Évaluation de $\mathrm{Ai}(x)$

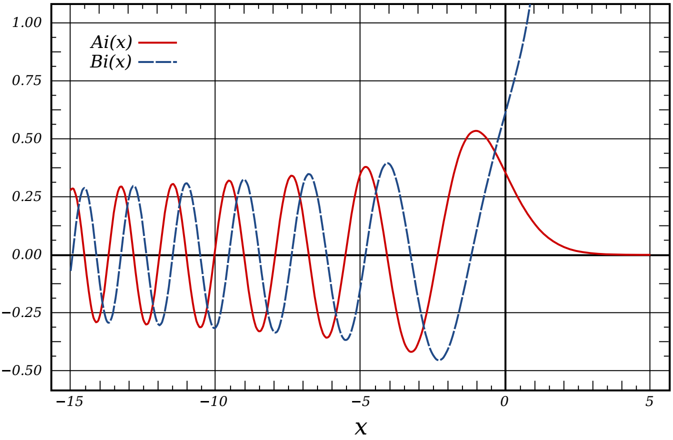
Cancellation catastrophique & comment y échapper

Marc Mezzarobba
projet AriC, Inria, ENS de Lyon

Sylvain Chevillard
projet Apics, Inria Sophia

Séminaire BiPoP-CASYS (LJK, Montbonnot), 5 avril 2013

The Airy Function $\mathrm{Ai}(x)$



$$\mathbf{Ai''}(x) = x\,\mathbf{Ai}(x) \qquad \mathrm{Ai}(0) = \frac{1}{3^{2/3}\,\Gamma(2/3)} \qquad \mathrm{Ai}'(0) = -\frac{1}{3^{1/3}\,\Gamma(1/3)}$$

Multiple-Precision Evaluation for $x > 0$

Standard Approach

"Small" x: Taylor Series at 0

- catastrophic cancellation for moderately large x
- need $p_{\text{work}} \gg p_{\text{res}}$

for $n = 0, 1, \ldots, N-1$

$$t_n := a_1(n) \cdot t_{n-1} \cdot x + a_2(n) \cdot t_{n-1} x^2 + \cdots + a_k(n) \cdot t_{n-k} \cdot x^k$$

$$s := s + t_n$$

(floating-point, precision p_{work})

"Large" x: Asymptotic Expansion at ∞

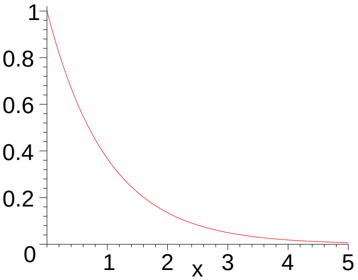
This talk

New evaluation algorithm for "small" x with $p_{\text{work}} \approx p_{\text{res}}$

Complete error analysis

Cancellation

A Simple Example



$$\exp(-x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} x^n$$

$$x = 20$$

```
> x := 20: N := 100:

> add((-20.)^n/n!, n=0..99);
```

$$-.12115250e-1$$

```
> exp(-20.);
```

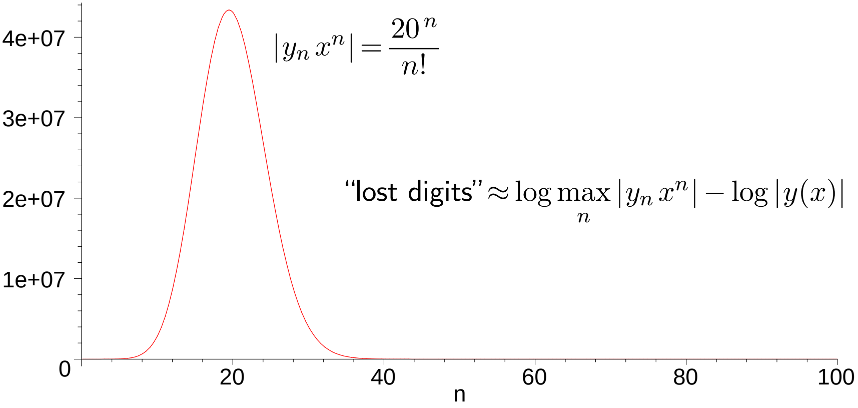
$$.2061153622e-8$$

```
> Digits := 30;
  add((-20.)^n/n!, n=0..99);
```

