On the Capacity of Non-Coherent Network Coding

Mahdi Jafari
EPFL
mahdi.jafarisiavoshani@epfl.ch

Soheil Mohajer
EPFL
soheil.mohajer@epfl.ch

Christina Fragouli
EPFL
christina.fragouli@epfl.ch

Suhas Diggavi
EPFL
suhas.diggavi@epfl.ch

Abstract— **The min-cut value towards a single receiver in a network with unit capacity edges can be achieved by routing a single bit. The multicast theorem in network coding shows that, the common min-cut value towards $N \geq 1$ receivers can also be achieved using packets of length $\log N$ bits, if the operations the intermediate nodes perform are deterministically known at the receivers. We here calculate the capacity in the case where these operations are unknown, and characterize how the capacity depends on the min-cut value and the packet length.**

I. Introduction

The min-cut value towards a single receiver in a network with unit capacity edges can be achieved by routing a single bit. Linear network coding, introduced in [1], demonstrated that with linear operations at intermediate nodes, one can achieve the common min-cut value when multicasting to $N \geq 1$ receivers by using packets of $\log N$ bits. However, this result assumes that the receivers know perfectly the operations that the network nodes perform. In practical networks, where such deterministic knowledge is not sustainable, the most popular approach is to append coding vectors at the headers of the packets to keep track of the linear combinations of the source packets they contain. This results in a loss of information rate with respect to the min-cut value. In a sense, this is akin to training symbols to learn the transformation induced by the network. Recently, algebraic subspace coding constructions have been proposed as a method that allows to achieve higher information rates by dispensing of the need for the coding vector overheads [3]. In this paper we examine what are the information theoretical rates that can be achieved in a network where the intermediate node operations are unknown.

We consider a network where neither the source nor the receivers have knowledge of the network topology or of the linear coding operations the network nodes perform. In [6] we proposed a model to capture this communication, where the source inserts in the network m packets of length T over some finite field $\mathbb{F}_q$, and each receiver collects n packets that consist of random combinations of the source packets. For this model, we proved that the source can communicate information to the receivers through the choice of the subspaces it employs, since subspaces are preserved under linear transformations, as was also observed in [3]. We also calculated the capacity for the case where $T > min(m, n) + n$. We here complete this work by determining the capacity for all values of m, n and T. This capacity is characterized for all regimes by:

$$\frac{1}{T}\left[\mathbb{1}_{\{T:\text{odd}\}} + i^*(T-i^*)\log_2 q + \mathcal{O}(q^{-1}\log q)\right],$$

where $i^* = \min\{m, n, \lfloor T/2 \rfloor\}$. Therefore, we see that for $q \gg 1$, the capacity behaves like $i^*(1-\frac{i^*}{T})\log_2 q$, for all ranges of T. Note that this rate is achievable for multicast information flow.

When T becomes very large, the capacity approaches the min-cut value $\min\{m,n\}\log_2 q$, as expected. Interestingly, when T is small, the capacity is achieved (for $q \gg 1$) by using subspaces across multiple dimensions. This is in direct contrast to the current constructions of subspace codes that utilize a single dimension subspaces to encode the information messages [2], [8], [9]. Our result demonstrates that such constructions are optimal only when $T \geq \min\{m,n\} + n$, while in regimes for small packet sizes, it is optimal to utilize subspaces of multiple dimensions, and the dimensions used vary with the relative values of m, n and T. Another direct implication of our work is that, for $\frac{T}{2} \geq n = m$, subspace coding does not offer benefits as compared to the coding vectors approach. Finally, our work can be directly extended in networks with packet erasures.

Recently, Silva *et al.* [10] independently and subsequent to our work in [6], considered a probabilistic model for noncoherent network coding which is an extension of the model introduced in [9]. In this model the transfer matrix is square ($m = n$) and is uniformly at random selected among all *full rank* $n \times n$ matrices. This is in contrast to our model, where the elements of the transfer matrix are chosen uniformly at random, and thus the transfer matrix itself may not have full rank. For the case where $T \geq 2n$, which is the only case considered in [10], and for $q \to \infty$, the capacity value of both approaches coincide.

