PEAK-TO-AVERAGE POWER RATIO REDUCTION IN OFDM SYSTEMS BY THE ADAPTIVE PROJECTED SUBGRADIENT METHOD

Renato L. G. Cavalcante and Isao Yamada

Department of Communications and Integrated Systems
Tokyo Institute of Technology, Tokyo 152-8552, Japan
Email: {renato,isao}@comm.ss.titech.ac.jp

ABSTRACT

One of the main drawbacks of the OFDM modulation is the high peak-to-average power ratio (PAPR) of the transmitted signal. In this study, we devise a low-complexity transmitter that mitigates the PAPR problem and satisfies multiple requirements of a given system, such as error vector magnitude (EVM) constraints, minimum transmitted power of the data symbols, among others. To reduce the PAPR, we use the adaptive projected subgradient method to suppress a sequence of convex cost functions over closed convex sets describing desired properties of the transmitted signal. Numerical examples show that the proposed scheme reduces the PAPR to reasonable levels with few iterations.

Index Terms— OFDM, adaptive projected subgradient method, optimization methods

1. INTRODUCTION

OFDM symbols have high power peaks compared with the mean power. Consequently, to prevent distortion of the transmitted signal, the power amplifiers of OFDM systems need to operate in the linear region for a wide range of instantaneous input power. Unfortunately, amplifiers with a large dynamic range are either costly or power inefficient. In general, the power consumption is more related to the peak power output rather than the average power output [1]. Therefore, peak-to-average power ratio (PAPR) reduction methods have been investigated to alleviate the dynamic range requirements of the amplifiers.

Techniques to mitigate the PAPR problem include, among others, amplitude clipping and filtering, coding, selected mapping, interleaving, tone reservation (TR), and active constellation extension (see [2] and the references therein for a good tutorial paper on this topic). No single technique is the best solution for all possible applications, so in this study the focus is on schemes in that all processing is done at the transmitter. More precisely, we propose transmitters based on a combination of the TR technique with others that slightly disturb the data symbols.

In systems applying the TR method, the transmitter reduces the PAPR problem by sending dummy symbols – i.e, symbols not carrying data – in some selected subcarriers [1, 3]. The optimal values of the dummy symbols are the solution of a certain convex optimization problem related to the PAPR. TR-based transmitters differ in the choice of the (convex) cost function, the possible constraint sets, and the algorithm used to find (or approximate) an optimal solution.

A promissing low-complexity TR-based algorithm is the active set approach [1], a special case of the subgradient method [3] based on Polyak's algorithm. The subgradient method is very efficient when a constraint set is imposed and the projection onto the set is computationally simple. However, when the constraint is the intersection of multiple sets, the projection onto the intersection is computationally expensive in general, so the subgradient method in [3] loses its basic advantage – the low complexity. Multiple constraint sets can be imposed to increase the PAPR reduction capability of the system. They can describe, for example, the allowed range of distortion of the data-bearing carriers [4]. Unfortunately, most proposals using multiple constraints solve the optimization problem with computationally complex interior point methods.

Recently, the adaptive projected subgradient method (APSM) has been introduced in [5] (see also [6]) to extend Polyak's algorithm to the case where we aim at suppressing a sequence of convex cost functions. In addition, the APSM can also asymptotically minimize (with low complexity) the sequence of cost functions over the intersection of many closed convex sets [7]. We devise a sequence of cost functions closely related to the PAPR and use the APSM to suppress this sequence over closed convex sets describing desired characteristics of the OFDM symbol such as those introduced in [4]. Numerical simulations show that the resulting transmitters have low complexity and reduce the PAPR to a reasonable level with few iterations.

