2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

A Novel Low-Complexity Method for Determining Nonadditive Interaction Measures Based on Least-Norm Learning

Wei An, Chunxiao Ren, Song Ci*, Dalei Wu, Haiyan Luo and Yanwei Liu

***Abstract*— Numerous research works have been done on the Choquet integral model due to the tremendous usage in many fields. However, the application is still significantly restricted by the curse of dimensionality, involved in determining the non-additive interaction measures, that can properly reflect the interactions among predictive attributes toward the objective. To this end, in this paper we propose a novel determination method for non-additive interaction measures by the way of solving a sequence of least norm problems and iteratively updating the values of interaction measures, namely least norm learning. This method can achieve a significant reduction on the computation time complexity from $O(m \times 2^n)$ to $O(mn)$ for solving the Choquet integral model, where m and n are the numbers of observations and attributes, respectively. Also we achieve to reduce the computation space complexity from $O(m \times 2^n)$ to $O(2^n)$. A case study on cross-layer optimized wireless multimedia communications is adopted to validate the proposed method. Both analytical and experimental results show the effectiveness of the proposed method.**

I. Introduction

The non-additive model based on Choquet integral (the *Choquet model* in short) has a great practical use in many fields, such as classification [1], multicriteria decision making [2], image and pattern recognition [3] and data modeling [4], due to its distinguished feature that the interaction among predictive attributes toward the objective attribute can be properly reflected through a set of non-additive measures. The Choquet integral is a generalization of the Lebesgue integral, defined over a set of non-additive measures (also called fuzzy measures). Let $X = \{x_1, x_2, \cdots, x_n\}$ be a set of attributes, $f(x)$ be the observated or partially evaluated value on each attribute $x \in X$, f be a tuple of observated or partially evaluated values on X, and y be an objective. The linear/additive multiregression model is traditionally represented as a weighted sum $y = \sum_{x\in X} w_x f(x)$, where the weight w_x is also regarded as a Lebesgue measure w on a singleton $\{x\}$, since the linear model is equivalent to a Lebesgue integral $y = (L)\int_X f dw$. Since the Choquet integral model breaks the restriction that the combined contribution of $\{x_i, x_j\}$ toward the objective y is the weighted sum of their respective contributions, and it uses a non-additive measure μ, which is defined over the powerset of X, and a Choquet integral, $y = (C)\int_X f d\mu$, it is clearly more powerful than the Lebesgue integral model because the non-additive measure μ considers the interaction among attributes toward the objective [5].

Wei An and Yanwei Liu are with the High Performance Network Lab, IOA, Chinese Academy of Sciences, Beijing 100190, China.

Chunxiao Ren is with the Administration Center, Shandong Academy of Information & Communication Technology, Jinan 250101, China.

Song Ci is with the Department of CEEN, University of Nebraska-Lincoln, NE 68182, USA, and the High Performance Network Lab, IOA, Chinese Academy of Sciences, Beijing 100190, China.

Haiyan Luo is with the Department of CEEN, University of Nebraska-Lincoln, NE 68182, USA.

Dalei Wu is with the Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

This work is supported in part by NSF under grant Nos.1145596 and 0830493, by NSFC under grant No. 61302031, and by Shandong Outstanding Young Scientists Foundation under Grant No. BS2013DX047.

*Corresponding author. Email: sci@engr.unl.edu

Numerous works have been done on the theory and application of the Choquet model over the past few decades. However, the practical use of the Choquet model is still significantly restricted due to the tremendous challenges when trying to use it in practical applications. One of the main challenges is how to determine the non-additive measure μ defined over the powerset of attributes X. Considering a set of sample data with the attribute set $\{x_1, x_2, \cdots, x_n\}$, according to the Choquet model, there are totally $2^n - 1$ non-additive measures, $\mu(\{x_1\})$, $\mu(\{x_2\})$, $\cdots$, $\mu(\{x_n\})$, $\mu(\{x_1, x_2\})$, $\cdots$, $\mu(\{x_1, x_n\})$, $\cdots$, $\mu(\{x_1, x_2, \cdots, x_n\})$, necessarily to be determined. As for non-additive measures, the basic idea to solve the Choquet model is to reduce the non-linear regression model to the traditional linear multiregression model, so that the Choquet model can be easily solved by using the least squares method in a quadratic running time [6]. Such idea was originally proposed in [7], and successfully applied on classification [8]. However, there is a problem that "bad" solutions are often generated [6], [4] because raw attribute data are often not good enough to identify all $2^n - 1$ coefficients. Thus, the obtained optimal solutions are sometimes too unreasonable to support the decision making. To get around such a bad-solution problem, a suboptimal algorithm, called HLMS (Heuristic Least Mean Squares) [9], is developed based on a gradient algorithm and the idea of equilibrium point, but the obtained suboptimal solution may be quite different from the optimal one. Literatures [10], [4] exploited genetic algorithm (GA) to determine the non-additive measures. However, due to randomness of chromosome generation and the huge 2^n dimensional search space, the obtained solutions at different running times are not unique, and actually quite different in general. Compared with these works utilizing the traditional heuristic strategies or genetic algorithms, our previous work [11], [12] proposed an efficient approach to determine the complete non-additive measure such that the solutions of the Choquet model is unique and deterministic.

Despite numerous works [5], [13], [14] having been done on solving the Choquet model, the use of non-additive measure in the practical application is yet seriously restricted by the exponential complexity [15], known as the curse of dimensionality, which involves solving a with time constraint problems which increase exponentially with respect to the number of attributes. In this paper, we consider the problem

978-1-4799-2072-3/14/$31.00 ©2014 IEEE

of exactly determining the non-additive interaction measures with efficient approaches. The main contributions of this paper are: 1) we develop a novel method, namely least norm learning, for efficient determination of non-additive interaction measures caused by the linear interactions among the predictive attributes on the basis of the Choquet integral; 2) the proposed method can significantly reduce the computation time complexity from $O(m \times 2^n)$ to $O(mn)$ for the Choquet integral, and remarkably reduce the computation space complexity from $O(m \times 2^n)$ to $O(2^n)$, where m is the number of observations and n is the number of predictive attributes; 3) we theoretically analyze the convergence in probability of our proposed algorithms; and 4) a case study on cross-layer optimized wireless multimedia communications is adopted to validate the effectiveness of our proposed method.

The rest of the paper is organized as follows. In Section II, we describe the non-additive interaction measure, restate our previous work [11], and analyze the existing challenges in our previous work. Section III devises efficient algorithms for determining non-additive interaction measures. In Section IV, a case study on cross-layer optimized wireless multimedia communications is conducted to validate the proposed method. Finally, Section V summarizes the paper.