II. The noncoherent finite field channel model

We consider a network where nodes perform random linear network coding over a finite field $\mathbb{F}_q$. We assume that time is slotted and the channel is block time-varying. At time slot t, the receiver observes

$$Y(t) = G(t)X(t), \qquad (1)$$

where $X(t) \in \mathbb{F}_q^{m\times T}$, $G(t) \in \mathbb{F}_q^{n\times m}$, and $Y(t) \in \mathbb{F}_q^{n\times T}$. At each timeslot, the receiver receives n packets of length T (captured as rows of matrix $Y(t)$) that are random linear combinations of the m packets injected by a source (captured as

This work was supported by the Swiss National Science Foundation under award PP002-110483 and by the EU projects NetReFound (FP6-IST-034413).

978-1-4244-4313-0/09/$25.00 ©2009 IEEE

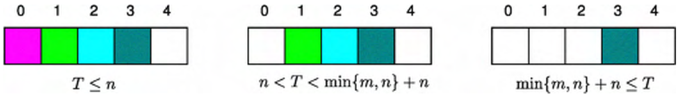
rows of matrix $X(t)$). The packet length T can be interpreted as the coherence time of the channel, during which the transfer matrix[1] G remains constant. Each element of the transfer matrix G is chosen uniformly at random from $\mathbb{F}_q$, changes independently from timeslot to timeslot, and is unknown to both the source and the receiver.[2]

The channel described by (1) can be interpreted as a discrete memoryless channel with input alphabet $\mathcal{X} \triangleq \mathbb{F}_q^{m\times T}$ and output alphabet $\mathcal{Y} \triangleq \mathbb{F}_q^{n\times T}$. As mentioned in [6] the model in (1) along with a uniform distribution for matrices G is information stable, so the capacity of this channel is given by

$$C = \frac{1}{T} \sup_{P_X(x)} I(X;Y), \qquad (2)$$

where $P_X(x)$ is the input distribution. To achieve the capacity a coding scheme may employ the channel (1) multiple times, and a codeword is a sequence of input matrices from $\mathcal{X}$. For a coding strategy that induces an input distribution $P_X(x)$, the achievable rate is

$$R = \frac{1}{T} I(X;Y).$$



$T \leq n$ $\quad$ $n < T < \min\{m,n\}+n$ $\quad$ $\min\{m,n\}+n \leq T$

Fig. 1. Active subspace dimensions for $m = 4$, $n = 3$.

III. Main Results

Our main theorem 1 allows to characterize the capacity for noncoherent network coding. We show that the capacity is achieved through subspace coding, where the information is communicated from the source to the receivers through the choice of subspaces.

Theorem 1: Consider the channel given in (1) and assume that G is drawn uniformly at random from $\mathbb{F}_q^{n\times m}$ and independently from block to block. Then there exists finite q_0 such that for $q > q_0$ the optimal input distribution is non-zero only for the matrices whose rank belongs to

$$\mathcal{A} = \{\min[(T-n)^+, m, n, T], \ldots, \min[m,n,T]\}, \qquad (3)$$

which we call the active set. The capacity of the channel is

$$\begin{aligned} C &= \frac{1}{T}\left[\log_2\left(\sum_{i\in\mathcal{A}} q^{i(T-i)}\right) + \mathbb{1}_{\{T:\text{odd}\}} + \mathcal{O}(q^{-1}\log q)\right] \\ &= \frac{1}{T}\left[\mathbb{1}_{\{T:\text{odd}\}} + i^*(T-i^*)\log_2 q + \mathcal{O}(q^{-1}\log q)\right], \qquad (4) \end{aligned}$$

where $i^* = \arg\min_{i\in\mathcal{A}} |T/2 - i| = \min\{m, n, \lfloor T/2 \rfloor\}$. Moreover, the optimal input distribution is uniform over all matrices X of the same dimension, and the probability of employing matrices X of rank i equals

$$\alpha_i^*(x) = 2^{-C} q^{i(T-i)}\left[1 + \mathcal{O}(q^{-1}\log q)\right], \quad \forall i \in \mathcal{A}, \qquad (5)$$

where C is the capacity of the channel.

