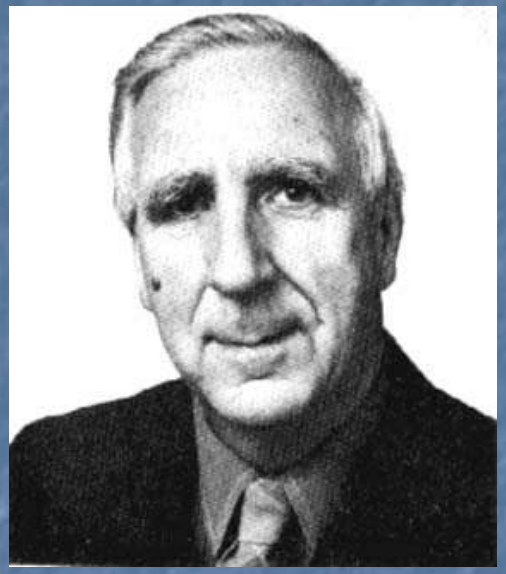
Quantitative Conditioning Criteria in Bayesian Automatic Adaptive Quadrature

Gheorghe Adam & Sanda Adam

LIT-JINR Dubna & IFIN-HH Bucharest

ROLCG 2012, Cluj-Napoca, Oct.25, 2012

Scientific floating point computation: numerical music or numerical noise?



Leslie Fox:

IMA Bull., **7**, 296-302 (1971),

"I have little doubt that **about 80 per cent of all the results printed from the computer are in error** to a much greater extent that the user would believe .."

Other enhancements of this statement:

- M.G. Cox and P.M. Harris, **Numerical analysis for algorithm design in metrology**, 83 pp., April 2004
- Sven Hammarling, **An introduction to the quality of computed solutions**, in B. Einarsson, Ed., *Accuracy and Reliability in Scientific Computing*, pp. 43-76, SIAM, Philadelphia, PA, USA, 2005
- N.S. Scott, et al., **e-Collisions using e-Science,** MMCP 2006, High Tatra Mountains, Slovakia, Aug. 28 – Sept. 01 2006

*The **modeling of physical phenomena** within **numerical experiments** often asks for the evaluation of huge numbers of **Riemann integrals** by numerical methods.*

*The study of the behavior of a system under **sudden change** of an **inner order parameter**, which results in drastic modification of the mathematical properties of the integrand (e.g., in phase transitions or processes involving fragmentation or fusion) **cannot be accommodated** within the **standard automatic adaptive quadrature** (**AAQ**) approach to the numerical solution due to the impossibility to decide in advance on the correct choice of the convenient library procedure.*

*The **Bayesian automatic adaptive quadrature** (**BAAQ**) tries to solve integrals by **merging rigorous mathematical criteria** with the **reality** of the **hardware** and **software environments**.*

*The BAAQ generates a **subrange binary tree** associated to the automatic subrange subdivision process under **probabilistic check** of a hierarchically ordered set of **conditioning criteria** which ensure the **exclusion** of **each** of the following five kinds of unacceptable integrand features at **each tree node**:*

1. ***Catastrophic cancellation** by subtraction.*
2. ***Integrand oscillations** at a rate beyond the resolving power of the current integrand profile.*
3. ***Rates of variation** of the integrand inside monotonicity subranges beyond maximally allowed polynomial thresholds.*
4. *Inner **isolated integrand discontinuities**.*
5. ***Irregular**, unsolvable by quadratures, **behavior** of the integrand function inside the integration domain.*

*The present paper derives **quantitative conditioning criteria** able to replace the empirical ones previously formulated for the **solution of the third feature** mentioned above.*

Overview

- Data sets subject to analysis
- Well-conditioning Ansatz
- Quantitative well-conditioning criteria
- Discussion

Gradual partition refinement

The AAQ approach results in an *integrand adapted* discretization of $[a,b]$ which defines a partition of $[a,b]$,

$$\Pi_N[a,b]=\{a=x^0<x^1<\cdots<x^i<\cdots<x^N=b\,|\,N\geq 1\}.$$

Over each subrange of the partition Π_N a *local quadrature rule* yields an approximate value q of the integral, together with a local error estimate $e>0$. The *global* approximate solution $(Q_N, E_N>0)$ is got by *summing up* the individual outputs over subranges.

A gradual refinement of Π_N is performed until some prescribed accuracy condition is fulfilled, or a failure diagnostic is issued. This can be described in terms of an evolving subrange binary tree over which BAAQ damps down the five unwanted features mentioned above.

Interpolatory quadrature sums

An interpolatory quadrature sum approximates a proper or improper one-dimensional Riemann integral,

$$I^w[\alpha,\beta]f = \int_\alpha^\beta w(x)\,f(x)dx$$

by means of an interpolatory algebraic polynomial,

$$q_{n+1}[\alpha,\beta]f = I^w[\alpha,\beta]p_n$$

the values of which equate those of the integrand function $f(x)$ at a specific set of quadrature knots x_k ,

$$p_n(x_k) = f(x_k),\ k = 0, 1, \ldots, n.$$

Integrand profiles

Definition 1. A local quadrature rule which asks for $2n+1$ inner quadrature knots over the node $[\alpha, \beta]$ of the subrange binary tree associates to the integrand function $f(x)$ the *current integrand profile* for $f_k = f(x_k)$,

$$\{(x_k, f_k), k = 0,1,\cdots,2n+2 | \alpha = x_0 < \cdots x_k < \cdots x_{2n+2} = \beta\}$$

Each *ancestor* of a terminal node in the subrange binary tree is characterized by its *own* current integrand profile set, which is *inherited* by its *descendents*.

