

機械学習を用いた量子状態異常検知 *

IBM 東京基礎研究所 原 聡 †

Satoshi Hara

IBM Research – Tokyo

概要

精緻な処理が必要となる量子情報処理において量子系の状態をあらわす密度行列の異常の検知は重要な課題である。本稿では機械学習を用いた密度行列の異常検知手法 ED^3 を紹介する。観測された密度行列は一般に統計的なゆらぎを含むが、 ED^3 を用いることでそのようなゆらぎから異常な変化を抽出することができる。シミュレーション及び実験データにおいて、 ED^3 が平均行列を使った単純な手法よりも高い精度を達成することを確認した。 ED^3 は密度行列の異常が問題となる多くの局面において有用な解決策となると考えている。

1 はじめに

精緻な処理が必要となる量子情報処理では所望の量子状態からのずれ、すなわち異常の検出が重要な課題である。しかしながら、異常の検知は量子系が持つ本質的なゆらぎのために困難な課題の一つでもある。通常、量子系の状態を表現する密度行列は量子トモグラフィー [2] を用いて推定される。この際、実験で得られる観測の総数は限られているために推定された密度行列には統計的なゆらぎが生じる。この統計的なゆらぎと量子系の変化とを区別することが異常検知における技術的な課題である。

本稿では機械学習を用いて密度行列の異常検知を行う方法 ED^3 (Erroneous Deviation Detector for Density matrices) [1] を紹介する。 ED^3 は密度行列の各要素の絶対値が変化するような量子系の異常を検知する手法である。このような異常にはデコヒーレンスなど実用上重要な例が数多く含まれる。量子もつれ光子対の量子トモグラフィーにおいて ED^3 と単純な手法とを比較したところ、シミュレーション・実験データの両方で異常検知性能の大きな向上が見られた。 ED^3 は密度行列を扱う問題には全て適用できるため、量子光学系に限らず超伝

* 本研究はブリストル大学の小野貴史氏、京都大学の竹内繁樹氏、岡本亮氏、及び大阪大学の鷲尾隆氏との共同研究である [1].

† satohara@jp.ibm.com

導回路やイオントラップを用いた系にも適用できる。このため、ED³ は密度行列の異常が問題となる多くの局面において有用な解決策となると考えている。

2 密度行列の異常検知手法

量子トモグラフィーにより K 個の観測密度行列 $\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_K$ が得られているとする。以下では k 番目の観測密度行列の (i, j) 要素の絶対値を $|\hat{\rho}_{k,ij}|$ とし、これら絶対値を要素として持つ観測密度行列を $|\hat{\rho}_k| \in \mathbb{R}^{d \times d}$ と表記する。本稿では、これら K 個の絶対値の観測密度行列 $|\hat{\rho}_1|, |\hat{\rho}_2|, \dots, |\hat{\rho}_K|$ から異常な行列のインデックス集合 $S \subseteq \{1, 2, \dots, K\}$ を探す問題を対象とする。ただし、異常行列のインデックス集合 S の補集合 S^c においては行列 $|\hat{\rho}_{k'}| (k' \in S^c)$ は全て正常であるとし、また異常行列は K 個の行列のごく一部、つまり $|S| \ll K$ であると仮定する。このような問題は機械学習において異常検知や外れ値検知として知られている [3]。

異常検知を最も直接的に実現する方法は、計測された密度行列を”正しい”密度行列と比較することである。量子トモグラフィー装置を物理モデルに基づいて表現することができれば、その物理モデルから”正しい”密度行列を計算することができる。そしてその”正しい”密度行列と各観測密度行列がどの程度異なるかを評価することで異常な行列を見つけることができる。このためには量子トモグラフィー装置内の光路に関する正確なパラメータを用いて物理モデルを計算する必要がある。しかし、現実的には実験装置の正確なパラメータを知ることは困難であり、その近似値を用いざるをえない。つまり、近似的なパラメータから計算された”正しい”密度行列は、量子トモグラフィーにより実際に得られる観測密度行列とは必ずしも一致しない。そのため、以下では物理モデルを仮定せず、統計的な手法を用いて異常検知を行うことを考える。

