Title
Probabilistic Error Analysis for Inner Products

Permalink
https://escholarship.org/uc/item/2bv5z74k

Journal
SIAM Journal on Matrix Analysis and Applications, 41(4)

ISSN
0895-4798

Authors
Ipsen, Ilse CF
Zhou, Hua

Publication Date
2020

DOI
10.1137/19m1270434

Peer reviewed

Published in final edited form as:
SIAM J Matrix Anal Appl. 2020 ; 41(4): 1726–1741. doi:10.1137/19m1270434.

PROBABILISTIC ERROR ANALYSIS FOR INNER PRODUCTS

ILSE C. F. IPSEN[†], **HUA ZHOU**[‡]

[†]Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205 USA.

[‡]Department of Biostatistics, University of California, Los Angeles, CA 90095-1772 USA.

Abstract

Probabilistic models are proposed for bounding the forward error in the numerically computed inner product (dot product, scalar product) between two real n-vectors. We derive probabilistic perturbation bounds as well as probabilistic roundoff error bounds for the sequential accumulation of the inner product. These bounds are nonasymptotic, explicit, with minimal assumptions, and with a clear relationship between failure probability and relative error. The roundoffs are represented as bounded, zero-mean random variables that are independent or have conditionally independent means. Our probabilistic bounds are based on Azuma's inequality and its associated martingale, which mirrors the sequential order of computations. The derivation of forward error bounds "from first principles" has the advantage of producing condition numbers that are customized for the probabilistic bounds. Numerical experiments confirm that our bounds are more informative, often by several orders of magnitude, than traditional deterministic bounds—even for small vector dimensions n and very stringent success probabilities. In particular the probabilistic roundoff error bounds are functions of $\sqrt{n}$ rather than n, thus giving a quantitative confirmation of Wilkinson's intuition. The paper concludes with a critical assessment of the probabilistic approach.

Keywords

roundoff errors; random variables; concentration bounds; martingale; Azuma's inequality

AMS subject classifications.

65G50; 65F30; 60G50; 60G42

1. Introduction.

Probabilistic approaches towards roundoff error analysis have been applied to matrix inversion by von Neumann and Goldstine [19] and Tienari [18], matrix addition and multiplication and Runge–Kutta methods by Hull and Swenson [15], solution of ordinary differential equations by Henrici [12], Gaussian elimination by Barlow and Bareiss [2, 3, 4], convolution and FFT by Calvetti [7, 8, 9], solution of eigenvalue problems by Chatelin and Brunet [5, 6, 10], and LU decomposition and linear system solution by Babuška and

ipsen@ncsu.edu.

Söderlind [1] and Higham and Mary [14]. Yet, the futility of probabilistic roundoff error analysis has also been pointed out [15, page 2], [16, page 17], since roundoffs apparently do not behave like random variables.

Nevertheless, we present probabilistic perturbation and roundoff error bounds for the forward error in the numerically computed inner product,[1]

$$\mathrm{x}^T \mathrm{y} = x_1 y_1 + \cdots + x_n y_n,$$

between two real n-vectors

$$\mathrm{x} = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix}^T \in \mathbb{R}^n \quad \text{and} \quad \mathrm{y} = \begin{pmatrix} y_1 & \cdots & y_n \end{pmatrix}^T \in \mathbb{R}^n .$$

Contributions.

The idea is to represent perturbations and roundoffs as random variables, express the total forward error as a sum of "local" forward errors, and then apply a concentration inequality to the sum. In contrast to some of the previous work, the roundoffs are not required to obey a particular probability distribution. We motivate the particular form of each probabilistic bound with a corresponding deterministic bound and interpret the various random variables in terms of particular forward errors.

Our probabilistic approach is most closely related to that of Higham and Mary [14] who derive backward error bounds. In contrast, our forward error bounds lead to condition numbers customized for the probabilistic bounds and avoid the potentially pessimistic union bound for the probabilities. The bounds are also rigorous, simple, and intuitive, with a clear relationship between failure probability and relative error.

Overview.

To facilitate the introduction of the probabilistic approach, we start as simply as possible, with probabilistic perturbation bounds (section 2). The perturbations are represented as independent, bounded, zero-mean random variables, and the forward error is bounded by Azuma's inequality. This is followed by probabilistic roundoff error bounds for the sequential accumulation of inner products (section 3). The roundoffs are represented as independent, bounded, zero-mean random variables, and the forward error is, again, bounded by Azuma's inequality. However, numerical experiments (section 5) illustrate that for nonnegative vectors of large dimension, the probabilistic expression stops being an upper bound. By way of an explanation, Henrici ends his 1963 paper [12, page 11] with the following:

> The crucial hypothesis for the above statistical theories is the hypothesis of independence of local errors. While this assumption seems to yield realistic results in many cases, some situations are known, [...], where local errors definitely cannot be considered to be independent. To elucidate the conditions under which local

[1] The superscript T denotes the transpose, and for relative bounds we assume $\mathbf{x}^T\mathbf{y} \neq 0$.

> errors act like independent variables would seem to be a fascinating if difficult problem.

We relax the independence assumption and derive a probabilistic error bound for roundoffs with conditonally independent means, based on an Azuma–Hoeffding martigale (section 4). As a consequence, we establish a quantitative confirmation of Wilkinson's intuition [20, section 1.33] that the roundoff error accumulated in n operations is proportional to $\sqrt{n}\ u$ rather than $n\ u$. The paper ends with a critical analysis of the probabilistic approach (section 6).

2. Perturbation bounds.

To motivate the roundoff error bounds and calibrate expectations, we review deterministic perturbation bounds (Lemma 2.1) and present the relevant concentration inequality (Lemma 2.2), followed by the probabilistic perturbation bound (Theorem 2.3).

The Hadamard product

$$\mathrm{x} \circ \mathrm{y} \equiv \left(x_1 y_1 \ \cdots \ x_n y_n\right)^T$$

allows a compact expression of componentwise relative perturbations as

$$\hat{\mathrm{x}} \equiv \begin{pmatrix} (1+\delta_1)x_1 \\ \vdots \\ (1+\delta_n)x_n \end{pmatrix} = \mathrm{x} + \delta \circ \mathrm{x}, \quad \hat{\mathrm{y}} \equiv \begin{pmatrix} (1+\theta_1)y_1 \\ \vdots \\ (1+\theta_n)y_n \end{pmatrix} = \mathrm{y} + \theta \circ \mathrm{y},$$

where $|\delta_k|, |\theta_k| \le u$, $1 \le k \le n$ for some $u > 0$, and the perturbation vectors are

$$\delta \equiv \left(\delta_1 \ \cdots \ \delta_n\right)^T, \quad \theta \equiv \left(\theta_1 \ \cdots \ \theta_n\right)^T.$$

Lemma 2.1. *If* $\frac{1}{p} + \frac{1}{q} = 1$, *then the relative forward error in the perturbed inner product is bounded by*

$$\left|\frac{\hat{\mathrm{x}}^T\hat{\mathrm{y}} - \mathrm{x}^T\mathrm{y}}{\mathrm{x}^T\mathrm{y}}\right| \le \frac{\|\mathrm{x} \circ \mathrm{y}\|_p}{\left|\mathrm{x}^T\mathrm{y}\right|} \|\delta + \theta + \delta \circ \theta\|_q.$$

The special case $p = q = 2$ *gives*

$$\left|\frac{\hat{\mathrm{x}}^T\hat{\mathrm{y}} - \mathrm{x}^T\mathrm{y}}{\mathrm{x}^T\mathrm{y}}\right| \le \sqrt{n}\frac{\|\mathrm{x} \circ \mathrm{y}\|_2}{\left|\mathrm{x}^T\mathrm{y}\right|} u(2+u). \tag{2.1}$$

Proof. The bounds follow from associativity, distributivity, the fact that all quantities are real, and the Hölder inequality. □

The subsequent concentration inequality bounds the deviation of a sum from its mean in terms of the deviations of the individual summands from their means.

