Report on Ding, Hickernell, Kritzer, and Mak:
"Adaptive Approximation for Multivariate Linear Problems with Inputs Lying in a Cone"

The paper studies the approximation of functions f from a Banach space $\mathcal{F}$ with respect to the norm of a target space $\mathcal{G}$, based on information from a finite sample of Fourier coefficients. While in classical theory, inputs are assumed to lie inside a ball $\mathcal{B}_R := \{f \in \mathcal{F} \mid \|f\|_{\mathcal{F}} \leq R\}$, and for any $\varepsilon > 0$ we can a priori determine a sample size $n \in \mathbb{N}$ such that an appropriate algorithm $\mathrm{APP}(f, n)$ guarantees $\|f - \mathrm{APP}(f, n)\|_{\mathcal{G}} \leq \varepsilon$, the authors consider unbounded input sets $\mathcal{C} \subset \mathcal{F}$ for which the sample size $n \in \mathbb{N}$ has to be adapted to the input. These unbounded input sets possess a cone property, namely, $f \in \mathcal{C}$ implies $cf \in \mathcal{C}$ for all $c \in \mathbb{R}$. Suitable cones $\mathcal{C}$ subsume conditions on inputs f such that from finitely many Fourier coefficients we may deduce upper bounds on the norm $\|f\|_{\mathcal{F}}$, from which a decision on a sample size $n(f, \varepsilon)$ can be made such that the algorithm guarantees an error of at most ε. The authors introduce the concept of *essential optimality*, stating that for a successful algorithm ALG guaranteeing $\|f - \mathrm{ALG}(f, \varepsilon)\|_{\mathcal{G}} \leq \varepsilon$ for all $f \in \mathcal{C}$, the worst case cost for inputs from the bounded subclass $\mathcal{C} \cap \mathcal{B}_R$ cannot be reduced for large R and small ε in a weak asymptotic sense. Three adaption approaches with corresponding input cones are studied and essential optimality is shown for two of them.

The authors provide interesting new perspectives on the topic of adaptivity. I recommend publishing it. However, I have several suggestions that should help improve the readability of the paper.

Major issue: Present Section 1 is long and should be split into two sections:
1. Introduction, and **2. Approximation with inputs lying in a ball**.
Section 1 should contain essential concepts of the underlying work in a more abstract way:

1.a) Function approximation with respect to the absolute error criterion is studied, see equation (2).
1.b) Algorithms are allowed to use coefficients from a particular basis representation $f = \sum_{i=1}^{\infty} \alpha_i(f) \cdot g_i$, that is, an oracle provides $\alpha_1(f), \ldots, \alpha_n(f)$ with $n = n(f, \varepsilon)$ beeing chosen based on a stopping rule, the output will be $\sum_{i=1}^{n(f,\varepsilon)} \alpha_i(f) \cdot g_i$.
1.c) The contents of Section 1.4 (Information Cost and Problem Complexity), including a definition for *non-adaptivity* $[n(f, \varepsilon) = n(\varepsilon)]$, a clarification on what is meant with *cones* of input functions (scaling property), and highlighting the difference of the presented theory to classical theory.

1.d) Some brief idea on how cone assumptions will look like that enable us to bound norms of functions $f \in \mathcal{F}$, maybe a simplified and more abstract version of the cone condition in (16) in the style

$$\|f\|_{\mathcal{F}} \leq h(\alpha_1(f), \ldots, \alpha_{n_0}(f))$$

with some homogeneous function $h : \mathbb{R}^{n_0} \to [0, \infty)$.
Some motivation for the other cones (without the details) and a summary of the main contributions of the paper.

Section 2 then should contain what is now to be found in Sections 1.1, 1.2, 1.3, and maybe also 1.6.

2.a) The problem you state in Section 1.2 is already quite specific, as it is admitted on page 4, lines 10–14, so the headline "General Linear Problem" is a bit exaggerated. It is essentially a diagonal operator between sequence spaces, motivated by what you present in Section 1.1. Maybe the illustrative example could be added right after the more general problem description. For the case study of your paper, without loss of generality, you might restrict to describing an approximation problem with $\mathrm{SOL}(f) = f$, thus $v_{\boldsymbol{k}}$ can be replaced by $u_{\boldsymbol{k}}$ for the sake of simplicity.

2.b) Despite the fact that the spaces you introduce in Sections 1.1 and 1.2 are at first only used in setting with balls as inputs, these are the spaces you have in mind for the rest of the paper. In that sense, the illustrative example of Section 1.1 is not just helpful to understand Section 1.2 with balls as inputs, but specifying $\mathcal{F}$ and $\mathcal{G}$ that way could also help to understand the cones you have in mind, giving some idea on how smoothness r plays a role in those situations. On page 18, after Remark 1, for example, you refer to the illustrative example again, but still you stay more general. It would be nice, though, if you could be more concrete in stating the results, with smoothness parameters r such as in Section 1.1.

