On the Expected Complexity of Integer Least-Squares Problems

Babak Hassibi

Department of Electrical Engineering

California Institute of Technology, Pasadena, CA 91125

Joint work with Haris Vikalo

DIMACS Workshop on Signal Processing for Wireless Transmission

October 7-9, 2002

Outline

- **Integer least-squares problem**
  - communications, cryptography, GPS
  - NP-hard
- **Sphere decoding**
- **Expected complexity**
  - $k = 2$ Waring problem, Jacobi's theta function
  - full lattice, subset of a lattice
  - polynomial-time for a wide range of SNRs
- **Simulation results**
- **Discussion and Conclusion**
  - connection to Shannon capacity

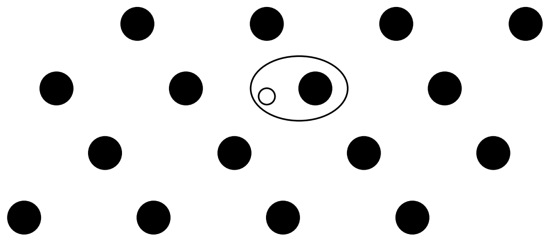
Integer Least-Squares Problems

The following problem is referred to as an *integer least-squares* problem

$$\min_{s \in \mathcal{D} \subset \mathcal{Z}^m} \quad \|x - Hs\|_2 \,,$$

where $x \in \mathcal{R}^n$, $H \in \mathcal{R}^{n \times m}$ and $\mathcal{D}$ is a subset of the integer lattice $\mathcal{Z}^m$.

- Note that the search space is discrete and can be infinite.
- Interpretation: Given the "skewed" lattice Hs, find the "closest" lattice point to a Given n-dimensional vector x.



This problem is well known to be NP-hard.

Applications

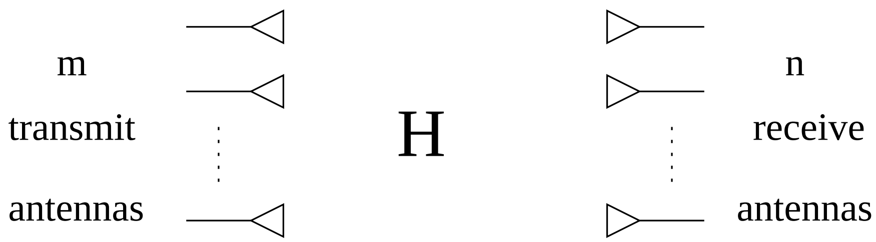
- **Communications**
  - trellis codes, lattice codes (Banihashemi-Khandani '98, Agrell-Vardy-Zeger '02)
  - CDMA systems (Brutel-Boutros '99)
  - fading channels (Viterbo-Boutros '00)
  - multi-antenna systems (Damen-Chkeif-Belfiore '00)
- **Cryptography**
  - lattice-based cryptosystems: Ajtai-Dwork '96, Goldreich-Goldwasser-Halevi '97, tensor-trapdoors '99
  - all these are public-key cryptosystems based on the NP-hardness of various lattice problems
- **GPS**
  - unknowns are *integer* multiples of the transmitted wavelengths and determine the distance to each satellite (Hassibi-Boyd '98)

ML detection leads to integer least-squares problems when:

- the signal constellation is asubset of the lattice
- the noise is AWG

Some examples:

- Multi-antenna systems

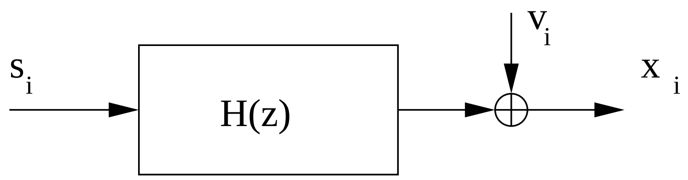


  - V-BLAST: H is the channel matrix and s the transmitted vector
  - linear space-time (LD) codes:

$$S = \sum_{q=1}^{Q} \left(A_q \alpha_q + B_q \beta_q \right),$$

  H is the *equivalent* channel matrix relating the $\{\alpha_q, \beta_q\}$ to the received signals

- FIR channel equalization



$$H = \begin{bmatrix} h_0 & & & \\ h_1 & h_0 & & \\ \vdots & \ddots & \ddots & \\ h_L & \ddots & \ddots & h_0 \\ & h_L & \ddots & h_1 \\ & & \ddots & \vdots \\ & & & h_L \end{bmatrix}$$

where $H(z) = h_0 + h_1 z^{-1} + \ldots + h_L z^{-L}$.

Some Heuristics

All practical systems employ approximations/heuristics:

- Invert and round to the closest integer (zero-forcing equalization):

$$\hat{s}_B = \left[H^\dagger x \right]_{\mathcal{Z}}.$$

  The above $\hat{s}_B$ is called the *Babai* estimate.

- Null and cancel (decision-feedback equalization):
  - only use the Babai estimate for one of the entries of s, say s_1
  - assume that s_1 is known and subtract out its effect to obtain a reduced integer least-squares problem with $m-1$ unknowns
  - solve similarly for s_2, etc.

- Nulling and cancelling with optimal ordering (V-BLAST):
  - perform nulling/cancelling from "strongest" to "weakest" signal

Lattice Reduction

Heuristic methods are exact only if the columns of H are orthogonal.

