Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational Physics 219 (2006) 7–12
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

Short note

A sharp error estimate for the fast Gauss transform

Xiaoliang Wan, George Em Karniadakis *

Division of Applied Mathematics, Brown University, 182 George Street, Box F, Providence, RI 02912, USA

Received 6 October 2004; received in revised form 19 April 2006; accepted 27 April 2006
Available online 19 June 2006

Abstract

We report an error estimate of the multi-dimensional fast Gauss transform (FGT), which is much sharper than that previously reported in the literature. An application to the Karhunen–Loeve decomposition in the three-dimensional physical space is also presented that shows savings of three orders of magnitude in time and memory compared to a direct solver.

Keywords: Fast Gauss transform; Error estimate

1. A new error estimate for the fast Gauss transform

The first version of the fast Gauss transform (FGT) method was proposed by Greengard and Strain [1], and it was later improved in [2,3] in terms of its computational speed-up. However, a wrong error estimate of truncated Hermite expansion was employed there. This was corrected later by Baxter and Roussos [4] but this new correct estimate overestimates the error and is not as useful in high-dimensional problems because of its restriction on the side length of mesh. Motivated by the work in [4], we have constructed a new error estimate for the original FGT method.

Let $R_{hh}(\boldsymbol{x}, \boldsymbol{y})$ denote the Gaussian kernel

$$R_{hh}(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{e}^{-\|\boldsymbol{x}-\boldsymbol{y}\|_2^2/\delta}, \tag{1}$$

where $\|\cdot\|_2$ indicates the L_2 norm. Following the notations of [1,4], we define for the one-dimensional case the Hermite function

$$h_n(x) = \mathrm{e}^{-x^2} H_n(x), \tag{2}$$

where $H_n(x)$ are one-dimensional Hermite polynomials. A d-dimensional multi-index $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d)$ is a d-tuple of nonnegative integers, playing the role of a multi-dimensional index. For any multi-index α and any $\boldsymbol{x} \in \mathbb{R}^d$, we define

* Corresponding author. Tel.: +1 401 863 1217; fax: +1 401 863 3369.
E-mail address: gk@dam.brown.edu (G.E. Karniadakis).

0021-9991/$ - see front matter
doi:10.1016/j.jcp.2006.04.016

$$|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_d, \quad \alpha! = \alpha_1!\alpha_2! \ldots \alpha_d!, \quad \boldsymbol{x}^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_d^{\alpha_d}. \tag{3}$$

If p is an integer, we say $\alpha \geqslant p$ if $\alpha_i \geqslant p$ for $1 \leqslant i \leqslant d$. We can now define the multi-dimensional Hermite polynomials and Hermite functions as

$$\begin{aligned} \boldsymbol{H}_\alpha(\boldsymbol{x}) &= H_{\alpha_1}(x_1) \ldots H_{\alpha_d}(x_d), \\ \boldsymbol{h}_\alpha(\boldsymbol{x}) &= \mathrm{e}^{-|\boldsymbol{x}|^2} \boldsymbol{H}_\alpha(\boldsymbol{x}) = h_{\alpha_1}(x_1) \ldots h_{\alpha_d}(x_d), \end{aligned} \tag{4}$$

where $|\boldsymbol{x}|^2 = x_1^2 + \cdots + x_d^2$.

Theorem 1. *If a source point $\boldsymbol{y} \in \mathbb{R}^d$ lies in a box B with center c_B and side length $r\sqrt{2\delta}$, then for any $\boldsymbol{x} \in \mathbb{R}^d$*

$$\mathrm{e}^{-\|\boldsymbol{x}-\boldsymbol{y}\|_2^2/\delta} = \sum_{\alpha<p} T_\alpha = \sum_{\alpha<p} \frac{1}{\alpha!} \left(\frac{\boldsymbol{y}-\boldsymbol{c}_B}{\sqrt{\delta}}\right)^\alpha \boldsymbol{h}_\alpha\left(\frac{\boldsymbol{x}-\boldsymbol{c}_B}{\sqrt{\delta}}\right) \tag{5}$$