Definition 2. For each descendent in the subrange binary tree, values of the integrand function are available for Bayesian analysis over the *extended integrand profile*,

$$\{(x_\mu, f_\mu = f(x_\mu)), \mu = 0,1,\cdots,M | \alpha = x_0 < \cdots x_\mu < \cdots x_M = \beta\}$$

Integrand profile distributions sampling

•The integrand profile subject to Bayesian analysis is not necessarily restricted to a single subrange only. It covers any convenient subset of the global integrand sampling associated to the partition Π_N.

• The most frequent local quadrature rules use integrand values at Gauss-Kronrod (GK) or Clenshaw-Curtis (CC) quadrature knots.

• Their property of being spanned by characteristic basis sets of orthogonal polynomials results in highly non-uniform distributions of the quadrature knots inside each subrange of Π_N: the highest around the subrange ends, the lowest around its centre.

• Thus, the reliability of the validation of a *tentative* inner discretization abscissa x^i of Π_N is significantly stregthened provided the Bayesian analysis setting the conditioning diagnostic of the integrand is done inside a neighborhood which contains pieces of integrand profiles from the two subranges adjacent at x^i.

Overview

- Data sets subject to analysis
- Well-conditioning Ansatz
- Quantitative well-conditioning criteria
- Discussion

The well-conditioning Ansatz

- ***Definition of monotonicity intervals*** within a piece of integrand profile of interest asks for the computation of the integrand variations $\delta_{k,k+1} = f_{k+1} - f_k$ in-between successive abscissas x_k and x_{k+1}. An abscissa x_k approximates an extremum $x^{(\lambda)}$ of the integrand provided $\delta_{k-1,k} \cdot \delta_{k,k+1} < 0$.

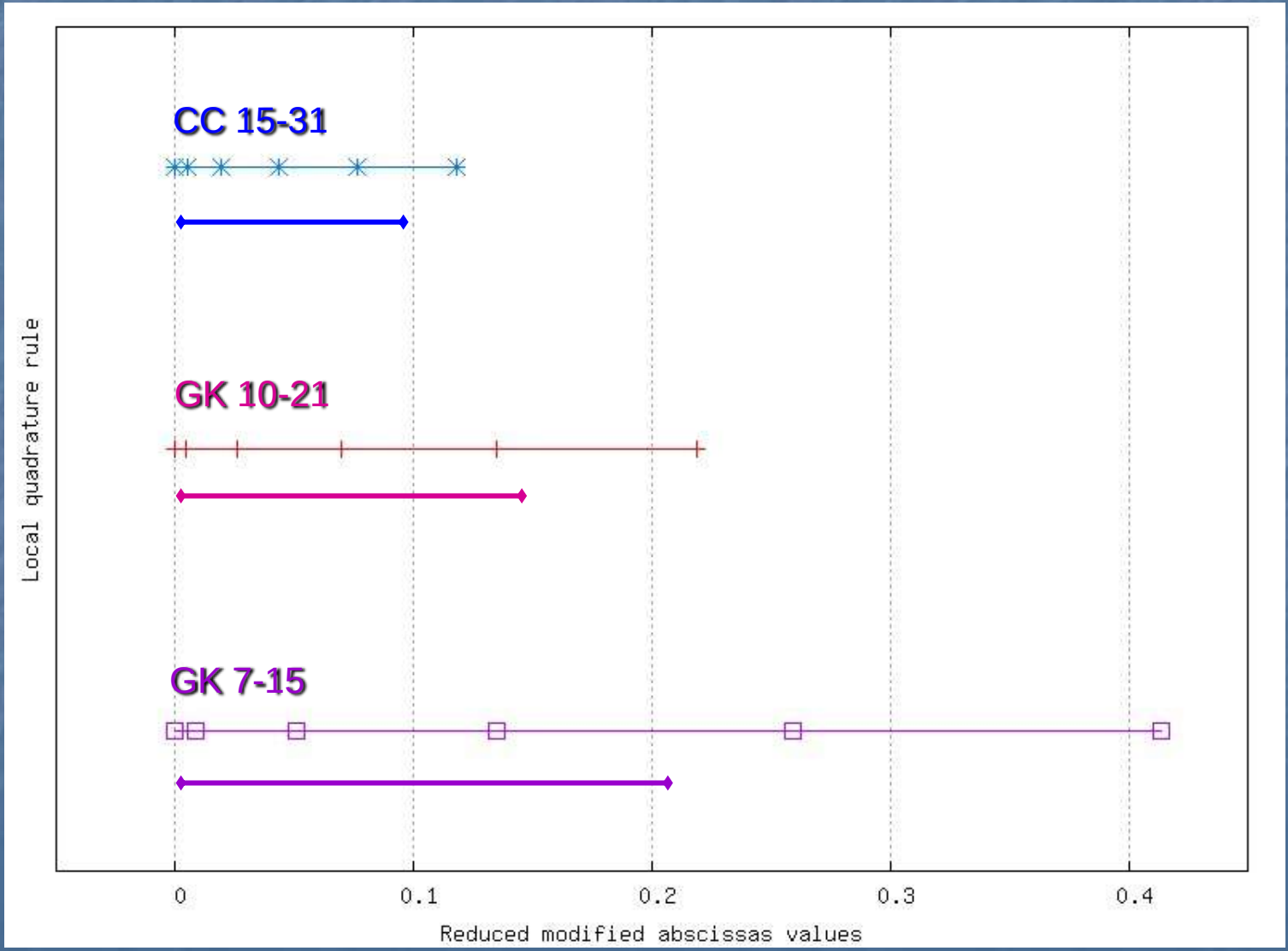
The set of intervals $[x^{(\lambda-1)}, x^{(\lambda)}]$ obtained in this way define the monotonicity intervals of interest.

