Democratic Representations

Christoph Studer, Tom Goldstein, Wotao Yin, and Richard G. Baraniuk

Abstract

Minimization of the ℓ_∞ (or maximum) norm subject to a constraint that imposes consistency to an underdetermined system of linear equations finds use in a large number of practical applications, including vector quantization, peak-to-average power ratio (PAPR) (or "crest factor") reduction in communication systems, approximate nearest neighbor search, and peak force minimization in robotics and control. This paper analyzes the fundamental properties of signal representations obtained by solving such a convex optimization problem. We develop bounds on the maximum magnitude of such representations using the uncertainty principle (UP) introduced by Lyubarskii and Vershynin, *IEEE Trans. IT, 2010*, and study the efficacy of ℓ_∞-norm-based PAPR reduction. Our analysis shows that matrices satisfying the UP, such as randomly subsampled Fourier or i.i.d. Gaussian matrices, enable the computation of what we call *democratic representations*, whose entries all have small and similar magnitude, as well as low PAPR. To compute democratic representations at low computational complexity, we present two new, efficient convex optimization algorithms. We finally demonstrate the efficacy of democratic representations for PAPR reduction in a DVB-T2-based broadcast system.

Index Terms

Convex optimization, democratic representations, first-order optimization methods, frames, ℓ_∞-norm minimization, peak-to-average power ratio (PAPR) (or "crest-factor") reduction, uncertainty principle.

C. Studer was with the Department of Electrical and Computer Engineering, Rice University and is now with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY (e-mail: studer@cornell.edu).

T. Goldstein and R. G. Baraniuk are with the Department of Electrical and Computer Engineering, Rice University, Houston, TX (e-mail: {tag7, richb}@rice.edu). Website: http://dsp.rice.edu

W. Yin was with the Department of Computational an Applied Mathematics, Rice University and is now with the Department of Mathematics, University of California, Los Angeles, CA (e-mail: wotaoyin@math.ucla.edu).

I. INTRODUCTION

In this paper, we analyze the properties of the solutions $\dot{\mathbf{x}} \in \mathbb{C}^N$ to the following convex minimization problem:

$$(\mathrm{P}_\infty^\varepsilon) \quad \underset{\tilde{\mathbf{x}} \in \mathbb{C}^N}{\text{minimize}} \; \|\tilde{\mathbf{x}}\|_\infty \quad \text{subject to } \|\mathbf{y} - \mathbf{D}\tilde{\mathbf{x}}\|_2 \leq \varepsilon.$$

Here, the vector $\mathbf{y} \in \mathbb{C}^M$ denotes the *signal* to be represented, $\mathbf{D} \in \mathbb{C}^{M \times N}$ is an overcomplete matrix (often called frame or dictionary) with $M < N$, and the real-valued approximation parameter $\varepsilon \geq 0$ determines the accuracy of the *signal representation* $\dot{\mathbf{x}}$.

As demonstrated in [2], certain matrices $\mathbf{D}$ enable the computation of signal representations $\dot{\mathbf{x}}$ whose entries all have magnitudes of the order $1/\sqrt{N}$. Since for such representations each entry is of approximately the same importance, we call them *democratic*.[1]

Figure 1 shows an example of three different representations of the same signal $\mathbf{y}$ using the columns of a subsampled discrete cosine transform (DCT) matrix.[2] In contrast to the (popular) least-squares and sparse representation, most of the entries of the democratic representation obtained via $(\mathrm{P}_\infty^\varepsilon)$ have the same (low) maximum magnitude. As a consequence of this particular magnitude property, the problem $(\mathrm{P}_\infty^\varepsilon)$ and the resulting signal representations feature prominently in a variety of practical applications.

A. Application Examples

1) Vector quantization: Element-wise quantization of democratic representations affects all entries of $\dot{\mathbf{x}}$ equally, which renders them less susceptible to quantization noise compared with direct quantization of the signal vector $\mathbf{y}$ [2]. Moreover, the corruption of a few entries of $\dot{\mathbf{x}}$ results in only a small error and, therefore, computing $\mathbf{y} = \mathbf{D}\dot{\mathbf{x}}$ after, e.g., transmission over an unreliable communication channel [5] or storage in unreliable memory cells [6], enables one to obtain a robust estimate of the signal vector $\mathbf{y}$.

2) Peak-to-average power ratio (PAPR) reduction: Wireless communication systems employing orthogonal frequency division multiplexing (OFDM) typically require linear and power-inefficient radio-frequency (RF) components (e.g., power amplifiers) to avoid unwanted signal distortions or out-of-band

---

[1]Other names for democratic representations have been proposed in the literature. The paper [2] uses "Kashin representations," whereas [3] uses both, "spread representations" and "anti-sparse representations." We also note that [4] used the term "democracy" for quantized representations where the individual bits have "equal-weight" in the context of sigma-delta conversion. Here, the signal representations $\dot{\mathbf{x}}$ are, in general, neither binary-valued nor quantized.

[2]The entries of the vector $\mathbf{y}$ are generated from a zero-mean i.i.d. Gaussian distribution with unit variance; the row indices of the DCT matrix have been chosen uniformly at random. All representations are computed via problems of the form $(\mathrm{P}_\infty^\varepsilon)$ with the ℓ_∞, ℓ_2, and ℓ_1-norm for the democratic, least-squares, and sparse representation, respectively, and $\varepsilon = 0$.
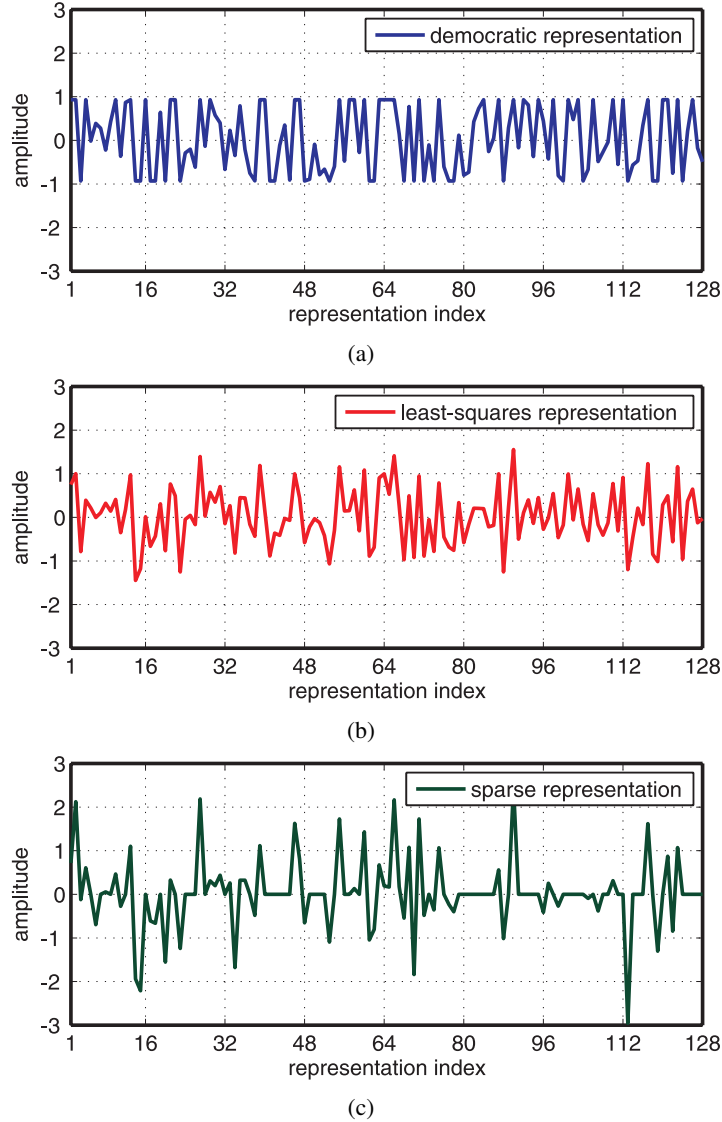
Fig. 1. There are, in general, infinitely many ways to represent a signal as a linear combination of columns from an overcomplete matrix. In this example, we compare the representations of a given signal vector that are optimal according to three different criteria, i.e., minimum ℓ_∞, ℓ_2, and ℓ_1-norm, using an overcomplete 64×128 subsampled DCT matrix; (a) democratic (ℓ_∞-norm) representation (11.1 dB PAPR); (b) least-squares (ℓ_2-norm) representation (16.2 dB PAPR); (c) sparse (ℓ_1-norm) representation (29.9 dB PAPR). We see that the democratic representation has (i) most of its entries' magnitudes at the same maximum level and (ii) very small dynamic range (also in terms of the PAPR).

radiation, as OFDM signals are prone to exhibit a large peak-to-average power ratio (PAPR) (also called the "crest factor") [7]; see (7) for the PAPR definition used in this paper. By allocating certain unused OFDM tones, which is known as tone reservation [8], or by exploiting the excess degrees-of-freedom in large-scale multi-antenna wireless systems (often called massive MIMO systems), one can transmit democratic representations, which significantly reduce the PAPR [9]. Hence, transmitting democratic

representations, instead of conventional (unprocessed) OFDM signals, substantially alleviates the need for expensive and power-inefficient RF components. The example in Fig. 1 confirms that the democratic representation exhibits substantially lower PAPR compared to a least-squares or sparse representation.

3) Approximate nearest neighbor search: Signal representations obtained from the problem $(\mathrm{P}_\infty^\varepsilon)$ also find use in the identification of approximate nearest neighbors in high-dimensional spaces [10]. The underlying idea is to compute a representation $\dot{\mathbf{x}}$ for the query vector $\mathbf{y}$. For certain matrices $\mathbf{D}$, the resulting representations are democratic and hence, resemble to an antipodal signal for which most coefficients take the values $-\alpha$ or $+\alpha$ for some $\alpha > 0$; see the democratic representation example in Fig. 1 where $\alpha \approx 1$. This property of the coefficients of $\dot{\mathbf{x}}$ can then be used to efficiently find approximate nearest vectors in an N-dimensional Hamming space.

4) Robotics and control: Kinematically redundant robots or manipulators admit infinitely many inverse solutions. Certain applications require a solution that minimizes the *maximum* force, acceleration, torque, or joint velocity, rather than minimizing the energy or power. Hence, in many practical situations, one is typically interested in solving problems of the form $(\mathrm{P}_\infty^\varepsilon)$ to obtain democratic representations with provably small ℓ_∞-norm, rather than minimum-energy (or least-squares) representations (the democratic representation in Fig. 1 has significantly smaller ℓ_∞-norm compared to the other two representations); corresponding practical application examples have been discussed in [11]–[13].

5) Recovery conditions for sparse signal recovery: As shown in [14], the ℓ_∞-norm of the representation obtained by solving a specific instance of $(\mathrm{P}_\infty^\varepsilon)$ with $\varepsilon = 0$ can be used to verify uniqueness and robustness conditions for ℓ_1-norm-based (analysis and synthesis) sparse signal recovery problems. Such recovery conditions are of particular interest in the emerging fields of sparse signal recovery [15]–[17] and compressive sensing (CS) [18]–[21].

B. What About Signal Recovery?

The problem $(\mathrm{P}_\infty^\varepsilon)$ with $\varepsilon = 0$ can be used to *recover* antipodal (or binary-valued) signals, i.e., vectors with coefficients belonging to the set $\{-\alpha, +\alpha\}$ for $\alpha > 0$ from the underdetermined system of linear equations $\mathbf{y} = \mathbf{Dx}$, provided that the matrix $\mathbf{D}$ meets certain conditions [22], [23]. The main focus here is, however, on (i) properties of democratic representations having minimal ℓ_∞-norm and small dynamic range, and (ii) their efficient computation, rather than on the recovery of a given antipodal vector $\mathbf{x}$ from the set of linear equations $\mathbf{y} = \mathbf{Dx}$ and corresponding uniqueness conditions. We refer the interested reader to [22], [23] for the details on antipodal signal recovery via ℓ_∞-norm minimization.

C. Relevant Prior Art on ℓ_∞-Norm Minimization

Results for minimizing the maximum amplitude of continuos, real-valued signals subject to linear constraints reach back to the 1960s, when Neustadt [24] studied the so-called *minimum-effort control problem*. In 1971, Cadzow proposed a corresponding practicable algorithm suitable for low-dimensional systems, where he proposed to solve the following real-valued, convex ℓ_∞-norm minimization problem [11]:

$$(\text{P}_\infty) \quad \underset{\tilde{\mathbf{x}} \in \mathbb{R}^N}{\text{minimize}} \, \|\tilde{\mathbf{x}}\|_\infty \quad \text{subject to } \mathbf{y} = \mathbf{D}\tilde{\mathbf{x}}.$$

Note that this problem coincides to a real-valued version of $(\text{P}_\infty^\varepsilon)$ with $\varepsilon = 0$. Specifically, in [11] it was shown that, for a large class of matrices $\mathbf{D}$, there exists a solution $\dot{\mathbf{x}}$ to (P_∞) for which a dominant portion of the entries' magnitudes correspond to $\|\dot{\mathbf{x}}\|_\infty$, whereas only a small fraction of the entries may have smaller magnitude; this result has been rediscovered recently by Fuchs in [3].

Another line of research that characterizes signal representations $\mathbf{x}$ with small (but not necessarily minimal) ℓ_∞-norm subject to $\mathbf{y} = \mathbf{D}\mathbf{x}$ have been established in 2010 by Lyubarskii and Vershynin [2]. In particular, [2] proves the existence of matrices $\mathbf{D}$ with arbitrarily small redundancy parameter $\lambda = N/M > 1$ for which every signal vector $\mathbf{y}$ has a democratic representation $\mathbf{x}$ satisfying

$$\|\mathbf{x}\|_\infty \leq \frac{K}{\sqrt{N}}\|\mathbf{y}\|_2. \tag{1}$$

Here, K is a (preferably small) constant that only depends on the redundancy parameter λ. The existence of such signal representations for certain sets of matrices can either be shown using fundamental results obtained by Kashin [25], Garnaev and Gluskin [26], or by analyzing the signal representations obtained via the iterative algorithm proposed in [2]. The latter (constructive) approach relies on an uncertainty principle (UP) for the matrix $\mathbf{D}$, which establishes a fundamental connection between the constant K in (1) and sensing matrices commonly used in sparse signal recovery and CS [18], [27].

D. Contributions

In this paper, we derive and investigate a host of fundamental properties for signal representations $\dot{\mathbf{x}}$ obtained from the ℓ_∞-norm minimization problem $(\text{P}_\infty^\varepsilon)$. In particular, we analyze its Lagrange dual problem to derive a refined and more general version of the bound on the ℓ_∞-norm of the signal representation $\dot{\mathbf{x}}$ established in [2]. We characterize magnitude properties of the signal representations $\dot{\mathbf{x}}$ obtained by solving $(\text{P}_\infty^\varepsilon)$, and we develop bounds on the resulting PAPR, which is of particular interest in OFDM-based communication systems. As a byproduct of our analysis, we present the Lagrange duals

to a variety of optimization problems, such as ℓ_1-norm minimization, which is often used for sparse signal recovery and compressive sensing. We then discuss classes of matrices that enable the computation of *democratic representations* via (P_∞^ε). Furthermore, we develop two computationally efficient algorithms to solve (P_∞^ε), referred to as CRAM (short for <u>c</u>onvex <u>r</u>eduction of <u>am</u>plitudes) and CRAMP (short for CRAM for Parseval frames). CRAM is suitable for arbitrary matrices $\mathbf{D}$ and approximation parameters $\varepsilon \in [0, \infty)$, whereas CRAMP exhibits lower complexity than CRAM for $\varepsilon = 0$ and Parseval frames. We provide numerical experiments to support our analysis and conclude by demonstrating the efficacy of democratic representations for PAPR reduction in a DVB-T2-based broadcast system [28].

