Mathematics of the Optical Fiber Communication:

Nonlinear Fourier Transforms and Information Theory

Mansoor I. Yousefi

Communications and Electronics Department
Télécom ParisTech, LTCI, Université Paris-Saclay, France

Horizon Maths 2017 - Mathématiques et réseaux

Télécom ParisTech, Paris, France
November 30, 2017

université PARIS-SACLAY

TELECOM ParisTech

Thanks to:

- Frank Kschischang (U. of Toronto)
- Gerhard Kramer (TU Munich)
- Hartmut Haffermann & Jan-Willem Goossens (Huawei Paris)
- Xianhe Yangzhang (University College London)

also the **organizers**!

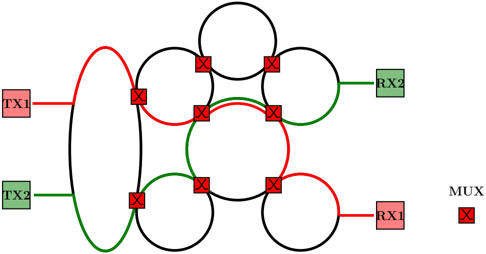
Table of Contents

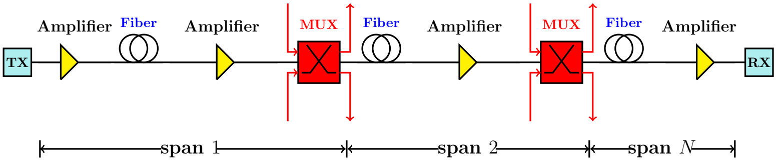
1 Introduction to Optical Networks

2 Nonlinear Fourier Transforms

3 Information Theory of Nonlinear Channels

Wired Communication Networks





(1) Multiple users with add-drop multiplexers (**ADMs**); (2) **interference unknown** to the user-of-interest (UOI); (3) network topology unknown

Optical Fiber

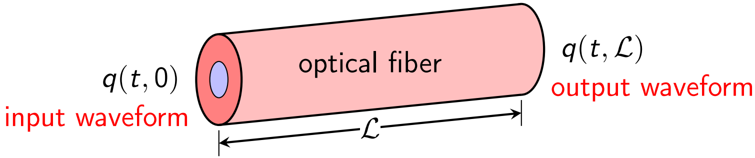
- **Adtantages:** Low-loss ($\sim$ 0.2 dB/km), huge bandwidth ($\sim$5-10 THz bandwidth), all-optical prosseing (laser sources, amplifiers, detectors)
- **Challenges:** Refractive index is a function of **frequency** and **intensity**

$$n(\omega, |q|^2) = \underbrace{n_0 + n_1(\omega - \omega_0) + n_2(\omega - \omega_0)^2 + \cdots}_{\text{dispersion}} + \underbrace{\gamma_0 |\mathbf{q}|^2}_{\text{Kerr nonlinearity}} + \cdots$$

intuition

1. **Dispersion**: n depends on frequency
2. **Kerr nonlinearity**: *the intensity of the signal modifies the refractive index!*
3. High reliability: $P_e = 10^{-15}$
4. High speed: 400 Gb/s

Stochastic Nonlinear Schrödinger Equation



Pulse propagation in optical fibers can be modeled by the *stochastic nonlinear Schrödinger (NLS) equation*:

$$\frac{\partial q(t,z)}{\partial z} = \underbrace{\frac{\partial^2 q(t,z)}{\partial t^2}}_{\text{dispersion}} \pm \underbrace{2j|q(t,z)|^2 q(t,z)}_{\text{nonlinearity}} + \underbrace{n(t,z)}_{\text{noise}}$$

- $q(t,z)$ is the signal, t is time, z is distance
- Distributed white Gaussian noise
- $+$ focusing regime, $-$ defocusing regime
- Vectorial generalizations exit

Fourier Analysis of the NLS Equation

Assume a Fourier series with variable coefficients for $q(t, z)$ at $z > 0$

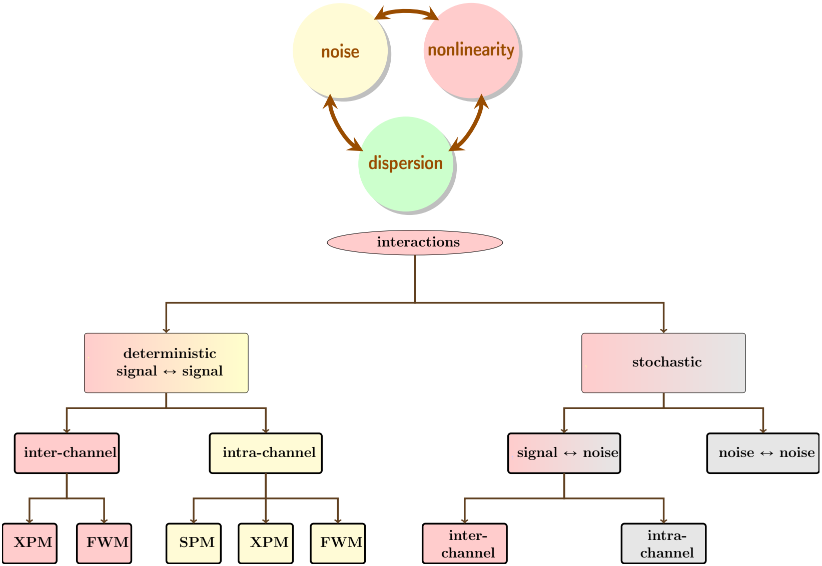
$$q(t, z) = \sum_{k=0}^{N-1} q_k(z) e^{j2\pi kWt}.$$