- ***Ansatz 1.*** Let $\{x_{k-1} < x_k < x_{k+1}\}$ denote three successive abscissas in the integrand sampling over a monotonicity interval.

(*a*) If $x_{k+1} - x_{k-1} < \nu_{\mathrm{B}}$, then the expected rate of variation of a well-conditioned integrand over $[x_{k-1}, x_{k+1}]$ cannot exceed that of a second degree polynomial.

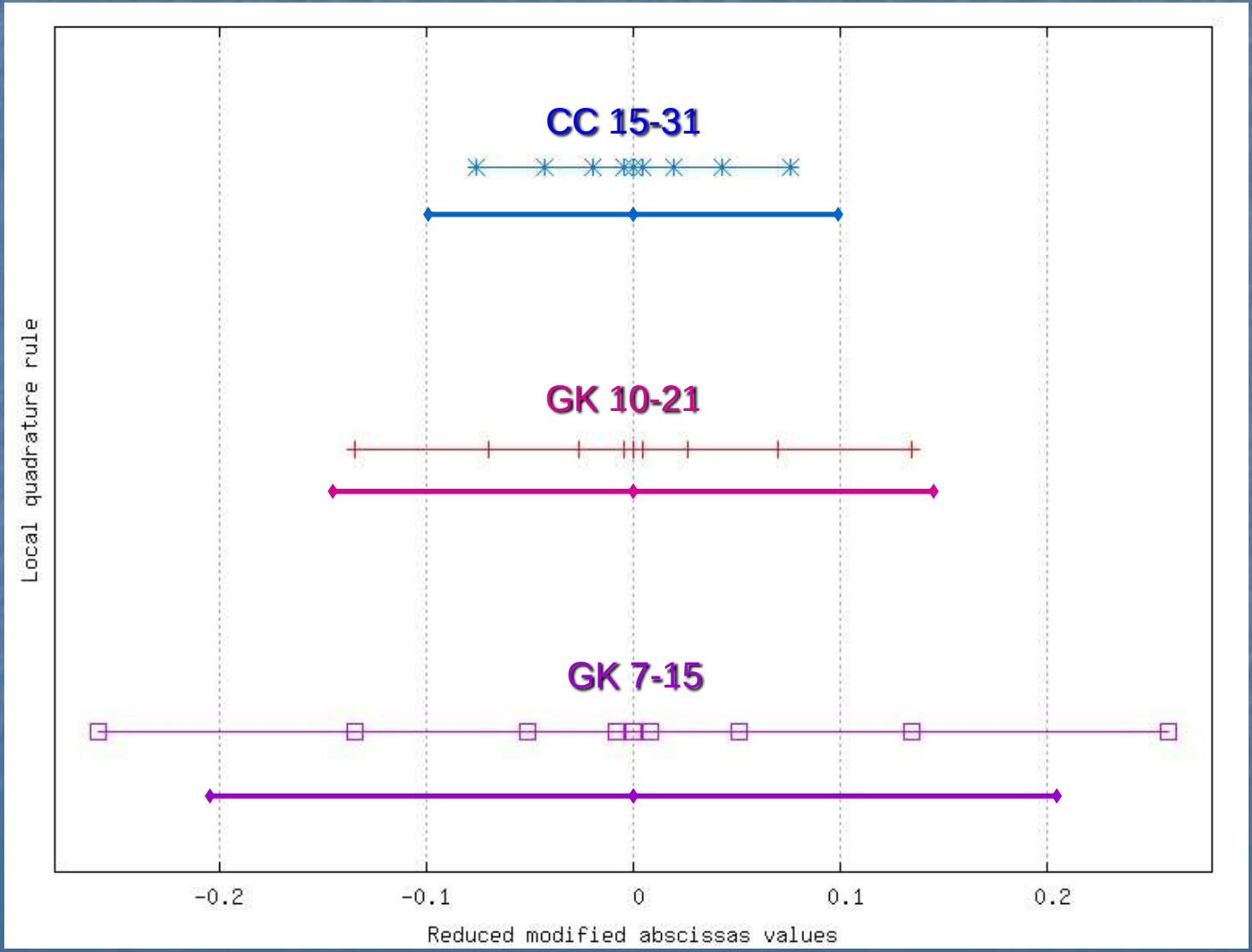
(*b*) If $x_{k+1} - x_{k-1} \geq \nu_{\mathrm{B}}$, then this rate of variation cannot exceed that of a third degree polynomial.

Lateral close proximity neighbourhoods of *terminal* subrange ends



*Contain **four** inner quadrature knots over distances equating the inter-knot distances at subrange centre.*

Central close proximity neighbourhoods of *two-sided* subrange ends



Contain ***seven*** *inner quadrature knots over distances equating the inter-knot distances at subrange centre.*

The Bayesian sensitivity norm, ν_B

- ***Definition***: $\nu_B = \max\{\nu_d, \nu_{fp}\}$.
- The ***algebraic degree of precision*** (d) ***related norm*** is defined over the current integrand profile as

$$\nu_d = \max_{1 \le k \le 2n+2} (x_k - x_{k-1}) = x_{n+1} - x_n = x_{n+2} - x_{n+1},$$

It sets an upper bound to the Bayesian analysis resolving power.

