Fast low rank approximation of matrices using Monte-Carlo techniques

Amir Niknejad
University of Illinois at Chicago

WVU, April 3, 2006

1 RANDOM k-SVD

Stable numerical algortihms of SVD introduced by Golub-Kahan 1965, Golub-Reinsch 1970:

Implicit QR Algo to reduce to upper bidiagonal form using Householder matrices, then Golub-Reinsch SVD algo to zero superdiagonal elements.

Complexity: $O(mn\min(m,n))$.

In applications for massive data:
$A \in \mathbb{R}^{m\times n}, m, n >> 1$ needed a good approximation $A_k = \sum_{i=1}^{k} \mathbf{x}_i \mathbf{y}_i^T, \mathbf{x}_i \in \mathbb{R}^m, \mathbf{y}_i \in \mathbb{R}^n, i = 1, \ldots, k << \min(m,n)$.

Random A_k approximation algo:

Find a good algo by reading l rows or columns of A at random and update the approximations.

Frieze-Kannan-Vempala FOCS 1998 suggest algo without updating.

2 FKNZ RANDOM ALGO [6]

Fast k-rank approximation and SVD algorithm

Input: positive integers m, n, k, l, N, $m \times n$ matrix A, $\epsilon > 0$.

Output: an $m \times n$ k-rank approximation B_f of A, with the ratios $\frac{||B_0||}{||B_t||}$ and $\frac{||B_{t-1}||}{||B_t||}$, approximations to k-singular values and k left and right singular vectors of A.

1. Choose k-rank approximation B_0 using k columns, (or rows), of A.

2. **for** $t = 1$ **to** N

- Select l columns, (or rows), from A at random and update B_{t-1} to B_t.

- Compute the approximations to k-singular values, and k left and right singular vectors of A.

- If $\frac{||B_{t-1}||}{||B_t||} > 1 - \epsilon$ let $f = t$ and finish.

Complexity: $O(mnk)$.

Each iteration $||A - B_{t-1}||_F \geq ||A - B_t||_F$.

3 DETAILS

Choose at random k columns of A. Apply modified Gram-Schmidt algo to obtain $\mathbf{x_1}, \ldots, \mathbf{x_q} \in \mathbb{R}^m, q \leq k$.
Set $B_0 := \sum_{i=1}^{q} \mathbf{x}_i (A^{\mathrm{T}} \mathbf{x}_i)^{\mathrm{T}}$.
$||A - B_0||_F^2 = \mathrm{tr}\, A^T A - \mathrm{tr}\, B_0^T B_0 =$
$\mathrm{tr}\, A^T A - \sum_{i=1}^{q} (A^{\mathrm{T}} \mathbf{x}_i)^{\mathrm{T}} (A^{\mathrm{T}} \mathbf{x}_i)$.

Choose at random another l columns of A: $\mathbf{w_1}, \ldots, \mathbf{w_l}$.
Apply modified Gram-Schmidt algo to
$\mathbf{x_1}, \ldots, \mathbf{x_q}, \mathbf{w_1}, \ldots, \mathbf{w_l}$ to obtain o.n.s.
$\mathbf{x_1}, \ldots, \mathbf{x_q}, \mathbf{x_{q+1}}, \ldots, \mathbf{x_p}$. Form
$C_0 := B_0 + \sum_{i=q+1}^{p} \mathbf{x}_i (A^T \mathbf{x}_i)^T$.

Find the first left k-o.n. left singular vectors $\mathbf{v_1}, \ldots, \mathbf{v_k}$ of C_0. Then $B_1 := \sum_{i=1}^{k} \mathbf{v}_i (A^T \mathbf{v}_i)^T$ and
$\mathrm{tr}\, B_0^T B_0 \leq \mathrm{tr}\, B_1^T B_1$.

Obtain B_t from B_{t-1} as above.