Lemma 2.2 (Azuma's inequality, Theorem 5.3 in [11]). *Let* $Z \equiv Z_1 + \cdots + Z_n$ *be a sum of independent random variables* $Z_1, \ldots, Z_n$ *with*

$$|Z_k - \mathbb{E}[Z_k]| \le c_k, \quad 1 \le k \le n.$$

Then for any $0 < \delta < 1$, *with probability at least* $1 - \delta$,

$$|Z - \mathbb{E}[Z]| \le \sqrt{\sum_{k=1}^{n} c_k^2}\sqrt{2\ln(2/\delta)}.$$

Proof. In [11, Theorem 5.3] set

$$\delta \equiv \Pr[|Z - \mathbb{E}[Z]| \ge t] \le 2\exp\left(-\frac{t^2}{2\sum_{k=1}^{n} c_k^2}\right),$$

and solve for t in terms of δ. If $|Z - \mathbb{E}[Z]| \ge t$ holds with probability at most δ, then the complementary event $|Z - \mathbb{E}[Z]| \le t$ holds with probability at least $1 - \delta$. □

Lemma 2.2 implies that, with high probability, a sum Z is close to its mean $\mathbb{E}[Z]$ if each summand Z_k is close to its mean $\mathbb{E}[Z_k]$. In the probabilistic perturbation bound below, the perturbations δ_k and θ_k are represented as independent, bounded random variables with zero mean.

Theorem 2.3. *Let* δ_k, θ_k *be independent random variables with* $\mathbb{E}[\delta_k] = \mathbb{E}[\theta_k] = 0$ *and* $|\delta_k|, |\theta_k| \le u$, $1 \le k \le n$. *Then for any* $0 < \delta < 1$, *with probability at least* $1 - \delta$,

$$\left|\frac{\hat{x}^T\hat{y} - x^Ty}{x^Ty}\right| \le \frac{\|x \circ y\|_2}{|x^Ty|}\sqrt{2\ln(2/\delta)}\,u(2+u).$$

Proof. Write the total forward error

$$Z \equiv \hat{x}^T\hat{y} - x^Ty = Z_1 + \cdots + Z_n$$

as a sum of independent random variables, where each summand Z_k represents a componentwise forward error,

$$Z_k \equiv x_k y_k((1+\delta_k)(1+\theta_k) - 1) = x_k y_k(\delta_k + \theta_k + \delta_k\theta_k), \quad 1 \le k \le n.$$

From the linearity of the mean, and δ_k and θ_k being independent zero-mean random variables, follows that the componentwise errors have zero mean,

$$\mathbb{E}[Z_k] = x_k y_k (\mathbb{E}[\delta_k] + \mathbb{E}[\theta_k] + \mathbb{E}[\delta_k]\mathbb{E}[\theta_k]) = 0, \quad 1 \le k \le n.$$

The boundedness of δ_k and θ_k implies that the deviation of Z_k from its mean $\mathbb{E}[Z_k] = 0$ is bounded by

$$|Z_k - \mathbb{E}[Z_k| = |Z_k| = |x_k y_k| \, |\delta_k + \theta_k + \delta_k \theta_k| \le c_k \equiv |x_k y_k| \tau, \quad 1 \le k \le n,$$

where $\tau \equiv 2u + u^2 = u(2 + u)$. Therefore, the conditions of Lemma 2.2 are satisfied, and we have

$$\sum_{k=1}^{n} c_k^2 = \sum_{k=1}^{n} |x_k y_k|^2 \tau 2 = \|\mathrm{x} \circ \mathrm{y}\|_2^2 \tau^2$$

Linearity implies that the total error also has zero mean,

$$\mathbb{E}[\hat{\mathrm{x}}^T \hat{\mathrm{y}} - \mathrm{x}^T \mathrm{y}] = \mathbb{E}[Z] = \mathbb{E}[Z_1] + \cdots + \mathbb{E}[Z_n] = 0.$$

Apply Lemma 2.2 to conclude that for any $0 < \delta < 1$, with probability at least $1 - \delta$,

$$|\hat{\mathrm{x}}^T \hat{\mathrm{y}} - \mathrm{x}^T \mathrm{y}| = |Z - \mathbb{E}[Z]| \le \|\mathrm{x} \circ \mathrm{y}\|_2 \sqrt{2 \ln (2/\delta)} \tau.$$

At last, divide both sides of the inequality by the constant $|\mathbf{x}^T \mathbf{y}|$. □

Remark 2.1. The probabilistic bound in Theorem 2.3 is by a factor of $\sqrt{n}$ tighter than the deterministic bound (2.1).

The two bounds differ in the factors $\sqrt{2 \ln (2/\delta)}$ versus $\sqrt{n}$, which implies the following:

1. The deterministic bound depends explicitly on the dimension n, while the probabilistic bound does not.
2. The probabilistic bound is tighter than the deterministic bound for $n > 2 \ln (2/\delta)$. Specifically, with a tiny failure probability of $\delta = 10^{-16}$, the probabilistic bound is tighter for $n > 76$, and $\sqrt{2 \ln (2/\delta)} \le 9$

3. Bounds for independent roundoffs.

After presenting the model for independent roundoffs (section 3.1), we derive a motivating deterministic bound (section 3.2), followed by our probabilistic bound (section 3.3).

3.1. Roundoff error model.

We assume that the elements of **x** and **y** are floating point numbers; thus can be stored exactly. The inner product is computed via recursive summation [13, section 4.1], by accumulating partial sums sequentially from left to right,

$$z_1 = x_1 y_1, \quad z_{k+1} = \sum_{j=1}^{k+1} x_j y_j, \quad 1 \le k \le n-1.$$

The roundoff error model in Table 3.1 corresponds to [13, (3.1) and (3.2)].

For $0 < u < 1$ and $k \geq 1$, we use the abbreviation

$$\gamma_k \equiv (1+u)^k - 1 = ku + \mathcal{O}\left(u^2\right). \tag{3.1}$$

If $ku < 1$, then [13, Lemma 3.1]

$$\gamma_k \le \frac{ku}{1-ku}. \tag{3.2}$$

In subsequent sections, we derive relative error bounds. Although Wilkinson [20, section I.25] remarks that “for extended sequences of additions it is no longer true that the computed value necessarily has a low relative error,” the numerical experiments in section 5 clearly illustrate that the computed inner product is accurate to several significant digits, even for very high dimensions.

3.2. A motivating deterministic bound.

We present an expression for the absolute forward error (Lemma 3.1), followed by two bounds for the relative error (Theorem 3.2).

The total forward error is expressed as a sum of “local” forward errors, one for each summand.

LEMMA 3.1. *The forward error for* $\hat{z}_n = \mathrm{fl}\left(\mathrm{x}^T \mathrm{y}\right)$ *in* Table 3.1 *equals*

$$\mathrm{fl}\left(\mathrm{x}^T \mathrm{y}\right) - \mathrm{x}^T \mathrm{y} = \hat{z}_n - z_n = Z_1 + \cdots + Z_n$$

with local forward errors

$$Z_1 \equiv x_1 y_1 \left((1+\theta_1) \prod_{\ell=2}^{n} (1+\delta_\ell) - 1 \right),$$

$$Z_k \equiv x_k y_k \left((1+\theta_k) \prod_{\ell=k}^{n} (1+\delta_\ell) - 1 \right), \quad 2 \le k \le n.$$

If $|\delta_k|, |\theta_k| \le u$ and γ_k *as in* (3.1), $1 \le k \le n$, *then*

$$|Z_1| \le c_1 \equiv |x_1 y_1| \gamma_n,$$

$$|Z_k| \leq c_k \equiv |x_k y_k| \gamma_{n-k+2}, \quad 2 \leq k \leq n.$$

Lemma 3.1 implies two bounds for the relative forward error, the first being the traditional bound [13, section 3.1], and the second being the motivation for our subsequent probabilistic bound.

