

On Recovering Latent Factors From Sampling And Firing Graph

Pierre Gouedard

20th September 2019

Abstract

Consider a set of latent factors whose observable effect of activation is caught on a measure space that appears as a grid of bits tacking value in $\{0, 1\}$. This paper intend to deliver a theoretical and practical answer to the question: Given that we have access to a perfect indicator of the activation of latent factors that label a finite dataset of grid's activity, can we imagine a procedure to build a generic identificator of factor's activations ?

1 Introduction

This paper starts by introducing a mathematical framework for our solution. Then it describes a procedure to build the generic factor's activations identificator. Finally it presents the result of the procedure for two particular statistical modelling of the measure grid's activity. This paper has been influenced by modern machine learning techniques, reviewed in [?], especially algorithm that perform automated feature engineering such as neural network and deep learning [?] as well as tree learning techniques [?] and improvements [?] and [?]. Finally, modern signal processing techniques, that I have been taught at the Ecole Polytechnique Fédérale De Lausanne, reviewed in [?], and recent work in statistics in large dimensions, to which I have been introduced during my stay in the Laboratoire d'Informatique Gaspard Monge, reviewed in [?], has been more than determinant for the conception of this paper. In order to make the core subject of this report more consistent, we introduce the following notations:

- n the size of the measure grid.
- \mathcal{G} the measure grid, composed of bits, $\mathcal{G} = \{b_1, \dots, b_n\}$.
- $S(\mathcal{G})$ the set of all possible permutations of \mathcal{G} , $|S(\mathcal{G})| = 2^n$.
- $S(\mathcal{G}, l)$ the set of all permutations of \mathcal{G} of size l .
- K the number of latent factors.
- \mathcal{F} the set of latent factors, $\mathcal{F} = \{f_1, \dots, f_K\}$.

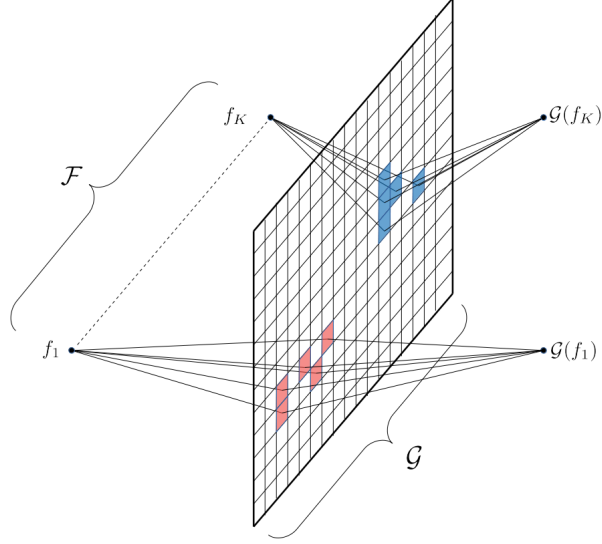


Figure 1: Measure grid model

- $S(\mathcal{F})$ the set of all possible permutations of \mathcal{F} , $|S(\mathcal{F})| = 2^K$.
- $S(\mathcal{F}, l)$ the set of all permutations of \mathcal{F} of size l .
- $\mathcal{G}(f)$ the set of bits activated by factor f , $\mathcal{G}(f) \in S(\mathcal{G})$ and $f \in \mathcal{F}$.
- $\mathcal{G}^{-1}(b)$ the set of factors that activate grid's bit b , $\mathcal{G}^{-1}(b) \in S(\mathcal{F})$ and $b \in \mathcal{G}$.
- $F(2)$ the field with elements $\{0, 1\}$, equipped with the logical XOR and logical AND respectively as the addition and multiplication.

2 Definitions And Properties

2.1 Statistical definitions

In this section we provide formalism for the statistical description of factors's activity and their signature on the measure grid.

Activation of factors

Each factor takes value in $\{0, 1\}$ at each instant of time. A factor with value 1 at some instant is active, otherwise it has value 0. At this stage of the paper we assume no particular statistical model for factors. Nevertheless, if we consider the set of all possible combination of active and unactive factors ($F(2)^K$), we assume that there is a well defined distribution d' such that

$$d' : F(2)^K \rightarrow [0, 1]$$

$$d'_x = \mathbb{P}(f_1 = x_1, \dots, f_K = x_K)$$

The statistical signature of a factor on the measure grid describes how the factor is linked to measure grid's bits. At this stage we simply assume that there is a well defined probability measure so that for any $I \in S(\mathcal{G})$

$$\begin{aligned} \mathbb{P}(\mathcal{G}(f) = I) &\in [0, 1] \\ \sum_{I \in S(\mathcal{G})} \mathbb{P}(\mathcal{G}(f) = I) &= 1 \end{aligned}$$

Latent factors's activations and signatures on the measure grid induce activations of measure grid's bits. We refer to this distribution over all possible combinations of activations of bits as d , and define it as

$$\begin{aligned} d &: F(2)^n \rightarrow [0, 1] \\ d_x &= \mathbb{P}(b_1 = x_1, \dots, b_n = x_n) \end{aligned}$$

Finally, we can also modelize the connection between factors and a measure grid's bit as a signature of the grid's bit on factor space. That is, for $I \in S(\mathcal{F})$, there is a well defined probability measure so that

$$\begin{aligned} \mathbb{P}(\mathcal{G}^{-1}(b) = I) &\in [0, 1] \\ \sum_{I \in S(\mathcal{F})} \mathbb{P}(\mathcal{G}^{-1}(b) = I) &= 1 \end{aligned}$$

Characteristic polynome

The activity of factors and grid's bits may be modelized using a set of multivariate polynomials whose fiber and image domain is respectively $F(2)^n$ and $F(2)$. The set of polynome associated with a set $I \in S(\mathcal{F})$ and $I' \in S(\mathcal{G})$ is denoted respectively $\{\mathcal{P}_{I,l}\}_{l \in \mathbb{N}}$ and $\{\mathcal{P}_{I',l}\}_{l \in \mathbb{N}}$. It represents a segmentation of states of respectively factors of I and measure grid's bits of I' .

$$\begin{aligned} \mathcal{P}_{I,l} &: F(2)^K \rightarrow F(2) \\ \mathcal{P}_{I,l}[\mathbf{x}] &= \begin{cases} \sum_{\pi \in S(I',l)} x_{\pi_1} \cdot \dots \cdot x_{\pi_l}, & \text{if } l \in \{1, \dots, |I|\} \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

and

$$\begin{aligned} \mathcal{P}_{I',l} &: F(2)^n \rightarrow F(2) \\ \mathcal{P}_{I',l}[\mathbf{x}'] &= \begin{cases} \sum_{\pi \in S(I,l)} x'_{\pi_1} \cdot \dots \cdot x'_{\pi_l}, & \text{if } l \in \{1, \dots, |I|\} \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

Where $S(I, l)$ and $S(I', l)$ are the set of all permutations of size l of respectively I and I' , $\mathbf{x} = [x_{f_1}, \dots, x_{f_K}]$ and $\mathbf{x}' = [x_{b_1}, \dots, x_{b_n}]$. Furthermore we define the characteristic polynomial of a set $I \in S(\mathcal{F})$ and $I' \in S(\mathcal{G})$ at level $l_0 \in \{0, \dots, |I|\}$ and $l'_0 \in \{0, \dots, |I'|\}$ as

$$\begin{aligned} \mathcal{P}_I^{l_0} : F(2)^K &\rightarrow F(2) & \mathcal{P}_{I'}^{l'_0} : F(2)^n &\rightarrow F(2) \\ \mathcal{P}_I^{l_0} &= \sum_{l=l_0}^{|I|} \mathcal{P}_{I,l} & \mathcal{P}_{I'}^{l'_0} &= \sum_{l=l'_0}^{|I'|} \mathcal{P}_{I',l} \end{aligned}$$

So far, the addition is set to be the logical XOR in the definition of fields $F(2)$. However, in the rest of this report, we will use symbol $+$ and \sum as representation of a logical OR in $F(2)$. This notation enables us to save a lot of time in writing complex polynomials. Denoting \oplus the logical XOR and \bar{x} the opposite of x , one have

$$(x \cdot \bar{y}) \oplus (\bar{x} \cdot y) = x + y$$

Operators on polynome

In order to qualify a set of factors and grid's bits, we define some basic operator. First, let $I_{\mathcal{G}}$ be a subset of $S(\mathcal{G})$, we denote by $F_2(I_{\mathcal{G}}, \mathcal{G})$ the operator that transforms $I_{\mathcal{G}}$ into a set of $|I_{\mathcal{G}}|$ vector in $F(2)^{|\mathcal{G}|}$.

$$F_2(I_{\mathcal{G}}, \mathcal{G}) : S(\mathcal{G}) \rightarrow F(2)^{|\mathcal{G}| \times |I_{\mathcal{G}}|}$$

For each vector $X \in F_2(\{I\}, \mathcal{G})$ such that $I \in I_{\mathcal{G}}$, an entry takes value 1 if the associated index belongs to I , 0 otherwise. This operator is convenient to evaluate the characteristic polynome. As an example, let $(I, I') \in S(\mathcal{G})^2$ and $l_0 \in \{1, \dots, |I|\}$ then

$$\sum_{x \in F_2(\{I'\}, \mathcal{G})} \mathcal{P}_I^{l_0} [x] = \begin{cases} 1, & \text{if } |I' \cap I| \geq l_0 \\ 0, & \text{otherwise} \end{cases}$$

Furthermore, given the distribution over measure grid's bits activation d , we define the norm of a characteristic Polynome $\mathcal{P}_I^{l_0}$ with respect to d as

$$\begin{aligned} \|\cdot\|_d : \mathcal{P}_{F_2(S(\mathcal{G}), \mathcal{G})} &\rightarrow [0, 1] \\ \|\mathcal{P}_I^{l_0}\|_d &= \sum_{x \in F_2(S(\mathcal{G}), \mathcal{G})} \mathcal{P}_I^{l_0} [x] \times d_x \end{aligned}$$

Where $\mathcal{P}_{F_2(S(\mathcal{G}), \mathcal{G})}$ the space of all polynomials with domain $F_2(S(\mathcal{G}), \mathcal{G})$ and \times is the simple multiplication in \mathbb{R} . Finally, keeping previous notations, let $\{\mathcal{P}_{I_i}^{l_i}\}_{i=1, \dots, k}$ a set of characteristic polynome for some integer $k \geq 2$, we define the product operator with respect to d as

$$\langle \cdot, \dots, \cdot \rangle_d : \mathcal{P}_{F_2(S(\mathcal{G}), \mathcal{G})}^k \rightarrow [0, 1]$$

$$\langle \mathcal{P}_{I_1}^{l_1}, \dots, \mathcal{P}_{I_k}^{l_k} \rangle_d = \sum_{x \in F_2(S(\mathcal{G}), \mathcal{G})} \left(\mathcal{P}_{I_1}^{l_1}[x] \cdot \dots \cdot \mathcal{P}_{I_k}^{l_k}[x] \right) \times d_x$$

Where \cdot denotes the usual multiplication in $F(2)$ and \times is the simple multiplication in \mathbb{R} . Finally, each operator specified above can also be defined in the factor space, using the characteristic polynomial in factor space and the distribution over factors's activations d' .

Stochastic processus induced by factor's activation

Factors's activations are observed as a strictly stationnary stochastic processus. That is for a couple $(I, l) \in \mathcal{F} \times \{1, \dots, |I|\}$, we associate a stochastic process $x_I^l[t]$ defined as

$$x_I^l[t] = \begin{cases} 1, & \text{with probability } \|\mathcal{P}_I^l\|_{d'} \\ 0, & \text{Otherwise} \end{cases}$$

with $\{x_I^l[t]\}_{t \in \mathbb{N}}$ 2-by-2 independant. Factors's signatures and their activations lead to bits's activations that are also observed as stricly stationary stochastic processus. Again for a couple $(I, l) \in \mathcal{G} \times \{1, \dots, |I|\}$, we associate a stochastic process $x_I^l[t]$ defined as

$$x_I^l[t] = \begin{cases} 1, & \text{with probability } \|\mathcal{P}_I^l\|_d \\ 0, & \text{Otherwise} \end{cases}$$

with $\{x_I^l[t]\}_{t \in \mathbb{N}}$ 2-by-2 independant.

2.2 Firing Graph

The firing graph is the main data structure used in our solution. In this section we propose a definition of it, as well as basic tools to support its analysis.

Graph specification

The algorithm presented in this report use a particular data structure that we refer as firing graph and that we denote $G(V, D_w)$.