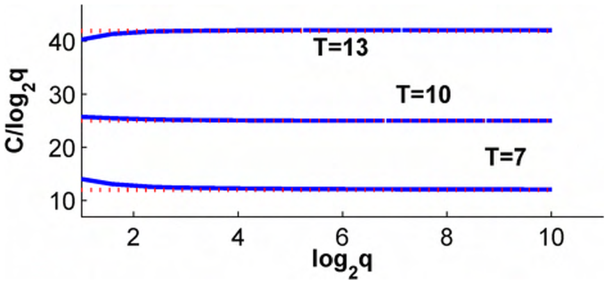


Fig. 2. Numerical calculation of the capacity for small values of q and $m = 11$, $n = 7$. The dotted line depicts $i^*(T - i^*)$.

This result can be directly extended to packet erasure networks. We model erasures by assuming that the receivers observes a number of packets n, where now n is a random variable with a given distribution.

Corollary 1: Consider the model in (1) but now assume that n is a random variable with a known distribution. Then

$$\begin{aligned} C = &\sum_{i=0}^{m} i(T-i)\log_2 q \Pr(n = i) \\ &+ m(T-m)\log_2 q \Pr(n > m). \end{aligned}$$

From Theorem 1, the capacity behaves as $i^*(1-i^*/T)\log q$, for large q. However, numerical simulations indicate a very fast convergence to this value as q increases. Fig. 2 depicts the capacity for small values of q, calculated using using the Differential Evolution toolbox for matlab [11].

We can now derive the following guidelines for network code design.

1) Choice of subspaces: The optimal input distribution uses subspaces of a single dimension equal to $\min\{m,n\}$ for $T \geq \min\{m,n\}+n$. As T reduces, the set of used subspaces gradually increases, by activating one by one smaller and smaller dimensional subspaces, until, for $T \leq n$, all subspaces are used with equal probability. Fig. 1 pictorially depicts this gradual inclusion of subspaces.

2) Values of m and n: For a given and fixed packet length T, the optimal value of m and n equals $m = n = \lfloor T/2 \rfloor$. (optimality is in the sense of minimum required to achieve the maximum information transfer for this T). For fixed T and m, the optimal value of n equals $n = \min\{m, \lfloor T/2 \rfloor\}$. For fixed T and n, the optimal value of m equals $n = \min\{n, \lfloor T/2 \rfloor\}$.

TABLE I

Information Loss from Using Coding Vectors When $n \geq m$.

	$T \leq m$	$m < T < 2m$	$T \geq 2m$
$C - R_{cv}$	C	$[m - \lfloor T/2 \rfloor]^2 \log_2 q$	0

[1] In the rest of the paper we will omit for convenience the time index t.

[2] In general, the topology of the network imposes some constraints on the transfer matrix G (see for example [4]). However, we believe that this is a reasonable model, especially for large scale dynamically changing networks.

3) Subspace coding vs. coding vectors: A natural question is, for what regimes using coding vectors [2] is far from the optimal solution. Table I summarizes this difference. We see that for $\lfloor T/2 \rfloor \geq m = n$, subspace coding does not offer benefits as compared to the coding vectors approach. Table I is calculated as follows. The achievable rate R_{cv} using coding vectors equals $R_{cv} \triangleq m(T-m)\log_2 q$, where each packet includes a coding vector of length m and $T-m>0$ information symbols. Clearly, R_{cv} is nonzero only for $T>m$, and equals zero for the cases $T \leq m$ or $n<m$. Assuming $T>m$ and $n \geq m$ for large q we can write

$$C - R_{cv} = [(m-\Delta)(\Delta+m-T)]\log_2 q,$$

where $\Delta = \min\{m, n, \lfloor T/2 \rfloor\}$. It can be easily shown that for $m > \lfloor T/2 \rfloor$ we have a loss of

$$C - R_{cv} = [m - \lfloor T/2 \rfloor]^2 \log_2 q,$$

while for $m \leq \lfloor T/2 \rfloor$ the loss is zero.