with error estimate

$$|E(p)| = \left|\sum_{\alpha \geqslant p} T_\alpha\right| \leqslant \sum_{i=1}^{d} \binom{d}{i} \left(Kp^{-1/4} \frac{r_p^p}{1-r_p}\right)^i, \tag{6}$$

where $K = 1.09(2\pi)^{-1/4}$ and $r_p = r\sqrt{\mathrm{e}/p} < 1$. Here e *denotes the base of natural logarithm and p the order of Hermite polynomials in the truncated expansion.*

Proof. Consider the ith components of x, y and c_B with $1 \leqslant i \leqslant d$. By the properties of one-dimensional Hermite polynomials, see [5, Eq. (22.9.17)], it is easy to obtain

$$\mathrm{e}^{-(x_i-y_i)^2/\delta} = \sum_{n_i=0}^{\infty} \frac{1}{n_i!} \left(\frac{y_i-c_i}{\sqrt{\delta}}\right)^{n_i} h_{n_i}\left(\frac{x_i-c_i}{\sqrt{\delta}}\right).$$

We introduce the following definitions as in [4]:

$$\begin{aligned} u_p^i(x_i, y_i, c_i) &= \sum_{n_i=0}^{p-1} \frac{1}{n_i!} \left(\frac{y_i-c_i}{\sqrt{\delta}}\right)^{n_i} h_{n_i}\left(\frac{x_i-c_i}{\sqrt{\delta}}\right), \quad 1 \leqslant i \leqslant d, \\ v_p^i(x_i, y_i, c_i) &= \sum_{n_i=p}^{\infty} \frac{1}{n_i!} \left(\frac{y_i-c_i}{\sqrt{\delta}}\right)^{n_i} h_{n_i}\left(\frac{x_i-c_i}{\sqrt{\delta}}\right), \quad 1 \leqslant i \leqslant d, \end{aligned}$$

which we can use to write the corresponding Gaussian kernel as

$$\mathrm{e}^{-\|\boldsymbol{x}-\boldsymbol{y}\|_2^2/\delta} = \prod_{i=1}^{d} (u_p^i + v_p^i).$$

Using the fact that

$$|u_p^i| - |v_p^i| \leqslant |u_p^i + v_p^i| = |\mathrm{e}^{-(x_i-y_i)^2/\delta}| = \mathrm{e}^{-(x_i-y_i)^2/\delta} \leqslant 1, \tag{7}$$

we have

$$|E(p)| = \left|\mathrm{e}^{-\|\boldsymbol{x}-\boldsymbol{y}\|_2^2/\delta} - \prod_{i=1}^{d} u_p^i\right| \leqslant f(|v_p^1|, |v_p^2|, \ldots, |v_p^d|), \tag{8}$$

where

$$f(|v_p^1|, |v_p^2|, \ldots, |v_p^d|) = \sum_{n=1}^{d} \sum_{i_1<\cdots<i_n} |v_p^{i_1}||v_p^{i_2}|, \ldots, |v_p^{i_n}|.$$

For example, if $d = 2$,

$$\begin{aligned}|E(p)| &= |(u_p^1 + v_p^1)(u_p^2 + v_p^2) - u_p^1 u_p^2| \\ &= |(u_p^2 + v_p^2)v_p^1 + u_p^1 v_p^2| \\ &\leqslant |u_p^2 + v_p^2||v_p^1| + |u_p^1||v_p^2| \\ &\leqslant |v_p^1| + (1 + |v_p^1|)|v_p^2| \quad \text{(by Eq. (7))} \\ &= |v_p^1| + |v_p^2| + |v_p^1||v_p^2|.\end{aligned}$$