E. Notation

Lowercase boldface letters stand for column vectors and uppercase boldface letters designate matrices. For a matrix $\mathbf{A}$, we denote its conjugate transpose and spectral norm by $\mathbf{A}^H$ and $\|\mathbf{A}\|_{2,2} = \sqrt{\lambda_{\max}(\mathbf{A}^H\mathbf{A})}$, respectively, where $\lambda_{\max}(\mathbf{A}^H\mathbf{A})$ denotes the maximum eigenvalue of $\mathbf{A}^H\mathbf{A}$. We use $\mathbf{0}_{M\times N}$ and $\mathbf{1}_{M\times N}$ to denote the all-zeros and all-ones matrix of dimension $M \times N$, respectively. The kth entry of a vector $\mathbf{a}$ is designated by $[\mathbf{a}]_k$, and $\Re\{\mathbf{a}\}$ and $\Im\{\mathbf{a}\}$ represent its real and imaginary part, respectively. We define the ℓ_p-norm of the vector $\mathbf{a} \in \mathbb{C}^N$ as follows:

$$\|\mathbf{a}\|_p = \begin{cases} \left(\sum_{k=1}^N |[\mathbf{a}]_k|^p\right)^{1/p} & \text{if } 1 \leq p < \infty \\ \max_{k \in \{1,...,N\}} |[\mathbf{a}]_k| & \text{if } p = \infty. \end{cases}$$

We also make use of the (non-standard) $\ell_{\widetilde{\infty}}$-norm [29] defined as $\|\mathbf{a}\|_{\widetilde{\infty}} = \max\{\|\Re\{\mathbf{a}\}\|_\infty, \|\Im\{\mathbf{a}\}\|_\infty\}$. The notation $\dot{\mathbf{x}}$ is used to refer to the solutions to the problem (P_∞^ε). Sets are designated by uppercase Greek letters; the cardinality of the set Ω is $|\Omega|$. The notation $\text{supp}(\mathbf{a})$ designates to the support set of the vector $\mathbf{a}$, i.e., the set of indices associated to non-zero entries in $\mathbf{a}$. The sign (or phase) of a complex-valued scalar $x \in \mathbb{C}$ is defined as

$$\text{sign}(x) = \begin{cases} x/|x| & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

We use $\text{sign}(\mathbf{a})$ and $\text{abs}(\mathbf{a})$ to denote the entry-wise application of the sign function and absolute value to the vector $\mathbf{a}$, respectively.

F. Organization of the Paper

The remainder of the paper is organized as follows. Section II introduces the essentials of frames and the uncertainty property (UP). In Section III, we develop the concept of democratic representations. Our main results are detailed in Section IV. Section V reviews suitable classes of matrices that satisfy the UP and enable the computation of democratic representations. Section VI develops computationally efficient algorithms for solving $(\text{P}_\infty^\varepsilon)$. Section VII provides numerical experiments and showcases the efficacy of democratic representations for PAPR reduction. We conclude in Section VIII. Most proofs are relegated to Appendices A through G.

II. FRAMES AND THE UNCERTAINTY PRINCIPLE

A. Frames

We often require the over-complete matrix $\mathbf{D} \in \mathbb{C}^{M \times N}$ with $M < N$ to satisfy the following property.

Definition 1 (Frame [30]): A matrix $\mathbf{D} \in \mathbb{C}^{M \times N}$ with $M \leq N$ is called a *frame* if

$$A\|\mathbf{w}\|_2^2 \leq \left\|\mathbf{D}^H\mathbf{w}\right\|_2^2 \leq B\|\mathbf{w}\|_2^2$$

holds for any vector $\mathbf{w} \in \mathbb{C}^M$ with $A \in \mathbb{R}$, $B \in \mathbb{R}$, and $0 < A \leq B < \infty$.

The tightest possible constants A and B are called the lower and upper *frame bounds*, respectively. The frame $\mathbf{D}$ is called a *tight frame* if $A = B$. Furthermore, if $A = B = 1$, then $\mathbf{D}$ is a *Parseval frame* [30]. In what follows, we exclusively study the finite-dimensional setting (i.e., where $M, N < \infty$) and thus $B < \infty$. Further, because our frame definition requires A to be strictly positive, $\mathbf{D}$ is guaranteed to be full rank. Thus, $(\text{P}_\infty^\varepsilon)$ is feasible for any frame $\mathbf{D}$ and $\varepsilon \geq 0$.

B. Full-Spark Frames

The next definition is concerned with the *spark* of a frame, which represents the cardinality of the smallest subset of linearly dependent frame columns.

Definition 2 (Full-spark frame [31]): A frame $\mathbf{D} \in \mathbb{C}^{M \times N}$ is called a *full-spark frame* if the columns of every $M \times M$ sub-matrix of $\mathbf{D}$ are linearly independent.

Full-spark frames have spark $M + 1$ and are ubiquitous in sparse signal recovery and CS (see [31] for a review). Even though verifying the full-spark property of an arbitrary matrix is, in general, a hard problem [32], many frames are known to be full spark. For example, any subset of rows from a Fourier matrix of prime dimension forms a full-spark frame [31], [33]. Further, Vandermonde matrices

with $M \leq N$ having distinct basis entries are known to be full spark frames [31], [34]. In addition, randomized constructions also exist that generate full-spark frames with high probability. In particular, if the entries of an $M \times N$ matrix with $M < N$ are generated from independent continuous random variables, then the resulting matrix is a full-spark frame with probability one (see [35] for a formal proof).

C. The Uncertainty Principle (UP)

Several of the results derived in this paper rely upon the uncertainty principle (UP) introduced in [2].

Definition 3 (Uncertainty principle [2]): We say that the frame $\mathbf{D} \in \mathbb{C}^{M \times N}$ satisfies the UP with parameters $\eta \in \mathbb{R}^+$ and $\delta \in (0,1)$ if

$$\|\mathbf{D}\mathbf{x}\|_2 \leq \eta \|\mathbf{x}\|_2 \tag{2}$$

holds for all (sparse) vectors $\mathbf{x} \in \mathbb{C}^N$ satisfying $|\mathrm{supp}(\mathbf{x})| \leq \delta N$.

We emphasize that (2) is trivially satisfied for $\eta = \|\mathbf{D}\|_{2,2}$ and for arbitrary vectors with $|\mathrm{supp}(\mathbf{x})| \leq N$. However, as in [2], we are particularly interested in frames satisfying the UP with parameters $\eta < \|\mathbf{D}\|_{2,2}$ and $\delta < 1$. For simplicity, we say that frames satisfying definition (3) with such non-trivial parameters "satisfy the UP."

Verifying the UP with parameters $\eta < \|\mathbf{D}\|_{2,2}$ and $\delta < 1$ for a given frame $\mathbf{D}$ requires, in general, a combinatorial search over all δN-sparse vectors [32]. Nevertheless, many classes of frames are known to satisfy the UP with high probability (see [2] and Section V for more details). We finally note that frames satisfying the UP are strongly related to sensing matrices with small restricted isometry constants; such matrices play a central role in CS [18], [21], [27].

III. DEMOCRATIC REPRESENTATIONS

We next introduce the concept of *democratic representations* and define the *democracy constants*.

A. Democratic Representations

For $M < N$ and $\varepsilon < \|\mathbf{y}\|_2$, there exist, in general, an infinite number of representations $\mathbf{x}$ for a given signal vector $\mathbf{y}$ that satisfy $\|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 \leq \varepsilon$.[3] In the remainder of the paper, we are particularly interested in representations for which every entry x_i, $i = 1, \ldots, N$ is of approximately the same importance. In particular, we seek so-called *democratic representations*, which have provably small ℓ_∞-norm and for

---

[3]Note in the case $\varepsilon \geq \|\mathbf{y}\|_2$, the problem $(\mathrm{P}_\infty^\varepsilon)$ returns the all-zeros vector and, hence, practically relevant choices of ε are in the range $0 \leq \varepsilon < \|\mathbf{y}\|_2$.

which all magnitudes are approximately equal. In order to make the concept of democratic representations more formal, we use the following definition.

Definition 4 (Democracy constants): Let $\mathbf{D} \in \mathbb{C}^{M \times N}$ be a given frame. Assume we obtain a signal representation $\dot{\mathbf{x}}$ for every vector $\mathbf{y} \in \mathbb{C}^M$ by solving $(\mathrm{P}_\infty^\varepsilon)$ with $\varepsilon \leq \|\mathbf{y}\|_2$. We define the *lower* and *upper democracy constants* $K_{\mathrm{l}} \in \mathbb{R}$ and $K_{\mathrm{u}} \in \mathbb{R}$ to be the largest and smallest constants for which

$$\frac{K_{\mathrm{l}}}{\sqrt{N}}\big(\|\mathbf{y}\|_2 - \varepsilon\big) \leq \|\dot{\mathbf{x}}\|_\infty \leq \frac{K_{\mathrm{u}}}{\sqrt{N}}\big(\|\mathbf{y}\|_2 - \varepsilon\big) \tag{3}$$

holds for every pair $\dot{\mathbf{x}}$ and $\mathbf{y}$.

We note that the democracy constants K_{l} and K_{u} depend *only* on properties of the frame $\mathbf{D}$ and the fact that all signal representations $\dot{\mathbf{x}}$ are obtained via $(\mathrm{P}_\infty^\varepsilon)$, and *not* on the signal vector $\mathbf{y}$. Note that Definition 4 enables us to analyze a generalized and refined setting of the special case (1) studied in [2] (see our results in Section IV).

In what follows, we are interested in (i) classes of frames for which the lower and upper democracy constants K_{l}, K_{u} are both close to 1, and (ii) computationally efficient algorithms that provably deliver such signal representations. In particular, if $K_{\mathrm{l}} \approx 1 \approx K_{\mathrm{u}}$, then all signal representations $\dot{\mathbf{x}}$ have similar ℓ_∞-norm and every entry will have a maximum magnitude of $1/\sqrt{N}$ (assuming $\varepsilon = 0$ and $\|\mathbf{y}\|_2 = 1$). Since this property *evenly* spreads the signal vector's energy $\|\mathbf{y}\|_2$ across all entries of $\dot{\mathbf{x}}$, we call such representations *democratic*.

Definition 5 (Democratic representations): Let $\mathbf{D} \in \mathbb{C}^{M \times N}$ be a given frame. If the associated democracy constants K_{l} and K_{u} are both close to 1, then the signal representations $\dot{\mathbf{x}}$ obtained by $(\mathrm{P}_\infty^\varepsilon)$ are called *democratic representations*.

B. Computing Democratic Representations

In order to compute representations $\mathbf{x}$ having small (but not necessarily minimal) ℓ_∞-norm subject to the set of linear equations $\mathbf{y} = \mathbf{D}\mathbf{x}$, one can use the iterative algorithm proposed in [2]. This method efficiently computes such representations for real-valued and approximate Parseval frames, i.e., frames $\mathbf{D} \in \mathbb{R}^{M \times N}$ satisfying the UP in [2] with Frame bounds $A = 1 - \xi$ and $B = 1 + \xi$ for some small $\xi \geq 0$. However, the algorithm in [2] (i) does not solve $(\mathrm{P}_\infty^\varepsilon)$ and is, in general, not guaranteed to find representations $\dot{\mathbf{x}}$ having the *smallest* ℓ_∞-norm, (ii) requires knowledge of the UP parameters η, δ, (iii) was introduced for real-valued systems,[4] and (iv) is only guaranteed to converge for approximate Parseval

---

[4]A corresponding generalization of the algorithm in [2] to the complex-valued case is straightforward.

frames. Moreover, if one is interested in *approximate* representations $\mathbf{x}$ for which $\|\mathbf{y} - \mathbf{Dx}\|_2 > 0$ rather than in *perfect* representations satisfying $\mathbf{y} = \mathbf{Dx}$, the algorithm in [2] must be modified accordingly. In order to overcome the limitations of the algorithm in [2], we propose to directly solve the convex problem $(\mathrm{P}_\infty^\varepsilon)$ instead; Section VI will detail two corresponding (and computationally efficient) algorithms.

IV. MAIN RESULTS

We now analyze several key properties of signal representations $\dot{\mathbf{x}}$ obtained from solving $(\mathrm{P}_\infty^\varepsilon)$. Section IV-A studies magnitude properties of the solutions to $(\mathrm{P}_\infty^\varepsilon)$. Section IV-B introduces the Lagrange dual problem to $(\mathrm{P}_\infty^\varepsilon)$, which is key in the proofs of Sections IV-C and IV-D, where we develop bounds on the lower and upper democracy constants K_l and K_u, respectively. Section IV-E analyzes the PAPR characteristics of democratic representations, and Section IV-F outlines an extension of our results to the $\ell_{\widetilde{\infty}}$-norm.

A. Extreme Values of Solutions to $(\mathrm{P}_\infty^\varepsilon)$

The magnitudes of signal representations obtained via $(\mathrm{P}_\infty^\varepsilon)$ exhibit specific and practically relevant properties. To study them, we need the following definition.

Definition 6 (Extreme and moderate entries): Given a vector $\mathbf{x} \in \mathbb{C}^N$, we call an entry x_i *extreme* if $|x_i| = \|\mathbf{x}\|_\infty$; we further call an entry x_i *moderate* if $|x_i| < \|\mathbf{x}\|_\infty$.

Without any specific assumptions on the overcomplete matrix $\mathbf{D}$ (apart from being full-rank), we next show that there always *exists* a solution to $(\mathrm{P}_\infty^\varepsilon)$ with a large portion of extreme entries (see Appendix A for the proof). In words, a democratic representation is one with a large portion of extreme entries.

Lemma 1 (Democratic representations exist): For any full-rank matrix $\mathbf{D} \in \mathbb{C}^{M \times N}$ with $M \leq N$, the problem $(\mathrm{P}_\infty^\varepsilon)$ admits a solution $\dot{\mathbf{x}}$ with at least $N - M + 1$ extreme entries.

This result implies that there exist signal representations $\dot{\mathbf{x}}$ for which a large number of (extreme) entries have equal magnitude. In particular, by increasing the redundancy $\lambda = N/M$ of $\mathbf{D}$, the problem $(\mathrm{P}_\infty^\varepsilon)$ admits representations for which the number of extreme values is arbitrarily close to N.

The next result shows that—given the matrix $\mathbf{D}$ is a full-spark frame—*every* solution to the problem $(\mathrm{P}_\infty^\varepsilon)$ has a bounded minimal number of extreme entries (see Appendix B for the proof).

Lemma 2 (All representations are democratic): If the frame $\mathbf{D}$ has full spark, then every solution to $(\mathrm{P}_\infty^\varepsilon)$ has at least $N - M + 1$ extreme entries.

