PROBABILISTIC ERROR ANALYSIS FOR INNER PRODUCTS*

ILSE C.F. IPSEN† AND HUA ZHOU‡

Abstract. Probabilistic models are proposed for bounding the forward error in the numerically computed inner product (dot product, scalar product) between of 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 non-asymptotic, explicit, and make minimal assumptions on perturbations and roundoffs.

The perturbations are represented as independent, bounded, zero-mean random variables, and the probabilistic perturbation bound is based on Azuma's inequality. The roundoffs are also represented as bounded, zero-mean random variables. The first probabilistic bound assumes that the roundoffs are independent, while the second one does not. For the latter, we construct a Martingale that mirrors the sequential order of computations.

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.

Key words. Perturbation bounds, roundoff errors, random variables, sums of random variables, Martingales

AMS subject classification. 65F30, 65G50, 60G42, 60G50

1. Introduction.

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

$$\mathbf{x}^T\mathbf{y} = x_1y_1 + \cdots + x_ny_n,$$

between two real n-vectors

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n \qquad \text{and} \qquad \mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n.$$

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

*Funding: The work of the first author was supported in part by National Science Foundation grants DMS-1745654 and DMS-1760374.

†Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205, USA, ipsen@ncsu.edu

‡Department of Biostatistics, University of California, Los Angeles, CA 90095-1772, USA, huazhou@ucla.edu

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

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 new condition numbers (Sections 2 and 3.3), and they are tighter because they avoid a union bound for the probabilities. Our bounds are also simple, intuitive, and easy to interpret, with a clear relationship between failure probability and relative error. Compared to [14, Theorem 3.1], our Corollary 4.8 is tighter and does not assume independence of roundoffs.

Overview. To facilitate the introduction of the probabilistic approach, we start as simple 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 inequaility. 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 inequaility. However, numerical experiments (Section 5) illustrate that for non-negative 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 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 errors act like independent variables would seem to be a fascinating if difficult problem.

As a consequence, and in contrast to [14], we relinquish the independence assumption and derive a general probabilistic roundoff error bound (Section 4). The roundoffs are represented as bounded, zero-mean random variables; and the forward error is bounded by an Azuma-Hoeffding Martingale. In particular, we present a quantitative confirmation of Wilkinson's intuition [20, Section 1.33] that the roundoff error in n operations is proportional to $\sqrt{n}\,u$ rather than $n\,u$. The paper ends with a critical analysis of the probabilistic approach, and a long list of future work (Section 6).

2. Perturbation bounds. To calibrate the roundoff error bounds and set the stage for the probabilistic approach, we start off with perturbation bounds: first, deterministic bounds that generalize the traditional bound and motivate the probabilistic bound (Section 2.1), and then the probabilistic bound (Section 2.2).

We use the Hadamard product

$$\mathbf{x} \circ \mathbf{y} \equiv \begin{pmatrix} x_1 y_1 & \cdots & x_n y_n \end{pmatrix}^T$$

to compactly express componentwise relative perturbations as

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

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

$$\boldsymbol{\delta} \equiv \begin{pmatrix}\delta_1 & \cdots & \delta_n\end{pmatrix}^T, \qquad \boldsymbol{\theta} \equiv \begin{pmatrix}\theta_1 & \cdots & \theta_n\end{pmatrix}^T.$$

2.1. Deterministic perturbation bound. We generalize the traditional perturbation bound to a whole class of bounds, and single out a specific bound to motivate the probabilistic bound in Section 2.2.

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

$$\left|\frac{\hat{\mathbf{x}}^T\hat{\mathbf{y}} - \mathbf{x}^T\mathbf{y}}{\mathbf{x}^T\mathbf{y}}\right| \leq \frac{\|\mathbf{x}\circ\mathbf{y}\|_p}{|\mathbf{x}^T\mathbf{y}|}\,\|\boldsymbol{\delta}+\boldsymbol{\theta}+\boldsymbol{\delta}\circ\boldsymbol{\theta}\|_q.$$

Proof. From associativity, distributivity and the fact that all quantities are real follows

$$\begin{aligned}\hat{\mathbf{x}}^T\hat{\mathbf{y}} - \mathbf{x}^T\mathbf{y} &= (\boldsymbol{\delta}\circ\mathbf{x})^T\mathbf{y} + \mathbf{x}^T(\boldsymbol{\theta}\circ\mathbf{y}) + (\boldsymbol{\delta}\circ\mathbf{x})^T(\boldsymbol{\theta}\circ\mathbf{y})\\ &= \sum_{k=1}^n x_k y_k\ (\delta_k + \theta_k + \delta_k\theta_k) = (\mathbf{x}\circ\mathbf{y})^T\,(\boldsymbol{\delta}+\boldsymbol{\theta}+\boldsymbol{\delta}\circ\boldsymbol{\theta}).\end{aligned}$$

The Hölder inequality implies

$$\left|(\mathbf{x}\circ\mathbf{y})^T\,(\boldsymbol{\delta}+\boldsymbol{\theta}+\boldsymbol{\delta}\circ\boldsymbol{\theta})\right| \leq \|\mathbf{x}\circ\mathbf{y}\|_p\,\|\boldsymbol{\delta}+\boldsymbol{\theta}+\boldsymbol{\delta}\circ\boldsymbol{\theta}\|_q.$$

□

Below is a specialization of Theorem 2.1 to popular p-norms.

COROLLARY 2.2. *Theorem 2.1 implies the following bounds.*

1. *Traditional bound* ($p = 1$)

$$\left|\frac{\hat{\mathbf{x}}^T\hat{\mathbf{y}} - \mathbf{x}^T\mathbf{y}}{\mathbf{x}^T\mathbf{y}}\right| \leq \frac{\|\mathbf{x}\circ\mathbf{y}\|_1}{|\mathbf{x}^T\mathbf{y}|}\,\|\boldsymbol{\delta}+\boldsymbol{\theta}+\boldsymbol{\delta}\circ\boldsymbol{\theta}\|_\infty \leq \frac{|\mathbf{x}|^T|\mathbf{y}|}{|\mathbf{x}^T\mathbf{y}|}\,u(2+u).$$

2. *Same amplifier as in Theorem 2.4* ($p = 2$)

$$\begin{aligned}\left|\frac{\hat{\mathbf{x}}^T\hat{\mathbf{y}} - \mathbf{x}^T\mathbf{y}}{\mathbf{x}^T\mathbf{y}}\right| &\leq \frac{\|\mathbf{x}\circ\mathbf{y}\|_2}{|\mathbf{x}^T\mathbf{y}|}\,\|\delta+\boldsymbol{\theta}+\boldsymbol{\delta}\circ\boldsymbol{\theta}\|_2 \qquad (2.1)\\ &\leq \sqrt{n}\,\frac{\|\mathbf{x}\circ\mathbf{y}\|_2}{|\mathbf{x}^T\mathbf{y}|}\,u(2+u).\end{aligned}$$

3. *Smallest amplifier* ($p = \infty$)

$$\left|\frac{\hat{\mathbf{x}}^T\hat{\mathbf{y}} - \mathbf{x}^T\mathbf{y}}{\mathbf{x}^T\mathbf{y}}\right| \leq \frac{\|\mathbf{x}\circ\mathbf{y}\|_\infty}{|\mathbf{x}^T\mathbf{y}|}\,\|\boldsymbol{\delta}+\boldsymbol{\theta}+\boldsymbol{\delta}\circ\boldsymbol{\theta}\|_1 \leq n\,\frac{\|\mathbf{x}\circ\mathbf{y}\|_\infty}{|\mathbf{x}^T\mathbf{y}|}\,u(2+u).$$

Proof. The traditional bound follows from

$$\|\mathbf{x}\circ\mathbf{y}\|_1 = \sum_{k=1}^n |x_k y_k| = \sum_{k=1}^n |x_k|\,|y_k| = |\mathbf{x}|^T|\mathbf{y}|.$$

□

The numerical experiments in Section 5.2.1 suggest that the three bounds tend to differ by at most an order of magnitude or so, with the traditional bound being the tightest.