$$\mathbf{Digits := 30}$$

$$.206115362243865948417e-8$$

Catastrophic Cancellation



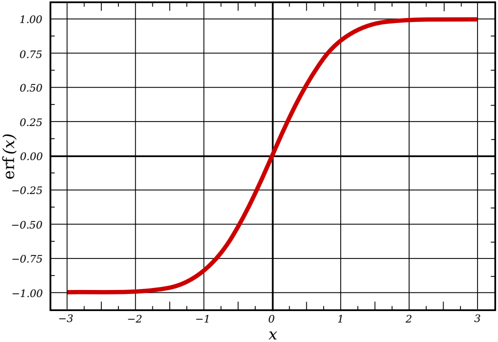
$$|y_n\, x^n| = \frac{20^{\,n}}{n!}$$

"lost digits" $\approx \log \max_n |y_n\, x^n| - \log |y(x)|$

A Better Way

$$\exp\left(-x\right) = \frac{1}{\exp\left(x\right)}$$

The Error Function



$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \left(x - \frac{1}{3} x^3 + \frac{1}{10} x^5 - \frac{1}{42} x^7 + \frac{1}{216} x^9 - \cdots \right)$$

catastrophic cancellation

But...

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \exp\left(-x^2\right) \underbrace{\sum_{n=0}^{\infty} \frac{2^n}{1 \cdot 3 \cdots (2\,n+1)}\, x^{2n+1}}_{G(x)}$$

(Abramowitz & Stegun, Eq. 7.1.6)

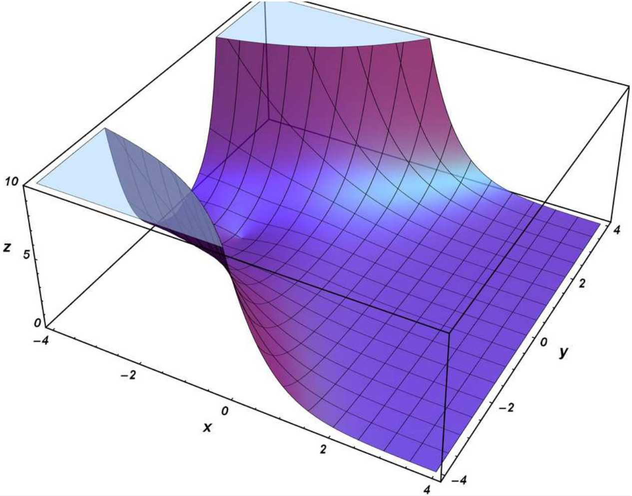
Algorithm

1. Compute $\frac{2}{\sqrt{\pi}} \boldsymbol{G(x)}$

   positive terms, minimal cancellation

2. Compute $\mathbf{exp}\,(\boldsymbol{x^2})$
3. Divide

Back to Ai



$$\mathrm{Ai}(x) = A - B\,x + \frac{A}{6}\,x^3 - \frac{B}{12}\,x^4 + \frac{A}{180}\,x^6 - \frac{B}{504}\,x^7 + \frac{A}{12960}x^9 - \cdots$$

$$= A \sum_{n=0}^{\infty} \frac{1 \cdot 4 \cdots (3\,n-2)}{(3\,n)!}\,x^{3n} - B \sum_{n=0}^{\infty} \frac{2 \cdot 5 \cdots (3\,n-1)}{(3\,n+1)!}\,x^{3n+1}$$

The GMR Method

The Gawronski-Müller-Reinhard Cancellation Reduction Method

Idea: **Find F and G** such that

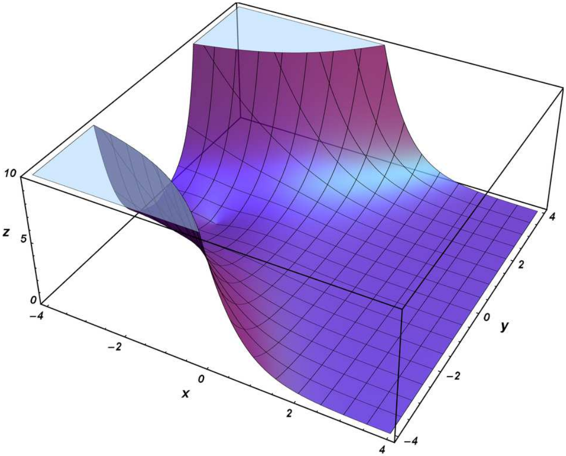
1. $y(x) = \dfrac{G(x)}{F(x)}$
2. F and G computable with little cancellation

