
DOI: https://doi.org/10.4213/tvp1955
Установлены некоторые общие свойства замкнутости сумм пространств измеримых функций. В качестве применения доказаны существование и единственность решений обобщенной задачи Шрёдингера при некотором условии интегрируемости, но без каких-либо предположений о топологии или ограниченности. Полученные свойства позволяют также доказать интересный результат о структуре законов с многомерными частными распределениями, установить существование оптимальных аппроксимаций в аддитивных статистических моделях и обобщить представление Колмогорова для непрерывных функций нескольких переменных. Из последнего результата вытекает, что любая локально ограниченная измеримая функция имеет точное представление нейронной сетью с одним скрытым слоем.

Ключевые слова и фразы: уравнение Шредингера, сумма пространств, аддитивные модели, многомерные частные распределения, представление Колмогорова.

1. Introduction

The main motivation of this paper is to derive existence and uniqueness results for some generalized versions of the Schrödinger equations. It turns out that these results are based on some crucial closedness properties of sum spaces of measurable functions, which are derived in Section 2. These closedness properties also allow to prove some interesting existence results for projections in additive statistical models, furthermore, to develop an interesting structural property of distributions with multivariate marginals, and to derive an extension of Kolmogorov’s famous representation theorem for continuous functions of $n$-variables.

Let $\mu$ be a probability measure on a product space $(E,A) = (E_1,A_1) \otimes (E_2,A_2)$ with marginals $\mu_i$ on $E_i$. Let $L^0(E_i,A_i,\mu_i) = L^0(\mu_i)$ denote the class of $A_i$-measurable real functions. It is clear from the context, if $f \in L^0(\mu_i)$ stands for some version or for the equivalence-class mod $\mu_i$-null-functions. Furthermore, let $\nu_i \in M^1(E_i,A_i)$ be probability measures on

*Institut für Mathematische Stochastik, Eckerstr. 1, 79104 Freiburg, Germany.
**Institut für Mathematische Statistik, Einsteinstr. 62, 48149 Münster, Germany.
Closedness of Sum Spaces and the Generalized Schrödinger Problem

$(E_i, A_i)$ continuous with respect to $\mu_i$ with densities $r_i = d\nu_i/d\mu_i$. Consider the following system of nonlinear integral equations which we denote as (generalized) Schrödinger-problem:

$$
E_\mu(a \otimes b | \pi_1 = x) = r_1(x) \quad \mu_1 \text{-a.e.}, \\
E_\mu(a \otimes b | \pi_2 = y) = r_2(y) \quad \mu_2 \text{-a.e.},
$$

(1.1)

where $a \otimes b \in L^0_+(\mu)$, $a \otimes b(x, y) := a(x)b(y)$, and $\pi_i : E_1 \times E_2 \to E_i$ are the projections on $E_i$.

If $\mu$ has a density $h$ with respect to $\mu_1 \otimes \mu_2$, i.e., $\mu = h \mu_1 \otimes \mu_2$, then (1.1) is equivalent to

$$
\int a(x) b(y) h(x, y) \mu_1(\text{d}y) = r_1(x) \quad \mu_1 \text{-a.e.}, \\
\int a(x) b(y) h(x, y) \mu_2(\text{d}x) = r_2(y) \quad \mu_2 \text{-a.e.},
$$

(1.2)

Equation (1.1) is equivalent to the following marginal problem. Let $M(\nu_1, \nu_2)$ denote the set of all probability measures on $(E, A)$ with marginals $\nu_1, \nu_2$. Then (1.1) is equivalent to:

Find $a \otimes b \in L^0_+(\mu)$ such that

$$
\nu = a \otimes b \mu \in M(\nu_1, \nu_2).
$$

(1.3)

Problem (1.1), (1.3) is of interest for the structure of (generalized) Schrödinger bridges (cf. [2], [3], [4], [6], [7], [11], [15], [21]). In our previous paper [18] we proved the existence and uniqueness of solutions of (1.2) with $a \in L^0_+(\mu_1)$, $b \in L^0_+(\mu_2)$ under an integrability condition but without any assumptions on the boundedness of $h$ or topological restrictions extending related results in [1], [8], [10], [11] and [20]. In this paper we investigate the general problem (1.1), (1.3). The progress is established by some crucial closedness properties of sum spaces of measurable functions in Section 2 which need the development of some new tools, and which also find some further applications in Sections 3 and 4. In particular we consider applications to optimal approximations in additive statistical models, some structure results for distributions with given marginals and a «generalization» of Kolmogorov’s representation theorem to measurable functions. A part of the problems is caused by the possibly complicated structure of the support of $\mu$.

2. Closedness Properties of Sum Spaces

Consider the sum space of measurable functions

$$
F = L^0(\mu_1) \oplus L^0(\mu_2) = \{ \varphi = f \oplus g; \ f \in L^0(\mu_1), \ g \in L^0(\mu_2) \},
$$

(2.1)
where \( f \oplus g(x, y) = f(x) + g(y) \). \( F \) is a subspace of
\[
G = (M(E_1) \oplus M(E_2)) \cap L^0(\mu),
\]
(2.2)
where \( M(E_i) \) is the class of all real valued functions (maps) on \( E_i \) and \( L^0(\mu) = L^0(E_1 \times E_2, A_1 \otimes A_2, \mu) \) is supplied with the complete metric of \( \mu \)-stochastic convergence. In this section we are concerned with a characterization of the closure of \( F \) in \( L^0(\mu) \) with respect to \( \mu \)-stochastic convergence (and related types of convergence). The following result is obtained by using essentially methods of the proofs in [18].

**Proposition 2.1.** \( G \) is closed in \( L^0(\mu) \) with respect to \( \mu \)-stochastic convergence. In particular, the closure of \( F \) in \( L^0(\mu) \) is a subset of \( G \).

