PART 1

Theory

1

Non-linear Signal Processing

We will summarize the evolution of signal processing by period, according to the technology then available. The periods range from 1910 until the present day, where technology evolved from vacuum tubes to FPGA, DSP and now QCA. The analytical evolution ranges from the study of transfer functions to inverse linear theory. From 1910 until 1940, Fourier [FOU 27], Laplace [LAP 14], Bode and Nyquist [NYQ 24] were the main authors. From 1940 until 1960, we will cite Gabor [GAB 46], Shannon [SHA 48] and Wiener [WIE 49]. From 1960 until 1980, Oppenheim [OPP 99], Kailath, Kalmann [KAL 60], Slepian [SLE 74] and Cooley and Tukey [COO 65] often appear in published articles. From 1980 until 2000, we will cite Kennedy [KEN 11] and Eberhart [EBE 95].

From 2000 until the present, technologically, the evolution has taken two forms, that of photonic crystals and that of quantum dot cells. [ANG 07] demonstrated the possibility of implanting quantum gates with photonic crystal wave guides. The photons confined in the photonic crystals and the wave guides formed by linear chains doped by default with atoms or quantum dots can generate strong non-linear interactions between the photons, thus enabling one or two qubit of quantum gates to be implemented. The simplicity of the gate-switching mechanism, the experimental feasibility of manufacturing two-dimensional photonic crystal devices and integrating such devices with optoelectronic components offer promising new possibilities for processing quantum and optic information networks.

[HEL 97] studied quantum dot cells (QCP) and demonstrated their convenience for non-linear signal processing. [LEN 97] provided the construction of a computing device with quantum dots. The base block of a QCa is a nanometric cell that contains five quantum dots. Each cell contains two electrons, the two electrons are on opposing dots in the cell due to Coulomb repulsion. The electrons can change place from one dot to another by tunneling, so the electrons can have two different states that correspond to the bits' Boolean values. The quantum point cells have three advantages: a very fast execution speed and infinitely reduced dimensions and energy consumption rates. The theoretical developments are sufficiently advanced. However, manufacturing them poses enormous difficulties. This is due to their quantum aspects, which should be taken into account. The only achievement to date is the quantum dot cellular automaton (QCA) [WHI 07].

Period	*Physics*	*Analytics*	*Names*
1910-1940	Empty tubes, localized circuits	Impulse response, Transfer function, transformation methods	Fourier, Laplace, Bode, Nyquist
1940-1960	Microwave circuits	Statistical concepts (correlation, adapted filters, information theory)	Gabor, Shannon, Wiener
1960-1980	Integrated computer, circuits, optical technology	Digital filters, spectrum estimation, fast Fourier transform, inverse linear theory.	Kailath, Oppenheim, Slepian, Tukey
1980-2000	FPGA, DSP, DSP in particle swarm	Ultra-fast signal processing	Kennedy, Eberhart
2000-2015	QCP, QCA	Quantum signal processing	Helsingius, Whitney, Athas

Table 1.1. *Devices, topics and main discoverers*

The simple QCAs at first reveal many problems in creating actual circuits. Thermal noise can alter the state of different cells, its effect increases with the cell's dimensions and the temperature. Other problems limit the maximal dimensions of individual circuits. Large circuits can themselves commute into undesirable states. However, the recent discovery of adiabatic pipelines seem to compensate this problem [ATH 94]. Large circuits can be divided into many small circuits. The dimension of each small circuit is matched such that the thermal noise does not disrupt the calculation. Another advantage of adiabatic pipelines lies in the manufacturing of delay lines and finite state machines, this being an essential advantage for most algorithms.

