Towards an

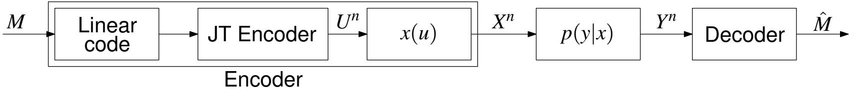
Algebraic Network Information Theory : Technical Lemmas

Sung Hoon Lim, Chen Feng, Adriano Pastore,
Bobak Nazer, Michael Gastpar

KIOST — UBC — CTTC — BU — EPFL

CISS 2018, Princeton, N.J., U.S.A. — March 23

Linear Coding + Multicoding Architecture



- Three components
  - (Auxiliary) linear code
  - Joint typicality encoder
  - Symbol-by-symbol mapping $x(u)$

Outline

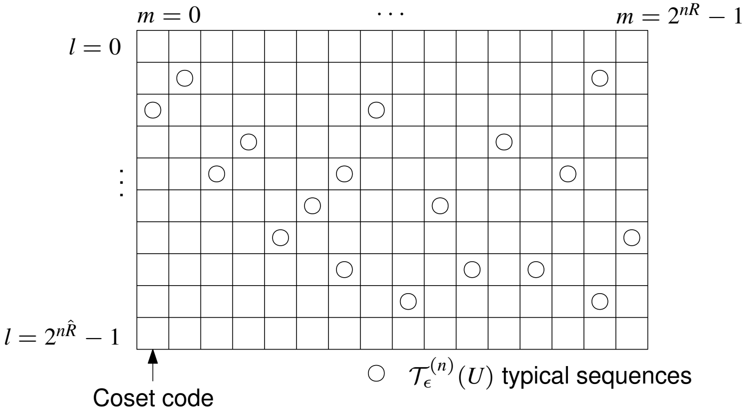
“Mismatched Typicality”

Consider a random coding argument where :

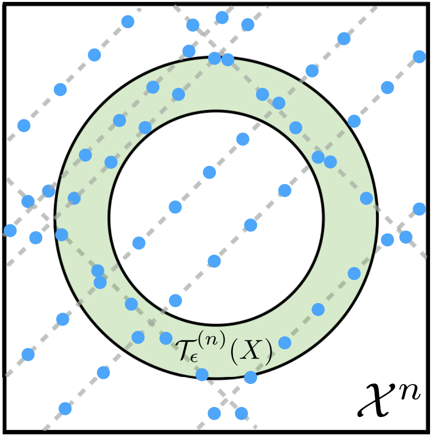
- first, a **base codebook** is drawn in such a way that every pair of codewords is drawn pairwise independently and that the (marginal) distribution of each codeword is IID $\prod \tilde{p}_X(\cdot)$
- then, in that codebook, we only actually use those codewords that lie in the typical set of a different distribution $p(x)$.

Note: The usual typicality argument simply has $p_X(x) = \tilde{p}_X(x)$.

Random Linear Codebooks



Random Linear Codebooks



Random Linear Codes

"Mismatched Typicality"

Lemma (Mismatched Typicality Lemma)

Let $X \sim p_X(x)$ *and let* $\tilde{p}_X(x)$ *be another distribution on* $\mathcal{X}$ *such that* $D_X = D(p_X \| \tilde{p}_X) < \infty$. *Then, for* $x^n \in \mathcal{T}_\epsilon^{(n)}(X)$,

$$2^{-n(D_X + H(X) + \delta(\epsilon))} \leq \prod_{i=1}^{n} \tilde{p}_X(x_i) \leq 2^{-n(D_X + H(X) - \delta(\epsilon))}.$$

Note: The usual typicality argument simply has $p_X(x) = \tilde{p}_X(x)$.

