Geometric Representation of Signals

Arnaldo Spalvieri

Dipartimento di Elettronica, Informazione e Bioingegneria

Politecnico di Milano, ITALY

Bit Rate and Symbol Rate

In digital communications, information is transmitted by randomly choosing a waveform in a set of waveforms, and by transmitting it through the channel. Consider a set of M waveforms, and suppose that the waveforms are chosen with uniform probability. With these assumptions, $\log_2 M$ bits are associated to the transmission of one waveform (one symbol). Transmission is repeated in time, sending through the channel a waveform every T seconds. The *bit rate* and the *symbol rate* are

$$R_b = \frac{\log_2 M}{T} \quad bit/second, \quad R_s = \frac{1}{T} \quad symbol/second.$$

Often symbols are complex. In this case, we call η the number of bits per symbol:

$$\eta = \frac{R_b}{R_s} = \log_2 M.$$

Signal Space

Let $s_i(t)$ denote the i-th complex waveform, and let

$$\mathcal{S} = \{s_1(t), s_2(t), \cdots, s_M(t)\}$$

denote the set of waveforms, which is often called *signal set*.

The *signal space* $\mathcal{X}$ is the set of all signals that can be expressed as a linear combination of the elements of the signal set:

$$\mathcal{X} = \{x(t)\}, \;\; x(t) = \sum_{i=1}^{M} c_i s_i(t),$$

where $\{c_i\}$ is any set of M complex scalars.

Geometric Representation of Signals

Consider an orthogonal basis of the signal space

$$\{\phi_1(t), \phi_2(t), \cdots, \phi_N(t)\},$$

$$\int_{-\infty}^{\infty} \phi_k^*(t)\phi_i(t)dt = \begin{cases} A, & k = i, \quad (normal), \\ 0, & k \neq i, \quad (orthogonal), \end{cases}$$

where A is a possibly complex scalar, the asterisk indicates the complex conjugate, and $N \leq M$ is the dimension of the signal space. When $A = 1$ the basis is said to be orthonormal. The above integral is called *correlation* between $\phi_k(t)$ and $\phi_i(t)$. Note that correlation between complex signals is not commutative.

Geometric Representation of Signals

Any signal $x(t)$ belonging to the signal space, including the signals that form the signal set, can be expressed as

$$x(t) = \sum_{k=1}^{N} x_k \phi_k(t).$$

This equation is the *geometric representation* of $x(t)$. It is easy to show that the complex scalars x_k are calculated as

$$x_k = \frac{1}{A} \int_{-\infty}^{\infty} \phi_k^*(t) x(t) dt.$$

Geometric Representation of Signals

The geometric representation is closely related to filtering. To see how, write

$$x_k = \frac{1}{A}\int_{-\infty}^{\infty} \phi_k^*(\tau)x(\tau)d\tau = \frac{1}{A}\int_{-\infty}^{\infty} \phi_k^*(t_0-\tau)x(t_0-\tau)d\tau.$$

Let $h_k(t) = A^{-1}\phi_k^*(t_0 - t)$ be the impulse responses of a bank of N filters. The correlation becomes

$$x_k = \int_{-\infty}^{\infty} h_k(\tau)x(t_0-\tau)d\tau,$$

that is the convolution between $x(t)$ and the k-th filter at time $t = t_0$.

Geometric Representation of Signals by Sampling

Consider the space of signals $\{x(t)\}$ bandlimited to B Hz. The sampling theorem says that one can recover the continuous-time signal from its samples $x_k = x(kT)$ as

$$x(t) = \sum_{k=-\infty}^{\infty} x_k sinc\left(\frac{t-kT}{T}\right), \quad sinc(u) = \frac{sin(\pi u)}{\pi u},$$

provided that $T^{-1} > 2B$. The sampling theorem is the geometrical representation of $x(t)$ by the (non-normal, exercise: find A.) orthogonal basis

$$\phi_k(t) = \frac{1}{T} sinc\left(\frac{t-kT}{T}\right), \quad k = \cdots, -1, 0, 1, \cdots,$$

$$x_k = \frac{1}{T}\int_{-\infty}^{\infty} sinc\left(\frac{\tau-kT}{T}\right) x(\tau)d\tau = \frac{1}{T}\int_{-\infty}^{\infty} sinc\left(\frac{kT-\tau}{T}\right) x(\tau)d\tau.$$

The last equation, where we use $sinc(u) = sinc(-u)$, shows the analogy with filtering through $T^{-1}sinc(t/T)$ and sampling at $t = kT$, $k = \cdots, -1, 0, 1, \cdots$. (Recall that $\mathcal{F}(T^{-1}sinc(t/T)) = rect(fT)$ to see what is happening in frequency domain.)