2. PRELIMINARIES

2.1. Adaptive projected subgradient method (APSM) [7]

For every vector $\boldsymbol{v} \in \mathbb{R}^M$, we define the norm of $\boldsymbol{v}$ by $\|\boldsymbol{v}\| := \sqrt{\boldsymbol{v}^T\boldsymbol{v}}$, which is the norm induced by the Euclidean inner product $\langle \boldsymbol{v}, \boldsymbol{y}\rangle := \boldsymbol{v}^T\boldsymbol{y}$ for every $\boldsymbol{v}, \boldsymbol{y} \in \mathbb{R}^M$. For a vector $\boldsymbol{v} \in \mathbb{C}^M$, the norm of $\boldsymbol{v}$ is defined by $\|\boldsymbol{v}\| := \sqrt{\boldsymbol{v}^H\boldsymbol{v}}$. For $\boldsymbol{v} = [v_1\ v_2\ \dots\ v_M]^T \in \mathbb{C}^M$, the infinity norm of $\boldsymbol{v}$ is defined by $\|\boldsymbol{v}\|_\infty := \max_{1\le i\le M}\{|v_i|\}$. A set is said to be convex if $\boldsymbol{v} = \alpha\boldsymbol{v}_1 + (1-\alpha)\boldsymbol{v}_2 \in C \subset \mathbb{R}^M$ for every $\boldsymbol{v}_1, \boldsymbol{v}_2 \in C$ and $0 \le \alpha \le 1$. Similarly, a set $\mathcal{C} \in \mathbb{C}^M$ is convex if $\boldsymbol{v} = \alpha\boldsymbol{v}_1 + (1-\alpha)\boldsymbol{v}_2 \in \mathcal{C}$ for every $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathcal{C}$ and $0 \le \alpha \le 1$ ($\alpha \in \mathbb{R}$). Let $C \subset \mathbb{R}^M$ be a nonempty closed convex set. The projection operator $P_C : \mathbb{R}^M \to C$ maps $\boldsymbol{v} \in \mathbb{R}^M$ to the unique vector $P_C(\boldsymbol{v}) \in C$ satisfying $\|\boldsymbol{v} - P_C(\boldsymbol{v})\| = \min_{\boldsymbol{y}\in C}\|\boldsymbol{v} - \boldsymbol{y}\|$.

Let the operator $T : \mathbb{R}^M \to \mathbb{R}^M$ be the concatenation (or composition) of projections onto m closed convex sets C_i ($i = 1, \dots, m$), i.e., $T := P_{C_m}P_{C_{m-1}}\cdots P_{C_1}$, where we assume $\cap_{i=1}^m C_i \neq \emptyset$, and $\Theta_n : \mathbb{R}^M \to [0, \infty)$ $(\forall n \in \mathbb{N})$ be a sequence of continuous convex functions. The subdifferential of Θ_n at a given point $\boldsymbol{y}$ is the set of all subgradients of Θ_n at this point, i.e.,

$$\partial\Theta_n(\boldsymbol{y}) := \{\boldsymbol{a} \in \mathbb{R}^M | \Theta_n(\boldsymbol{y}) + \langle \boldsymbol{x} - \boldsymbol{y}, \boldsymbol{a}\rangle \le \Theta_n(\boldsymbol{x}), \forall \boldsymbol{x} \in \mathbb{R}^M\} \neq \emptyset.$$

For an arbitrarily given $\boldsymbol{w}_0 \in \mathbb{R}^M$, the adaptive projected subgradient method minimizes asymptotically Θ_n over $\bigcap_{i=1}^m C_i$ with

the sequence $(\boldsymbol{w}_n)_{n\in\mathbb{N}}$ given by

$$\boldsymbol{w}_{n+1} := \begin{cases} T\left(\boldsymbol{w}_n - \lambda_n \dfrac{\Theta_n(\boldsymbol{w}_n)}{\|\Theta_n'(\boldsymbol{w}_n)\|^2}\Theta_n'(\boldsymbol{w}_n)\right), \\ \qquad \text{if} \quad \Theta'_n(\boldsymbol{w}_n) \neq \boldsymbol{0} \\ T(\boldsymbol{w}_n) \quad \text{otherwise,} \end{cases} \tag{1}$$

where $\Theta_n^{'}(\boldsymbol{w}_n) \in \partial\Theta_n(\boldsymbol{w}_n)$, $\partial\Theta_n(\boldsymbol{w}_n)$ is the subdifferential of Θ_n at $\boldsymbol{w}_n$, and $0 \leq \lambda_n \leq 2$. For the convergence properties of the algorithm in (1), the reader is referred to [5, 7].

2.2. System Model

Assume a transmitter sending N symbols with the OFDM modulation. We denote the OFDM constellation (the OFDM symbol in frequency domain) by $\boldsymbol{c} := [c_1 \ \cdots c_N]^T \in \mathbb{C}^N$. A discrete-time approximation of the continuous-time OFDM symbol can be obtained by applying the inverse discrete Fourier transform (IDFT) with L-times oversampling to $\boldsymbol{c}$ [1, 3, 4], i.e.,

$$\boldsymbol{x} = \boldsymbol{F}\boldsymbol{c},$$

where $\boldsymbol{x} := [x_1 \ \ldots \ x_{NL}]^T \in \mathbb{C}^{NL}$ is the discrete time-domain OFDM symbol and the element of the lth row and mth column of the matrix $\boldsymbol{F} \in \mathbb{C}^{NL\times N}$ is given by

$$\boldsymbol{F}_{lm} = \frac{1}{\sqrt{NL}} e^{j2\pi(l-1)[m-(N/2)]/(NL)} \quad \begin{matrix} l = 1, \ldots, NL \\ m = 1, \ldots, N. \end{matrix}$$