- Orthogonalizing the columns of H destroys the lattice structure.
- *Lattice reduction* attempts to alleviate this: Find an invertible matrix T, such that T and T^{-1} have integer entries, and such that the matrix $G = HT$ is as "orthogonal as possible". Now solve

$$\min_{t \in \mathcal{Z}^m} \quad \|x - Gt\|_2 \,,$$

  using the earlier heuristics and set $s = T^{-1}t$.
- Lattice reduction itself is NP-hard. A common heuristic is the LLL (Lenstra, Lenstra and Lovasz, 1982) algorithm—essentially, Gram-Schmidt over integers.
- However, lattice reduction is not useful when solving over subsets of a lattice since it destroys the properties of the subset.

How Good are the Heuristics?

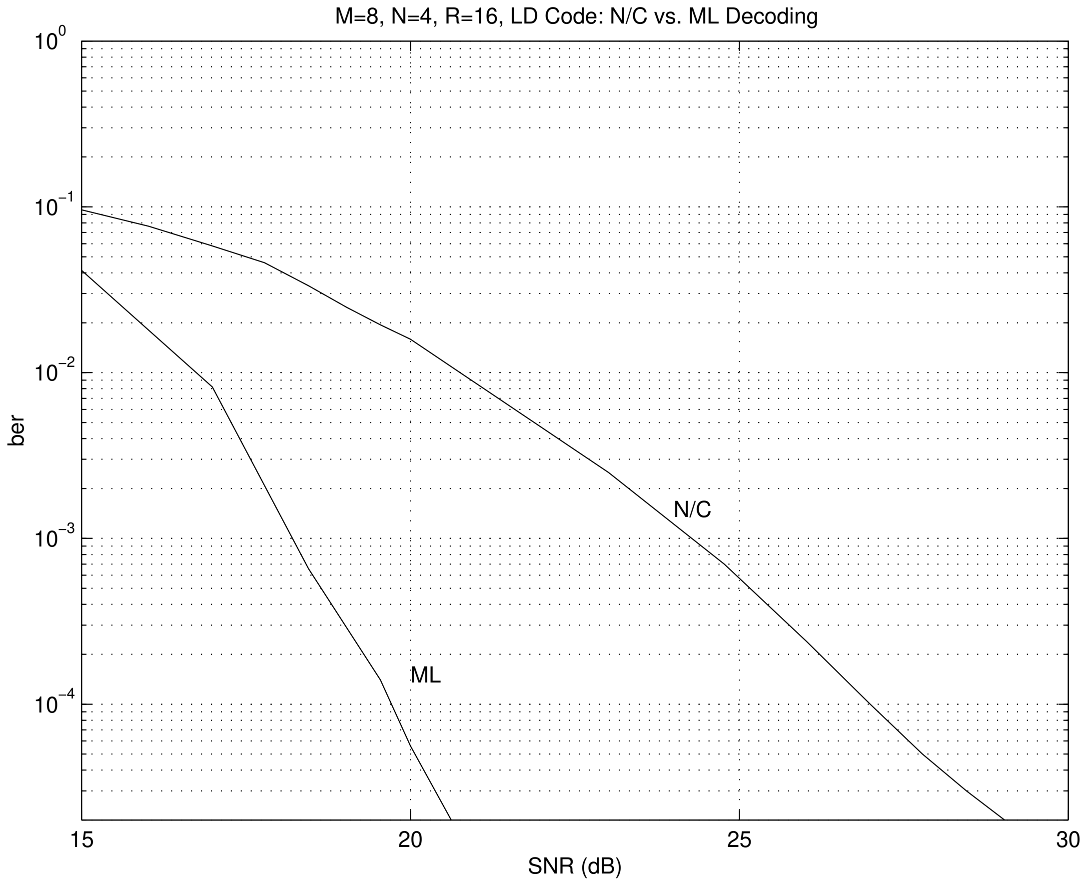
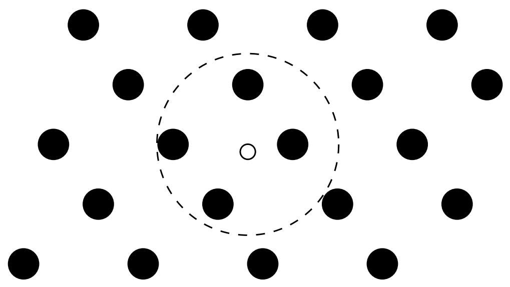


Figure 1: Bit error performance of a rate 16 linear space-time code, corresponding to $m = 64$. ML vs. nulling/cancelling with optimal ordering. (No. of lattice points$= 2^{128} \approx 3.4 \times 10^{38}$).

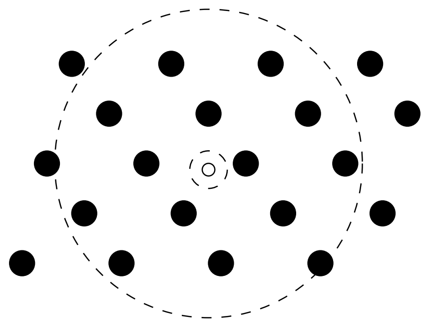
Exact Methods

- All the heuristic methods require $O(m^3)$ computations.
- There exist several exact methods that are a bit more sophisticated than doing a full search over the integer lattice (Kannan's algorithm, KZ algorithm, etc.)
- One is *sphere decoding* (Fincke-Post '85): search only over lattice points lying in a certain hypersphere centered around x.



Seems like a neat idea. But there are two important questions:

1. *How to choose r?* Clearly, if r is too big, we get too many points, but if r is too small, we get no points.



