Supporting Information for "Arbitrary-Order Derivatives of Quantum Chemical Methods via Automatic Differentiation"

Adam S. Abbott, Boyi Z. Abbott, Justin M. Turney, and Henry F. Schaefer III*

Center for Computational Quantum Chemistry, University of Georgia, Athens, GA, 30602

E-mail: ccq@uga.edu

Approximating memory use of partial derivative computations

The multivariate Faà di Bruno formula generalizes the chain rule for higher order derivatives in the case of a composite function in which the inner function is multivariate and outer function is univariate. Let $G = g(x_1, ..., x_n)$ and consider the function $F(G)$. Then,

$$\frac{\partial^n}{\partial x_1 \cdots \partial x_n} F(G) = \sum_{\pi \in P} F^{(|\pi|)}(G) \cdot \prod_{B \in \pi} \frac{\partial^{|B|} G}{\prod_{j \in B} \partial x_j}$$

where P is the set of all partitions of the set $\{1, 2, ..., n\}$ and π is a particular partition in P. B is a particular block in the partition π, so $B \in \pi$ runs through all blocks of partition π. Finally, $|\pi|$ and $|B|$ denote the cardinality of π and B, respectively.

The first partial derivative is just the familiar chain rule, and when produced according to the above equation, it is trivial since P in this case has just one partition, with one block.

$$\frac{\partial}{\partial x_1} F(G) = F'(G) \cdot \frac{\partial G}{\partial x_1}$$

In the case of two partial derivatives $\frac{\partial^2}{\partial x_1 \partial x_2}$, P is the set of all partitions of the set $\{1, 2\}$, so $P = \{\{\{1\}, \{2\}\}, \{\{1, 2\}\}\}$ which by counting the unique blocks that appear in each partition, indicates that our final result should involve the derivatives $\frac{\partial}{\partial x_1}$, $\frac{\partial}{\partial x_2}$, $\frac{\partial^2}{\partial x_2 \partial x_1}$. Once worked out, we find that,

$$\frac{\partial^2}{\partial x_2 \partial x_1} F(G) = F''(G) \cdot \frac{\partial G}{\partial x_2} \cdot \frac{\partial G}{\partial x_1} + F'(G) \cdot \frac{\partial^2 G}{\partial x_2 \partial x_1}$$

In the case of three partial derivatives $\frac{\partial^3}{\partial x_1 \partial x_2 \partial x_3}$, P contains 5 partitions, with a total of 7 unique blocks.

$$P = \{\{\{1\}, \{2\}, \{3\}\}, \{\{1, 2\}, \{3\}\}, \{\{1, 3\}, \{2\}\}, \{\{1\}, \{2, 3\}\}, \{\{1, 2, 3\}\}\}$$

Without even working out the result, we know by inspecting the unique blocks of the parti-

tions that we will require the derivatives: $\frac{\partial}{\partial x_1}$, $\frac{\partial}{\partial x_2}$, $\frac{\partial}{\partial x_3}$, $\frac{\partial^2}{\partial x_1 \partial x_2}$, $\frac{\partial^2}{\partial x_1 \partial x_3}$ $\frac{\partial^2}{\partial x_2 \partial x_3}$, $\frac{\partial^3}{\partial x_1 \partial x_2 \partial x_3}$ The full derivative expansion is:

$$\begin{aligned}\frac{\partial^3}{\partial x_1 \partial x_2 \partial x_3} F(G) = F'(G) &\frac{\partial^3 G}{\partial x_1 \partial x_2 \partial x_3} \\ &+ F''(G) \left(\frac{\partial G}{\partial x_1} \frac{\partial^2 G}{\partial x_2 \partial x_3} + \frac{\partial G}{\partial x_2} \frac{\partial^2 G}{\partial x_1 \partial x_3} + \frac{\partial G}{\partial x_3} \frac{\partial^2 G}{\partial x_1 \partial x_2} \right) \\ &+ F'''(G) \frac{\partial G}{\partial x_1} \frac{\partial G}{\partial x_2} \frac{\partial G}{\partial x_3}\end{aligned}$$

In the context of estimating peak memory use, the vast majority of memory use comes from the two-electron integrals (TEIs), which for n basis functions are collected in an array of shape (n, n, n, n). Each partial derivative of the two-electron integrals w.r.t. Cartesian coordinates of the molecule $\frac{\partial^n}{\partial x_1 \cdots \partial x_n}$ will likewise be packed into an array of shape (n, n, n, n). Therefore, counting the number of partial derivatives which appear in the above expansions gives you the number of arrays of size (n, n, n, n) which will be held in memory. For example, the second partial derivative of an energy computation will at some point involve 3 TEI derivative arrays in addition to the standard TEIs, so this will be 4 arrays of shape (n, n, n, n).

