MINIMAL ANTENNA-SUBSET SELECTION UNDER CAPACITY CONSTRAINT FOR POWER-EFFICIENT MIMO SYSTEMS: A RELAXED $\ell_1$ MINIMIZATION APPROACH

Masahiro Yukawa
RIKEN
BSI, Mathematical Neuroscience Lab.
2-1 Hirosawa, Wako, Saitama, 351-0198 Japan

Isao Yamada
Tokyo Institute of Technology
Dept. Commun. Integrated Systems
S3-60 Tokyo, 152-8550 Japan

ABSTRACT
This paper addresses the minimal subset selection of antennas achieving designated channel capacity. This is one of the most natural approaches to alleviating the power consumption in MIMO systems, while it is a mathematically challenging nonlinearly-constrained sparse optimization ($\ell_0$-norm minimization) problem. We present an efficient algorithmic solution, to this highly combinatorial problem, using convex and differentiable relaxations of the $\ell_0$-norm. The proposed algorithm is based on the hybrid steepest descent method for the subgradient projection operator together with the soft-thresholding technique, minimizing the Moreau envelope of the $\ell_1$-norm subject to the capacity constraint. The simulation results show that the proposed algorithm realizes a near optimal solution to the original nonlinearly-constrained sparse optimization problem.

Index Terms— Antenna selection, MIMO systems, $\ell_1$ minimization, convex optimization

1. INTRODUCTION
Multiple antenna systems, broadly-termed MIMO (multiple-input multiple-output) systems, have given paramount impacts to a wide range of research fields including communications, signal processing, and information theory because of its potential to increase the data rate without additional bandwidth [1, 2]. The gain, however, comes at the price of hardware and signal processing complexity, power consumption etc. [3]. One of the main causes for the complexity-increase is the cost of multiple RF (radio frequency) chains. Antenna selection has been considered as an attractive approach to reduce the hardware complexity without severely losing the advantages of MIMO systems (see [4–7] and references therein). In particular, it has been shown that the antenna selection retains the diversity degree compared to the full-complexity system [4]. The complexity reduction is achieved by equipping fewer RF chains than the antenna elements at the receiver/transmitter, and the same number of antennas as the RF chains are selected so that the achieved channel capacity is maximized.

Differently from the prior works, we consider power-limited systems in which it is desired to consume the minimum amount of power with the designated channel capacity achieved. At the receiver, for instance, each antenna element requires a ‘power-consuming’ RF chain that comprises a low noise amplifier, a frequency down-converter (a mixer), and an analog-to-digital converter. Also the signal processing complexity may seriously increase with the number of antenna elements. Therefore, it would be a natural requirement to select the minimal antenna subset that achieves the designated channel capacity; the cardinality of such a subset depends highly on the channel state, signal to noise ratio (SNR) etc.

In this paper, we address the minimal antenna-subset selection problem under the capacity constraint. For simplicity, we consider the case where the number of RF chains equipped is the same as that of antenna elements. The problem is mathematically challenging because it is nonlinearly-constrained sparse optimization ($\ell_0$-norm minimization). We present an efficient algorithmic solution using convex and differentiable relaxations of the $\ell_0$-norm. The proposed algorithm is based on the hybrid steepest descent method (HSDM) [8] for the subgradient projection operator together with the soft-thresholding technique, minimizing the Moreau envelope of the $\ell_1$-norm subject to the capacity constraint. The simulation results show that the proposed algorithm realizes a near optimal solution to the original nonlinearly-constrained sparse optimization problem.

2. SYSTEM MODEL
For a MIMO system with $N_T$ transmit antennas and $N_R$ receive antennas, the received signal can be represented as

$$ r_i := \sqrt{E_i} g_i s_i + n_i \in \mathbb{C}^{N_R}. $$

Here, $r_i$ represents the ith sample of the signals measured at the $N_R$ receive antennas, $s_i \in \mathbb{C}^{N_T}$ the ith symbol transmitted from the $N_T$ transmit antennas, $E_i > 0$ the average energy at each receive antenna, $G \in \mathbb{C}^{N_R \times N_T}$ the channel matrix whose $(p,q)$th component represents the channel characteristics between the $p$th receive antenna and the $q$th transmit antenna, and $n_i$ the additive white Gaussian noise with energy $N_0/2$ per complex dimension. We make the standard assumptions that the channel has frequency-flat fading and $G$ is perfectly known at the receiver. Also we assume that $G$ is totally unknown at the transmitter, therefore choosing $s_i$ such that its covariance matrix is $I_{N_T}/N_T$ [7]; we denote by $I_m$ the $m \times m$ identity matrix. In this case, it is known that the channel capacity (mutual information) is given as follows [1]:

$$ C_{\text{full}} := \log_2 \det \left( I_{N_T} + \frac{\rho}{N_T} G^H G \right) \text{ bps/Hz}, $$

where $\rho := E_i/N_0$ is the average SNR, $(\cdot)^H$ stands for the Hermitian transposition.

