Rome-Moscow school of Matrix Methods and Applied Linear Algebra
August 20 – September 18, 2016

Linear Algebra Issues in Wireless Communications

Vladimir Lyashev

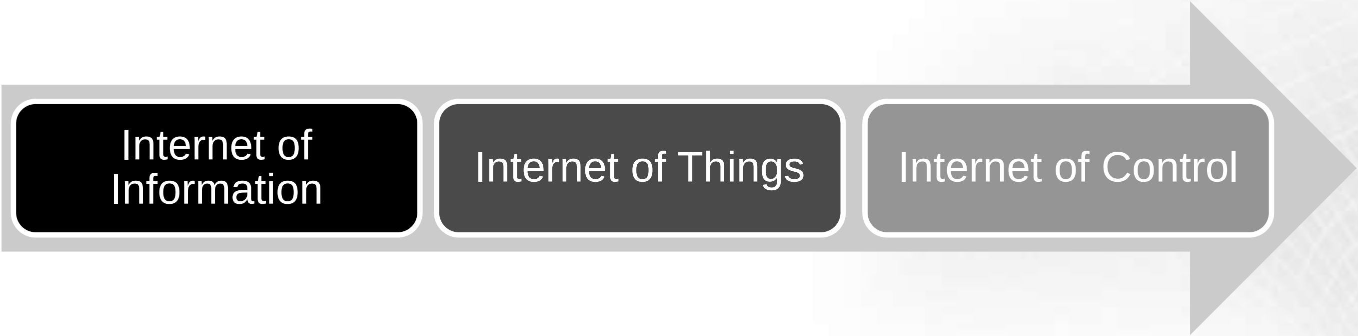
Russian Research Center
[vladimir.lyashev@huawei.com]

HUAWEI

About me

Head of *Interference Cancellation Lab.*
at Huawei Technologies,
Russian Research Center,
Wireless Research Division.

Research Fellow
at Southern Federal University,
the Institute of Signal Processing and Control Systems



Vladimir Lyashev

Aim & Short Outline

Brief introduction in the state-of-the-art problems of matrix calculations in wireless communication should allow participants to understand importance of mathematics and main problems here, which become part of our life.

The introduction is presented by 3 parts:

Introduction in wireless communication and its relation to linear algebra problems

- System model of wireless communication
- Capacity of the system and linear space extension
- Definition of system matrix and its properties discussion
- Outlook of matrix methods that are playing important role in telecom.

Matrix methods and requirements to them from the industry

- Not well-posed problems in the sense of Hadamard (ill-posed problems)
- QR, LU, iterative methods and requirements to them from wireless systems
- Important role of SVD in ultra-high rate communications
- Belief propagation approach in detection problems

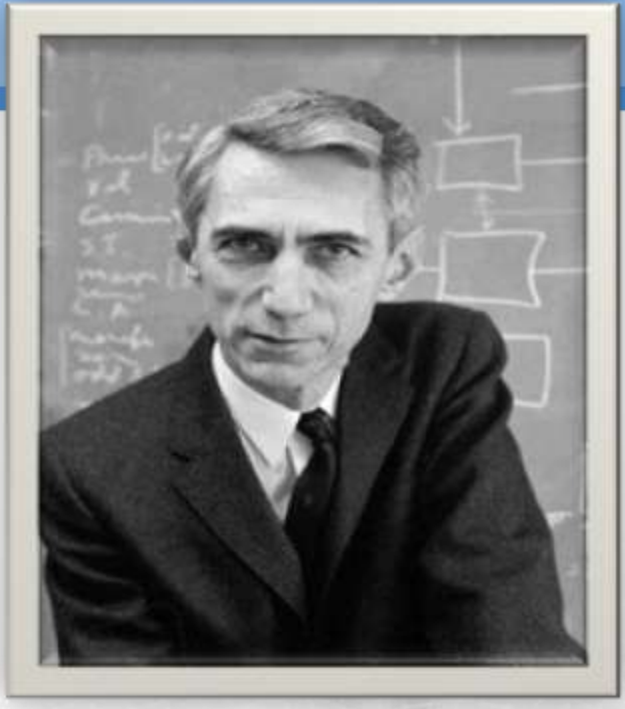
Implementation issues

- Robustness & accuracy
- The simplest one doesn’t mean the fastest one // free discussion

