Delay Minimization for NOMA-MEC Offloading

Zhiguo Ding, Derrick Wing Kwan Ng, Robert Schober, and H. Vincent Poor

***Abstract*—This letter considers the minimization of the offloading delay for nonorthogonal multiple access assisted mobile edge computing (NOMA-MEC). By transforming the delay minimization problem into a form of fractional programming, two iterative algorithms based on, respectively, Dinkelbach's method and Newton's method are proposed. The optimality of both methods is proved and their convergence is compared. Furthermore, criteria for choosing between three possible modes, namely orthogonal multiple access, pure NOMA, and hybrid NOMA, for MEC offloading are established.**

***Index Terms*—Non-orthogonal multiple access (NOMA) and mobile edge computing (MEC) offloading.**

I. Introduction

THE application of non-orthogonal multiple access (NOMA) to mobile edge computing (MEC) has received considerable attention recently [1]–[5]. In particular, the superior performance of NOMA-MEC with fixed resource allocation was illustrated in [1]. In [2], a weighted sum-energy minimization problem was investigated in a multi-user NOMA-MEC system. In [3], the energy consumption of NOMA-MEC networks was minimized assuming that each user has access to multiple bandwidth resource blocks. In [4], joint power and time allocation was designed for NOMA-MEC, again with the objective of minimizing the offloading energy consumption. To the best of the authors' knowledge, the minimization of the offloading delay for NOMA-MEC has not yet been studied.

The aim of this letter is to study delay minimization for NOMA-MEC offloading. Compared to the energy minimization problems studied in [2]–[5], minimizing the offloading delay is more challenging, since the delay is the ratio of two rate-related functions. We first transform the delay minimization problem into a form of fractional programming. However, the transformed problem is fundamentally different from the original fractional programming problem in [6]. To this end, two algorithms based on respectively Dinkelbach's method and Newton's method are proposed, and their optimality is proved. While the two methods are equivalent for conventional fractional programming, Newton's method is shown to converge faster than Dinkelbach's method for the problem considered here. In addition, criteria for choosing between three possible modes, namely orthogonal multiple access (OMA), pure NOMA, and hybrid NOMA (H-NOMA), for MEC offloading are established. Interestingly, we find that pure NOMA can outperform H-NOMA when there is sufficient energy for MEC offloading, whereas H-NOMA always outperforms pure NOMA if the objective is to reduce the energy consumption.

Manuscript received July 18, 2018; revised August 31, 2018; accepted October 3, 2018. Date of publication October 15, 2018; date of current version November 8, 2018. The work of Z. Ding was supported in part by the UK EPSRC under Grant EP/L025272/2 and in part by EC-H2020 under Grant 690750. The work of D. W. K. Ng was supported by the Australian Research Councils Discovery Early Career Researcher Award funding scheme (Project DE170100137). The work of H. V. Poor was supported by the U.S. National Science Foundation under Grant CNS-1702808. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Feifei Gao. (*Corresponding author: Zhiguo Ding.*)

Z. Ding is with the School of Electrical and Electronic Engineering, the University of Manchester, Manchester, U.K. (e-mail: zhiguo.ding@manchester.ac.uk).

D. W. K. Ng is with the School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, Australia (e-mail: w.k.ng@unsw.edu.au).

R. Schober is with the Institute for Digital Communications, Friedrich-Alexander-University Erlangen-Nurnberg (FAU) Germany (e-mail: robert.schober@fau.de).

H. V. Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

Digital Object Identifier 10.1109/LSP.2018.2876019

II. System Model

Consider a MEC offloading scenario, in which two users, denoted by user m and user n, offload their computation tasks to a MEC server [7]. Without loss of generality, assume that the two users' tasks contain the same number of nats, denoted by N, and user m's computation deadline, denoted by D_m, is shorter than user n's, denoted by D_n, i.e., $D_m \leq D_n$.