3. MINIMAL ANTENNA-SUBSET SELECTION UNDER CAPACITY CONSTRAINT

3.1. Problem Statement
We focus on the receive antenna selection. Let $C^* \in \{0, C_{\text{full}}\}$ be the designated channel capacity to be ensured. The problem is to select the minimal antenna subset that achieves the capacity $C^*$. Let $x := [x_1, x_2, \ldots, x_{N_R}]^T \in \{0, 1\}^{N_R}$ represent an antenna subset in such a way that $x_j = 1$ ($x_j = 0$) indicates that the $j$th antenna is selected (not selected); $(\cdot)^T$ stands for the transposition. Then, the channel capacity with the antenna subset represented by $x$ is given by

$$ C(x) := \log_2 \det \left( I_{N_T} + \frac{\rho}{N_T} G^H X G \right) \text{ bps/Hz}, $$

where $G^H X G$ represents the channel matrix whose $(p,q)$th component represents the channel characteristics between the $p$th receive antenna and the $q$th transmit antenna, and $n_i$ the additive white Gaussian noise with energy $N_0/2$ per complex dimension. We make the standard assumptions that the channel has frequency-flat fading and $G$ is perfectly known at the receiver. Also we assume that $G$ is totally unknown at the transmitter, therefore choosing $s_i$ such that its covariance matrix is $I_{N_T}/N_T$ [7]; we denote by $I_m$ the $m \times m$ identity matrix. In this case, it is known that the channel capacity (mutual information) is given as follows [1]:

$$ C_{\text{full}} := \log_2 \det \left( I_{N_T} + \frac{\rho}{N_T} G^H G \right) \text{ bps/Hz}, $$

where $\rho := E_i/N_0$ is the average SNR, $(\cdot)^H$ stands for the Hermitian transposition.

1When the multiple antennas are exploited for spatial multiplexing or the space-time trellis codes are adopted, the complexity increases sometimes exponentially [3].

2The channel could be moderately frequency-selective [4, 5].
where \( \mathbf{X} := \text{diag}(\mathbf{x}) \). The minimal antenna-subset selection problem is thus formulated as follows:

\[
\min_{x \in [0,1]^n} \| x \|_0 \quad \text{s.t.} \quad C(x) \geq C^*,
\]

(4)

where \( \| \cdot \|_0 \) denotes the \( \ell_0 \)-norm that counts the number of nonzero components. The problem in (4) is mathematically challenging, because it is nonlinearly-constrained sparse optimization. In general, finding its optimal solution involves exhaustive search. In the following, we present an efficient algorithmic solution using convex and differentiable relaxations of the \( \ell_0 \) norm.

### 3.2. Convex and Differentiable Relaxations

In recent years, it has been proven both theoretically and experimentally that sparse recovery is possible in many cases by means of the \( \ell_1 \)-norm [9, 10]. To alleviate the difficulty in the combinatorial nature of the problem, we reformulate (4) into the following:

\[
\min_{x \in [0,1]^n} \psi(x) := \| x \|_1 \quad \text{s.t.} \quad \varphi(x) := C^* - C(x) \leq 0,
\]

(5)

where \( \| \cdot \|_1 \) denotes the \( \ell_1 \)-norm that sums up the absolute value of each component. Because the function \( C \) is concave on \( \mathbb{R}^n_+ \), \( \varphi \) is convex on \( \mathbb{R}^n_+ \); \( \mathbb{R}_+ \) denotes the set of all nonnegative real numbers. Unfortunately, we can still not find any computationally efficient solver for the reformulated problem in (5) because (i) the function \( \psi \) is (convex but) neither differentiable nor strictly-convex and (ii) the metric projection onto the constraint set (i.e., the zero level set of \( \varphi \)) is not efficiently computable (For instance, the generalized Haugazeau’s scheme [12] cannot be applied directly because of the non-strict-convexity of \( \psi \)). We therefore introduce the Moreau envelope as another relaxation for differentiability, thereby we can use the hybrid steepest descent method.