Bandwidth and Number of Dimensions

Suppose one wants to transmit information using R_s independent symbols per second, that is using R_s orthogonal dimensions per second. The continuous-time signal can be generated by filtering the symbols sequence through a filter with impulse response $\sqrt{R_s}sinc(R_s t)$. In the practice, one renounces to the impulse response $\sqrt{R_s}sinc(R_s t)$, replacing it by some other impulse response whose Fourier transform has a smooth transition in the frequency domain, thus using some excess bandwidth, that is

$$R_s = \frac{number\ of\ dimensions}{number\ of\ seconds} < 2B,$$

where the signal bandwidth B includes the excess bandwidth.

Example: the Fourier Series

The analogy with the Fourier analysis of a periodic time-domain signal $x(t)$ is apparent. In that case, the geometrical representation is the set of Fourier coefficients

$$X_k = \int_{-\infty}^{\infty} \phi_k^*(t)x(t)dt,$$

and the (non-normal) basis functions are

$$\phi_k(t) = \frac{1}{T} \cdot rect\left(\frac{t}{T}\right) \cdot e^{j\frac{2\pi kt}{T}},$$

where T is the period.

Example: the Dual of the Fourier Series

From the periodic spectrum

$$X(e^{j2\pi fT}) = \frac{1}{T}\sum_k X\left(f - \frac{k}{T}\right),$$

where $X(f)$ is the Fourier transform of $x(t)$, one can get its geometrical representation as the time-domain sequence

$$x_k = \int_{-\infty}^{\infty} \Phi_k^*(f) X(e^{j2\pi fT}) df = x(kT),$$

with

$$\Phi_k(f) = T rect(fT) e^{-j2\pi fkT}, \quad \forall k.$$

Again, the basis functions form an orthogonal basis, not normal.

Example: 8-PSK

Let the signal set be constituted by the eight waveforms

$$\{s_1(t), s_2(t), \cdots, s_8(t)\}$$

with

$$s_k(t) = \sqrt{\frac{2}{T}} \cdot rect\left(\frac{t}{T}\right) \cdot cos\left(\frac{2\pi t}{T} + \frac{k\pi}{4}\right), \quad k = 1, 2, \cdots, 8.$$

Show that a set of orthonormal basis functions is

$$\left\{\phi_1(t) = \sqrt{\frac{2}{T}} \cdot rect\left(\frac{t}{T}\right) \cdot cos\left(\frac{2\pi t}{T}\right), \phi_2(t) = \sqrt{\frac{2}{T}} \cdot rect\left(\frac{t}{T}\right) \cdot sin\left(\frac{2\pi t}{T}\right)\right\},$$

and find out the coefficients s_{ki}, $k = 1, 2, \cdots, 8$; $i = 1, 2$; that express the eight waveforms as linear combination of the basis functions.

Geometric Representation of Noise

Consider a complex and stationary noise process with zero mean and autocorrelation

$$\psi_n(t) = E\{n^*(\tau)n(t+\tau)\}.$$

Projecting $n(t)$ onto the signal space we have a vector $\mathbf{n}$ whose k-th entry is

$$n_k = \int_{-\infty}^{\infty} \phi_k^*(t)n(t)dt.$$

The expected value of n_k is

$$E\{n_k\} = \int_{-\infty}^{\infty} \phi_k^*(t)E\{n(t)\}dt = 0.$$

The covariances are

$$E\{n_k^* n_j\} = E\left\{\left(\int_{-\infty}^{\infty} n^*(t_1)\phi_k(t_1)dt_1\right)\left(\int_{-\infty}^{\infty} n(t_2)\phi_j^*(t_2)dt_2\right)\right\}.$$

Geometric Representation of Noise

Introducing the autocorrelation one writes

$$E\{n_k^* n_j\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi_n(t_2 - t_1)\phi_k(t_1)\phi_j^*(t_2)dt_1 dt_2.$$