A History

[1948] Groundbreaking article by Claude E. Shannon

"A mathematical theory of communication"



introducing the definition of capacity
in communication systems:

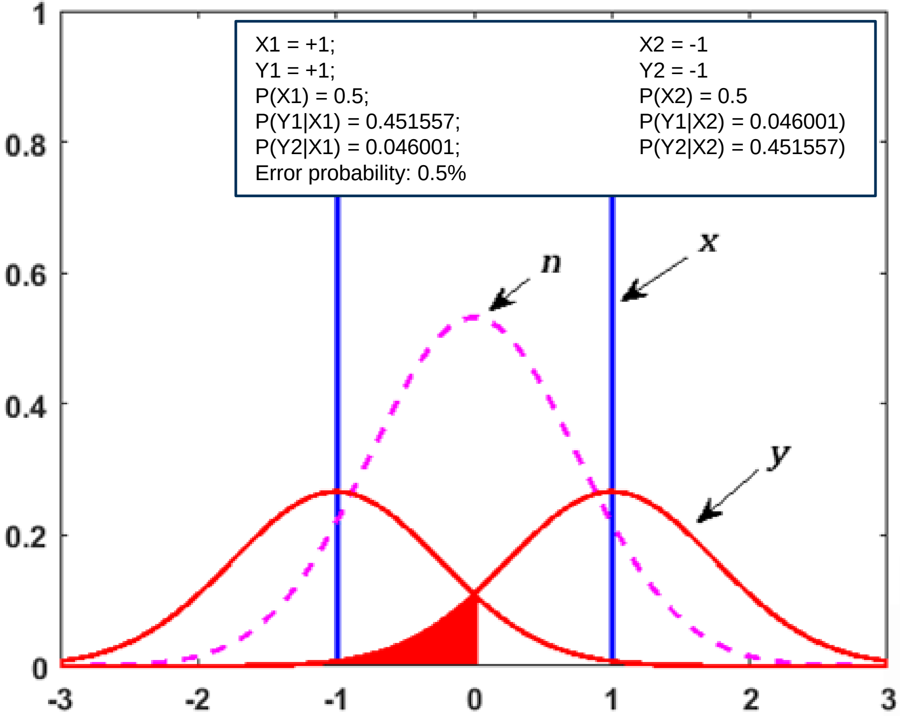
The channel capacity is a measurement of the maximum amount of information that can be transmitted over a channel and received with a negligible probability of error at the receiver.

"Information is the resolution of uncertainty."
- Claude Elwood Shannon
Father of Information Theory