- Based on complex analysis
- Starting point: asymptotic behaviour of y at complex ∞

W. Gawronski, J. Müller, M. Reinhard. SIAM J. Num. An., 2007.

M. Reinhard. Phd thesis, Universität Trier, 2008.

Asymptotics



$$\mathrm{Ai}(z) \sim \frac{\exp\left(-\frac{2}{3} z^{3/2}\right)}{2\sqrt{\pi}\, z^{1/4}}$$

as $z \to \infty$

in any sector

$\{z \in \mathbb{C} | -\varphi < \arg z < \varphi\}$

with $\varphi > 0$

The Indicator of an Entire Function

$|y(r\,e^{i\theta})| \approx \exp\left(\boldsymbol{h}(\theta)\,r^{\boldsymbol{\rho}}\right)$
for large r

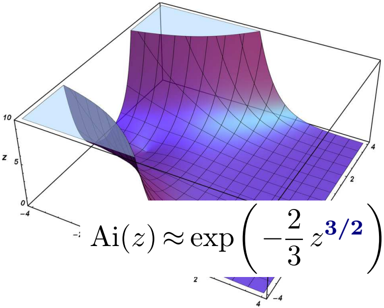
$$M(r) = \sup_{|z|=r} |y(z)|$$

Order

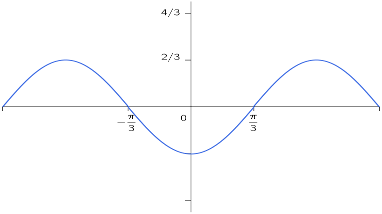
$$\rho = \limsup_{r\to+\infty} \frac{\ln\ln M(r)}{\ln r} = \frac{3}{2}$$

Indicator

$$h(\theta) = \limsup_{r\to+\infty} \frac{\ln|y(r\,e^{i\theta})|}{r^{\rho}} = -\frac{2}{3}\cos\left(\frac{3}{2}\,\theta\right)$$



$$\mathrm{Ai}(z) \approx \exp\left(-\frac{2}{3}\,z^{\mathbf{3/2}}\right)$$



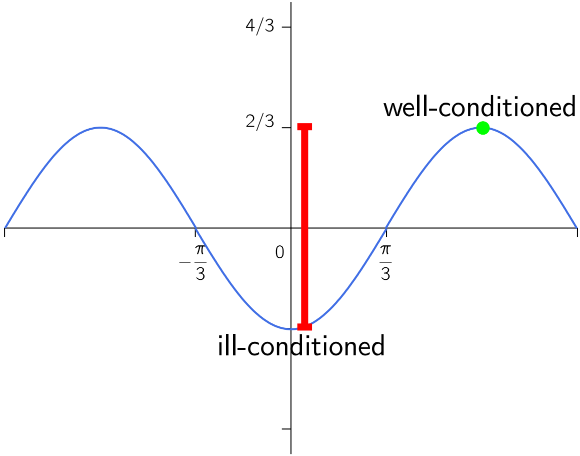
Lost in Cancellation

$|y(r\,e^{i\theta})| \approx \exp\left(\boldsymbol{h}(\theta)\, r^{\boldsymbol{\rho}}\right)$
for large r

$\max_n |y_n\, z^n| = M(|z|)^{1+o(1)}$

$$\begin{aligned}
\textbf{“lost” digits} &\approx \log_{10}\Big(\max_n |y_n\, z^n|\Big) - \log_{10}|y(z)| \\
&\approx \log_{10}\frac{M(|z|)}{|y(z)|} \\
&\approx \ln\frac{M(|z|)}{|y(z)|} \\
&\approx \big(r^\rho \max_\varphi h(\varphi)\big) - r^\rho\, h(\theta) \qquad (z = r\,e^{i\theta}) \\
&= r^\rho\,(\boldsymbol{\max h - h(\theta)})
\end{aligned}$$

Lost Digits



The GMR Method

- "lost" digits $\approx r^{\rho}\,(\max h - h(\theta))$

- same $\rho \qquad \Rightarrow \qquad h_{G/F} = h_G - h_F$

$$\begin{cases} F(z) \approx e^{h_F(\theta) r^{\rho}} \\ G(z) \approx e^{h_G(\theta) r^{\rho}} \end{cases} \quad \Rightarrow \quad \frac{G(z)}{F(z)} \approx \exp\left[(h_G(\theta) - h_F(\theta))\, r^{\rho}\right]$$