The covariance $E\{n_k^* n_j\}$ is the element (k, j) of the so-called *covariance matrix*. A very important case is that where the noise is white with power spectral density N_0, that is

$$\psi_n(t) = N_0 \delta(t),$$

where $\delta(t)$ is the Dirac delta. In this case it is easy to see that the covariance matrix is diagonal, all the elements along the main diagonal being equal to N_0:

$$E\{n_k^* n_j\} = \begin{cases} N_0, & k = j, \\ 0, & k \neq j. \end{cases}$$

Geometric Representation of Noise by Sampling

Consider the space of complex signals $\{n(t)\}$ bandlimited to $B/2$ Hz, and sample $n(t)$ with sampling rate $T^{-1} > B$, leading to the samples at $t = kT$

$$n_k = \frac{1}{T}\int_{-\infty}^{\infty} sinc\left(\frac{kT-\tau}{T}\right) n(\tau)d\tau = n(kT).$$

(Again, we are not using a normal basis.) Then one can prove that

$$E\{n_i^* n_{k+i}\} = \psi_n(kT).$$

For $i = 1, 2, \cdots, N$, the i-the row of covariance matrix is built from the samples of the autocorrelation function starting from $k = i - 1$ to $k = N - i - 1$. The main diagonal of the covariance matrix is $\psi_n(0)$, that is the power $P_{n_i} = E\{n_i^* n_i\}$ of the discrete-time noise sequence. In the common case of continuous-time white noise with p.s.d. N_0 in the range $(-B/2, B/2]$, one has $\psi_n(t) = N_0 B sinc(Bt)$, $P_{n(t)} = E\{n_i^* n_i\} = N_0 B$.

Gaussian Noise

When the noise $n(t)$ is Gaussian, the entries of $\mathbf{n}$ are Gaussian random variables, whose properties are completely described by the vector of the mean values and by the covariance matrix. Assuming zero-mean noise, the joint probability density function of $\mathbf{n}$ is

$$f(\mathbf{n}) = \frac{1}{(2\pi)^{N/2} \cdot \sqrt{det(\Sigma_{\mathbf{n}})}} \cdot \exp\left(-\frac{1}{2}\mathbf{n}^H \Sigma_{\mathbf{n}}^{-1} \mathbf{n}\right),$$

where $\Sigma_{\mathbf{n}}$ is the covariance matrix of $\mathbf{n}$. Note that, whenever possible, we use the shorthand notation for probabilities, skipping the subscripts.

Inner Product

Let the boldface character represent column vectors, that is

$$\mathbf{x} = (x_1, x_2, \cdots, x_N)^T,$$

where the superscript T denotes the transposed vector. The *inner product* between $\mathbf{x}$ and $\mathbf{y}$ is the complex scalar

$$\mathbf{x}^H \mathbf{y} = \sum_{k=1}^{N} x_k^* y_k,$$

where the superscript H denotes the transposed and conjugate vector. Note that the inner product between vectors with complex entries is not commutative.

Inner Product

A consequence of geometric representation of signals is that, given two waveforms $x(t)$ and $y(t)$ belonging to the signal space, their correlation is

$$\int_{-\infty}^{\infty} x^*(t)y(t)dt = \int_{-\infty}^{\infty} \sum_{k=1}^{N} x_k^* \phi_k^*(t) \sum_{i=1}^{N} y_i \phi_i(t)dt = \sum_{k=1}^{N} x_k^* y_k.$$

The sum in the right side of the above equation is the inner product between the complex vectors that express $x(t)$ and $y(t)$ in the signal space, therefore

$$\mathbf{x}^H \mathbf{y} = \int_{-\infty}^{\infty} x^*(t)y(t)dt.$$

Energy

As a special case of correlation, we mention that the inner product $\mathbf{x}^H\mathbf{x}$ is the energy of $x(t)$, that is hereafter denoted E_x:

$$E_x = \int_{-\infty}^{\infty} |x(t)|^2 dt = \mathbf{x}^H\mathbf{x}.$$

The Schwarz Inequality

It is a basic result of vector spaces that

$$|\mathbf{x}^H\mathbf{y}| \leq ||\mathbf{x}|| \cdot ||\mathbf{y}||,$$

where

$$||\mathbf{x}|| = \sqrt{\sum_{i=1}^{N} |x_i|^2}$$

is the norm, or the length in the Euclidean space, of vector $\mathbf{s}$. Equality holds only when

$$\mathbf{x} = a \cdot \mathbf{y},$$

where a is a scalar. In time-domain the inequality reads

$$\left|\int_{-\infty}^{\infty} x^*(t)y(t)dt\right| \leq \sqrt{\int_{-\infty}^{\infty} |x(t)|^2 dt} \cdot \sqrt{\int_{-\infty}^{\infty} |y(t)|^2 dt}.$$

