Supplementary Information to

Spectrally accurate, reverse-mode differentiable bounce-averaging algorithm and its applications

Kaya Unalmis

February 1, 2026

1 Introduction

In this note, we discuss how we ensure automatic differentiation (AD) accurately computes derivatives of integrals with singularities and roots of polynomials as they relate to bounce-averaging. Some care must be taken to enforce automatic differentiability of the discretized system.

We focus on the differentiation of an integral with boundary singularities.

$$I(\boldsymbol{x}_{\text{opt}}) = \int_{a(\boldsymbol{x}_{\text{opt}})}^{b(\boldsymbol{x}_{\text{opt}})} \mathrm{d}\zeta \; f(\zeta, \boldsymbol{x}_{\text{opt}}) \tag{1}$$

We discretize the system with a spectrally accurate quadrature using a change of variable

$$\zeta = \zeta(z, \boldsymbol{x}_{\text{opt}}) \tag{2}$$

such that the Jacobian vanishes at the boundaries to the proper order.

$$I(\boldsymbol{x}_{\text{opt}}) = \int_{-1}^{1} \mathrm{d}z \; f(\zeta(z, \boldsymbol{x}_{\text{opt}}), \boldsymbol{x}_{\text{opt}}) \left(\frac{\partial \zeta}{\partial z}\right)_{\boldsymbol{x}_{\text{opt}}} \tag{3}$$

We approximate this integral with samples at z_i with weights ω_i. Defining the Jacobian of the affine map $(a, b) \to (-1, 1)$ as

$$c(\boldsymbol{x}_{\text{opt}}) = [b(\boldsymbol{x}_{\text{opt}}) - a(\boldsymbol{x}_{\text{opt}})]/2 \tag{4}$$

and the Jacobian of the transformation that cancels the singularity as J, the quadrature becomes

$$I_{\mathrm{d}}(\boldsymbol{x}_{\text{opt}}) = c(\boldsymbol{x}_{\text{opt}}) \sum_{i=1}^{N} f(\zeta(z_i, \boldsymbol{x}_{\text{opt}}), \boldsymbol{x}_{\text{opt}}) \underbrace{J(z_i)\omega_i}_{W_i}. \tag{5}$$

2 Potential failure modes

In this section, we discuss failure modes that are common with naïve use of automatic differentiation. Johnson et al. present the limitations of finite-precision arithmetic as a flaw of automatic differentiation in [1]. The issue discussed here is more fundamental in that it persists in infinite-precision.

The first problem that may arise in the differentiation of (1) is that AD differentiates the quadrature, whereas the goal is to differentiate the integral.

The second problem that may arise is due to a limitation of automatic differentiation. Automatic differentiation can compute the derivative of a map

$$F \colon \boldsymbol{x}_{\text{opt}} \mapsto \prod_i \sum_j f_{ij}(\boldsymbol{x}_{\text{opt}}) \tag{6}$$

only if each f_{ij} is differentiable with respect to $\boldsymbol{x}_{\text{opt}}$. On the other hand, finite-differencing can bypass this requirement for it only requires F to be differentiable. We will refer to this stricter form of differentiability required by AD as atomic differentiability, where each atom of computation must be differentiable. If F is differentiable, but there exists an f_{ij} which is not, then one must analytically combine atoms of computation by imposing the AD tool groups operations with a custom derivative. For example, this may be done by defining the vector Jacobian product of $x^2 = x^{-1/2} x^{5/2}$ in the form $\langle \cdot, 2x \rangle$ instead of composing two separate derivatives $\langle \cdot, -x^{-3/2}/2 \rangle$ and $\langle \cdot, 5x^{3/2}/2 \rangle$.

2.1 Example

In this section, an example is shown which demonstrates the issues discussed earlier.

