

Faster IBP solving via RATRACER and tricks

Vitaly Magerya

Institute for Theoretical Physics (ITP),
Karlsruhe Institute of Technology (KIT)

Siegen,
February 14, 2023

IBP relations

An *IBP integral family* with L *loop momenta* l_i , and E *external momenta* p_i , is the set of Feynman integrals

$$I_{\nu_1, \nu_2, \dots, \nu_N} \equiv \int \frac{d^d l_1 \cdots d^d l_L}{D_1^{\nu_1} \cdots D_N^{\nu_N}}, \quad D_i \equiv (l_j \pm p_k \pm \dots)^2 - m_i^2 + i0,$$


where ν_i are the “indices”, the D_i are the “denominators”.

The idea: shifting l_k by *any vector* v should not change I :

$$\lim_{\alpha \rightarrow 0} \frac{\partial}{\partial \alpha} I(l_k \rightarrow l_k + \alpha v) = \int d^d l_1 \cdots d^d l_L \frac{\partial}{\partial l_k^\mu} \frac{v^\mu}{D_1^{\nu_1} \cdots D_N^{\nu_N}} = 0.$$

These are the *IBP relations*, valid for all k and all v , L ($L + E$) in total.

* * *

Example. For $I_{a,b,c} \equiv$ , if we choose $k = 1$ and $v = l_1$, we will get

$$(d - 2a - b - c)I_{a,b,c} - cI_{a-1,b,c+1} - bI_{a-1,b+1,c} = 0.$$

Laporta algorithm

To solve IBP relations in practice use the *Laporta algorithm*: [Laporta '00]

1. Substitute integer values for the indices ν_i into the IBP relations, obtaining a large linear system with many different $I_{\nu_1 \dots \nu_N}$.
2. Define an ordering on $I_{\nu_1 \dots \nu_N}$ from “simple” to “complex” integrals.
3. Perform Gaussian elimination on the linear system, eliminating the most “complex” integrals first.
4. A small number of “simple” integrals will remain uneliminated.
⇒ These are the *master integrals*. The rest will be expressed as their linear combinations.

Available software

IBP solvers not using modular arithmetic:

- * **LITERED** (useful Mathematica functions, required by FIRE). [Lee '13]
- * **FORCER** (for massless 2-point functions). [Ruijl, Ueda, Vermaseren '17]

IBP solvers that use modular arithmetic:

- * **FINRED** (a private implementation). [von Manteuffel et al]
- * **FIRE6**. [Smirnov, Chuharev '19]
 - * Does not provide multivariate reconstruction.
- * **KIRA** when used with **FIREFLY**.

[Klappert, Lange, Maierhöfer, Usovitsch '20; Klappert, Klein, Lange '20]

- * **FINITEFLOW** (a library for arbitrary computations). [Peraro '19]
- * **CARAVEL** (a library for amplitude computations). [Cordero, Sotnikov et al '20]

... and multiple others.

Now also introducing: **RATRACER** (with KIRA and FIREFLY). [V.M. '22]

Modular arithmetic methods

To find a symbolic form of a rational function $f(x_1, \dots, x_N)$:

- * *Evaluate* f modulo a prime number *many times*, with x_i set to integers.
- * *Reconstruct* the exact symbolic form of f from the obtained values.

Example: if we have an unknown $f(x)$, and we have evaluated

$$\begin{aligned} f(11) &= 139 \pmod{997}, & f(65) &= 479 \pmod{997}, \\ f(38) &= 350 \pmod{997}, & f(92) &= 115 \pmod{997}, \end{aligned}$$

then we can use *polynomial interpolation* to find a polynomial form of f :

$$f(x) = 618 + 979x + 486x^2 + 41x^3 \pmod{997},$$

and then *rational function reconstruction* to find an equivalent rational form:

$$f(x) = \frac{996 + 333x}{1 + x} \pmod{997},$$

and finally *rational number reconstruction* to find the rational coefficients:

$$f(x) = \frac{-1 + \frac{2}{3}x}{1 + x} \pmod{997}.$$

Guess that this is the true form of $f(x)$; evaluate more times to verify.