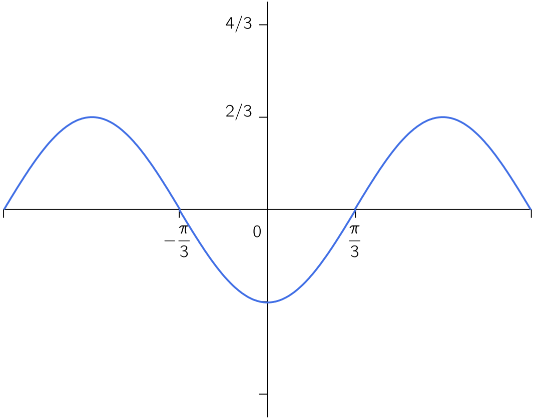
Idea (refined): look for

- an auxiliary series F,
- a modified series $G = y\,F$,

both of order ρ, such that h_F and $h_G \approx$ their max for $\theta = 0$

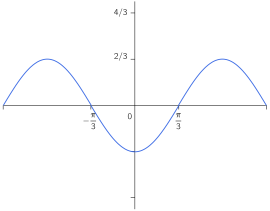
Auxiliary Series for $\mathrm{Ai}(x)$

A First Try

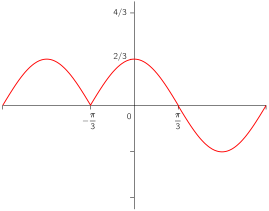


$$\mathrm{Ai}(x) = \frac{G(x)}{\exp\left(\alpha\, x^{\mathbf{3/2}}\right)}?$$

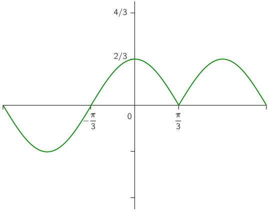
Indicators



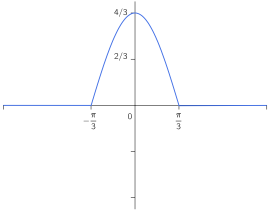
$\mathrm{Ai}(x)$



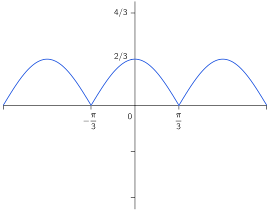
$\mathrm{Ai}(j^{-1}\,x)$



$\mathrm{Ai}(j\,x)$

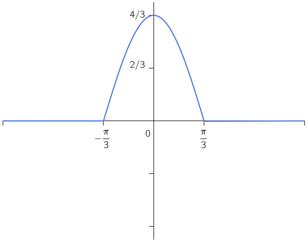


$\mathrm{Ai}(j\,x)\,\mathrm{Ai}(j^{-1}\,x)$

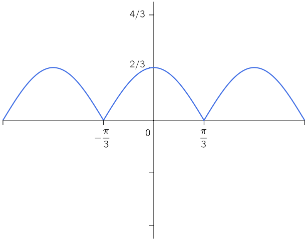


$\mathrm{Ai}(x)\,\mathrm{Ai}(j\,x)\,\mathrm{Ai}(j^{-1}\,x)$

Auxiliary & Modified Series



$$F(x) = \mathrm{Ai}(j\,x)\,\mathrm{Ai}(j^{-1}\,x)$$
$$= \frac{1}{4}\left(\mathrm{Ai}(x)^2 + \mathrm{Bi}(x)^2\right)$$



$$G(x) = \mathrm{Ai}(x)\,F(x)$$

D-Finiteness

A function y is **D-finite** (holonomic) when it satisfies a linear ODE with polynomial coefficients.

Examples: $\mathbf{Ai}(\boldsymbol{x}), \exp(x), \operatorname{erf}(x)$... $\quad \operatorname{Ai}''(x) = x\operatorname{Ai}(x)$

If $f(x)$, $g(x)$ are D-finite functions, then:

- For any algebraic function a, the composition $f(a(x))$ is D-finite
  $y(x) = \operatorname{Ai}(j\,x) \quad y''(x) = x\,y(x)$
- The sum $f(x) + g(x)$, the product $f(x) \cdot g(x)$ are D-finite
  $F(x) = y(x) \cdot \operatorname{Ai}(j^{-1}x) \quad F'''(x) = 4\,x\,F'(x) + 2\,F(x)$

D-Finiteness and Recurrences

$$F'''(x) = 4\,x\,F'(x) + 2\,F(x)$$

If $f(x)$ is a D-finite function, then:

- The Taylor coefficients of $f(x)$ obey a linear recurrence relation with polynomial coefficients

$$F(x) = \sum_{n=0}^{\infty} F_n\,x^n \qquad \boldsymbol{F_{n+3}} = \frac{2\,(2\,n+1)}{(n+1)\,(n+2)\,(n+3)}\,\boldsymbol{F_n}$$

The Auxiliary Series $F(x)$

$$F(x) = \mathrm{Ai}(j\,x)\,\mathrm{Ai}(j^{-1}\,x) = \sum_{n=0}^{\infty} F_n\,x^n$$

D-Finiteness

$$\mathrm{Ai}''(x) - x\,\mathrm{Ai}(x) = 0 \quad \rightsquigarrow \quad \boldsymbol{F_{n+3}} = \frac{2\,(2\,n+1)}{(n+1)\,(n+2)\,(n+3)}\,\boldsymbol{F_n}$$

$$\mathrm{Ai}(0) = A \quad \mathrm{Ai}'(0) = B \qquad F_0 = \frac{1}{3^{4/3}\,\Gamma\left(\frac{2}{3}\right)^2} \qquad F_1 = \frac{1}{2\sqrt{3}\,\pi}$$

$$F_2 = \frac{1}{3^{2/3}\,\Gamma\left(\frac{1}{3}\right)^2}$$

- Two-term recurrence $\Rightarrow$ Easy to evaluate
- Obviously $F_n > 0$ $\Rightarrow$ Minimal cancellation

The Modified Series $G(x)$

$$G(x) = \mathrm{Ai}(x)\,F(x) = \sum_{n=0}^{\infty} G_n\,\boldsymbol{x^{3n}}$$

D-Finiteness

$$\boldsymbol{G_{n+2}} = \frac{10\,(n+1)^2\,\boldsymbol{G_{n+1}} - \boldsymbol{G_n}}{(n+1)\,(n+2)\,(3\,n+4)\,(3\,n+5)}$$

$$G_0 = \frac{1}{9\,\Gamma\left(\frac{2}{3}\right)^3}$$

$$G_1 = \frac{1}{18\,\Gamma\left(\frac{2}{3}\right)^3} - \frac{1}{3\,\Gamma\left(\frac{1}{3}\right)^3}$$

$$G(x) = 0.44749\cdot 10^{-1} + 0.50371\cdot 10^{-2}\,x^3 + .14053\cdot 10^{-3}\,x^6 + .17388\,10^{-5}\,x^9 + .12091\cdot 10^{-7}\,x^{12} + .53787\cdot 10^{-10}\,x^{15} + \cdots$$

Observe that $\boldsymbol{G_n > 0}$ (proof?)

Minimality

Are We Done Yet?

G_n is one of the solutions of

$$\boldsymbol{u_{n+2}} = \frac{10\,(n+1)^2\,\boldsymbol{u_{n+1}} - \boldsymbol{u_n}}{(n+1)\,(n+2)\,(3\,n+4)\,(3\,n+5)}$$

Perron-Kreuser Theorem

$$u_n = \frac{v_n}{n!^2} \qquad \frac{v_{n+1}}{v_n} \to \begin{cases} \text{either } 1 & \text{dominant solution (generic case)} \\ \text{or } \ 1/9 & \text{minimal solution (non-generic)} \end{cases}$$

Experimentally $G_n \approx \dfrac{1}{9^n\,n!^2}$ (minimal) (proof?)