II. The Choquet Model for Determining Non-Additive Measures

In this section, we give a brief introduction on non-additive measure, the Choquet model, and how to determine the non-additive measures with the Choquet model.

A. Non-Additive Measure

Let $X = \{x_1, x_2, \cdots, x_n\}$ be a set of attributes with $n = |X|$. A *fuzzy measure* on X is a set function $\mu : \mathcal{P}(X) \to \mathcal{R}$ with a constraint $\mu(\emptyset) = 0$, where $\mathcal{P}(X)$ is the powerset of X, and $\mathcal{R}$ is the real domain. According to our previous work [11], the following two traditional restrictions on fuzzy measures can be relaxed: (i) the co-domain of the set function μ is $\mathcal{R}$ instead of $\mathcal{R}^+$, where $\mathcal{R}^+$ represents the positive values in the real domain $\mathcal{R}$; (ii) the monotonicity, $A \subset B \subseteq X$ implies $\mu(A) \leq \mu(B)$, is not necessary.

The main characteristic of fuzzy measures is that they can express interactions among attributes being aggregated in a more flexible and precise manner. A fuzzy measure is said to be *additive* if $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever $A \cap B = \emptyset$, otherwise, it is *non-additive*. In the applications of multicreteria decision making [16], the case of $\mu(A \cup B) > \mu(A) + \mu(B)$ is often called *positive interaction* or *positive synergy* between criteria A and B; whereas the case of $\mu(A \cup B) < \mu(A) + \mu(B)$ is called *negative interaction* or *negative synergy*, where the union of criteria (or attributes) does not bring anything more to the objective [12], [11].

B. The Choquet Model

The interaction among attributes toward the objective can be properly measured through a non-additive fuzzy measure, which is the main feature of the Choquet integral model. Assume the data consists of m observations of the attributes $x_1, x_2, \cdots, x_n$ and the objective y, in the following form:

x_1	x_2	$\cdots$	x_n	y
f_{11}	f_{12}	$\cdots$	f_{1n}	y_1
f_{21}	f_{22}	$\cdots$	f_{2n}	y_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
f_{m1}	f_{m2}	$\cdots$	f_{mn}	y_q

where each row is an observation of attributes $x_1, x_2, \cdots, x_n$ and y. The observation of $x_1, x_2, \cdots, x_n$ can be regarded as a function $f : X \to \mathcal{R}$; hence the j-th observation of $x_1, x_2, \cdots, x_n$ is denoted by f_j, and we write $f_{ji} = f_j(x_i)$ where $1 \leq i \leq n$ and $1 \leq j \leq m$.

The interaction among predictive attributes X toward the objective y is described by a set function μ defined on the power set of X satisfying the condition of vanishing at the empty set, i.e., $\mu : \mathcal{P}(X) \to \mathcal{R}$ with $\mu(\emptyset) = 0$. The new non-additive multi-regression model is expressed as

$$y = e + \int_{(c)} f d\mu + \mathcal{N}(0, \delta^2), \qquad (1)$$

where e is a regression constant, $\int_{(c)}$ is the Choquet integral, f is an asseveration of $x_1, x_2, \cdots, x_n$, μ is a fuzzy measure, and $\mathcal{N}(0, \delta^2)$ is a normally distributed random perturbation with expection 0 and variance δ^2. The Choquet integral $\int_{(c)}$ of the data observation f, w.r.t. a fuzzy measure μ, is defined as:

$$\int_{(c)} f d\mu = \int_{-\infty}^{0} [\mu(F_\alpha) - \mu(X)] d\alpha + \int_{0}^{+\infty} \mu(F_\alpha) d\alpha \qquad (2)$$

where $F_\alpha = \{x \mid f(x) \geq \alpha\}$ for any $\alpha \in (-\infty, +\infty)$, and is called the α-cut set of f.

C. Precision Determination of Non-Additive Measures

According to literatures [6], [11], the basic idea to solve the Choquet model is a two-step procedure. The first step is to reduce the non-linear multiregression model to the traditional linear multiregression model by converting each n-dimensional vector attribute datum to a 2^n-dimensional vector datum, which is defined over the powerset of attributes, according to Eq.(2). The readers who are interested in the detail of the data transformation can refer to our previous work [11]. Then, for all m observations, we have the following linear equation system:

$$y = e + A\mu + \mathcal{N}(0, \delta^2), \qquad (3)$$

where $A = (a_{ij})_{m \times N}$, $N = 2^{n-1}$, and $\mu = (\mu(\{x_1\}), \cdots, \mu(\{x_n\}), \cdots, \mu(\{x_1, \cdots, x_n\}))^T$. $y = (y_1, y_2, \cdots, y_m)^T$ and y^T denotes the transpose of y. Based on the first step, the second step is to solve the linear model Eq.(3) by using the standard least-square method. For the given observation data, the optimal regression coefficients μ can be determined by using the least squares method in order to make δ^2 minimal. Then, we use μ_k to denote $\mu(X'_k)$, where X'_k is a subset of X, i.e. $X'_k \in \mathcal{P}(X)$. Readers who are interested in the detail procedure for determining $\mu_k (k = 1, 2, \cdots, N = 2^n - 1)$ can refer to [11].

In this algorithm, according to our previous work [11], constructing the $m \times N$ augmented matrix $\bar{A} = [a_{lk}]$ takes the time of $O(mN)$. The least-square solution to the linear equations Eq.(3) is usually calculated with

$$\mu = (A^T A)^{-1} A^T y, \tag{4}$$

where A^T represents the transpose of A. Thus, the computation complexity is $O(mN^2)$. Calculating the regression residual error $\hat{\delta}^2$ expends the time of $O(mN)$. Therefore, the overall computation complexity is $O(mN) + O(mN^2) + O(mN) = O(mN^2) = O(m \times 2^{2n})$, where $N = 2^n - 1$.

Once μ_k has been determined, we can identify which subset of attributes has the most significant impact on the objective. So we can further fine-tune those attributes to improve the objective under the current system and conditions.

D. Challenges

To achieve the optimal objective, the prerequisite is to evaluate contributions/interaction measures, made by one or multiple attributes in $X = \{x_1, x_2, \cdots, x_n\}$, to the objective y. Based on the derived interaction measures, the subset in $\mathcal{P}(X)$ making the dominating contributions can be found and used for further fine-tuning those attributes to improve the objective with the available systems and conditions. With this purpose, in the last subsection we provide an approach of precision determination of the non-additive interaction measures of attributions. It is worthy to be mentioned that this approach is targeted to address those application issues which involve not more than 10 attributes. Otherwise the $2^n - 1$ fuzzy measures made by the $n (n > 10)$ attributes is too large to determine by solving the Choquet model with the least squares method whose computation complexity increases exponentially as the number n of attributes increases. On the other hand, due to the limited computation capability and storage space of most current electronic devices, developing efficient approaches is an urgent requirement for quantitatively measuring the contribution of each subset of attributes in $\mathcal{P}(X)$.

