In this paper we consider a class of nonlinear estimators that are the solutions to a certain variational problem. These estimators generalize the concept of conditional expectation; we investigate their continuity and convergence properties.
ABSTRACT

In this paper we consider a class of nonlinear estimators that are the solutions to a certain variational problem. These estimators generalize the concept of conditional expectation; we investigate their continuity and convergence properties.

I. INTRODUCTION

In this paper we will develop some machinery that generalizes some well known approximation and continuity properties of minimum mean square error analysis. These will allow us to examine estimation theory in a much more general context and give us a much wider choice of criteria for analyzing and penalizing error. We will be required to forego some of the main convenient accoutrements of minimum mean square error analysis: the Hilbert space structure of $L^2$, the projection theorem, and the fact that minimum norm projection onto a closed subspace is a nonexpansive linear operator. However, in return we receive a method that allows us to customize our notion of error to our particular model. Oftentimes, mean square error is not the appropriate fidelity criterion.

II. PRELIMINARIES

Throughout this paper let $(\Omega, \mathcal{F}, P)$ denote a probability space and let $L^0(\Omega, \mathcal{F}, P)$ denote the set of all random variables $X: \Omega \rightarrow \mathbb{R}$ modulo a.s. equivalence equipped with the topology of convergence in probability.

Let $\phi: [0, \infty) \rightarrow [0, \infty)$ be convex, increasing, and satisfy $\phi(0)=0$. We define the Orlicz space

$$L^\phi(\Omega, \mathcal{F}, P) = \{X \in L^0(\Omega, \mathcal{F}, P): \int_\Omega \phi(|X|) dP < \infty\}.$$

The function $\phi$ is said to satisfy the doubling condition if there exist $C, M > 0$ such that $x \geq M \Rightarrow \phi(2x) \leq C \phi(x)$. If $\phi$ satisfies the doubling condition, $L^\phi(\Omega, \mathcal{F}, P)$ is a vector space. Moreover, it is a Banach space with the Luxemburg norm

$$||X||_{L^\phi} = \inf \{\lambda > 0: \int_\Omega \phi\left(\frac{|X|}{\lambda}\right) dP \leq \phi(1)\}.$$

Note that this is simply the Minkowski functional of a certain subset of $L^\phi$.

Also, if $\{X_n\}_{n=1}^\infty$ and $X$ are in $L^\phi$, $X_n \xrightarrow{L^\phi} X$ if and only if $\lim_{n \rightarrow \infty} \int_\Omega \phi(|X_n - X|) dP = 0$.

For the basic facts on Orlicz spaces see [3]. Details on Minkowski functions can be found in [1], pp. 294-295 or [6], pp. 23-26.

Let $\phi: [0, \infty) \rightarrow [0, \infty)$ be increasing and have a strictly increasing first derivative $\phi'$ with $\phi(x) \rightarrow \infty$ as $x \rightarrow \infty$. Since $\phi: [0, \infty) \rightarrow [0, \infty)$ is a homeomor-


The primary result of [5] is contained in the

**Theorem 1:** Let \( \Phi, \Psi \) be an Orlicz pair, \((\Omega, \mathcal{F}, P)\) be a nonatomic probability

space, \( M \) be a separable metric space and \( \{X(t) : t \in [0, T]\} \) be a stochastically

continuous process on \((\Omega, \mathcal{F}, P)\) taking values in \( M \). Then for any round off

scheme \( \{Q_n\}_{n=1}^{\infty} \) on \( M \) and any increasing sequence \( \{P_m\}_{m=1}^{\infty} \) of partitions of

\([0, T]\) whose meshes decrease to zero we have for any \( Y \in L^\Phi(\Omega, \mathcal{F}, P) \)

\[ \| E_\Phi(Y|Q^*_n(X(t))): t\in\mathcal{T} \|_{L^\Phi} \to 0 \]
as \( m,n \to \infty \); or, equivalently,
\[ 0 = \lim_{m,n \to \infty} \int_\Omega \Phi(\| E_\Phi(Y|Q^*_n(X(t))): t\in\mathcal{T} \|_{L^\Phi}) d\mathcal{P} \]
where \( E_\Phi(\cdot|\mathcal{F}): L^\Phi(\Omega, \mathcal{F}, \mathcal{P}) \to L^\Phi(\Omega, \mathcal{F}, \mathcal{P}) \) denotes metric projection and \( \mathcal{F} \) is any \( \sigma \)-subalgebra of \( \mathcal{P} \).