Projection onto a Subspace

Let $\mathcal{X}'$ be a subspace of the signal space $\mathcal{X}$ spanned by the basis functions $\{\phi_1(t), \phi_2(t), \cdots, \phi_{N'}(t)\}$, $N' \leq N$. Then the projection of $x(t)$ onto $\mathcal{X}'$ is

$$x'(t) = \sum_{i=1}^{N'} \phi_i(t) \cdot \int_{-\infty}^{\infty} \phi_i^*(t) x(t) dt = \sum_{i=1}^{N'} x_i \phi_i(t).$$

The geometric representation of the projection $x'(t)$ is the vector

$$\mathbf{x}' = (x_1, x_2, \cdots, x_{N'})^T.$$

A basic property of vector spaces is that the geometric representation $\mathbf{x}'$ of the projection $x'(t)$ is the closest point in $\mathcal{X}'$ to $\mathbf{x}$, and that the projection error

$$\mathbf{e} = \mathbf{x} - \mathbf{x}'$$

is orthogonal to $\mathbf{x}'$.

Pythagorean Theorem

Consider a point $\mathbf{x}$ in $\mathcal{X}$ and a point $\mathbf{y}$ in some subspace $\mathcal{X}'$ of $\mathcal{X}$. Let $\mathbf{x}'$ be the projection of $\mathbf{x}$ onto $\mathcal{X}'$. Since

$$\mathbf{x} - \mathbf{y} = (\mathbf{x} - \mathbf{x}') + (\mathbf{x}' - \mathbf{y}),$$

by orthogonality between $(\mathbf{x}' - \mathbf{y}) \in \mathcal{X}'$ and the projection error $(\mathbf{x} - \mathbf{x}')$, we see that the three points

$$\mathbf{x}, \mathbf{y}, \mathbf{x}'$$

form a right-angle triangle where $(\mathbf{x} - \mathbf{y})$ is the hypotenuse and

$$(\mathbf{x} - \mathbf{x}') \perp (\mathbf{x}' - \mathbf{y}).$$

Pythagorean Theorem

Writing

$$||\mathbf{x}-\mathbf{y}||^2 = ||(\mathbf{x}-\mathbf{x}')+(\mathbf{x}'-\mathbf{y})||^2 = ||\mathbf{x}-\mathbf{x}'||^2+||\mathbf{x}'-\mathbf{y}||^2+2\Re\{(\mathbf{x}'-\mathbf{y})^H(\mathbf{x}-\mathbf{x}')\},$$

and observing that the last inner product is zero by orthogonality between the two vectors, we conclude that

$$||\mathbf{x}-\mathbf{y}||^2 = ||\mathbf{x}-\mathbf{x}'||^2+||\mathbf{x}'-\mathbf{y}||^2,$$

which is the Pythagorean theorem.

Gram-Schmidt Orthonormalization

The geometric representation is most useful when $N < M$. Then two questions arise. How small can N be? How can a set of basis functions be constructed? The answer to these questions is the Gram-Schmidt orthonormalization procedure. Take

$$\phi_1(t) = \frac{s_1(t)}{\sqrt{\int_{-\infty}^{\infty} |s_1(t)|^2 dt}}.$$

Project $s_2(t)$ over the first basis function

$$s_2'(t) = \phi_1(t) \cdot \int_{-\infty}^{\infty} \phi_1^*(\tau) s_2(\tau) d\tau.$$

Gram-Schmidt Orthonormalization

By construction, the error between $s_2(t)$ and its projection over $\phi_1(t)$ is orthogonal to $\phi_1(t)$. If the projection error is zero, then $s_2(t)$ can be expressed using only $\phi_1(t)$, and the procedure continues looking for $\phi_2(t)$ by projecting $s_3(t)$ over $\phi_1(t)$. If the projection error is nonzero, then one can take the normalized error as the second basis function:

$$\phi_2(t) = \frac{s_2(t) - s_2'(t)}{\sqrt{\int_{-\infty}^{\infty} |s_2(t) - s_2'(t)|^2 dt}},$$

and the procedure continues by taking as third basis function the normalized error between $s_3(t)$ and the sum of the projections of $s_3(t)$ over $\phi_1(t)$ and $\phi_2(t)$. The procedure stops when all the projection errors are zero. At the end, $N \leq M$ orthonormal basis functions are obtained, where N is the dimension of the signal space.