By using $L \geq 4$, the power peaks of the vector $\boldsymbol{x}$ are good approximations of the power peaks of the continuous-time transmitted signal [4]. For simplicity, we assume that D subcarriers are used for data transmission and the remaining $N - D$ subcarriers can be used for PAPR reduction. Let $\mathcal{I} := \{i_1, \ldots, i_D\}$ be the set of indices indicating the subcarriers available for data transmission, i.e., $i \in \mathcal{I}$ iff c_i is encoded data. As in [4], we denote the diagonal carrier selection matrix by $\boldsymbol{S} = \mathrm{diag}(S_1, \ldots, S_N) \in \mathbb{R}^{N\times N}$, where

$$S_i = \begin{cases} 1, & \text{if} \quad i \in \mathcal{I} \\ 0, & \text{otherwise.} \end{cases}$$

The basic idea of PAPR reduction algorithms is to transmit an OFDM constellation $\widetilde{\boldsymbol{c}} = [\widetilde{c}_1 \ \ldots \ \widetilde{c}_N]^T$, which is a modified version of $\boldsymbol{c}$, so that the PAPR of $\widetilde{\boldsymbol{x}} = \boldsymbol{F}\widetilde{\boldsymbol{c}}$ is lower than that of $\boldsymbol{x} = \boldsymbol{F}\boldsymbol{c}$. For a possibly modified discrete signal $\widetilde{\boldsymbol{x}} = \boldsymbol{F}\widetilde{\boldsymbol{c}} \in \mathbb{C}^{NL}$, we define the PAPR as

$$\mathrm{PAPR} = \frac{\|\widetilde{\boldsymbol{x}}\|_\infty^2}{\frac{1}{NL}E\left(\|\boldsymbol{F}\boldsymbol{S}\widetilde{\boldsymbol{c}}\|^2\right)} = \frac{\|\widetilde{\boldsymbol{x}}\|_\infty^2}{\frac{1}{NL}E\left(\|\boldsymbol{S}\widetilde{\boldsymbol{c}}\|^2\right)}. \tag{2}$$

In this study, we only consider PAPR reduction methods that do not require any modification in the receiver, i.e., the receiver is able to recover the transmitted symbol c_i ($i \in \mathcal{I}$) directly from $\widetilde{c}_i$.

2.3. PAPR optimization based on Polyak's method

In [3], the transmitted constellation $\widetilde{\boldsymbol{c}} = \boldsymbol{F}^H\widetilde{\boldsymbol{x}}$ is decomposed into two vectors $\boldsymbol{\rho}$ and $\boldsymbol{\varphi}$ such that

$$\widetilde{\boldsymbol{x}} = \boldsymbol{\Gamma}\boldsymbol{\rho} + \boldsymbol{U}\boldsymbol{\varphi}, \tag{3}$$

where $\boldsymbol{\rho} \in \mathbb{C}^{N-D}$ is the vector of tones c_i ($i \notin \mathcal{I}$) to be used for PAPR reduction and $\boldsymbol{\Gamma} \in \mathbb{C}^{NL\times(N-D)}$ is a matrix with the corresponding columns of $\boldsymbol{F}$. Similarly, the vector $\boldsymbol{\varphi} \in \mathbb{C}^D$ is the vector containing the data symbols c_i ($i \in \mathcal{I}$), and $\boldsymbol{U} \in \mathbb{C}^{NL\times D}$ is the matrix with the corresponding columns of $\boldsymbol{F}$.

With the decomposition in (3), if only $\boldsymbol{\rho}$ is allowed to change, ideal PAPR reduction algorithms should minimize $J(\boldsymbol{\rho}) := \|\widetilde{\boldsymbol{x}}\|_\infty^2$ subject to a constraint describing desired characteristics of the vector $\boldsymbol{\rho}$. However, instead of minimizing $J(\boldsymbol{\rho})$ directly, iterative algorithms can reduce the PAPR to reasonable levels with few low-complexity iterations when an alternative cost function is used. Such a cost function is used in the optimization problem given by [3]

$$\boldsymbol{\rho}^\star \in \arg\min_{\boldsymbol{\rho}\in C_s} \sum_{l=1}^{K} \omega_l |\widetilde{x}_{k_l}|^p, \tag{4}$$

where K is the number of peaks to be considered, $\widetilde{\boldsymbol{x}} := [\widetilde{x}_1 \ \ldots \ \widetilde{x}_{NL}]^T$, $|\widetilde{x}_{k_1}| \geq |\widetilde{x}_{k_2}| \geq \cdots \geq |\widetilde{x}_{k_{NL}}|$, the ordered tuple $(k_1, \ldots, k_{NL})$ is a permutation of $(1, \ldots, NL)$, ω_l is a non-negative weight, p is the integer exponent, and C_s is a closed convex set that describes desired characteristics of the signal (e.g., power spectrum density masks). If the projection onto the set C_s is simple, Polyak's algorithm can be used to approximate $\boldsymbol{\rho}^\star$ with few iterations and low complexity [3].

