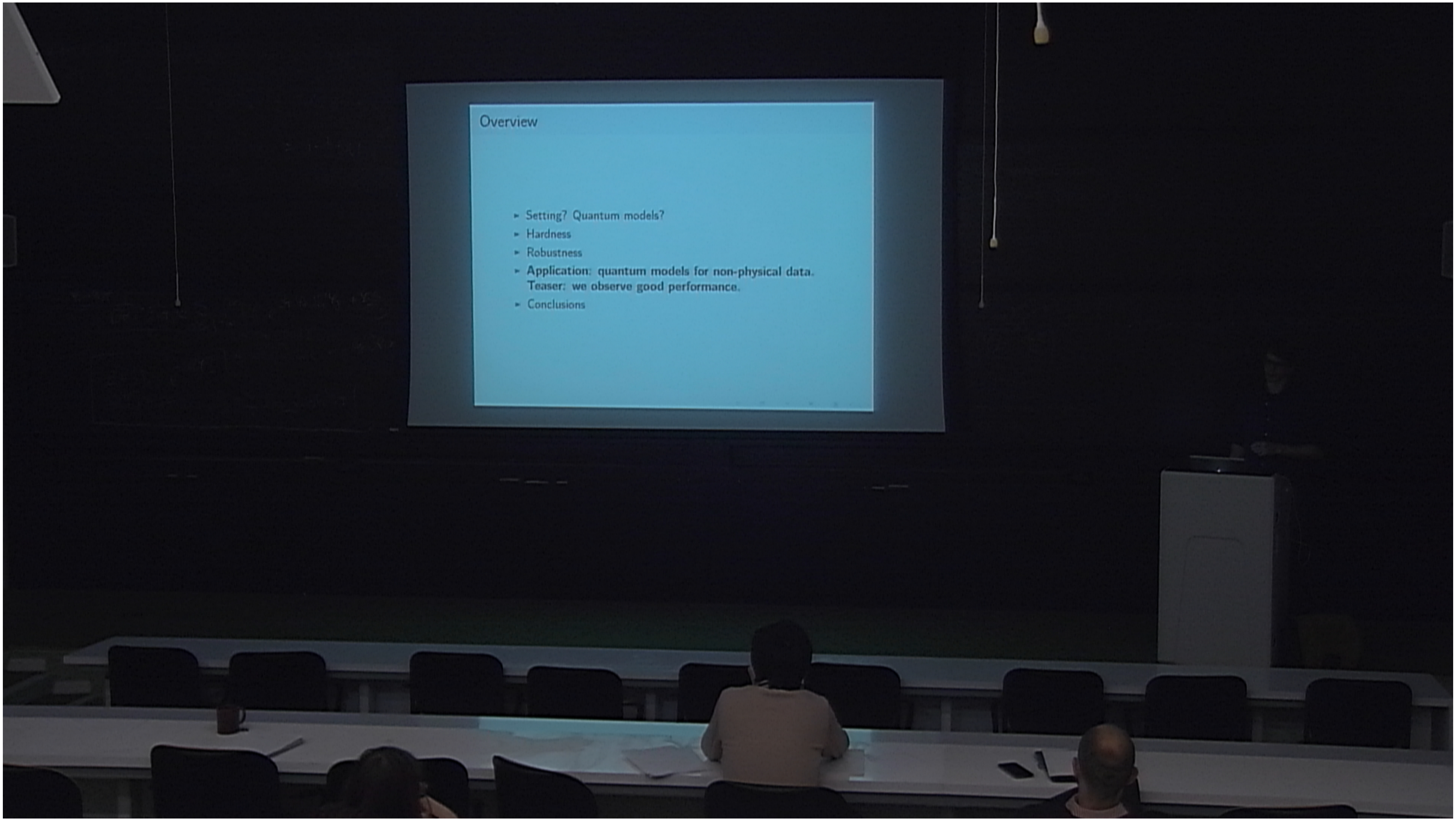
Title: Learning quantum models for physical and non-physical data

Date: Oct 28, 2015 04:15 PM

URL: http://pirsa.org/15100121

Abstract: <p>In this talk I address the problem of simultaneously inferring unknown quantum states and unknown quantum measurements from empirical data. This task goes beyond state tomography because we are not assuming anything about the measurement devices. I am going to talk about the time and sample complexity of the inference of states and measurements, and I am going to talk about the robustness of the minimal Hilbert space dimension. Moreover, I will describe a simple heuristic algorithm (alternating optimization) to fit states and measurements to empirical data. For this algorithm the dataset does not need to be quantum. Hence, the proposed algorithm enables us to interpret general datasets from a quantum perspective. By analyzing movie ratings, we demonstrate the power of quantum models in the context of item recommendation which is a key discipline in machine learning. We observe that quantum models can compete with state-of-the-art algorithms for item recommendation. Based on joint work with Aram Harrow. Relevant preprints: arXiv:1412.7437 and arXiv:1510.02800.</p>

<p> </p>


Overview
Setting? Quantum models?
Hardness
Robustness
Application: quantum models for non-physical data.
Teaser: we observe good performance.
Conclusions

Overview

- Setting? Quantum models?
- Hardness
- Robustness
- **Application: quantum models for non-physical data. Teaser: we observe good performance.**
- Conclusions

Setting?