- V is the set of vertices $V = \{v_1, \dots, v_{|V|}\}$
- D_w is the weighted direct link matrix, $D_w \in \mathbb{N}^{|V| \times |V|}$ and $[D_w]_{i,j} = w$ indicate an edge of weight w from vertex v_i to vertex v_j if $w > 0$

G is a directed weighted graph whose vertices are organized in layer. A vertex v of some layer $i \in \mathbb{N}$ must have at least one incoming edge from a vertex of layer $i - 1$. It may also have incoming edges from any vertices of layer $k \in \mathbb{N}, k < i$. Such a set of vertices will be referred as the input domain of v . Vertices of layer 0 have empty input domains, they correspond to bits of the measure grid \mathcal{G} . Each vertex stores the tuple (I, l_0)

- I the set of vertices at the tail of incoming edge of the vertex, referred as input set
- l_0 the firing rate's lower bound of the vertex, referred as level, $l_0 \in \{1, \dots, |I|\}$

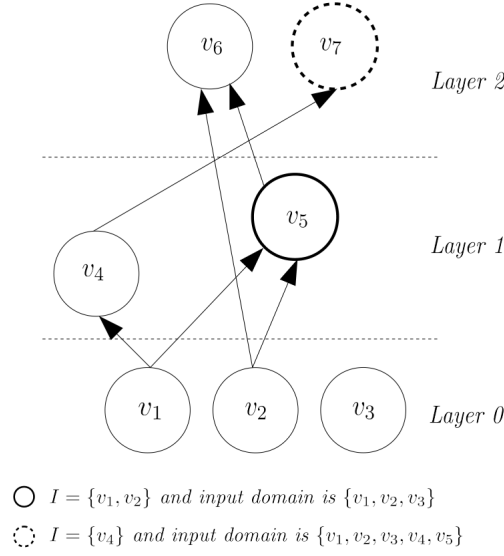


Figure 2: Firing graph

Graph Polynomials

As for bits of the measure grid and factors, a vertex $v(I, l_0)$ of a firing graph is associated with the set of polynomes $\{\mathcal{P}_{I,l}\}_{l \in \{l_0, \dots, |I|\}}$. Each polynome is a segment of its characteristic polynome \mathcal{P}_v that describes activation, at instant t , of v , given its input domain's activations at instant $t - 1$. If we denote by n , I and l_0 respectively the size of the input domain of v , the set of vertex that has a link toward v and the level of v , then

$$\begin{aligned}
 \mathcal{P}_{v,l} : F(2)^n &\rightarrow F(2) & \mathcal{P}_v : F(2)^n &\rightarrow F(2) \\
 \mathcal{P}_{v,l}[\mathbf{x}] &= \begin{cases} \sum_{\pi \in S(I,l)} x_{\pi_1} \cdot x_{\pi_2} \cdot \dots \cdot x_{\pi_l}, & \text{if } l \in \{l_0, \dots, |I|\} \\ 0, & \text{otherwise} \end{cases} & \mathcal{P}_v[\mathbf{x}] &= \sum_{l=l_0}^{|I|} \mathcal{P}_{v,l}[\mathbf{x}]
 \end{aligned}$$

Where $S(I, l)$ is the set of all permutations of size l of elements of I and $\mathbf{x} \in F_2(\{I\}, D_v)$, where D_v is the input domain of v . Furthermore, all operators on polynome defined previously is applicable. Let v, v_1, \dots, v_k be some vertices of the firing graph with the same input domain and d a distribution over activations of their input domain's vertices. Then the norm and the product with respect to distribution d are defined as

$$\|\mathcal{P}_v\|_d = \sum_{x \in F_2(S(\mathcal{G}), \mathcal{G})} \mathcal{P}_v[x] \times d_x \quad \langle \mathcal{P}_{v_1}, \dots, \mathcal{P}_{v_k} \rangle_d = \sum_{x \in F_2(S(\mathcal{G}), \mathcal{G})} (\mathcal{P}_{v_1}[x] \cdot \dots \cdot \mathcal{P}_{v_k}[x]) \times d_x$$

Finally, activations of vertices are observed as stochastic processus. Given a vertex $v(I, l)$ we

define

$$x_v[t] = \begin{cases} 1, & \text{with probability } \|\mathcal{P}_v\|_d \\ 0, & \text{Otherwise} \end{cases}$$

The stochastic process that takes value 1 if the vertex v activates and 0 otherwise, at each instant of time. If measure grid's bits compound layer 0 of the firing graph, then, from definition of bit's stochastic process and linearity of state's propagations, $x_v[t]$ is strictly stationary.

Connection to grid's bit

The firing graph is a convenient data structure to measure activity of a complex group of measure grid's bits. When the firing graph's layer 0 is composed of measure grid's bits, the characteristic polynome of each vertex can be represented as a characteristic polynome in the measure grid's space, without consideration of time and delay. Let G be such a firing graph, then for any vertex of layer 1, $v(I, |I|)$, the characteristic polynome v is equal to the characteristic polynome of the set of bits $I \in \mathcal{G}$ with level $|I|$.

$$\begin{aligned} \mathcal{P}_v &= \mathcal{P}_{I,|I|} \\ x_v[t] &= x_I^{|I|}[t-1] \end{aligned}$$

Furthermore if we set the level of v to 1 its characteristic polynome become the logical *or*-sum of the characteristic polynome of each bits of I

$$\begin{aligned} \mathcal{P}_v &= \sum_{b \in I} \mathcal{P}_{\{b\},1} \\ x_v[t] &= \begin{cases} 1, & \text{if } \sum_{b \in I} x_{\{b\}}^1[t-1] > 0 \\ 0, & \text{Otherwise} \end{cases} \end{aligned}$$

Besides, one can design more complexe arrangements of vertices that enable to model activations of multiple sets of measure grid's bits. Let G be a firing graph with its layer 0 composed of \mathcal{G} , let $u(I, |I|)$ and $v(I', 1)$, such that $I \cap I' = \emptyset$, be vertices of layer 1 and $w(\{u, v\}, 2)$ a vertex of layer 2. Then one can see that that characteristic polynome of w verifies

$$\begin{aligned} \mathcal{P}_w &= \sum_{b \in I'} \mathcal{P}_{I \cup \{b\}, |I|+1} \\ x_w[t] &= \begin{cases} 1, & \text{if } \sum_{b \in I'} x_{I \cup \{b\}}^{|I|+1}[t-2] > 0 \\ 0, & \text{Otherwise} \end{cases} \end{aligned}$$

2.3 Evaluation of measure grid's bits

A perfect indicator of the activation of a given factor f can be used to evaluate the possibility of any set of bits to be part of f 's signature on the measure grid.

Factor's signature

One way to describe the activity of a factor f on the measure grid is to associate it to a polynome in the measure grid's space

$$\begin{aligned}\mathcal{P}_{\mathcal{G}(f)} &: F(2)^n \rightarrow F(2) \\ \mathcal{P}_{\mathcal{G}(f)} &= \mathcal{P}_{\mathcal{G}(f), |\mathcal{G}(f)|}\end{aligned}$$

$\mathcal{P}_{\mathcal{G}(f)}$ is refered as the polynomial signature of f on \mathcal{G} . Anytime f is active then its polynomial signature takes value 1. Yet under particular modelling of factor's links to measure grid, the polynomial signature of f can take value 1 while f is not active. More formally let $f \in \mathcal{F}$, $\forall I \in S(\mathcal{F})$ such that $x \in F_2(\{I\}, \mathcal{F})$ and $x' \in F_2(\{\cup_{f \in I} \mathcal{G}(f)\}, \mathcal{G})$

$$\mathcal{P}_f[x] = 1 \Rightarrow \mathcal{P}_{\mathcal{G}(f)}[x'] = 1$$

Furthermore if $\exists J \in S(\mathcal{F} \setminus \{f\})$ such that $\mathcal{G}(f) \subset \bigcup_{f' \in J} \mathcal{G}(f')$ then

$$\mathcal{P}_f[x] = 1 \Leftrightarrow \mathcal{P}_{\mathcal{G}(f)}[x'] = 1$$

basic metrics

Let $I \in S(\mathcal{G})$, $l \in \{1, \dots, |I|\}$, $f \in \mathcal{F}$ and e the event "factor f is active". Then we define the recall coefficient of couple (I, l) with respect to f as

$$\mu_{I,l,f} = \langle \mathcal{P}_I^l, \mathcal{P}_{\mathcal{G}(f)} \rangle_{d|e} + \langle \mathcal{P}_I^l, \bar{\mathcal{P}}_{\mathcal{G}(f)} \rangle_{d|e}$$

Where $d|e$ is the distribution over bit's activations given event e and $\bar{\mathcal{P}}_{\mathcal{G}(f)}$ is the complement of $\mathcal{P}_{\mathcal{G}(f)}$ in $F(2)$. Furthermore we define the precision coefficient of couple (I, l) with respect to f as

$$\nu_{I,l,f} = \langle \mathcal{P}_I^l, \mathcal{P}_{\mathcal{G}(f)} \rangle_{d|\bar{e}} + \langle \mathcal{P}_I^l, \bar{\mathcal{P}}_{\mathcal{G}(f)} \rangle_{d|\bar{e}}$$

Where $d|\bar{e}$ is the distribution over bit's activations given not event e . Finally we define the purity coefficient of couple (I, l) with respect to f as

$$\omega_{I,l,f} = \frac{\nu_{I,l,f}}{\mu_{I,l,f}}$$

The lower $\omega_{I,l,f}$ is, the purer is the couple (I, l) with respect to f . The recall, precision and purity coefficient can be defined for any vertex v of a firing graph where vertices of layer 0 are composed by measure grid's bit and are denoted respectively $\mu_{v,f}$, $\nu_{v,f}$ and $\omega_{v,f}$. The latter are computed by using the representation of \mathcal{P}_v as a characteristic polynomial in the measure grid's space.

advanced metrics

Let $I \in S(\mathcal{G})$, $l \in \{1, \dots, |I|\}$, $f \in \mathcal{F}$ and e the event "factor f is active". We define the precision of the couple (I, l) with respect to factor f as

$$\phi_{I,l,f} = \frac{\|\mathcal{P}_I^l\|_{d,e}}{\|\mathcal{P}_I^l\|_d}$$

We also define the recall of the couple (I, l) with respect to factor f as

$$\psi_{I,l,f} = \frac{\|\mathcal{P}_I^l\|_{d,e}}{\|\mathcal{P}_{\mathcal{G}(f)}\|_{d,e}}$$

Where d, e , the distribution over the combination of activations of measure grid's bits that intersect with event e . The precision and the recall are defined for any vertex v of a firing graph where vertices of layer 0 are composed by measure grid's bit and are denoted respectively $\phi_{v,f}$ and $\psi_{v,f}$. Again, The latter are computed by using the representation of \mathcal{P}_v as a characteristic polynomial in the measure grid's space.

Advanced stochastic process induced by vertex

Given a firing graph with its layer 0 composed of measure grid's bits, we have seen that the propagation of activations induces a stochastic process at each vertex. Here we introduce some more complex stochastic processus at each vertex of G . Given a vertex v at layer $k \geq 0$, its characteristic polynome \mathcal{P}_v , a factor $f \in \mathcal{F}$ and e , the event "factor f is active", we define the score process of v with respect to factor f as

$$s_{v,f} [N, T, p, q] = N + \sum_{t=1}^T s_{v,p,q,t,f}$$

Where $(N, T, p, q) \in \mathbb{N}^4$ and $\{s_{v,p,q,t,f}\}_{t \in \mathbb{N}}$ a set of i.i.d random variable. $s_{v,p,q,t,f}$ takes value q if the event e was true at instant $t - k$ and value $-p$ if it was false, given that v activates at instant t . That is, $\forall t < k$, $s_{v,p,q,t,f} = 0$ and $\forall t \geq k$

$$s_{v,p,q,t,f} = \begin{cases} q, & \text{with probability } q_s \\ -p, & \text{with probability } 1 - q_s \end{cases}$$

Where $q_s = \frac{q_r}{q_r + q_p}$ with $q_r = \|\mathcal{P}_v\|_{d,e}$ and $q_p = \|\mathcal{P}_v\|_d - q_r$. d, e is the distribution over measure grid's activations that intersect with the event e

2.4 Properties

This paragraph intend to deliver useful properties for the analysis of the algorithm. The proof of every properties can be found in the appendix A at the end of this paper.

Polynomial decomposition

Partition

Let $v_1(I, l_0)$, $v_2(J, 0)$ and $v_3(K, 0)$, be three vertices at the layer 1 of some firing graph, with the same input domain \mathcal{G} . If $I = J \cup K$ and $J \cap K = \emptyset$, then, $\forall x \in F_2(S(\mathcal{G}), \mathcal{G})$

$$\mathcal{P}_I^{l_0}[x] = \sum_{l=l_0}^{|I|} \sum_{j=0}^{|J|} \mathcal{P}_{J,j}[x] \cdot \mathcal{P}_{K,l-j}[x] \quad (1)$$

In particular for $b \in I$

$$\mathcal{P}_{I,l}[x] = \mathcal{P}_{I \setminus \{b\}, l}[x] \cdot \mathcal{P}_{\{b\}, 0}[x] + \mathcal{P}_{I \setminus \{b\}, l-1}[x] \cdot \mathcal{P}_{\{b\}, 1}[x] \quad (2)$$

Decomposition

Let G be a firing graph with layer 0 composed of \mathcal{G} . Let $u(I, l_u)$, $v(I', l_v)$ such that $I \cap I' = \emptyset$ as vertices of layer 1 and $w(\{u, v\}, 2)$ as vertex of layer 2. Let $K \in \cup_{l \in \{l_u, \dots, |I'|\}} S(I', l)$, $x \in F_2(S(\mathcal{G}), \mathcal{G})$ and $x' = \begin{bmatrix} \mathcal{P}_u[x] & \mathcal{P}_v[x] \end{bmatrix}$ then

$$\mathcal{P}_{K, |K|}[x] \cdot \mathcal{P}_{\{u, v\}, 2}[x'] = \sum_{l=l_u}^{|I|} \sum_{J \in S(I, l)} \mathcal{P}_{J \cup K, l+|K|}[x] \quad (3)$$

In particular if $l_u = |I|$ and $l_v = 1$, then for any vertex of layer 0, $b \in I'$

$$\mathcal{P}_{b, 1}[x] \cdot \mathcal{P}_{\{u, v\}, 2}[x'] = \mathcal{P}_{I \cup \{b\}, |I|+1}[x] \quad (4)$$

Metrics

Throughout this section, we consider G to be a firing graph with layer 0 composed by measure grid's bits \mathcal{G} and $f \in \mathcal{F}$ denote some target factor that is linked to some bit of the measure grid. The distribution of activation of latent factors and measure grid's bits will be respectively denoted d and d' and e is the event "factor f is active". Furthermore we use v to denote some vertex of G whose characteristic polynome respects $\mathcal{P}_v = \mathcal{P}_I^l$ with $(I, l) \in S(\mathcal{G}), \{1, \dots, |I|\}$ and $f \in \mathcal{F}$ some factor

Precision of vertex

The precision of v with respect to f is

$$\phi_{v, f} = \frac{\|\mathcal{P}_f\|_{d'}}{\|\mathcal{P}_f\|_{d'} + (1 - \|\mathcal{P}_f\|_{d'}) \times \omega_{I, l, f}} \quad (5)$$

Furthermore, if $\mu_{v, f} = 1$ we have

$$\phi_{v, f} \leq \frac{\|\mathcal{P}_f\|_{d'}}{\|\mathcal{P}_f\|_{d'} + (1 - \|\mathcal{P}_f\|_{d'}) \times \omega_{\mathcal{G}(f), |\mathcal{G}(f)|, f}} \quad (6)$$

Recall of vertex

The recall of v with respect to f is

$$\psi_{v,f} = \mu_{I,l,f} \quad (7)$$

Furthermore,

$$0 \leq \phi_{v,f} \leq 1 \quad (8)$$

Where right equality is reached whenever v is connected to a set of measure grid's bit $I \in \mathcal{G}$, with level $l_0 = |I|$ such that $I \subset \mathcal{G}(f)$.

vertex's score process

If $s_{v,f}[N, T, p, q]$ denotes the score process of v with respect to f , with $N, T, p, q \in \mathbb{N}^4$, then

$$\mathbb{E}[s_{v,f}[N, T, p, q]] = N + T \times (\phi_{I,l,f} \times (p + q) - p) \quad (9)$$

Furthermore,

$$\text{Var}[s_{v,p,q,t,f}] = (q + p)^2 \times \phi_{I,l,f} \times (1 - \phi_{I,l,f}) \quad (10)$$

3 Identification of Latent Factor

In this section, we present a procedure to identify a latent factor's activation. The procedure consists of two steps:

- Sampling: Sample the measure grid and build a firing graph.
- Draining: Drain the firing graph to exclude high purity coefficient's vertices.