III. Non-Additive Measure Determination with Least Norm Learning

In Section II-C, matrix $A \in R^{m \times N}$ $(N = 2^n - 1)$ is derived for the determination of fuzzy measures. Then, the least squares problem Eq.(3) can be written as the following equivalent form:

$$J_1 = \min_{\mu} \|A\mu - y\|, \tag{5}$$

where $\mu = (\mu_1, \mu_2, \cdots, \mu_N)^T$. According to [17], [18], the least squares problem Eq. (5) has the same solution to the following equation

$$A^T A \mu = A^T y. \tag{6}$$

Now, let us assume A is skinny, i.e., $m \geq N$. This is reasonable because to determine N non-additive measures, more than N observations are usually sampled. Thus, if A is full column rank, i.e., **rank**$(A) = N$, the solution Problem Eq. (5) can be written as

$$\mu = (A^T A)^{-1} A^T y. \tag{7}$$

However, if A is not full column rank, i.e., **rank**$(A) < N$, $A^T A$ is singular and then $(A^T A)^{-1}$ can be replaced by the Moore-Penrose pseudoinverse $(A^T A)^+$, where $(A^T A)^+$ denotes the Moore-Penrose pseudoinverse of $A^T A$. As for Eq.(7), considering the limited computation capability of computers, the solution Eq.(7) cannot be derived when A is a large matrix in size. To address this issue, our goal is to develop an effective method, which runs in low computation complexity and small memory space compared with our previous method proposed in [5], [12].

A. Singleton Least-Norm Learning

Before going further, we first consider the following system of linear equations,

$$A\mu = y, \tag{8}$$

where $A \in R^{N \times N}$. Here, matrix A can also be written as $A = (\alpha_1^T, \alpha_2^T, \cdots, \alpha_m^T)^T$, i.e.,

$$\alpha_i^T = (a_{i1}, a_{i2}, \cdots, a_{iN}) \neq 0, \tag{9}$$

where $\alpha_i \in R^N$ $(i = 1, 2, \cdots, m)$. Here, we write the row vectors of A as the column vectors α_i $(i = 1, 2, \cdots, m)$ in consistence with that all the vectors are column in this work. Assume **rank**$(A) = N$ and for $1 \leq i \neq j \leq N$, we have $\alpha_i^T \alpha_j = 0$. For each α_i of A, we consider the following least squares problem as

$$\begin{aligned} J_i &= \min_{\mu} \left\{ \|\alpha_i^T \mu - y_i\| + \sqrt{\lambda} \|\mu\| \right\} \\ &= \min_{\mu} \left\| \begin{pmatrix} \alpha_i^T \\ \sqrt{\lambda} \cdot I_N \end{pmatrix} \mu - \begin{pmatrix} y_i \\ 0_{N \times 1} \end{pmatrix} \right\|, \end{aligned} \tag{10}$$

where λ is a factor with $\lambda > 0$, and I_N is the identity matrix of order N. According to Eq.(7), we have the solution of Eq.(10) as follows

$$\begin{aligned} \mu_{\alpha_i}(\lambda) &= \left((\alpha_i, \sqrt{\lambda} I_N) \begin{pmatrix} \alpha_i^T \\ \sqrt{\lambda} I_N \end{pmatrix} \right)^{-1} \times \\ &\quad (\alpha_i, \sqrt{\lambda} I_N) \begin{pmatrix} y_i \\ 0_{N \times 1} \end{pmatrix} \\ &= \left(\alpha_i \alpha_i^T + \lambda I_N \right)^{-1} (\alpha_i \cdot y_i + 0_{N \times 1}) \\ &= \left(\alpha_i \alpha_i^T + \lambda I_N \right)^{-1} \cdot \alpha_i \cdot y_i. \end{aligned} \tag{11}$$

Here, it is easy to derive **rank**$(\alpha_i \alpha_i^T) = 1$, which implies the inverse of $\alpha_i \alpha_i^T$ does not exist. According to the Theory of Matrix Analysis [19], we rewrite $\mu_{\alpha_i}(\lambda)$ as

$$\begin{aligned} \mu_{\alpha_i}(\lambda) &= \lambda^{-1} (I + \lambda^{-1} \alpha_i \alpha_i^T)^{-1} \alpha_i \cdot y_i \\ &= \lambda^{-1} (I - \lambda^{-1} \alpha_i \alpha_i^T + \lambda^{-2} \alpha_i \alpha_i^T \alpha_i \alpha_i^T + \cdots) \cdot \alpha_i \cdot y_i \\ &= \lambda^{-1} \alpha_i (1 - \lambda^{-1} \alpha_i^T \alpha_i + \lambda^{-2} \alpha_i^T \alpha_i \alpha_i^T \alpha_i + \cdots) \cdot y_i \\ &= \alpha_i (\lambda + \alpha_i^T \alpha_i)^{-1} y_i \end{aligned} \tag{12}$$

Now, let $\lambda \to 0$, then we have

$$\mu_{\alpha_i} = \mu_{\alpha_i}(\lambda = 0) = \alpha_i (\alpha_i^T \alpha_i)^{-1} y_i. \tag{13}$$

For any α_i and α_j $(1 \leq i, j \leq N,\ i \neq j)$, we have

$$\alpha_i^T \cdot \mu_{\alpha_j} = \alpha_i^T \cdot \alpha_j (\alpha_j^T \alpha_j)^{-1} y_j = 0. \tag{14}$$

When $1 \leq i = j \leq n$, we have

$$\alpha_i^T \cdot \mu_{\alpha_i} = \alpha_i^T \alpha_i (\alpha_i^T \alpha_i)^{-1} y_i = y_i, \tag{15}$$

i.e.,

$$\alpha_i^T \cdot \mu_{\alpha_j} = \begin{cases} 0 & \text{for } i \neq j \\ y_i & \text{for } i = j \end{cases}. \tag{16}$$

Further, we have

$$A\mu = A \cdot \sum_{j=1}^{N} \mu_{\alpha_j} = \begin{pmatrix} \alpha_1^T \mu_{\alpha_1} + 0 + \cdots + 0 \\ 0 + \alpha_2^T \mu_{\alpha_2} + \cdots + 0 \\ \vdots \\ 0 + 0 + \cdots + \alpha_N^T \mu_{\alpha_N} \end{pmatrix} = y. \tag{17}$$

Obviously, $\mu = \sum_{j=1}^{N} \mu_{\alpha_j}$ is the solution of Problem Eq. (8).

