Correntropy Induced Joint Power and Admission Control Algorithm for Dense Small Cell Network

Zhirong Luan¹, Hua Qu¹, Jihong Zhao¹, Badong Chen¹, Jose. C. Principe²

¹School of Electronic and Information Engineering, Xi’an Jiaotong University, Xi’an, 710049, P.R.China.
²Computational NeuroEngineering Laboratory, Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL. 32611 USA.
E-mail: luanzhirong@stu.xjtu.edu.cn

Abstract: We consider the joint admission and power control problem in a dense small cell network, which contains multiple interference links. The goal is to mainly maximize the number of the admitted links, and at the same time minimize the transmit power. We formulate the admission control and power control problem as a joint optimization problem, which is however NP-hard. Such NP-hard problem can be relaxed to a p-norm problem (0<p<1) by using the correntropy induced metric. The correntropy is a novel nonlinear similarity measure, which has been successfully used in the robust and spares signal processing, especially when the data contain large outliers. Thus, in this work we propose a new correntropy induced joint power and admission control algorithm (CJPA). To achieve a faster convergence speed, we also propose an adaptive kernel size method, in which the kernel size is determined by the error so that the convergence speed is the fastest during the iterations. Simulation results show that the proposed approach can achieve much better results than the existing works.

1. Introduction

The increasing demands on wireless data transmission create explosive requirement for the wireless network capacity [1-3]. The dense small cell network (DSCN) is envisioned as one of the key solutions for enhancing network capacity [4]. The DSCN is composed of mixed lower power base stations (BS), i.e. femtocells and picocells. Users will enjoy better wireless services due to the short transmission distance [5]. However, the high spectrum cost requires intense spectrum reuse [4]. The dense deployment of low power BSs will significantly increase the inter-cell interference. To increase the network capacity, it is necessary to management the interference.

Power control is an effective tool for inter-cell and intra-cell interference management [6] [7]. Originally, to support the requirement of wireless services, the power control problem is formulated to ensure a certain value of the signal to interference plus noise ratio (SINR). However, in DSCN, the strong interference may lead to the problem that not all links can be supported simultaneously regardless of power control used [8]. This brings the issue of the admission control (or link removal). The goal of the admission control is to maximize the number of the admitted links. Since the decoupling of the power control and admission control may remove unnecessary links, it is preferable to combine the optimization processes of these two problems [9]. Such combined optimization process can search the maximum admitted link solution and determine which link to be turned off [9].
In [10], the joint power and admission control protects the active users in the network. In [11], the algorithm optimize the ratio between the number of the users and the transmit power to maximize the overall throughput. But these algorithms are far from maximizing the number of the admitted links in the network [8]. In [8], a linear programming deflation algorithm is proposed to approximate the joint power and admission control problem, which is NP-hard. And its solution can be used to iteratively remove the interfering links. The algorithm will keep removing the links until all the remaining links can be simultaneously supported.

A recent work [9] proposes a linear programming algorithm for the joint power and admission control problem. The joint problem is first formulated to minimize a sparse $l_0$-norm problem. Since it is NP-hard, this $l_0$-norm minimization problem is then relaxed to an $l_1$-norm convex problem, which can be solved by linear programming. The solution guides the link removal procedure just as [8] does.

The proposed algorithm in this work is closely related to [9], where an $l_0$-minimization problem is formulated to solve the joint power and admission control problem, which is NP-hard. Then the $l_0$-norm problem is relaxed for effectively searching the optimal. The main contributions of this paper are basically two. 1) We adopt the correntropy [12][13][14][15] to formulate the joint power and admission control problem. Recent studies [16][17] show that the correntropy has an excellent performance for approximating the $l_0$-norm problem. 2) We also develop a new adaptive kernel size method to improve the convergence speed and compute the kernel size using an efficient fixed-point iteration. The solution of correntropy cost function can be used to iteratively remove the strong interference links. Simulation results show that the proposed algorithm can support more links than the existing approaches in [9].

The rest of this paper is organized as follows. Section II introduced the system model and the admission and power control problem. In Section III, we first formulate the NP-hard joint optimization problem. And then we relax the problem with a correntropy induced cost function, which can be solved by the gradient descent algorithm. We also propose the adaptive kernel size method for fast convergence speed. Simulation results are provided in Section IV. Conclusion is given in Section V.