4 Lifting body original

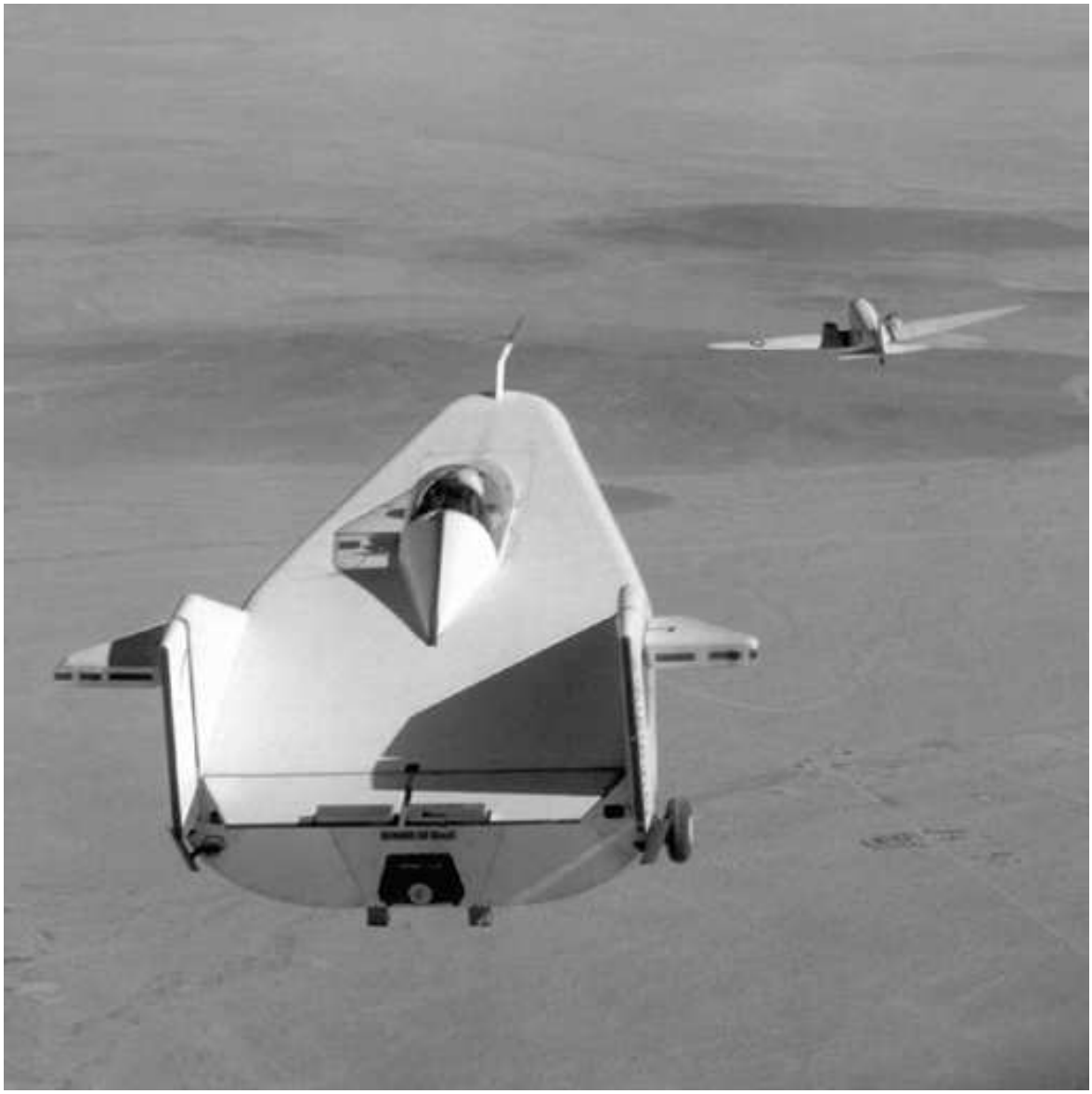

Figure 1: Lifting body image $\mathbf{512} \times \mathbf{512}$.