Considering the inequality for Hermite functions [5, Eq. (22.14.17)]

$$\frac{1}{n!}|h_n(x)| \leqslant K_1 \frac{2^{n/2}}{\sqrt{n!}} \mathrm{e}^{-x^2/2}, \qquad n \geqslant 0 \quad \text{and} \quad x \in \mathbb{R} \tag{9}$$

with K_1 being a numerical constant less than 1.09 in value, and the Stirling's formula [5, Eq. (6.1.38)] for the factorial $n!$, we can bound each $|v_p^i|$ as

$$\begin{aligned}|v_p^i| &= \left| \sum_{n_i=p}^{\infty} \frac{1}{n_i!} \left(\frac{y_i - c_i}{\sqrt{\delta}} \right)^{n_i} h_{n_i} \left(\frac{x_i - c_i}{\sqrt{\delta}} \right) \right| \\ &\leqslant K_1 \sum_{n_i=p}^{\infty} \frac{1}{\sqrt{n_i!}} \left(\frac{r}{\sqrt{2}} \right)^{n_i} 2^{n_i/2} \mathrm{e}^{-x_i^2/2} \\ &\leqslant K_1 \sum_{n_i=p}^{\infty} \frac{1}{\sqrt{n_i!}} r^{n_i} \\ &= K_1 \sum_{n_i=p}^{\infty} r^{n_i} (2\pi)^{-1/4} n_i^{-\frac{n_i}{2}-\frac{1}{4}} \mathrm{e}^{\frac{n_i}{2}-\frac{\theta}{24n_i}}, \quad 0 < \theta < 1 \text{ (Stirling)} \\ &\leqslant K_1 (2\pi)^{-1/4} \sum_{n_i=p}^{\infty} r^{n_i} \left(\frac{\mathrm{e}}{n_i} \right)^{n_i/2} n_i^{-1/4} \\ &\leqslant K_1 (2\pi p)^{-1/4} \sum_{n_i=p}^{\infty} r^{n_i} \left(\frac{\mathrm{e}}{p} \right)^{n_i/2} \leqslant K_1 (2\pi p)^{-1/4} \frac{r_p^p}{1 - r_p},\end{aligned} \tag{10}$$

where $r_p = r\sqrt{\mathrm{e}/p}$. Since $f(|v_p^1|, |v_p^2|, \ldots, |v_p^d|)$ is an increasing function in terms of each $|v_p^i|$, it can be bounded by replacing each $|v_p^i|$ with the bound given in Eq. (10). Then the truncation error estimate (6) follows immediately. □

Theorem 1 is based on a simple case with only one source point in a box; it can be easily generalized for the case with many source points in a box. Now we restate the lemma given by Greengard and Strain in [1] without proof, and apply a new error bound for the truncated Hermite expansion.

Lemma 2 (Greengard and Strain, 1991). *Let N_B sources y_i with weights q_i lie in a box B with center c_B and side length $r\sqrt{2\delta}$. Then the Gaussian field due to the sources in B,*

$$G(\boldsymbol{x}) = \sum_{i=1}^{N_B} \mathrm{e}^{-\|\boldsymbol{x}-\boldsymbol{y}_i\|_2^2/\delta} q_i, \tag{11}$$

is equal to a single Hermite expansion about c_B:

$$G(\boldsymbol{x}) = \sum_{\alpha \geqslant 0} B_\alpha \boldsymbol{h}_\alpha \left(\frac{\boldsymbol{x} - \boldsymbol{c}_B}{\sqrt{\delta}} \right). \tag{12}$$

The coefficients B_α are given by

$$B_\alpha = \frac{1}{\alpha!} \sum_{i=1}^{N_B} q_i \left(\frac{\boldsymbol{y}_i - \boldsymbol{c}_B}{\sqrt{\delta}} \right)^\alpha. \tag{13}$$

If the Hermite expansion is truncated at order p, the error E(p) satisfies the bound:

$$|E(p)| = \left|\sum_{\alpha \geqslant p} B_\alpha \boldsymbol{h}_\alpha\left(\frac{\boldsymbol{x}-\boldsymbol{c}_B}{\sqrt{\delta}}\right)\right| \leqslant Q_B \sum_{i=1}^{d}\binom{d}{i}\left(Kp^{-1/4}\frac{r_p^p}{1-r_p}\right)^i, \tag{14}$$

where $Q_B = \sum |q_i|$, $K = 1.09(2\pi)^{-1/4}$, *and* $r_p = r\sqrt{e/p}$.