THEOREM 3.2. *Let* $|\delta_k|, |\theta_k| \leq u$, γ_k *as in* (3.1), *and* c_k *as in Lemma* 3.1, $1 \leq k \leq n$ *Then the relative error for* $\hat{z}_n = \mathrm{fl}(\mathrm{x}^T\mathrm{y})$ *in* Table 3.1 *is bounded by*

$$\left|\frac{\mathrm{fl}(\mathrm{x}^T\mathrm{y}) - \mathrm{x}^T\mathrm{y}}{|\mathrm{x}^T\mathrm{y}|}\right| \leq \frac{|\mathrm{x}|^T|\mathrm{y}|}{|\mathrm{x}^T\mathrm{y}|}\gamma_n \tag{3.3}$$

and by

$$\left|\frac{\mathrm{fl}(\mathrm{x}^T\mathrm{y}) - \mathrm{x}^T\mathrm{y}}{|\mathrm{x}^T\mathrm{y}|}\right| \leq \frac{\sqrt{\sum_{k=1}^n c_k^2}}{|\mathrm{x}^T\mathrm{y}|}\sqrt{n}. \tag{3.4}$$

Proof. Lemma 3.1 implies for the absolute error that

$$|\mathrm{fl}(\mathrm{x}^T\mathrm{y}) - \mathrm{x}^T\mathrm{y}| = |\hat{z}_n - z_n| \leq \sum_{k=1}^{n} c_k = |x_1 y_1|\gamma_n + \sum_{k=2}^{n} |x_k y_k|\gamma_{n-k+2}.$$

For the first bound, apply the Hölder inequality to $\sum_{k=1}^n c_k = \mathrm{v}^T\mathrm{g}$, where

$$\mathrm{v} \equiv \left(|x_1 y_1| \;\cdots\; |x_n y_n|\right)^T, \quad \mathrm{g} \equiv \left(\gamma_n \;\gamma_n \;\gamma_{n-1} \;\cdots\; \gamma_2\right)^T.$$

For the second bound, apply the norm relation $\sum_{k=1}^n c_k = \|\mathrm{c}\|_1 \leq \|\mathrm{c}\|_2\sqrt{n}$ to

$$\mathbf{c} \equiv \left(c_1 \;\cdots\; c_n\right)^T.$$

3.3. Probabilistic bound.

We model the roundoffs as independent, bounded zero-mean random variables and use Azuma's inequality in Lemma 2.2.

THEOREM 3.3. *Let* δ_k, θ_k *be independent random variables with* $\mathbb{E}[\delta_k] = \mathbb{E}[\theta_k] = 0$ *and* $|\delta_k|, |\theta_k| \leq u$, γ_k *as in* (3.1)*; and* c_k *as in Lemma* 3.1, $1 \leq k \leq n$. *Then for any* $< \delta < 1$, *with probability at least* $1 - \delta$,

$$\left|\frac{\mathrm{fl}(\mathrm{x}^T\mathrm{y}) - \mathrm{x}^T\mathrm{y}}{\mathrm{x}^T\mathrm{y}}\right| = \left|\frac{\hat{z}_n - z_n}{z_n}\right| \leq \frac{\sqrt{\sum_{k=1}^n c_k^2}}{|\mathrm{x}^T\mathrm{y}|}\sqrt{2\ln(2/\delta)}.$$

Proof. Since the roundoffs are independent random variables, so is the total forward error in Lemma 3.1,

$$Z \equiv Z_1 + \cdots + Z_n = \mathrm{fl}\left(\mathrm{x}^T \mathrm{y}\right) - \mathrm{x}^T \mathrm{y}\,.$$

The random variables

$$Z_1 \equiv x_1 y_1 \left((1+\theta_1) \prod_{\ell=2}^{n} (1+\delta_\ell) - 1 \right),$$

$$Z_k \equiv x_k y_k \left((1+\theta_k) \prod_{\ell=k}^{n} (1+\delta_\ell) - 1 \right), \quad 2 \le k \le n,$$

represent the local forward errors and have zero mean, $\mathbb{E}[Z_k] = 0$. By linearity, the total forward error also has zero mean, $\mathbb{E}$ [Z]=0. The deviations of the local errors from their means are bounded by

$$|Z_k - \mathbb{E}[Z_k]| = |Z_k| \le c_k, \quad 1 \le k \le n\,.$$

Apply Lemma 2.2 to *Z*, and divide both sides by the constant $/\mathbf{x}^T\mathbf{y}/$. □

Remark 3.1. The probabilistic bound in Theorem 3.3 tends to be tighter than the deterministic bound (3.4) in Theorem 3.2.

As in Remark 2.1, the two bounds differ in the factors $\sqrt{2\ln(2/\delta)}$ versus $\sqrt{n}$, and the same conclusions apply. This is also illustrated by numerical experiments in section 5.2.

4. Bounds for roundoffs with conditionally independent means.

After presenting the roundoff error model (section 4.1), we derive a motivating deterministic bound (section 4.2) and then present a probabilistic bound with relaxed assumptions (section 4.3), followed by a simpler bound (section 4.4).

4.1. Roundoff error model.

As in section 3.1, we assume that the elements of **x** and **y** are floating point numbers, hence can be stored exactly. Our model in Table 4.1 differs from the traditional model in Table 3.1 only in the bookkeeping. It distinguishes each step that introduces a roundoff and explicitly separates additions (+) from multiplications (*). There are n multiplications and $n - 1$ additions, $2n - 1$ distinct roundoffs.

The model in Table 4.1 is designed to do without additional intermediate factors like $x_k y_k(1 + \delta_{2k-2})$ and is expressed solely in terms of partial sums. Since we assume a guard digit model without fused multiply-add, the roundoff for addition can be recorded in a subsequent step. The very first partial sum incurs no addition, so we allocate the roundoff to the second partial sum for easier indexing.

4.2. A motivating deterministic bound.

We bound the incremental errors in the partial sums (Lemma 4.1) and then present a deterministic bound (Theorem 4.2) to motivate our subsequent probabilistic bounds.

The *total* forward error is

$$Z_{2n} \equiv \hat{s}_{2n} - s_{2n} = \mathrm{fl}\left(\mathrm{x}^T\mathrm{y}\right) - \mathrm{x}^T\mathrm{y}, \tag{4.1}$$

while the *partial sum forward errors* are

$$Z_1 = 0 \quad \text{and} \quad Z_k \equiv \hat{s}_k - s_k, \quad 1 \le k \le 2n\,.$$

The partial sum errors Z_k distinguish the newly arrived roundoff from the previous roundoffs. The difference between two successive partial sum errors Z_k and Z_{k-1} captures the most recent roundoff on the way from Z_{k-1} to Z_k. Below we state the recursion for Z_k and bound the "incremental error" by $c_k\, u$, where the c_k are different from but the same in spirit as the ones in Lemma 3.1.

Lemma 4.1. *The forward errors for the partial sums in* Table 4.1 *satisfy*

$$Z_{2k} = Z_{2k-1} + \hat{s}_{2k-1}\delta_{2k-1}, \quad 1 \le k \le n,$$

$$Z_{2k-1} = Z_{2k-2} + x_k y_k \delta_{2k-2}, \quad 2 \le k \le n\,.$$

If $|\delta_k|, \le u$, $1 \le k \le 2n - 1$, *then the incremental errors satisfy*

$$|Z_{k+1} - Z_k| \le c_k\, u, \quad 1 \le k \le 2n - 1,$$

where

$$c_{2k-1} \equiv |x_1 y_1|(1+u)^{k-1} + \sum_{j=2}^{k} |x_j y_j|(1+u)^{k-j+1}, \quad 1 \le k \le n,$$

$$c_{2k-2} \equiv |x_k y_k|, \quad 2 \le k \le n\,.$$

Proof. A first induction establishes bounds for the partial sums,

$$|\hat{s}_{2k-1}| \le |x_1y_1|(1+u)^{k-1} + |x_2y_2|(1+u)^{k-1} + \cdots + |x_ky_k|(1+u) = |x_1y_1|(1+u)^{k-1} + \sum_{j=2}^{k} |x_jy_j|(1+u)^{k-j+1}, \quad 1 \le k \le n,$$

$$|\hat{s}_{2k}| \le |x_1y_1|(1+u)^{k} + |x_2y_2|(1+u)^{k} + \cdots + |x_ky_k|(1+u)^2 = |x_1y_1|(1+u)^{k} + \sum_{j=2}^{k} |x_jy_j|(1+u)^{k-j+2}, \quad 1 \le k \le n,$$

while a second induction establishes bounds for the incremental errors. □

Lemma 4.1 implies a deterministic bound for the relative forward error.