III. A CONVEXITY CONDITION FOR CONVEX FUNCTIONALS ON REFLEXIVE BANACH SPACES

Let \( B \) be a reflexive Banach space and \( \psi: B \to \mathbb{R} \) be a convex, norm continuous map satisfying
\[ \lim_{||x|| \to \infty} \psi(x) = +\infty. \]
Then using the Smul'lyan theorems we see that \( \forall x \in B \)
\[ \min_{z \in K} \psi(x-z) \]
possesses a closed convex nonvoid subset \( L_x \) of \( K \) of solutions. Furthermore, if \( \psi \) is strictly convex each \( L_x \) is a singleton \( \{p(x)\} \). We call the map \( p: B \to K \) the solution map of (1). Let \( \delta > 0 \); define
\[ K_{\delta,x} = \{ z \in K: \psi(x-z) < \delta + \psi(x-p(x)) \}. \]
We say \( \psi \) is locally uniformly convex if \( \lim_{\delta \to 0} \text{diam} K_{\delta,x} = 0 \) for all \( x \in B \) and closed convex \( K \subseteq B \).

Proposition 2: If \( \psi: B \to \mathbb{R} \) is locally uniformly convex and uniformly continuous on bounded subsets of \( B \), \( K \subseteq B \) is nonvoid convex and closed and \( p: B \to K \) is the solution map of (1) then \( p \) is norm continuous.

Proof: Let \( x_n \to x \) in \( B \) and \( \delta > 0 \). Take \( z \in K_{\delta,x} \). Then
\[ \psi(x-z) < \delta + \psi(x-p(x)). \]
Since \( x_n \to x \) there exists \( N \in \mathbb{N} \) so that \( n \geq N \Rightarrow \psi(x_n-z) < \delta + \psi(x_n-p(x)). \)
By minimality
\[ \psi(x_n-p(x_n)) \leq \psi(x_n-z) \]
\[ < \delta + \psi(x_n-p(x)). \]
Taking limit suprema,
\[ \lim_{n \to \infty} \sup \psi(x_n-p(x_n)) < \psi(x-p(x)). \]  \( \quad (2) \)
Suppose by way of the contrapositive that \( \{p(x_n)\}_{n=1}^\infty \) is unbounded.
Since \( x_n \to x \), \( \{x_n\}_{n=1}^\infty \) is bounded; this forces \( \{x_n-p(x_n)\}_{n=1}^\infty \) to be unbounded.
Observing that \( \lim_{||x|| \to \infty} \psi(x) = +\infty \) it now follows that
\[ \lim_{n \to \infty} \sup \psi(x_n-p(x_n)) = +\infty, \]
a flagrant violation of (2). This insures that \( \{p(x_n)\}_{n=1}^\infty \) is bounded.
Suppose that for some \( z_0 \in B \), \( p(x_n) \to z_0 \). Then \( x_n-p(x_n) \to x-z_0 \); \( \psi \) is convex and norm continuous and therefore weakly lowersemicontinuous. Thus
\[
\psi(x-z_0) \leq \liminf_{n \to \infty} \psi(x_n-p(x_n)) \\
\leq \limsup_{n \to \infty} \psi(x_n-p(x_n)) \leq \psi(x-p(x)).
\]

Next observe that \(K\) is closed and convex and therefore weakly closed. Thus \(p(x_n) \in K,\ \text{ne}\mathbb{N},\ p(x_n) \rightarrow z_0 \Rightarrow z_0 \in K\). By minimality, \(z_0 = p(x)\).