2. System model
We adopt the same system model as [9]. Consider $K$ SISO interference links, denoted by $L_i (i \in \{1, 2, \ldots, K\})$, in a DSCN, where the transmitters are placed randomly and the receivers’ locations are placed randomly around each in a small disc area [9]. The signal to interference plus noise ratio (SINR) of the $i^{th}$ user is defined as
\[
SINR_i = \sum_{j \neq i} \frac{p_j g_{i,j}}{p_j g_{i,j} + \sigma^2},
\]

(1)

where \(p_i\) is the transmit power of the link \(i\), and \(j\) denotes index of other links in the network, \(\sigma^2\) is the noise power and \(g_{i,j}\) is the path gain from the transmitter of link \(j\) to the receiver of the link \(i\). Denote the set of the admitted links’ transmit powers by \(\mathbf{P}\) and the set of admitted links by \(S\). The joint power and admission control problem can be formulated as a two-stage problem [9]. The first stage is the admission control, which maximizes the number of admitted links:

\[
\begin{align*}
\max \quad |S| \\
\text{s.t.} \quad & \quad \text{SINR}_i \geq \gamma_i, \\
& \quad 0 \leq p_i \leq p_i^{\text{max}} \quad i \in S
\end{align*}
\]

(2)

where \(|S|\) is the number of links in \(S\); \(\gamma_i\) is the SINR’s lower bound for \(L_i\) and \(p_i^{\text{max}}\) is maximum transmit power of \(L_i\).

The second stage is the power control, which minimizes the total transmit power:

\[
\begin{align*}
\min \quad & \quad \sum_{i \in S} p_i \\
\text{s.t.} \quad & \quad \text{SINR}_i \geq \gamma_i, \\
& \quad 0 \leq p_i \leq p_i^{\text{max}} \quad i \in S \subseteq N
\end{align*}
\]

(3)

Note that, the first stage optimization should be given priority since improving the network capacity is more important than saving transmit power in this problem.

3. Correntropy induced joint power control and admission control algorithm (CJPA)

According to (1), SINR is a nonlinear function of the transmit powers. To facilitate the transmission power updating in the proposed CJPA, we first linearize the constraint. In particular,

\[
\begin{align*}
\text{SINR}_i \geq \gamma_i, \\
\Rightarrow p_i g_{i,i} - \gamma_i \sum_{j \neq i} p_j g_{i,j} - \gamma_i \eta_i \geq 0
\end{align*}
\]

(4)

Since \(g_{i,j} > 0\), we have

\[
p_k - \gamma_k \sum_{j \neq k} p_j g_{k,j} g_{k,k}^{-1} \gamma_i \eta_k \geq 0.
\]

(5)

We define a matrix \(\mathbf{A} \in \mathbb{R}^{K \times K}\) by

\[
a_{i,j} = \begin{cases} 
1, & \text{if } i = j, \\
\gamma_i g_{i,j} g_{i,j}^{-1}, & \text{if } i \neq j.
\end{cases}
\]

(6)

where \(a_{i,j}\) is the \((i, j)\)-th entry of \(\mathbf{A}\).
And we define a vector \( \mathbf{c} \in \mathbb{R}^{K \times 1} \) whose \( i \)-th entry is
\[
   c_i = \frac{y_i \eta_i}{g_{i,j}} \quad (7)
\]

Then, the problems (2) and (3) becomes
\[
\begin{align*}
\max & \quad |S| \\
\text{s.t.} & \quad \sum_{i \in S} a_{k,j}p_i - c_k \geq 0 \\
& \quad 0 \leq p_i \leq p_i^{\max} \quad i \in S \subseteq N
\end{align*} \quad (8)
\]

and
\[
\begin{align*}
\min & \quad \sum_{i \in S} p_i \\
\text{s.t.} & \quad \sum_{i \in S} a_{k,j}p_i - c_k \geq 0 \\
& \quad 0 \leq p_i \leq p_i^{\max} \quad i \in S \subseteq N
\end{align*} \quad (9)
\]

Note that [9] also transform the constraints of the formulated problem, but the definition of normalized channel matrix (also denoted by \( \mathbf{A} \)) in [9] is totally different from the \( \mathbf{A} \) in our work and the purpose is different. We want to get a linear constraint but [9] want to normalize the constraint. Although there are differences, the properties of these two definitions are similar. In particular, \( \mathbf{A} \) is a square matrix, whose diagonal entries are 1 and other entries are non-positive.

