NEURAL LATTICE DECODERS

Vincent Corlay
Mitsubishi Electric R&D Centre Europe

In collaboration with Joseph Boutros, Philippe Ciblat and Loïc Brunel.

November 29, 2018

Overview

1. Context & Motivation
2. Lattices
3. Neural Lattice Decoders
4. Conclusions

Context

- Lattices are mathematical objects suited to model MIMO systems.
  - Good solution for lattice decoding $\Rightarrow$ more efficient MIMO systems.

- Lattices can also be used for channel coding, shaping and cryptography.

- Neural Networks and Deep Learning are emerging technologies with many advantages.

Open problem:
Low complexity near-optimal lattice decoding algorithm ?

The Decoding Problem

- $z \in \mathcal{M} \subseteq \mathbb{Z}^n$ is the input uncoded message.
- $y = g(z) \in \mathbb{R}^n$ is the received message.
- $g(\cdot)$: coding and/or channel.

MAP decoding

$$\text{Find } f(\cdot) \text{ s.t. } f(y) \approx \arg\max_{z \in \mathcal{M}} P(z|y)$$

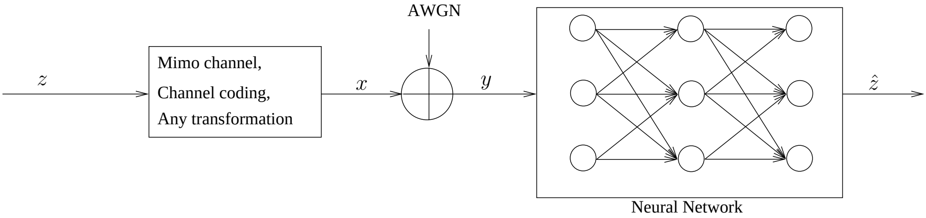


Figure: Discrete-time baseband channel model.

MAP: Maximum A Posteriori

Deep Learning: what for?

DL for Computer Vision

- Minimum error rate for the classification on a set of images is unknown.
- People have been struggling just to outperform humans until the use of convolutional networks.
- **Performance is more important than complexity.**

DL for Decoding

- Since 1948 the limit for reliable communication over a noisy channel is known.
- There exist algorithms to get really close to this limit (LDPC, Polar) and to optimally decode (Max. Likelihood).
- **Complexity is the main goal in Coding Theory when applying DNNs.**

DL: Deep Learning, DNN: Deep Neural Network

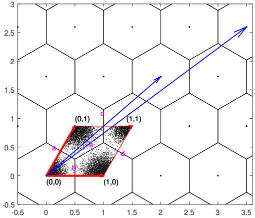
Overview

1. Motivation
2. **Lattices**
3. Neural Lattice Decoders
4. Conclusions

Lattices

A lattice is a discrete additive subgroup of $\mathbb{R}^n$:

- There are n basis vectors, $\mathcal{B} = \{g_i\}_{i=1}^{n}$.
- The lattice is given by all their integer linear combinations.
  E.g. $\{x = z \cdot G,\ z \in Z^n\}$.
- Lattices are the real Euclidean counterpart of error-correcting codes.
  - Codes are vector spaces over a finite field.
  - Lattices are modules over a real or a complex ring, e.g. $\mathbb{Z}$, $\mathbb{Z}[i]$, $\mathbb{Z}[\omega]$.



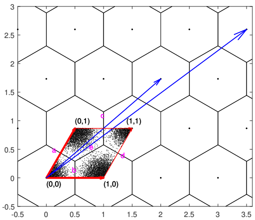
Fundamental regions

- Voronoi cell.
- Fundamental parallelotope $\mathcal{P}(\mathcal{B})$.
  - Good and bad bases.

