Exploring Glitchland I

A Few Answers and Many Questions

Innocenzo M. Pinto

University of Sannio at Benevento and INFN

Institute of Cosmic Ray Research, University of Tokyo, May 23, 2014

Outline

Glitches, Vetos, and Dreams

A Toy-Model for (Linear) Glitches

Glitch Manifold Dimensions : PCA

Clustering

Natural Bases and Sparsity

Dictionary Learning or the Best Model of a Cat

Is it Hopeless ?

Glitches ?

Transient disturbances of environmental (exogenous, e.g. lightning) and/or instrumental (endogenus, e.g. laser) origin;

Appear *ubiquitously* in the data gathered by interferometric GW detectors, with a *wide range* of energies;

Idiosyncratic signals: exhibit a *wide variety of shapes*; still mostly visually *similar* and *erratically recurrent*.

Glitch *rate* roughly inversely proportional to glitch *strenght*;

Important impact on instrument's noise (*non-stationarity, heavy-tails*) -> glitches *spoil* naïve (Gaussian) detectors.

… word comes from Yiddish term גליטש

Virgo-LIGO Glitchology

A. DiCredico et al., "Gravitational Wave Burst Vetoes in the LIGO S2 and S3 Data Analysis," Class. Quantum Grav. 22 (2005) S1051.

L. Blackburn et al., "Monitoring Noise Transients During the Fifth LIGO Science Run," Class. Quantum Grav. 25 (2008) 184004.

N. Christensen et al., "LIGO S6 Detector Characterization Studies," Class. Quantum Grav. 27 (2010) 194010.

F. Acernese et al., "Noise Studies During the First Virgo Science Run and After," Class. Quantum Grav. 25 (2008) 184003.

M. Del Prete et al., "Characterization of a Subset of Large Amplitude Noise Events in VIRGO Science Run 1 (VSR1)," Class. Quantum Grav. 26 (2009) 204022 .

J. Aasi et al., "The Characterization of Virgo Data and its Impact on Gravitational Wave searches," Class. Quantum Grav. 29 (2012) 155002.]

Glitch Hunting Goals

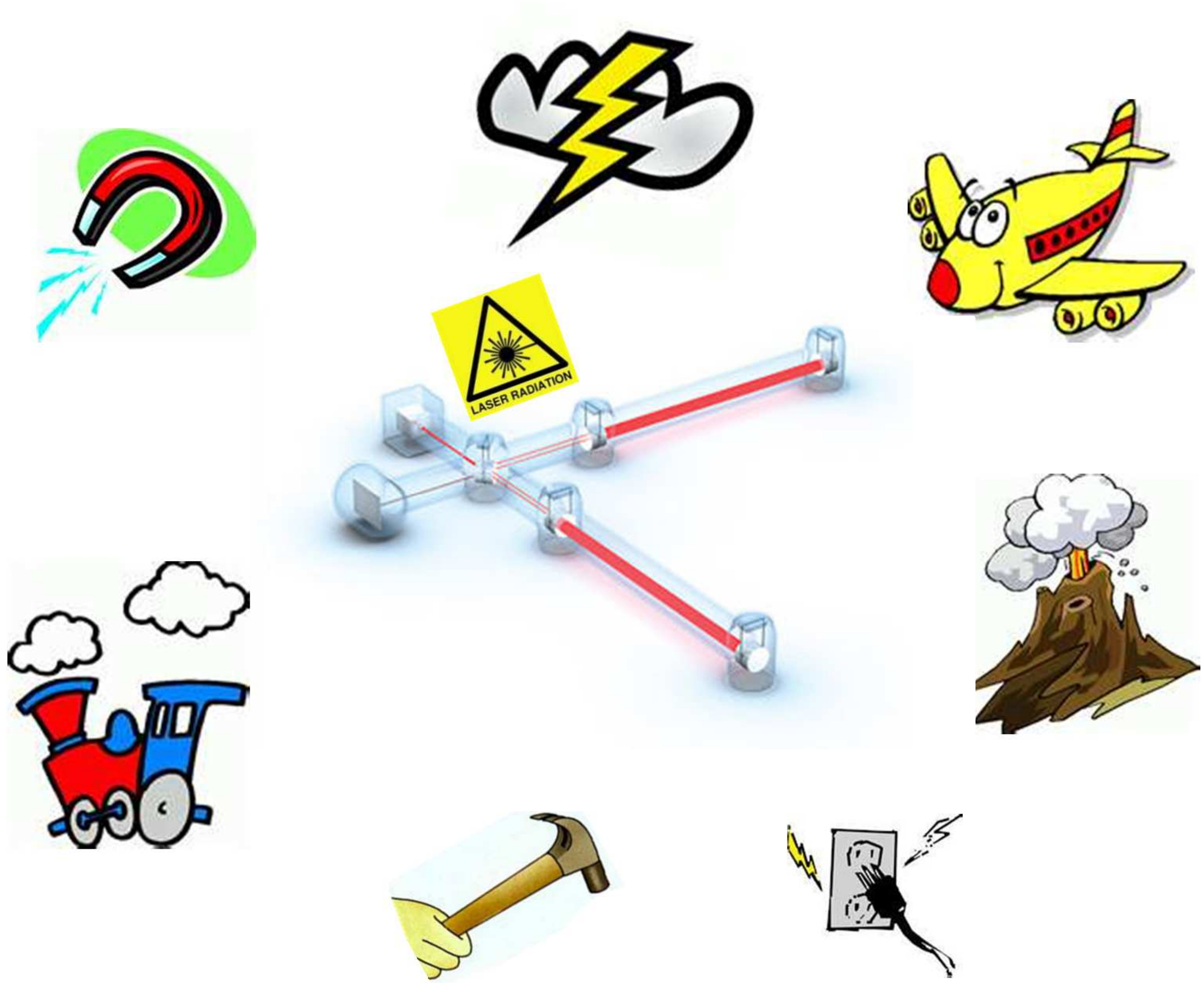
Tracing out the *origin* of typical glitches (glitch classes), and tweaking the machine design so as to *suppress* or *mitigate* them;