This result implies that for full-spark frames with large redundancy $\lambda = M/N$, every solution to $(\mathrm{P}_\infty^\varepsilon)$ is a democratic representation (with most entries being extreme). As noted in Section II-B, a large

number of deterministic and random constructions of full-spark frames are known. Hence, solving $(\mathrm{P}_\infty^\varepsilon)$ allows the computation of democratic representations for a large number of frames. In addition, Lemma 2 enables us to obtain the following ℓ_∞/ℓ_2-norm inequality.

Theorem 3 (Democratic ℓ_∞/ℓ_2-norm inequality): If $\mathbf{D}$ is a full-spark frame, then

$$\sqrt{N - M + 1}\|\dot{\mathbf{x}}\|_\infty \leq \|\dot{\mathbf{x}}\|_2 \tag{4}$$

holds for every representation $\dot{\mathbf{x}}$ obtained by solving $(\mathrm{P}_\infty^\varepsilon)$.

Proof: The proof immediately follows from Lemma 2 and the straightforward inequality $\|\dot{\mathbf{x}}\|_2^2 = \sum_{i=1}^N |\dot{x}_i|^2 \geq (N - M + 1)\|\dot{\mathbf{x}}\|_\infty^2$. ∎

We note that the norm inequality (4) is stronger than the standard norm bound $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2$ for $M < N$ (which holds for arbitrary vectors $\mathbf{x}$). More importantly, as we show below in Section IV-E, the refined norm inequality (4) is particularly useful for characterizing the limits of PAPR reduction methods that rely on democratic representations.

We conclude by noting that results related to Lemmata 1 and 2 for the special problem $(\mathrm{P}_\infty^\varepsilon)$ with $\varepsilon = 0$ have been developed in the literature [3], [11]. In particular, [11] establishes bounds on the minimum number of entries that satisfy $x_i \leq \|\mathbf{x}\|_\infty$ (i.e., entries that are not necessarily extremal). This result, however, does not allow us to extract bounds on the number of extremal values, which is in contrast to Lemmata 1 and 2. Reference [3] mentions that signal representations obtained from (P_∞) have, in general, exactly $N - M + 1$ extreme entries. This result, however, is stated without proof and, more importantly, without explicitly specifying conditions on the classes of the matrices for which it is supposed to hold.

B. Lagrange Dual Problem

In order to derive bounds on the lower and upper democracy constants K_l and K_u, respectively, and to study the PAPR behavior of solutions to $(\mathrm{P}_\infty^\varepsilon)$, we make use of the following theorem (see Appendix C for the proof).

Theorem 4 (Lagrange dual problem): Let the ℓ_p-norm primal problem (with $1 \leq p \leq \infty$) be

$$(\mathrm{P}_p^\varepsilon) \quad \underset{\tilde{\mathbf{x}} \in \mathbb{C}^N}{\mathrm{minimize}} \ \|\tilde{\mathbf{x}}\|_p \quad \text{subject to} \ \|\mathbf{y} - \mathbf{D}\tilde{\mathbf{x}}\|_2 \leq \varepsilon.$$

Then, the corresponding Lagrange dual problem is given by

$$(\mathrm{D}_p^\varepsilon) \quad \begin{cases} \underset{\tilde{\mathbf{z}} \in \mathbb{C}^M}{\mathrm{maximize}} & \Re\big(\mathbf{y}^H \tilde{\mathbf{z}}\big) - \varepsilon\|\tilde{\mathbf{z}}\|_2 \\ \text{subject to} & \|\mathbf{D}^H \tilde{\mathbf{z}}\|_d \leq 1 \end{cases}$$

with $1/p + 1/d = 1$; for $p = 1$ we have $q = \infty$ and vice versa. The norm $\| \cdot \|_d$ corresponds to the dual norm of $\| \cdot \|_p$.

Note that Theorem 4 includes not only the Lagrange dual to the problem $(\text{P}_\infty^\varepsilon)$, but also other frequently studied optimization problems, such as the Lagrangian dual to (P_1^ε), which is often used for sparse signal recovery or CS.

C. Bound on Lower Democracy Constant K_l

In order to characterize the lower democracy constant K_l for a given frame $\mathbf{D}$, we next derive a corresponding lower bound.

Lemma 5 (Lower democracy bound): Let $\mathbf{D} \in \mathbb{C}^{M \times N}$ be a frame with upper frame bound B. Then, every vector $\mathbf{y} \in \mathbb{C}^M$ admits a signal representation $\dot{\mathbf{x}}$ with the following lower bound $\widetilde{K}_\mathrm{l}$ on the lower democracy constant K_l (see Appendix D for the proof):

$$\widetilde{K}_\mathrm{l} = \frac{1}{\sqrt{B}} \leq K_\mathrm{l}. \tag{5}$$

The representations $\dot{\mathbf{x}}$ obtained from the problem $(\text{P}_\infty^\varepsilon)$ are guaranteed to satisfy (5), as the proof of Lemma 5 exploits properties of its solution. In the special case $\varepsilon = 0$ and for Parseval frames, Lemma 5 guarantees that all vectors $\mathbf{y}$ admit a signal representation satisfying

$$\frac{\|\mathbf{y}\|_2}{\sqrt{N}} \leq \|\dot{\mathbf{x}}\|_\infty \,.$$

This lower bound was established previously in [2, Obs. 2.1b]. Lemma 5 generalizes this result to arbitrary frames and to representations for which $\varepsilon > 0$. It is furthermore interesting to observe that the lower democracy bound in (5) only depends on the upper frame constant B (and implicitly on the fact that frames satisfy $A > 0$); this is in contrast to the bound on the upper democracy constant K_u derived next.

D. Bound on Upper Democracy Constant K_u

In order to characterize the upper democracy constant K_u, we next derive an upper bound $\widetilde{K}_\mathrm{u}$ on K_u by using the uncertainty principle (UP) for frames (see Appendix E for the proof).

Theorem 6 (Upper democracy bound): Let $\mathbf{D} \in \mathbb{C}^{M \times N}$ be a frame with frame bounds A, B that satisfies the uncertainty principle (UP) with parameters η, δ. Then, every signal vector $\mathbf{y}$ admits a signal representation $\dot{\mathbf{x}}$ with the following upper bound on the upper democracy constant:

$$K_\mathrm{u} \leq \widetilde{K}_\mathrm{u} = \frac{\eta}{(A - \eta\sqrt{B})\sqrt{\delta}}, \tag{6}$$

provided $A > \eta\sqrt{B}$.

The representations $\dot{x}$ obtained from the problem (P_∞^ε) are guaranteed to satisfy (6), as the proof exploits properties of its solution. In addition, Theorem 6 shows that if a frame $\mathbf{D}$ satisfies (i) $A > \eta\sqrt{B}$ and (ii) $\delta > 0$, then one can compute democratic representations for every signal vector $\mathbf{y}$ by solving (P_∞^ε). In addition, the condition $A > \eta\sqrt{B}$ indicates that the use of Parseval frames is beneficial in practice, i.e., leads to democratic representations with smaller ℓ_∞-norm—an observation that was made empirically by Fuchs [36]; corresponding simulation results are provided in Section VII. In order to achieve representations having provably small ℓ_∞-norm (close to 1), one is typically interested in finding frames satisfying the UP with small η and large δ. Both properties can be achieved simultaneously for certain classes of frames (see [2] and Section V for corresponding examples).

We note that Theorem 6 improves upon the results in [2]. In particular, the bound $\widetilde{K}_u$ in (6) is strictly smaller than the bound obtained in [2, Thms. 3.5 and 3.9]. To see this, consider the case of $\mathbf{D}$ being a Parseval frame and $\varepsilon = 0$; this enables us to establish the following relation between the upper democracy bound $\widetilde{K}_u$ in (6) and the bound K from [2, Thm. 3.5]:

$$\widetilde{K}_u = \frac{\eta}{(1-\eta)\sqrt{\delta}} < \frac{1}{(1-\eta)\sqrt{\delta}} = K.$$

The strict inequality follows from the fact that η is required to be smaller than one, which is a consequence of $A > \eta\sqrt{B}$. Hence, by solving (P_∞^ε) rather than using the algorithm proposed in [2], we arrive at an upper bound that is more tight (i.e., by a factor of η). For approximate Parseval frames satisfying $A = 1-\xi$ and $B = 1+\xi$ with $0 \le \xi < 1$, the upper democracy bound in (6) continues to be superior to that in [2, Thm. 3.9]. Furthermore, Theorem 6 also encompasses approximate representations ($\varepsilon > 0$) and the case of complex-valued vectors and frames, which is in contrast to the results developed in [2].

E. PAPR Properties of Democratic Representations

1) PAPR reduction via democratic representations: Democratic representations can be used to (often substantially) reduce a signal's dynamic range, which is typically characterized in terms of its PAPR (or "crest factor") defined below. For example, the transmission of information-bearing signals over frequency-selective channels typically requires sophisticated equalization schemes at the receive side. Orthogonal frequency-division multiplexing (OFDM) [7] is a well-established way of reducing the computational complexity of equalization (compared to conventional equalization schemes). Unfortunately, OFDM signals are known to suffer from a high PAPR, which requires linear RF components (e.g., power amplifiers). Since linear RF components are, in general, more costly and less power efficient compared

to their non-linear counterparts, practical transceiver implementations often deploy sophisticated PAPR-reduction schemes [37]. Prominent approaches for reducing the PAPR exploit either certain reserved OFDM tones [8] or the excess degrees-of-freedom in large-scale multi-antenna wireless systems [9]. As we will show next, democratic representations computed via (P_∞^ε) have intrinsically low PAPR.

We start by defining the PAPR of arbitrary vectors $\mathbf{x}$.

Definition 7 (Peak-to-average power ratio): Let $\mathbf{x} \in \mathbb{C}^N$ be a nonzero vector. Then, the peak-to-average power ratio (PAPR) (or "crest factor") of $\mathbf{x}$ is defined as

$$PAPR(\mathbf{x}) = \frac{N\|\mathbf{x}\|_\infty^2}{\|\mathbf{x}\|_2^2}. \tag{7}$$

Note that for arbitrary vectors $\mathbf{x} \in \mathbb{C}^N$, the PAPR satisfies the following inequalities:

$$1 \le PAPR(\mathbf{x}) \le N, \tag{8}$$

an immediate consequence of standard norm bounds. The lower bound (best case) is achieved for signals having constant amplitude (or modulus), whereas the upper bound (worst case) is achieved by vectors having a single nonzero entry. As we will show next, the worst-case PAPR of signal representations obtained through (P_∞^ε) is typically much smaller than the upper bound in (8) suggests. To show this, we next bound the PAPR of signal representations obtained through (P_∞^ε) with the aid of (i) the democratic ℓ_∞/ℓ_2-norm inequality in (4) or (ii) the upper democracy bound in (6).

The following PAPR bound only depends on the dimensions of the full-spark frame $\mathbf{D}$.

Theorem 7 (Full-spark PAPR bound): Let $\mathbf{D} \in \mathbb{C}^{M \times N}$ be a full-spark frame. Then, the PAPR of every signal representation $\dot{\mathbf{x}}$ obtained by solving (P_∞^ε) satisfies

$$PAPR(\dot{\mathbf{x}}) \le \frac{N}{N - M + 1} \tag{9}$$

for vectors $\mathbf{y} \in \mathbb{C}^M$ satisfying $\|\mathbf{y}\|_2 \ne 0$ and $\varepsilon < \|\mathbf{y}\|_2$.

Proof: The proof directly follows from Lemma 2 and the PAPR definition (7) by replacing $\|\dot{\mathbf{x}}\|_2^2$ by $(N - M + 1)\|\dot{\mathbf{x}}\|_\infty^2$. ∎

Theorem 7 implies that ℓ_∞-norm minimization can be used as a practical and efficient substitute for minimizing the PAPR in (7) *directly*, which is difficult to achieve in practice. In addition, we observe that frames with large redundancy parameter $\lambda = N/M$ enable the computation of representations with arbitrary low PAPR (by increasing the dimension N). To see this, let $\lambda \to \infty$, which results in the following asymptotic bound: $PAPR(\dot{\mathbf{x}}) \le (1 + 1/N)^{-1}$.

The following theorem provides a PAPR bound for frames that satisfy the UP with parameters η, δ (see Appendix F for the proof).

Theorem 8 (UP-based PAPR bound): Let $\mathbf{D} \in \mathbb{C}^{M \times N}$ be a frame with the upper democracy bound $\widetilde{K}_u$ in (6). Then, the PAPR of every signal representation $\dot{\mathbf{x}}$ obtained via $(\mathrm{P}_{\infty}^{\varepsilon})$ satisfies the following bound:

$$PAPR(\dot{\mathbf{x}}) \leq \widetilde{K}_{\mathrm{u}}^2 B \qquad (10)$$

for vectors $\mathbf{y} \in \mathbb{C}^M$ with $\|\mathbf{y}\|_2 \neq 0$ and $\varepsilon < \|\mathbf{y}\|_2$.

Theorem 8 reveals that frames satisfying the UP and having a small upper democracy bound $\widetilde{K}_{\mathrm{u}}$ are particularly effective in terms of reducing the PAPR. A practically relevant example of frames satisfying these properties are randomly subsampled discrete Fourier transform (DFT) matrices, which naturally appear in OFDM-based tone-reservation schemes for PAPR reduction (see, e.g., [8] for the details). A corresponding application example is shown below in Section VII-C.

It is worth mentioning that the PAPR bounds (9) and (10) do *not* depend on the approximation parameter ε. Hence, in practice, an increase in ε (as long as $\varepsilon < \|\mathbf{y}\|_2$) is expected[5] to not affect the PAPR of the democratic representation $\dot{\mathbf{x}}$, which is in contrast to (3).

2) Transmit power increase: If democratic representations are used for PAPR reduction, e.g., in an OFDM-based communication system, then it is important to realize that transmitting $\dot{\mathbf{x}}$ instead of the minimum-power (or least squares) solution $\hat{\mathbf{x}}$ obtained from

$$(\mathrm{P}_2^{\varepsilon}) \quad \underset{\tilde{\mathbf{x}}}{\text{minimize}} \ \|\tilde{\mathbf{x}}\|_2 \quad \text{subject to} \ \|\mathbf{y} - \mathbf{D}\tilde{\mathbf{x}}\|_2 \leq \varepsilon,$$

may result in a larger transmit power. Therefore, it is of practical interest to study the associated power increase (PI), defined as

$$PI = \frac{\|\dot{\mathbf{x}}\|_2^2}{\|\hat{\mathbf{x}}\|_2^2}, \qquad (11)$$

when transmitting democratic representations $\dot{\mathbf{x}}$ instead of the least-squares representation $\hat{\mathbf{x}}$ of $(\mathrm{P}_2^{\varepsilon})$. The following result provides an upper bound on the PI (see Appendix G for the proof).

Theorem 9 (Power increase): Let $\mathbf{D} \in \mathbb{C}^{M \times N}$ be a frame with the upper frame bound B and upper democracy constant K_u. Then, the power increase, *PI*, in (11) of every signal representation $\dot{\mathbf{x}}$ obtained

---

[5]We note that our own experiments show virtually no impact of the approximation parameter ε to the PAPR, which confirms this behavior empirically.

by solving $(\mathrm{P}_\infty^\varepsilon)$ satisfies

$$1 \leq PI \leq \widetilde{K}_\mathrm{u}^2 B \tag{12}$$

for vectors $\mathbf{y} \in \mathbb{C}^M$ satisfying $\|\mathbf{y}\|_2 \neq 0$ and $\varepsilon < \|\mathbf{y}\|_2$.