The above method for deriving the solution of Problem Eq.(8) provides a new perspective on solving the Problem Eq.(5). However, in Problem Eq.(5), the row vectors, α_1^T, α_2^T, $\cdots$, α_m^T, of A are usually non-orthogonal. Then, for any α_i and α_j ($1 \leq i = j \leq N$), we have

$$\alpha_i^T \cdot \mu_{\alpha_i} = \alpha_i^T \alpha_i (\alpha_i^T \alpha_i)^{-1} y_i = y_i. \tag{18}$$

But when $1 \leq i \neq j \leq N$,

$$\alpha_i^T \cdot \mu_{\alpha_j} = \alpha_i^T \cdot \alpha_j (\alpha_j^T \alpha_j)^{-1} y_j \neq 0. \tag{19}$$

This implies

$$A \cdot \sum_{j=1}^{m} \mu_{\alpha_j} \neq y. \tag{20}$$

Thus, $\mu = \sum_{j=1}^{m} \mu_{\alpha_j}$ is no longer the solution of Problem Eq. (8). To address such a general case, given the current solution μ_i and the μ_{α_j} derived from the observation (α_j, y_j) with $\alpha_j \mu_{\alpha_j} = y_j$, we adopt the strategy of updating the projection of the μ_i on the direction of the μ_{α_j} using the μ_{α_j}, and thereby obtain the new solution μ_{i+1}. First, we derive the solution $\mu_{\alpha_1} = \alpha_1(\alpha_1^T \alpha_1)^{-1} y_1$ and let $\mu_1 = \mu_{\alpha_1}$. Now, for $\alpha_2^T \mu = y_2$ ($\alpha_2 \neq 0$), we derive the solution $\mu_{\alpha_2} = \alpha_2(\alpha_2^T \alpha_2)^{-1} y_2$. The vector projection of the μ_1 onto the μ_{α_2} is $\mu_{1,\alpha_2} = \frac{\langle \mu_1, \mu_{\alpha_2} \rangle}{\langle \mu_{\alpha_2}, \mu_{\alpha_2} \rangle} \mu_{\alpha_2}$, where $\langle \cdot, \cdot \rangle$ denotes the inner product of two vectors, and the vector projection of the μ_1 onto the μ_{α_2} is the orthogonal projection of the μ_1 onto the straight line defined by the μ_{α_2}. Second, we update the component μ_{1,α_2} of the μ_1 projected on the μ_{α_2} as

$$\mu'_{1,\alpha_2} = \gamma \frac{\langle \mu_1, \mu_{\alpha_2} \rangle}{\langle \mu_{\alpha_2}, \mu_{\alpha_2} \rangle} \mu_{\alpha_2} + (1-\gamma)\mu_{\alpha_2}, \tag{21}$$

where the iterative factor γ is the number with $0 < \gamma < 1$. For vector μ_{α_2}, μ_1 can be written as the sum of the vector projection μ_{1,α_2} and the vector rejection $\mu_{1,\alpha_2}^{\perp}$, i.e.,

$$\mu_1 = \mu_{1,\alpha_2}^{\perp} + \mu_{1,\alpha_2}, \tag{22}$$

where the vector rejection is the orthogonal projection of μ_1 onto a plane orthogonal to μ_{α_2}. Thus, we calculate the μ_2 by replacing the μ_{1,α_2} by μ'_{1,α_2} as

$$\mu_2 = \mu_{1,\alpha_2}^{\perp} + \mu'_{1,\alpha_2} = \mu_1 + (1-\gamma)\left[1 - \frac{\langle \mu_1, \mu_{\alpha_2} \rangle}{\langle \mu_{\alpha_2}, \mu_{\alpha_2} \rangle}\right] \mu_{\alpha_2}. \tag{23}$$

Repeat above procedure, and for α_i, we calculate the μ_i as

$$\mu_i = \mu_{i-1} + (1-\gamma)\left[1 - \frac{\langle \mu_{i-1}, \mu_{\alpha_i} \rangle}{\langle \mu_{\alpha_i}, \mu_{\alpha_i} \rangle}\right] \mu_{\alpha_i}. \tag{24}$$

With this procedure, one observation is processed with the least norm approach once a time. Assume m observations is enough to approximate the μ. This is reasonable because the number of rows in A can be any number in the way of constructing matrix A by randomly sampling the m observations. Here, we call this procedure as the *Singleton Least Norm Learning (SLNL)* with the details shown in Algorithm 1.

Algorithm 1 Solving the least squares problem Eq.(5) with the singleton least-norm learning(SLNL).

Require: observations $\{(\alpha_i, y_i)\}_{i=1}^m$, γ and ϵ

$\mu_0 = 0$
while $i < m$ **do**
  Derive the solution of $\alpha_i^T \mu = y_i$ with the Eq.(13), i.e., $\mu_{\alpha_i} = \alpha_i^T(\alpha_i \alpha_i^T)^{-1} y_i$.
  Calculate the solution of $A\mu = b$ with

$$\mu_i = \mu_{i-1} + (1-\gamma)\left[1 - \frac{\langle \mu_{i-1}, \mu_{\alpha_i} \rangle}{\langle \mu_{\alpha_i}, \mu_{\alpha_i} \rangle}\right] \mu_{\alpha_i}. \tag{25}$$

  if the last 100 μs do not change **then**
    Exit;
  end if
  $i = i + 1$
end while
return μ_i

Remarks: According to Eq.(13), the Eq.(24) cannot be updated when $y_i = 0$ for the observation (α_i, y_i). Assume $\mu_0 \neq 0$ is the solution such that $A\mu = 0$. It is obviously observed that the observation α_i makes the contributions to μ_0. Thus, it is necessary to be used for training μ_0. Let another observation be $(\alpha_j, y_j \neq 0)$. Then, to address the above problem, we construct a new observation as $\alpha'_i = \alpha_i + \alpha_j$ with $y_i + y_j \neq 0$. Therefore, we derive the new least norm solution for α'_i as

$$\begin{aligned} \mu_{\alpha'_i} =& \alpha'_i(\alpha_i'^T \alpha'_i)^{-1}(y_i + y_j) \\ =& (\alpha_i + \alpha_j)((\alpha_i + \alpha_j)^T(\alpha_i + \alpha_j))^{-1} y_j. \end{aligned} \tag{26}$$

Through this way, we conquer the problem of $y_i = 0$ and update the μ with both observations $(\alpha_i, y_i = 0)$ and $(\alpha_j, y_j \neq 0)$.