**Proof.** 1. For any limit point \( \varphi = \varphi(x, y) \) of \( G \) with respect to \( \mu \)-stochastic convergence, there exists a sequence \( f_n \oplus g_n \) in \( G \) such that
\[
\lim_{n \to \infty} (f_n(x) + g_n(y)) = \varphi(x, y), \quad \mu \text{-a.e.}
\]
Then \( C := \{(x, y) \in E; \lim_{n}(f_n(x) + g_n(y)) \text{ exists}\} \in A_1 \otimes A_2, \mu(C) = 1 \) and without loss of generality
\[
\varphi(x, y) = \lim_{n}(f_n(x) + g_n(y)) 1_C(x, y).
\]
2. Choose \((x_1, y_1) \in C \) and define
\[
\begin{align*}
  f^1_n(x) &:= f_n(x) - f_n(x_1), \quad g^1_n(y) := g_n(y) + f_n(x_1), \\
  C^1_1 &:= \{x \in E_1; f^1(x) \text{ exists}\}, \\
  C^1_2 &:= \{y \in E_2; g^1(y) \text{ exists}\}.
\end{align*}
\]
(2.3)
Then \((x_1, y_1) \in C^1_1 \times C^1_2 \) and for \((x, y) \in C^1_1 \times C^1_2 \) it holds
\[
\begin{align*}
  f_n(x) + g_n(y) &= (f_n(x) - f_n(x_1)) + (g_n(y) + f_n(x_1)) \to f^1(x) + g^1(y),
\end{align*}
\]
i.e., \((x, y) \in C \). Therefore, \( C^1_1 \times C^1_2 \subset C \).
3. Choose \((x_2, y_2) \in C \setminus C^1_1 \times C^1_2 \) and define \( f^2_n, g^2_n, f^2, g^2, C^2_1, C^2_2 \) as in 2. with \((x_1, y_1) \) replaced by \((x_2, y_2) \). Then
\[
\begin{align*}
  C^1_1 \cap C^2_2 &= \emptyset, \quad C^1_2 \cap C^2_2 = \emptyset.
\end{align*}
\]
(2.5)
To prove (2.5) let \( x \in C^1_1 \cap C^2_2 \). Then \( f_n(x) - f_n(x_1) \to f^1(x) \) and \( f_n(x) - f_n(x_2) \to f^2(x) \) and, therefore,
\[
\begin{align*}
  f_n(x_2) - f_n(x_1) \to f^1(x) - f^2(x) = f^1(x_2).
\end{align*}
\]
This implies that \( x_2 \in C^1_1 \). From
\[
\begin{align*}
  g_n(y_2) + f_n(x_1) &= (g_n(y_2) + f_n(x_2)) - (f_n(x_2) - f_n(x_1)) \\
  &\to g^2(y_2) - f^1(x_2)
\end{align*}
\]
(2.6)
it follows by (2.3) that \( y_2 \in C_2^1 \). This contradicts \((x_2, y_2) \notin C_1^1 \times C_2^1 \). By (transfinite) induction we obtain a (possibly uncountable) family \((C_i^1 \times C_i^2, i \in I)\), such that \((x_i, y_i) \in C_i^1 \times C_i^2, C_i^1 \cap C_j^1 = \emptyset, C_i^2 \cap C_j^2 = \emptyset\) for all \(i, j \in I, i \neq j\), and \(C = \sum_{i \in I} C_i^1 \times C_i^2\). Furthermore, as in point 2 we obtain functions \(f \circ 1_{C_i^1}\) and \(g \circ 1_{C_i^2}\) such that

\[
\varphi(x, y) 1_{C}(x, y) = (f(x) + g(y)) 1_{C}(x, y),
\]

where \(f(x) := \sum_{i \in I} f^i(x) 1_{C_i^1}(x)\) and \(g(y) := \sum_{i \in I} g^i(y) 1_{C_i^2}(y)\).

Remarks. a) Additionally, we point out that the components \(C_i^1 \times C_i^2\) of the convergence set \(C = \sum_{i \in I} C_i^1 \times C_i^2\) are in \(A_1 \otimes A_2\). Indeed, \(C_{x_i} = C_i^2\), where \(C_{x_i}\) denotes the \(x_i\)-section of \(C \in A_1 \otimes A_2\), hence

\[
C_i^2 \in A_2 \text{ and } C_i^1 \in A_1 \text{ for all } i \in I
\]

as well. Furthermore, the corresponding components \(f \circ 1_{C_i^1}\) and \(g \circ 1_{C_i^2}\) of \(f\) and \(g\) are measurable. Clearly, the \(x_i\)-section of \(\varphi, \varphi(x_i, \cdot) = f(x_i) \circ 1_{C_i^2} + g \circ 1_{C_i^2}\), is \(A_2\)-measurable, hence

\[
g \circ 1_{C_i^2} \text{ is } A_2\text{-measurable and } f \circ 1_{C_i^1} \text{ is } A_1\text{-measurable}
\]

as well.

b) H. Kellerer pointed out to us a very simple and nice argument which yields the stated representation \(C = \sum_{i \in I} C_i^1 \times C_i^2, C_i^1 \cap C_j^2 = \emptyset\) for \(i, j \in I, i \neq j\), and \(\nu = 1, 2\) in the proof of Proposition 2.1: The two-dimensional differences of \(\varphi_n\) disappear. Therefore, given \(x_1, x_2 \in E_1\) and \(y_1, y_2 \in E_2\), it follows that \((x_2, y_2) \in C\) if \((x_1, y_1), (x_1, y_2), (x_2, y_1) \in C\). This implies that the sections \(C_{x_1}\) and \(C_{x_2}\) are either disjoint or coincide, which yields the assertion.

c) \(F\) is closed, if \(I\) is countable, or if all components \(C_i^1 \times C_i^2\) are one-point sets. But \(F\) is not closed in general. The following beautiful counterexample due to N. Gantert (private communication) has surprised us and in fact it pointed out a mistake in a first version of this paper.

Example (N. Gantert). Consider \(E_1 = E_2 = \{0, 1\}^\mathbb{N}, A_1 = A_2 = (\mathcal{P}((0, 1)))^\mathbb{N}\) where \(\mathcal{P}(A)\) is the set of all subsets of \(A\) and \(\mu_1 = (B(1, \frac{1}{2}))^\mathbb{N}\), define the Markov-kernel \(K\) from \(E_1\) to \(E_2\) by \(K(x, \cdot) = \sum_{k=1}^{\infty} \frac{1}{2k} \varepsilon_{x(k)}\), where \(x(k) = (x_1, \ldots, x_{k-1}, x_{k+1}, x_{k+2}, \ldots)\), and let \(\mu = \mu_1 \otimes K\) be the product on \(E_1 \times E_2\). Furthermore, consider the measurable functions

\[
f_n(x) = \sum_{i=1}^{n} x_i, \quad g_n(y) = - \sum_{i=1}^{n-1} y_i, \quad n \geq 2.
\]

Obviously, \(f_n \to \infty (\mu\text{-a.e.}), g_n \to -\infty (\mu\text{-a.e.})\) and

\[
f_n \oplus g_n \to \varphi = \varphi 1_C \in L^0(\mu) \text{ } \mu\text{-a.e.},
\]

5
where \( C = \{ (x, y) \in E; \lim_{n \to \infty} (f_n(x) + g_n(y)) \text{ exists} \} \). Clearly, \( (x, y) \in C \) if and only if \( x_{i+1} = y_i \) eventually, and \( x, x' \) are in the same component if and only if \( x_i = x'_i \) eventually, i.e., the components are the atoms of the tail \( \sigma \)-field on \( E_1 \).