5 Lifting body compressed

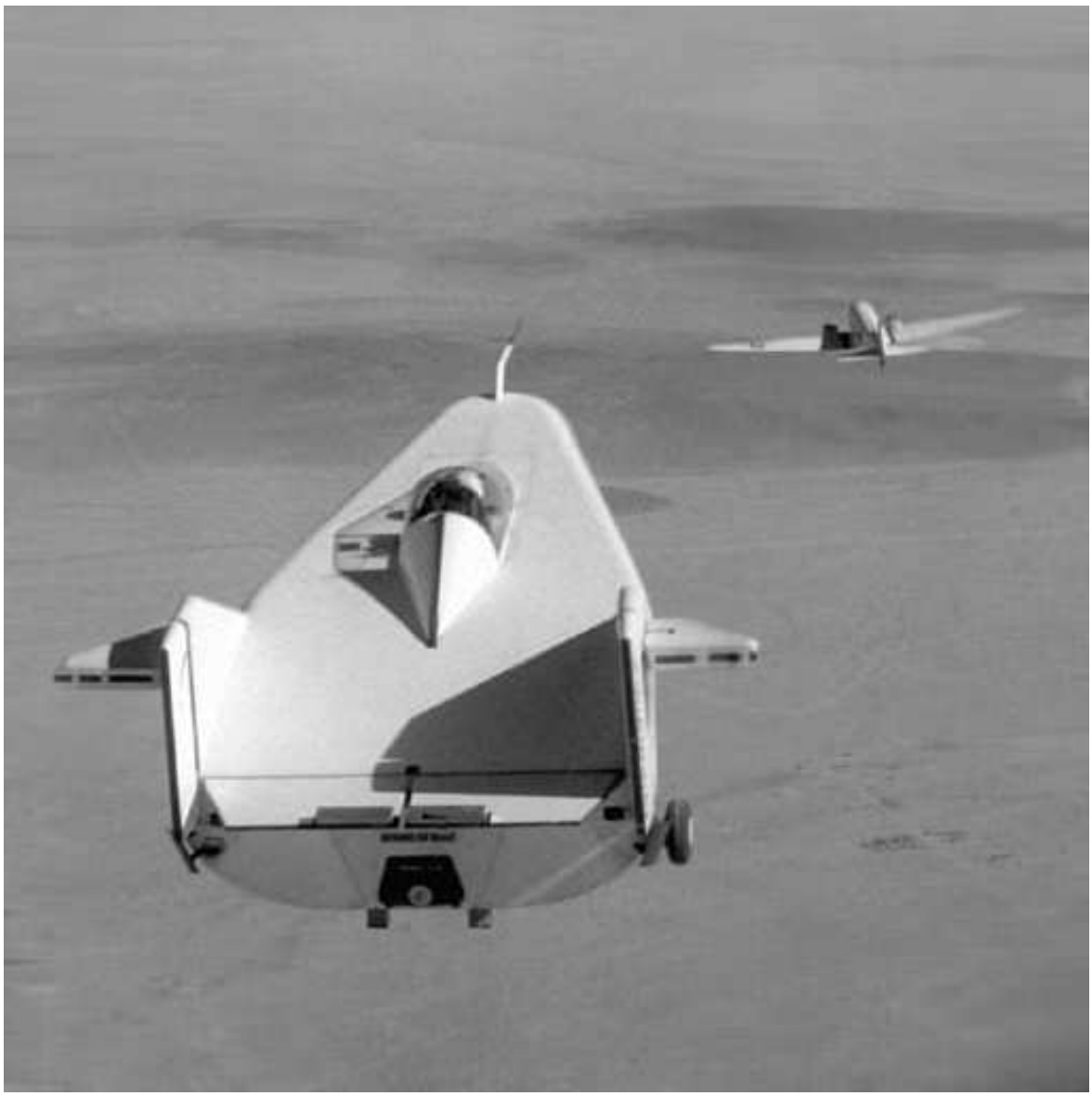

Figure 2: 80-rank approximation of Lifting body image 512×512.