**Definition 1** Let \( \langle \cdot , \cdot \rangle \) denote the inner product defined on \( \mathbb{R}^n \) as \( \langle \mathbf{x} , \mathbf{y} \rangle := \mathbf{x}^T \mathbf{y}, \mathbf{x} , \mathbf{y} \in \mathbb{R}^n \), and \( \| \cdot \|_2 \), its induced norm (or the \( \ell_2 \) norm). Given a lower semi-continuous and proper convex function \( f : \mathbb{R}^n \rightarrow ( -\infty , \infty ) \), its Moreau envelope of index \( \gamma \in (0, \infty ) \) is defined as follows [13]:

\[
\gamma f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad \mathbf{x} \mapsto \min_{y \in \mathbb{R}^n} \left( f(y) + \frac{1}{2\gamma} \| y - x \|_2^2 \right).
\]

(6)

The function \( \gamma f \) is convex and differentiable with its gradient

\[
\nabla \gamma f(x) = \frac{1}{\gamma} (x - \text{prox}_{\gamma f}(x)),
\]

(7)

where \( \text{prox}_{\gamma f}(x) \in \mathbb{R}^n \) is the unique minimizer of \( g_{\gamma f} : \mathbb{R}^n \rightarrow \mathbb{R}, \mathbf{y} \mapsto f(y) + \frac{1}{2\gamma} \| y - x \|_2^2 \) and \( \text{prox}_{\gamma f} : \mathbb{R}^n \rightarrow \mathbb{R}^n \) is called the proximity operator of index \( \gamma \) of \( f \). An important property is that \( \nabla \gamma f(x) \) is Lipschitz continuous.

Given a closed convex set \( C \subset \mathbb{R}^n \), we denote by \( P_C(x) \) the (Euclidean) metric projection of \( x \in \mathbb{R}^n \) onto \( C \). The projection \( P_C \) can be characterized as the proximity operator of an arbitrary index \( \gamma \in (0, \infty ) \) of the following indicator function:

\[
i_C(x) := \begin{cases} 0 & \text{if } x \in C \setminus C, \\ 1 & \text{if } x \not\in C. \end{cases}
\]

(8)

**Algorithm 1**

1. For an initial vector \( \mathbf{x}_0 \in \mathbb{R}^n \), generate \( \{ \mathbf{x}_k \}_{k=1}^\infty \) recursively by HSDM to solve (9) (\( \mathcal{Q} \): the prespecified number of iterations), and let \( \mathbf{x}_Q := [x_1^{(1)}, x_2^{(1)}, \ldots , x_n^{(1)}]^T \).

2. Compute the arithmetic mean for generality in place of the non-negative \( \mathbf{X} \) offers an approximate solution to (4) by \( \psi \).

3. Choose the indices corresponding to the components no smaller than \( \mathcal{E} \) as a temporary antenna subset.

4. Choose the minimal antenna subset.

**Fig. 1.** The \( \ell_0 \)-norm, the \( \ell_1 \)-norm, and its Moreau envelope of index \( \gamma = 1 \).

In (5), the minimization of \( \psi \) is equivalent to the minimization of \( \psi_\omega : \mathbb{R}^n \rightarrow \mathbb{R}_+, \mathbf{x} \mapsto \omega \| x \|_1 \) for an arbitrary constant \( \omega > 0 \) (\( \psi = \psi_{\omega=1} \)). Therefore, rather than the Moreau envelope of \( \psi \), we use the Moreau envelope of \( \psi_\omega \) for generality in place of the non-differentiable \( \psi \). The Moreau envelope of the \( \ell_1 \)-norm is illustrated in Fig. 1. Thus, our optimization problem to solve is the following:

\[
\min_{x \in [0,1]^n} \gamma \psi_\omega(x) \quad \text{s.t.} \quad \varphi(x) \leq 0.
\]

(9)

### 3.3. Proposed Antenna-Subset Selection Algorithm

Our basic strategy is the following: (i) compute the solution \( \mathbf{x}^* \) to the problem in (9) by HSDM and (ii) choose the antenna subset associated with the indices of (the minimum number of) the largest components of \( \mathbf{x}^* \) such that the designated capacity \( C^* \) is achieved. Letting \( \mathcal{I} := \{ 1, 2, \ldots , N_R \} \), the proposed algorithm is given as below.

1. Set \( \mathcal{J} := \emptyset \).

2. For each \( j \in \mathcal{I} \):

   a. \( \mathcal{J} := \mathcal{J} + \{ j \} \).