- The ***floating point degree of precision*** (d_{fp}) ***related norm*** over a subrange $[\alpha, \beta] \subseteq [a, b]$ is defined as

$$\nu_{fp} = (\beta - \alpha)/(d_{fp} - 1).$$

- ***Discussion***: ν_d prevails over usual subranges. However, ν_{fp} is the appropriate quantity on both very small or very large subranges, or on subranges very far apart from the origin, where $d_{fp} \gg d$.

Algebraic Degree of Precision

The quadrature sum $q_{n+1}[\alpha,\beta]f$ solves exactly the polynomial integrals over the fundamental power set,

$$q_{n+1}[\alpha,\beta]x^k = I[\alpha,\beta]x^k, \forall k = 0,1,\cdots,d, \ \forall[\alpha,\beta] \subset \mathbf{R}$$

The maximum degree d, at which these identities hold, defines the *algebraic degree of precision* of the quadrature sum $q_{n+1}[\alpha,\beta]f$.

In the literature, the algebraic degree of precision, d, is considered to be a *specific universal parameter* of a given interpolatory quadrature sum, irrespective of the extent and localization of the integration domain on the real axis.

Floating Point Degree of Precision (1)

In the calculation over $\mathbf{R}$ of the set of probe integrals

$$\sigma_m = I[0,\beta]\pi_m,\ \pi_m(x) = \sum_{l=0}^{m} x^l,\ \beta > 0,\ m = 0,1,\cdots,d$$

each monomial x^l entering the integrand $\pi_m(x)$ brings a ***distinct, non-negligible***, contribution to σ_m .

In floating point computations, the above property of the monomials x^l of bringing distinct, non-negligible contributions to σ_m may get infringed both at integration limits $\beta << 1$ and $\beta >> 1$.

The maximum degree $d_{\mathrm{fp}} \le d$ at which the identity of the individual monomial contributions is preserved in floating point computations defines the *floating point degree of precision* of the quadrature sum.

Its definition is formalized in the next two slides.

Floating Point Degree of Precision (2*a*)

1. Let $[\alpha, \beta] \subset \mathbf{R}$ denote the integration range of interest.
2. Let $q[\alpha, \beta] f$, a quadrature sum of algebraic degree of precision d , be computed over a set of t-bit floating point machine numbers ($t = 52$ in double precision).
3. Let $\xi > 0$, let $fl(a)$ denote the floating point approximate of $a \in \mathbf{R}$, and let $[a]$ denote the ceiling of $fl(a)$.
4. Let

$$d_\xi = \begin{cases} d & \text{iff } \xi \in [x_m, x_M] \\ [\ln \varepsilon_0 / \ln \xi] & \text{iff } \xi < x_m \\ [-\ln \varepsilon_0 / \ln \xi] & \text{iff } \xi > x_M \end{cases}$$

where $\varepsilon_0 = 2^{-t}$, $x_m = \varepsilon_0^{1/d}$, $x_M = x_m^{-1}$.

Floating Point Degree of Precision ($2b$)

5. For the integration range $[\alpha, \beta]$ we define

$$X = \max\{ fl(|\alpha|), fl(|\beta|)\}, \quad X > 0,$$

$$\rho = fl(|\beta - \alpha| / X), \quad 0 < \rho \leq 2.$$

The quantities d_X and d_ρ are computed from 4.

6. Then the *floating point degree of precision*, $d_{\text{fp}} \leq d$ associated to $q[\alpha, \beta] f$ is the positive integer