In this work, as user m has a more stringent delay requirement than user n, user n will be served in an opportunistic manner as described in the following. In particular, user m's transmit power, denoted by P_m, is set to the same value as in OMA, i.e., P_m satisfies $D_m \ln(1 + P_m|h_m|^2) = N$, where h_k denotes user k's channel gain, $k \in \{m, n\}$. User n is allowed to access the time slot of D_m seconds allocated to user m, under the condition that user m experiences the same rate as for OMA. As shown in [4], this can be realized if user m's message is decoded after user n's at the MEC server and user n's data rate, denoted by R_n, during D_m is set as follows:

$$R_n = \ln\left(1 + \frac{P_{n,1}|h_n|^2}{P_m|h_m|^2 + 1}\right), \tag{1}$$

where $P_{n,1}$ denotes the power used by user n during D_m. If user n cannot finish its offloading within D_m, a dedicated time slot, denoted by T_n, is allocated to user n, and the user's transmit power during T_n is denoted by $P_{n,2}$. Note that the three cases with $\{P_{n,1} = 0, P_{n,2} \neq 0\}$, $\{P_{n,1} \neq 0, P_{n,2} = 0\}$, and $\{P_{n,1} \neq 0, P_{n,2} \neq 0\}$ correspond to OMA, pure NOMA, and H-NOMA, respectively. We note that both OMA and pure NOMA can be viewed as special cases of H-NOMA. However, in this paper, the three modes are considered separately and H-NOMA is restricted to the case with $\{P_{n,1} \neq 0, P_{n,2} \neq 0\}$.

III. NOMA-Assisted MEC Offloading

In this paper, we assume that NOMA-MEC offloading is transparent to user m, i.e., user m's delay performance in NOMA remains the same as that in OMA while user n is admitted to D_m with its achievable data rate constrained by (1).

Therefore, P_m and D_m are assumed to be fixed, and our objective is to minimize user n' delay by optimizing $P_{n,1}$ and $P_{n,2}$ as follows:

$$\underset{P_{n,1},P_{n,2}}{\text{minimize}} \quad D_m + T_n \tag{2a}$$

$$\text{s.t.} \quad \{P_{n,1}P_{n,2}\} \in \mathcal{S}, \tag{2b}$$

where $T_n = \frac{N - D_m \ln\left(1+\frac{P_{n,1}|h_n|^2}{P_m|h_m|^2+1}\right)}{\ln(1+|h_n|^2 P_{n,2})}$,

$$\mathcal{S} = \{D_m P_{n,1} + T_n P_{n,2} \le E, P_{n,1} \ge 0, P_{n,2} \ge 0, T_n \ge 0\}, \tag{3}$$

and E is user n's energy constraint. We note that T_n is zero if pure NOMA is used, and the constraint in Eq. (2b) implies that user n's power during D_m needs to ensure that user m experiences the same rate as in OMA.

Define $E_1 = D_m\left(e^{\frac{N}{D_m}} - 1\right)|h_n|^{-2}$ and $E_2 = E_1 e^{\frac{N}{D_m}}$. E_1 and E_2 are the energy thresholds to determine the use of OMA, pure NOMA, and H-NOMA, as shown in the following.

A. Case $E \ge E_2$

This corresponds to the case with sufficient energy at user n for MEC offloading, and the minimal D_m can be achieved by using pure NOMA, i.e., $P_{n,2} = 0$. The condition for adopting pure NOMA is shown in the following. Since $P_{n,2} = 0$, all the energy is consumed during the NOMA phase to minimize the delay, which means $P_{n,1} = \frac{E}{D_m}$. Hence, user n is able to offload its task within D_m if

$$\begin{aligned} N &\le D_m \ln\left(1 + \frac{P_{n,1}|h_n|^2}{P_m|h_m|^2 + 1}\right) \\ &\overset{(a)}{=} D_m \ln\left(1 + e^{-\frac{N}{D_m}} \frac{E}{D_m}|h_n|^2\right), \end{aligned} \tag{4}$$

where step (a) is obtained by assuming that user m's power, P_m, satisfies the constraint $D_m \ln\left(P_m|h_m|^2 + 1\right) = N$. By solving the inequality in (4), the condition $E \ge E_2$ can be obtained.