"Mismatched Typicality"

To prove the first statement, observe that, $\prod_{i=1}^{n} \tilde{p}_X(x_i) = \prod_{x \in \mathcal{X}} \tilde{p}_X(x)^{n\pi(x|x^n)}$, where $\pi(x|x^n)$ is the empirical pmf of x^n. Then,

$$\begin{aligned}
\log \tilde{p}_X(x^n) &= \sum_{x \in \mathcal{X}} n\pi(x|x^n) \log \tilde{p}_X(x) \\
&= \sum_{x \in \mathcal{X}} n(\pi(x|x^n) - p_X(x) + p_X(x)) \log \tilde{p}_X(x) \\
&= n \sum_{x \in \mathcal{X}} p_X(x) \log \tilde{p}_X(x) - n \sum_{x \in \mathcal{X}} (\pi(x|x^n) - p_X(x)) (-\log \tilde{p}_X(x)) \\
&= -n (D(p_X \| \tilde{p}_X) + H(X)) - n \sum_{x \in \mathcal{X}} (\pi(x|x^n) - p_X(x)) (-\log \tilde{p}_X(x))
\end{aligned}$$

"Mismatched Typicality"

Since $x^n \in \mathcal{T}_\epsilon^{(n)}(X)$,

$$\begin{aligned}
\left|\sum_{x\in\mathcal{X}} \left(\pi(x|x^n) - p_X(x)\right)\left(-\log \tilde{p}_X(x)\right)\right| &\le \sum_{x\in\mathcal{X}} \left|\pi(x|x^n) - p_X(x)\right|\left(-\log \tilde{p}_X(x)\right) \\
&\le -\epsilon \sum_{x\in\mathcal{X}} p_X(x)\log \tilde{p}_X(x) \\
&= \epsilon\left(D(p_X\|\tilde{p}_X) + H(X)\right)
\end{aligned}$$

"Mismatched Typicality"

Lemma (Mismatched Joint Typicality Lemma)

Let $(X, Y) \sim p_{X,Y}(x, y)$ and $\tilde{p}_X(x)$ be another distribution on $\mathcal{X}$ such that $D(p_X \| \tilde{p}_X) < \infty$. Let $\epsilon' < \epsilon$. Then, there exists $\delta(\epsilon) > 0$ that tends to zero as $\epsilon \to 0$ such that the following statement holds:

1. *If $\tilde{y}^n$ is an arbitrary sequence and $\tilde{X}^n \sim \prod_{i=1}^n \tilde{p}_X(\tilde{x}_i)$, then*

$$\mathsf{P}\{(\tilde{X}^n, \tilde{y}^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y)\} \leq 2^{-n(I(X;Y)+D(p_X\|\tilde{p}_X)-\delta(\epsilon))}$$

2. *If $\tilde{y}^n \in \mathcal{T}_{\epsilon'}^{(n)}(Y)$ and $\tilde{X}^n \sim \prod_{i=1}^n \tilde{p}_X(\tilde{x}_i)$, then for n sufficiently large,*

$$\mathsf{P}\{(\tilde{X}^n, \tilde{y}^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y)\} \geq 2^{-n(I(X;Y)+D(p_X\|\tilde{p}_X)+\delta(\epsilon))}$$

The proof follows from the Mismatched Typicality Lemma and standard cardinality bounds on the conditional typical set $\mathcal{T}_\epsilon^{(n)}(X|y^n)$.

Packing Lemma

Packing Lemma for mismatched distributions

- $(X, Y) \sim p_{X,Y}(x, y)$
- $\tilde{p}_X(x)$ is another distribution on $\mathcal{X}$
- $\tilde{Y}^n$ be an arbitrarily distributed random sequence
- Codebook $\mathcal{C}$: $\tilde{X}^n(m) \sim \prod_{i=1}^n \tilde{p}_X(\tilde{x}_i)$, $m \in [2^{nR}]$
- Codewords in $\mathcal{C}$ are pairwise independent of Y^n

Then,

$$\lim_{n \to \infty} \mathsf{P}\{(\tilde{X}^n(m), \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y) \text{ for some } m \in \mathcal{C}\} = 0,$$

if $R < I(X; Y) + D(p_X \| \tilde{p}_X) - \delta(\epsilon)$

Covering Lemma

Covering Lemma for mismatched distributions

- $(X, \hat{X}) \sim p_{X,\hat{X}}(x, \hat{x})$
- $\tilde{p}_{\hat{X}}(\hat{x})$ is another distribution on $\hat{\mathcal{X}}$
- X^n is a random sequence with $\lim_{n\to\infty} \mathsf{P}\{X^n \in \mathcal{T}_\epsilon^{(n)}(X)\} = 1$
- Codebook $\mathcal{C}$: $\tilde{X}^n(m) \sim \prod_{i=1}^n \tilde{p}_{\hat{X}}(\hat{x}_i)$, $m \in [2^{nR}]$
- Codewords in $\mathcal{C}$ are pairwise independent and independent of X^n

