

## Original software publication

## ACORNS: An easy-to-use code generator for gradients and Hessians

Deshana Desai<sup>1</sup>, Eta Shuchatowitz<sup>1</sup>, Zhongshi Jiang, Teseo Schneider, Daniele Panozzo<sup>\*</sup>

New York University, 60 5th Ave, New York, NY 10011, United States of America



## ARTICLE INFO

## Article history:

Received 9 July 2021

Received in revised form 10 November 2021

Accepted 12 November 2021

## Keywords:

Code generation

Automatic differentiation

## ABSTRACT

The computation of first and second-order derivatives is a staple in many computing applications, ranging from machine learning to scientific computing. We propose an algorithm to automatically differentiate algorithms written in a subset of C99 code and its efficient implementation as a Python script. We demonstrate that our algorithm enables automatic, reliable, and efficient differentiation of common algorithms used in physical simulation and geometry processing.

© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Code metadata

## Current code version

Permanent link to code/repository used for this code version

Code Ocean compute capsule

Legal Code License

Code versioning system used

Software code languages, tools, and services used

Compilation requirements, operating environments &amp; dependencies

If available Link to developer documentation/manual

Support email for questions

v3.0

<https://github.com/ElsevierSoftwareX/SOFTX-D-21-00129>

N/A

MIT

git

Python, C99

Conda Package Manager

<https://github.com/deshanadesai/acorns>

panozzo@nyu.edu

## 1. Motivation and significance

The calculation of derivatives is used in many applications, including machine learning, scientific computing, geometry processing, computer vision, natural language processing, and many more. The calculation of derivatives of large and complex computational graphs is a computational bottleneck for these applications. Thus the efficient computation of derivatives is an area of active research interest. Extensive research and engineering efforts have been spent in the last years to support applications in deep learning, that led to the development of libraries such as TensorFlow or PyTorch, which target applications with large and dense tensors. The availability of easy-to-use differentiation libraries is one of the reasons for the massive advancement in deep learning. They allow researchers to focus on the design of new network architectures and loss functions, while automating and ensuring correctness in the low-level tasks required to minimize these functionals. However, they are specialized for computing

first derivatives only, and they are optimized for large, dense tensors, which are uncommon in applications outside of deep learning. The novelty in these libraries is not in the automatic differentiation technique used, but in the ease of use, cross-platform availability, and ease of integration in new research projects.

Our goal is to provide a similar solution for scientific applications requiring first and second derivatives of expressions involving small to medium tensors, using an approach based on symbolic differentiation. When paired with the optimization algorithms available in modern compilers, our approach generates code around an order of magnitude faster than existing methods for computing Hessians while allowing users to write their functions directly in the C language. The algorithm is easy to integrate into existing build systems and produces dependency-free code.

The core parts of the library are implemented in Python using a C99 parser [1]. We perform symbolic differentiation with unrolled derivatives, stored as expression trees, and generate an efficient kernel to evaluate them using an existing C99 compiler, addressing the potentially long compilation times with a file-splitting approach. State-of-the-art compilers are capable of sophisticated optimizations such as loop and expression trees vectorization, elimination of redundancy in expressions, and

\* Corresponding author.

E-mail address: [panozzo@nyu.edu](mailto:panozzo@nyu.edu) (Daniele Panozzo).

<sup>1</sup> Both authors contributed equally to this research.

reusing computations performed in previous iterations of loops. Our approach benefits from these features, and the performance of our method will likely increase as compiler technology progresses. The expression unrolling also allows us to trivially parallelize computations not only for different data points but also for a single data point over multiple CPU cores, which are common in modern architectures. Finally, we provide support for the efficient differentiation of algorithms with nested loops, arrays, and functions. A key difference with existing libraries, is that we create kernels to efficiently evaluate the expression itself, its gradient, and its Hessian, all optionally supporting parallel evaluation.

Our open-source implementation makes the integration of the algorithm straightforward in modern C and C++ applications, thanks to its ability to directly differentiate code and to produce standard, dependency-free C99 code that can be easily used in existing applications.

To demonstrate the practical impact of our approach, we integrated it into two applications, one in geometry processing and one in scientific computing. We show that our library provides important speedups in the overall running time of mesh parametrization algorithms and in the quasi-static simulation of non-linear elastic deformations.

### 1.1. Related work

Automatic (or algorithmic) differentiation has been introduced in the pioneering works of [2–4]. We briefly review the most relevant work and refer readers to various books and survey for further detail [5–8] and [www.autodiff.org](http://www.autodiff.org) for a collection of implementations. We provide comparisons against representative methods for operator overloading and source transformation for both gradients and higher order derivatives in Section 4.

### 1.2. Operator overloading

The most common way of implementing automatic differentiation is through defining a new object class, which associates the gradient information to the data. In forward mode, the gradient information is updated throughout the trace of the computation. And in reverse mode, a tape of computation is recorded, and the gradient information is then (back)propagated following the recorded tape. Due to the versatility and freedom of prototyping, operator overloading based approaches have gained a lot of attraction, and notable examples include [9–14]. However, the efficiency of the code often suffers due to the additional runtime overhead, which is especially noticeable when the same computation is executed repeatedly.

### 1.3. Source transformation

Source transformation approaches parse the input source code (most commonly C and Fortran [15–18]) and generate a new algorithm that computes the derivatives of the input function. Compared to operator overloading approaches, obtaining the source code file *a priori* opens the doors to global data flow analysis and optimization, thus reducing the running time. Additionally, this approach enables the developers to directly debug the code for derivatives computation, a crucial feature to identify and address potential issues with numerical round-off or overflows. The reliance on compilers to compile the generated source-code makes these approaches future-proof, in the sense that the generated code will benefit from progress on compiler design and optimization.