$$C = B_f \log_2(1 + \mathrm{SNR})$$

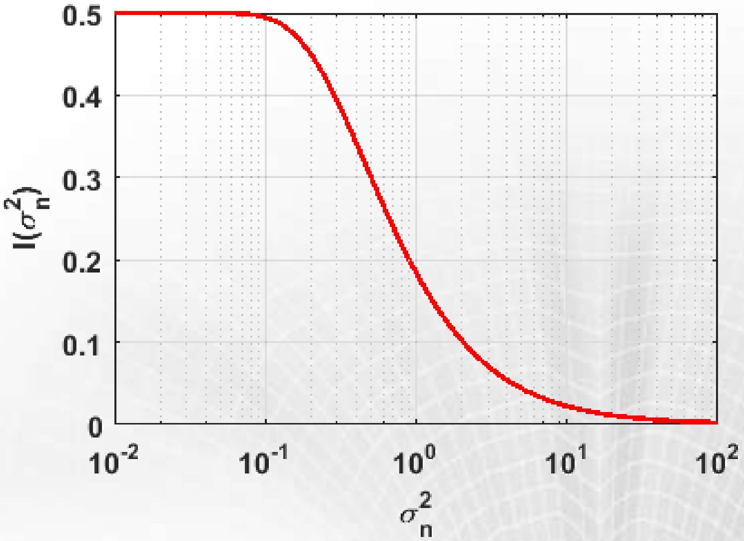
PDF: additive noise

$$y = x + n$$



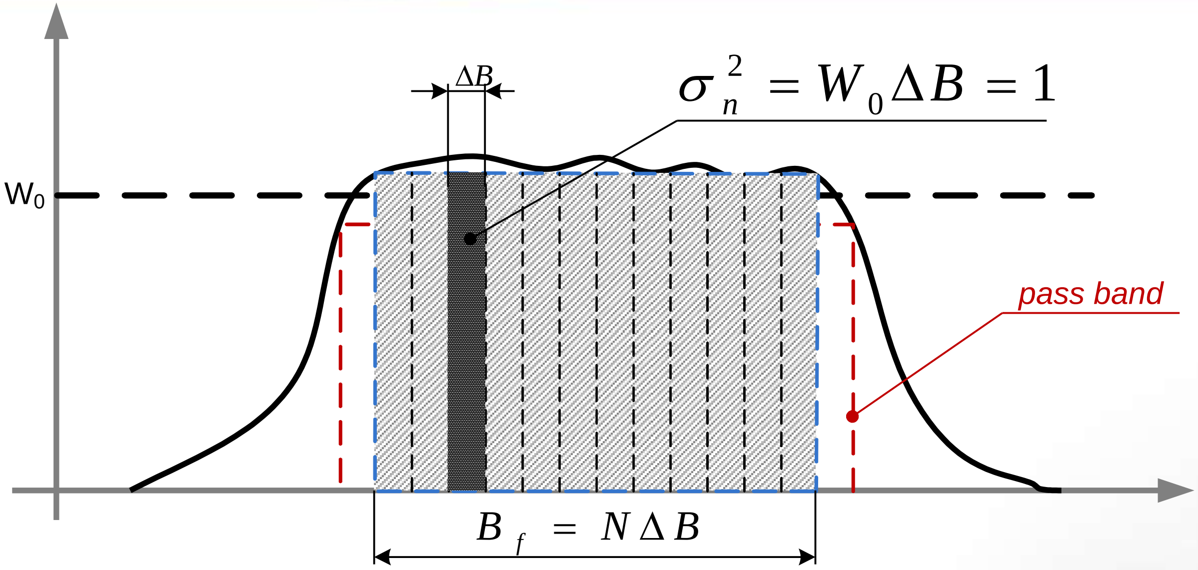
$$H(Y) = -\sum_n P(Y_n)\log_2 P(Y_n)$$

$$H(Y \mid X) = -\sum_k P(X_k)\sum_n P(Y_n \mid X_k)\log_2 P(Y_n \mid X_k)$$



Vladimir Lyashev

Signal-to-Noise Ratio



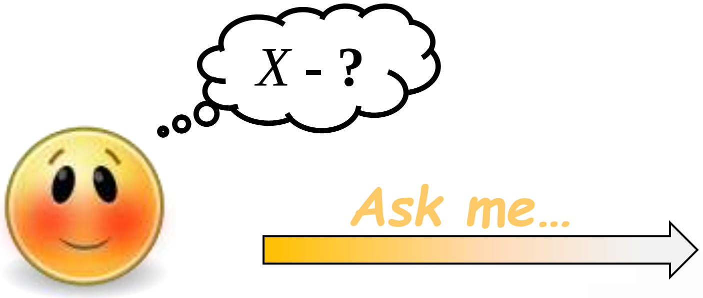
$$y = x + n$$

$$P_y = yy^H = (x+n)(x+n)^H = xx^H + \underbrace{xn^H}_{0} + \underbrace{nx^H}_{0} + nn^H = P_x + \sigma_n^2;$$

$$P_y = \sigma_n^2\left(1 + \frac{P_x}{\sigma_n^2}\right) \Rightarrow N\Delta B \log_2\left(P_y\right) = N\Delta B \log_2\left(\sigma_n^2\left(1 + \frac{P_x}{\sigma_n^2}\right)\right) = N\Delta B \log_2\left(1 + \frac{P_x}{\sigma_n^2}\right)$$

Detection problem

$$X \in [1, 2, \ldots, 32]$$



Search constraint: answer can be YES/NO

Search strategy: bisection method or interval halving or binary search or dichotomy method.