The performance gain of NOMA-MEC over OMA-MEC is obvious in this case since user n's delay in OMA is $D_m + \frac{N}{\ln(1+|h_n|^2 P_{n,2})}$, which is strictly larger than D_m. However, the comparison between OMA and NOMA becomes more complicated for the energy-constrained cases.

B. Case $E_1 < E < E_2$

This corresponds to the case in which there is not sufficient energy at user n to support pure NOMA. Note that both H-NOMA and OMA are still applicable. Due to space limitations, we focus on H-NOMA ($P_{n,i} \ne 0$, $i \in \{1,2\}$), as the OMA solution can be obtained in a straightforward manner.

Note that T_n is the ratio of two functions of $P_{n,1}$ and $P_{n,2}$, respectively, which motivates the use of fractional programming. However, compared to conventional fractional programming in [6], the problem in (2) is more challenging since the fractional function T_n does not only appear in the objective function but also in the constraint. Two iterative algorithms will be developed based on the following Dinkelbach auxiliary function parametrized by μ:

$$\begin{aligned} F(\mu) = \underset{P_{n,1},P_{n,2}}{\text{maxmize}} \quad & \ln\left(1 + |h_n|^2 P_{n,2}\right) \\ & - \mu\left(N - D_m \ln\left(1 + e^{-\frac{N}{D_m}} P_{n,1}|h_n|^2\right)\right), \end{aligned} \tag{5}$$

where $\{P_{n,1}, P_{n,2}\} \in \tilde{\mathcal{S}}(\mu)$ and[1]

$$\tilde{\mathcal{S}}(\mu) = \left\{D_m P_{n,1} + \mu^{-1} P_{n,2} \le E, P_{n,1} \ge 0, P_{n,2} \ge 0\right\}. \tag{6}$$

Different from the original form in [6], the constraint set $\tilde{\mathcal{S}}(\mu)$ for the auxiliary function is also a function of μ and $F(\mu)$ might have more than one root.

For a fixed μ, the following lemma provides the optimal H-NOMA solution for problem (5).

Lemma 1: For a fixed μ, the optimal H-NOMA power allocation policy for problem (5) is given by

$$\begin{cases} P_{n,1}^*(\mu) = \frac{E - \mu^{-1}\left(e^{\frac{N}{D_m}} - 1\right)|h_n|^{-2}}{D_m + \mu^{-1}} \\ P_{n,2}^*(\mu) = \frac{E + D_m\left(e^{\frac{N}{D_m}} - 1\right)|h_n|^{-2}}{D_m + \mu^{-1}} \end{cases}. \tag{7}$$

Proof: Due to space limitations, only a sketch of the proof is provided. Problem (5) is convex for a fixed μ, which leads to the following Karush-Kuhn-Tucker (KKT) conditions [8]:

$$\begin{cases} -\frac{\mu D_m e^{-\frac{N}{D_m}}|h_n|^2}{1+e^{-\frac{N}{D_m}}P_{n,1}|h_n|^2} + \lambda_1 D_m - \lambda_2 - \lambda_3 = 0 \\ -\frac{|h_n|^2}{1+|h_n|^2 P_{n,2}} + \lambda_1 \mu^{-1} - \lambda_2 - \lambda_3 = 0 \\ D_m P_{n,1} + \mu^{-1} P_{n,2} \le E \\ \lambda_1\left(D_m P_{n,1} + \mu^{-1} P_{n,2} - E\right) = 0 \\ P_{n,i} \ge 0, \forall i \in \{1,2\}, \quad \lambda_i P_{n,i-1} = 0, \forall i \in \{2,3\} \\ \lambda_i \ge 0, \forall i \in \{1,2,3\} \end{cases},$$