In our algorithm, we restrict to a subset of C99 covering common use-cases in scientific computing, finite element analysis, and computer graphics.

A minor, yet practically relevant, drawback of source code transformation is the presence of the additional source files, posing challenges on the build system and version control of the evolving software. We show that our library, thanks to its permissive open-source license, minimal dependencies, and availability in the Conda package system, is effortless to integrate into a CMake-based build system [19]. We show how to automatically generate the required files as part of the build process, making the automatic differentiation transparent to the user (Section 3.3).

### 1.4. Higher order derivatives

Although valuable in scientific computing and numerical optimization, computing higher-order derivatives is challenging due to the need of memory management and sparsity pattern detection [20,21]. Therefore, only few existing libraries support second or higher-order derivatives computation. Our method supports dense Hessian computation, taking advantage of the symmetric structure to reduce computational times [22].

## 2. Software description

The *input* of our algorithm is an algorithm written in a subset of C99 [23]: we support arrays, compile time known range (nested) loops, binary assignments, functions, and variable declarations. The *output* is a set of self-contained, multi-threaded C99 functions to evaluate the function in the input file, its first derivative, and its second derivative. We designed our library to become part of a build system, seamlessly generating the derivative code as part of the build procedure of a software package (Section 3.3).

We overview our algorithm in Section 2.1, and show a complete execution on a prototypical expression in Section 3. We then discuss how to parallelize the evaluation in Section 3.1 and how to make our approach scale to large expressions in Section 3.2. We show the performance of the algorithm on synthetic expressions (Section 4), comparing it with state-of-the-art differentiation libraries, and integrate it into a physical simulation library for a realistic benchmark.

### 2.1. Overview

*Primer on automatic differentiation.* Given a target function  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , the corresponding Jacobian matrix has entries  $J_{ij} = \frac{\partial f_i}{\partial x_j}$ . We use the pyc parser [1] to parse the input C99 code to produce an Abstract Syntax Tree (AST) of the input algorithm. The leaves of the tree are either numerical constants or variables and the internal nodes follow a hierarchical structure built from the code. Let us consider an example: if  $f$  is a composite function:  $f(x) = h \circ g(x) = h(g(x))$ , with  $x \in \mathbb{R}^n$ ,  $g: \mathbb{R}^n \rightarrow \mathbb{R}^k$ ,  $h: \mathbb{R}^k \rightarrow \mathbb{R}^m$ , then

$$J_{ij} = \frac{\partial f_i}{\partial x_j} = \frac{\partial h_i}{\partial g_1} \frac{\partial g_1}{\partial x_j} + \frac{\partial h_i}{\partial g_2} \frac{\partial g_2}{\partial x_j} + \cdots + \frac{\partial h_i}{\partial g_k} \frac{\partial g_k}{\partial x_j}. \quad (1)$$

From the example we see that the auto-differentiation of a function  $f$  breaks down the action of the Jacobian matrix on a vector into simple components, which are evaluated sequentially. We store the operations required for the calculation of the derivative of each of these pieces as a string. The final derivative is the accumulation of the piece-wise derivatives of the sub-operations following the chain rule.

For computing second order derivatives, we differentiate the tree twice. We further optimize over the number of calculations by taking advantage of the symmetry of the Hessian: we compute only the lower triangular part of the matrix and copy it over the upper triangular half.

*Implementation specific details.* The derivative formulas are unrolled in a sequence of assignments (no loops or conditional are used) and saved on a C file. The file is then compiled with full code optimization (-O3 with gcc [24] and clang [25]). This leverages the compiler’s ability to perform redundancy elimination on trees, reusing computations (especially memory loads and stores), and automatically vectorizing the code when appropriate.

In the case of loops or conditional statements being present in the program, we perform an additional parsing step to expand these structures. We perform loop unrolling to produce a sequence of operations and eliminate other instructions that control the loop. Conditional statements are removed by evaluating their condition during parsing. The output of this phase is a list of assignment instructions, which are stored in an intermediate binary format. This file is then parsed to create the AST, differentiated, and the differentiated AST is exported as a C file.

### 3. Illustrative example

To further illustrate the mechanisms of ACORNS, we consider a typical example from statistics. We will compute the gradient of the *cross entropy* loss function, between two probability distributions  $a$  and  $b$ . The loss function can be given by the following piece of C99 code:

```
double cross_entropy(const double **a, const double **b){  
    double loss = 0;  
    for(int i = 0; i < 2; i++){  
        for(int j = 0; j < 2; j++ ){  
            loss = loss - (b[i][j] * log(a[i][j] + 0.00001));  
        }  
    }  
    return loss;  
}
```

The function is unrolled to produce the following code:

```
    double loss = 0;
    loss = (loss)-((b[0][0])*log((a[0][0])+(0.00001)));
    loss = (loss)-((b[0][1])*log((a[0][1])+(0.00001)));
    loss = (loss)-((b[1][0])*log((a[1][0])+(0.00001)));
    loss = (loss)-((b[1][1])*log((a[1][1])+(0.00001)));
    return loss;
}
```

An Abstract Syntax Tree is constructed from the unrolled code. The syntax tree is differentiated using back-propagation to construct the derivative equations. This unrolled code is finally compiled by a C compiler with full code optimization enabled (see Fig. 1).