Vladimir Lyashev

Resume

5 answers YES/NO ⟺ 5 bits of information

$$I = \log_2 M = \log_2\left(1 + \frac{P}{\sigma^2}\right)$$

Hartley's law: quantity of information *M*, which is necessary for detection the specific value, is the base-2-logarithm of the number of distinct messages *M* that could be sent.

Open Problems in Wireless Communication

- MULTI-USERS SYSTEMS

  (CDMA, SCMA, MU-MIMO, etc.)

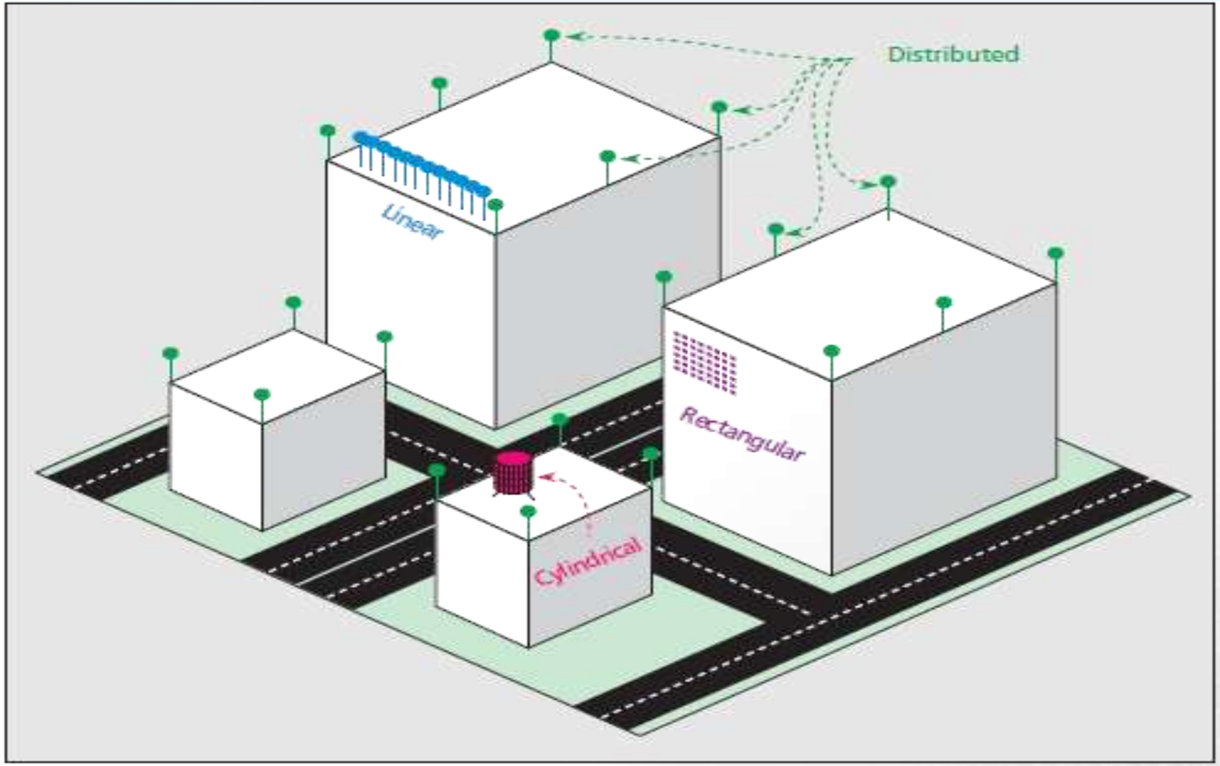
- MULTI-ANTENNA SYSTEMS

  (massive-MIMO,

  cell splitting/antenna selection,

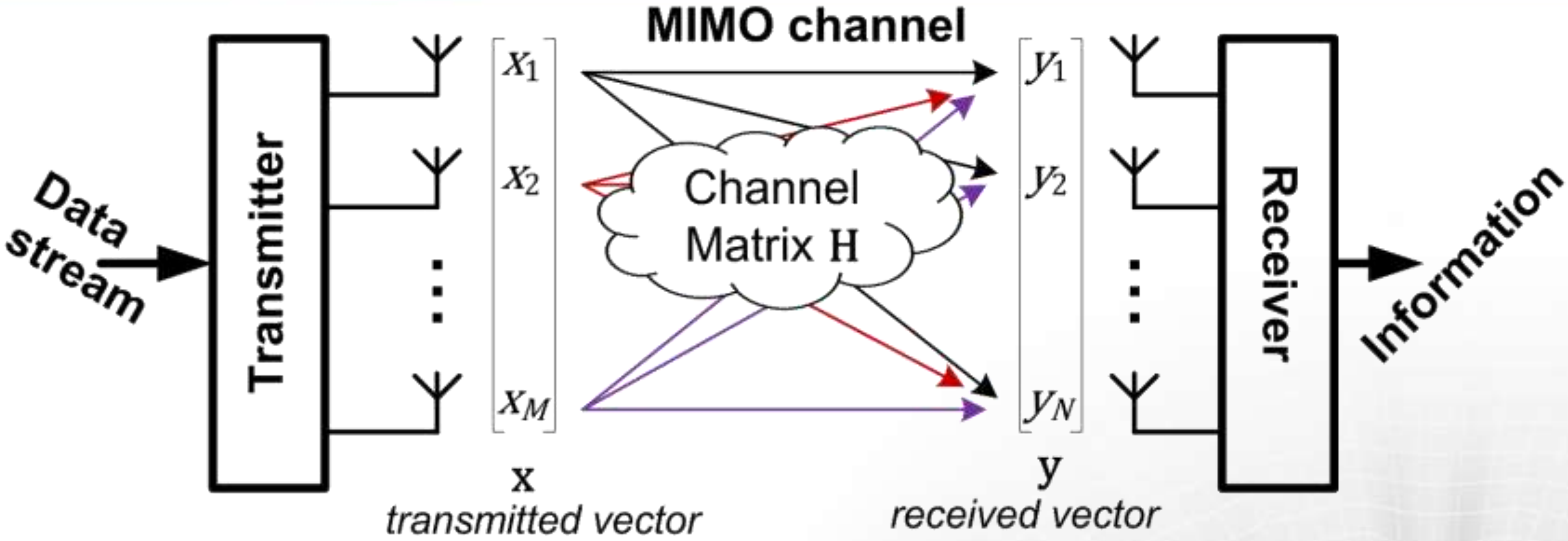
  precoding/beamforming)

MIMO System Evolution



MORE spectrum efficiency → MORE antenna elements

Multiple Input – Multiple Output Communications



$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$$

$$\mathbf{x} \in \mathbb{C}^{M\times 1}, \mathbf{y} \in \mathbb{C}^{N\times 1}, \mathrm{H} = \mathbf{R}_{TX}^{1/2}\mathcal{N}(0, \mathbf{I}) \in \mathbb{C}^{N\times M}$$

Capacity of the system and linear space extension

$$C_{SISO} = B_f \log_2\left(1 + \frac{P_{TX}\|\mathbf{H}\|}{\sigma_n^2}\right) = B_f \log_2\left(1 + \frac{P_{TX}\lambda_1^2}{\sigma_n^2}\right)$$

$$C_{MIMO} = B_f \log_2 \det(\mathbf{I} + \mathbf{HQH}^H)$$

$$\mathbf{Q} \geq 0, \text{trace}\,\mathbf{Q} \leq \frac{P_{TX}}{N_{TX}\sigma_n^2}$$

P_{TX} – power of single transmitter;
N_{TX} – number of transmission antennas;
Q – covariance matrix of transmitted signal.

In current product it is required matrix operations for
N_{RB} (6…110) matrices with size 4x4 … 128x128 per 1ms or less
In future N_{RB} can be extended to 500, and time scale can be reduced 10 times!!!

