Randomized algorithms for the low-rank approximation of matrices

Roberta Graff[*], **Javier de Ruiz Garcia**[†], **and Franklin Sonnery**[†]

[*]University of Cambridge, Cambridge, United Kingdom, and [†]Universidad de Murcia, Bioquimica y Biologia Molecular, Murcia, Spain

Submitted to Proceedings of the National Academy of Sciences of the United States of America

We use heat kernels or eigenfunctions of the Laplacian to construct local coordinates on large classes of Euclidean domains and Riemannian manifolds (not necessarily smooth, e.g. with $\mathcal{C}^\alpha$ metric). These coordinates are bi-Lipschitz on large neighborhoods of the domain or manifold, with constants controlling the distortion and the size of the neighborhoods that depend only on natural geometric properties of the domain or manifold. The proof of these results relies on novel estimates, from above and below, for the heat kernel and its gradient, as well as for the eigenfunctions of the Laplacian and their gradient, that hold in the non-smooth category, and are stable with respect to perturbations within this category. Finally, these coordinate systems are intrinsic and efficiently computable, and are of value in applications.

monolayer | structure | x-ray reflectivity | molecular electronics

Abbreviations: SAM, self-assembled monolayer; OTS, octadecyltrichlorosilane

Low-rank approximation in this article we study the evolution of "almost-sharp" fronts for the surface quasi-geostrophic equation. This 2-D active scalar equation reads for the surface quasi-geostrophic equation.

$$\frac{D\theta}{Dt} = \frac{\partial\theta}{\partial t} + u\cdot\nabla\theta = 0 \qquad \mathbf{(1)}$$

where,

$$u = (u_1, u_2) = (-\frac{\partial\psi}{\partial y}, \frac{\partial\psi}{\partial x}) \qquad \mathbf{(2)}$$

and,

$$(-\triangle)^{\frac{1}{2}}\psi = \theta\,, \qquad \mathbf{(3)}$$

For simplicity we are considering fronts on the cylinder, i.e. we take (x, y) in $\mathbb{R}/_{\mathbb{Z}} \times \mathbb{R}$. In this setting we define $-\triangle^{-\frac{1}{2}}$ that comes from inverting the third equation by convolution with the kernel. To avoid irrelevant considerations at ∞ we will take η to be compactly supported

$$\frac{\chi(u,v)}{(u^2+v^2)^{\frac{1}{2}}} + \eta(u,v)$$

where $\chi(x,y)\;\epsilon\;C_0^\infty,\;\chi(x,y) = 1\;$ in $|x-y| \le r$ and $supp\chi$ is contained in $\{|x-y| \le R\}$ with $0 < r < R < \frac{1}{2}$. Also $\eta\;\epsilon\;C_0^\infty$, $\eta(0,0) = 0$.

Section 1: Preliminaries

The main mathematical interest in the quasi-geostrophic equation lies in its strong similarities with the 3-Euler equations. These results were first proved by Constantin, Majda and Tabak, see (10), (11) and (12) for more details. There are several other research lines for this equation, both theoretical and numerical. See (13), (14), (15), (6), (3), and (4). The question about the regularity of the solutions for QG remains as an open problem.

Subsection 1.1: Interpolative Decompositions. Recently one of the authors has obtained the equation for the evolution of sharp fronts (in the periodic setting), proving its local well-posedness for that equation. (see (16) and (10) for more details). This is a problem in contour dynamics. Contour dynamics for other fluid equations has been studied extensively. (see (17) (15) and (14)).

We begin our analysis on almost sharp fronts for the quasi-geostrophic equation recalling the notion of weak solution.

Lemma 1. *A bounded function θ is a weak solution of QG if for any $\phi\,\epsilon\,C_0^\infty(\mathbb{R}/_{\mathbb{Z}}\times\mathbb{R}\times[0,\varepsilon])$ we have*

$$\int_{\mathbb{R}^+\times\mathbb{R}/_{\mathbb{Z}}\times\mathbb{R}} \theta(x,y,t)\,\partial_t\phi\,(x,y,t)dydxdt + \int_{\mathbb{R}^+\times\mathbb{R}/_{\mathbb{Z}}\times\mathbb{R}} \theta\,(x,y,t)u(x,y,t)\cdot\nabla\phi\,(x,y,t)dydxdt = 0 \qquad \mathbf{[4]}$$

where u is determined by equations **(2)** *and* **(3)**.