Layman's IBP performance checklist

To improve IBP performance:

1. Use modular arithmetic methods. [von Manteuffel, Schabinger '14; Peraro '16]
2. *Make the result smaller:*
 - 2.1 Reduce whole amplitudes (not individual integrals).
 - 2.2 Choose master integrals that minimize the result size.
 - * Use *d-factorizing bases* that ensure the factorization of d in the denominators of IBP coefficients. [Usovitsch '20; Smirnov, Smirnov '20]
 - * Consider *quasi-finite bases*. [von Manteuffel, Panzer, Schabinger '14]
 - * Consider *uniform transcendentality bases*, if possible. [Bendle et al '19]
 - 2.3 Construct a smaller ansatz for the result. [Abreu et al '19; De Laurentis, Page '22]
 - 2.4 Set some of the variables to fixed numbers.
 - * E.g. reduce with m_H^2/m_t^2 set to 12/23.
 - * Or perform IBP reduction separately for each phase-space point, and interpolate in between. [Jones, Kerner et al '18; Chen, Heinrich et al '19, '20]
3. Improve the *evaluation performance:*
 - 3.1 Combine IBP relations (using syzygies) to eliminate integrals with raised (or lowered) indices. [Gluza, Kajda, Kosower '10; Schabinger '11]
 - 3.2 *Just solve the equations faster?*

Optimizing the modular Gaussian elimination

When performing Gaussian elimination one needs to:

- * Represent the equation set as a sparse matrix data structure.
 - * Keep the equations sorted.
 - * Keep terms in each equation sorted.
 - * Adjust the layout (and maybe reallocate memory) after each operation.
 - * This is not much work, but so is modular arithmetic!

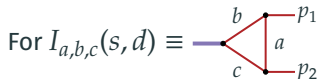
IBP solvers using modular arithmetics will:

- * Recreate the same data structures, same memory allocations, in the same order during each evaluation, many times.
 - * Only the modular values change between evaluations.
- * Spend relatively little time on actual modular arithmetic.
 - * Because it is so fast!

How to speed this up? Eliminate the data structure overhead:

- * *Record the list of arithmetic operations* performed during the first evaluation (“*a trace*”).
- * Simply *replay this list* for subsequent evaluations.

Rational traces



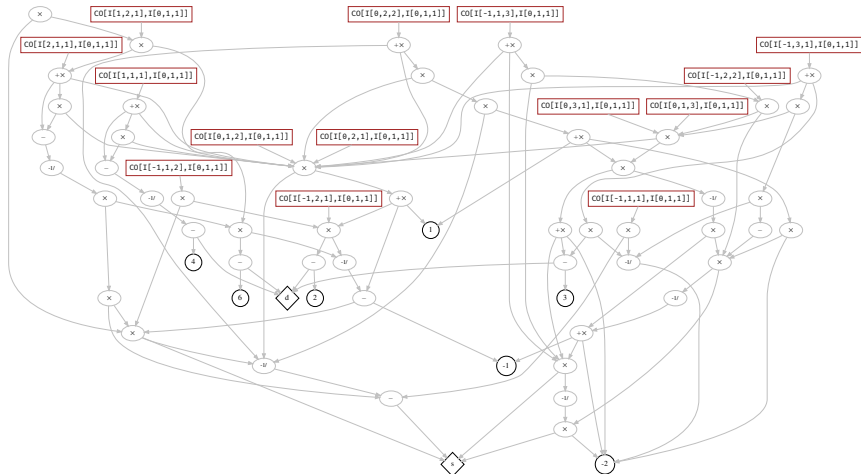
the *trace* of the IBP solution might look like:

```
t0 = var 'd'
t1 = int 4
t2 = sub t0 t1
t3 = int 1
t4 = var 's'
t5 = neg t4
t6 = int 6
t7 = sub t0 t6
t8 = int -1
t9 = int 2
t10 = int -2
t11 = sub t0 t9
t12 = int 3
t13 = sub t0 t12
t14 = mul t4 t10
t15 = neginv t5
t16 = mul t4 t15
t17 = sub t8 t16
t18 = mul t5 t16
t19 = neginv t17
t20 = mul t7 t19
[...]

t54 = addmul t53 t27 t44
t55 = mul t25 t44
t56 = addmul t55 t25 t44
t57 = mul t23 t44
t58 = addmul t57 t23 t44
t59 = mul t20 t58
t60 = mul t16 t59
save t60 as CO[I[1,1,2],I[0,1,1]]
save t59 as CO[I[1,2,1],I[0,1,1]]
save t58 as CO[I[2,1,1],I[0,1,1]]
save t56 as CO[I[1,1,1],I[0,1,1]]
save t54 as CO[I[-1,1,3],I[0,1,1]]
save t52 as CO[I[-1,2,2],I[0,1,1]]
save t51 as CO[I[-1,3,1],I[0,1,1]]
save t46 as CO[I[0,1,3],I[0,1,1]]
save t49 as CO[I[0,2,2],I[0,1,1]]
save t47 as CO[I[-1,1,2],I[0,1,1]]
save t46 as CO[I[0,3,1],I[0,1,1]]
save t42 as CO[I[-1,2,1],I[0,1,1]]
save t44 as CO[I[0,1,2],I[0,1,1]]
save t44 as CO[I[0,2,1],I[0,1,1]]
save t45 as CO[I[-1,1,1],I[0,1,1]]
```

