Science & Technology Facilities Council

Technical Report

RAL-TR-2013-006

Tensor product of random orthogonal matrices

M Arioli

May 2013

RAL Library
STFC Rutherford Appleton Laboratory
R61
Harwell Oxford
Didcot
OX11 0QX

Tel: +44(0)1235 445384
Fax: +44(0)1235 446403
email: libraryral@stfc.ac.uk

Science and Technology Facilities Council reports are available online at: http://epubs.stfc.ac.uk

ISSN 1358- 6254

Version May 22, 2013

TENSOR PRODUCT OF RANDOM ORTHOGONAL MATRICES

M. ARIOLI[1,2]

Abstract. In this short note, we introduce a class of orthogonal matrices of order n for which the matrix by vector product can be be computed in $\mathcal{O}(n \log n)$ instead of $\mathcal{O}(n^2)$. The matrices in this class form a proper Lie subgroup of the set of the orthogonal matrices random generated following the Haar's measure distribution. Given a vector that has the absolute values of its entries presenting large variations of magnitude, the product of a matrix in the subgroup by this vector will produce a new vector where the magnitude of the absolute values of the entries does not vary by a very large amount.

Key words. Random orthogonal matrices, sparse matrices, Gaussian factorization, pivoting

AMS subject classifications. 65F05

1. Introduction. Let $\mathbf{A}$ be an $n \times q$ matrix with real entries $a_{i,j} \in [-1, 1]$. We assume that the entries of the rows, or the columns, are varying in size by several orders of magnitude and we seek orthogonal matrices that can homogenize the entries in $\mathbf{A}$ without changing either the condition number or if $n = q$ the eigenvalues eigenvectors properties. Similar work was described in [2, 6, 7]. In the following, we will denote by $\mathbf{G}(\theta)$ a Givens rotation of order 2 and by $\mathbf{H}_m$ an orthogonal matrix of order m. In particular, we will assume that the $\mathbf{G}(\theta) \in \mathsf{SO}(2)$ (i.e. $\mathbf{det}(\mathbf{G}_i) = 1$) but we will leave $\mathbf{H}_m \in \mathsf{O}(m)$ (i.e. $\mathbf{det}(\mathbf{H}_i) = \pm 1$):

- $$\mathbf{G}(\theta) = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}, \qquad \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}];$$
- $\mathsf{SO}(2)$ is the standard special orthogonal Lie group;
- $\mathsf{O}(m)$ the orthogonal Lie group of the orthogonal matrices (we recall that $\mathsf{O}(m)$ has two connected components).

Finally, we will use the Frobenius (euclidean) topology on the Lie groups of matrices throughout the paper, i.e. let $\mathbf{A} \in \mathbb{R}^n$ and $\mathbf{B} \in \mathbb{R}^n$

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{trace}(\mathbf{A}^T\mathbf{B})$$
$$||\mathbf{A}||_F^2 = \mathbf{trace}(\mathbf{A}^T\mathbf{A}).$$

Moreover, given a matrix $\mathbf{X}(\theta)$ with entries that are functions of a set of parameters $\theta \in \mathbb{R}^p$ we will denote by $\mathbf{dX}$ its differential.

2. Random orthogonal matrices. The need for generating random orthogonal matrices is widely diffused in applied mathematics and physics and a very good survey on the topic can be found in [3]. Here, given a $\mathsf{O}(k)$, we want to generate a matrix that is a random choice with respect to the Haar's measure on $\mathsf{O}(k)$. In [8] and [1], two approaches are presented that achieve this task. Given the property that $\mathsf{O}(k)$

[1]Rutherford Appleton Laboratory, Chilton, Didcot, Oxon, OX11 0QX, UK. e-mail: m.arioli@stfc.ac.uk , Phone: +44 1235 445332
[2]The work was supported by the EPSRC Grant EP/E053351/1.