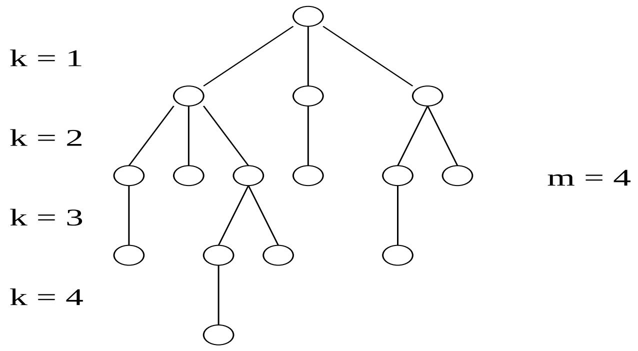
A natural candidate is the:

- *covering radius*, defined to be the smallest radius of spheres centered at the lattice points that cover the entire space

Unfortunately, computing the covering radius is NP-hard.

2. *How can we tell which lattice points are inside the sphere?* If this requires testing each lattice point, then there is no point in sphere decoding.

- When $m = 1$ the answer to the second question is an interval.
- We can use this observation to go from dimension k to $k+1$: for every k-dimensional point in a sphere of radius r, the values of the $k+1$-th dimensional coordinate that lie in the higher dimensional sphere of radius r form an interval.
- Therefore the algorithm searches over all lattice points in spheres of radius r and dimensions $1, 2, \ldots m$.
- The algorithm constructs a tree, where the branches in the k-th level of the tree correspond to the lattice points inside the sphere of radius r and dimension k.



- The complexity of the algorithm depends on the *size* of the tree.

To be more specific, introducing the QR decomposition $H = QR$, the lattice point s lies in a sphere of radius r around x iff,

$$r^2 > \|x - Hs\|^2 = \|x - QRs\|^2 = \|\underbrace{Q^* x}_{y} - Rs\|^2.$$

Expanding this last expression and using the fact that R is upper triangular

$$\begin{aligned} r^2 &> \sum_{i=1}^{M} (y_i - \sum_{j=i}^{m} r_{ij} s_j)^2 \\ &= (y_m - r_{mm} s_m)^2 + (y_{m-1} - r_{m-1,m} s_m - r_{m-1,m-1} s_{m-1})^2 + \ldots \end{aligned}$$

The condition $r^2 > (y_m - r_{mm} s_m)^2$ yields an interval for s_m. For every value of s_m in this interval, the condition

$$r^2 > (y_m - r_{mm} s_m)^2 + (y_{m-1} - r_{m-1,m} s_m - r_{m-1,m-1} s_{m-1})^2,$$

yields an interval for s_{m-1}.

And so on...

The Sphere Decoder Algorithm

Input: $H = QR$, x, $\hat{s} = R^{-1}Q^*x$, r.

1. Set $k = m$, $r_m'^2 = r^2 - \|x\|^2 + \|H\hat{s}\|^2$, $\hat{s}_{m|m+1} = \hat{s}_m$
2. (Bounds for s_k) Set $z = \frac{r_k'}{r_{kk}}$, $UB(s_k) = \lfloor z + \hat{s}_{k|k+1} \rfloor$, $s_k = \lceil -z + \hat{s}_{k|k+1} \rceil - 1$
3. (Increase s_k) $s_k = s_k + 1$. If $s_k \leq UB(s_k)$ go to 5, else to 4.
4. (Increase k) $k = k + 1$ and go to 3.
5. (Decrease k) If $k = 1$ go to 6. Else $k = k - 1$, $\hat{s}_{k|k-1} = \hat{s}_k + \sum_{j=k+1}^{m} \frac{r_{kj}}{r_k k}(s_j - \hat{s}_j)$, $r_k'^2 = r_{k+1}'^2 - r_{k+1,k+1}^2(s_{k+1} - \hat{s}_{k+1|k+2})$.
6. Solution found. Save s_k and go to 3.

A First Look at the Expected Complexity

Here is a very handwavy argument (that can be made rigorous):

- For an arbitrary point x, the expected number of lattice points inside a k-dimensional sphere of radius r is proportional to the volume

  $$\frac{\pi^{k/2}}{\Gamma(k/2+1)} r^k.$$

  Therefore the expected total number of points visited is

  $$\sum_{k=1}^{m} \frac{\pi^{k/2}}{\Gamma(k/2+1)} r^k > \sum_{k=1}^{\frac{m}{2}} \frac{\pi^{k}}{\Gamma(k+1)} r^{2k} \approx e^{\pi r^2}, \quad \text{for large } m.$$

- To have a nonvanishing probability of finding a point in the m-dimensional sphere, its volume must be $\frac{\pi^{m/2}}{(m/2)!} r^m = O(1)$. But from Stirling's formula this implies that $r^2 = O(m)$ and that the *expected* complexity of the algorithm is exponential, $e^{O(m)}$.

A Random Model

- Though not unexpected, this is a discouraging result.

Often, however, the vector x is not arbitrary, but is a lattice point perturbed by additive noise with known statistical properties:

$$x = Hs + v,$$

say, where the entries of v are independent $N(0, \sigma^2)$ random variables.

We further assume, for simplicity, that $m = n$.

- A first by-product is a method to obtain the radius r. Note that $\|v\|^2 = \|x - Hs\|^2$ is a Ξ^2 random variable with $m/2$ degrees of freedom. Then we may choose $r^2 = \alpha m \sigma^2$ so that we find a lattice point with high probability:

$$\int_0^{\alpha m} \frac{\lambda^{m/2-1}}{\Gamma(m/2)} e^{-\lambda} d\lambda = 0.9, \quad \text{say.}$$

  Thus, we determine the radius based on the noise, *not* on the lattice.