Substituting (1) into the NLS equation

$$\begin{aligned} j\frac{\partial q_k(z)}{\partial z} = & \underbrace{-4\pi^2 W^2 k^2 q_k(z)}_{\text{dispersion}} + \underbrace{2|q_k(z)|^2 q_k(z)}_{\text{SPM}} \\ & + \underbrace{4q_k(z) \sum_{\ell \neq k} |q_\ell(z)|^2}_{\text{XPM}} \\ & + \underbrace{2 \sum_{\substack{\ell \neq m \\ \ell \neq k}} q_\ell(z) q_m^*(z) q_{k+m-\ell}(z)}_{\text{FWM}} + n_k(z), \end{aligned}$$

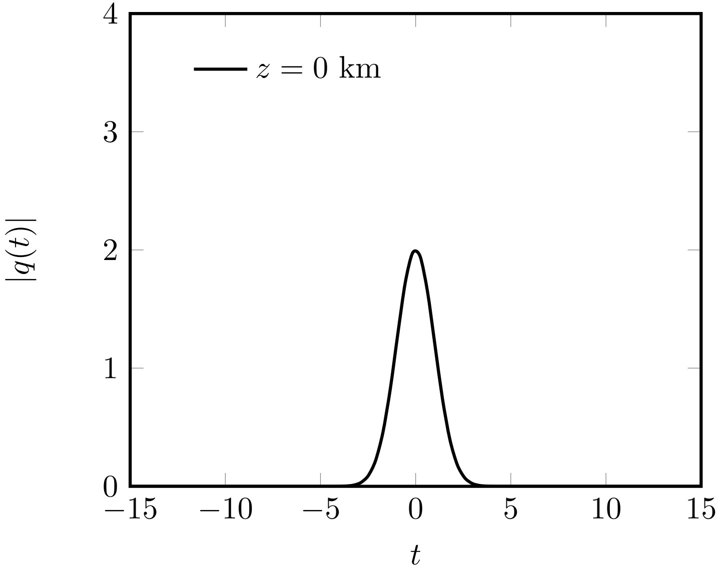
in which n_k are the noise coordinates and where we have identified the dispersion, self-phase modulation (SPM), cross-phase modulation (XPM) and four-wave mixing (FWM) terms.

Nonlinear Effects in Fibers

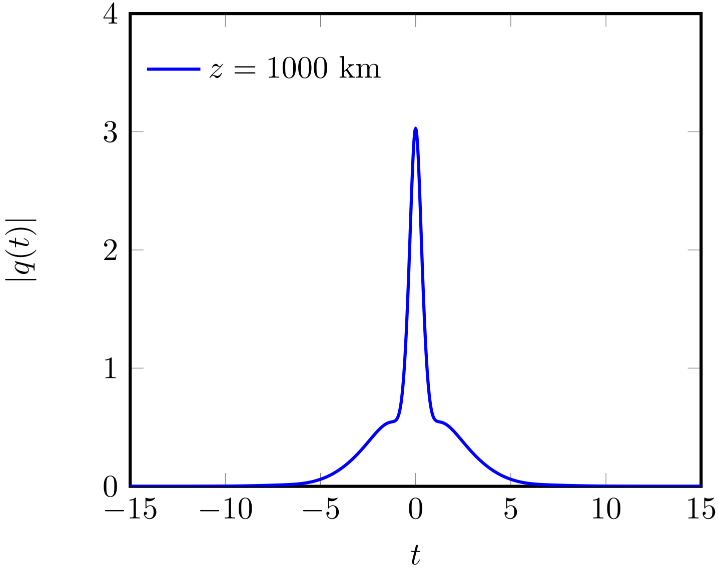


SPM & XPM= self- & cross- phase modulation; FWM = four-wave mixing.