Theorem 4.2. *If* $|\delta_k|, \le u$, *and* c_k *as in Lemma* 4.1, $1 \le k \le 2n-1$, *then the relative error for* $\hat{s}_{2n} = \mathrm{fl}(\mathrm{x}^T\mathrm{y})$ *in* 4.1 *bounded by*

$$\left|\frac{\mathrm{fl}(\mathrm{x}^T\mathrm{y}) - \mathrm{x}^T\mathrm{y}}{\mathrm{x}^T\mathrm{y}}\right| \le \sqrt{2n-1}\,\frac{\sqrt{\sum_{k=1}^{2n-1} c_k^2}}{|\mathrm{x}^T\mathrm{y}|}\,u\,.$$

Proof. Represent the total error (4.1) as a telescoping sum of incremental errors

$$\mathrm{fl}(\mathrm{x}^T\mathrm{y}) - \mathrm{x}^T\mathrm{y} = Z_{2n} = (Z_{2n} - Z_{2n-1}) + (Z_{2n-1} - Z_{2n-2}) + \cdots + (Z_2 - Z_1),$$

where $Z_1 = 0$. The expressions for c_k from Lemma 4.1 give the bound

$$|Z_{2n}| \le \underbrace{|Z_{2n} - Z_{2n-1}|}_{\le c_{2n-1}u} + \underbrace{|Z_{2n-1} - Z_{2n-2}|}_{\le c_{2n-2}u} + \cdots + \underbrace{|Z_2 - Z_1|}_{\le c_1 u} \le \sum_{k=1}^{2n-1} c_k\,u\,.$$

At last, apply the relation between the vector one- and two-norms. □

4.3. Probabilistic bound.

After defining an Azuma martingale (Definition 4.3) and customizing it to our context (Lemma 4.4), we present a probabilistic bound that only requires the mean of roundoffs to be independent of previous roundoffs (Theorem 4.5), followed by a comparison with the deterministic bounds (Remark 4.1).

Definition 4.3 (martingale, Definition 12.1 in [17]). *A sequence of random variables* Z_1, Z_2 … *is a martingale with respect to a sequence* $\delta_1, \delta_2, \ldots$ *if for* $k \ge 1$

1. Z_k *is a function of* $\delta_1, \delta_2, \ldots \delta_{k-1}$,
2. $\mathbb{E}[|Z_k|] < \infty$,
3. $\mathbb{E}[Z_{k+1} | \delta_1, \ldots, \delta_{k-1}] = Z_k$.

The version of the martingale below is tailored to our context.

Lemma 4.4 (Azuma–Hoeffding martingale, Theorem 12.4 in [17]). *Let $Z_1, \ldots, Z_{2n}$ be a martingale with*

$$|Z_k - Z_{k-1}| \le c_{k-1}, \quad 2 \le k \le 2n.$$

Then for any $0 < \delta < 1$ *with probability at least* $1 - \delta$,

$$|Z_{2n} - Z_1| \le \sqrt{\sum_{k=1}^{2n-1} c_k^2}\sqrt{2\ln(2/\delta)}.$$

Proof. In [17, Theorem 12.4], set $m = 2n - 1$ and

$$\delta \equiv \Pr[|Z_m - Z_0| \ge t] \le 2\exp\left(-\frac{t^2}{2\sum_{k=1}^{m} c_k^2}\right).$$

Solve for t in terms of δ. If $|Z - \mathbb{E}[Z]| \ge t$ holds with probability at most δ, then the complementary event $|Z - \mathbb{E}[Z]| \le t$ holds with probability at least $1 - \delta$. □

Again, we represent the roundoffs as bounded, zero-mean random variables, but now the assumption on independence is relaxed: The conditional mean of a roundoff does not depend on the previous roundoffs. The following bound resembles Theorem 3.3 but contains more summands.

Theorem 4.5. *Let δ_k be random variables with $|\delta_k| \le u$ and*

$$0 = \mathbb{E}[\delta_k] = \mathbb{E}[\delta_k | \delta_1, \ldots, \delta_{k-1}], \quad 2 \le k \le 2n - 1, \tag{4.2}$$

and c_k as in Lemma 4.1. *Then for any* $0 < \delta < 1$, *with probability at least* $1 - \delta$,

$$\left|\frac{\mathrm{fl}\left(\mathrm{x}^T\mathrm{y}\right) - \mathrm{x}^T\mathrm{y}}{\mathrm{x}^T\mathrm{y}}\right| \le \frac{\sqrt{\sum_{k=1}^{2n-1} c_k^2}}{|\mathrm{x}^T\mathrm{y}|}\sqrt{2\ln(2/\delta)}u.$$

Proof. Since $Z_1 = 0$, Table 4.1 implies for the total forward error (4.1) that

$$|\mathrm{fl}\left(\mathrm{x}^T\mathrm{y}\right) - \mathrm{x}^T\mathrm{y}| = |\hat{s}_{2n} - s_{2n}| = |Z_{2n}| = |Z_{2n} - Z_1|.$$

To apply Lemma 4.4, we show that the partial sum forward errors $Z_1, Z_2, \ldots, Z_{2n}$ form a martingale with respect to the roundoffs $\delta_1, \ldots, \delta_{2n-1}$. To this end, we need to check the conditions in Definition 4.3 and Lemma 4.4.

1. The recursions in Lemma 4.1 show that Z_k is a function of the roundoffs $\delta_1, \ldots, \delta_{k-1}$, $2 \le k \le 2n$.

2. The expectation of $|Z_k|$ is finite summands, roundoffs because $/Z_k/$ is a finite sum of bounded and the have zero mean.

3. Lemma 4.1 implies $Z_2 = Z_1 + x_1 y_1\ \delta_1$ with $Z_1 = 0$ a constant. From (4.2) follows

$$\mathbb{E}[Z_2] = \mathbb{E}[Z_1 + x_1 y_1 \delta_1] = Z_1 + x_1 y_1 \mathbb{E}[\delta_1] = Z_1 .$$

More generally, consider the first set of recursions from Lemma 4.1,

$$\begin{aligned}&\mathbb{E}[Z_{2k-1} | \delta_1, \ldots, \delta_{2k-3}] = \mathbb{E}[Z_{2k-2} + x_k y_k \delta_{2k-2} | \delta_1, \ldots, \delta_{2k-3}] \\ &= \mathbb{E}[Z_{2k-2} | \delta_1, \ldots, \delta_{2k-3}] + x_k y_k \mathbb{E}[\delta_{2k-2} | \delta_1, \ldots, \delta_{2k-3}] = Z_{2k-2}, \quad 2 \le k \le n,\end{aligned}$$

where the last equality follows from (4.2) and from the dependence of Z_{2k-2} on $\delta_1, \ldots, \delta_{2k-3}$ in item 1.

Now consider the second set of recursions from Lemma 4.1,

$$\begin{aligned}&\mathbb{E}[Z_{2k} | \delta_1, \ldots, \delta_{2k-2}] = \mathbb{E}[Z_{2k-1} + \hat{s}_{2k-1}\delta_{2k-1} | \delta_1, \ldots, \delta_{2k-2}] \\ &= \mathbb{E}[Z_{2k-1} | \delta_1, \ldots, \delta_{2k-2}] + \mathbb{E}[\hat{s}_{2k-1}\delta_{2k-1} | \delta_1, \ldots, \delta_{2k-2}] \\ &= Z_{2k-1}, \quad 1 \le k \le n,\end{aligned}$$

where as above the last equality follows from (4.2) and from the dependence of Z_{2k-1} on $\delta_1, \ldots, \delta_{2k-2}$ in item 1.

Thus, $Z_1, Z_2, \ldots, Z_{2n}$ form a martingale with respect to $\delta_1, \ldots, \delta_{2n-1}$.

4. Lemma 4.1 implies $/Z_{k+1} - Z_k| \le c_k u$, $1 \le k \le 2n - 1$.

Thus, the assumptions for Lemma 4.4 are satisfied, and we can use it to bound $/Z_{2n} - Z_1/$ with the c_k from Lemma 4.1. □

Remark 4.1. The probabilistic bound in Theorem 4.5 tends to be tighter than the deterministic bound in Theorem 4.2.

The two bounds differ in the factors $\sqrt{2\ln(2/\delta)}$ versus $\sqrt{2n-1}$, which implies the following:

1. The deterministic bound increases with the dimension n, while the probabilistic bound does not.

2. The probabilistic bound is tighter for $n > \ln(2/\delta) + \frac{1}{2}$. Specifically, with a tiny failure probability of $\delta = 10^{-16}$, the probabilistic bound is tighter for $n \ge 39$, and $\sqrt{2\ln(2/\delta)} \le 9$.