IV. The Channel Capacity

We will use the following notation. Let the Grassmannian $\text{Gr}(i,V)$ denotes the set of all i-dimensional subspaces of a finite-dimensional vector space V. We use $\mathcal{G}_q(T,d)$, or more conveniently $\mathcal{G}(T,d)$, to denote the Gaussian binomial, the number of distinct d-dimensional subspaces of $\mathbb{F}_q^T$.

A. Simplified Mutual Information

In this subsection we express the mutual information in a simplified form. Since the rows of G are chosen independent of each other, conditioned on sending some matrix $X=x$, the rows of the received matrix Y are independent of each other among all the vectors in the row span of x. The independence of rows of Y let us write the conditional probability of Y given X as

$$P_{Y|X}(y|x) = \begin{cases} q^{-n\dim(\langle x \rangle)} & \langle y \rangle \subseteq \langle x \rangle, \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

where $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and $\langle y \rangle \subseteq \langle x \rangle$ states that $\langle y \rangle$ is a subspace of $\langle x \rangle$.

The mutual information $I(X;Y)$ between X and Y can be written as a function of $P_X(x)$ and $P_{Y|X}(y|x)$ as

$$I(X;Y) = \sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y}}} P_X(x)P_{Y|X}(y|x)\log_2\left(\frac{P_{Y|X}(y|x)}{P_Y(y)}\right). \tag{7}$$

It is clear from (6) that $P_{Y|X}(y|x_1) = P_{Y|X}(y|x_2)$ for all $x_1, x_2 \in \mathcal{X}$ such that $\langle x_1 \rangle = \langle x_2 \rangle$. Similarly the transition probabilities in (6) can be rewritten as

$$P_{Y|X}(y|\pi_x) = \begin{cases} q^{-n\dim(\pi_x)} & \langle y \rangle \subseteq \pi_x, \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

where $P_{Y|X}(y|\pi_x) \triangleq \Pr(Y=y|\langle X \rangle = \pi_x)$. Here $\pi_x \in \widetilde{\mathcal{X}}$ where

$$\widetilde{\mathcal{X}} \triangleq \bigcup_{i=0}^{T} \text{Gr}\left(i, \mathbb{F}_q^T\right).$$

Note that with an abuse of notation we have used $P_{Y|X}(y|\cdot)$ to denote two different functions (6) and (8). These two properties allow us to express the mutual information in (7) as stated in the following lemma

Lemma 1: Finding the capacity of the channel in (1) is equivalent to maximizing

$$I(X;Y) = \sum_{\substack{\pi_x\in\widetilde{\mathcal{X}},\\ y\in\mathcal{Y}}} P_X(\pi_x)P_{Y|X}(y|\pi_x)\log_2\left(\frac{P_{Y|X}(y|\pi_x)}{P_Y(y)}\right), \tag{9}$$

over all choices of $P_X(\pi_x) \triangleq \Pr(\langle X \rangle = \pi_x)$.

The following lemma states that the optimal solution should be uniform over all subspaces with the same dimension, as expected from the symmetry of the channel.

Lemma 2: The input distribution that maximizes $I(X;Y)$ is the one which is uniform over all subspaces having the same dimension.

Proof: Let $P_X(\pi_x)$ be the optimal input distribution of the channel with transition probabilities given in (8). For a fix dimension $0 \leq d \leq \min(m,T)$, and an arbitrary permutation

$$\sigma : \{1,2,\ldots,\mathcal{G}(T,d)\} \to \{1,2,\ldots,\mathcal{G}(T,d)\}$$

which acts on subspaces of dimension d, define $P_\sigma(\pi_x)$ as

$$P_\sigma(\pi_x) = \begin{cases} P_X(\sigma(\pi_x)) & \text{if } \dim(\pi_x) = d, \\ P_X(\pi_x) & \text{if } \dim(\pi_x) \neq d. \end{cases}$$