Consider experiment allowing...

- preparation of states $S_1, \ldots, S_X$,
- performance of Y measurements $(O_{y1}, \ldots, O_{yZ})$, $y \in \{1, \ldots, Y\}$.

S_x, O_{yz} are *instructions*:
docs to prepare state or perform measurement
(incl. temporal and spatial info)

For example:
S_1 : detailed instructions for preparing a s.c. qubit
$(O_{1z})_z$: detailed instructions for readout

Setting?

Consider experiment allowing...

- preparation of states $S_1, \ldots, S_X$,
- performance of Y measurements $(O_{y1}, \ldots, O_{yZ})$, $y \in \{1, \ldots, Y\}$.

S_x, O_{yz} are *instructions*:
docs to prepare state or perform measurement
(incl. temporal and spatial info)

For example:

S_x: detailed instructions for preparing a s.c. qubit
$(O_{yz})_z$: detailed instructions for readout

Quantum models?

We measure

$$p_{xyz} := \mathbb{P}[z|xy]$$

for all $(xyz) \in \Omega \subseteq \{1, ..., X\} \times \{1, ..., Y\} \times \{1, ..., Z\}$

Quantum models?

We measure

$$p_{xyz} := \mathbb{P}[z|xy]$$

for all $(xyz) \in \Omega \subseteq \{1, ..., X\} \times \{1, ..., Y\} \times \{1, ..., Z\}$

To interpret the $(p_{xyz})_{xyz\in\Omega}$ quantumly,

$$S_x \mapsto \rho_x, \quad (O_{yz})_z \mapsto (E_{yz})_z$$

such that

$$p_{xyz} = \mathrm{tr}(\rho_x E_{yz}).$$

$((\rho_x)_x, (E_{yz})_{yz})$: **quantum model for data** $(p_{xyz})_{xyz\in\Omega}$

Trivial solution

Let $(|x\rangle)_{x=1}^{X}$ basis in $\mathbb{C}^X$. If

$$\rho_x = |x\rangle\langle x|,$$

$$E_{yz} = \sum_{x:(xyz)\in\Omega} p_{xyz}|x\rangle\langle x| + \sum_{x:(xyz)\notin\Omega} \delta_{z1}|x\rangle\langle x|$$

then

$$\mathrm{tr}(\rho_x E_{yz}) = p_{xyz}, \quad \forall (xyz) \in \Omega.$$

Perfect fit **but** for outcomes $(xyz) \notin \Omega$

$$\mathbb{P}[z|xy] = \begin{cases} 1, & \text{if } z = 1, \\ 0, & \text{otherwise.} \end{cases}$$

$\Rightarrow$ no predictive power on Ω^c.

Example for overfitting: good fit but poor predictions.

Trivial solution

Let $(|x\rangle)_{x=1}^{X}$ basis in $\mathbb{C}^X$. If

$$\rho_x = |x\rangle\langle x|,$$
$$E_{yz} = \sum_{x:(xyz)\in\Omega} p_{xyz}|x\rangle\langle x| + \sum_{x:(xyz)\notin\Omega} \delta_{z1}|x\rangle\langle x|$$

then

$$\mathrm{tr}(\rho_x E_{yz}) = p_{xyz}, \quad \forall (xyz) \in \Omega.$$

Perfect fit **but** for outcomes $(xyz) \notin \Omega$

$$\mathbb{P}[z|xy] = \begin{cases} 1, & \text{if } z = 1, \\ 0, & \text{otherwise.} \end{cases}$$

$\Rightarrow$ no predictive power on Ω^c.

Example for *overfitting*; good fit but poor predictions.

MinDim

Overfitting because $\#(\text{dof}) \propto X \rightarrow$ *Want to minimize* $\#(\text{dof})$

MinDim:

$$\begin{aligned} \min \quad & d \\ \text{s.t.} \quad & \exists\ d\text{-dimensional states } \rho_x \text{ and measurements } (E_{yz})_z \\ & \text{s.t.} \quad p_{xyz} = \operatorname{tr}(\rho_x E_{yz}) \ \ \forall\, xyz \in \Omega. \end{aligned}$$

Describes data-driven *learning of quantum models*.