Both processus will be described and the efficiency of the draining algorithm quantified.

3.1 Sampling

Sampling the measure grid consists in following a procedure to select some bits of it. This procedure is usually designed to be the most efficient in the fullfilment of specific quantitative objective. First, we assume that we have access to a determinist exact indicator of f 's activations with $f \in \mathcal{F}$. Then, the objective of sampling is to maximize the probability that we sample a bit whose purity coefficient with respect to f is lower or equal to some positive constant ω . That is, if we denote s the random variable of the outcome of a single sampling, the objective is to maximize

$$\mathbb{P}(\omega_{\{s\},1,f} \leq \omega)$$

Again, if we have a set $I \subset \mathcal{G}$ of bits, the objective of sampling is to maximize the probability of selecting a bit b , for which the purity of $I \cup \{b\}$ at level $|I|+1$ is lower to a given positive constant ω . That is, if we denote by s the random variable of the outcome of a single sampling, the objective is to maximize

$$\mathbb{P}(\omega_{I \cup \{s\}, |I|, f} \leq \omega)$$

We propose a very intuitive sampling method based on the indicator of activation of target factor f . Given parameters $p_s \in [0, 1]$ and S_p respectively the probability of picking a bit and a set of pre-selected measure grid's bits, the sampling procedure writes

Algorithm 1 Sampling

Input: p_s, S_p

Output: S

```

 $S \leftarrow \{\}, x_f \leftarrow nextFactorState(), X_G \leftarrow nextGridState()$ 
while  $S$  is empty do
  if  $x_f = 1$  and  $\forall b \in S_p, X_G[b] = 1$  then
    for all  $b \in \mathcal{G} \setminus S \cup S_p$  do
      if  $X_G[b] = 1$  then
         $S \leftarrow S \cup \{b\}$  with probability  $p_s$ 
      end if
    end for
  end if
   $x_f \leftarrow nextFactorState()$ 
   $X_G \leftarrow nextGridState()$ 
end while

```

Where x_f and X_G are respectively a scalar that takes value 1 when factor f is active, 0 otherwise, and a mapping with measure grid's bits as keys and their states as values (0 or 1). The second mean of the sampling procedure is to build a firing graph. The construction of the firing graph requires to set a parameter $N \in \mathbb{N}$ that corresponds to the initial weight of edges that will be drained. In addition we set a mask matrix $G_{mask} \in \{0, 1\}^{|V|}$ that controls which vertex is allowed to have their outgoing edges updated during draining. We consider two kind of firing graphs.

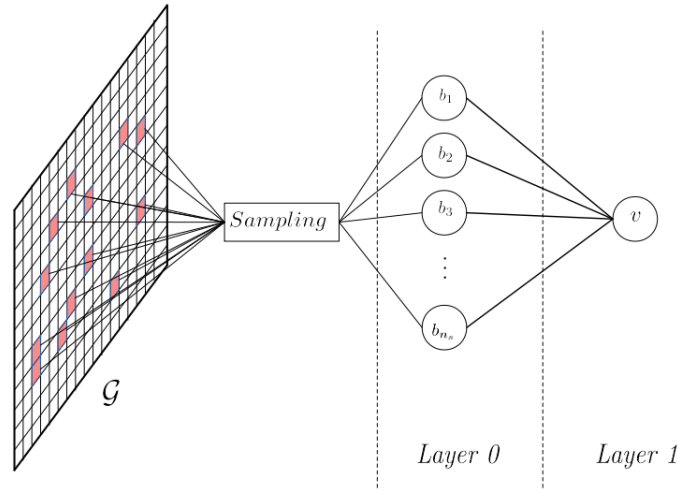


Figure 3: Single sampled firing graph

In figure ??, sampled bits $\{b_1, \dots, b_{n_s}\}$ are used as vertices of the layer 0 of a firing graph G , $n_S = |S|$. Then vertex $v(\{b_1, \dots, b_{n_s}\}, 1)$ is added at the layer 1 of G . Furthermore, we set G_{mask} so to allow only layer 0's outgoing edges to be updated through draining.

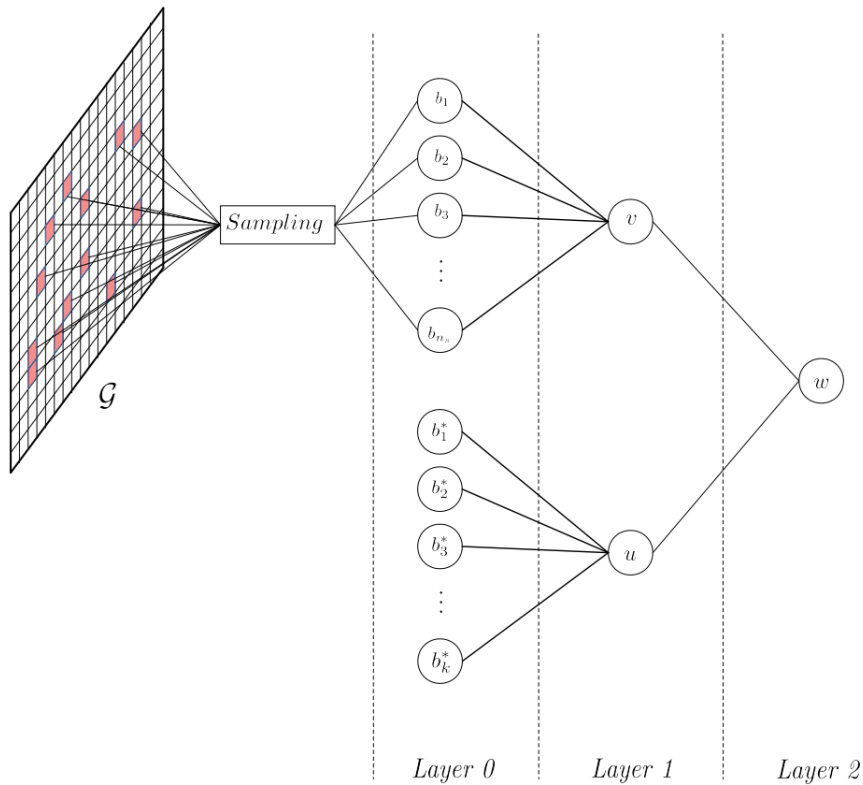


Figure 4: Joint sampled firing graph

In figure ??, sampled bits $\{b_1, \dots, b_{n_s}\}$ and pre-selected bits $\{b_1^*, \dots, b_k^*\}$ for some $k \in \mathbb{N}^*$ compound the layer 0 of the firing graph G , $n_S = |S|$. Then, vertices $v(\{b_1, \dots, b_{n_s}\}, 1)$ and

$u(\{b_1, \dots, b_k\}, k)$ are added at layer 1 of G and vertex $w(\{u, v\}, 2)$ at layer 2 of G . Finally, we set G_{mask} so that only b_1, \dots, b_{n_s} 's outgoing edges are allowed to be updated through draining.

3.2 Draining

Draining the firing graph consists in iterating a forward propagation of bits's activations and a backward propagation of feedback generated by factor's activations through the firing graph. Feedback are meant to increment or decrement the weight of unmasked vertices's outgoing edges. Given that an edge with a null or negative weight vanishes, at the end of the routine, connections of the graph differentiate between vertices's purity. To ease understanding of the algorithm, we split vertices of the firing graph into input and core vertices which are respectively vertices of layer 0 and vertices of layers > 0 . Furthermore, we introduce a new type of vertices that can only have incoming edges from core vertices. We refer to those vertices as outputs.

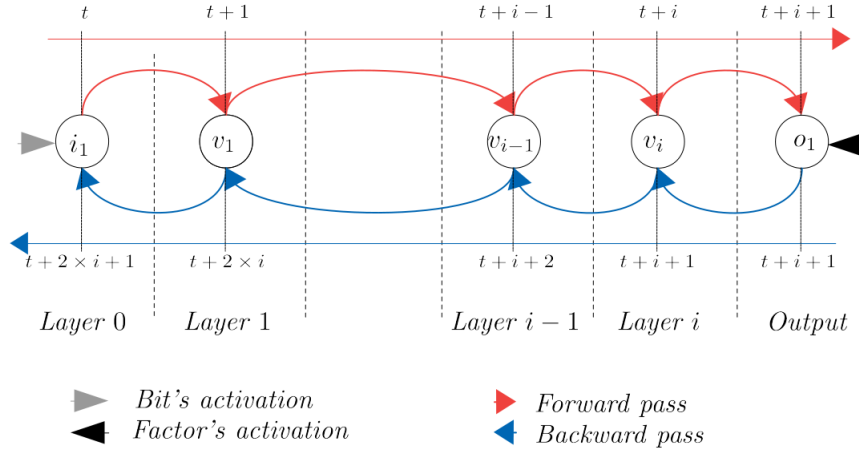


Figure 5: Draining diagram

We use n_i , n_c and n_o to refer to the number of respectively input, core and output vertices. Furthermore, we define $I_w \in \mathbb{N}^{n_i \times n_c}$, $C_w \in \mathbb{N}^{n_c \times n_c}$ and $O_w \in \mathbb{N}^{n_c \times n_o}$ that correspond to the weighted direct link matrices respectively from input toward core vertices, core toward core vertices and core toward output vertices. Furthermore we will use $A = A_w > 0$, $A \in \{I, C, O\}$ to denote the corresponding unweighted direct link matrices. Finally, in order to represent in a more convenient way stochastic processus induced by measure grid's activations, we define the following stochastics vectors

- $x_i^{(t)} \in \{0, 1\}^{1 \times n_i}$ the vector of activations of input vertices at instant t
- $x_c^{(t)} \in \{0, 1\}^{1 \times n_c}$ the vector of activations of core vertices at instant t
- $x_o^{(t)} \in \{0, 1\}^{1 \times n_o}$ the vector of activations of output vertices at instant t

The propagation of activations through the firing graph can be represented with two equations:

Forward processing (FP)

Forward transmitting (FT)

$$\begin{aligned}\tilde{x}_c^{(t)} &= x_i^{(t-1)} \cdot I + x_c^{(t-1)} \cdot C \\ \tilde{x}_o^{(t)} &= x_c^{(t-1)} \cdot O\end{aligned}$$

$$\begin{aligned}[x_c^{(t)}]_i &= \begin{cases} 1, & \text{if } [\tilde{x}_c^{(t)}]_i > l_i \\ 0, & \text{Otherwise} \end{cases} \\ [x_o^{(t)}]_j &= \begin{cases} 1, & \text{if } [\tilde{x}_o^{(t)}]_j > 1 \\ 0, & \text{Otherwise} \end{cases}\end{aligned}$$

Where \cdot is the usual matrix multiplication, $(i, j) \in \{1, \dots, n_c\} \times \{1, \dots, n_o\}$ and l_i is the level of the i^{th} core vertex. An output vertex of the firing graph is fed with the activation of a targeted factor decayed in time by the number of layer - 1. That is, for single and joint sampled firing graphs, the decay is respectively set to 1 and 2. Factor's activations generate a feedback to the output that is back propagated through the firing graph. Supposing that we set the factor's decay to $d \geq 1$, the feedback is defined as

$$x_{b,o}^{(t)} = x_o^{(t)} \circ \left((p + q) \times x_f^{(t-d)} - p \right)$$

Where \circ denotes the Hadamard product, $x_f^{(t-d)}$ is the vector of states of factors at instant $t - d$ and p, q are pre-defined positive integers. A correct backpropagation of $x_{b,o}^{(t)}$ up to the input vertices is made possible by using time and space coherence of firing graph's forward states. We denote by $V_i, i \in \mathbb{N}$ the set of vertices that has a path, composed of i vertices, toward an output vertex. Let G be a firing graph with $k \in \mathbb{N}^*$ layers augmented with a layer of output vertices. Let V_o the set of output vertices, $\forall v, o \in V_0 \times V_o$, v is eligible to o 's feedback at instant t if and only if