Suppose that \( \varphi = f \oplus g \) (\( \mu \)-a.s.) for some \( f \oplus g \in L^0(\mu_1) \oplus L^0(\mu_2) \). Let \( X = (X_i; i \in \mathbb{N}) = \text{id}_{E_1} \) and \( X^k = (X_i^k; k \in \mathbb{N}), X_i^k = X_i \) for \( i > k, X_i^k = 0 \) for \( i \leq k \). \( X \) and \( X^k \) are in the same component, hence \( f \circ X - f \circ X^k = \lim_{n \to \infty} (f_n \circ X - f_n \circ X^k) = \sum_{i=1}^k X_i, \) i.e., \( f \circ X = \sum_{i=1}^k X_i + f \circ X^k \). Note that \( \sum_{i=1}^k X_i, f \circ X^k \) are independent and that \( \sum_{i=1}^k X_i \) is \( B(k, \frac{1}{2}) \)-distributed.

Given \( 0 < \varepsilon < \frac{1}{50} \), choose \( a \in \mathbb{R} \) with \( \mu\{f \circ X > a\} < \varepsilon \) and choose \( b < 0 \) (by the central limit theorem) such that \( \lim_{n \to \infty} \mu\{\sum_{i=1}^n X_i < 1/2(n + b \sqrt{n})\} = \frac{1}{2} \), hence \( \mu\{\sum_{i=1}^n X_i \geq b_n\} > \frac{3}{4} (\forall n \geq n_b) \), \( b_n := \frac{1}{2}(n + b \sqrt{n}) \). Then, on one hand,

\[
\varepsilon > \mu\{f \circ X > a\} \geq \mu\{f \circ X^n > a - b_n\} \mu\left\{ \sum_{i=1}^n X_i \geq b_n \right\}
\]

\[
> \frac{3}{4} \mu\{f \circ X^n > a - b_n\}
\]

yields \( \mu\{f \circ X^n > a - b_n\} < \frac{3}{4} \varepsilon (\forall n \geq n_b) \). On the other hand,

\[
\varepsilon > \mu\{f \circ X \leq -a\} \geq \mu\{f \circ X^n \leq a - b_n\} \mu\left\{ \sum_{i=1}^n X_i \leq b_n - 2a \right\}
\]

\[
\geq \left(1 - \frac{4}{3} \varepsilon \right) \frac{1}{16} (\forall n \geq m_b \geq n_b) \text{ for some } m_b \in \mathbb{N}
\]

yields \( \varepsilon > \frac{3}{5} \frac{1}{2} \), which contradicts \( \varepsilon < \frac{1}{50} \). Therefore, the limit \( \varphi \) having a decomposition in \( G \) can not have a decomposition in \( F \).

We next state some sufficient conditions which ensure the existence of a measurable decomposition of a limit point \( \varphi \in F \), i.e.,

\[ \varphi = f \oplus g \in L^0(\mu_1) \oplus L^0(\mu_2). \]

If \( \varphi = f \oplus g 1_C = \tilde{f} \oplus \tilde{g} 1_C \in G \) are two different representations in \( G \), then \( f 1_{C_1} = \tilde{f} 1_{C_1} + \alpha 1_{C_1} \) and \( g 1_{C_2} = \tilde{g} 1_{C_2} - \alpha 1_{C_2} \) for some \( \alpha \in \mathbb{R} \) and all \( i \in I \), i.e., the additive representation of \( \varphi \) is unique up to constants on the components \( C_1 \times C_2 \) and we call \( \tilde{f}, \tilde{g} \) a modification of \( f, g \).

Define the \( \mu \)-completions of \( A \) and \( A_i \),

\[ A' := (A_1 \otimes A_2)_\mu, \]

\[ A'_1 := \{ A_1 \subset E_1; A_1 \times E_2 \in A' \}, \]

\[ A'_2 := \{ A_2 \subset E_2; E_1 \times A_2 \in A' \}. \] (2.10)

Then \( A'_1 \otimes A'_2 \) is the largest product-\( \sigma \)-algebra in \( A' \), and \( A'_i \) depend on \( \mu \), not merely on the marginal measures. \( \mu \) extends uniquely to a measure \( \mu' \)
on $A'$. Let $\mu'_i$ denote the marginals of $\mu'$ on $A'_i$ and

$$F' := L^0(\mu'_1) \oplus L^0(\mu'_2).$$  \hfill (2.11)

The proof of the following closedness theorem and proposition shows that centering at the conditional median is the appropriate tool to obtain closedness. We formulate however the theorem with an easy to verify sufficient condition. For the appropriate and nearly necessary but not in general easy to check condition for closedness see the following remarks.

**Theorem 2.1.** Suppose $\varphi_n = f_n \oplus g_n \in F'$ is $\mu'$-stochastically convergent to $\varphi$. If

$$\sup_n |\varphi_n| \text{ is real-valued,}$$

then $\varphi$ is in $F'$.

**Proof.** We adhere to the notation in the proof of Proposition 2.1. Without loss of generality, $f_n \oplus g_n \to \varphi$ ($\mu'$-a.e.).

1. We can assume that $\sum_{i \in J} C^1_i = E_1$ and $\sum_{i \in J} C^2_i = E_2$. Indeed, by $C \subset \pi_1(C) \times E_2$, and $\mu'(C) = 1$ it follows that $\pi_1(C) \times E_2 \in A'$, hence $\pi_1(C) \in A'_1$ and $\pi_2(C) \in A'_2$ as well. Thus $\varphi(x,y) = f(x)\mathbb{1}_{\pi_1(C)}(x) + g(y)\mathbb{1}_{\pi_2(C)}(y)$ ($\mu'$-a.e.) and $E_1 = \sum_{i \in J} C^1_i + (\pi_1(C))^c$, $E_2 = \sum_{i \in J} C^2_i + (\pi_2(C))^c$.

2. Define $T_1: (E_1, A'_1) \to (I, J)$, $T_1(x) = i$ if $x \in C^1_i$, and $J$ the $\sigma$-algebra generated by the one-point sets $\{i\}$, $i \in I$. Let $F_n(i, \cdot)$ denote the (right continuous) distribution function of the conditional distribution $\mu'_n|T_1=i$, and let $F^{-1}_n(i, \cdot)$ denote the (left continuous) inverse distribution function (the quantile function) of $F_n(i, \cdot)$. Then $m_n(i) := F^{-1}_n(i, \frac{1}{2})$ is the smallest conditional median of $f_n$ given $T_1 = i$. Since $F^{-1}_n(i, \cdot)$ is left continuous $\forall i \in I$, and $F^{-1}_n(\cdot, u)$ is $J$-measurable $\forall u \in \mathbb{[0,1]}$, we obtain that $F^{-1}_n$ is $J \otimes (\mathbb{B} \cap [0,1])$-measurable, where $\mathbb{B} \cap [0,1]$ is the Borel $\sigma$-algebra on $[0,1]$. This implies that $m_n$ is $J$-measurable and, therefore, $m_n \circ T_1$ is $\sigma(T_1)$-measurable.