In certain cases, replacing C_s by a set that is the intersection of many closed convex sets is useful to reduce the PAPR problem further. In these situations, Polyak's method can only be applied if the exact projection onto this intersection is computed, a potentially computationally expensive task. In addition, for large K, finding good values for the weights can be challenging. The weights should the same for all iterations of the algorithm, otherwise Polyak's method does not guarantee convergence because the cost function changes at each iteration. In Sect. 3 we introduce an algorithm that can deal with multiple sets and can give larger weights to larger peaks at each iteration. Before proceding with the proposed algorithm, we give examples of sets that can be used to reduce the PAPR problem further.

2.4. PAPR problem with multiple convex constraint sets

The problem in (4) does not allow any distortion in the data symbols c_i ($i \in \mathcal{I}$). However, most standards take into account the imperfections of practical transmitters. A constellation $\widetilde{\boldsymbol{c}}$ is acceptable if it belongs to the error vector magnitude (EVM) constraint set [4, 8]

$$C_{\mathrm{EVM}} := \{\widetilde{\boldsymbol{c}} \in \mathbb{C}^N | \quad \|\boldsymbol{S}(\widetilde{\boldsymbol{c}} - \boldsymbol{c})\| \leq \epsilon\}, \tag{5}$$

where $\epsilon \leq \mathrm{EVM}_{\max}\sqrt{DP_0}$, P_0 is the average power of the carrier modulation, and $\mathrm{EVM}_{\max}$ is maximum allowed distortion, usually specified by the standard being implemented.

To reduce the PAPR without reducing the average power of the data, we can impose the constraint $\|\boldsymbol{S}\widetilde{\boldsymbol{c}}\|^2 \geq \|\boldsymbol{S}\boldsymbol{c}\|^2$, which defines a nonconvex set. A convex relaxation of this constraint is obtained by expanding $\|\boldsymbol{S}(\widetilde{\boldsymbol{c}} - \boldsymbol{c})\|^2$ (as in [4])

$$\begin{aligned} 2\mathrm{Re}(\widetilde{\boldsymbol{c}}^H\boldsymbol{S}\boldsymbol{c}) &= \|\boldsymbol{S}\widetilde{\boldsymbol{c}}\|^2 + \|\boldsymbol{S}\boldsymbol{c}\|^2 - \|\boldsymbol{S}(\widetilde{\boldsymbol{c}} - \boldsymbol{c})\|^2 \\ &\geq 2\|\boldsymbol{S}\boldsymbol{c}\|^2 - \|\boldsymbol{S}(\widetilde{\boldsymbol{c}} - \boldsymbol{c})\|^2 \\ &\geq 2\|\boldsymbol{S}\boldsymbol{c}\|^2 - \epsilon^2, \end{aligned}$$

where the last inequality comes from the set C_{EVM}. The power constraint set is thus

$$C_P := \left\{\widetilde{\boldsymbol{c}} \in \mathbb{C}^N | \mathrm{Re}(\widetilde{\boldsymbol{c}}^H\boldsymbol{S}\boldsymbol{c}) \geq \|\boldsymbol{S}\boldsymbol{c}\|^2 - \frac{\epsilon^2}{2}\right\}. \tag{6}$$

Ideally, the proposed algorithm in Sect. 3 should find $\widetilde{\boldsymbol{c}}^\star$ given by

$$\widetilde{\boldsymbol{c}}^\star \in \arg\min_{\widetilde{\boldsymbol{c}}\in C_I} \|\boldsymbol{F}\widetilde{\boldsymbol{c}}\|_\infty^2, \tag{7}$$

where $C_I := C_{\rm EVM} \cap C_P$. Note that additional constraints (such as those considering power spectral density masks and power overheads [4]) can be easily introduced in the proposed algorithm. For brevity of exposition, we only consider the sets $C_{\rm EVM}$ and C_P in this study.

Next, we devise a related cost function and a low-complexity algorithm that can provide a good estimate of $\widetilde{\boldsymbol{c}}^\star$ with few iterations.

