MPPM Constellation Selection for Free-Space Optical Communications

Trung Thanh Nguyen, *Student Member, IEEE*, and Lutz Lampe, *Senior Member, IEEE*

***Abstract*— We consider the problem of designing multipulse pulse-position modulation (MPPM) constellations whose sizes are powers of two. This problem amounts to selecting a subset from the collection of all $\binom{n}{w}$ possible signal points of MPPM with w pulses in n time slots. In a previous work, we have tackled this selection using combinatorial heuristics. In this letter, we further explore two new continuous optimization approaches. The first one is a modified Blahut-Arimoto algorithm. The second one is inspired from compressed sensing. Using the constellation-constrained channel capacity as the figure of merit, numerical results from a relevant free-space optical communication example suggest that simple combinatorial heuristics yield practically the best designs.**

***Index Terms*— Multipulse pulse-position modulation (MPPM), free-space optical (FSO) communications, discrete optimization, compressed sensing.**

I. Introduction

Multipulse pulse-position modulation (MPPM) is a promising signaling scheme for free-space optical (FSO) communications. It is a generalization of pulse-position modulation (PPM) such that there are $w > 1$ pulses per symbol consisting of $n > w$ slots. Compared to PPM (i.e., when $w = 1$), MPPM enables a favorable trade-off between power and bandwidth efficiencies [1]–[4].

The number of all possible (n, w)-MPPM signal points is $M = \binom{n}{w}$, which is not a power of two. Utilizing the full constellation would require a complicated bit-to-symbol mapping and demapping. A power-of-two constellation size is therefore often preferred, especially if error-control coding (ECC) is applied. Thus, an important task in designing a coded MPPM system is to select a subset of $K < M$ symbols for transmission. There are different criteria to select a subset, e.g., to achieve the smallest symbol error rate. However, with the increasing popularity of modern capacity-approaching codes, maximizing the constellation-constrained channel capacity seems to be the most relevant criterion. In [5], using this criterion, we have formulated the subset selection as a combinatorial optimization and proposed a number of enumeration heuristics. Since the solutions of such heuristics are deemed suboptimal, we are motivated to explore other alternative approaches.

Manuscript received August 25, 2010; revised July 26, 2011. This work was supported by the National Sciences and Engineering Research Council (NSERC) of Canada.

Trung Thanh Nguyen was with the Deptartment of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada. He is now with McKinsey & Company Vietnam. Lutz Lampe is with the Deptartment of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada. (e-mail: trungn@ece.ubc.ca, Lampe@ece.ubc. ca).

In this letter, we revisit the problem of MPPM constellation selection and formulate it as a continuous optimization problem. To this end, we again aim to maximize the constrained channel capacity, but relax the condition of uniformly distributed signal points during constellation selection. For MPPM FSO transmission over discrete channels, we propose two algorithms to solve this relaxed problem. The first is a Blahut-Arimoto-type algorithm, and the second is based on compressed sensing. Perhaps surprisingly, numerical results for $(16, 4)$-MPPM transmission over quantum-limited channels demonstrate that the constellation designs obtained in [5] through enumeration techniques can hardly be improved by relaxation and continuous optimization. This in turn leads us to the nontrivial conclusion that the former is already an adequate approach for the considered MPPM constellation design. Even though this result cannot directly be generalized, the presented algorithms are readily applicable for optimization for other constellations and channel models.

Notation: Matrices and vectors are written in bold face, e.g., $\boldsymbol{X}$ and $\boldsymbol{x}$. The i-th element of $\boldsymbol{x}$ is x_i. X_{ij} is the element of matrix $\boldsymbol{X}$ in row i and column j. $\|\boldsymbol{x}\|_1$ is the ℓ_1-norm and $\|\boldsymbol{x}\|_0$ denotes the number of nonzero components of $\boldsymbol{x}$, respectively. $D(\boldsymbol{x}\|\boldsymbol{y})$ for $\|\boldsymbol{x}\|_1 = \|\boldsymbol{y}\|_1 = 1$ denotes the Kullback-Leibler divergence between $\boldsymbol{x}$ and $\boldsymbol{y}$. $\lambda_{\min}(\boldsymbol{X})$ denotes the minimal eigenvalue of $\boldsymbol{X}$, $\mathbf{1}_M$ is the length-M all-one column vector, and $\operatorname{diag}\{\boldsymbol{x}\}$ is the diagonal matrix with elements of $\boldsymbol{x}$ on its main diagonal. Finally, $\Pr\{\cdot\}$ denotes the probability of the event in brackets.