2.2. Probabilistic perturbation bound. We derive a probabilistic bound corresponding to the deterministic bound (2.1), and then compare the two bounds.

The basis for the probabilistic bounds is a concentration inequality, which bounds the deviation of a sum from its mean in terms of the deviations of the individual summands from their means.

LEMMA 2.3 (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*

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

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

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

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 $1 - \delta$. ∎

Thus, if each summand Z_k is close to its mean $\mathbb{E}[Z_k]$, then with high probability, the sum Z is also close to its mean $\mathbb{E}[Z]$.

In the probabilistic perturbation bound below, the perturbations δ_k and θ_k are represented as independent, bounded, zero-mean random variables.

THEOREM 2.4. *Let the perturbations* 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* $1 - \delta$, *the relative forward error in the perturbed inner product is bounded by*

$$\begin{aligned}\left|\frac{\hat{\mathbf{x}}^T\hat{\mathbf{y}} - \mathbf{x}^T\mathbf{y}}{\mathbf{x}^T\mathbf{y}}\right| &\le \frac{\|\mathbf{x}\circ\mathbf{y}\|_2}{|\mathbf{x}^T\mathbf{y}|}\,\sqrt{2\,\ln(2/\delta)}\,u(2+u)\\ &= \frac{\sqrt{\sum_{k=1}^{n}|x_k y_k|^2}}{|\mathbf{x}^T\mathbf{y}|}\,\sqrt{2\,\ln(2/\delta)}\,u(2+u).\end{aligned}$$

Proof. Write the total forward error

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

as a sum of independent random variables, where each summand represents a "local" forward error,

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

From the linearity of the mean and δ_k, θ_k being independent random variables with $\mathbb{E}[\delta_k] = \mathbb{E}[\theta_k] = 0$ follows

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

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

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

where $\tau \equiv 2u + u^2 = u(2+u)$. Therefore, the conditions of Lemma 2.3 are satisfied, and we have

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

The linearity of the expected value implies

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

Apply Lemma 2.3 to conclude that for any $0 < \delta < 1$, with probability at least $1 - \delta$,

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

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

Remark 2.1 (Comparsion). *The probabilistic bound in Theorem 2.4 is by a factor of $\sqrt{n}$ tighter than the deterministic bound (2.1) in Corollary 2.2.*

The probabilistic bound in Theorem 2.4 holds with probability at least $1 - \delta$,

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

while the deterministic bound (2.1) equals

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

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

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 $\sqrt{2\, \ln{(2/\delta)}} \le 9$.*

The numerical experiments in Section 5.2.2 illustrate that the probabilistic bound tends to be at least two orders of magnitude tighter than the deterministic bound.

Example 2.1. *We illustrate the behaviour of the amplifier*

$$\kappa_2 \equiv \|\mathbf{x} \circ \mathbf{y}\|_2 / |\mathbf{x}^T \mathbf{y}|$$

in the probabilistic bound in Theorem 2.4 with three very special cases.

1. *No cancellation:*

   If all $x_k y_k$ have the same sign, then $\kappa_2^2 = \frac{\sum_{k=1}^{n} |x_k y_k|^2}{\left(\sum_{k=1}^{n} |x_k y_k|\right)^2} \le 1$, so that

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

   If also $x_k y_k = w \ne 0$ for $1 \le k \le n$, then $\kappa_2^2 = \frac{nw^2}{(nw)^2} = \frac{1}{n}$, so that κ_2 decreases with increasing dimension n,

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

2. *Severe cancellation:*
If $x_k y_k = (-1)^k w$ *for* $1 \le k \le n$, $w \ne 0$, *and* n *is odd, then* $\kappa_2^2 = \frac{nw^2}{w^2} = n$, *so that* κ_2 *increases with increasing dimension* n,

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

3. Probabilistic roundoff error bound, assuming independence of roundoff. After presenting the model for independent roundoffs (Section 3.1), we derive a motivating deterministic bound (Section 3.2), followed by the probabilistic bound (Section 3.3).

3.1. Roundoff error model. We assume that the elements of $\mathbf{x}$ and $\mathbf{y}$ are floating point numbers, and 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,

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

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

Table 3.1
Traditional roundoff error model (guard digits, no fused multiply-add)

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

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

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

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

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

3.2. A motivating deterministic bound. First we unravel the expressions for the computed partial sums, and then bound the sums in terms of inputs and the roundoffs.

Lemma 3.1. *The partial sums in Table 3.1 are equal to*

$$\hat{z}_1 = x_1 y_1 (1+\theta_1)$$

$$\hat{z}_k = x_1 y_1 (1+\theta_1) \prod_{\ell=2}^{k} (1+\delta_\ell) + \sum_{j=2}^{k} x_j y_j (1+\theta_j) \prod_{\ell=j}^{k} (1+\delta_\ell), \qquad 2 \le k \le n.$$

If $|\delta_k|, |\theta_k| \le u$, $1 \le k \le n$, *then the partial sums are bounded by*

$$|\hat{z}_1| \le |x_1 y_1| (1+u)$$

$$|\hat{z}_k| \le |x_1 y_1| (1+u)^k + \sum_{j=2}^{k} |x_j y_j| (1+u)^{k-j+2}, \qquad 2 \le k \le n.$$

LEMMA 3.2. *The total forward error for the computed inner product* $\hat{z}_n = \mathrm{fl}(\mathbf{x}^T\mathbf{y})$ *in Table 3.1 is expressed as a sum of "local forward errors",*

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

with a local forward error for each summand,

$$Z_1 \equiv x_1 y_1 \left((1+\theta_1) \prod_{\ell=2}^{n} (1+\delta_\ell) - 1 \right)$$
$$Z_k \equiv x_k y_k \left((1+\theta_k) \prod_{\ell=k}^{n} (1+\delta_\ell) - 1 \right), \qquad 2 \leq k \leq n$$

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

$$|Z_1| \leq c_1 \equiv |x_1 y_1|\, \gamma_n$$
$$|Z_k| \leq c_k \equiv |x_k y_k|\, \gamma_{n-k+2}, \qquad 2 \leq k \leq n.$$

Proof. This is analogous to [13, Lemma 3.1]. □

Now we can bound the total forward error.

THEOREM 3.3. *Let the roundoffs satisfy* $|\delta_k|, |\theta_k| \leq u$, $1 \leq k \leq n$, *with* γ_k *as in (3.1).*

Then the forward error of the computed inner product $\hat{z}_n = \mathrm{fl}(\mathbf{x}^T\mathbf{y})$ *in Table 3.1 is bounded by*

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

Proof. Applying the triangle inequality to the total forward error in Lemma 3.2 gives

$$|\hat{z}_n - z_n| \leq \sum_{k=1}^{n} |Z_k| \leq \sum_{k=1}^{n} c_k.$$

□

The first consequence is the traditional forward error bound [13, Section 3.1].