$\Rightarrow$ numerically **unstable** recursion

Miller's Method

Idea

"Unroll" the recurrence backwards for stability

...starting from arbitrary "initial" values

Algorithm

Choose $N \gg 0$

Set $u_N = 1$, $u_{N+1} = 0$

Compute $u_{N-1}, \ldots, u_1, u_0$

using the recurrence

Return the list of $\tilde{G}_n^{(N)} = \frac{G_0}{u_0} u_n$

$$
\uparrow \quad \begin{array}{lcl}
u_0 = 5.045\,10^{22} & \rightarrow & G_0 = 4.475\,10^{-2} \\
u_1 = 5.679\,10^{21} & & G_1 = 5.039\,10^{-3} \\
u_2 = 1.584\,10^{20} & & G_2 = 1.405\,10^{-4} \\
u_3 = 1.960\,10^{18} & & G_3 = 1.739\,10^{-5} \\
u_4 = 1.363\,10^{16} & & G_4 = 1.209\,10^{-8} \\
u_5 = 6.064\,10^{13} & \rightarrow & G_5 = 5.379\,10^{-11} \\
u_6 = 1.873\,10^{11} & & G_6 = 1.661\,10^{-13} \\
u_7 = 4.248\,10^{8} & & G_7 = 3.768\,10^{-16} \\
u_8 = 7.369\,10^{5} & & G_8 = 6.538\,10^{-19} \\
u_9 = 1000. & & G_9 = 8.869\,10^{-22} \\
\boldsymbol{u_{10} = 1.} & \rightarrow & G_{10} = 8.869\,10^{-25} \\
\boldsymbol{u_{11} = 0.} & & G_{11} = 0
\end{array}
$$

Convergence of Miller's Method

Algorithm

Choose $N \gg 0$

Set $\boldsymbol{u_N = 1,\ u_{N+1} = 0}$ $\leftarrow$ same starting values for all N

Compute $u_{N-1}, \ldots, u_1, u_0$
(using the recurrence)

Return the list of $\tilde{G}_n^{(N)} = \frac{G_0}{u_0} u_n$

Theorem (classical)

For fixed n, we have $\boldsymbol{\tilde{G}_n^{(N)} \to G_n}$ as $\boldsymbol{N \to \infty}$

Evaluation Algorithm

Complete Algorithm

1. Choose working precision, series truncation orders
2. Compute $F(x)$ by direct recurrence
3. Compute $G(x)$ using Miller's method
4. Divide

Works well in practice. (proof?)

Proofs & Error Bounds

What Remains To Do

- Prove that (G_n) is a minimal solution

  i.e., the one to which Miller's method converges

- Prove that $G_n \geqslant 0$

  so that the summation is numerically stable

- Bound the tails of the series F and G [easy]
- Bound the roundoff errors in $\sum F_n x^n$ [tedious but routine]
- Bound the method error of Miller's algorithm (i.e., $|G_n - \tilde{G}_n^{(N)}|$)

  $\rightsquigarrow$ Main issue: **need bounds on G_n**

- Bound the corresponding additional roundoff errors [M&vdS 1976]

R.M.M. Matthiej & A. van der Sluis, Numerische Mathematik, 1976

Controlling G_n

Proposition

$G_n \sim \gamma_n = \frac{1}{4\sqrt{3}\,\pi\,9^n\,n!^2}$ with $\left|\frac{G_n}{\gamma_n} - 1\right| \leqslant 2.4\,n^{-1/4}$ for all $n \geqslant 1$

Corollary: $G_n > 0$ (for large n, then for all n)

Idea of the proof

- $G_n = \frac{1}{2\,\pi\,i} \oint \frac{G(z)}{z^{3n+1}}\,\mathrm{d}z$
- saddle-point method
- $\mathrm{Ai}(z) \sim \frac{e^{-\frac{2}{3}z^{3/2}}}{2\sqrt{\pi}\,z^{1/4}} =: \widetilde{\mathrm{Ai}}(z)$, $\left|\frac{\mathrm{Ai}(z)}{\widetilde{\mathrm{Ai}}(z)} - 1\right| \leqslant r^{-3/2}\,\frac{5}{48}\cos\frac{\theta}{2}$

Conclusion

Summary

- New well-conditioned formula for $\mathrm{Ai}(x)$,
  obtained by an extension of the GMR method
- Detailed example of how to make the method rigorous
- Ready-to-use multiple-precision algorithm for $\mathrm{Ai}(x)$

Next question: How much of this is specific to $\mathrm{Ai}(x)$?

- Entire function
- Ability to find auxiliary series
- D-finiteness [constraints on the order of the recurrences?]
- Asymptotic estimate with error bound

Credits & Public Domain Dedication

This document uses

- the following images from Wikimedia Commons, all by **User:Inductiveload** and placed in the public domain
  - `http://commons.wikimedia.org/wiki/File:Airy_Functions.svg`
  - `http://commons.wikimedia.org/wiki/File:AiryAi_Arg_Contour.svg`
  - `http://commons.wikimedia.org/wiki/File:AiryAi_Abs_Surface.png`
  - `http://commons.wikimedia.org/wiki/File:Error_Function.svg`
- icons from the Oxygen icon set (`http://www.oxygen-icons.org/`), distributed under the Creative Commons Attribution-ShareAlike 3.0 license (`http://creativecommons.org/licenses/by-sa/3.0/`).