Lattices

A lattice is a discrete additive subgroup of $\mathbb{R}^n$:

- There are n basis vectors, $\mathcal{B} = \{g_i\}_{i=1}^{n}$.
- The lattice is given by all their integer linear combinations.
  E.g. $\{x = z \cdot G,\ z \in Z^n\}$.
- Lattices are the real Euclidean counterpart of error-correcting codes.
  - Codes are vector spaces over a finite field.
  - Lattices are modules over a real or a complex ring, e.g. $\mathbb{Z}$, $\mathbb{Z}[i]$, $\mathbb{Z}[\omega]$.



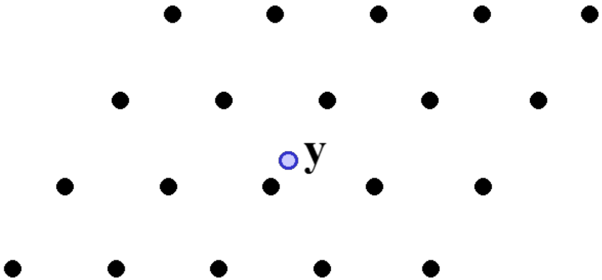
Fundamental regions

- Voronoi cell.
- Fundamental parallelotope $\mathcal{P}(\mathcal{B})$.
  - Good and bad bases.

The Closest Vector Problem

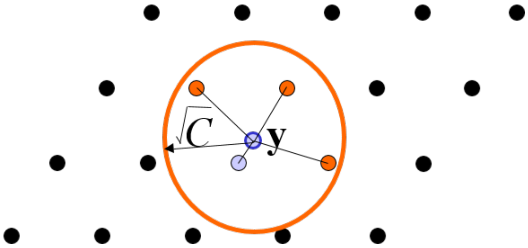
The CVP (lattice decoding)

Given a point in $\mathbb{R}^n$ find the closest lattice point.



CVP: Closest Vector Problem

Solving the CVP: Sphere Decoding



Drawbacks

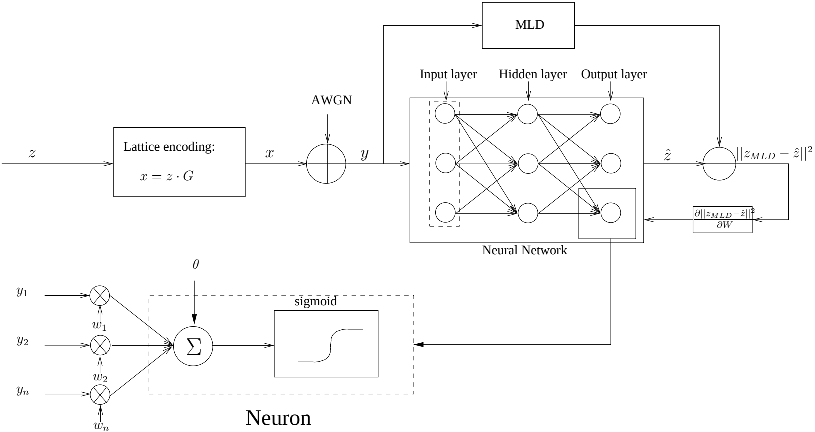
- Not hardware-friendly.
- Complexity highly variable.
- Complexity not well understood.
- Can not be parallelized: high latency.
- Can not take advantage of possible approximations.

CVP: Closest Vector Problem

Overview

1. Motivation
2. Lattices
3. **Neural Lattice Decoders**
4. Conclusions

Neural Lattice Decoding (1)



MLD: Maximum Likelihood Decoding

Neural Lattice Decoding (2)

Theorem (Anthony & Bartlett 1999)

The two-layer sigmoid networks are "universal approximators", in a sense that, given any continuous function f defined on some compact subset $\mathcal{S}$ of $\mathbb{R}^n$, and any desired accuracy ϵ, there is a two-layer sigmoid network computing a function that is within ϵ of f at each point of $\mathcal{S}$.

The fundamental parallelotope $\mathcal{P}(\mathcal{B})$ is a compact set that can be used for neural lattice decoding.