The new error bound given in Theorem 1 is much sharper, especially in three-dimensions, than the one given by Baxter and Roussos in [4]; the latter has the form:

$$|E(p)| \leqslant (1-r)^{-d}\sum_{i=0}^{d-1}\binom{d}{i}(1-r^p)^i\left(\frac{K_1 r^p}{\sqrt{p!}}\right)^{d-i}. \tag{15}$$

If we ignore the high-order terms in both error bounds, we can get the corresponding simple versions

$$|E_{\text{new}}(p)| \leqslant dKp^{-1/4}\frac{r_p^p}{1-r_p} \tag{16}$$

and

$$|E_{\text{BR}}(p)| \leqslant dK_1(1-r)^{-d}(1-r^p)^{d-1}\frac{r^p}{\sqrt{p!}} \tag{17}$$

of our error estimate and Baxter and Roussos's error estimate, respectively, which are the dominant parts of the two error estimates. To compare the two error bounds, we define a function ρ as

$$\rho(p) = \frac{E_{\text{BR}}(p)}{E_{\text{new}}(p)}. \tag{18}$$

It is easy to verify that ρ is an increasing function in terms of p. In Fig. 1, we show the values of $\rho(p)$ at different p when $d = 3$. We can see that $E_{\text{new}}(p)$ is much sharper than $E_{\text{BR}}(p)$, especially when p is relatively large and r is close to 1. Next we check the accuracy of the two error bounds by reconsidering the example in [4]. We assume that there is a single Gaussian at source point $\boldsymbol{y} = (0,0,0)$ with parameter $\delta = 1$ and weight $q = 1$. We evaluate $G(\boldsymbol{x})$ on a $50 \times 50 \times 50$ grid in the unit box $[0,1]^3$. In Table 1, we compare the two error bounds with the L_∞ error on the grid points. It can be seen that $E_{\text{new}}(p)$ is about four times as big as the L_∞ error for different polynomial orders; however, the ratio of $E_{\text{BR}}(p)$ over the L_∞ error keeps growing as p increases. For this case, $E_{\text{BR}}(p)$ overestimates the error by two orders of magnitude while $E_{\text{new}}(p)$ is of the same order as the actual error. Thus, the new error bound is more reliable to be used in controlling the accuracy of FGT.

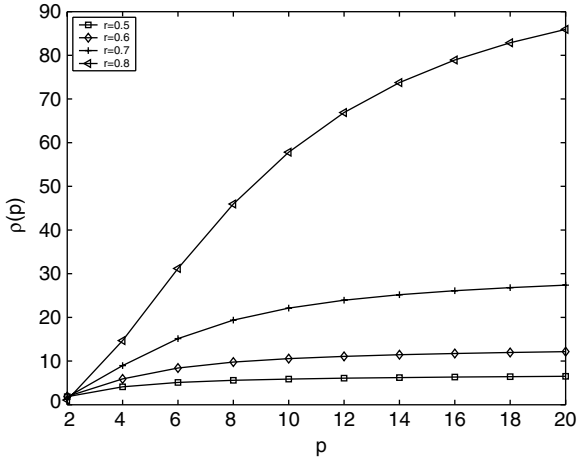

Fig. 1. Ratio of the two error bounds at different polynomial orders for spatial dimension $d = 3$.

