Hessians and Higher-Order Derivatives

SIAM CSE 2021

Hessians

Options for computing second derivatives of scalar functions $\mathbb{R}^N \Rightarrow \mathbb{R}$ by applying AD twice:

- Forward over forward ($O(N^2)$)
- Forward over reverse ($O(N)$)
- Reverse over forward ($O(N)$)
- Reverse over reverse ($O(N)$)

Hessians

Options for computing second derivatives of scalar functions $\mathbb{R}^N \Rightarrow \mathbb{R}$ by applying AD twice:

- ~~Forward over forward ($O(N^2)$)~~
- Forward over reverse ($O(N)$)
- Reverse over forward ($O(N)$)
- ~~Reverse over reverse ($O(N)$)~~

Hessians

Options for computing second derivatives of scalar functions $\mathbb{R}^N \Rightarrow \mathbb{R}$:

- ~~Forward over forward~~ $\cancel{(O(N^2))}$
- Forward over reverse $(O(N))$
- Reverse over forward $(O(N))$
- ~~Reverse over reverse~~ $\cancel{(O(N))}$
- Edge-pushing algorithm
- Univariate Taylor series

Hessians: practicalities

- Not all AD tools can be applied multiple times
- Some AD tools can be applied twice, but only if invoked with additional options
- Successive application of AD rarely exploits symmetry
- Some AD tools have direct support for second- and higher-order derivatives, without needing to be applied multiple times
- Hv using forward+reverse is cheap

Hessians: edge-pushing algorithm

- Gower, Robert M., and Margarida P. Mello. "A new framework for the computation of Hessians." OMS 27(2): 251-273.
- Compute reverse-mode gradient and apply reverse mode to the (implicit) gradient graph, exploiting symmetry at every step
- Pathological counter-examples exist, but in practice edge pushing is flop efficient
- More bookkeeping than other methods; performance is implementation-dependent

High-order derivatives

- As differentiation order increases, benefits of reverse mode decrease
- As differentiation order increases, benefits of exploiting symmetry increase
- Symmetry-aware high-order tensors require extensive bookkeeping
- Univariate Taylor series offer an alternative

High-order derivatives: univariate Taylor series

- Instead of computing mixed partials such as $\frac{\partial^2 f}{\partial x \partial y}$ in addition to $\frac{\partial^2 f}{\partial x^2}$, $\frac{\partial^2 f}{\partial y^2}$, compute only univariate derivatives, but introduce new variables such as $u = x + y$
- Use interpolation to recover $\frac{\partial^2 f}{\partial x \partial y}$, etc. from $\frac{\partial^2 f}{\partial x^2}$, $\frac{\partial^2 f}{\partial y^2}$, $\frac{\partial^2 f}{\partial u^2}$
- Pros: highly structured, easily parallelized
- Cons: non-trivial setup, interpolation roundoff
- Available in ADOL-C, Rapsodia

High-order derivatives: sparsity

- Can employ coloring-based Hessian compression, analogous to Jacobians
- Symmetry in Hessians potentially reduces number of colors, since only H_{ij} **or** H_{ji} is needed
- Cost of univariate Taylor series method proportional to number of Hessian / HOD-tensor entries required