II. MPPM Transmission and Problem Definition

We consider the transmission of MPPM signal vectors $\boldsymbol{x}$ of length n with $\|\boldsymbol{x}\|_0 = \|\boldsymbol{x}\|_1 = w$ "on" slots corresponding to $x_i = 1$ and $(n - w)$ "off" slots corresponding to $x_i = 0$. The collection of all such vectors $\boldsymbol{\mathcal{X}} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\}$ has size $M = \binom{n}{w}$, but only a subset $\boldsymbol{\mathcal{S}} \subset \boldsymbol{\mathcal{X}}$ of size $K = |\boldsymbol{\mathcal{S}}|$ is desired for transmission, where $K = 2^L$ and $L \in \mathbb{N}$. Each component of $\boldsymbol{x}$ is transmitted over a memoryless channel, producing output vector $\boldsymbol{y}$ from the set[1] $\boldsymbol{\mathcal{Y}} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_N\}$. The channel is fully described by its $N \times M$ transition matrix $\boldsymbol{Q}$ with entries

$$Q_{ij} = \Pr\{\boldsymbol{y}_i|\boldsymbol{x}_j\} \,. \tag{1}$$

Let us denote the probability of selecting the input $\boldsymbol{x}_j$ by p_j and the probability of observing the channel output $\boldsymbol{y}_i$ by q_i,

[1] Throughout this letter, we consider discrete channels with discrete outputs $\boldsymbol{y}$ as a result of quantization at the receiver, which in FSO transmission is often explicitly included in the channel model. However, the algorithms proposed in this letter can also be applied to continuous-output channels using Monte Carlo techniques to perform integration, cf. e.g. [6].

respectively. The average mutual information between channel input $\boldsymbol{x}$ and output $\boldsymbol{y}$, i.e., the constellation-constrained channel capacity, is given by

$$I(\boldsymbol{p}) = \sum_{j=1}^{M}\sum_{i=1}^{N} p_j Q_{ij} \log_2\left(\frac{Q_{ij}}{q_i}\right) \;, \tag{2}$$

where we explicitly expressed the dependency on the input distribution $\boldsymbol{p} = [p_1, p_2, \ldots, p_M]^T$. We note that the output distribution $\boldsymbol{q} = [q_1, q_2, \ldots, q_N]^T$ follows as $\boldsymbol{q} = \boldsymbol{Q}\boldsymbol{p}$.

The problem that we address in this letter is to find the subset $\boldsymbol{\mathcal{S}}$ of input vectors that maximizes $I(\boldsymbol{p})$ from (2). We assume that the L bits mapped to the K signal points are uniformly distributed. As a result, the distribution of signal points is also uniform. Hence, we seek the input distribution that maximizes $I(\cdot)$,

$$\boldsymbol{p}^* = \underset{\boldsymbol{p}\in\boldsymbol{\mathcal{P}}}{\operatorname{argmax}}\{I(\boldsymbol{p})\} \;, \tag{3}$$

where the maximization is performed on the constrained set of probability vectors

$$\boldsymbol{\mathcal{P}} = \{\boldsymbol{p} \in \mathbb{R}^M : \|\boldsymbol{p}\|_0 = K,\; \|\boldsymbol{p}\|_1 = 1,\; p_j \in \{0, 1/K\}\, \forall j\} \;. \tag{4}$$

III. Discrete and Continuous Optimization

In this section, we first briefly review the discrete random search from [5], which directly performs the optimization (3). To overcome the difficulty of enumeration in discrete optimization, we then propose two relaxed formulations which lead to continuous optimization problems. The first uses a Blahut-Arimoto-type algorithm to find solutions in the full set of probability vectors. Hence, the associated capacity performances serve as benchmark. The second approach is inspired by recent work on recovering sparse signals using regularization with penalty functions. In this case, the probability vector delivered by the optimization is projected into $\boldsymbol{\mathcal{P}}$ through quantization.

A. Discrete Optimization

The direct optimization of $I(\boldsymbol{p})$ over the set $\boldsymbol{\mathcal{P}}$ is a 0-1 integer optimization problem with $\binom{M}{K}$ enumerations. While various partial enumeration or heuristic procedures are applicable to the problem at hand, we found in [5] that the following random search with acceptance and rejection based on $I(\boldsymbol{p})$ can hardly be improved upon by more elaborate (meta-)heuristical searches.