COROLLARY 3.4 (Traditional bound). *Let the roundoffs satisfy* $|\delta_k|, |\theta_k| \leq u$, $1 \leq k \leq n$, *with* γ_k *as in (3.1).*

Then the relative forward error of the computed inner product $\hat{z}_n = \mathrm{fl}(\mathbf{x}^T\mathbf{y})$ *in Table 3.1 is bounded by*

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

Proof. Define the vectors

$$\mathbf{v} \equiv \begin{pmatrix} |x_1 y_1| & \cdots & |x_n y_n| \end{pmatrix}^T, \qquad \mathbf{g} \equiv \begin{pmatrix} \gamma_n & \gamma_n & \gamma_{n-1} & \cdots & \gamma_2 \end{pmatrix}^T,$$

and apply the Hölder inequality to

$$\sum_{k=1}^{n} c_k = \mathbf{v}^T \mathbf{g} \leq \|\mathbf{v}\|_1 \, \|\mathbf{g}\|_\infty = \sum_{k=1}^{n} |x_k y_k| \, \gamma_n = |\mathbf{x}|^T |\mathbf{y}| \, \gamma_n.$$

□

The second consequence is the motivation for the probabilistic bound to follow.

COROLLARY 3.5 (Deterministic version of Theorem 3.6). *Let the roundoffs satisfy* $|\delta_k|, |\theta_k| \leq u$, $1 \leq k \leq n$, *with* γ_k *as in (3.1).*

Then the relative forward error of the computed inner product $\hat{z}_n = \mathrm{fl}(\mathbf{x}^T\mathbf{y})$ *in Table 3.1 is bounded by*

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

where $c_1 \equiv |x_1 y_1| \, \gamma_n$, *and* $c_k \equiv |x_k y_k| \, \gamma_{n-k+2}$, $2 \leq k \leq n$.

Proof. Define the non-negative vector $\mathbf{c} \equiv \begin{pmatrix} c_1 & \cdots & c_n \end{pmatrix}^T$ and use the relation between vector norms

$$\sum_{k=1}^{n} c_k = \|\mathbf{c}\|_1 \leq \|\mathbf{c}\|_2 \sqrt{n} = \sqrt{\sum_{k=1}^{n} c_k^2} \, \sqrt{n}.$$

□

3.3. Probabilistic forward error bound. Since the roundoffs are independent, bounded zero-mean random variables, we can use Azuma's inequality in Lemma 2.3.

THEOREM 3.6. *Let the roundoffs* δ_k, θ_k *be independent random variables with* $\mathbb{E}[\delta_k] = \mathbb{E}[\theta_k] = 0$ *and* $|\delta_k|, |\theta_k| \leq u$, $1 \leq k \leq n$, *and let* γ_k *as in (3.1).*

Then for any $0 < \delta < 1$, *with probability at least* $1 - \delta$, *the relative forward error in the computed inner product* $\hat{z}_n = \mathrm{fl}(\mathbf{x}^T\mathbf{y})$ *in Table 3.1 is bounded by*

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

where $c_1 \equiv |x_1 y_1| \, \gamma_n$, *and* $c_k \equiv |x_k y_k| \, \gamma_{n-k+2}$, $2 \leq k \leq n$.

Proof. Since the roundoffs are independent random variables, so is the total forward error in Lemma 3.2,

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

The random variables

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

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

represent the local forward errors and have zero mean, $\mathbb{E}[Z_k] = 0$. By linearity, the total forward error has zero mean as well,

$$\mathbb{E}[Z] = \mathbb{E}\left[Z_1 + \cdots + Z_n\right] = \mathbb{E}[Z_1] + \cdots + \mathbb{E}[Z_n] = 0.$$

The deviations of the local errors from their means are bounded by

$$|Z_k - \mathbb{E}[Z_k]| = |Z_k| \leq c_k, \qquad 1 \leq k \leq n,$$

with c_k as in Lemma 3.2. Thus we can apply Lemma 2.3 to Z, and then divide both sides by the constant $|\mathbf{x}^T\mathbf{y}|$. ▯

Remark 3.1 (Comparison). *The probabilistic bound in Theorem 3.6 tends to be tighter than the corresponding deterministic bound in Corollary 3.5.*

The probabilistic bound in Theorem 3.6 holds with probability at least $1-\delta$,

$$\left|\frac{\hat{\mathbf{x}}^T\hat{\mathbf{y}} - \mathbf{x}^T\mathbf{y}}{\mathbf{x}^T\mathbf{y}}\right| \leq \frac{\sqrt{\sum_{k=1}^n c_k^2}}{|\mathbf{x}^T\mathbf{y}|}\sqrt{2\,\ln(2/\delta)},$$

while the deterministic bound in Corollary 3.5 equals

$$\left|\frac{\hat{\mathbf{x}}^T\hat{\mathbf{y}} - \mathbf{x}^T\mathbf{y}}{\mathbf{x}^T\mathbf{y}}\right| \leq \frac{\sqrt{\sum_{k=1}^n c_k^2}}{|\mathbf{x}^T\mathbf{y}|}\sqrt{n},$$

where $c_1 \equiv |x_1y_1|\,\gamma_n$, and $c_k \equiv |x_ky_k|\,\gamma_{n-k+2}$, $2 \leq k \leq n$, with γ_k as in (3.1).

As in Remark 2.1, the two bounds differ in the factors $\sqrt{2\,\ln(2/\delta)}$ versus $\sqrt{n}$, which implies:

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 $\sqrt{2\,\ln(2/\delta)} \leq 9$.*

The numerical experiments in Section 5.3 illustrate that the probabilistic expression can be as much as two orders of magnitude tighter then the deterministic bound, but stops being an upper bound for non-negative vectors of large dimension.

4. General probabilistic roundoff error bound.

In contrast to the previous section, we make no assumptions on the independence of roundoffs. After presenting the roundoff error model (Section 4.1), we derive a motivating deterministic bound (Section 4.2), and then present the probabilistic bound (Section 4.3), followed by two upper bounds that take a simpler form (Section 4.4).

4.1. Roundoff error model.

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

The model in Table 4.1 is designed to do without additional intermediate factors like $x_ky_k(1+\delta_{2k-2})$, 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.

First we bound the computed partial sums in terms of the inputs, and the unit roundoff u.