is a locally compact matrix Lie group, Haar's measure is the natural generalization of the Lebesgue to the orthogonal matrices. The Lebesgue measure is invariant for translations. Given a fixed orthogonal matrix $\mathbf{Q}$, a left (or right) Haar's measure is invariant for the transformations

$$\begin{cases} \mathsf{O}(k) \ni \mathbf{H} \to \mathbf{QH} & \text{(left)} \\ \mathsf{O}(k) \ni \mathbf{H} \to \mathbf{HQ} & \text{(right)} \end{cases}$$

Hereafter, we will use Haar's measures that are both left and right invariant. In particular, we will follow the framework given in [8] in order to generate a realization of an orthogonal stochastic matrix following a density function based on the Haar's measure. In Figure 2.1 we describe the algorithm (ROMG) in detail. We assume that a random number generator for a stochastic real variable following a normal law $\mathbf{N}(0, I)$ is available.

Random Orthogonal Matrix Generator (ROMG) algorithm Let k be order of the matrix to compute and $\mathbf{H} = \mathbf{I}$.
for $i = k : -1 : 2$ **do**
  Generate i realizations of a random normally distributed ($\mathbf{N}(0, I)$) stochastic variable ;
  Organize them as the entries of the vector $\widetilde{\mathbf{v}}_i \in \mathbb{R}^{k-i+1}$
  and let $\mathbf{v} = \begin{bmatrix} 0 \\ \widetilde{\mathbf{v}} \end{bmatrix} \in \mathbb{R}^k$;
  Compute the Householder matrix $\mathbf{H}_i = I - 2\dfrac{\mathbf{v}_i\mathbf{v}_i^T}{||\mathbf{v}_i||_2^2}$
  $\mathbf{H} = \mathbf{H}_i\mathbf{H}$.
end do.

FIG. 2.1. *Random Orthogonal Matrix Generator*

Taking into account the property that every orthogonal matrix $\mathbf{H}$ where each entry is an independent variable, satisfies

$$\mathbf{H}^T\mathbf{H} = \mathbf{I}$$

we have

$$\mathbf{H}^T\mathbf{dH} = -(\mathbf{H}^T\mathbf{dH})^T$$

i.e. $\mathbf{H}^T\mathbf{dH}$ is skew-symmetric. Then, we will assume that [4]

$$\mathbf{H}^T\mathbf{dH} \equiv \bigwedge_{i=1}^{k} \bigwedge_{j=i+1}^{k} \mathbf{H}_{\cdot,j}^T\mathbf{dH}_{\cdot,i},$$

and that $\mathbf{H}^T\mathbf{dH}$ is the differential form linked to the manifold of the orthogonal matrices of order k. Finally, we remember that the differential form $\mathbf{H}^T\mathbf{dH}$ is of maximum degree and this coincides with the number of degrees of freedom of the orthogonal group $k(k-1)/2$.

3. Tensor product Lie subgroup. In order to homogenize the columns and rows of a $n \times n$ matrix A, we can operate the following transformation where H_c and H_r are two realizations of the random orthogonal group with ROMG algorithm:

$$\mathbf{A} \to \mathbf{H}_c \mathbf{A} \mathbf{H}_r. \tag{3.1}$$

It is obvious that the benefit of homogenization will be negligible vis-a-vis the excessive computational cost of (3.1).

Therefore, we propose to use a subgroup of $\mathsf{O}(n)$ by which we can preserve the relevant properties of homogenization of $\mathbf{A}$ at a lower computational cost and with the possibility of reducing further the computational cost by an efficient parallel implementation of the algorithm.