3.1 \( l_0 \)-norm minimization problem formulation

As has been proved in [9], the above (8) and (9) can be reformulated as a signal stage optimization problem as follow (Please refer to [9] for details.)
\[
\begin{align*}
\min & \quad |X|_0 + \alpha \sum_{i \in S} p_i \\
\text{s.t.} & \quad X = \mathbf{A}P - \mathbf{C} \\
& \quad 0 \leq P \leq P^{\max}
\end{align*} \quad (10)
\]

where \( |X|_0 \) is the \( l_0 \)-norm of \( X \) and \( \alpha \) is a weight value. The goal of the joint optimization problem is to maximize the number of admitted links. When there are more than one maximum admissible set, we will pick the one with the minimum total transmit power by the second term of the cost in (10). Since \( |X|_0 \) has priority and it is a integer, \( \alpha \) should satisfy
\[
\alpha \sum_{i \in S} p_i < 1 \\
\Rightarrow \alpha < \frac{1}{\sum_{i \in S} p_i},
\]

such that the solution is unique.

### 3.2 Correntropy induced cost function for \(l_0\)-norm approximation

The \(l_0\)-norm problem (10) is NP-hard [9]. We consider using the correntropy [12] to relax the problem, which approximates \(l_0\)-norm problem accurately when the kernel size is close to zero. Correntropy between two variables \(Y_i = \{y_{1,i}\}_{i=1}^{N}\) and \(Y_j = \{y_{2,j}\}_{j=1}^{N}\) is a measure defined as [12]

\[
\hat{v}(Y_i, Y_j) = E[\mathbf{\kappa}_\sigma(Y_i - Y_j)] = \frac{1}{N} \sum_{i=1}^{N} \mathbf{\kappa}_\sigma(y_{1,i} - y_{2,i}),
\]

where \(\mathbf{\kappa}_\sigma(y_{1,i} - y_{2,i})\) is the Gaussian kernel function defined as

\[
\mathbf{\kappa}_\sigma(y_{1,i} - y_{2,i}) = \exp\left(\frac{-(y_{1,i} - y_{2,i})^2}{2\sigma^2}\right),
\]

and \(\sigma\) is the kernel size. Thus, correntropy is a measurement of how similar two variables are with respect to the kernel size \(\sigma\). Assume that \(\sigma\) is small, \(\mathbf{\kappa}_\sigma(y_{1,i} - y_{2,i}) \approx 1\) only when \(y_{1,i}\) is very close to \(y_{2,i}\), otherwise \(\mathbf{\kappa}_\sigma(y_{1,i} - y_{2,i}) \approx 0\). Thus when \(\sigma \to 0\), \(N[1 - \hat{v}(Y_i, Y_j)]\) can be used to approximate \(l_0\)-norm problem \(|Y_i - Y_j|_0\).

Correntropy approximates the \(l_0\)-norm only when the kernel size approaches 0. But a too small kernel will lead to very slow convergence speed (this can be solved by the adaptive kernel size method in the following) and large gradient when approaching optimal, which may cause fluctuation. So we do not directly adopt the Correntropy with \(\sigma \to 0\). To get better results when the kernel size is larger, inspired by the c-loss function in [16], we introduce a positive scaling constant to modify the cost function, which is defined as

\[
\beta = \frac{1}{1 - \exp\left(-\frac{1}{2\sigma^2}\right)}.
\]

We defined \(e = y_{1,i} - y_{2,i}\). Fig.1 plot the functions \(f_1(e) = 1 - \mathbf{\kappa}_\sigma(e)\) and \(f_2(e) = \beta[1 - \mathbf{\kappa}_\sigma(e)]\) with different kernel sizes. When kernel is small, i.e. \(\sigma = 0.1\), \(f_1(e)\) and \(f_2(e)\) almost overlaps. But when the kernel size is a little larger, i.e. \(\sigma = 1\), \(\beta\) increases the gradient of the cost function. As a result, if \(f_1(e)\)
and $f_i(e)$ have the same error, the value of $f_2(e)$ is closer to 1. And this means that $f_2(e)$ has a better performance for approximating the $l_0$-norm. In addition, $f_2(e)$ behaves like $l_2$-norm when $\sigma \to +\infty$ [16].