where the λ_i are Lagrange multipliers. For the H-NOMA case, $P_{n,1} > 0$ and $P_{n,2} > 0$, and hence $\lambda_2 = 0$ and $\lambda_3 = 0$ due to the constraints $\lambda_i P_{n,i} = 0, \forall i \in \{2,3\}$. Therefore, the KKT conditions can be simplified as follows:

$$\begin{cases} -\frac{\mu D_m e^{-\frac{N}{D_m}}|h_n|^2}{1+e^{-\frac{N}{D_m}}P_{n,1}|h_n|^2} + \lambda_1 D_m = 0 \\ -\frac{|h_n|^2}{1+|h_n|^2 P_{n,2}} + \lambda_1 \mu^{-1} = 0 \\ D_m P_{n,1} + \mu^{-1} P_{n,2} \le E \\ \lambda_1\left(D_m P_{n,1} + \mu^{-1} P_{n,2} - E\right) = 0 \\ P_{n,i} \ge 0, \forall i \in \{1,2\}, \quad \lambda_1 \ge 0 \end{cases}. \tag{8}$$

Due to the constraint $-\frac{|h_n|^2}{1+|h_n|^2 P_{n,2}} + \lambda_1 \mu^{-1} = 0$, we have $\lambda_1 \ne 0$ for the optimal solution, since otherwise $\frac{|h_n|^2}{1+|h_n|^2 P_{n,2}} = 0$, which cannot be true. Since $\lambda_1 \ne 0$ and $\lambda_1(D_m P_{n,1} + \mu^{-1} P_{n,2} - E) = 0$, we have $D_m P_{n,1} + \mu^{-1} P_{n,2} - E = 0$. With some algebraic manipulations, the lemma can be proved.

Remark 1: By substituting $P_{n,1}^*(\mu)$ and $P_{n,2}^*(\mu)$ into (5), $F(\mu)$ can be expressed as an explicit function of μ. Unlike [6], $F(\mu)$ is not convex and may have multiple roots, which results in fundamental changes to the convergence proof.

[1]For the case $E < E_2$, $T_n > 0$ as $N > D_m \ln\left(1 + e^{-\frac{N}{D_m}} \frac{E}{D_m}|h_n|^2\right)$, and hence the constraint, $T_n \ge 0$, can be omitted.

Algorithm 1: Dinkelbach's Method Based Algorithm.

b
1: Set $t=0$, $\mu_0=+\infty$, $\delta \to 0$.
2: **while** $F(\mu_t) < -\delta$ **do**
3: $t=t+1$.
4: Update $P_{n,1}^t$ and $P_{n,2}^t$ by using $\mu=\mu_{t-1}$ in (7).
5: Update $F(\mu_t)$ by using $P_{n,1}^t$ and $P_{n,2}^t$.
6: Update μ_t as $\mu_t = \frac{\ln\left(1+|h_n|^2 P_{n,2}^t\right)}{N-D_m \ln\left(1+e^{-\frac{N}{D_m}} P_{n,1}^t |h_n|^2\right)}$
7: **end**
8: $P_{n,1}^* = P_{n,1}^t$ and $P_{n,2}^* = P_{n,2}^t$.

Algorithm 2: Newton's Method Based Algorithm.

1: Set $t=0$, $\mu_0=+\infty$, $\delta \to 0$.
2: **while** $F(\mu_t) < -\delta$ **do**
3: $t=t+1$.
4: Update $\mu_{t+1} = \mu_t - \frac{F(\mu_t)}{F'(\mu_t)}$.
5: **end**
6: $P_{n,1}^{*,N}$ and $P_{n,2}^{*,N}$ are obtained by using $\mu=\mu_t$ in (7).

