Differential programming demo

- Background
- Using Zygote
- Simple functions
- Toy problems

Intro (Wiki)

- Code as composition of simple elementary operations and functions
- Accurate to working precision

Fundamental to AD is the decomposition of differentials provided by the chain rule. For the simple composition

$$y = f(g(h(x))) = f(g(h(w_0))) = f(g(w_1)) = f(w_2) = w_3$$

$$w_0 = x$$
$$w_1 = h(w_0)$$
$$w_2 = g(w_1)$$
$$w_3 = f(w_2) = y$$

the chain rule gives

$$\frac{dy}{dx} = \frac{dy}{dw_2}\frac{dw_2}{dw_1}\frac{dw_1}{dx} = \frac{df(w_2)}{dw_2}\frac{dg(w_1)}{dw_1}\frac{dh(w_0)}{dx}$$

1. **forward accumulation** computes the recursive relation: $\frac{dw_i}{dx} = \frac{dw_i}{dw_{i-1}}\frac{dw_{i-1}}{dx}$ with $w_3 = y$, and,
2. **reverse accumulation** computes the recursive relation: $\frac{dy}{dw_i} = \frac{dy}{dw_{i+1}}\frac{dw_{i+1}}{dw_i}$ with $w_0 = x$.

Zygote

DON'T UNROLL ADJOINT: DIFFERENTIATING SSA-FORM PROGRAMS

Michael J Innes [1]

ABSTRACT

This paper presents reverse-mode algorithmic differentiation (AD) based on source code transformation, in particular of the Static Single Assignment (SSA) form used by modern compilers. The approach can support control flow, nesting, mutation, recursion, data structures, higher-order functions, and other language constructs, and the output is given to an existing compiler to produce highly efficient differentiated code. Our implementation is a new AD tool for the Julia language, called Zygote, which presents high-level dynamic semantics while transparently compiling adjoint code under the hood. We discuss the benefits of this approach to both the usability and performance of AD tools.

A supervised learning approach

Algorithm 2: basic learning framework

Given labeled data

model $Y = \mathcal{F}$

while $1 \leq i \leq epochs$ **do**

Predict: $\hat{Y}(\theta) = \mathcal{F}(\theta)$

Minimize: $L(\hat{Y}(\theta), Y)$

e.g $\theta_{i+1} = \theta_i - lr * \nabla_\theta L(\theta)$

end

Result: $\hat{\theta}$

Pendulum Example

$$\frac{dx}{dt} = v$$

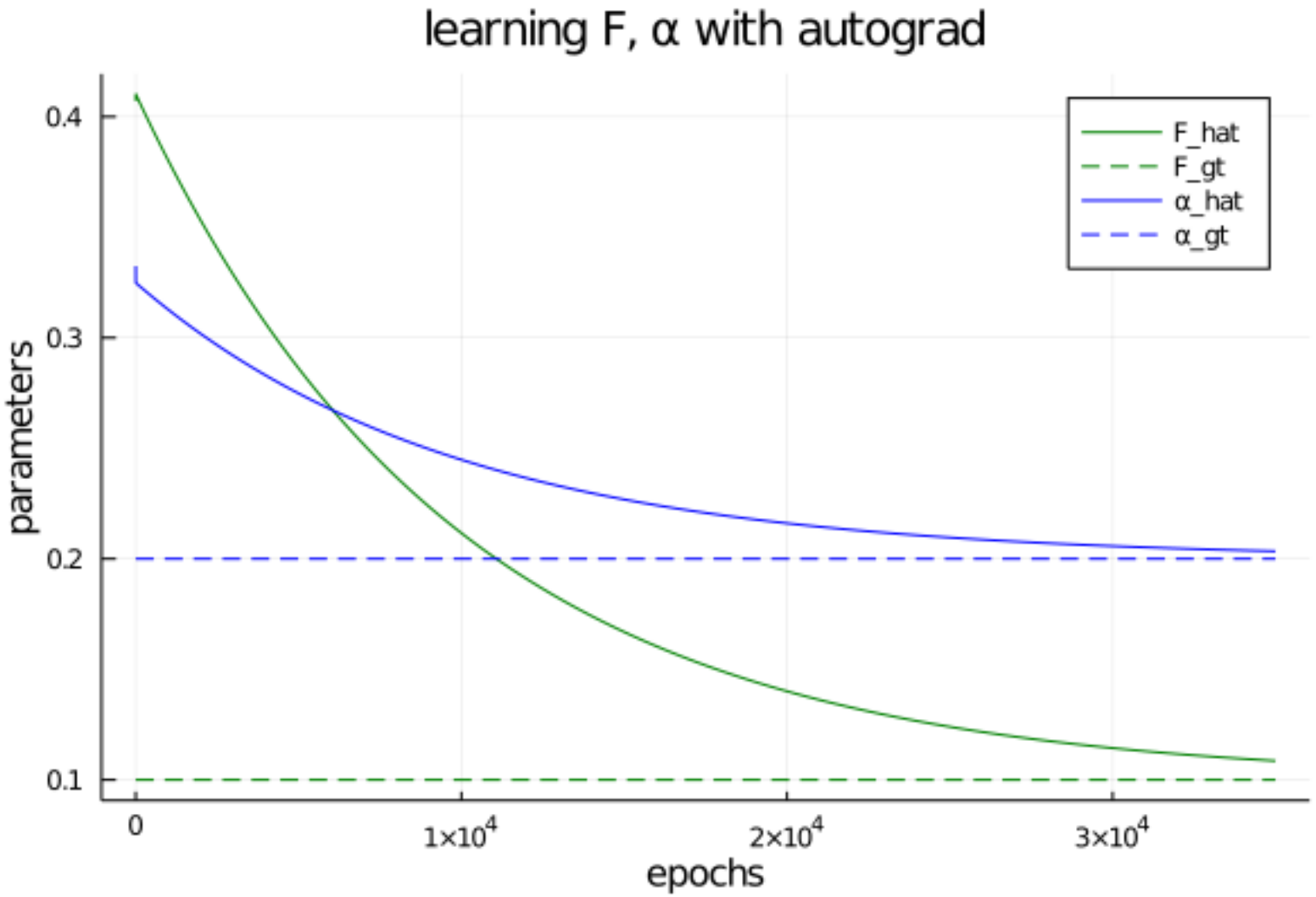
$$\frac{dv}{dt} = -\sin(x) - \alpha v + F$$

$$\{\hat{x}(\theta), \hat{v}(\theta)\} = \int_0^\tau RHS dt$$

$$L(\hat{x}(\theta), \hat{v}(\theta), x, v) = ||(\hat{x}(\theta) - x, \hat{v}(\theta) - v)||_2$$

$$\theta = \theta - lr * \nabla L$$

Two parameters



Analytic method (two params)