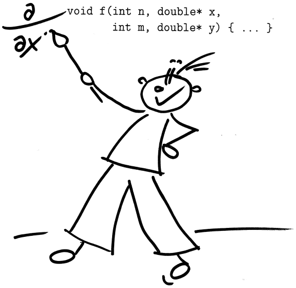


Chain Rule Differentiation is NP-Complete

Uwe Naumann

RWTH Aachen University, Aachen, Germany

Outline

History

Revised Proof

Conclusion

Outline

History

Revised Proof

Conclusion

... path elimination: W. Baur, V. Strassen: *The Complexity of Partial Derivatives.* Theoretical Computer Science, 1983.

... vertex elimination: A. Griewank, S. Reese: *On the Calculation of Jacobian Matrices by the Markowitz Rule.* AD @ Breckenridge, 1991.

... edge elimination: U. N.: *Elimination Techniques for Cheap Jacobians.* AD @ Nice, 2000.

... face elimination: U. N.: *Optimal Accumulation of Jacobian Matrices by Elimination Methods on the Dual Computational Graph.* Mathematical Programming, 2004.

$\not\Rightarrow$ U. N.: *Optimal Jacobian accumulation is NP-complete.* Mathematical Programming, 2008.

Outline

History

Revised Proof

Conclusion

Chain Rule of Differentiation at its Simplest

(Layered Composite) Primal

$$F : \boldsymbol{R}^n \rightarrow \boldsymbol{R}^m \quad : \quad \mathrm{x} \mapsto \mathrm{y} = F(\mathrm{x}) = F_q(F_{q-1}(\ldots F_1(\mathrm{x}) \ldots)) \tag{1}$$

Jacobian

$$F' \equiv \frac{dF}{d\mathrm{x}} = \prod_{i=1}^{q} F_i' \equiv F_q' \cdot F_{q-1}' \cdot \ldots \cdot F_1' \, .$$

Hessian (index notation; ... and higher derivative tensors)

$$[F'']_{\delta,\alpha_1,\alpha_2} = \sum_{j=1}^{q} \left(\left[\prod_{i=j+1}^{q} F_i' \right]_{\delta,\gamma} [F_j'']_{\gamma,\beta_1,\beta_2} \left[\prod_{k=1}^{j-1} F_k' \right]_{\beta_1,\alpha_1} \left[\prod_{k=1}^{j-1} F_k' \right]_{\beta_2,\alpha_2} \right)$$

- U.N.: *On the Computational Complexity of the Chain Rule of Differential Calculus*. arXiv:2107.05355.
- U.N.: *On Sparse Matrix Chain Products*. SIAM CSC 2020.

Complexity Analysis

Definition (Chain Rule Differentiation)

INSTANCE: A composite function as in Equation (1) with given elemental derivatives up to order p and a positive integer K.

QUESTION: Can the p-th derivative of F be computed with at most K fma[1] operations?

Theorem

Chain Rule Differentiation *is NP-complete.*

Numerical (assuming infinite precision) validation of given solution against any feasible evaluation of $F^{[p]} \equiv \frac{d^p F}{dx^p}$

[1] fused multiply-add

Definition

INSTANCE: A collection $C = \{C_\nu \subseteq A : \nu = 1, \ldots, |C|\}$ of subsets $C_\nu = \{c_i^\nu : i = 1, \ldots, |C_\nu|\}$ of a finite set A and a positive integer K.

QUESTION: Is there a sequence $u_i = s_i \cup t_i$ for $i = 1, \ldots, k$ of $k \leq K$ union operations, where each s_i and t_i is either $\{a\}$ for some $a \in A$ or u_j for some $j < i$, such that s_i and t_i are disjoint for $i = 1, \ldots, k$ and such that for every subset $C_\nu \in C$, $\nu = 1, \ldots, |C|$, there is some u_i, $1 \leq i \leq k$, that is identical to C_ν?

Example: $A = \{a_1, a_2, a_3, a_4\}$, $C = \{\{a_1, a_2\}, \{a_2, a_3, a_4\}, \{a_1, a_3, a_4\}\}$, $K = 4$
$\Rightarrow$ YES as $C_1 = u_1 = \{a_1\} \cup \{a_2\}$; $u_2 = \{a_3\} \cup \{a_4\}$; $C_2 = u_3 = \{a_2\} \cup u_2$; $C_3 = u_4 = \{a_1\} \cup u_2$.

Lemma

EC *is NP-complete.*

- M.R. Garey, D.S. Johnson: *Computers and Intractability: A Guide to the Theory of NP-Completeness.* Freeman, 1979.