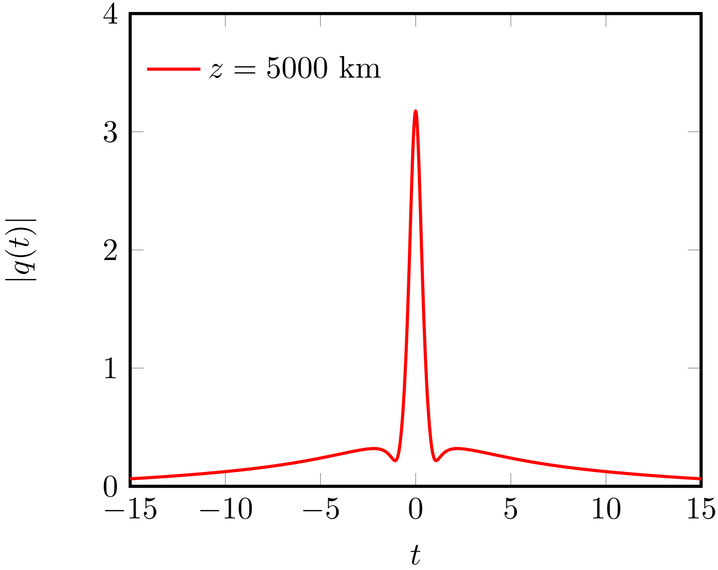
Example of Signal Propagation



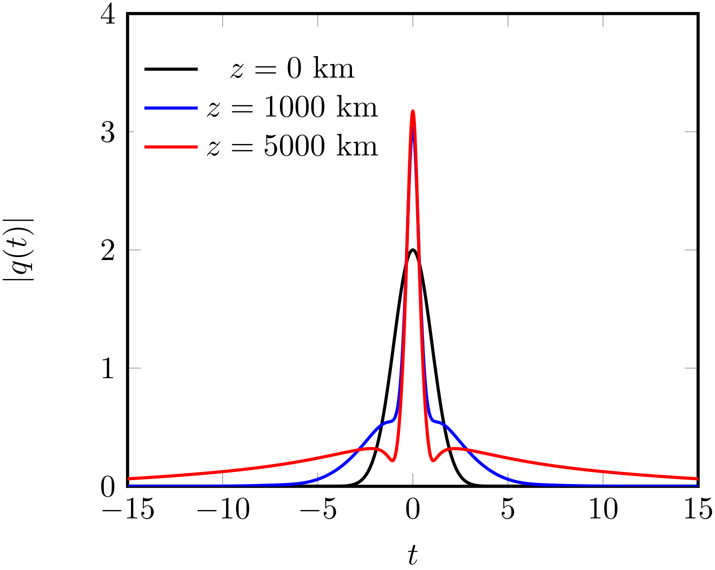
Example of Signal Propagation



Example of Signal Propagation

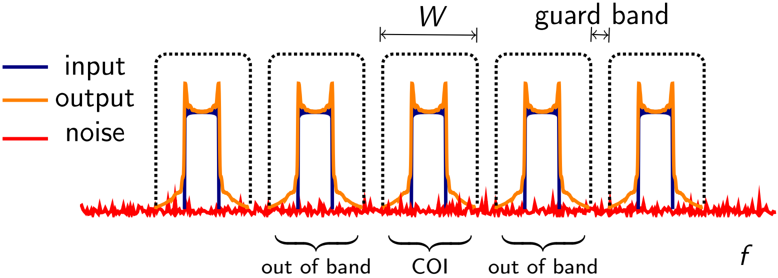


Example of Signal Propagation



The State-of-the-Art Approach

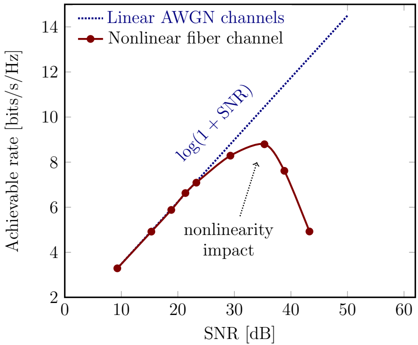
1. **TX**: Wavelength-division multiplexing (WDM)



2. **RX**: Digital back-propagation (BP)

$$q(t,0) \longrightarrow \boxed{q(t,\mathcal{L}) = K_{\mathrm{NLS}}(q(t,0))} \longrightarrow \boxed{q(t,0) = K_{\mathrm{NLS}}^{-1}(q(t,\mathcal{L}))} \longrightarrow q(t,0)$$

Current Achievable Rates



Fiber nonlinearity places an upper limit on capacity

Nonlinear Shannon limit in fiber

Capacity crunch in fiber!

Central Question:

Does fiber nonlinearity really place an upper limit on achievable spectral efficiency?

Origin of the Capacity Limitation – 1

Let $T : \mathcal{H} \mapsto \mathcal{H}$ be a *linear map*:

$$y = T(x) + n,$$

where x and y are input and output signals and n is noise.