The discrete random search is formalized in the following pseudo-code.

Algorithm 1: Uniformly randomly select $\hat{\boldsymbol{p}} \in \boldsymbol{\mathcal{P}}$
- **for** number of searches
  - Uniformly randomly select $\boldsymbol{p} \in \boldsymbol{\mathcal{P}}$
  - **if** $(I(\boldsymbol{p}) > I(\hat{\boldsymbol{p}}))$
    - $\hat{\boldsymbol{p}} = \boldsymbol{p}$
  - **end if**
- **end for**
- **Return** $\hat{\boldsymbol{p}}$

The above pseudo-code represents uninformed search. In [5], we showed that focusing the search to a "good" region of $\boldsymbol{\mathcal{P}}$ can improve the quality of the solution. To this end, the MPPM signal points can be considered as the codewords of a length-n weight-w constant-weight code (CWC). Let $\boldsymbol{\mathcal{C}}$ denote an (n, w) CWC with the largest possible size A for a given minimum Hamming distance d_{min}. We argue that optimized CWCs, i.e., codes with large d_{min} for given parameters (n, w), represent good search regions. Since CWCs are only available for certain sizes A (see tables in [7], [8]), the following approach is applied: (i) find a CWC with large d_{min} and A close to K. (ii) If $A \geq K$, we make the constellation always a subset of $\boldsymbol{\mathcal{C}}$. Otherwise, we make the constellation always a superset of $\boldsymbol{\mathcal{C}}$, i.e., $\boldsymbol{\mathcal{C}}$ and additional $K - A$ randomly selected vectors from $\boldsymbol{\mathcal{P}}\backslash\boldsymbol{\mathcal{C}}$.

B. Blahut-Arimoto-type Optimization

The main problem when directly tackling problem (3) are the "discrete" constraints $\|\boldsymbol{p}\|_0 = K$ and $p_j \in \{0, 1/K\}$ in (4). We therefore define the new set

$$\tilde{\boldsymbol{\mathcal{P}}} = \{\boldsymbol{p} \in \mathbb{R}^M : H(\boldsymbol{p}) \leq L,\; \|\boldsymbol{p}\|_1 = 1,\; p_j \geq 0\, \forall j\} \tag{5}$$

with continuous constraints, where

$$H(\boldsymbol{p}) = -\sum_{j=1}^{M} p_j \log_2(p_j) \tag{6}$$

is the average entropy associated with $\boldsymbol{p}$. Clearly, $\boldsymbol{\mathcal{P}} \subset \tilde{\boldsymbol{\mathcal{P}}}$, and thus

$$\boldsymbol{p}^* = \underset{\boldsymbol{p}\in\tilde{\boldsymbol{\mathcal{P}}}}{\operatorname{argmax}}\{I(\boldsymbol{p})\} \tag{7}$$

is a relaxed version of (3). We can look at (7) as a problem of nonuniform signaling investigated for Gaussian and optical wireless intensity channels in [9], [10].

Since both mutual information and entropy are convex-$\cap$ functions of the input probability, problem (7) is a non-convex optimization. We thus resort to finding local maxima of $I(\boldsymbol{p})$. Similarly to [11], we consider the Lagrangian

$$J_\mu(\boldsymbol{p}) = I(\boldsymbol{p}) - \mu H(\boldsymbol{p}) \;, \tag{8}$$

whose (possibly local) maximization for different values of μ returns $\boldsymbol{p}_\mu^*$ with associated mutual information-entropy pair $(I(\boldsymbol{p}_\mu^*), H(\boldsymbol{p}_\mu^*))$. If $J_\mu(\boldsymbol{p}_{\boldsymbol{\mu}}^*)$ is the global maximum, $(I(\boldsymbol{p}_\mu^*), H(\boldsymbol{p}_\mu^*))$ supports the convex-$\cap$ hull of $I(H)$. Clearly, we have $0 < \mu < 1$. The choice $\mu = 0$ corresponds to the unconstrained optimization of mutual information (3). Furthermore, $(I(\boldsymbol{p}_1^*), H(\boldsymbol{p}_1^*)) = (0, 0)$ for $\mu \geq 1$.

We propose the following variant of the Blahut-Arimoto algorithm [12], [13] to find maxima of $J_\mu(\boldsymbol{p})$ (notation is adopted from [14]).

