Error Accumulation in Summation

Let E_N be the total rounding error after summing N terms, where δ_i is the error introduced at the i-th step:

$$E_N = \sum_{i=1}^{N} \delta_i$$

We assume the individual rounding errors δ_i are independent random variables with mean $\mu = \mathbb{E}[\delta_i]$ and variance $\sigma^2 = \mathbb{V}[\delta_i]$.

Error Metrics

The total error E_N is characterized by:

- **Bias**: $B = \mathbb{E}[E_N]$
- **Variance**: $\mathbb{V}[E_N] = \mathbb{E}[(E_N - B)^2]$
- **Standard Deviation**: $\sqrt{\mathbb{V}[E_N]}$
- **Root Mean Square Error (RMSE)**: $\text{RMSE} = \sqrt{\mathbb{E}[E_N^2]} = \sqrt{B^2 + \mathbb{V}[E_N]}$

Error Statistics per Rounding Mode (Without Algorithmic Tricks)

Assuming independence and identical distribution for δ_i:

- The total bias is $B = \sum_{i=1}^{N} \mathbb{E}[\delta_i] = N\mu$.
- The total variance is $\mathbb{V}[E_N] = \sum_{i=1}^{N} \mathbb{V}[\delta_i] = N\sigma^2$.

Comparing Deterministic Rounding (Det) and Stochastic Rounding (SR):

- **Deterministic Rounding (Det)**:
  - Individual errors δ_i may have non-zero mean: $\mu_{\text{det}} \neq 0$.
  - Individual errors have variance σ^2_{det}.
  - Total Bias: $B_{\text{det}} = N\mu_{\text{det}}$.
  - Total Variance: $\mathbb{V}[E_N]_{\text{det}} = N\sigma^2_{\text{det}}$.
  - Total RMSE: $\text{RMSE}_{\text{det}} = \sqrt{(N\mu_{\text{det}})^2 + N\sigma^2_{\text{det}}}$.
- **Stochastic Rounding (SR)**:
  - Designed to be unbiased: $\mu_{\text{sr}} = \mathbb{E}[\delta_i] = 0$.
  - Individual errors typically have higher variance than Det: $\sigma^2_{\text{sr}} \geq \sigma^2_{\text{det}}$. (SR introduces randomness where Det might have a fixed offset).
  - Total Bias: $B_{\text{sr}} = N\mu_{\text{sr}} = 0$.
  - Total Variance: $\mathbb{V}[E_N]_{\text{sr}} = N\sigma^2_{\text{sr}}$.
  - Total RMSE: $\text{RMSE}_{\text{sr}} = \sqrt{0^2 + N\sigma^2_{\text{sr}}} = \sqrt{N}\sigma_{\text{sr}}$.

Comparison without tricks: SR outperforms Det ($\text{RMSE}_{\text{sr}} < \text{RMSE}_{\text{det}}$) when the deterministic bias term $N|\mu_{\text{det}}|$ grows faster or is significantly larger than the difference in standard deviation terms. This occurs when N is large and μ_{det} is non-negligible, as RMSE_{det} grows roughly linearly with N due to bias, while RMSE_{sr} grows as $\sqrt{N}$.

Effect of DDF-Shifting and Alternating Summation

These algorithmic techniques modify the sequence of operations and the values being summed, thereby changing the characteristics of the individual rounding errors δ_i under deterministic rounding. Let the modified errors be δ'_i.

- They reduce the magnitude of numbers involved in sums/differences.
- They increase the symmetry of rounding errors (less systematic bias).

The net effect is that for deterministic rounding, these tricks significantly reduce the bias per operation and potentially the variance:

- $\mu'_{\text{det}} \approx 0$ (deterministic bias is largely eliminated).
- $\sigma'_{\text{det}} \leq \sigma_{\text{det}}$ (variance may also decrease).

Error Statistics per Rounding Mode (With Algorithmic Tricks)

Let the statistics under the modified algorithm be denoted with a prime (').

- **Deterministic Rounding (Det) with Tricks**:
  - Total Bias: $B'_{\text{det}} = N\mu'_{\text{det}} \approx 0$.
  - Total Variance: $\mathbb{V}[E_N]'_{\text{det}} = N(\sigma'_{\text{det}})^2$.
  - Total RMSE: $\text{RMSE}'_{\text{det}} = \sqrt{(N\mu'_{\text{det}})^2 + N(\sigma'_{\text{det}})^2} \approx \sqrt{N}\sigma'_{\text{det}}$.
- **Stochastic Rounding (SR) with Tricks**:
  - SR remains unbiased by design: $\mu'_{\text{sr}} = 0$.
  - The tricks might reduce the scale of numbers, possibly reducing SR variance compared to the raw case: $\sigma'_{\text{sr}} \leq \sigma_{\text{sr}}$. However, SR still introduces randomness, so it's expected that $\sigma'_{\text{sr}} > \sigma'_{\text{det}}$.
  - Total Bias: $B'_{\text{sr}} = 0$.
  - Total Variance: $\mathbb{V}[E_N]'_{\text{sr}} = N(\sigma'_{\text{sr}})^2$.
  - Total RMSE: $\text{RMSE}'_{\text{sr}} = \sqrt{N}\sigma'_{\text{sr}}$.

Comparison with tricks: Since the algorithmic tricks reduce the deterministic bias μ'_{det} to near zero, the primary advantage of SR (eliminating bias) is nullified. The comparison now hinges primarily on the variances. If the deterministic scheme with tricks achieves near-zero bias ($\mu'_{\text{det}} \approx 0$) and its inherent variance per step (σ'_{det}) is lower than the variance introduced by SR ($\sigma'_{\text{sr}} > \sigma'_{\text{det}}$), then:

$$\text{RMSE}'_{\text{det}} \approx \sqrt{N}\sigma'_{\text{det}} < \sqrt{N}\sigma'_{\text{sr}} = \text{RMSE}'_{\text{sr}}$$

In this scenario, deterministic rounding combined with the algorithmic improvements outperforms stochastic rounding. SR becomes detrimental because it adds variance without providing a significant bias-reduction benefit (as the bias is already handled algorithmically).