Let us assume that

$$n = 2^p 3^{q_1} 5^{q_2}, \tag{3.2}$$

with p, q_1 and q_2 positive integer numbers, then, using ROMG algorithms, we can generate $G_i \in \mathsf{O}(2)$ for $i = 1, \ldots, p$, and $H_j \in \mathsf{O}(3)$ for $j = 1, \ldots, q_1$ and $W_j \in \mathsf{O}(5)$ for $j = 1, \ldots, q_2$. Then, we propose to use the following matrix Lie subgroup of $\mathsf{O}(n)$:

$$\widetilde{\mathsf{O}}(n) = \left\{ \mathbf{H} \mid \mathbf{H} = \Big(\bigotimes_{i=1}^{p} \mathbf{G}_i \Big) \otimes \Big(\bigotimes_{j=1}^{q_1} \mathbf{H}_j \Big) \otimes \Big(\bigotimes_{i=1}^{q_2} \mathbf{W}_i \Big) \right\} \tag{3.3}$$

The Lie subgroup $\widetilde{\mathsf{O}}(n)$ as a linear vector space has a dimension lower than the full space $\mathsf{O}(n)$ which is $n(n-1)/2$. For the specific choice in (3.2), we have that the dimension is $2p + 3q_1 + 10q_2$.

Remark 3.1. *The choice of the prime numbers $2, 3, 5$ is quite arbitrary and more prime numbers or non prime numbers can be used. Our choice is purely motivated by the desire to give a taste for a general example, without the use of a very complex notation.*

3.1. Factorizations. The matrices of $\widetilde{\mathsf{O}}(n)$ can be factorized as a product of simpler factors. This property will be the key to achieve the target of reducing the complexity. Hereafter, we will denote by $\mathbf{I}_t$ the identity matrix of order t. Each of the three blocks in (3.3), can be factorized as

$$\begin{aligned} \textstyle\bigotimes_{i=1}^{p} \mathbf{G}_i &= \textstyle\prod_{j=1}^{p} \mathbf{I}_{2^{j-1}} \otimes \mathbf{G}_j \otimes \mathbf{I}_{2^{p-j}} \\ \textstyle\bigotimes_{i=1}^{q_1} \mathbf{H}_i &= \textstyle\prod_{j=1}^{q_1} \mathbf{I}_{3^{j-1}} \otimes \mathbf{H}_j \otimes \mathbf{I}_{3^{q_1-j}} \\ \textstyle\bigotimes_{i=1}^{q_2} \mathbf{W}_i &= \textstyle\prod_{j=1}^{q_2} \mathbf{I}_{5^{j-1}} \otimes \mathbf{W}_j \otimes \mathbf{I}_{5^{q_2-j}}; \end{aligned} \tag{3.4}$$

and, then, we have that each $\mathbf{H} \in \widetilde{\mathsf{O}}(n)$ can be written as

$$\begin{aligned} \mathbf{H} = &\Big[\prod_{j=1}^{p} I_{2^{j-1}} \otimes \mathbf{G}_j \otimes \mathbf{I}_{2^{p-j}} \otimes \mathbf{I}_{3^{q_1} 5^{q_2}} \Big] \Big[\mathbf{I}_{2^p} \otimes \prod_{j=1}^{q_1} \mathbf{I}_{3^{j-1}} \otimes \mathbf{H}_j \otimes \mathbf{I}_{3^{q_1-j}} \otimes \mathbf{I}_{5^{q_2}} \Big] \\ &\Big[\mathbf{I}_{2^p 3^{q_1}} \otimes \prod_{j=1}^{q_2} \mathbf{I}_{5^{j-1}} \otimes \mathbf{W}_j \otimes \mathbf{I}_{5^{q_2-j}} \Big] \end{aligned} \tag{3.5}$$

We remark that we could choose a different order in combining the three blocks. However, we think that the increasing order we chose will be easier to implement. The last matrix in (3.5) is block diagonal and, thus, we can use a block version of the matrix by vector product.

3.2. Haar's measure. The topological properties of $\mathsf{O}(n)$ are transferred to $\widetilde{\mathsf{O}}(n)$ that, therefore, is a proper Lie subgroup of matrices. In the following, we prove that the Haar's measure in the smallest atomic spaces are generating a global Haar's measure on $\widetilde{\mathsf{O}}(n)$. First of all, we compute the differential of $H \in \widetilde{\mathsf{O}}(n)$