3. Let $(x_i, y_i) \in C^1_i \times C^2_i$, $i \in I$, be as in the proof of Proposition 2.1. Then

$$m(i) := \lim_n (m_n(i) - f_n(x_i)) \text{ is finite } \forall i \in I.$$  \hfill (2.13)

Indeed, if $\psi := \sup |\varphi|_n$ and $(x_0, y_0) \in E$, then by assumption $|f_n(x) + g_n(y_0)| \leq \psi(x_0, y_0)$ and $|f_n(x_0) + g_n(y)| \leq \psi(x_0, y) \forall x \in E_1$, $y \in E_2$. Consider the modification $\tilde{f}_n(x) = f_n(x) - \frac{1}{2}(f_n(x_0) - g_n(y_0))$, $\tilde{g}_n(y) = g_n(y) + \frac{1}{2}(f_n(x_0) - g_n(y_0))$. Then $|\tilde{f}_n| \leq F := \psi(\cdot, y_0) + \frac{1}{2}\psi(x_0, y_0) \in L^0(\mu'_1)$, $|\tilde{g}_n| \leq G := \psi(x_0, \cdot) + \frac{1}{2}\psi(x_0, y_0) \in L^0(\mu'_2)$ and $\varphi_n = \tilde{f}_n \oplus \tilde{g}_n$. Hence we assume without loss of generality that $f_n = \tilde{f}_n$ and $g_n = \tilde{g}_n$. Then $|m_n(i)| = |\text{med}(f_n \mid T_1 = i)| \leq \text{med}(F \mid T_1 = i) < \infty$ and $|f_n(x_i)| \leq F(x_i) < \infty \forall n \in \mathbb{N}$, $\forall i \in I$. This yields that $m(i) = \lim_n (m_n(i) - f_n(x_i))$ is finite.

4. Define

$$\hat{f}(x) := \lim_n (f_n(x) - m_n \circ T_1(x)) = f(x) - m \circ T_1(x).$$  \hfill (2.14)
By (2.7) and (2.13) the function $\hat{f}$ is real valued and $\hat{f}$ is $A'_1$-measurable, since $f_n$, $m_n \circ T_1$ are $A'_1$-measurable. Define $T_2$: $(E_2, A'_2) \rightarrow (I, J)$, $T_2(y) = i$ if $y \in C_i^2$. Then $m_n \circ T_1(x) = m_n \circ T_2(y) \forall (x, y) \in C$. Since

$$\hat{g}(y) := \lim(g_n(y) + m_n(i))$$
$$= \lim(g_n(y) + f_n(x) - f_n(x) + m_n(i))$$
$$= \varphi(x, y) - \hat{f}(x) \quad \text{for all } (x, y) \in C_i^1 \times C_i^2,$$
and all $i \in I$

we obtain the $A'_2$-measurable function

$$\hat{g}(y) = \lim(g_n(y) + m_n \circ T_2(y)). \quad (2.15)$$

and, by (2.14), (2.15), the desired decomposition

$$\varphi \cdot 1_C = (\hat{f} + \hat{g}) \cdot 1_C, \quad \text{i.e., } \varphi \in L^0(\mu_1') \otimes L^0(\mu_2').$$

Let us point out that if a measurable decomposition of the limit point $\varphi$ in Proposition 2.1 exists at all, then the centering on the conditional median (as in the proof of Theorem 2.1) yields a measurable decomposition. To be precise:

**Proposition 2.2.** Let $\varphi(x, y) \cdot 1_C(x, y) = (f(x) + g(y)) \cdot 1_C(x, y)$ be a limit point of $L^0(\mu_1') \otimes L^0(\mu_2')$ in $L^0(\mu)$ as in Proposition 2.1. If there exists a measurable decomposition $\varphi \cdot 1_C = \hat{f} \otimes \hat{g} \cdot 1_C \in L^0(\mu_1') \otimes L^0(\mu_2')$, then

$$\tilde{f} := f - \sum_{i \in I} \text{med}(f \cdot 1_{C_i^1} | T_1 = i) \cdot 1_{C_i^1} \quad \text{is $A'_1$-measurable,}$$

$$\tilde{g} := g + \sum_{i \in I} \text{med}(f \cdot 1_{C_i^2} | T_1 = i) \cdot 1_{C_i^2} \quad \text{is $A'_2$-measurable, and}$$

$$\varphi \cdot 1_C = \tilde{f} \otimes \tilde{g} \cdot 1_C,$$

where $T_1$: $(E_1, A_1) \rightarrow (I, J)$, $T_1(C_i^1) = i$, and $J$ is generated by the one-point sets.

**Proof.** The additive representation of $\varphi \cdot 1_C$ is unique up to constants on the components $C_i^1 \times C_i^2 \in A_1 \otimes A_2$, hence $\tilde{f} = f + \sum_{i \in I} \alpha_i1_{C_i^1}$ for certain $\alpha_i \in \mathbb{R}$, $i \in I$. Since $f \cdot 1_{C_i^1}$ is $A_1$-measurable, we obtain

$$\hat{f} \cdot 1_{C_i^1} - \text{med}(\hat{f} \cdot 1_{C_i^1} | T_1 = i) \cdot 1_{C_i^1}$$
$$= f \cdot 1_{C_i^1} + \alpha_i1_{C_i^1} - \text{med}(f \cdot 1_{C_i^1} + \alpha_i1_{C_i^1} | T_1 = i) \cdot 1_{C_i^1}$$
$$= f \cdot 1_{C_i^1} - \text{med}(f \cdot 1_{C_i^1} | T_1 = i) \cdot 1_{C_i^1},$$

and, therefore,

$$\tilde{f} = \hat{f} - \text{med}(\hat{f} | T_1) = f - \sum_{i \in I} \text{med}(f \cdot 1_{C_i^1} | T_1 = i) \cdot 1_{C_i^1}.$$
is $A_1$-measurable. Define
\[
\overline{g}(y) := g(y) + \sum_{i \in I} \text{med}(f 1_{C_i} | T_1 = i) 1_{C_i}.
\]

Then $\overline{g}$ is $A_2$-measurable. Indeed, $\overline{g}(y)1_C(x, y) = (\varphi(x, y) - \overline{f}(x))1_C(x, y)$ is $A_1 \otimes A_2$-measurable and for any $\alpha \in \mathbb{R}$ it holds $\{\overline{g}1_C \leq \alpha\} = E_1 \times \{\overline{g} \leq \alpha\} \cap C \subset E_1 \times \{\overline{g} \leq \alpha\} \subset E_1 \times \{\overline{g} \leq \alpha\} \cap C + C^c$ with $\mu(C^c) = 0$. We obtain by definition of $A_1$ that $\{\overline{g} \leq \alpha\} \in A_1$. Hence $\overline{g}$ is $A_2$-measurable and $\varphi1_C = \overline{f} \otimes \overline{g}1_C$.