Expected Complexity as a Function of σ^2

- Clearly when $\sigma^2 = 0$, the exact solution can be found in $O(m^3)$ time. (The pseudo-inverse does the trick).
- When $\sigma^2 \to \infty$, the received vector x becomes arbitrary, for which we know the expected complexity is exponential.
- But what happens at intermediate noise levels?

This is what we shall study.

We shall also find it useful to assume that the lattice-generating-matrix H is random and comprised of independent $N(0,1)$ entries.

- This makes the problem analytically tractable.
- It is also a very reasonable assumption for large, unstructured, matrices H. (Wigner's semi-circle law, mixing conditions, etc.)

The expected complexity is given by

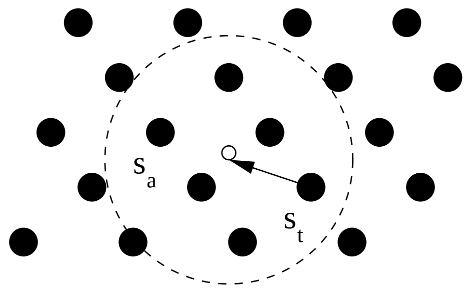
$$\sum_{k=1}^{m} (\text{expected \# of points in } k\text{-sphere of radius } r) \cdot (\text{flops/point})$$

$$= \sum_{k=1}^{m} E_p(k, r^2 = \alpha m \sigma^2) \cdot (2k + 17)$$

- The question is how to compute $E_p(k, r^2)$?

Suppose that the lattice point s_t was transmitted and that the vector $x = Hs_t + v$ was observed. The probability that an arbitrary lattice point s_a lies in a hypersphere of radius r around x can be computed to be

$$\gamma\left(\frac{r^2}{\sigma^2 + \|s_a - s_t\|^2}, \frac{k}{2}\right) = \int_0^{\frac{r^2}{\sigma^2+\|s_a - s_t\|^2}} \frac{\lambda^{k/2-1}}{\Gamma(k/2)} e^{-\lambda} d\lambda.$$



Note that the probability just computed depends only on $\|s_a - s_t\|^2 = \|s\|^2$, i.e., on the squared norm of an arbitrary lattice point in the k-dimensional lattice. It is thus straightforward to see that

$$E_p(k, r^2) = \sum_{n=0}^{\infty} \gamma\left(\frac{r^2}{\sigma^2 + n}, \frac{k}{2}\right) \cdot (\text{\# of lattice points with } \|s\|^2 = n).$$

Since

$$\|s\|^2 = s_1^2 + \ldots + s_k^2,$$

we basically, need to figure out how many ways a non-negative integer n can be represented as the sum of k squared integers.

But this is a classic problem in number theory, related to the $k = 2$ Waring problem.

Representing Integers as Sum of Squares

The number of ways a non-negative integer n can be expressed as the sum of k squared integers is denoted by $r_k(n)$. The problem of computing $r_k(n)$ (and, in fact, a much more general problem) was first posed by Waring in 1770.

- Euler introduced the following (now called Jacobi) theta function

$$\theta(x) = \sum_{m=-\infty}^{\infty} x^{m^2} = 1 + 2\sum_{m=1}^{\infty} x^{m^2},$$

  and noted that

$$\theta^k(x) = 1 + \sum_{n=1}^{\infty} r_k(n) x^n. \qquad (1)$$

- Using the relationship between theta functions and elliptic functions Jacobi showed

$$r_2(n) = 4\left(d_1(n) - d_3(n)\right),$$

where $d_1(n)$ and $d_3(n)$ are the number of divisors of n congruent to 1 and 3 mod 4, respectively. Jacobi also obtained a similar formula for $r_4(n)$.

- Similar methods have been used by Liouville and Eisenstein to compute $r_k(n)$ for $k = 6, 8, 10, 12$. Ramanujan (and Hardy and Littlewood) computed explicit formulas for even $k \leq 24$. For odd k, Dirichlet found $r_3(k)$, and Eisentein, Smith and Minkowski found $r_5(k)$ and $r_7(k)$. But that's about as far as it goes.

- A plethora of asymptotic results (in k and n) are available.

- In anycase, for any given k and n the value of $r_k(n)$ can be numerically computed from (1). It is also a built-in function in *Mathematica*, `SumOfSquaresR[k,n]`.

The Expected Complexity Over the Full Lattice

As a function of m, the lattice dimension, and σ^2, the noise variance, the expected complexity of sphere decoding therefore becomes