Consider for $x_{\text{opt}} \in (-1, 1)$, the integral

$$I(x_{\text{opt}}) = \int_{-1}^{1} \mathrm{d}z \; f(z, x_{\text{opt}}) \tag{7}$$

$$f(z, x_{\text{opt}}) = (z - x_{\text{opt}}) \sin(z - x_{\text{opt}})^{-1}. \tag{8}$$

Note $f(x_{\text{opt}}, x_{\text{opt}}) = 0$. Using the change of variables $u = z - x_{\text{opt}}$ so that

$$I(x_{\text{opt}}) = \int_{-1-x_{\text{opt}}}^{1-x_{\text{opt}}} \mathrm{d}u \; u \sin u^{-1} \tag{9}$$

and defining $g(u) = f(z(u, x_{\text{opt}}), x_{\text{opt}})$, one obtains

$$\frac{\mathrm{d}I}{\mathrm{d}x_{\text{opt}}} = -g(1 - x_{\text{opt}}) + g(-1 - x_{\text{opt}}). \tag{10}$$

Now an AD program which computes (8) will differentiate beneath the integral (7), which yields a non-integrable term:

$$\int_{-1}^{1} \mathrm{d}z \, |(z - x_{\text{opt}})^{-1} \cos(z - x_{\text{opt}})^{-1}| = \infty. \tag{11}$$

Consequently, the quadrature will yield noise and the discretized system will be ill-conditioned.

2.2 Derivatives of bounce integrals

In this section, we show analytically that AD is suitable for differentiating bounce integrals.

We seek the following derivative.

$$\frac{\mathrm{d}I}{\mathrm{d}\boldsymbol{x}_{\text{opt}}} = \int_{-1}^{1} \mathrm{d}z \, \frac{\partial}{\partial \boldsymbol{x}_{\text{opt}}} f(\zeta(z, \boldsymbol{x}_{\text{opt}}), \boldsymbol{x}_{\text{opt}}) \left(\frac{\partial \zeta}{\partial z}\right)_{\boldsymbol{x}_{\text{opt}}} \tag{12}$$

In our application, if the quantity on the right side of (12) is integrable, then the equality in (12) holds. AD as implemented in our code computes the derivative of the discretization (5).

$$\begin{aligned}\frac{\mathrm{d}I_{\mathrm{d}}}{\mathrm{d}\boldsymbol{x}_{\text{opt}}} &= \frac{\mathrm{d}c}{\mathrm{d}\boldsymbol{x}_{\text{opt}}} \sum_{i=1}^{N} W_i f(\zeta(z_i, \boldsymbol{x}_{\text{opt}}), \boldsymbol{x}_{\text{opt}}) \\ &\quad + c(\boldsymbol{x}_{\text{opt}}) \sum_{i=1}^{N} W_i \frac{\partial}{\partial \boldsymbol{x}_{\text{opt}}} f(\zeta(z_i, \boldsymbol{x}_{\text{opt}}), \boldsymbol{x}_{\text{opt}})\end{aligned} \tag{13}$$

The latter sum is a quadrature of

$$\int_{-1}^{1} \mathrm{d}z \, J(z) \frac{\partial}{\partial \boldsymbol{x}_{\text{opt}}} f(\zeta(z, \boldsymbol{x}_{\text{opt}}), \boldsymbol{x}_{\text{opt}}). \tag{14}$$

We proceed to verify the quadrature yields a spectrally accurate approximation of (14).

For bounce integrals weighted by $|v_{\parallel}|^{-1}$, we write

$$f \colon \zeta, \boldsymbol{x}_{\text{opt}} \mapsto g(\zeta, \boldsymbol{x}_{\text{opt}})(1 - \lambda |B|(\zeta, \boldsymbol{x}_{\text{opt}}))^{-1/2} \tag{15}$$

where λ is the pitch angle and $g(\zeta, \boldsymbol{x}_{\text{opt}}) \in C^{\infty}$ for all $\zeta, \boldsymbol{x}_{\text{opt}}$. For fixed $\boldsymbol{x}_{\text{opt}}$, such a quantity f has a boundary singularity that asymptotically matches the behavior of $y^{-1/2}$ near $y = 0$. This statement holds everywhere except when a boundary point is at a local maximum of $|B|$, in which case the singularity changes form to y^{-1} and becomes non-integrable.

