ANALYTIC VARIATIONS ON QUADTREES

Philippe Flajolet
INRIA, Rocquencourt (France)

√ Tree data structure
√ Multidimensional data in $\mathbf{R}^d$

♡ *Analytic approach:*
— Recurrences and generating functions,
— Complex asymptotics,
— Limit distributions

• PF + G. Gonnet (ETH), C. Puech (Grenoble), M. Robson (Canberra): *Algorithmica* **10** (1993).

• PF + M. Hoshi (Japan): *BIT* **32** (1992).

• PF + T. Lafforgue (Paris): *Discrete Comput. Geometry* **12** (1994).

• PF + B. Salvy (INRIA), L. Laforest & G. Labelle (Montréal): *Random Str. & Algorithms*, **7** (1995-6).

DIVIDE & CONQUER PARADIGM

To `Solve`(E):
— `Split` $E \mapsto E_1, E_2, \ldots, E_r$;
— for each j do `Solve`(E_j);
— `Weave` partial solutions.

$$C_n = t_n + r \sum_{k=0}^{n} \pi_{n,k} C_k.$$

Binary search trees: $\pi_{n,k} = 1/n$.

Digital trees: $\pi_{n,k} = 2^{-n}\binom{n}{k}$.

Quadtrees, $d = 2$: $\pi_{n,k} = [H_n - H_k]/n$,

$$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}.$$

Binary Search Tree

$x < \texttt{root} < y$

E=(jan,feb,mar,apr,may,jun,jul)

$$\pi_{n,k} = \frac{1}{n}$$

Binary Search Tree $\pi_{n,k} = 1/n$.

$$C_n = t_n + \frac{2}{n}\sum_{k=0}^{n-1} C_k$$

$$C(z) := \sum_n C_n z^n$$

$$C(z) = t(z) + \int_0^z C(u)\,\frac{du}{1-u}$$

$$C(z) = \frac{1}{(1-z)^2}\int_0^z \left(\frac{d}{dx}t(x)\right)(1-x)^2\,dx.$$

Theorem. Search cost = 38% loss

$$C_n \sim 2n\log n \doteq 1.38\log_2 n$$

QUADTREES: COST FUNCTIONS

.   & Probabilities

Let p_{n_1,n_2,n_3,n_4} *be the probability that the four root subtrees of a quadtree built on* $n = 1+n_1+n_2+n_3+n_4$ *points have sizes* n_1, n_2, n_3, n_4.

$$p_{n_1,n_2,n_3,n_4} = \frac{1}{n \cdot n!} \frac{(n_1+n_2)!\,(n_3+n_4)!\,(n_1+n_3)!\,(n_2+n_4)!}{n_1!\,n_2!\,n_3!\,n_4!}.$$

$$p_{n_1,n_2,n_3,n_4} = \binom{n-1}{n_1,n_2,n_3,n_4} \times \int_0^1 \int_0^1 [(uv)^{n_1}((1-u)v)^{n_2}(u(1-v))^{n_3}((1-u)(1-v))^{n_4}]\, du\, dv.$$

$$\int_0^1 x^\alpha (1-x)^\beta \, dx = \frac{\alpha!\,\beta!}{(\alpha+\beta+1)!},$$

Lemma. *Additive Cost Functions,*

$$C_n = t_n + 4 \sum_{k=0}^{n-1} \pi_{n,k} C_k,$$

$$\pi_{n,k} = \frac{1}{n} \left[H_n - H_k \right].$$

$$\begin{aligned}\varpi_{n,k,\ell} &= \binom{n-1}{k, \ell - k, n-1-\ell} \times \\ &\quad \int_0^1 \int_0^1 (uv)^k (u(1-v))^{\ell-k} (1-u)^{n-1-\ell} \, du \, dv \\ &= \frac{1}{n(\ell+1)}.\end{aligned}$$

$$C(z) = t(z) + 4 \int_0^z \frac{dt}{t(1-t)} \int_0^t C(u) \frac{du}{(1-u)}.$$

Differential equation for $C(z)$:

$$z(1-z)^2 \frac{d^2 f}{dz^2} + (1-2z)(1-z) \frac{df}{dz} - 4f = \delta t(z),$$

$$\delta t(z) = \frac{d}{dz} z(1-z) \frac{d}{dz} t(z).$$

HYPERGEOMETRICS, $_2F_1$

Hypergeometric function (Euler)

$$ {}_2F_1[a,b;c;z] := 1 + \frac{a \cdot b}{c}\,\frac{z}{1!} + \frac{a(a+1)\cdot b(b+1)}{c(c+1)}\,\frac{z^2}{2!} + \cdots $$

Hypergeometric equation

$$ z(1-z)\frac{d^2F}{dz^2} + (c+1-a-b)\frac{dF}{dz} - abF = 0 $$

3 regular singular points (Riemann), $0, 1, \infty$.

Additive cost, $C(z)$:

$$z(1-z)^2\frac{d^2f}{dz^2}+(1-2z)(1-z)\frac{df}{dz}-4f=\delta t(z),$$

$$\delta t(z)=\frac{d}{dz}z(1-z)\frac{d}{dz}t(z).$$

Try $\sim(1-z)^\beta$, *cf.* **singularity analysis**

Find $\beta=-2$ (indicial equation)

Get $\phi(z)=\frac{1+2z}{(1-z)^2}$ (special hypergeometric)

"Variation of constant"

$$f(z)=\frac{1+2z}{(1-z)^2}\int_0^z\frac{(1-u)^3}{u(1+2u)^2}\int_0^u\frac{(1+2v)}{(1-v)^2}\delta t(v)\,dv\,dt.$$

With $t_n=n$, $t(z)=z/(1-z)^2$, explicit form.

Theorem 1 *Mean search cost has closed form*

$$\left(1+\frac{1}{3n}\right)H_n-\frac{n+1}{6n}.$$

PAGING.

Stop when hitting cells with less than b points.

$$f(z) = \frac{1+2z}{(1-z)^2}\int_0^z \frac{(1-u)^3}{u(1+2u)^2}\int_0^u \frac{(1+2v)}{(1-v)^2}\delta t(v)\,dv\,dt.$$

$$\delta t(z) = \frac{d}{dz}z(1-z)\frac{d}{dz}t(z).$$

$t_n = [\![n > b]\!],\ t(z) = z^{b+1}(1-z)^{-1}.$

Singularity analysis. Coefficient is γ_b such that

$$f(z) \sim \frac{\gamma_b}{(1-z)^2} \qquad (z \to 1),$$

which requires definite integrals.

$$\text{Data (?)} \begin{cases} b=1 & 120 - 12\pi^2 \\ b=2 & 534 - 54\pi^2 \\ b=3 & 1422 - 144\pi^2 \\ b=4 & \frac{5923}{2} - 300\pi^2 \end{cases}$$

Take Generating double function w.r.t. b (!)

Use **computer algebra**

First integral has size $\approx$ **2,000**,
Second integral has size $\approx$ **6,000**.

$$\log(1-z), \log(1-yz), \log(1+2z), Li_2(z), Li_2(yz)$$

$$Li_2(z) = \sum_{n=1}^{\infty} \frac{z^n}{n^2}.$$

+ Transformation formulæ

Theorem 2 *Number of pages* $\sim \gamma_b n$,

$$\gamma_b = 18b^2 + 27b + 1 - 18b(b+1)^2 \left[\frac{\pi^2}{6} - \sum_{j=1}^{b} \frac{1}{j^2} \right]$$

Page filling is about 33%,

$$\sim \frac{1}{3} + \frac{4}{15b} + \frac{2}{25b^2} - \frac{54}{875b^3} + \cdots$$

Fragmentation: $b = 60$, $\widehat{b} = 20$
$\Longrightarrow$ Filling ratio ≈ 0.67, Access ≈ 1.42.

SINGULARITY ANALYSIS

Mapping between (singular) function scale and coefficient scale

$$[z^n]\frac{1}{(1-z)^\alpha} = \binom{n+\alpha-1}{n-1} \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}.$$

Cf. Darboux's method, Tauberian theorems.