Lemma 4.1. *Let the roundoffs satisfy $|\delta_k| \leq u$, $1 \leq k \leq 2n-1$.*

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	$\hat{s}_{2n} = \mathrm{fl}(\mathbf{x}^T \mathbf{y})$	$s_{2n} = \mathbf{x}^T \mathbf{y}$

Then the partial sums computed in Table 4.1 are bounded by

$$
\begin{aligned}
|\hat{s}_{2k-1}| &\le |x_1 y_1| (1+u)^{k-1} + |x_2 y_2| (1+u)^{k-1} + \cdots + |x_k y_k| (1+u) \\
&= |x_1 y_1| (1+u)^{k-1} + \sum_{j=2}^{k} |x_j y_j| (1+u)^{k-j+1}, \qquad 1 \le k \le n,
\end{aligned}
$$

and

$$
\begin{aligned}
|\hat{s}_{2k}| &\le |x_1 y_1| (1+u)^{k} + |x_2 y_2| (1+u)^{k} + \cdots + |x_k y_k| (1+u)^2 \\
&= |x_1 y_1| (1+u)^{k} + \sum_{j=2}^{k} |x_j y_j| (1+u)^{k-j+2}, \qquad 1 \le k \le n.
\end{aligned}
$$

Proof. The proof is by induction, starting with the basis for $k = 1$,

$$
\begin{aligned}
|\hat{s}_1| &= |x_1 y_1| = |x_1 y_1| (1+u)^0, \\
|\hat{s}_2| &= |\hat{s}_1 (1 + \delta_1)| \le |x_1 y_1| (1+u).
\end{aligned}
$$

Assuming, as the hypothesis, that the statement of the lemma is correct, the induction step gives for $1 \le k \le n-1$,

$$
\begin{aligned}
|\hat{s}_{2k+1}| &= |\hat{s}_{2k} + x_{k+1} y_{k+1} (1 + \delta_{2k})| \le |\hat{s}_{2k}| + |x_{k+1} y_{k+1}| (1+u) \\
&\le |x_1 y_1| (1+u)^k + \sum_{j=2}^{k} |x_j y_j| (1+u)^{k-j+2} + |x_{k+1} y_{k+1}| (1+u) \\
&= |x_1 y_1| (1+u)^k + \sum_{j=2}^{k} |x_j y_j| (1+u)^{(k+1)-j+1} + |x_{k+1} y_{k+1}| (1+u) \\
&= |x_1 y_1| (1+u)^k + \sum_{j=2}^{k+1} |x_j y_j| (1+u)^{(k+1)-j+1},
\end{aligned}
$$

and for $1 \le k \le n-1$,

$$
\begin{aligned}
|\hat{s}_{2k+2}| &= |\hat{s}_{2k+1} (1 + \delta_{2k+1})| \le |\hat{s}_{2k+1}| (1+u) \\
&= |x_1 y_1| (1+u)^{k+1} + \sum_{j=2}^{k+1} |x_j y_j| (1+u)^{(k+1)-j+2}.
\end{aligned}
$$

$\square$

The *total* forward error is

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

while the *partial sum forward errors* are

$$Z_k \equiv \hat{s}_k - s_k, \qquad 1 \leq k \leq 2n,$$

where $Z_1 = 0$. We use these partial sum errors to distinguish the newly arrived roundoff from the previous roundoffs. Then we establish a recursion for the partial sum errors Z_k, and bound the difference between two successive partial sum errors Z_k and Z_{k-1} by the "incremental error" $c_k\, u$. This incremental error $c_k\, u$ captures the most recent roundoff introduced when moving from Z_{k-1} to Z_k.

LEMMA 4.2. *The forward errors for the partial sums in Table 4.1 satisfy the recursions*

$$\begin{aligned} Z_{2k} &= Z_{2k-1} + \hat{s}_{2k-1}\, \delta_{2k-1}, \qquad 1 \leq k \leq n, \\ Z_{2k-1} &= Z_{2k-2} + x_k y_k\, \delta_{2k-2}, \qquad 2 \leq k \leq n. \end{aligned}$$

If $|\delta_k| \leq u$, $1 \leq k \leq 2n-1$, *then*

$$|Z_{2k} - Z_{2k-1}| \leq c_{2k-1}\, u, \qquad 1 \leq k \leq n,$$

where

$$|\hat{s}_{2k-1}| \leq c_{2k-1} \equiv |x_1 y_1|\, (1+u)^{k-1} + \sum_{j=2}^{k} |x_j y_j|\, (1+u)^{k-j+1},$$

and for $2 \leq k \leq n$,

$$|Z_{2k-1} - Z_{2k-2}| \leq c_{2k-2}\, u, \qquad \textit{where} \quad c_{2k-2} \equiv |x_k y_k|.$$

Proof. The proof is by induction, following the recursions in Table 4.1. Since $Z_1 = 0$, the induction starts one step later than the one in Lemma 4.2, and the induction basis is

$$\begin{aligned} Z_2 &= \hat{s}_2 - s_2 = \hat{s}_1(1+\delta_1) - s_1 = Z_1 + \hat{s}_1\, \delta_1, \\ Z_3 &= \hat{s}_3 - s_3 = \hat{s}_2 + x_2 y_2\, (1+\delta_2) - (s_2 + x_2 y_2) = Z_2 + x_2 y_2\, \delta_2. \end{aligned}$$

Assuming, as the hypothesis, that the statement of the lemma is correct, the induction step gives for $1 \leq k \leq n-1$,

$$Z_{2k+2} = \hat{s}_{2k+2} - s_{2k+2} = \hat{s}_{2k+1}\, (1+\delta_{2k+1}) - s_{2k+1} = Z_{2k+1} + \hat{s}_{2k+1}\, \delta_{2k+1},$$

and for $2 \leq k \leq n-1$,

$$\begin{aligned} Z_{2k+1} &= \hat{s}_{2k+1} - s_{2k+1} = \hat{s}_{2k} + x_{k+1} y_{k+1}\, (1+\delta_{2k}) - (s_{2k} + x_{k+1} y_{k+1}) \\ &= Z_{2k} + x_{k+1} y_{k+1}\, \delta_{2k}. \end{aligned}$$

Lemma 4.1 and the above recursions imply the bounds

$$\begin{aligned} |Z_2 - Z_1| &= |\hat{s}_1\, \delta_1| \leq |\hat{s}_1|\, u \leq c_1\, u \qquad \text{where} \quad c_1 = |x_1 y_1|, \\ |Z_3 - Z_2| &= |x_2 y_2\, \delta_2| \leq |x_2 y_2|\, u \qquad \text{where} \quad c_2 = |x_2 y_2|. \end{aligned}$$

In general,

$$|Z_{2k} - Z_{2k-1}| = |\hat{s}_{2k-1}\,\delta_{2k-1}| \le |\hat{s}_{2k-1}|\,u \le c_{2k-1}\,u, \qquad 2 \le k \le n,$$

where $c_{2k-1} = |x_1 y_1|\,(1+u)^{k-1} + \sum_{j=2}^{k} |x_j y_j|\,(1+u)^{k-j+1}$, and

$$|Z_{2k+1} - Z_{2k}| = |x_{k+1}y_{k+1}\,\delta_{2k}| \le |x_{k+1}y_{k+1}|\,u \le c_{2k}\,u, \qquad 2 \le k \le n-1,$$

where $c_{2k} = |x_{k+1}y_{k+1}|$. ∎

THEOREM 4.3 (Deterministic version of Theorem 4.6). *Let the roundoffs satisfy $|\delta_k| \le u$, $1 \le k \le 2n-1$.*

Then the relative forward error of the computed inner product $\hat{s}_{2n} = \mathrm{fl}(\mathbf{x}^T\mathbf{y})$ in Table 4.1 is bounded by

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

where

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

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

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

where $Z_1 = 0$. With the expressions for c_k from Lemma 4.2,

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

As in the proof of Corollary 3.5, the relation between the vector one- and two-norms implies

$$\sum_{k=1}^{2n-1} c_k \le \sqrt{2n-1}\,\sqrt{\sum_{k=1}^{2n-1} c_k^2}.$$

∎

4.3. Probabilistic forward error bound. We derive a probabilistic bound based on an Azuma Martingale, which does not require independence of roundoffs, and then compare the probabilistic and deterministic bounds.