- v was active at instant $t - 1$
- v has an edge toward o

The same principle can be used to backpropagate the feedback from vetices of V_0 towards vertices of V_1 and so on. Generally speaking, the back propagation from vertices of V_i towards V_{i+1} respects $\forall v, v' \in V_i \times V_{i-1}$, v is eligible to feedback of v' at instant t if and only if

- v was active at instant $t - (2 \times i + 1)$
- v has an edge toward v'

Finally we can encode the backpropagation equations as

Backward transmitting (BT)

$$\begin{aligned}\tilde{X}_{b,c}^{(t)} &= (O \cdot X_{b,o}^{(t-1)} + C \cdot X_{b,c}^{(t-1)}) \circ X_{m,c}^{(t)T} \\ X_{b,i}^{(t)} &= (I \cdot X_{b,c}^{(t-1)}) \circ X_{m,i}^{(t)T}\end{aligned}$$

Backward processing (BP)

$$\begin{aligned}X_{b,o}^{(t)} &= \begin{bmatrix} \mathbf{0}_{n_o \times 1} & x_{b,o}^{(t)} & \mathbf{0}_{n_o \times d_{max} - 2} \end{bmatrix} \\ X_{b,c}^{(t)} &= \begin{bmatrix} \mathbf{0}_{n_c \times 2} & [\tilde{X}_{b,c}^{(t)}]_{(:n_c, : (d_{max} - 2))} \end{bmatrix}\end{aligned}$$

Structure updates (SU)

$$\begin{aligned} O_w &= O_w + O \circ (X_{b,o}^{(t-1)} \cdot X_{m,c}^{(t)})^T \\ C_w &= C_w + C \circ (X_{b,c}^{(t-1)} \cdot X_{m,c}^{(t)})^T \\ I_w &= I_w + I \circ (X_{b,c}^{(t-1)} \cdot X_{m,i}^{(t)})^T \end{aligned}$$

Where $X_{m,c}^{(t)} = [x_c^{(t)} \dots x_c^{(t-d_{max})}]^T$ and $X_{m,i}^{(t)} = [x_i^{(t)} \dots x_i^{(t-d_{max})}]^T$, $X_{c,b}^{(t)} \in \{0, q, -p\}^{n_c \times d_{max}}$ for $t \in \mathbb{N}^*$ and $X_{c,b}^{(0)} = \mathbf{0}_{n_c \times d_{max}}$. Furthermore $d_{max} \geq (l-1) \times 2 + 1$ where l is the number of layers of the firing graph. Finally we provide a parameter $T \in \mathbb{N}$ to the draining algorithm. It controls the targeted number of feedback that an edge should receive before disabling its update. Maintaining update's permissions for each edge requires an operation similar to structure updates. Finally, the draining algorithm iterates forward and backward pass until either G is composed of two distinct connexe components, no structure update is enabled or the maximum number of iterations $T_{max} \in \mathbb{N}$ has been reached.

Algorithm 2 Draining

Input: $G, \mathbb{T}, T_{max}, p, q, \text{decay}$

Output: G drained

```

i ← 0                                     ▷ Initialisation
Xb,c, xb,o, xi, xc, Xm,c, Xm,i ← InitSignals()
while i < Tmax do                                     ▷ Core loop
    xi ← nextGridState()
    xc, xo ← FT(G, xi, xc)                                     ▷ Forward pass
    Xm,c, Xm,i, xc, xo ← FP(xc, xo)
    if i ≥ decay then
        xf ← nextFactoreState()
        Xb,c, Xb,o ← BP(Xb,c, Xb,o, xf, p, q)                                     ▷ Backward pass
        G' ← SU(T, G, Xb,c, Xb,o, Xm,c, Xm,i)
        Xb,c, Xb,i ← BT(G, Xb,c, Xb,o, Xm,c, Xm,i, )
        G ← G'
    end if
    if G.cc == 2 or not G.mask.any() then                                     ▷ Stop conditions
        break
    end if
    i ← i + 1
end while

```

Clearly, the complexity of the algorithm is dominated by the backward transmit and structure updates operations. A standard worst case analysis of those operations gives $\mathcal{O}(n^4 \times d_{max}^2)$, where n is the total number of vertices in the firing graph. Yet this analysis relies on standard complexity time for dense matrix operations, and does not take into account neither the sparsity

of signals and direct link matrices nor the distribution of input vertices's activations. In practice, we have found that the forward and backward propagation of bits and factors's activations is time consuming, especially when both N and T are large numbers. Thus, to reduce running time, batch_size successive bits and factors's states are forward and backward propagated with an efficient vectorization of the equation. The decrease in time complexity of this practical trick is impressive and worth the gain in space complexity of the algorithm. Finally this trick may requires to dynamically change the batch_size so that treshhold for the number of updates at each edges is respected.

3.3 Analysis of the algorithm

Theorem 3.1 *Given a set of sampled bits S , a set of pre-selected bits $I = \{b_1^*, \dots, b_i^*\}$ a target factor f and G , the firing graph built after sampling algorithm. A 5-tuple (ω, N, T, p, q) exists such that the probability of event E : "no input vertices of G have outcoming edges at the end of the draining" is upper bounded. More specifically*

$$\mathbb{P}(E) \leq \sum_{j=0}^{|S|} p_-^j \times \mathbb{P}_{\mathcal{S}}(|\{s \in S \mid \omega_{I \cup \{s\}, |I|+1, f} < \omega\}| = j)$$

Where $p_- = \mathbb{P}(s_{v,f}[N, T, p, q] < 0 \mid \omega_{I \cup \{s\}, i+1, f} < \omega)$. Where $v(I \cup \{s\}, i+1)$, for any $s \in S$, is a vertex of layer 1 of a firing graph G of 2 layers. Furthermore

$$\mathbb{P}(s_{v,f}[N, T, p, q] < 0 \mid \omega_{I \cup \{s\}, i+1, f} < \omega) \leq C \times \max\left(\exp\left(-T \times \left(\frac{\delta_f c}{\sigma}\right)^2\right), \exp(-T \times \delta_f c)\right)$$

With δ_f , C and c are positive constants that depends on ω and i and f . $\text{Var}[s_{v,p,q,f,t}] = \sigma^2$.

Proof. As a reminder, in the core of this proof, we refer to d and d' respectively to the distribution over bits's activations and factors's activations. Given the arrangement of vertices of graph G and the forward equations of the draining algorithm, the activation of any vertices $b \in S$ that will be propagated toward an output vertex, is modelled by the following characteristic polynomial

$$\mathcal{P}_{\{b\},1} \cdot \mathcal{P}_{\{u,v\},2}$$

With $v(S, 1)$ and $u(\{b_1^*, \dots, b_i^*\}, i)$. Thus, using (??), the activity of b that is propagated to the ouput vertex is the same than the activity of a vertex $v(\{b_1^*, \dots, b_i^*, b\}, i+1)$ at the layer 1 of a firing graph G' where b_1^*, \dots, b_i^* and b compound its layer 0. Furtermore, given the time and space consistency of the backpropagation of the feedback from the output vertex, the weight of the outcoming edge of b , at the convergence of the draining algorithm, is either 0 or equal to the score process of vertex v in G' with respect to f , $s_{v,f}[N, T, p, q]$. Then, the first inequality is obtained by developeping

$$\begin{aligned}\mathbb{P}(E) &= \sum_{j=0}^{|S|} p_-^j \times p_+^{|S|-j} \times \mathbb{P}_{\mathcal{S}}(|\{s \in S \mid \omega_{I \cup \{s\}, i+1, f} < \omega\}| = j) \\ &\leq \sum_{j=0}^{|S|} p_-^j \times \mathbb{P}_{\mathcal{S}}(|\{s \in S \mid \omega_{I \cup \{s\}, i+1, f} < \omega\}| = j)\end{aligned}$$

Where

- $I = \{b_1^*, \dots, b_i^*\}$
- $p_- = \mathbb{P}(s_{v,f}(N, T, p, q) < 0 \mid \omega_{I \cup \{s\}, i+1, f} < \omega)$
- $p_+ = \mathbb{P}(s_{v,f}(N, T, p, q) < 0 \mid \omega_{I \cup \{s\}, i+1, f} \geq \omega)$

Then, we choose the value of the positive real ω such that a measure grid's bit b^+ verifies

$$\begin{aligned}b^+ &= \arg \min_{b \in \mathcal{G}} |\omega - \omega_{I \cup \{b\}, i+1, f}| \\ &\text{such that } \omega - \omega_{I \cup \{b\}, i+1, f} > 0\end{aligned}$$

And we define the vertex $v^+(I \cup \{b^+\}, i+1)$ and $\delta^+ = |\omega - \omega_{v^+, f}|$. If vertex v is such that $\omega_{v,f} < \omega$ then using (??) one gets

$$\underbrace{\frac{\|\mathcal{P}_f\|_{d'}}{\|\mathcal{P}_f\|_{d'} + (\omega - \delta) \times (1 - \|\mathcal{P}_f\|_{d'})}}_{\phi_{v,f}} \geq \underbrace{\frac{\|\mathcal{P}_f\|_{d'}}{\|\mathcal{P}_f\|_{d'} + (\omega - \delta^+) \times (1 - \|\mathcal{P}_f\|_{d'})}}_{\phi_{v^+, f}} > \underbrace{\frac{\|\mathcal{P}_f\|_{d'}}{\|\mathcal{P}_f\|_{d'} + \omega \times (1 - \|\mathcal{P}_f\|_{d'})}}_{\phi}$$

For some real $\delta \geq \delta^+ > 0$. Then, We choose the 4-tuple (N, T, p, q) as follow:

$$\begin{aligned}(p, q) &\in \mathbb{N}^2 \text{ such that } \phi \times (p + q) - p < 0 \\ N &= -T \times (\phi \times (p + q) - p) \\ T &\in \mathbb{N} \text{ such that } N \text{ large enough}\end{aligned}$$

Thus, given $\omega_{v,f} < \omega$ one can write

$$\begin{aligned}\mathbb{P}(s_{v,f}[N, T, p, q] < 0) &= \mathbb{P}\left(N + \sum_{t=1}^T s_{v,p,q,t,f} < 0\right) \\ &= \mathbb{P}\left(\sum_{t=1}^T s_{v,p,q,t,f} - T \times \mathbb{E}[s_{v,p,q,1,f}] < -N - T \times \mathbb{E}[s_{v,p,q,1,f}]\right)\end{aligned}$$

Furthermore from the definition of ϕ and $\phi_{v,f}$ we have

$$\phi_{v,f} = \phi + \underbrace{\delta \times \phi \times \phi_{v,f}}_{\delta_{v,f}} \times \frac{1 - \|\mathcal{P}_f\|_{d'}}{\|\mathcal{P}_f\|_{d'}}$$

Yet using equation (??) one have

$$\begin{aligned}\mathbb{E}[s_{v,p,q,1,f}] &= \phi_{v,f} \times (p+q) - p \\ &= (\phi + \delta_{v,f}) \times (p+q) - p\end{aligned}$$

Using $N = -T \times (\phi \times (p+q) - p)$ and the definition of ϕ one have

$$-N - T \times \mathbb{E}[s_{v,p,q,1,f}] = -T \times (p+q) \times \delta_{v,f}$$

Thus

$$\begin{aligned}\mathbb{P}(s_{v,f}[N, T, p, q] < 0) &= \mathbb{P}\left(\sum_{t=1}^T s_{v,p,q,t,f} - \mathbb{E}[s_{v,p,q,t,f}] < -T \times (p+q) \times \delta_{v,f}\right) \\ &\leq \mathbb{P}\left(\left|\sum_{t=1}^T s_{v,p,q,t,f} - \mathbb{E}[s_{v,p,q,t,f}]\right| > T \times (p+q) \times \delta_{v,f}\right) \\ &\leq \mathbb{P}\left(\left|\sum_{t=1}^T s_{v,p,q,t,f} - \mathbb{E}[s_{v,p,q,t,f}]\right| > T \times \delta_f\right)\end{aligned}$$

With $\delta_f = (p+q) \times \delta^+ \times \phi^2 \times \frac{1-\|\mathcal{P}_f\|_{d'}}{\|\mathcal{P}_f\|_{d'}}$.

At this point we have to notice that $\{s_{v,p,q,t,f}\}_{t=1,\dots,T}$ is a sequence of i.i.d random variables with mean μ and variance σ^2 that verifies $|s_{v,p,q,t,f}| \leq \max(p, q)$. Thus one can apply the Chernoff inequality as formulated in [?]. In particular, taking $\lambda = \sigma^{-1}\delta_f$ we obtain

$$\begin{aligned}\mathbb{P}\left(\left|\sum_{t=1}^T s_{v,p,q,t,f} - \mathbb{E}[s_{v,p,q,t,f}]\right| > T\delta_f\right) &= \mathbb{P}\left(\left|\sum_{t=1}^T s_{v,p,q,t} - \mathbb{E}[s_{v,p,q,t}]\right| > \lambda\sigma\sqrt{T}\right) \\ &\leq C \times \max\left(\exp\left(-T \times \left(\frac{\delta_f c}{\sigma}\right)^2\right), \exp(-T \times \delta_f c)\right)\end{aligned}$$

With C, c some positive constant and $\text{Var}[s_{v,p,q,t,f}] = \sigma^2$ for $t \in \{1, \dots, T\}$. Q.E.D.

Theorem 3.2 *Given a set of sampled bits S , a set of pre-selected bits $I = \{b_1^*, \dots, b_i^*\}$ a target factor f and G the firing graph built after sampling algorithm. A sequence of 5-tuple (ω, N, T, p, q) exists such that for each input vertex v of G , from which the output is reachable, we have*

$$\mathbb{P}(\omega_{v,f} > \omega) \leq C \times \max\left(\exp\left(-T \times \left(\frac{\delta_f c}{\sigma}\right)^2\right), \exp(-T \times \delta_f c)\right) \quad (11)$$

Where $v(I \cup \{s\}, i+1)$, for any $s \in S$, is a vertex of layer 1 of a firing graph G of 2 layers and δ_f, C and c are positive constants that depends on ω and i and $\text{Var}[s_{v,p,q,t}] = \sigma^2$.

