

課題 1.

研究テーマ：再帰的構造を有する密行列を利用した高速な行列の sketching

(本文)

私が修士課程において行いたい研究テーマは様々あるが、その内の一つが**行列の sketching**である。

このテーマの背景

初めに、このテーマの背景である sketching とは、一体どのようなものなのかについて述べる。これは、近年発展してきたデータ圧縮技法の一つであり、非常に大規模な行列 $A \in \mathbb{R}^{n \times d}$ ($n \gg d$) に対して、一定の確率、および、一定の割合のもとでその性質 (ベクトル積との L^2 ノルムなど) を保持するような、より小さい行列 B を求めることを指す。そのような操作をすることで、大規模なデータを、現実的に計算可能なサイズにまで落とした上で、より高速に最適化などを行うことが可能となる [2]。

代表的な手法の一つに、**sparse matrix** を用いた sketching が挙げられる [10]。sparse matrix とは、Fig2 に示したような、各列に ± 1 の内のどちらかが、ランダムに丁度一つ存在するような疎行列である。そして、このような data oblivious な行列であっても、一定のパラメータの下で、Johnson-Lindenstrauss Transform (JLT) になるということが知られている。つまり、

$$S \text{ is JLT}(\epsilon, \delta, f) \Leftrightarrow |\langle Sv, Sv' \rangle - \langle v, v' \rangle| \leq \epsilon \|v\|_2 \|v'\|_2 \quad (\text{with probability } 1 - \delta) \quad (v, v' \in f\text{-element subset } V \subset \mathbb{R}^n)$$

となる [10]。このような性質を持つ行列 S (Fig2) などを用いて、行列 A (Fig1) の性質を、その部分空間である行列 $B = SA$ (Fig3) に埋め込む。これが、sparse matrix を用いた sketching の基本的な考え方である。

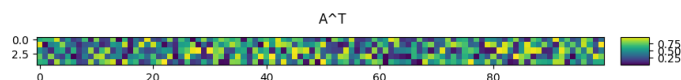


Figure 1: sketching の対象となる行列 A の転置

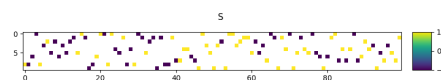


Figure 2: sparse matrix S

近年の進展状況

先程は sparse matrix を掛け合わせる手法を取り上げたが、matrix sketching には他にも様々な手法がある。一つは、element-wise に行列の要素を取って来て、全体の形はそのままに疎行列にする方法がある [6]。また、後述のように、幾つかの列を組み合わせながら作成していく方法などもある。そして、それぞれの方面で、盛んな研究が行われており、相互に影響を与えながら進展しているといった印象を受けている。

また、特に近年で言えば、機械学習分野など、大規模なデータを扱う分野の盛り上がりと呼応して、pass-efficient や streaming などといった語が登場するようになってきている。則ち、データを何回か、あるいは、1 回だけしか読み込まないということである。全体をそのままは保持できないような非常に大規模なデータを扱う際、情報の収集と同時に sketching をしてしまえば、メモリを節約し、かつ、全体の性質を保持したまま、データを扱うことが可能となる [7]。なので、このような状況下での sketching も盛んになり始めているようである。

大規模データの効率的な取り扱いの重要性は日ごとに増しているが、そのような場面においてこそ、行列の sketching は真価を発揮すると考えており、今後のさらなる発展が期待される分野であると言えるだろう。そして、それ故に、この分野の研究を視野に入れている。

自分なりの着想・構想

最後に、私個人の観点からの展望を述べる。

一つ大きく注目している点が、**Walsh 行列** (Fig4) を利用した sketching である。詳細は省くが、この Walsh 行列 H_n に対し、対角成分にランダムに ± 1 が入る対角行列 D との積 $H_n D$ を考えると、各列は独立な ± 1 からなる vector であり、かつ、行列全体としても full rank なので、Chernoff の不等式 (適切な設定の下で、 $P(X \geq a) \leq \frac{E(e^{tX})}{e^{ta}}$) と似た議論により、 $\max_{\|x\|_2=1} \|H_n D x\|_\infty$ がとある定数オーダーで抑えられることが分かる。直感的には、 H_n が密行列であることから、 $H_n D x$ の要素は、 x の各成分が互いに打ち消し合うように合算された値となり、比較的小さい値に抑えられていると解釈出来る。そして、この性質と、Walsh 行列の高速な演算を可能にさせる FFT like なアルゴリズムである高速アダマール変換を用いた sketching が、Subsampled Randomized Hadamard Transform という従来手法である [2][10][1]。