DEFINITION 4.4 (Martingale, Definition 12.1 in [17]). *A sequence of random variables $Z_1, Z_2 \ldots$ 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 $\delta_1, \ldots, \delta_{k-1}$,*
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 4.5 (Azuma-Hoeffding Martingale, Theorem 12.4 in [17]). *Let $Z_1, \ldots, Z_{2n}$ be a Martingale with*

$$|Z_k - Z_{k-1}| \leq c_{k-1}, \qquad 2 \leq k \leq 2n.$$

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

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

Proof. In [17, Theorem 12.4], set

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

and $m = 2n - 1$, and then solve for t in terms of δ. If $|Z - \mathbb{E}[Z]| \geq t$ holds with probability at most δ, then the complementary event $|Z - \mathbb{E}[Z]| \leq t$ holds with probability at least $1 - \delta$. ☐

Again, the roundoffs are represented as bounded, zero-mean random variables, but now they are not required to be independent. The following bound resembles the one in Theorem 2.1, but contains more summands.

THEOREM 4.6. *Let the roundoffs δ_k be random variables with $\mathbb{E}[\delta_k] = 0$ and $|\delta_k| \leq u$, $1 \leq k \leq 2n - 1$.*

Then for any $0 < \delta < 1$, with probability at least $1 - \delta$, the relative forward error of the computed inner product $\hat{s}_{2n} = \mathrm{fl}(\mathbf{x}^T\mathbf{y})$ in Table 4.1 is bounded by

$$\left|\frac{\mathrm{fl}(\mathbf{x}^T\mathbf{y}) - \mathbf{x}^T\mathbf{y}}{\mathbf{x}^T\mathbf{y}}\right| = \left|\frac{\hat{s}_{2n} - s_{2n}}{s_{2n}}\right| \leq \frac{\sqrt{\sum_{k=1}^{2n-1} c_k^2}}{|\mathbf{x}^T\mathbf{y}|}\, \sqrt{2\, \ln (2/\delta)}\, u,$$

where

$$c_{2k-1} = |x_1 y_1|\,(1+u)^{k-1} + \sum_{j=2}^{k} |x_j y_j|\,(1+u)^{k-j+1}, \qquad 1 \leq k \leq n,$$
$$c_{2k-2} = |x_k y_k|, \qquad 2 \leq k \leq n.$$

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

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

To apply Lemma 4.5, 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 4.4 and Lemma 4.5.

1. The recursions in Lemma 4.2 show that Z_k is a function of the roundoffs $\delta_1, \ldots, \delta_{k-1}$, $2 \leq k \leq 2n$.
2. The expectation of $|Z_k|$ is finite because $|Z_k|$ is a finite sum of bounded summands, and the roundoffs have zero mean.

3. Lemma 4.2 implies $Z_2 = Z_1 + x_1 y_1\, \delta_1$ where $Z_1 = 0$. The linearity of expectation and zero-mean property of the roundoffs implies

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

More generally, item 1 implies that Z_{2k-2} depends on $\delta_1, \ldots, \delta_{2k-3}$, $2 \leq k \leq n$. Conditioning on all of these roundoffs removes the randomness and produces a fixed value,

$$\mathbb{E}[Z_{2k-2}\,|\delta_1, \ldots, \delta_{2k-3}] = Z_{2k-2}, \qquad 2 \leq k \leq n,$$

Combine the above with the zero-mean property of the roundoffs

$$\mathbb{E}[\delta_{2k-2}\,|\delta_1, \ldots, \delta_{2k-3}] = \mathbb{E}[\delta_{2k-2}] = 0$$

and Lemma 4.2 to conclude

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

Now consider the remaining recursions $Z_{2k} = Z_{2k-1} + \hat{s}_{2k-1}\, \delta_{2k-1}$, $1 \leq k \leq n$. Item 1 and Table 4.1 show that Z_{2k-1} and $\hat{s}_{2k-1}$ depend only on the roundoffs $\delta_1, \ldots, \delta_{2k-2}$. Conditioning Z_{2k-1} and $\hat{s}_{2k-1}$ on all of these roundoffs removes the randomness and produces fixed values,

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

Arguing as above shows

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

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

4. Lemma 4.2 implies

$$|Z_{2k} - Z_{2k-1}| \leq c_{2k-1}\, u, \qquad 1 \leq k \leq n,$$

where

$$|\hat{s}_{2k-1}| \leq c_{2k-1} \equiv |x_1 y_1|\, (1+u)^{k-1} + \sum_{j=2}^{k} |x_j y_j|\, (1+u)^{k-j+1},$$

and for $2 \leq k \leq n$,

$$|Z_{2k-1} - Z_{2k-2}| \leq c_{2k-2}\, u, \qquad \text{where} \quad c_{2k-2} \equiv |x_k y_k|.$$

Thus, the conditions for Lemma 4.2 are satisfied, and we can use it to bound $|Z_{2n} - Z_1|$ with the above c_k from Lemma 4.2. ☐

Remark 4.1 (Comparison). *The probabilistic bound in Theorem 4.6 tends to be tighter than the deterministic bound in Theorem 4.3.*

The probabilistic bound in Theorem 4.6 holds with probability at least $1-\delta$,

$$\left|\frac{\mathrm{fl}(\mathbf{x}^T\mathbf{y})-\mathbf{x}^T\mathbf{y}}{\mathbf{x}^T\mathbf{y}}\right| = \left|\frac{\hat{s}_{2n}-s_{2n}}{s_{2n}}\right| \leq \frac{\sqrt{\sum_{k=1}^{2n-1} c_k^2}}{|\mathbf{x}^T\mathbf{y}|}\,\sqrt{2\,\ln{(2/\delta)}}\,u,$$

while the deterministic bound in Theorem 4.3 is

$$\left|\frac{\mathrm{fl}(\mathbf{x}^T\mathbf{y})-\mathbf{x}^T\mathbf{y}}{\mathbf{x}^T\mathbf{y}}\right| = \left|\frac{\hat{s}_{2n}-s_{2n}}{s_{2n}}\right| \leq \frac{\sqrt{\sum_{k=1}^{2n-1} c_k^2}}{|\mathbf{x}^T\mathbf{y}|}\,\sqrt{2n-1}\,u,$$

where $c_{2k-1}=|x_k y_k|$ *and* $c_{2k}=\sum_{j=1}^{k}|x_j y_j|(1+u)^{k-j+1}$, $1\leq k\leq n$.

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

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* $n\geq 39$, *and* $\sqrt{2\,\ln{(2/\delta)}}\leq 9$.

4.4. Simpler forward error bounds. We derive two upper bounds for Theorem 4.3 and 4.6 that have a simpler form, and then conform Wilkinson's intuition [20, Section 1.33].

The first bound is more compact than Theorem 4.6, and makes use of abbreviations for the leading subvectors of $|\mathbf{x}|\circ|\mathbf{y}|$, and vectors containing powers of $1+u$.