Let $\psi$ be a real valued function on $E$, i.e., $\psi \in M(E)$. Then denote by $F'_\psi$ the subset of $F'$ bounded by $\psi$, i.e.,
\[
F'_\psi = \{ h \in L^0(\mu'_1) \oplus L^0(\mu'_2); |h| \leq \psi \}.
\]

**Corollary 2.1.** $F'_\psi$ is closed in $L^0(\mu')$ with respect to $\mu'$-stochastic convergence.

**Proof.** For any limit point $\varphi$ of $F'_\psi$ there exists a sequence $\varphi_n$ in $F'_\psi$ such that $\varphi_n \to \varphi$ ($\mu'$-a.e.). Since $\sup_n |\varphi_n| \leq \psi$, the assertion follows from Theorem 2.1.

**Corollary 2.2.** Suppose $(\varphi_n)$ is a sequence in $L^1(\mu'_1) \oplus L^1(\mu'_2) \subset L^1(\mu')$ and $|\varphi_n| \leq \psi \in M(E)$.

a) If $\varphi_n \to \varphi$ in $L^1(\mu)$, then $\varphi \in L^0(\mu'_1) \oplus L^0(\mu'_2)$. Moreover, $\varphi \in L^1(\mu'_1) \oplus L^1(\mu'_2)$ if $\psi(\cdot, y_0) \in L^1(\mu'_1)$ and $\psi(x_0, \cdot) \in L^1(\mu'_2)$ for some $(x_0, y_0) \in E$.

b) If $\varphi_n \to \varphi$ with respect to the weak topology $\sigma(L^1(\mu'), L^\infty(\mu'))$, then $\varphi \in L^0(\mu'_1) \oplus L^0(\mu'_2)$.

**Proof.** a) follows from Theorem 2.1. The integrability follows from majorized convergence.

b) follows from a) observing that $(L^1(\mu'_1) \oplus L^1(\mu'_2)) \cap \{ h \in L^1(\mu'); |h| \leq \psi \}$ is convex and, therefore, the weak and the norm-closure of this set coincide.

**Remarks.** a) The limit $\varphi$ in Corollary 2.2 a) is not in $L^1(\mu'_1) \oplus L^1(\mu'_2)$ in general (cf. [18]).

b) An inspection of the proof of Theorem 2.1 shows that the boundedness condition (2.12) is only used to prove the finiteness condition (2.13), which is violated in the counterexample. So we can reformulate Theorem 2.1 in a more general form substituting (2.12) by (2.13), but in general it seems to be difficult to check the latter condition directly. By Proposition 2.2 it is seen that condition (2.13) is «nearly» necessary, too.

c) Let $L^0_{++}(\mu'_1) := \{ f \in L^0(\mu'_1); f > 0 \ (\mu'_1 \text{-a.e.}) \}$ and define the product $f \otimes g(s, t) := f(s)g(t)$. Then the limit points of sequences in $L^0_{++}(\mu'_1) \otimes L^0_{++}(\mu'_2)$ bounded in $L^0_{++}(\mu')$ are in $L^0_{++}(\mu'_1) \otimes L^0_{++}(\mu'_2)$ again. The proof is immediate from Theorem 2.1 by using logarithms.
(2.18) (cf. [19]). Consider the following example:

\[ E_1 = E_2 = \{0, 1\}, \quad 0 < \varepsilon < \frac{1}{4}, \]

\[ \mu := \left( \frac{1}{2} - \varepsilon \right) (\delta_{\{(0,0)\}} + \delta_{\{(1,1)\}}} + \varepsilon (\delta_{\{(0,1)\}} + \delta_{\{(1,0)\}}) \]  

and \( \varphi(x, y) := f(x) + g(y) \) with \( f(0) = g(0) = 0, f(1) = -g(1) = 1 \). Then

\[ S_1\varphi(x) = \text{med}_\mu(\varphi | \pi_1 = x) = 0 \quad \text{and} \]

\[ S_2\varphi(y) = \text{med}_\mu(\varphi | \pi_2 = y) = 0. \]  

In particular, the alternating projection algorithm does not converge to an optimal approximation and it seems impossible to identify \( f, g \) in terms of \( S_1, S_2 \). The optimal approximation is given by the projection equations

\[ S_1(\varphi - g) = f, \quad S_2(\varphi - f) = g. \]  

But it is not clear how to solve them.

Theorem 2.1 shows that the limit point in \( L^0(\mu') \) of a sequence \( f_n \oplus g_n \) in \( F' \) satisfying (2.12) or (2.13) is of the form \( f \oplus g \), where \( f, g \) are measurable only with respect to the \( \mu \)-completions \( A'_1 \). Therefore, let us point out that even in fairly general situations equality \( A'_1 = A_1 \) holds if \( A_i \) are assumed to be complete in the usual sense. Concerning the notion of perfect measure spaces we refer to [16, Definition 2.1].

**Theorem 2.2.** If \( (E_i, A_i, \mu_i) \) are complete perfect measure spaces, then \( A_i = A'_i \). Therefore, given \( \psi \in M(E) \), the set \( F_\psi \subset (L^0(\mu_1) \oplus L^0(\mu_2)) \) is closed in \( L^0(\mu) \).

**Proof.** A basic result in perfect measure spaces (cf. [16, Theorem 12.2.1]) implies that \( \mu_*(A_1 \times E_2) = (\mu_1)_*(A_1) \) for any \( A_1 \subset E_1 \), where \( \mu_*, (\mu_1)_* \) are the inner measures of \( \mu, \mu_1 \). Given \( A_1 \in A'_1, (\mu_1)_*(A_1) + (\mu_1)_*(A'_1) = \mu_*(A_1 \times E_2) + (\mu_1)_*(A'_1 \times E_2) = \mu'(A_1 \times E_2) + \mu'(A'_1 \times E_2) = 1 \). Hence, there are \( B_i \in A_1, B_1 \subset A_1, B_2 \subset A'_1 \) with \( \mu_1(B_1) + \mu_1(B_2) = 1 \). Therefore, \( B_1 \subset A_1 \subset B_2 \) and \( \mu_1(B_1) = \mu(B_2) \). This implies \( A_1 \in A_1 \), since \( A_1 \) is \( \mu_1 \)-complete.

**Remark.** Note that perfectness of a measure space is preserved by completion (cf. [16, 2.2, p. 5]). In particular, Theorem 2.2 implies the closedness property of \( F_\psi \) for polish spaces and, more general, for analytic spaces. The necessity to enlarge \( A_i \) to \( A'_i \) in Theorem 2.1 arises from the fact that
projections of product measurable sets onto the marginal spaces are known to be in the completions of the marginal spaces only under additional assumptions like analyticity. Therefore, we conjecture that the enlargement of $A_i$ to $A_i'$ in Theorem 2.1 can not be omitted in general.