Rational traces

For $I_{a,b,c}(s, d) \equiv \begin{array}{c} b \\ \diagdown \quad \diagup \\ \text{---} \quad \text{---} \\ \diagup \quad \diagdown \\ c \end{array} \begin{array}{c} p_1 \\ a \\ p_2 \end{array}$ the *trace* of the IBP solution might look like:



RATRACER overview

RATRACER (“Rational Tracer”): a program for *solving systems of linear equations* using modular arithmetic based on rational traces. [V.M. '22]

- * Can trace the solution of arbitrary systems of linear equations: IBP relations, dimensional recurrence relations, amplitude definitions, etc.
- * Can optimize and transform traces.
- * Uses FIREFLY for reconstruction. [Klappert, Klein, Lange '20, '19]
- * Initially created for solving IBPs for massive 5-point 2-loop diagrams.
- * Available at github.com/magv/ratracer.

Intended usage:

1. Use KIRA (or LITERED, or custom code) to export IBP relations.
2. Use RATRACER to load them and solve them.

Trace optimizations

Given a trace, RATRACER can optimize it using:

- * *Constant propagation:*

$$\begin{cases} t11 = \text{int } 2 \\ t12 = \text{int } 3 \\ t13 = \text{mul } t11 \ t12 \end{cases} \Rightarrow \begin{cases} t11 = \text{int } 2 \\ t12 = \text{int } 3 \\ t13 = \text{int } 6 \end{cases}$$

- * *Trivial operation simplification:*

$$\begin{cases} t11 = \text{int } -1 \\ t12 = \text{mul } t11 \ t7 \end{cases} \Rightarrow \begin{cases} t11 = \text{int } -1 \\ t12 = \text{neg } t7 \end{cases}$$

- * *Common subexpression elimination:*

$$\begin{cases} t11 = \text{add } t5 \ t7 \\ t12 = \text{add } t5 \ t7 \end{cases} \Rightarrow \begin{cases} t11 = \text{add } t5 \ t7 \\ t12 = t11 \end{cases}$$

- * *Dead code elimination:*

$$\begin{cases} t11 = \text{add } t5 \ t7 \\ [\dots, t11 \text{ is unused}] \end{cases} \Rightarrow \begin{cases} \text{nop} \\ [\dots] \end{cases}$$

- * Especially useful if a user wants to select a subset of the outputs.





- * *“Finalization”:*

$$\begin{cases} t11 = \text{add } t5 \ t6 \\ t12 = \text{add } t11 \ t7 \\ [\dots, t11 \text{ is unused}] \end{cases} \Rightarrow \begin{cases} t11 = \text{add } t5 \ t6 \\ t11 = \text{add } t11 \ t7 \\ [\dots] \end{cases}$$

- * Needed to minimize the temporary memory needed for the evaluation.

RATRACER benchmarks

For IBP reduction of every integral (i.e. not single amplitudes):

	Evaluation speedup vs. KIRA+FIREFLY	$\frac{t_{\text{reconstruction}}}{t_{\text{evaluation}}}$	Total speedup vs. KIRA+FIREFLY	Total speedup vs. KIRA+FERMAT	Total speedup vs. FIRE6
	20	3.3	5.2	1.2	$\infty?$
	7.8	1/3.3	6.0	37	$\infty?$
	26	25	1.7	1/3.3	1.8
	9.6	8.8	5.2	2.6	8.8

[github.com/magv/ibp-benchmark]

Resulting performance:

- * Consistent ~10x speedup in modular evaluation over KIRA+FIREFLY.
- * Up to ~5x *speedup in total reduction time* over KIRA+FIREFLY for complicated examples, 1x-30x over KIRA+FERMAT, ∞ x over FIRE6.

Problems with big traces

Problem:

- * The size of a trace is proportional to the number of operations.
 - * I.e. $\sim N_{\text{integrals}}^2$ for sparse IBP systems.
 - \Rightarrow Megabytes to gigabytes for problems of interest.
- * Computer memory is expensive and limited.

Solution:

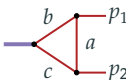
1. Always keep the trace on disk, never load it fully into memory.
 - * Compress it on disk for storage (via ZSTD, GZIP, BZIP2, or LZMA).
 2. During the evaluation read the trace piece by piece.
 3. During the optimization make sure the algorithms have bounded memory usage.
- \Rightarrow Multi-GB traces are supported easily in RATRACER.