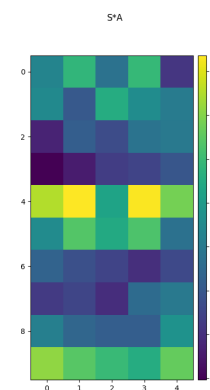
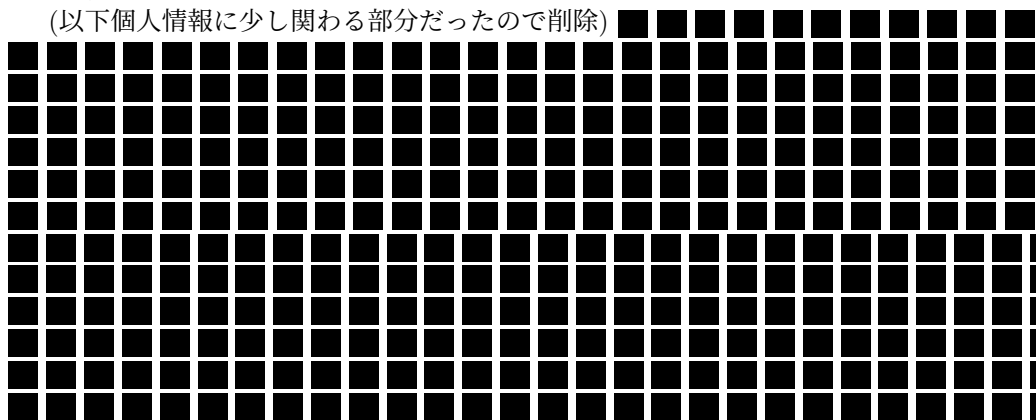


Figure 3: embedding された行列 SA

しかし、Walsh 行列以外にも、高速に演算可能な密行列は数多く存在している。そして、それらの行列の利用はまだあまり検討されていないのではないかと感じた。疎行列が高速な演算を行えることは自明だが、密行列であっても、場合によっては非常に高速というのは、なかなか非自明で興味深いことであるように思われる。なので、今後の研究の一つの方向性として、そのような、密行列を利用した sketching の研究を行いたいと考えている。



課題 2.

重要な事項：特異値分解

(1) 数理的な詳細

特異値分解とは、行列分解の手法の一つであり、任意の実行列 $A \in \mathbb{R}^{n \times m}$ を $A = U\Sigma V^T$ と分解することを指す。ただし、 $U \in \mathbb{R}^{n \times n}$ と $V \in \mathbb{R}^{m \times m}$ は直交行列であり、 Σ は特異値を (一般的には降順に) 対角成分に並べた行列である (なお、行列は必ずしも実行列である必要はなく、体 K の元であれば良いが、ここでは実数に限りて議論を進める)。ここで登場した特異値にはいくつか重要な性質が存在し、行列 A の階数と非零の特異値の個数が一致すること、特異値が $A^T A$ の固有値と一致すること、特異値の二乗和はフロベニウスノルム $\|A\|_F$ の二乗に一致すること ($\because \|A\|_F = \sum_{i,j} |a_{ij}|^2 = \text{tr}(A^T A)$) などが挙げられる。他にも重要な性質は多数存在するが、実用と結びつけた説明の方が良いと考え、本問 (2) にその他多くを記した。また、特異値分解は、一般には $\mathcal{O}(n^3)$ の計算量を要するが、近年では実質的に $\mathcal{O}(n^2)$ の計算量で求める手法なども提案されている [13]。

(2) 数理情報学における意義

特異値分解は、数理情報学において極めて頻出の手法であり、実際、計数工学科における授業内でも、幾度となく登場している。本節では、各分野に沿ってその意義を述べる。

線形代数-条件数 行列の作用素ノルム $\|A\|_p = \sup_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}$ は、 $p = 2$ の場合、 A の最大特異値に一致することが知られている。これは、特異値分解した際の直交行列 U, V に関して、それらをベクトルに掛けてもノルムが不変という性質から導かれる。また、行列の条件数 $\kappa_p(A) = \|A\|_p \|A^{-1}\|_p$ は、同様の理論から A の最大特異値と最小特異値の比に一致することが分かる。ここで、条件数は、 $\sup_{b \neq 0} \frac{\|b\|_2}{\|A^{-1}b\|_2} \cdot \sup_{\Delta b \neq 0} \frac{\|A^{-1}\Delta b\|_2}{\|\Delta b\|_2}$ とも表せ、これは連立方程式 $A(x_b + x_{\Delta b}) = b + \Delta b$ における誤差項 Δb の鋭敏性を評価している。このように、条件数は数値解析において行列の安定性を表す指標として用いられており、特異値分解と合わせて重要と言える [12]。