Theorem. *MinDim* is *NP*-hard.

Proof: reduction from *3-coloring*. See [CS, arXiv:1510.02800].

MinDim

Overfitting because $\#(\mathrm{dof}) \propto X$ → *Want to minimize* $\#(\mathrm{dof})$

> *MinDim*:
>
> $$\begin{aligned} \min \quad & d \\ \text{s.t.} \quad & \exists\ d\text{-dimensional states } \rho_x \text{ and measurements } (E_{yz})_z \\ & \text{s.t.} \quad p_{xyz} = \mathrm{tr}(\rho_x E_{yz}) \ \ \forall\, xyz \in \Omega. \end{aligned}$$

Describes data-driven *learning of quantum models*.

> **Theorem.** *MinDim* is *NP*-hard.

Proof: reduction from *3-coloring*. See [CS, arXiv:1510.02800].

Trivial solution

Let $(|x\rangle)_{x=1}^{X}$ basis in $\mathbb{C}^X$. If

$$\rho_x = |x\rangle\langle x|,$$
$$E_{yz} = \sum_{x:(xyz)\in\Omega} p_{xyz}|x\rangle\langle x| + \sum_{x:(xyz)\notin\Omega} \delta_{z1}|x\rangle\langle x|$$

then

$$\mathrm{tr}(\rho_x E_{yz}) = p_{xyz}, \quad \forall (xyz) \in \Omega.$$

Perfect fit **but** for outcomes $(xyz) \notin \Omega$

$$\mathbb{P}[z|xy] = \begin{cases} 1, & \text{if } z = 1, \\ 0, & \text{otherwise.} \end{cases}$$

$\Rightarrow$ no predictive power on Ω^c.

Example for *overfitting*; good fit but poor predictions.

MinDim

Overfitting because $\#(\text{dof}) \propto X$ → *Want to minimize* $\#(\text{dof})$

MinDim:

$$\begin{aligned} \min \quad & d \\ \text{s.t.} \quad & \exists\ d\text{-dimensional states } \rho_x \text{ and measurements } (E_{yz})_z \\ & \text{s.t.} \quad p_{xyz} = \operatorname{tr}(\rho_x E_{yz}) \quad \forall\, xyz \in \Omega. \end{aligned}$$

Describes data-driven *learning of quantum models*.

Theorem. *MinDim* is *NP*-hard.

Proof: reduction from *3-coloring*. See [CS, arXiv:1510.02800].

MinDim

Overfitting because $\#(\text{dof}) \propto X \rightarrow$ *Want to minimize* $\#(\text{dof})$

> *MinDim*:
>
> min d
>
> s.t. $\exists$ d-dimensional states ρ_x and measurements $(E_{yz})_z$
>
> s.t. $p_{xyz} = \mathrm{tr}(\rho_x E_{yz}) \;\; \forall xyz \in \Omega$.

Describes data-driven *learning of quantum models*.

> **Theorem.** *MinDim* is *NP*-hard.

Proof: reduction from *3-coloring*. See [CS, arXiv:1510.02800].

$MinDim^{AB}$

A natural 2-party version of *MinDim*:

$MinDim^{AB}$:

min d

s.t. $\exists$ d^2-dimensional state ρ and d-dimensional measurements $(E_{yz})_z$ and $(F_{yz})_z$ satisfying $p_{yzy'z'} = \mathrm{tr}(\rho E_{yz} \otimes F_{y'z'})\ \forall (yzy'z') \in \Omega$

Theorem. $MinDim^{AB}$ is *NP*-hard

Proof: reduction from *3-coloring*. See [CS, arXiv:1510.02800].

Trivial solution

Let $(|x\rangle)_{x=1}^{X}$ basis in $\mathbb{C}^X$. If

$$\rho_x = |x\rangle\langle x|,$$
$$E_{yz} = \sum_{x:(xyz)\in\Omega} p_{xyz}|x\rangle\langle x| + \sum_{x:(xyz)\notin\Omega} \delta_{z1}|x\rangle\langle x|$$

then

$$\mathrm{tr}(\rho_x E_{yz}) = p_{xyz}, \quad \forall (xyz) \in \Omega.$$

Perfect fit **but** for outcomes $(xyz) \notin \Omega$

$$\mathbb{P}[z|xy] = \begin{cases} 1, & \text{if } z = 1, \\ 0, & \text{otherwise.} \end{cases}$$

$\Rightarrow$ no predictive power on Ω^c.

Example for *overfitting*; good fit but poor predictions.