Corollary 4.7 (Compact upper bound). *Define the* k*-vectors*

$$(\mathbf{x}\circ\mathbf{y})_k \equiv \begin{pmatrix}|x_1y_1|\\ |x_2y_2|\\ \vdots\\ |x_ky_k|\end{pmatrix}, \qquad \mathbf{u}_k \equiv \begin{pmatrix}(1+u)^{k-1}\\ (1+u)^{k-1}\\ \vdots\\ 1+u\end{pmatrix}, \qquad 2\leq k\leq n.$$

If $\frac{1}{p}+\frac{1}{q}=1$, *then in Theorems 4.3 and 4.6 we have*

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

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 $c_1=|x_1y_1|$ and $c_{2k-2}=|x_ky_k|$, $2\leq k\leq n$, follows

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

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

$$\begin{aligned} c_{2k-1} &= |x_1y_1|\,(1+u)^{k-1} + \sum_{j=2}^{k}|x_jy_j|\,(1+u)^{k-j+1} \\ &= (\mathbf{x}\circ\mathbf{y})_k^T\,\mathbf{u}_k \leq \|(\mathbf{x}\circ\mathbf{y})_k\|_p\,\|\mathbf{u}_k\|_q, \qquad 2\leq k\leq n. \end{aligned}$$

□

The second bound, below, takes a much simpler form.

Corollary 4.8 (Simplest upper bound for Theorem 4.6). *Let the roundoffs δ_k be random variables with $\mathbb{E}[\delta_k]=0$ and $|\delta_k|\le u$, $1\le k\le 2n$; and let γ_k as in (3.1).*

Then for any $0<\delta<1$, with probability at least $1-\delta$, the relative forward error of the computed inner product $\hat{s}_{2n}=\mathrm{fl}(\mathbf{x}^T\mathbf{y})$ in Table 4.1 is bounded by

$$\left|\frac{\mathrm{fl}(\mathbf{x}^T\mathbf{y})-\mathbf{x}^T\mathbf{y}}{\mathbf{x}^T\mathbf{y}}\right| \le \frac{|\mathbf{x}|^T|\mathbf{y}|}{|\mathbf{x}^T\mathbf{y}|}\sqrt{2\ln(2/\delta)}\sqrt{\frac{u\,\gamma_{2n}}{2}}. \tag{4.2}$$

Proof. In Corollary 4.7, choose $p=1$ and $q=\infty$, so that

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

The relation between vector norms implies $\|\mathbf{x}\circ\mathbf{y}\|_2\le\|\mathbf{x}\circ\mathbf{y}\|_1$. Insert the preceding two inequalities into Corollary 4.7,

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

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

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

At last substitute this into Theorem 4.6. □

Remark 4.2 (Comparison with traditional bound). *We quantify and confirm Wilkinson's intuition [20, Section 1.33], by illustrating that the probabilistic bounds in Theorem 4.6, and Corollaries 4.7 and 4.8 are proportional to $\sqrt{n}\,u$, while the traditional bound in Corollary 3.4 is proportional to $n\,u$.*

Let $\gamma_k=(1+u)^k-1$, $k\ge 1$, be as in (3.1). The probabilistic bound in Corollary 4.8 holds with probability at least $1-\delta$,

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

while the deterministic bound in Corollary 3.4 equals

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

For large n, the bounds behave asymptotically like their first order terms,

$$\gamma_n\approx n\,u, \qquad \sqrt{\frac{u\,\gamma_{2n}}{2}}\approx\sqrt{n}\,u.$$

For small n with $2n\,u < 1$, one can bound [13, Lemma 3.1],

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

Thus, the probabilistic bound 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 $n > 80$.

5. Numerical experiments. After describing the setup for the experiments (Section 5.1), we present experiments for the perturbation bounds (Section 5.2), the roundoff error bounds assuming independence (Section 5.3), and the general roundoff error bounds (Section 5.4).

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 $\mathbf{x}$ and $\mathbf{y}$ of dimension up to $n = 10^8$ will be considered:

- The elements of $\mathbf{x}$ and $\mathbf{y}$ can have different signs. Specifically, x_j and y_j are iid[2] standard normal random variables with mean 0 and variance 1, and $\mathbf{x}$ and $\mathbf{y}$ are generated with the Matlab commands
```
x = single(rand(n, 1)), y = single(rand(n, 1))
```
- The elements of $\mathbf{x}$ and $\mathbf{y}$ all have the same sign. Specifically, x_j and y_j are absolute values of iid standard normal random variables, and $\mathbf{x}$ and $\mathbf{y}$ are generated with the Matlab commands
```
x = single(abs(rand(n, 1))), y = single(abs(rand(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 $2^{-53} \approx 1.11 \cdot 10^{-16}$. Bounds are computed in double precision. Computations were performed in Matlab R2017a, on a 3.1GHz Intel Core i7 processor.

5.2. Experiments for the perturbation bounds. We illustrate the perturbation bounds in Section 2. Here the vectors $\mathbf{x}$ and $\mathbf{y}$ are perturbed, while the computations are exact.

We select single precision perturbations δ_j and θ_j that are uniformly distributed in $[-u, u]$, where $u = 2^{-24} \approx 5.96 \cdot 10^{-8}$ is the single precision roundoff, and generate the perturbation vectors $\boldsymbol{\delta}$ and $\boldsymbol{\theta}$ each with the Matlab command
```
u * (2 * double(single(rand(n, 1))) - ones(n, 1)).
```
The inner product of the perturbed vectors $\hat{\mathbf{x}}^T\hat{\mathbf{y}}$ is represented by the double precision computation `dot(double(xh), double(yh))`.

5.2.1. Amplifiers in Corollary 2.2. We compare the amplifiers of $u(2+u)$ in the upper bounds of Corollary 2.2, listed again below,

$$\kappa_1 \equiv \frac{\|\mathbf{x} \circ \mathbf{y}\|_1}{|\mathbf{x}^T\mathbf{y}|} = \frac{|\mathbf{x}|^T|\mathbf{y}|}{|\mathbf{x}^T\mathbf{y}|} = \frac{\sum_{j=1}^n |x_j y_j|}{|\mathbf{x}^T\mathbf{y}|} \tag{5.1}$$

$$\kappa_2 \equiv \sqrt{n}\,\frac{\|\mathbf{x} \circ \mathbf{y}\|_2}{|\mathbf{x}^T\mathbf{y}|} = \sqrt{n}\,\frac{\sqrt{\sum_{j=1}^n |x_j y_j|^2}}{|\mathbf{x}^T\mathbf{y}|}$$

$$\kappa_\infty \equiv n\,\frac{\|\mathbf{x} \circ \mathbf{y}\|_\infty}{|\mathbf{x}^T\mathbf{y}|} = n\,\frac{\max_{1\le j\le n} |x_j y_j|}{|\mathbf{x}^T\mathbf{y}|}.$$

[2] independent identically distributed

Figure 5.1 illustrates that, among the three amplifiers in (5.1), the traditional κ_1 tends to be the lowest. It also illustrates that amplification of roundoff can be orders of magnitude larger for vector elements with different signs, compared to vectors where all elements have the same sign.

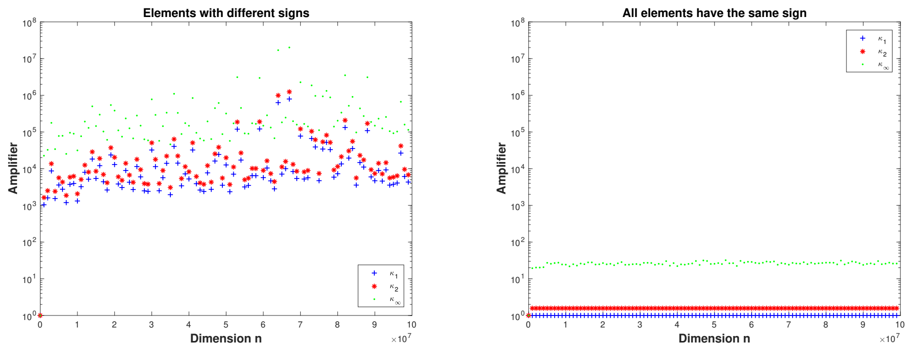

FIG. 5.1. *Comparison of amplifiers in (5.1): κ_1 (blue), κ_2 (red), and κ_∞ (green) versus vector dimensions $1 \leq n \leq 10^8$ in steps of 10^6. Vertical axis starts at 1 and ends at 10^8. Left panel: Elements can have different signs. Right panel: All elements have the same sign.*

5.2.2. Probabilistic perturbation bound in Theorem 2.4 and Remark 2.1. This experiment follows up on Remark 2.1, where we compare the probabilistic bound from Theorem 2.4 to the corresponding deterministic bound from Corollary 2.2.

- Deterministic bound

$$\left| \frac{\hat{\mathbf{x}}^T\hat{\mathbf{y}} - \mathbf{x}^T\mathbf{y}}{\mathbf{x}^T\mathbf{y}} \right| \leq \frac{\|\mathbf{x} \circ \mathbf{y}\|_2}{|\mathbf{x}^T\mathbf{y}|} \sqrt{n}\, u(2+u). \tag{5.2}$$

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

$$\left| \frac{\hat{\mathbf{x}}^T\hat{\mathbf{y}} - \mathbf{x}^T\mathbf{y}}{\mathbf{x}^T\mathbf{y}} \right| \leq \frac{\|\mathbf{x} \circ \mathbf{y}\|_2}{|\mathbf{x}^T\mathbf{y}|} \sqrt{2\,\ln(2/\delta)}\, u(2+u). \tag{5.3}$$

Figure 5.2 illustrates that the probabilistic bound (5.3) tends to be at least two orders orders of magnitude tighter than the deterministic bound (5.2).

5.3. Experiments for the roundoff error bounds based on independent roundoff. We illustrate the roundoff error bounds in Section 3.

The inner products $\mathrm{fl}(\mathbf{x}^T\mathbf{y})$ are computed in single precision with unit roundoff, 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.

Specifically, we compare the probabilistic bound in Theorem 3.6 with the corresponding deterministic bound in Corollary 3.5.

- Deterministic bound

$$\left| \frac{\mathrm{fl}(\mathbf{x}^T\mathbf{y}) - \mathbf{x}^T\mathbf{y}}{|\mathbf{x}^T\mathbf{y}|} \right| \leq \frac{\sqrt{\sum_{k=1}^n c_k^2}}{|\mathbf{x}^T\mathbf{y}|} \sqrt{n} \tag{5.4}$$

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

$$\left| \frac{\mathrm{fl}(\mathbf{x}^T\mathbf{y}) - \mathbf{x}^T\mathbf{y}}{\mathbf{x}^T\mathbf{y}} \right| \leq \frac{\sqrt{\sum_{k=1}^n c_k^2}}{|\mathbf{x}^T\mathbf{y}|} \sqrt{2\,\ln(2/\delta)}, \tag{5.5}$$

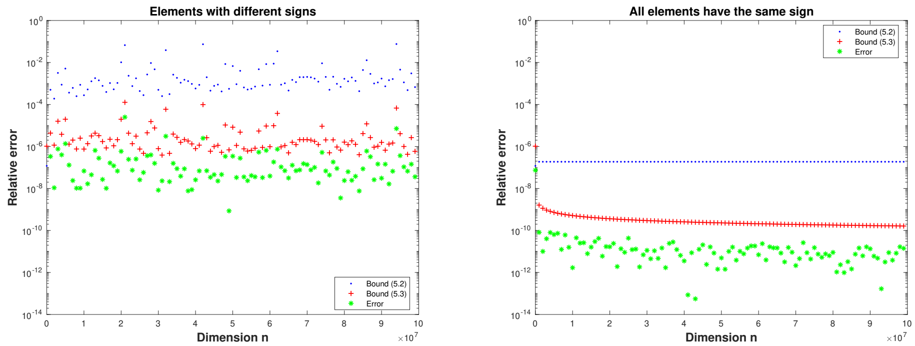

FIG. 5.2. *Comparison of probabilistic bound (red 5.3) with deterministic bound (blue 5.2), and relative error (green) versus vector dimensions* $1 \le n \le 10^8$ *in steps of* 10^6*. Vertical axis starts at* 10^{-14} *and ends at 1. Left panel: Elements can have different signs. Right panel: All elements have the same sign.*

where $c_1 \equiv |x_1 y_1|\, \gamma_n$, and $c_k \equiv |x_k y_k|\, \gamma_{n-k+2}$, $2 \le k \le n$, and $\gamma_k = (1+u)^k - 1$ as in (3.1).

Figure 5.3 illustrates that the probabilistic result (5.5) tends to be two orders of magnitude tighter than the deterministic bound (5.4) for vectors whose elements can have different signs. However, (5.5) stops being a bound for vectors of large dimension all of whose elements have the same sign.

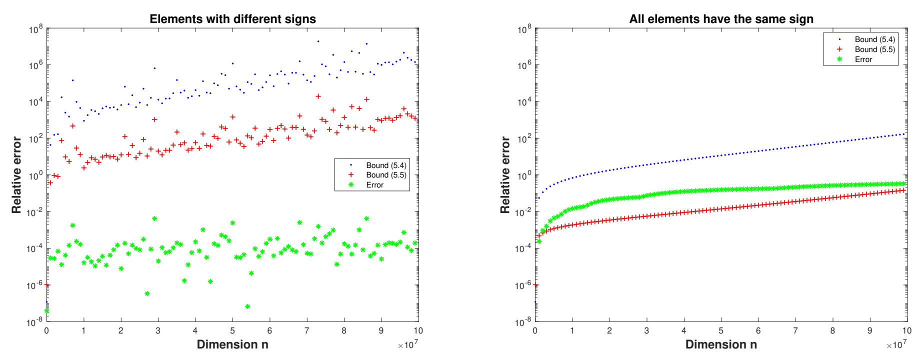

FIG. 5.3. *Comparison of probabilistic bound (red 5.5) with deterministic bound (blue 5.4), and relative error (green) versus vector dimensions* $1 \le n \le 10^8$ *in steps of* 10^6*. Vertical axis starts at* 10^{-8} *and ends at* 10^8*. Left panel: Elements can have different signs. Right panel: All elements have the same sign.*