It is interesting to observe that the RHS of the bound (12) in Theorem 9 coincides with the RHS of the PAPR bound (10). As a consequence, the use of frames that yield good PAPR reduction properties also guarantee a small power increase compared to directly transmitting a least-squares representation.

F. The $\ell_{\widetilde{\infty}}$-Norm and Its Implications

In certain applications, one might be interested in minimizing the $\ell_{\widetilde{\infty}}$-norm rather than the ℓ_∞-norm of the signal representation. Such representations can be useful if the PAPR of both the real and imaginary parts need to be minimized individually (see, e.g., [9]). In order to derive properties of signal representations obtained by solving

$$(\mathrm{P}_{\widetilde{\infty}}^\varepsilon) \quad \underset{\tilde{\mathbf{x}} \in \mathbb{C}^N}{\text{minimize}} \; \|\tilde{\mathbf{x}}\|_{\widetilde{\infty}} \quad \text{subject to } \|\mathbf{y} - \mathbf{D}\tilde{\mathbf{x}}\|_2 \leq \varepsilon,$$

we can use the following inequalities developed in [29, Eq. 78]:

$$\frac{1}{\sqrt{2}}\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_{\widetilde{\infty}} \leq \|\mathbf{x}\|_\infty. \tag{13}$$

These inequalities imply that all properties derived from the original problem $(\mathrm{P}_\infty^\varepsilon)$ hold as well for democratic representations obtained by $(\mathrm{P}_{\widetilde{\infty}}^\varepsilon)$ up to a factor of at most two.

V. Frames that Enable Democratic Representations

As shown in [2], random orthogonal matrices, random partial DFT matrices, and random sub-Gaussian matrices satisfy the UP in Definition 3 with high probability. Hence, matrices drawn from such classes are particularly suitable for the computation of democratic representations with small ℓ_∞-norm and for applications requiring low PAPR. As an example, we briefly restate a result obtained in [2] for matrices whose entries are chosen i.i.d. sub-Gaussian.

Definition 8 (Sub-Gaussian RV [2, Def. 4.5]): A random variable X is called sub-Gaussian with parameter β if

$$\Pr\{|X| > u\} \leq \exp\!\left(1 - u^2/\beta^2\right) \quad \text{for all} \quad u > 0.$$

For matrices having i.i.d. sub-Gaussian entries, the following result has been established in [2].

Theorem 10 ([2, Thm. 4.6]: UP for sub-Gaussian Matrices): Let $\mathbf{A}$ be a $M \times N$ matrix whose entries are i.i.d. zero-mean sub-Gaussian RVs with parameter β. Assume that $\lambda = N/M$ for some $\lambda \geq 2$. Then, with probability at least $1 - \lambda^{-M}$, the random matrix $\mathbf{D} = \frac{1}{\sqrt{N}}\mathbf{A}$ satisfies the UP with parameters

$$\eta = C_0 \beta \sqrt{\frac{\log(\lambda)}{\lambda}} \quad \text{and} \quad \delta = \frac{C_1}{\lambda},$$

where C_0, $C_1 > 0$ are absolute constants.

Theorem 10 implies that, for random sub-Gaussian matrices, the UP with parameters η and δ is satisfied with high probability. Moreover, the UP parameters η, δ only depend on the redundancy $\lambda = N/M$ of $\mathbf{D}$. Since $\mathbf{D} = \frac{1}{\sqrt{N}}\mathbf{A}$ is not, in general, a tight frame, it was furthermore shown in [2, Cor. 4.9] that $\mathbf{D}$ is a so-called approximate Parseval frame with high probability, i.e., $\mathbf{D}$ satisfies the frame bounds $A = 1 - \xi$ and $B = 1 + \xi$ for some small $\xi > 0$. Hence, random sub-Gaussian matrices can be used to efficiently compute democratic representations with democracy bounds K_l and K_u in (5) and (6) by solving ($\mathrm{P}_\infty^\varepsilon$).

Reference [2] established results similar to that of Theorem 10 for random orthogonal and random partial DFT matrices. Partial (or randomly sub-sampled) DFT matrices have two key advantages (over sub-Gaussian matrices): (i) they are Parseval frames, which typically yield better democracy bounds (see (6) and Section VII for numerical experiments), and (ii) the product of a vector with the matrix $\mathbf{D}$ or its Hermitian transpose $\mathbf{D}^H$ can be computed at low computational complexity, i.e., with roughly $N \log_2(N)$ operations using fast Fourier transforms. The latter property is of significant practical relevance as it enables one to compute democratic representations with low computational complexity; the next section details new algorithms for solving ($\mathrm{P}_\infty^\varepsilon$) that are able to exploit such fast transforms.

VI. EFFICIENT ALGORITHMS FOR SOLVING ($\mathrm{P}_\infty^\varepsilon$)

In order to compute the solution to ($\mathrm{P}_\infty^\varepsilon$), general-purpose solvers for convex optimization problems can be used (see, e.g., [38], [39]). For large-dimensional problems, however, more efficient methods become necessary. A Lagrange formulation of ($\mathrm{P}_\infty^\varepsilon$) that leads to a computationally more efficient method, called the fast iterative truncation algorithm (FITRA), was proposed in [9]. However, to solve ($\mathrm{P}_\infty^\varepsilon$) exactly, new algorithms are required.

Efficient methods for solving ($\mathrm{P}_\infty^\varepsilon$) should be capable of exploiting fast transforms for computing $\mathbf{D}\mathbf{p}$ and $\mathbf{D}^H\mathbf{q}$ (for two vectors $\mathbf{p}$ and $\mathbf{q}$ of appropriate dimension). Hence, we next propose two new algorithms that directly solve ($\mathrm{P}_\infty^\varepsilon$) and are able to exploit fast transforms. The first method, referred to as CRAM (short for <u>c</u>onvex <u>r</u>eduction of <u>am</u>plitudes) directly solves ($\mathrm{P}_\infty^\varepsilon$) at low computational complexity.

Algorithm 1 CRAM: Convex Reduction of Amplitudes

1: **inputs**: $\mathbf{x}_0 \in \mathbb{C}^N, \mathbf{y}, \mathbf{v}_0, \boldsymbol{\lambda}_0 \in \mathbb{C}^M, \tau, \sigma \in (0, \infty), k = 0$
2: **while** not converged **do**
3:     $\mathbf{x}_{k+1} \leftarrow \text{PROXINF}(\mathbf{x}_k - \tau \mathbf{D}^T \boldsymbol{\lambda}_k, \tau)$
4:     $\mathbf{v}_{k+1} \leftarrow \varepsilon(\mathbf{v}_k - \tau \boldsymbol{\lambda}_k)./\max\{\text{abs}(\mathbf{v}_k - \tau \boldsymbol{\lambda}_k), \varepsilon\}$
5:     $\boldsymbol{\lambda}_{k+1} \leftarrow \boldsymbol{\lambda}_k + \sigma(\mathbf{D}\mathbf{x}_{k+1} - \mathbf{v}_{k+1} - \mathbf{y})$
6:     $k \leftarrow k + 1$
7: **end while**

The second method, referred to as CRAMP (short for CRAM for Parseval frames) is particularly suited for Parseval frames and for $\varepsilon = 0$, which results in even lower computational complexity than CRAM.

A. CRAM: Convex Reduction of Amplitudes

To solve $(\text{P}_\infty^\varepsilon)$, we use the adaptive primal-dual hybrid gradient (PDHG) scheme proposed in [40]. To this end, we rewrite $(\text{P}_\infty^\varepsilon)$ as the following constrained convex program:

$$(\text{P}_\infty^\varepsilon) \quad \begin{cases} \displaystyle\operatorname*{minimize}_{\tilde{\mathbf{x}} \in \mathbb{C}^N, \mathbf{v} \in \mathbb{C}^M} & \|\tilde{\mathbf{x}}\|_\infty \\[2mm] \text{subject to} & \mathbf{v} = \mathbf{y} - \mathbf{D}\tilde{\mathbf{x}}, \; \|\mathbf{v}\|_2 \leq \varepsilon. \end{cases}$$

The constraint $\|\mathbf{v}\|_2 \leq \varepsilon$ can be removed by introducing the characteristic function $\chi_\varepsilon(\mathbf{v})$, which is zero when $\|\mathbf{v}\|_2 \leq \varepsilon$ and infinity otherwise. Additionally, we enforce the linear constraint $\mathbf{v} = \mathbf{y} - \mathbf{D}\tilde{\mathbf{x}}$ using the Lagrange multiplier vector $\boldsymbol{\lambda} \in \mathbb{C}^M$, which yields the (equivalent) saddle-point formulation

$$\max_{\boldsymbol{\lambda} \in \mathbb{C}^M} \min_{\tilde{\mathbf{x}} \in \mathbb{C}^N, \mathbf{v} \in \mathbb{C}^M} \|\tilde{\mathbf{x}}\|_\infty + \langle \mathbf{D}\tilde{\mathbf{x}} - \mathbf{v} - \mathbf{y}, \boldsymbol{\lambda} \rangle + \chi_\epsilon(\mathbf{v}),$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product. We emphasize that a saddle point of this problem formulation corresponds to a minimizer of $(\text{P}_\infty^\varepsilon)$.

We compute the saddle point of this formulation using the PDHG scheme detailed in Algorithm 1. Note that the operators $\max\{\cdot, \cdot\}$ and $\text{abs}(\cdot)$, as well as the division operation $./$ on line 4 of Algorithm 1 operate element-wise on vector entries. Algorithm 1 converges for constant step-sizes $\tau, \sigma \in (0, \infty)$ satisfying $\tau\sigma(\|\mathbf{D}\|_{2,2}^2 + 1) < 1$ (see [40] for the details). To achieve fast convergence of CRAM in Algorithm 1, we adaptively select the step-size parameters τ, σ using the recently developed method proposed in [40]. We conclude by noting that CRAM is advantageous over other splitting methods, such as ADMM [41], which require the solution of computationally complex minimization sub-steps, such as the solution of (possibly) high-dimensional least-squares problems.

Algorithm 2 Proximal Operator for the ℓ_∞-Norm

1: **procedure** PROXINF($\mathbf{z}, \tau$)
2:     **inputs**: $\mathbf{z} \in \mathbb{C}^N$, $\tau \in (0, \infty)$
3:     $\mathbf{a} \leftarrow \mathrm{abs}(\mathbf{z})$
4:     $\mathbf{s} \leftarrow \mathrm{sort}(\mathbf{a}, \text{'descending'})$
5:     **for** $k = 1, \ldots, N$ **do**
6:        $c_k \leftarrow \frac{1}{k} \sum_{i=1}^{k} (s_i - \tau)$
7:     **end for**
8:     $\alpha \leftarrow \max \big\{ 0, \max_i \{ c_i \} \big\}$
9:     **for** $k = 1, \ldots, N$ **do**
10:       $u_k \leftarrow \min\{ a_k, \alpha \} \, \mathrm{sign}(z_k)$
11:    **end for**
12:    **return** $\mathbf{u}$
13: **end procedure**

CRAM, as detailed in Algorithm 1, requires the evaluation of the *proximal* operator of the ℓ_∞-norm, which is given by

$$\mathrm{PROXINF}(\mathbf{z}, \tau) = \underset{\tilde{\mathbf{x}} \in \mathbb{C}^N}{\arg\min} \, \|\tilde{\mathbf{x}}\|_\infty + \frac{1}{2\tau} \|\tilde{\mathbf{x}} - \mathbf{z}\|_2^2, \tag{14}$$

where $\tau > 0$. The minimization in (14) does not have a closed form solution. Nevertheless, one can exactly compute $\mathrm{PROXINF}(\mathbf{z}, \tau)$ with low computational cost using the program detailed in Algorithm 2.

B. CRAMP: CRAM for Parseval Frames

The CRAM algorithm detailed above is suitable for arbitrary Frames and approximation parameters $\varepsilon \geq 0$. We next detail an algorithm that is computationally more efficient than CRAM for the special case of Parseval frames and $\varepsilon = 0$.

CRAMP (short for CRAM for Parseval frames) directly solves the complex-valued version of (P_∞) by alternating between projections onto the linear constraint $\mathbf{Dx} = \mathbf{y}$ and the evaluation of the *proximal* operator of ℓ_∞-norm as in (14). For general frames, the projection of $\mathbf{x}$ onto the linear constraint $\mathbf{Dx} = \mathbf{y}$ is given by

$$\Pi(\mathbf{x}) = \mathbf{x} - \mathbf{D}^H (\mathbf{DD}^H)^{-1} (\mathbf{Dx} - \mathbf{y}).$$

This projection, however, requires the computation of the inverse $(\mathbf{DD}^H)^{-1}$, which may result in significant computational costs. When $\mathbf{D}$ is a Parseval frame, we have $\mathbf{DD}^H = \mathbf{I}_{M \times M}$. Consequently,

---

Algorithm 3 CRAMP: CRAM for Parseval Frames

1: **inputs:** $\mathbf{z}_0 \in \mathbb{C}^N$, $\tau \in (0, \infty)$, $k = 0$
2: **while** not converged **do**
3: $\quad \hat{\mathbf{x}}_k \leftarrow \text{ProxInf}(\mathbf{z}_k, \tau)$
4: $\quad \hat{\mathbf{z}}_k \leftarrow 2\hat{\mathbf{x}}_k - \mathbf{z}_k$
5: $\quad \mathbf{x}_{k+1} \leftarrow \Pi(\hat{\mathbf{z}}_k)$
6: $\quad \mathbf{z}_{k+1} \leftarrow \mathbf{z}_k + \mathbf{x}_{k+1} - \hat{\mathbf{x}}_k$
7: $\quad k \leftarrow k + 1$
8: **end while**

---

the above projection simply corresponds to

$$\Pi(\mathbf{x}) = \mathbf{x} - \mathbf{D}^H(\mathbf{D}\mathbf{x} - \mathbf{y}), \tag{15}$$

which can be carried out at low computational complexity. Hence, CRAMP is particularly suited for Parseval frames.[6]

CRAMP is obtained by applying Douglas-Rachford splitting to the equivalent optimization problem

$$(\text{P}_\infty) \quad \underset{\tilde{\mathbf{x}} \in \mathbb{C}^N}{\text{minimize}} \ \|\tilde{\mathbf{x}}\|_\infty + \chi_\Pi(\tilde{\mathbf{x}}),$$

where the proximal of the indicator function $\chi_\Pi(\tilde{\mathbf{x}})$ is simply the projection onto the constraint $\mathbf{D}\tilde{\mathbf{x}} = \mathbf{y}$. By using the ℓ_∞-norm proximal operator (14) and the projection operator for Parseval frames (15), we arrive at CRAMP as summarized in Algorithm 3. We note that convergence of Douglas-Rachford splitting has been proved for arbitrary convex functions and any positive stepsize $\tau > 0$ [42].