Resulting questions and problems

- **Approximation** algorithms?
- **Robustness of** d under measurement uncertainty?
  ($\rightarrow$ *up next*)
- Worst case analysis $\rightarrow$ Want to illuminate **tradeoff** between
  - relevance of the class $\mathcal{C}$ of $(p_{xyz})_{xyz \in \Omega}$,
  - computational complexity of $\mathcal{C}$.
- **Heuristic** algorithms?
  ($\rightarrow$ some results; e.g., alternating optimization[1])

[1] CS, arXiv:1210.1105 (2012).

Resulting questions and problems

- **Approximation** algorithms?
- **Robustness of** d under measurement uncertainty?
  ($\rightarrow$ *up next*)
- Worst case analysis $\rightarrow$ Want to illuminate **tradeoff** between
  - relevance of the class $\mathcal{C}$ of $(p_{xyz})_{xyz\in\Omega}$,
  - computational complexity of $\mathcal{C}$.
- **Heuristic** algorithms?
  ($\rightarrow$ some results; e.g., alternating optimization[1])

[1] CS, arXiv:1210.1105 (2012).

Resulting questions and problems

- **Approximation** algorithms?
- **Robustness of** d under measurement uncertainty? ($\rightarrow$ *up next*)
- Worst case analysis $\rightarrow$ Want to illuminate **tradeoff** between
  - relevance of the class $\mathcal{C}$ of $(p_{xyz})_{xyz\in\Omega}$,
  - computational complexity of $\mathcal{C}$.
- **Heuristic** algorithms? ($\rightarrow$ some results; e.g., alternating optimization[1])

[1] CS, arXiv:1210.1105 (2012).

Resulting questions and problems

- **Approximation** algorithms?
- **Robustness of** d under measurement uncertainty? (→ *up next*)
- Worst case analysis → Want to illuminate **tradeoff** between
  - relevance of the class $\mathcal{C}$ of $(p_{xyz})_{xyz\in\Omega}$,
  - computational complexity of $\mathcal{C}$.
- **Heuristic** algorithms? (→ some results; e.g., alternating optimization[1])

[1] CS, arXiv:1210.1105 (2012).

Related work

- Related to *MinDim*, *MinDim*^{AB}:
  - N. Harrigan, T. Rudolph, and S. Aaronson (2007)
  - S. Wehner, M. Christandl, and A. Doherty (2008)
  - R. Gallego, N. Brunner, C. Hadley, and A. Acin (2010)
  - M. Navascues, A. Felix, M. Araujo, and T. Vertesi (2015)
- *MinDim* with application of sequences of gates:
  - S.T. Merkel et al (2012)
  - R. Blume-Kohout et al (2013)
  - S. Kimmel et al (2014, 2015)
  - A. Dugas, J. Wallman, and J. Emerson (2015)
  - A. Monras and A. Winter (2014)
  - C. Jackson and S.J. van Enk (2015)
- Related to extension complexity:
  - S. Fiorini et al (2012)
  - J. Gouveia, P. Parrilo, and R. Thomas (2013)
  - H. Fawzi et al (2014)
  - Z. Ji (2013)
  - S. Burgdorf, M. Laurent, and T. Piovesan (2015)

Related work

- Related to *MinDim*, $MinDim^{AB}$:
  - N. Harrigan, T. Rudolph, and S. Aaronson (2007)
  - S. Wehner, M. Christandl, and A. Doherty (2008)
  - R. Gallego, N. Brunner, C. Hadley, and A. Acin (2010)
  - M. Navascues, A. Felix, M. Araujo, and T. Vertesi (2015)
- *MinDim* with application of sequences of gates:
  - S.T. Merkel et al (2012)
  - R. Blume-Kohout et al (2013)
  - S. Kimmel et al (2014, 2015)
  - A. Dugas, J. Wallman, and J. Emerson (2015)
  - A. Monras and A. Winter (2014)
  - C. Jackson and S.J. van Enk (2015)
- Related to extension complexity:
  - S. Fiorini et al (2012)
  - J. Gouveia, P. Parrilo, and R. Thomas (2013)
  - H. Fawzi et al (2014)
  - Z. Ji (2013)
  - S. Burgdorf, M. Laurent, and T. Piovesan (2015)

Related work

- Related to *MinDim*, $MinDim^{AB}$:
  - N. Harrigan, T. Rudolph, and S. Aaronson (2007)
  - S. Wehner, M. Christandl, and A. Doherty (2008)
  - R. Gallego, N. Brunner, C. Hadley, and A. Acin (2010)
  - M. Navascues, A. Felix, M. Araujo, and T. Vertesi (2015)
- *MinDim* with application of sequences of gates:
  - S.T. Merkel et al (2012)
  - R. Blume-Kohout et al (2013)
  - S. Kimmel et al (2014, 2015)
  - A. Dugas, J. Wallman, and J. Emerson (2015)
  - A. Monras and A. Winter (2014)
  - C. Jackson and S.J. van Enk (2015)
- Related to extension complexity:
  - S. Fiorini et al (2012)
  - J. Gouveia, P. Parrilo, and R. Thomas (2013)
  - H. Fawzi et al (2014)
  - Z. Ji (2013)
  - S. Burgdorf, M. Laurent, and T. Piovesan (2015)