$$\begin{aligned} dH = H \Bigg\{ & \sum_{j=1}^{p} \Big(\mathbf{I}_{2^{j-1}} \otimes \mathbf{G}_j^T \mathbf{dG}_j \otimes \mathbf{I}_{2^{p-j}} \Big) \otimes \mathbf{I}_{3^{q_1}5^{q_2}} + \\ & \sum_{j=1}^{q_1} \mathbf{I}_{2^p} \otimes \Big(\mathbf{I}_{3^{j-1}} \otimes \mathbf{H}_j^T \mathbf{dH}_j \otimes \mathbf{I}_{3^{q_1-j}} \Big) \otimes \mathbf{I}_{5^{q_2}} + \\ & \sum_{j=1}^{q_2} \mathbf{I}_{2^p 3^{q_1}} \otimes \Big(\mathbf{I}_{5^{j-1}} \otimes \mathbf{W}_j^T \mathbf{dW}_j \otimes \mathbf{I}_{5^{q_2-j}} \Big) \Bigg\} \end{aligned} \tag{3.6}$$

Each $\mathbf{G}_j^T \mathbf{dG}_J$, $\mathbf{H}_j^T \mathbf{dH}_J$, and $\mathbf{W}_j^T \mathbf{dW}_J$ is a skew-symmetric matrix, thus, we have that $\mathbf{H}^T \mathbf{dH}$ is a skew-symmetric matrix. Next, we observe that the left invariance is straight-forward: let $\mathbf{B} \in \widetilde{\mathsf{O}}(n)$ a fixed matrix then

$$\mathbf{H}^T \mathbf{dH} \to (\mathbf{BH})^T \mathbf{d}(\mathbf{BH}) = \mathbf{H}^T \mathbf{B}^T \mathbf{B} \mathbf{dH} = \mathbf{H}^T \mathbf{dH}.$$

The right invariance of $\mathbf{H}^T \mathbf{dH}$ is also a straight-forward consequence of Theorem 2.1.4 in [4]:

$$\mathbf{d}(\mathbf{HB}) = \mathbf{det}(\mathbf{B})^{n-1} \mathbf{dH} = \pm 1 \mathbf{dH}.$$

Therefore, we can conclude that

THEOREM 3.1. *The matrix* $\mathbf{H}$ *produced by ROMG algorithm is a realization of a random orthogonal matrix distributed according to the Haar's measure on* $\widetilde{\mathsf{O}}(n)$.

Finally, we remark that the differential form associated with $\mathbf{H}^T \mathbf{dH}$ is the wedge (external product) of the differential forms of the atomic parts of the Kronecker product. In particular if $n = 2^p$ then we have

$$\mathbf{H}^T \mathbf{dH} = \bigwedge_{i=1}^{p} d\theta_i.$$

3.3. Density functions. The determination of the density function in the general case can be quite tricky. However, if $n = 2^p$ several simplifications are available. In this case, each row (or column) has entries that, neglecting the sign, are a permutation of the entries of the first column. In particular, the $\mathbf{G}_j$ are Givens rotations $\mathbf{G}(\theta_j)$ with $-\pi \le \theta_j \le \pi$. Owing to the independence of each of them, we have that the entries of the first column of H are a product of independent cosine and sine of the stochastic variables θ_j. Therefore, each entry has a density function that is the product of the density functions of the sines and the cosines. Following [5, page 133], the density functions are

$$f_{\cos\theta_j}(y) = \frac{1}{\pi\sqrt{1-y^2}} \qquad \text{and} \qquad f_{\sin\theta_j}(y) = \frac{1}{\pi\sqrt{1-y^2}}.$$

Thus, we have that the density function ρ of the products is

$$\rho = \frac{1}{\pi^p} \prod_{i=1}^{p} \frac{1}{\sqrt{1-y_i^2}}. \tag{3.7}$$

We observe that ρ is defined if and only if $-1 \leq y_i \leq 1$. Therefore we assume that $\rho = 0$ for values of y outside the interval $[-1, 1]$.