Proof

Reduction from EC

Given an instance $I_1 = (A, C, K)$ of EC

$\Rightarrow$ bijection $A \leftrightarrow \tilde{A}$
(requires $|A|$ mutually distinct primes to avoid $42 = 6 \cdot 7 = 3 \cdot 14 ...$)

$\Rightarrow$ bijection $C \leftrightarrow \tilde{C}$

$\Rightarrow$ extended instance $I_2 = (\tilde{A} \cup \tilde{B}, \tilde{C}, K + |\tilde{B}|)$ of EC, s.t. same cardinalities of $\tilde{C}_j$ through addition of unique primes not in $\tilde{A}$.

$\Rightarrow$ Solution for I_2 implies solution of I_1.

Example:

$$A = \{a_1, a_2, a_3, a_4\} \Rightarrow \tilde{A} = \{2, 3, 5, 7\}$$
$$\tilde{B} = \{11\}$$
$$C = \{\{a_1, a_2\}, \{a_2, a_3, a_4\}, \{a_1, a_3, a_4\}\} \Rightarrow \tilde{C} = \{\{2, 3, 11\}, \{3, 5, 7\}, \{2, 5, 7\}\}$$
$$K + |\tilde{B}| = K + 1 = 5\ .$$

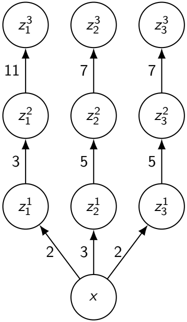
From

$$F_1 : \boldsymbol{R} \to \boldsymbol{R}^{|\tilde{C}|} : \qquad z_1 = F_1(x) : \quad z_j^1 = \frac{\tilde{c}_1^j}{p!} \cdot x^p$$

$$F_i : \boldsymbol{R}^{|\tilde{C}|} \to \boldsymbol{R}^{|\tilde{C}|} : \quad z_i = F_i(z_{i-1}) : \; z_j^i = \tilde{c}_i^j \cdot z_j^{i-1}$$

follows (by induction on order)

$$F^{[p]} = \prod_{i=2}^{q} F_i^{[1]} \cdot F_1^{[p]} ,$$



$$\text{where} \quad F_i^{[1]} = F_i' = (d_{j,k}^i) \in \boldsymbol{R}^{|\tilde{C}|\times|\tilde{C}|} , s.t. \; d_{j,k}^i = \begin{cases} \tilde{c}_i^j & \text{if } j = k \\ 0 & \text{otherwise} \end{cases} ,$$

$$\text{and} \quad F_1^{[p]} = \left(\tilde{c}_1^j\right) \in \boldsymbol{R}^{|\tilde{C}|} = \boldsymbol{R}^{|\tilde{C}|\times 1\times\ldots(p \text{ times})\ldots\times 1} .$$

Proof

Example

Software and Tools for Computational Engineering

RWTH AACHEN UNIVERSITY

$$F^{[p]} = F_3^{[1]} \cdot F_2^{[1]} \cdot F_1^{[p]} = \begin{pmatrix} 11 & & \\ & 7 & \\ & & 7 \end{pmatrix} \cdot \begin{pmatrix} 3 & & \\ & 5 & \\ & & 5 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 3 \\ 2 \end{pmatrix}$$

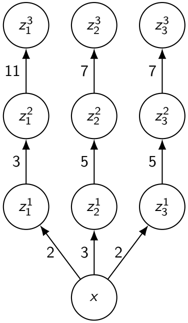
$$\frac{dz_1^2}{dx} = \frac{dz_1^2}{dz_1^1} \cdot \frac{d^p z_1^1}{dx^p} = 3 \cdot 2 = 6$$

$$\frac{dz_2^3}{dz_2^1} = \frac{dz_3^3}{dz_3^1} = \frac{dz_3^3}{dz_3^2} \cdot \frac{dz_3^2}{dz_3^1} = 7 \cdot 5 = 35$$

$$F_1^{[p]} = \tilde{b}_1 \cdot \left(\frac{dz_1^2}{dz_1^1} \cdot \frac{d^p z_1^1}{dx^p} \right) = 11 \cdot 6 = 66$$

$$F_2^{[p]} = \frac{dz_2^3}{dz_2^1} \cdot \frac{d^p z_2^1}{dx^p} = 35 \cdot 3 = 105$$

$$F_3^{[p]} = \frac{dz_3^3}{dz_3^1} \cdot \frac{d^p z_3^1}{dx^p} = 35 \cdot 2 = 70\ .$$



Outline

History

Revised Proof

Conclusion

Conclusion

- theoretically relevant BUT inappropriate granularity
- tensor-free elimination techniques
  - U.N.: *Optimization of Generalized Jacobian Chain Products without Memory Constraints*. arXiv:2003.05755. Under review.
  - U.N.: *Hessian Chain Bracketing*. arXiv:2103.09480. Under review.
- AD mission planning ($\Rightarrow$ ADMission software)