Although $F(\mu)$ is different from the form in [6], a modified Dinkelbach's method can still be developed as shown in Algorithm 1. In Algorithm 1, δ denotes a small positive threshold. In addition, a Newton's method-based iterative algorithm can also be developed as in Algorithm 2, where $F'(\mu)$ denotes the first-order derivative of $F(\mu)$.

Theorem 1 shows the optimality of the two algorithms.

Theorem 1: The two iterative algorithms shown in Algorithms 1 and 2 converge to the same optimal solution of problem (2).

Proof: We start the proof by first studying the roots of the function in (5) which can potentially be the optimal solution. We note that the steps provided in [6] cannot be straightforwardly applied to prove the optimality of the modified Dinkelbach's method since $F(\mu)$ may have multiple roots.

Step 1 of the proof is to show the existence and uniqueness of the root of $F(\mu)$, for $\left(e^{\frac{N}{D_m}}-1\right)|h_n|^{-2}E^{-1} < \mu < \infty$. To prove this, it is sufficient to show that 1): $F(\mu)$ is strictly concave, 2): $F(\mu)=-\infty$ for $\mu \to +\infty$ and 3): $F(\mu)>0$ for $\mu \to \left(e^{\frac{N}{D_m}}-1\right)|h_n|^{-2}E^{-1}$.

The first order derivative of $F(\mu)$ is given by

$$F'(\mu) = \frac{|h_n|^2 E + D_m\left(e^{\frac{N}{D_m}}-1\right)}{e^{\frac{N}{D_m}} D_m \mu + |h_n|^2 E\mu + 1} - \left[N - D_m \times \ln\left(1+e^{-\frac{N}{D_m}}|h_n|^2 \frac{E - \frac{\left(e^{\frac{N}{D_m}}-1\right)|h_n|^{-2}}{\mu}}{D_m + \frac{1}{\mu}}\right)\right]. \quad (9)$$

Unlike [6], $F'(\mu)$ is neither strictly negative nor positive. The second-order derivative of $F(\mu)$ is given by

$$F''(\mu) = \frac{\left(D_m^2 - \left(e^{\frac{N}{D_m}} D_m + |h_n|^2 E\right)^2\right)}{\left(e^{\frac{N}{D_m}} D_m \mu + |h_n|^2 E\mu + 1\right)^2 (\mu D_m + 1)}. \quad (10)$$

Since $e^{\frac{N}{D_m}} > 1$ and $|h_n|^2 E > 0$, we have

$$F''(\mu) < 0, \quad (11)$$

which means that $F(\mu)$ is a strictly concave function of μ.

When $\mu \to \left(e^{\frac{N}{D_m}}-1\right)|h_n|^{-2}E^{-1}$, the H-NOMA solution approaches $P_{n,1}^*(\mu) \to 0$ and $P_{n,2}^*(\mu) \to \left(e^{\frac{N}{D_m}}-1\right)|h_n|^{-2}$. Consequently, $F(\mu)$ can be approximated as follows:

$$F(\mu) \to \ln\left(1+|h_n|^2 \frac{E\mu + \mu D_m\left(e^{\frac{N}{D_m}}-1\right)|h_n|^{-2}}{D_m\mu+1}\right) - \mu N \to \frac{N}{D_m} - \frac{N\left(e^{\frac{N}{D_m}}-1\right)|h_n|^{-2}}{E}.$$

As suggested by the title of Section III-B, it is assumed that $E > E_1$, which means

$$F(\mu) > 0, \quad (12)$$

for $\mu \to \left(e^{\frac{N}{D_m}}-1\right)|h_n|^{-2}E^{-1}$.