Algorithm 2: Initial probability vector $\boldsymbol{p}^{(0)}$ with positive elements, $k = 0$

1: **do**
2: $\quad D_j^{(k)} = D(\boldsymbol{Q}_j \| \boldsymbol{Q}\boldsymbol{p}^{(k)})$
3: $\quad p_j^{(k+1)} = (p_j^{(k)})^{1/(1-\mu)} \times \dfrac{\exp(D_j^{(k)}/(1-\mu))}{\sum_{l=1}^{M} (p_l^{(k)})^{1/(1-\mu)} \exp(D_l^{(k)}/(1-\mu))}$
4: $\quad k \leftarrow k + 1$
5: **while** $(D(\boldsymbol{p}^{(k)} \| \boldsymbol{p}^{(k-1)}) > t)$
6: **Return** $\boldsymbol{p}^{(k)}$

In Algorithm 2, $\boldsymbol{Q}_j$ denotes the jth column of $\boldsymbol{Q}$ and $t > 0$ is a small threshold value to terminate the iteration. We have the following result for Algorithm 2.

Lemma 1: $J_\mu(\boldsymbol{p}^{(k)})$ is monotonically increasing with k.

Proof: We can write (cf. [13, Theorem 1])

$$\begin{aligned} \max_{\boldsymbol{p} \in \breve{\boldsymbol{\mathcal{P}}}} \{J_\mu(\boldsymbol{p})\} &= \max_{\boldsymbol{p} \in \breve{\boldsymbol{\mathcal{P}}}} \{I(\boldsymbol{p}) - \mu H(\boldsymbol{p})\} \\ &= \max_{\boldsymbol{p} \in \breve{\boldsymbol{\mathcal{P}}}} \max_{\boldsymbol{P} \in \boldsymbol{\mathcal{T}}} \{I(\boldsymbol{p}, \boldsymbol{P}) - \mu H(\boldsymbol{p})\} \ , \end{aligned} \tag{9}$$

where

$$\breve{\boldsymbol{\mathcal{P}}} = \{\boldsymbol{p} \in \mathbb{R}^M : \|\boldsymbol{p}\|_1 = 1,\ p_j \geq 0 \,\forall j\} \tag{10}$$

$$I(\boldsymbol{p}, \boldsymbol{P}) = \sum_{j=1}^{M} \sum_{i=1}^{N} p_j Q_{ij} \log_2(P_{ji}/p_j) \tag{11}$$

and $\boldsymbol{\mathcal{T}}$ is the set of all $M \times N$ transition probability matrices. Successive optimization with respect to $\boldsymbol{P}$ for given $\boldsymbol{p}^{(k)}$ and with respect to $\boldsymbol{p}$ for given $\boldsymbol{P}^{(k)}$ leads to the update equations in Algorithm 2. Thus, $J_\mu(\boldsymbol{p}^{(k)})$ monotonically increases with k. ■

We note that for *symmetric* transmission channels the uniform distribution maximizes mutual information. Therefore, $\boldsymbol{p}_0^* = \frac{1}{M}\boldsymbol{1}_M$ is the global optimum for $J_\mu(\boldsymbol{p})$ for $\mu = 0$. Furthermore, we observe from Algorithm 2, Lines 2 and 3, that $\boldsymbol{p}_0^*$ is always a fixed point of the recursion. Since the second derivative of $J_\mu(\boldsymbol{p})$ is given by

$$\nabla^2 J_\mu(\boldsymbol{p}) = -\boldsymbol{Q}^T (\mathrm{diag}\{\boldsymbol{q}\})^{-1} \boldsymbol{Q} + \mu (\mathrm{diag}\{\boldsymbol{p}\})^{-1} \ , \tag{12}$$

$\boldsymbol{p}_0^*$ is an attractive fixed point as long as

$$\mu < \mu_t = \lambda_{\min}(\boldsymbol{Q}^T (\mathrm{diag}\{\boldsymbol{q}\})^{-1} \boldsymbol{Q})/M \ . \tag{13}$$

Hence, $\boldsymbol{p}_0^*$ is a (local) maximum of $J_\mu(\boldsymbol{p})$ for $0 \leq \mu < \mu_t$.

Considering that $J_\mu(\boldsymbol{p})$ is continuous in μ and $\boldsymbol{p}$, one way to initialize Algorithm 2 is to use a slightly innovated solution from the previous optimization as the start vector $\boldsymbol{p}^{(0)}$ when incrementing $\mu > \mu_t$. The innovation is needed to move away from $\boldsymbol{p}_0^*$.