Reserved for Publication Footnotes

Subsection 1.2: Data Sources. We are interested in studying the evolution of almost sharp fronts for the QG equation. These are weak solutions of the equation with large gradient ($\sim \frac{1}{\delta}$, where 2δ is the thickness of the transition layer for θ).

Subsection 1.3: Cylindrical Case. We are going to consider the cylindrical case here. We consider a transition layer of thickness smaller than 2δ in which θ changes from 0 to 1. (see Figure 1).That means we are considering θ of the form

$$\begin{array}{ccl} \theta & = & 1 \text{ if } \quad y \geq \varphi(x,t) + \delta \\ \theta & \text{bounded if} & |\varphi(x,t) - y| \leq \delta \\ \theta & = & 0 \text{ if } \quad y \leq \varphi(x,t) - \delta \end{array} \quad [5]$$

where φ is a smooth periodic function and $0 < \delta < \frac{1}{2}$.

Theorem 1. *If the active scalar θ is as in* **(5)** *and satisfies the equation* **(4)***, then φ satisfies the equation*

$$\begin{aligned} \frac{\partial \varphi}{\partial t}(x,t) &= \int_{\mathbb{R}/\mathbb{Z}} \frac{\frac{\partial \varphi}{\partial x}(x,t) - \frac{\partial \varphi}{\partial u}(u,t)}{[(x-u)^2 + (\varphi(x,t) - \varphi(u,t))^2]^{\frac{1}{2}}} \\ &\quad \chi(x-u, \varphi(x,t) - \varphi(u,t))du \;+ \\ &+ \int_{\mathbb{R}/\mathbb{Z}} \Big[\frac{\partial \varphi}{\partial x}(x,t) - \frac{\partial \varphi}{\partial u}(u,t)\Big] \\ &\quad \eta(x-u, \varphi(x,t) - \varphi(u,t))du + Error \end{aligned} \quad [6]$$

with $|Error| \leq C\,\delta|log\delta|$ where C depends only on $\|\theta\|_{L^\infty}$ and $\|\nabla\varphi\|_{L^\infty}$.

Remark 1. *Note that equation* **(5)** *specifies the function φ up to an error of order δ. Theorem 1 provides an evolution equation for the function φ up to an error of order $\delta|log\delta|$.*

In order to analyze the evolution of the "almost-sharp" front we substitute the above expression for θ in the definition of a weak solution (see **(4)**). We use the notation $X = O(Y)$ to indicate that $|X| \leq C|Y|$ where the constant C depends only on $\|\theta\|_{L^\infty}$, $\|\nabla\varphi\|_{L^\infty}$ and $\|\phi\|_{C^1}$, where ϕ is a test function appearing in Definition 1.

We consider the 3 different regions defined by the form on θ. Since $\theta = 0$ in the region I the contribution from that region is 0, i.e.

$$\int_{I\times\mathbb{R}} \theta(x,y,t)\,\partial_t\phi\,(x,y,t)\,dydxdt + \int_{I\times\mathbb{R}} \theta\,(x,y,t)u(x,y,t)\cdot\nabla\phi\,(x,y,t)\,dydxdt = 0$$

As for the region II

$$\int_{II\times\mathbb{R}} \theta(x,y,t)\partial_t\phi(x,y,t)dxdydt = O(\delta)$$

since θ is bounded and hence $O(1)$, and the area of the region II is $O(\delta)$. As for the second term

$$\int_{II\times\mathbb{R}} u\theta\nabla\phi dxdydt = O(\delta log(\delta))$$

To see this, we fix t. We must estimate

$$\int_{\mathbb{R}^2} u\cdot(\mathbb{1}_{II}\theta\nabla\phi)dxdy$$

Using **(2)** and **(3)** we obtain

$$u = \nabla^{\perp}(\triangle^{-\frac{1}{2}}\theta)$$

and hence $u(x,y,t) = K * \theta(x,y,t)$ where K looks locally like the orthogonal of the Riesz transform, precisely

$$K(u,v) = \nabla^{\perp}\Big\{\frac{\chi(u,v)}{(u^2+v^2)^{\frac{1}{2}}} + \eta(u,v)\Big\} \quad (7)$$

