Chapter 11

Conclusions

Sparse polynomial approximation is a useful tool in parametric model problems, including surrogate model construction in UQ. The theory of best s-term approximation supports the use of polynomial-based methods, and techniques such as least squares and compressed sensing are known to have desirable sample complexity bounds for obtaining polynomial approximations. In this work, we have closed a key gap between these two areas of research, by showing the existence of algorithms that achieve the algebraic and exponential rates of the best s-term approximation with respect to the number of samples m. Thus, sparse polynomial approximation can be practically realized in a provably sample-efficient manner. As our numerical experiments confirm, our algorithms are practical, and actually perform better than the theory suggests.

There are a number of avenues for further research. First, this work has focused on Chebyshev and Legendre polynomials on the hypercube $[-1, 1]^d$. It is plausible that it can be extended to general ultraspherical or Jacobi polynomials. A more significant challenge involves Hermite or Laguerre polynomials on $\mathbb{R}^d$ or $[0, \infty)^d$, respectively. This is an interesting problem for future research.

It is notable that the algorithms developed in this work do not generally compute m-term polynomial approximations. Indeed, (inexact) minimizers of the SR-LASSO problem will generally be nonsparse vectors of length $N = \Theta(n, d)$. It is interesting to investigate whether one can develop algorithms that achieve the same error bounds while computing m-term polynomial approximations. In classical compressed sensing, one can typically compute sparse solutions by using a greedy or iterative procedure (see, e.g., [61]). Unfortunately, it is not clear how to extend these procedures to the weighted case with theoretical guarantees. Nonetheless, certain weighted greedy methods appear to work well in practice for sparse polynomial approximation [4].

Another motivation for considering different algorithms is to see if the computational cost estimates can be reduced. While this is often not the main computational bottleneck in parametric model problems (generally, computing the samples is the most computationally intensive step), it is still an important issue. We have shown that the computational cost is at worst subexponential in m in infinite dimensions, and algebraic in m (for fixed d) in finite dimensions. Whether these are optimal is an interesting open problem. Here, ideas from sublinear-time algorithms [38,39] may be particularly useful.

In the case of the exponential rates, it is notable that the best s-term approximation error is exponentially small in $\gamma \cdot s^{1/d}$ (see Theorem 2.6), whereas the exponents in

Section 3.3.3 are $(\gamma/2) \cdot (m/(c_0 L))^{1/d}$ (Chebyshev) and $\gamma \cdot (m/(c_0 L))^{1/(2d)}$ (Legendre). The reason for this can be traced to the sample complexity estimate for computing a sparse (and lower) polynomial approximations via compressed sensing with Monte Carlo sampling, i.e., $m \approx c_0 \cdot 2^d \cdot s \cdot L$ (Chebyshev) or $m \approx c_0 \cdot s^2 \cdot L$ (Legendre). To see why this is the case, combine Lemma 8.1 with (10.2). In the setting of least squares, in which the desired polynomial subspace is known, it is possible to change the sampling measure to obtain sample complexity bounds that are log-linear in s and therefore near optimal. See, e.g., [9, 44, 71]. More recently, several works [21, 54, 56, 85, 93, 137] have also introduced sampling schemes that achieve linear sample complexity in s – i.e., optimal up to a constant. Unfortunately, it is unknown whether linear or log-linear sample complexity possible in the compressed sensing setting, where the target subspace is unknown. See [10] for further discussion on this issue.

Finally, as previously noted in Section 1.5, this work focuses on polynomial approximation, and not on fundamental issues pertaining to tractability and the information complexity of the classes of multivariate holomorphic functions considered. For some related work in this direction, see [81, 112, 147] and references therein. A question of particular interest is whether pointwise samples (i.e., *standard information*), and more specifically, i.i.d. pointwise samples (i.e., *random information*) constitutes optimal or near-optimal information for these classes of functions. These questions have recently been considered in a broader context in [79, 88]. See also [87] for the case of functions in Sobolev spaces. As we observed in Section 1.5, in a recent work [12] we derived lower bounds for the (adaptive) m-widths for classes of $(\boldsymbol{b}, \varepsilon)$-holomorphic functions in infinite dimensions. Showing that the algorithms (or small modifications thereof) developed in this work also attain (nearly) matching upper bounds – and, consequently, that i.i.d. pointwise samples constitute (near) optimal information – is an interesting problem for future work.