Further remarks:

i) page 3, line 1: "implies an ordering of the wavenumbers" – the ordering is not necessarily unique?

ii) page 3, lines 5/6: Did you mix up *less restrictive* and *more restrictive*?

iii) page 3, line 17: "This algorithm contains a rule" – Do you mean: "Such an algorithm will contain ..."?

iv) page 3, line 18: "The objectives of *this chapter*" – Do you mean *this paper*?

v) page 3, items in lines 19/18: Is there a way to motivate $\mathcal{C}$ first – maybe with a rough idea about what kind of adaption should be possible – and then find an algorithm which adapts in the best possible way?

vi) page 4, lines 4–5: mathematical symbols belonging to different clauses clash, only separated by a comma. It might be improve readability to include some english words inbetween.

vii) page 7, last sentence of section 1.3: "*Choosing* that subset well is both an art and a science." To me, the art lies in *finding* a reasonable subset that subsumes essential properties of typical inputs while having a fairly easy description. It is not that much a matter of *choice* — the description of an algorithm might involve some choice for us, though.

viii) page 8, equation (14): could the conditions on ε and R be simplified or generalized as $R/\varepsilon \in [C, \infty)$ for some $C > 0$?

ix) page 8, Theorem 2: Does this have to be a *theorem*? As part of the introduction, it seems more like an *example* illustrating that the concept of *essential optimality* simplifies a lot for balls as inputs.

x) Section 1.5 provides ideas on how to generalize concepts of tractability when — besides ε^{-1} and d — a quantity $R > 0$ has to be included (or essentially the ratio R/ε instead of ε^{-1}). In its generality, it could be included in the introduction, but since all major results on it are presented in Section 2.2 and mainly build on previous work, tractability concepts could be introduced there on the fly, together with section 1.6 on tractability for the illustrative example with a ball as input.

xi) Section 2, page 12: The cone in (16) depends on A. Maybe this could be made clear in the notation, $\mathcal{C} = \mathcal{C}_A$ or similar.

xii) page 15, lines 16–24: Proposition 1 is not very surprising, it could be removed. (Besides, in (21) on the left-hand side it should be $(\ldots)_{i=1}^n$ instead of $(\ldots)_{k_i=1}^n$?) Concerning the cone from (16), it is a bit unsatisfactory that its definition already contains a specification of the one and only pilot sample. Are there different easily motivated assumptions for cones $\mathcal{C}'$ that imply usefulness of a pilot sample, i.e., $\mathcal{C}' \subseteq \mathcal{C}_A$? Maybe conditions on the ratio of different "natural" norms? It would be already satisfying if such an example is given for two distinct norms in the style of the approximation problem from Section 1.1, say, with different smoothness.

xiii) on page 18, after Remark 1, or on page 19 after Corollary 1: As already mentioned, I suggest to state results with smoothness parameters r in the style of Section 1.1.

xiv) page 18, last two formulas: it is unclear what the role of κ_i and ℓ_i is. Further, where does the the special structure of $\boldsymbol{k}_i = (k, 0, \ldots, 0)$ come into play? On the right-hand side?

xv) Section 3, page 20, cone in equation (28): Your concept of assumptions that allow to track the decay rate is quite interesting. It would be helpful to relate this to the illustrative example, in particular: What will the blocks $\mathcal{K}_j$ look like? Then also: What kind of smoothness is implied for functions from that cone? If it simplifies the description, a setting with equal importance of the coordinates would already be a good example.

Finally, the cone seems to be defined with fixed parameters a and b which could be made clear by denoting $\mathcal{C} = \mathcal{C}_{a,b}$.

xvi) Regarding the cone in (28), it seems that knowing $\sigma_1(f)$ will already suffice to decide on a sample size that guarantees an error at most ε. You should point out that checking each $\sigma_j(f)$ will allow us to stop earlier.

xvii) On page 21 you talk about *reasonable* functions. I would like to understand why this is *reasonable*. What motivation do you have in mind? Does r_Δ relate in any way to smoothness in the style of the illustrative example? In the sequel, you give some idea on the size of n_j, but it is unclear what the blocks of Fourier coefficients $\mathcal{K}_j$ could look like in an application.

xviii) Can the rates of Theorem 6 be made explicit with a particular example of a space and a choice of blocks $\mathcal{K}_j$, being as illustrative as section 1.1?

xix) page 29, line 4: "only a small number of inputs in f are important" — do you mean "small number of coefficients of f" or "inputs in $\mathcal{F}$"?

xx) On page 29, you list guiding principles from the experimental design literature. Can there something be said on which of these assumptions describe cones of inputs? How do they relate to the cone defined in (42)?

xxi) on page 30, the cones $\mathcal{C}_{\boldsymbol{\lambda}}$ are all defined with the same constant A? I suggest to denote them $\mathcal{C}_{\boldsymbol{\lambda},A}$.

xxii) In formula (40), the only non-zero entry might be at any position?

xxiii) Is it clear that $\overline{\mathcal{C}}$ from (42) is not contained in $\mathcal{C}_{\boldsymbol{\lambda}}$ for one fixed $\boldsymbol{\lambda}$? To me it does not seem to be fully understood, that this set is really a new cone.