Also define $P^*(\pi_x) = \frac{1}{\mathcal{G}(T,d)!}\sum_\sigma P_\sigma(\pi_x)$ where the summation is over all possible permutations. Rewriting the mutual information in (9) as a function of the input distribution and the transition probabilities, $I(P_X(\pi_x), P_{Y|X}(y|\pi_x))$, we have

$$\begin{aligned} I(P^*(\pi_x),&P_{Y|X}(y|\pi_x)) \\ &= I\left(\frac{1}{\mathcal{G}(T,d)!}\sum_\sigma P_\sigma(\pi_x), P_{Y|X}(y|\pi_x)\right) \\ &\overset{(a)}{\geq} \frac{1}{\mathcal{G}(T,d)!}\sum_\sigma I(P_\sigma(\pi_x), P_{Y|X}(y|\pi_x)) \\ &\overset{(b)}{=} I(P_X(\pi_x), P_{Y|X}(y|\pi_x)) \end{aligned}$$

where (a) is due to concavity of the mutual information with respect to the input distribution, and (b) holds because $I(P_\sigma(\pi_x), P_{Y|X}(y|\pi_x)) = I(P_X(\pi_x), P_{Y|X}(y|\pi_x))$ for all σ, since the permutation only permutes the terms in a summation in (9).

Note that $P^*(\pi_x)$ assigns equal probabilities to all subspaces with dimension d, and above-mentioned inequality shows that it is as good as the optimal input distribution. A similar argument holds for all $0 \leq d \leq \min(m,T)$. Therefore, a dimensional-uniform distribution achieves the capacity of the channel. ■

Lemma 2 shows that the optimal input distribution can be expressed as

$$\Pr(\langle X \rangle = \pi_x) = \frac{\alpha_r}{\mathcal{G}(T,r)}, \tag{10}$$

where $r = \dim(\pi_x)$ and $\alpha_r = \Pr(\dim(\langle X \rangle) = r)$ and we have $\sum_{r=0}^{\min(m,T)} \alpha_r = 1$.

Assuming the optimal input probability distribution of the form (10), the probability of receiving a specific matrix $Y = y$ at the receiver can be written as

$$P_Y(y) = \frac{1}{\mathcal{G}(T, d_y)} \sum_{d_x=d_y}^{\min(m,T)} \mathcal{G}(d_x, d_y) q^{-nd_x} \alpha_{d_x}, \qquad (11)$$

that shows $P_Y(y)$ only depends on $d_y = \dim(\langle y \rangle)$. Therefore, having $P_Y(y)$ only depends on d_y and replacing (10) in (9), we get

$$\begin{aligned} I = &- \sum_{d_x=0}^{\min(m,T)} \alpha_{d_x} n d_x \log_2 q \qquad (12) \\ &- \sum_{d_x=0}^{\min(m,T)} \left[\alpha_{d_x} \sum_{d_y=0}^{\min(n,d_x)} S_{d_y} \mathcal{G}(d_x, d_y) q^{-nd_x} \log_2 (P_Y(y)) \right], \end{aligned}$$

where S_{d_y} is the number of different $n \times T$ matrices over $\mathbb{F}_q$ that their rows span a specific subspace $\pi \in \mathbb{F}_q^T$ with dimension $0 \le d_y \le \min(n, T)$.

B. The Optimal Solution: Approach

As stated in the last subsection, the problem of finding the optimal input distribution is reduced to finding the optimal choice for α_i, $i = 0, \ldots, \min(m, T)$. Note that the mutual information is a concave function with respect to α_i's. This allows us to use the Kuhn-Tucker theorem [5] to solve the convex optimization problem. According to this theorem, the maximizing values, denoted by α_i^* satisfy

$$\begin{cases} \left.\frac{\partial I(X;Y)}{\partial \alpha_k}\right|_{\alpha_i^*} = \lambda \quad \forall k : \ \alpha_k^* > 0, \\ \\ \left.\frac{\partial I(X;Y)}{\partial \alpha_k}\right|_{\alpha_i^*} \le \lambda \quad \forall k : \ \alpha_k^* = 0, \end{cases} \qquad (13)$$

for some constant λ where $\sum_{i=0}^{\min(m,T)} \alpha_i^* = 1$.