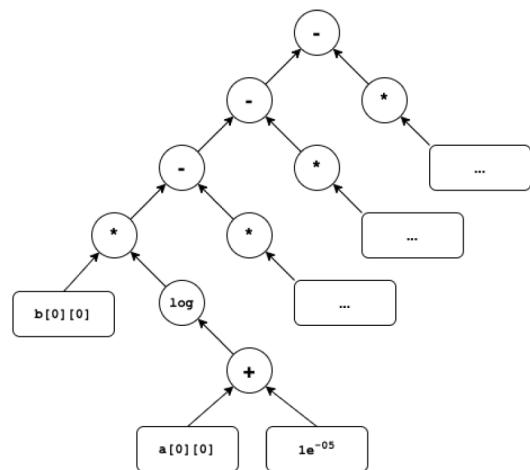
```

for(int p = 0; p < num_points; ++p)
{
    ders[p*2+0] += (((((0) - (((log((a[0][0] + 0.00001)) * (b[0][0]) + b[0][0] * ((1/(a[0][0] + 0.00001))*0)))) - (((log((a[0][1] + 0.00001)) * (b[0][1]) + b[0][1] * ((1/(a[0][1] + 0.00001))*0)))))) - (((log((a[1][0] + ...
}

```

### 3.1. Parallelization

Parallel computation can be used to accelerate the evaluation of large gradients or Hessians [26]. We observe that, in our case, the evaluation of the different entries is independent, and can thus be trivially parallelized using OpenMP [27] directives. We experimentally study the scaling of our generated code in Section 4.



**Fig. 1.** Expression graph for *cross entropy* loss function.

### 3.2. Large expressions

As the complexity of the expressions and/or the number of outputs grows, the size of the generated C file grows accordingly. While this opens more opportunities for the compilers to optimize the evaluation, it also increases memory consumption and compilation time. For large expressions, such as the Hessian of a cubic Lagrangian element, this might become impractical. We propose a simple, yet effective and practical, strategy to overcome this issue: we split the output into multiple C files that are compiled independently. This has the downside of reducing the optimization opportunities of compilers. However, in our experiments, the performance drop is negligible and we thus use this option aggressively to reduce compilation times.

To evaluate the effect of the split on performances, we consider a function for calculating the Neo-Hookean Elasticity energy [28] for a linear, quadratic, and cubic Lagrange element (12, 30, and 60 variables respectively). The code generated by these functions is large (29 GB for the function with 60 variables) and the compilation with `gcc9` does not terminate in 160 h.

We tested splitting into different number of files, measuring compilation time, evaluation time, and binary size after compilation.

As expected, the compilation time decreases as the files become smaller ([Fig. 2 Top](#)) while the size of the binary generated is not noticeably affected by the split ([Fig. 2 Middle](#)). Surprisingly, the evaluation time of the generated code is also not noticeably affected by the split ([Fig. 2 Bottom](#)). As a default, we thus opt for splitting so that each file had a size of around 16 MB, which is a good compromise between number of files and compilation time.

### 3.3. Integration in a CMake build system

One of the key design goal of our system is the easy and fast deployment in numerical based methods. We show how ACORNS can be integrated within a CMake build system for quick and easy deployment. In the following we describe the process and we refer to our github repository for the complete example <https://github.com/deshanadesai/acorns>. Let us consider a simple problem: we want to minimize the following function (saved in the file `function.c`) using gradient descent.

```
int function_0(double x){  
    double energy = pow(x, 4) - 3*pow(x, 3) + 2;  
    return energy;  
}
```

To use ACORNS to its derivative, we need to add the code generation to the CMake project file:

```
cmake_minimum_required(VERSION 3.1)
project(SimpleProject)

file(MAKE_DIRECTORY ${PROJECT_SOURCE_DIR}/ders)
add_custom_command(
OUTPUT
  ${PROJECT_SOURCE_DIR}/ders/der_0.c
  ${PROJECT_SOURCE_DIR}/ders/der_0.h
COMMAND
  acorns_autodiff
  ${PROJECT_SOURCE_DIR}/functions/function_0.c
  energy --vars x --func function_0
  --output_filename ${PROJECT_SOURCE_DIR}/ders/der_0
DEPENDS
  ${PROJECT_SOURCE_DIR}/functions/function_0.c
)
add_library(derivative "ders/der_0.c")
target_include_directories(derivative PUBLIC ders)

add_executable(main_target src/example.c)

# Link target derivative to main_executable
target_link_libraries(main_target PUBLIC derivative)
```

This will generate the `ders_0.h` and `ders_0.c` file, which will be directly compiled by CMake and linked with the other files in the project. To use the derivatives it is then sufficient to include the generated header file. We show an example of using the computed derivative to find the root by using gradient descent.

```
#include <der_0.h>
#include <stdio.h>
#include <math.h>

int main(int argc, char** argv) {
    double vals[1] = {6.0};
    double ders[1] = {0.0};

    int iteration;
    for (iteration = 0; iteration < 10000; iteration++) {
        // calculate derivative using ACORNS
        compute(vals, 1, ders);

        float step = 0.01 * ders[0];
        vals[0] -= step;

        if (fabs(step) <= 1e-5)
            break;
    }

    printf("Minimum: %f\n", vals[0]);
    printf("Iterations: %d\n", iteration);
    return 0;
}
```

## 4. Impact

We compare our generated code against 4 state-of-the-art differentiation algorithms:

1. Tapenade [15] shares many similarities with our system. It converts C code into a C program computing its derivatives and Hessians.<sup>2</sup>
2. PyTorch [9] is a backward auto-differentiation library tailored for machine learning. It provides both a CPU and a GPU version, we compare against the CPU version only.

<sup>2</sup> Hessian computation in Tapenade is not officially supported, we used the semi-manual construction proposed in the Q&A section of the documentation for the comparison.