4.4. Simpler probabilistic bound.

With the help of a bound on the sum of incremental errors (Lemma 4.6), we derive a simpler probabilistic bound (Corollary 4.7) and compare it to the corresponding deterministic bound (Remark 4.2), thereby confirming Wilkinson’s intuition [20, section 1.33].

The following compact bound for the incremental errors makes use of abbreviations for the leading subvectors of $|\mathbf{x}| \circ |\mathbf{y}|$ and vectors with powers of $1 + u$.

LEMMA 4.6. *Define the k-vectors*

$$(\mathrm{x} \circ \mathrm{y})_k \equiv \begin{pmatrix} |x_1 y_1| \\ |x_2 y_2| \\ \vdots \\ |x_k y_k| \end{pmatrix}, \quad \mathbf{u}_k \equiv \begin{pmatrix} (1+u)^{k-1} \\ (1+u)^{k-1} \\ \vdots \\ 1+u \end{pmatrix}, \quad 2 \le k \le n.$$

If $\frac{1}{p} + \frac{1}{q} = 1$, *then the* c_k *in Lemma* 4.1 *satisfy*

$$\sum_{k=1}^{2n-1} c_k^2 \le \|\mathrm{x} \circ \mathrm{y}\|_2^2 + \sum_{k=2}^{n} \|(\mathrm{x} \circ \mathrm{y})_k\|_p^2 \|\mathbf{u}_k\|_q^2 .$$

Proof. Partition

$$\sum_{k=1}^{2n-1} c_k^2 = \sum_{k=1}^{n} c_{2k-1}^2 + \sum_{k=2}^{n} c_{2k-2}^2 = \sum_{k=2}^{n} c_{2k-1}^2 + c_1^2 + \sum_{k=2}^{n} c_{2k-2}^2 .$$

From $c_1 = |x_1 y_1|$ and $c_{2k-2} = |x_k y_k|$, $2 \le k \le n$, follows

$$c_1^2 + \sum_{k=2}^{n} c_{2k-2}^2 = \sum_{k=1}^{n} |x_k y_k|^2 = \|x \circ y\|_2^2 .$$

Thus $\sum_{k=1}^{2n-1} c_k^2 = \|\mathrm{x} \circ \mathrm{y}\|_2^2 + \sum_{k=2}^{n} c_{2k-1}^2$. In the remaining sum, apply Hölder's inequality to each summand,

$$\begin{aligned} c_{2k-1} &= |x_1 y_1| (1+u)^{k-1} + \sum_{j=2}^{k} |x_j y_j| (1+u)^{k-j+1} \\ &= (\mathrm{x} \circ \mathrm{y})_k^T \mathbf{u}_k \le \|(\mathrm{x} \circ \mathrm{y})_k\|_p \|\mathbf{u}_k\|_q, \quad 2 \le k \le n. \end{aligned}$$

□

Lemma 4.6 implies a simple bound for Theorem 4.5.

COROLLARY 4.7. *Let* δ_k *be random variables as in Theorem* 4.5, *and* γ_k *as in* (3.1), $1 \le k \le 2n - 1$. *Then for any* $0 < \delta < 1$, *with probability at least* $1 - \delta$,

$$\left| \frac{\mathrm{fl}\left(\mathrm{x}^T \mathrm{y}\right) - \mathrm{x}^T \mathrm{y}}{\mathrm{x}^T \mathrm{y}} \right| \le \frac{|\mathrm{x}|^T |\mathrm{y}|}{|\mathrm{x}^T \mathrm{y}|} \sqrt{2 \ln (2/\delta)} \sqrt{\frac{u\, \gamma_{2n}}{2}} . \tag{4.3}$$

Proof. Choose $p = 1$ and $q = \infty$ in Lemma 4.6,

$$\|(x \circ y)_k\|_1 \|\mathbf{u}_k\|_\infty \le \left\| x \circ y \right\|_1 (1+u)^{k-1}, \quad 2 \le k \le n.$$

Insert $\|\mathbf{x} \bigcirc \mathbf{y}/_2 \le \|\mathbf{x} \bigcirc \mathbf{y}\|_1$ into Lemma 4.6,

$$\sum_{k=1}^{2n-1} c_k^2 \le \|x \circ y\|_2^2 + \sum_{k=2}^{n} \|(x \circ y)_k\|_1^2 \|\mathbf{u}_k\|_\infty^2 \le \|x \circ y\|_1^2 \left(1 + \sum_{k=2}^{n} (1+u)^{2(k-1)}\right).$$

The second factor is a geometric sum,

$$1 + \sum_{k=1}^{n-1} (1+u)^{2k} = \sum_{k=0}^{n-1} (1+u)^{2k} = \frac{(1+u)^{2n} - 1}{(1+u)^2 - 1} = \frac{\gamma_{2n}}{u^2 + 2u}.$$

Combining the preceding inequalities gives

$$\sqrt{\sum_{k=1}^{2n-1} c_k^2} \le \|x \circ y\|_1 \sqrt{\frac{\gamma_{2n}}{u^2 + 2u}} \le \|x \circ y\|_1 \sqrt{\frac{\gamma_{2n} u}{2}}.$$

At last, substitute this into Theorem 4.5. □

Remark 4.2. We quantify and confirm Wilkinson's intuition [20, section 1.33], by illustrating that the probabilistic bound in Corollary 4.7 is proportional to $\sqrt{n u}$, while the deterministic bound (3.3) in Theorem 3.2 is proportional to $n\,u$.

The bounds have the same condition number but differ in the other factor: γ_n in Theorem 3.2 versus $\sqrt{2\ln(2/\delta)}\sqrt{u\,\gamma_{2n}/2}$ in Corollary 4.7. For large n, (3.1) implies that the bounds behave asymptotically like their first order terms,

$$\gamma_n \approx n\,u, \quad \sqrt{u\,\gamma_{2n}/2} \approx \sqrt{n}\,u.$$

For small n with $2n\,u < 1$, (3.2) implies

$$\gamma_n \le \frac{nu}{1 - nu}, \quad \sqrt{\frac{u\,\gamma_{2n}}{2}} \le \frac{\sqrt{n}\,u}{\sqrt{1 - 2n\,u}}.$$

Thus, the probabilistic bound in Corollary 4.7 is proportional to $\sqrt{n}\;u$.

Furthermore, $\gamma_n > \sqrt{u\,\gamma_{2n}/2}$ for $n \ge 2$. With a failure probability of $\delta = 10^{-16}$, the probabilistic bound is tighter than the deterministic bound for $n > 76$.

5. Numerical experiments.

After describing the setup for the experiments (section 5.1), we present experiments for the probabilistic bound with independent roundoffs (section 5.2) and the one with more relaxed assumptions (section 5.3).

5.1. Experimental setup.

We use a tiny failure probability of $\delta = 10^{-16}$, which gives a probabilistic factor of $\sqrt{2\ln(2/\delta)} \le 8.7$.

Two types of vectors **x** and **y** of dimension up to $n = 10^8$ will be considered:

- The elements of **x** and **y** can have different signs. Specifically, x_j and y_j are iid[2] standard normal random variables with mean 0 and variance 1, and **x** and **y** are generated with the MATLAB commands

$$\mathrm{x} = \mathrm{single}\,(\mathrm{randn}\,(\mathrm{n},\ 1)),\ \mathrm{y} = \mathrm{single}\,(\mathrm{randn}\,(\mathrm{n},\ 1))\,.$$

- The elements of **x** and **y** all have the same sign. Specifically, x_j and y_j are absolute values of iid standard normal random variables, and **x** and **y** are generated with the MATLAB commands

$$\mathrm{x} = \ \mathrm{single}\,(\mathrm{abs}\,(\mathrm{randn}\,(\mathrm{n},\ 1))),\ \mathrm{y} = \ \mathrm{single}\,(\mathrm{abs}\,(\mathrm{randn}\,(\mathrm{n},\ 1)))\,.$$

The “exact” inner products $\mathbf{x}^T\mathbf{y}$ are represented by the double precision computation dot(double(x), double(y)) with unit roundoff $2^{-3} \approx 1.11 \cdot 10^{-16}$. The “computed” inner products fl($\mathbf{x}^T\mathbf{y}$) are represented by the single precision computation with unit roundoff $u = 2^{-24} \approx 5.96 \cdot 10^{-8}$, in a loop that explicitly stores the products $x_k y_k$ before adding them to the partial sum, so as to bypass the fused multiply-add. All bounds are computed in double precision. Computations were performed in MATLAB R2017a on a 3.1GHz Intel Core i7 processor.

