OPTIMIZATION VIA SEPARATED REPRESENTATIONS AND THE CANONICAL TENSOR DECOMPOSITION

MATTHEW J REYNOLDS[∗], GREGORY BEYLKIN[†], AND ALIREZA DOOSTAN[∗]

Abstract. We introduce a new, quadratically convergent algorithm for finding maximum absolute value entries of tensors represented in the canonical format. The computational complexity of the algorithm is linear in the dimension of the tensor. We show how to use this algorithm to find global maxima of non-convex multivariate functions in separated form. We demonstrate the performance of the new algorithms on several examples.

1. Introduction

Finding global extrema of a multivariate function is an ubiquitous task with many approaches developed to address this problem (see e.g. [19]). Unfortunately, no existing optimization method can guarantee that the results of optimization are true global extrema unless restrictive assumptions are placed on the function. Assumptions on smoothness of the function do not help since it is easy to construct an example of a function with numerous local extrema “hiding” the location of the true one. While convexity assumptions are helpful for finding global maxima, in practical applications there are many non-convex functions. For non-convex functions various randomized search strategies have been suggested and used but none can assure that the results are true global extrema (see, e.g., [27, 19]).

We propose a new approach to the problem of finding global extrema of a multivariate function under the assumption that the function has certain structure, namely, a separated representation with a reasonably small separation rank. While this assumption limits the complexity of the function, i.e. number of independent degrees of freedom in its representation, there is no restriction on its convexity.

where $s_l > 0$ are referred to as s-values. In this approximation the functions $u_j^{(l)}(x_j)$, $j = 1, \dots, d$ are not fixed in advance but are optimized in order to achieve the accuracy goal with (ideally) a minimal *separation rank* r. In (2.1) we set $x_j \in \mathbb{R}$ while noting that in general the variables x_j may be complex-valued or low dimensional vectors. Importantly, a separated representation is not a projection onto a subspace, but rather a nonlinear method to track a function in a high-dimensional space using a small number of parameters. We note that the separation rank indicates just the nominal number of terms in the representation and is not necessarily minimal.

Any discretization of the univariate functions $u_j^{(l)}$

(ALS) which was introduced originally for data fitting as the PARAFAC (PARAllel FACtor) [18] and the CANDECOMP [10] models. ALS has been used extensively

Algorithm 2

To find the entries with maximum absolute value of a tensor $\mathbf{U}$

the largest s-value. However, we have not encountered such a situation using the s-norm in practical problems.

3.3. Numerical demonstration of convergence. We construct an experiment using random tensors, with low separation rank ($r = 4$ in what follows), in dimen-

Another interesting case occurs when a tensor has multiple maximum entry candidates, either close to or exactly the same in magnitude. To explore this case, we construct an experiment similar to the previous one, namely, find the maximum absolute value of a CTD of dimension $d = 6$ and $M = 32$ samples in each direction.

to yield a CTD appropriate for use with Algorithm 2. The second, more general problem, is to construct a separated representation given data or an analytic expression of the function, from which a CTD can then be built.

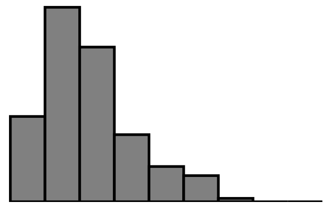
In the first problem the objective function is already represented in separated form as in (2.1), and an interpolation scheme is associated with each direction. Such problems have been subject to growing interest due to the use of CTDs for solving operator equations, e.g., deterministic or stochastic PDE/ODE systems (see, for example, [5, 6, 13, 24, 11, 12]).

In the second problem the goal is to optimize a multivariate function where no separated representation is readily available. What typically is available is a data scale linearly with respect to the number of samples in each direction [6]. However, the total number of samples needed to satisfy the local interpolation requirement may be large and, thus, additional steps to accelerate the reduction algorithm τ in Algorithm 2 may be required. One such technique consists of using the *QR* factorization as explained in [8, Section 2.4].

An example of applying analytic techniques to put a function in separated form and subsequently using Algorithm 2 to optimize the separated function is provided in Section 5.2.

5. Numerical examples

5.1. Comparison with power method algorithm from Section 2.2. To test Algorithm 2, we construct tensors from random factors and find the entries with the largest absolute value. For these examples we select dimension d



for $0 < \delta < x < \infty$. We arrive at the approximation of the first term in (5.1),

$$\left| e^{-b\left(\frac{1}{d}\sum_{i=1}^{d} x_i^2\right)^{1/2}} - \sum_{j=0}^{R} w_j \prod_{i=1}^{d} \exp\left(-x_i^2 e^{s_j}\right) \right| \le \tag{5.4}$$

where $w_j = \frac{hb}{2\sqrt{\pi d}} \exp\left(-\frac{b^2}{4d} e\right.$

the genetic algorithm uses different initial populations for each test. The largest maximum found by Algorithm 2 in the 500 tests was 8.744×10^{-1}, with mean 8.731×10^{-1} and standard deviation 5.7×10^{-4}. For comparison, the largest maxima found by the genetic algorithm in these tests was 8.750×10^{-1}, with mean 8.747×10^{-1} and standard deviation 3.6×10^{-4}. Taking into account that the accuracy of solving (5.5) and the RMS validation error of the separated representation are both of the same order as the computed standard deviations, the results

References

[1] B. Alpert. A class of bases in L^2 for the sparse representation of integral operators. *SIAM J. Math. Anal*, 24(1):246–262, 1993.

[2] B. Alpert, G. Beylkin, D. Gines, and L. Vozovoi. Adaptive solution of partial differential equations in multiwavelet bases. *J. Comput. Phys.*, 182(1):149–190, 2002.

[23] A. Nouy. Generalized spectral decomposition method for solving stochastic finite element equations: Invariant subspace problem and dedicated algorithms. *Computer Methods in Applied Mechanics and Engineering*, 197:4718–4736, 2008.

[24] A. Nouy. Proper generalized decompositions and separated representations for the numerical solution of high dimensional stochastic problems. *Archives of Computational Methods in Engineering*, 17:403–434, 2010.

[25] Y.S. Ong, P.B. Nair, and A.J. Keane. Evolutionary optimization of computationally expensive problems via surrogate modeling. *American Institute of Aeronautics and Astronautics Journal*, 41(4):687–696, 2003.

[26] M. Reynolds, A. Doostan, and G. Beylkin. Randomized alternating least squares for canonical tensor decompositions: application to a pde with random data. *SIAM J. Sci. Comput.*, accepted 7.9701007.9701144-0.7566.361.324Tm6.1X.9701144.0755[(a)-0.Ty21Tf1(p)-0.144.0750.1(i)-0.1(o)-0.154(a)-0.Ty21TfrX