3. Mitsuba Autodiff [14,29] compute automatic differentiation using operator overloading and depends on Eigen [30] for the linear algebra and storage. We run all our comparisons using only static stack-allocated matrices, since the dynamic mode is slower and unnecessary for the expressions we use for benchmarking.
4. Enoki [31] is a new library developed specifically to support differentiable rendering. Similarly to PyTorch, it supports both CPU and GPU evaluation, and we compare only with the CPU version and only for gradients, since it does not support Hessian computation.
5. Adept [32] is a C++ library that uses operator overloading to perform automatic differentiation. It does not support higher-order derivatives, so we restrict our comparison to gradients. Adept does not support parallelization, so we run our comparison on a single thread.

We run our experiments on a 2.35 GHz AMD EPYC™ 7452 running Ubuntu 19.10 GNU/Linux 5.3.0-29-generic x86\_64.

*Expression types.* We use three expression to evaluate different realistic scenarios where autodifferentiation is used:

1. An high-order polynomial, a non-polynomial function, and procedural expressions with large number of variables:

$$f(x) = \frac{x^2 + 3x - x/4}{x} + x^4 + 22/7x^3 + x^9. \quad (2)$$

2. A scalar trigonometric function:

$$g(x) = \sin(x) + \cos(x) + x^2. \quad (3)$$

3. A vector-valued polynomial with  $s$  variables:

$$h_s(x) = 4^s \prod_{i=1}^s x_i(1 - x_i), \quad x = (x_1, \dots, x_s). \quad (4)$$

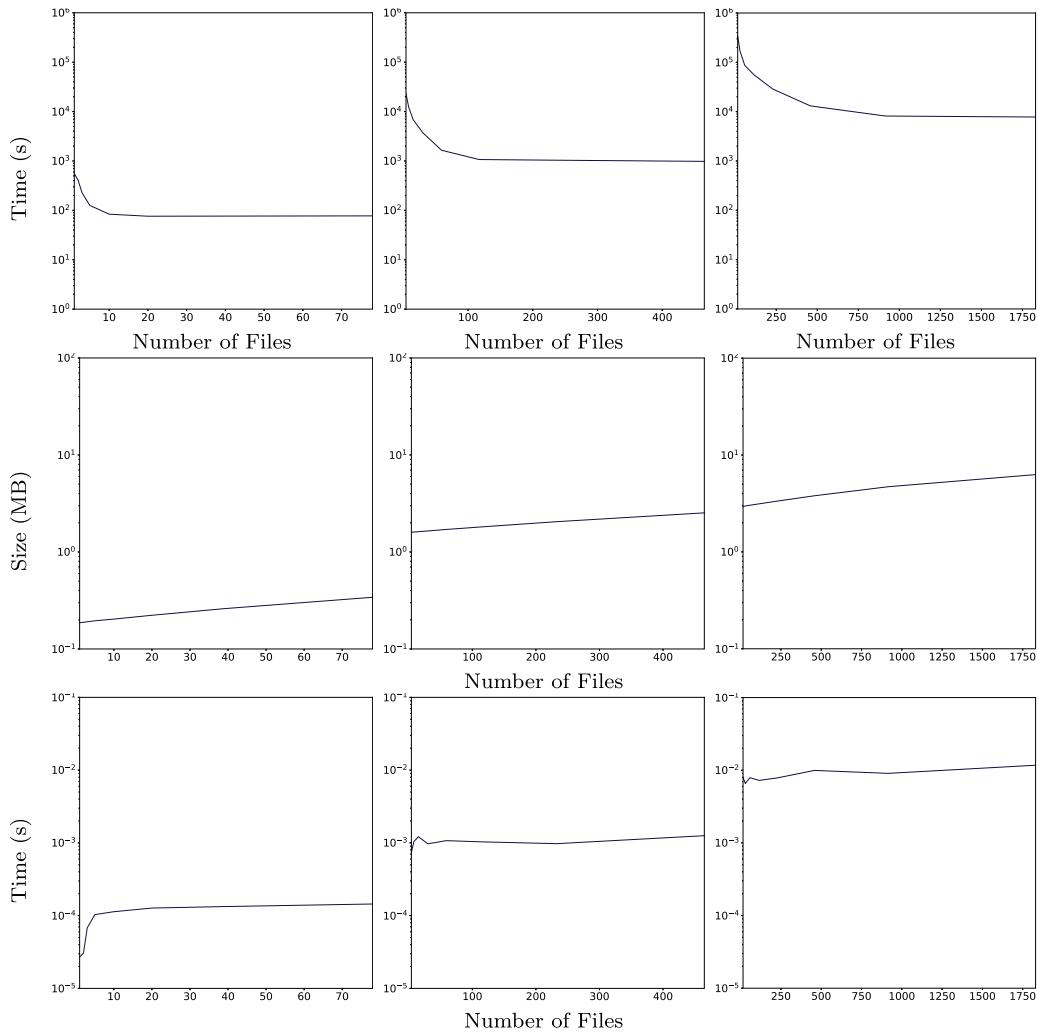
For each expression, we randomly generate evaluation points and plot the average running times over 10 runs. ACORNS generates C code, Mitsuba and Enoki are both running in C++. All the C and C++ code is compiled with `gcc9/g++9` with the flags `-O3`, `-ffast-math`, `-fno-finite-math-only` and, only for C++, `-std=c++11`.

*Gradient.* In Fig. 3 (top) we report the timings, for different number of evaluation points, for the three functions. For the two polynomial functions we are roughly  $5\times$  faster than Mitsuba, Adept, and Enoki and  $2.9\times$  faster than PyTorch. For the trigonometric function, due to the increase in the cost of evaluating the trigonometric function, our advantage is slightly reduced. For gradients, our approach is very similar to Tapenade, and we obtain comparable timings.