There are additional considerations to predict the memory use of a partial derivative evaluation with our scheme. First, our Python library requires about 1 gigabyte (GB) of memory just to import due to it containing an interface to a rather large Libint library. Second, our Hartree-Fock implementation uses a JIT-compiled construction of the Fock matrix (specifically, the contraction of the density matrix with the two electron integrals to form the coulomb and exchange terms). The JIT compilation causes a further intermeidate to be formed for performance reasons, which doubles memory use. This can be turned off, but if one goes on to do MP2 and CCSD(T), which require just as much memory for the TEI transformation.

Therefore, if N is the number of GB required to store a single (n, n, n, n) TEI or TEI derivative array, and k is the number of partial derivatives appearing in the multivariate Faà

di Bruno expansions, the estimated peak memory use is

$$2 \cdot ((k+1) \cdot N) + 1$$

where the factor of two comes from the JIT-compiled Fock matrix formation step, the $k+1$ is accounting for the standard (non-differentiated) TEI's, and the final addition of 1 is from standard Python overhead of importing the package.

For a concrete example, the H_2CO Hartree Fock cc-pVTZ system is 100 Cartesian basis functions, so the size of the TEI array or any TEI partial derivative array is $100^4 * 8$ bytes, or 0.8 GB. The number of partial derivative arrays for first, second, and third derivatives will be 1, 3, and 7 respectively, as we found earlier. For fourth, fifth, and sixth derivatives, there are 15, 31, and 63 partial derivatives which appear in the multivariate Faá Di Bruno expansion, respectively, which tells us how many derivative TEI arrays will appear in the computation. We can then compute the predicted peak memory use and compare to the actual results, assuming of course the contribution from other stored arrays is negligible.

Derivative	Number of Partial Derivatives	Predicted Peak GB	Actual Peak GB
Gradient	1	4.2	4.6
Hessian	3	7.4	7.8
Cubic	7	13.8	14.1
Quartic	15	26.6	26.7
Quintic	31	52.2	51.9
Sextic	63	103.4	102.2

We note that the series 1,3,7,15,31,63 obeys the simple recursive formula $k_1 = 1$, $k_{n+1} = 2 \cdot k_{n-1} + 1$.

Performance Data

Each nth order partial derivative value corresponds to the derivative $\frac{\partial^n E}{\partial z_1^n}$ in units of Hartrees/Bohr^n where z_1 is the z-coordinate on the first atom for the following geometries (in Bohr). We note that each partial derivative takes the same amount of time, though diagonal partial derivatives such as $\frac{\partial^n E}{\partial z_1^n}$ require the least amount of integral derivatives that are needed. Each method (HF, MP2, CCSD(T)) used the same geometry for a given basis, which is the result of optimizing at the CCSD(T) level of theory.

```
H2O  cc-pVDZ
O          0.000000436927    0.128648202865    0.000000000000
H         -1.417747097703   -1.020872149778    0.000000000000
H          1.417740163355   -1.020868187071    0.000000000000
H2O  cc-pVTZ
O         -0.000007070942    0.125146536460    0.000000000000
H         -1.424097055410   -0.993053750648    0.000000000000
H          1.424209276385   -0.993112599269    0.000000000000
N2  cc-pVDZ
N  0.000000000000    0.000000000000    -1.056883600544
N  0.000000000000    0.000000000000     1.056883600544
N2  cc-pVTZ
N   0.000000000000    0.000000000000    -1.040129860737
N   0.000000000000    0.000000000000     1.040129860737
H2CO  cc-pVDZ
C          0.000000000000    0.000000000000   -1.148089555770
O          0.000000000000    0.000000000000    1.148667224612
H         -0.000000000000    1.785840975984   -2.280038383525
H          0.000000000000   -1.785840975984   -2.280038383525
H2CO  cc-pVTZ
C          0.000000000000    0.000000000000   -1.141817200432
```

O	0.000000000000	0.000000000000	1.138747061525
H	−0.000000000000	1.762942010951	−2.238660220220
H	0.000000000000	−1.762942010951	−2.238660220220