3. PROPOSED ALGORITHM

For any vector $\boldsymbol{v} \in \mathbb{C}^N$, we define the "underline" operator as

$$\underline{\boldsymbol{v}} := \left[\mathrm{Re}(\boldsymbol{v}^T) \quad \mathrm{Im}(\boldsymbol{v}^T)\right]^T =: [\underline{v}_1 \ \ldots \ \underline{v}_{2N}]^T \in \mathbb{R}^{2N}.$$

The "underline" version of a set $C \subset \mathbb{C}^N$ is given by

$$\underline{C} := \{\underline{\boldsymbol{v}} = [\boldsymbol{v}_R^T \ \boldsymbol{v}_I^T]^T \in \mathbb{R}^{2N} | \\ \boldsymbol{v}_R + j\boldsymbol{v}_I \in C, \ \boldsymbol{v}_R, \boldsymbol{v}_I \in \mathbb{R}^N\},$$

where $j = \sqrt{-1}$.

Therefore, the "underline" versions of the sets $C_{\rm EVM}$ and C_P are equivalently expressed by

$$\underline{C}_{\rm EVM} := \left\{\widetilde{\underline{\boldsymbol{c}}} \in \mathbb{R}^{2N} | \left\| \begin{bmatrix} \boldsymbol{S} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S} \end{bmatrix} (\widetilde{\underline{\boldsymbol{c}}} - \underline{\boldsymbol{c}}) \right\|^2 \le \epsilon^2 \right\}$$

and

$$\underline{C}_P = \left\{\widetilde{\underline{\boldsymbol{c}}} \in \mathbb{R}^{2N} | \widetilde{\underline{\boldsymbol{c}}}^T \begin{bmatrix} \boldsymbol{S} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S} \end{bmatrix} \underline{\boldsymbol{c}} \ge \|\boldsymbol{S}\boldsymbol{c}\|^2 - \frac{\epsilon^2}{2} \right\}.$$

We also define the function

$$\phi_k(\widetilde{\underline{\boldsymbol{c}}}) := \widetilde{\underline{\boldsymbol{c}}}^T \boldsymbol{M}_k \widetilde{\underline{\boldsymbol{c}}}, \tag{8}$$

where

$$\boldsymbol{M}_k = \begin{bmatrix} \mathrm{Re}(\boldsymbol{f}_k^H \boldsymbol{f}_k) & -\mathrm{Im}(\boldsymbol{f}_k^H \boldsymbol{f}_k) \\ \mathrm{Im}(\boldsymbol{f}_k^H \boldsymbol{f}_k) & \mathrm{Re}(\boldsymbol{f}_k^H \boldsymbol{f}_k) \end{bmatrix},$$

$\boldsymbol{f}_k \in \mathbb{C}^{1\times N}$ is the kth row of the matrix $\boldsymbol{F}$, and $k \in \{1, \ldots, NL\}$. Note that, for given k and $\widetilde{\underline{\boldsymbol{c}}}$, the function $\phi_k(\widetilde{\underline{\boldsymbol{c}}})$ satisfies $\phi_k(\widetilde{\underline{\boldsymbol{c}}}) = |\widetilde{x}_k|^2$.

At each iteration, the proposed iterative algorithm aims at approaching a minimizer of

$$\Theta_n(\widetilde{\underline{\boldsymbol{c}}}) := h\left(\sum_{j=1}^K \omega_j^{(n)} \phi_{k_j}(\widetilde{\underline{\boldsymbol{c}}}) - \gamma\right), \tag{9}$$

where $h(x) = \max(x, 0)$, $(k_1, \ldots, k_K)$ are K indices corresponding to the elements of $\widetilde{\boldsymbol{x}} = \boldsymbol{F}\widetilde{\boldsymbol{c}}$ with the K largest magnitude, $\omega_j^{(n)}$ are non-negative weights (at iteration n) satisfying $\sum_{j=1}^K \omega_j^{(n)} = 1$, and γ is a non-negative constant (c.f. Proposition 1). The cost function $\Theta_n(\widetilde{\underline{\boldsymbol{c}}})$ is essentially a measure of the (weighted) total energy of the K largest magnitude components of the vector $\widetilde{\boldsymbol{x}}$ (recall that $\phi_k(\widetilde{\underline{\boldsymbol{c}}}) = |\widetilde{x}_k|^2$). Note that, unlike the cost function in (4), the weights in (9) are allowed to change at each iteration, hence the cost function is time varying (see also [5, 6]).

For a given estimate $\widetilde{\boldsymbol{c}}_n$ of $\widetilde{\boldsymbol{c}}^\star$, to decrease rapidly the largest peak of $\widetilde{\boldsymbol{x}}_n = \boldsymbol{F}\widetilde{\boldsymbol{c}}_n$ considered in (9), we assign larger weights to peaks with larger magnitude. We propose the following weighting scheme:

$$\omega_j^{(n)} = \frac{\phi_{k_j}(\widetilde{\underline{\boldsymbol{c}}}_n)}{\sum_{l=1}^K \phi_{k_l}(\widetilde{\underline{\boldsymbol{c}}}_n)}, \quad j = 1, \ldots, K, \tag{10}$$

where $(k_1, \ldots, k_K)$ are as defined in (9). The summation in (10) is only a constant used to guarantee that $\sum_{j=1}^K \omega_j^{(n)} = 1$.