By taking the partial derivative of the mutual information with respect to α_k, we have

$$\begin{aligned} I'_k \triangleq & \frac{\partial I(X;Y)}{\partial \alpha_k} = -nk \log_2 q \qquad (14) \\ &- \sum_{d_y=0}^{\min(n,k)} S_{d_y} \mathcal{G}(k, d_y) q^{-nk} \log_2 (P_Y(y)) - \log_2 e, \end{aligned}$$

where $P_Y(y)$ is given in (11). Multiplying both sides of (14) by α_k and summing over k we get

$$I - \log_2 e = \sum_{k=0}^{\min(m,T)} \alpha_k I'_k.$$

By choosing the optimal values $\alpha_k = \alpha_k^*$ for $0 \le k \le \min(m, T)$, the RHS becomes λ, and the mutual information increases to C. So we may write

$$\lambda = C - \log_2 e.$$

C. Solution for Large Field Size

For the rest of this paper, we focus on large size fields, $q \gg 1$. This assumption allows us to use some approximations to simplify the conditions in (13). For example, by absorbing $\log_2 e$ in λ, one can rewrite $\tilde{I}_k \triangleq I'_k + \log_2 e$ for large q as

$$\begin{aligned} \tilde{I}_k = & -nk \log_2 q \qquad (15) \\ &- \sum_{d_y=0}^{\min(n,k)} \left(1 + \mathcal{O}(q^{-1})\right) q^{-(n-d_y)(k-d_y)} \log_2 (P_Y(y)), \end{aligned}$$

where we have used the asymptotic expressions $\mathcal{G}(k, d_y) = q^{d_y(k-d_y)} \left(1 + \mathcal{O}(q^{-1})\right)$ and $S_{d_y} = q^{nd_y} \left(1 + \mathcal{O}(q^{-1})\right)$ [7]. Using similar approximations, $\log_2 P_Y(y)$ in (11) can be rewritten as

$$\begin{aligned} \log_2 (P_Y(y)) = & - d_y T \log_2 q + \mathcal{O}(q^{-1}) \\ & + \log_2 \left(\sum_{d_x=d_y}^{\min(m,T)} q^{-(n-d_y)d_x} \alpha_{d_x} \right) \\ = & \Theta(\log q). \qquad (16) \end{aligned}$$

Using (16) one can conclude that the dominating term in the summation in (15) is the one obtained for $d_y = \min(n, k)$. Since, the remaining terms are of order $q^{-1} \log q$, we can write

$$\begin{aligned} \tilde{I}_k = & [T \min(n,k) - nk] \log_2 q + \mathcal{O}(q^{-1} \log q) \\ & - \log_2 \left(\sum_{d_x=\min(n,k)}^{\min(m,T)} q^{-[n-\min(n,k)]d_x} \alpha_{d_x} \right). \qquad (17) \end{aligned}$$

Therefore, the Kuhn-Tucker conditions can be rewritten as

$$\sum_{d_x=\min(n,k)}^{\min(m,T)} q^{-[n-\min(n,k)]d_x} \alpha_{d_x} \ge 2^{-C+\mathcal{O}(q^{-1} \log q)} q^{[T \min(n,k) - nk]},$$

where the inequality holds with equality for all k with $\alpha_k^* > 0$.

Let $\delta \triangleq \min(m, T)$ and define the $(\delta + 1) \times (\delta + 1)$ matrix $\mathbf{A}$ with elements

$$\mathbf{A}_{ij} \triangleq \begin{cases} q^{-[n-\min(n,i)]j} & \min(n,i) \le j \le \delta, \\ 0 & \text{otherwise.} \end{cases} \qquad (18)$$

We also define the column vector $\mathbf{b}$ with elements $\mathbf{b}_i \triangleq q^{[T \min(n,i) - ni]}$ for $0 \le i \le \delta$. Note that for convenience the indices of matrix $\mathbf{A}$ and vector $\mathbf{b}$ start from 0. Using these definitions, we are able to rewrite the Kuhn-Tucker conditions in the matrix form as

$$\mathbf{A} \boldsymbol{\alpha}^* \succeq 2^{-C+\mathcal{O}(q^{-1} \log q)} \mathbf{b}, \qquad (19)$$

where we use "$\succeq$" to denote element-wise inequality for the vectors, and $\boldsymbol{\alpha}^*$ is the vector of the optimum probabilities of choosing subspaces of certain dimension. In the following, we consider two cases for $\delta \le n$ and $\delta > n$, and find $\boldsymbol{\alpha}^*$ for each of them, separately.