When $\mu \to \infty$, $F(\mu)$ can be approximated as follows:

$$F(\mu) \to \ln\left(1+|h_n|^2 \frac{E + D_m\left(e^{\frac{N}{D_m}}-1\right)|h_n|^{-2}}{D_m}\right) - \mu\left(N - D_m \ln\left(1+e^{-\frac{N}{D_m}}|h_n|^2\frac{E}{D_m}\right)\right) \to -\infty, \quad (13)$$

where the last step is obtained since for H-NOMA $\left(N - D_m \ln\left(1+e^{-\frac{N}{D_m}}|h_n|^2\frac{E}{D_m}\right)\right) > 0$. Combining (11), (12), and (13), the proof for the first step is complete. The unique root of $F(\mu)$ for $\mu > \left(e^{\frac{N}{D_m}}-1\right)|h_n|^{-2}E^{-1}$ is denoted by μ^*, where $\mu^* > \left(e^{\frac{N}{D_m}}-1\right)|h_n|^{-2}E^{-1}$.

Step 2 is to show that $T_n^* \triangleq \frac{1}{\mu^*}$ is the optimal solution of problem (2). This step can be proved by using contradiction. Assume that the optimal solution of problem (2) is smaller than T_n^*, denoted by $\bar{T}_n^*$. Denote $\bar{\mu}^* = \frac{1}{\bar{T}_n^*}$ and hence $\bar{\mu}^* > \mu^*$. Following steps similar to those in [6], we have $F(\bar{\mu}^*)=0$, i.e., $\bar{\mu}^*$ is also a root of $F(\mu)$ and $\bar{\mu}^* > \mu^* > \left(e^{\frac{N}{D_m}}-1\right)|h_n|^{-2}E^{-1}$, which contradicts the uniqueness of the root of $F(\mu)$. Step 2 means that finding the optimal solution of problem (2) is equivalent to finding the largest root of $F(\mu)$. Therefore, the optimality of Newton's method can be straightforwardly shown, and in the following, we only focus on the modified Dinkelbach's method.

Step 3 is to show $F(\mu)$ is a strictly decreasing function of μ, for $\mu^* \le \mu \le \infty$, which can be proved by using the facts that μ^* is the unique root of $F(\mu)$ for $\mu > \left(e^{\frac{N}{D_m}}-1\right)|h_n|^{-2}E^{-1}$, and $F(\mu)$ is concave.

Step 4 is to show $F(\mu_{t+1}) < 0$ if $F(\mu_t) < 0$ for the modified Dinkelbach's method. Since $F(\mu_t) < 0$ and $F(\mu)$ is a strictly decreasing function for $\mu \ge \mu^*$, we have $\mu_t > \mu^*$. The convergence analysis for Newton's method is provided first to facilitate that of Dinkelbach's method. The quadratic convergence analysis for Newton's method yields the following [9]:

$$\mu^* - \mu_{t+1} = -\frac{F''(\mu_\xi)}{2F'(\mu_t)}(\mu^* - \mu_t)^2, \quad (14)$$

where $\mu^* \le \mu_\xi \le \mu_t$. Note that $F'(\mu_t) < 0$ and $F''(\mu_\xi) < 0$. Therefore, we have

$$\mu^* - \mu_{t+1} < 0 \quad (15)$$

which means $F(\mu_{t+1}) < 0$ for Newton's method.

For Dinkelbach's method, $F(\mu)$ is first expressed as $F(\mu) = A(\mu) - \mu B(\mu)$, which means that μ is updated as follows:

$$\mu_{t+1} = \frac{A(\mu_t)}{B(\mu_t)} = \mu_t + \frac{F(\mu_t)}{B(\mu_t)}. \tag{16}$$

Recall that Newton's method updates μ as follows:

$$\mu_{t+1} = \mu_t - \frac{F(\mu_t)}{F'(\mu_t)} = \mu_t + \frac{F(\mu_t)}{B(\mu_t) - (A'(\mu_t) - \mu_t B'(\mu_t))}. \tag{17}$$

From (9), $A'(\mu_t) - \mu_t B'(\mu_t)$ can be expressed as follows:

$$A'(\mu_t) - \mu_t B'(\mu_t) = \frac{|h_n|^2 E + D_m\left(e^{\frac{N}{D_m}} - 1\right)}{e^{\frac{N}{D_m}} D_m \mu + |h_n|^2 E\mu + 1} > 0. \tag{18}$$

Note that $B(\mu) > (A'(\mu_t) - \mu_t B'(\mu_t)) > 0$ since $F'(\mu_t) < 0$. Therefore, the step size of Newton's method is larger than that of Dinkelbach's method. In other words, starting with the same μ_t, μ_{t+1} obtained from Newton's method is smaller than that of Dinkelbach's method, which means $F(\mu_{t+1}) < 0$ also holds for Dinkelbach's method.