Property 1. Assume $e = y_{1,i} - y_{2,i}$. The correntropy induced $l_0$-norm approximation problem with respect to the error, i.e. $f_1(e) = 1 - \kappa_\sigma(e)$ or $f_2(e) = \beta[1 - \kappa_\sigma(e)](f_2(e) = \beta f_1(e)$ and $\beta$ is a constant to $e$), is a quasi-convex optimization problem with any given finite positive value of the kernel size $\sigma$. [17]

Proof. With (13), $f_1(e) = 1 - \exp\left(-\frac{e^2}{2\sigma^2}\right)$. $f_1(e)$ reaches the minimum value 0 when $e=0$. Meanwhile, for $e<0$, $f_1(e)$ is nonincreasing, and for $e>0$, $f_1(e)$ is nondecreasing. According to the necessary and sufficient conditions of quasi-convex on P.99 in [18], $f_1(e)$ is quasi-convex, as well as $f_2(e)$.

The quasi-convexity of the sub-function cannot guarantee the quasi-convexity of the cost function, which is the sum of the quasi-convex sub-function. However, we can design a gradient descent to reach the global with a sufficient small step size. [17]

Then, the modified correntropy cost function for approximating (10) is formulated as

$$ F(P, \sigma) = K \beta [1 - \tilde{v}(AP, C)] + \alpha \sum_{i \in S} p_i $$

s.t. $0 \leq P \leq P_{\text{max}}$ \hfill (15)

where $\sigma = \{\sigma_1, \ldots, \sigma_{|\mathcal{S}|}\}$ is the set of the kernel size of all the links. To be more specific, (15) can be written as

$$ F(P, \sigma) = \beta \sum_{i \in S} [1 - \kappa_{\sigma_i} \left( \sum_{j \in S} a_{i,j} p_j - c_i \right)] + \alpha \sum_{i \in S} p_i $$

s.t. $0 \leq P \leq P_{\text{max}}$ \hfill (16)
where $\beta_i$ is the positive scaling constant of link $i$. With (15) and (16), we transform a mixed-integer NP-Hard problem in a continuously differentiable problem.

### 3.3 Adaptive kernel size gradient decent searching method

Since the modified correntropy induced cost function is always a smooth function, we consider searching the optimal by the gradient decent method. The gradient the cost function with respect to the power is

$$\frac{dF(P, \sigma)}{dp_i} = \sum_{j \in S} \beta_j \exp\left(-\frac{e_j^2}{2\sigma_j^2}\right) \frac{e_j a_{ij}}{\sigma_j} + \alpha,$$  

(17)

where $e_j = \sum_{j \in S} a_{i,j} p_j - c_i$. Then the adaptive process of the transmit power is

$$p_{i,t+1} = p_{i,t} + \mu\left(-\frac{dF(P, \sigma)}{dp_i}\right).$$  

(18)

From (17), when the power is small but $\alpha$ is relatively big, the step size may be too large and causes fluctuation. Meanwhile, when transmit power is very small, keeping reducing the transmit power makes little contribution for decreasing the cost function. So a lower bound for transmit power is necessary, which is denoted by $P_{\text{min}} = \{p_{t,\text{min}}, \ldots, p_{[t]_{\text{min}}}\}$.

As fig.1 illustrated, when the error is large and the kernel is small, the gradient will be very small, which requires too many steps for searching the optimum. On the other hand, a large kernel size will result in a faster converging process. But the performance for approximating $l_0$-norm becomes worse. Considering both the convergence speed and the $l_0$-norm approximation performance, we propose an adaptive kernel size gradient descent searching algorithm.