Identifying surviving glitches in the GW channel, capitalizing on information from the instrumental / monitoring channels, and tagging/vetoing the data appropriately ;

Subtracting identified glitches from the GW channel (data cleaning);

Characterizing statistically the resulting non Gaussian noise, and devising robust detection algorithm (noise modeling).

Glitch Origin



GW Detectors are MIMO

MIMO = Multiple Input, Multiple Output

disturbances enter IFO
at some "susceptible"
instrument entry points

…

"GW" channel

y_0

…

"AUX" channels (do *not* couple to GW)

$\{y_i | i = 1,2,\ldots,K\}$

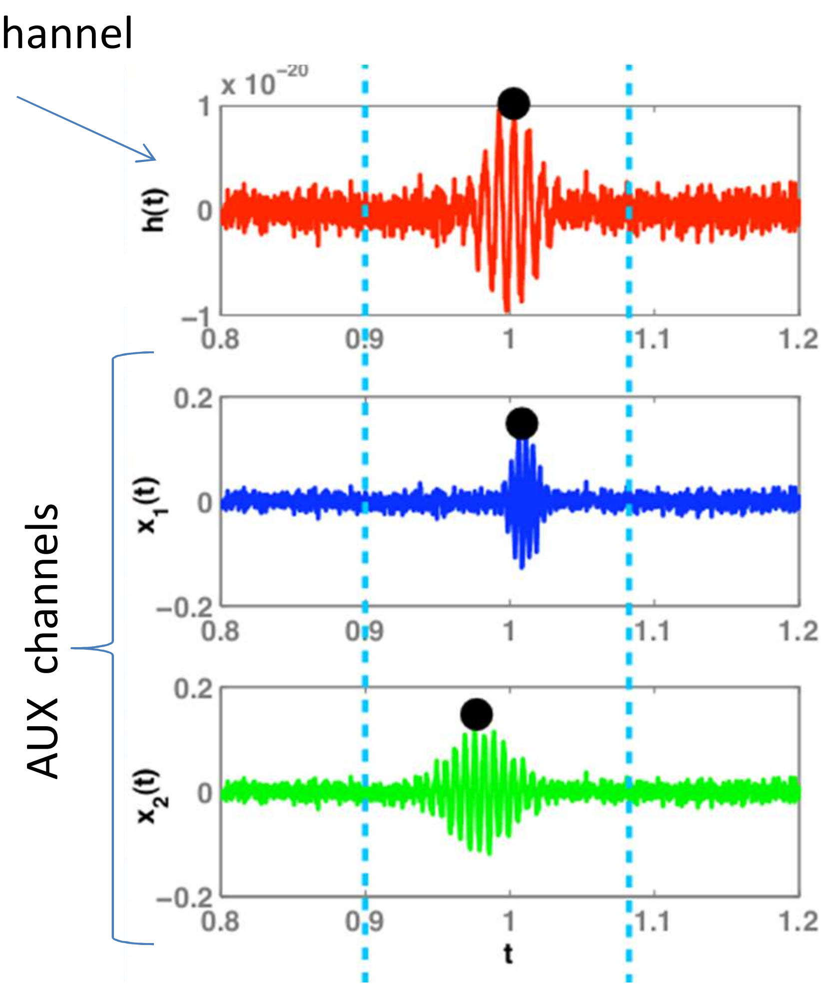
the "input ports" are usually non-accessible

Vetoing

Check whether a trigger in the GW channel is ***time – coincident*** with a trigger in one (or more) AUX channel(s), *at a given significance level.*

Characterize preliminarily *accidental coincidence statistics* (time slide experiments).