Then,

$$\lim_{n\to\infty} \mathsf{P}\{(X^n, \tilde{X}^n(m)) \in \mathcal{T}_\epsilon^{(n)}(X, \hat{X}) \text{ for some } m \in \mathcal{C}\} = 1,$$

if $R > I(X; \hat{X}) + D(p_X \| \tilde{p}_X) + \delta(\epsilon)$

Covering Lemma — Proof

Let $\mathcal{A} = \{m \in [1:2^{nR}] : (X^n, \tilde{X}^n(m)) \in \mathcal{T}_\epsilon^{(n)}(X, \hat{X})\}$. Then, by the Chebyshev lemma,

$$\mathsf{P}\{|\mathcal{A}| = 0\} \leq \frac{\operatorname{Var}(|\mathcal{A}|)}{(\mathsf{E}\,|\mathcal{A}|)^2}.$$

For $m \in [1:2^{nR}]$, define the indicator random variables

$$E(m) = \begin{cases} 1 & \text{if } (X^n, \tilde{X}^n(m)) \in \mathcal{T}_\epsilon^{(n)}(X, \hat{X}), \\ 0 & \text{otherwise,} \end{cases}$$

and let $p_1 := \mathsf{P}\{E(1) = 1\}$ and $p_2 := \mathsf{P}\{E(1) = 1, E(2) = 1\} = p_1^2$.

Covering Lemma — Proof

Then,

$$\begin{aligned}
\mathsf{E}(|\mathcal{A}|) &= \sum_m \mathsf{P}\{(X^n, \tilde{X}(m)) \in \mathcal{T}_\epsilon^{(n)}(X, \hat{X})\} = 2^{nR} p_1, \\
\mathsf{E}(|\mathcal{A}|^2) &= \sum_m \mathsf{P}\{(X^n, \tilde{X}(m)) \in \mathcal{T}_\epsilon^{(n)}(X, \hat{X})\} \\
&\quad + \sum_m \sum_{m' \neq m} \mathsf{P}\{(X^n, \tilde{X}(m)) \in \mathcal{T}_\epsilon^{(n)}(X, \hat{X}), (X^n, \tilde{X}(m')) \in \mathcal{T}_\epsilon^{(n)}(X, \hat{X}) \\
&\le 2^{nR} p_1 + 2^{n2R} p_2 = 2^{nR} p_1 + 2^{n2R} p_1^2.
\end{aligned}$$

Thus, $\mathrm{Var}(|\mathcal{A}|) \le 2^{nR} p_1$.

Covering Lemma — Proof

From the Joint Typicality Lemma, for sufficiently large n, we have

$$p_1 \leq 2^{-n(I(X;\hat{X})+D(p_X\|\tilde{p}_X)-\delta(\epsilon))},$$
$$p_1 \geq 2^{-n(I(X;\hat{X})+D(p_X\|\tilde{p}_X)+\delta(\epsilon))},$$

and hence,

$$\frac{\mathrm{Var}(|\mathcal{A}|)}{(\mathsf{E}\,|\mathcal{A}|)^2} \leq \frac{1}{2^{nR}p_1} \leq 2^{-n(R-I(X;\hat{X})-D(p_X\|\tilde{p}_X)-\delta(\epsilon))},$$

which tends to zero as $n \to \infty$ if

$$R > I(X;\hat{X}) + D(p_X\|\tilde{p}_X) + \delta'(\epsilon).$$

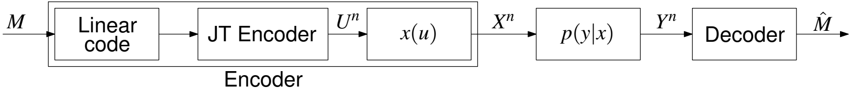
Outline

Mismatched Typicality

Nested Linear Codes

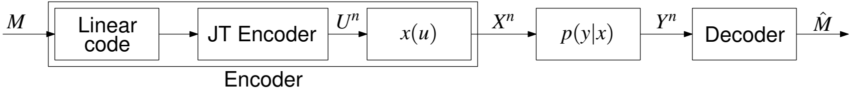
A Markov Lemma

Linear Coding + Multicoding Architecture



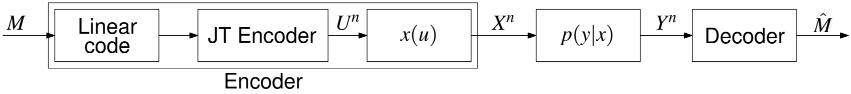
- Three components
  - (Auxiliary) linear code
  - Joint typicality encoder
  - Symbol-by-symbol mapping $x(u)$