As for the theoretical analysis of Algorithm 1, the convergence of Algorithm 1 is given with the following theorem.

Theorem 1. Algorithm 2 converges in probability to the optimal solution of Problem Eq.(5).

Here, we omit the details of proof due to the page limitation.

Computation Complexity Analysis of Algorithm 1. For calculating $\mu_{\alpha_i} = \alpha_i(\alpha_i^T \alpha_i)^{-1} y_i$, the computation complexity is $O(N)$, where N is the number of elements in μ. Updating the μ_i with $\mu_i = \mu_{i-1} + (1-\gamma)\left[1 - \frac{\langle \mu_{i-1}, w_i \rangle}{\langle w_i, w_i \rangle}\right] \cdot w_i$, it is required to run m times. Therefore, the overall computation complexity of Algorithm 2 is $O(mN)$.

B. Solving the Choquet Model with the SLNL Approach

In this section, we solve the Choquet model with Algorithm 1. The basic idea for solving the Choquet model is a two-step procedure. The first step is to reduce α_i of size N to ξ_i of size n. In Algorithm 1, it is worth noting that only one observation (α_i, y_i) is utilized for calculating μ_{α_i} once a time. According to [20], the discrete Choquet integral of a function $f : X \to R^+$ with respect to μ is defined by

$$C_\mu(f) := \sum_{i=1}^{n} (f(x_{(i)}) - f(x_{(i-1)}))\mu(\Omega_{(i)}), \tag{27}$$

where $\cdot_{(i)}$ indicates that the indices have been permuted so that $0 \le f(x_{(1)}) \le \cdots \le f(x_{(n)})$, and $\Omega_{(i)} := \{x_{(i)}, \cdots, x_{(n)}\}$, and $f(x_{(0)}) = 0$. Thus, for α_i^T, we have at most n non-zero elements in α_i^T of matrix A. Let ξ_i^T be the corresponding vector of α_i^T, that has at most n non-zero elements, where

$$\begin{aligned}\xi_i^T =& ((f(x_{(1)}) - f(x_{(0)}), (f(x_{(2)}) - f(x_{(1)}), \cdots, \\ & (f(x_{(n)}) - f(x_{(n-1)})).\end{aligned} \tag{28}$$

Then the Eq.(13) can be rewritten with ξ_i^T as

$$\mu_{\xi_i} = \xi_i(\xi_i^T \xi_i)^{-1} y_i. \tag{29}$$

However, it is worthy to be mentioned that when we use the Eq.(29) for calculating μ_{ξ_i}, we have to encounter another problem, i.e., given the μ_{ξ_i}, how to modify μ with μ_{ξ_i}, which is also the issue addressed in the second step. As for this issue, we first build the index for each subset of X. Meanwhile, each non-zero element ξ_{ij} in ξ_i also indicates that the corresponding subset appears in the current observation. For each non-zero element of the derived μ_{ξ_i}, find the corresponding index $\nu_{\xi_{ij}}$ of the element ξ_{ij} and modify the $\mu_{\nu_{\xi_{ij}}}$ with $\mu_{\xi_{ij}}$ as follows

$$\mu_{\nu_{\xi_{i,j}}} = \gamma \cdot \mu_{\nu_{\xi_{i,j}}} + (1-\gamma) \cdot \mu_{\xi_{i,j}}. \tag{30}$$

With the above two-step procedure, we successfully decrease the size of the row vectors α_i of the matrix A to ξ_i, and then the size of matrix A is significantly reduced. The details for solving the Choquet model can be found in Algorithm 2.

Obviously, the above two-step procedure does not change the convergence of the Algorithm 1. Now, the computation complexity of Algorithm 2 is given as follows.

Computation Complexity Analysis of Algorithm 2: For solving the Choquet model, with Eq.(29), the computation complexity of deriving $\mu(\alpha_i)$ will be reduced from $O(N)$ to $O(n)$. Since Algorithm 1 runs at $O(mN)$ time, then the overall computation complexity of Algorithm 2 is $O(mn)$.

As for storage space involved in solving the Choquet model, applying our previous method [11] to derive the non-additive measures, we have to store the matrix A of size $m \times N$, which is the main part for storage. Thus, the required storage space for our previous work is $O(mN)$. With the proposed method, we only need to make the storage of ξ_i. When μ_{ξ_i} is derived, we free the storage and read the ξ_{i+1}. However, we have to make the storage space of N for μ. Therefore, we significantly reduce the required storage space from $O(mN)$ to $O(N) = O(2^n)$ for solving the Choquet model.

Algorithm 2 The SLNL approach for the Choquet Model

Require: observations $\{(\xi_i, y_i)\}_{i=1}^m$, γ
1: $\mu_0 = 0$
2: **for** i from 1 to m **do**
3: Derive the solution of $\xi_i^T \mu = y_i$ with the Eq.(29), i.e., $\mu_{\xi_i} = \xi_i^T(\xi_i \xi_i^T)^{-1} y_i$.
4: **for** j from 1 to n **do**
5: Find the corresponding subset of ξ_{ij} and its index $\nu_{\xi_{ij}}$ and update $\mu_{\nu_{\xi_{i,j}}}$ with Eq.(30)
6: **end for**
7: **end for**
8: **return** μ

IV. A Case Study: Cross-Layer Optimized Multimedia Delivery over Wireless Networks

In this section, we will adopt cross-layer optimized wireless multimedia communications as a case study to validate the proposed method and to illustrate the application of the Choquet integral for optimizing the video transmission over wireless networks.

A. Experiment Environment

To describe the wireless channel quality, a Finite-State Markov Channel model (FSMC) is proposed for Rayleigh fading channel in [21], [22], in which each state corresponds to a specific channel quality. The signal-to-interference-noise ratio (SINR) of channel are quantized to q states, where s_i is the state and $\mathcal{P}_{ij}$ denotes the transit probability from state s_i to s_j. Then, the transition probability matrix is $\mathcal{P} = (\mathcal{P}_{ij})_{q \times q}$. According to the state probability and one step transfer matrix of the channel, it outputs quantized channel state s_i. The channel is a discrete memoryless channel in each state and a Markovian genie determines the state of the channel. In this paper, the channel conditions are assumed to be composed of good state s_1 (SINR=30 dB) and bad state s_2 (SINR=15 dB). We consider two scenarios: the good channel with the transition probability matrix with $\mathcal{P}_{11} = \mathcal{P}_{21} = 0.1$, $\mathcal{P}_{12} = \mathcal{P}_{22} = 0.9$ and the bad channel with transition probability matrix with $\mathcal{P}_{11} = \mathcal{P}_{21} = \mathcal{P}_{12} = \mathcal{P}_{22} = 0.5$.