C. Recovering Sparsity

Our second continuous optimization approach is inspired from compressed sensing. A problem related to (3) is given by considering maximization of the cost function

$$I(\boldsymbol{p}) - \mu \|\boldsymbol{p}\|_0 \tag{14}$$

over probability vectors $\boldsymbol{p}$ and $\mu > 0$. Due to the second term, sparse solutions are favored, which is desirable for the selection of subsets $\boldsymbol{\mathcal{S}}$. The recovery of sparse solution has received great attention in several application fields, cf. e.g. [15]. To render optimization problems tractable, it has often been proposed to replace the $\|\cdot\|_0$ penalty function by the ℓ_1-norm. This is obviously not an option for the problem considered here, as $\|\boldsymbol{p}\|_1 = 1$. However, there are several other penalty functions that are suitable for achieving sparsity in the solution, among which the ℓ_q pseudo-norm with $0 < q < 1$ is very popular [16].

Adopting an ℓ_q penalty function, we obtain the problem formulation

$$\boldsymbol{p}^* = \underset{\boldsymbol{p} \in \breve{\boldsymbol{\mathcal{P}}}}{\mathrm{argmax}} \{I(\boldsymbol{p}) - \mu \sum_{j=1}^{M} (p_j)^q\} \ . \tag{15}$$

The cost function in (15) is a difference of convex-∩ (DC) functions and thus its optimization is generally difficult. We can however make use of an iterative procedure known as DC programming to find extreme points of $I(\boldsymbol{p}) - \mu \sum_{j=1}^{M} (p_j)^q$.

Following the presentation in [16, Algorithm 2], we arrive at the following algorithm.

Algorithm 3: Initial probability vector $\boldsymbol{p}^{(0)}$, $k = 0$

1: **do**
2: $\quad \alpha_j = \dfrac{q}{(p_j^{(k)})^{1-q} + \delta}$
3: $\quad \boldsymbol{p}^{(k+1)} = \underset{\boldsymbol{p} \in \breve{\boldsymbol{\mathcal{P}}}}{\mathrm{argmax}} \{I(\boldsymbol{p}) - \mu \sum_{j=1}^{M} \alpha_j p_j\}$
4: $\quad k \leftarrow k + 1$
5: **while** $(D(\boldsymbol{p}^{(k)} \| \boldsymbol{p}^{(k-1)}) > t)$
6: $\boldsymbol{i} =$ **index_sort**$(\boldsymbol{p}^{(k)})$
7: **for** $j = 1 : K$
8: $\quad \hat{p}_{\boldsymbol{i}(l)} = 1/K$
9: **end**
10: **for** $j = K + 1 : M$
11: $\quad \hat{p}_{\boldsymbol{i}(l)} = 0$
12: **end**
13: **Return** $\hat{\boldsymbol{p}}$

The value δ in Line 2 is a very small positive constant for numerical stability. The maximization of the convex-∩ objective function in Line 3 can be performed running Algorithm 2 with the modified update

$$p_j^{(m+1)} = p_j^{(m)} \frac{\exp(D_j^{(m)} - \mu\alpha_i)}{\sum_{l=1}^{M} p_l^{(m)} \exp(D_l^{(m)} - \mu\alpha_i)} \ . \tag{16}$$

The value $t > 0$ in Line 5 is a threshold for terminating the iteration, as in Algorithm 2. The function **index_sort**$(\boldsymbol{x})$ in Line 6 sorts the elements of $\boldsymbol{x}$ in descending order and returns the corresponding indexes. The final two for-loops in Algorithm 3 assign uniform probabilities to the selected MPPM signal elements.

By changing α_j in Algorithm 3 other penalty functions can be applied. This includes entropy $H(\boldsymbol{p})$ (which does not enforce sparsity) and Algorithm 3 can be applied to maximize the Lagrangian (8). Numerical experiments showed that Algorithm 2 gives slightly better results than Algorithm 3 when applied to maximize (8).

IV. Results and Discussion

As an interesting example of FSO transmission, we consider the quantum-limited channel and MPPM with parameters $n = 16$ and $w = 4$, for which the constellation size is $M = 1820$. These parameters were also considered, for example, in [17]. Similar trends to those shown in the following have also been observed with the case $(n, w, M) = (12, 3, 220)$.