Table S1: Partial derivatives performance data for H_2O

Method	Basis	Deriv	Result	Timing (s)	Memory (GB)
Hartree Fock	cc-pvdz	gradient	0.0	5	1.511
Hartree Fock	cc-pvdz	hessian	0.02625379	8	1.567
Hartree Fock	cc-pvdz	cubic	0.0	15	1.663
Hartree Fock	cc-pvdz	quartic	1.0975729	32	1.825
Hartree Fock	cc-pvdz	quintic	0.0	74	2.026
Hartree Fock	cc-pvdz	sextic	-17.52607173	203	2.513
MP2	cc-pvdz	gradient	0.0	6	1.527
MP2	cc-pvdz	hessian	0.00156076	10	1.573
MP2	cc-pvdz	cubic	0.0	19	1.67
MP2	cc-pvdz	quartic	1.07697246	39	1.816
MP2	cc-pvdz	quintic	0.0	89	2.064
MP2	cc-pvdz	sextic	-17.10615297	241	2.744
CCSD(T)	cc-pvdz	gradient	0.0	36	1.724
CCSD(T)	cc-pvdz	hessian	0.00013651	73	1.933
CCSD(T)	cc-pvdz	cubic	0.0	161	2.284
CCSD(T)	cc-pvdz	quartic	1.07991147	382	3.18
CCSD(T)	cc-pvdz	quintic	0.0	969	5.998
CCSD(T)	cc-pvdz	sextic	-17.15941458	2809	19.061
Hartree Fock	cc-pvtz	gradient	0.0	12	2.06
Hartree Fock	cc-pvtz	hessian	0.02426231	25	2.652
Hartree Fock	cc-pvtz	cubic	0.0	55	3.825
Hartree Fock	cc-pvtz	quartic	1.14709206	134	6.141
Hartree Fock	cc-pvtz	quintic	0.0	373	10.745
Hartree Fock	cc-pvtz	sextic	-18.54498022	1048	19.926
MP2	cc-pvtz	gradient	0.0	14	2.057
MP2	cc-pvtz	hessian	0.00024441	28	2.652
MP2	cc-pvtz	cubic	0.0	62	3.826
MP2	cc-pvtz	quartic	1.11209909	151	6.142
MP2	cc-pvtz	quintic	0.0	375	11.037
MP2	cc-pvtz	sextic	-17.81590525	1112	19.929
CCSD(T)	cc-pvtz	gradient	0.0	59	2.365
CCSD(T)	cc-pvtz	hessian	1.04e-05	131	3.48
CCSD(T)	cc-pvtz	cubic	0.0	295	5.753
CCSD(T)	cc-pvtz	quartic	1.11552515	682	9.788
CCSD(T)	cc-pvtz	quintic	0.0	1704	19.561
CCSD(T)	cc-pvtz	sextic	-17.86693418	4621	37.69

Table S2: Partial derivatives performance data for N_2

Method	Basis	Deriv	Result	Timing (s)	Memory (GB)
Hartree Fock	cc-pvdz	gradient	-0.13575134	5	1.53
Hartree Fock	cc-pvdz	hessian	1.51493309	8	1.612
Hartree Fock	cc-pvdz	cubic	5.67655058	15	1.748
Hartree Fock	cc-pvdz	quartic	17.49151553	29	2.02
Hartree Fock	cc-pvdz	quintic	51.81217508	65	2.31
Hartree Fock	cc-pvdz	sextic	143.43064277	161	2.969
MP2	cc-pvdz	gradient	0.02516452	6	1.533
MP2	cc-pvdz	hessian	1.36731575	11	1.611
MP2	cc-pvdz	cubic	5.84397824	19	1.75
MP2	cc-pvdz	quartic	17.62660051	37	2.02
MP2	cc-pvdz	quintic	51.12189468	83	2.376
MP2	cc-pvdz	sextic	141.92170812	206	3.195
CCSD(T)	cc-pvdz	gradient	-0.00182839	46	1.739
CCSD(T)	cc-pvdz	hessian	1.44855411	95	1.949
CCSD(T)	cc-pvdz	cubic	5.69461741	208	2.305
CCSD(T)	cc-pvdz	quartic	17.69872414	481	3.218
CCSD(T)	cc-pvdz	quintic	51.89103652	1218	6.082
CCSD(T)	cc-pvdz	sextic	139.48286122	3420	19.195
Hartree Fock	cc-pvtz	gradient	-0.11316389	13	2.252
Hartree Fock	cc-pvtz	hessian	1.56725202	25	3.04
Hartree Fock	cc-pvtz	cubic	5.73697081	55	4.596
Hartree Fock	cc-pvtz	quartic	17.64393845	125	7.683
Hartree Fock	cc-pvtz	quintic	56.9763774	339	13.825
Hartree Fock	cc-pvtz	sextic	204.06021251	897	26.092
MP2	cc-pvtz	gradient	0.02027611	15	2.253
MP2	cc-pvtz	hessian	1.39803794	29	3.037
MP2	cc-pvtz	cubic	5.91082225	60	4.619
MP2	cc-pvtz	quartic	17.93602461	142	7.683
MP2	cc-pvtz	quintic	53.74918513	371	14.388
MP2	cc-pvtz	sextic	187.48080839	919	26.09
CCSD(T)	cc-pvtz	gradient	-0.00451517	71	2.646
CCSD(T)	cc-pvtz	hessian	1.4735436	163	4.369
CCSD(T)	cc-pvtz	cubic	5.78386961	377	7.237
CCSD(T)	cc-pvtz	quartic	18.05590271	908	13.522
CCSD(T)	cc-pvtz	quintic	54.10024003	2293	25.669
CCSD(T)	cc-pvtz	sextic	183.63134711	6300	49.971