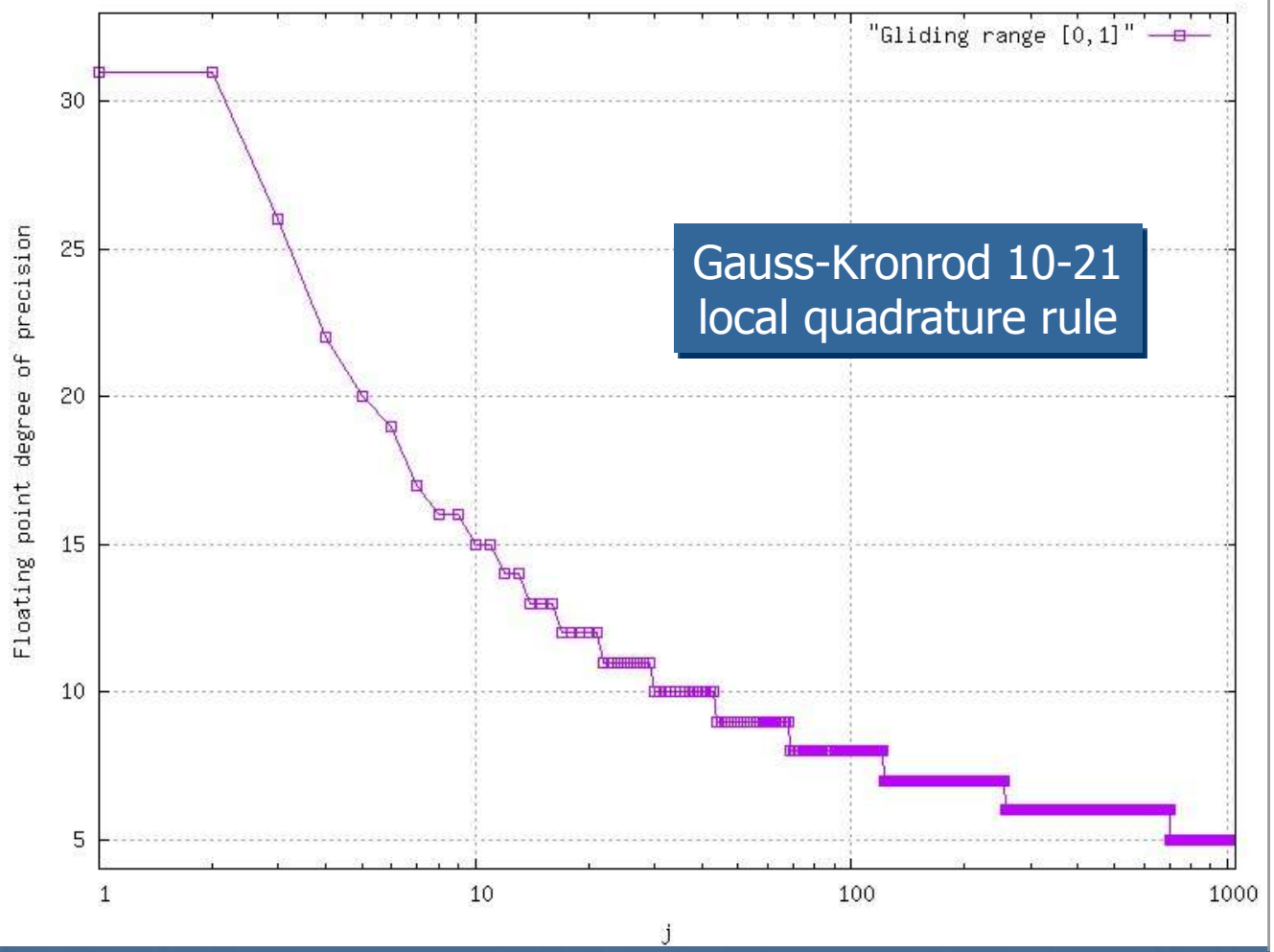
$$d_{\text{fp}} = \max\{\min\{d_X, d_\rho\}, 3\}$$

Features of the Floating Point Degree of Precision

- *Gliding integration range* [0,1] *on the real axis.*

The following plot gives outputs for the family of 1024 integration ranges

$$\{[j\alpha, j\alpha + \beta], \alpha = \beta = 1; j = 0, 1, ..., 1023\}$$



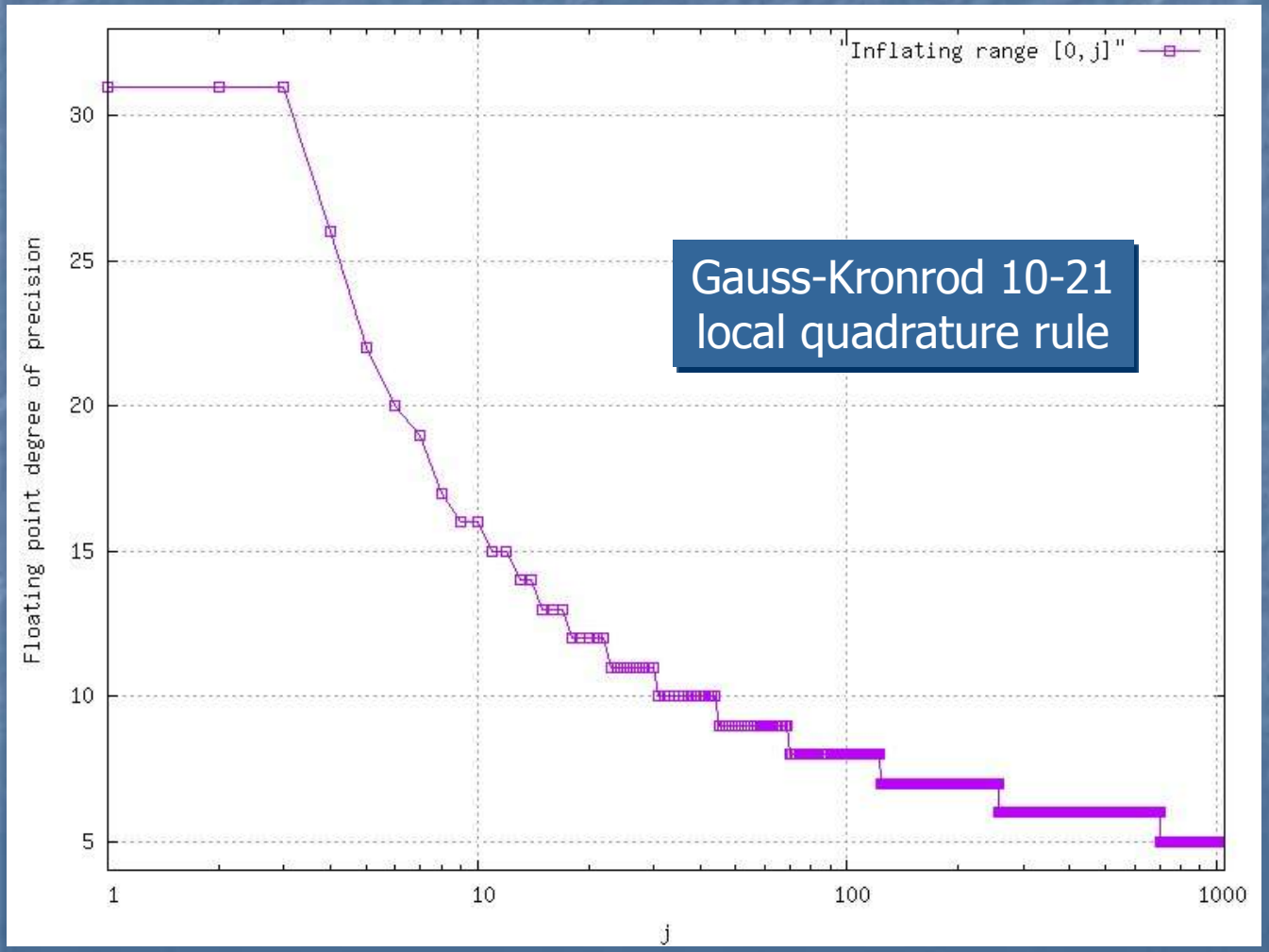
Variation of the floating point degree of precision of the GK 10-21 local quadrature rule over the gliding range* [0, 1] *versus its distance j from the origin. It is shown that $d_{fp} = d = 31$ at low j values ($j = 0, 1, 2$), then d_{fp} abruptly decreases at larger but small enough j, to show slower decreasing rates under the displacement of [0,1] far away from the origin, reaching a bottom value $d_{fp} = 5$ at $701 \leq j \leq 1023$.