2.1 平均行列を使う手法（平均値法）

物理モデルを用いずに正しい密度行列を観測密度行列から近似的に表現することを考える。最も単純な方法の一つとして K 個の観測密度行列の平均値を使う方法が考えられる。

$$|\bar{\rho}| = \frac{1}{K} \sum_{k=1}^K |\hat{\rho}_k|.$$

K 個の観測密度行列のうち大半は正常な行列だと仮定しているため、この平均行列は正しい密度行列の近似値として使うことができる。そして、各行列のこの平均行列からの偏差 $\tilde{\omega}_k = |\hat{\rho}_k| - |\bar{\rho}|$ に基づいて、異常の度合いを評価する。ここでは量子力学的な差異を測る一般的な方法としてトレース距離 [4] を用いる。

$$\tilde{e}_k = \|\tilde{\omega}_k\|_{\text{tr}}. \quad (1)$$

ここで、 $\|\cdot\|_{\text{tr}}$ は行列の特異値の和をあらわしておりトレースノルムと呼ばれる。計算された $\tilde{e}_k$ が一定値以上大きければ、密度行列は正しい密度行列からの乖離が大きい、つまり異常だと判定できる。

平均行列を使う方法は単純で直感的にもわかりやすいが、問題点として平均行列 $|\bar{\rho}_k|$ の計算に数が少ないとは言え異常な行列が混入して偏りを生じさせていることがあげられる。この偏りが異常検知の性能を低下させてしまう要因となる。

2.2 機械学習を用いた異常検知手法 ED³ [1]

k 番目の絶対値の密度行列の真の値を $|\rho_k|$ とすると、これは”正しい”密度行列 θ とそこからの偏差 ω_k を使って $|\rho_k| = \theta + \omega_k$ と表現できる。また、観測密度行列 $|\hat{\rho}_k|$ はこの $|\rho_k|$ に観測に伴う統計的ゆらぎが加わったものとして解釈できる。このとき、正常な行列については”正しい”密度行列 θ からの偏差については $\omega_k \approx 0$ と想定できる。他方、異常な行列については $\omega_k \neq 0$ となる。そのため、偏差の行列の集合 $\{\omega_1, \omega_2, \dots, \omega_K\}$ は多くが零行列であり、ごく一部に非零行列を含むスパースな構造を持つと仮定できる。このようにデータの背後にスパースな構造が隠れている場合には機械学習のスパース推定の技術が有効である [5, 6, 7]。そこで、ED³ では以下の正則化付きの最小二乗問題を解くことで、このスパースな偏差行列の集合を推定する。

$$\min_{\theta, \{\omega_k\}_{k=1}^K} \frac{1}{2} \sum_{k=1}^K \|\hat{\rho}_k - (\theta + \omega_k)\|_F^2 + \gamma \sum_{k=1}^K \|s \odot \omega_k\|_F. \quad (2)$$

ここで $\|\cdot\|_F$ は行列のフロベニウスノルム、 $\odot$ はアダマール積、 γ は非負のパラメータである。また、行列 $s \in \mathbb{R}^{d \times d}$ は

$$s_{ij} = \left(\frac{1}{K} \sum_{k=1}^K \tilde{\omega}_{k,ij}^2 \right)^{-1/2}$$

により定義される。式 (2) の第一項はなるべく各観測密度行列 $|\hat{\rho}_k|$ に近い $\theta + \omega_k$ を推定すること、第二項は行列の各要素の大きさを正規化したうえでできる限りスパースでありゼロに近い偏差行列の集合を推定することに対応する。最後に問題 (2) を解くことによって得られる ω_k から、式 (1) と同じくトレース距離 $e_k = \|\omega_k\|_{\text{tr}}$ を計算し、 e_k が一定値以上大きければ異常と判断する。式 (2) による推定では、 θ と ω_k の同時推定により平均行列を使う場合に比べて”正しい”密度行列 θ の偏りが抑えられること、スパースな偏差行列を推定できることの2点により、精度の高い異常検知が期待できる。