Table S3: Partial derivatives performance data for H_2CO

Method	Basis	Deriv	Result	Timing (s)	Memory (GB)
Hartree Fock	cc-pvdz	gradient	-0.04663682	8	1.637
Hartree Fock	cc-pvdz	hessian	1.0017135	16	1.855
Hartree Fock	cc-pvdz	cubic	2.5516519	33	2.212
Hartree Fock	cc-pvdz	quartic	7.5452544	80	2.789
Hartree Fock	cc-pvdz	quintic	31.66771523	210	3.404
Hartree Fock	cc-pvdz	sextic	103.89265322	562	5.328
MP2	cc-pvdz	gradient	-0.00024591	10	1.637
MP2	cc-pvdz	hessian	0.95653754	20	1.809
MP2	cc-pvdz	cubic	2.7183241	40	2.154
MP2	cc-pvdz	quartic	8.10820552	90	2.767
MP2	cc-pvdz	quintic	31.9622877	222	3.597
MP2	cc-pvdz	sextic	89.45987355	601	5.415
CCSD(T)	cc-pvdz	gradient	-0.00206666	61	1.797
CCSD(T)	cc-pvdz	hessian	0.95827377	139	2.014
CCSD(T)	cc-pvdz	cubic	2.69027372	305	2.49
CCSD(T)	cc-pvdz	quartic	8.0135445	702	3.418
CCSD(T)	cc-pvdz	quintic	31.36844546	1685	6.467
CCSD(T)	cc-pvdz	sextic	88.4419578	4381	19.984
Hartree Fock	cc-pvtz	gradient	-0.04576046	80	4.629
Hartree Fock	cc-pvtz	hessian	1.02043268	214	7.792
Hartree Fock	cc-pvtz	cubic	2.60109703	566	14.107
Hartree Fock	cc-pvtz	quartic	7.41378113	1516	26.701
Hartree Fock	cc-pvtz	quintic	30.77945081	3904	51.871
Hartree Fock	cc-pvtz	sextic	105.99285715	11045	102.178
MP2	cc-pvtz	gradient	0.00275325	83	4.631
MP2	cc-pvtz	hessian	0.95610567	231	7.796
MP2	cc-pvtz	cubic	2.70002679	589	14.108
MP2	cc-pvtz	quartic	7.79808327	1547	26.696
MP2	cc-pvtz	quintic	31.09498435	4022	53.39
MP2	cc-pvtz	sextic	100.93064724	11164	102.178
CCSD(T)	cc-pvtz	gradient	0.00029838	170	6.994
CCSD(T)	cc-pvtz	hessian	0.95775115	447	13.287
CCSD(T)	cc-pvtz	cubic	2.66957093	1103	21.154
CCSD(T)	cc-pvtz	quartic	7.73799153	2902	47.028
CCSD(T)	cc-pvtz	quintic	30.81469245	7862	90.949
CCSD(T)	cc-pvtz	sextic	100.60116905	22741	152.98