Robustness

- Often: ∃ low dimensional solutions
- May suggest d not robust
- I.e.,

$$\forall\ (\mathbb{C}^D, (\rho_x)_x, (E_{yz})_{yz})$$
$$\exists\ (\mathbb{C}^d, (\rho'_x)_x, (E'_{yz})_{yz}) \text{ with } d \ll D$$
$$\text{s.t. } \mathrm{tr}(\rho'_x E'_{yz}) \approx \mathrm{tr}(\rho_x E_{yz}).$$

- *True?*

Incompressibility

- $\mathrm{tr}(\rho_x E_{yz}) \le \|\rho_x\| \|E_{yz}\|_1 \le \mathrm{tr}(E_{yz}) \quad \forall x$

  $\max_x\{\mathrm{tr}(\rho_x E_{yz})\} \le \mathrm{tr}(E_{yz})$

  $\sum_{yz} \max_x\{\mathrm{tr}(\rho_x E_{yz})\} \le \sum_{yz} \mathrm{tr}\Big(E_{yz}\Big) = \mathrm{tr}(I) = d$

  [Lee, Wei, de Wolf 2014], [CS, Harrow 2014]

  $\max_x\{\mathrm{tr}(\rho_x E_{yz})\} = 1$ possible for all z

  $Z \le d$

  In separate result about compression of "psd factorizations":
  error $\sim \mathrm{tr}(E_{yz}) \sim \mathrm{rank}(E_{yz})$

Incompressibility

- $\mathrm{tr}(\rho_x E_{yz}) \leq \|\rho_x\| \|E_{yz}\|_1 \leq \mathrm{tr}(E_{yz}) \quad \forall x$
- $\max_x\{\mathrm{tr}(\rho_x E_{yz})\} \leq \mathrm{tr}(E_{yz})$
- $\sum_z \max_x\{\mathrm{tr}(\rho_x E_{yz})\} \leq \sum_z \mathrm{tr}\Big(E_{yz}\Big) = \mathrm{tr}(I) = d$

[Lee, Wei, de Wolf, 2014], [CS, Harrow, 2014]

$\max_x\{\mathrm{tr}(\rho_x E_{yz})\} = 1$ possible for all z

$Z \leq d$

In separate result about compression of 'psd factorizations':
error $\sim \mathrm{tr}(E_{yz}) \sim \mathrm{rank}(E_{yz})$

Compression of quantum models

Hence, large Z and large $\mathrm{rank}(E_{yz})$ appear to lead to robustness.

This motivates consideration of ...

Pseudo low-rank: $(\mathbb{C}^D, (\rho_x)_x, (E_{yz})_{yz})$ *pseudo low-rank* if

- Z small, i.e., $Z \sim \mathcal{O}(1)$ in D
- $\underbrace{E_{y1}, \ldots, E_{y,z'-1}}_{\text{low-rk}}, E_{yz'}, \underbrace{E_{y,z'+1}, \ldots, E_{y,Z}}_{\text{low-rk}}$

Compression of quantum models

Theorem. Let

$$J := X + YZ,$$
$$d = \mathcal{O}\Big(\tfrac{1}{\varepsilon^2}\ \ln(4JD)\Big),$$
$$(\mathbb{C}^D, (\rho_x)_x, (E_{yz})_{yz})$$ *pseudo low-rank.*

Then, $\exists\ (\rho'_x)_x\ (E'_{yz})_{yz}$ d-dimensional s.t.

$$\big|\mathrm{tr}(\rho_x E_{yz}) - \mathrm{tr}(\rho'_x E'_{yz})\big| = \mathcal{O}(\varepsilon)$$

- Proof: analysis of explicit compression scheme; see [CS Harrow 2014]
- Generalization of [Winter, quant-ph/0401060] (specific to rk-1 measurements; quadratic compression)
- Implications in *learning*, *communication complexity* and *extension complexity*

Overview

- Quantum models?
- Hardness
- Robustness
- **Application: quantum models for non-physical data.**
- Conclusions

Movie recommendation

Given:

	Jurassic Park	*Despicable Me*	*Airplane!*	*Notting Hill*
Alice	⋆	⋆⋆⋆⋆⋆	⋆⋆⋆⋆	⋆⋆⋆
Bob	⋆⋆⋆⋆	⋆⋆⋆⋆	?	⋆
Charlie	?	⋆⋆⋆	⋆⋆⋆	⋆⋆⋆
Donald	?	?	?	⋆⋆⋆⋆⋆

Algorithms for analysis: "*recommender systems*"

Good recommender systems:

- *predictive power* (despite very sparse and noisy data)
- *computational tractability* (despite quickly growing numbers of users and items)
- *interpretability* (to allow for feedback, market analysis and visual representations)

Movie recommendation

Given:

	Jurassic Park	*Despicable Me*	*Airplane!*	*Notting Hill*
Alice	⋆	⋆⋆⋆⋆⋆	⋆⋆⋆⋆	⋆⋆⋆
Bob	⋆⋆⋆⋆	⋆⋆⋆⋆	?	⋆
Charlie	?	⋆⋆⋆	⋆⋆⋆	⋆⋆⋆
Donald	?	?	?	⋆⋆⋆⋆⋆

- Algorithms for analysis: "*recommender systems*"
- *Good* recommender systems:
  - *predictive power* (despite very sparse and noisy data)
  - *computational tractability* (despite quickly growing numbers of users and items)
  - *interpretability* (to allow for feedback, market analysis and visual representations)

Movie recommendation: quantum models

Adopt q-models through "*system-state-measurement paradigm*"

- System: abstract part of thinking determining movie taste
- State: individual preference (i.e., manifestation of the system)
- Measurement: "Do you like item i?"

System-state-measurement paradigm through quantum models:

$$\begin{aligned} \min \quad & d \\ \text{s.t.} \quad & \exists\ d\text{-dimensional states } \rho_u \text{ and measurements } (E_{iz})_z \\ & \text{s.t. } \delta_{z,R_{ui}} \approx \mathrm{tr}(\rho_u E_{iz}) \quad \forall\, uiz \in \Omega. \end{aligned}$$

Q: Does there even exist a low-dimensional quantum model fitting real-world data?

Movie recommendation: quantum models

Adopt q-models through "*system-state-measurement paradigm*"

- System: abstract part of thinking determining movie taste
- State: individual preference (i.e., manifestation of the system)
- Measurement: "Do you like item i?"

System-state-measurement paradigm through quantum models:

$$\begin{aligned} &\min \quad d \\ &\text{s.t.} \quad \exists\ d\text{-dimensional states } \rho_u \text{ and measurements } (E_{iz})_z \\ &\qquad\quad \text{s.t. } \delta_{z,R_{ui}} \approx \operatorname{tr}(\rho_u E_{iz}) \quad \forall\, uiz \in \Omega. \end{aligned}$$

Q: Does there even exist a low-dimensional quantum model fitting real-world data?

∃? quantum model

Observation: real-world data is *sparse*.

Dataset	items	U	I	Density
Douban	movies	129,490	58,541	0.22%
Epinions (665K)	cars, books, ...	40,163	139,738	0.012%
Flixster	movies	147,612	48,794	0.11%
FilmTrust	movies	1,508	2,071	1.14%
MovieLens 1M	movies	6,040	3,706	4.47%
Netflix	movies	480,189	17,770	1.18%

Q: Low-dimensional quantum model fitting **sparse** data?

∃? quantum model

Theorem. Let $J = U + IZ$. Assume number of ratings per item is constant in U. Then, for

$$d = \mathcal{O}\Big(\frac{1}{\varepsilon^2}\ \mathrm{ln}(4JU)\Big),$$

there exists a d-dimensional quantum model $((\rho_x)_x\ (E_{yz})_{yz})$ satisfying

$$\big|\delta_{z,R_{ui}} - \mathrm{tr}(\rho_u E_{iz})\big| = \mathcal{O}(\varepsilon)$$

for all $(u,i) \in \Omega$ and for every $z \in \{1, ..., Z\}$.

∃? quantum model

Theorem. Let $J = U + IZ$. Assume number of ratings per item is constant in U. Then, for

$$d = \mathcal{O}\left(\frac{1}{\varepsilon^2} \ln(4JU)\right),$$

there exists a d-dimensional quantum model $((\rho_x)_x \ (E_{yz})_{yz})$ satisfying

$$\left|\delta_{z,R_{ui}} - \mathrm{tr}(\rho_u E_{iz})\right| = \mathcal{O}(\varepsilon)$$

for all $(u, i) \in \Omega$ and for every $z \in \{1, ..., Z\}$.

- Proof: compression of trivial quantum model.
- *No predicitve power but great fit* → overfitting.
- Hence, *there exists a quantum model* but it is *easy to overfit sparse data quantumly*.