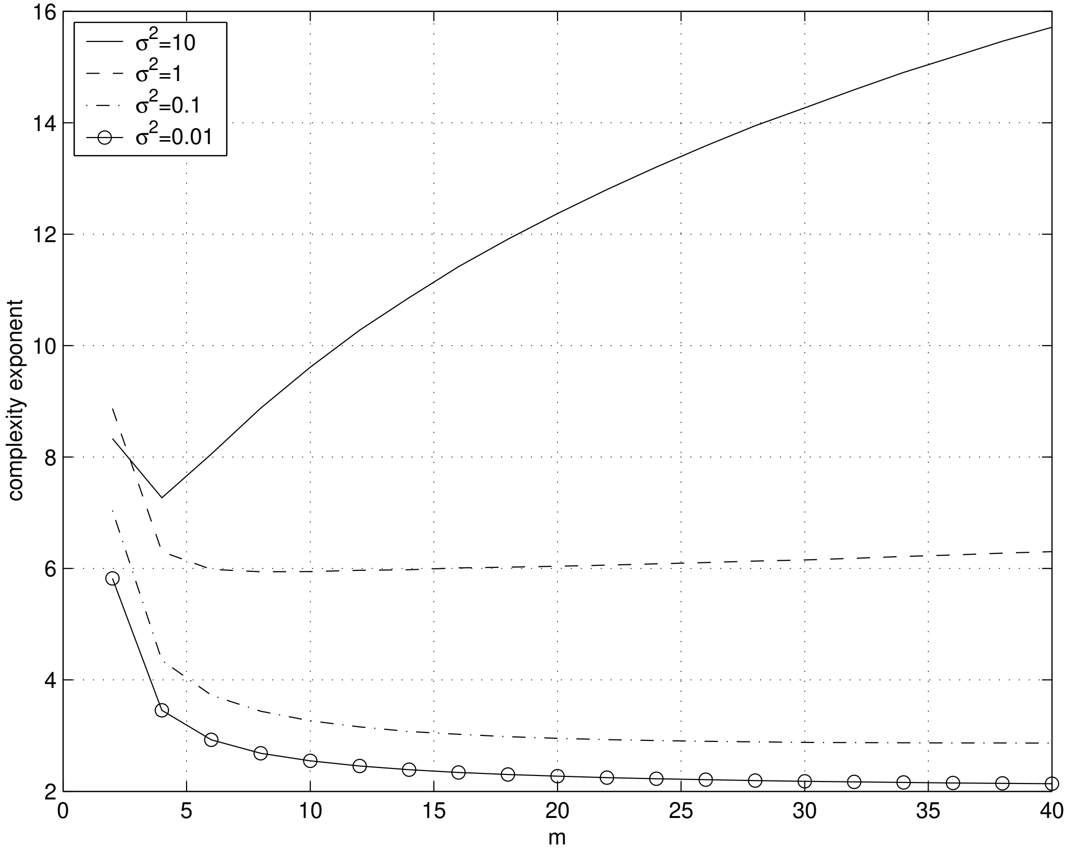
$$C(m,\sigma^2) = \sum_{k=1}^{m}(2k+17)\sum_{n=0}^{\infty} r_k(n)\gamma\left(\frac{m\sigma^2}{\sigma^2+n},\frac{k}{2}\right).$$

It is often useful to look at the *complexity exponent*:

$$C_e = \frac{\log C(m,\sigma^2)}{\log M}.$$

- When $C = O(m^\alpha)$, then $C_e \approx \alpha$.
- When $C = O(\beta^m)$, then $C_e \approx \frac{m}{\log m}$.

Expected Complexity as a Function of m



Sphere Decoding Over a Subset of the Lattice

All this is nice, but in communications problems the transmitted signal is chosen from a subset of the lattice, not the full lattice.

- The subsets of the lattice used are most often L-PAM constellations

$$\mathcal{D}_L^m = \left\{ -\frac{L-1}{2}, -\frac{L-3}{2}, \ldots, \frac{L-3}{2}, \frac{L-1}{2} \right\}^m .$$

- L is usually a power of two:

$$\mathcal{D}_2^m = \left\{ -\frac{1}{2}, \frac{1}{2} \right\}^m \quad , \quad \mathcal{D}_4^m = \left\{ -\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2} \right\}^m , \quad \text{etc.}$$

- Moreover, the interesting statistical quantity is not the noise variance σ^2, but the signal-to-noise ratio (SNR):

$$\rho = \frac{m(L^2-1)}{12\sigma^2}.$$

Computing the expected complexity now becomes tricky, since the expected number of subset lattice points in the k-dimensional sphere is given by

$$\frac{1}{L^k} \sum_n \sum_{s_t, s_a \in \mathcal{D}_L^k, \|s_t - s_a\|^2 = n} \gamma\left(\frac{\alpha m \sigma^2}{\sigma^2 + n}, \frac{k}{2}\right),$$

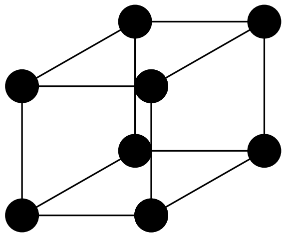
and enumerating the set

$$\left\{(s_t, s_a) \mid s_t, s_a \in \mathcal{D}_L^k, \|s_t - s_a\|^2 = n\right\},$$

appears to be complicated.

$\boxed{L = 2}$

The constellation consists of the corners of a k-dimensional hypercube.



Due to symmetry, all points are the same, and so we may choose

$$s_t = \left[\ -\tfrac{1}{2}, \ldots, -\tfrac{1}{2}\ \right]^T.$$

Then $s_a - s_t$ is a vector zeros and ones. Clearly the number of such vectors whose squared norm is n is given by $\binom{k}{n}$. Therefore

$$C(m, \rho) = \sum_{k=1}^{m} (2k + 17) \sum_{n=0}^{k} \binom{k}{n} \gamma\left(\frac{\alpha m}{1 + \frac{12\rho n}{m(L^2 - 12)}}, \frac{k}{2}\right).$$