MAP Detection

Let

$$\mathbf{x} = \mathbf{s} + \mathbf{n}$$

be the geometric representation of a symbol randomly drawn from the signal set plus additive noise. We want to detect the transmitted symbol embedded in noise making errors with probability as low as possible. This is guaranteed by the *maximum a posteriori* (MAP) probability detection rule, which is

$$\hat{\mathbf{s}} = arg \max_{\mathbf{s}_i \in \mathcal{S}} P(\mathbf{s}_i|\mathbf{x}),$$

where now $\mathcal{S}$ is the set of vectors obtained from the geometrical representation of the signal set. By Bayes rule one writes

$$\hat{\mathbf{s}} = arg \max_{\mathbf{s}_i \in \mathcal{S}} \frac{P(\mathbf{s}_i) f(\mathbf{x}|\mathbf{s}_i)}{f(\mathbf{x})} = arg \max_{\mathbf{s}_i \in \mathcal{S}} P(\mathbf{s}_i) f(\mathbf{x}|\mathbf{s}_i).$$

ML Detection

When the a priori probability of symbols $P(\mathbf{s}_i)$ is uniform, the MAP detection rule becomes

$$\hat{\mathbf{s}} = arg \max_{\mathbf{s}_i \in \mathcal{S}} f(\mathbf{x}|\mathbf{s}_i),$$

which takes the name of *maximum likelihood* (ML) detection rule. Often, the conditional probability dominates over the a priori probability, therefore the performance of ML detection is close to the performance of MAP detection. Also, sometimes the a priori probability is unknown, hence one is led to assume that it is uniform. Moreover ML detection is simpler than MAP detection. For these reasons, ML detection finds wide application.

ML Detection in Presence of Additive Gaussian Noise

In presence of additive Gaussian Noise one has

$$f(\mathbf{x}|\mathbf{s}_i) = \frac{1}{(2\pi)^{N/2} \cdot \sqrt{det(\Sigma_{\mathbf{n}})}} \cdot \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{s}_i)^H \Sigma_{\mathbf{n}}^{-1}(\mathbf{x}-\mathbf{s}_i)\right),$$

therefore the ML detection rule is

$$\hat{\mathbf{s}} = arg \min_{\mathbf{s}_i \in \mathcal{S}} (\mathbf{x}-\mathbf{s}_i)^H \Sigma_{\mathbf{n}}^{-1}(\mathbf{x}-\mathbf{s}_i).$$

Detection is greatly simplified when the noise is white. In this case, the covariance matrix is diagonal, leading to

$$\hat{\mathbf{s}} = arg \min_{\mathbf{s}_i \in \mathcal{S}} (\mathbf{x}-\mathbf{s}_i)^H (\mathbf{x}-\mathbf{s}_i) = arg \min_{\mathbf{s}_i \in \mathcal{S}} \sum_{j=1}^{N} |x_j - s_{ij}|^2.$$

The rule says: decide in favor of the vector $\mathbf{s}_i$ at minimum (squared) distance from the observation $\mathbf{x}$.

Projection Detection

Computing the squared distance one finds

$$\hat{\mathbf{s}} = arg \min_{\mathbf{s}_i \in \mathcal{S}} (\mathbf{x}^H \mathbf{x} + \mathbf{s}_i^H \mathbf{s}_i - 2\Re\{\mathbf{s}_i^H \mathbf{x}\}).$$

Disregarding the common term $\mathbf{x}^H \mathbf{x}$ the decision rule becomes

$$\hat{\mathbf{s}} = arg \max_{\mathbf{s}_i \in \mathcal{S}} \left(2 \cdot \Re\{\mathbf{s}_i^H \mathbf{x}\} - \mathbf{s}_i^H \mathbf{s}_i\right).$$

The above detection rule is called *projection detection* because it requires the projection $\mathbf{x}$ over the orthonormal basis.