In the protocol stack of wireless multimedia, each layer has one or multiple key design variables which significantly affect the overall system function. For instance, at the application layer, prediction mode and quantization parameter (QP) in video encoding are two critical design variables [23]. At the data link layer, automatic repeat request (ARQ), media access control protocols, and packetization are often used to maintain a low packet loss rate. At the physical layer, modulation and coding schemes (MCSs) have been adopted to achieve a good tradeoff between transmission rate and transmission reliability. In this case study, we investigate the real-time transmission of an individual video bitstream across a multi-hop IEEE 802.11a/e wireless network, in which contention-free access to the medium provided by the HCF controlled channel access protocol (HCCA) [24] is assumed.

In the latest H.264/AVC standard, the allowed QP values are $\{0, 1, 2, \cdots, 51\}$. To simplify the problem, we compress the video by choosing different QPs. To achieve a good tradeoff between performance and computation complexity, without loss of generality, in this work we choose $\{5, 10, 15, 20, 25, 30, 35, 40\}$ as the operating points of the system parameter QP. For another system parameter MCS, without losing generality, the schemes in the modulation and coding schemes at the Physical Layer specified in IEEE 802.11a/e networks are adopted as its operating points [24]. The system model for the cross-layer optimized multimedia transmission is shown in Figure 1, where the expected received video quality is used as the objective function, and the design variables include QP at the application layer and MCS at the physical layer. It is assumed that the controller is able to acquire the corresponding system information, such as the expected video quality from the encoder and the network conditions from lower layers by interacting with each layer.

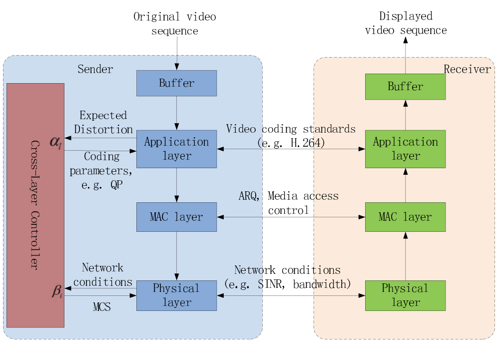


Fig. 1. The system model for sampling the controlled parameters involved in the cross layer design of video transmission over the wireless network.

Let $\Pi = \{\pi_1, \pi_2, \cdots, \pi_I\}$ be the set of I packets that compose the current video frame to be transmitted. Let $E[\mathcal{D}_i]$ be the expected video quality of packet π_i. To provide a smooth video display experience to end users, each frame is associated with a frame decoding deadline T^{budget} [25], [27]. T^{budget} imposes a delay constraint on the transmission of each packet composing the current frame as

$$\sum_{i=1}^{I} T_i \leq T^{\text{budget}}, \tag{31}$$

where T_i is the end-to-end delay of packet π_i transmitted from the sender to the receiver.

Let α_l denote the QP for the lth coding unit of the current frame and β_i the MCS for transmitting packet π_i. Here, the coding unit could be a video frame, a slice, or a macroblock, depending on different adaptation time intervals. Denote $\mathbb{A}$ and $\mathbb{B}$ as the sets of all operating points of α_l and β_i, respectively. We assume that $|\mathbb{A}| = \mathcal{A}$ and $|\mathbb{B}| = \mathcal{B}$, where $|\mathbb{A}|$ is the cardinality of the set $\mathbb{A}$. Thus, the goal of cross-layer design is to find the optimal operating point vector $\{\alpha_l, \beta_i\}$ such that the received video quality is minimized under the constraint of packet delay deadline, i.e.,

$$\min_{\alpha_l \in \mathbb{A}, \beta_i \in \mathbb{B}} \sum_{i=1}^{I} E[\mathcal{D}_i], \text{ s.t.} : \sum_{i=1}^{I} T_i \leq T^{\text{budget}}.$$

The experiments are designed using H.264/AVC JM 12.2 [28]. We encode the Y-component of the first 100 frames of the QCIF video sequence "Foreman" at different frame rates.

B. Validation of the Proposed Method

In the following, we will validate the proposed method for cross-layer design. For video transmission, one video frame is usually divided into multiple slices and delivered slice by slice over wireless networks. Considering the dependency between the current video frame and previous video frame in video encoding, we treat one video frame as a whole to control the video quality. The video quality (PSNR, i.e., y) of frames are observed as the design objective function on all the operating points of two system parameters QP and MCS. Figure 2 is the path graph for F-frame delivery, where each path, also namely operation path, is used to describe the operation sequence for all frames, and each option is composed of operation points of both QP and MCS. Here we denote $Opt_i(j)$ the jth option for frame i, which is an operating point pair of QP and MCS. As mentioned in the last subsection, we have 8 operation points for both QP and MCS. Thus, as shown in Figure 2, each frame has 64 options. Let $z = g(Opt_i(j))$ be the function of video quality (PSNR) of frame i under the option $Opt_i(j)$. Each path is an option sequence which is composed of the options on all F frames. Let $x_{2(i-1)+1}$ and $x_{2(i-1)+2}$ be the options for frame i, where $x_{2(i-1)+1}$ and $x_{2(i-1)+2}$ represent the parameters of QP and MCS, respectively. Then, for F frames, we have $X = (x_1, x_2, \cdots, x_{2(i-1)+1}, x_{2(i-1)+2}, \cdots, x_{2(F-1)+1}, x_{2(F-1)+2})$. To measure the video quality (PSNR) $\mathcal{D}$ of F frames, we calculate the average video quality of all frames on the operation path Pa as:

$$\mathcal{D}(Pa) = \frac{\sum\limits_{Opt_i(j) \in Pa} g(Opt_i(j))}{F}. \tag{32}$$

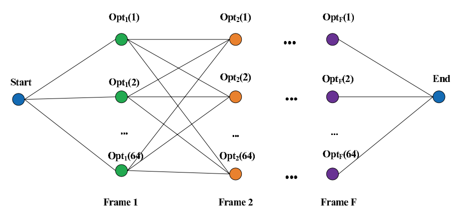


Fig. 2. The path graph for F-frame video transmission.