Step 5 is to show $\mu_{t+1} < \mu_t$, which can be proved by using (16) and (17). An interesting property deduced from Steps 4 and 5 is that if the initial value of μ is set as ∞, μ_t is decreasing and approaches μ^* as the number of iterations increases, but will never pass μ^*, as $F(\mu_t)$ is always negative.

Step 6 is to show that Dinkelbach's method converges to μ^*. This can be proved by contradiction. Assume that the method converges to $\breve{\mu}$, i.e., $\lim_{\to\infty}\mu_t = \breve{\mu}$ and $\breve{\mu} \neq \mu^*$. Although $F(\mu)$ might have more than one root, $F(\mu_t)$ is always negative if $\mu_0 = \infty$, as illustrated by Step 4. Therefore, μ_t is always larger than μ^*, i.e., $\breve{\mu} > \mu^*$. Since $F(\mu)$ is strictly decreasing for $\mu > \mu^*$, we have the conclusion that $0 = F(\breve{\mu}) < F(\mu^*) = 0$, which cannot be true. The proof is complete. ■

Corollary 1: The algorithm based on Newton's method converges faster than the one based on Dinkelbach's method.

Proof: The corollary can be proved by using Step 4 in the proof for Theorem 1. ■

Remark 2: The rationale behind the existence condition of the H-NOMA solution, i.e., $E > E_1$, can be explained as follows. From the proof of Theorem 1, we learn that $F(\mu)$ is a decreasing function of μ for $\mu > \mu^*$. On the other hand, $P^*_{n,1}(\mu)$ and $P^*_{n,2}(\mu)$ are increasing functions of μ. So it is important to ensure that when μ is reduced to a value $\bar{\mu}$, such that $P^*_{n,1}(\bar{\mu}) = 0$, i.e., $\bar{\mu} = \frac{e^{\frac{N}{D_m}} - 1}{E|h_n|^2}$, $F(\mu)$ is positive. Otherwise, a positive root of $F(\mu)$ does not exist, and there is no feasible H-NOMA solution. By substituting $\bar{\mu}$ into (7) and solving the inequality $F(\bar{\mu}) > 0$, the existence condition is obtained. Similarly, one can find that $E \geq N|h_n|^{-2}$ is the condition under which OMA is feasible. Hence, OMA-MEC is the only feasible option to minimize the delay if $N|h_n|^{-2} \leq E \leq E_1$, i.e., there is very limited energy available for MEC offloading.

Remark 3: For the case $E_1 < E < E_2$, both OMA and NOMA are applicable. Simulation results show that H-NOMA always yields less delay than OMA, although we have yet to obtain a formal proof for this conjecture.

IV. Numerical Studies

In this section, the performance of the proposed MEC offloading scheme is studied and compared by using computer simulations, where normalized channel gains are adopted for the purpose of clearly demonstrating the impact of the channel conditions on the delay. In Fig. 1a, the impact of NOMA on the MEC offloading delay is shown as a function of the energy consumption. Note that the curves for NOMA-MEC are generated based on the combination of H-NOMA and pure NOMA, i.e., if $E \geq D_m\left(e^{\frac{N}{D_m}} - 1\right)e^{\frac{N}{D_m}}|h_n|^{-2}$, pure NOMA is used, otherwise H-NOMA is used. For a given amount of energy consumed, Fig. 1a shows that the use of NOMA reduces the delay significantly. Particularly when there is plenty of energy available at user n, i.e., $E \geq D_m\left(e^{\frac{N}{D_m}} - 1\right)e^{\frac{N}{D_m}}|h_n|^{-2}$, the use of NOMA ensures that D_m is sufficient for offloading and there is no need to utilize extra time. Fig. 1b provides a comparison of the convergence rates of the two proposed iterative algorithms. Note that both algorithms start with a delay of 0 ($\mu_0 = \infty$) and only the delay after convergence in the figure is achievable. As shown in the figure, Dinkelbach's method generally converges more slowly than Newton's method, as predicted by Corollary 1, although they perform the same in conventional scenarios [6].