Projecting signals onto an orthonormal basis $\{\phi_k\}_{k\in\mathbb{N}}$:

$$\{x, y, n\} = \sum_{k=1}^{\infty} \{x_k, y_k, n_k\}\phi_k, \quad \text{Thus:}$$

$$\boxed{y_k = x_k\langle T\phi_k, \phi_k\rangle + \underbrace{\sum_{i\neq k} x_i\langle T\phi_i, \phi_k\rangle}_{\text{linear interactions}} + n_k}$$

However, if $\{\phi_k(t)\}_k$ is the set of eigenvectors of T, then

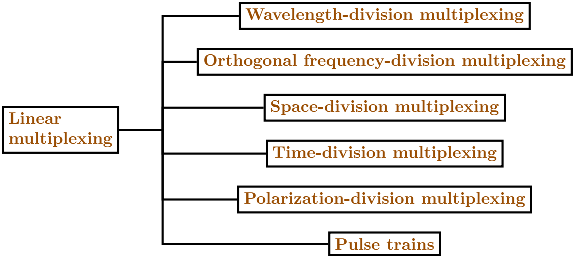
$$y_k = \lambda_k x_k + n_k$$

where λ_k is eigenvalue.

Origin of the Capacity Limitations – 2

Capacity crunch occurs if the **basis** used for communication is **not compatible** with the channel.

- Deterministic nonlinear effects are not a fundamental limitation. It is the **method of the communication** causing the problem
- After abstracting away non-essential aspects, current methods, in essence, modulate **linear-algebraic modes**
- In nonlinear channels, this introduces **interference** and **ISI**
- *BP cannot remove the interference in a network scenario*



Nonlinear frequency-division multiplexing (NFDM)

- It was realized that the NLS equation supports **nonlinear eigenfunctions** which have a crucial **independence property**, the key to build a multiuser system

- The tool necessary to reveal signal degrees of freedom is

  Nonlinear Fourier Transform

- Based on NFT, we constructed an *NFDM*, which can be viewed as a generalization of OFDM to optical fiber

- Exploiting the integrability, NFDM modulates **non-interacting degrees-of-freedom**

- Capacity of the NFDM in the deterministic model is **infinite**

Table of Contents

1 Introduction to Optical Networks

2 Nonlinear Fourier Transforms

3 Information Theory of Nonlinear Channels

Linear Convolutional Channel

Let $T_h(x) = h * x$

$$\boxed{y(t) = h(t) * x(t) + n(t)}, \quad 0 \leqslant t < T.$$

The eigenvectors and eigenvalues of $T_h(x) = h * x$ are

$$\phi_k(t) = \frac{1}{\sqrt{T}} \exp(-jk\omega_0 t), \quad \omega_0 = \frac{2\pi}{T},$$

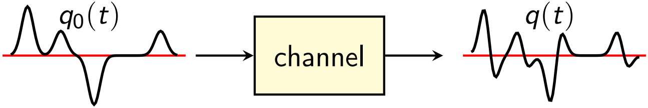
and $\lambda_k = \mathcal{F}(h(t))(k\omega_0)$.
Fourier transform maps convolution into a multiplication operator

$$\boxed{Y(\omega) = H(\omega) \bullet X(\omega) + N(\omega)}, \quad \omega = k\omega_0.$$

1. Frequency ω is **conserved** in the channel
2. Channel is decomposed into **parallel independent** channels
3. **OFDM**: information is encoded in spectral amplitudes $X(\omega)$

General Waveform Channels

- *Instantaneous waveform channel*



$$q(t) = K(q_0(t)) + n(t)$$

- *Evolutionary channel*. Here the signal evolves according to an *evolution equation* in 1+1 dimensions (time t, distance z)

$$\frac{\partial q}{\partial z} = K(q(t,z)) + n(t)$$

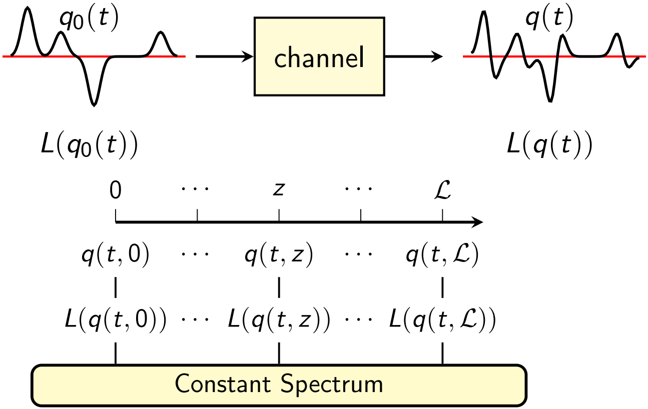
Examples: ($q_t := \partial_t q$)

- $K(q) = j|q|^2$ (memoryless)
- $K(q) = q_{tt}$ (heat eq.)
- $K(q) = -j(q_{tt} + 2|q|^2 q)$(NLS)
- $K(q) = q_{ttt} + 6qq_t$ (KdV)

Isospectral Flow

A Key Idea

We seek an **invariant** under evolution (in the absence of noise). Let L be a linear differential operator (depending on $q(t,z)$). It **may** be possible to find an L whose (eigenvalue) **spectrum** remains constant, even as q evolves (in z).