The evolution of signal processing to the non-linear domain is motivated by technological advances in CMOS [HUN 10], new perspectives in micro-electronics [LUR 04] and new logic gates [MEY 07] as well as nano-networks [AKY 11]. The main works to tackle signal processing are the following: Kurth describes other approaches to identify non-linearities in time series in geophysics and astrophysics. His work consists of a data pretreatment and tests on signal stationarity and artifacts, for which he provides a robust non-linear method for reducing noise [KUR 94]. [KAT 06] discuss the problems and methodologies for systems and signals, with particular attention on modeling, identifying and processing signals. It reveals some common precepts between technologies. In his work, he summarizes the new directions and predictions on systems and signals. In his published technical note, Kijewski reports the increasing use of frequency conversion time for analyzing and interpreting non-stationary and non-linear signals in a wide range of scientific and technical techniques. He focuses his attention on continuous wavelet transform and mode decomposition in tandem with Hilbert transform. His study evaluates the performance of two approaches in analyzing a large variety of classic non-linear signals. The fundamental difference between the two approaches appears in the instantaneous frequency obtained using Hilbert transform, which characterizes sub- and super-cyclic nonlinearities simultaneously, while the wavelet based on the instantaneous frequency captures the super-cyclic nonlinearities with an additional measure of the bandwidth characterizing the sub-cyclic nonlinearities [KIJ 07].

Among the works applied to the nonlinear analysis of medical signals using data mining, we will cite [BOG 10]. Dougherty’s work on determining coefficients to decide on the signal’s non-linear character deserves to be mentioned [DOU 00]. Diversity has been handled by Modarres using a Hammerstein filter [MOD 09]. Perez has detailed all the characteristics of Gaussian processes in a review article [PER 13]. As for new developments, the future predictions expected in non-linear signal processing are summarized by Gao [GAO 12].

Finally, we will cite a series of books that tackle non-linear signal processing with a variety of methods [ARC 05, MOO 00, STO 05, PRA 08, ROC 04].

1.1. Distributions

When processing the signal, using distributions is fundamental. However, this usage is not clearly defined: we must work with functions or distributions. Of course this depends on the level of abstraction we are working with. Considering what is known as Dirac comb or pseudo function, the best-adapted formalism is that of distributions. Distribution theory in the mathematical sense of the term has been developed by Schwartz [SCH 51]. The objective has therefore been to generalize the notion of function, in order to give a correct mathematical meaning sense to the objects handled manipulated by physicists, additionally retaining the possibility of carrying out operations such as derivations, convolutions or Fourier or Laplace transforms. This generalization of the notion of function has been pursued in various directions, in particular the notion of the hyper-function, thanks to Sato [SAT 59, SAT 60]. We will examine this notion in detail in the context of the evolution of

Fourier transform, which will enable us to identify a significant aspect of non-linear signal processing [SHA 13]. However, it is Gumbel's works that have enabled the distribution aspect, in the sense of the term relating to probability, to be clarified [GUM 35, GUM 50, GUM 53, GUM 54, GUM 60, GUM 61, GUM 67].

A link exists between the probabilities and the distributions. In fact, if p is a positive function defined over $\Re$, such as:

$$\int_{-\infty}^{+\infty} p(x)dx = 1 \qquad [1.1]$$

knowing the probability density enables the expected value of any function f of the random variable X to be calculated.

$$E\left(\phi(X)\right) = \int_{-\infty}^{+\infty} \phi(x)p(x)dx \qquad [1.2]$$

we can say that p defines a regular distribution since p is therefore *a fortiori* summable in $L^1_{loc}(\Re)$, and we have:

$$E(\phi(x)) = \left\langle p, \phi \right\rangle \qquad [1.3]$$

Holmes, in his technical report on the role of group theory in signal processing's mathematical foundations, shows the usefulness and benefit of group theory, through the Plancherel formula. The essential idea is to use special unit transformation for compression and decorrelation. The suggestions given show that nonabelian group filters can improve the standard methods of discreet Fourier transform and fast Fourier transform (DFT and FFT) without increasing the calculation's complexity [HOL 87].

1.2. Variance

Variance is a measure that serves to characterize a distribution's dispersion. It indicates how the statistical series or the random variable is dispersed around its mean or its expected value. A zero variance shows that all the values are identical. A small variance is a sign that the values are close to one another, while a high variance is a sign that they are very distant.

1.3. Covariance

Covariance is an extension of variance. Correlation is a normalized form of covariance. This concept is naturally generalized to many variables (random vector) by the covariance matrix.

For the stochastic processes that cover the evolution of a random variable, covariance gives way to the concepts of auto-covariance and autocorrelation, then to the estimation of spectral density for stationary processes.

Space time adaptive processing (STAP) normally requires knowledge of the inverse covariance matrix (ICM) of unwanted signals to detect signals from the desired target. The computational load for generating a reliable inverse covariance matrix prevents an adaptive processing from being implemented in radar systems in real time [DON 05].

