Review of: "A class of non-parametric deformed exponential statistical models" by L. Montrucchio and G. Pistone

The paper re-derives the Hilbert manifold of reference [13] using the formalism of *general deformed exponential statistical manifolds* as developed in Naudts [12], chapter 10. This approach provides an *interpretation* of the Hilbert manifold, but does not provide any new proofs in its construction. The proofs it contains, in this respect, differ from those in [13] and [B] (below) only in notation. More significantly, the paper generalises the class of deformed exponentials to incorporate "sub-exponential" behaviour for negative values of the argument.

The paper introduces local charts, each centred on a probability measure in the Hilbert manifold. However, since these charts are affinely related to the *global* chart of [13] and [B], they do not introduce any new geometry. (The metric and affine connections induced by the local charts are precisely those induced by the global chart of [13].) In fact, the relation between the local chart at P (s_p as defined in Proposition 9) and the global chart ϕ of [13] and [B] is as follows:

$$\begin{aligned} s_p(Q) &= \phi(Q) - \phi(P) - E_{\tilde{P}}(\phi(Q) - \phi(P)) \qquad (1) \\ \phi(Q) &= \phi(P) + s_p(Q) - E_\mu s_p(Q), \qquad (2) \end{aligned}$$

where $\tilde{P}$ is the "escort" probability of section 3.1. The "A-exponential" $\exp_A$ is denoted ψ in [13] and [B]. The key step in the construction of the Hilbert manifold in [13] and [B] is to use the implicit mapping theorem to define a *normalisation function* on the model space $Z : H \to R$, such that $\psi(a + Z(a))$ is a probability density. Much of the subsequent development in [13] and [B] concerns the regularity of this function and its effect on the geometry. In the paper under review a "local" normalisation function, $K_p : H_p \to R$, is defined at each base point P in the manifold, such that $\exp_A(u - K_p(u) + \log_A p)$ is a probability density. Now

$$\exp_A(u - K_p(u) + \log_A p) = \psi(b - E_{\tilde{P}}(b - a) - K_p(u) + Z(a)) = \psi(b + Z(b)), \qquad (3)$$

where $u = s_p(Q)$, $a = \phi(P)$, and $b = \phi(Q)$. So the normalisation functions are are related as follows,

$$K_p(u) = Z(a) - Z(b) - E_{\tilde{P}}(b - a), \qquad (4)$$

and all properties of K_p (convexity, differentiability, etc.) can be obtained from those of $-Z$, which are fully developed in [13] and [B]. No new methods of proof are needed or provided. By way of example:

- Part 1 of Proposition 5 is a special case of Lemma 3.1 in [B], which establishes the smoothness properties of all the superposition operators associated with Amari's embedding maps (ξ_α in [B]), not just the case $\alpha = -1$ treated here.

- Part 2 of Proposition 5 is contained within Proposition 4.1 of [B]. The latter establishes the regularity of the inclusion map of the statistical manifold in the embedding manifold of finite measures. See, also Lemma 2.1 in [13].

Since the transition maps between local and global charts are affine, the Fenchel convex duality associated with the local normalisation function, K_p, is wholly equivalent to that associated with the global normalisation function, $-Z$. Moreover, it is not clear to me that this duality has any statistical significance. (See the concluding remarks in [B].) For example, it generates a Riemannian metric that differs from the Fisher-Rao metric at every base point of the manifold, except μ. This suggests that the local charts, as well as having no new geometrical meaning, have no new statistical meaning either, and calls into question the meaning of the "A-scores". This is in contrast with the exponential Orlicz manifold, from which the formalism is taken. In that context, local charts are defined at each base point of a "maximal exponential model", and the transition maps between them are affine. So the local (to P) duality relations are once again geometrically equivalent to the global duality relations, defined in terms of a single, preferred chart. The key difference in that context is that the local charts have significant *statistical* meaning. (The normalisation function is the cumulant generating function so that, for example, tangent vectors in the local chart are statistical scores.) This *statistical* significance more than justifies the introduction of local charts to the exponential Orlicz manifold, even though no new geometry is introduced by their use. (The Fisher-Rao metric and Amari-Chentsov tensor can be obtained in a global chart from derivatives of the global convex functions.)

A non-parametric statistical manifold, on which local charts play a more significant role, was constructed by Loaiza and Quiceno in [A] (below). As with the exponential Orlicz manifold, each connected component of the "q-exponential statistical Banach manifold" constructed is covered by any single chart. However, the transition maps between local charts, while diffeomorphic, are not affine, and so the local convex dualities at each point differ from one another. The "local" Fenchel duality in [A] is Amari's $\pm\alpha$-duality, which has important statistical meaning in Physics and non-extensive statistics. A large class of non-parametric statistical manifolds, in which local charts play an essential role of this nature, is developed in [18].