Isospectral Families of Operators

If the eigenvalues of $L(z)$ do not depend on z, then we refer to $L(z)$ as an **isospectral family** of operators.

Example: The operator L can be a matrix

$$L(z) = \begin{pmatrix} \cos(z) & \sin(z) \\ \sin(z) & -\cos(z) \end{pmatrix}, \quad \lambda = \pm\mathbf{1}, \quad L(z) = G(z)\Lambda G^{-1}(z),$$

where $\Lambda = \mathrm{diag}(1, -1)$.

□

Compact self-adjoint operators can be diagonalized similarly, via *Hilbert-Schmidt Spectral Theorem*. Here, Λ is a *multiplication operator*.

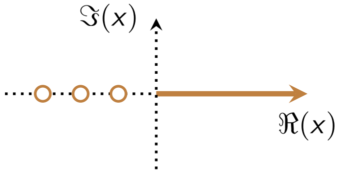
Spectrum of Bounded Linear Operators

- **Spectrum** of an operator is defined as

$$\sigma(L) = \{\lambda \,\big|\, L - \lambda I \text{ is not invertible}\}$$

Example: Linear Schrödinger operator

$$L(q(t,z)) = -\frac{\partial^2}{\partial t^2} + q(t,z).$$



- **Classification**: Spectrum can be *discrete* (like matrices), *continuous*, residual, essential, etc.

The Lax Equation

We have $L(z) = G(z)\Lambda G^{-1}(z)$, where Λ does not depend on z. Assuming that $L(z)$ varies smoothly with z, we can form

$$\begin{aligned} \frac{\mathrm{d}L(z)}{\mathrm{d}z} &= G'\Lambda G^{-1} + G\Lambda\left(-G^{-1}G'G^{-1}\right) \\ &= \underbrace{G'G^{-1}}_{M(z)}\underbrace{\left(G\Lambda G^{-1}\right)}_{L(z)} - \underbrace{\left(G\Lambda G^{-1}\right)}_{L(z)}\underbrace{G'G^{-1}}_{M(z)} \\ &= M(z)L(z) - L(z)M(z) = [M, L], \end{aligned} \tag{1}$$

where $[M, L] \triangleq ML - LM$ is the *commutator bracket*. In other words, every diagonalizable isospectral operator $L(z)$ satisfies the differential equation (1).
The converse is also true.

Lax Pairs

Lemma

Let $L(z)$ be a diagonalizable family of operators. Then $L(z)$ is an isospectral family if and only if it satisfies

$$\frac{dL}{dz} = [M, L], \tag{2}$$

for some operator M, where $[M, L] = ML - LM$.

Definition

The operators L and M satisfying (2) are called a *Lax Pair* (after Peter D. Lax, who introduced the concept [1968]).

Integrable System

Let L and M be operators (depending on $q(t,z)$).

$$\frac{\partial L}{\partial z} = [M, L] \quad \Longleftrightarrow \quad \frac{\partial q}{\partial z} = K(q)$$

operator form — signal form

Example [KdV]: Let $q(t,z)$ be a real-valued function and choose

$$L = \partial_t^2 + q/3, \quad M = 4\partial_t^3 + q_t + q\partial_t.$$

Then: $$\frac{\partial L}{\partial z} = [M, L] \quad \Longleftrightarrow \quad \boxed{q_z = q_{ttt} + qq_t}.$$

NLS Equation

For the normalized nonlinear Schödinger equation

$$jq_z = q_{tt} + 2|q|^2 q,$$

Zakharov and Shabat (1972) found a Lax pair:

$$L = j \begin{pmatrix} \frac{\partial}{\partial t} & -q(t,z) \\ -q^*(t,z) & -\frac{\partial}{\partial t} \end{pmatrix},$$

$$M = \begin{pmatrix} 2j\lambda^2 - j|q(t,z)|^2 & -2\lambda q(t,z) - jq_t(t,z) \\ 2\lambda q^*(t,z) - jq_t^*(t,z) & -2j\lambda^2 + j|q(t,z)|^2 \end{pmatrix}.$$

As $q(t,z)$ evolves according to the NLS equation, the spectrum of L is preserved.

Thus the NLS equation is indeed generated by a Lax pair!