$$\begin{aligned}\left(\frac{\partial f}{\partial \boldsymbol{x}_{\text{opt}}}\right)_z &: z, \boldsymbol{x}_{\text{opt}} \mapsto \left(\frac{\partial \zeta}{\partial \boldsymbol{x}_{\text{opt}}}\right)_z \left(\frac{\partial f}{\partial \zeta}\right)_{\boldsymbol{x}_{\text{opt}}} + \left(\frac{\partial f}{\partial \boldsymbol{x}_{\text{opt}}}\right)_\zeta \\ &= \frac{g(\zeta(z, \boldsymbol{x}_{\text{opt}}), \boldsymbol{x}_{\text{opt}})}{2(1 - \lambda|B|)^{3/2}} \left(\frac{\partial \lambda |B|}{\partial \boldsymbol{x}_{\text{opt}}}\right)_z + \frac{(\partial g / \partial \boldsymbol{x}_{\text{opt}})_z}{(1 - \lambda|B|)^{1/2}}\end{aligned} \tag{16}$$

Recalling that

$$\zeta(z = -1, \boldsymbol{x}_{\text{opt}}) = a(\boldsymbol{x}_{\text{opt}}) \tag{17}$$
$$\zeta(z = 1, \boldsymbol{x}_{\text{opt}}) = b(\boldsymbol{x}_{\text{opt}}) \tag{18}$$

and differentiating the bounce point constraint involving $a(\boldsymbol{x}_{\text{opt}})$, i.e. $\lambda |B|(a(\boldsymbol{x}_{\text{opt}}), \boldsymbol{x}_{\text{opt}}) = 1$, and likewise with the constraint involving $b(\boldsymbol{x}_{\text{opt}})$, with respect to $\boldsymbol{x}_{\text{opt}}$ gives the following constraint at $z = \pm 1$.

$$\left[\left(\frac{\partial \zeta}{\partial \boldsymbol{x}_{\text{opt}}}\right)_z \left(\frac{\partial}{\partial \zeta}\right)_{\boldsymbol{x}_{\text{opt}}} + \left(\frac{\partial}{\partial \boldsymbol{x}_{\text{opt}}}\right)_\zeta\right] \lambda |B| = 0 \tag{19}$$

Thus, (16) is integrable. Moreover, differentiation of the quadrature with respect to the controllable parameters remains a spectrally accurate approximation of the derivative of the integral. Likewise, because the quadrature nodes are sufficiently far from the singularity, the quantity I in (12) is "atom"-ically differentiable. Therefore, we do not expect catastrophic cancellation in double-precision to prevent accurate removal of the non-integrable behavior.

3 Roots of polynomials

In this section, we discuss how we ensure differentiability of (6) when one of the intermediate operations involves computing the bounce points in (1) as the roots of a nonlinear equation.

A smooth perturbation to the controllable parameters can lead to a singular perturbation in a bounce orbit. That is a non-differentiable change with respect to the controllable parameters. These occur, for example, at the bounce orbits with pitch angles such that bounce points are close to local maxima of $|B|$. Typically, optimization objectives are formulated to integrate over a pitch angle coordinate, so these non-differentiable changes do not have measurable contribution to the true continuous system.[1] However, they will have measurable contribution to the discretized system.[2]

To resolve this, a robust approach is to first find the local maxima of $|B|$ in each region, then compute the quadrature over the pitch angles with open rules that have sufficient distance in the pitch angle coordinate from

[1] We refer to the continuous system as the system where the objective for the bounce integrals is computed at an infinite number of pitch angles, and likewise for the discretization of the other coordinates.