Neural Lattice Decoding (2)

Theorem (Anthony & Bartlett 1999)

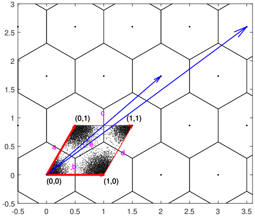
The two-layer sigmoid networks are "universal approximators", in a sense that, given any continuous function f defined on some compact subset $\mathcal{S}$ of $\mathbb{R}^n$, and any desired accuracy ϵ, there is a two-layer sigmoid network computing a function that is within ϵ of f at each point of $\mathcal{S}$.

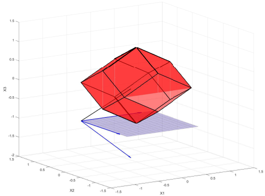
The fundamental parallelotope $\mathcal{P}(\mathcal{B})$ is a compact set that can be used for neural lattice decoding.

Voronoi-Reduced Basis (1)

(New) Definition (Voronoi-Reduced Basis)

Let $\mathcal{B}$ be the $\mathbb{Z}$-basis of a rank-n lattice Λ in $\mathbb{R}^n$. $\mathcal{B}$ is said Voronoi-reduced if, for any point $y \in \mathcal{P}(\mathcal{B})$, the closest lattice point $\hat{x}$ to y is one of the 2^n corners of $\mathcal{P}(\mathcal{B})$, i.e. $\hat{x} = \hat{z}G$ where $\hat{z} \in \{0,1\}^n$.





A VR-basis induces:

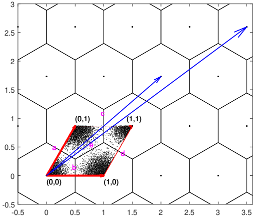
- Binary outputs network.
- Less Voronoi-Partitions within the fundamental parallelotope: The function to learn is simpler.

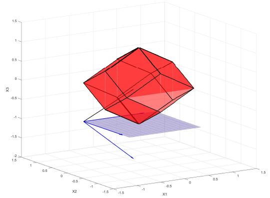
VR: Voronoi-Reduced

Voronoi-Reduced Basis (1)

(New) Definition (Voronoi-Reduced Basis)

Let $\mathcal{B}$ be the $\mathbb{Z}$-basis of a rank-n lattice Λ in $\mathbb{R}^n$. $\mathcal{B}$ is said Voronoi-reduced if, for any point $y \in \mathcal{P}(\mathcal{B})$, the closest lattice point $\hat{x}$ to y is one of the 2^n corners of $\mathcal{P}(\mathcal{B})$, i.e. $\hat{x} = \hat{z}G$ where $\hat{z} \in \{0, 1\}^n$.



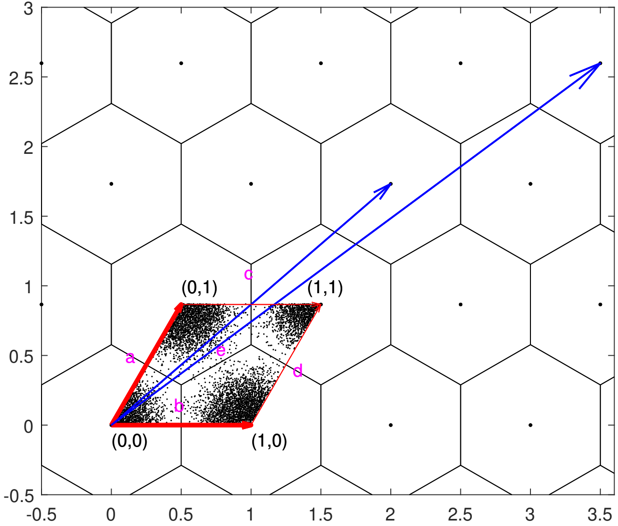


A VR-basis induces:

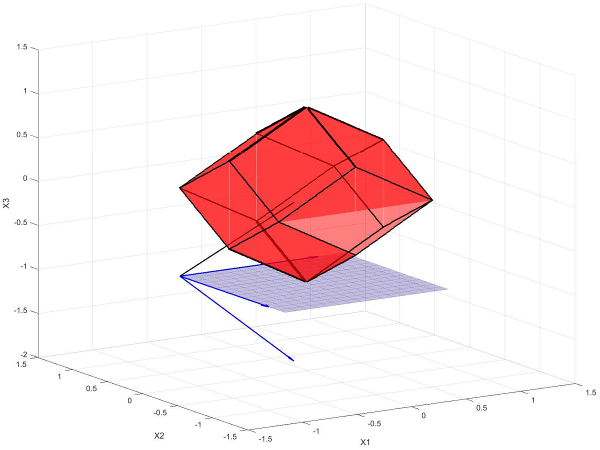
- Binary outputs network.
- Less Voronoi-Partitions within the fundamental parallelotope:
  The function to learn is simpler.

VR: Voronoi-Reduced

Voronoi-Reduced Basis (2)



Voronoi-Reduced Basis (3)



Learning is easier with a good basis

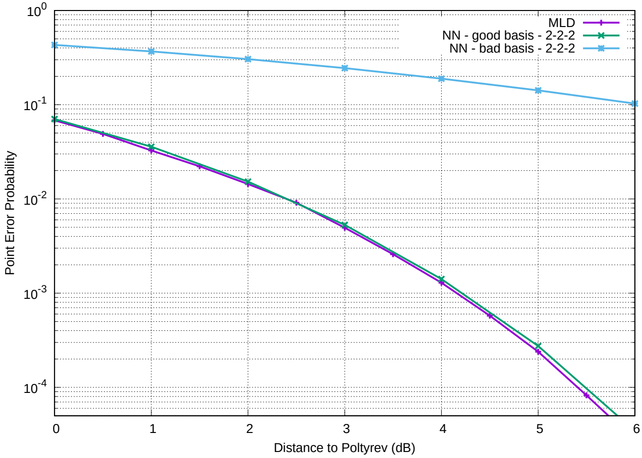
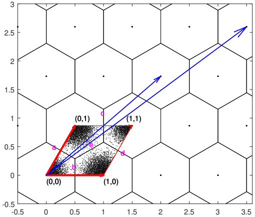


Figure: Feed-forward neural network applied to the lattice A_2, with a good and a bad basis.

MLD: Maximum Likelihood Decoding

The HLD: a hand-made NLD



A Boolean equation to decode

- $x = z \cdot G + \eta,\, x \in \mathcal{P}(\mathcal{B})$
- $\hat{z} = (\hat{z}_1, \hat{z}_2)$
- $\hat{z}_1 = c + b \cdot e$

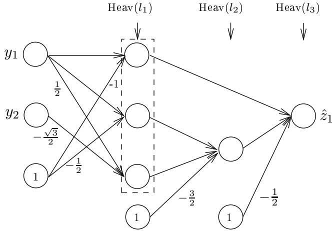


Figure: Neural network computing the Boolean equation.

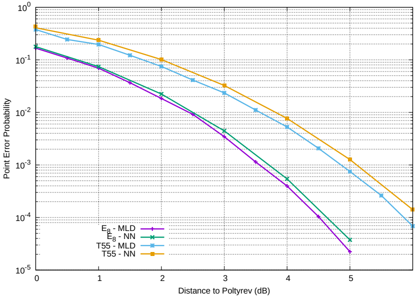
The HLD is MLD when used on a lattice with a Voronoi-reduced basis.

HLD: Hyperplane Logical Decoder, NLD: Neural Lattice Decoder, Heav: Heaviside function

Learning to decode (1)

Settings

- Standard fully-connected network with 3 hidden layers (No constraint on the architectures).
- Dense lattice E_8 & MIMO Lattice $T55$ ($n = 16$).
- Size of first hidden layer $\approx$ kissing number: $\tau(E_8) = 240$, $\tau(T55) = 30$.
- Nb params: W=83200 for E_8, W=6280 for $T55$*.



For E_8: $\frac{\log_2(W)}{n} = 2.0$ (supra-lin.), for $T55$: $\frac{\log_2(W)}{n} = 0.78$ (sub-lin.).