The cost function can be treated as the sum of $|S|$ sub-cost functions, which is defined as

$$f_i(P, \sigma_i) = \beta_i \left[ 1 - \exp\left(\frac{-e_i^2}{2\sigma_i^2}\right) \right] + \alpha p_i.$$  

(19)

Then

$$F(P, \sigma) = \sum_{i \in S} f_i(P, \sigma_i),$$  

(20)

Property 2. Both the cost function $F(P, \sigma)$ and the sub-cost $f_i(P, \sigma_i)$ function are invex for any given finite positive kernel size $\sigma_i$. [17]
The proof can be found in [17].

Remark 1. If there exists $P^*$ such that $\frac{dF(P, \sigma)}{dP^*} = 0$, then $P^*$ is the global optimal solution of problem (16).

Assume $e_i = \sum_{j \in S} a_{i,j} p_j - c_i$, we have

$$f_i(e_i, \sigma_i) = \beta [1 - \exp(\frac{-e_i^2}{2\sigma_i^2})] + \alpha p_i.$$  \hspace{1cm} (21)

And (17) becomes

$$\frac{dF(P, \sigma)}{dp_j} = \sum_{j \in S} \frac{df_j(e_j, \sigma_j)}{dp_j} + \alpha = \sum_{j \in S} \beta \frac{df_j(e_j, \sigma_j)}{de_j} \frac{de_j}{dp_j} + \alpha$$ \hspace{1cm} (22)

According to (14), since $\beta_j$ is a function of $\sigma_j$, $\beta_j$ is unique for each link as well as $\sigma_j$. Regardless of $\alpha$ (which is a constant in the gradient), we first study the property of

$$\hat{f}_i(e_i, \sigma_i) = \beta [1 - \exp(\frac{-e_i^2}{2\sigma_i^2})].$$ \hspace{1cm} (23)

Fig.2 plots the $\hat{f}_i(e_i, \sigma_i)$ with respect to $\sigma_i$ and $e_i$ in the 3D space. When the error changes from 2 to 0, for a large kernel size, the norm of the gradient is large at first and then becomes small (from arrow 1 to arrow 3). Meanwhile, for a small kernel size, the norm of the gradient is almost 0 at first and then becomes large (from arrow 2 to arrow 4). Then we fix the error to be a small value, i.e. $e=0.5$. By reducing the kernel size from 2.5 to 0.01, we found that the norm of the gradient first increases. It will reach a maximum value and then reduces. According to this, we conclude that there exists a best kernel size that brings the fastest convergence speed for a fixed error. If the initial kernel size is large, with the reduction of the error, the kernel size should also be reduced to improve the convergence speed. We can search the best kernel size with the gradient method.

We take the gradient of $\frac{dF(p_i, \sigma_i)}{dp_i}$ with respect to $\sigma_i$ to find the best kernel size of a fixed error.

$$d \frac{df(p_i, \sigma_i)}{dp_i d\sigma_i} = d \beta \exp(\frac{-e_i^2}{2\sigma_i^2}) \frac{e_i}{\sigma_i} d\sigma$$

$$\frac{dF(p_i, \sigma_i)}{dp_i d\sigma_i} = \frac{e \exp(\frac{-e_i^2}{2\sigma_i^2})}{\sigma_i^3 (1 - \exp(\frac{-1}{2\sigma_i^2}))} \frac{1}{1 - \exp(\frac{-1}{2\sigma_i^2})} \sigma_i^2 + (\frac{e_i^2}{\sigma_i^2} - 2)$$ \hspace{1cm} (24)
Fig. 2 Correntropy induced sub cost function

where $e_j = \sum_{j \in S_i} a_{i,j} p_j - c_j$. Since 
\[
e^{\frac{-e_j^2}{2\sigma_i^2}} - \frac{1}{\sigma_i^3 (1 - \exp(-1/2\sigma_i^2))} > 0, \text{ if } \frac{d}{dp_i d\sigma_i} df(p_i, \sigma_i) = 0, \text{ then }
\]
\[
e^{\frac{-1}{2\sigma_i^2}} + \frac{1}{1 - \exp\left(-\frac{1}{2\sigma_i^2}\right)} = 0 \Rightarrow \frac{\exp\left(-\frac{1}{2\sigma_i^2}\right)}{1 - \exp\left(-\frac{1}{2\sigma_i^2}\right)} + e_i^2 = 2\sigma_i^2 \tag{25}
\]

(note that \(d \frac{dF(P, \sigma)}{dp_i d\sigma_i} = d \frac{df(p_i, \sigma_i)}{dp_i d\sigma_i}\) because \(\sum_{j \in i} \beta_j \frac{-e_j^2}{2\sigma_j^2} \frac{e_j a_{i,j}}{\sigma^2_j} + \alpha\) in \(F(P, \sigma)\) is a constant to \(\sigma_i\)

Property 3. The kernel size in Eq. (25) has a solution only when \(e_i^2 < 0.5\). If \(e_i^2 \geq 0.5\), the kernel size for fastest convergence speed will be positive infinity.