Since θ is bounded, from the above expression for K we obtain that u is of exponential class (see (11)) and hence

$$\int_{\mathbb{R}^2} u\cdot(\mathbb{1}_{II}\theta\nabla\phi)dxdy$$

is the integral of a function of exponential class over a set of measure δ, the set II. Recalling that the dual of the set of functions of exponential class is $LlogL$ we obtain

$$\int_{\mathbb{R}^2} u\cdot(\mathbb{1}_{II}\theta\nabla\phi)dxdy = O(ln)$$

As for the region III we proceed as follows.

We decompose $u = u_f + u_b$, where $u_f = K * \mathbb{1}_{III}$ and $u_b = K * \theta\mathbb{1}_{II}$. Notice that both u_f and u_b are divergence free.

For a fixed t we have

$$\begin{aligned}\int_{III} u_b \cdot \nabla\phi dx &= \int_{\mathbb{R}^2} \left[K * (\theta\mathbb{1}_{II})\right] \cdot \left[\mathbb{1}_{III}\nabla\phi\right] dxdy \\ &= \int_{\mathbb{R}^2} \theta\mathbb{1}_{II} K * \cdot [\mathbb{1}_{III}\nabla\phi] dxdy = O(\delta log\delta)\end{aligned}$$

since the area where we are integrating is $O(\delta)$ and the function $K * [\mathbb{1}_{III}\nabla\phi]$ is of exponential class, where we have used the notation $K * \cdot$ to denote the sum of the application of the operator to each component.

We are left to estimate the terms

$$\int_{III\times\mathbb{R}} \theta\partial_t\phi dxdydt + \int_{III\times\mathbb{R}} \theta u_f \cdot \nabla\phi dxdydt =: A + B$$

The first term gives

$$\begin{aligned}A &= \int_{III\times\mathbb{R}} \theta\, \partial_t\phi\, dydxdt = \int_{y>\varphi(x,t)+\delta} \partial_t\phi\, dxdydt \\ &= \int_{y=\varphi(x,t)+\delta} \phi\partial_t\varphi dxdt\end{aligned}$$

In order to analyze B, fix t and consider only the space integration

$$\begin{aligned}\int_{III} \theta\, u_f \cdot \nabla\phi\, dy\, dx &= \lim_{\epsilon\to 0^+} \int_{y>\varphi(x,t)+\delta+\epsilon} u_f \cdot \nabla\phi\, dxdy \\ &= \lim_{\epsilon\to 0^+} \int_{y=\varphi(x,t)+\delta+\epsilon} u_f\phi \cdot (\frac{\partial\varphi}{\partial x}, -1)dx\end{aligned}$$

Now we look more closely to the integrand of the above expression. We have

$$u_f\phi \cdot (\frac{\partial\varphi}{\partial x}, -1) = \phi(x,y,t) \int_{v>\varphi(u,t)+\delta} K(x-u, y-v) \cdot (\frac{\partial\varphi}{\partial x}, -1)dudv$$

Now, considering all the contributions from all the regions, we obtain that equation **(4)** is equivalent to

$$\begin{aligned}&\int_{y=\varphi(x,t)+\delta} \phi(x,y,t) \frac{\partial\varphi}{\partial t}(x,t)\, dx\, dt \\ +\ &\int_{y=\varphi(x,t)+\delta} -\phi(x,y,t) \int_{v=\varphi(u,t)+\delta} \frac{\frac{\partial\varphi}{\partial u} - \frac{\partial\varphi}{\partial x}}{((x-u)^2 + (y-v)^2)^{\frac{1}{2}}} \chi(x-u, y-v) du\, dx\, dt \\ +\ &\int_{y=\varphi(x,t)+\delta} -\phi(x,y,t) \int_{v=\varphi(u,t)+\delta} \left[\frac{\partial\varphi}{\partial u} - \frac{\partial\varphi}{\partial x}\right] \eta(x-u, y-v) dudxdt + O(\delta|log\delta|) = 0\end{aligned}$$