In the quantum-limited channel an "on"-pulse is erased with probability

$$\epsilon = \mathrm{e}^{-\lambda_{\mathrm{on}}} \tag{17}$$

and λ_{on} is the mean number of photons counted at the detector during an "on"-pulse. The channel output vectors $\boldsymbol{y}_i$ correspond to all binary words of length n with at most w ones, and thus

$$N = \sum_{k=0}^{w} \binom{n}{k} \,. \tag{18}$$

The elements of the channel transition matrix are given by

$$Q_{ij} = \begin{cases} 0\,, & \text{if } \min\{\boldsymbol{x}_j - \boldsymbol{y}_i\} < 0 \\ (1-\epsilon)^{w-d_{ij}}\epsilon^{d_{ij}}\,, & \text{otherwise} \end{cases} \tag{19}$$

where $d_{ij} = \|\boldsymbol{x}_j - \boldsymbol{y}_i\|_1$.

We run Algorithm 1 with the number of searches equal to 10^3, and Algorithms 2 and 3 with the iteration terminating threshold $t = 10^{-6}$. These values result in comparable run times for all of our algorithm implementations. However, we note that constellation selection is performed only once in the design stage, and therefore run time is not a concern, as long as it is practicable. Figure 1 shows the results obtained with Algorithms 1 and 2 for different values of ϵ. For the discrete optimization Algorithm 1, we consider both uninformed random search and random search making use of the $(16, 4)$ CWC with 140 codewords [7], for $K = [64, 128, 256, 512]$. For the continuous optimization with Algorithm 2, successive initialization as described in Section III-B was used, and the obtained pairs $(I(\boldsymbol{p}_\mu), H(\boldsymbol{p}_\mu))$ are connected to a line (each point on this line is achievable by timesharing of two solutions). Also included in Figure 1 are the $(I(\boldsymbol{p}), H(\boldsymbol{p}))$ points for a particular constellation $\boldsymbol{\mathcal{S}}_1$, which was obtained as the first K elements of $\boldsymbol{\mathcal{X}}$, where $\boldsymbol{\mathcal{X}}$ with parameters (n, w) was constructed from two MPPM subsets with parameters $(n-1, w-1)$ and $(n-1, w)$ and so forth. Furthermore, the capacity for 4PPM (in bits per four 4PPM symbols, which occupy $n = 16$ time slots), which has the same n/w ratio as the considered MPPM transmission, is shown as a reference at $H(\boldsymbol{p}) = 4 \times 2 = 8$ bits.

We make the following observations. Firstly, a "poorly" chosen MPPM constellation, such as $\boldsymbol{\mathcal{S}}_1$, can result in significant losses in achievable data rate. Secondly, discrete optimization benefits noticeably from the availability of a "good" set of MPPM signal vectors, such as the codewords of an CWC, to drive the search towards favorable solutions. Furthermore, the restriction to 0-$(1/K)$ solutions $\boldsymbol{p}$ does only entail negligible losses. Or in other words, there is hardly any benefit in

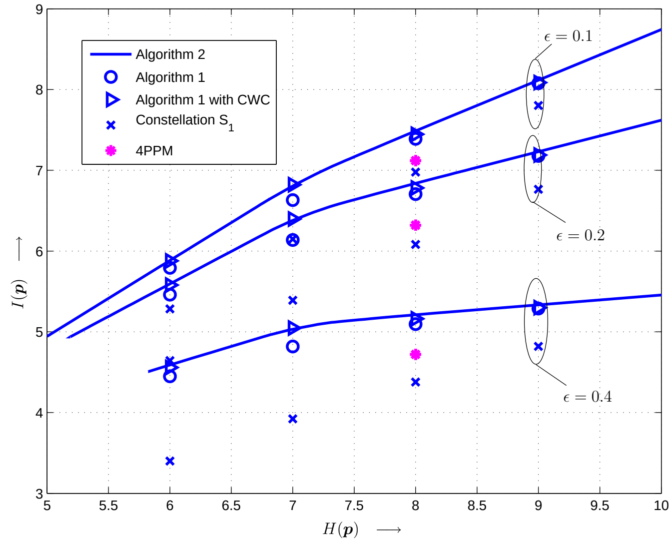

Fig. 1. $I(\boldsymbol{p})$ vs. $H(\boldsymbol{p})$ for probability vectors $\boldsymbol{p}$ optimized with Algorithms 1 and 2. $(16, 4)$-MPPM transmission over a quantum-limited channel with erasure probability ϵ. For comparison: capacity for 4PPM transmission in bits/($n = 16$ time slots).