Proof:

Assume

\[
h(\sigma_i) = \frac{\exp\left(-\frac{1}{2\sigma_i^2}\right)}{1 - \exp\left(-\frac{1}{2\sigma_i^2}\right)} + e_i^2 - 2\sigma_i^2, \tag{26}
\]

where \(e_i^2\) is a constant to \(h(\sigma_i)\) and \(h(\sigma_i)\) will always be positive if \(e_i^2 > \min_{\sigma_i} \frac{\exp\left(-\frac{1}{2\sigma_i^2}\right)}{1 - \exp\left(-\frac{1}{2\sigma_i^2}\right)} - 2\sigma_i^2\).
Fig. 3 plots the $h(\sigma_i)$ for different error ($e_i = 0.4$ and 1). From fig. 3, $h(\sigma_i)$ is monotonic decreasing.

Thus, \[
\frac{\exp\left(-\frac{1}{2\sigma_i^2}\right)}{1 - \exp\left(-\frac{1}{2\sigma_i^2}\right)} - 2\sigma_i^2 \text{ can reach its minimum value when the kernel size is positive infinity.}
\]

In Fig. 3, we can observe the behavior of $h(\sigma_i)$ for different error (error=0.4 and 1).

Then we compute the minimum value of \[
\frac{\exp\left(-\frac{1}{2\sigma_i^2}\right)}{1 - \exp\left(-\frac{1}{2\sigma_i^2}\right)} - 2\sigma_i^2,
\]
which can be achieved when \[\sigma_i \to +\infty.\]

\[
\frac{\exp\left(-\frac{1}{2\sigma_i^2}\right)}{1 - \exp\left(-\frac{1}{2\sigma_i^2}\right)} - 2\sigma_i^2 = \frac{\exp\left(-\frac{1}{2\sigma_i^2}\right) - 2\sigma_i^2 \left(1 - \exp\left(-\frac{1}{2\sigma_i^2}\right)\right)}{1 - \exp\left(-\frac{1}{2\sigma_i^2}\right)}
\]

Then, we assume \[g_1 = \exp\left(-\frac{1}{2\sigma_i^2}\right) - 2\sigma_i^2 \left(1 - \exp\left(-\frac{1}{2\sigma_i^2}\right)\right)\]
and \[g_2 = 1 - \exp\left(-\frac{1}{2\sigma_i^2}\right)\].

Using L'Hospital rule,
\[
\lim_{\sigma_i \to \infty} \frac{g_1(\sigma_i)}{g_2(\sigma_i)} = \lim_{\sigma_i \to \infty} \left(-1 + \frac{4\left(1 - \exp\left(-\frac{1}{2\sigma_i^2}\right)\right) + 2\left(-\exp\left(-\frac{1}{2\sigma_i^2}\right)\right) \frac{1}{\sigma_i^2}}{-\frac{1}{\sigma_i^2} \exp\left(-\frac{1}{2\sigma_i^2}\right)}\right)
\]
Then assume

\[ g_3(\sigma) = 4(1 - \exp(-\frac{1}{2\sigma^2})) + 2(-\exp(-\frac{1}{2\sigma^2})) \frac{1}{\sigma^2} \]

\[ g_4(\sigma) = -\frac{1}{\sigma^3} \exp(-\frac{1}{2\sigma^2}) \]