1.4. Stationarity

It is often convenient to consider stationarity in a random process, this constitutes a simplification if the process linked to the signal is considered as a stationary process. Suppressing stationary and non-stationary noise in the case of cardiac signals has been developed by [RAH 11].

1.5. Bayes inference

The procedures for statistical inference are applied when the available information is less complete than that usually studied. In this case, the initial information is taken to be a series of probability measures P. With an initial probability measure, an estimation of the corresponding Bayes can be found. The inference procedure recommended, when an initial group of probabilities P is available, is to find a set of estimations corresponding to P. This is called an achievable set of estimates [POT 83]. A practical case is developed in this book on minimal mean square error (MMSE) [LAM 13]. Modern statistical approaches accepted in communication theory use statistical inference in designing and evaluating statistical tests [MET 54]. Farina *et al*. developed an algorithm built around Bayes inference in identifying and tracking radar [FAR 02]. [TAK 05] show that the Jeffreys prior plays an important role in statistical inference.

The statistical inference consists of estimating unknown characteristics of a population from a sample taken from this population. The statistical inference is therefore a set of methods enabling reliable conclusions to be drawn from data from statistical samples.

Methods of statistical inference initially enabled fundamental notions of probability to be deduced, as well as hypothesis tests and confidence intervals as stipulated by [PEA 94], [PEA 01] and [WAL73]. Thereafter, re-sampling techniques came to light with [ULA 04], [EFR 82] and [MIS 57].

Algorithms are sought for processing data or signals using models whose parameters will be adjusted by the statistical use of data. The statistical approach enables robust methods applicable to broad signal categories to be developed with the processing of massive data. On the other hand, modeling complex space–time phenomena is a difficult operation. Many processes can only be partially observed. The latent state's statistical inference (by data assimilation) becomes a problem of paramount importance. Two large categories of space-time models are widespread in the literature [MAN 12]:

– geostatistical models that use a statistical description of covariance functions as a starting point;

– space-time models that enable partial differential equations or their stochastic analogue to be generated from a physical description of the mechanism.

The calculation of the statistical inference can also be formulated as shown by Pereyra *et al*. This has the advantage of resolving the difficulties posed by the Cramer–Rao limit.

The image segmentation obtained by tomography in the case of the brain is based on the differential evolution of the Bayes inference as shown by [WAN 11a]. Another example of the application of Bayes inference is given by [WAN 11b] on the exponential distribution function or risk function in establishing Bayes estimators.

Beyerer gives a generalized Bayesian inference process as shown in Figure 1.1.

Takahashi introduces a universal Bayes which is a Bayesian version of the Martin-Löf test, it establishes a series of theorems linked to Bayes statistical inference in terms of a random sequence [TAK 06].

Another example of inference is given us by mixing models in the statistical segmentation of medical images. A Monte Carlo sampling combined with Markov chains is used [WOO 06].

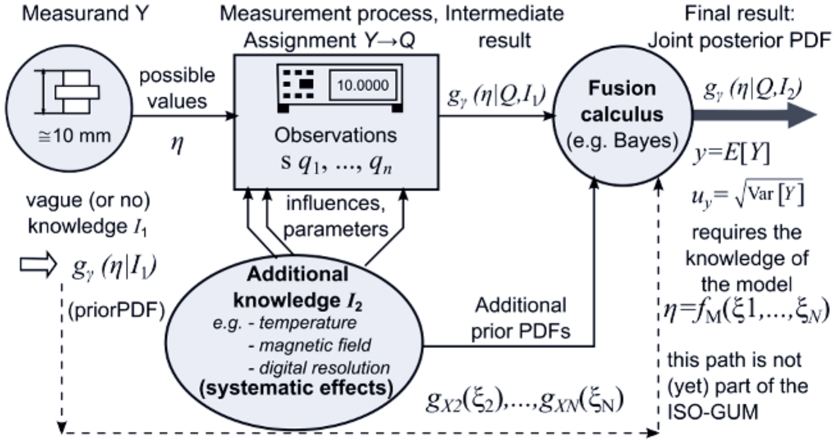

Figure 1.1. *Generalized Bayesian inference process*

Qu and Hu, in a Bayesian framework for generalizing associative functional networks (GAFN), detail a variational Bayes learning algorithm for assimilating later distributions of associative network parameters. This algorithm means statistical inferences can be avoided [QU 07].