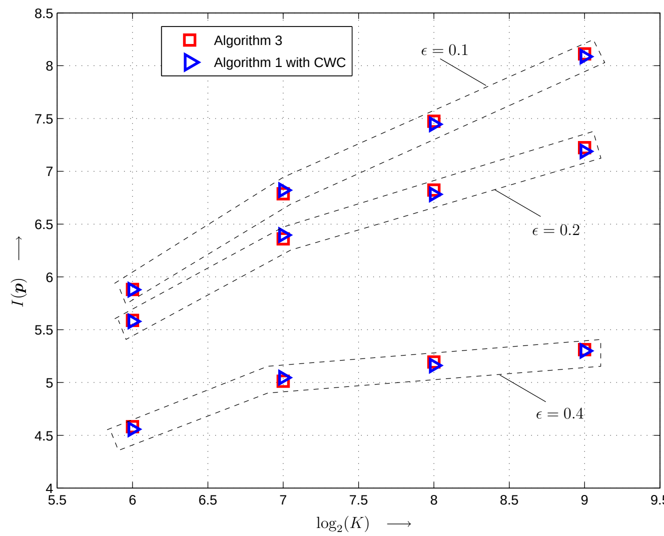

Fig. 2. $I(\boldsymbol{p})$ vs. $\log_2(K)$ for optimized $(16, 4)$-MPPM constellations and transmission over a quantum-limited channel with erasure probability ϵ.

allowing $\boldsymbol{p}$ being chosen from a continuous set and using a nonuniform MPPM signal set. Finally, we note that the properly designed $(16, 4)$ MPPM signal subset outperforms 4PPM in terms of achievable data rate. The comparison with 4PPM is appropriate and fair, as $(16, 4)$ MPPM and 4PPM have the same same peak power, average power, and bandwidth requirements.

Next, Figure 2 provides a comparison for the MPPM constellations obtained with Algorithms 1 (using CWC initialization) and 3. For Algorithm 3, a grid search for the regularization parameter was performed. We observe a close match between the mutual information attained for the constellations optimized by discrete search and DC programming. Together with the results from Figure 1, this suggest that the proposed

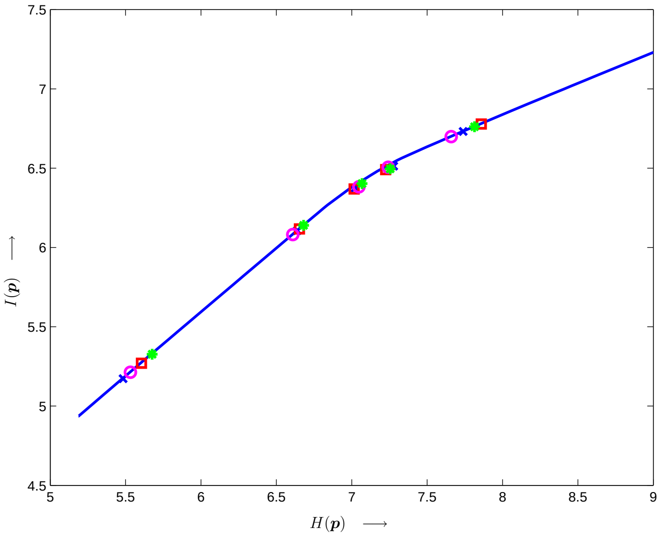

Fig. 3. $I(\boldsymbol{p})$ vs. $H(\boldsymbol{p})$ obtained with Algorithm 2. $(16, 4)$-MPPM transmission over a quantum-limited channel with erasure probability $\epsilon = 0.2$. Line: successive initialization starting from $\mu = 0$, $0 \leq \mu < 0.82$. Markers: $\boldsymbol{p}^{(0)}$ is chosen randomly, for $\mu = [0.41, 0.57, 0.61, 0.73, 0.81]$.

DC programming is a workable approach for MPPM signal design and that the subsets found with informed random search are close to optimum.