By contrast, the local charts defined on the Hilbert manifold introduce neither new geometry nor discernable statistical meaning. The important convex duality on the Hilbert manifold is that associated with the α-divergences, which is developed fully in [B] for all $\alpha \in [-1, 1]$. Derivatives of these *global* convex functions, around any point in the manifold, yield the Fisher-Rao metric and the Amari-Chentsov tensor (if $\lambda \geq 3$) at that point [B]. Local charts are not needed.

The paper under review contains two new developments.

1. It *explicitly* constructs the Fenchel duality associated with the normalisation function $K_p(\equiv -Z)$. In particular, it identifies the dual variable, u^*, and dual convex function, K_p^*. It does this in the context of the natural dual domains, $L^1(\mu)$ and $L^\infty(\mu)$. However, as pointed out above, this has questionable statistical meaning, and does not require the introduction of local charts. The more statistically important $\pm\alpha$-dualities on the Hilbert manifold have already been analysed in [B].

2. It generalises the results of [13] to a wider class of manifolds based on generalised exponential functions $\exp_A$ with "sub-exponential" behaviour for negative values

of the argument. However, as pointed out in Proposition 8 (2), a large class of these generalised exponentials (those for which $A'(0) > 0$) give rise to manifolds that coincide as sets with the Hilbert manifold of [13]. It would be remarkable if any manifolds in this class were not C^k-isomorphic to that constructed in [13], and therefore wholly equivalent to it. (In fact any such deviation would call into question the usefulness of the particular generalised exponential.) The particular choice of chart made in [13] has many advantages since it is the sum of the mixture and exponential representations of probability measures in the manifold. This fact leads to important global relations between the chart and the α-divergences; for example

$$D^{(-1)}(P|Q) + D^{(+1)}(P|Q) \leq \frac{1}{2}\|\phi(P) - \phi(Q)\|^2. \tag{5}$$

It would therefore seem natural to regard this as the *canonical* chart of the class for which $A'(0) > 0$.

The second equivalence class of manifolds constructed (those for which $A'(0) = 0$ and $A''(0) > 0$) is more restrictive since charts involve a term behaving as $1/p$. For example, if X is the real line and μ is the standard Gaussian measure, then the only Gaussians belonging to the manifold are those with variances in the range $2/3 < \sigma^2 < 2$. The Kaniadakis example is in this class and would be a strong candidate for a canonical chart, and may be of interest in statistical mechanics. However, as pointed out by the authors, the Tsallis exponential is not included in this or any other of the classes. Beyond this second class, where more derivatives of A are zero at zero, membership restrictions become tighter still.

It seems highly likely that the first class (those equivalent to the manifold of [13]) will be by far the most significant and applicable. (Witness the straightforward product space result in Proposition 8.)

Summarising the above.

- The results presented concerning the construction of the Hilbert manifold, and their proofs, differ significantly from those in [13] and [B] only in the notation used.

- The Fenchel duality associated with the normalisation functions (whether expressed in local or global charts) appears to have no statistical meaning. In particular, it does not generate the Fisher-Rao metric. It is not clear that the A-scores defined have any meaning.

- The introduction of local charts greatly complicates the presentation, and does not give any significant insight. One of the key advantages of the Hilbert manifold is that it is covered by a single chart, which maps *onto* the model space. This is clearly important in applications. The statistically relevant α-divergences can be fully analysed in this chart.

- The fibres of the Hilbert statistical bundle constructed are modelled on subspaces of $L^2(\mu)$, which is not toplinear isomorphic with $L^2(P)$ unless $p, p^{-1} \in L^\infty(\mu)$ (as the authors point out in the introduction). By contrast, the fibres in Amari's Hilbert statistical bundles [1] are the Hilbert spaces $L^2(P)$, in which the integrating measure is the base measure. This difference is testament to the special status of the reference measure μ in the Hilbert manifold, and reinforces the above criticism of multiple charts.

- The generalisation of the deformed exponential functions to incorporate those with strict "sub-exponentiality" (class 2 in the above discussion) may be of interest in statistical mechanics. However, it seems likely that those in class 1 (which seem wholly equivalent to that in [13] and [B]) will be far more significant and useful.

In addition to the above comments on the paper's content, I also have some criticisms of its style. It develops a particular *interpretation* of the Hilbert manifold: as a deformed exponential model. However, this is not the way the paper is written. After a brief reference to [13] (but not [B], which is central to this topic), it goes on to construct the manifold from scratch, using different notation, but essentially equivalent proofs. This gives the impression that the paper contains important new constructions or proofs, rather than an interpretation, and slight extension of, existing results. The casual reader is left believing that the Hilbert manifold is being constructed here, for the first time. The absence of any cross-referencing to concepts or results in the original papers (for example the equivalence between K_p here and $-Z$ in [13]) is striking.

References

[**A**] Loaiza, G. and Quiceno, H.R., A q-exponential statistical Banach manifold, J. Math. Anal. Appl. 398, 466-476 (2013).

[**B**] Newton, N.J.: Infinite-dimensional statistical manifolds based on a balanced chart, Bernoulli 22(2), 711–731 (2016). arXiv:1308.3602 (v1 August 2013).