Proof. As in the proof of the previous theorem, using the arrangement of vertices of G , the property (??) and the forward and backward equations of the draining algorithm, one can show that the weight of the outgoing edge of any vertices $b \in S$ of G is either equal to 0 or to the score process $s_{v,f}[N, T, p, q]$ where $v(\{b_1^*, \dots, b_i^*, b\}, i+1)$ is a vertex at the layer 1 of a firing graph G' where b_1^*, \dots, b_i^* and b compound its layer 0. Furthermore, if sample b still have outgoing edges after draining, then

$$\mathbb{P}(\omega_{v,f} > \omega) = \mathbb{P}_S(\omega_{I \cup \{b\}, i+1, f} > \omega) \times \mathbb{P}(s_{v,f}(N, T, p, q) > 0 | \omega_{I \cup \{b\}, i+1, f} > \omega)$$

Then, we choose the value of the positive real ω such that a bit b^- verifies

$$b^- = \arg \min_{b \in \mathcal{G}} |\omega - \omega_{I \cup \{b\}, f}|$$

such that $\omega - \omega_{I \cup \{b\}, f} < 0$

And we define the vertex $v^-(I \cup \{b^-\}, i+1)$ and $\delta^- = |\omega - \omega_{v^-, f}|$. If v is such that $\omega_{v,f} < \omega$ then using (??) we have

$$\underbrace{\frac{\|\mathcal{P}_f\|_{d'}}{\|\mathcal{P}_f\|_{d'+(\omega+\delta)} \times (1 - \|\mathcal{P}_f\|_{d'})}}_{\phi_{v,f}} \leq \underbrace{\frac{\|\mathcal{P}_f\|_{d'}}{\|\mathcal{P}_f\|_{d'+(\omega+\delta^-)} \times (1 - \|\mathcal{P}_f\|_{d'})}}_{\phi_{v^-,f}} < \underbrace{\frac{\|\mathcal{P}_f\|_{d'}}{\|\mathcal{P}_f\|_{d'+\omega} \times (1 - \|\mathcal{P}_f\|_{d'})}}_{\phi}$$

for some $\delta \geq \delta^- > 0$. Then defining the 4-tuple (N, T, p, q) as

$$(p, q) \in \mathbb{N}^2 \text{ such that } \phi \times (p+q) - p < 0$$

$$N = -T \times (\phi \times (p+q) - p)$$

$$T \in \mathbb{N} \text{ such that } N \text{ large enough}$$

Then, reproducing the same development as it was done in the proof of previous theorem, one can derive a convenient form to easily apply the Chernoff inequality.

$$\mathbb{P}(s_{v,f}(N, T, p, q) > 0 | \omega_{v,f} > \omega) \leq \mathbb{P}\left(\left|\sum_{t=0}^{T-1} s_{v,p,q,t,f} - \mathbb{E}[s_{v,p,q,t,f}]\right| > T \times \delta_f\right)$$

With $\delta_f = (p+q) \times \delta^- \times \phi^2 \times \frac{1 - \|\mathcal{P}_f\|_{d'}}{\|\mathcal{P}_f\|_{d'}}$. Then using the Chernoff inequality as written in [?] using $\lambda = \sigma^{-1} \delta_f$ we obtain

$$\begin{aligned} \mathbb{P}\left(\left|\sum_{t=0}^{T-1} s_{v,p,q,t,f} - \mathbb{E}[s_{v,p,q,t,f}]\right| > T \times \delta_f\right) &= \mathbb{P}\left(\left|\sum_{t=0}^{T-1} s_{v,p,q,t,f} - \mathbb{E}[s_{v,p,q,t,f}]\right| > \lambda \sigma \sqrt{T}\right) \\ &\leq C \times \max\left(\exp\left(-T \times \left(\frac{\delta_f c}{\sigma}\right)^2\right), \exp(-T \times \delta_f c)\right) \end{aligned}$$

With C, c some positive constant and $\text{Var}[s_{v,p,q,t,f}] = \sigma^2$. Q.E.D

3.4 Limit of the generic case

The combination of theorems shows that the association of sampling and draining with the right choice of 5-tuple (ω, N, T, p, q) gives a convenient tool to select measure grid's bits with purity coefficient lower than a target ω . Furthermore, when $T \rightarrow +\infty$, the correct selection is almost certain, which highlights the trade-off between efficiency and complexity of the algorithm that is embedded in the choice of ω and T , on which depends N, p and q . This generic procedure and its analysis deliver a strong framework that eases the derivation of more specific results that may be obtained under specific modelling of latent factors's activations and measure grid signatures. Nevertheless, it leaves two fundamental points clueless

- No possibility to quantify further the effectiveness of the sampling strategy
- No specific procedure or heuristics to choose positive real value ω

In the rest of this paper, we present two particular cases of factor's and measure grid's modelling that enables a better quantification of the sampling strategy and stronger heuristics for the choice of ω .

4 Case of signal plus noise

This particular case is designed to be easy to analyze. We first define the statistical modelling of factors and bits's activations. Then, we quantify the sampling strategy and justify a choice for the 5-tuple (ω, T, N, p, q) . Finally, we present simulations and provide discussion of the results obtained with this special case.

4.1 Statistical modelling

In this particular case, we assume that the target factor f is linked to some $|\mathcal{G}(f)|=k$ measure grid's bits and activates with probability p_f . We also assume that bits of the measure grid are identically and independently subject to a noisy activation with probability p_N . We may see noisy activations as the result of n noisy latent factors, linked to exactly 1 bit of the measure grid, that is $K = n + 1$. Under this model, the probability for a bit $b \in \mathcal{G}$ to activate is defined as

$$\mathbb{P}(\text{"}b \text{ active"}) = \begin{cases} p_f + p_N \times (1 - p_f), & \text{if } b \in \mathcal{G}(f^*) \\ p_N, & \text{Otherwise} \end{cases}$$

As a consequence, for any $I \in S(\mathcal{G})$ such that $|I \cap \mathcal{G}(f)|=i$ and $j = |I|-i$ if we set $x \in F_2(\{I\}, \mathcal{G})$, the distribution over measure grid bits's activations is defined as

$$d_x = \begin{cases} p_N^{i+j} \times (1 - p_N)^{n-i-j} \times (1 - p_f) + p_N^j \times (1 - p_N)^{n-k-j} \times p_f, & \text{if } i = k \\ p_N^{i+j} \times (1 - p_N)^{n-i-j} \times (1 - p_f), & \text{Otherwise} \end{cases}$$

In the rest of the section, we will always refer to this distribution as d .

4.2 Evaluation of bits

Let G be a firing graph with a layer 0 composed of measure grid's bits. Then, the precision of a vertex $v(I, |I|)$ of layer 1 of G , with respect to f , depends only on $|I|$ and $|I \cap \mathcal{G}(f)|$. Indeed, if $|I \cap \mathcal{G}(f)| = i$ then

$$\phi_{v,f} = \frac{pf}{p_f + (1 - p_f) \times p_N^i}$$

With identification of terms using (??) we have $\omega_{v,f} = p_N^i$ and using previously defined distribution, one finds that $\mu_{v,f} = p_N^{|I|-i}$, $\nu_{v,f} = p_N^{|I|}$. Besides, given a set of bits I such that $I \subset \mathcal{G}(f)$, if $b \in \mathcal{G}(f) \setminus I$, the precision of vertex $v(I \cup \{b\}, |I|+1)$ with respect to f is

$$\phi_v = \frac{pf}{p_f + (1 - p_f) \times p_N^{|I|+1}}$$

if $b \notin \mathcal{G}(f)$

$$\phi_v = \frac{pf}{p_f + (1 - p_f) \times p_N^{|I|}}$$

4.3 Sampling Strategy

In this particular case we follow the generic sampling procedure \mathcal{S} with parameter p_S . Thus, using the previously defined statistical distribution of bits's activations, if we denote S , the set of sampled bits using \mathcal{S} , the distribution of the cardinal of S is

$$\mathbb{P}(|S|=s) = \begin{cases} \binom{n-k}{s-k} \times p_N^{s-k} \times (1-p_N)^{n-k-s} \times p_S, & \text{if } s \geq k \\ 0, & \text{otherwise} \end{cases}$$

Thus its expected size is $\mathbb{E}[|S|] = k + (n-k) \times p_N \times p_S$. Furthermore if $I = \{b_1^*, \dots, b_i^*\} \in S(\mathcal{G})$ is some set of pre-selected bits and S is a set of bits sampled using \mathcal{S} , a positive real ω_i exists such that

$$\begin{aligned} \mathbb{P}_{\mathcal{S}}(|\{s \in S \setminus \omega_{I \cup \{s\}, |I|+1, f} < \omega_i\}| = j) &= \mathbb{P}_{\mathcal{S}}(|\{s \in S \setminus s \in \mathcal{G}(f)\}| = j) \\ &= \binom{|\mathcal{G}(f)|-i}{j} \times p_S^j \times (1-p_S)^{|\mathcal{G}(f)|-i-j} \end{aligned}$$

4.4 Identification of factors

First, in the case of a single sampled firing graph, one can see that bits's purity coefficients take only two values with respect to f

$$\omega_{\{b\}, 1, f} = \begin{cases} p_N, & \text{if } b \in \mathcal{G}(f) \\ 1, & \text{otherwise} \end{cases}$$

Thus if we choose

$$\omega_0 = \frac{(1 + p_N)}{2}$$

It maximizes the purity margin defined as

$$\delta_0 = \frac{(\omega_0 - \omega_{\{b\},1,f}) + (\omega_{\{b'\},1,f} - \omega_0)}{2} = \frac{(1 - p_N)}{2}$$

Where $b \in \mathcal{G}(f)$ and $b' \notin \mathcal{G}(f)$. In the case of a joint sampled firing graph in which a set $I = \{b_1^*, \dots, b_i^*\}$ of $i \in \mathbb{N}^*$ pre-selected bits that verify $\forall b \in I, b \in \mathcal{G}(f)$, remaining bit's purity coefficients with respect to f can take again two values

$$\omega_{I \cup \{b\},i+1,f} = \begin{cases} p_N^{i+1}, & \text{if } b \in \mathcal{G}(f) \\ p_N^i, & \text{otherwise} \end{cases}$$

Thus if we choose

$$\omega_i = \frac{(1 + p_N) \times p_N^i}{2}$$

it maximizes the purity margin defined as

$$\delta_i = \frac{(\omega_i - \omega_{I \cup \{b\},i+1,f}) + (\omega_{I \cup \{b'\},i+1,f} - \omega_i)}{2} = \frac{(1 - p_N) \times p_N^i}{2}$$

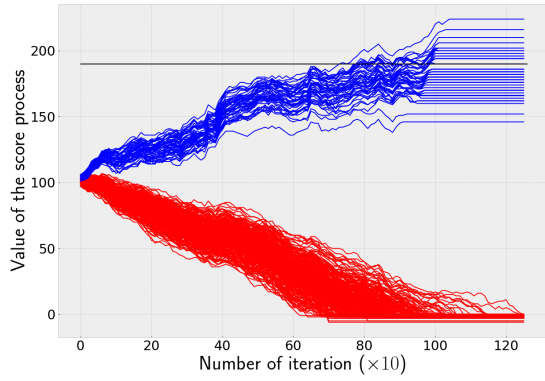
Where $b \in \mathcal{G}(f)$ and $b' \notin \mathcal{G}(f)$. Finally we define the 5-tuple (N_i, T_i, p, q) as

$$\begin{aligned} (p, q) &\in \mathbb{N}^2 \text{ such that } \phi_i \times (p + q) - p \leq 0 \text{ and } \phi'_i \times (p + q) - p > 0 \\ N &= -T \times (\phi \times (p + q) - p) \\ T &\in \mathbb{N} \text{ such that } N \text{ large enough} \end{aligned}$$

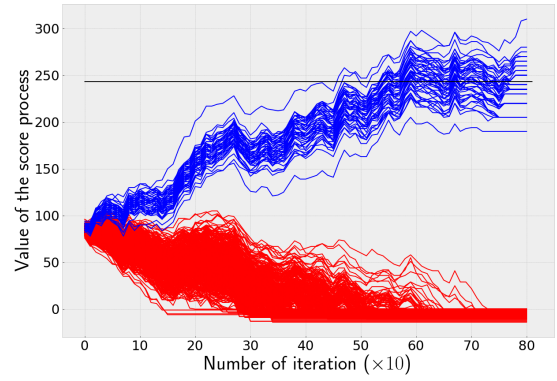
Where $t \in \mathbb{N}$, $\phi_i = \frac{p_f}{p_f + \omega_i \times (1 - p_f)}$ and $\phi'_i = \frac{p_f}{p_f + (\omega_i - \delta_i) \times (1 - p_f)}$.

4.5 Simulation

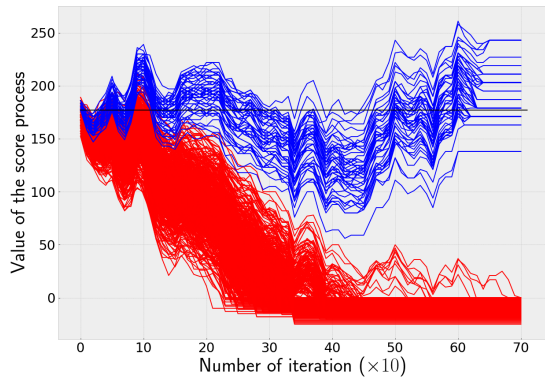
The signal plus noise model is implemented in python and mainly uses standard numpy and scipy modules to generate random signal that fit its probabilistic model. More details about the implementation can be found in appendix B. We generate $n = 1000$ bits that randomly activate with probability p_N and we choose randomly $|\mathcal{G}(f)| = 50$ bits that are linked to a latent factor that activates with probability $p_f = 0.3$. Finally we build the single sampled firing graph using $p_S = 1$.