[K.C. Cannon LIGO P070085]



time window
(markers identify peak-times)

Popular Wisdom

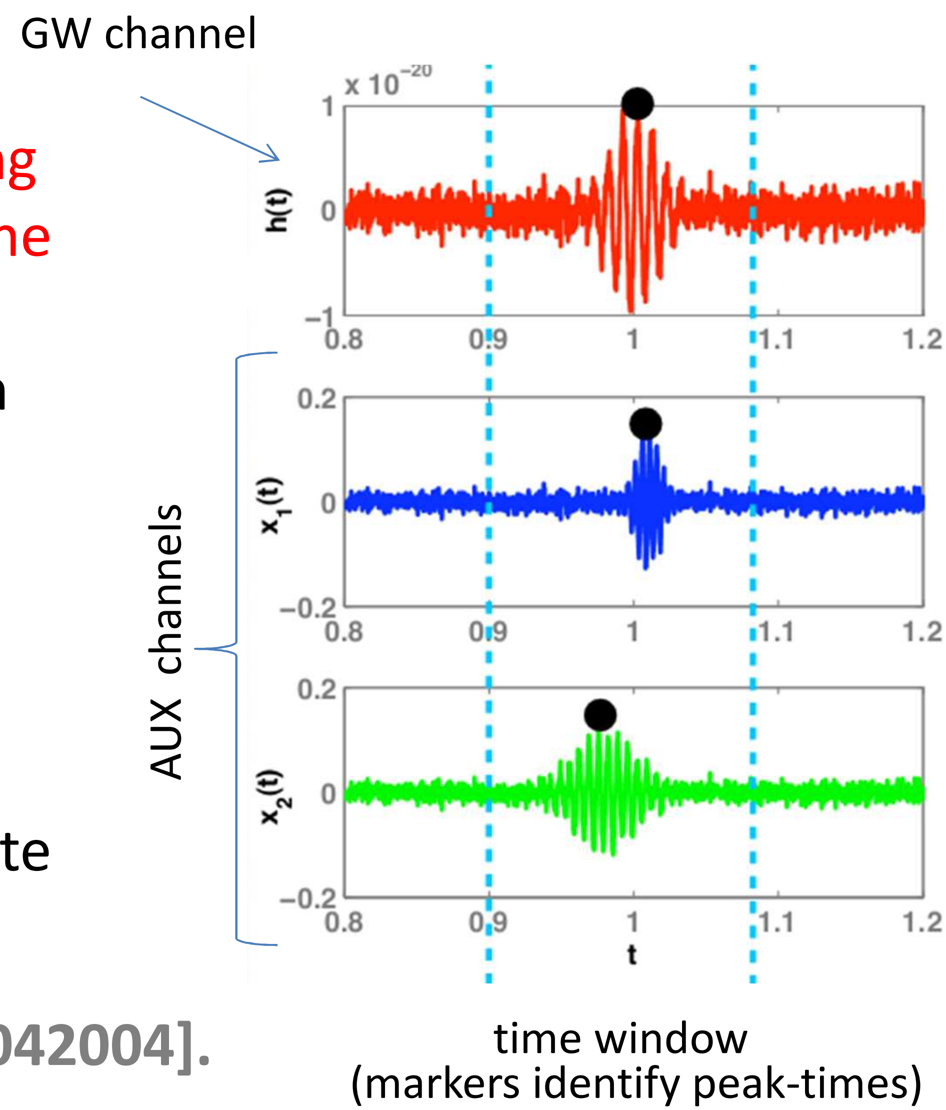
We should refrain from
throwing the baby out with the bath water…

Ajith's Improved Veto Strategy

Use knowledge of the coupling (*transfer functions*) between the AUX channels and the GW one to check *consistency* between transients occurring in the GW and instrumental channels .

Better veto efficiency, lower rate of accidental vetos.

[P. Ajith et al., PRD 76 (2007) 042004].



"Dream" Strategy

-*Combine* AUX channel data $\{y_k | k = 1,2,..,K\}$ so as to obtain a *basis (orthogonal* and *complete* set), $\{z_h | h = 1,2,..,M\}$;

Non redundant : *M* is *minimal*
Orthogonal : the same trigger may not appear in two channels
Complete : $z_k = 0 \ \forall k \leftrightarrow$ *instrument is quiet* (no glitches)

-Derive the transfer functions $\{H_{0h} | h = 1,2,..,M\}$ connecting the above z_h to the GW channel y_0 (assume linearity);

-Clean-up (!) the GW channel easily (frequency domain)

$$clean[Y_0] = Y_0(f) - \sum_{h=1}^{M} H_{0h}(f)\, z_h(f)$$

… to be recognized as a (tough !) ICA problem…

[A. Hyvärinen et al., *Independent Component Analysis,* Wiley (2001)]

A Toy Glitch Model

Focus on the GW channel (analysis can be rephrased for AUX channels similarly);

Assume a *single* exogenous/endogenous disturbance entering IFO via *several* “susceptible” entry points connected to the GW channel by (unknown) transfer functions ${H_i}^{(0)}$;

Assume disturbance bandwidth as *large* compared to spectral support of the ${H_i}^{(0)}$.

coupling amplitude of disturbance to entry point ♯ i

GW signal

$$Y_0(f) = Y_{GW}(f) + \sum_i A_i {H_i}^{(0)}(f) \exp(\imath\theta_i)$$

Linear combination of transfer functions (impulse responses, in time domain) connecting the disturbance-susceptible IFO entry points to the GW channel,

coupling phase $\theta_i = \omega\tau_i$, τ_i = disturbance propagation delay to entry point ♯ i

A Toy Glitch Model, contd.