Table 1
Comparison of L_∞ error $\epsilon_{\max}$ and error bounds in a cube with a single Gaussian at one vertex of the cube; $r = 1/\sqrt{2}$

p	6	8	10	12	14	16	18
$E_{\text{new}}(p)/\epsilon_{\max}$	4.03	3.89	3.86	3.88	3.92	3.97	4.02
$E_{\text{BR}}(p)/\epsilon_{\max}$	68.02	79.06	87.06	93.46	98.62	103.19	107.16

In addition, the mesh restriction for the new error estimate is $r_p = r\sqrt{\mathrm{e}/p} < 1$ instead of $r < 1$, which means that we can group the points in a relative bigger box if the polynomial order p is sufficiently large. For example, consider the three-dimensional case with $r = 1$ and $\delta = 1$. $E_{\text{BR}}(p)$ will be infinity for this case; however, $E_{\text{new}}(p) = 0.012$ for $p = 8$. This is very important for three-dimensional problems because if we can increase the side length of mesh by a factor η, where $\eta > 1$, the total number of boxes in a cube can be reduced significantly by a factor $1/\eta^3$.

Remark 3. During the review process, John Strain pointed out another paper, see [2], where he derived a similar error estimate of the form

$$|E_p| \leqslant \frac{r_p^p}{1 - r_p}, \tag{19}$$

where $r_p = r\sqrt{\mathrm{e}/(p+1)}$. It can be seen that this error estimate is similar with the error bound of $|v_p^i|$ in Eq. (10) for the *one-dimensional* case. In fact, our error estimate is somewhat sharper in the one-dimensional case. However, this error estimate cannot be applied in high dimensions as it is not mathematically correct when $d > 1$ since it does not contain any information about the dimensionality d. This can be easily verified by an exponential function e^{-x^2} in a box $[-a, a]^d$, whose approximation by low-order Hermite polynomials is

$$\mathrm{e}^{-x^2} = \prod_{i=1}^{d}(1 - x_i^2). \tag{20}$$

Let $a = 0.1$ and check the error at point $\hat{\boldsymbol{x}}$ with $\hat{x}_i = a$, $i = 1, 2, \ldots, d$. The approximation error at $\hat{\boldsymbol{x}}$ is larger than the error estimate given by (19) for $d \geqslant 3$!

2. Application: a fast three-dimensional eigensolver

We present an eigensolver [6] accelerated by FGT for

$$\int_D R_{hh}(\boldsymbol{x}, \boldsymbol{y})\phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \lambda\phi(\boldsymbol{y}), \tag{21}$$

where D is a three-dimensional domain. Such a problem is extensively used in applications, e.g., the Karhunen–Loeve expansion of random inputs for the polynomial chaos methods [7]. Using the Nyström method and a proper set $\{\boldsymbol{y}_i\}$ of quadrature points [8], we obtain the following general eigenproblem

$$\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{B}^{-1}\boldsymbol{v}, \tag{22}$$

where

$$v_i = w_i\phi(\boldsymbol{y}_i), \quad A_{i,j} = R_{hh}(\boldsymbol{y}_i, \boldsymbol{y}_j), \quad B_{i,j} = \delta_{ij}w_i, \quad i, j = 1, 2, \ldots, M.$$

Here $M = 190{,}825$ is the number of unknowns, δ_{ij} is Kronecker delta and w_i are the corresponding integration weights on quadrature points $\boldsymbol{y}_i$.

For an iterative eigensolver, e.g., the implicitly restarted Arnoldi method [9], the essential part is the matrix–vector multiplication, which is also the most time-consuming part. We employ FGT to accelerate the computation of $\boldsymbol{A}\boldsymbol{v}$ in each Arnoldi iteration. According to Lemma 2, we can rewrite the matrix–vector multiplication in Eq. (22) as

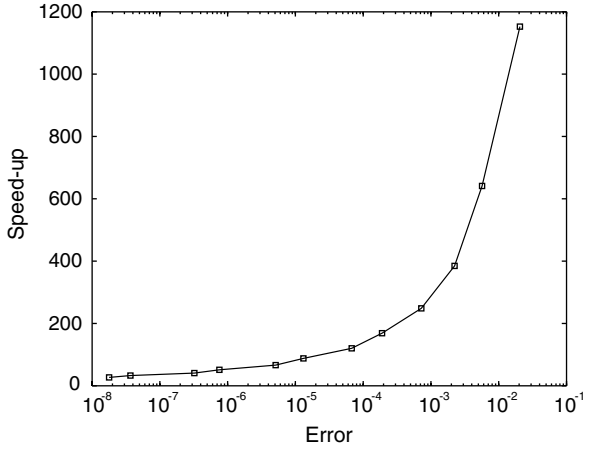

Fig. 2. Speed-up of the fast eigensolver. The solutions given by the direct parallel solver serve as reference ones.