The CRAMP algorithm exhibits a practically relevant advantage over CRAM: Every iterate produced by the CRAMP algorithm is *feasible* (i.e., $\mathbf{D}\mathbf{x}_k = \mathbf{y}$ for all k). This property is particularly important in real-time signal processing systems where algorithms are terminated after a pre-determined number of iterations to meet tight throughput constraints. Because all iterates are feasible, CRAMP is guaranteed to terminate with an exact representation $\mathbf{x}$ of the signal vector $\mathbf{y}$, regardless of whether convergence to a *minimum* ℓ_∞-norm solution has been reached.

VII. NUMERICAL EXPERIMENTS

We next provide numerical results that empirically characterize the key properties of democratic representations shown in Section IV. In particular, we simulate a lower bound on K_u in (6) and evaluate

---

[6]Note that the projection (15) can easily adapted to the case of tight frames, i.e., where $A = B$. The resulting projection operator is simply given by $\Pi(\mathbf{x}) = \mathbf{x} - A^{-1}\mathbf{D}^H(\mathbf{D}\mathbf{x} - \mathbf{y})$.
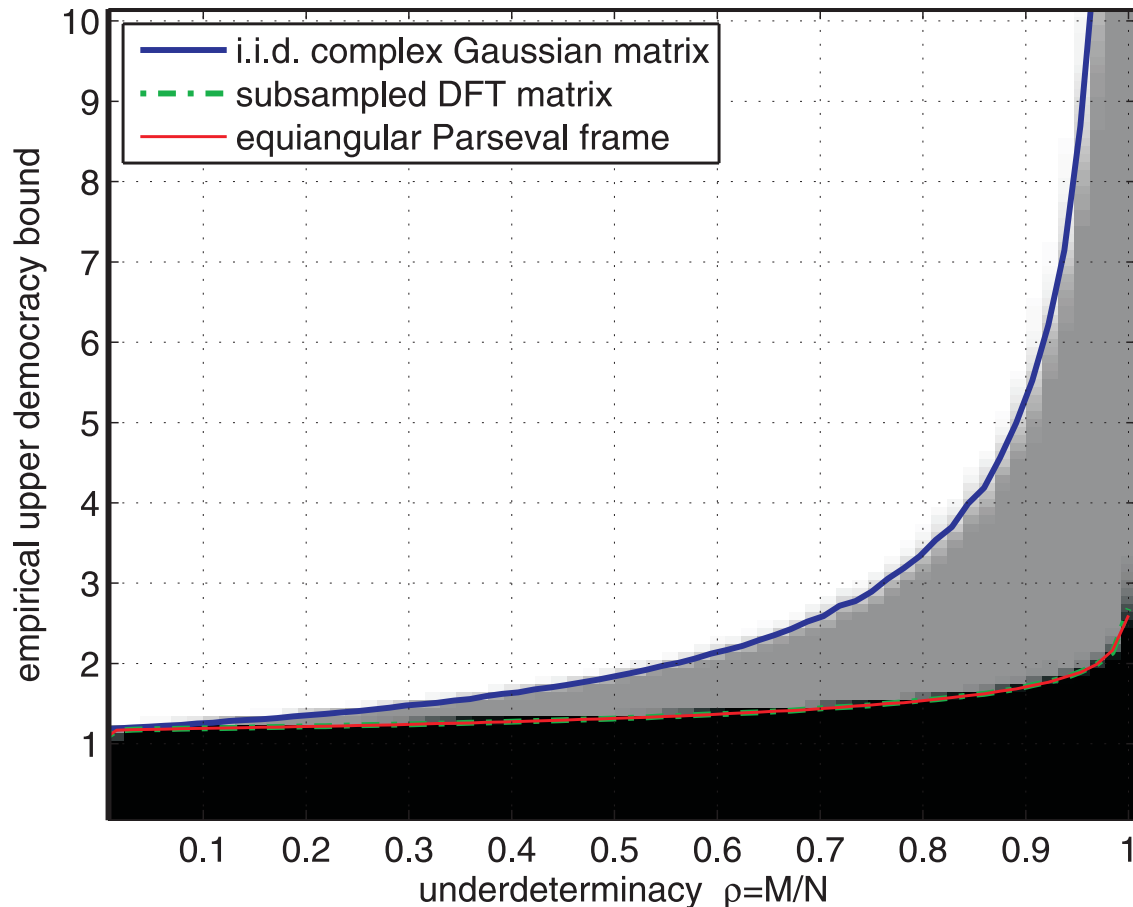
Fig. 2. Empirical phase diagram depending on the underdeterminacy $\rho = M/N$ for the upper Kashin bound K_{u} using $(\mathrm{P}_\infty^\varepsilon)$ with $\varepsilon = 0$ and for (i) i.i.d. complex Gaussian matrices, (ii) randomly subsampled DFT matrices, and (iii) equiangular Parseval frames. The curves represent the individual, sharp 50% phase-transition boundaries. (Note that the curves for subsampled DFT matrices and equiangular Parseval frames overlap.)

the PAPR behavior of solutions to $(\mathrm{P}_\infty^\varepsilon)$ for complex i.i.d. Gaussian and randomly subsampled discrete Fourier transform (DFT) bases. We finally show an application example of democratic representations for PAPR reduction in an OFDM-based DVB-T2 broadcast system.

A. Impact of Frame Properties on the Upper Democracy Constant

In Fig. 2, we show empirical phase diagrams that characterize the upper democracy bound K_{u} for i.i.d. Gaussian matrices, randomly subsampled DCT matrices, and equiangular Parseval frames constructed using the algorithm of [43].

1) Simulation procedure: We fix $N = 512$ and vary M from 1 to 512. For each measurement/dimension pair (N, M), we perform 100 Monte-Carlo trials, and for each trial we generate a frame $\mathbf{D}$ from each
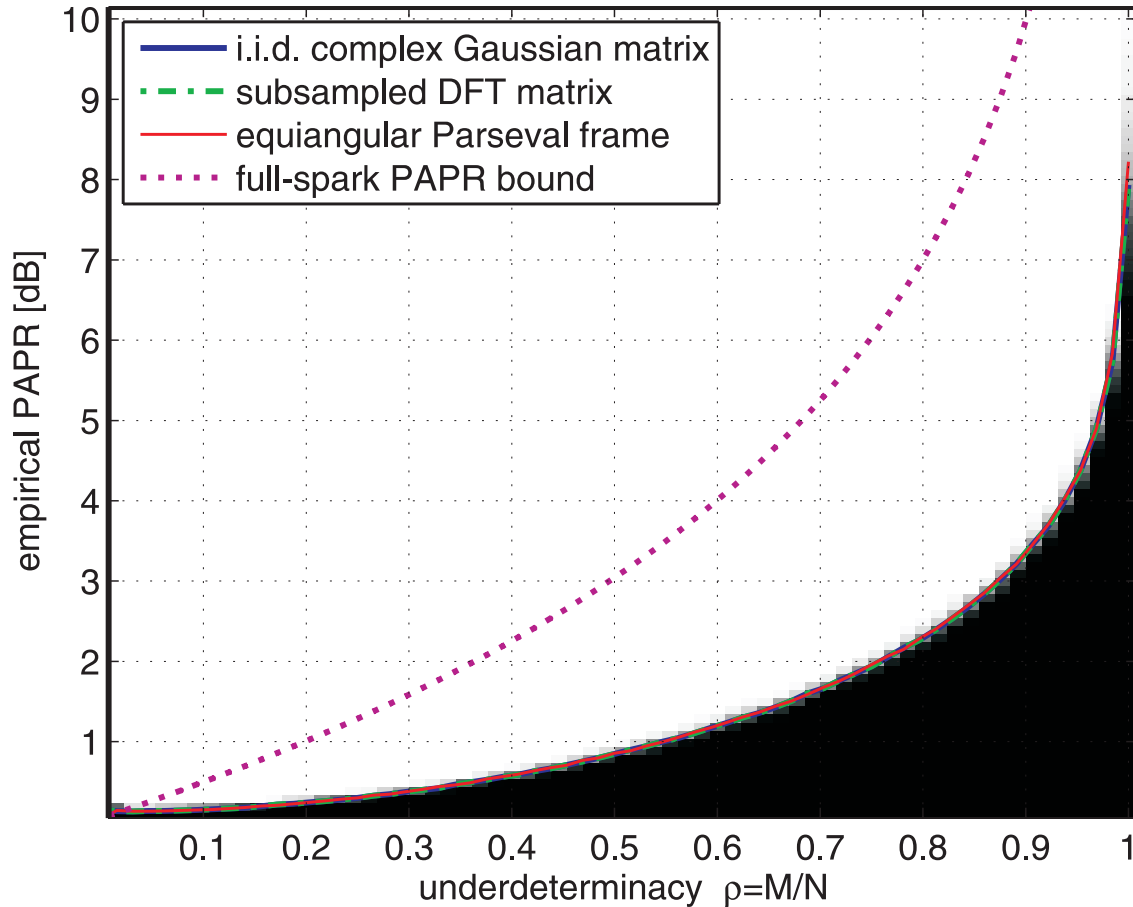
Fig. 3.   Empirical phase diagram depending on the underdeterminancy $\rho = M/N$ for the PAPR using $(\mathrm{P}_\infty^\varepsilon)$ with $\varepsilon = 0$ and for (i) i.i.d. complex Gaussian matrices, (ii) randomly subsampled DFT matrices, and (iii) equiangular Parseval frames. The curves represent the individual, sharp 50% phase-transition boundaries. We also show the full-spark PAPR bound (9). (Note that the curves for i.i.d. complex Gaussian matrices, subsampled DFT matrices, and equiangular Parseval frames overlap.)

matrix/frame class specified above. We furthermore generate a complex i.i.d. zero-mean Gaussian vector $\mathbf{y}$ and normalize it to $\|\mathbf{y}\|_2 = 1$. We use CRAMP from Section VI to compute signal representations $\dot{\mathbf{x}}$ from $(\mathrm{P}_\infty^\varepsilon)$ with $\varepsilon = 0$ for each instance of $\mathbf{D}$ and $\mathbf{y}$. We then compute an empirical lower bound $\hat{K}_\mathrm{u}$ on the upper democracy constant using the obtained representations $\dot{\mathbf{x}}$ for each trial as follows:

$$\hat{K}_\mathrm{u} = \frac{\sqrt{N}\|\dot{\mathbf{x}}\|_\infty}{\|\mathbf{y}\|_2 - \varepsilon} \leq K_\mathrm{u} \quad \text{with} \quad \varepsilon = 0. \tag{16}$$

We finally generate phase diagrams, which show the empirical probability for which $\hat{K}_\mathrm{u}$ is larger or smaller than a given empirical upper democracy constant (given by the y-axis).

   2) Discussion: The empirical phase diagram shown in Fig. 2 shows a sharp transition between the values of $\hat{K}_\mathrm{u}$ that have been realized (for a given under determinacy $\rho = M/N$) and the values that were not achieved. Moreover, we see that subsampled DFTs and equiangular Parseval frames have smaller

(empirical) upper democracy constant than that of i.i.d. Gaussian matrices. This behavior is predicted by (6) and observed previously [36], and can be attributed to the fact that subsampled DFT matrices are Parseval frames, whereas i.i.d. Gaussian matrices are, in general, not tight frames (see also Section V). Hence, the use of Parseval frames tends to yield democratic representations with smaller ℓ_∞-norm than general (non-tight) frames, which is reflected by the upper democracy bound of (6) that explicitly depends on the frame bounds A and B.

B. Impact of Frame Properties on PAPR

In Fig. 3, we characterize the impact of frame properties on the PAPR of signal representations obtained by solving $(\text{P}_\infty^\varepsilon)$.

1) Simulation procedure: We carry out a similar simulation procedure as in Section VII-A1, but instead we compute $PAPR(\dot{\mathbf{x}})$ for each instance of $\mathbf{D}$ and $\mathbf{y}$.

2) Discussion: The phase diagram shown in Fig. 3 exhibits a sharp transition between the (empirical) PAPR values achieved in this simulation and the values that were not achieved. It is interesting to see that all 50% phase transitions overlap, which is in stark contrast to the transition behavior of the upper democracy bound discussed above. We, hence, conclude that the particular choice of the frame has a negligible impact for PAPR-reduction. In addition, Fig. 3 also shows the full-spark PAPR bound (9). We note that the PAPR bound in Theorem 8 depends on the upper frame bound B, which is not reflected in this simulation; an investigation of a tighter PAPR bound is part of ongoing work. Furthermore, the gap between the 50% phase transition and the full-spark PAPR bound appears to rather large. Nevertheless, we note that the full-spark PAPR bound does neither depend on the signal to be represented nor on the specifics of the used frame (apart from its dimensions). Hence, one can imagine that for certain frames one might be able to construct adversarial signals whose representations exhibit high PAPR.

C. Application Example: PAPR Reduction in DVB-T2

We now demonstrate the efficacy of democratic representations for PAPR reduction via tone reservation in an OFDM-based DVB-T2 broadcast system [28].

1) Algorithm details and simulation procedure: We consider a simplified[7] DVB-T2 system, where we use the tones reserved for PAPR reduction to generate OFDM time-domain signals $\mathbf{x}$ having low PAPR. In particular, we generate the entries of the frequency-domain (signal) vector $\mathbf{y}$ by inserting i.i.d. random

---

[7]We ignore DVB-T2-specific OFDM frame structures, such as pilot tones. For the sake of simplicity, we generate 256-QAM symbols for all used tones.

256-QAM symbols into the data-carrying tones and by inserting 0's into the specified zero-tones. The set of entries in $\mathbf{y}$ containing the constellation symbols and the zero-tones is denoted by Ω; the complement set Ω^c contains the tones reserved for PAPR reduction. We can now write the time-domain vector $\mathbf{x}$ as $\mathbf{Fx} = \mathbf{y}$, where $\mathbf{F}$ is a DFT matrix of appropriate size that satisfies $\mathbf{FF}^H = \mathbf{I}$. In the following experiment, we use a DFT of dimension $N = 32\,768$ as specified in [28]. For this particular DFT size, we have $|\Omega^c| = 288$ tones reserved for PAPR reduction. By separating the signal vector $\mathbf{y}$ into two disjoint parts $\mathbf{y}_\Omega$ and $\mathbf{y}_{\Omega^c}$, we can rewrite the time-domain vector as $\mathbf{x} = \mathbf{F}_\Omega^H \mathbf{y}_\Omega + \mathbf{F}_{\Omega^c}^H \mathbf{y}_{\Omega^c}$. Hence, for the OFDM tones in Ω, we have $\mathbf{y}_\Omega = (\mathbf{F}_\Omega^H)^H \mathbf{x}$; here, $(\mathbf{F}_\Omega^H)^H$ is a subsampled DFT matrix having $M = 32\,480$ rows from the set Ω and all $N = 32\,768$ columns. Since $\mathbf{x}$ is the time-domain vector to be transmitted, we can reduce its PAPR by solving the following problem:

$$(\text{PR}) \quad \underset{\tilde{\mathbf{x}} \in \mathbb{C}^N}{\text{minimize}} \, \|\tilde{\mathbf{x}}\|_\infty \quad \text{subject to } \mathbf{y}_\Omega = (\mathbf{F}_\Omega^H)^H \tilde{\mathbf{x}},$$

which is a specific instance of $(\text{P}_\infty^\varepsilon)$ with $\mathbf{D} = (\mathbf{F}_\Omega^H)^H$ and $\varepsilon = 0$. By solving (PR) we obtain a time-domain signal $\dot{\mathbf{x}}$ that has (i) low PAPR and (ii) a frequency-domain representation that corresponds to $\mathbf{y}_\Omega$ on the set of used OFDM tones.