As aforementioned, to deliver F-frame video, we have to adjust $2 \times F$ parameters in total, including both the QP and MCS, as $X = (x_1, x_2, \cdots, x_n)$ with $n = 2F$. Due to the complex interactions among frames, according to Eq.(32), choosing too many frames will dwarf the contribution of operations on each frame. Thus, taking into account the contribution of each frame and without loss of generality, we first choose $F = 5$ frames. Then, for $n = 2 \times 5 = 10$

parameters, we total have $2^{10} - 1 = 1023$ non-additive measures. Hence, we use the proposed method to derive these 1023 non-additive measures based on the observation data. The details of above procedure are stated as follows:

- Assume we have a history database of observation data for the past video delivery. Sample the M paths from the database and denote by Ω the set of all sampled paths. Since each path contains 10 option points of QP and MCS, we derive the coefficient matrix A with Algorithm 1. Thus, we have a least squares problem as

$$\min_{\mu} \|A\mu - y\|, \tag{33}$$

  where A is the $m \times 1023$ matrix. μ is the 1023×1 vector and y the $m \times 1$ vector.
- Implement the least norm learning with Algorithm 3 and derive the non-additive measure vector μ. Then, the components of vector μ show the contributions of these non-additive measures.

We select the bad and good channel conditions to derive the non-additive vector μ. Usually, the observed data the number of which is far more than the number of parameters is usually collected. This makes it intractable for directly solving the least square problem Eq. (33), with applying Algorithm 1, because of the required huge storage space and computation capability. Thus, our proposed method of Algorithm 3 is applied in the following sections. Figure 3 shows the results of contributions of non-additive measures under bad and good channel conditions. Compared with the contributions under the bad channel condition, the contributions under good channel condition have bigger amplitude from around -15 to around 60, while the amplitude from around 4 to around 20 under bad channel condition. This implies that the adjustments of parameters under the good channel condition can have much more improvement than that under bad channel condition. Meanwhile, this also implies that, under the bad channel condition, a considerable number of non-additive measures achieve less or even no contribution for improving the received video quality.

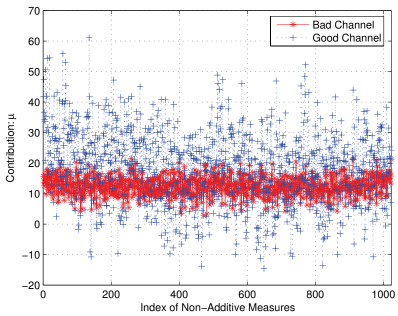


Fig. 3. The contributions of non-additive measures.

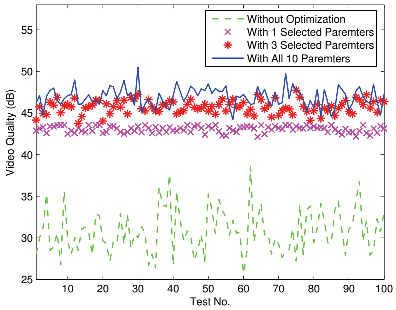
To optimize the number of adjustable parameters involved in the cross-layer design of video transmission, the remarkable parameters are required to be found such that the video quality can be almost maintained or well approximated by only adjusting selected parameters. Thus, we select the non-additive measure that achieves the most contribution among all non-additive measures, and then find the involved parameters of this non-additive measure. As shown in Figure 3, for the good channel condition, we find the set $\{x_3, x_8, x_9, x_{10}\}$ whose non-additive measure is $\mu(\{x_3, x_8, x_9, x_{10}\}) = 61.08$. Except these selected parameters, the options for the other parameters are randomly selected from the related ranges. By only adjusting 4 in 10 of the selected parameters, the mean PSNR of 5 frames is 53.48 dB, 94.04% of the optimal frame quality (56.87 dB) by adjusting all 10 parameters. With respect to the bad channel condition, we derive the set $\{x_2, x_9, x_{10}\}$ with the non-additive measure $\mu(\{x_2, x_9, x_{10}\}) = 21.47$. Here and in the following sections, the searching space spanning by n parameters is $8 \times 8 \times \cdots \times 8 = 8^n$. Obviously, we significantly reduce the searching space composed of $8^{10} = 1.07 \times 10^9$ paths into the space with $8^4 = 4096$ paths. By only adjusting 3 in 10 of the selected parameters, the mean PSNR of 5 frames is 20.592 dB, 99.99% of the optimal frame quality (20.594 dB) by adjusting all 10 parameters. Moreover, for both bad and good channel conditions, our proposed method provides an efficient way for optimizing the cross-layer design in video transmission because adjusting the selected parameters encounters a much smaller searching space and greatly reduces the amount of calculations. In sum, our proposed method of adjusting the remarkable parameters can effectively and efficiently optimize the video transmission in cross-layer design.

1) Performance Evaluation in Statistics: To validate the performance of our proposed method for video transmission of cross-layer design in statistics, let us consider a scenario of student's t test that is implemented to find the confidence intervals for the video quality. In order to apply student t test reasonably, our procedure adopts the following two assumptions as stated in [29]: (1) the operation path is randomly drawn from the operation path space; (2) the operation path can be reasonably supposed to have a normal distribution. Suppose we have the optimal values of $\mathscr{X}_1, \mathscr{X}_2, \cdots, \mathscr{X}_q$ with the mean (μ) after doing q tests. The standard deviation of their parameter distribution (standard error) is σ. According to literature [29], we can compute the $(1-\epsilon)$ confidence interval of the optimal value $\mathscr{X}$ as:

$$\mu - \frac{B\sigma}{\sqrt{q}} \leq \mathscr{X} \leq \mu + \frac{B\sigma}{\sqrt{q}}. \tag{34}$$

where B is the critical value, which can be found from Appendix C "Critical Values of t" in [29].

We have carried out 100 tests under the good channel condition. Under the good channel, the most contributed fuzzy measures are $\mu(\{x_6\}) = 40.14$, $\mu(\{x_6, x_8\}) = 28.84$, $\mu(\{x_3, x_6, x_8\}) = 21.75$. Based on the contributions of these fuzzy measures, we collect the quality (PSNR) of the video frames by adopting three ways for adjusting the parameters: (i) optimizing with one parameter x_6; (ii) with three parameters $\{x_3, x_6, x_8\}$; (iii) with all the 10 parameters. For comparison, the video quality of frames without optimization are also collected. Figure 4(a) shows the results of the average video quality of frames under the good channel condition. In this figure, our proposed method achieves the mean value of 45.76 dB (i.e., 97.49% of the optimal PSNR 46.94 dB) with 3 selected parameters, 43.06 dB (i.e., 91.73% of the optimal PSNR 46.94 dB) by only adjusting one selected parameter,



(a) Under good channel condition

Fig. 4. The statistical results for the received video quality (PSNR) by adjusting the selected parameters and all parameters under the good and bad channel conditions.

and 30.82 dB (i.e., 65.66% of the optimal PSNR 46.94 dB) without optimization, respectively. By adjusting a part of parameters, the video quality over wireless networks is significantly maintained. Also, according to Eq. (34), we have 95%, 99% and 99.5% confidence intervals of average video quality (PSNR) in the confidence intervals [45.58, 45.94], [45.52, 45.99] and [45.49, 46.02], under the good channel condition. Thus, our proposed method can effectively optimize the video quality. More specifically, we have 99% confidence that the received video qualities are better than 45.52 dB (i.e., 96.97% of the PSNR 46.94 dB derived from the optimization with all the 10 parameters). Therefore, under the good channel condition, the results in statistics sufficiently illustrate the effectiveness of our proposed method.