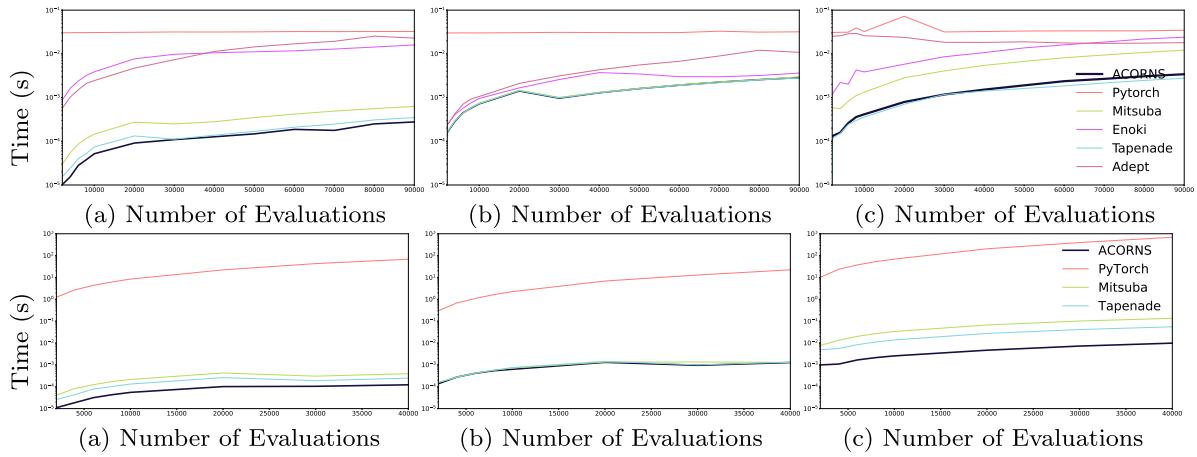
In Fig. 4 (left) we fix the number of evaluations and change the parameter  $s$  in Eq. (4), to evaluate how the methods scale with the number of variables. As expected, all methods scale linearly, but our method is still  $3.97\times 7.95\times$ , and  $3\times$  faster than Mitsuba, Enoki, and Pytorch respectively. As for Fig. 3, Tapenade's code has very similar running time as ours. Adept scales similarly to Enoki.

*Hessian.* We repeat the same experiments as in previous section to compare ACORNS with Mitsuba, PyTorch, and Tapenade to compute Hessians. Fig. 3 (bottom) shows that our code is very efficient,  $12.8\times$  faster than Mitsuba and  $1336\times$  faster than PyTorch (which is not optimized to evaluate Hessians). Tapenade is faster than the other alternatives, but it is still around 5 times slower than our code. Note also that while the Hessian generation is completely automatic with ACORN, it does require manual interaction with Tapenade, since Hessian computation is not fully supported.

The scaling with respect to the number of variables for Hessian evaluation shown in Fig. 4 (right) has a similar trend as for the gradients.



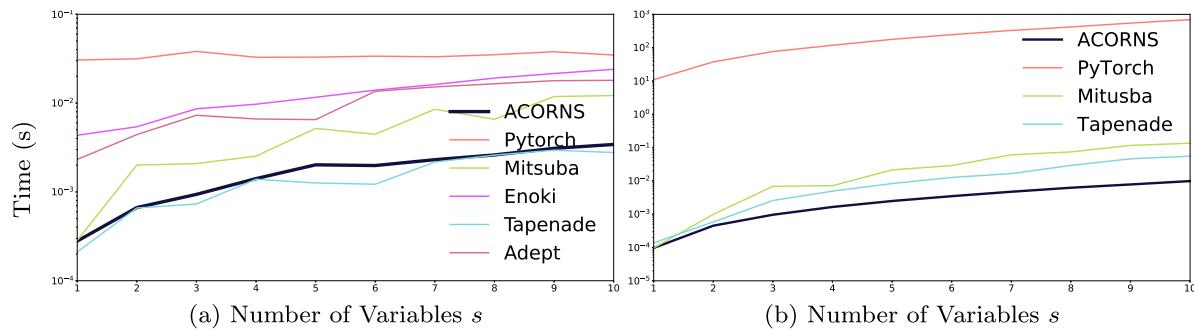
**Fig. 2.** (Top) Compile times for Hessians of a function of 12, 30 and 60 variables, respectively. (Middle) Storage sizes of the generated binary files (.o) for Hessians of a function of 12, 30, and 60 variables, respectively. (Bottom) Run times for Hessians of a function of 12, 30, and 60 variables, respectively.