6 SIMULATIONS 1

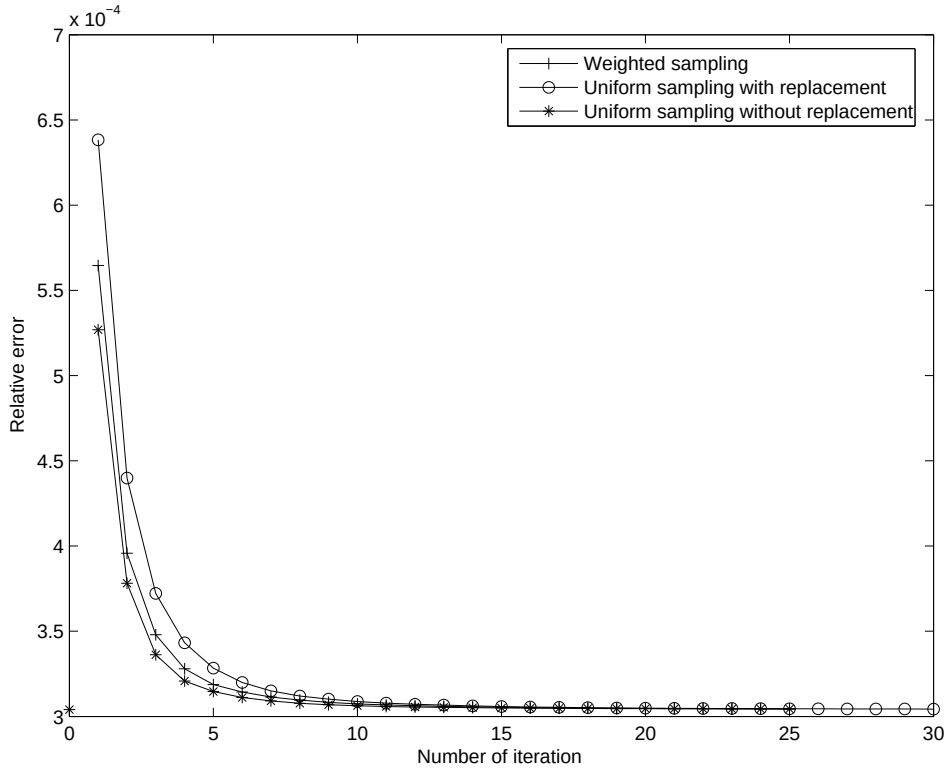

Figure 3: Convergence property of the Monte-Carlo method for Liftingbody image($\mathbf{512 \times 512}$), $\boldsymbol{k = 80}$.

7 SIMULATIONS 2

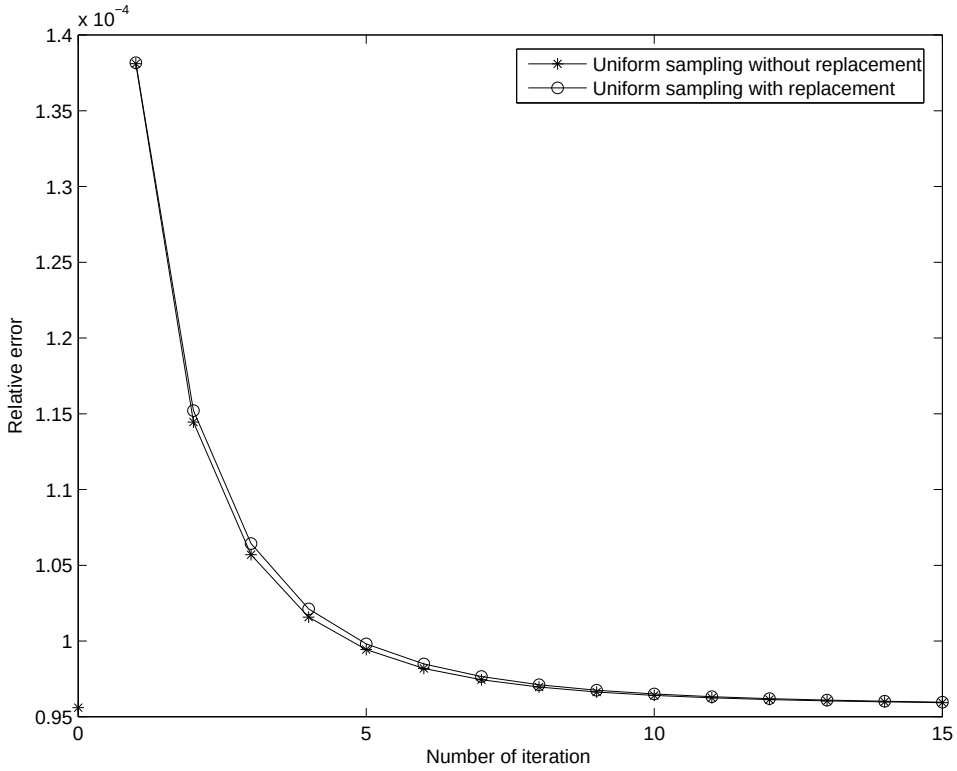

Figure 4: Convergence property of the Monte-Carlo method for random data matrix($\mathbf{3000 \times 500}$) $\boldsymbol{k = l = 100}$.

8 COMPARISONS

Table 1: Comparison of relative error and speed up of our algorithm with optimum k-rank approximation algorithm

Data sets	Speed up	Re. ratio
Cameraman($\mathbf{256 \times 256}$), $k = 80$	1.145	1.083
Liftingbody ($\mathbf{512 \times 512}$), $k = 100$	8	1.08
Map image($\mathbf{627 \times 865}$) $k = 200$	3.33	1.067
Random matrix($\mathbf{8000 \times 200}$) $k = 100$	42	1.1

9 Choosing columns of A

Frieze, Kannan and Vempala [10] suggest to choose column $\mathbf{c}_i(A)$ with probability $\frac{||\mathbf{c}_i(A)||^2}{||A||_F^2}$.