1. *Basic function scale.* Asymptotic scale for **functions**

$$\mathcal{S} = \left\{ A(\frac{1}{1-z}) \ \mid \ A(u) = u^\beta (\log u)^\gamma (\log\log u)^\delta \right\}$$

gives asymptotic scale for **coefficients**

$$[z^n]A(\frac{1}{1-z}) \approx \frac{1}{\Gamma(\beta)}\frac{A(n)}{n}.$$

2. *Transfers.* For $\sigma(z)$ in $\mathcal{S}$

$$\begin{array}{ccccccc}
f(z) & = & O(\sigma(z)) & \Rightarrow & f_n & = & O(\sigma_n) \\
f(z) & = & o(\sigma(z)) & \Rightarrow & f_n & = & o(\sigma_n) \\
f(z) & \sim & \sigma(z) & \Rightarrow & f_n & \sim & \sigma_n \\
f(z) & \sim & \sum_j \sigma^{\langle j \rangle}(z) & \Rightarrow & f_n & \sim & \sum_j \sigma_n^{\langle j \rangle}
\end{array}$$

3. *Conditions.* The *Camembert Condition* (strong regularity!).

Proof technique. Cauchy's coefficient formula and Hankel contours.

$$f(z) = \frac{1}{(1-z)^\beta}\left(\frac{1}{z}\log(1-z)^{-1}\right)^{-1/2}.$$

Step 1. Hankel contour: $z = 1 - \frac{t}{n}$

$$f_n = \frac{1}{2i\pi}\int_{\mathcal{H}}\left(\frac{n}{t}\right)^\beta (\log n - \log t)^{-1/2}\,(..)$$

$$\left(1-\frac{t}{n}\right)^{-n-1}\,\frac{dt}{n}$$

Step 2. Approximation.

$$n.f_n \approx \left[n^\beta(\log n)^{-1/2}\right]\cdot\frac{1}{2i\pi}\int_{\mathcal{H}}\frac{e^t}{t^\beta}\,dt.$$

Integral gives $1/\Gamma(\beta)$!!!

PARTIAL MATCH

`Search(x)`: Compare to root;

- `Search`(NW),`Search`(SW) OR `Search`(NE),`Search`(SE)

Perfectly balanced partition $\mapsto \sqrt{n}$.

$$z(1-z)^2\frac{d^2}{dz^2}Q(z) + 2(1-z)^2\frac{d}{dz}Q(z) - 4Q(z) = \frac{2}{1-z}$$

Theorem 3 *The expected cost Q_n of a partial match query is*

$$1 + Q_n = \sum_{k=0}^{n} \binom{\alpha - 1 + n - k}{n-k}\binom{k-\alpha-1}{k}\binom{k-\alpha}{k}\frac{1}{k+1},$$

$$\alpha = \frac{\sqrt{17}-1}{2} \approx 1.56155$$

.

$$Q_n \sim \gamma n^{\alpha-1}, \qquad \gamma = \frac{1}{2}\frac{\Gamma(2\alpha)}{\Gamma(\alpha)^3}.$$

PERSPECTIVE: HOLONOMY

Holonomic functions. [Zeilberger] *Vector Space of* $\frac{\partial^{i+j+k}}{\partial x^i \partial y^j \partial z^k} F(x,y,z)$ *is finite dimensional.*

Holonomic sequences. Coeffs of holonomic functions $=$ P-recursive.

Contains rational, algebraic, exponential, logarithmic, hypergeometric functions.

Closure. by sums, products, convolutions, diagonals, integration, differentiation, Laplace transforms, algebraic substitution, etc.

Here: Univariate **Holonomic asymptotics**

- use classification of singularities (reg., irreg.)
- transfer to coefficients.

Holonomic elements:

$$\lambda \omega^n n^\alpha (\log n)^k \cdot \{e^{n^{r/s}}\}.$$

ω, α algebraic numbers, k, r, s integers.