問題 (2) の各項は凸関数なので全体としても凸関数であり、各種凸最適化アルゴリズムによって $\theta, \{\omega_k\}_{k=1}^K$ の大域最適解を得ることができる。本研究では最適化手法として ADMM

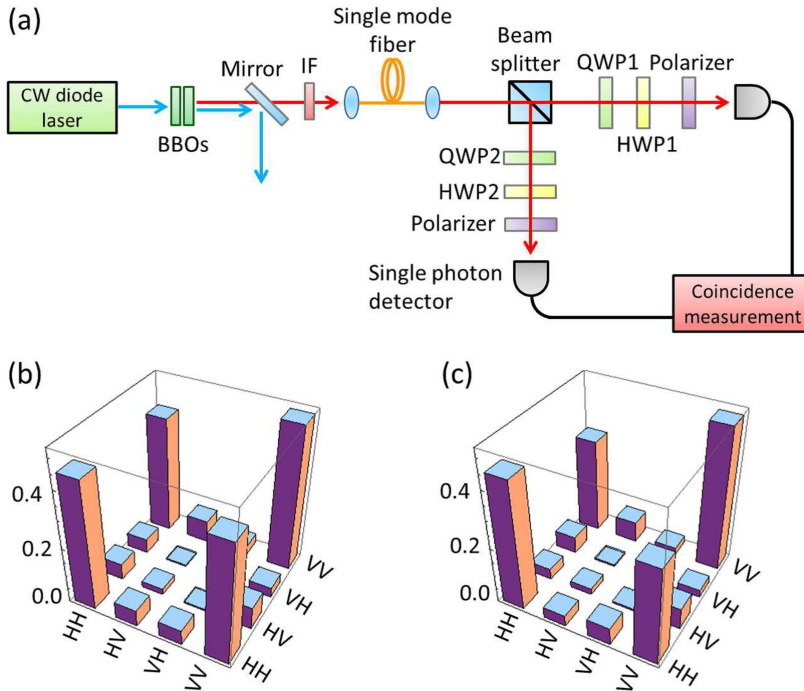


図 1 (a) 量子もつれ光子対の量子トモグラフィー装置, (b) 正常な密度行列の各要素の平均のヒストグラム, (c) 異常な密度行列の各要素の平均のヒストグラム

(Alternating Direction Method of Multipliers) [8] を用いた。このアルゴリズムは問題 (2) の構造を活用することで効率的に大域最適解を求めることができる。

3 評価実験

ED³ と平均値法の性能比較実験を行った。実験では図 1(a) に概要を示した量子トモグラフィー装置を用い、1 個の観測密度行列を 1000 回の量子もつれ光子対入射の計数から得た。観測を繰り返すことで、図 1(b) の正常な密度行列及び図 1(c) の異常な密度行列をそれぞれ複数生成した。なお、各密度行列を計測した前後に装置のパラメータが変動していないことを可能な限り確認して、極力所望の密度行列が計測されるようにした。異常な密度行列では (1, 4) 及び (4, 1) の非対角要素 (図中 (HH, VV) に相当) の値が 0.421 から 0.339 まで低下している。この異常状態は正常な場合よりも 2 光子間の量子もつれの程度が少ない状態である。本評価実験では、実際の実験データ及び図 1(a) の実験系を模した計算機シミュレーションによるデータの 2 種類を用意した。