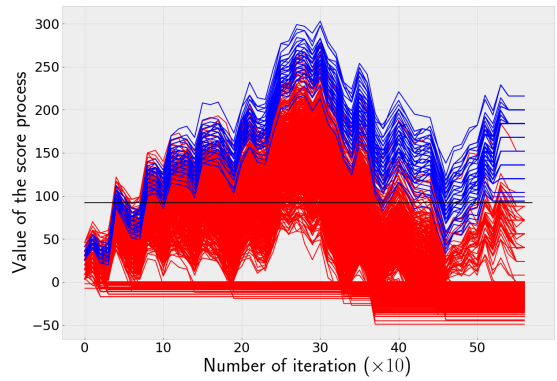
(a) $p_N = 0.3, (p, q) = (1, 1)$



(b) $p_N = 0.5, (p, q) = (2, 3)$



(c) $p_N = 0.7, (p, q) = (3, 5)$



(d) $p_N = 0.9, (p, q) = (5, 11)$

Figure 6: Observation of the score process for different SNR models $T = 500$

Each subplot of figure ?? shows the weight of outgoing edges of sampled vertices. Blue lines show the weight of edges outgoing from sampled bit $b \in \mathcal{G}(f)$ and red lines correspond to the weight of edges outgoing from sampled bit $b \notin \mathcal{G}(f)$. Finally the black horizontal line represents the theoretical mean value of $s_{v,f}[N, T, p, q]$ of a vertex with characteristic polynome $\mathcal{P}_v = \mathcal{P}_{\{b\},1}$, with $b \in \mathcal{G}(f)$. As theory suggests, we can see two distinct phenomenons, blues lines converge around theoretical mean for process of bits linked to the target factor and red lines converge to 0. However, the higher is p_N , the less noticeable is the distinction between each process. This is explained by the fact that the higher is p_N , the closer are the precision of bits linked to target factor f and the precision of noisy bits. Futhermore the later observation induces a high value of $p + q$ which result in a more volatile score process. For the second simulation we use $n = 1000$, $|\mathcal{G}(f)| = 50$, $p_f = 0.3$, $T = 200$ and $p_S = 0.5$. Yet at the end of the draining we choose all the input vertices of the firing graph that still have an outgoing edge and use their combined activation as an estimator of the target factor's activation. We then measure their precision and recall over 100 repetition for each SNR ratio

P_N	Mean ϕ	Standard deviation ϕ	Mean ψ	Standard deviation ψ	Number of fails
0.3	1.0	0.00	0.87	0.30	0
0.5	1.0	0.03	0.66	0.42	0
0.7	0.97	0.13	0.43	0.46	4
0.9	0.75	0.14	0.13	0.25	19

Table 1: Evaluation of naive factor’s activation estimation, $T = 200$, 100 repetitions

Table ?? shows quality indicators of the estimator for different SNR ratio. The two first columns give respectively the mean and standard deviation of the precision of the estimator. The two following columns are respectively the mean and the standard deviation of the recall of the estimator. Finally the last column is the number of experiments that ended without any input vertices having a path towards the output, so that the construction of an estimator is not possible. Again, we see that the quality of the estimator drops as the theoretical precision between noisy bits and factor’s bits are close to each other. Yet it reveals that this naive estimator, for a reasonable SNR ratio, is still efficient to predict the activation of target latent factor. Finally, we simulate the signal plus noise model in the settings of joint sampled firing graph. We use a measure grid of $n = 1000$ bits from which we sampled randomly $|\mathcal{G}(f)| = 50$ bits linked to target factor f that activates with probability $p_f = 0.3$ and we set $p_N = 0.6$. Finally we built the joint sampled firing graph by pre-selecting randomly 5 bits linked to the factor and running the sampling algorithm described previously using $p_S = 1$.

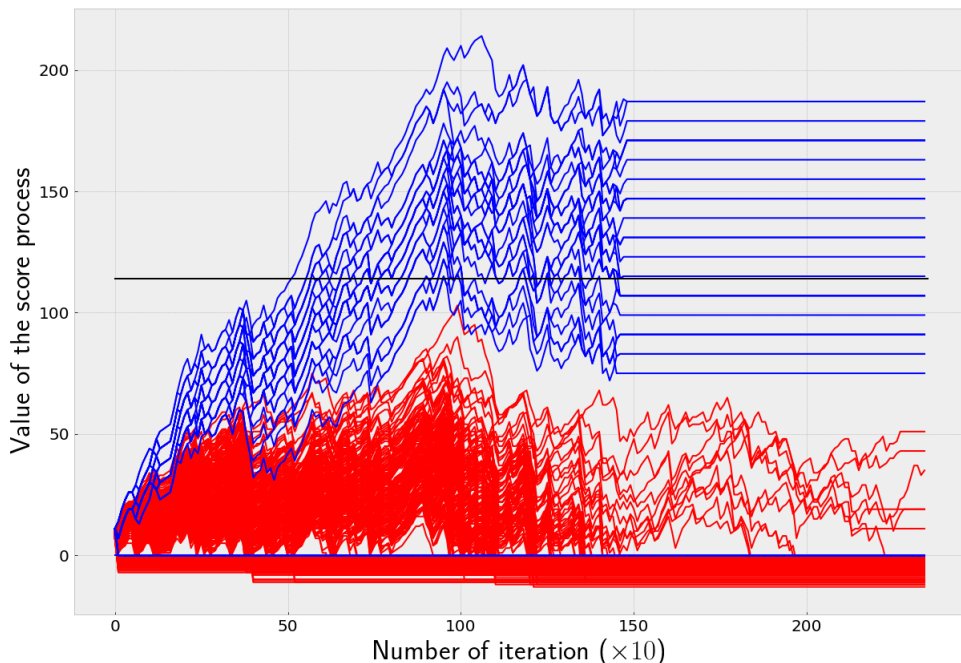


Figure 7: Observation of the score process in a joint sampled firing graph with $T = 500$ and $(p, q) = (7, 1)$

In this case, we obtain $N = 7$ and $\omega_5 \simeq 0.062$ when following the procedure described in previous section. As for the first experiment, blue lines show the weight of edges outcoming from sampled bit $b \in \mathcal{G}(f)$ and red lines correspond to the weight of edges outcoming from sampled bit $b \notin \mathcal{G}(f)$. The black horizontal line represents the theoretical mean value of $s_{v,f}[N, T, p, q]$, where v has characteristic polynome $\mathcal{P}_v = \mathcal{P}_{\{b_1, \dots, b_5, b\}, 5}$, with $\{b_1, \dots, b_5\}$ the set of pre-selected bits and $b \in \mathcal{G}(f)$. The simulation validate the expectation from theory and the high value of $p + q$ explains the high volatility of score processus.

5 Case of sparse measure grid

This particular case is more complex than the previous one. We first define the statistical signature of factors and bits's activations. Then we quantify the sampling strategy and justify a choice for the 5-tuple (ω, T, N, p, q) . Finally, we present simulations and provide discussion of results obtained with this particular case.

5.1 Statistical modelling

Latent factor activation

We assume that each of the K latent factors activates independantly with probability p_f . As a consequence, for any $I \in S(\mathcal{F})$, if we define x such that $x \in F_2(\{I\}, \mathcal{F})$, we can define the distribution of factor's activation as

$$d'_x = \binom{|I|}{K} \times p_f^{|I|} \times (1 - p_f)^{K - |I|}$$

Measure grid activation

We assume two major properties of activations of measure grid's bits.

- For each factor $f \in \mathcal{F}$, each bit $b \in \mathcal{G}$ has equal probability p_g to belong to $\mathcal{G}(f)$.
- For each factor $f \in \mathcal{F}$, for each couple $b_1, b_2 \in \mathcal{G}^2$, events " $b_1 \in \mathcal{G}(f)$ " and " $b_2 \in \mathcal{G}(f)$ " are independent.

As a consequence the probability for a bit b to activate, given that every factor of some set $\{f_1, \dots, f_k\} \subset \mathcal{F}$ is active, writes

$$\begin{aligned} \mathbb{P}(b \text{ active} \mid f_1, \dots, f_k \text{ active}) &= \sum_{i=1}^k \binom{k}{i} \times p_g^i \times (1 - p_g)^{k-i} \\ &= 1 - (1 - p_g)^k \end{aligned}$$

The above quantity depends only on the number of active latent factors. Thus, for any $I \in S(\mathcal{G})$, if we define $x \in F_2(S(\mathcal{G}), \mathcal{G})$, we can define the distribution of bits's activations as

$$d_x = \sum_{k=1}^K \left[\binom{K}{k} \times p_f^k \times (1 - p_f)^{K-k} \right] \times \left[p_{g|k}^i \times (1 - p_{g|k})^{n-i} \right]$$

With $i = |I|$ and $p_{g|k} = \mathbb{P}(b \text{ active} | f_1, \dots, f_k \text{ active})$.

5.2 Evaluation of bits

Let G be a firing graph whose layer 0 is composed of measure grid's bits. Given a target factor f , the precision with respect to f of a vertex $v(\{b\}, 1)$ of the layer 1 of G depends on whether $b \in \mathcal{G}(f)$ and on $|\mathcal{G}^{-1}(b)|$. Indeed, if $b \in \mathcal{G}(f)$ and $|\mathcal{G}^{-1}(b)| = l$, $l \in \{1, \dots, K\}$, it will be said to have a purity rank of l and its precision with respect to f writes

$$\phi_{v,f} = \frac{pf}{p_f + (1 - p_f) \times \omega_l}$$

where $\omega_l = 1 - (1 - p_f)^{l-1}$. If $b' \notin \mathcal{G}(f)$ the precision of $v'(\{b'\}, 1)$ with respect to f writes

$$\phi_{v',f} = p_f$$

Futhermore, if we have a vertex $v(I, |I|)$ such that $\forall b \in I$, $b \in \mathcal{G}(f)$ and $\min_{b \in I} |\mathcal{G}^{-1}(b)| = l$, then the precision of v with respect to f verifies

$$\phi_{v,f} \leq \frac{pf}{p_f + \omega_l^- \times (1 - pf)}$$

With

$$\omega_l^- = \sum_{k=K-l-1}^K \binom{K}{k} p_f^k \times (1 - p_f)^{K-k}$$

The minimum purity coefficient one can obtained with bits that verifies $b \in \mathcal{G}(f)$ and $|\mathcal{G}^{-1}(b)| = l$. That is, the case of a vertex $v(I, |I|)$ with I composed of every possible $\binom{K}{l}$ such bits.

5.3 Sampling Strategy

We follow the generic sampling procedure \mathcal{S} with parameter $p_{\mathcal{S}}$. Although it is not hard to derive key quantification such as $\mathbb{E}[|S|]$ or probabilities to sample bits linked to a target factor f under this modelling, generic formulas are not elegant and present not much interest in this simulation.

5.4 Identification of factors

First, in the case of a single sampled firing graph, for any grid's bits linked to factor f , there is only K different purity coefficients possible. Thus we may set ω to ω_l , using l reasonably small to differentiate lower purity rank from greater purity rank samples. In the case of a joint sampled graph, where a set of $I = \{b_1^*, \dots, b_l^*\}$ were pre-selected then the choice of ω is not trivial and is hard to be efficiently and generically derived. Let $\omega_{I,|I|,f}$ the purity coefficient of

the pre-selected set of bits we set $\omega = \omega_{I,|I|,f} - \delta$ where $\delta \in \mathbb{R}$ should be chosen with caution. Finally we choose the 5-tuple $\{(\omega, N, T, p, q)$ as

$$\begin{aligned} (p, q) &\in \mathbb{N}^2 \text{ such that } \phi \times (p + q) - p < 0 \\ N &= -T \times (\phi \times (p + q) - p) \\ T &\in \mathbb{N} \text{ such that } N \text{ large enough} \end{aligned}$$

Where $\phi = \frac{p_f}{p_f + \omega \times (1 - p_f)}$.

5.5 Simulation

The sparse measure grid model is implemented using python and the standard python numpy and scipy modules to generate random signal that fit its probabilistic model. In our case we generate $n = 1000$ bits with $K = 10$ latent factors that activate with probability $p_f = 0.3$ and we link measure grid's bits independently with probability $p_g = 0.3$. Finally we built the single sampled firing graph running the sampling algorithm described previously, using $p_S = 1$. Finally, we set $\omega = \omega_{10}$, the higher purity coefficient for bits linked to the target factor f .