Applying the scheme in (1) to suppress the sequence $\Theta_n(\widetilde{\underline{\boldsymbol{c}}})$ in (9) over $\underline{C}_I := \underline{C}_{\rm EVM} \cap \underline{C}_P$, for an arbitrary $\widetilde{\underline{\boldsymbol{c}}}_0 \in \mathbb{R}^{2N}$, we arrive at sequence given by

$$\widetilde{\underline{\boldsymbol{c}}}_{n+1} := \begin{cases} P_{C_{\rm EVM}} P_{C_P}\left(\widetilde{\underline{\boldsymbol{c}}}_n - \lambda_n \dfrac{\Theta_n(\widetilde{\underline{\boldsymbol{c}}}_n)}{\|\Theta_n'(\widetilde{\underline{\boldsymbol{c}}}_n)\|^2} \Theta_n'(\widetilde{\underline{\boldsymbol{c}}}_n)\right), \\ \qquad \text{if} \quad \Theta_n'(\widetilde{\underline{\boldsymbol{c}}}_n) \neq \boldsymbol{0} \\ P_{C_{\rm EVM}} P_{C_P}(\widetilde{\underline{\boldsymbol{c}}}_n) \quad \text{otherwise,} \end{cases} \tag{11}$$

where

$$\Theta_n'(\widetilde{\underline{\boldsymbol{c}}}_n) = \begin{cases} \dfrac{1}{\sum_{l=1}^K \phi_{k_l}(\widetilde{\underline{\boldsymbol{c}}}_n)} \sum_{j=1}^K \phi_{k_j}(\widetilde{\underline{\boldsymbol{c}}}_n)(\boldsymbol{M}_{k_j} + \boldsymbol{M}_{k_j}^T)\widetilde{\underline{\boldsymbol{c}}}_n, \\ \qquad \text{if} \quad \Theta_n(\widetilde{\underline{\boldsymbol{c}}}_n) > 0 \\ \boldsymbol{0}, \quad \text{otherwise,} \end{cases}$$

$$P_{\underline{C}_P}(\widetilde{\underline{\boldsymbol{c}}}) = \widetilde{\underline{\boldsymbol{c}}} - \left(\widetilde{\underline{\boldsymbol{c}}}^T \underline{\boldsymbol{r}} - \|\boldsymbol{S}c\|^2 + \frac{\epsilon^2}{2}\right) \frac{\underline{\boldsymbol{r}}}{\|\underline{\boldsymbol{r}}\|^2},$$

$$\underline{\boldsymbol{r}} = \begin{bmatrix} \boldsymbol{S} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S} \end{bmatrix} \underline{\boldsymbol{c}},$$

$(k_1, \ldots, k_K)$ are as defined in (9), and the projection onto $\underline{C}_{\rm EVM}$ is shown below.

Fact 1 *Let $\widetilde{\underline{\boldsymbol{c}}}$ be a point outside $\underline{C}_{\rm EVM}$. Then the projection onto $\underline{C}_{\rm EVM}$, denoted by $\underline{\boldsymbol{g}} := [\underline{g}_1 \ \ldots \underline{g}_{2N}]^T$, is given by*

$$\underline{g}_i = \begin{cases} \dfrac{\widetilde{\underline{c}}_i + \kappa \underline{c}_i}{1+\kappa}, \quad \text{if} \quad i \in \mathcal{I} \quad \text{or} \quad i - N \in \mathcal{I}, \\ \widetilde{\underline{c}}_i, \quad \text{otherwise,} \end{cases} \tag{12}$$

where

$$\kappa = \frac{\left\| \begin{bmatrix} \boldsymbol{S} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S} \end{bmatrix} (\widetilde{\underline{\boldsymbol{c}}} - \underline{\boldsymbol{c}}) \right\|}{\epsilon} - 1.$$

Proof: The proof follows from [9, Th. 3.4-1]. ■

(NOTE: i) Multiplying $\boldsymbol{S}$ by a vector is a simple operation because we only need to set to zero some elements of the vector. Therefore, by also using $(\boldsymbol{M}_k + \boldsymbol{M}_k^T)\widetilde{\underline{\boldsymbol{c}}} = 2[\mathrm{Re}(\boldsymbol{f}_k^H \widetilde{x}_k)^T \mathrm{Im}(\boldsymbol{f}_k^H \widetilde{x}_k)^T]^T$ and $\phi_k(\widetilde{\underline{\boldsymbol{c}}}) = |\widetilde{x}_k|^2$, there is no need to perform any matrix–vector multiplication in (11) once $\boldsymbol{F}\widetilde{\boldsymbol{c}}_n$ is known. The calculation of $\boldsymbol{F}\widetilde{\boldsymbol{c}}_n$, which is an IDFT, dominates the complexity of the algorithm, but it can be computed with efficient algorithms such as the inverse fast Fourier transform. ii) Owing to the special combination of $C_{\rm EVM}$ and C_P, it is guaranteed that $\widetilde{\underline{\boldsymbol{c}}}_n \in \underline{C}_I$.)