We note that, in practice, the time-domain signals pass through a digital-to-analog converter, which typically applies a low-pass reconstruction filter to the resulting time-domain signal. To accurately assess the PAPR of the resulting analog (filtered) time-domain signal, one typically considers the PAPR of an oversampled system, which is achieved by an appropriate zero-padding of frequency-domain vector $\mathbf{y}$ (see [44] and the references therein). As in [44] we compute the PAPR using $4\times$ oversampling.

In the following experiments, we generate 10^5 OFDM signals $\mathbf{y}$ as specified above and compare the PAPR of the following methods/algorithms: (i) conventional OFDM transmission (where $y_i = 0$ for $i \in \Omega^c$); (ii) PAPR-reduced OFDM transmission using the algorithm detailed in the DVB-T2 standard[8] [28, Sec. 9.6.2.1]; (iii) PAPR-reduced OFDM transmission as by solving (PR) via CRAMP[9]; (iv) PAPR-reducd OFDM transmission by solving a variant of (PR) via CRAMP that directly operates on the $4\times$ oversampled system. As a performance measure, we compare the complementary cumulative distribution function (CDF) of the oversampled PAPR values (in decibel) obtained in all simulation trials [44].

2) Discussion: From Fig. 4, we see that conventional OFDM transmission exhibits the largest PAPR. The algorithm in [28] is able to reduce the PAPR by roughly $1.5\,\text{dB}$ (corresponding to a complementary

---

[8]We perform $1\,000$ algorithm iterations and use a set of optimized algorithm parameters to achieve minimal PAPR.

[9]The maximum number of iterations is set to $1\,000$; on average, CRAMP terminates after 250 iterations.
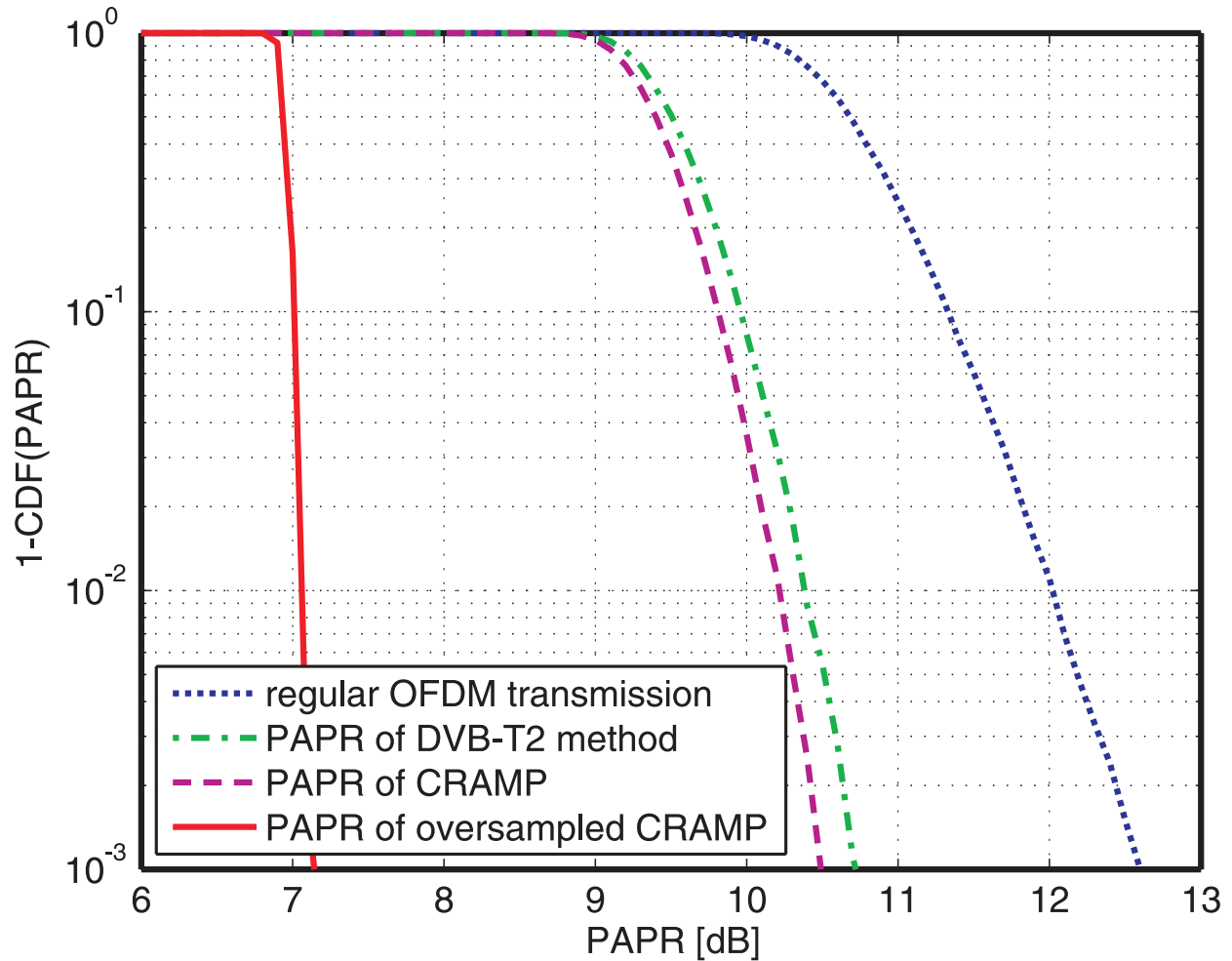
Fig. 4.   PAPR reduction performance for a DVB-T2-based broadcast system with $32\,768$ OFDM tones and 256-QAM modulation [28]. We compare the PAPR reduction performance of (i) regular OFDM transmission (no PAPR reduction use), (ii) the algorithm specified in the DVB-T2 standard, (iii) CRAMP, and (iv) CRAMP directly applied to the $4\times$ oversampled system.

CDF of 10^{-2}). Solving (PR) via CRAMP reduces the PAPR by roughly $1.7\,\mathrm{dB}$. Solving (PR) directly on the oversampled system leads to a significant PAPR reduction of about $4.5\,\mathrm{dB}$ (or $3\,\mathrm{dB}$ more than conventional schemes). Hence, PAPR reduction using ℓ_∞-norm minimization is able to significantly outperform existing methods.

We conclude by noting that CRAMP-based PAPR reduction exhibits, in general, higher complexity than the algorithm specified in the DVB-T2 standard [28], but requires no parameter tuning. Since CRAMP does not exploit the fact that the effective matrix $(\mathbf{F}_\Omega^H)^H$ in the considered application has a very specific structure and extremely low redundancy (i.e., $\lambda \approx 1.008$), we are convinced that more efficient algorithms can be developed for solving (PR) in this particular setting.

VIII. CONCLUSIONS

In this paper, we have analyzed a host of fundamental properties of signal representations with minimum ℓ_∞ (or maximum) norm. Specifically, we have developed properties on the magnitudes of such representations, and we characterized their peak-to-average power (PAPR) properties, which is of practical interest for OFDM-based communication systems. We have demonstrated the existence of matrices for which *democratic representations* with small ℓ_∞-norm and small PAPR exist. We have furthermore developed two new and computationally efficient algorithms for solving $(\mathrm{P}_\infty^\varepsilon)$. To support our analysis, we have conducted a set of numerical experiments, which highlight that (i) Parseval frames lead to democratic representations with smaller ℓ_∞-norm compared to general frames and (ii) democratic representations offer tremendous PAPR reduction gains over existing approaches.

There are many avenues for follow-on research. An analytical characterization of the sharp phase transitions for $(\mathrm{P}_\infty^\varepsilon)$ observed in Section VII-A, e.g., using techniques developed in [22], [45], is an interesting open research problem. In addition, the development of algorithms particularly suited for PAPR reduction in OFDM-based communication systems is left for future work. We also believe that assessing the efficacy of democratic representations in other practical applications, such as vector quantization, approximate nearest neighbor search, filter design, or robotics and control, is an interesting research direction.

APPENDIX A

PROOF OF LEMMA 1

We proceed in two steps: We first show that there exists a solution with at least $N - M$ extreme values. We then use this result to prove the existence of a solution with at least $N - M + 1$ extreme values.

Assume for contradiction that every solution to $(\mathrm{P}_\infty^\varepsilon)$ has less than $N - M$ extreme values. Let $\dot{\mathbf{x}}$ be a solution to $(\mathrm{P}_\infty^\varepsilon)$ with the largest possible number of extreme entries. Let $\Sigma \subseteq \{1, \ldots, N\}$ denote the set of indices corresponding to the extreme entries of $\dot{\mathbf{x}}$ and Σ^c denote the complement set, i.e., the set of indices corresponding to *moderate* entries. By assumption $|\Sigma| < N - M$. We will now create a contradiction by constructing a solution with more extreme values than $\dot{\mathbf{x}}$.

Let $\mathbf{D}_{\Sigma^c}$ denote the columns of $\mathbf{D}$ corresponding to moderate entries of $\mathbf{x}$. There are more than M moderate entries in $\dot{\mathbf{x}}$, so the columns of $\mathbf{D}_{\Sigma^c}$ (which are each of dimension M) must be linearly dependent. It follows that there exists a non-zero vector $\mathbf{v}$ with $\mathrm{supp}(\mathbf{v}) \subseteq \Sigma^c$ and $\mathbf{D}\mathbf{v} = \mathbf{0}$. Let $\alpha \in \mathbb{R}$ be the largest scalar with $\|\dot{\mathbf{x}} + \alpha\mathbf{v}\|_\infty = \|\dot{\mathbf{x}}\|_\infty$. Note the vector $\mathbf{v}$ is only non-zero on the moderate coordinates of $\dot{\mathbf{x}}$. Because α is maximal, $|[\dot{\mathbf{x}} + \alpha\mathbf{v}]_i| = \|\dot{\mathbf{x}}\|_\infty$ for some $i \in \Sigma^c$. The vector $\dot{\mathbf{x}} + \alpha\mathbf{v}$ has

the same ℓ_∞-norm as $\dot{\mathbf{x}}$ and $\mathbf{D}\dot{\mathbf{x}} = \mathbf{D}(\dot{\mathbf{x}} + \alpha\mathbf{v})$. As a consequence, $\dot{\mathbf{x}} + \alpha\mathbf{v}$ is also a solution to $(\mathrm{P}_\infty^\varepsilon)$. However, by construction, the vector $\dot{\mathbf{x}} + \alpha\mathbf{v}$ has more extreme entries than $\dot{\mathbf{x}}$, which contradicts our assumption. We have thus established the existence of a solution with at least $N - M$ extreme values.

We now show the existence of a solution with $N - M + 1$ extreme values. Assume for contradiction that no such solution exists. Then, the argument above guarantees the existence of a solution $\dot{\mathbf{x}}$ with exactly $N - M$ extreme values. Proceeding as before, let $\Sigma \subseteq \{1, \ldots, N\}$ contain the indices of extreme entries in $\dot{\mathbf{x}}$, and let Σ^c contain the moderate indices. The matrix $\mathbf{D}_{\Sigma^c}$ has dimension $M \times M$. If the columns of $\mathbf{D}_{\Sigma^c}$ are linearly dependent, then the argument above can be repeated to construct a solution with more than $N - M$ extreme values. We thus only need to consider the case where $\mathbf{D}_{\Sigma^c}$ has full rank. In this case, there is a vector $\mathbf{v} \in \mathbb{C}^N$ with $\operatorname{supp}(\mathbf{v}) \subseteq \Sigma^c$ and $\mathbf{D}\mathbf{v} = \mathbf{D}\dot{\mathbf{x}}_\Sigma$. Note that $\|\dot{\mathbf{x}}_{\Sigma^c}\|_\infty < \|\dot{\mathbf{x}}\|_\infty$, and so we can choose some $0 < \delta < 1$ to be small enough that $\|\dot{\mathbf{x}}_{\Sigma^c} + \delta\mathbf{v}\|_\infty < \|\dot{\mathbf{x}}\|_\infty$.

Now consider the vector $\mathbf{y} = \dot{\mathbf{x}} - \delta\dot{\mathbf{x}}_\Sigma + \delta\mathbf{v}$. This vector has the following ℓ_∞-norm bound:

$$
\begin{aligned}
\|\mathbf{y}\|_\infty &= \|\dot{\mathbf{x}} - \delta\dot{\mathbf{x}}_\Sigma + \delta\mathbf{v}\|_\infty \\
&= \|\dot{\mathbf{x}}_\Sigma - \delta\dot{\mathbf{x}}_\Sigma + \dot{\mathbf{x}}_{\Sigma^c} + \delta\mathbf{v}\|_\infty \\
&= \|(1-\delta)\dot{\mathbf{x}}_\Sigma + \dot{\mathbf{x}}_{\Sigma^c} + \delta\mathbf{v}\|_\infty \\
&= \max\{(1-\delta)\|\dot{\mathbf{x}}_\Sigma\|_\infty, \|\dot{\mathbf{x}}_{\Sigma^c} + \delta\mathbf{v}\|_\infty\} < \|\dot{\mathbf{x}}\|_\infty. \quad (17)
\end{aligned}
$$

At the same time, we have $\mathbf{D}\mathbf{y} = \mathbf{D}(\dot{\mathbf{x}} - \delta\dot{\mathbf{x}}_\Sigma + \delta\mathbf{v}) = \mathbf{D}\dot{\mathbf{x}}$. It follows that $\dot{\mathbf{x}}$ is not a true solution to $(\mathrm{P}_\infty^\varepsilon)$, which is a contradiction. Hence, there exists a solution with at least $N - M + 1$ extreme values.

Appendix B

Proof of Lemma 2

Assume for contradiction that $(\mathrm{P}_\infty^\varepsilon)$ admits a solution $\dot{\mathbf{x}}$ with $|\Sigma| < N - M + 1$. Let $\dot{\mathbf{x}}_\Sigma$ be the vector $\dot{\mathbf{x}}$ with support restricted to Σ. Consider the columns of $\mathbf{D}$ corresponding to the moderate entries of $\dot{\mathbf{x}}$, i.e., $\mathbf{D}_{\Sigma^c}$. Because $\mathbf{D}$ is a full-spark frame and $|\Sigma^c| \geq M$, the set of columns $\mathbf{D}_{\Sigma^c}$ has full rank. The remainder of the proof mirrors that of Lemma 1. Because $\mathbf{D}_{\Sigma^c}$ is invertible, there is a vector $\mathbf{v} \in \mathbb{C}^N$ with $\operatorname{supp}(\mathbf{v}) \subseteq \Sigma^c$ and $\mathbf{D}\mathbf{v} = \mathbf{D}\dot{\mathbf{x}}_\Sigma$. Note that $\|\dot{\mathbf{x}}_{\Sigma^c}\|_\infty < \|\dot{\mathbf{x}}\|_\infty$, and so we can choose some small $0 < \delta < 1$ with $\|\dot{\mathbf{x}}_{\Sigma^c} + \delta\mathbf{v}\|_\infty < \|\dot{\mathbf{x}}\|_\infty$.