Nonlinear Fourier Transform. Summary-1

$$L = j\begin{pmatrix} \frac{\partial}{\partial t} & -\mathbf{q}(\mathbf{t}) \\ -\mathbf{q}^*(\mathbf{t}) & -\frac{\partial}{\partial t} \end{pmatrix}$$

- *Generalized frequencies:* **eigenvalues λ of L**
- *Nonlinear Fourier coefficients:* $\mathbf{a}$, $\mathbf{b}$ where

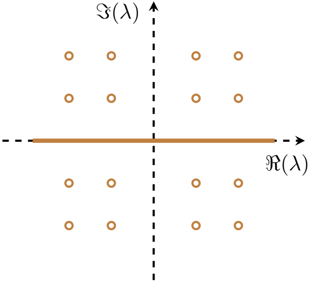
$$V(\lambda) = \begin{pmatrix} a \\ b \end{pmatrix}$$

is a normalized eigenvector of L

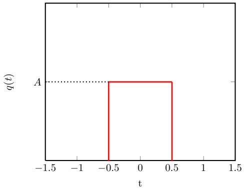
Nonlinear Fourier Transform. Summary-2

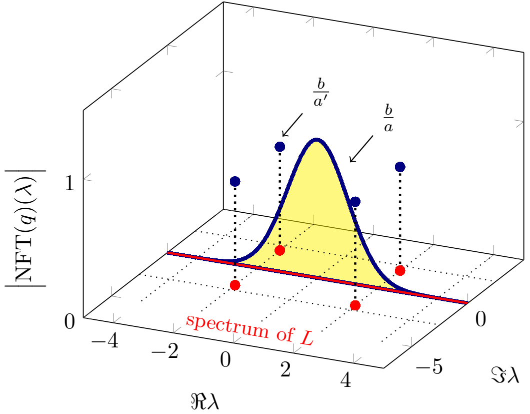
The Zakharov-Shabat operator has two types of spectra:

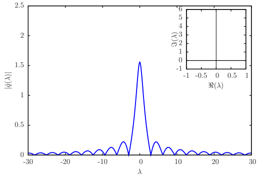
- A *discrete (or point) spectrum* which occurs in $\mathbb{C}^+$ and corresponds to *solitons*
- A *continuous spectrum*, which in general includes the whole real line $\mathbb{R}$

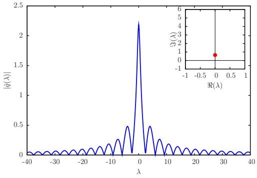


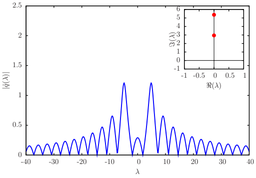
Nonlinear Fourier Transform. Summary-3











Properties of the NFT

1. The NFT shares some of properties of the Fourier transform (FT)
2. FT is a special case of the NFT if $||q||_{L_1} \ll 1$
3.

$$\text{Linear: } y(t) = h(t) * x(t) \longleftrightarrow Y(\omega) = H(\omega)X(\omega)$$

$$\text{Integrable: } y(t) = x(t) * (L, M; \mathcal{L}) \longleftrightarrow \text{NFT}(y)(\lambda) = H(\lambda, \mathcal{L})\,\text{NFT}(x)(\lambda)$$

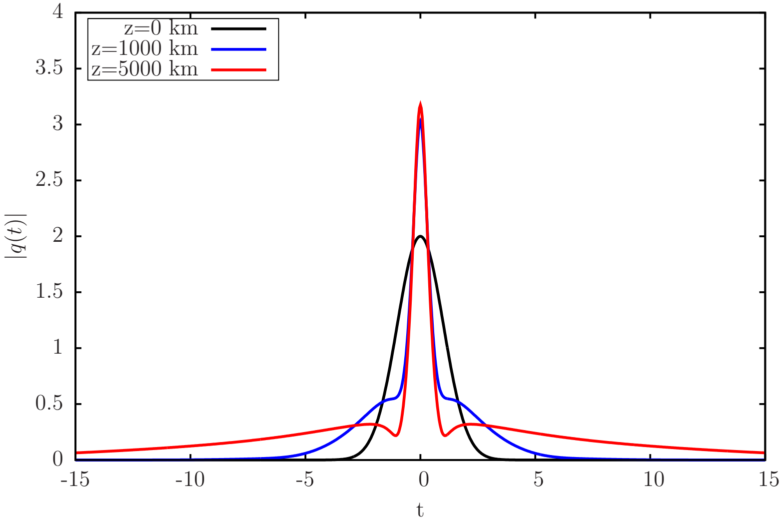
where $H(\lambda, \mathcal{L}) = e^{-4j\lambda^2\mathcal{L}}$ is the channel filter. *The generalized frequencies are invariant in the channel.*

4. When there are a finite number of parameters, the solutions can be expressed via theta functions.

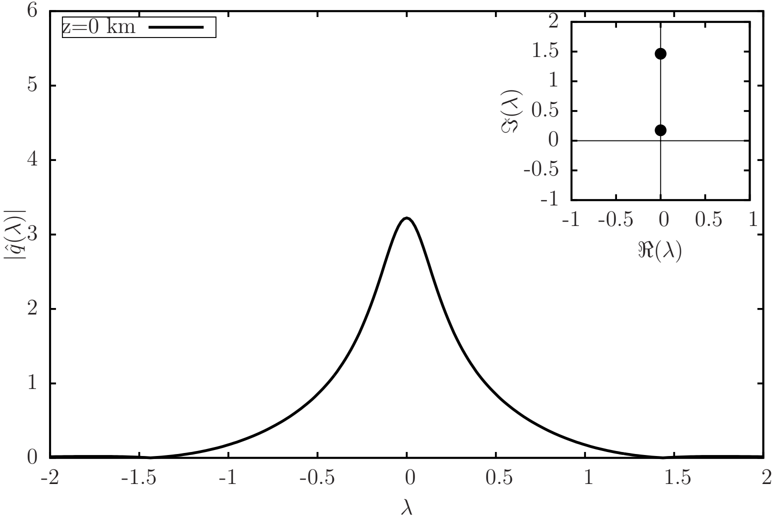
Let $\mathbb{K}$ be an $N \times N$ complex matrix with $\Im(\mathbb{K}) > 0$. The Riemann theta function is defined by