First case: $\delta \le n$**.** In this case we can explicitly write the matrix $\mathbf{A}$ and vector $\mathbf{b}$ as

$$\mathbf{A} = \begin{bmatrix} 1 & q^{-n} & \cdots & q^{-(\delta-1)n} & q^{-\delta n} \\ 0 & q^{-(n-1)} & \cdots & q^{-(\delta-1)(n-1)} & q^{-\delta(n-1)} \\ 0 & 0 & \cdots & q^{-(\delta-1)(n-2)} & q^{-\delta(n-2)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & q^{-(\delta-1)(n-\delta+1)} & q^{-\delta(n-\delta+1)} \\ 0 & 0 & \cdots & 0 & q^{-\delta(n-\delta)} \end{bmatrix},$$

and

$$\mathbf{b} = \begin{bmatrix} 1 & q^{(T-n)} & \cdots & q^{\delta(T-n)} \end{bmatrix}^{\mathrm{T}}.$$

The fact that the expression inside the $\log(\cdot)$ function in (17) is non-zero for $k = \delta$, forces α^*_δ to be positive. Thus the last row of the matrix inequality in (19) should be satisfied as an equality. Therefore,

$$\alpha^*_\delta = q^{\delta(T-\delta)} 2^{-C+\mathcal{O}(q^{-1}\log q)}.$$

The following lemma helps to find the behavior of the optimal input distribution. We include the proof in [7].

Lemma 3: Let $\delta \le n$ and $\boldsymbol{\alpha}^*$ be the optimal solution of the Kuhn-Tucker conditions in (19). Then $\alpha^*_j > 0$ implies $\alpha^*_i > 0$ for $j \le i \le \delta$.

Using this lemma, it is easy to verify that there exists some $0 \le \kappa \le \delta$, where the inequalities in (19) indexed by $\kappa \le j \le \delta$ hold as equality. Moreover, $\alpha^*_i = 0$ for $0 \le i < \kappa$. Therefore, one can solve the set of equations recursively, and show that

$$\alpha^*_i = \begin{cases} q^{i(T-i)} 2^{-C+\mathcal{O}(q^{-1}\log q)} & : \ \kappa \le i \le \delta, \\ 0 & : \ 0 \le i < \kappa, \end{cases} \tag{20}$$

and it only remains to determine κ. Since $\alpha^*_{\kappa-1} = 0$, we can rewrite the inequality indexed by $\kappa - 1$ as

$$\sum_{j=\kappa}^{\delta} q^{-(n-\kappa+1)j} \alpha^*_j \ge q^{(\kappa-1)(T-n)} 2^{-C+\mathcal{O}(q^{-1}\log q)}.$$

Replacing α^*_j from (20), we get

$$q^{(\kappa-1)(T-n)} 2^{-C+\mathcal{O}(q^{-1}\log q)} \left[\sum_{j=\kappa}^{\delta} q^{(T-n-j)(j-\kappa+1)} - 1 \right] \ge 0,$$

which holds if and only if $(T-n-j)|_{j=\kappa} \ge 0$. Since κ is the largest ℓ where $\alpha^*_{\ell-1} = 0$, we have $\kappa = \min[(T-n)^+, \delta]$.

Second case: $\delta > n$**.** We now write matrix $\mathbf{A}$ and vector $\mathbf{b}$ as

$$\mathbf{A} = \left[\begin{array}{ccccccc} 1 & q^{-n} & \cdots & \cdots & \cdots & \cdots & q^{-\delta n} \\ 0 & q^{-(n-1)} & \cdots & \cdots & \cdots & \cdots & q^{-\delta(n-1)} \\ \vdots & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & q^{-(n-1)} & q^{-n} & \cdots & q^{-\delta} \\ \hline 0 & \cdots & 0 & 0 & 1 & \cdots & 1 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 1 \end{array}\right],$$

and

$$\mathbf{b} = \left[1 \ q^{(T-n)} \ \ldots \ q^{(n-1)(T-n)} \ q^{n(T-n)} \ q^{n(T-n-1)} \ \ldots \ q^{n(T-\delta)} \right]^{\mathrm{T}}.$$