As in Appendix A, we consider the vector $\mathbf{y} = \dot{\mathbf{x}} - \delta\dot{\mathbf{x}}_\Sigma + \delta\mathbf{v}$, which satisfies the bound established in (17). At the same time, we have $\mathbf{D}\mathbf{y} = \mathbf{D}(\dot{\mathbf{x}} - \delta\dot{\mathbf{x}}_\Sigma + \delta\mathbf{v}) = \mathbf{D}\dot{\mathbf{x}}$. Hence, it follows that $\dot{\mathbf{x}}$ is not a true solution to $(\mathrm{P}_\infty^\varepsilon)$, which contradicts our assumption.

APPENDIX C

PROOF OF THEOREM 4

Let $\|\mathbf{w}\|_p$ and $\|\mathbf{v}\|_d$ denote the primal and dual norm of the vectors $\mathbf{w}$ and $\mathbf{x}$ satisfying

$$\|\mathbf{w}\|_p = \max_{\mathbf{v}} \left\{ \Re\left(\mathbf{v}^H \mathbf{w}\right) : \|\mathbf{v}\|_p \leq 1 \right\}$$

with $1/p + 1/q = 1$. Then, for primal and dual norms, we have the following result [38]:

$$\min_{\mathbf{x}} \left\{ \|\mathbf{x}\|_p - \Re\left(\mathbf{z}^H \mathbf{D} \mathbf{x}\right) \right\} = \begin{cases} 0, & \|\mathbf{D}^H \mathbf{z}\|_d \leq 1 \\ -\infty, & \text{otherwise.} \end{cases} \tag{18}$$

We are now ready to derive the Lagrange dual problem $(\mathrm{D}_p^\varepsilon)$ to the primal problem $(\mathrm{P}_p^\varepsilon)$. To this end, we introduce the auxiliary vector $\mathbf{r} \in \mathbb{C}^M$ to rewrite $(\mathrm{P}_p^\varepsilon)$ as

$$\min_{\mathbf{x}} \left\{ \|\mathbf{x}\|_p : \|\mathbf{D}\mathbf{x} - \mathbf{y}\|_2 \leq \varepsilon \right\}$$
$$= \min_{\mathbf{x},\mathbf{r}} \left\{ \|\mathbf{x}\|_p : \mathbf{D}\mathbf{x} + \mathbf{r} = \mathbf{y}, \|\mathbf{r}\|_2 \leq \varepsilon \right\}.$$

By introducing the Lagrange dual variable $\mathbf{z} \in \mathbb{C}^M$, we obtain

$$\min_{\mathbf{x},\mathbf{r}} \left\{ \|\mathbf{x}\|_p : \mathbf{D}\mathbf{x} + \mathbf{r} = \mathbf{y}, \|\mathbf{r}\|_2 \leq \varepsilon \right\}$$
$$= \min_{\mathbf{x},\mathbf{r}} \max_{\mathbf{z}} \left\{ \|\mathbf{x}\|_p - \Re\left(\mathbf{z}^H(\mathbf{D}\mathbf{x} + \mathbf{r} - \mathbf{y})\right) : \|\mathbf{r}\|_2 \leq \varepsilon \right\}$$
$$= \max_{\mathbf{z}} \min_{\mathbf{x},\mathbf{r}} \left\{ \|\mathbf{x}\|_p - \Re\left(\mathbf{z}^H(\mathbf{D}\mathbf{x} + \mathbf{r} - \mathbf{y})\right) : \|\mathbf{r}\|_2 \leq \varepsilon \right\}. \tag{19}$$

For a given $\mathbf{z}$, the inner minimization problem of (19) is separable in the unknown vectors $\mathbf{x}$ and $\mathbf{r}$. The optimal auxiliary vector $\mathbf{r}$ is given by

$$\mathbf{r} = \begin{cases} \varepsilon \mathbf{z} / \|\mathbf{z}\|_2, & \mathbf{z} \neq \mathbf{0}_{M \times 1} \\ \mathbf{0}_{M \times 1}, & \text{otherwise,} \end{cases}$$

and, in either case, we have $\Re\left(\mathbf{z}^H \mathbf{r}\right) = \varepsilon \|\mathbf{z}\|_2$. Together with (18), we find that (19) is equal to

$$\max_{\mathbf{z}} \left\{ \Re\left(\mathbf{y}^H \mathbf{z}\right) - \varepsilon \|\mathbf{z}\|_2 : \|\mathbf{D}^H \mathbf{z}\|_d \leq 1 \right\},$$

which corresponds to the Lagrange dual problem $(\mathrm{D}_p^\varepsilon)$. Note that since in the derivation of $(\mathrm{D}_p^\varepsilon)$ all intermediate steps hold with equality, there is no duality gap.

APPENDIX D

PROOF OF LEMMA 5

The proof follows from a lower bound on the value of the dual problem (D_∞^ε). Specifically, we have

$$\|\dot{\mathbf{x}}\|_\infty = \max_{\mathbf{z}} \left\{ \Re(\mathbf{y}^H \mathbf{z}) - \varepsilon \|\mathbf{z}\|_2 : \left\| \mathbf{D}^H \mathbf{z} \right\|_1 \leq 1 \right\}, \tag{20}$$

which we bound from below by replacing the optimal solution $\dot{\mathbf{z}}$ by the following estimate:

$$\hat{\mathbf{z}} = \frac{\mathbf{y}}{\left\| \mathbf{D}^H \mathbf{y} \right\|_1}, \tag{21}$$

which satisfies the constraint $\left\| \mathbf{D}^H \hat{\mathbf{z}} \right\|_1 \leq 1$. Hence, inserting (21) in the right-hand side (RHS) of (20) leads to the following lower bound:

$$\|\dot{\mathbf{x}}\|_\infty \geq \frac{\|\mathbf{y}\|_2^2 - \varepsilon \|\mathbf{y}\|_2}{\left\| \mathbf{D}^H \mathbf{y} \right\|_1}. \tag{22}$$

To further bound the RHS of (22) from below, we use standard norm bounds and the upper frame bound B of $\mathbf{D}$ to compute an upper bound to $\left\| \mathbf{D}^H \mathbf{y} \right\|_1$ as follows:

$$\left\| \mathbf{D}^H \mathbf{y} \right\|_1 \leq \sqrt{N} \left\| \mathbf{D}^H \mathbf{y} \right\|_2 \leq \sqrt{NB} \|\mathbf{y}\|_2. \tag{23}$$

Combining (23) with (22) finally yields

$$\|\dot{\mathbf{x}}\|_\infty \geq \frac{\|\mathbf{y}\|_2 - \varepsilon}{\sqrt{NB}}. \tag{24}$$

Note that in (21) we assumed that $\left\| \mathbf{D}^H \mathbf{y} \right\|_1 > 0$. Since $\mathbf{D}$ is a frame with lower frame bound $A > 0$, we have

$$\left\| \mathbf{D}^H \mathbf{y} \right\|_1 \geq \left\| \mathbf{D}^H \mathbf{y} \right\|_2 \geq \sqrt{A} \|\mathbf{y}\|_2 > 0,$$

which is satisfied whenever $\|\mathbf{y}\|_2 > 0$. In the case $\|\mathbf{y}\|_2 = 0$ the bound (24) continues to hold.

APPENDIX E

PROOF OF THEOREM 6

The proof proceeds in two stages. First, we separate the objective function of the Lagrange dual problem (D_∞^ε) into two independent terms and second, we derive an upper bound on the ℓ_2-norm of the solution $\dot{\mathbf{z}}$ to (D_∞^ε).

A. Separating the Result of the Lagrange Dual Problem

From the Lagrange dual problem (D^ε_∞), we have

$$\|\dot{\mathbf{x}}\|_\infty = \Re(\mathbf{y}^H\dot{\mathbf{z}}) - \varepsilon\|\dot{\mathbf{z}}\|_2 \le \left|\mathbf{y}^H\dot{\mathbf{z}}\right| - \varepsilon\|\dot{\mathbf{z}}\|_2$$

$$\le \|\dot{\mathbf{z}}\|_2(\|\mathbf{y}\|_2 - \varepsilon), \tag{25}$$

as an immediate consequence of the Cauchy-Schwarz inequality.[10] In the remaining steps of the proof, we derive an upper bound on $\|\dot{\mathbf{z}}\|_2$ in (25). To this end, we first expand

$$\|\dot{\mathbf{z}}\|_2 = \left\|(\mathbf{DD}^H)^{-1}\mathbf{DD}^H\dot{\mathbf{z}}\right\|_2, \tag{26}$$

where $\mathbf{DD}^H$ is invertible since $\mathbf{D}$ is a frame with lower frame bound satisfying $A > 0$. Application of the Rayleigh-Ritz theorem [46, Thm. 4.2.2] to the right-hand side (RHS) of (26) leads to the following upper bound:

$$\|\dot{\mathbf{z}}\|_2 \le \left\|(\mathbf{DD}^H)^{-1}\right\|_{2,2}\left\|\mathbf{DD}^H\dot{\mathbf{z}}\right\|_2 \le \frac{1}{A}\left\|\mathbf{DD}^H\dot{\mathbf{z}}\right\|_2, \tag{27}$$

where the second inequality is a result of

$$\left\|(\mathbf{DD}^H)^{-1}\right\|_{2,2} = \frac{1}{\|(\mathbf{DD}^H)^{-1}\|_{2,2}} \le \frac{1}{A}$$

and the assumption that $\mathbf{D}$ is a frame with lower frame bound $A > 0$. We next derive an upper bound on $\left\|\mathbf{DD}^H\mathbf{z}\right\|_2$ in (27).

Note that one can straightforwardly arrive at an upper bound on $\|\dot{\mathbf{x}}\|_\infty$ as follows:

$$\left\|\mathbf{DD}^H\dot{\mathbf{z}}\right\|_2 \le \|\mathbf{D}\|_{2,2}\left\|\mathbf{D}^H\dot{\mathbf{z}}\right\|_1 \le \|\mathbf{D}\|_{2,2} \tag{28}$$

using $\left\|\mathbf{D}^H\dot{\mathbf{z}}\right\|_2 \le \left\|\mathbf{D}^H\dot{\mathbf{z}}\right\|_1$ and the constraint $\left\|\mathbf{D}^H\dot{\mathbf{z}}\right\|_1 \le 1$ of the dual problem (D^ε_∞). Hence, by combining (25), (27), and (28) one would arrive at the following result:

$$\|\dot{\mathbf{x}}\|_\infty \le \frac{\|\mathbf{D}\|_{2,2}}{A}\left(\|\mathbf{y}\|_2 - \varepsilon\right). \tag{29}$$

This bound is, however, overly pessimistic and does not exploit additional properties of the frame $\mathbf{D}$. Note that for Parseval frames, the result (29) leads to the bound $\|\dot{\mathbf{x}}\|_\infty \le \|\mathbf{y}\|_2 - \varepsilon$.

---

[10]Note that the bound (25) appears to be tight for $\varepsilon = 0$, i.e., we were able to construct signal and frame instances for which we have $\|\dot{\mathbf{x}}\|_\infty = \|\mathbf{y}\|_2\|\dot{\mathbf{z}}\|_2$ up to machine precision. A systematic characterization of such signal and frame instances is left for future work.

B. Refined Upper Bound

In order to arrive at a refined bound on $\left\|\mathbf{DD}^H\dot{\mathbf{z}}\right\|_2$, we define an N-dimensional vector $\mathbf{v} = \mathbf{D}^H\dot{\mathbf{z}}$ and divide its coefficients into $S = \lceil 1/\delta \rceil$ disjoint support sets, each[11] of cardinality δN such that

$$\Omega_1 \cup \cdots \cup \Omega_\ell \in \{1, \ldots, N\},$$

where $\lceil x \rceil$ rounds the scalar $x \in \mathbb{R}$ to the nearest integer towards infinity. Moreover, the magnitudes of the entries in $\mathbf{v}$ associated to set Ω_ℓ are no smaller than the magnitudes associated with the sets Ω_k, $k > \ell$. In other words, Ω_1 contains the indices associated to the largest δN entries in $\mathbf{v}$, Ω_2 the δN coefficients associated to the second largest entries, etc. This partitioning scheme now allows us to rewrite $\left\|\mathbf{DD}^H\dot{\mathbf{z}}\right\|_2$ as

$$\left\|\mathbf{DD}^H\dot{\mathbf{z}}\right\|_2 = \left\|\mathbf{D}\sum_{i=1}^{S}\mathbf{P}_{\Omega_i}\mathbf{D}^H\dot{\mathbf{z}}\right\|_2,$$

where the matrix $\mathbf{P}_{\Omega_i}$ realizes a projection onto the set Ω_i. Application of the triangle inequality, followed by using properties of the UP with parameters η, δ leads to the following:

$$\left\|\mathbf{DD}^H\dot{\mathbf{z}}\right\|_2 \leq \sum_{i=1}^{S}\left\|\mathbf{DP}_{\Omega_i}\mathbf{D}^H\dot{\mathbf{z}}\right\|_2 \leq \sum_{i=1}^{S}\eta\left\|\mathbf{P}_{\Omega_i}\mathbf{D}^H\dot{\mathbf{z}}\right\|_2$$
$$= \eta\left\|\mathbf{P}_{\Omega_1}\mathbf{D}^H\dot{\mathbf{z}}\right\|_2 + \sum_{i=2}^{S}\eta\left\|\mathbf{P}_{\Omega_i}\mathbf{D}^H\dot{\mathbf{z}}\right\|_2. \tag{30}$$

Since the sets Ω_i order the entries of $\mathbf{v} = \mathbf{D}^H\dot{\mathbf{z}}$ according to their magnitudes, we can use a technique developed in [47], which states that for $i \in \{2, \ldots, S\}$ we have

$$\left\|\mathbf{P}_{\Omega_i}\mathbf{v}\right\|_2 \leq \sqrt{\delta N}\left\|\mathbf{P}_{\Omega_i}\mathbf{v}\right\|_\infty \leq \frac{1}{\sqrt{\delta N}}\left\|\mathbf{P}_{\Omega_{i-1}}\mathbf{v}\right\|_1.$$

This result in combination with the RHS of (30) leads to

$$\left\|\mathbf{DD}^H\dot{\mathbf{z}}\right\|_2 \leq \eta\left\|\mathbf{P}_{\Omega_1}\mathbf{D}^H\dot{\mathbf{z}}\right\|_2 + \sum_{i=1}^{S}\frac{\eta}{\sqrt{\delta N}}\left\|\mathbf{P}_{\Omega_i}\mathbf{D}^H\dot{\mathbf{z}}\right\|_1$$
$$= \eta\left\|\mathbf{P}_{\Omega_1}\mathbf{D}^H\dot{\mathbf{z}}\right\|_2 + \frac{\eta}{\sqrt{\delta N}}\left\|\mathbf{D}^H\dot{\mathbf{z}}\right\|_1$$
$$= \eta\left\|\mathbf{P}_{\Omega_1}\mathbf{D}^H\dot{\mathbf{z}}\right\|_2 + \frac{\eta}{\sqrt{\delta N}}, \tag{31}$$

---

[11]Note that the last support set Ω_ℓ can have a cardinality that is smaller than δN; such cases, however, leave the proof unaffected.

where the first equality follows from the fact that $\left\|\mathbf{D}^H \dot{\mathbf{z}}\right\|_1 \leq 1$ for any solution $\dot{\mathbf{z}}$ to the dual problem ($\mathrm{D}_\infty^\varepsilon$).