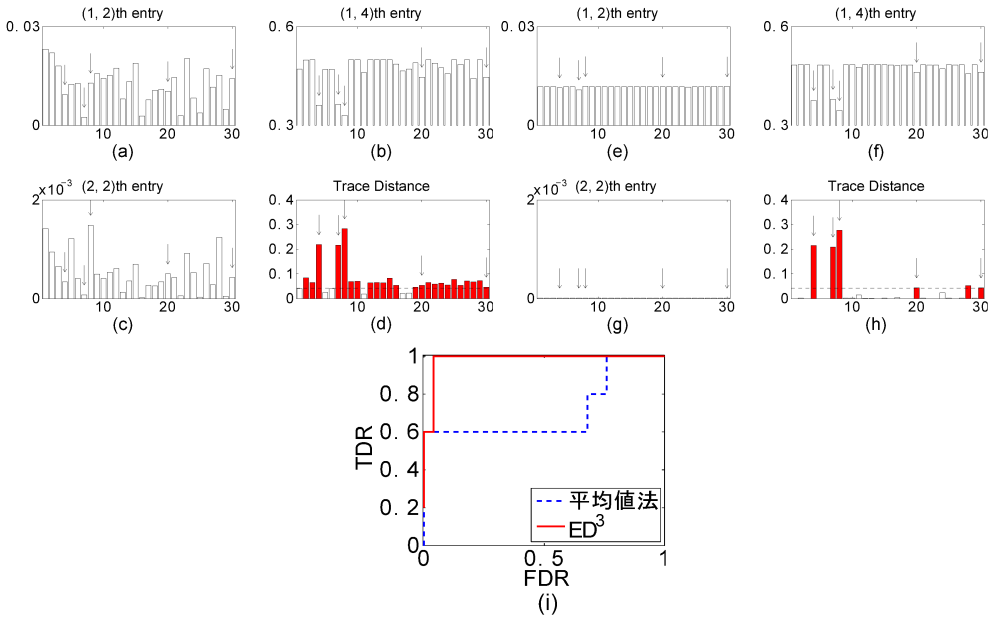


図2 シミュレーションデータの結果一例：(a)–(d) 観測行列の (1, 2), (1, 4) 及び (2, 2) 要素と平均値法により計算されたトレース距離（図中の矢印は異常行列），(e)–(h) ED^3 により計算された行列の (1, 2), (1, 4) 及び (2, 2) 要素とトレース距離（図中の矢印は異常行列），(i) 平均値法と ED^3 の ROC 曲線

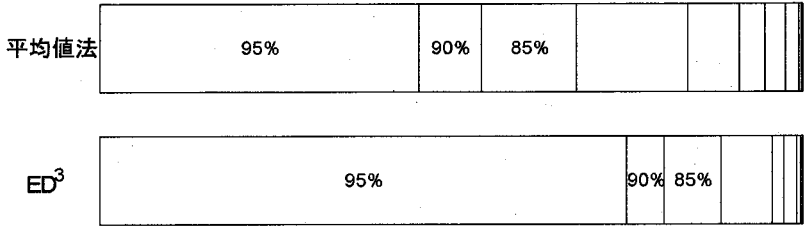


図3 シミュレーションデータでの 1000 回の実験における AUC の分布

3.1 結果：シミュレーションデータ

シミュレーションデータでは正常な密度行列を 25000 個，異常な密度行列を 5000 個生成した．これらを 1000 個の組にわけ，それぞれ正常行列 25 個，異常行列 5 個の計 30 個を 1 つのデータセット ($K = 30$) とした．これら 1000 個のデータセットそれぞれで平均値法と ED^3 の性能を比較した．なお， ED^3 のパラメータ γ は事前に 10 個のデータセットで複数の γ の値を試し，その結果が一番良かった値を 1000 個のデータセット実験で使った．

図 2 にシミュレーションデータ及びその結果の一例を示す。図 2(a)–(c) はある 1 つのデータセットの 30 個の観測行列の (1, 2), (1, 4) 及び (2, 2) 要素である。また、図 2(e)–(g) は ED^3 で推定された行列 $\theta + \omega_k$ の (1, 2), (1, 4) 及び (2, 2) 要素である。矢印がついた項目が 5 個の異常行列をあらわしている。観測行列の要素 (図 2(a)–(c)) と ED^3 により推定された値 (図 2(e)–(g)) とを比較すると、元の行列にはあった統計的なゆらぎが ED^3 により抑えられ小数の変動だけが残っていることがわかる。特に (1, 4) 要素でこの傾向は顕著であり、矢印のついた異常行列を除き、他の正常行列では要素がほぼ一定の値になっている。つまり、 ED^3 により確かに期待した通りのスパースな偏差行列が得られていることがわかる。図 2(d), (h) はそれぞれ平均値法、 ED^3 から計算されたトレース距離である。トレース距離が 0.04 (破線) を超える行列は赤で強調表示してある。図 2(d) では 0.04 前後の値を取っていた正常行列の多くが図 2(h) では 0 になっている。他方、図 2(d) では他の行列に埋もれていた 2 つの異常行列が図 2(h) では埋もれずに非零のトレース距離として抽出されている。これは平均値法では検知できなかった 2 つの異常が ED^3 ではきちんと検知できたことを示している。図 2(i) は 2 つの手法それぞれについて ROC 曲線を描画したものである。 ED^3 のほうが曲線がより左上にあり、平均値法よりも高い異常検知精度を達成できていることがわかる。