Allow for *several* exogenous/endogenous disturbances occurring *at random times, with random strenghts*;

Switch to time domain

coupling amplitude of disturbance ♯ m to entry point ♯ i (random)

$$y_0(t) = y_{GW}(t) + \sum_m \left[\sum_i A_i^{(m)} h_i^{(0)}\left(t - \tau_i^{(m)}\right) \right]$$

Propagation delay of disturbance ♯ m to entry point ♯ i

Sum over external disturbances

Impulse response of "channel" connecting entry point ♯ i to GW channel

Linear combination of impulse responses of (linear, time invariant) "channels" connecting the disturbance-susceptible IFO entry points to the GW channel,

A Toy Glitch Model, contd.

According to the above toy model, glitches are linear superpositions involving *always the same* (but basically *unknown*) *wave-Forms, i.e.,* the transfer functions $h_i^{(0)}(t)$, that appear in the data combined *with random amplitudes and delays*.

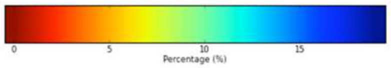
If the $h_i^{(0)}(t)$ were known, glitches *could be effectively subtracted* from the data in (almost) real time. **Questions** :

- **In what sense do the $\boldsymbol{h_i^{(0)}(t)}$ form a "natural" *dictionary* ?**
- **How many are there / how many do we need ?**
- **Is there a way to retrieve the $\boldsymbol{h_i^{(0)}(t)}$ from glitchy datasets ?**

Warnings:

The above toy model assumes *linearity*. A nonlinear generalization (based on Volterra-Wiener series) is possible. Weak *time-invariance* is also assumed…