∃? quantum model

Theorem. Let $J = U + IZ$. Assume number of ratings per item is constant in U. Then, for

$$d = \mathcal{O}\Big(\frac{1}{\varepsilon^2} \ln(4JU)\Big),$$

there exists a d-dimensional quantum model $((\rho_x)_x\ (E_{yz})_{yz})$ satisfying

$$\left|\delta_{z,R_{ui}} - \mathrm{tr}(\rho_u E_{iz})\right| = \mathcal{O}(\varepsilon)$$

for all $(u, i) \in \Omega$ and for every $z \in \{1, ..., Z\}$.

- Proof: compression of trivial quantum model.
- *No predicitve power but great fit* $\rightarrow$ overfitting.
- Hence, *there exists a quantum model* but it is *easy to overfit sparse data quantumly*.

Why quantum?

Q: Why quantum? *In [CS, December 2015]:*

- high predictive power on real-life datasets
- Practical relaxation of compressed classical models
- extraction of hierarchical orderings of properties

Next: *high predictive power.*

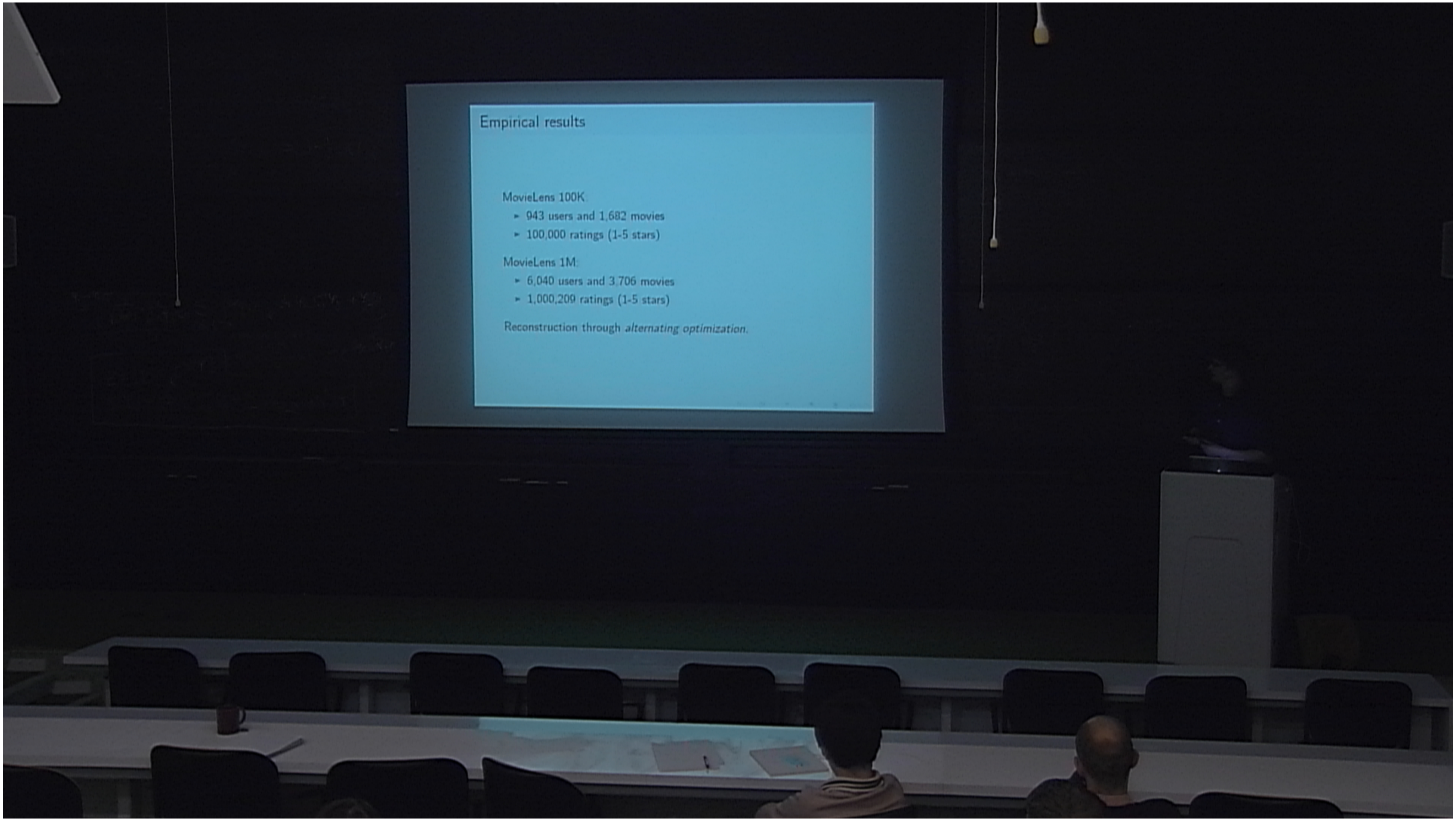
Evaluation of predictive power

- **Claim**: q-models are good at predicting unknown entries.
- **Q**: How do we decide "model A is more predicitve than model B"?