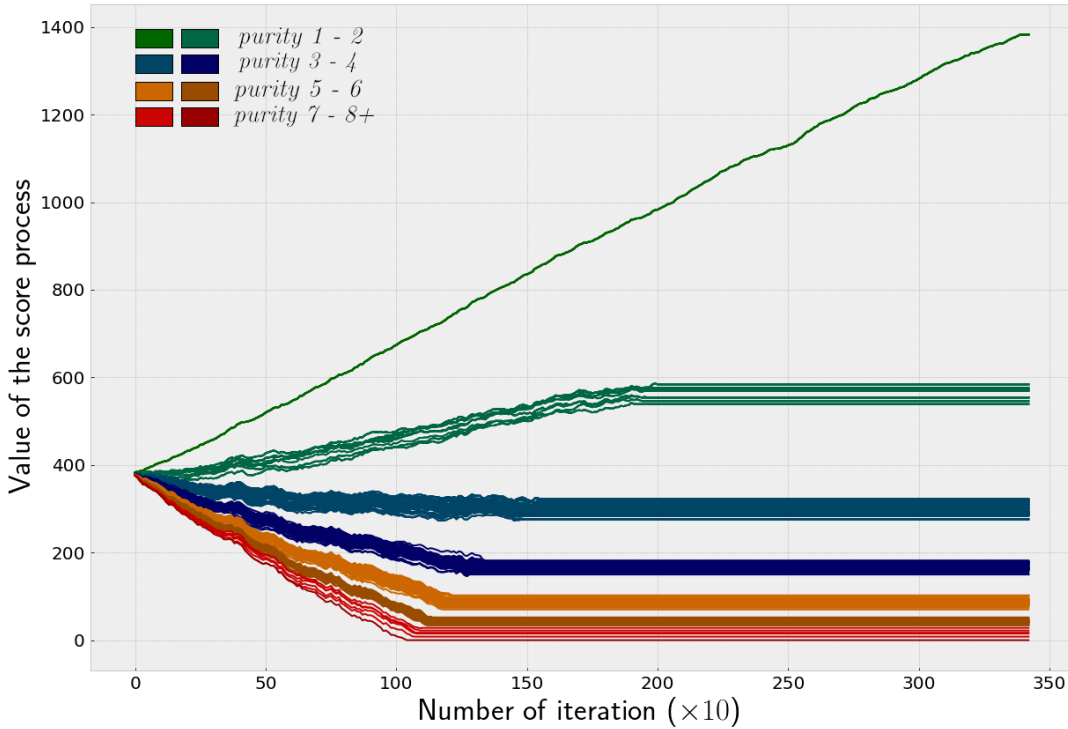
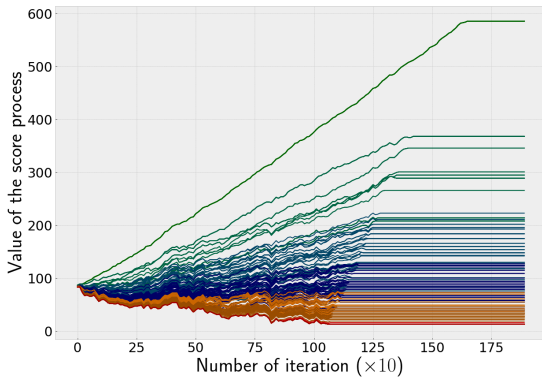


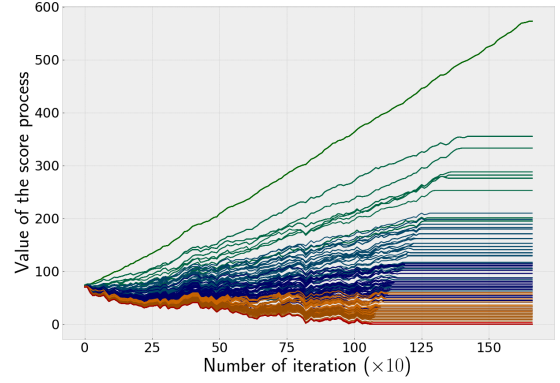
Figure 8: Observation of the score process in a single sampled firing graph with $T = 1000$ and $(p, q) = (1, 1)$

We clearly see a rapid differentiation of score processus according to their purity rank. We can also observe that, at the end of draining, the higher the purity coefficient is, the closer are weights

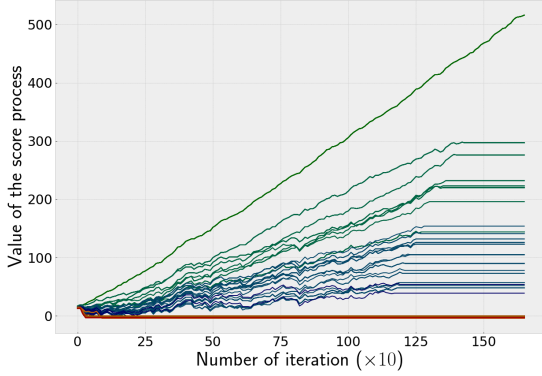
of corresponding edges. Finally, the behaviour of score processus validates the efficiency of the draining algorithm to rank bits of the measure grids blindly, in an attempt to identify latent factors. The second experiment with the sparse measure grid model aims to give intuition on the choice of δ used for draining a joint sampled firing graph. As for the previous simulation, we generate $n = 1000$, bits with $K = 10$ latent factors that activate with probability $p_f = 0.3$ and we link measure grid's bits independently with probability $p_g = 0.3$. Then we choose randomly $i = 5$ bits, denoted by $I = \{b_1^*, \dots, b_i^*\}$, with purity rank 4 with respect to the target factor f . Finally we sampled and built the joint sampled firing graph using $p_S = 1$ and the set of pre-selected bits I . The procedure described in the previous section to choose the target purity coefficient consists in estimating the purity coefficient $\hat{\omega}_{I,|I|,f}$ and to set δ so that $\omega = \hat{\omega}_{I,|I|,f} - \delta$.



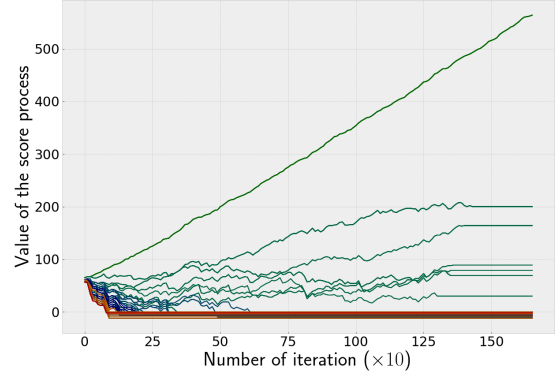
(a) $\delta = 0$.



(b) $\delta = 10^{-2}$



(c) $\delta = 5 \times 10^{-2}$



(d) $\delta = 10^{-1}$

Figure 9: Observation of sampled bit's score processus in a joint sampled firing graph $i = 5$, $T = 500$ and $(p, q) = (1, 1)$

In each simulation, $\hat{\omega}_{I,|I|,f}$ has been estimated using 1000 samples and we use $T = 500$. Furthermore, each figure corresponds to a different value of δ that induces different values of ω , set as $\omega = \hat{\omega}_{I,|I|,f} - \delta$. As for the first experiment, the different colored lines in each subfigure show the weight of edges outcoming from sampled bits with different purity ranks. As expected, we

can see that the higher δ is, the more discriminative the draining procedure is. If δ is set to 0, then every sampled bits will remain connected in the firing graph after draining, which is not of great interest. Yet, if δ is set too high we may end with two connexe components, which is not desirable neither. Thus, the experiment confirms the difficulties that we may face choosing the right value for δ .

6 Discussion

This paper has presented an algorithm that consists in a generic optimisation of a firing graph, in an attempt to solve the abstract task of identifying latent factor's activations. Furthermore it has provided theoretical certitude on the effectiveness of the procedure. However, the iterative optimisation method associated with the diversity and flexibility of the architecture of a firing graph opens doors to further applications, notably in the field of inverse problem and in the very hype field of machine learning. Indeed in supervised classification, we are given a dataset composed of features that may be numerical or categorical description of samples and targets that specify the class of samples. If we assume that the activation of a target is a combination of latent factors's activations and that we operate the minimum transformation of features so that they take the form of a measure grid, a light layer of procedures could turn our solution into a supervised classifier. The specificity of such a learner would give it an interesting position in the supervised learning landscape. Indeed, its iterative optimisation and flexible architecture could make it an adaptative learner, that scale to large dataset, with minimum processing work on raw data, in the manner of a neural network. Yet unlike neural network the algorithm handle very efficiently categorical or sparse feature space. Furthermore, compared to the most advanced tree based classification, its flexible architecture is more suitable to learning update and on-the-fly evaluation or addition of new features. Finally, given the hype granted to the field of machine learning nowadays, both in the scientific community and civil society, it would be common sense to orient this piece of research to this field.

Appendices

A Properties

Partition

Let $v_1(I, l_0)$, $v_2(J, 0)$ and $v_3(K, 0)$, be three vertices at the layer 1 of some firing graph, with the same input domain \mathcal{G} such that $I = J \cup K$ and $J \cap K = \emptyset$. result (??) stands that $\forall x \in F_2(S(\mathcal{G}), \mathcal{G})$

$$\mathcal{P}_I^{l_0}[x] = \sum_{l=l_0}^{|I|} \sum_{j=0}^{|J|} \mathcal{P}_{J,j}[x] \cdot \mathcal{P}_{K,l-j}[x]$$

Proof. The statement above can also be written

$$\mathcal{P}_{I,l}[x] = \sum_{j=0}^{|J|} \mathcal{P}_{J,j}[x] \cdot \mathcal{P}_{K,l-j}[x]$$

$\forall l \in \{l_0, \dots, |I|\}$, now we propose a simple proof by contradiction. Let $l \in \{l_0, \dots, |I|\}$, $X \in S(\mathcal{G})$ and $x \in F_2(\{X\}, \mathcal{G})$ such that

$$\begin{aligned} \mathcal{P}_{I,l}[x] &= 1 \\ \sum_{j=0}^{|J|} \mathcal{P}_{J,j}[x] \cdot \mathcal{P}_{K,l-j}[x] &= 0 \end{aligned}$$

Yet, if J and K is a partition of I and $|I \cap X| = l$, then $(j^*, k^*) \in \{0, \dots, |J|\} \times \{0, \dots, |K|\}$ exists such taht

$$\begin{aligned} |X \cap J| &= j^* \\ |X \cap K| &= k^* \\ j^* + k^* &= l \end{aligned}$$

Thus for $x \in F_2(\{X\}, \mathcal{G})$

$$\mathcal{P}_{J,j^*}[x] \cdot \mathcal{P}_{K,k^*}[x] = 1$$

which contradicts our first assumption. Let $l \in \{l_0, \dots, |I|\}$, $X \in S(\mathcal{G})$ and $x \in F_2(\{X\}, \mathcal{G})$ such that

$$\mathcal{P}_{I,l}[x] = 0$$

$$\sum_{j=0}^{|J|} \mathcal{P}_{J,j}[x] \cdot \mathcal{P}_{K,l-j}[x] = 1$$

Thus above statement implies that $j^* \in \{0, \dots, \min(l, |J|)\}$ exists such that

$$\mathcal{P}_{J,j^*}[x] \cdot \mathcal{P}_{K,l-j^*}[x] = 1$$

Thus $|X \cap J| = j^*$ and $|X \cap K| = l - j^*$. Since J and K is a partion of I we must have

$$|X \cap I| = |X \cap J| + |X \cap K| = l$$

As a consequence for $x \in F_2(\{X\}, \mathcal{G})$, $\mathcal{P}_{I,l}[x] = 1$ and give us the contradiction.

Result (??) is a particular case of result (??)

Decomposition

Let G be a firing graph with layer 0 composed of \mathcal{G} . Let $u(I, l_u), v(I', l_v)$ such that $I \cap I' = \emptyset$ be vertices of layer 1 and $w(\{u, v\}, 2)$ be a vertex of layer 2. Let $K \in \cup_{l \in \{l_v, \dots, |I'|\}} \mathcal{S}(I', l)$, $x \in F_2(\mathcal{S}(\mathcal{G}), \mathcal{G})$ and $x' = \begin{bmatrix} \mathcal{P}_u[x] & \mathcal{P}_v[x] \end{bmatrix}$, the result (??) stands that

$$\mathcal{P}_{K,|K|}[x] \cdot \mathcal{P}_{\{u,v\},2}[x'] = \sum_{l=l_u}^{|I|} \sum_{J \in \mathcal{S}(I,l)} \mathcal{P}_{J \cup K, l+|K|}[x]$$

Proof. The proof the above statement is derived by a straight forward development of the equation, first using result (??) and the fact that K and $I' \setminus K$ is a partition of I' we can write

$$\begin{aligned} \mathcal{P}_{\{u,v\},2}[x'] &= \left(\sum_{l=l_u}^{|I|} \mathcal{P}_{I,l}[x] \right) \cdot \left(\sum_{l=l_v}^{|I'|} \mathcal{P}_{I',l}[x] \right) \\ &= \left(\sum_{l=l_u}^{|I|} \mathcal{P}_{I,l}[x] \right) \cdot \left(\sum_{l=l_v}^{|I'|} \sum_{k=0}^{|K|} \mathcal{P}_{K,k}[x] \cdot \mathcal{P}_{I' \setminus K, l-k}[x] \right) \end{aligned}$$

Thus

$$\begin{aligned}
\mathcal{P}_{K,|K|}[x] \cdot \mathcal{P}_{\{u,v\},2}[x'] &= \mathcal{P}_{K,|K|}[x] \cdot \left(\sum_{l=l_u}^{|I|} \mathcal{P}_{I,l}[x] \right) \cdot \left(\sum_{l=l_v}^{|I'|} \sum_{k=0}^{|K|} \mathcal{P}_{K,k}[x] \cdot \mathcal{P}_{I' \setminus K, l-k}[x] \right) \\
&= \left(\sum_{l=l_u}^{|I|} \mathcal{P}_{I,l}[x] \right) \cdot \left(\sum_{l=l_v}^{|I'|} \mathcal{P}_{K,|K|}[x] \cdot \mathcal{P}_{I' \setminus K, l-|K|}[x] \right) \\
&= \left(\sum_{l=l_u}^{|I|} \mathcal{P}_{I,l}[x] \cdot \mathcal{P}_{K,|K|}[x] \right) \cdot \underbrace{\sum_{l=0}^{|I'|-|K|} \mathcal{P}_{I' \setminus K, l}[x]}_{=1} \\
&= \sum_{l=l_u}^{|I|} \mathcal{P}_{I,l}[x] \cdot \mathcal{P}_{K,|K|}[x]
\end{aligned}$$

The last line is equal to $\sum_{l=l_u}^{|I|} \sum_{J \in S(I,l)} \mathcal{P}_{J \cup K, l+|K|}[x]$ and thus the proof is achieved.