We can now bound the first RHS term in (31) as

$$\left\|\mathbf{P}_{\Omega_1}\mathbf{D}^H \dot{\mathbf{z}}\right\|_2 \leq \left\|\mathbf{D}^H \dot{\mathbf{z}}\right\|_2 \leq \sqrt{B}\|\dot{\mathbf{z}}\|_2 \tag{32}$$

using the facts that (i) $\mathbf{P}_{\Omega_1}$ is a projector and (ii) $\mathbf{D}$ is a frame with (upper) frame bound B. By combining (27), (31), and (32) we arrive at

$$\|\dot{\mathbf{z}}\|_2 \leq \frac{1}{A}\left(\eta\sqrt{B}\|\dot{\mathbf{z}}\|_2 + \frac{\eta}{\sqrt{\delta N}}\right),$$

which can be rewritten as

$$\|\dot{\mathbf{z}}\|_2 \leq \frac{\eta}{(A - \eta\sqrt{B})\sqrt{\delta N}} \tag{33}$$

provided that $A > \eta\sqrt{B}$ holds. Combining (25) with (33) finally yields

$$\|\dot{\mathbf{x}}\|_\infty \leq \frac{\eta}{(A - \eta\sqrt{B})\sqrt{\delta N}}\left(\|\mathbf{y}\|_2 - \varepsilon\right), \tag{34}$$

which concludes the proof. We finally note that (34) is able to scale in $1/\sqrt{N}\|\mathbf{y}\|_2$ for certain frames (see Section V).

APPENDIX F

PROOF OF THEOREM 8

The proof follows from separately bounding the numerator and denominator of the PAPR defined in (7). We first bound $N\|\dot{\mathbf{x}}\|_\infty^2$ using (6) to arrive at

$$N\|\dot{\mathbf{x}}\|_\infty^2 \leq \widetilde{K}_{\mathrm{u}}^2(\|\mathbf{y}\|_2 - \varepsilon)^2. \tag{35}$$

The second part of the proof bounds $\|\dot{\mathbf{x}}\|_2^2$ from below. To this end, it is important to realize that

$$\|\dot{\mathbf{x}}\|_2 \geq \min_{\mathbf{x}}\left\{\|\mathbf{x}\|_2 : \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 \leq \varepsilon\right\} \tag{36}$$

because $\dot{\mathbf{x}}$ satisfies $\|\mathbf{y} - \mathbf{D}\dot{\mathbf{x}}\|_2 \leq \varepsilon$ and the RHS is the minimizer for all vectors $\mathbf{x} \in \mathbb{C}^N$ satisfying $\|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 \leq \varepsilon$. We next compute a lower bound on the RHS of (36). From Theorem 4 with $p = 2$ and

$q = 2$, we have

$$\min_{\mathbf{x}} \left\{ \|\mathbf{x}\|_2 : \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 \leq \varepsilon \right\}$$
$$= \max_{\mathbf{z}} \left\{ \Re(\mathbf{y}^H \mathbf{z}) - \varepsilon \|\mathbf{z}\|_2 : \left\| \mathbf{D}^H \mathbf{z} \right\|_2 \leq 1 \right\}. \tag{37}$$

Using a similar strategy as in Appendix D, we replace the optimal solution $\dot{\mathbf{z}}$ of the dual problem in (37) by the estimate

$$\dot{\mathbf{z}} = \frac{\mathbf{y}}{\left\| \mathbf{D}^H \mathbf{y} \right\|_2}, \tag{38}$$

which satisfies the constraint $\left\| \mathbf{D}^H \mathbf{y} \right\|_2 \leq 1$. Hence, by inserting the estimate (38) into the RHS of (37), we obtain the following lower bound:

$$\max_{\mathbf{z}} \left\{ \Re(\mathbf{y}^H \mathbf{z}) - \varepsilon \|\mathbf{z}\|_2 : \left\| \mathbf{D}^H \mathbf{z} \right\|_2 \leq 1 \right\}$$
$$\geq \frac{\|\mathbf{y}\|_2^2 - \varepsilon \|\mathbf{y}\|_2}{\left\| \mathbf{D}^H \mathbf{y} \right\|_2}. \tag{39}$$

The upper frame bound $\left\| \mathbf{D}^H \mathbf{y} \right\|_2 \leq \sqrt{B} \|\mathbf{y}\|_2$ enables us to further bound the RHS of (39) from below as

$$\frac{\|\mathbf{y}\|_2^2 - \varepsilon \|\mathbf{y}\|_2}{\left\| \mathbf{D}^H \mathbf{y} \right\|_2} \geq \frac{\|\mathbf{y}\|_2 - \varepsilon}{\sqrt{B}}. \tag{40}$$

By combining (36), (37), (39), and (40), we finally obtain

$$\|\dot{\mathbf{x}}\|_2^2 \geq \frac{\left(\|\mathbf{y}\|_2 - \varepsilon \right)^2}{B}. \tag{41}$$

Consequently, if $\varepsilon < \|\mathbf{y}\|_2$, then we can bound the PAPR of the representation $\dot{\mathbf{x}}$ obtained from (P_∞^ε) using (35) and (41) as $PAPR(\dot{\mathbf{x}}) \leq \widetilde{K}_{\mathrm{u}}^2 B$. Note that in (38) we assumed that $\dot{\mathbf{x}} \neq 0$, i.e., we require $A > 0$ and $\|\mathbf{y}\|_2 \neq 0$.

APPENDIX G

PROOF OF THEOREM 9

For the lower bound in (12), we simply recall the fact that the ℓ_2-norm of the LS solution $\|\hat{\mathbf{x}}\|_2$ is, by definition of the optimization problem (P_2^ε), smaller than the ℓ_2-norm of a democratic representation $\|\dot{\mathbf{x}}\|_2$. Consequently, we get the trivial bound $PI = \|\dot{\mathbf{x}}\|_2^2 / \|\hat{\mathbf{x}}\|_2^2 \geq 1$.

To arrive at the upper bound in (12), we follow closely the proof in Appendix F and individually

bound the numerator and denominator of (12). First, we bound $\|\dot{\mathbf{x}}\|_2^2$ from above as

$$\|\dot{\mathbf{x}}\|_2^2 \leq N\|\dot{\mathbf{x}}\|_\infty^2 \leq \widetilde{K}_{\mathrm{u}}^2(\|\mathbf{y}\|_2 - \varepsilon)^2, \tag{42}$$

which is a consequence of standard norm bounds and the results of (3) and (6). We next obtain a lower bound on the denominator $\|\hat{\mathbf{x}}\|_2^2$ of (12). To this end, we carry out the steps in (37)–(41) for $\|\hat{\mathbf{x}}\|_2^2$ to arrive at the following lower bound:

$$\|\hat{\mathbf{x}}\|_2^2 \geq \frac{\left(\|\mathbf{y}\|_2 - \varepsilon\right)^2}{B}.$$

Finally, combining (42) with (42) yields $PI \leq \widetilde{K}_{\mathrm{u}}^2 B$. We conclude by noting that the steps in (37)–(41) require $\|\mathbf{y}\|_2 \neq 0$, $A > 0$, and $\varepsilon < \|\mathbf{y}\|_2$.

ACKNOWLEDGMENTS

REFERENCES

[1] C. Studer, W. Yin, and R. G. Baraniuk, "Signal representations with minimum ℓ_∞-norm," in *50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Allerton, IL, Oct. 2012, pp. 1270–1277.

[2] Y. Lyubarskii and R. Vershynin, "Uncertainty principles and vector quantization," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3491–3501, Jul. 2010.

[3] J.-J. Fuchs, "Spread represenations," in *Proc. 45th Asilomar Conf. on Signals, Systems, and Comput.*, Pacific Grove, CA, USA, 2011.

[4] A. R. Calderbank and I. Daubechies, "The pros and cons of democracy," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1721–1725, Jun. 2002.

[5] M. Püschel and J. Kovačević, "Real, tight frames with maximal robustness to erasures," in *Proc. IEEE Data Compression Conf. (DDC)*, Mar. 2005, pp. 63–72.

[6] C. Novak, C. Studer, A. Burg, and G. Matz, "The effect of unreliable LLR storage on the performance of MIMO-BICM," in *Proc. of 44th Asilomar Conf. on Signals, Systems, and Comput., Pacific Grove, CA, USA*, Nov. 2010, pp. 736–740.

[7] R. van Nee and R. Prasad, *OFDM for wireless multimedia communications*. Artech House Publ., 2000.

[8] J. Illic and T. Strohmer, "PAPR reduction in OFDM using Kashin's representation," in *Proc. IEEE 10th Workshop on Sig. Proc. Advances in Wireless Comm. (SPAWC)*, Perugia, Italy, June 2009, pp. 444–448.

[9] C. Studer and E. G. Larsson, "PAR-aware large-scale multi-user MIMO-OFDM downlink," *IEEE J. Sel. Areas Comm.*, vol. 31, no. 2, pp. 303–313, Feb. 2013.

[10] H. Jégou, T. Furon, and J.-J. Fuchs, "Anti-sparse coding for approximate nearest neighbor search," *arXiv:1110.3767v2*, Oct. 2011.

[11] J. A. Cadzow, "Algorithm for the minimum-effort problem," *IEEE Trans. Autom. Control.*, vol. 16, no. 1, pp. 60–63, Feb. 1971.

[12] A. S. Deo and I. D. Walker, "Minimum effort inverse kinematics for redundant manipulators," *IEEE Trans. Robotics and Automation*, vol. 13, no. 5, pp. 767–775, Oct. 1997.

[13] Y. Zhang, J. Wang, and Y. Xu, "A dual neural network for bi-criteria kinematic control of redundant manipulators," *IEEE Trans. Robotics and Automation*, vol. 18, no. 6, pp. 923–931, Dec. 2002.

[14] H. Zhang, M. Yan, and W. Yin, "One conition for all: solution uniqueness and robustness of ℓ_1-synthesis and ℓ_1-analysis minimizations," Rice University, CAAM Technical Report 13-10, Apr. 2013.

[15] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[16] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, 1st ed. Springer, Aug. 2010.

[17] C. Studer and R. G. Baraniuk, "Stable restoration and separation of approximately sparse signals," *Appl. Comput. Harmon. Anal.*, in press, 2013.

[18] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[19] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[20] E. J. Candès and M. B. Wakin, "An introdutction to compressive sampling," *IEEE Sig. Proc. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[21] R. G. Baraniuk, M. Davenport, R. A. DeVore, and M. B. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constr. Approx.*, vol. 28, pp. 253–263, Jan. 2008.

[22] D. L. Donoho and J. Tanner, "Precise undersampling theorems," *Proc. of the IEEE*, vol. 98, no. 6, pp. 913–924, June 2010.

[23] O. Mangasarian and B. Recht, "Probability of unique integer solution to a system of linear equations," *Europ. J. of Operational Research*, vol. 214, no. 1, pp. 27–30, 2011.

[24] L. W. Neustadt, "Minimum effort control systems," *J. Soc. Indus. and Appl. Math. Ser*, vol. 1, no. 1, pp. 16–31, Apr. 1962.

[25] B. Kashin, "Sections of some finite dimensional sets and classes of smooth functions," *Izv. Acad. Nauk SSSR*, vol. 41, no. 2, pp. 334–351, 1977.

[26] A. Y. Garnaev and E. D. Gluskin, "On widths of the Euclidean ball," *Soviet Math. Dokl.*, vol. 30, no. 1, pp. 200–204, 1984.

[27] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections and universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, pp. 5406–5425, 2006.

[28] ETSI EN 302 755, "Digital video broadcasting (DVB); frame structure channel coding and modulation for a second generation digital terrestrial television broadcasting system (DVB-T2)," Version 1.3.1, Tech. Rep., Apr. 2012.

[29] D. Seethaler and H. Bölcskei, "Performance and complexity analysis of infinity-norm sphere-decoding," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1085–1105, Mar. 2010.

[30] V. I. Morgenshtern and H. Bölcskei, *A short course on frame theory*, E. Serpedin, T. Chen, and D. Rajan, Ed.   Chapter in Mathematical Foundations for Signal Processing, Communications, and Networking: CRC Press, 2012.

[31] B. Alexeev, J. Cahill, and D. Mixon, "Full spark frames," *Journal of Fourier Analysis and Applications*, vol. 18, no. 6, pp. 1167–1194, 2012. [Online]. Available: http://dx.doi.org/10.1007/s00041-012-9235-4

[32] M. E. Pfetsch and A. M. Tillmann, "The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing," *arXiv:1205.2081v2*, May 2012.

[33] ——, "Chebotarev and his density theorem," *Math. Intelligencer*, vol. 18, no. 26, pp. 26–37, 1996.

[34] J.-J. Fuchs, "Sparsity and uniqueness for some specific under-determined linear systems," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, vol. 5, 2005, pp. v/729–v/732 Vol. 5.

[35] T. Blumensath and M. E. Davies, "Sampling theorems for signals from the union of finite-dimensional linear subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1872–1882, Apr. 2009.

[36] J.-J. Fuchs, Personal communication, 2011.

[37] S. H. Han and J. H. Lee, "An overview of peak-to-average power ratio reduction techniques for multicarrier transmission," *IEEE Wireless Comm.*, vol. 12, no. 2, pp. 1536–1284, Apr. 2005.

[38] S. Boyd and L. Vandenberghe, *Convex Optimization*.   New York, NY, USA: Cambridge Univ. Press, 2004.

[39] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.21," http://cvxr.com/cvx/, Apr. 2011.

[40] T. Goldstein, E. Esser, and R. Baraniuk, "Adaptive Primal-Dual Hybrid Gradient Methods for Saddle-Point Problems," *Available on Arxiv.org (arXiv:1305.0546)*, 2013.

[41] T. Goldstein, B. O'Donoghue, S. Setzer, and R. Baraniuk, "Fast alternating direction optimization methods," UCLA, CAM Technical Report 12–35, 2012.

[42] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, pp. 293–318, 1992.

[43] J. A. Tropp, I. S. Dhillon, R. W. Heath Jr., and T. Strohmer, "Designing structured tight frames via an alternating projection method," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 188–209, Jan. 2005.

[44] M. Mroué, A. Nafkha, J. Palicot, B. Gavalda, and N. Dagorne, "Performance and implementation evaluation of TR PAPR reduction methods for DVB-T2," *Hindawi Intl. J. of Digital Multimedia Broadcasting*, pp. 1–10, Aug. 2010, doi:10.1155/2010/797393.

[45] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Natl. Acad. Sci. USA*, vol. 106, no. 45, pp. 18 914–18 919, Nov. 2009.

[46] R. A. Horn and C. R. Johnson, *Matrix Analysis*.   New York, NY: Cambridge Press, 1985.

[47] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *C. R. Acad. Sci. Paris, Ser. I*, vol. 346, pp. 589–592, 2008.