$$\boxed{\theta(\mathbf{t}|\mathbb{K}) = \sum_{\mathbf{m}\in\mathbb{Z}^N} \exp\Big(2\pi j(\mathbf{m}^T\mathbf{t} + \frac{1}{2}\mathbf{m}^T\mathbb{K}\mathbf{m})\Big), \quad \mathbf{t} \in \mathbb{C}^N.}$$

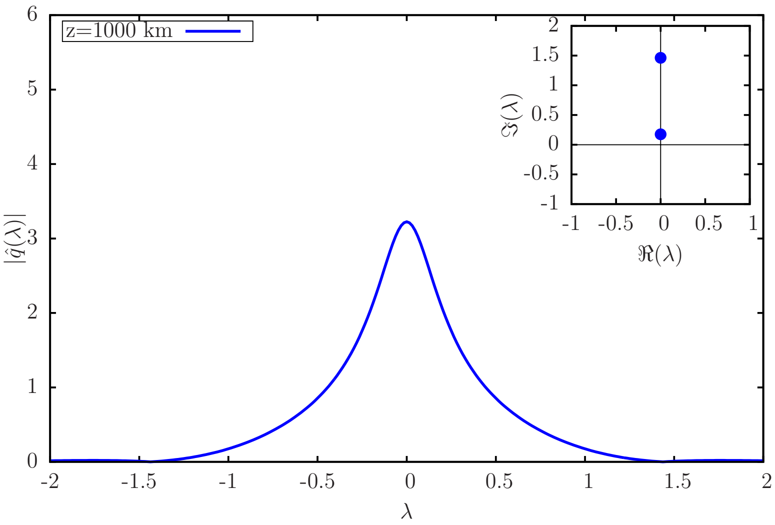
Launching a Pulse (revisited)



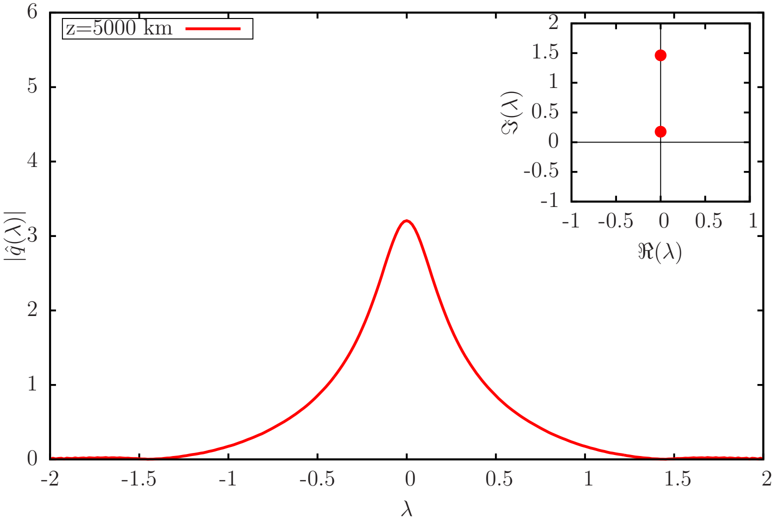
Launching a Pulse (revisited)



Launching a Pulse (revisited)

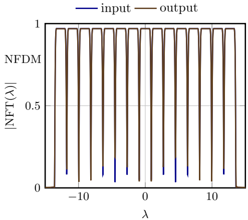


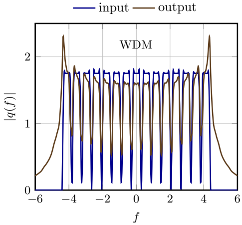
Launching a Pulse (revisited)

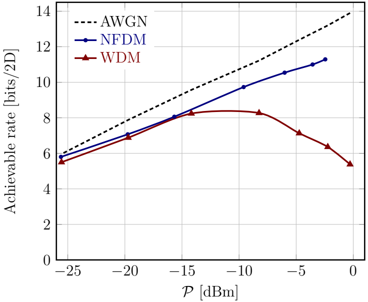


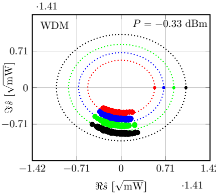
NFDM vs WDM

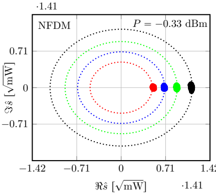
W = 60 GHz, z = 2000 km, 15 users, one symbol per user, defocusing.