5.2. Experiments for independent roundoffs.

We illustrate the roundoff error bounds in section 3 by following up on Remark 3.1 and comparing the probabilistic bound in Theorem 3.3 with the deterministic bound (3.4) in Theorem 3.2.

- Deterministic bound

$$\left|\frac{\mathrm{fl}\left(\mathrm{x}^T\mathrm{y}\right) - \mathrm{x}^T\mathrm{y}}{|\mathrm{x}^T\mathrm{y}|}\right| \le \frac{\sqrt{\sum_{k=1}^{n} c_k^2}}{|\mathrm{x}^T\mathrm{y}|}\sqrt{n}\,. \tag{5.1}$$

- Probabilistic bound holding with probability at least $1 - \delta$,

[2] Independent identically distributed.

$$\left|\frac{\mathrm{fl}\left(x^T y\right) - x^T y}{x^T y}\right| \le \frac{\sqrt{\sum_{k=1}^{n} c_k^2}}{|x^T y|}\sqrt{2\ln(2/\delta)}, \tag{5.2}$$

where $c_1 \equiv |x_1 y_1|\,\gamma_n$, and $c_k \equiv |x_k y_k|\,\gamma_{n-k+2}$, and $\gamma_k = (1+u)^k - 1$, $2 \le k \le n$.

Figures 5.1 and 5.2 illustrate that the probabilistic result (5.2) tends to be two orders of magnitude tighter than the deterministic bound (5.1) for vectors whose elements can have different signs.

5.3. Experiments for roundoffs with conditionally independent means.

We illustrate the roundoff error bounds in section 4 by following up on Remark 4.2 and comparing the probabilistic bound in Corollary 4.7 with the deterministic bound (3.3) in Theorem 3.2.

- Traditional bound

$$\left|\frac{\mathrm{fl}\left(x^T y\right) - x^T y}{x^T y}\right| \le \frac{|x|^T |y|}{|x^T y|}\gamma_n\,. \tag{5.3}$$

- Probabilistic bound

$$\left|\frac{\mathrm{fl}\left(x^T y\right) - x^T y}{x^T y}\right| \le \frac{|x|^T |y|}{|x^T y|}\sqrt{\ln(2/\delta)}\sqrt{u\,\gamma_{2n}}, \tag{5.4}$$

where $\gamma_k = (1+u)^k - 1$ as in (3.1).

Figures 5.3 and 5.4 illustrate that the probabilistic result (5.4) tends to be at least two orders of magnitude tighter than the deterministic bound (5.3) for vectors whose elements can have different signs. However, (5.4) stops being a bound for vectors of large dimension all of whose elements have the same sign.

6. Conclusions.

We presented derivations and numerical experiments for probabilistic roundoff error bounds for the sequentially accumulated inner product of two real n-vectors. The probabilistic bounds are tighter than the deterministic bounds, often by several orders of magnitude.

Issues.

For vectors of dimension $n \ge 10^7$ and a tiny failure probability of $\delta = 10^{-16}$, the probabilistic results are not entirely satisfactory: On the one hand, they are still too pessimistic for vectors whose elements have different signs, while on the other hand they stop being upper bounds for vectors all of whose elements have the same sign. We believe that this is not a numerical artifact: It occurs in different implementations, in different languages (MATLAB and Julia), and on different processors.

A simple fix would be to adjust the failure probability, making it even more stringent when elements can differ in sign, while relaxing it when all elements have the same sign. However, this does not get to the heart of the problem. Should the failure probability be explicitly and systematically tied to the dimension *n*? This would be inconsistent with concentration inequalities, which do not explicitly depend on the number of summands. Alternatively, should one not model roundoffs as zero-mean random variables, but instead introduce a bias, possibly dimension-dependent, for vectors with structure, such as those where all elements have the same sign; see also [14, section 4.2].

Acknowledgments.

We thank Jack Dongarra, Nick Higham, and Cleve Moler for helpful discussions.

Funding:

The work of the first author was partially supported by National Science Foundation grants DMS-1745654, DMS-1760374.

REFERENCES

[1]. Babuška I and Söderlind G, On roundoff error growth in elliptic problems, ACM Trans. Math. Software, 44 (2018), 33.

[2]. Bareiss EH and Barlow JL, Roundoff error distribution in fixed point multiplication, BIT, 20 (1980), pp. 247–250.

[3]. Barlow JL and Bareiss EH, On roundoff error distributions in floating point and logarithmic arithmetic, Computing, 34 (1985), pp. 325–347.

[4]. Barlow JL and Bareiss EH, Probabilistic error analysis of Gaussian elimination in floating point and logarithmic arithmetic, Computing, 34 (1985), pp. 349–364.

[5]. Bennani M, Brunet M-C, and Chatelin F, De l’utilisation en calcul matriciel de modèles probabilistes pour la simulation des erreurs de calcul, C. R. Math. Acad. Sci. Paris, 307 (1988), pp. 847–850.

[6]. Brunet M-C and Chatelin F, CESTAC, a tool for a stochastic round-off error analysis in scientific computing, in Numerical Mathematics and Applications (Oslo, 1985), IMACS Trans. Sci. Comput 85, I, North-Holland, Amsterdam, 1986, pp. 11–20.

[7]. Calvetti D, Roundoff error for floating point representation of real data, Comm. Statist. Theory Methods, 20 (1991), pp. 2687–2695.

[8]. Calvetti D, A stochastic roundoff error analysis for the fast Fourier transform, Math. Comp, 56 (1991), pp. 755–774.

[9]. Calvetti D, A stochastic roundoff error analysis for the convolution, Math. Comp, 59 (1992), pp. 569–582.

[10]. Chatelin F and Brunet M-C, A probabilistic round-off error propagation model. Application to the eigenvalue problem, in Reliable Numerical Computation, Oxford University Press, New York, 1990, pp. 139–160.

[11]. Chung F and Lu L, Concentration inequalities and martingale inequalities: A survey, Internet Math, 3 (2006), pp. 79–127.

[12]. Henrici P, Problems of stability and error propagation in the numerical integration of ordinary differential equations, in Proceedings of the International Congress of Mathematicians (Stockholm 1962), Institut Mittag-Leffler, Djursholm, 1963, pp. 102–113.

[13]. Higham NJ, Accuracy and Stability of Numerical Algorithms, 2nd ed., SIAM, Philadelphia, PA, 2002.

[14]. Higham NJ and Mary T, A new approach to probabilistic rounding error analysis, SIAM J. Sci. Comput, 41 (2019), pp. A2815–A2835.

[15]. Hull TE and Swenson JR, Tests of probabilistic models for the propagation of roundoff errors, Comm. ACM, 9 (1966), pp. 108–113.

[16]. Kahan W, The improbability of probabilistic error analyses for numerical computations, 3 1996.

[17]. Mitzenmacher M and Upfal E, Probability and Computing: Randomized Algorithms and Probabilistic Analysis, Cambridge University Press, Cambridge, UK, 2005.

[18]. Tienari M, A statistical model of roundoff error for varying length floating-point arithmetic, BIT, 10 (1970), pp. 355–365.

[19]. von Neumann J and Goldstine HH, Numerical inverting of matrices of high order, Bull. Amer. Math. Soc, 53 (1947), pp. 1021–1099.

[20]. Wilkinson JH, Rounding errors in algebraic processes, Dover Publications, Inc., New York, 1994. Reprint of the 1963 original (Prentice-Hall, Englewood Cliffs, NJ).