Non-Linear Glitches

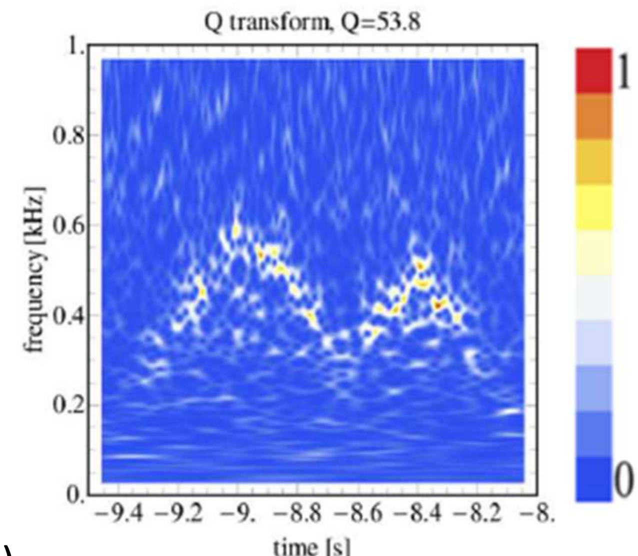


"Fast" channels

"Slow" channels

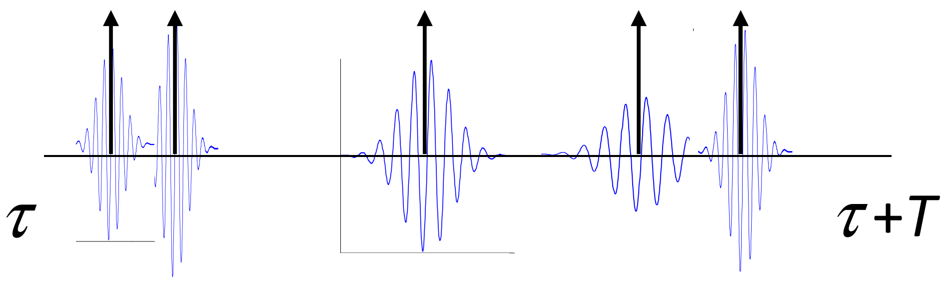
Channel Name	H1:ASC-ETMX_P	H1:ASC-ETMX_Y	H1:ASC-ETMY_P	H1:ASC-ETMY_Y	H1:ASC-ITMX_P	H1:ASC-ITMX_Y	H1:ASC-ITMY_P	H1:ASC-ITMY_Y	H1:LSC-MICH_CTRL	H1:LSC-PRC_CTRL
H1:LINEAR	6.92%	10.10%	6.31%	11.27%	11.33%	10.29%	8.70%	11.08%	11.27%	15.74%
H1:ASC-QPDX_P	6.74%	10.10%	6.67%	7.41%	8.57%	7.72%	8.14%	9.06%	13.60%	14.21%
H1:ASC-QPDX_Y	10.23%	10.23%	2.08%	7.17%	8.82%	9.06%	4.04%	6.06%	13.84%	15.19%
H1:ASC-QPDY_P	9.55%	8.94%	3.98%	6.92%	9.74%	8.45%	5.21%	6.80%	12.19%	14.45%
H1:ASC-QPDY_Y	6.80%	5.57%	6.67%	5.94%	7.66%	8.14%	6.00%	7.35%	10.84%	8.82%
H1:ASC-WFS1_QP	5.63%	5.08%	10.04%	6.12%	9.12%	8.51%	9.80%	9.00%	10.65%	12.12%
H1:ASC-WFS1_QY	8.63%	14.21%	10.10%	9.86%	7.96%	8.88%	7.59%	8.21%	5.08%	8.45%
H1:ASC-WFS2_IP	9.25%	8.76%	8.02%	9.68%	9.49%	9.19%	12.25%	10.35%	10.29%	13.41%
H1:ASC-WFS2_IY	6.43%	10.35%	7.72%	9.49%	7.59%	8.02%	8.39%	10.59%	8.08%	10.04%
H1:ASC-WFS2_QP	9.68%	11.70%	10.96%	14.45%	8.70%	10.84%	9.06%	12.19%	14.21%	11.14%
H1:ASC-WFS2_QY	8.08%	9.55%	7.35%	9.43%	9.25%	9.31%	8.70%	10.41%	5.57%	13.84%
H1:ASC-WFS3_IP	12.74%	7.78%	7.35%	9.37%	12.86%	5.08%	7.23%	8.45%	10.65%	15.74%
H1:ASC-WFS3_IY	10.35%	13.35%	8.27%	14.15%	12.92%	13.90%	10.84%	12.98%	11.88%	15.86%
H1:ASC-WFS4_IP	10.96%	9.49%	9.86%	12.55%	12.86%	8.88%	8.33%	10.10%	12.68%	14.82%
H1:ASC-WFS4_IY	10.59%	9.19%	8.39%	11.27%	12.25%	5.14%	9.98%	8.76%	12.62%	13.90%

Bi-linear glitch zoo table (LIGO)



A GEO600 "arch" glitch
TF plot (courtesy M. Was)

Glitchy Noise Statistics



$$n_g(t \in \Theta) = \sum_{k=1}^{K_\Theta(\Theta)} \psi(t - t_\Theta{}^{(k)}; \tilde{a}_\Theta{}^{(k)})$$

$$prob[K_\Theta[T] = K] = \frac{(N_\Theta)^K e^{-N_\Theta}}{K!}$$

$N_\Theta = \lambda_\Theta T =$ expected number of glitches in Θ

local firing-rate (may fluctuate adiabatically , Cox process)

$\Theta = [\tau, \tau + T]$, the analysis window

$\left\{ t_\Theta^{(k)} \mid k = 1, 2, \ldots, K_\Theta[T] \right\}$, a set of $i.i.d.$ firing-times, $t_\Theta^{(k)} \in [\tau, \tau + T]$

$K_\Theta[T]$, the random (Poisson–distributed) number of glitches in Θ

$\psi(t; \vec{a})$, *a time-frequency atom* with time - barycenter at $t = 0$, and shape parameters $\vec{a} = \{a_1, a_2, \ldots, a_N\}$

$\left\{ \vec{a}_\Theta^{(k)} \mid k = 1, 2, \ldots, K_\Theta[T] \right\}$, a set of $i.i.d.$ shape params w. *known* priors

[M. Principe and I. Pinto, Class. Quantum Grav. **25** (2008) 075013]

Modeling a Glitchy Noise Component, contd.

- The proposed glitch noise model belongs to Middleton's class of generalized shot noises. [D Middleton IEEE T-EMC-21 (1979) 209]

- Characteristic functions for Middleton noise can be derived in closed form *up to any order* [D. Middleton, J. Appl. Phys. 22 (1951) 1143] – depend on *coarse* statistical properties of the glitch atoms;

- Locally optimum detectors (LOD) for such noise can be implemented easily [M. Principe and I.M. Pinto, CQG, 26 (2009) 204001, M. Principe and I.M. Pinto, LIGO-P1000134 (2010)]

A Toy Glitch Model, contd.

According to the above toy model, glitches are linear superpositions involving *always the same* (but basically *unknown*) *wave-Forms, i.e.,* the transfer functions $h_i^{(0)}(t)$, that appear in the data combined *with random amplitudes and delays*.