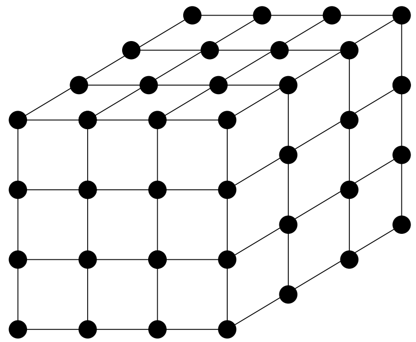
$\boxed{L = 4}$

- In this case, all points in the constellation are *not* the same. For each entry of s_t, we can distinguish between the "corner" points $\pm 3/2$ and the "center" points $\pm 1/2$.
- If a corner point is chosen then the corresponding entry of $s_t - s_a$ can take on the following values

$$\{0, \pm 1, \pm 2, \pm 3\},$$

with which we can identify the polynomial

$$\phi_0(x) = 1 + x + x^4 + x^9.$$



- Likewise, for a center point we have

$$\{\mp 1, 0, \pm 1, \pm 2\},$$

and

$$\phi_1(x) = 1 + 2x + x^4.$$

Thus, if among the k entries of s_t, we choose a corner point l times, the number of ways $\|s_t - s_a\|^2$ can add up to n is given by the coefficient of x^n in the polynomial

$$\begin{pmatrix} k \\ l \end{pmatrix} \phi_0^l(x)\phi_1^{k-l}(x).$$

This implies that the expected number of points in the k-dimensional sphere is given by

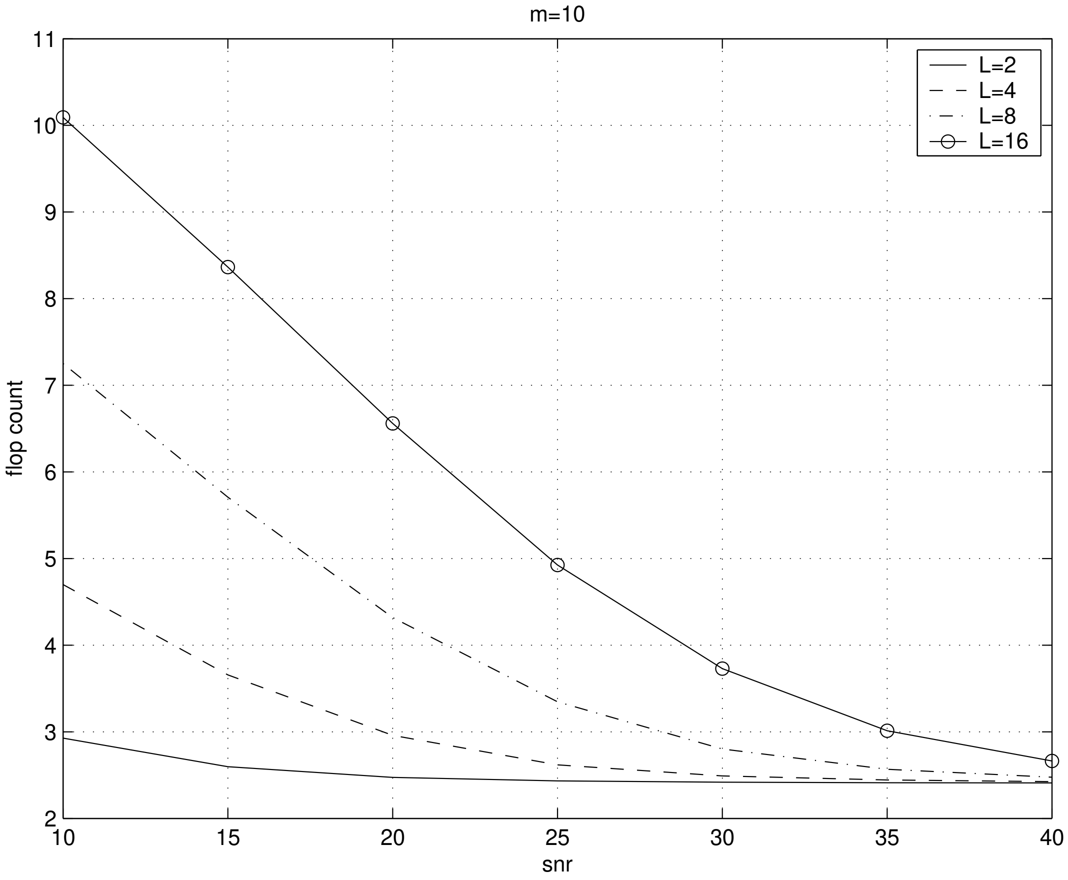
$$E_p(k,\rho) = \sum_n \frac{1}{2^k} \sum_{l=0}^{k} \begin{pmatrix} k \\ l \end{pmatrix} \left\{\phi_0^l(x)\phi_1^{k-l}(x)\right\}_n \gamma\left(\frac{\alpha m}{1 + \frac{12\rho n}{m(L^2-12)}}, \frac{k}{2}\right),$$

so that the expected complexity is

$$C(m,\rho) = \sum_{k=1}^{m} (2k+17)E_p(k,\rho).$$

Similar formulas can be developed for $L = 8, 16$, etc.