3. Choose the indices corresponding to the components no smaller than \( \mathcal{E} \) as a temporary antenna subset.

4. Choose the minimal antenna subset.

5. Output \( \mathcal{J} \) as the selected antenna subset (which is ensured to achieve \( C^* \) because of Step 4).
The following subsection is devoted to explain how to solve the problem in (9) by HSDM.

3.4. Optimization by Hybrid Steepest Descent Method

The problem in (9) has two constraints: the capacity constraint \( x \in \text{lev}_{\leq 0} \varphi \) and the box constraint \( x \in K := [0, 1]^n \) \( \cap \{ x \in \mathbb{R}^n : 0 \leq x_j \leq 1, \forall j \in J \} \). We mention that the projection \( P_{\text{lev}_{\leq 0} \varphi} \) has no closed-form solution, whereas \( P_K \) can be easily computed. Since \( \varphi_1(x_1) = C^* - C_{\text{all}} < 0 \) for \( x_1 \in [1, \ldots, 1] \), we have \( K \cap \text{lev}_{\leq 0} \varphi \neq \emptyset \). Thus, the problem in (9) is equivalent to the following variational inequality problem [8]: find \( x^* \in K \cap \text{lev}_{\leq 0} \varphi \) such that

\[
\langle x - x^*, \nabla \varphi(x) \rangle \geq 0, \quad \forall x \in K \cap \text{lev}_{\leq 0} \varphi.
\]  

This problem can be solved by the hybrid steepest descent method.

**Proposition 1 (Hybrid Steepest Descent Method) [8, Proposition 6]**

Let \( T_{x_0}(f) : \mathbb{R}^n \rightarrow \mathbb{R}^n \) denote the subgradient projection relative to \( \varphi \) (see below), and define

\[
\hat{T}_{x_0} := P_K \left[ \left( 1 - \alpha \right) I + \alpha T_{x_0}(f) \right] \quad \alpha \in (0, 2),
\]

where \( I : \mathbb{R}^n \rightarrow \mathbb{R}^n \) represents the identity mapping. Then, for any \( x_0 \in \mathbb{R}^n \) and \( (\lambda_k)_{k \geq 1} \subset \mathbb{R}_+ \) satisfying (i) \( \lim_{k \rightarrow \infty} \lambda_k = 0 \) and (ii) \( \sum_{k=1}^\infty \lambda_k = \infty \), \( (x_k)_{k \geq 0} \) generated by

\[
x_{k+1} := \left( I - \lambda_{k+1} \nabla \varphi(x_k) \right) \hat{T}_{x_k}(x_k), \quad k \geq 0, \quad \alpha \in (0, 2),
\]

converges to the solution to the variational inequality problem shown above.

**Definition 2** For a continuous (but not necessarily differentiable) convex function \( f : \mathbb{R}^n \rightarrow \mathbb{R} \), a vector \( \bar{x} \in \mathbb{R}^n \) is said to be a subgradient of \( f \) at \( x \in \mathbb{R}^n \) if \( (z - x, \bar{x}) + f(x) \leq f(z), \forall z \in \mathbb{R}^n \). The set of all subgradients of \( f \) at \( x \) is denoted by \( \partial f(x) \). Suppose that \( \text{lev}_{\leq 0} f := \{ x \in \mathbb{R}^n : f(x) \leq 0 \} \neq \emptyset \). Then, the mapping \( T_{x_0}(f) : \mathbb{R}^n \rightarrow \mathbb{R}^n \) defined as

\[
T_{x_0}(f)(x) := \begin{cases} x & \text{if } f(x) > 0 \\ x - \frac{f(x)}{\| f(x) \|_2} & \text{otherwise} \end{cases}
\]

is called subgradient projection relative to \( f \), where \( f'(x) \in \partial f(x) \).

Since \( \varphi \) is differentiable on \( \mathbb{R}^n \), its gradient \( \nabla \varphi(x) := \left[ \frac{\partial \varphi(x)}{\partial x_1}, \frac{\partial \varphi(x)}{\partial x_2}, \ldots, \frac{\partial \varphi(x)}{\partial x_n} \right]^T \) is the unique subgradient at any \( x \in \mathbb{R}^n \); i.e., \( \partial \varphi(x) = \{ \nabla \varphi(x) \} \). Letting \( G^\varphi := [g_1, g_2, \ldots, g_n] \), we have

\[
I_{N_T} + \frac{\rho}{N_T} G^\varphi X G = I_{N_T} + \sum_{j=1}^{N_T} x_j \left( \frac{\rho}{N_T} g_j g_j^H \right),
\]

which is positive definite. Therefore, \( \forall x \in \mathbb{R}^n, \forall j \in I \), we have