Using L'Hospital rule again

\[ \lim_{\sigma \to \infty} \frac{g_3'(\sigma)}{g_4'(\sigma)} = \lim_{\sigma \to \infty} \frac{\frac{2}{\sigma^2}(-\exp(-\frac{1}{2\sigma^2}))}{\frac{1}{\sigma^3} \exp(-\frac{1}{2\sigma^2}) + \frac{4}{\sigma^3} \exp(-\frac{1}{2\sigma^2})} = 0.5 \]

\[ \exp(-\frac{1}{2\sigma^2}) \]

\[ \Rightarrow \min_{\sigma} \left( \frac{-1}{\sigma^2} - 2\sigma^2 \right) = -1 + 0.5 = -0.5 \]

So \( h(\sigma) = 0 \) has a solution only when \( e_i^2 < 0.5 \). And when \( e_i^2 \geq 0.5 \), \( h(\sigma) \) is always positive, which means that \( \frac{df(p_i, \sigma)}{dp_i} \) is monotonic increasing with respect to \( \sigma \). Thus, the \( \frac{df(p_i, \sigma)}{dp_i} \) will achieve the largest value when \( \sigma_i \to +\infty \).

Property 4. The feasible kernel size is a monotonically increasing function with respect to the error when \( e_i^2 < 0.5 \).

From fig.3, when \( e \) is increased, the whole curve of \( h(\sigma) \) will move upward. Thus the solution of \( h(\sigma) = 0 \) will become larger if the error is larger.

When \( e_i^2 \leq 0.5 \), we treat \( \sigma_i^2 \) as a variable and use the fixed-point iteration to solve (25). In the simulations, the fix-point iteration converges in at most 10 steps. And it can be found that, a smaller error requires a smaller kernel size. And when \( e_i^2 \geq 0.5 \), let \( \sigma_i \to +\infty \) for fast convergence speed, as stated above, \( \hat{f}(e_i) \) will behave like the \( l_2 \)-norm, which can be denoted as

\[ \hat{f}(e_i) = e_i^2 = (\sum_{j \in S} a_{i,j}p_j - c_i)^2 \]  

(28)

The gradient of (21) becomes

\[ \frac{df_j(e_i)}{dp_i} = 2e_ja_{j,i} + \alpha \]  

(29)
And when \( e_j^2 < 0.5 \),

\[
\frac{df_j(e_j)}{dp_i} = \beta_j \left( \frac{-e_j^2}{2\sigma_j^2} \right) \frac{e_j a_{j,i}}{\sigma_j^2} + \alpha
\]

(30)

From (30), the step size will become large when kernel size is small. To prevent the fluctuation, we define a lower bound \( \sigma_{\text{min}} \) for the kernel size.

Then we can update the transmit power by (18) and (22). The correntropy is a measure of how similar two variables are with respect to the kernel size. The correntropy induced cost function will converge when the error is much smaller than the kernel size, but not when the error is zero. In our problem, the results of the SINRs of the links will converge to a small range around the target SINR \((\gamma_i - \Delta\gamma, \gamma_i + \Delta\gamma)\), where \( \Delta\gamma \) is defined as the allowable error. We assume that the link can be supported when \( \text{SINR} \geq \gamma_i - \Delta\gamma \). And the result will be feasible we set \( \Delta\gamma \) to be a very small value. Here, we set \( \Delta\gamma \) to be 0.002.

The solution of (10) can be used to guide the link removal procedure. With these solutions we know if all the links are supported. Meanwhile, we can use the efficient link removal strategy in [19]. If not all the link can be supported, link \( k_0 \) will be dropped by the following rule

\[
k_0 = \arg \max_{k \in K} \left\{ \sum_{j \neq k} g_{j,k}P_k + \sum_{j \neq k} g_{k,j}P_j \right\}
\]

(31)

To be more specific, the proposed CJPA algorithm is concluded in Algorithm 1.

**Algorithm 1 CJPA**

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>For each link ( i )</td>
</tr>
<tr>
<td>Loop</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Repeat until all SINRs converge.</td>
</tr>
<tr>
<td>3.</td>
<td>Calculate ( A ) and ( C ) by (6) and (7);</td>
</tr>
<tr>
<td>4.</td>
<td>Calculate ( e = { e_j, j \in S } )</td>
</tr>
<tr>
<td>5.</td>
<td>For ( j \in S )</td>
</tr>
<tr>
<td>6.</td>
<td>If ( e_j^2 \leq 0.5 )</td>
</tr>
<tr>
<td>7.</td>
<td>Calculate the kernel size by (25) with the fixed-point iteration;</td>
</tr>
<tr>
<td>8.</td>
<td>Calculate ( \frac{df_j(e_j)}{dp_i} ) By (30);</td>
</tr>
<tr>
<td></td>
<td>Else</td>
</tr>
</tbody>
</table>
9. Calculate \( \frac{df(e_i)}{dp_i} \) By (29);