And hence we obtain the equation

$$\begin{aligned}\frac{\partial\varphi}{\partial t}(x,t) &= \int_{\mathbb{R}/\mathbb{Z}} \frac{\frac{\partial\varphi}{\partial x}(x,t) - \frac{\partial\varphi}{\partial u}(u,t)}{[(x-u)^2 + (\varphi(x,t) - \varphi(u,t))^2]^{\frac{1}{2}}} \chi(x-u, \varphi(x,t) - \varphi(u,t)) du \\ &+ \int_{\mathbb{R}/\mathbb{Z}} \left[\frac{\partial\varphi}{\partial x}(x,t) - \frac{\partial\varphi}{\partial u}(u,t)\right] \eta(x-u, \varphi(x,t) - \varphi(u,t)) du + O(\delta|log\delta|)\end{aligned}$$

which proves the theorem.

It would be interesting to give a rigorous construction of an almost sharp front solving the surface quasi-geostrophic equation for given initial data φ and arbitrarily small thickness δ.

The problem of the evolution of almost sharp fronts considered here could be a simple model for a very interesting and hard problem of justifying rigorously the equation for the evolution of a vortex line.

If one imagines the vorticity as a delta function supported along a curve, the attempt of recovering the velocity by using the Biot-Savart law shows a singularity proportional to the inverse of the distance to the curve. The heuristic derivations of an equation for the evolution of the curve simply remove that singularity from the equation. The main problem faced in a rigorous derivation is that a vortex line, as described above, fails to be a weak solution of the Euler equation.

Modifying the definition of weak solution may introduce objects unrelated to physically meaninful solutions of the 3D Euler equation (see (12)) Instead one could try to consider solutions supported on a very small neighborhood of the "vortex line" and obtain an equation for the evolution of such a solution based on the core line and the thickness, hoping that some limit could be found when sending the thickness to 0.

The analysis we have presented here is the analog of the proposed strategy for the surface quasi-geostrophic equation. This equation also contains a singularity in the velocity as we approach the front. The singularity is only logarithmic, weaker than the one in 3D Euler. Nevertheless we hope this might provide some insights for the understanding of the vortex line problem.

Subsection 1.4: Observations. The following observations can be made form the numerical experiments described in this section, and are consistent with the results of more extensive experimentation performed by the authors:

- The CPU times in Tables 1–3 are compatible with the estimates in formulae **14**, **16**, and **23**.
- The precision produced by each of Algorithms **I** and **II** is similar to that provided by formula **3**, even when ω_{k+1} is close to the machine precision.

Section 2: Conclusions

This note describes two classes of randomized algorithms for the compression of linear operators of limited rank, as well as applications of such techniques to the construction of singular value decompositions of matrices. Obviously, the algorithms of this note can be used for the construction of other matrix decompositions, such as the Schur decomposition. In many situations, the numerical procedures described in this note are faster than the classical ones, while ensuring comparable accuracy.

Materials and Methods

Digital RNA SNP Analysis. A real-time PCR assay was designed to amplify *PLAC4* mRNA, with the two SNP alleles being discriminated by TaqMan probes. *PLAC4* mRNA concentrations were quantified in extracted RNA samples followed by dilutions to approximately one target template molecule of either type (i.e., either allele) per well. Details are given in the *SI Materials and Methods*.

Digital RCD Analysis. Extracted DNA was quantified by spectrophotometry (NanoDrop Technologies, Wilmington, DE) and diluted to a concentration of approximately one target template from either chr21 or ch1 per well.

Appendix: Estimating the Spectral Norm of a Matrix

In this appendix we describe a method for the estimation of the spectral norm of matrix A. The method does not require access to the individual entries of A; it requires only applications of A and A^* to vectors. It is a version of the classical power method. Its probabilistic analysis summarized below was introduced fairly recently in refs. 13 and 14. This appendix is included here for completeness.

Appendix

This is an example of an appendix without a title.

ACKNOWLEDGMENTS. This work was partially supported by Spanish Ministry of Science and Technology Grant BFM2002-02042 (to D.C. and J.L.R.) and by National Science Foundation Grand DMS-0245242 (to C.F.).