\[
\frac{\partial \varphi(x)}{\partial x_j} = -\frac{1}{\ln 2} \left( I_{N_T} + \frac{\rho}{N_T} G^\varphi X G \right)^{-1} \frac{\rho}{N_T} g_j g_j^H = -\frac{\rho}{N_T \ln 2} g_j ^H \left( I_{N_T} + \frac{\rho}{N_T} G^\varphi X G \right)^{-1} g_j.
\]

where \( I_n \) stands for the trace of matrix. Note that, since \( \left( I_{N_T} + \frac{\rho}{N_T} G^\varphi X G \right)^{-1} g_j > 0, \forall j \in I \), this is the soft-thresholding algorithm is given approximately by

\[
\text{pro}_{\gamma, \omega}(x) := \text{sign}(x) \max\{ |(x, e_j)| - \gamma, 0 \} e_j,
\]

where \( \gamma, \omega \) are positive constants. Then, the upper bound for the soft-thresholding technique proposed for denoising in [14].

**Remark 1** The operator \( I - \lambda_{k+1} \nabla \varphi \) in (12) can be written as

\[
I - \lambda_{k+1} \nabla \varphi = I - \lambda_{k+1} \nabla \varphi + \lambda_{k+1} \nabla \varphi - \lambda_{k+1} \nabla \varphi,
\]

which is the soft-thresholding algorithm is given approximately by

\[
\text{pro}_{\gamma, \omega}(x) := \text{sign}(x) \max\{ |(x, e_j)| - \gamma, 0 \} e_j.
\]

4. SIMULATION RESULTS

Simulations are performed to show the efficacy of the proposed minimal antenna-subset selection algorithm. We consider the Rayleigh channel where the elements of \( G \) are independently drawn from a complex zero-mean Gaussian distribution. For all the simulations, the HSDM parameters are set to \( \alpha = 1, \gamma = 1.2, \omega = 0.8, Q = 20, \) and \( \lambda_k = 1 (k = 1, 2, \ldots, Q) \). All the simulated points are calculated by averaging over 2000 independent realizations of the channel matrix \( G \).

First, Fig. 2 depicts results for \( N_T = 16, N_R = 4, \) and \( C^* = 10, 20 \). Figure 2a describes the average number \( \bar{L}_H \) of antennas selected by the proposed algorithm. As a reference, we also plot the optimal solution to the original problem in (4); the optimal is computed by computationally-exhaustive full search. It is seen that the results of the proposed algorithm are comparable to the optimal; this suggests the reasonability of the relaxations introduced in Section 3.2. Figure 2b describes the ergodic capacity of the proposed algorithm. With \( L_H \) denoting the number of antennas selected by the proposed algorithm, we also plot \( C_{\text{max}}(L_H) \), the maximum achievable capacity with the subset of \( L_H \) antennas, which is computed by exhaustive search. It is seen that the performance of the proposed algorithm is approximately the same as \( C_{\text{max}}(L_H) \); this is the side effects of the proposed algorithm.

Second, Fig. 3 illustrates the results for (a) \( N_R = 16, N_T = 16, \) \( C^* = 20, 40 \) and (b) \( N_R = 64, N_T = 16, 64, \) and \( C^* = 60, 120 \). From Fig. 2a and Fig. 3, it is seen that the number of antennas to be used can significantly be reduced particularly for high SNR. Moreover, in Fig. 3b, we observe no distinct difference between \( N_T = 16 \) and \( N_T = 64 \) for SNR higher than 15 dB. Finally, Fig. 4 plots \( L_R \) against \( N_R \) for \( N_T = 30, 40, 50 \), and \( C^* = 20. \) The result shows that an increase of the number of antenna elements equipped could yield reduction of the number of antennas used.
Fig. 2. Comparisons with the optimal selection for $N_R = 16$, $N_T = 4$, and $C^* = 40, 20$.

5. CONCLUSION

This paper has investigated the minimal antenna-subset selection problem, which is of great importance in communications but, at the same time, mathematically challenging nonlinearly-constrained sparse optimization. We have presented an efficient algorithm based on the hybrid steepest descent method, which has been employed to minimize the Moreau envelope of the $t_1$-norm subject to the designated capacity constraint. The proposed algorithm is composed of the subgradient projection and the soft-thresholding technique; the former is responsible for the capacity constraint and the latter promotes sparsity. The simulation results have shown that the proposed algorithm has realized (i) the near-minimal antenna subset and (ii) the near-maximum capacity achievable with the same number of antennas as selected by the algorithm.

6. REFERENCES