Linear Coding + Multicoding Architecture



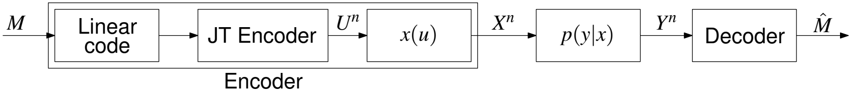
- Messages $m \in [2^{nR}]$, auxiliary indices $l \in [2^{n\hat{R}}]$

Linear Coding + Multicoding Architecture



- Messages $m \in [2^{nR}]$, auxiliary indices $l \in [2^{n\hat{R}}]$
- Represented in $\mathbb{F}_\mathsf{q}$: $\big[\boldsymbol{\nu}(m), \boldsymbol{\nu}(l)\big] \in \mathbb{F}_\mathsf{q}^\kappa$

Linear Coding + Multicoding Architecture



- Messages $m \in [2^{nR}]$, auxiliary indices $l \in [2^{n\hat{R}}]$
- Represented in $\mathbb{F}_\mathsf{q}$: $\big[\boldsymbol{\nu}(m), \boldsymbol{\nu}(l)\big] \in \mathbb{F}_\mathsf{q}^\kappa$
- Codebook construction:

$$u^n(m,l) = \big[\boldsymbol{\nu}(m), \boldsymbol{\nu}(l)\big]\,\mathsf{G} \oplus d^n, \quad m \in [2^{nR}], l \in [2^{n\hat{R}}]$$

- Generator matrix $\mathsf{G} \in \mathbb{F}_\mathsf{q}^{\kappa \times n}$, $g_{ij} \sim p_\mathsf{q}(g_{ij}) := \mathrm{Unif}(\mathbb{F}_\mathsf{q})$
- Dither $d^n \in \mathbb{F}_\mathsf{q}^n$, $d_i \sim p_\mathsf{q}(d_i)$

Joint Typicality Encoding

- (Almost) all codewords are typical in the uniform typical set

$$u^n(m, l) \in \mathcal{T}_\epsilon^{(n)}(p_{\mathsf{q}})$$

Joint Typicality Encoding

- (Almost) all codewords are typical in the uniform typical set

$$u^n(m,l) \in \mathcal{T}_\epsilon^{(n)}(p_{\mathsf{q}})$$

- "Shaping": Use codewords that are typical with respect to p_U

Joint Typicality Encoding

- (Almost) all codewords are typical in the uniform typical set

$$u^n(m,l) \in \mathcal{T}_{\epsilon}^{(n)}(p_{\mathsf{q}})$$

- "Shaping": Use codewords that are typical with respect to p_U

Joint typicality encoding

Fix $p(u)$ and $x(u)$. For each m, find an index l such that $u^n(m,l) \in \mathcal{T}_{\epsilon'}^{(n)}(U)$ and transmit $x_i = x(u_i(m,l))$:

Joint Typicality Encoding

- (Almost) all codewords are typical in the uniform typical set

$$u^n(m,l) \in \mathcal{T}_{\epsilon}^{(n)}(p_{\mathsf{q}})$$

- "Shaping": Use codewords that are typical with respect to p_U

Joint typicality encoding

Fix $p(u)$ and $x(u)$. For each m, find an index l such that $u^n(m,l) \in \mathcal{T}_{\epsilon'}^{(n)}(U)$ and transmit $x_i = x(u_i(m,l))$: successful w.h.p. if

$$\hat{R} > D(p_U \| p_{\mathsf{q}})$$

Covering Lemma

Covering Lemma for mismatched distributions

- $(X, \hat{X}) \sim p_{X,\hat{X}}(x, \hat{x})$
- $\tilde{p}_{\hat{X}}(\hat{x})$ is another distribution on $\hat{\mathcal{X}}$
- X^n is a random sequence with $\lim_{n\to\infty} \mathsf{P}\{X^n \in \mathcal{T}_\epsilon^{(n)}(X)\} = 1$
- Codebook $\mathcal{C}$: $\tilde{X}^n(m) \sim \prod_{i=1}^n \tilde{p}_{\hat{X}}(\hat{x}_i)$, $m \in [2^{nR}]$
- Codewords in $\mathcal{C}$ are pairwise independent and independent of X^n