Figure 5.4 zooms in on the left panel in Figure 5.3 and illustrates that (5.5) remains an upper bound for vector dimensions up to about $n = 10^6$. The fact that it ceases to be an upper bound for $n > 10^6$ does not appear to be a numerical issue, as nothing changes when the products $|x_k y_k|$ are sorted in increasing or in decreasing order of magnitude.

5.4. Experiments for the general roundoff error bounds. We illustrate the roundoff error bounds in Section 4.

As in the previous section, the inner products $\mathrm{fl}(\mathbf{x}^T\mathbf{y})$ are computed in single precision with unit roundoff, in a loop that explicitly stores the products $x_k y_k$ before

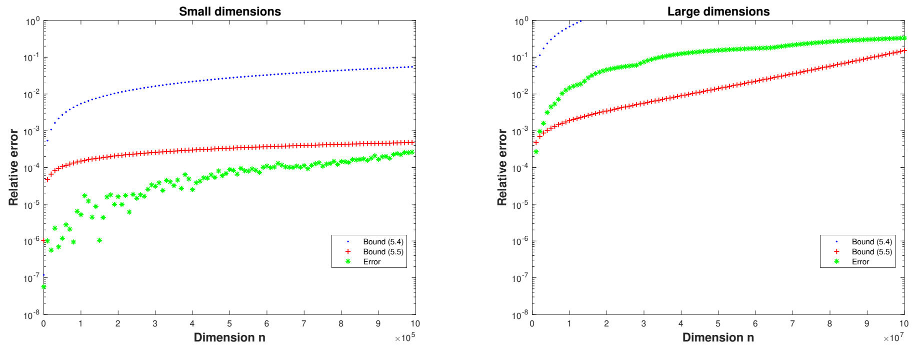

FIG. 5.4. *Comparison of probabilistic bound (red 5.5) with deterministic bound (blue 5.4), and relative error (green) versus vector dimensions when all elements have the same sign. Vertical axis starts at* 10^{-8} *and ends at* 1. *Left panel: Small dimensions* $1 \leq n \leq 10^6$ *in steps of* 10^4. *Right panel: Large dimensions* $10^6 \leq n \leq 10^8$ *in steps of* 10^6.

adding them to the partial sum, so as to bypass the fused multiply-add.

This experiment follows up on Remark 4.2, where we compare the probabilistic bound in Corollary 4.8 to the corresponding deterministic bound in Corollary 3.4.

- Traditional bound

$$\left|\frac{\mathrm{fl}(\mathbf{x}^T\mathbf{y}) - \mathbf{x}^T\mathbf{y}}{\mathbf{x}^T\mathbf{y}}\right| \leq \frac{|\mathbf{x}|^T|\mathbf{y}|}{|\mathbf{x}^T\mathbf{y}|}\,\gamma_n, \tag{5.6}$$

- Probabilistic bound

$$\left|\frac{\mathrm{fl}(\mathbf{x}^T\mathbf{y}) - \mathbf{x}^T\mathbf{y}}{\mathbf{x}^T\mathbf{y}}\right| \leq \frac{|\mathbf{x}|^T|\mathbf{y}|}{|\mathbf{x}^T\mathbf{y}|}\,\sqrt{\ln(2/\delta)}\,\sqrt{\frac{u\,\gamma_{2n}}{2}}, \tag{5.7}$$

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

Figure 5.5 illustrates that the probabilistic result (5.7) tends to be at least two orders of magnitude tighter than the deterministic bound (5.6) for vectors whose elements can have different signs. However, unfortunately, (5.7) stops being a bound for vectors of large dimension all of whose elements have the same sign.

6. Conclusions, and future work. We presented derivations and numerical experiments for probabilistic perturbation and roundoff error bounds for the sequentially accumulated inner product of two real n-vectors, assuming a guard digit model and no fused multiply-add. The probabilistic bounds are tighter than the corresponding deterministic bounds, often by several orders of magnitude.

Issues. However, for vectors of dimension $n \geq 10^7$ 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 stops being upper bounds for vectors all of whose elements have the same sign –regardless of whether roundoffs are assumed to be independent or not. The latter phenomenon does not appear to be a numerical artifact.

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

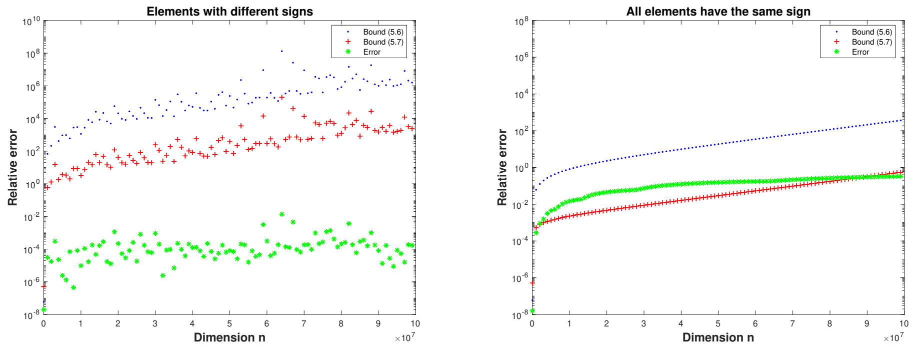

FIG. 5.5. *Comparison of probabilistic bound (red 5.7) with deterministic bound (blue 5.6), and relative error (green) versus vector dimensions $1 \leq n \leq 10^8$ in steps of 10^6. Vertical axis starts at 10^{-8} and ends at 10^8. Left panel: Elements can have different signs. Right panel: All elements have the same sign.*

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].

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

REFERENCES

[1] I. Babuška and G. Söderlind, *On roundoff error growth in elliptic problems*, ACM Trans. Math. Software, 44 (2018), pp. Art. 33, 22.

[2] E. H. Bareiss and J. L. Barlow, *Roundoff error distribution in fixed point multiplication*, BIT, 20 (1980), pp. 247–250.

[3] J. L. Barlow and E. H. Bareiss, *On roundoff error distributions in floating point and logarithmic arithmetic*, Computing, 34 (1985), pp. 325–347.

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

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

[6] M.-C. Brunet and F. Chatelin, *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] D. Calvetti, *Roundoff error for floating point representation of real data*, Comm. Statist. Theory Methods, 20 (1991), pp. 2687–2695.

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

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

[10] F. Chatelin and M.-C. Brunet, *A probabilistic round-off error propagation model. Application to the eigenvalue problem*, in Reliable numerical computation, Oxford Sci. Publ., Oxford Univ. Press, New York, 1990, pp. 139–160.

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

[12] P. Henrici, *Problems of stability and error propagation in the numerical integration of ordinary differential equations*, in Proc. Internat. Congr. Mathematicians (Stockholm 1962), Inst. Mittag-Leffler, Djursholm, 1963, pp. 102–113.

[13] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM, Philadelphia, second ed., 2002.
[14] N. J. Higham and T. Mary, *A new approach to probabilistic rounding error analysis*, MIMS EPrint 2018.33, University of Manchester, 2018.
[15] T. E. Hull and J. R. Swenson, *Tests of probabilistic models for the propagation of roundoff errors*, Comm. ACM, 9 (1966), pp. 108–113.
[16] W. Kahan, *The improbability of probabilistic error analyses for numerical computations*, March 1996.
[17] M. Mitzenmacher and E. Upfal, *Probability and Computing*, Cambridge University Press, Cambridge, 2005. Randomized Algorithms and Probabilistic Analysis.
[18] M. Tienari, *A statistical model of roundoff error for varying length floating-point arithmetic*, Nordisk Tidskr. Informationsbehandling (BIT), 10 (1970), pp. 355–365.
[19] J. von Neumann and H. H. Goldstine, *Numerical inverting of matrices of high order*, Bull. Amer. Math. Soc., 53 (1947), pp. 1021–1099.
[20] J. H. Wilkinson, *Rounding errors in algebraic processes*, Dover Publications, Inc., New York, 1994. Reprint of the 1963 original [Prentice-Hall, Englewood Cliffs, NJ].