Table S4: Energy, full Gradient, full Hessian performance data

Molecule	Method	Basis	Deriv	Timing (s)	Memory (GB)
N_2	Hartree Fock	cc-pvdz	energy	3.23	1.491
H_2O	Hartree Fock	cc-pvdz	energy	2.94	1.485
H_2CO	Hartree Fock	cc-pvdz	energy	3.91	1.531
N_2	Hartree Fock	cc-pvtz	energy	6.66	1.855
H_2O	Hartree Fock	cc-pvtz	energy	5.21	1.754
H_2CO	Hartree Fock	cc-pvtz	energy	27.38	3.041
N_2	MP2	cc-pvdz	energy	3.82	1.49
H_2O	MP2	cc-pvdz	energy	3.3	1.482
H_2CO	MP2	cc-pvdz	energy	4.62	1.532
N_2	MP2	cc-pvtz	energy	7.2	1.856
H_2O	MP2	cc-pvtz	energy	5.81	1.759
H_2CO	MP2	cc-pvtz	energy	29.07	3.039
N_2	CCSD(T)	cc-pvdz	energy	15.23	1.589
H_2O	CCSD(T)	cc-pvdz	energy	12.93	1.584
H_2CO	CCSD(T)	cc-pvdz	energy	19.73	1.645
N_2	CCSD(T)	cc-pvtz	energy	24.35	2.048
H_2O	CCSD(T)	cc-pvtz	energy	20.22	1.91
H_2CO	CCSD(T)	cc-pvtz	energy	55.37	4.608
N_2	Hartree Fock	cc-pvdz	gradient	36.31	1.571
H_2O	Hartree Fock	cc-pvdz	gradient	7.8	1.597
H_2CO	Hartree Fock	cc-pvdz	gradient	17.55	2.685
N_2	Hartree Fock	cc-pvtz	gradient	28.94	5.264
H_2O	Hartree Fock	cc-pvtz	gradient	43.28	5.591
H_2CO	Hartree Fock	cc-pvtz	gradient	551.79	31.274
N_2	MP2	cc-pvdz	gradient	9.52	1.652
H_2O	MP2	cc-pvdz	gradient	9.13	1.643
H_2CO	MP2	cc-pvdz	gradient	19.88	2.792
N_2	MP2	cc-pvtz	gradient	31.08	5.363
H_2O	MP2	cc-pvtz	gradient	41.37	5.657
H_2CO	MP2	cc-pvtz	gradient	452.92	32.134
N_2	CCSD(T)	cc-pvdz	gradient	58.5	1.819
H_2O	CCSD(T)	cc-pvdz	gradient	48.42	1.793
H_2CO	CCSD(T)	cc-pvdz	gradient	98.25	3.188
N_2	CCSD(T)	cc-pvtz	gradient	132.39	7.894
H_2O	CCSD(T)	cc-pvtz	gradient	120.91	8.379
H_2CO	CCSD(T)	cc-pvtz	gradient	998.65	51.556
N_2	Hartree Fock	cc-pvdz	hessian	16.85	2.351
H_2O	Hartree Fock	cc-pvdz	hessian	22.58	2.353
H_2CO	Hartree Fock	cc-pvdz	hessian	118.86	10.535
N_2	Hartree Fock	cc-pvtz	hessian	143.68	25.079
H_2O	Hartree Fock	cc-pvtz	hessian	292.21	37.157
H_2CO	Hartree Fock	cc-pvtz	hessian	5432.0	336.582
N_2	MP2	cc-pvdz	hessian	21.85	2.61
H_2O	MP2	cc-pvdz	hessian	24.86	2.621
H_2CO	MP2	cc-pvdz	hessian	128.69	13.536
N_2	MP2	cc-pvtz	hessian	141.21	27.648
H_2O	MP2	cc-pvtz	hessian	303.06	40.494
H_2CO	MP2	cc-pvtz	hessian	5292.0	372.114
N_2	CCSD(T)	cc-pvdz	hessian	172.48	3.627
H_2O	CCSD(T)	cc-pvdz	hessian	144.41	3.683
H_2CO	CCSD(T)	cc-pvdz	hessian	793.84	22.732
N_2	CCSD(T)	cc-pvtz	hessian	797.15	53.596
H_2O	CCSD(T)	cc-pvtz	hessian	960.49	83.79
H_2CO	CCSD(T)	cc-pvtz	hessian	15165.0	796.16