1000 個のデータセットについて ROC 曲線下の面積 AUC (Area Under the Curve) を 0% から 100% で評価した結果が図 3 である。平均値法では約半数でしか $AUC=95\%$ が達成できなかったのに対し、 ED^3 では約 7 割で $AUC=95\%$ が達成できた。この結果からも ED^3 の優位性が確認できる。

3.2 結果：実験データ

実験データでは正常な密度行列を 250 個、異常な密度行列を 50 個生成した。ここからランダムに正常な行列を 25 個、異常な行列を 5 個取り出し計 30 個を 1 つのデータセット ($K = 30$) とした。これを 1000 回繰り返して 1000 個のデータセットを作り、それぞれで平均値法と ED^3 の性能を比較した。なお、シミュレーションデータのときと同様に ED^3 のパラメータ γ は事前に 10 個のデータセットで複数の γ の値を試し、その結果が一番良かった値を 1000 個のデータセット実験で使った。

図 4(a)–(c) に実験データの一例を示す。図 4(a)–(c) はある 1 つのデータセットの 30 個の観測行列の (1, 2), (1, 4) 及び (2, 2) 要素である。シミュレーションデータ (図 2(a)–(c)) と実験データ (図 4(a)–(c)) とを比較すると、どちらも同じ実験系を対象にしているにも関わらず得られる密度行列が異なることがわかる。例えば、シミュレーションデータでは (2, 2) 要素は 10^{-3} 前後と小さな値を取っているのに対し、実験データでは 0.02 前後の少し大きめの値を取っている。これは実験系がシミュレーションの系とは異なること、つまり理想的な物理モデルとは完全に一致しないことを示している。先にも述べたとおり、この結果は物理モデルを