4. Homogenisation of a vector. In this Section, we study the problem of the homogenization of vectors having entries the absolute value of which can vary by several orders of magnitude. In this case the meaning of homogenization is to linearly transform these vectors into vectors having entries the absolute value of which is almost constant. Let $\mathbf{u} \in \mathbb{R}^n$, $n = 2^p$, be a real vector with entries u_j and let

$$\mathcal{J} = \{i_1, i_2, \ldots, i_k\} \subset \mathcal{I} = \{1, 2, \ldots, n\}$$

a subset of the indexes for which we have $|u_i| \gg |u_j|$ for $i \in \mathcal{I}$ and $j \in \mathcal{I} \setminus \mathcal{J}$.

The set of vectors $\mathbf{u}$ that have the previous distribution of values for the entries can be homogenized by the use of the random orthogonal matrices introduced above.

Taking into account that

$$\mathbf{u} = \sum_{j \in \mathcal{J}} \mathbf{e}_j u_j + \sum_{k \in \mathcal{I} \setminus \mathcal{J}} \mathbf{e}_k u_k$$

the problem can be reduced to prove that given $\mathbf{G} \in \widetilde{\mathsf{O}}(n)$ the product $\mathbf{g} = \mathbf{G}\mathbf{e}_j$ has elements that are approximatively of the same order of magnitude in absolute values. This is an easy consequence of the results of Section 3.3. The density function ρ is given by (3.7) and we can easily compute for the column $\mathbf{G}_j$ of $\mathbf{G}$:

- the probability Pr that X, an entry in $\mathbf{g}$, be in the interval $[-\tau, \tau]$, $\tau > 0$,

$$Pr\{-\tau \leq X \leq \tau\} = \int_{-\tau}^{\tau} \rho(y)\mathbf{d}y = \frac{2^p}{\pi^p}\big(\arcsin(\tau)\big)^p \Big(\approx \frac{2^p}{\pi^p}\tau^p \text{ if } \tau \ll 1\Big),$$

- the **mean** value μ of one entry in $\mathbf{g}$

$$\mu = \int_{-1}^{1} y\rho(y)\mathbf{d}y = 0,$$

- and the standard deviation σ^2 of one entry in $\mathbf{g}$

$$\sigma^2 = \int_{-1}^{1} (y - \mu)^2 \rho(y)\mathbf{d}y \approx \frac{1}{2^p}.$$

Therefore $\mathbf{v} = \mathbf{G}\mathbf{u}$ is the linear combination of vectors with entries having a distribution with zero mean and $\sigma^2 = 1/2^p$.

5. Computational complexity. The computational complexity for the matrix by vector product of a matrix $\mathbf{H} \in \widetilde{\mathsf{O}}(n)$ with $n = \prod_{j=1}^{k} b_j^{p_j}$, with b_j and p_j positive integer numbers, by a vector $\mathbf{y} \in \mathbb{R}^n$ is given by:

$$n\Big(\sum_{j=1}^{k} b_j p_j\Big) \approx \mathcal{O}(n \ln n).$$

If we compare this complexity with the one of a general full matrix by a vector, i.e. n^2, then we have a substantial advantage.

6. Numerical experiments. We remark that the choice of using bases $2, 3, 5$ is not the only possibility. Other choices such as base 10 can be easier and adequate. In our numerical experiments we have chosen as dimensions $n = 20^2$, $n = 30^2$, and $n = 10^3$ where the basis is respectively 20, 30, and 10. There is another advantage in making this choice: the values of the entries in the random orthogonal matrix are less small in absolute value than those obtained using the base 2. This give a better distribution of the values in the columns and rows of $\mathbf{H} \in \widetilde{\mathsf{O}}(n)$ and in the vector $\mathbf{z} = \mathbf{Hu}$. In order to test the numerical results, we have chosen to compare the results of the matrix by vector product of 100 random generated matrix $\mathbf{H} \in \widetilde{\mathsf{O}}(n)$ by the vector $\mathbf{v} \in \mathbb{R}^n$

$$\mathbf{v}_i = \begin{cases} 100 & i = 10 \\ 450 & i = 45 \\ 1000 & i = n \\ 1 & \text{otherwise,} \end{cases}$$

with the results of the matrix by vector product of a random matrix $\mathbf{M} \in \mathsf{O}(n)$ by the same $\mathbf{v}$. The computational complexity in the two cases is summarized in Table 6.1.