Vladimir Lyashev

What is channel matrix ?

$$\mathbf{H} = \mathbf{U}\Lambda\mathbf{V}^H \Rightarrow \mathbf{H}\mathbf{V} = \mathbf{U}\Lambda \quad \Rightarrow \quad \mathbf{u}_1 = \frac{1}{\lambda_1}\mathbf{H}\mathbf{v}_1 = \frac{1}{\lambda_1}[\mathbf{h}_1 \ \mathbf{h}_2 \ldots \mathbf{h}_M]\mathbf{v}_1$$

$$\mathbf{u}_i = \tfrac{1}{\lambda_i}\mathbf{R}_{TX}^{1/2}\left[\mathcal{N}^{N\times1}(0,\mathbf{I}) \ \mathcal{N}^{N\times1}(0,\mathbf{I}) \ \ldots \ \mathcal{N}^{N\times1}(0,\mathbf{I})\right]\mathbf{v}_i =$$
$$= \tfrac{1}{\lambda_i}\mathbf{R}_{TX}^{1/2}\left[v_{i1}\mathcal{N}^{N\times1}(0,\mathbf{I}) + v_{i2}\mathcal{N}^{N\times1}(0,\mathbf{I}) + \cdots + v_{iM}\mathcal{N}^{N\times1}(0,\mathbf{I})\right] =$$
$$= \tfrac{1}{\lambda_i}\mathbf{R}_{TX}^{1/2}\mathcal{N}^{N\times1}(0,\mathbf{I}) = \tfrac{1}{\lambda_i}\mathcal{N}^{N\times1}(0,\mathbf{R}_{TX}).$$

Actually, the spatial correlation matrix $\mathbf{R}_{TX}$ is depends of angle of destination (AoD). Thus we can build egenspace matrix $\hat{\mathbf{U}}$ from defined vectors $\mathbf{u}_i$ and correlation matrix $\hat{\mathbf{U}}\hat{\mathbf{U}}^H$, which are similar to Wishart matrices with some constrain on eigenvalues λ_i distribution.

In our assumption, egenvalues can be defined as

$$\frac{\lambda_2}{\lambda_1} \in [0...0.9] \text{ and } \left.\frac{\lambda_i}{\lambda_{i-1}}\right|_{i>2} < \varepsilon.$$

State-of-the-Art Problem

- To solve the matrix equation as detection problem
- Maximize Signal-to-Noise Ratio for better performance
- User grouping for better throughput

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$$

$$\frac{\|\mathbf{w}\mathbf{H}\mathbf{x}\|_2}{\|\mathbf{n}\|_2} \to \max$$

$$\gamma \log_2 \det(\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^H) \to \max$$

Vladimir Lyashev

Matrix applications in MIMO

- Eigenvalue problems for beamforming and interference cancellation
- Reduced-rank signal processing for regularization of inverse problems for antenna combining algorithms and distributed MIMO/scheduling systems
- Low complexity algorithms for adaptive spatial filtering (dirty paper coding, STBC)
- System identification (pilot contamination problem in massive MIMO, type of traffic)
- Tensor Algebra for joint signal processing (channel estimation, multiuser detection)
- Matrix volume maximization for MU-MIMO and non-orthogonal access
- Polynomial series and approximation theory in intermodulation components suppression

Matrix Algebra for MIMO Problems

Matrix Algebra is convenient instrument for problem formulations in information and communications systems and to apply the concepts and algorithms of numerical linear algebra for optimal design and operation of these systems

Important math. areas for MIMO Systems

- QR factorization and least squares problems
- Conditioning and stability issues in numerical linear algebra
- Numerical methods for solving systems of linear equations
- Eigenvalue problems and the computation of the singular value decomposition
  - Arnoldi algorithm
  - Lanczos iterations
- CUR matrix decomposition
- Gradients based approaches for non-linear problems

Vector Precoding

The set of equations that describe the MMSE solution for vector precoding are:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$$

Set $\mathbf{x} = \mathbf{p}s$ that

$$\mathbf{p} = \arg\min_{\hat{p}} \left\| \mathbf{H}^H \left(\mathbf{H}\mathbf{H}^H + \mathbf{R}_{uu} \right)^{-1} \mathbf{y} - \hat{\mathbf{p}}s \right\|_2^2$$

Using SVD, the matrix **H** can be diagonalized by orthogonal matrices **U** and **V**

$$\mathbf{H} = \mathbf{U} \times \mathbf{\Lambda} \times \mathbf{V}^H$$

Thus, $\mathbf{y} = [\mathbf{U} \times \mathbf{\Lambda} \times \mathbf{V}]x + \mathbf{n}$ by transmitting $\dot{\mathbf{x}} = x\mathbf{V}$