Trace transformations

Given a trace, RATRACER can:

- * *Set some of the variables to expressions* or numbers.
 - * E.g. set mh_2 to “ $12/23*mt_2$ ”, d to “ $4-2*eps$ ”, or s to “13600”.
 - * No need to remake the IBP system just to set a variable to a number.
- * *Select any subset of the outputs*, and drop operations that don't contribute to them (via dead code elimination).
 - * Can be used to split the trace into parts.
 - * Each part can be reconstructed separately (e.g. on a different machine).
 - * See master-wise and sector-wise reduction in other solvers.
- * *Expand the result into a series* in any variable.
 - * By evaluating the trace while treating each value as a series, and saving the trace of that evaluation.
 - * Done before the reconstruction, so one less variable to reconstruct in, but potentially more expressions (depending on the truncation order).
 - * In practice only few leading orders in ε are needed, so expand in ε up to e.g. $\mathcal{O}(\varepsilon^0)$, and *don't waste time on reconstructing the higher orders*.

Truncated series expansion

For $I_{a,b,c}(s, d) \equiv$  before expansion:

- * Variables to reconstruct in: s and d .
- * Trace outputs: “CO [I [1, 1, 1], I [0, 1, 1]]”, etc:

$$I_{1,1,1} = \text{CO} [I [1, 1, 1], I [0, 1, 1]] I_{0,1,1}.$$





After expansion in ε to $\mathcal{O}(\varepsilon^0)$:

- * Variables to reconstruct in: only s .
- * Trace outputs: “ORDER [CO [I [1, 1, 1], I [0, 1, 1], eps⁻¹]]”, etc:

$$I_{1,1,1} = \text{ORDER} [\text{CO} [I [1, 1, 1], I [0, 1, 1], \text{eps}^{-1}] \varepsilon^{-1} I_{0,1,1} \\ + \text{ORDER} [\text{CO} [I [1, 1, 1], I [0, 1, 1], \text{eps}^0] \varepsilon^0 I_{0,1,1}.$$

- * Might be slower to evaluate, but fewer evaluations are needed.
⇒ The more complicated the problem, the higher the speedup.

RATRACER + series expansion benchmarks

	probe time speedup			total time speedup		
	$\mathcal{O}(\varepsilon^0)$	$\mathcal{O}(\varepsilon^1)$	$\mathcal{O}(\varepsilon^2)$	$\mathcal{O}(\varepsilon^0)$	$\mathcal{O}(\varepsilon^1)$	$\mathcal{O}(\varepsilon^2)$
	1/1.3	1/1.5	1/1.8	3.2	2.4	1.9
	1/2.0	1/2.5	1/3.0	2.7	1.4	1/1.3
	1/1.4	1/2.4	1/2.9	2.3	1.7	1.4
	1/1.0	1/1.6	1/2.1	4.3	2.3	1.6

[github.com/magv/ibp-benchmark]

Resulting performance:

- * A *~3x speedup* with ε expansion up to $\mathcal{O}(\varepsilon^0)$.
- * The higher the expansion, the less the benefit.

Guessing the denominators

The *denominators of IBP coefficients factorize* into few unique factors.

If some candidate factors are known, then we can find the powers of those factors in each coefficient: [Abreu et al '18; Heller, von Manteuffel '21]

1. Choose a factor to search for, e.g. $(d - 6)$.
2. Set all variables to random values, e.g. $d = 95988281$, $s = 75579811$.
 $\Rightarrow (d - 6) = 95988275 = 5^2 \cdot 103 \cdot 37277$.
3. Evaluate the IBP solution using these numbers.

$$* \text{ E.g. } \text{CO}[I_{2,1,1}, I_{0,1,1}] = \frac{383953112}{548314574947073136171275} = \frac{2^3 \cdot 1117 \cdot 42967}{5^2 \cdot 103 \cdot 37277 \cdot 75579811^2}.$$

4. Find common prime factors, identify their powers.

$$* \text{CO}[I_{2,1,1}, I_{0,1,1}] \sim (d - 6)^{-1} s^{-2}.$$

Automated implementation: `toos/guessfactors` from RATRACER.

To find the set of possible factors:

- * Reconstruct a simpler subset of the coefficients. (A few per sector).
- \Rightarrow Easy with RATRACER, just select individual outputs.

Once the factors are found, *speedup the reconstruction* by dividing them out from the expressions.

- * I.e. reconstruct $\text{CO}[I_{2,1,1}, I_{0,1,1}] / (d - 6) / s^2$, not just $\text{CO}[I_{2,1,1}, I_{0,1,1}]$.