Features of the Floating Point Degree of Precision

- *Inflating integration range* [0,j] *on the real axis.*

The following plot gives outputs for the family of 1023 integration ranges

$\{[0, j], j = 1, 2, ..., 1023\}$

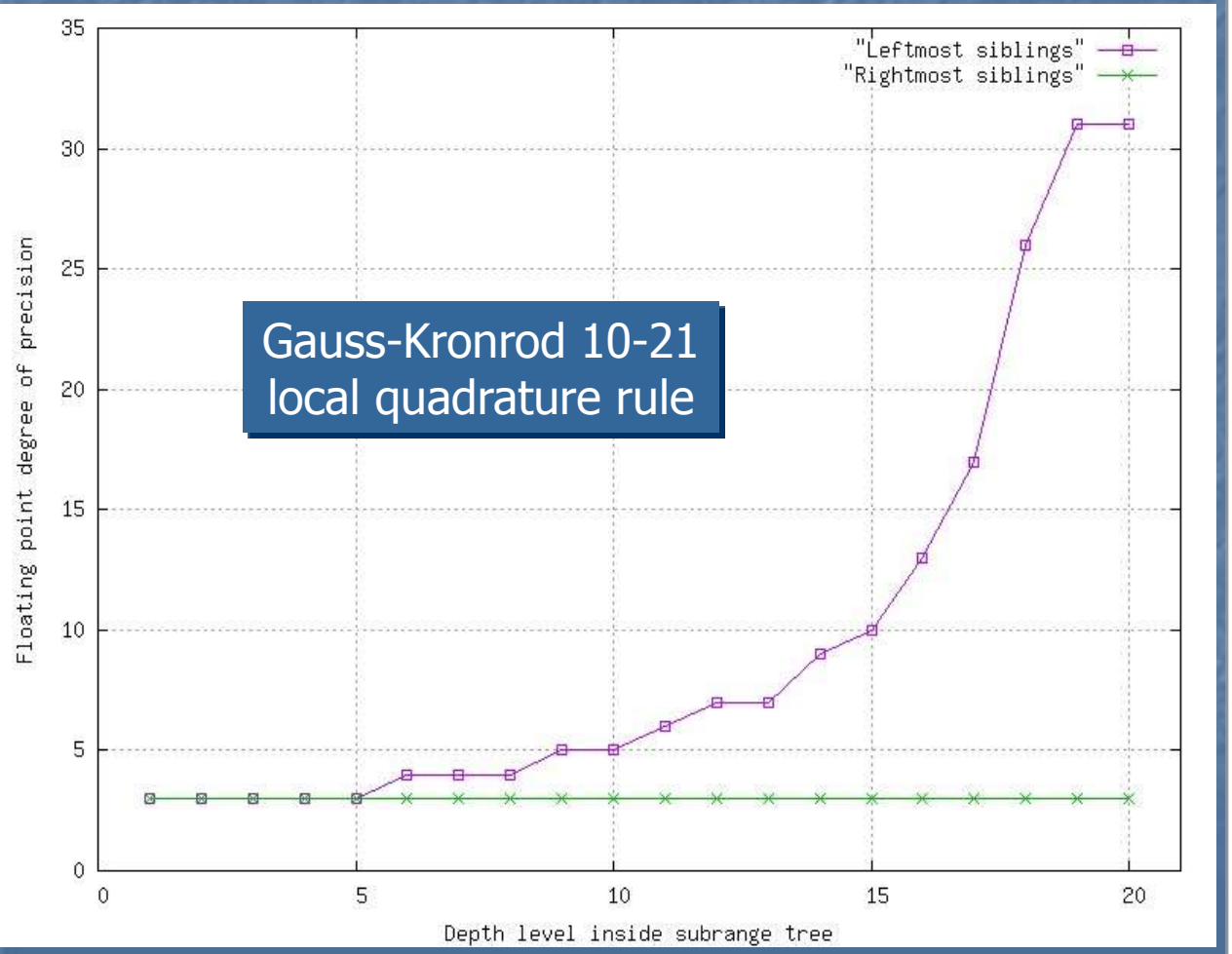


Variation of the floating point degree of precision of the GK 10-21 local quadrature rule over the inflating range* [0, j] *versus its width j. The plot of the computed values of d_{fp} points to a behaviour of d_{fp} which is similar to that reported in the previous case $d_{fp} = d = 31$ at $j = 1, 2, 3$; ***abrupt*** and then ***milder decreasing*** rate down to $d_{fp} = 5$ at $702 \leq j \leq 1023$.