3. On the Generalized Schrödinger Problem

Let $M(\mu_1, \mu_2)$ be the set of all probability measures on $(E_1 \times E_2, A_1 \otimes A_2)$ with marginals $\mu_i$. Let $I(\nu \mid \mu) = \int \ln(d\nu/d\mu) d\nu$ denote the Kullback-Leibler distance and define $\nu^* \in M(\nu_1, \nu_2)$ to be the $I$-projection of $\mu$ on $M(\nu_1, \nu_2)$ if

$$I(\nu^* \mid \mu) = \inf \{I(\nu \mid \mu); \nu \in M(\nu_1, \nu_2), \nu \ll \mu\} =: m. \quad (3.1)$$

From the closedness results in Section 2 we obtain the existence of a solution of the generalized Schrödinger equation (1.1), (1.3), if $m$ is finite.

**Theorem 3.1.** Assume that $\nu_i \ll \mu_i$, $i = 1, 2$, and $m < \infty$.

a) There exists a uniquely determined $I$-projection $\nu^*$ on $M(\nu_1, \nu_2)$.

b) There exists $a \otimes b \in L^0_+(\nu^*)$ such that $(d\nu^*/d\mu)(x, y) = a(x)b(y)$, $(\nu^*-a.e.)$.

**Proof.** a) $M(\nu_1, \nu_2)$ is closed in variation norm and so the assertion follows from [5, Theorem 2.1].

b) By Theorem 3.1 of [5], $\ln(d\nu^*/d\mu)$ belongs to the closure of $L^1(\nu_1) \oplus L^1(\nu_2)$ in $L^1(\nu^*)$. By Proposition 2.1, therefore, $\ln(d\nu^*/d\mu)(x, y) = f(x) + g(y)$ $(\nu^*-a.e.)$ for some $f \oplus g \in G$ and the result follows with $a = \exp f$, $b = \exp g$.

**Remarks.** a) The functions $a$, $b$ are not unique in general, but only the product $a \otimes b$ is unique $\nu^*$-a.e. It is easily seen that $a$ and $b$ are unique up to constants if and only if any two points $(x', y')$ and $(x'', y'')$ in the support $A$ of $\nu^*$ are finitely connected in $A$, i.e., there are finitely many points $(x_1, y_1) = (x', y'), (x_2, y_2), \ldots, (x_n, y_n) = (x'', y'')$ in $A$, such that $x_i = x_{i+1}$ or $y_i = y_{i+1}$ for all $i = 1, \ldots, n - 1$. For example, this is true if $\mu = h\mu_1 \otimes \mu_2$, $h > 0$, $\mu_1 \otimes \mu_2$-a.e.

b) Theorem 3.1 is of importance for the derivation of characterizations and properties of (generalized) Schrödinger bridges. To be precise, let $\nu_0$ be the Wiener measure on $S := C([0, T], \mathbb{R}^d)$, $0 < T < \infty$. Furthermore, with $\Omega = S^N$ and with $P = P_0^\mathbb{N}$ let $X_i$ be the projections. Föllmer [7] noted that Schrödinger bridges have a large deviation interpretation. If $\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ is the empirical measure of $X_1, \ldots, X_n$, then, by Sanov's large deviation theorem, $(\mu_n)$ satisfies a LD-principle with rate function $I(Q_0 \mid P_0)$.

For a subset $\Gamma$ of probability measures on $S$ it holds:

$$P\{\mu_n \in \Gamma\} \sim \exp \left\{ -n \inf_{Q_0 \in \Gamma} I(Q_0 \mid P_0) \right\}. \quad (3.2)$$
If $\mu = P_0(X(0), X(T))$ is the joint distribution of $(X(0), X(T))$ with respect to Wiener measure and
\[
\Gamma := \{ Q_0 \in M^1(S); Q_0^X = \nu_1, Q_0^{X(T)} = \nu_2 \}
\] (3.3)
with $\nu_i \ll \mu_i$, then $\Gamma$ describes large deviations at the initial and final time points.

For any $Q_0 \in \Gamma$ it holds
\[
I(Q_0 \mid P_0) = I(\nu \mid \mu) + \int I(Q_0^x \mid P_0^x) \nu(dx, dy),
\] (3.4)
where $\nu = Q_0(X(0), X(T))$ and $Q_0^x, P_0^x$ are the conditional distributions of $Q_0, P_0$ given $X(0) = x, X(T) = y$. This implies
\[
I(Q_0 \mid P_0) \geq I(\nu \mid \mu) \geq I(\nu^* \mid \mu) = I(P^* \mid P_0),
\] (3.5)
where $P^* := \int P_0^y \nu^*(dx, dy)$ is the Schrödinger bridge. This minimum entropy interpretation of $P^*$ yields via Sanov’s theorem (3.2) that $P^*$ is «favoured» in the sense of large deviations by the empirical distribution in $\Gamma$.

Related results can be derived for modifications of $\Gamma$ (like Prokhorov-type neighbourhoods) and also for general $P_0$, replacing the Wiener measure, and thus leading to generalized Schrödinger bridges $P^*$. For the ample literature on these kind of problems we refer to [7], [14], [6], [4], [2], [20], [3], [11].

The factorization property $(d\nu^* d\mu)(x, y) = a(x) b(y), (\nu^*-a.e.)$ is crucial for the derivation of characteristic properties of the Schrödinger bridges (like the (reciprocal) Markov property, diffusion property, $h$-path process). Theorem 3.1 gives a product form of the density but in general with nonmeasurable factors $a, b$ only. It is however not clear whether this factorization is enough to derive the above mentioned properties of Schrödinger bridges.

4. Some Extensions and Applications

Clearly, the statements in Section 2 carry over to product spaces of $k \in \mathbb{N}$ factors. In particular, let $(E, A) = \otimes_{i=1}^k (E_i, A_i)$, let $\mu$ be a probability measure on $(E, A)$ with marginals $\mu_i$ and $\psi \in M(E)$.

**Proposition 4.1.** a) $(\otimes_{i=1}^k M(E_i)) \cap L^0(\mu')$ is closed in the metric space $L^0(\mu')$ with respect to $\mu'$-stochastic convergence.

b) $(\otimes_{i=1}^k L^0(\mu_i'))\psi$ is closed in $L^0(\mu')$.

For another aspect let $(\Omega, B, P)$ be a complete probability space. Let $B_1, \ldots, B_k \subset B$ be sub-$\sigma$-algebras of $B$ containing the $P$-null sets. $P_i := P \mid B_i$ denote the restriction of $P$ to $B_i$ and $L^0(P_i) = L^0(\Omega, B_i, P_i)$ are the real-valued $B_i$-measurable functions. Define $\mu$ on $\otimes_{i=1}^k (\Omega, B_i)$ by $\mu(B_1 \times \cdots \times B_k) = P(B_1 \cap \cdots \cap B_k)$ and let $T_i : (\Omega, B) \longrightarrow (\Omega, B_i), T_i = \text{id}_\Omega$. Note that
Closedness of Sum Spaces and the Generalized Schrödinger Problem

p(T_1,...,T_k) = \mu, P(T_i) = P_i = \mu_i, and that B_i coincides with its \mu-completion. This leads to the following

Proposition 4.2. \((\otimes_{i=1}^k L^0(P_i)\psi)\) is closed in \(L^0(P)\) for any \(\psi \in M(\Omega)\).