Finally, Figure 3 compares the solutions found with Algorithm 2 using different initial vectors $\boldsymbol{p}^{(0)}$ for the case of $\epsilon = 0.2$. More specifically, the results obtained with successive initialization are shown as a continuous line, and the results obtained with different randomly chosen initial vectors are shown as markers for $\mu = [0.41, 0.57, 0.61, 0.73, 0.81]$. Firstly, it can be seen that different initializations result in different solutions of the cost function $J_\mu(\boldsymbol{p})$, i.e., different local maxima are obtained for different starting solutions. Secondly, we observe that the solutions for random initialization are fairly similar to those obtained with successive initialization (in fact, a closer look at the results reveals a small advantage for successive initialization especially around $H(\boldsymbol{p}) = 7$ bits). This provides some indication that the results obtained with successive initialization are close to the global maximum of $J_\mu(\boldsymbol{p})$, which in turn emphasizes the conclusion that restriction to 0-$(1/K)$ vectors $\boldsymbol{p}$ does hardly entail any performance losses.

V. Conclusions

The selection of good (n, w)-MPPM constellations of size $K < \binom{n}{w}$ is a non-trivial problem encountered in FSO communications. In this letter, we have addressed this problems from different angles. In addition to revisiting the discrete random search from [5], we have developed two alternative approaches based on continuous optimization. We have presented numerical results for MPPM transmission over quantum-limited channels which demonstrate that the solutions from appropriately initialized discrete and from continuous optimization lead to constellations with very similar performances in terms of constrained capacity. We thus conjecture that conceptually relatively simple combinatorial optimization is an effective tool for the considered selection problem. The proposed continuous optimization algorithms can readily be applied for optimization for other constellations and channel transition matrices.

References

[1] H. Sugiyama and K. Nosu, "MPPM: a method for improving the band-utilization efficiency in optical PPM," *J. Lightwave Tech.*, vol. 7, no. 3, pp. 465–472, Mar. 1989.

[2] J. M. Budinger, M. J. Vanderaar, P. K. Wagner, S. B. Bibyk, and G. S. Mecherle, "Combinatorial pulse position modulation for power-efficient free-space laser communications," in *SPIE Proc.*, vol. 1866, Jan. 1993, pp. 214–225.

[3] J. Hamkins and B. Moision, "Multipulse pulse-position modulation on discrete memoryless channels," *Interplanetary Network Progress Report*, vol. 42, p. 161, May 2005.

[4] T. T. Nguyen and L. Lampe, "Coded multipulse pulse-position modulation for free-space optical communications," *IEEE Trans. Commun.*, vol. 58, no. 4, pp. 1036–1041, Apr. 2010.

[5] ——, "Capacity-maximized MPPM constellations for free-space optical communications," in *Proc. 6th Symp. Commun. Sys., Networks Digital Sig. Processing*, Graz, Austria, Jul. 2008, pp. 97–101.

[6] J. Dauwels, "Numerical computation of the capacity of continuous memoryless channels," in *Proc. 26th Symp. Inform. Theory in the BENELUX*, Brussels, Belgium, May 2005, pp. 221–228.

[7] A. Brouwer, J. Shearer, N. Sloane, and W. Smith, "A new table of constant weight codes," *IEEE Trans. Inform. Theory*, vol. 36, no. 6, pp. 1334–1380, 1990.

[8] D. H. Smith, L. A. Hughes, and S. Perkins, "A new table of constant weight codes of length greater than 28," *Electron. J. Combinatorics*, vol. 13, no. A2, 2006.

[9] F. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 913–929, May 1995.

[10] A. Farid and S. Hranilovic, "Design of non-uniform capacity-approaching signaling for optical wireless intensity channels," in *Proc. IEEE Intl. Symp. Inform. Theory*, Toronto, Canada, Jul. 2008, pp. 2327–2331.

[11] P. Chou, T. Lookabaugh, and R. Gray, "Entropy-constrained vector quantization," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 1, pp. 31–42, Jan. 1989.

[12] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. 18, no. 1, pp. 14–20, Jan. 1972.

[13] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. 18, no. 4, pp. 460–473, Jul. 1972.

[14] G. Matz and P. Duhamel, "Information-geometric formulation and interpretation of accelerated Blahut-Arimoto-type algorithms," in *Inform. Theory Workshop (ITW)*, San Antonion, Texas, Oct. 2004, pp. 66–70.

[15] E. Candès and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[16] G. Gasso, A. Rakotomamonjy, and S. Canu, "Recovering sparse signals with a certain family of nonconvex penalties and DC programming," *IEEE Trans. Signal Processing*, vol. 57, no. 12, pp. 4686–4698, Dec. 2009.

[17] C. Georghiades, "Modulation and coding for throughput-efficient optical systems," *IEEE Trans. Inform. Theory*, vol. 40, no. 5, pp. 1313–1326, Sep. 1994.