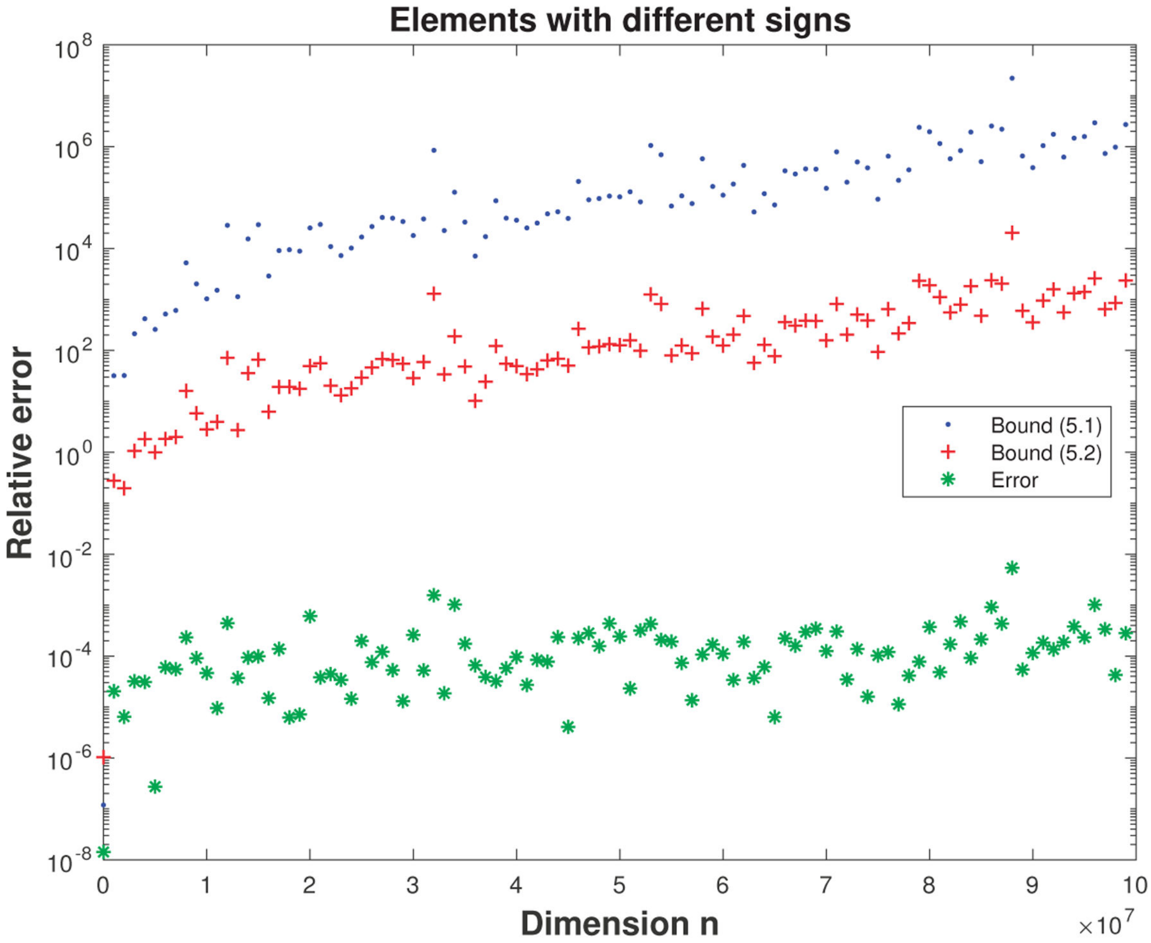

Fig. 5.1.

Comparison of probabilistic bound (red 5.2*) with deterministic bound (blue* 5.1*) and relative error (green) versus vector dimensions* $1 \le n \le 10^8$ *in steps of* 10^6*. Vertical axis starts at* 10^{-8} *and ends at* 10^8*. Elements can have different signs.*

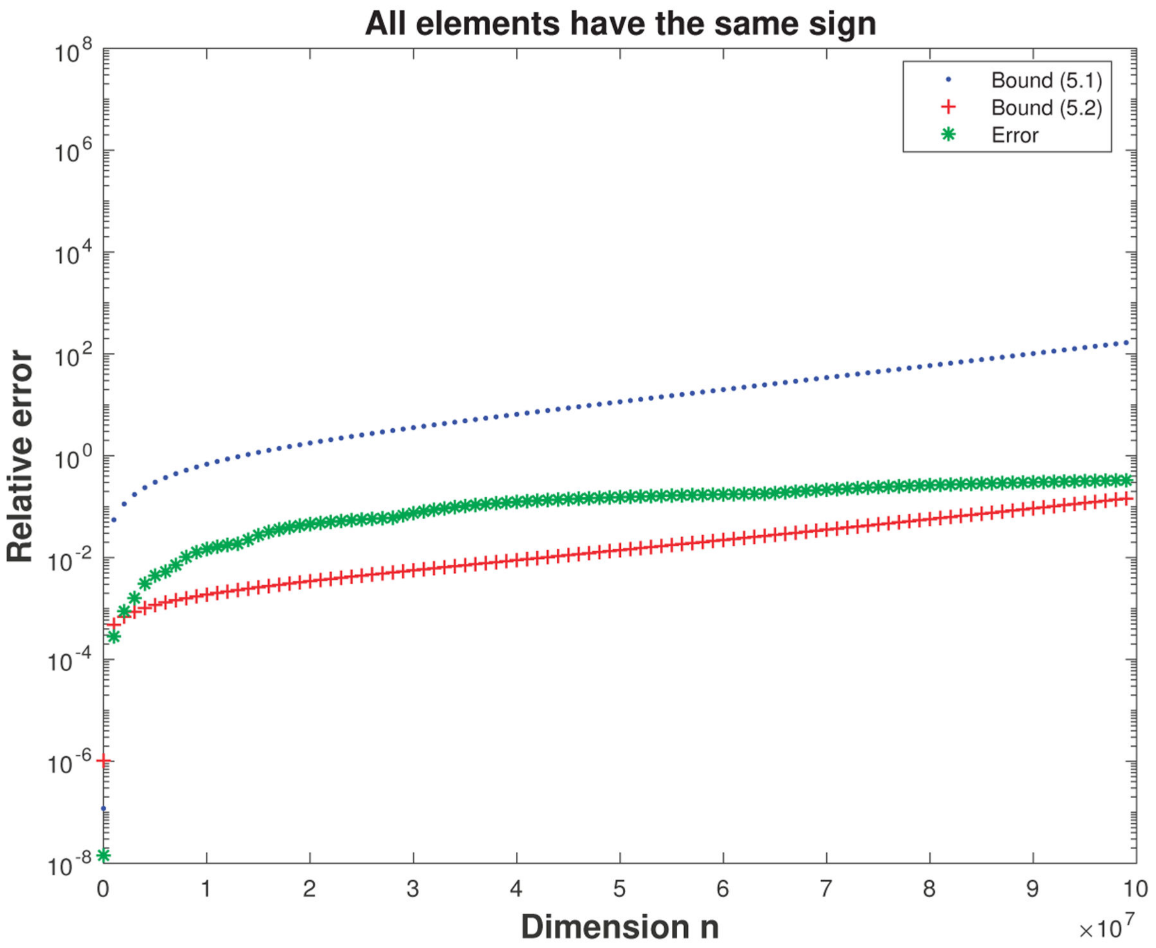

Fig. 5.2.

Comparison of probabilistic bound (red 5.2*) with deterministic bound (blue* 5.1*) and relative error (green) versus vector dimensions* $1 \le n \le 10^8$ *in steps of* 10^6*. Vertical axis starts at* 10^{-8} *and ends at* 10^8*. All elements have the same sign.*