V. Conclusions

One of the major challenges in the application of the Choquet integral is that the determination of non-additive interaction measures is seriously restricted by the curse of dimensionality. In this paper, we have developed an effective determination method on the basis of the least norm learning for efficient determination of non-additive interaction measures. The proposed method has significantly reduced the computation time and space complexity involved in determining the non-additive interaction measures. Further, the proposed method has been proved theoretically convergent in probability. Finally, a case study has been carried out based on cross-layer optimized wireless multimedia communications. Extensive experiment results have validated the effectiveness of our proposed method.

References

[1] M. Grabisch, *Fuzzy Measures and Integrals–Theory and Applications*. New York: Physica Verlag, 2000, ch. Fuzzy integral for classification and feature extraction, pp. 415–434.

[2] ——, "Fuzzy integral in multicriteria decision making," *Fuzzy Sets and Systems*, vol. 69, pp. 279–298, 1995.

[3] K. Xu, Z. Wang, M. L. Wong, and K.-S. Leung, "Discover dependency pattern among attributes by using a new type of nonlinear multiregression," *International Journal of Intelligent Systems*, vol. 16, no. 8, pp. 949–962, 2001.

[4] M. Grabisch, *Information fusion in data mining*. Physica-Verlag, 2003, ch. Modelling data by the Choquet integral, pp. 135–148.

[5] W. Zheng, H.-F. Guo, and S. Ci, "Precise determination of non-additive measures," in *IEEE Conference on Cybernetics and Intelligent Systems*, Chengdu, China, Jun. 2008, pp. 911–916.

[6] P. Miranda and M. Grabisch, "Optimization issues for fuzzy measures," *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, vol. 7, no. 6, pp. 545–560, 1999.

[7] T. Mori and T. Murofushi, "An analysis of evaluation model using fuzzy measure and the Choquet integral," in *Proceedings of the 5th Fuzzy System Symposium*, Japan, 1989, pp. 207–212.

[8] M. Grabisch and J. M. Nicolas, "Classification by fuzzy integral–performance and tests," *Fuzzy Sets and Systems, Special issues on Pattern Recognition*, vol. 65, no. 2–3, pp. 255–271, 1994.

[9] M. Grabisch, "A new algorithm for identifying fuzzy measures and its application to pattern recognition," in *International Joint Conference of 4th IEEE International Conference on Fuzzy Systems and 2nd International Fuzzy Engineering Symposium*, Yokohama, Japan, March 1995, pp. 145–150.

[10] Z. Wang, K. S. Leung, and J. Wang, "A genetic algorithm for determining nonadditive set functions in information fusion," *Fuzzy Sets and Systems*, vol. 102, pp. 463–469, 1999.

[11] S. Ci and H. F. Guo, "Quantitative dynamic interdependency measure and significance analysis for cross-layer design under uncertainty," in *16th International Conference on Computer Communications and Networks*, Honolulu, Hawaii, USA, Aug. 2007.

[12] Dalei Wu, Song Ci, Haiyan Luo, and Hai-Feng Guo, "A theoretical framework for interaction measure and sensitivity analysis in cross-layer design[j]," *ACM Transactions on Modeling and Computer Simulation*, vol. 21, no. 1, pp. 6:1–25, 2010.

[13] M. Grabisch and E. Raufaste, "An empirical study of statistical properties of the Choquet and Sugeno integrals," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 4, p. 839, 2008.

[14] M. Grabisch and C. Labreuche, "Bi-capacities–I: definition, Mobius transform and interaction," *Fuzzy Sets and Systems*, vol. 151, no. 2, pp. 211–236, 2005.

[15] I. Kojadinovic and C. Labreuche, "Partially bipolar Choquet integrals," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 4, pp. 839–850, 2009.

[16] M. Grabisch and M. Roubens, "Application of the Choquet integral in multicriteria decision making," in *Fuzzy Measures and Integrals*. Physica Verlag, 2000, pp. 348–374.

[17] Å. Björck, *Numerical Methods for Least Squares Problems*. SIAM: Society for Industrial and Applied Mathematics, 1996.

[18] G. Golub, "Numerical methods for solving linear least squares problems," *Numerische Mathematik*, vol. 7, pp. 206–216, 1965.

[19] X. Zhang, *Matrix analysis and applications*. Tsinghua University Press & Springer, 2004.

[20] G. Choquet, "Theory of capacities," *Annales de l'Institut Fourier*, vol. 5, pp. 131–295, 1953.

[21] H. S. Wang and N. Moayeri, "Finite-state markov channel-a useful model for radio communication channels," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 1, pp. 163–171, 1995.

[22] Q. Pan, S. Liu, M. Xu, and C. Jia, "Finite-state Markov model for the aeronautical channel," in *Proc. of the 5th International Conference on Wireless Communications, Networking and Mobile Computing*, Beijing, China, 2009.

[23] H.264, *Advanced video coding for generic audiovisual services*. ITU-T Rec. H.264/ISO/IEC 14496-10(AVC), 2005.

[24] 802.11e, *IEEE Standard for Information Technology–Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements*. IEEE, 2005.

[25] P. Pahalawatta, R. Berry, T. Pappas, and A. Katsaggelos, "Content-aware resource allocation and packet scheduling for video transmission over wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 4, pp. 749–759, May 2007.

[26] E. Maani, P. Pahalawatta, R. Berry, T. Pappas, and A. Katsaggelos, "Resource allocation for downlink multiuser video transmission over wireless lossy networks," *IEEE Transactions on Image Processing*, vol. 17, no. 9, pp. 1663–1671, 2008.

[27] D. Wu, S. Ci, H. Luo, H. Wang, and A. Katsaggelos, "Application-centric routing for video streaming over multi-hop wireless networks," in *Proceedings of the Sixth Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, Rome, Italy, 2009.

[28] JM12.2, *H.264/AVC Software Coordination*. http://iphome.hhi.de/suehring/tml/, 2010.

[29] R. Lowry, *Concepts and Applications of Inferential Statistics*. http://faculty.vassar.edu/lowry/webtext.html, 2010.