Features of the Floating Point Degree of Precision

- *Non-equivalence of the siblings in the binary subrange tree.*

Case study of the root domain $[0, 2^{20}]$



A binary subrange tree is built up to n-th depth level by bisection of the parent ranges. ***Comparison of the dependencies of the floating point degrees of precision of the GK 10-21 local quadrature rule on the depth level in the binary subrange tree generated by the root range*** $[0, 2^n]$***, for the leftmost and the rightmost siblings are plotted for*** $n=20$***.***

While the floating point degree of precision of the rightmost siblings in the binary subrange tree ***keeps the minimal*** d_{fp} value of the root range $[0, 2^n]$, the values of the floating point degree of precision of GK 10-21 for the leftmost siblings ***increases*** from the initial minimal d_{fp} value up to the maximally possible value $d_{fp}=d=31$ at the last depth levels.

Overview

- Data sets subject to analysis
- Well-conditioning Ansatz
- Quantitative well-conditioning criteria
- Discussion

Quantitative well-conditioning criteria (*Preliminaries*)

- ***Monotonicity interval subject to analysis:*** $x_0 < x_1 < \cdots < x_{\Lambda+1}$.
- *Definition 3. The above sequence is specified as follows:*
  - ***closed****, if both* x_0 *and* $x_{\Lambda+1}$ *are extrema of* $f(x)$
  - ***open****, if neither* x_0 *nor* $x_{\Lambda+1}$ *are extrema of* $f(x)$
  - ***closed to the left****, if* x_0 *is an extremum, while* $x_{\Lambda+1}$ *is not;*
  - ***closed to the right****, if* x_0 *is an extremum, while* $x_{\Lambda+1}$ *is not.*
- *Definition 4. The ends of the monotonicity interval are* ***isolated*** *from each other provided we can define* ***two second order divided differences****,* $d_{\gamma-2,\gamma-1,\gamma}$ *and* $d_{\delta,\delta+1,\delta+2}$, *such that* $\delta \geq \gamma$.
- *Corollary.* The *Definition* 4 results in the condition $\Lambda \geq 3$ i.e., that the monotonicity interval extends over ***five consecutive abscissas at least***.

Consistency criterion #1
(*local*, over a *triplet* of abscissas)

- *Let* $\{(x_\lambda, f_\lambda),\ \lambda = k-1, k, k+1\}$ *denote a piece of integrand profile spanned over the open sequence* $\{x_{k-1} < x_k < x_{k+1}\}$.
- *The following conditions are consistent with the integrand well-conditioning inside* (x_{k-1}, x_{k+1}).

(*i*1) *If* $|d_{k-1,k}| < |d_{k,k+1}|$ *and* $x_{k+1} - x_{k-1} < \nu_B$, *then* $1 < |d_{k,k+1}| / |d_{k-1,k}| < 2 + \rho_k$.

(*i*2) *If* $|d_{k-1,k}| < |d_{k,k+1}|$ *and* $x_{k+1} - x_{k-1} \geq \nu_B$,

then $1 < |d_{k,k+1}| / |d_{k-1,k}| < 3 + 3\rho_k + \rho_k^2$.

(*ii*1) *If* $|d_{k-1,k}| > |d_{k,k+1}|$ *and* $x_{k+1} - x_{k-1} < \nu_B$ *then* $1 < |d_{k-1,k}| / |d_{k,k+1}| < 2 + \rho_k^{-1}$

(*ii*2) *If* $|d_{k-1,k}| > |d_{k,k+1}|$ *and* $x_{k+1} - x_{k-1} \geq \nu_B$,

then $1 < |d_{k-1,k}| / |d_{k,k+1}| < 3 + 3\rho_k^{-1} + \rho_k^{-2}$.

In these equations, $\rho_k = (x_{k+1} - x_k)/(x_k - x_{k-1})$.

Consistency criterion #2 (*global*, over a *closed* sequence)

- *Let* $x_0 < x_1 < \cdots < x_{\Lambda+1}$ *denote a closed sequence defining a piece of integrand profile* $\{(x_\lambda, f_\lambda),\ \lambda = 0,1,\cdots,\Lambda+1\}$ *and let* $\Pi_{0,\Lambda+1} = d_{0,1,2} \cdot d_{\Lambda-1,\Lambda,\Lambda+1}$.

(*i*) *The condition* $\Pi_{0,\Lambda+1} < 0$ *is consistent with the integrand well-conditioning both in the right lateral neighborhood of* x_0 *and the left lateral neighborhood of* $x_{\Lambda+1}$.