An essential application for Bayesian inference is mitigating breakdowns in launching space vehicles. It is cited under the subject "Bayesian anomaly". Contrary to classical inference, the mean estimations differ critically. The difference in the probability of failure in the system's performance is deduced from estimations taken before the test or from real experiments. The probabilities calculated initially differ drastically from the probabilities obtained afterwards [PHI 08].

A second application for Bayesian inference is made on an FPGA architecture. This is made in checking spam in email content. The inference engine uses the logarithmic number system (LNS) to simplify the naive Bayesian calculation. A noise model for the inference engine is developed and the noise limits have been analyzed in order to determine the accuracy of the inference [MAR 08].

Where non-Bayesian signals are concerned, two fairly efficient methods can be used:

– the Cramer–Rao inequality (some call it the Cramer–Rao bound), expresses a lower bound on the variance of an unbiased estimator based on Fisher information.

It states that the reverse of the Fisher information $\Im(\theta)$, of a parameter θ, is a lower bound of the variance of an unbiased estimator of this parameter, written $\hat{\theta}$.

$$\mathrm{var}(\hat{\theta}) \geq \Im(\theta)^{-1} = E\left[\left(\frac{\partial}{\partial\theta}\ln L(X;\theta\right)\right]^{-1} \qquad [1.4]$$

If the model is regular, the Cramer–Rao bound can be written:

$$\Im(\theta)^{-1} = -E\left[\frac{\partial^2}{\partial\theta^2}\ln L(X;\theta\right]^{-1} \qquad [1.5]$$

where L(X; θ) is the probability function;

– the maximum likelihood method [MOO 99]:

If a sample has produced the finite sequence of numbers x^*_1, x^*_2,x^*_n and if we have chosen to model this situation using a n-sample X_1, …, X_n of random independent variables of the law $\Im(\theta)$, and if choosing the value of the parameter q is the problem we are confronted with, we can consider the event:

$$E^* = \{X_1 = x_1^*,...,X_n = x_n^*\}, \qquad [1.6]$$

Generally:

$$E(x_1,..,x_n) = \{X_1 = x_1,..,X_n = x_n\} = \{X_1 = x_1\} \cap ... \cap \{X_n = x_n\} \qquad [1.7]$$

and its probability:

$$L(x_1,...,x_n;\theta) = P_\theta\left(E(x_1,...,x_n)\right) = P_\theta\left(\{X_1 = x_1\} \cap ..| \cap \{X_n = x_n\}\right) \qquad [1.8]$$

$$P_\theta\left(\{X_1 = x_1\}\right)..P_\theta\left(\{X_n = x_n\}\right) \qquad [1.9]$$

where this last inequality results from the independence hypothesis of the random variables X_i. The idea is that the choice θ* which it is sensible to make for θ, is that for which this probability is maximal for the values x^*_1,…x^*_n obtained and to ask:

$$\theta^* = Arg\max_\theta\left\{L\left(x_1^*,..,x_n^*;\theta\right)\right\} \qquad [1.10]$$

If θ exists, it is unique, and it is the value for which $\theta \mapsto L\left(x_1^*,..,x_n^*;\theta\right)$ is maximal. This comes to resolve the following equation in θ:

$$\frac{\partial L}{\partial \theta}\left(x_1^*,..,x_n^*;\theta\right)=0 \qquad [1.11]$$

By definition, the function L_n (x_1,..,x_n ; θ) is:

for $X_i \mapsto L(\theta)$ is called the probability of law L:

$$L_n\left(x_1,...,x_n;\theta\right)=\prod_{i=1}^{n} P_\theta\left(\left\{X_i=x_i\right\}\right) \qquad [1.12]$$

1.6. Tensors in signal processing

Tensor formulations are not usually used in signal processing. This is due to the fact that signal processing practitioners have found solutions on second order tensors using a symbolic matrix notation as an invariant representation. Examples are provided by linear operators, dyadic products (correlation matrixes) and vector functions derived by Jacobian matrixes [RUI 07]. The operation shown most often is Kronecker's tensor product; a high-dimension signal is obtained using two low dimensions. This is commonly used to obtain a separable multidimensional base.