Since \(B\) is reflexive and \(\{p(x_n)\}_{n=1}^{\infty}\) is bounded each subsequence of \(\{p(x_n)\}_{n=1}^{\infty}\) must have a further weakly convergent subsequence. But the foregoing discussion tells us that this further subsequence must converge to \(p(x)\). Immediately it follows \(p(x_n) \rightarrow p(x)\).

We may apply the weak lower semicontinuity of \(\psi\) to see that
\[
\psi(x-p(x)) \leq \liminf_{n \to \infty} \psi(x_n-p(x_n)).
\]

Combine this with (2) to get
\[
\psi(x-p(x)) = \lim_{n \to \infty} \psi(x_n-p(x_n)).
\]

Put \(M = 1 + \sup \{ ||x_n-p(x) || \}; \) note \(M < \infty\). Take \(\varepsilon > 0\), choose \(n > 0\) s.t. \(n < 1\) and \( ||x-y|| < n,\ ||x||, ||y|| \leq M \Rightarrow \{ |\psi(x)-\psi(y)| < \varepsilon\). Because \(x_n \rightarrow x\) there exists \(N \in \mathbb{N}\) such that \(n \geq N \Rightarrow ||x_n-x|| < \varepsilon\). Then \(n \geq N \Rightarrow
\]
\[
||x_n-p(x_n)-(x-p(x_n))|| = ||x_n-x|| < \varepsilon
\]
so \(n \geq N \Rightarrow
\]
\[
|\psi(x_n-p(x_n)) - \psi(x-p(x))| < \varepsilon.
\]

This forces \(\lim_{n \to \infty} \psi(x_n-p(x_n)) - \psi(x-p(x)) = 0\) and hence \(\lim_{n \to \infty} \psi(x_n-p(x_n)) = \psi(x-p(x))\).

Noting \(\lim_{n \to \infty} \psi(x-p(x_n)) = \psi(x-p(x))\) there exists \(N \in \mathbb{N}\) so that \(n \geq N \Rightarrow
\]
\[
\psi(x-p(x_n)) < \delta + \psi(x-p(x)).
\]

But this says that for \(n \geq N,\ p(x_n) \in K_{\delta,x}\). Obviously \(p(x) \in K_{\delta,x}\), so for \(n \geq N,\ ||p(x_n) - p(x)|| \leq \text{dia } K_{\delta,x}\). Since \(\lim_{\delta \to 0} \text{dia } K_{\delta,x} = 0\) we must have \(p(x_n) \rightarrow p(x)\).

QED

**Proposition 3:** Suppose \(\{K_n\}_{n=1}^{\infty}\), \(K\) are nonvoid closed convex subsets of \(B\) with \(K_n \subseteq K_{n+1}\), \(n \in \mathbb{N}\) and \(K = \bigcup_{n=1}^{\infty} K_n\). Let \(\psi : B \rightarrow \mathbb{R}\) be locally uniformly convex and norm continuous. Finally denote by \(p_n\) the solution map of (1) corresponding to \(K_n,\ n \in \mathbb{N}\) and \(p\) denote the solution map of \(K\). Then for \(x \in B,\ p_n(x) \rightarrow p(x)\).

**Proof:** Let \(\delta > 0,\ x \in B\). By the continuity of \(\psi,\ K_{\delta,x}\) is a relative neighborhood of \(p(x)\) in \(K\). Noting \(K = \bigcup_{n=1}^{\infty} K_n\), we see there must be an \(N \in \mathbb{N}\) so that
\(n \geq N \Rightarrow K_n \cap K_{\delta,x} \neq \emptyset\).

Fix \(n \geq N,\ z \in K_{\delta,x} \cap K_n\). Then we know that (i) \(\psi(x-z) < \delta + \psi(x-p(x));\)
(ii) \(\psi(x-p_n(x)) < \psi(x-z)\). Assembling the facts, \(n \geq N \Rightarrow
\]
\[
\psi(x-p_n(x)) < \delta + \psi(x-p(x)).
\]

Letting \(n \to \infty\),
\[ \limsup_{n \to \infty} \psi(x-p_n(x)) \leq \psi(x-p(x)). \quad (3) \]

We may imitate a familiar argument in Proposition 2 to see that \( \{p_n(x)\}_{n=1}^{\infty} \) is bounded.