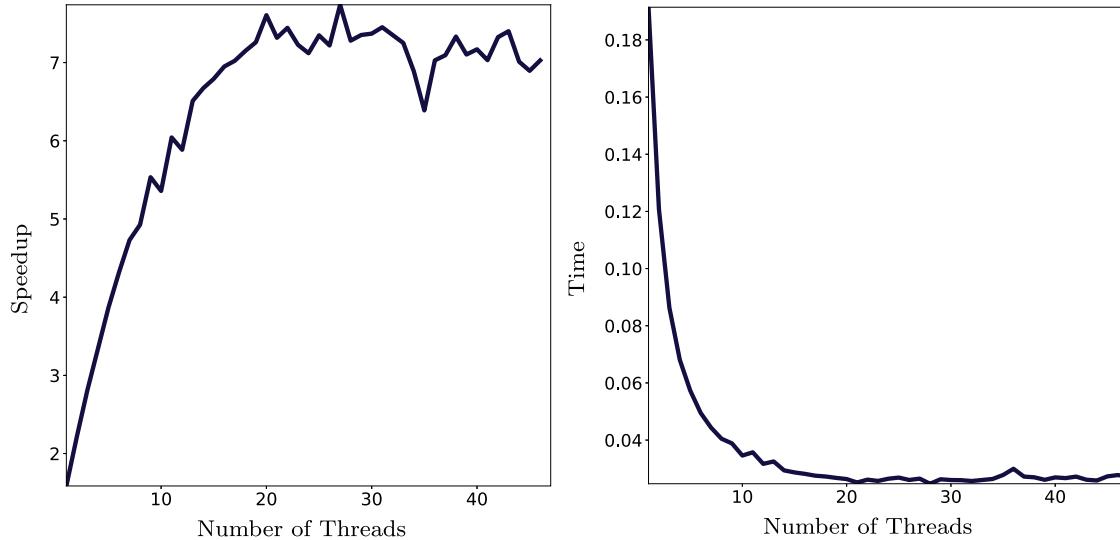
**Fig. 3.** Average runtimes in log scale for computing the gradient (top) or Hessian (bottom) for different test functions. (a) Long Polynomial in Eq. (2) (b) Trigonometric Function in Eq. (3) (c) Polynomial in Eq. (4) with  $s = 10$ .

**Parallel evaluation.** We ran the function in Eq. (4) for  $s = 25$  and computed the gradient for 100000 evaluation points allowing OpenMP to use a different number of threads. We observe an almost linear scaling up to 15 threads as shown in Fig. 5 (right) on

our workstation, and it then flattens as more threads are added. We speculate that this is due to saturation of the memory bandwidth of the workstation, since our task is inherently massively parallelizable.



**Fig. 4.** Run times with respect to the number of variables for the polynomial in Eq. (4) for the different methods. (a) Gradient (b) Hessian.



**Fig. 5.** Time and speedup of our method over number of threads for computing the Hessian expression of  $h_{25}$  Eq. (4) for 100,000 points.

## 5. Conclusions

We introduced ACORNS, a software library for automatic differentiation that generates efficient kernels for computing both gradients and Hessian and can be easily integrated in existing C or C++ projects. The core idea of our algorithm is to fully unroll the computational graph, and rely on modern compilers to optimize the resulting code. Compared with alternatives, our algorithm is faster at evaluation time for Hessian computation, but slower during compilation: we believe that this trade-off is very interesting for many scientific computing applications where the evaluation of small dense Hessian is repeated millions of times. The compilation time could be reduced by performing symbolic simplification on the code before generating the .c files. In addition, we could avoid the parsing time of the compiler by directly generating byte-code (for example for the LLVM compiler [25]) and compiling it in memory, avoiding unnecessary disk access. Variable conditional used in branches and loops are not supported by ACORNS. Support for it could be added, but the performance will likely decrease, due to the reduced optimization options and more complex parallelization.

We would like to extend ACORNS to directly support the generation of sparse Hessians [21,33] and also extend the code generation to target GPU accelerators.

Our open-source reference implementation and a set of examples on how to use it is available at <https://github.com/deshanadessai/acorns>, and it is also released as a Conda package on conda-forge to allow an easy installation on all major operating systems.

We also release our benchmark at <https://github.com/geometryprocessing/acorns-benchmark> to foster replicability of our results.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

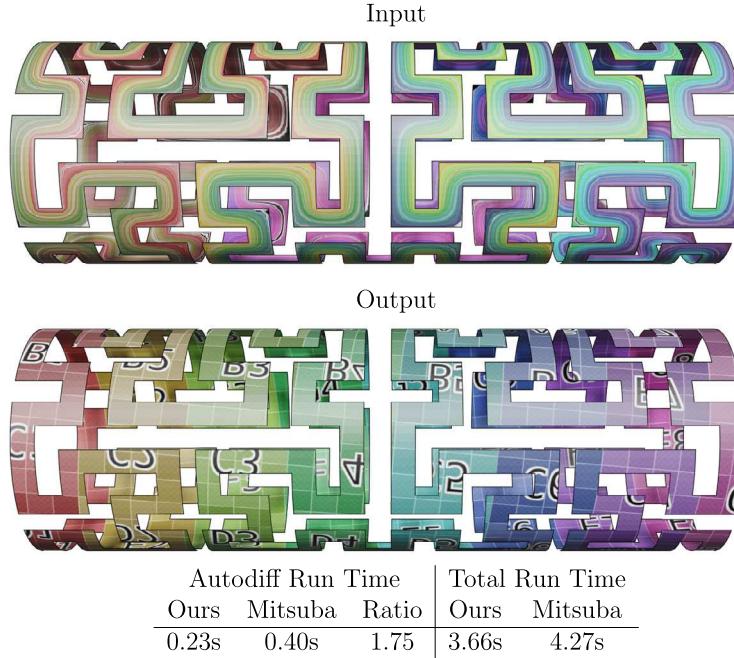
We thank NYU High Performance Computing for resources, services, and staff expertise. This work was partially supported by the National Science Foundation, United States of America CAREER award under Grant No. 1652515, the National Science Foundation, United States of America grants OAC-1835712, OIA1937043, CHS-1908767, CHS-1901091, the NSERC, Canada grants RGPIN-2021-03707 and DGECR-2021-00461, a Sloan Fellowship, a gift from Adobe Research, a gift from nTopology, and a gift from Advanced Micro Devices, Inc.