The systematic use of tensorial concepts in signal processing is motivated by the field of higher order statistics. Thus, entities of higher order statistics such as higher order moments and cumulants [GIA 87, CAR 90] and [CAR 91] are higher order tensors. The mathematical framework is based on multilinear algebra, that is to say generalizing matrix algebra to high-order tensors.

Interesting advances have been reported in this context in blind identification and blind source separation. These advances are based on recent works on value decomposition [BAS 07] and linked algebraic approaches, such as principal component analysis and independent component analysis.

In recent years, interest in developing tensorial methods has grown. This interest is boosted by new medical imaging methods, such as the diffusion tensor in magnetic resonance imaging and the need to obtain these tensors and visualize them as images. Currently, detection methods (sensing) provide tensorial data that are usually arranged as sampled, multidimensional signals. However, none of these measures is entirely reliable, since every tensor gives degraded and noisy data.

Processing for multidimensional data generally begins with cutting the tensor into vectors or observation matrixes, so that second order methods are applicable. These methods are essentially based on the covariance matrix, and more recently on higher order statistics. The processed data are then fused to find the initial tensor's dimension [MUT 07].

The tensors produced by diffusion tensor magnetic resonance imaging represent the covariance in a Brownian model. In this physical interpretation, the diffusion tensors should be defined as symmetric and positive. However, this current approach to the statistical analysis of tensor diffusion, which handles these linear items do not take account of the positive symmetry constraint. The difficulty results from the fact that the diffusion tensor space is not the same shape as a vectorial space. It has been demonstrated that diffusion tensor space is a type of curve known as a symmetric Riemann space. Methods have been developed to produce statistics called means and modes variety in this space. It is also shown that these statistics conserve these tensors' geometric properties by including the constraints that are true positive values. Formulating symmetrical space also leads us to a natural definition of diffusion tensors and to new anisotropic measurement methods [FLE 07].

The benefit of measuring anisotropy is also stated in Castano's article [CAS 07]. A useful generalization is also suggested. In fact, his approach is articulated theoretically properties of the space of multivariate normal distributions where it is possible to define a invariant Riemanian and affine metric and to express statistics on the varieties of defined positive and symmetric matrixes. The contribution gives tools for anisotropic filtering and regularizing tensor fields. Real and synthetic diffusion tensor data are validated.

Other very important works on theory have been produced by the following authors [WES 94, KOL 08, LAT 97, LAT 00b, CIC 08, CIC 14a, CIC 14b]. These works define the framework of a new era for tensorial signal processing on the basis of what are called tensor networks. The applications obtained by tensorial signal processing also cover CDMA, radar and wireless communications [NIO 10, LAT 07, ALM 07].

1.7. Processing the quantum signal

Quantum signal processing comprises two parts, classic processing and tensorial processing. For classic processing, we will refer to Eldar's work [ELD 02], in this article, on the basis of existing or new signal processing algorithms.

For tensorial processing, Zanardi *et al.* show that the division of quantum systems into subsystems is dictated by the measures and interactions accessible. The emergence of a multi-part tensorial structure of the state-space and the notion of quantum entanglement are therefore the observables induced. A general algebraic framework is developed, it formalizes the concept of multi-part tensors. Two essential aspects in the quantum domain are analyzed, the quantum information processing and decoherence control [ZAN 04]. Quantum tensor formulation was established by Hardy [HAR 12].

Another aspect that can ensure the transition from the classic to the quantum aspect is tackled by Le Bihan in his works on the contribution to the processing of valued-symbols on non-communicative algebraic structures. These works are articulated on the signals' quaternion character, which ensures the transition from classic to Clifford algebra, which is a fundamental element for developing calculations in quantum mechanics [BIH 11a, BIH 11b, BIH 11c].

The targeted applications are the quantum information processing (quantum memories and calculators).

Processing quantum signals promises to contribute to designing the first generation of quantum information processors, noting what represents the biggest challenge at the moment for pioneers in quantum information: understanding electronic noise [KNI 02, KVE 03].

We are interested in the way in which the signals' quantum properties, arising from such devices, can be interpreted by classic instruments. Experiments will enable us the question of compatibility between quantum processors and the computers of today to be examined more deeply [BEH 94, KEI 06].