1. M. Belkin and P. Niyogi, *Using manifold structure for partially labelled classification*, Advances in NIPS, 15 (2003).
2. P. Bérard, G. Besson, and S. Gallot, *Embedding Riemannian manifolds by their heat kernel*, Geom. and Fun. Anal., 4 (1994), pp. 374–398.
3. R.R. Coifman and S. Lafon, *Diffusion maps*, Appl. Comp. Harm. Anal., 21 (2006), pp. 5–30.
4. R.R. Coifman, S. Lafon, A. Lee, M. Maggioni, B. Nadler, F. Warner, and S. Zucker, *Geometric diffusions as a tool for harmonic analysis and structure definition of data. Part I: Diffusion maps*, Proc. of Nat. Acad. Sci., (2005), pp. 7426–7431.
5. P. Das, M. Moll, H. Stamati, L. Kavraki, and C. Clementi, *Low-dimensional, free-energy landscapes of protein-folding reactions by nonlinear dimensionality reduction*, P.N.A.S., 103 (2006), pp. 9885–9890.
6. D. Donoho and C. Grimes, *Hessian eigenmaps: new locally linear embedding techniques for high-dimensional data*, Proceedings of the National Academy of Sciences, 100 (2003), pp. 5591–5596.
7. D. L. Donoho and C. Grimes, *When does isomap recover natural parameterization of families of articulated images?*, Tech. Report Tech. Rep. 2002-27, Department of Statistics, Stanford University, August 2002.
8. M. Grüter and K.-O. Widman, *The Green function for uniformly elliptic equations*, Man. Math., 37 (1982), pp. 303–342.
9. R. Hempel, L. Seco, and B. Simon, *The essential spectrum of neumann laplacians on some bounded singular domains*, 1991.
10. Kadison, R. V. and Singer, I. M. (1959) Extensions of pure states, *Amer. J. Math.* 81, 383-400.
11. Anderson, J. (1981) A conjecture concerning the pure states of $B(H)$ and a related theorem. in *Topics in Modern Operator Theory*, Birkhaüser, pp. 27-43.
12. Anderson, J. (1979) Extreme points in sets of positive linear maps on $B(H)$. *J. Funct. Anal.* 31, 195-217.
13. Anderson, J. (1979) Pathology in the Calkin algebra. *J. Operator Theory* 2, 159-167.
14. Johnson, B. E. and Parrott, S. K. (1972) Operators commuting with a von Neumann algebra modulo the set of compact operators. *J. Funct. Anal.* 11, 39-61.
15. Akemann, C. and Weaver, N. (2004) Consistency of a counterexample to Naimark's problem. *Proc. Nat. Acad. Sci. USA* 101, 7522-7525.
16. J. Tenenbaum, V. de Silva, and J. Langford, *A global geometric framework for nonlinear dimensionality reduction*, Science, 290 (2000), pp. 2319–2323.
17. Z. Zhang and H. Zha, *Principal manifolds and nonlinear dimension reduction via local tangent space alignement*, Tech. Report CSE-02-019, Department of computer science and engineering, Pennsylvania State University, 2002.

Fig. 1. Almost Sharp Front

Fig. 2. Full width illustration

Table 1. Repeat length of longer allele by age of onset class. This is what happens when the text continues.

Age of onset,	Repeat length				
years	n	Mean	SD	Range	Median
Juvenile, 2−20	40	60.15	9.32	43−86	60
Typical, 21−50	377	45.72	2.97	40−58	45
Late, $>$50	26	41.85	1.56	40−45	42*

*The no. of wells for all samples was 384. Genotypes were determined by mass spectrometric assay. The m_t value indicates the average number of wells positive for the over represented allele.

Table 2. Summary of the experimental results

Parameters		Averaged Results							Comparisons			
n	S^{*}_{MAX}	t_1	r_1	m_1	t_2	r_2	m_2	t_{lb}	t_1/t_2	r_1/r_2	m_1/m_2	t_1/t_{lb}
10*	1	4	.0007	4	4	.0020	4	4	1.000	.333	1.000	1.000
10†	5	50	.0008	8	50	.0020	12	49	.999	.417	.698	1.020
100‡	20	2840975	.0423	95	2871117	.1083	521	—	.990	.390	.182	—

*Stanford Synchrotron Radiation Laboratory (Stanford University, Stanford, CA)

†$R_{\mathrm{FREE}} = R$ factor for the ~ 5% of the randomly chosen unique reflections not used in the refinement.

‡Calculated for all observed data