(*ii*) *Otherwise, the clarification of the integrand conditioning inside the two above-mentioned lateral neighborhoods asks for further analysis of the integrand behavior over finer samplings inside* (x_0, x_1) *and* $(x_\Lambda, x_{\Lambda+1})$ *respectively.*

Consistency criterion #3
(*local*, at terminal *closed* endpoints)

- *If* $x_0 < x_1 < \cdots < x_{\Lambda+1}$ *is closed to the left and/or to the right, then the following consistency criteria hold.*
- *If it is closed to the left, let* $\Pi_{0,1} = d_{-1,0,1} \cdot d_{0,1,2}$.

(*i*1) *The condition* $\Pi_{0,1} > 0$ *is consistent with the integrand well-conditioning inside the right lateral neighborhood of* x_0.

(*i*2) *Otherwise, further integrand analysis over a finer sampling inside* (x_0, x_1) *is needed.*

- *If it is closed to the right, let* $\Pi_{\Lambda,\Lambda+1} = d_{\Lambda-1,\Lambda,\Lambda+1} \cdot d_{\Lambda,\Lambda+1,\Lambda+2}$.

(*ii*1) *The condition* $\Pi_{\Lambda,\Lambda+1} > 0$ *is consistent with the integrand well-conditioning inside the left lateral neighborhood of* $x_{\Lambda+1}$.

(*ii*2) *Otherwise, further integrand analysis over a finer sampling inside* $(x_\Lambda, x_{\Lambda+1})$ *is needed.*

Overview

- Data sets subject to analysis
- Well-conditioning Ansatz
- Quantitative well-conditioning criteria
- Discussion

Quantities involving differences

- ***First order divided differences:***

The first order divided difference of the integrand $f(x)$, *over the pair* $\{(x_\lambda, f_\lambda),\ \lambda = k-1, k\}$, *is given by*

$$d_{k-1,k} = \delta_{k-1,k} / (x_k - x_{k-1}), \quad \text{where} \quad \delta_{k\text{-}1,k} = f_k - f_{k-1},$$

and it approximates the integrand slope over (x_{k-1}, x_k).

- ***Second order divided differences:***

The second order divided difference of the integrand $f(x)$, *over the triplet* $\{(x_\lambda, f_\lambda),\ \lambda = k-1, k, k+1\}$, *is given by*

$$d_{k-1,k,k+1} = (d_{k,k+1} - d_{k-1,k}) / (x_{k+1} - x_{k-1}),$$

and it approximates the integrand curvature over (x_{k-1}, x_{k+1}).

- ***Interval ratios:*** $\rho_k = (x_{k+1} - x_k) / (x_k - x_{k-1})$.

Redefinitions for accurate computation of differences

- *Let* $x \in [\alpha,\beta] = [\alpha,\gamma] \cup [\gamma,\beta],\ \gamma = (\beta+\alpha)/2,\ h = (\beta-\alpha)/2.$

If $x \in [\alpha,\gamma]$, *then* $x = \alpha + h\xi$; *If* $x \in [\gamma,\beta]$, *then* $x = \beta - h\xi$; $\xi \in [0,1]$.

- ***Redefined first order divided differences:***

Given the pair $\{(x_\lambda, f_\lambda),\ \lambda = k-1,k\}$, *then*

$$\bar{d}_{k-1,k} = \delta_{k-1,k}/(\xi_k - \xi_{k-1}),\ \text{where}\ \delta_{k\text{-}1,k} = f_k - f_{k-1}.$$

- ***Redefined second order divided differences & interval ratios:***

If $\{(x_\lambda, f_\lambda),\ \lambda = k-1,k,k+1\}$, *refer to a* ***same*** *subrange,*

$$\bar{d}_{k-1,k,k+1} = (\bar{d}_{k,k+1} - \bar{d}_{k-1,k})/(\xi_{k+1} - \xi_{k-1}),$$

$$\rho_k = (\xi_{k+1} - \xi_k)/(\xi_k - \xi_{k-1}).$$

If $x_k \equiv x^i$, *a discretization abscissa of the partition* Π_N, *then* $\rho_k = \rho$,

(***i***) <u>*inside*</u> (x_k, x_{k+1}): $\bar{d}_{k-1,k,k+1} = (\bar{d}_{k,k+1} - \rho\bar{d}_{k-1,k})/[\xi_1(1-\rho)]$,

(***ii***) <u>*inside*</u> (x_{k-1}, x_k): $\bar{d}_{k-1,k,k+1} = (\rho^{-1}\bar{d}_{k,k+1} - \bar{d}_{k-1,k})/[\xi_1(\rho^{-1}-1)]$,

where $\rho = h_r/h_l,\ h_r = (x^{i+1} - x^i)/2,\ h_l = (x^i - x^{i-1})/2.$

Discussion

- *The reported Bayesian well-conditioning criteria over monotonicity intervals of the integrand put on firm ground and complete similar empirical previously defined criteria.*
- *We have shown that accurate computation of the involved divided differences can be done, provided we redefine them in terms of standard modified reduced abscissas. This result allows numerical definition to machine accuracy of the isolated integrand discontinuity points.*
- *The quantitative criteria look different for the densely sampled, respectively sparsely sampled regions of quadrature knots for local quadrature rules spanned by orthogonal polynomials.*
- *Numerical estimates of the derived bounds are in good agreement with the previously derived empirical values.*
- *The set of reported Bayesian criteria exhausts all the cases concerning integrand conditioning over monotonicity intervals.*

Thank you for your attention !