In summary, the iteration in (11) approximates $\widetilde{\underline{\boldsymbol{c}}}^\star$ in (7) by suppressing the time-varying cost function (9) over $\underline{C}_I$. In the next proposition, we show that the estimate of $\widetilde{\underline{\boldsymbol{c}}}^\star$ can be improved at each iteration.

Proposition 1 *(On the convergence of the algorithm)*
Let $\mu = \min_{\widetilde{\underline{\boldsymbol{c}}} \in C_I} \|\boldsymbol{F}\widetilde{\underline{\boldsymbol{c}}}\|_\infty^2$, then following holds.

- *(a) If $\gamma \geq \mu$ and $\lambda_n \in (0,2)$, then the iteration in (11) satisfies $\|\widetilde{\underline{\boldsymbol{c}}}_{n+1} - \widetilde{\underline{\boldsymbol{c}}}^\star\| \leq \|\widetilde{\underline{\boldsymbol{c}}}_n - \widetilde{\underline{\boldsymbol{c}}}^\star\|$ for any $\boldsymbol{c}^\star$ in (7).*
- *(b) For a given estimate $\widetilde{\underline{\boldsymbol{c}}}_n$ of any $\widetilde{\underline{\boldsymbol{c}}}^\star$ in (7) such that $\sum_{j=1}^{K} \omega_j^{(n)} \phi_{k_j}(\widetilde{\underline{\boldsymbol{c}}}_n) - \mu > 0$, if γ in (9) is set to zero, the iteration in (11) with a step size within the range*

$$0 < \lambda_n < 2\left(1 - \frac{\mu}{\Theta_n(\widetilde{\underline{\boldsymbol{c}}}_n)}\right) \tag{13}$$

guarantees that $\|\widetilde{\underline{\boldsymbol{c}}}_{n+1} - \widetilde{\underline{\boldsymbol{c}}}^\star\| < \|\widetilde{\underline{\boldsymbol{c}}}_n - \widetilde{\underline{\boldsymbol{c}}}^\star\|$.

Proof: The proof is omitted due to lack of space. ■

Knowledge of $\min_{\widetilde{\boldsymbol{c}} \in C_I} \|\boldsymbol{F}\widetilde{\boldsymbol{c}}\|_\infty^2$ is not available in general, but Proposition 1(a) suggests practical transmitters using γ with a value satisfying $\Theta_n(\widetilde{\underline{\boldsymbol{c}}}^\star) = 0$ for most transmitted symbols. Alternatively, Proposition 1(b) suggests the choice $\gamma = 0$ with a decreasing step size such as $\lambda_n < 2/n$. Even if the conditions of Proposition 1(b) are violated with such a choice, practically, after some iterations we can expect that the step size will be small enough so that the conditions are again satisfied. We show in the next section that the second choice of parameters ($\gamma = 0$ and $\lambda_n < 2/n$) gives excellent performance with few iterations in practice.

4. SIMULATIONS AND FINAL REMARKS

In Fig. 1, we compare the following approaches:

- (subgradient) the subgradient method applied to the cost function in (4) with $\omega_l = 1$, $p = 2$, $K = 20$, $C_s := \mathbb{C}^{N-D}$, and step size $\lambda_n = 2/n$ (except for the step size[1], other parameters of the simulations are the same as in [3, Sect. 4]);
- (proposed-uniform) the proposed algorithm with $K = 20$ and uniform weighting (so that the gain of applying (10) is highlighted);
- (proposed) the proposed algorithm in (11) with $K = 20$ (weights as in (10));
- (subgradient-optimal) $\widetilde{\boldsymbol{x}} = \boldsymbol{\Gamma}\boldsymbol{\rho}^\star + \boldsymbol{U}\boldsymbol{\varphi}$, where $\boldsymbol{\rho}^\star = \arg\min_{\boldsymbol{\rho} \in \mathbb{C}^{N-D}} \|\boldsymbol{\Gamma}\boldsymbol{\rho} + \boldsymbol{U}\boldsymbol{\varphi}\|_\infty$, which is the best performance that can be obtained with the subgradient method in [3];
- (proposed-optimal) the ideal performance of the proposed algorithm, i.e., the performance of $\boldsymbol{c}^\star$ in (7) (see also [4]).