If the $h_i^{(0)}(t)$ were known, glitches *could be effectively subtracted* from the data in (almost) real time. **Questions** :

- **In what sense do the $\boldsymbol{h_i^{(0)}(t)}$ form a "natural" *dictionary* ?**
- **How many are there / how many do we need ?**
- **Is there a way to retrieve the $\boldsymbol{h_i^{(0)}(t)}$ from glitchy datasets ?**

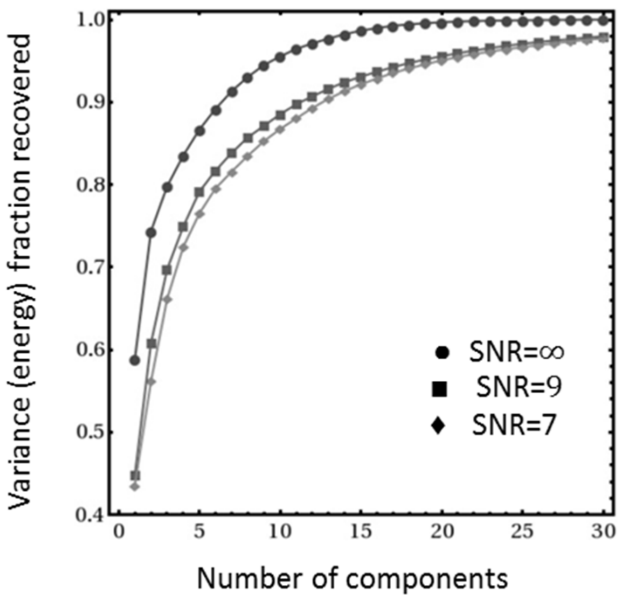
Warnings:

The above toy model assumes *linearity*. A nonlinear generalization (based on Volterra-Wiener series) is possible. Weak *time-invariance* is also assumed...

Answers: Glitch Manifold Dimension

Principal Component Analysis (PCA) may answer the question ***"how many basis elements do we need"*.**
[I. T Jolliffe, "Principal Component Analysis," Springer, 2002**]**

Independent PCA implementations on different (LIGO) glitch datasets indicate that the number of needed basis elements is $\sim \mathbf{20}$ to account for $\sim \mathbf{90\%}$ of the glitch energies, for all (typical) glitches in the datasets.



[I.M. Pinto, L. Troiano et al., Int. J. Mod. Phys. C24 (2013) 1350084;
M. Cavaglia and D. Trifiro', LIGO Document G1300368-v1]

PCA in a Nutshell

Glitches are transient waveforms with time - limited support T, sampled at some frequency f_s . Any glitch g is thus represented as *a point* $\boldsymbol{g}$ (vector) in Euclidean space R^N , $N = f_s T$.

Let $\mathfrak{D}_g = \{g_i | i = 1,2, \dots , G\}$ our collection (dataset) of glitches. Let $\boldsymbol{D}_g$ the $N \times G$ matrix whose columns are the $\boldsymbol{g_i}$ vectors ;

1) "Standardize" matrix $\boldsymbol{D}_g$, subtracting from each column its average:

$$\tilde{g}_{ij} = g_{ij} - (1/G) \sum_j g_{ij}$$

2) Compute covariance matrix $\boldsymbol{\Sigma}$ among standardized column vectors :

$$\Sigma_{hk} = (\tilde{\boldsymbol{g}}_h, \tilde{\boldsymbol{g}}_k)$$

3) Diagonalize $\boldsymbol{\Sigma}$, and enumerate the eigenvalues in order of decreasing (absolute) value. *The corresponding ordered eigenvectors yield the PCA basis* $\{\boldsymbol{\pi_i} | i = 1,2, \dots , G\}$.

PCA in a Nutshell, contd.

The magnitude of the PCA eigenvalues drops steeply at a certain order *N*.*

Correspondingly, the energy of *all* glitches is fully recovered by using only the first *N** vectors from the PCA basis (addition of further terms does *not* improve representation accuracy to any sensible extent).

N* thus represents a sort of effective dimension of the manifold spanned by the glitch data set.

PCA appears as a *compressive* coding where the vectors $g_i \in R^N$ get represented by the vectors $\alpha_h \in R^{N^*}$, with $\alpha_{hk} = (g_h, \pi_k)$.

The (truncated) PCA basis $\{\pi_i | i = 1, 2, \ldots, N^*\}$ is *not unique.*

Is there a *best / natural* minimal $(\sim N^*)$ - size "basis" ?