Result (??) is a particular case of result (??)

Let G be a firing graph with layer 0 composed by measure grid's bits \mathcal{G} and $f \in \mathcal{F}$ denote some target factor that is linked to some bit of the measure grid. The distribution of activation of latent factors and measure grid's bits will be denoted d and d' and the event "factor f is active" will be denoted by e . Furthermore, let v be some vertex of G whose characteristic polynome respects $\mathcal{P}_v = \mathcal{P}_I^l$ with $(I, l) \in S(\mathcal{G}), \{1, \dots, |I|\}$ and $f \in \mathcal{F}$ some factor.

Precision of vertex

The result (??) stands that the precision of v with respect to f writes

$$\phi_{v,f} = \frac{\|\mathcal{P}_f\|_{d'}}{\|\mathcal{P}_f\|_{d'} + (1 - \|\mathcal{P}_f\|_{d'}) \times \omega_{I,l,f}}$$

Proof. First, starting from the definition of $\phi_{v,f}$

$$\phi_{v,f} = \phi_{I,l,f} = \frac{\|\mathcal{P}_I^l\|_{d,e}}{\|\mathcal{P}_I^l\|_d}$$

Thus using $\mathcal{P}_* = \mathcal{P}_{\mathcal{G}(f), |\mathcal{G}(f)|}$ one have

$$\phi_{v,f} = \frac{\langle \mathcal{P}_I^l, \mathcal{P}_* \rangle_{d,e} + \langle \mathcal{P}_I^l, \bar{\mathcal{P}}_* \rangle_{d,e}}{\langle \mathcal{P}_I^l, \mathcal{P}_* \rangle_{d,e} + \langle \mathcal{P}_I^l, \bar{\mathcal{P}}_* \rangle_{d,e} + \langle \mathcal{P}_I^l, \mathcal{P}_* \rangle_{d,\bar{e}} + \langle \mathcal{P}_I^l, \bar{\mathcal{P}}_* \rangle_{d,\bar{e}}}$$

Yet $\langle \mathcal{P}_I^l, \bar{\mathcal{P}}_* \rangle_{d,e} = 0$

$$\phi_{v,f} = \frac{\langle \mathcal{P}_I^l, \mathcal{P}_* \rangle_{d|e} \times \|\mathcal{P}_f\|_{d'}}{\langle \mathcal{P}_I^l, \mathcal{P}_* \rangle_{d|e} \times \|\mathcal{P}_f\|_{d'} + (\langle \mathcal{P}_I^l, \mathcal{P}_* \rangle_{d|\bar{e}} + \langle \mathcal{P}_I^l, \bar{\mathcal{P}}_* \rangle_{d|\bar{e}}) \times \|\bar{\mathcal{P}}_f\|_{d'}}$$

Finally by identification of term

$$\phi_{v,f} = \frac{\mu_{I,l,f} \times \|\mathcal{P}_f\|_{d'}}{\mu_{I,l,f} \times \|\mathcal{P}_f\|_{d'} + \nu_{I,l,f} \times (1 - \|\mathcal{P}_f\|_{d'})}$$

Which gives the expected result.

The result (??) stands that if $\mu_{v,f} = 1$ we have

$$\phi_{v,f} \leq \frac{\|\mathcal{P}_f\|_{d'}}{\|\mathcal{P}_f\|_{d'} + (1 - \|\mathcal{P}_f\|_{d'}) \times \omega_{\mathcal{G}(f),|\mathcal{G}(f)|,f}}$$

Proof. The result (??) can be proven by simple contradiction, suppose there is a tuple $(I, l_0) \neq (\mathcal{G}(f), |\mathcal{G}(f)|)$ such that

$$\begin{aligned} \mu_{I,l_0,f} &= 1 \\ \nu_{I,l_0,f} &< \nu_{\mathcal{G}(f),|\mathcal{G}(f)|,f} \end{aligned}$$

First, denoting $\mathcal{P}_{\mathcal{G}(f),|\mathcal{G}(f)|} = \mathcal{P}_*$ and using (??) we have

$$\begin{aligned} \langle \mathcal{P}_I^{l_0}, \mathcal{P}_* \rangle_d &= \sum_{l=l_0}^{|I|} \sum_{k=0}^{I \cap \mathcal{G}(f)} \langle \mathcal{P}_{I \cap \mathcal{G}(f),k} \cdot \mathcal{P}_{I \setminus \mathcal{G}(f),l-k}, \mathcal{P}_* \rangle_d \\ &= \sum_{l=l_0}^{|I|} \sum_{k=0}^{I \cap \mathcal{G}(f)} \sum_{x \in F_2(S(\mathcal{G}), \mathcal{G})} \mathcal{P}_{I \cap \mathcal{G}(f),k}[x] \cdot \mathcal{P}_{I \setminus \mathcal{G}(f),l-k}[x] \cdot \mathcal{P}_*[x] \times d_x \\ &= \sum_{x \in F_2(S(\mathcal{G}), \mathcal{G})} \mathcal{P}_*[x] \cdot \left(\sum_{l=l_0}^{|I|} \mathcal{P}_{I \setminus \mathcal{G}(f),l-|I \cap \mathcal{G}(f)|}[x] \right) \times d_x \end{aligned}$$

Where d is any well defined distribution on $F_2(S(\mathcal{G}), \mathcal{G})$. Thus, we have

$$\mu_{I,l_0,f} = \sum_{x \in F_2(S(\mathcal{G}), \mathcal{G})} \mathcal{P}_*[x] \cdot \left(\sum_{l=l_0}^{|I|} \mathcal{P}_{I \setminus \mathcal{G}(f),l-|I \cap \mathcal{G}(f)|}[x] \right) \times d_{x|e}$$

As a consequence, in order to have $\mu_{I,l_0,f} = 1$ we must have $\{\mathcal{P}_{I \setminus \mathcal{G}(f),l-|I \cap \mathcal{G}(f)|}\}_{l \in \{l_0, \dots, |I|\}}$ to be a partion of $F_2(S(\mathcal{G}), \mathcal{G})$. Thus $l_0 \leq |I \setminus \mathcal{G}(f)|$. On the other end the precision coefficient writes

$$\nu_{I,l_0,f} = \sum_{x \in F_2(S(\mathcal{G}), \mathcal{G})} \mathcal{P}_*[x] \cdot \left(\sum_{l=l_0}^{|I|} \mathcal{P}_{I \setminus \mathcal{G}(f),l-|I \cap \mathcal{G}(f)|}[x] \right) \times d_{x|\bar{e}} + \langle \mathcal{P}_I^{l_0}, \bar{\mathcal{P}}_* \rangle_{d|\bar{e}}$$

so if $l_0 \leq |I \setminus \mathcal{G}(f)|$

$$\begin{aligned} \nu_{I,l_0,f} &= \nu_{\mathcal{G}(f),|\mathcal{G}(f)|,f} + \langle \mathcal{P}_I^{l_0}, \bar{\mathcal{P}}_* \rangle_{d|\bar{e}} \\ &\geq \nu_{\mathcal{G}(f),|\mathcal{G}(f)|,f} \end{aligned}$$

Since $\langle \mathcal{P}_I^{l_0}, \bar{\mathcal{P}}_* \rangle_{d|\bar{e}} \geq 0$, which lead to a contradiction.

Recall of vertex

The result (??) stands that the recall of v with respect to f is

$$\psi_{v,f} = \mu_{I,l,f}$$

Furthermore, the result (??) stands that

$$0 \leq \phi_{v,f} \leq 1$$

Where right equality is reached whenever v is connected to a set of measure grid's bit $I \in \mathcal{G}$, with level $l_0 = |I|$ such that $I \subset \mathcal{G}(f)$.

Proof. From the definition of $\psi_{v,f}$ we have

$$\psi_{vf} = \psi_{I,l,f} = \frac{\|\mathcal{P}_I^l\|_{d,e}}{\|\mathcal{P}_{\mathcal{G}(f)}\|_{d,e}}$$

Thus using $\mathcal{P}_* = \mathcal{P}_{\mathcal{G}(f),|\mathcal{G}(f)|}$ one have

$$\psi_{vf} = \frac{\langle \mathcal{P}_I^l, \mathcal{P}_* \rangle_{d,e} + \langle \mathcal{P}_I^l, \bar{\mathcal{P}}_* \rangle_{d,e}}{\|\mathcal{P}_f\|_{d'}}$$

Yet $\langle \mathcal{P}_I^l, \bar{\mathcal{P}}_* \rangle_{d,e} = 0$, thus

$$\psi_{vf} = \frac{\langle \mathcal{P}_I^l, \mathcal{P}_* \rangle_{d,e}}{\|\mathcal{P}_f\|_{d'}} = \frac{\langle \mathcal{P}_I^l, \mathcal{P}_* \rangle_{d,e} \times \|\mathcal{P}_f\|_{d'}}{\|\mathcal{P}_f\|_{d'}} = \mu_{I,l,f}$$

Finally the result (??) directly comes with the definition.

vertex's score process

Let $s_{v,f}[N, T, p, q]$ be the score process of v with respect to f , for some $N, T, p, q \in \mathbb{N}^4$. The result (??) stands that

$$\mathbb{E}[s_{v,f}[N, T, p, q]] = N + T \times (q_s \times (p + q) - p) = N + T \times (\phi_{I,l,f} \times (p + q) - p)$$

Proof. From the definition of the score process we have

$$s_{v,f}[N, T, p, q] = N + \sum_{t=1}^T s_{v,p,q,f,t}$$

with $\{s_{v,p,q,f,t}\}_{t=1,\dots,T}$ a sequence of i.i.d such that

$$s_{v,p,q,t,f} = \begin{cases} q, & \text{with probability } q_s = \frac{q_r}{q_r+q_p} \\ -p, & \text{with probability } 1 - q_s \end{cases}$$

Thus we can write

$$\begin{aligned} \mathbb{E}[s_{v,f}[N, T, p, q]] &= N + \sum_{t=1}^T \mathbb{E}[s_{v,p,q,t,f}] \\ &= N + \sum_{t=1}^T q \times q_s - p \times (1 - q_s) \\ &= N + \sum_{t=1}^T q_s \times (p + q) - p \\ &= N + T \times (q_s \times (p + q) - p) \end{aligned}$$

Yet $q_s = \frac{q_r}{q_r+q_p}$, with $q_r = \|\mathcal{P}_v\|_{d,e}$ and $q_p = \|\mathcal{P}_v\|_{d,\bar{e}}$, thus

$$q_s = \frac{\|\mathcal{P}_v\|_{d,e}}{\|\mathcal{P}_v\|_{d,e} + \|\mathcal{P}_v\|_{d,\bar{e}}} = \frac{\|\mathcal{P}_v\|_{d,e}}{\|\mathcal{P}_v\|_d}$$

Which is the definition of $\phi_{v,f}$ that is equal to $\phi_{I,l,f}$ by definition.

Furthermore the result (??) stands that,

$$\text{Var}[s_{v,p,q,t,f}] = (p + q)^2 \times \phi_{I,l,f} \times (1 - \phi_{I,l,f})$$

Proof. Using the result of previous proof we first compute $\mathbb{E}[s_{v,p,q,t,f}^2]$

$$\begin{aligned} \mathbb{E}[s_{v,p,q,t,f}^2] &= q^2 \times q_s + p^2 \times (1 - q_s) \\ &= q^2 \times \phi_{v,f} + p^2 \times (1 - \phi_{v,f}) \\ &= \phi_{v,f} \times (q + p) \times (q - p) + p^2 \end{aligned}$$

Furthermore we have seen previously that $\mathbb{E}[s_{v,p,q,t,f}] = \phi_{v,f} \times (p + q) - p$, thus

$$\mathbb{E}[s_{v,p,q,t,f}]^2 = \phi_{v,f} \times (p + q) \times (\phi_{v,f} \times (p + q) - 2 \times p) + p^2$$

Finally

$$\text{Var}[s_{v,p,q,t,f}] = \mathbb{E}[s_{v,p,q,t,f}^2] - \mathbb{E}[s_{v,p,q,t,f}]^2 = (q + p)^2 \times \phi_{v,f} \times (1 - \phi_{v,f})$$

Which gives the expected result since $\phi_{v,f} = \phi_{I,l,f}$ by definition.

B Implementation

The code that has been used to obtain results of simulations can be found on github at <https://github.com/pierreGouedard/deyep> under the branch *publi_1*. The code is exclusively written in python, is compatible with interpreter python2.7 and python3 and requires python modules numpy and scipy. Code for simulation can be found under

- tests/signal_plus_noise_1.py
- tests/signal_plus_noise_2.py
- tests/signal_plus_noise_3.py
- tests/sparse_identification.py
- tests/sparse_identification_2.py

where the list below are relative to the root directory of the project. The code in the branch *publi_1* has not changed since the submission of this paper, however, the code in other branch, notably master, may have been optimised, augmented or refactored.

References

- [1] Christopher M Bishop. *Pattern recognition and machine learning*. Information science and statistics. Springer, New York, NY, 2006. Softcover published in 2016.
- [2] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 5 2015.
- [3] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, CA, 1984.
- [4] Andy Liaw and Matthew Wiener. Classification and Regression by randomForest. *R News*, 2(3):18–22, 2002.
- [5] Ji Zhu, Hui Zou, Saharon Rosset, and Trevor Hastie. Multi-class adaboost, 2009.
- [6] Martin Vetterli, Jelena Kovačević, and Vivek K Goyal. *Foundations of Signal Processing*. Cambridge University Press, 2014.
- [7] T. Tao. *Topics in Random Matrix Theory, ser. Graduate Studies in Mathematics*. Providence, Rhode Island: American Mathematical Society, 2012.