Focusing regime, vectorial models, experiments, robustness to perturbations, ...

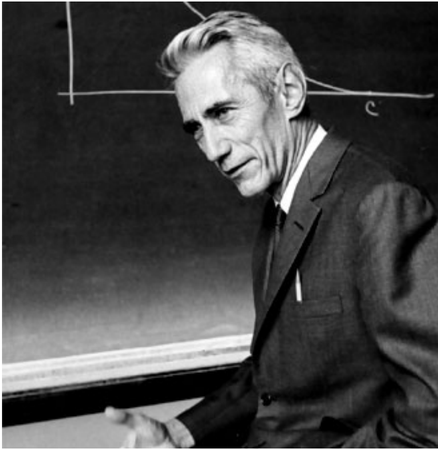
Table of Contents

1 Introduction to Optical Networks

2 Nonlinear Fourier Transforms

3 Information Theory of Nonlinear Channels

Shannon's Formula for Linear Channels



The capacity of a channel $p_{Y|X}(y|x)$

$$C = \sup_{p_X(x)} I(X;Y)$$

The mutual information is defined as

$$I(X;Y) = h(Y) - h(Y|X)$$

where

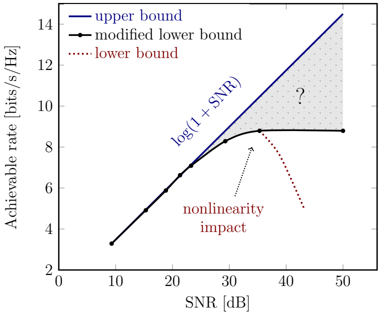
$$h(X) = -\int p_x(X)\log(p_X(x))dx.$$

For a linear channels

$$\boxed{C = \log(1+\text{SNR}), \quad \text{bits/s/Hz}}$$

The capacity of optical fiber is unknown, for about 50 years. Even $p(y|x)$ is unknown!

Upper Bound



Theorem

Consider the discrete-time periodic model $\mathbb{C}^n \mapsto \mathbb{C}^n$. We have

$$\mathcal{C}(\mathcal{P}) \leqslant \log(1 + \mathsf{SNR}).$$

The proof combines:

- Energy and entropy conservation
- Shannon's entropy power inequality

Proof: Invariant Measures for PDEs

Lemma (Volume Preservation in NLS)

Let $\Omega = (\ell^2, \mathcal{E}, \mu)$ be a measure space, where $\ell^2 \triangleq \{\mathbf{q}^n \mid \sum |q_k|^2 < \infty\}$ and

$$\mu(A) = \mathrm{vol}(A) = \int_A \left(\prod_{k=1}^{n} \mathrm{d}q_k \mathrm{d}q_k^* \right), \quad \forall A \in \mathcal{E},$$

is the Lebesgue measure. Transformation T_z underlying the NLS equation, as a dynamical system on Ω, is ***measure-preserving****. That is to say*

$$\mu(T_z(A)) = \mu(A), \quad \forall A \in \mathcal{E}.$$

Application 1: Theorem. The flow of T_z is entropy preserving!
Application 2: There are invariant measures. Gibbs measure:

$$\mathrm{d}\mu_x = \frac{1}{Z} \exp\left\{ -\alpha \left(\sum_{i=1}^{m} |q_i|^4 - |q_i - q_{i-1}|^2 \right) \right\} \prod_{i=1}^{m} \mathrm{d}q_i \; \chi_{||q|| \leqslant 1}$$

where $\alpha > 0$, Z is the partition function, and χ_S is the indicator function.

Asymptotic Capacity

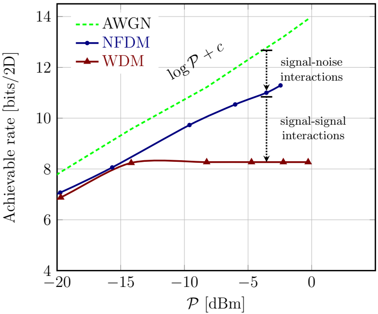
Theorem

Discrete-time periodic model $\mathbb{C}^n \mapsto \mathbb{C}^n$:

$$\mathcal{C}(\mathcal{P}) = \frac{1}{n}\log(\log\mathcal{P}) + c,$$

where $c \triangleq c(n,\mathcal{P}) < \infty$.

With n signal DOFs, $n-1$ DOFs are asymptotically lost to **signal-noise interactions**.



Conclusions

We showed examples where advanced mathematics help make progress in long-standing engineering problems.

- Nonlinear Fourier transforms could be used for data transmission

- The **growth** of the capacity is **too small** compared to the linear channel

$$C = \frac{1}{n}\log(\log \mathcal{P}) + c$$