Hints: Clustering

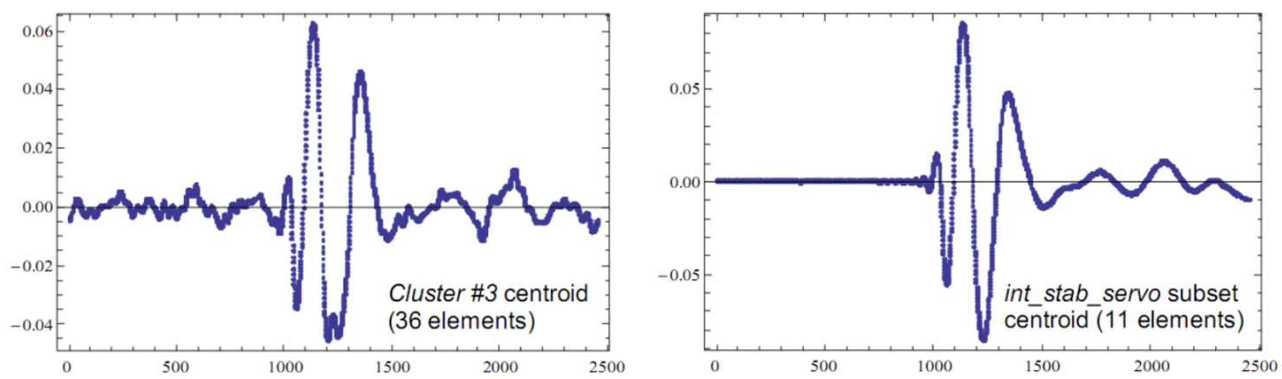
Once a glitch dataset $\{\boldsymbol{g_i}\}$ has been represented using a non-redundant (e.g., PCA) "basis" *clustering algorithms* can be used to identify families of *similar glitches.*

Several glitch clustering algorithms have been proposed/used, based, e.g., on *proximity* in coarse-feature space [S. Mukherjee et al., CQG 24 (2007) S701], longest-common-subsequences [S. Mukherjee et al., J. Phys. Conf. Ser. 243 (2010) 012006], Kohonen self-organizing maps [S. Rampone et al., Int. J. Mod. Phys. C24 (2013) 1350085], and GANN [S.M. Kim et al., LIGO-G1201110]

The clustering goodness can be gauged using different metrics (e.g., the Davies-Bouldin cluster separation measure [IEEE T-PAMI, 1 (1979) 224])

Clustering, contd.

In many cases, the cluster-centroid waveforms are found to be almost *coincident with the centroids of glitch datasets whose origin is known*, resulting from injection of external noise transients at *specific* instrument entry points.



[I.M. Pinto, L. Troiano et al., Int. J. Mod. Phys. C24 (2013) 1350084

This is *suggestive* that *glitch cluster centroids may correspond* to the *pure* “canonical” responses $h_i^{(0)}(t)$.

Answers: Best/Natural "Basis"

- One may argue that the key requirement of a best / natural N* - size "basis" would be that of allowing to represent each glitch in the data set $\{\boldsymbol{g_i}\}$ using *a fewest significant (nonzero) coefficients* .
- Such a requirement can be stated formally as follows

$$min\|\boldsymbol{D}_\alpha\|_0 \text{ subject to } \left\|\boldsymbol{D}_g - \boldsymbol{D}_\alpha \cdot \boldsymbol{D}_\pi\right\|_2 < \varepsilon$$

and is technically a *maximal sparsity requirement* .

- The above L_0 (constrained) optimization problem is *NP-hard.* But under broad assumptions *one may use the L_1 norm* to apply convex optimization [D.L. Donoho et al., PNAS 100 (2003) 2197]
- Usually, the basis is *given*, and the above is known as a "pur-suit" algotihm, of which several flavors exist [ibid]. In our case, we would like to *find out* the "best" or natural "basis" *as part of our sparsest representation problem* .

Dictionary Learning

The sought natural "basis" needs not to be technically a *basis* (it may be overcomplete, and contain linearly dependent subsets). Following Gabor, we call it *a dictionary.*

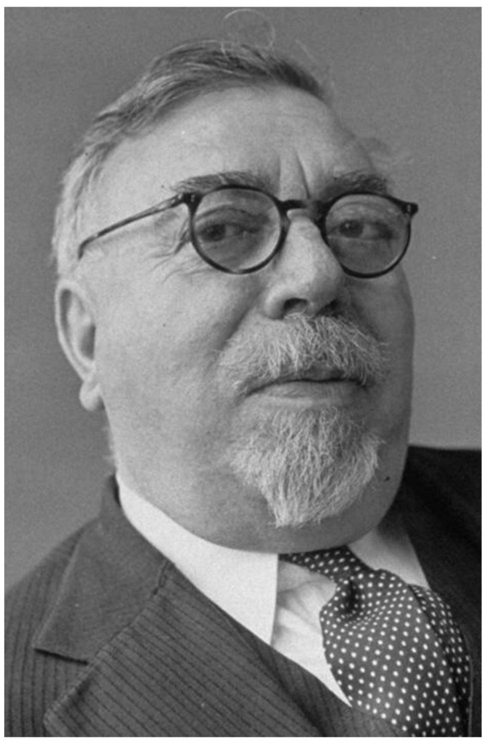
Goal: Given the dataset $\mathfrak{D}_{\mathbf{g}}$ of glitch waveforms, *construct* a ***data-adapted*** dictionary $\mathfrak{D}_{\mathbf{w}}$ under a ***sparsity constraint***.