Since \( K_n \subseteq K_{n+1} \subseteq K \), \( n \in \mathbb{N} \) it is easy to see that

\[ \psi(x-p_n(x)) \geq \psi(x-p_{n+1}(x)) \geq \psi(x-p(x)). \]

With the assistance of (3) we obtain

\[ \psi(x-p(x)) = \lim_{n \to \infty} \psi(x-p_n(x)) . \]

Pick \( \delta > 0 \). There exists \( N \in \mathbb{N} \) such that \( n \geq N \Rightarrow \psi(x-p_n(x)) < \psi(x-p(x)) + \delta \).

We may now conclude \( p_n(x) \in K_{\delta, x}, \ n \geq N \). Since \( \text{diam} K_{\delta, x} \to 0 \) as \( \delta \to 0 \) we conclude that \( p_n(x) \to p(x) \) in norm. \( \text{QED} \)

IV. SOME FACTS ABOUT ORLICZ SPACES

**Theorem 4:** Let \( (\Omega, \mathcal{F}, P) \) be a probability space and \( \Phi, \Psi \) be an Orlicz pair.

Then \( L_\Phi^\Psi(\Omega, \mathcal{F}, P) \) is uniformly convex if and only if

\[ \forall \varepsilon \text{ such that } 0 < \varepsilon < \frac{1}{4} \text{ there exists } R_\varepsilon > 0 \text{ so that} \]

\[ \liminf_{n \to \infty} \frac{\psi(x)}{\psi((1-\varepsilon)x)} > R_\varepsilon. \]

**Proof:** This is the central result of [4].

Define \( \xi: L_\Phi^\Psi(\Omega, \mathcal{F}, P) \to [0, \infty) \) by \( \xi(X) = \int_{\Omega} \Phi(|X|)dP \). Then \( \xi \) is a strictly convex norm continuous functional and \( \lim ||X||_\Phi^\Psi = \infty \).

Let \( 0 < \varepsilon < 1 \), \( X, Y \in L_\Phi^\Psi(\Omega, \mathcal{F}, P) \) and \( X \neq Y \) a.s. \([P]\). Then \( ||X-Y||_\Phi^\Psi \neq 0 \). Using the convexity of \( \Phi \),

\[ \Phi(|X-Y|) = \Phi\left(\varepsilon \left(\frac{1}{\varepsilon} |X-Y|\right) + (1-\varepsilon)(0)\right) \]

\[ \leq \varepsilon \Phi\left(\frac{|X-Y|}{\varepsilon}\right) + (1-\varepsilon)\Phi(0) \]

\[ = \varepsilon \Phi\left(\frac{|X-Y|}{\varepsilon}\right). \]

Integrating,

\[ \int_{\Omega} \Phi(|X-Y|)dP \leq \varepsilon \int_{\Omega} \Phi\left(\frac{|X-Y|}{\varepsilon}\right)dP. \]

Now suppose \( ||X-Y||_\Phi^\Psi < 1 \). Then

\[ \int_{\Omega} \Phi(|X-Y|)dP \leq ||X-Y||_\Phi^\Psi \int_{\Omega} \Phi\left(\frac{|X-Y|}{\varepsilon}\right)dP \]

\[ = ||X-Y||_\Phi^\Psi \Phi(1), \]

amply demonstrating the global uniform continuity of \( \xi \).

We will not address the question of the local uniform convexity of \( \xi \) here. The hypothesis of Theorem 4 or possibly something similar might
guarantee the local uniform convexity of $\xi$.