画像処理-逆問題

特異値分解の重要な応用の一つに、Moore-Penrose の一般化逆行列が挙げられる。この行列は、 $A^+ = V\Sigma^+U^T$ と計算することが出来る。ただし、 Σ^+ は、長方対角行列であり、 Σ の対角成分の非零要素を逆数にして転置したものである。このような行列は、最小二乗問題において、ユークリッドノルム最小の解を与えることなどで知られ、画像のポケ除去などの不良設定問題で応用がある。また、同じく逆問題として、数理工学特論第二など、データ同化やカルマンフィルタなどの文脈においてもこの一般化逆行列は登場することが知られている [11]。

なお、特異値分解は低ランク近似にも用いられており、ある階数 $r'(\leq r)$ での、2 ノルム、および、フロベニウスノルムの意味での最良近似行列は、 A の特異値分解において、最大 r' 個の特異値を残したものであることが知られている (Eckart-Young-Mirsky Theorem)[9]。このことから、画像の圧縮にも使用することが可能である。Fig5 に示した特異値分解による画像の低ランク近似の実験結果では、rank=16 程度でも十分復元出来ており、rank=128 では元画像とほぼ同等になる。元画像は情報理工学系研究科 HP より。コードは GitHub 上に載せてある [3]。

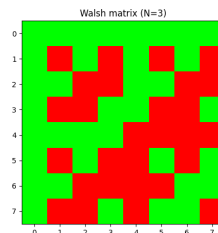


Figure 4:
Walsh matrix
green: $+\frac{1}{\sqrt{2^n}}$, red: $-\frac{1}{\sqrt{2^n}}$

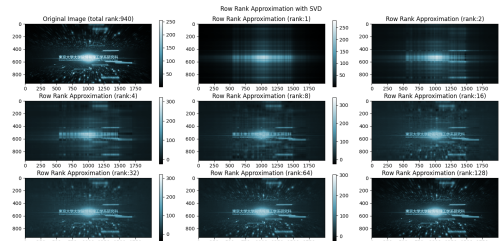


Figure 5: 画像の低ランク近似

機械学習-グラフネットワークにおける異常検知 時系列上の時刻 t におけるグラフネットワーク G_t に対して、隣接行列 A_t と、対角成分に次数を持つ D_t から、ラプラシアン行列 $L_t = D_t - A_t$ を定義することが出来るが、その特異値や特異ベクトルをネットワーク構造の異常検知に用いる手法 (Laplacian Anomaly Detection (LAD)) が知られている [5]。また、参考に挙げた論文中でも述べられている、特に注目すべき点として、SVD が “node permutation invariant”、つまり、行や列の入れ替え操作に対して、特異値が不変ということがある。これは特異値分解の U, V が直交行列ということから直ちに従う。このように、直交変換という工学全体で頻出の操作に対する不変性も、特異値分解の意義をより示していると言える。

統計学-主成分分析 最後に扱うのが、主成分分析 (PCA) である。これは次元削減の最も代表的な例と言え、特異値分解を用いて行う事が可能である。具体的には、 $A = U\Sigma V^T$ と分解されている時、 AV の第 k 列目が第 k 主成分である [14]。詳細は、後述の応用例に譲る。

このように、特異値分解は様々な分野の礎とも言える手法であり、その意義は計り知れない。

(3) 応用例

特異値分解は単にこれらの分野に役立つのみならず、行列の sketching においても極めて重要である。ここでは、特異値分解を利用した行列の sketching として、“Simple and deterministic matrix sketching” に掲載されている **Frequent Direction** という手法を取り上げる [7]。この論文は、データマイニングについての国際会議 SIGKDD において 2013 年度の BestPaperAward を受賞していることから、その重要性は何える。

この手法では、特異値分解を sketching で高速に行うことを目指している。大まかには、sketching の対象となる $A \in \mathbb{R}^{n \times m}$ に対し、その出力結果を $B \in \mathbb{R}^{l \times m}$ とすると、 A の各列を B の空列に追加し、 B を特異値分解などして変形する、という操作を n 回繰り返す。この時、 B は非常に小さいサイズであるので、計算量を抑えることが出来る。そして、理論的な精度保証として重要なのが、以下の不等式評価である。

$$\|A^T A - B^T B\| \leq 2 \frac{\|A\|_F^2}{l}$$

証明を述べる。尤も、細部は元論文を見れば良いだけなので、ここでは証明の要点に絞って記述する。まず、ループ毎の評価値を三角不等式でばらすことにより、また、行列の作用素ノルムが最大特異値と一致することなどにより、 $\|A^T A - B^T B\| \leq \sum_{i=1}^n \delta_i$ と評価することが出来る。ただし、 δ_i はループ毎に出現するとある特異値である。そして、特異値の二乗和はフロベニウスノルムの二乗に一致することなどから、 $\|B^n\|_F^2 \leq \|A\|_F^2 - (\frac{l}{2}) \cdot \sum_{i=1}^n \delta_i$ と抑えられる。以上の二式を合わせると、目的の不等式評価が得られる。そして、この不等式より、 l に対して線形に B の近似精度が向上する為、これはかなり良い近似だと言うことが出来る。