Correlation Detection

Going back from the signal space to the time-domain waveforms one has

$$\hat{s}(t) = arg \max_{s_i(t)\in\mathcal{S}} \left(2 \cdot \Re\left\{\int_{-\infty}^{\infty} s_i^*(t)x(t)dt\right\} - \int_{-\infty}^{\infty} s_i^*(t)s_i(t)dt\right).$$

This detection rule is known as *correlation detection.* It is implemented by finding the maximum between the biased versions of the sampled outputs of a bank of M filters with impulse responses $\{s_i^*(-t)\}$, the bias being the energy of $s_i(t)$. Note that in this case there is no need of specifying an orthonormal set of basis function. Also note that with the correlation detector M correlations between time-domain waveforms are required, while the projection detector requires only $N \leq M$ correlations between time-domain waveforms.

Detection of an Impulse with Random Amplitude Embedded in White Noise

Let the signal set be $\{s_i(t) = a_i h(t)\}$, where $\mathcal{A} = \{a_i\}$ is a set of M complex scalars that are associated to the information coming from the source. Applying the Gram-Schmidt orthonormalization to the signal set, one finds that the basis is made by only one function:

$$\phi(t) = \frac{h(t)}{\sqrt{E_h}}.$$

The geometric representation of $s(t)$ is the complex scalar

$$s = \int_{-\infty}^{\infty} \phi^*(t) s(t) dt = a\sqrt{E_h}.$$

Detection of an Impulse with Random Amplitude Embedded in White Noise

Consider

$$x(t) = ah(t) + n(t),$$

where $n(t)$ is white complex noise with p.s.d. N_0. Project $x(t)$ over the space spanned by the signal set. The geometric representation of $x(t)$ is the complex scalar

$$x = a\sqrt{E_h} + n,$$

where n, the projection of $n(t)$ over $\phi(t)$, is a random variable with zero mean and variance N_0.

Detection of an Impulse with Random Amplitude Embedded in White Noise

Note that we have projected $x(t)$ over the space of $s(t)$. One could conjecture that, due to the presence of the noise, the signal space of $x(t)$ is not the same as the signal space of $s(t)$, hence detection could be improved by adding more basis functions with the hope of gaining information about $n(t)$ and of using this information in the detection activity. Suppose therefore of adding $N-1$ orthogonal dimensions to the one associated to the noiseless signal. Of course, the projection of the noiseless signal over the $N-1$ new dimensions is zero, therefore the projection of the noisy signal along these new dimensions is the projection of noise. When the noise is white, the entries of the noise vector obtained by projecting the noise along these new dimensions are independent among them, hence they do not bring information about the noise that affects the first dimension. We hasten to point out that the picture is quite different when the noise is non-white, because the noise that is present along the new dimensions can bring information about the noise that affects the signal in the first dimension.

Sampled Matched Filter, Sufficient Statistics, and SNR

When projection is implemented as filtering, the impulse response of the optimal filter is

$$p(t) = \phi^*(t_0 - t) = \frac{h^*(t_0 - t)}{\sqrt{E_h}},$$

and the output of the filter is sampled at $t = t_0$. The optimal filter followed by optimal sampling takes the name of *Sampled Matched Filter* (SMF). The signal-to-noise ratio (SNR) after the sampled matched filter when a_i is extracted is

$$\text{SNR}_i = \frac{|a_i|^2 \cdot E_h}{N_0} = \frac{E_{s_i}}{N_0}.$$

Since a_i is a random variable, it is convenient to define the SNR as

$$\text{SNR} = \frac{E\{|a|^2\} \cdot E_h}{N_0} = \frac{E_s}{N_0}.$$

Sampled Matched Filter, Sufficient Statistics, and SNR

Since the input noise is white, the power of the noise after sampling is depends only on the energy of the impulse response used to implement correlation. Moreover, invoking the Schwarz inequality it is easy to see that any other impulse response with unit energy that is not $h^*(t_0 - t)/\sqrt{E_h}$ leads to a lower signal power at $t = t_0$. Therefore the SNR at the sampling instant $t = t_0$ is maximized by the sampled matched filter.

Moreover, it is worth observing that, since adding orthogonal dimensions to the one spanned by the SMF brings only new samples of independent noise, the output of the SMF contains all what is needed to perform ML detection. Technically speaking, it is said that the output of the SMF is a *sufficient statistics*.