Proof. Let \(\varphi\) be a limit point and \(\varphi_n = \sum_{i=1}^k h_{ni}\) be a sequence in \((\otimes_{i=1}^k L^0(P_i)\psi)\) such that \(\varphi_n \rightarrow \varphi\) (\(P\)-a.e.). Clearly, \(h_{ni}\) are \(\sigma(T_i) = \mathcal{B}_i\)-measurable and \(\varphi\) is \(\sigma(T_1,...,T_k) = \sigma(\bigcup_{i=1}^k \mathcal{B}_i)\)-measurable. By the factorization theorem, there are measurable functions \(k_n : \Omega, \mathcal{B}_i \rightarrow [0,\infty)\) such that \(k_n \circ T_i = h_{ni}\) and \(k \circ (T_1,...,T_k) = \varphi\). This yields \(P\{\lim_{n \rightarrow \infty} \sum_{i=1}^k k_{ni} = \varphi\} = \mu\{\lim_{n \rightarrow \infty} \sum_{i=1}^k k_{ni} = k\},\) since \(P(T_i) = \mu_i\). By assumption, \(|\sum_{i=1}^k k_{ni}(\omega_i)| \leq \sum_{i=1}^k \psi(\omega_i) < \infty\) for all \((\omega_1,...,\omega_k) \in \Omega^k\), hence Theorem 2.1 or Proposition 4.1 yields \(k(\omega_1,...,\omega_k) = \sum_{i=1}^k k_i(\omega_i)\) for some \(k_i \in L^0(\mu_i) = L^0(P_i)\). This implies \(\varphi = \sum_{i=1}^k h_{ni}, (P\text{-a.e.}), \) where \(h_i = k_i \circ T_i\).

Remark. In particular, Proposition 4.2 applies to functions of several variables like \(\sum_{i=1}^k f(x_{i+1})\) or as ridge regression theory to functions of the type \(\sum_{i=1}^k f(A_i x_{i+1})\), where \(A_i\) are some fixed functions on \(E\).

From Theorem 3.1 and these propositions we obtain an interesting structure theorem for distributions with multivariate marginals. Let \((E,\mathcal{A}) = \otimes_{i=1}^k (E_i,\mathcal{A}_i)\), let \(\mathcal{H} \subset \mathcal{P}\{1,...,k\}\) be a system of subsets of \(\{1,...,k\}\) and let \(\mu \in M^1(E,\mathcal{A})\) have \(H\)-marginals \(\mu_H := \mu_n\), \(H \in \mathcal{H}\), where \(\pi_H : E \rightarrow \prod_{j \in H} E_j =: E_H\) is the projection on the \(H\)-component. Let \(\nu_H \in M^1(E_H,\mathcal{A}_H), \mathcal{A}_H := \otimes_{j \in H} \mathcal{A}_j\), be continuous with respect to \(\mu_H, H \in \mathcal{H}\), and let \(M_\mathcal{H} := M((\nu_H), H \in \mathcal{H})\) be the set of all probabilities on \((E,\mathcal{A})\) with \(H\)-marginals \(\nu_H, H \in \mathcal{H}\). The following proposition gives an interesting relation between \(M(\mu_H, H \in \mathcal{H})\) and \(M(\nu_H, H \in \mathcal{H})\).

Proposition 4.3. If \(M_\mathcal{H} := M(\nu_H, H \in \mathcal{H}) \neq \emptyset\) and \(\inf I(\nu | \mu); \nu \in M_\mathcal{H} < \infty, then there exists \(\otimes_{H \in \mathcal{H}} a_H \mu \in L^0(\mu)\) such that \(\nu = (\otimes_{H \in \mathcal{H}} a_H) \mu \in M_\mathcal{H}\), where \(\otimes_{H \in \mathcal{H}} a_H(x) = \prod_{H \in \mathcal{H}} a_H(x_H), x_H := \pi_H(x)\).

Proof. The proof is similar to that of Theorem 3.1 applying Proposition 4.1 with sub-\(\sigma\)-algebras \(\sigma(\pi_H) \subset \mathcal{A} = \otimes_{j=1}^k \mathcal{A}_j, H \in \mathcal{H}\).

The following proposition gives a basic existence result for additive statistical models. Consider \((E,\mathcal{A}) = \otimes_{i=1}^k (E_i,\mathcal{A}_i)\) and a probability measure \(\mu\) on \((E,\mathcal{A})\) with marginals \(\mu_i\). In additive statistical models the systematic part of the observations is assumed to be in \(F^\alpha := \otimes_{i=1}^k L^\alpha(\mu_i) \subset L^\alpha(\mu)\), where \(0 < \alpha \leq \infty\).

Proposition 4.4. a) For any \(0 < \alpha < \infty\) there exists a probability space \((E,\mathcal{A},\mu)\) such that \(F^\alpha\) is not closed in \(L^\alpha(\mu)\).

b) For \(1 < \alpha < \infty\) and any \(\varphi \in L^\alpha(\mu)\) there exists a unique optimal approximation \(\bar{\varphi}\) of \(\varphi\) in the closure of \(F^\alpha\) in \(L^\alpha(\mu)\) and \(\bar{\varphi} \in (\otimes_{i=1}^k M(E_i)) \cap L^0(\mu)\).

Proof. a) If \(0 < \alpha < \infty\), then consider \(E_1 = E_2 = \mathbb{N}\) and probabilities \(\mu_1 = \mu_2\) with density \(p(2n-1) = p(2n) = \frac{1}{2}2^{-n}\) and the probability measure \(\mu \ll \mu_1 \otimes \mu_2\) with density \(q(2n,2n-1) = q(2n-1,2n) = \frac{1}{2}2^{-2n}\),
Ruschendorf L., Thomsen W.