また、この特異値分解の性能を検証するために、iris データセットを用いて PCA を行った結果が、以下である。元々のデータは、Fig6 のように多次元データ (図では 3 次元に制限) であるが、それを、PCA によって二次元に圧縮している。そして、各クラスター毎の境界は、Fig7 でも Fig8 でも、ほぼ同様にくっきりと分かるようになっており、両方の PCA の成功を裏付けている。

なお、これらの実行結果は、いくつかの日本語記事 [4][8] を参照した上で、自ら実装したものを使用した。コードは GitHub 上に載せてある [3]。

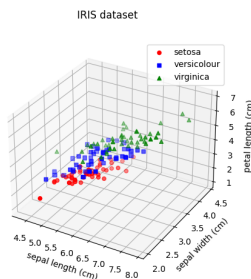


Figure 6: 解析する元データ (3d)

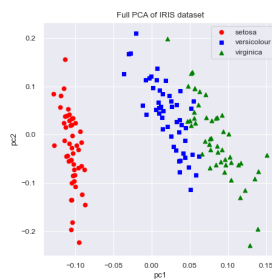


Figure 7: 通常の PCA 結果



Figure 8: sketching による PCA 結果

このコードの計算量はほぼ nnz オーダーであるが、先述の通り、他の手法でも特異値分解が高速に出来てしまうので、速度面で圧倒的な優位性があるとは言えないかも知れない。しかし、先程も触れた “streaming” という特性を持ち、並列に実行可能である点で、大規模データに対しては、より優位性を持つと言えるだろう。

このように、本手法は、sketching と特異値分解の双方にとって、重要な応用例であると言える。

以上で、本課題への解答を終了する。

参考文献

- [1] Nir Ailon and Bernard Chazelle. The Fast Johnson–Lindenstrauss Transform and Approximate Nearest Neighbors. *SIAM Journal on Computing*, Vol. 39, No. 1, pp. 302–322, January 2009.
- [2] Roberta Falcone, Angela Montanari, and Laura Anderlucchi. Matrix sketching for supervised classification with imbalanced classes, December 2019.
- [3] Hiroki Hamaguchi. applyingForGraduateSchool, June 2023.
- [4] Shohei Hido. 今年の SIGKDD ベストペーパーを実装・公開してみました, August 2013.
- [5] Shenyang Huang, Yasmeen Hitti, Guillaume Rabusseau, and Reihaneh Rabbany. Laplacian Change Point Detection for Dynamic Graphs. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 349–358, August 2020.
- [6] Abhisek Kundu and Petros Drineas. A Note on Randomized Element-wise Matrix Sparsification, April 2014.
- [7] Edo Liberty. Simple and deterministic matrix sketching. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13, pp. 581–588, New York, NY, USA, August 2013. Association for Computing Machinery.
- [8] Shinichi Takayanagi. 論文読んだ「Simple and Deterministic Matrix Sketching」. <https://speakerdeck.com/stakaya/lun-wen-du-nda-simple-and-deterministic-matrix-sketching>, April 2019.
- [9] Prof Richard Wilkinson. *3.5 Low-rank Approximation — Multivariate Statistics*.
- [10] David P. Woodruff. Sketching as a Tool for Numerical Linear Algebra. *Foundations and Trends® in Theoretical Computer Science*, Vol. 10, No. 1-2, pp. 1–157, 2014.
- [11] 伊藤耕介, 藤井陽介. 逆問題としての4次元データ同化. ながれ : 日本流体力学会誌 = Nagare : journal of Japan Society of Fluid Mechanics, Vol. 39, No. 3, pp. 167–179, June 2020.
- [12] 宇泰松尾. 数値解析第2章 線形方程式: 連立一次方程式 (数値解析 講義資料), September 2021.
- [13] 山本有作. 特異値分解アルゴリズムの最近の発展. システム／制御／情報, Vol. 58, No. 8, pp. 315–320, 2014.
- [14] 啓人關戸. データ分析基礎 主成分分析.