Complexity as a Function of SNR



Complexity as a Function of m

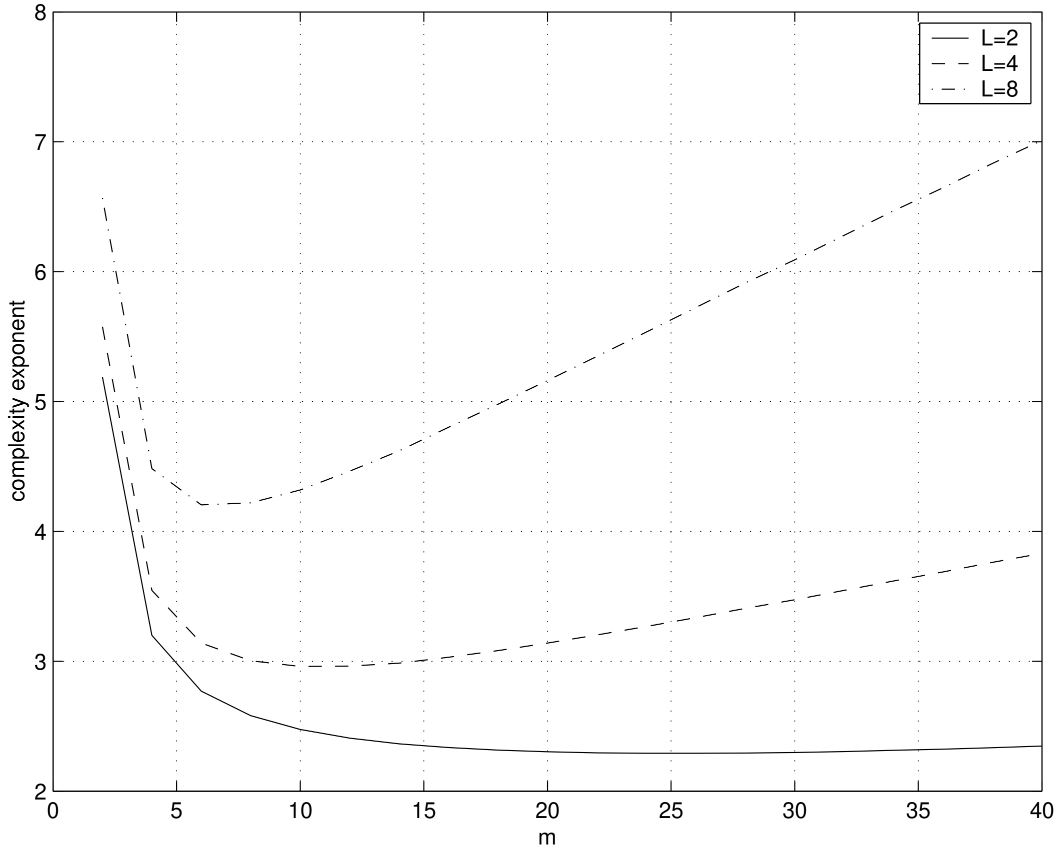


Figure 2: SNR = 10db.

Computational Complexity and Shannon Capacity

Each entry of s carries $\log L$ bits of information so that the data transmission rate is

$$R = m \log L \quad \text{bits/channel use}$$

The simulations presented show that for a fixed m and L (and hence fixed transmission rate R) the computational complexity increases as we decrease the SNR.

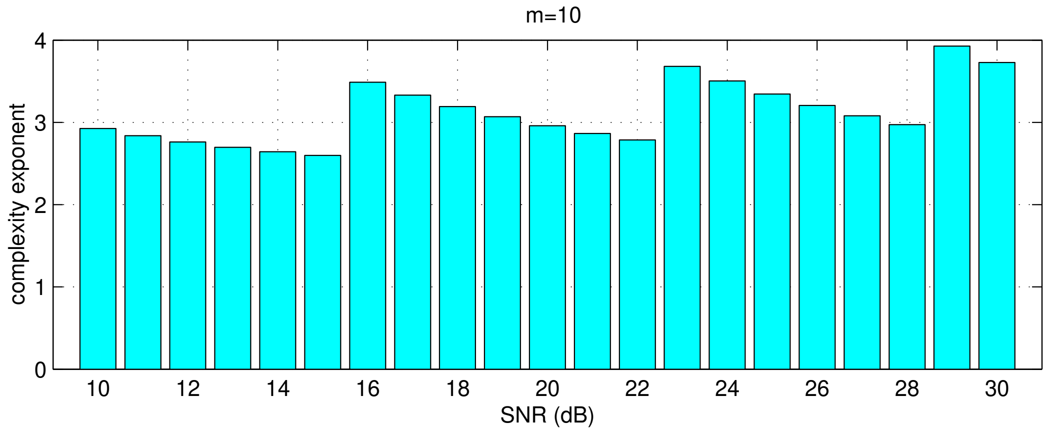
- In practice, however, we would never want the SNR to be such that the Shannon capacity

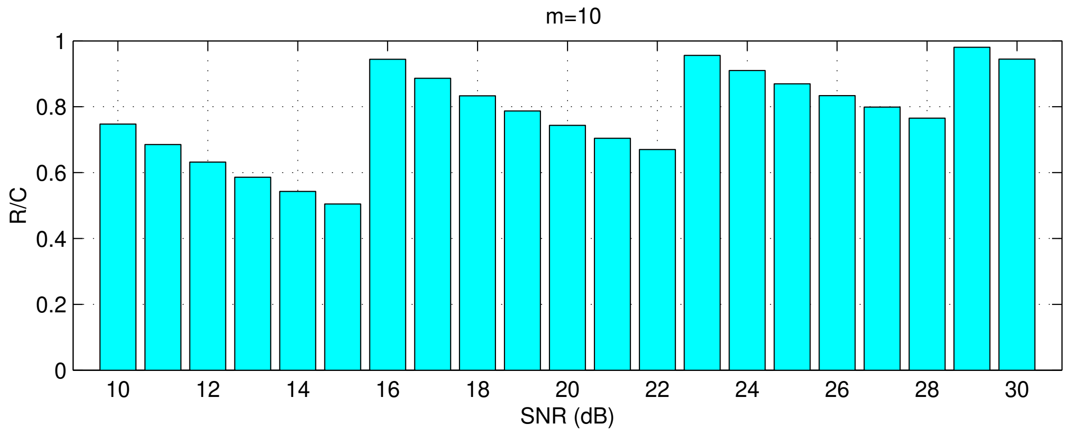
  $$C_{shannon}(m, \rho) = \frac{1}{2} E \log \det \left(I_m + \frac{\rho}{M} H H^T \right),$$

  is not able to support the rate R.