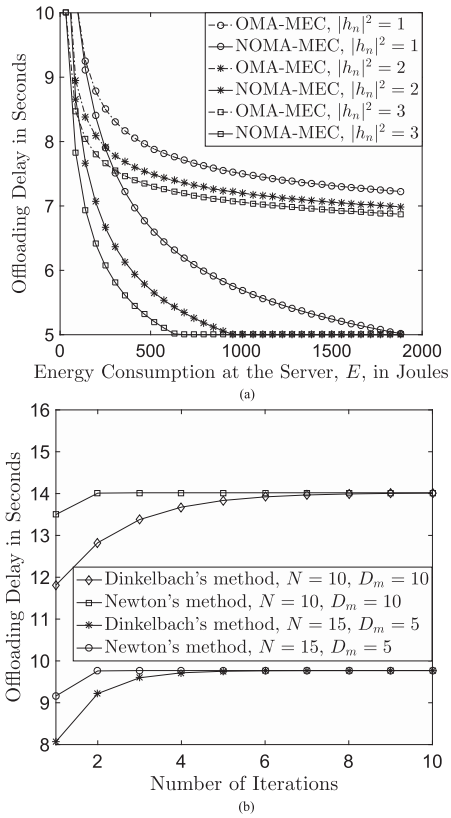

Fig. 1. Impact of NOMA on MEC offloading and the convergence of the two proposed iterative method.

V. Conclusion

In this paper, two iterative algorithms have been developed to minimize the offloading delay of NOMA-MEC. The optimality of the algorithms has been proven and their rates of convergence have been analyzed. Furthermore, criteria for choosing between the three possible modes, OMA, pure NOMA, and H-NOMA, for MEC offloading have been established. An important topic for future research is to investigate NOMA-MEC for general scenarios with more than two users.

REFERENCES

[1] Z. Ding, P. Fan, and H. V. Poor, "Impact of non-orthogonal multiple access on the offloading of mobile edge computing," *IEEE Trans. Commun.*, to be published, arXiv:1804.06712.

[2] F. Wang, J. Xu, and Z. Ding, "Optimized multiuser computation offloading with multi-antenna NOMA," in *Proc. IEEE Globecom Workshops*, Singapore, Dec. 2017, pp. 1–6.

[3] A. Kiani and N. Ansari, "Edge computing aware NOMA for 5G networks," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1299–1306, Apr. 2018.

[4] Z. Ding, J. Xu, O. A. Dobre, and H. V. Poor, "Joint power and time allocation for NOMA-MEC offloading," *IEEE Trans. Veh. Tech.*, submitted, arXiv:1807.06306.

[5] Y. Wu and L. P. Qian, "Energy-efficient NOMA-enabled traffic offloading via dual-connectivity in small-cell networks," *IEEE Commu. Lett.*, vol. 21, no. 7, pp. 1605–1608, Jul. 2017.

[6] W. Dinkelbach, "On nonlinear fractional programming," *Manage. Sci.*, vol. 13, no. 7, pp. 492–298, 1967.

[7] 3rd Generation Partnership Project, "Study on downlink multiuser superposition transmission for LTE," Sophia Antipolis, France, Mar. 2015.

[8] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[9] E. Hildebrand, *Introduction to Numerical Analysis*. New York, NY, USA: Dover, 1987.