Instead of **x** and pre-multiply the receive signal by vector $\mathbf{U}^H$, the transformed received signal vector becomes:

$$\tilde{\mathbf{y}} = \mathbf{U}^H \mathbf{y} = \mathbf{U}^H \left[\mathbf{U} \times \mathbf{\Lambda} \times \mathbf{V}^H \right] \underbrace{\dot{\mathbf{x}}}_{x\mathbf{V}} + \mathbf{U}^H \mathbf{n} = \mathbf{\Lambda} x + \tilde{\mathbf{n}}$$

In TDD systems, the channel is the same on transmitter and receiver, but it is changing in time and robust precoding of the channel is challengeable problem in wireless comm.

SVD

is very important operation !!!

Eigenvector calculation algorithm

We have applied Lanczos tridiagonalization before eigenvalue computing to accelerate convergence, reduce dimension $\mathbb{C} \to \mathbb{R}$, and implimented parallel processing based on *Sturm algorithm*. In figure 1 is shown basic scheme.

$\mathbf{T} = \mathbf{V}_L^H \mathbf{R} \mathbf{V}_L$ is tridiagonal matrix, where $\mathbf{T} \in \mathbb{R}^{n \times n}$, $\mathbf{V}_L \in \mathbb{C}^{N \times n}$, $\mathbf{R} \in \mathbb{C}^{N \times N}$, $n \leq N$.

$$\mathbf{T} = \begin{pmatrix} a_1 & b_1 & & & \\ b_1 & a_2 & b_2 & & \\ & b_2 & a_3 & \ddots & \\ & & \ddots & \ddots & b_{n-1} \\ & & & b_{n-1} & a_n \end{pmatrix}$$

We can compute all of eigenvalues and eigenvectors or only required number. Finally, we have set of eigenvalues and using definition

$$\mathbf{T}\mathbf{v}_i = \lambda_i \mathbf{v}_i \quad \text{and} \quad \lambda_i \mathbf{v}_i = \lambda_i \mathbf{I} \mathbf{v}_i$$

can write equation:

$$(\mathbf{T} - \lambda_i \mathbf{I})\mathbf{v}_i = 0.$$

Solving this matrix equation by LU or QR numeric method we can obtain eigenvectors for matrix $\mathbf{T}$ as $\mathbf{V}_T = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]$, and eigenvectors of initial matrix $\mathbf{R}$ can be found via transformation

$$\mathbf{V}_R = \mathbf{V}_L \mathbf{V}_T,$$

where $\mathbf{V}_R \in \mathbb{C}^N$, $\mathbf{V}_T \in \mathbb{R}^n$, $\mathbf{V}_L \in \mathbb{C}^{N \times n}$.

Tikhonov Regularization in Inverse Problem

Each least squares problem has to be regularized. In the linear case,

$$\mathbf{Ax} = \mathbf{b} + \boldsymbol{\varepsilon}$$

we want to solve minimization problem

$$\|\mathbf{Ax} - \mathbf{b}\|_2 \le \|\boldsymbol{\varepsilon}\|_2$$

after regularization

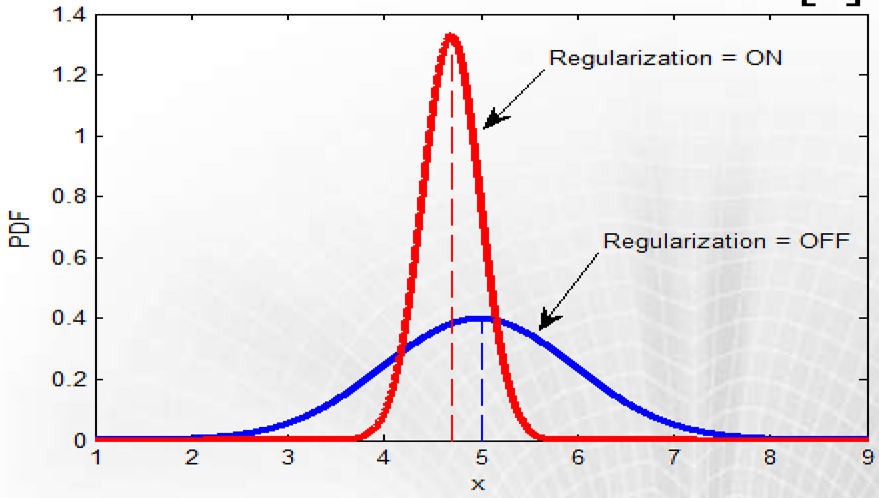
$$\|\mathbf{Ax} - \mathbf{b}\|_2 + \|\boldsymbol{\Gamma}\mathbf{x}\|_2 \to \min$$