$$\boldsymbol{A}\boldsymbol{v} \mapsto \sum_{j=1}^{M} \mathrm{e}^{-\|\boldsymbol{y}_i-\boldsymbol{y}_j\|_2^2/\delta} q_j = \sum_{\alpha \geqslant 0} B_\alpha \boldsymbol{h}_\alpha\left(\frac{\boldsymbol{y}_i - \boldsymbol{c}_B}{\sqrt{\delta}}\right), \tag{23}$$

where $q_j = \phi(\boldsymbol{y}_j)w_j$. If the Hermite series is truncated after p^d terms, we can form the coefficients B_α first at a cost of $O(p^d M)$ work and evaluate the Hermite series at all quadrature points, with a total cost of $O(2p^d M)$. We note that the memory cost is $O(nM)$ with n being the number of Lanczos basis vectors. To this end, we obtain an $O(M)$ algorithm in both time and memory.

We use boxes with side length $r = 1.06$ to group the grid points. We note that Baxter and Roussos's error estimate does not work for this case since $r > 1$ here. Based on such a "coarse" mesh, savings of about *three orders of magnitude* in time and memory are obtained compared to a direct eigensolver, which is an $O(M^2)$ algorithm due to the direct matrix–vector multiplication. Let $\boldsymbol{\lambda}$ be a vector consisting of the approximate eigenvalues. We assume that $\lambda_i \geqslant \lambda_j$, if $i < j$. We define a normalized error for $\boldsymbol{\lambda}$ as

$$\epsilon = \frac{\|\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathrm{ref}}\|_2}{\|\boldsymbol{\lambda}_{\mathrm{ref}}\|_2}, \tag{24}$$

where $\boldsymbol{\lambda}_{\mathrm{ref}}$ is a reference solution. The speed-up is shown in Fig. 2 with respect to the normalized error ϵ of the first 38 eigenvalues, where the reference solutions are obtained from a parallel eigensolver on 256 processors with direct matrix–vector multiplications.

Acknowledgements

This work was supported by NSF (AMC-SS) and AFOSR.

References

[1] L. Greengard, J. Strain, The fast Gauss transform, SIAM J. Sci. Stat. Comput. 12 (1) (1991) 79–94.
[2] J. Strain, The fast Gauss transform with variable scales, SIAM J. Sci. Stat. Comput. 12 (5) (1991) 1131–1139.
[3] L. Greengard, X. Sun, A new version of the fast Gauss transform, in: Proceedings of the International Congress of Mathematicians III, Extra vol. III, 1998, pp. 575–584.
[4] B.J.C. Baxter, G. Roussos, A new error estimate of the fast Gauss transform, SIAM J. Sci. Comput. 24 (1) (2002) 257–259.
[5] M. Abramowitz, I.A. Stegun, Handbook of Mathematical Functions, Dover, New York, 1970.
[6] X. Wan, Multi-element generalized polynomial chaos for differential equations with random inputs: algorithms and applications, Ph.D. Thesis, Brown University, Division of Applied Mathematics (in preparation).
[7] R.G. Ghanem, P. Spanos, Stochastic Finite Elements: A Spectral Approach, Springer, New York, 1991.
[8] K.E. Atkinson, The Numerical Solution of Integral Equations of the Second Kind, Cambridge University Press, Cambridge, 1997.
[9] R.B. Lehoucq, D.C. Sorensen, Deflation techniques for an implicitly restarted Arnoldi iteration, SIAM Matrix Anal. Applicat. 17 (4) (1996) 789–821.