## Appendix. Applications

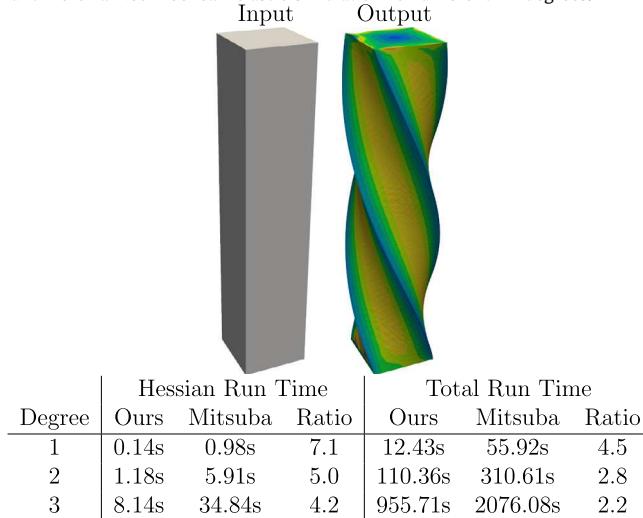
We integrated ACORNS in two applications in geometry processing and in physical simulation to evaluate the gain in performance in a realistic setting. In both these cases, the algorithm needs to assemble a dense gradient and a sparse Hessian. For both applications, ACORNS is used to compute small dense

**Table A.1**

Optimize the parametrization of a surface mesh.

**Table A.2**

Runtime of a Neo-Hookean elastic simulation for different FE degrees.



blocks, which are then assembled in a larger sparse Hessian matrix. Directly supporting the automatic construction of the sparse Hessian is an interesting avenue for future work.

#### A.1. Parametrization

Given a surface embedded in three dimensional space, an important problem in geometry processing is to assign parametric values to it, i.e. computing a flattening of the surface on a plane. We refer the readers to [34,35] for a comprehensive survey. In our application, we consider minimizing the non-convex symmetric Dirichlet energy [36] with projected Newton's method [37]: when the Hessian is not positive (semi-)definite, possibly at saddle points, we project it to the closest positive definite matrix through polar decomposition. The algorithm involves computing the full Jacobians and Hessians per element: our original implementation uses Mitsuba, and we replaced it with ACORNS. We

summarize the statistics in **Table A.1**. The automatic differentiation run time is reduced considerably, but in this application it is not the bottleneck, thus resulting in a small reduction of the overall running time.

#### A.2. Finite element simulation

The finite element method aims at solving partial differential equations (PDEs) describing physical phenomena. A typical example is to simulate the deformation of an elastic body subject to some external forces (boundary conditions); the algorithm involves minimizing an elastic energy, and requires computing derivatives. We integrate ACORNS in PolyFEM [38] and compared, for different finite element discretization order (ranging from linear to cubic), the timings of a Neo-Hookean simulation using the Mitsuba autodiff and with ACORNS to compute Hessian and gradient of the energy. In the finite element method, all computations are performed locally on each tetrahedron. The different

order will have different local number of degree of freedom (which determine the size of the gradient and Hessian): a linear tetrahedron has 12, a quadratic has 30, and a cubic has 60. Our method provides an overall speedup of 2 to 4 times, depending on the degree, over the entire simulation time (Table A.2).