For the first three methods, the iterative algorithms, the number of iterations is indicated in the figure. The set C_{EVM} allows an EVM up to -20dB. We do not assume knowledge of $\min_{\widetilde{\boldsymbol{c}} \in C_I} \|\boldsymbol{F}\widetilde{\boldsymbol{c}}\|_\infty^2$, so we set the parameter γ of the proposed algorithms to zero and the step size to $\lambda_n = 1.9/n$ (see discussion after Proposition 1), where n is the iteration number.

The simulated system is based on the 802.11a standard [8], where $D = 52$ out of $N = 64$ carriers contain 16-QAM modulated data or pilot symbols. The empirical PAPR complementary cumulative distribution function (CCDF) is obtained by simulating 100,000 OFDM symbols, the oversampling factor of which is set to $L = 4$ (for the algorithms and the CCDF plot).

Compared to the subgradient approach of [3], the proposed algorithms are able to find an OFDM symbol with much lower PAPR because all components of the vector $\boldsymbol{c}$ are allowed to change. The uniform weighting scheme is clearly outperformed by the weighting scheme in (10), the transmitter of which achieves the performance limit of the subgradient method in only three iterations. Note that most of the performance gains are obtained in three iterations because the step size becomes too small for $n \geq 4$.

[1] In the present simulation, the step size used in [3, Sect. 4] resulted in worse performance than $\lambda_n = 2/n$.

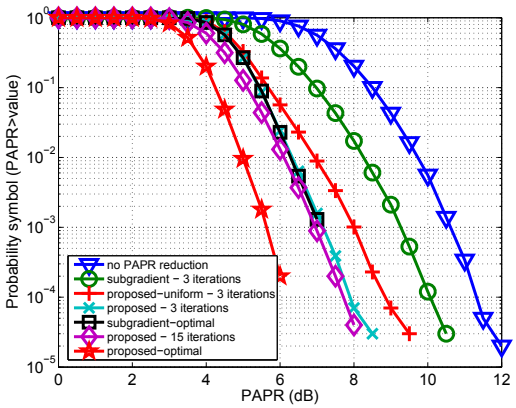

Fig. 1. Empirical PAPR CCDF of the algorithms.

To reduce the complexity of the proposed transmitters, we can easily modify them in such a way that only a few selected data carriers are distorted. Finally, different (nonsmooth) cost functions, possibly taking into account the nonlinearities of the amplifiers, and more sets can also be easily applied to the proposed algorithm.

Acknowledgment: The authors would like to thank Prof. Kohichi Sakaniwa of Tokyo Institute of Technology for helpful suggestions and comments. This work was partially supported by JSPS Grants-in-Aid (C-19500186).

5. REFERENCES

[1] B. S. Krongold and D. L. Jones, "An active-set approach for OFDM PAR reduction via tone reservation," *IEEE Trans. Signal Processing*, vol. 52, no. 2, pp. 495–509, Feb. 2004.

[2] S. H. Han and J. H Lee, "An overview of peak-to-average power ratio reduction techniques for multicarrier transmission," *IEEE Wireless Commun. Mag.*, vol. 12, no. 2, pp. 56–65, April 2005.

[3] A. T. Erdogan, "A low complexity multicarrier PAR reduction approach based on subgradient optimization," *Signal Processing*, vol. 86, no. 12, pp. 3890–3903, Dec. 2006.

[4] A. Aggarwal and T. H. Meng, "Minimizing the peak-to-average power ratio of OFDM signals using convex optimization," *IEEE Trans. Signal Processing*, vol. 54, no. 8, pp. 3099–3110, Aug. 2006.

[5] I. Yamada and N. Ogura, "Adaptive projected subgradient method for asymptotic minimization of sequence of nonnegative convex functions," *Numerical Functional Analysis and Optimization*, vol. 25, no. 7/8, pp. 593–617, 2004.

[6] D. Mandic and I. Yamada, "Machine learning and signal processing applications of fixed point theory," IEEE ICASSP 2007: Tutorial Textbook, April 2007.

[7] K. Slavakis, I. Yamada, and N. Ogura, "The adaptive projected subgradient method over the fixed point set of strongly attracting nonexpansive mappings," *Numerical functional analysis and optimization*, vol. 27, no. 7-8, pp. 905–930, Dec. 2006.

[8] *Wireless LAN Medium Access Control (MAC) and physical layer (PHY) specifications: High-speed physical layer in the 5 Ghz Band*, IEEE std., Sept. 1999.

[9] H. Stark and Y. Yang, *Vector Space Projections – A Numerical Approach to Signal and Image Processing, Neural Nets and Optics*, Wiley, New York, 1998.