Then,

$$\lim_{n\to\infty} \mathsf{P}\{(X^n, \tilde{X}^n(m)) \in \mathcal{T}_\epsilon^{(n)}(X, \hat{X}) \text{ for some } m \in \mathcal{C}\} = 1,$$

if $R > I(X; \hat{X}) + D(p_X \| \tilde{p}_X) + \delta(\epsilon)$

Joint Typicality Decoding

Joint typicality decoding

Find the unique index $\hat{m}$ such that

$$(u^n(\hat{m}, \hat{l}), y^n) \in \mathcal{T}_\epsilon^{(n)}(U, Y)$$

for some $\hat{l}$

Joint Typicality Decoding

Joint typicality decoding

Find the unique index $\hat{m}$ such that

$$(u^n(\hat{m}, \hat{l}), y^n) \in \mathcal{T}_\epsilon^{(n)}(U, Y)$$

for some $\hat{l}$: successful w.h.p. if

$$R + \hat{R} < I(U; Y) + D(p_U \| p_{\mathsf{q}})$$

Joint Typicality Decoding

Joint typicality decoding

Find the unique index $\hat{m}$ such that

$$(u^n(\hat{m}, \hat{l}), y^n) \in \mathcal{T}_\epsilon^{(n)}(U, Y)$$

for some $\hat{l}$: successful w.h.p. if

$$R + \hat{R} < I(U; Y) + D(p_U \| p_{\mathsf{q}})$$

- Joint typicality lemmas for mismatched distributions
- Covering and packing lemmas for mismatched distributions

Linear Coding + Multicoding Architecture

- Eliminate $\hat{R}$ in encoding and decoding conditions

$$\hat{R} > D(p_U \| p_{\mathsf{q}}), \qquad R + \hat{R} < I(U;Y) + D(p_U \| p_{\mathsf{q}})$$

Linear Coding + Multicoding Architecture

- Eliminate $\hat{R}$ in encoding and decoding conditions

$$\hat{R} > D(p_U \| p_{\mathsf{q}}), \qquad R + \hat{R} < I(U;Y) + D(p_U \| p_{\mathsf{q}})$$

Capacity

$$R < \max_{p(u),x(u)} I(U;Y)$$

- Observed by Miyake ('10), Padakandla-Pradhan ('13), in our work, plus probably elsewhere.
- "Shaping" p_X with $p_U = p_X$ and $U = X$
- We only need $\mathsf{q} \geq |\mathcal{X}|$
- Analysis of linear codes for JT encoding/decoding is not so different from analysing IID codes

Outline

Mismatched Typicality

Nested Linear Codes

A Markov Lemma

A Markov Lemma

Given a distribution

$$p(x, u_1, u_2, \dots, u_K) = p(x) \prod_{k=1}^{K} p(u_k|x),$$

and a sequence x^n, consider K encoders, each selecting a codeword index ℓ_k so that

$$(x^n, U_k^n(\ell_k)) \in \mathcal{T}_\epsilon^{(n)}(X, U_k).$$

We would like to infer that

$$(x^n, U_1^n(\ell_1), \dots, U_K^n(\ell_K)) \in \mathcal{T}_\epsilon^{(n)}(X, U_1, \dots, U_K).$$

A Markov Lemma

If we look at a random coding argument ("code construction") for which it can be proved that each of the L codwords is selected uniformly and *independently* from the respective (conditionally) typical sets, we could use Problem 2.9 from Csiszar & Körner's textbook:

Lemma

Let $V_1, \ldots, V_K$ be random variables that are conditionally independent given the random variable X. Then, for sufficiently small $\epsilon' < \epsilon$ and $x^n \in \mathcal{T}_{\epsilon'}^{(n)}(X)$,

$$\lim_{n\to\infty} \frac{\left|\mathcal{T}_{\epsilon'}^{(n)}(V_1|x^n) \times \cdots \times \mathcal{T}_{\epsilon'}^{(n)}(V_K|x^n) \cap \left(\mathcal{T}_{\epsilon}^{(n)}(V_1, \ldots, V_K|x^n)\right)^{\mathsf{c}}\right|}{\left|\mathcal{T}_{\epsilon'}^{(n)}(V_1|x^n) \times \cdots \times \mathcal{T}_{\epsilon'}^{(n)}(V_K|x^n)\right|} = 0.$$

For the nested linear code construction, the generator matrix G is shared between all users. Therefore, this cannot be used directly.

A Markov Lemma

Lemma (Markov Lemma for Nested Linear Codes)

For sufficiently small $\epsilon' < \epsilon$ *and any* $x^n \in \mathcal{T}_{\epsilon'}^{(n)}(X)$*,*

$$\lim_{n\to\infty} \mathsf{P}\{(x^n, U_1^n(L_1), \ldots, U_K^n(L_k)) \in \mathcal{T}_{\epsilon}^{(n)}(X, U_1, \ldots, U_K)\} = 1,$$

if

$$\hat{R}_k > I(U_k; X) + D(p_{U_k} \| p_{\mathsf{q}}) + \delta(\epsilon'), \quad k \in [1:K].$$

A Markov Lemma

We prove this by establishing that :

- When the indices $L_1, L_2, \ldots, L_K$, expressed as vectors over $\mathbb{F}_{\mathsf{q}}^n$, are *linearly independent*, then even though we use the same generator matrix, the codewords are chosen indepedently and uniformly.
- Then, we show that "there are not too many cases" where the indices are not independent.

A Markov Lemma

Let $\mathcal{S}_k$ be a subset of $\mathbb{F}_{\mathsf{q}}^n$. For any subset $\mathcal{S}$ of $\mathcal{S}_1 \times \cdots \times \mathcal{S}_K$, define

$$Z_{\mathcal{S}} := \sum_{(l_1,\ldots,l_K)} \mathbf{1}((U_1^n(l_1), \ldots, U_K^n(l_K)) \in \mathcal{S}),$$

i.e., the number of codeword tuples that fall in $\mathcal{S}$. Since the codewords are uniformly distributed, the mean of $Z_{\mathcal{S}}$ is

$$\mu_{\mathcal{S}} = \frac{|\mathcal{S}|}{\mathsf{q}^{Kn-(n_1+\cdots+n_K)}},$$

where q^{n_k} is the size of the kth codebook.

A Markov Lemma

Then, we establish via Chebyshev that

$$\mathsf{P}\left\{|Z_{\mathcal{S}} - \mu_{\mathcal{S}}| \geq \frac{\gamma|\mathcal{S}_1|\cdots|\mathcal{S}_K|}{\mathsf{q}^{Kn-(n_1+\cdots+n_K)}}\right\}$$

$$\leq \frac{1}{\gamma^2}\left(\frac{\mathsf{q}^{Kn-(n_1+\cdots+n_K)}}{|\mathcal{S}_1|\cdots|\mathcal{S}_K|} + \mathsf{q}^{K^2}\sum_{t=1}^{K-1}\sum_{1\leq j_1<\cdots<j_t\leq K}\frac{\mathsf{q}^{n-n_{j_1}}}{|\mathcal{S}_{j_1}|}\cdots\frac{\mathsf{q}^{n-n_{j_t}}}{|\mathcal{S}_{j_t}|}\right).$$

The key ingredient is

$$\mathsf{E}(Z_{\mathcal{S}}^2) = \sum_{l_1,\ldots,l_K,\tilde{l}_1,\ldots,\tilde{l}_K} \mathsf{P}\left\{(U_1^n(l_1),\ldots,U_K^n(l_K)) \in \mathcal{S}, (U_1^n(\tilde{l}_1),\ldots,U_K^n(\tilde{l}_K)) \in \mathcal{S}\right\}$$

Some Concluding Thoughts

- *Mismatched typicality* can serve as a first tool to analyze nested linear codes.
- It exactly parallels the standard typicality methodology.
- In a multi-user setting, it appears that a more fine-grained analysis of the (nested linear) code construction is necessary.