Usage for IBP reduction

1. Use KIRA to generate the IBP equations.

```
$ cat >config/integralfamilies.yaml <<EOF
integralfamilies:
  - name: "I"
    loop_momenta: [1]
    top_level_sectors: [b111]
    propagators:
      - ["1", 0]
      - ["1-p1", 0]
      - ["1+p2", 0]
EOF
$ cat >config/kinematics.yaml <<EOF
kinematics:
  outgoing_momenta: [p1, p2]
  kinematic_invariants: [[s, 2]]
  scalarproduct_rules:
    - [[p1,p1], 0]
    - [[p2,p2], 0]
    - [[p1,p2], "s/2"]
# symbol_to_replace_by_one: s
EOF
$ cat >export-equations.yaml <<EOF
jobs:
  - reduce_sectors:
      reduce:
        - {sectors: [b111], r: 4, s: 1}
      select_integrals:
        select_mandatory_recursively:
          - {sectors: [b111], r: 4, s: 1}
      run_symmetries: true
      run_initiate: input
EOF
$ kira export-equations.yaml
```

2. Use RATRACER to create a trace with the solution.

```
$ ratracer \
  load-equations input_kira/I/SYSTEM_I_0000000007.kira.gz \
  load-equations input_kira/I/SYSTEM_I_0000000006.kira.gz \
  solve-equations choose-equation-outputs --maxr=4 --maxs=1 \
  optimize finalize save-trace I.trace.gz
```

3. Optionally expand the outputs into a series in ϵ .

```
$ ratracer \
  set d '4-2*eps' load-trace I.trace.gz \
  to-series eps 0 \
  optimize finalize save-trace I.eps0.trace.gz
```

4. Use RATRACER (+FIREFLY) to reconstruct the solution.

```
$ ratracer \
  load-trace I.eps0.trace.gz \
  reconstruct --to=I.solution.txt --threads=8 --inmem
```

Usage as a library

RATRACER is built to support custom user-defined traces.

Any rational algorithm can be turned into a trace (via the C++ API).

Usage:

```
#include <ratracer.h>
int main() {
    Tracer tr = tracer_init();

    Value x = tr.var(tr.input("x"));
    Value y = tr.var(tr.input("y"));

    Value x_sqr =
        tr.pow(x, 2);
    Value expr =
        tr.add(x_sqr, tr.mulint(y, 3));

    /* expr = x^2 + 3y */
    tr.add_output(expr, "expr");

    tr.save("example.trace.gz");
    return 0;
}
```

API:

```
struct Value { uint64_t id; uint64_t val; };
struct Tracer {
    Value var(size_t idx);
    Value of_int(int64_t x);
    Value of_fmpz(const fmpz_t x);
    bool is_zero(const Value &a);
    bool is_minus1(const Value &a);
    Value mul(const Value &a, const Value &b);
    Value mulint(const Value &a, int64_t b);
    Value add(const Value &a, const Value &b);
    Value addint(const Value &a, int64_t b);
    Value sub(const Value &a, const Value &b);
    Value addmul(const Value &a,
                const Value &b1,
                const Value &b2);
    Value inv(const Value &a);
    Value neginv(const Value &a);
    Value neg(const Value &a);
    Value pow(const Value &base, long exp);
    Value div(const Value &a, const Value &b);
    void assert_int(const Value &a, int64_t n);
    void add_output(const Value &src, const char *name);
    size_t input(const char *name, size_t len);
    size_t input(const char *name);
    int save(const char *path);
    void clear();
};
Tracer tracer_init();
```

RATRACER future plans

For very large examples *main memory speed becomes the bottleneck* for the modular evaluation. So:

- * Investigate optimizing traces to improve memory access patterns.

For other examples RATRACER speeds up the evaluation enough that the *modular reconstruction in FIREFLY becomes the bottleneck*. So:

- * Reduce the overhead in FIREFLY to speed up simpler examples.
 - * Improve parallelizability in FIREFLY to help with complicated examples.
- ⇒ Ongoing collaboration with FIREFLY authors.

Beyond guessing the denominators:

- * Investigate smaller ansätze for the results (via e.g. partial fractioning).

[Abreu et al '19; De Laurentis, Page '22]

Summary

RATRACER:

- * Practical faster modular evaluation of linear system solutions.
- * Trace optimization, transformation, slicing and dicing.
- * Coefficient expansion in ε (and not only).
- * Denominator guessing.
- * Available at github.com/magv/ratracer.
 - * Benchmark code & results at github.com/magv/ibp-benchmark.
- * TODO: faster evaluation, faster reconstruction, more tricks.