V. A CONVERGENCE PRINCIPLE FOR A GENERALIZED CONDITIONAL EXPECTATION

**Definition:** Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{F}$ be a sub $\sigma$-algebra of $\mathcal{F}$ and $\phi, \psi$ be an Orlicz pair. For $X \in L^\phi(\Omega, \mathcal{F}, P)$, we define $\hat{E}_\phi(X|\mathcal{F})$ to be the unique solution to the variational problem

$$\min \{ \int_{\Omega} \phi(|X-Z|)dP : Z \in L^\phi(\Omega, \mathcal{F}, P) \}.$$ 

Recall from [5] that if $\phi(x) = \frac{x^2}{2}$ for $x > 0$, then $\hat{E}_\phi(\cdot|\mathcal{F})$ is ordinary conditional expectation given $\mathcal{F}$.

**Remark:** If the functional $X \mapsto \int_{\Omega} \phi(|X|)dP$ is locally uniformly convex then $\hat{E}_\phi(\cdot|\mathcal{F})$ is norm continuous on $L^\phi(\Omega, \mathcal{F}, P)$.

**Theorem 5:** Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\{\mathcal{F}_n\}_{n=1}^\infty$ be an increasing family of sub $\sigma$-algebras of $\mathcal{F}$ and put $\mathcal{F} = \bigvee_{n=1}^\infty \mathcal{F}_n$. Suppose $\phi, \psi$ is an Orlicz pair so that $X \mapsto \int_{\Omega} \phi(|X|)dP$ is locally uniformly convex. Then for $X \in L^\phi(\Omega, \mathcal{F}, P)$,

$$\hat{E}_\phi(X|\mathcal{F}_n) \xrightarrow{L^\phi} \hat{E}_\phi(X|\mathcal{F}) \quad \text{as } n \to \infty,$$

or equivalently,

$$\lim_{n \to \infty} \int_{\Omega} \phi(|\hat{E}_\phi(X|\mathcal{F}_n) - \hat{E}_\phi(X|\mathcal{F})|)dP = 0.$$ 

**Proof:** This is an immediate consequence of Proposition 3. QED

**Theorem 6:** Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\{X(t) : t \in [0, T]\}$ be a process on $(\Omega, \mathcal{F}, P)$ taking values in a separable metric space $M$ and $\phi, \psi$ be an Orlicz pair with $X \mapsto \int_{\Omega} \phi(|X|)dP$ locally uniformly convex. Then if $\{Q_n\}_{n=1}^\infty$ is a round off scheme for $M$ and $\{P_m\}_{m=1}^\infty$ is an increasing sequence of partitions of $[0, T]$ whose meshes decrease to zero, we have for $X \in L^\phi(\Omega, \mathcal{F}, P)$

$$\hat{E}_\phi(Y|Q_n(X(t)) : t \in P_m) \xrightarrow{L^\phi} \hat{E}_\phi(Y|X(t) : t \in [0, T])$$

as $m, n \to \infty$, or equivalently,

$$\lim_{m, n \to \infty} \int_{\Omega} \phi(|\hat{E}_\phi(Y|Q_n(X(t)) : t \in P_m) - \hat{E}_\phi(Y|X(t) : t \in [0, T])|)dP = 0.$$ 

VI. CONCLUSIONS AND A PARTING SHOT AT EARLIER WORK

The abstract principle developed in Section III subsumes the Banach space principle developed in [5]. Simply take the functional to be the norm.

We have enhanced the feasibility of studying the operators $\hat{E}_\phi$ numerically in two ways. First, knowing that $\hat{E}_\phi$ is continuous assures us that
it will tolerate small $L^\Phi$-perturbations. Second, the result of Theorem 6 shows us we may approximate $\hat{E}_\Phi(Y|X(t); t\in[0,T])$ by $\hat{E}_\Phi(Y|Q_n(X(t)); t\in P_m)$. Since $\sigma(Q_n(X(t)); t\in P_m)$ is finite, calculation of $\hat{E}_\Phi(Y|Q_n(X(t)); t\in P_m)$ consists of solving a finite dimensional nonlinear optimization problem. A large body of knowledge now exists about the numerical solution of such problems.

ACKNOWLEDGEMENTS

This research was supported by the Air Force Office of Scientific Research under Grants AFOSR-81-0047 and its successor. The authors would like to thank the Banach Space and Approximation Theory groups at the University of Texas at Austin for many helpful suggestions.

REFERENCES

END
10-86
DTIC