Can be implemented **efficiently** as an iterative algorithm that switches between **sparse-coding** (using a given dictionary), and **adaptive dictionary updating** (using clustering) [Kreutz et al., Neural Comp. 15 (2003) 349; Aharon et al., IEEE T-SP 54 (2006) 4311].

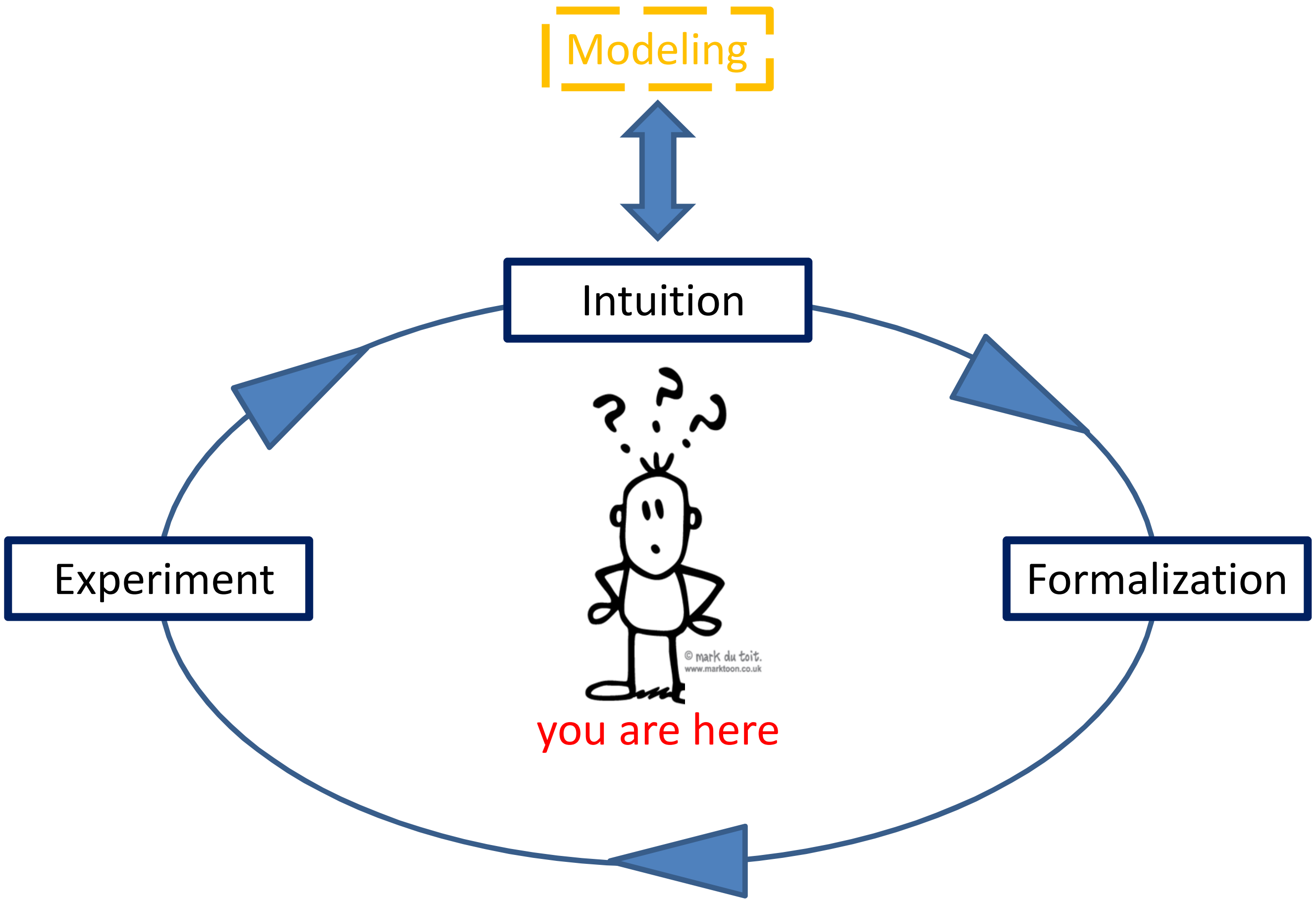
Preliminary results on reduced LIGO datasets indicate a substantial representation compression [Matta and Pinto, 2014, in progress].

Best/Natural "Basis" contd.

- One may further argue that the above sought *dictionary* goes *closest* to the sought $\{\boldsymbol{h_i}^{(\mathbf{0})}(\boldsymbol{t})\}$ set.
  In the words of Norbert Wiener, *"the best model for a cat is another cat, or preferably, the same cat"*



Conclusions



Hopeless ?

BSS (Blind Source Separation) - a special kind of ICA, [S. Choi et al., Neural Info. Lett, 7 (2005) 1] also based on sparse dictionary learning [M. Zibulewski, Neur. Comp. 13 (2001) 863] is now used routinely for this:

 =  +  + ...

(courtesy P. Bofill , UPC)