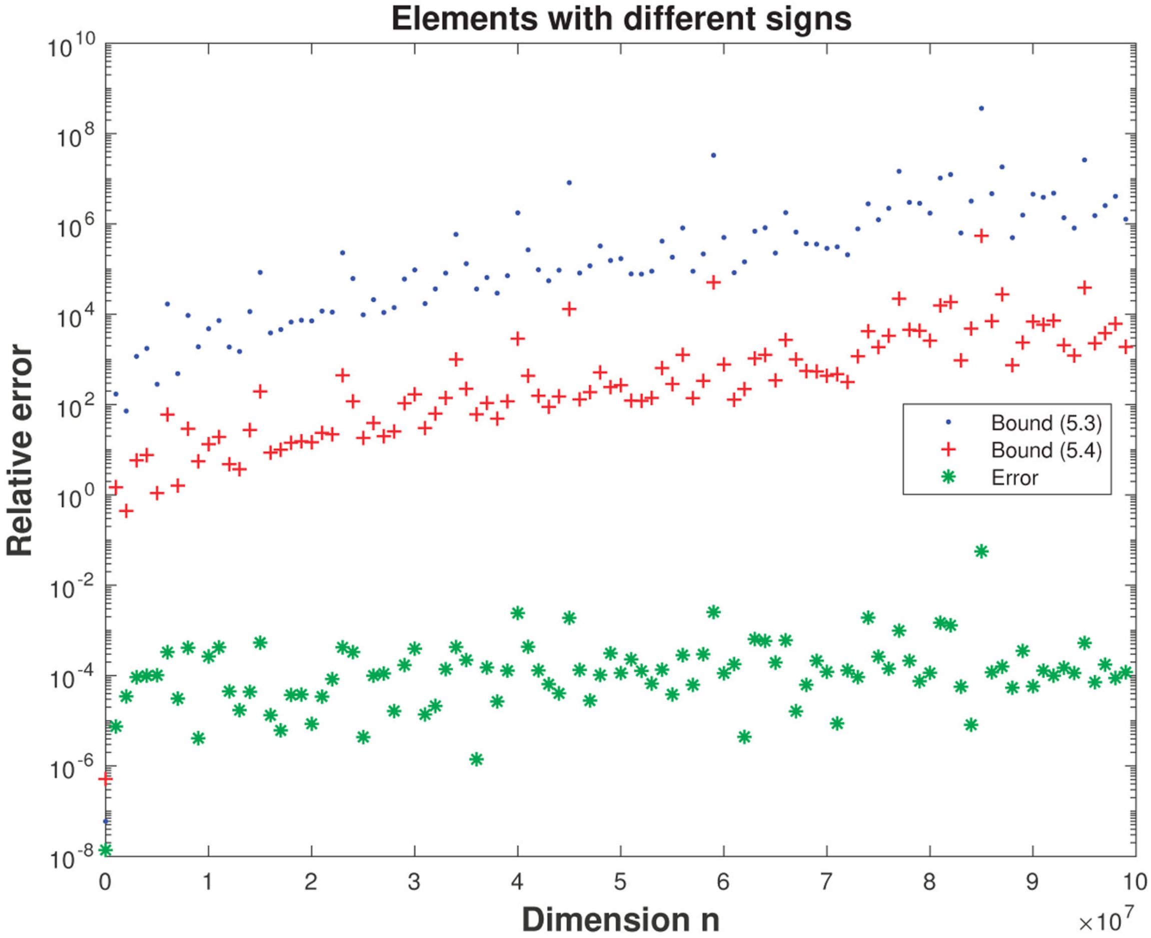

Fig. 5.3.

Comparison of probabilistic bound (red 5.4*) with deterministic bound (blue* 5.3*) and relative error (green) versus vector dimensions* $1 \leq n \leq 10^8$ *in steps of* 10^6*. Vertical axis starts at* 10^{-8} *and ends at* 10^8*. All elements have the same sign.*

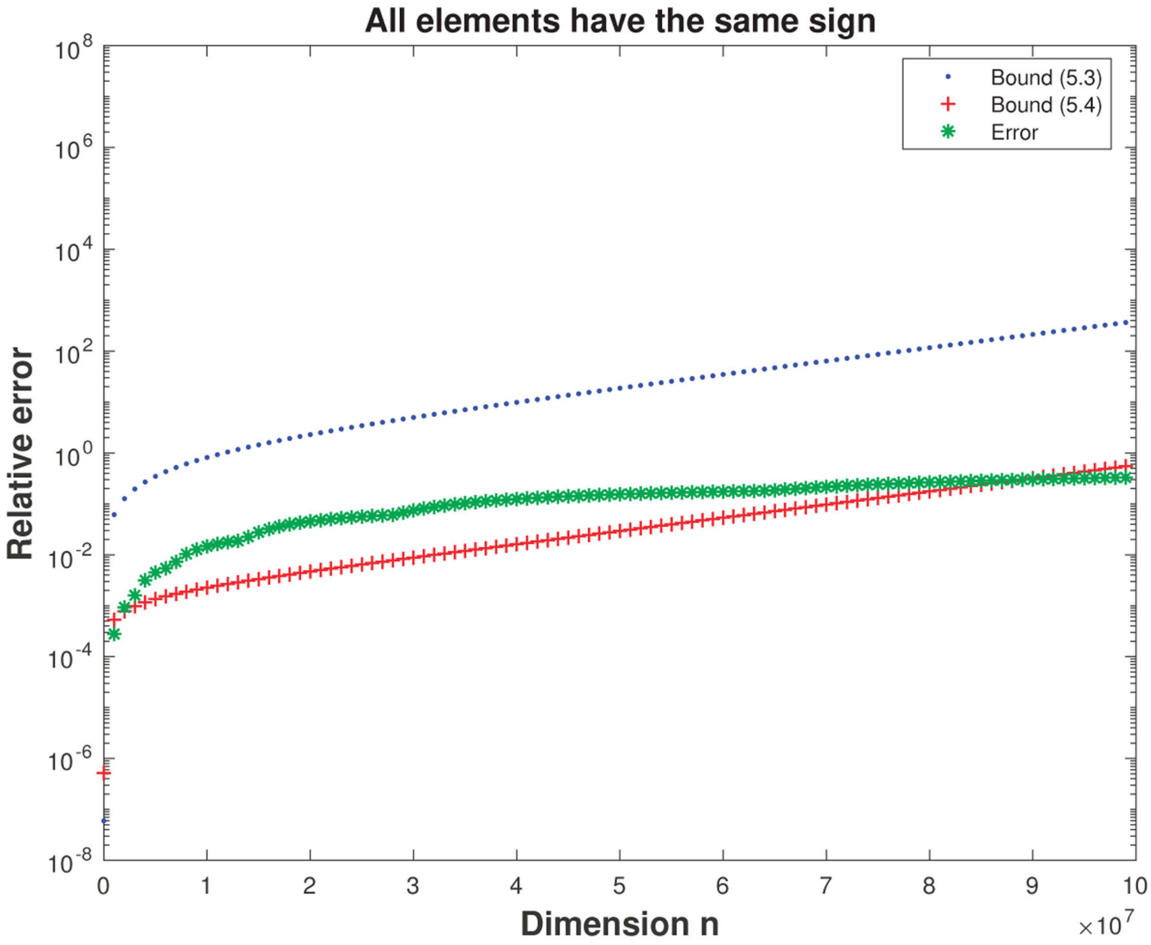

Fig. 5.4.

Comparison of probabilistic bound (red 5.4*) with deterministic bound (blue* 5.3*) and relative error (green) versus vector dimensions* $1 \leq n \leq 10^8$ *in steps of* 10^6*. Vertical axis starts at* 10^{-8} *and ends at* 10^8*. All elements have the same sign.*

Table 3.1

Traditional roundoff error model (guard digits, no fused multiply-add).

Floating point arithmetic	Exact computation
$\hat{z}_1 = x_1 y_1 (1 + \theta_1)$	$z_1 = x_1 y_1$
$\hat{z}_{k+1} = (\hat{z}_k + x_{k+1} y_{k+1} (1 + \theta_{k+1}))(1 + \delta_{k+1})$	$z_{k+1} = z_k + x_{k+1} y_{k+1}$
$\hat{z}_n = \mathrm{fl}\left(\mathrm{x}^T \mathrm{y}\right)$	$z_n = \mathbf{x}^T \mathbf{y}$

Table 4.1

Our roundoff error model (guard digits, no fused multiply-add).

Operation	Floating point arithmetic	Exact computation
*	$\hat{s}_1 = x_1 y_1$	$s_1 = x_1 y_1$
	$\hat{s}_2 = \hat{s}_1(1 + \delta_1)$	$s_2 = s_1$
*	$\hat{s}_{2k-1} = \hat{s}_{2k-2} + x_k y_k(1 + \delta_{2k-2})$	$s_{2k-1} = s_{2k-2} + x_k y_k$
+	$\hat{s}_{2k} = \hat{s}_{2k-1}(1 + \delta_{2k-1})$	$s_{2k} = s_{2k-1}$
Output	$\hat{s}_{2n} = \mathrm{fl}\left(\mathrm{x}^T \mathrm{y}\right)$	$s_{2n} = \mathrm{x}^T \mathrm{y}$