$\mathbf{H}$	$\mathbf{Hv}$	$\mathbf{M}$	$\mathbf{Mv}$
$3\frac{4}{3}10^3 = 4,000$	$30,000$	$\approx 10^9$	$\approx 10^6$

TABLE 6.1
Computational complexity (number of sum+addition operations) for computing $\mathbf{H}$ *and* $\mathbf{M}$ *and their product by* $\mathbf{v}$ *with* $n = 10^3$.

We have chosen as comparison parameters the ratio ρ between the two following quantities that measure how the first entry in $\mathbf{Mv}$ and in $\mathbf{Hv}$ is acceptable as pivot in the Gaussian factorization of an hypothetical matrix where the vector $\mathbf{v}$ is the first column of the Schur factor after some steps. For each of our 100 random matrices, we compute the two parameters as

$$\rho_{\mathbf{M}} = \frac{|(\mathbf{Mv})_1|}{\max_i |(\mathbf{Mv})_i|} \qquad \text{and} \qquad \rho_{\mathbf{H}} = \frac{|(\mathbf{Hv})_1|}{\max_i |(\mathbf{Hv})_i|}.$$

We remark that both $\rho_{\mathbf{M}}$ and $\rho_{\mathbf{H}}$ are independent of the norm of $\mathbf{v}$. Finally, we compute ρ as

$$\frac{\rho_{\mathbf{H}}}{\rho_{\mathbf{M}}}.$$

In Table 6.2, we report the percentages of the cases when $\rho > 1$, $\rho < 1$ and $0.1 \leq \rho \leq 2$. The numerical results show that the tensor approach can be less effective in homogenising an unbalanced vector than using a general random orthogonal matrix. However, both methods are almost equivalent if we look at the percentage of the cases when $\rho \in [0.1, 2]$. In these cases we do not see any major difference in the quality of the result justifying the much larger computational cost of the construction $\mathbf{M}$ and of the corresponding cost for the product $\mathbf{Mv}$.

n	$\rho > 1$	$\rho < 1$	$0.1 \le \rho \le 2$	$\rho < 0.1$
400	41%	59%	68%	12%
900	28%	72%	72%	10%
1000	20%	80%	58%	30%

Table 6.2
Percentages of successful and unsuccessful cases for ρ.

7. Conclusions. We have proposed a method for the homogenisation of an unbalanced vector of order n that has a computational cost much lower than the cost of a classical method using a random orthogonal matrix of order n.

We observed that this tensor product can be less effective than the method generating the full random orthogonal matrix, even if is frequently comparable.

REFERENCES

[1] T. W. Anderson, I. Olkin, and L. G. Underhill, *Generation of random orthogonal matrices*, SIAM J. Sci. Stat. Comput., 8 (1987), pp. 625–629.

[2] M. Baoulin, J. Dongarra, J. Herrmann, and S. Tomov, *Accelerating linear system solutions usin randomization techniques*, Trans. Math. Soft., (to appear).

[3] A. Edelman and N. R. Rao, *Random matrix theory*, Acta Numerica, 14 (2005), pp. 233–297.

[4] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, Wiley, New York, 1982.

[5] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes with Errata Sheet*, McGraw Hill Higher Education, 2002.

[6] D. Parker, *Random butterfly transformations with applications in computational linear algebra.*, Tech. Rep. CSD-950023, Computer Science Department, UCLA., 1995.

[7] D. Parker and B. Pierce, *The randomizing FFT: an alternative to pivoting in Gaussian elimination.*, Tech. Rep. CSD-950037, Computer Science Department, UCLA., 1995.

[8] G. W. Stewart, *The efficient generation of random orthogonal matrices with an application to condition estimators*, SIAM J. Numer. Anal., 17 (1980), pp. 403–409.