The last $\delta-n+1$ rows of $\mathbf{A}$ are the same while b_i is decreasing with i for $i \ge n$. Thus, the last $\delta - n$ inequalities are strict and therefore,

$$\alpha^*_{n+1} = \cdots = \alpha^*_\delta = 0. \tag{21}$$

The remaining equations can simply be reduced to the fist case. Define

$$\tilde{\mathbf{A}} = \begin{bmatrix} 1 & q^{-n} & \cdots & q^{-(n-1)n} & q^{-n^2} \\ 0 & q^{-(n-1)} & \cdots & q^{-(n-1)(n-1)} & q^{-n(n-1)} \\ 0 & 0 & \cdots & q^{-(n-1)(n-2)} & q^{-n(n-2)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & q^{-(n-1)} & q^{-n} \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix},$$

and

$$\tilde{\mathbf{b}} = \begin{bmatrix} 1 & q^{(T-n)} & \cdots & q^{n(T-n)} \end{bmatrix}^{\mathrm{T}}.$$

The remaining conditions in this case can be written as

$$\tilde{\mathbf{A}}\alpha^* \succeq 2^{-C+\mathcal{O}(q^{-1}\log q)} \tilde{\mathbf{b}}, \tag{22}$$

which is exactly similar to (19), for $\delta = n$. Therefore, the optimal solution for the first case will also satisfy these conditions, *i.e.*,

$$\alpha^*_i = \begin{cases} q^{i(T-i)} 2^{-C+\mathcal{O}(q^{-1}\log q)} & \kappa \le i \le n, \\ 0 & 0 \le i < \kappa, \end{cases} \tag{23}$$

with $\kappa = \min[(T-n)^+, n]$. Summarizing (21) and (23), we can obtain the optimal solution for this regime, as

$$\alpha^*_i = \begin{cases} 0 & n < i \le \delta, \\ q^{i(T-i)} 2^{-C+\mathcal{O}(q^{-1}\log q)} & \kappa \le i \le n, \\ 0 & 0 \le i < \kappa, \end{cases}$$

where $\kappa = \min[(T-n)^+, n]$. This concludes the proof.

References

[1] S.-Y. R. Li, N. Cai, and R. W. Yeung, "Linear network coding", *IEEE Transactions on Information Theory*, Volume 49, Feb. 2003.

[2] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding", *Allerton Conference on Communication, Control, and Computing*, IL, October 2003.

[3] R. Koetter and F. Kschischang, "Coding for errors and erasures in random network coding", *IEEE Transactions on Information Theory*, Volume 54, Issue 8, August 2008.

[4] M. Jafari Siavoshani, C. Fragouli, and S. Diggavi, "Passive topology discovery for network coded systems", *Information Theory Workshop (ITW)*, Bergen, Norway, July 2007.

[5] S. Boyd and L. Vandenberghe, "Convex Optimization", *Cambridge University Press*, 2004.

[6] M. Jafari Siavoshani, C. Fragouli, and S. Diggavi, "Noncoherent multi-source network coding", *IEEE International Symposium on Information Theory (ISIT)*, Page(s) 817–821, Canada, Toronto, July 2008.

[7] M. Jafari, S. Mohajer, C. Fragouli, and S.Diggavi, "Non-Coherent Network Coding Capacity", *EPFL Technical Report*, December 2008.

[8] D. Silva and F. R. Kschischang, "Using rank-metric codes for error correction in random network coding", *IEEE International Symposium on Information Theory*, Nice, France, June 2007.

[9] A. Montanari and R. Urbanke, "Coding for network coding", December 2007, available online : http://arxiv.org/abs/0711.3935/.

[10] D. Silva, F. R. Kschischang, and R. Koetter, "Capacity of Random Network Coding under a Probabilistic Error Model", July 2008, *available online at http://arxiv.org/pdf/0807.1372/.*

[11] K. Price and R. Storn, "Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces", *Journal of Global Optimization*, Volume 11, Page 341–359, 1997.