GENERAL DIMENSION $d \geq 2$

Splitting probabilities.

$$C_n = t_n + 2^d \sum_{k=0}^{n-1} \pi_{n,k} C_k.$$

$$\begin{cases} \pi^{(1)}_{n,k} & = & \frac{1}{n} \\ \pi^{(2)}_{n,k} & = & \frac{1}{n}[H_n - H_k] \\ \pi^{(3)}_{n,k} & = & \frac{1}{2n}[(H_n - H_k)^2 + (H_n^{(2)} - H_k^{(2)}] \end{cases}$$

$$\mathbf{I}f(z) = \int_0^z f(u)\,\frac{du}{1-u}, \qquad \mathbf{I}f(z) = \int_0^z f(u)\,\frac{du}{u(1-u)}$$

$$C(z) = t(z) + 2^d \mathbf{J}^{d-1} \mathbf{I} C(z).$$

Theorem 4 *The expected cost $C_n^{(d)}$ of a fully specified search in a d-dimensional quadtree of size n satisfies*

$$C_n^{(d)} \sim \frac{2}{d} \log n.$$

$\diamondsuit$ Homogeneous equation, singular solutions at $z = 1$.

$\spadesuit$ Matrix "variation of constant" method

$\heartsuit$ + Singularity analysis.

Partial Match

Q_n = *expected cost in a quadtree with* $\frac{s}{d}$ *known coordinates*

$$Q_n = 1 + \frac{2^d}{n(n+1)} \sum_{\mathcal{L}} \left[\frac{1}{(\ell_1+2)\cdots(\ell_{s-1}+2)} \right] \cdot$$

$$\left[\frac{1}{(\ell_{s+1}+1)\cdots(\ell_{d-1}+1)} \right] Q_{\ell_d},$$

$$\mathcal{L} = \left\{ (\ell_1, \ell_2, \ldots, \ell_d) \mid n > \ell_1 \geq \ell_2 \geq \cdots \geq \ell_{d-1} \geq \ell_d \geq 0 \right\}.$$

Theorem 5 *Expected cost* $Q_n^{(s,d)}$ *of type* s/d *is*

$$Q_n^{(s,d)} \sim \gamma_{s,d} n^{1-s/d+\theta(s/d)}.$$

Algebraic number $\theta \in [0,1]$ *is*

$$(\theta+3-x)^x(\theta+2-x)^{1-x} - 2 = 0.$$

Propagate: (i) the structure of the system; (ii) the **reduced** singular structure.

$$\theta(\frac{1}{3}) = \sqrt[3]{\frac{107}{27} + \frac{2\sqrt{318}}{9}} + \frac{1}{9\sqrt[3]{\frac{107}{27} + \frac{2\sqrt{318}}{9}}} - 2 \approx 0.04952.$$

PROBABILITY DISTRIBUTIONS.

Univariate problem: $\mathcal{L}f(z) = 0$.

$$f(z) \approx (1-z)^{-\alpha}$$

where $\alpha:\ I(\alpha) = 0$, is an algebraic number.

Bivariate problem: $\mathcal{L}_u f(z, u) = 0$.

— a perturbation near $u = 1$.

$$f(z) \approx (1-z)^{-\alpha(u)}$$

where $\alpha(u)$ is an algebraic functions

Singularity perturbation analysis $\Rightarrow$ characteristic function + moment generating function.

+ continuity theorems, etc.

Random search:

$$\Phi(u, z) = 1 + 2^d u \mathbf{J}^{d-1} \mathbf{I} \Phi(u, z).$$

Theorem 6 *Mean and standard deviation of search costs:*

$$\mu_n \sim \frac{2}{d} \log n, \quad \sigma_n^2 \sim \frac{2}{d^2} \log n.$$

Distribution tends to

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-w^2/2}\, dw.$$

Central + Local Limit Theorem.

Exponential tails.

- cycles in permutations,
- components in functional graphs
- irreducible factors in polynomilas $GF(q)[x]$, etc
- many multivariate holonomic recurrences