Solution:

1. decide on error measure $\mathrm{dist}(\text{prediction}, \text{ground truth})$
2. split Ω; $\Omega = \Omega_{\mathrm{train}} \dot\cup \Omega_{\mathrm{test}}$
3. use $(p_{uiz})_{(uiz)\in\Omega_{\mathrm{train}}}$ to compute models A and B.
4. Model A is better than B if

$$\mathrm{dist}(\text{prediction}_A, \text{truth}) < \mathrm{dist}(\text{prediction}_B, \text{truth})$$


Empirical results
MovieLens 100K:
943 users and 1,682 movies
100,000 ratings (1-5 stars)
MovieLens 1M:
6,040 users and 3,706 movies
1,000,209 ratings (1-5 stars)
Reconstruction through alternating optimization.

Empirical results

	100K MAE	*100K* RMSE	*1M* MAE	*1M* RMSE
UserKNN[2]	0.737	0.944	0.703	0.905
ItemKNN[3]	0.723	0.924	0.688	0.876
NMF[4]	0.752	0.955	0.727	0.920
SVD++[5]	0.719	**0.912**	0.668	**0.851**
Quantum	**0.703**	0.994	**0.641**	0.917

[2] Resnick et al 1994
[3] Rendle et al 2009
[4] Lee and Seung 2001
[5] Koren, 2008

Empirical results

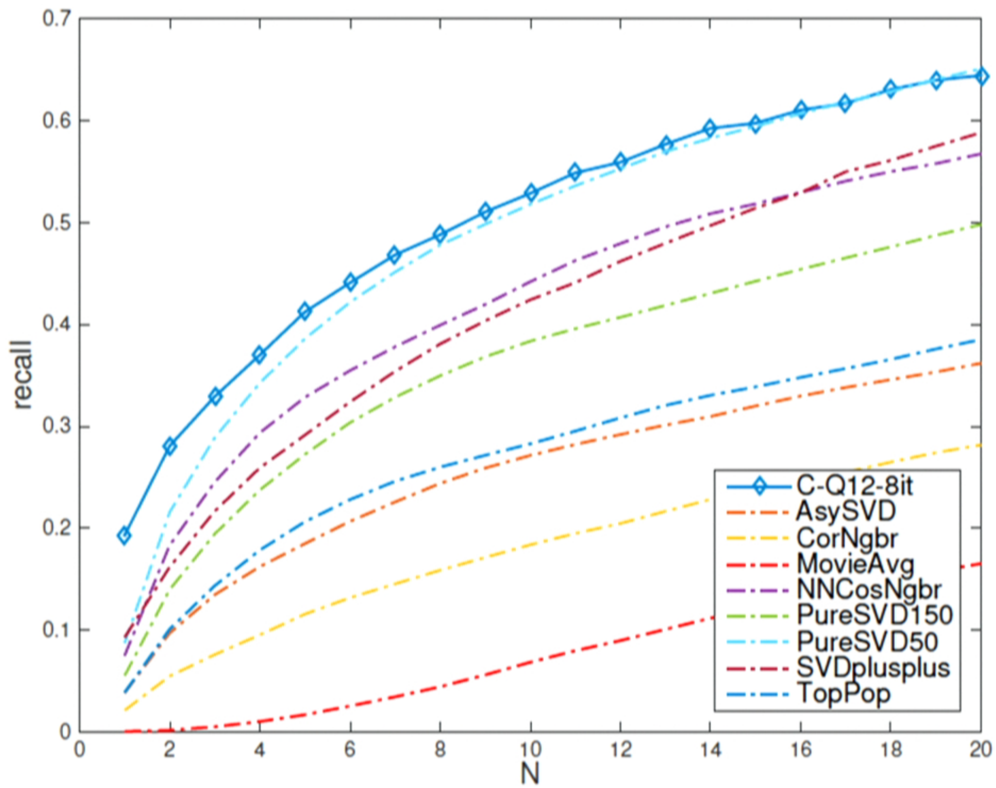


Figure: Recall at N. C-Q12-8it is a 12-dimensional quantum model.

Conclusions

- Considered problem of learning of quantum models
- *NP*-hard.
- For some data, d is robust. But pseudo-low rank models admit exponential compression.

  Demonstration of power of quantum models for non-physical data in the context of item recommendation.