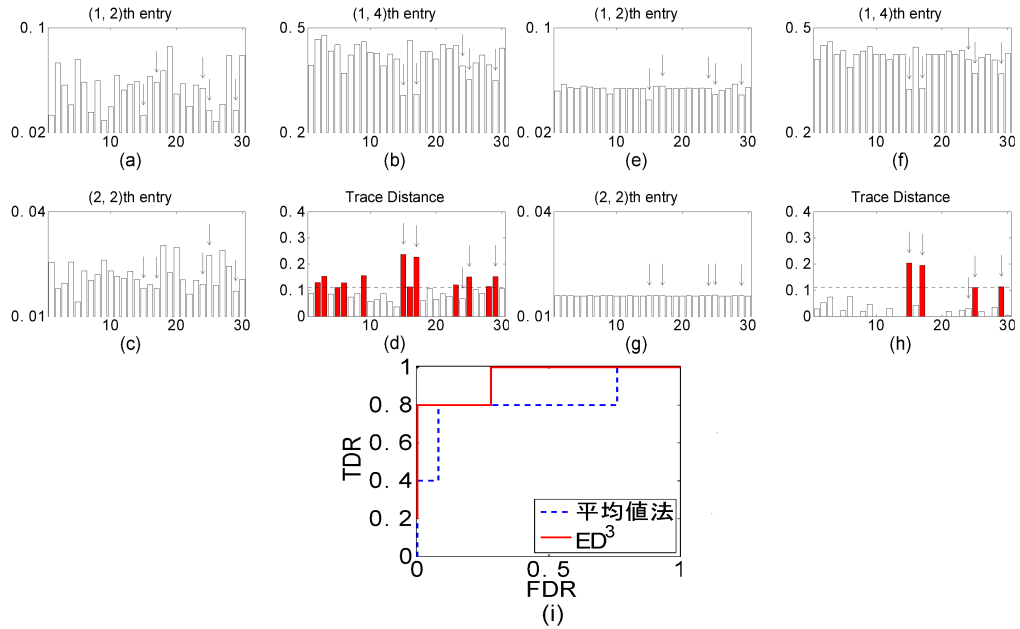


図 4 実験データの結果一例: (a)–(d) 観測行列の (1, 2), (1, 4) 及び (2, 2) 要素と平均値法により計算されたトレース距離 (図中の矢印は異常行列), (e)–(h) ED^3 により計算された行列の (1, 2), (1, 4) 及び (2, 2) 要素とトレース距離 (図中の矢印は異常行列), (i) 平均値法と ED^3 の ROC 曲線

平均値法	95%	90%	85%			
ED^3	95%	90%	85%			

図 5 実験データでの 1000 回の実験における AUC の分布

使った異常検知は真の実験パラメータが得られないためにうまくいかないことを示している。

図 4(e)–(g) は ED^3 で推定された行列 $\theta + \omega_k$ の (1, 2), (1, 4) 及び (2, 2) 要素である。図 4(a)–(c) と比較すると、シミュレーションデータでの結果 (図 2) と同様に、 ED^3 により元データの統計的ゆらぎが抑えられ、異常に伴う変動が抽出できていることがわかる。図 4(d), (h) はそれぞれ平均値法、 ED^3 から計算したトレース距離である。こちらもシミュレーションデータの結果と同様に、平均値法では埋もれてしまっていた異常が ED^3 により検知できていることがわかる。2 つの手法の結果を ROC 曲線で比較したものが図 4(i) であり、ここからも ED^3 の優位性がわかる。

1000 個のデータセットについて ROC 曲線下の面積 AUC (Area Under the Curve) を 0% から 100% で評価した結果が図 5 である。平均値法では約 3 割でしか $AUC=95\%$ が達成できなかったのに対し、 ED^3 では半数以上で $AUC=95\%$ が達成できた。この結果も ED^3 の優位性を示している。

4 まとめ

本稿では機械学習を用いて密度行列の異常を検知する方法 ED^3 について紹介した。 ED^3 では観測行列の偏差の集合がスパースになることに着目し、機械学習のスパース正則化の技術を導入することで異常検知性能の向上を達成した。実際に、 ED^3 を使うことでシミュレーションデータ、実験データの両方において平均値法よりも高い精度で異常が検知できることを確認した。

ED^3 は量子情報と機械学習という異なる分野の組み合わせから生まれた新しい技術である。今後、これらの領域の課題・技術の交流が深まることでより良い・新しい技術が生み出される可能性がある。例えば ED^3 は密度行列の絶対値の変化に着目した技術であるが、これを位相変化の検知まで広げることができればより多くの局面で異常検知が可能となる。また、密度行列でなく物理観測系をモデル化してそこに機械学習の技術を導入することで、観測データからより直接的に異常を検知できるようになる可能性も考えられる。

参考文献

- [1] Satoshi Hara, Takafumi Ono, Ryo Okamoto, Takashi Washio, and Shigeki Takeuchi. Anomaly detection in reconstructed quantum states using a machine-learning technique. *Physical Review A*, 89(2):022104, 2014.
- [2] Daniel FV James, Paul G Kwiat, William J Munro, and Andrew G White. Measurement of qubits. *Physical Review A*, 64(5):052312, 2001.
- [3] Victoria J Hodge and Jim Austin. A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22(2):85–126, 2004.
- [4] Michael A Nielsen and Isaac L Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010.
- [5] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.
- [6] Pradeep Ravikumar, John Lafferty, Han Liu, and Larry Wasserman. Sparse additive models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(5):1009–1030, 2009.

- [7] Ali Jalali, Sujay Sanghavi, Chao Ruan, and Pradeep K Ravikumar. A dirty model for multi-task learning. In *Advances in Neural Information Processing Systems*, pages 964–972, 2010.
- [8] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.