[2] To explain this point in more detail, note that if the computation is discretized such that a singular perturbation is possible, e.g. if a bounce point is found which lies exactly on a local maxima, then, because the corresponding integral has finite measure in the discretized system, the objective would not be differentiable. Likewise, if the bounce point is very close to the maxima, in particular within a few times the machine precision of the computer, then the derivative will be inaccurate.

these singular points.[3] In practice, it is also useful to regularize the discretized system to better approximate the continuous one. For that approach, when defining the derivative of the roots of a polynomial with respect to its coefficients, instead of relying on AD to differentiate through the many unstable operations that produce the root iteratively from given coefficients, we perform a regularized differentiation through the fixed point of the iteration using the implicit function theorem.

In this case, using the implicit function theorem also provides a physically interpretable mechanism to regularize (2) to ensure differentiability $(\partial\zeta/\partial\boldsymbol{x}_{\text{opt}})_z$, which encodes how the bounce points change with respect to the optimizable parameters. Any time a bounce point lies at a bifurcation where a perturbation to the controllable parameters will move the location of the bounce point significantly, the derivative

$$\left(\frac{\partial\lambda|B|}{\partial\zeta}\right)_{\boldsymbol{x}_{\text{opt}}} = \lambda\left(\frac{\partial|B|}{\partial\zeta}\right)_{\boldsymbol{x}_{\text{opt}}} \tag{20}$$

which requires inversion to compute $(\partial\zeta/\partial\boldsymbol{x}_{\text{opt}})_z$ in the implicit function theorem (19) will tend to zero. When computing $(\partial\zeta/\partial\boldsymbol{x}_{\text{opt}})_z$, we regularize (20) with some small number $\sim 10^{-12}$ near machine precision, which is large enough that the subsequent operations that build the objective's derivative via chain rule will not suffer from loss of numerical precision.

Physically, this corresponds to computing the derivative of the objective as if the bounce point was close to the local maxima instead of exactly on it. This can be interpreted as if we had discretized the system such that the chosen pitch angle yielded a bounce point slightly below the local maxima, or equivalently as if we are smoothing the discretized system so that any observed non-differentiable samples are recognized to have zero measure in the continuous system.

4 Finite-differencing

In this section, we tabulate comparisons with finite-differencing. As AD computes the exact derivative of the discretized problem to the accuracy allowed by double-precision arithmetic, and we have shown the discretized problem is a spectrally accurate approximation, we assume the AD derivative is the true derivative. Note that we have validated our AD differentiates bounce averages with sufficient accuracy using the model in Appendix C of the main text.

In our analysis, the error in the finite-difference derivative is minimized using a step size of $h = 10^{-11}$. For example, with an effective ripple objective $f\colon \mathbb{R}^{4074} \to \mathbb{R}^{10}$ on a W7-X stellarator equilibrium, the maximum error in any element of the Jacobian matrix is at least 10^{-4} with a step size of $h = 10^{-11}$.

$$\inf_{h\in\mathbb{R}} \sup_j \left\| \frac{\partial f}{\partial \boldsymbol{x}_{\text{opt}}}\boldsymbol{e}_j - \frac{f(\boldsymbol{x}_{\text{opt}} + h\boldsymbol{e}_j) - f(\boldsymbol{x}_{\text{opt}})}{h} \right\|_\infty \geq 10^{-4} \tag{21}$$

In table 1, we show how the finite-difference derivative converges to the automatic derivative.

step size h	error
10^{-2}	278.70
10^{-4}	0.30
10^{-6}	0.15
10^{-8}	0.03
10^{-9}	0.002
10^{-10}	1.74×10^{-4}
10^{-11}	6.56×10^{-5}
10^{-12}	3.46×10^{-3}
10^{-13}	0.02

Table 1: Finite-difference error for the derivative with respect to a randomly selected optimizable parameter. The first-order convergence visible after step size 10^{-8} is halted by catastrophic cancellation.

References

[1] Daniel Johnson et al. "Software-based automatic differentiation is flawed". *Journal of Computational Physics* 542 (2025), p. 114319. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2025.114319.

[3] The AD tool should be informed to not differentiate the task of finding these quadrature points for the integration in the pitch angle coordinate.