So what is the expected complexity like for $R < C_{shannon}(m, \rho)$?

Complexity vs. R/C





Back to Earlier Example

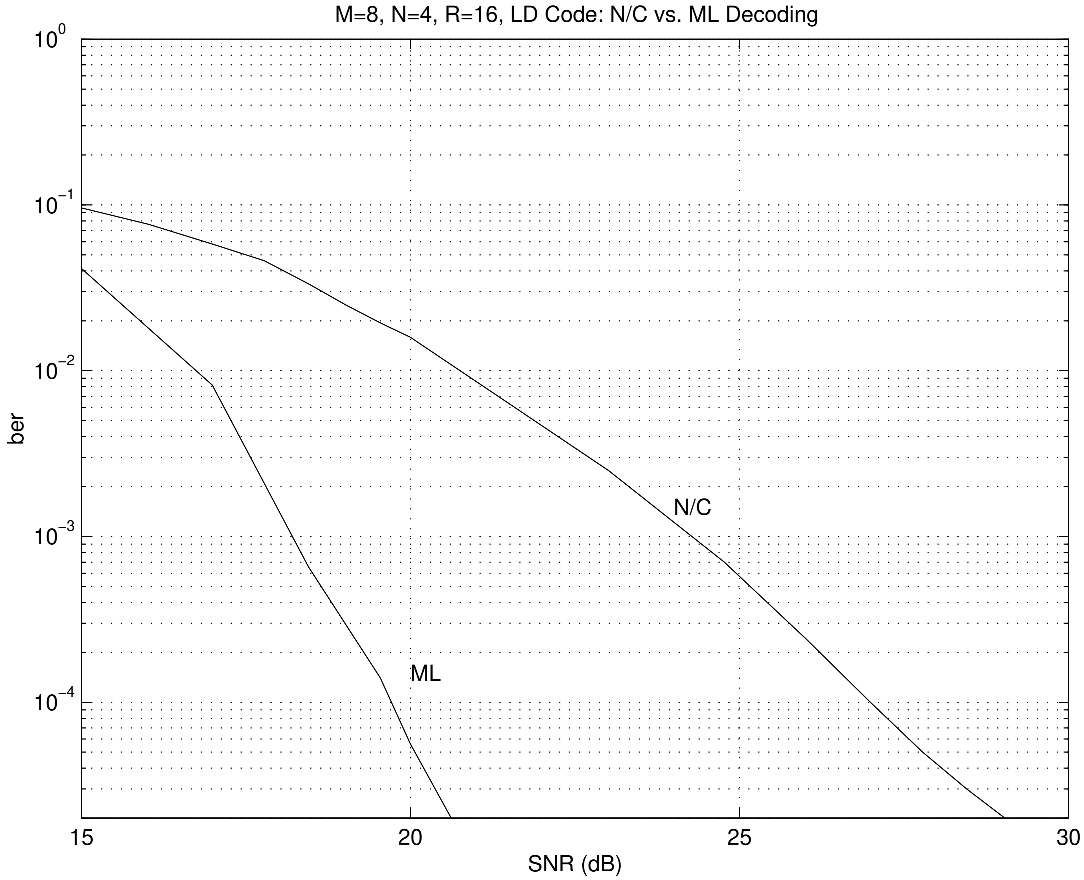


Figure 3: Bit error performance of a rate 16 linear space-time code, corresponding to $m = 64$. (No. of lattice points$= 2^{16\cdot 8} \approx 3.4 \times 10^{38}$). The ML complexity was between once and twice that of nulling/cancelling.

Further Work

The fact that, over a wide range of rates and SNR's, ML decoding can be done in polynomial (often sub-subic) time, has many practical ramifications. It also suggests generalizations in several directions:

- equalization of FIR channels
  - the channel now has Toeplitz strucure—this leads to a Toeplitz version of the Waring problem!
  - it so turns out that the more ISI, the merrier!
- iterative decoding
  - turbo-codes, LDPC codes, etc.
- extensions to Galois fields
- trade-off between computational complexity and decoding performance

Speculation

Suppose we have a set $\mathcal{S} = \{S_1, \ldots, S_L\}$ of $L = 2^{RT}$ input signals of length T. Suppose further that we send these through a communications channel to obtain the output signal X and perform maximum-likelihood (ML) decoding at the output of the channel to recover the input signal

$$\hat{S} = \arg\max_{S \in \mathcal{S}} \log p(X|S).$$

Then information theory states that, as long as R is strictly less than the Shannon capacity C, then as $T \to \infty$ the probability of error in the ML decoder goes to zero.

- In practice ML decoding is rarely used since it requires an exponential search over the set $\mathcal{S}$.
- However, given that we are transmitting below capacity, is it possible that ML decoding can be done in polynomial-time?

Discussion and Conclusion

- Introduced the integer least-squares problem and its applications
- Problem is worst-case (and average) NP-hard
- Considered the average complexity of sphere decoding as a function of the noise variance
  - connections to number theory
  - full lattice, subset of a lattice
  - polynomial complexity for a wide range of SNRs
- In many applications average complexity makes more sense than worst-case complexity (Levin 1986, many researchers since)
  - needs a natural stochastic setting
- Future work
  - connections to Shannon capacity? phase transitions?...
  - further applications, equalization,...