Competitive decoding algorithm only for non-dense lattices.

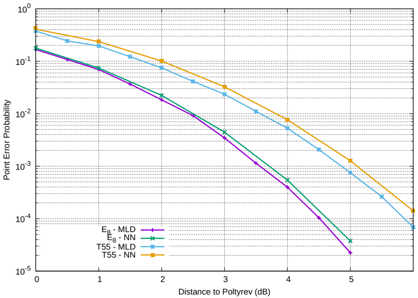
* For $T55$ it is possible to reach MLD performance with a slight increase in the number of parameters W.

MLD: Maximum Likelihood Decoding, lin.: linear

Learning to decode (1)

Settings

- Standard fully-connected network with 3 hidden layers (No constraint on the architectures).
- Dense lattice E_8 & MIMO Lattice $T55$ ($n = 16$).
- Size of first hidden layer $\approx$ kissing number: $\tau(E_8) = 240$, $\tau(T55) = 30$.
- Nb params: W=83200 for E_8, W=6280 for $T55$*.



For E_8: $\frac{\log_2(W)}{n} = 2.0$ (supra-lin.), for $T55$: $\frac{\log_2(W)}{n} = 0.78$ (sub-lin.).

Competitive decoding algorithm only for non-dense lattices.

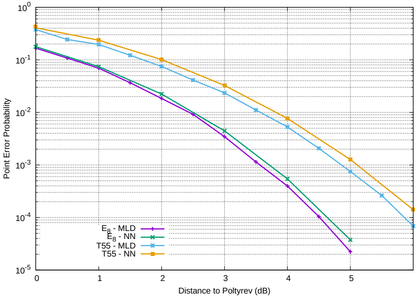
* For $T55$ it is possible to reach MLD performance with a slight increase in the number of parameters W.

MLD: Maximum Likelihood Decoding, lin.: linear

Learning to decode (1)

Settings

- Standard fully-connected network with 3 hidden layers (No constraint on the architectures).
- Dense lattice E_8 & MIMO Lattice $T55$ ($n = 16$).
- Size of first hidden layer $\approx$ kissing number: $\tau(E_8) = 240$, $\tau(T55) = 30$.
- Nb params: W=83200 for E_8, W=6280 for $T55$*.



For E_8: $\frac{\log_2(W)}{n} = 2.0$ (supra-lin.), for $T55$: $\frac{\log_2(W)}{n} = 0.78$ (sub-lin.).

Competitive decoding algorithm only for non-dense lattices.

* For $T55$ it is possible to reach MLD performance with a slight increase in the number of parameters W.

MLD: Maximum Likelihood Decoding, lin.: linear

Learning to decode (2)

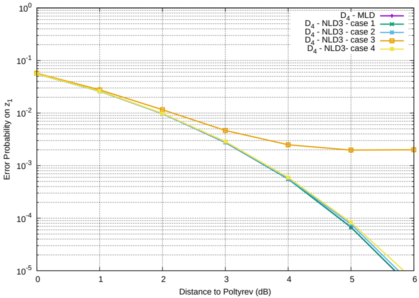
Settings

- HLD network with L1 regularization.
- Dense lattice D_4 .
- HLD equation for D_4:

$$z_1 = u_1 + u_2 \cdot u_3 \cdot u_4 \cdot u_5 \cdot u_6 + u_4 \cdot u_7 \cdot u_8 + u_4 \cdot u_7 \cdot u_9 + \cdot u_4 \cdot u_{10}.$$

The equation has 5 logical OR
(5 neurons in the second hidden layer).

- We fix the projections (first layer).
- We learn the rest with L1 regu.
- The number of neuron in the second layer decreases from 5 to 2.



Learning with L1 regularization enables to factorize equations.

Conclusion

- Machine learning in the heart of Telecommunications systems.
- Easy implementation of models in large companies thanks to parallel computing.
- Embed Neural Network inside hardware: analog or digital solutions?