$q(2n, 2n) = q(2n - 1, 2n - 1) = \frac{1}{2}(2^{-n} - 2^{-2n})$. $\mu$ has marginals $\mu_1, \mu_2$. Define $f, g: \mathbb{N} \rightarrow \mathbb{R}^1, f(2n - 1) = 2^{n/\alpha}, f(2n) = 0$, and $g(2n - 1) = -2^{n/\alpha}, g(2n) = 0$. Then $f \notin L^\alpha(\mu_1), g \notin L^\alpha(\mu_2)$ but $\varphi = f \oplus g \in L^\alpha(\mu)$ and $\int |\varphi|^\alpha d\mu = 1$. With $f_n := f 1_{\{1, \ldots, n\}}, g_n := g 1_{\{1, \ldots, n\}}, f_n \in L^\alpha(\mu_1), g_n \in L^\alpha(\mu_2)$ and $\varphi_n := f_n \oplus g_n \rightarrow \varphi$ in $L^\alpha(\mu)$. Therefore, $\varphi$ is in the closure of $F^\alpha$ in $L^\alpha(\mu)$. For any representation $\varphi = f' \oplus g'(\mu\text{-a.e.})$, it holds $f'(i) = f(i) + c_i, g'(j) = g(j) - c_j$ for some constants $c_i$ and $(i, j)$ with $q(i, j) > 0$. With $d_n := c_{2n-1} = -c_{2n}$ and using the inequality $|a + b|^\alpha \leq 2^\alpha(|a|^\alpha + |b|^\alpha)$ we obtain $\int |f'|^\alpha d\mu_1 = \sum_{n=1}^{\infty} (|2^n|^\alpha + d_n|\alpha\cdot \frac{1}{2}\cdot 2^{-n} + | - d_n|^\alpha\frac{1}{2}2^{-n}) = \infty$. This implies that $\varphi$ is not in $F^\alpha$.

If $\alpha = \infty$, then consider $E_1 = E_2 = \mathbb{R}^1, A_1 = A_2 = B^1$. Let $Q_1 = Q_2 = N(0, 1), A = \{(x, y); x - 1 \leq y \leq x + 1\} \in B^2$ and let $\mu$ be the product $Q_1 \otimes Q_2$ conditional on $A$. Furthermore, let $f = \text{id}_{\mathbb{R}^1}, g = -\text{id}_{\mathbb{R}^1}$ and $\varphi := f \oplus g$. $\varphi = f \oplus g$ is the unique measurable additive decomposition of $\varphi$ but $f \notin L^\infty(\mu_1), g \notin L^\infty(\mu_2), \varphi \in L^\infty(\mu)$, since $|\varphi| \leq 1 (\mu\text{-a.e.}).$

Let $(\varphi_n) \subset F^\alpha$ be a sequence such that $||\varphi - \varphi_n||_\alpha \rightarrow \inf\{||\varphi - h||_\alpha; h \in F^\alpha\} =: a$. Then $(\varphi_n)$ is norm-bounded and, therefore, relatively weakly sequentially compact. Since $F^\alpha$ is convex the norm-closure and the weak closure coincide and, therefore, $||\varphi - \varphi|| = a$ for any accumulation point $\varphi$ of $(\varphi_n)$. Obviously, $\varphi$ is uniquely determined and by Proposition 4.1, $\varphi \in (\oplus_{\alpha=1}^\infty M(E_i)) \cap L^\alpha(\mu)$ and $\varphi$ is the optimal approximation of $\varphi$.

The final application will be an extension of a famous representation theorem of Kolmogorov (1957) for continuous functions $f: [0, 1]^n \rightarrow \mathbb{R}$ of $n$ variables, disproving conjecture no. 13 of Hilbert's list of 23 problems. Kolmogorov's theorem states the existence of real continuous functions $\alpha_{ij}: [0, 1] \rightarrow \mathbb{R}^1, 1 \leq i \leq 2n + 1, 1 \leq j \leq n$, such that any continuous function $f = f(x_1, \ldots, x_n)$ has a representation of the form

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{2n+1} g \left( \sum_{j=1}^{n} \alpha_{ij}(x_j) \right) \quad (4.1)$$

with some continuous functions $g$ depending on $f$. The functions $\alpha_{ij}$ even may be chosen to be Lipschitz-continuous, and $(\sum_j \alpha_{ij})$ to be one-to-one. As reference for this result and several modifications we refer to Chapter 11 of [13]. (4.1) is considered to be a theoretical justification of neural networks. We obtain a «generalization» to measurable functions as follows.

**Proposition 4.5. Let $\mu \in M^1(\mathbb{R}^n, B^\alpha)$ and let $f \in L^0(\mu)$ be locally bounded. Then**

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{2n+1} g \left( \sum_{j=1}^{n} \alpha_{ij}(x_j) \right) \quad (\mu\text{-a.e.})$$

**for some measurable functions $g: \mathbb{R}^1 \rightarrow \mathbb{R}^1$ and $\alpha_{ij}: \mathbb{R}^1 \rightarrow \mathbb{R}^1$ piecewise Lipschitz-continuous.**
Proof. Since $f$ is assumed to be locally bounded on $\mathbb{R}^n$, it is majorized by a continuous function $h$ on $\mathbb{R}^n$, $|f| \leq h$. Let without loss of generality $f \geq 0$, then decomposing $\mathbb{R}^n$ into disjoint unions of cubes we can approximate $f$ by a sequence $0 \leq f_m \leq n$ of continuous functions with respect to $\mu$-stochastic convergence. By (4.1) $f_m$ has a representation of the form

$$f_m(x) = \sum_{i=1}^{2n+1} \sum_{j=1}^n g_m \left( \sum_{j=1}^n \alpha_{ij}(X_j) \right),$$

where $g_m$ are continuous on the cubes and, in particular, measurable on $\mathbb{R}^n$ and the $(\alpha_{ij})$ are as in (4.1). From the proof of (4.1) in [14] it is seen that a representing function $g = g_f$ in (4.1) can be chosen such that $f_1 \leq f_2$ implies $g_{f_1} \leq g_{f_2}$. This yields $0 \leq g_m \leq g_h$ for all $m \in \mathbb{N}$. Then, by Proposition 4.2 with $B_i = \sigma(T_i)$, Remark b) implies the existence of a measurable function $g$ such that $f(x) = \sum_{i=1}^{2n+1} g(\sum_{j=1}^n \alpha_{ij}(x_j))$ ($\mu$-a.e.).

Remarks. a) Kolmogorov's theorem is considered as a theoretical foundation of neural networks (cf. [9, p. 122–137], [17]). There are many papers on generalizations of this theorem to non continuous functions and to variants of the theorem with special kind of functions $\alpha_{ij}$ in the literature on neural networks. Proposition 4.5 improves the known representation and approximation results for neural networks like the back propagation theorem (cf. [9, p. 132] and references given there). For the first time it states an exact representation result for standard feed forward networks with one hidden layer. It was known up to now that two hidden layers are sufficient (cf. the discussion of this point in [17] and [9]).

b) By the monotonicity property of the representing function $g$ in (4.1) it is clear that the representation (4.1) holds for any continuous function $f$ on $\mathbb{R}^n$. Therefore, it is possible to extend the representation property to measurable functions majorized on open subsets $U_j$ by continuous functions $f_j$ such that $\mu(\partial U_j) = 0$, $(U_j)$ pairwise disjoint and $\cup U_j = \mathbb{R}^n$.

Acknowledgement. We are grateful to H. Föllmer and H. Kellerer for critical and stimulating remarks on this paper. N. Gantert informed us on the counterexample in Section 2.

REFERENCES

4. Brunaud M. Finite Kullback information diffusion laws with fixed marginals and

Поступила в редакцию 10.II.1995