## References

- [1] Bendersky E. Pycparser. GitHub Repository 2020. <https://github.com/eliben/pycparser>.
- [2] Wengert RE. A simple automatic derivative evaluation program. *Commun ACM* 1964;7(8):463–4. <http://dx.doi.org/10.1145/355586.364791>.
- [3] Speelpenning B. Compiling fast partial derivatives of functions given by algorithms. Tech. rep., Illinois Univ., Urbana (USA). Dept. of Computer Science; 1980, <http://dx.doi.org/10.2172/5254402>.
- [4] Beda LM, Korolev LN, Sukkikh NV, Frolova TS. *Programs for automatic differentiation for the machine BESM*. Technical Report, Moscow, USSR: Institute for Precise Mechanics and Computation Techniques, Academy of Science; 1959, (In Russian).
- [5] Naumann U. The art of differentiating computer programs: An introduction to algorithmic differentiation. vol. 24, SIAM; 2012, <http://dx.doi.org/10.1137/1.9781611972078>.
- [6] Griewank A, Walther A. Evaluating derivatives: Principles and techniques of algorithmic differentiation. vol. 105, Siam; 2008, <http://dx.doi.org/10.1137/1.9780898717761>.
- [7] Bischof CH, Bücker HM. Computing derivatives of computer programs. In: Grotendorst J, editor. *Modern methods and algorithms of quantum chemistry: Proceedings*. NIC Series, Second edition. vol. 3, Jülich: NIC-Directors; 2000, p. 315–27, URL <http://hdl.handle.net/2128/6053>.
- [8] Margossian CC. A review of automatic differentiation and its efficient implementation. *Wiley Interdiscip Rev Data Min Knowl Discov* 2019;9(4):e1305. <http://dx.doi.org/10.1002/widm.1305>.
- [9] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: An imperative style, high-performance deep learning library. In: *Advances in neural information processing systems*. 2019, p. 8024–35.
- [10] Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, et al. Automatic differentiation in pytorch. In: NIPS Autodiff Workshop; 2017.
- [11] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. Tensorflow: A system for large-scale machine learning. In: 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16). 2016, p. 265–83. <http://dx.doi.org/10.1145/2450153.2450158>.
- [12] Bell BM. CppAD: a package for C++ algorithmic differentiation. *Comput. Infrast. Oper. Res.* 2012;57(10). URL <https://coin-or.github.io/CppAD/doc/CppAD.htm>.
- [13] Griewank A, Juedes D, Utke J. Algorithm 755: ADOL-C: a package for the automatic differentiation of algorithms written in C/C++. *ACM Trans Math Softw* 1996;22(2):131–67. <http://dx.doi.org/10.1145/229473.229474>.
- [14] Jakob W. Mitsuba renderer. 2010, <http://www.mitsuba-renderer.org>.
- [15] Hascoet L, Pascual V. The tapenade automatic differentiation tool: Principles, model, and specification. *ACM Trans Math Softw* 2013;39(3):1–43. <http://dx.doi.org/10.1145/2450153.2450158>.
- [16] Voßbeck M, Giering R, Kaminski T. Development and first applications of TAC++. In: *Advances in automatic differentiation*. Springer; 2008, p. 187–97. [http://dx.doi.org/10.1007/978-3-540-68942-3\\_17](http://dx.doi.org/10.1007/978-3-540-68942-3_17).
- [17] Vassilev V, Vassilev M, Penev A, Moneta L, Ilieva V. Clad – automatic differentiation using clang and LLVM. 608, (1):IOP Publishing; 2015, 012055. <http://dx.doi.org/10.1088/1742-6596/608/1/012055>.
- [18] Moses WS, Churavy V. Instead of rewriting foreign code for machine learning, automatically synthesize fast gradients. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, editors. *Advances in neural information processing systems*. 33, 2020, p. 12472–85, URL <https://proceedings.neurips.cc/paper/2020/file/9332c513ef44b682e9347822c2e457ac-Paper.pdf>.
- [19] Martin K, Hoffman B. *Mastering cmake: A cross-platform build system*. Kitware; 2010.
- [20] Gower RM, Mello MP. *Hessian Matrices Via Automatic Differentiation*. Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica; 2010.
- [21] Walther A. Computing sparse hessians with automatic differentiation. *ACM Trans Math Softw* 2008;34(1):1–15. <http://dx.doi.org/10.1145/1322436.1322439>.
- [22] Petra CG, Qiang F, Lubin M, Huchette J. On efficient hessian computation using the edge pushing algorithm in julia. *Optim Methods Softw* 2018;33:1010–29. <http://dx.doi.org/10.1080/10556788.2018.1480625>.
- [23] Ritchie DM, Kernighan BW, Lesk ME. *The C programming language*. Prentice Hall Englewood Cliffs; 1988.
- [24] Stallman RM. *GNU Compiler collection internals*. Free Softw Found 2002.
- [25] Lattner C, Adve V. LLVM: A compilation framework for lifelong program analysis & transformation. In: *International symposium on code generation and optimization*. 2004. CGO 2004. IEEE; 2004, p. 75–86. <http://dx.doi.org/10.1109/CGO.2004.1281665>.
- [26] Bücker HM, Rasch A, Vehreschild A. Automatic generation of parallel code for hessian computations. In: *Proceedings of the 2005 and 2006 international conference on OpenMP shared memory parallel programming*. IWOMP'05/IWOMP'06, Berlin, Heidelberg: Springer-Verlag; 2005, p. 372–81. [http://dx.doi.org/10.1007/978-3-540-68555-5\\_30](http://dx.doi.org/10.1007/978-3-540-68555-5_30).
- [27] Chandra R, Dagum L, Kohr D, Menon R, Maydan D, McDonald J. *Parallel programming in OpenMP*. Morgan Kaufmann; 2001.
- [28] Rivlin RS. Large elastic deformations of isotropic materials. 1. Fundamental concepts. *Philos Trans R Soc Lond Ser A Math Phys Eng Sci* 1948;240(822):459–508. [http://dx.doi.org/10.1007/978-1-4612-2416-7\\_5](http://dx.doi.org/10.1007/978-1-4612-2416-7_5).
- [29] Grinspun E, Hirani AN, Desbrun M, Schröder P. Discrete shells. In: *Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on computer animation*. Citeseer; 2003, p. 62–7.
- [30] Guennebaud G, Jacob B, et al. Eigen v3. 2010, <http://eigen.tuxfamily.org>.
- [31] Jakob W. Enoki: structured vectorization and differentiation on modern processor architectures. 2019, <https://github.com/mitsuba-renderer/enoki>.
- [32] Hogan RJ. Fast reverse-mode automatic differentiation using expression templates in C++. *ACM Trans Math Softw* 2014;40(4):1–16. <http://dx.doi.org/10.1145/2560359>.
- [33] Gebremedhin AH, Tarafdar A, Pothen A, Walther A. Efficient computation of sparse hessians using coloring and automatic differentiation. *INFORMS J Comput* 2009;21(2):209–23. <http://dx.doi.org/10.1287/ijoc.1080.0286>.
- [34] Floater MS, Hormann K. Surface parameterization: a tutorial and survey. In: *Advances in multiresolution for geometric modelling*. Springer; 2005, p. 157–86. [http://dx.doi.org/10.1007/3-540-26808-1\\_9](http://dx.doi.org/10.1007/3-540-26808-1_9).
- [35] Sheffer A, Hormann K, Levy B, Desbrun M, Zhou K, Praun E, et al. Mesh parameterization: Theory and practice. In: *ACM SIGGRAPP course notes*. 10, (1281500.1281510). 2007, <http://dx.doi.org/10.1145/1508044.1508091>.
- [36] Smith J, Schaefer S. Bijective parameterization with free boundaries. *ACM Trans Graph* 2015;34(4):1–9. <http://dx.doi.org/10.1145/2766947>.
- [37] Teran J, Sifakis E, Irving G, Fedkiw R. Robust quasistatic finite elements and flesh simulation. In: *Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on computer animation*. 2005, p. 181–90. <http://dx.doi.org/10.1145/1073368.1073394>.
- [38] Schneider T, Dumas J, Gao X, Zorin D, Panizzo D. PolyFEM. 2019, <https://polyfem.github.io/>.