If $s \geq k$ are chosen then the k-approximation satisfies A_k

$$||A - A_k||_F^2 \leq \sum_{i=k+1}^{m} \sigma_i(A)^2 + \frac{10k}{s}||A||_F^2.$$

If $s \geq \frac{k}{10\epsilon}$ then

$$||A - A_k||_F^2 \leq \sum_{i=k+1}^{m} \sigma_i(A)^2 + \epsilon||A||_F^2.$$

Deshpande, Rademacher, Vempala and Wang [4] improved the sampling by modifying the sampling $\mathbf{c}_i(A)$ according to new probabilities $\frac{||\mathbf{c}_i(A-A_k)||^2}{||A-A_k||_F^2}$.

Perhaps our algorithm can be combined with above sampling of columns to get better results.

References

[1] O. Alter, P.O. Brown and D. Botstein, Singular value decomposition for genome-wide expression data processing and modelling, *Proc. Nat. Acad. Sci. USA* 97 (2000), 10101-10106.

[2] O. Alter, P.O. Brown and D. Botstein, Generalized singular decomposition for comparative analysis of genome-scale expression data sets of two different organisms, *Proc. Nat. Acad. Sci. USA* 100 (2003), 3351-3356.

[3] O. Alter, G.H. Golub, P.O. Brown and D. Botstein, Novel genome-scale correlation between DNA replication and RNA transcription during the cell cycle in yeast is predicted by data-driven models, 2004 *Miami Nature Winter Symposium*, Jan. 31 - Feb. 4, 2004, to appear.

[4] A. Deshpande, L. Rademacher, S. Vemapala and G. Wang, Matrix Approximation and Projective Clustering via Volume Sampling, *SODA*, 2006.

[5] S. Friedland, A New Approach to Generalized Singular Value Decomposition, to appear in *SIMAX*.

[6] S. Friedland, M. Kaveh, A. Niknejad and H. Zare, Fast Monte-Carlo Low Rank Approximations for Matrices, preprint 2005, submitted, (Arxiv...)

[7] S. Friedland, M. Kaveh, A. Niknejad and H. Zare, An Algorithm for Missing Value Estimation for DNA Microarray Data, preprint 2005, submitted, (Arxiv...)

[8] S. Friedland, A. Niknejad and L. Chihara, A simultaneous reconstruction of missing data in DNA microarrays, to appear in *Linear Algebra and Its Applications*.

[9] S. Friedland, J. Nocedal and M. Overton, The formulation and analysis of numerical methods for inverse eigenvalue problems, SIAM J. Numer. Anal. 24 (1987), 634-667.

[10] A. Frieze, R. Kannan and S. Vempala, Fast Monte-Carlo algorithms for finding low rank approximations, *Proceedings of the 39th Annual Symposium on Foundation of Computer Science*, 1998.

[11] G.H. Golub and C.F. Van Loan, *Matrix Computation*, John Hopkins Univ. Press, 3rd Ed., 1996.

[12] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge Univ. Press, 1987.

[13] R.A. Johnson, D. W. Wichern, Applied Multivariate Statistical Analysis, Prentice Hall, New Jersey, 4th edition (1998).

[14] H. Kim, G.H. Golub and H. Park, Missing value estimation for DNA microarray gene expression data: local least squares imputation, *Bioinformatics* 21 (2005), 187-198.

[15] S. Oba, M. Sato, I. Takemasa, M. Monden, K. Matsubara and S. Ishii, A Baesian missing value estimation method for gene expression profile data, *Bioinformatics* 19 (2003), 2088-2096.

[16] C.C. Paige and M. A. Saunders, Towards a generalized singular value decomposition, *SIAM J. Numer. Anal.* 18 (1981), 398-405.

[17] P.T. Spellman, G. Sherlock, M.Q. Zhang, V.R. Iyer, K. Anders, M.B. Eisen, P.O. Brown, D. Botstein and B. Futcher, Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces*

cerevisiae by microarray hybridization, Mol. Biol. Cell, **9** (1998), 3273-3297.

[18] G.W. Stewart, A method for computing the generalized singular value decomposition, *Matrix Pencils*, B. K**a**gström and A. Ruhe, *Lecture Notes in Mathematics*, 973 (1982), 207-220.