10. End if

11. End for

12. Calculate \( \frac{dF(P, \sigma)}{dp_i} \) by (22)

13. Update \( p_{i,t+1} = middle(p_{min}, p_{max} + \mu(-\frac{dF(P)}{dp_i}), p_{max}) \).

14. End repeat

15. If \( SINR_j > \gamma_j - \Delta \gamma, j \in S \)

16. Break

17. Else

18. Remove link \( k_0 \) by (31), \( \{S\} = \{S\} / k_0 \)

19. End if

End loop

4. Simulation Results

We first compare the performance of the CJPA with one of the latest joint power and admission control algorithm, the NLPD in [9]. The simulation is carried out in both small scale network and large scale network as [9]. These simulation results are averaged over 200 Monte-Carlo runs. And we also compare the convergence speed of CJPA with and without the adaptive kernel method.

In our simulations, we use the familiar channel parameters as in [9]. The transmitter’s location is randomly located in a 2Km \( \times \) 2Km area. The receiver is served by the transmitter that provides the strongest signal. This is different from the setting in [9], which assumes that the users are randomly located in a radius around the transmitter. By this setting in [9], some of the receivers may suffer strong interference that the receivers may not choose the corresponding transmitters in reality. We first generate multi-users around each transmitter in a radius from 10m to 400m and assume that the initial transmit powers of all links are 10mw. Then remove the receivers whose received interference is stronger than the serving signal strength and randomly choose one of the left receivers as the one for further simulation. The channel gains are \( g_{i,j} = \frac{1}{d_{i,j}^4} \), where \( d_{i,j} \) is the Euclidean distance from transmitter of link j to the receiver of link i [9]. The SINR targets of all links are \( \gamma_j = 2 \) and the noise power level \( \sigma^2 \) is -90dBm. The transmit power budget of link i is \( p_{i}^{max} = 100mW \) and \( p_{i}^{min} = 10^{-5} mw \).
Fig. 4 compares the performance of the CJPA with different smallest kernel sizes and the NLPD [9]. The vertical axis is the average number of supported links. In small networks, CJPA significantly support more links than the NLPD since the correntropy can closely approximate the $l_0$-norm problem. For different kernel sizes, a smaller kernel size will result in more admitted links due to the same reason. Meanwhile, the differences among the kernel sizes {0.25, 0.5, 0.75} are small. This means that the performance is robust to kernel size when the kernel size is small.

The comparisons between NLPD and the CJPA in large network are illustrated as Fig. 5. We can see from the results that the CJPA supports much more links than the NLPD.

A typical result that compares the convergence speeds of the CJPA with and without the adaptive kernel method of one link is shown in Fig. 6. The kernel size of the fixed kernel method is 0.25 and the minimum kernel size of the adaptive kernel size is also 0.25. From Property 3, the adaptive kernel size will converge to minimum kernel size when the link is feasible. Thus the kernel size of both two methods will be 0.25 after the convergence. The learning rate is 0.05. For the adaptive kernel size method, the kernel size is initialized to be 5. The adaptive kernel size method significantly increases the convergence speed. The SINR converges very fast at the beginning of the process and then achieve the optimal in 100 steps. By the fixed kernel method, the convergence requires about 450 steps.
5. Conclusion

In this paper, we consider solving a NP-hard problem of the joint power and admission control as a sparse optimization problem. We adopt the correntropy to approximate the $l_0$-norm as the tool for sparse optimization and propose the proposed correntropy induced joint power and admission control algorithm (CJPA). We also propose an adaptive kernel size method to improve the convergence speed. Simulation results show that the CJPA can significantly increase the number of the available links in both small scale and large scale DSCN.
6. Acknowledgments
This work is supported by the National Natural Science Foundation of China (61371087, 61372152), “863” Program under Grants 2014AA01A701 and “973” Program under Grants 2015CB351703.

7. References