the solution is

$$\mathbf{x} = \left(\mathbf{A}^H\mathbf{A} + \boldsymbol{\Gamma}^H\boldsymbol{\Gamma}\right)^{-1}\mathbf{A}^H\mathbf{b} =$$

$$= \left(\mathbf{A}^H\mathbf{A} + \gamma\mathbf{I}\right)^{-1}\mathbf{A}^H\mathbf{b}$$

$$\mathbf{x} = [5]$$



How to define Tikhonov matrix Γ ?

Resume

Typical size of the matrix is less 64×64 elements. For such matrix we need fast algorithms for

- eigenvector decomposition;
- matrix inversion (current baseline is classical Cholesky decomposition algorithm)

The questions are:

1. Can we define some less complexity algorithm for matrix inversion and eigenvector calculation than provided baseline algorithms?
2. Do some fast algorithms (approaches) exist in modern linear algebra to compute such specific matrices?

Homework

1. Generate Wishart-matrix $\mathbf{A}_k$ with following parameters of distribution:

Sigma =

1.0000	0.1000	0.1000	0.1000
0.1000	1.0000	0.1000	0.1000
0.1000	0.1000	1.0000	0.1000
0.1000	0.1000	0.1000	1.0000

df = 8

```
df = 8;
Sigma = 0.1*ones(4) + 0.9*eye(4);
A = wishrnd(Sigma,df)/df
```

2. Set up 2000 samples of equation:

$$\mathbf{A}_k \in \mathbb{R}^{4\times 4}$$

$$\mathbf{A}_k \mathbf{x}_k = \mathbf{b}_k + \boldsymbol{\varepsilon}_k, \text{ where } \mathbf{x}_k = \mathbf{p}_k s_k, \; s_k \in [-1,+1];$$

$$\boldsymbol{\varepsilon}_k \in \mathrm{N}(0,\sigma^2), \; \mathbf{p}_k \text{ is subject to } \left\|\mathbf{A}_k \hat{\mathbf{x}}_k - \mathbf{b}_k\right\|_2 < \sigma^2.$$

Homework

3. Solve noisy equation sample by sample and define probability of right solution averaged over all samples.
// default vector $\mathbf{p}_k = \frac{1}{\sqrt{4}}(1,1,1,1)^T$

$\|\mathbf{p}_k\|_2 = 1 \quad (!)$

4. Repeat item #3 for $\mathbf{p}_k = \mathbf{u}_k^{(1)}$- eigenvector of matrix $\mathbf{A}_k$, corresponded to the largest singular value.

5. How stochastic information about additive noise $\boldsymbol{\varepsilon}_k$ can be utilized for minimization of error probability?..

Homework

I. Generate 2000 samples of metrix **A** and keep in memory for all numerical experiments.

II. Set mapping vector **p** (two ways).

III. Map one-bit symbol (-1/+1) from 1x1 to 4x1: **x** = **p***s*.

IV. Compute **b** = **Ax**.

V. Add gaussian noise (sigma is variation parameter for analysis):
b’ = **b** + **n**.

VI. Solve noise equation: **Ax’** = **b**’, that define **x**’ as estimation value.

VII. Find a way to define *s’* by known **p** and estimated **x’**.

VIII. Check: how many *s’* = *s* ?...
P_{error} = 1 - <right **s’**> / <number of samples>

IX. P_{error} can be defined as function of deviation of noise (sigma).

 Vladimir Lyashev